

(72) **Inventors; and**
(75) **Inventors/Applicants** (*for US only*): **KALYANARAMAN, Vaniambadi** [US/US]; 6050 California Circle, Apt. 101, Rockville, MD 20852 (US). **WHITNEY, Stephen** [US/US]; 2910 Lindell Court, Silver Springs, MD 20902 (US). **VANCOTT, Thomas, C.** [US/US]; 21428 Denit Estates Drive, Brookeville, MD 20833 (US). **POLONIS, Victoria** [US/US]; 1340 Locksley Lane, Silver Spring, MD 20904 (US). **ALVING, Carl** [US/US]; 7118 Glenbrook Road, Bethesda, MD 20814 (US). **MATYAS, Gary, R.** [US/US]; 1845 Snowberry Way, Olney, MD 20832 (US). **RAO, Mangala** [US/US]; 607 Hyde Road, Silver Spring, MD 20902 (US). **MAROVICH, Mary** [US/US]; 8713 Ridge Road, Bethesda, MD 20817 (US). **MCCUTCHAN, Francine** [US/US]; 1430 First Avenue North #2, Seattle, WA 98019 (US). **TOVANABUTRA, Sodsai** [US/US]; 8 Silver Kettle Court, Gaithersburg, MD 20878 (US).

(54) Title: TRUNCATED HIV ENVELOPE PROTEINS (ENV), METHODS AND COMPOSITIONS RELATED THERETO

[illegible]

(57) Abstract: The instant application provides methods and related compositions pertaining to novel HIV envelope proteins. In some embodiments, the invention relates to methods and compositions for the preparation, production, and administration of isolated novel HIV envelope nucleic acid and protein sequences suitable, for example, as vaccines against HIV.

WO 2012/149038 A1



SANDERS-BUELL, Eric [US/US]; 1824 Horseback Trail, Vienna, VA 22182 (US).

(74) **Agents:** EHRHARD, Kathleen 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, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, 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, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

TITLE OF THE INVENTION**TRUNCATED HIV ENVELOPE PROTEINS (ENV) , METHODS AND COMPOSITIONS RELATED THERETO****FEDERAL FUNDING**

This invention was made, in part, with government support under Cooperative Agreement Number W81XWH-07-2-0067. The Federal Government has certain rights in the invention.

CROSS-REFERENCE TO RELATED APPLICATION

This application claims the benefit of priority of U.S. Provisional Application Serial No. 61/478,857 filed April 25, 2011, which is incorporated herein by reference in its entirety.

FIELD OF THE INVENTION

The invention relates generally to novel HIV envelope proteins and to methods and compositions related thereto. More particularly, the invention relates to methods and compositions for the preparation, production, and administration of isolated novel HIV envelope nucleic acid and protein sequences suitable, for example, in certain embodiments, as vaccines against HIV.

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 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.

The form of HIV-1 that dominates the global epidemic is called the major group of HIV-1. There are three HIV-1 groups, the major group (M group), the outlier group (O group), and the non-M/non-O group (N group). There is also the P group. The M group is further divided into nine distinct genetic subtypes, which are commonly referred to as clades and circulating recombinant forms (CRFs). HIV-1 M group subtypes/clades are labelled A, B, C, D, F, G, H, J, and K. Clade B is the most prevalent in the United States, while clade C is the most prevalent worldwide. CRF01_AE or former clade E and CRF02_AG are the most prevalent inter-subtype recombinant strains in the HIV-1 epidemic. Geographic distribution of genetic subtypes and inter-subtype recombinant forms is continually changing, and current data offers incomplete estimates.

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 structural proteins of the nucleocapsid core and matrix. The *pol* gene encodes reverse transcriptase (RT), integrase (IN), and viral protease (PR) 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 et al., Cellular and Molecular Immunology, 4th edition, W.B. Saunders Company, 2000, pp. 454-456). gp140 is a modified form of the Env glycoprotein, which contains the external 120-kDa envelope glycoprotein portion and the extracellular 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 downregulation 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 important 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.

Research on the Env glycoprotein has shown that the virus has many effective protective mechanisms with few vulnerabilities (Wyatt & Sodroski, *Science*. 1998 Jun 19;280(5371):1884-8). For fusion with its target cells, HIV-1 uses a trimeric Env complex containing gp120 and gp41 subunits (Burton et al., *Nat Immunol*. 2004 Mar;5(3):233-6). The fusion potential of the Env complex is triggered by engagement of the CD4 receptor and a coreceptor, usually CCR5 or CXCR4. Neutralizing antibodies seem to work either by binding to the mature trimer on the virion surface and preventing initial receptor engagement events, or by binding after virion attachment and inhibiting the fusion process (Parren & Burton, *Adv Immunol*. 2001;77:195-262). In the latter case, neutralizing antibodies may bind to epitopes whose exposure is enhanced or triggered by receptor binding. However, given the potential antiviral effects of neutralizing antibodies, it is not unexpected that HIV-1 has evolved multiple mechanisms to protect it from antibody binding (Johnson & Desrosiers, *Annu Rev Med*. 2002;53:499-518).

Most experimental HIV-1 vaccines tested in human and/or non-human primate suggests that a successful vaccine will incorporate immunogens that elicit broad neutralizing antibodies (bNabs) and robust cell-mediated immunity. HIV-1 envelope glycoprotein (Env) is the main viral protein involved in the entry of the virus and is also the primary target for neutralizing antibodies, but due to immune evasion strategies and extreme sequence variability of Envs, generation of bNabs has been a daunting task (Phogat S, Wyatt R. *Curr Pharm Des*. 2007, 13:213-27; Phogat S, et al. *J Intern Med*. 2007 262:26-43, Karlsson Hedestam GB, et al *Nat Rev Microbiol*. 2008, 6:143-55).

Developing effective vaccines to prevent HIV infection or neutralize HIV infection has been difficult. The ability to elicit broad and potent neutralizing antibodies is a major challenge in the development of an HIV-1 vaccine. It is a primary goal to develop an HIV vaccine that can effectively elicit specific anti-viral neutralizing antibodies as well as cell-mediated immune responses to prevent infection and control the spread of HIV, with a potential for considerable breadth of reactivity across genetic clades. The extraordinary degree of genetic diversity of HIV has been problematic for vaccine development.

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

SUMMARY OF THE INVENTION

In certain embodiments, the instant application provides an isolated peptide comprising a truncated HIV Env protein, wherein the HIV Env protein is mutated in the native gp120/gp41 cleavage site to prevent protease cleavage, comprises the MPER of gp41, and is truncated prior to the transmembrane domain.

In some embodiments, the HIV Env protein comprises about 1-10 hydrophilic amino acids at its C-terminus. In certain embodiments, the about 1-10 hydrophilic amino acids are three lysines.

In some embodiments, the MPER of gp41 comprises the 4E10 epitope. In certain embodiments, the MPER of gp41 comprises the amino acid sequence: LWYIK (SEQ ID NO: 24) at its C-terminus. In further embodiments, the HIV Env protein comprises about 1-10 non-native hydrophilic amino acids C-terminal to and contiguous with the LWYIK (SEQ ID NO: 24) amino acid sequence. In certain embodiments, the HIV Env protein binds integrin $\alpha 4\beta 7$.

In some embodiments, the HIV Env protein is derived from an HIV-1 strain classified in a group selected the group consisting of: M, O, N, and P. In certain embodiments, the HIV-1 strain is isolated from an individual with an acute HIV-1 infection. In other embodiments, the HIV-1 strain is isolated from an individual with a chronic HIV-1 infection. In certain embodiments, the HIV Env protein is derived from an HIV-1 group M strain. In further embodiments, the HIV-1 group M strain is a subtype (clade) selected from the group consisting of: A, B, C, D, F, G, H, J, and K. In a particular embodiment, the subtype (clade) is clade B. In another embodiment, the subtype (clade) is clade D. In yet another embodiment, the subtype (clade) is clade C.

In certain embodiments, the HIV Env protein comprises an amino acid sequence having 85 % or greater identity to the amino acid sequence depicted in SEQ ID NO: 1.

In further embodiments, the HIV Env protein comprises 90 % or greater identity to the amino acid sequence depicted in SEQ ID NO: 1. In yet other embodiments, the peptide comprises an amino acid sequence having 95 % or greater identity to the amino acid sequence depicted in SEQ ID NO: 1. In certain embodiments, the peptide comprises an amino acid sequence having 98 % or greater identity to the amino acid sequence depicted in SEQ ID NO: 1. In other embodiments, the peptide comprises an amino acid sequence having 99 % or greater identity to the amino acid sequence depicted in SEQ ID NO: 1. In a particular embodiment, the peptide comprises the amino acid sequence depicted in SEQ ID NO: 1.

In some embodiments, the application pertains to an isolated nucleic acid comprising a nucleic acid sequence encoding an amino acid sequence having 85 %, 90 %, 95 %, 98 %, 99 %, or greater identity to the amino acid sequence depicted in SEQ ID NO: 1. In certain embodiments, the nucleic acid sequence encodes the amino acid sequence depicted in SEQ ID NO: 1. In a particular embodiment, the isolated nucleic acid comprises the nucleic acid sequence depicted in SEQ ID NO: 20.

In yet other embodiments, the instant application pertains to a vector comprising nucleic acid encoding an amino acid sequence having 85 %, 90 %, 95 %, 98 %, 99 %, or greater identity to the amino acid sequence depicted in SEQ ID NO: 1. In certain embodiments, the application relates to a host cell comprising the vector comprising nucleic acid encoding an amino acid sequence having 85 %, 90 %, 95 %, 98 %, 99 %, or greater identity to the amino acid sequence depicted in SEQ ID NO: 1. In a particular embodiment, the host cell is a CHO cell.

In yet other embodiments, the instant application relates to a method of making a peptide comprising an amino acid sequence having at least 85 % or greater identity to the amino acid sequence depicted in SEQ ID NO: 1, comprising culturing a host cell comprising a vector comprising nucleic acid encoding an amino acid sequence having 85 %, 90 %, 95 %, 98 %, 99 %, or greater identity to the amino acid sequence depicted in SEQ ID NO: 1 under conditions suitable for protein expression and isolating the peptide.

In certain embodiments, the instant application provides a composition comprising an isolated HIV Env protein, such as an isolated HIV Env protein comprising an amino acid sequence having 85 %, 90 %, 95 %, 98 %, 99 %, or greater identity to the amino acid sequence depicted in SEQ ID NO: 1 and pharmaceutically acceptable carrier.

In yet other embodiments, the instant application relates to a method of generating antibodies against HIV in a mammal, comprising administering to the mammal a composition comprising an isolated HIV Env protein, such as an isolated HIV Env protein comprising an amino acid sequence having 85 %, 90 %, 95 %, 98 %, 99 %, or greater identity to the amino acid sequence depicted in SEQ ID NO: 1 and a pharmaceutically acceptable carrier. In certain embodiments, the composition further comprises an adjuvant. In certain embodiments, the adjuvant comprises a liposome formulation. In further embodiments, the liposome formulation comprises one or more of: dimyristoyl phosphatidylcholine, dimyristoyl phosphatidylglycerol, cholesterol, and phospholipid. In a particular embodiment, the liposome formulation comprises phospholipid A. In certain embodiments, the antibodies generated in the mammal are antibodies that compete with the peptide comprising the truncated HIV Env protein for binding integrin $\alpha 4\beta 7$.

In some embodiments, the instant application relates to a method of conferring immunity against HIV in a mammal, comprising administering to the mammal a composition comprising an isolated HIV Env protein, such as an isolated HIV Env protein comprising an amino acid sequence having 85 %, 90 %, 95 %, 98 %, 99 %, or greater identity to the amino acid sequence depicted in SEQ ID NO: 1 and a pharmaceutically acceptable carrier. In certain embodiments, the composition further comprises an adjuvant. In certain embodiments, the adjuvant comprises a liposome formulation. In further embodiments, the liposome formulation comprises one or more of: dimyristoyl phosphatidylcholine, dimyristoyl phosphatidylglycerol, cholesterol, and phospholipid. In a particular embodiment, the liposome formulation comprises phospholipid A. In further embodiments, the method comprises administering the composition to the mammal by injection.

Examples of mammals to which the compositions of the invention can be administered include human, non-human primates, dogs, rabbits, guinea pigs, and mice.

In yet other embodiments, the instant application relates to a subunit vaccine comprising an HIV Env protein of the invention, such as an isolated peptide comprising a truncated HIV Env protein, wherein the HIV Env protein is mutated in the native gp120/gp41 cleavage site to prevent protease cleavage, comprises the MPER of gp41, and is truncated prior to the transmembrane domain. In some embodiments, the HIV Env protein comprises about 1-10 hydrophilic amino acids at its C-terminus. In certain embodiments, the about 1-10 hydrophilic amino acids are three lysines. In other embodiments, the subunit vaccine comprises an isolated HIV Env protein, such as an isolated HIV Env protein comprising an amino acid sequence having 85 %, 90 %, 95 %, 98 %, 99 %, or greater identity to the amino acid sequence depicted in SEQ ID NO: 1

In some embodiments, the instant application relates to a nucleic acid vaccine comprising an isolated nucleic acid comprising a nucleic acid sequence encoding an amino acid sequence having 85 %, 90 %, 95 %, 98 %, 99 %, or greater identity to the amino acid sequence depicted in SEQ ID NO: 1. In certain embodiments, the nucleic acid sequence encodes the amino acid sequence depicted in SEQ ID NO: 1..

In yet other embodiments, the instant application pertains to an isolated peptide comprising an amino acid sequence having 90 % or greater identity to the amino acid sequence depicted in SEQ ID NO: 3, SEQ ID NO: 4, or SEQ ID NO: 5. In certain embodiments, the peptide comprises an amino acid sequence having 98 % or greater identity to an amino acid sequence depicted in SEQ ID NO: 3, SEQ ID NO: 4, or SEQ ID NO: 5. In further embodiments, the peptide comprises an amino acid sequence having 99 % or greater identity to an amino acid sequence depicted in SEQ ID NO: 3, SEQ ID NO: 4, or SEQ ID NO: 5. In a particular embodiment, the peptide comprises the amino acid sequence depicted in SEQ ID NO: 3, SEQ ID NO: 4, or SEQ ID NO: 5.

In another embodiment, the application relates to an isolated nucleic acid sequence comprising a nucleic acid sequence encoding an amino acid sequence having 90 %, 95 %, 98 %, 99 %, or greater identity to the amino acid sequence depicted in SEQ ID

NO: 3, SEQ ID NO: 4, or SEQ ID NO: 5. In certain embodiments, the nucleic acid sequence encodes the amino acid sequence depicted in SEQ ID NO: 3, SEQ ID NO: 4, or SEQ ID NO: 5.

In some embodiments, the instant application relates to a kit comprising (a) a composition comprising an isolated peptide comprising a truncated HIV Env protein, wherein the HIV Env protein is mutated in the native gp120/gp41 cleavage site to prevent protease cleavage, comprises the MPER of gp41, and is truncated prior to the transmembrane domain and a pharmaceutically acceptable carrier and (b) instructions for administration of the composition to a mammal. In some embodiments, the HIV Env protein comprises about 1-10 hydrophilic amino acids at its C-terminus. In certain embodiments, the about 1-10 hydrophilic amino acids are three lysines.

In some embodiments, the application relates to a kit comprising (a) a composition comprising an isolated HIV Env protein, such as an isolated HIV Env protein comprising an amino acid sequence having 85 %, 90 %, 95 %, 98 %, 99 %, or greater identity to the amino acid sequence depicted in SEQ ID NO: 1 and (b) instructions for administration of the composition to a mammal.

In other embodiments, the application relates to a kit comprising (a) a composition comprising an isolated nucleic acid comprising a nucleic acid sequence encoding an amino acid sequence having 85 %, 90 %, 95 %, 98 %, 99 %, or greater identity to the amino acid sequence depicted in SEQ ID NO: 1 and a pharmaceutically acceptable carrier and (b) instructions for administration of the composition to a mammal. In certain embodiments, the nucleic acid sequence encodes the amino acid sequence depicted in SEQ ID NO: 1

In certain embodiments, the application relates to an isolated peptide comprising a truncated HIV Env protein, wherein the HIV Env protein is mutated in the native gp120/gp41 cleavage site to prevent protease cleavage, comprises the MPER of gp41, and is truncated prior to the transmembrane domain, wherein the HIV Env protein is mutated in the leader sequence. In some embodiments, the native signal peptide is replaced with a tPA signal peptide. In certain embodiments, the tPA signal peptide

comprises a sequence selected from the group consisting of: SEQ ID NO: 21 and SEQ ID NO: 22.

In yet other embodiments, the instant application provides an isolated peptide comprising an amino acid sequence having 90 % or greater identity to the amino acid sequence depicted in SEQ ID NO: 7 or SEQ ID NO: 9. In further embodiments, the peptide comprises an amino acid sequence having 98 % or greater identity to an amino acid sequence depicted in SEQ ID NO: 7 or SEQ ID NO: 9. In still further embodiments, the peptide comprises an amino acid sequence having 99 % or greater identity to an amino acid sequence depicted in SEQ ID NO: 7 or SEQ ID NO: 9. In a particular embodiment, the peptide comprises the amino acid sequence depicted in SEQ ID NO: 7 or SEQ ID NO: 9.

In some embodiments, the application relates to an isolated nucleic acid sequence comprising a nucleic acid sequence encoding an amino acid sequence having 85 %, 90 %, 95 %, 98 %, 99 %, or greater identity to the amino acid sequence depicted in SEQ ID NO: 7 or SEQ ID NO: 9. In certain embodiments, the isolated nucleic acid sequence comprises a nucleic acid sequence depicted in SEQ ID NO: 3 or SEQ ID NO: 5.

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

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic of the vector, pJWIRESpuro.

Figure 2: Immunoprecipitation Western Blot of 293/pJW Ba-L gp140 DC 4E10 Puro transfection. Conditioned media of 293 cells transfected with pJW Ba-L gp140 DC 4E10 Puro and naïve cells are immunoprecipitated with Human monoclonal antibodies to the MPER region (2F5 and 4E10), HIV-1 (+) human serum and normal human serum. Precipitated proteins are resolved on 12.5% SDS-PAGE and transferred to PVDF. Ba-L gp140 DC 4E10 proteins are detected with MoAb to gp41 (M25) followed by Goat anti-mouse IgG AP conjugate, and BCIP/NBT substrate. Ba-L gp140 DC 4E10 is detected

with 4E10 (lane 1), 2F5 (lane 2) and HIV-1(+) human serum (lane 3), and not detected with normal human serum (lane 4). No band corresponding to Ba-L gp140 DC 4E10 is detected in the naïve sample with 4E10 (lane 5), 2F5 (lane 6) and HIV-1(+) human serum (lane 7) or normal human serum (lane 8).

Figure 3 depicts the nucleic acid sequence of HIV-1 Ba-L gp140 DC 4E10 (SEQ ID NO: 6). The tPa signal is highlighted.

Figure 4 depicts the amino acid sequence of Ba-L gp140 DC 4E10 protein (SEQ ID NO: 7). The tPa signal is highlighted.

Figure 5 depicts the nucleic acid sequence of HIV-1 Ba-L gp145 (SEQ ID NO: 8). Highlighted is the tPa signal peptide. This sequence is identical in the gp145 region to Figure 3, differing only in the tPa signal sequence.

Figure 6 depicts the amino acid sequence of HIV-1 Ba-L gp145 protein (SEQ ID NO: 9). The tPa signal is highlighted.

Figure 7: Mammalian expression plasmid pJWTCDE-N.

Figure 8: HIV-1 subtype C gp160 expression plasmid. HIV-1 gp160 genes are ligated into pSWTIPK3 at the NheI and EcoRI sites in frame with the t-Pa signal peptide.

Figure 9 depicts the nucleic acid sequence for Clade C, C3728v2c6 gp160 (SEQ ID NO: 10). The HIV-1 C3728v2c6 gp160 nucleic acid sequence is codon optimized. The tPa signal is highlighted

Figure 10 depicts the nucleic acid sequence for Clade C, C3728v2c6 gp160 DC (SEQ ID NO: 11). The HIV-1 C3728v2c6 gp160 DC nucleic acid sequence is codon optimized.

The tPa signal is highlighted. The gp120/gp41 cleavage site is mutated to prevent cleavage.

Figure 11 depicts the nucleic acid sequence for Clade C, C06838v1c48 gp160 (SEQ ID NO: 12). The HIV-1 C06838v1c48 gp160 nucleic acid sequence is codon optimized. The tPa signal is highlighted.

Figure 12 depicts the nucleic acid sequence for Clade C, C06838v1c48 gp160 DC (SEQ ID NO: 13). The HIV-1 C06838v1c48 gp160 DC nucleic acid sequence is codon optimized. The tPa signal is highlighted. The gp120/gp41 cleavage site is mutated to prevent cleavage.

Figure 13 depicts the nucleic acid sequence for Clade C, C06980v1c3 gp160 (SEQ ID NO: 14). The HIV-1 C06980v1c3 gp160 nucleic acid sequence is codon optimized. The tPa signal is highlighted.

Figure 14 depicts the nucleic acid sequence for Clade C, C06980v1c3 gp160 DC (SEQ ID NO: 15). The HIV-1 C06980v1c3 gp160 DC nucleic acid sequence is codon optimized. The tPa signal is highlighted. The gp120/gp41 cleavage site is mutated to prevent cleavage.

Figure 15 depicts the nucleic acid sequence for Clade C, C06980v0c22 gp160 (SEQ ID NO: 16). The HIV-1 C06980v0c22 gp160 nucleic acid sequence is codon optimized. The tPa signal is highlighted.

Figure 16 depicts the nucleic acid sequence for Clade C, C06980v0c22 gp160 DC (SEQ ID NO: 17). The HIV-1 C06980v0c22 gp160 DC nucleic acid sequence is codon optimized. The tPa signal is highlighted. The gp120/gp41 cleavage site is mutated to prevent cleavage.

Figure 17 is a schematic of the pSWTIPK3 vector.

Figure 18 depicts the nucleic acid sequence for pSWTIPK3 (SEQ ID NO: 18).

Figure 19: IP western blot of CHO-K1 cells transfected with HIV-1 subtype C gp160 and gp160 DC expression plasmids. Env proteins are immunoprecipitated from the 48 hr. post-transfection cell lysates using HIV-1 (+) human serum, resolved on 4-15% SDS-PAGE and transferred to PVDF. Proteins from the following constructs were detected using rabbit antibodies to HIV-1 subtype B gp160: C06838v1c48 gp160 (lane 1), C06980v1c3 gp160 (lane 2), C06980v0c22 gp160 (lane 3), C3728v2c6 gp160 (lane 4), C06838v1c48 gp160DC (lane 5), C06980v1c3 gp160DC (lane 6), C06980v0c22 gp160DC (lane 7), C3728v2c6 gp160DC (lane 8), naïve CHO-K1 (-) control (lane 9), Ba-L gp145 (+) control (lane 10) and subtype C 96ZM651 gp140 (+) control (lane 11). Molecular weight protein markers are run in lane 12.

Figure 20: IP western blot of HEK293 cells transfected with HIV-1 subtype C gp160 and gp160 DC expression plasmids. Env proteins are immunoprecipitated from the 48 hr. post-transfection cell lysates using huMAb to gp41 (4E10), resolved on 4-15% SDS-PAGE and transferred to PVDF. Proteins from the following constructs were detected using rabbit antibodies to HIV-1 subtype B gp160: C06838v1c48 gp160 (lane 1), C06980v1c3 gp160 (lane 2), C06980v0c22 gp160 (lane 3), C3728v2c6 gp160 (lane 4), C06838v1c48 gp160DC (lane 5), C06980v1c3 gp160DC (lane 6), C06980v0c22 gp160DC (lane 7), C3728v2c6 gp160DC (lane 8), naïve HEK293 (-) control (lane 9), Ba-L gp160 (+) control (lane 10), naïve CHO-K1 (-) control (lane 11) and CHO-K1/Ba-L gp160 (+) control (lane 12). Molecular weight protein markers are run in lane 13.

Figure 21: HIV-1 C06980v0c22 gp145 expression plasmid. The HIV-1 gp145 gene is ligated into pSWTIPK3 at the NheI and EcoRI sites in frame with the t-Pa signal peptide.

Figure 22 depicts the nucleic acid sequence for pSWC06980v0c22 gp145 (SEQ ID NO: 19).

Figure 23 depicts the codon optimized nucleic acid sequence for HIV-1 C06980v0c22 gp145 (SEQ ID NO: 20). The tPa signal is highlighted.

Figure 24 is the protein sequence of HIV-1 C06980v0c22 gp145 (SEQ ID NO: 32). The tPa signal is highlighted.

Figure 25A (SEQ ID NO: 21) and B (SEQ ID NO: 22) depicts tPA sequences employed in the Env proteins of the invention.

Figure 26: IP western blot of CHO-K1 cells transfected with pSWC06980v0c22 gp145. Env proteins are immunoprecipitated from the 48 hr. post-transfection conditioned media and cell lysates using HIV-1 (+) human serum, resolved on 4-15% SDS-PAGE and transferred to PVDF. The gp145 was detected using rabbit antibodies to HIV-1 subtype B gp160 and Subtype C gp120: naïve CHO-K1 (-) control media (lane 1), supercoiled pSWC06980v0c22 gp145 media (lane 2), linearized pSWC06980v0c22 gp145 media (lane 3), naïve CHO-K1 (-) control cell lysate (lane 4), supercoiled pSWC06980v0c22 gp145 cell lysate (lane 5), linearized pSWC06980v0c22 gp145 cell lysate (lane 6). Molecular weight protein markers are run in lane 7.

Figure 27: 4-15% SDS-PAGE of C06980v0c22 gp145 purified from the conditioned media of CHO cell lines H-73-9-2-8 and H-73-9-3-9. 5µg protein is resolved under reducing and nonreducing conditions and stained with coomassie blue R250: H-73-9-2-8 nonreduced (lane 1), H-73-9-3-9 nonreduced (lane 2), H-73-9-2-8 reduced (lane 3) and H-73-9-3-9 reduced (lane 4). A molecular weight protein marker is run in lane 5.

Figure 28: Western blot of C06980v0c22 gp145 purified from the conditioned media of CHO cell lines H-73-9-2-8 and H-73-9-3-9. 0.5µg protein is resolved under reducing and nonreducing conditions on 4-15% SDS-PAGE, transferred to PVDF and detected with an HIV-1 (+) serum: H-73-9-2-8 nonreduced (lane 1), H-73-9-3-9 nonreduced (lane 2), H-73-9-2-8 reduced (lane 3) and H-73-9-3-9 reduced (lane 4). A molecular weight protein marker is run in lane 5.

Figure 29: Flow chart of downstream purification methods for gp145

Figure 30. SE-HPLC analysis of purified Recombinant HIV-1C06980v0c22 gp145 (lot 112009). 1:10 dilution of purified protein was prepared in 1X PBS and 20 µL was loaded on the TSK-GEL 3000SWXL Column (TOSOH BIOSEP). The column was eluted with isocratic gradient of 1X PBS at flow rate of 1.0 mL/min, resulting in the identification of 4 gp145 species.

Figure 31: Flow chart of downstream purification methods for gp145.

Figure 32: SE-HPLC analysis of purified Recombinant HIV-1C06980v0c22 gp145 lot 120710A. 1:10 dilution of purified protein was prepared in 1X PBS and 20 µL was loaded on the TSK-GEL 3000SWXL Column (TOSOH BIOSEP). The column was eluted with isocratic gradient of 1X PBS at flow rate of 1.0 mL/min, resulting in the identification of 4 gp145 species.

Figure 33. Flow chart of downstream purification methods for gp145.

Figure 34: SE-HPLC analysis of purified Recombinant HIV-1C06980v0c22 gp145 lot 120710B. 1:10 dilution of purified protein was prepared in 1X PBS and 20 µL was loaded on the TSK-GEL 3000SWXL Column (TOSOH BIOSEP). The column was eluted with isocratic gradient of 1X PBS at flow rate of 1.0 mL/min, resulting in the identification of 4 gp145 species.

Figure 35: HIV-1 Subtype C Env Sequence Alignment. Boxed region represents the gp41/gp120 cleavage domain. The amino terminal amino acid is a serine derived from the NheI cloning site at the tPa signal terminus.

Figure 36: C06980v0c22 gp145 amino acid sequence (SEQ ID NO: 32). The tPA signal, cleavage site mutations, and C-terminal triple lysine are as indicated in the boxed regions.

Figure 37: C06980v0c22 gp145 nucleotide sequence (SEQ ID NO: 20) and translation (SEQ ID NO: 32). The tPA leader sequence is indicated between the nucleic acid sequence and translation. The cleavage site mutations and terminal lysine repeat are as indicated in the boxed regions.

Figure 38 depicts the C-terminal residues of an HIV Env protein according to the invention (SEQ ID NOS: 25 and 43).

Figure 39 Antigenicity: 4E10 and VRC01 bind to CO6980 gp145 by ELISA.

Figure 40 Neutralization: the CO6980 PV is Sensitive to the 4E10 and VRC01 mAbs.

Figure 41 depicts the Rabbit clade C gp145 Study design.

Figure 42 gp145 clade C immunized rabbit sera neutralize Tier 1 pseudoviruses from clade B and C.

Figure 43 Development of neutralizing antibodies against the HIV-1 clade C Tier 1 pseudovirus post-immunization.

Figure 44 Cross-clade neutralization of B, C and AE IMC in the PBMC assay using gp145 immunized rabbit sera.

Figure 45 Development of neutralizing antibodies against the HIV-1 clade B BaL IMC post-immunization.

Figure 46 The gp145 immunized rabbit sera bind clade C Envs.

Figure 47 Neutralization of GS015 is IgG Mediated (TZMbl Neutralization assay).

Figure 48 Composite of neutralization values by immunogen.

Figure 49 depicts the results of an I.P. Western blot of protein-free media adapted CHO C06980v0c22 gp145 cell. The gp145 is precipitated from the conditioned media using human antibodies, resolved on 4-15 % SDS-PAGE, transferred to PVDF and detected using rabbit antisera to gp120 and gp160.

Figure 50 Homologies and Glycosylation Sites.

Figure 51 Amino acid sequences of clade D gp140 and clade C gp145 after codon optimization.

Figure 52 Binding of sera from immunized rabbits to different Envs.

Figure 53: IFN γ ELISPOT results in the A) lymph node and B) spleen, shown as spot count after stimulation with HIV-1 antigen.

Figure 54: Detection of IL-2 expression by intracellular staining.

Figure 55: ELISA binding titers of all groups against HIV-1 envelope protein gp145 and gp140.

Figure 56: Neutralization results of all groups in two assay platforms, A) TZMbl and B) PBMC.

Figure 57. Flow-cytometry based $\alpha 4\beta 7$ binding and inhibition assay.

Figure 58. Induction of $\alpha 4\beta 7$ expression on primary T cells. CD4⁺ (upper panels) and CD8⁺ (lower panels) T cells isolated from PBMC by magnetic bead separation were cultured with anti-CD3/CD28, IL-2 and retinoic acid for 5d. The primary CD4⁺ (upper panels) and CD8⁺ (lower panels) T cells were cultured to express $\alpha 4\beta 7$ bound to recombinant gp120 or gp145 protein, or a cyclic peptide containing the V2 loop region of Env. Bound protein/peptide (blue histogram) and no-protein neutravidin-PE control (green histogram) are shown. The gp145 panels show that CD8⁺ cells are 89.7% positive, and CD4⁺ T cells are 93.3% positive for $\alpha 4\beta 7$ binding to gp145.

Figure 59. Binding of HIV-1 Env to $\alpha 4\beta 7$ expressing T cells. Primary CD4⁺ (upper panels) and CD8⁺ (lower panels) T cells cultured to express $\alpha 4\beta 7$ bound to recombinant gp120 or gp145 protein, or a cyclic peptide containing the V2 loop region of Env. Bound protein/peptide (blue histogram) and no-protein neutravidin-PE control (green histogram) are shown.

Figure 60. Blocking interactions between V2 and $\alpha 4\beta 7$. Primary isolated T cells were cultured to express $\alpha 4\beta 7$ as described. Anti-V2 monoclonal antibodies were preincubated with biotinylated (A) gp120 or (B) cyclic-V2 peptide prior to binding to cells. Anti- $\alpha 4$ monoclonal antibody was pre-bound to cells prior to protein addition as a positive control.

Figure 61: Blue Native PAGE of C06980v0c22 gp145 proteins. 5 μ g of the following proteins are resolved on a 4-16% Novex Bis-Tris gel using Invitrogen's Native PAGE system: lot 112009 (lane 1), lot 120710A (lane 2) and lot 120710B (lane 3). Molecular

weight protein markers are run in lane 4. For each lot of gp145, 3 multimeric species (A, B and C) are evident. Multimer B is predominant.

Figure 62: SDS-PAGE of EGS crosslinked C06980v0c22 gp145 (lot 120710A) under nonreducing conditions. 5µg of gp145 was treated with 12.5, 5, 1, 0.2 and 0 mM EGS and resolved on a 3-8 % NuPAGE Tris Acetate polyacrylamide gel under nonreducing conditions: 12.5 mM EGS (lane 1), 5 mM EGS (lane 2), 1 mM EGS (lane 3), 0.2 mM EGS (lane 4) and 0 mM EGS (lane 5). EGS crosslinked phosphorylase B was run in lane 6 as a molecular weight protein marker.

Figure 63: SDS-PAGE of EGS crosslinked C06980v0c22 gp145 (lot 120710A) purified on Superose 6. 10 µg of gp145 from the column load and eluted fractions was crosslinked with 5mM EGS. Crosslinked and noncrosslinked gp145 was resolved on a 3-8 % NuPAGE Tris Acetate polyacrylamide gel: 5 µg noncrosslinked gp145 column load (A), 10 µg crosslinked gp145 column load (B) and 10 µg crosslinked eluted fractions 26-32 (C). EGS crosslinked phosphorylase B was run as a molecular weight protein marker (D).

DETAILED DESCRIPTION

Many candidate HIV vaccines do not interact with the natural neutralizing antibodies in humans. As described herein, Applicants have demonstrated that the HIV-1 Env can be modified to bind broadly reactive antibodies. Accordingly, the instant invention provides methods and related compositions pertaining to novel HIV Env proteins.

The novel HIV Env proteins of the invention comprise the entire ectodomain of an HIV Env protein, including the membrane proximal external region (MPER) of gp41. The gp41 protein consists of three main domains, namely, the ectodomain, the transmembrane domain, and the cytoplasmic tail. The ectodomain consists of the fusion peptide, N-terminal heptad repeat, C-terminal heptad repeat, and the MPER.

The MPER of gp41 typically comprises the last 24-28 C-terminal amino acids of the gp41 ectodomain. The MPER is a highly conserved region of the HIV Env protein and contains epitopes for broadly neutralizing human monoclonal antibodies, in particular, the 2F5, Z13, and 4E10 monoclonal antibodies.

The inventive HIV Env proteins described herein are truncated HIV Env proteins that are mutated in the gp120/gp41 cleavage site to prevent protease cleavage, comprise the MPER of gp41, and are truncated prior to the transmembrane domain. The HIV Env proteins of the invention may comprise any native MPER of the gp41 of an HIV Env protein.

In certain embodiments, the HIV Env proteins of the invention comprise an MPER sequence comprising the amino acid sequence, ALDSWNNLWNWFDIS (SEQ ID NO: 23). In certain embodiments, the HIV Env proteins of the invention comprise an MPER sequence comprising the amino acid sequence, LWYIK (SEQ ID NO: 24). In some embodiments, the MPER sequence comprises the amino acid sequence, ELLALDSWNNLWNWFDISNWLWYIK (SEQ ID NO: 25). In other embodiments, the MPER sequence comprises the amino acid sequence, DLLALDSWKNLWNWFDITNWLWYIK (SEQ ID NO: 26).

Typically, an HIV Env protein of the invention will comprise at least one of the epitopes for the monoclonal antibodies, 2F5, Z13, and 4E10. Examples of 2F5 epitopes include ALDSWN (SEQ ID NO: 27) as disclosed herein, ELDKWA (SEQ ID NO: 28), and EKNEQELLELDKWASLW (SEQ ID NO: 29) (*see, e.g.*, Montero, M, et al., Microbiology and Molecular Biology Reviews (2008) 72(1):54-84 and references cited therein). Examples of 4E10 epitopes include NWFDIS (SEQ ID NO: 30) as disclosed herein and NWFDIT (SEQ ID NO: 31). *Id.*

The HIV Env proteins of the instant invention lack the transmembrane domain and cytoplasmic tail but comprise the entire ectodomain of gp41. In certain embodiments, the ectodomain is modified to comprise about 1-10 hydrophilic amino acids at its C-terminus. The hydrophilic amino acid residues are typically added to the ectodomain of a truncated HIV Env protein of the invention in order to, in certain embodiments, improve

exposure of this region by making it more hydrophilic. In embodiments wherein the HIV Env protein comprises about 1-10 hydrophilic amino acids at its C-terminus, the hydrophilic amino acids are typically contiguous with the final amino acid residue of the native MPER sequence. Thus, the 1-10 hydrophilic amino acids typically comprise the final amino acid residues at the C-terminus of the HIV Env protein. In certain embodiments, the 1-10 hydrophilic amino acids are one or more lysine residues.

In certain embodiments, the HIV Env proteins of the instant invention are derived from an HIV strain isolated from an individual with an acute HIV infection. In other embodiments, the HIV infection is chronic. In certain embodiments, the HIV Env protein is derived from an HIV-1 strain classified in a group that is M, O, N, or P. In a particular embodiment, the HIV Env protein is derived from an HIV-1 group M strain. In further embodiments, the HIV-1 group M strain is a subtype (clade) selected from A, B, C, D, F, G, H, J, K, and hybrids thereof. In further embodiments, the HIV Env proteins are derived from an HIV-1 Group M strain that is a Clade B, Clade C, or Clade D strain. In some embodiments, the Clade B, C, or D strain is isolated from an individual with an acute infection. In other embodiments, the Clade B, C, or D strain is isolated from an individual with a chronic infection. Examples of suitable parent HIV strains from which the Env proteins of the instant invention can be derived include the HIV-1 Clade D sequences depicted in GenBank under Accession Nos. AF484477, AF484511, and AF484502 and the HIV-1 Clade C sequences depicted in GenBank under Accession Nos. HM215344 and HM215345.

In certain embodiments, the HIV Env proteins described herein are useful as immunogens in different forms to use as HIV vaccine components to elicit bNabs, e.g., against HIV-1. The different forms of the HIV Env can be used in a prime, as DNA/vector expressing the protein/protein and/or as a boost as protein. For example, in some embodiments, an HIV Env protein of the invention is administered to a mammal as a DNA vaccine, followed by administration of a boost as protein. In further embodiments, the HIV Env protein is administered as nucleic acid in a plasmid, followed by administration in a viral vector (e.g., as nucleic acid in an MVA), followed by

administration as a protein. In some embodiments, the inventive HIV Env proteins could also be used as particulate immunogens by crosslinking to virus particles like Qbeta, cow pea mosaic virus, CRM, HPV, HBsAg, etc.

In certain embodiments, HIV Env proteins of the instant invention are utilized as reagents for screening of new broad neutralizing antibodies and/or mapping of human sera with broad neutralizing serum activity and animal sera following immunization studies. In other embodiments, HIV Env proteins of the instant invention are utilized for screening of small molecules that compete for binding of broad neutralizing antibodies. The identified small molecules could be used as immunogens or anti-viral compounds.

As described herein, Applicants have generated recombinant Env proteins with unique sequences in which Applicants have modified the leader, modified the cleavage site for gp120/gp41, added a hydrophilic amino acid-tail and terminated the sequence before the transmembrane domain such that it comprises the full ectodomain of gp41. The DNA sequences are unique as they are codon optimized.

In another advantageous embodiment, the HIV Env proteins have substantially similar sequences to the HIV Env protein sequences depicted in Figures 4, 6, 24, 35, 36, and/or 37. In another particularly advantageous embodiment, the HIV Env proteins have a substantially similar MPER sequence to the MPER sequence depicted in Figure 38.

In a particularly advantageous embodiment, the HIV Env proteins of the present invention have about 75%, about 76%, about 77%, about 78%, about 79%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99% or about 100% sequence identity to SEQ ID NO: 1 or any of the HIV Env protein sequences depicted in Figures 3-6, 23, 24, and 35-37.

In one embodiment, the HIV Env proteins of the present invention may be used as reagents to screen for and identify new broadly neutralizing antibodies. Assays for screening for neutralizing antibodies are known in the art. A neutralization assay approach has been described previously (Binley JM, et al., (2004). Comprehensive Cross-

Clade Neutralization Analysis of a Panel of Anti-Human Immunodeficiency Virus Type 1 Monoclonal Antibodies. *J. Virol.* 78: 13232-13252). Pseudotyped viruses may be generated by co-transfecting cells with at least two plasmids encoding the soluble Env cDNA of the present invention and the rest of the HIV genome separately. In the HIV genome encoding vector, the Env gene may be replaced by the firefly luciferase gene. Transfectant supernatants containing pseudotyped virus may be co-incubated overnight with B cell supernatants derived from activation of an infected donor's primary peripheral blood mononuclear cells (PBMCs) or with monoclonal or polyclonal (serum) antibodies. Cells stably transfected with and expressing CD4 plus the CCR5 and CXCR4 coreceptors may be added to the mixture and incubated for 3 days at 37° C. Infected cells may be quantified by luminometry.

In some embodiments, for the screening of broad neutralizing antibodies, an envelope-enzyme fusion protein may be constructed by attaching an enzyme to the C-terminal end of an envelope protein. Virus particles comprising of the fusion protein and wild type and/or soluble envelope glycoprotein may be generated and used to infect target cells in the presence of a patients' sera. Activities of enzyme measured in such infected cells are measures of virus binding and entry to the target cells that are mediated by the wild type viral envelope protein. Examples of enzymes that can be used to generate the fusion protein include, but are not limited to, luciferase, bacterial or placental alkaline phosphatase, β -galactosidase, and fluorescent proteins such as Green fluorescent protein or toxins. The assay, in general, can also be carried out in 96-well plate. Decreased enzyme activities in the presence of the sera indicate that there are neutralizing antibodies in the sera.

As used herein, the terms "drug," "agent," and "compound" encompass any composition of matter or mixture which provides some pharmacologic effect that can be demonstrated in-vivo or in vitro. This includes small molecules, antibodies, microbiologicals, vaccines, vitamins, and other beneficial agents. As used herein, the terms further include any physiologically or pharmacologically active substance that produces a localized or systemic effect in a patient.

Nucleic Acids, Proteins, and Recombinant Technology

The present invention employs, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA and immunology, which are within the capabilities of a person of ordinary skill in the art. Such techniques are explained in the literature. See, for example, J. Sambrook, E. F. Fritsch, and T. Maniatis, 1989, *Molecular Cloning: A Laboratory Manual*, Second Edition, Books 1-3, Cold Spring Harbor Laboratory Press; Ausubel, F. M. et al. (1995 and periodic supplements; *Current Protocols in Molecular Biology*, ch. 9, 13, and 16, John Wiley & Sons, New York, N.Y.); B. Roe, J. Crabtree, and A. Kahn, 1996, *DNA Isolation and Sequencing: Essential Techniques*, John Wiley & Sons; M. J. Gait (Editor), 1984, *Oligonucleotide Synthesis: A Practical Approach*, Irl Press; and, D. M. J. Lilley and J. E. Dahlberg, 1992, *Methods of Enzymology: DNA Structure Part A: Synthesis and Physical Analysis of DNA* Methods in Enzymology, Academic Press. Each of these general texts is herein incorporated by reference.

The term “nucleic acid” encompasses DNA, RNA (e.g., mRNA, tRNA), heteroduplexes, and synthetic molecules capable of encoding a polypeptide and includes all analogs and backbone substitutes such as PNA that one of ordinary skill in the art would recognize as capable of substituting for naturally occurring nucleotides and backbones thereof. Nucleic acids may be single stranded or double stranded, and may be chemical modifications. The terms “nucleic acid” and “polynucleotide” are used interchangeably. Because the genetic code is degenerate, more than one codon may be used to encode a particular amino acid, and the present compositions and methods encompass nucleotide sequences which encode a particular amino acid sequence.

Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation, respectively.

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. The conventional one-letter or three-letter code for amino acid residues are used herein.

As used herein, a "synthetic" molecule is produced by *in vitro* chemical or enzymatic synthesis rather than by an organism.

As used herein, the term "expression" refers to the process by which a polypeptide is produced based on the nucleic acid sequence of a gene. The process includes both transcription and translation.

A "gene" refers to the DNA segment encoding a polypeptide or RNA.

An "isolated" polynucleotide or polypeptide is one that is substantially free of the materials with which it is associated in its native environment. By substantially free, is meant at least 50%, advantageously at least 70%, more advantageously at least 80%, and even more advantageously at least 90% free of these materials.

An "isolated" nucleic acid molecule is a nucleic acid molecule separate and discrete from the whole organism with which the molecule is found in nature; or a nucleic acid molecule devoid, in whole or part, of sequences normally associated with it in nature; or a sequence, as it exists in nature, but having heterologous sequences in association therewith.

"Native" proteins or polypeptides refer to proteins or polypeptides isolated from the source in which the proteins naturally occur. "Recombinant" polypeptides refer to polypeptides produced by recombinant DNA techniques; *e.g.*, produced from cells transformed by an exogenous DNA construct encoding the desired polypeptide. "Synthetic" polypeptides include those prepared by chemical synthesis as well as the synthetic antigens described above.

By "homolog" is meant an entity having a certain degree of identity with the subject amino acid sequences and the subject nucleotide sequences. As used herein, the term "homolog" covers identity with respect to structure and/or function, for example, the expression product of the resultant nucleotide sequence has the enzymatic activity of a subject amino acid sequence. With respect to sequence identity, preferably there is at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or even 99% sequence identity. These terms also encompass allelic variations of the sequences. The term, homolog, may apply to the relationship between genes separated by the event of speciation or to the relationship between genes separated by the event of genetic duplication.

Relative sequence identity can be determined by commercially available computer programs that can calculate % identity between two or more sequences using any suitable algorithm for determining identity, using, for example, default parameters. A typical example of such a computer program is CLUSTAL. Advantageously, the BLAST algorithm is employed, with parameters set to default values. The BLAST algorithm is described in detail on the National Center for Biotechnology Information (NCBI) website.

The homologs of the peptides as provided herein typically have structural similarity with such peptides. A homolog of a polypeptide includes one or more conservative amino acid substitutions, which may be selected from the same or different members of the class to which the amino acid belongs.

In one embodiment, the sequences may also have deletions, insertions or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent substance. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the secondary binding activity of the substance is retained. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include leucine,

isoleucine, valine, glycine, alanine, asparagine, glutamine, serine, threonine, phenylalanine, and tyrosine.

The present invention also encompasses conservative substitution (substitution and replacement are both used herein to mean the interchange of an existing amino acid residue with an alternative residue) that may occur e.g., like-for-like substitution such as basic for basic, acidic for acidic, polar for polar, etc. Non-conservative substitution may also occur e.g., from one class of residue to another or alternatively involving the inclusion of unnatural amino acids such as ornithine (hereinafter referred to as Z), diaminobutyric acid ornithine (hereinafter referred to as B), norleucine ornithine (hereinafter referred to as O), pyridylalanine, thienylalanine, naphthylalanine and phenylglycine. Conservative substitutions that may be made are, for example, within the groups of basic amino acids (Arginine, Lysine and Histidine), acidic amino acids (glutamic acid and aspartic acid), aliphatic amino acids (Alanine, Valine, Leucine, Isoleucine), polar amino acids (Glutamine, Asparagine, Serine, Threonine), aromatic amino acids (Phenylalanine, Tryptophan and Tyrosine), hydroxyl amino acids (Serine, Threonine), large amino acids (Phenylalanine and Tryptophan) and small amino acids (Glycine, Alanine).

Many methods of amplifying DNA are known in the art, and any such method can be used, see for example Sambrook et al., *Molecular Cloning; A Laboratory Manual* 2d ed. (1989). For example, a DNA fragment of interest can be amplified using the polymerase chain reaction or some other cyclic polymerase mediated amplification reaction.

The amplified region of DNA can then be sequenced using any method known in the art. Advantageously, the nucleic acid sequencing is by automated methods (reviewed by Meldrum, *Genome Res.* September 2000;10(9):1288-303, the disclosure of which is incorporated by reference in its entirety), for example using a Beckman CEQ 8000 Genetic Analysis System (Beckman Coulter Instruments, Inc.). Methods for sequencing nucleic acids include, but are not limited to, automated fluorescent DNA sequencing (see, e.g., Watts & MacBeath, *Methods Mol Biol.* 2001;167:153-70 and MacBeath et al., *Methods Mol Biol.* 2001;167:119-52), capillary electrophoresis (see, e.g., Bosserhoff et

al., Comb Chem High Throughput Screen. December 2000;3(6):455-66), DNA sequencing chips (see, e.g., Jain, Pharmacogenomics. August 2000;1(3):289-307), mass spectrometry (see, e.g., Yates, Trends Genet. January 2000;16(1):5-8), pyrosequencing (see, e.g., Ronaghi, Genome Res. January 2001;11(1):3-11), and ultrathin-layer gel electrophoresis (see, e.g., Guttman & Ronai, Electrophoresis. December 2000; 21 (18):3952-64), the disclosures of which are hereby incorporated by reference in their entirety. The sequencing can also be done by any commercial company. Examples of such companies include, but are not limited to, the University of Georgia Molecular Genetics Instrumentation Facility (Athens, Ga.) or SeqWright DNA Technologies Services (Houston, Tex.).

Any one of the methods known in the art for amplification of DNA may be used, such as for example, the polymerase chain reaction (PCR), the ligase chain reaction (LCR) (Barany, F., Proc. Natl. Acad. Sci. (U.S.A.) 88:189-193 (1991)), the strand displacement assay (SDA), or the oligonucleotide ligation assay ("OLA") (Landegren, U. et al., Science 241:1077-1080 (1988)). Nickerson, D. A. et al. have described a nucleic acid detection assay that combines attributes of PCR and OLA (Nickerson, D. A. et al., Proc. Natl. Acad. Sci. (U.S.A.) 87:8923-8927 (1990)). Other known nucleic acid amplification procedures, such as transcription-based amplification systems (Malek, L. T. et al., U.S. Pat. No. 5,130,238; Davey, C. et al., European Patent Application 329,822; Schuster et al., U.S. Pat. No. 5,169,766; Miller, H. I. et al., PCT Application W089/06700; Kwoh, D. et al., Proc. Natl. Acad. Sci. (U.S.A.) 86:1173 (1989); Gingeras, T. R. et al., PCT Application W088/10315)), or isothermal amplification methods (Walker, G. T. et al., Proc. Natl. Acad. Sci. (U.S.A.) 89:392-396 (1992)) may also be used.

To perform a cyclic polymerase mediated amplification reaction according to the present invention, the primers are hybridized or annealed to opposite strands of the target DNA, the temperature is then raised to permit the thermostable DNA polymerase to extend the primers and thus replicate the specific segment of DNA spanning the region between the two primers. Then the reaction is thermocycled so that at each cycle the

amount of DNA representing the sequences between the two primers is doubled, and specific amplification of gene DNA sequences, if present, results.

Any of a variety of polymerases can be used in the present invention. For thermocyclic reactions, the polymerases are thermostable polymerases such as Taq, KlenTaq, Stoffel Fragment, Deep Vent, Tth, Pfu, Vent, and UITma, each of which are readily available from commercial sources. For non-thermocyclic reactions, and in certain thermocyclic reactions, the polymerase will often be one of many polymerases commonly used in the field, and commercially available, such as DNA pol I, Klenow fragment, T7 DNA polymerase, and T4 DNA polymerase. Guidance for the use of such polymerases can readily be found in product literature and in general molecular biology guides.

Typically, the annealing of the primers to the target DNA sequence is carried out for about 2 minutes at about 37-55° C, extension of the primer sequence by the polymerase enzyme (such as Taq polymerase) in the presence of nucleoside triphosphates is carried out for about 3 minutes at about 70-75° C, and the denaturing step to release the extended primer is carried out for about 1 minute at about 90-95° C. However, these parameters can be varied, and one of skill in the art would readily know how to adjust the temperature and time parameters of the reaction to achieve the desired results. For example, cycles may be as short as 10, 8, 6, 5, 4.5, 4, 2, 1, 0.5 minutes or less.

Also, "two temperature" techniques can be used where the annealing and extension steps may both be carried out at the same temperature, typically between about 60-65° C, thus reducing the length of each amplification cycle and resulting in a shorter assay time.

Typically, the reactions described herein are repeated until a detectable amount of product is generated. Often, such detectable amounts of product are between about 10 ng and about 100 ng, although larger quantities, e.g. 200 ng, 500 ng, 1 mg or more can also, of course, be detected. In terms of concentration, the amount of detectable product can be from about 0.01 pmol, 0.1 pmol, 1 pmol, 10 pmol, or more. Thus, the number of cycles of the reaction that are performed can be varied, the more cycles are performed, the more

amplified product is produced. In certain embodiments, the reaction comprises 2, 5, 10, 15, 20, 30, 40, 50, or more cycles.

For example, the PCR reaction may be carried out using about 25-50 μ l samples containing about 0.01 to 1.0 ng of template amplification sequence, about 10 to 100 pmol of each generic primer, about 1.5 units of Taq DNA polymerase (Promega Corp.), about 0.2 mM dDATP, about 0.2 mM dCTP, about 0.2 mM dGTP, about 0.2 mM dTTP, about 15 mM MgCl₂, about 10 mM Tris-HCl (pH 9.0), about 50 mM KCl, about 1 μ g/ml gelatin, and about 10 μ l/ml Triton X-100 (Saiki, 1988).

Those of ordinary skill in the art are aware of the variety of nucleotides available for use in the cyclic polymerase mediated reactions. Typically, the nucleotides will consist at least in part of deoxynucleotide triphosphates (dNTPs), which are readily commercially available. Parameters for optimal use of dNTPs are also known to those of skill, and are described in the literature. In addition, a large number of nucleotide derivatives are known to those of skill and can be used in the present reaction. Such derivatives include fluorescently labeled nucleotides, allowing the detection of the product including such labeled nucleotides, as described below. Also included in this group are nucleotides that allow the sequencing of nucleic acids including such nucleotides, such as chain-terminating nucleotides, dideoxynucleotides and boronated nuclease-resistant nucleotides. Commercial kits containing the reagents most typically used for these methods of DNA sequencing are available and widely used. Other nucleotide analogs include nucleotides with bromo-, iodo-, or other modifying groups, which affect numerous properties of resulting nucleic acids including their antigenicity, their replicatability, their melting temperatures, their binding properties, etc. In addition, certain nucleotides include reactive side groups, such as sulfhydryl groups, amino groups, N-hydroxysuccinimidyl groups, that allow the further modification of nucleic acids comprising them.

The term "oligonucleotide" is defined as a molecule comprised of two or more deoxyribonucleotides, preferably more than three. Its exact size will depend upon many factors which, in turn, depend upon the ultimate function and use of the oligonucleotide.

The term "primer" as used herein refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product, which is complementary to a nucleic acid strand, is induced, i.e., in the presence of nucleotides and an inducing agent such as a DNA polymerase and at a suitable temperature and pH. The primer may be either single-stranded or double-stranded and must be sufficiently long to prime the synthesis of the desired extension product in the presence of the inducing agent. The exact length of the primer will depend upon many factors, including temperature, source of primer and use for the method. In certain embodiments, oligonucleotides that can be used as primers to amplify specific nucleic acid sequences of a gene in cyclic polymerase-mediated amplification reactions, such as PCR reactions, consist of oligonucleotide fragments. Such fragments should be of sufficient length to enable specific annealing or hybridization to the nucleic acid sample. The sequences typically will be about 8 to about 44 nucleotides in length, but may be longer. Longer sequences, e.g., from about 14 to about 50, are advantageous for certain embodiments.

In embodiments where it is desired to amplify a fragment of DNA, primers having contiguous stretches of about 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, or 24 nucleotides from a gene sequence are contemplated.

As used herein, "hybridization" refers to the process by which one strand of nucleic acid base pairs with a complementary strand, as occurs during blot hybridization techniques and PCR techniques.

Whichever probe sequences and hybridization methods are used, one ordinarily skilled in the art can readily determine suitable hybridization conditions, such as temperature and chemical conditions. Such hybridization methods are well known in the art. For example, for applications requiring high selectivity, one will typically desire to employ relatively stringent conditions for the hybridization reactions, e.g., one will select relatively low salt and/or high temperature conditions, such as provided by about 0.02 M to about 0.10 M NaCl at temperatures of about 50° C to about 70° C. Such high

stringency conditions tolerate little, if any, mismatch between the probe and the template or target strand. It is generally appreciated that conditions can be rendered more stringent by the addition of increasing amounts of formamide. Other variations in hybridization reaction conditions are well known in the art (see for example, Sambrook et al., *Molecular Cloning; A Laboratory Manual* 2d ed. (1989)).

Hybridization conditions are based on the melting temperature (T_m) of the nucleic acid binding complex, as taught, e.g., in Berger and Kimmel (1987, *Guide to Molecular Cloning Techniques, Methods in Enzymology*, Vol 152, Academic Press, San Diego CA), and confer a defined "stringency" as explained below.

Maximum stringency typically occurs at about $T_m - 5^\circ\text{C}$ (5°C below the T_m of the probe); high stringency at about 5°C to 10°C below T_m ; intermediate stringency at about 10°C to 20°C below T_m ; and low stringency at about 20°C to 25°C below T_m . As will be understood by those of ordinary skill in the art, a maximum stringency hybridization can be used to identify or detect identical nucleotide sequences while an intermediate (or low) stringency hybridization can be used to identify or detect similar or related polynucleotide sequences.

In one aspect, the present invention employs nucleotide sequences that can hybridize to another nucleotide sequence under stringent conditions (e.g., 65°C and $0.1\times\text{SSC}$ { $1\times\text{SSC} = 0.15\text{ M NaCl}$, $0.015\text{ M Na}_3\text{ Citrate pH } 7.0$). Where the nucleotide sequence is double-stranded, both strands of the duplex, either individually or in combination, may be employed by the present invention. Where the nucleotide sequence is single-stranded, it is to be understood that the complementary sequence of that nucleotide sequence is also included within the scope of the present invention.

Stringency of hybridization refers to conditions under which polynucleic acid hybrids are stable. Such conditions are evident to those of ordinary skill in the field. As known to those of ordinary skill in the art, the stability of hybrids is reflected in the melting temperature (T_m) of the hybrid which decreases approximately 1 to 1.5°C with every 1% decrease in sequence homology. In general, the stability of a hybrid is a function of sodium ion concentration and temperature. Typically, the hybridization

reaction is performed under conditions of higher stringency, followed by washes of varying stringency.

As used herein, high stringency includes conditions that permit hybridization of only those nucleic acid sequences that form stable hybrids in 1 M Na⁺ at 65-68 °C. High stringency conditions can be provided, for example, by hybridization in an aqueous solution containing 6x SSC, 5x Denhardt's, 1 % SDS (sodium dodecyl sulphate), 0.1 Na⁺ pyrophosphate and 0.1 mg/ml denatured salmon sperm DNA as non-specific competitor. Following hybridization, high stringency washing may be done in several steps, with a final wash (about 30 minutes) at the hybridization temperature in 0.2 - 0.1x SSC, 0.1 % SDS.

It is understood that these conditions may be adapted and duplicated using a variety of buffers, e.g., formamide-based buffers, and temperatures. Denhardt's solution and SSC are well known to those of ordinary skill in the art as are other suitable hybridization buffers (see, e.g., Sambrook, et al., eds. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York or Ausubel, et al., eds. (1990) *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc.). Optimal hybridization conditions are typically determined empirically, as the length and the GC content of the hybridizing pair also play a role.

Nucleic acid molecules that differ from the sequences of the primers and probes disclosed herein, are intended to be within the scope of the invention. Nucleic acid sequences that are complementary to these sequences, or that are hybridizable to the sequences described herein under conditions of standard or stringent hybridization, and also analogs and derivatives are also intended to be within the scope of the invention. Advantageously, such variations will differ from the sequences described herein by only a small number of nucleotides, for example by 1, 2, or 3 nucleotides.

Nucleic acid molecules corresponding to natural allelic variants, homologues (i.e., nucleic acids derived from other species), or other related sequences (e.g., paralogs) of the sequences described herein can be isolated based on their homology to the nucleic acids disclosed herein, for example by performing standard or stringent hybridization

reactions using all or a portion of the known sequences as probes. Such methods for nucleic acid hybridization and cloning are well known in the art.

Similarly, a nucleic acid molecule detected in the methods of the invention may include only a fragment of the specific sequences described. Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids, a length sufficient to allow for specific hybridization of nucleic acid primers or probes, and are at most some portion less than a full-length sequence. Fragments may be derived from any contiguous portion of a nucleic acid sequence of choice. Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below.

Derivatives, analogs, homologues, and variants of the nucleic acids of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids of the invention, in various embodiments, by at least about 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or even 99% identity over a nucleic acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art.

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 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 (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; 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).

In all search programs in the suite the gapped alignment routines are integral to the database search itself. Gapping can be turned off if desired. The default penalty (Q) for a gap of length one is Q=9 for proteins and BLASTP, and Q=10 for BLASTN, but may be changed to any integer. The default per-residue penalty for extending a gap (R) is R=2 for proteins and BLASTP, and R=10 for BLASTN, but may be changed to any integer. Any combination of values for Q and R can be used in order to align sequences so as to maximize overlap and identity while minimizing sequence gaps. The default amino acid comparison matrix is BLOSUM62, but other amino acid comparison matrices such as PAM can be utilized.

Alternatively or additionally, the term "homology" or "identity", for instance, with respect to a nucleotide or amino acid sequence, can indicate a quantitative measure of homology between two sequences. The percent sequence homology can be calculated as $(N_{\text{ref}} - N_{\text{dif}}) * 100 / N_{\text{ref}}$, wherein N_{dif} is the total number of non-identical residues in the two sequences when aligned and wherein N_{ref} is the number of residues in one of the sequences. Hence, the DNA sequence AGTCAGTC will have a sequence identity of 75% with the sequence AATCAATC ($N_{\text{ref}} = 8$; $N_{\text{dif}} = 2$). "Homology" or "identity" can refer to the number of positions with identical nucleotides or amino acids divided by the number of nucleotides or amino acids in the shorter of the two sequences wherein

alignment of the two sequences can be determined in accordance with the Wilbur and Lipman algorithm (Wilbur & Lipman, Proc Natl Acad Sci USA 1983;80:726, incorporated herein by reference), for instance, using a window size of 20 nucleotides, a word length of 4 nucleotides, and a gap penalty of 4, and computer-assisted analysis and interpretation of the sequence data including alignment can be conveniently performed using commercially available programs (e.g., Intelligenetics.TM. Suite, Intelligenetics Inc. CA). When RNA sequences are said to be similar, or have a degree of sequence identity or homology with DNA sequences, thymidine (T) in the DNA sequence is considered equal to uracil (U) in the RNA sequence. Thus, RNA sequences are within the scope of the invention and can be derived from DNA sequences, by thymidine (T) in the DNA sequence being considered equal to uracil (U) in RNA sequences. Without undue experimentation, the skilled artisan can consult with many other programs or references for determining percent homology.

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. 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(dot)com). Thus, the nucleotide sequences of the invention can readily be codon optimized.

As used herein, the term "probe" refers to a molecule (e.g., an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly or by PCR amplification), that is capable of hybridizing to another molecule of interest (e.g., another oligonucleotide). When probes are oligonucleotides they may be single-stranded or double-stranded. Probes are useful in the detection, identification and isolation of particular targets (e.g., gene sequences). As described herein, it is contemplated that probes used in the present invention may be labelled with a label so that is detectable in any detection system, including, but not limited to enzyme

(e.g., ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, and luminescent systems.

The primers and probes described herein may be readily prepared by, for example, directly synthesizing the fragment by chemical means or by introducing selected sequences into recombinant vectors for recombinant production. Methods for making a vector or recombinants or plasmid for amplification of the fragment either in vivo or in vitro can be any desired method, e.g., a method which is by or analogous to the methods disclosed in, or disclosed in documents cited in: U.S. Pat. Nos. 4,603,112; 4,769,330; 4,394,448; 4,722,848; 4,745,051; 4,769,331; 4,945,050; 5,494,807; 5,514,375; 5,744,140; 5,744,141; 5,756,103; 5,762,938; 5,766,599; 5,990,091; 5,174,993; 5,505,941; 5,338,683; 5,494,807; 5,591,639; 5,589,466; 5,677,178; 5,591,439; 5,552,143; 5,580,859; 6,130,066; 6,004,777; 6,130,066; 6,497,883; 6,464,984; 6,451,770; 6,391,314; 6,387,376; 6,376,473; 6,368,603; 6,348,196; 6,306,400; 6,228,846; 6,221,362; 6,217,883; 6,207,166; 6,207,165; 6,159,477; 6,153,199; 6,090,393; 6,074,649; 6,045,803; 6,033,670; 6,485,729; 6,103,526; 6,224,882; 6,312,682; 6,348,450 and 6,312,683; U.S. patent application Ser. No. 920,197, filed Oct. 16, 1986; WO 90/01543; W091/11525; WO 94/16716; WO 96/39491; WO 98/33510; EP 265785; EP 0 370 573; Andreansky et al., Proc. Natl. Acad. Sci. USA 1996;93:11313-11318; Ballay et al., EMBO J. 1993;4:3861-65; Felgner et al., J. Biol. Chem. 1994;269:2550-2561; Frolov et al., Proc. Natl. Acad. Sci. USA 1996;93:11371-11377; Graham, Tibtech 1990;8:85-87; Grunhaus et al., Sem. Virol. 1992;3:237-52; Ju et al., Diabetologia 1998;41:736-739; Kitson et al., J. Virol. 1991;65:3068-3075; McClements et al., Proc. Natl. Acad. Sci. USA 1996;93:11414-11420; Moss, Proc. Natl. Acad. Sci. USA 1996;93:11341-11348; Paoletti, Proc. Natl. Acad. Sci. USA 1996;93:11349-11353; Pennock et al., Mol. Cell. Biol. 1984;4:399-406; Richardson (Ed), Methods in Molecular Biology 1995;39, "Baculovirus Expression Protocols," Humana Press Inc.; Smith et al. (1983) Mol. Cell. Biol. 1983;3:2156-2165; Robertson et al., Proc. Natl. Acad. Sci. USA 1996;93:11334-11340; Robinson et al., Sem. Immunol. 1997;9:271; and Roizman, Proc. Natl. Acad. Sci. USA 1996;93:11307-11312. Strategies for probe design are described in WO95/11995, EP 717,113 and WO97/29212.

The present invention further contemplates direct and indirect labelling techniques. For example, direct labelling incorporates fluorescent dyes directly into the nucleotide sequences that hybridize to the array-associated probes (e.g., dyes are incorporated into nucleotide sequence by enzymatic synthesis in the presence of labelled nucleotides or PCR primers). Direct labelling schemes yield strong hybridization signals, typically using families of fluorescent dyes with similar chemical structures and characteristics, and are simple to implement. In some embodiments comprising direct labelling of nucleic acids, cyanine or alexa analogs are utilized in multiple-fluor comparative array analyses. In other embodiments, indirect labelling schemes can be utilized to incorporate epitopes into the nucleic acids either prior to or after hybridization to the microarray probes. One or more staining procedures and reagents are used to label the hybridized complex (e.g., a fluorescent molecule that binds to the epitopes, thereby providing a fluorescent signal by virtue of the conjugation of dye molecule to the epitope of the hybridised species).

Oligonucleotide sequences used as probes according to the present invention may be labeled with a detectable moiety. Various labeling moieties are known in the art. Said moiety may be, for example, a radiolabel (e.g., ^3H , ^{125}I , ^{35}S , ^{14}C , ^{32}P , etc.), detectable enzyme (e.g. horse radish peroxidase (HRP), alkaline phosphatase etc.), a fluorescent dye (e.g., fluorescein isothiocyanate, Texas red, rhodamine, Cy3, Cy5, Bodipy, Bodipy Far Red, Lucifer Yellow, Bodipy 630/650-X, Bodipy R6G-X and 5-CR 6G, and the like), a colorimetric label such as colloidal gold or colored glass or plastic (e.g. polystyrene, polypropylene, latex, etc.), beads, or any other moiety capable of generating a detectable signal such as a colorimetric, fluorescent, chemiluminescent or electrochemiluminescent (ECL) signal.

Probes may be labeled directly or indirectly with a detectable moiety, or synthesized to incorporate the detectable moiety. In one embodiment, a detectable label is incorporated into a nucleic acid during at least one cycle of a cyclic polymerase-mediated amplification reaction. For example, polymerases can be used to incorporate fluorescent nucleotides during the course of polymerase-mediated amplification

reactions. Alternatively, fluorescent nucleotides may be incorporated during synthesis of nucleic acid primers or probes. To label an oligonucleotide with the fluorescent dye, one of conventionally-known labeling methods can be used (Nature Biotechnology, 14, 303-308, 1996; Applied and Environmental Microbiology, 63, 1143-1147, 1997; Nucleic Acids Research, 24, 4532-4535, 1996). An advantageous probe is one labeled with a fluorescent dye at the 3' or 5' end and containing G or C as the base at the labeled end. If the 5' end is labeled and the 3' end is not labeled, the OH group on the C atom at the 3'-position of the 3' end ribose or deoxyribose may be modified with a phosphate group or the like although no limitation is imposed in this respect.

Spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means can be used to detect such labels. The detection device and method may include, but is not limited to, optical imaging, electronic imaging, imaging with a CCD camera, integrated optical imaging, and mass spectrometry. Further, the amount of labeled or unlabeled probe bound to the target may be quantified. Such quantification may include statistical analysis. In other embodiments the detection may be via conductivity differences between concordant and discordant sites, by quenching, by fluorescence perturbation analysis, or by electron transport between donor and acceptor molecules.

In yet another embodiment, detection may be via energy transfer between molecules in the hybridization complexes in PCR or hybridization reactions, such as by fluorescence energy transfer (FET) or fluorescence resonance energy transfer (FRET). In FET and FRET methods, one or more nucleic acid probes are labeled with fluorescent molecules, one of which is able to act as an energy donor and the other of which is an energy acceptor molecule. These are sometimes known as a reporter molecule and a quencher molecule respectively. The donor molecule is excited with a specific wavelength of light for which it will normally exhibit a fluorescence emission wavelength. The acceptor molecule is also excited at this wavelength such that it can accept the emission energy of the donor molecule by a variety of distance-dependent energy transfer mechanisms. Generally the acceptor molecule accepts the emission

energy of the donor molecule when they are in close proximity (e.g., on the same, or a neighboring molecule). FET and FRET techniques are well known in the art. See for example U.S. Pat. Nos. 5,668,648, 5,707,804, 5,728,528, 5,853,992, and 5,869,255 (for a description of FRET dyes), Tyagi et al. *Nature Biotech.* vol. 14, p 303-8 (1996), and Tyagi et al., *Nature Biotech.* vol 16, p 49-53 (1998) (for a description of molecular beacons for FET), and Mergny et al. *Nucleic Acid Res.* vol 22, p 920-928, (1994) and Wolf et al. *PNAS* vol 85, p 8790-94 (1988) (for general descriptions and methods for FET and FRET), each of which is hereby incorporated by reference.

The nucleotide sequences of the present invention may be inserted into vectors. The term "vector" is widely used and understood by those of ordinary skill in the art, and as used herein the term "vector" is used consistent with its meaning to those of ordinary skill in the art. For example, the term "vector" is commonly used by those ordinarily 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.

For example, a vector is a replicon, such as plasmid, phage or cosmid, to which another DNA segment may be attached so as to bring about the replication of the attached segment. A "replicon" is any genetic element (e.g., plasmid, chromosome, virus) that functions as an autonomous unit of DNA replication *in vivo*; i.e., capable of replication under its own control. An "origin of replication" refers to those DNA sequences that participate in DNA synthesis. An "expression control sequence" is a DNA sequence that controls and regulates the transcription and translation of another DNA sequence. A coding sequence is "operably linked" and "under the control" of transcriptional and translational control sequences in a cell when RNA polymerase transcribes the coding sequence into mRNA, which is then translated into the protein encoded by the coding sequence.

In general, expression vectors containing promoter sequences which facilitate the efficient transcription and translation of the inserted DNA fragment are used in connection with the host. The expression vector typically contains an origin of

replication, promoter(s), terminator(s), as well as specific genes which are capable of providing phenotypic selection in transformed cells. When the polynucleotide encodes a polyprotein fragment, advantageously, in the vector, an initiation codon (ATG) is placed at 5' of the reading frame and a stop codon is placed at 3'. Other elements for controlling expression may be present, such as enhancer sequences, stabilizing sequences and signal sequences permitting the secretion of the protein. The transformed hosts can be fermented and cultured according to means known in the art to achieve optimal cell growth.

Any vector that allows expression of the immunogens of the present invention may be used in accordance with the present invention. In certain embodiments, the immunogens of the present invention may be used *in vitro* (such as using cell-free expression systems) and/or in cultured cells grown *in vitro*. For such applications, any vector that allows expression of the immunogens *in vitro* and/or in cultured cells may be used.

A DNA "coding sequence" is a double-stranded DNA sequence which is transcribed and translated into a polypeptide *in vivo* when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxyl) terminus. A coding sequence can include, but is not limited to, prokaryotic sequences, cDNA from eukaryotic mRNA, genomic DNA sequences from eukaryotic (e.g., mammalian) DNA, and even synthetic DNA sequences. A polyadenylation signal and transcription termination sequence will usually be located 3' to the coding sequence. A "cDNA" is defined as copy-DNA or complementary-DNA, and is a product of a reverse transcription reaction from an mRNA transcript.

Transcriptional and translational control sequences are DNA regulatory sequences, such as promoters, enhancers, ribosome binding sites, upstream regulatory domains, polyadenylation signals, terminators, and the like, that provide for the expression of a coding sequence in a host cell. A "cis-element" is a nucleotide sequence, also termed a "consensus sequence" or "motif", that interacts with other proteins which

can upregulate or downregulate expression of a specific gene locus. A "signal sequence" can also be included with the coding sequence. This sequence encodes a signal peptide, N-terminal to the polypeptide, that communicates to the host cell and directs the polypeptide to the appropriate cellular location. Signal sequences can be found associated with a variety of proteins native to prokaryotes and eukaryotes. Not all of these control sequences need always be present in a recombinant vector so long as the desired gene is capable of being transcribed and translated.

A "promoter sequence" is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. The promoter sequence is typically bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence is a transcription initiation site, as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase. Eukaryotic promoters often, but not always, contain "TATA" boxes and "CAT" boxes. Prokaryotic promoters contain Shine-Dalgarno sequences in addition to the -10 and -35 consensus sequences.

As used herein, the terms "restriction endonucleases" and "restriction enzymes" refer to enzymes which cut double-stranded DNA at or near a specific nucleotide sequence.

"Recombinant DNA technology" refers to techniques for uniting two heterologous DNA molecules, usually as a result of in vitro ligation of DNAs from different organisms. Recombinant DNA molecules are commonly produced by experiments in genetic engineering. Synonymous terms include "gene splicing", "molecular cloning" and "genetic engineering". The product of these manipulations results in a "recombinant" or "recombinant molecule".

A cell has been "transformed" or "transfected" with exogenous or heterologous DNA when such DNA has been introduced inside the cell. The transforming DNA may or may not be integrated (covalently linked) into the genome of the cell. In prokaryotes,

yeast, and mammalian cells for example, the transforming DNA may be maintained on an episomal element such as a vector or plasmid. With respect to eukaryotic cells, a stably transformed cell is one in which the transforming DNA has become integrated into a chromosome so that it is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter cells containing the transforming DNA. A "clone" is a population of cells derived from a single cell or ancestor by mitosis. A "cell line" is a clone of a primary cell that is capable of stable growth in vitro for many generations. An organism, such as a plant or animal, that has been transformed with exogenous DNA is termed "transgenic".

As used herein, the term "host" is meant to include not only prokaryotes but also eukaryotes such as yeast, plant and animal cells. Prokaryotic hosts may include *E. coli*, *S. typhimurium*, *Serratia marcescens* and *Bacillus subtilis*. Eukaryotic hosts include yeasts such as *Pichia pastoris*, mammalian cells and insect cells and plant cells, such as *Arabidopsis thaliana* and *Tobaccum nicotiana*. A number of mammalian cell lines are known in the art and include immortalized cell lines available from the American Type Culture Collection (ATCC), such as, but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (e.g., Hep G2), Mandin-Darby bovine kidney ("MDBK") cells, as well as others. Similarly, bacterial hosts such as *E. coli*, *Bacillus subtilis*, and *Streptococcus* spp., will find use with the present expression constructs. Yeast hosts useful in the present invention include inter alia, *Saccharomyces cerevisiae*, *Candida albicans*, *Candida maltosa*, *Hansenula polymorpha*, *Kluyveromyces fragilis*, *Kluyveromyces lactis*, *Pichia guillermondii*, *Pichia pastoris*, *Schizosaccharomyces pombe* and *Yarrowia lipolytica*. Insect hosts useful in the present invention include, but are not limited to, *Spodoptera frugiperda* cells.

A "heterologous" region of the DNA construct is an identifiable segment of DNA within a larger DNA molecule that is not found in association with the larger molecule in nature. Thus, when the heterologous region encodes a mammalian gene, the gene will

usually be flanked by DNA that does not flank the mammalian genomic DNA in the genome of the source organism. In another example, the coding sequence is a construct where the coding sequence itself is not found in nature (e.g., a cDNA where the genomic coding sequence contains introns, or synthetic sequences having codons different than the native gene). Allelic variations or naturally-occurring mutational events do not give rise to a heterologous region of DNA as defined herein. For example, a polynucleotide, may be placed by genetic engineering techniques into a plasmid or vector derived from a different source, and is a heterologous polynucleotide. A promoter removed from its native coding sequence and operatively linked to a coding sequence other than the native sequence is a heterologous promoter.

As used herein, "fragment" or "portion" as applied to a gene or a polypeptide, will ordinarily be at least 10 residues, more typically at least 20 residues, and preferably at least 30 (e.g., 50) residues in length, but less than the entire, intact sequence. Fragments of these genes can be generated by methods known to those skilled in the art, e.g., by restriction digestion of naturally occurring or recombinant fiber or fibrin genes, by recombinant DNA techniques using a vector that encodes a defined fragment of the fiber or fibrin gene, or by chemical synthesis.

Methods for making and/or administering a vector or recombinants or plasmid for expression of gene products of genes of the invention either *in vivo* or *in vitro* can be any desired method, e.g., a method which is by or analogous to the methods disclosed in, or disclosed in documents cited in: U.S. Patent Nos. 4,603,112; 4,769,330; 4,394,448; 4,722,848; 4,745,051; 4,769,331; 4,945,050; 5,494,807; 5,514,375; 5,744,140; 5,744,141; 5,756,103; 5,762,938; 5,766,599; 5,990,091; 5,174,993; 5,505,941; 5,338,683; 5,494,807; 5,591,639; 5,589,466; 5,677,178; 5,591,439; 5,552,143; 5,580,859; 6,130,066; 6,004,777; 6,130,066; 6,497,883; 6,464,984; 6,451,770; 6,391,314; 6,387,376; 6,376,473; 6,368,603; 6,348,196; 6,306,400; 6,228,846; 6,221,362; 6,217,883; 6,207,166; 6,207,165; 6,159,477; 6,153,199; 6,090,393; 6,074,649; 6,045,803; 6,033,670; 6,485,729; 6,103,526; 6,224,882; 6,312,682; 6,348,450 and 6; 312,683; U.S. patent application Serial No. 920,197, filed October 16, 1986; WO 90/01543; W091/11525; WO 94/16716; WO 96/39491; WO

98/33510; EP 265785; EP 0 370 573; Andreansky et al., Proc. Natl. Acad. Sci. USA 1996;93:11313-11318; Ballay et al., EMBO J. 1993;4:3861-65; Felgner et al., J. Biol. Chem. 1994;269:2550-2561; Frolov et al., Proc. Natl. Acad. Sci. USA 1996;93:11371-11377; Graham, Tibtech 1990;8:85-87; Grunhaus et al., Sem. Virol. 1992;3:237-52; Ju et al., Diabetologia 1998;41:736-739; Kitson et al., J. Virol. 1991;65:3068-3075; McClements et al., Proc. Natl. Acad. Sci. USA 1996;93:11414-11420; Moss, Proc. Natl. Acad. Sci. USA 1996;93:11341-11348; Paoletti, Proc. Natl. Acad. Sci. USA 1996;93:11349-11353; Pennock et al., Mol. Cell. Biol. 1984;4:399-406; Richardson (Ed), Methods in Molecular Biology 1995;39, "Baculovirus Expression Protocols," Humana Press Inc.; Smith et al. (1983) Mol. Cell. Biol. 1983;3:2156-2165; Robertson et al., Proc. Natl. Acad. Sci. USA 1996;93:11334-11340; Robinson et al. Sem. Immunol. 1997;9:271; and Roizman, Proc. Natl. Acad. Sci. USA 1996;93:11307-11312.

The invention also provides for transformed host cells comprising a vector of the invention. In one embodiment, the vector is introduced into the cell by transfection, electroporation or infection. The invention also provides for a method for preparing a transformed cell expressing an immunogen of the present invention comprising transfecting, electroporating or infecting a cell with an expression vector (e.g., a DNA vaccine) to produce an infected producing cell and maintaining the host cell under biological conditions sufficient for expression of the immunogen in the host cell.

According to another embodiment of the invention, the expression vectors are expression vectors used for the in vitro expression of proteins in an appropriate cell system. The expressed proteins can be harvested in or from the culture supernatant after, or not after secretion (if there is no secretion a cell lysis typically occurs or is performed), optionally concentrated by concentration methods such as ultrafiltration and/or purified by purification means, such as affinity, ion exchange or gel filtration-type chromatography methods.

It is understood to one of skill in the art that conditions for culturing a host cell varies according to the particular gene and that routine experimentation is necessary at times to determine the optimal conditions for culturing the vector depending on the host

cell. A "host cell" denotes a prokaryotic or eukaryotic cell that has been genetically altered, or is capable of being genetically altered by administration of an exogenous polynucleotide, such as a recombinant plasmid or vector. When referring to genetically altered cells, the term refers both to the originally altered cell and to the progeny thereof.

Polynucleotides comprising a desired sequence can be inserted into a suitable cloning or expression vector, and the vector in turn can be introduced into a suitable host cell for replication and amplification. Polynucleotides can be introduced into host cells by any means known in the art. The vectors containing the polynucleotides of interest can be introduced into the host cell by any of a number of appropriate means as described above, including direct uptake, endocytosis, transfection, f-mating, electroporation, transfection employing calcium chloride, rubidium chloride, calcium phosphate, DEAE-dextran, or other substances; microprojectile bombardment; lipofection; and infection (where the vector is infectious, for instance, a retroviral vector). The choice of introducing vectors or polynucleotides will often depend on features of the host cell.

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

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 immunogens of the invention can be expressed.

For example, when the aim is to express the immunogens of the invention *in vitro*, or in cultured cells, or in any prokaryotic or eukaryotic system for the purpose of producing the protein(s) encoded by that immunogen, 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 immunogens under the identified circumstances.

When the aim is to express the immunogens of the invention *in vivo* in a subject, for example in order to generate an immune response against an HIV antigen and/or protective or therapeutic immunity against HIV, 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 immunogens of the invention in a laboratory animal, such as for pre-clinical testing of HIV immunogenic compositions and vaccines of the invention. In other embodiments, it will be desirable to express the immunogens 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 be attenuated to prevent 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.

Any vector suitable for administration as a vaccine may be employed in the instant invention. In certain embodiments of the instant invention, vectors suitable for use as DNA vaccines are used, such as pVAX and pcDNA vectors (Invitrogen).

In other 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 (e.g., adenovirus subtypes Ad5, Ad11, Ad26, Ad35, Ad48 and Ad49), adeno-associated viruses (AAV), alphaviruses, retroviruses and poxviruses, including avipox viruses, attenuated poxviruses, and vaccinia viruses, such as the modified vaccinia Ankara virus (MVA). In certain embodiments, a vaccine of the invention comprises an adenovirus selected from Ad5, Ad11, Ad26, Ad35, Ad48 and

Ad49. 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.

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 certain 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 immunogens of the invention in a laboratory animal, such as for pre-clinical testing of HIV immunogenic compositions and vaccines of the invention. In other embodiments, it will be desirable to express the immunogens 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 certain embodiments the subject is a human, for example a human that is infected with, or is at risk of infection with, an HIV.

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 and may be used as one or more components of a prophylactic or therapeutic vaccine against HIV for the prevention, amelioration or treatment of HIV. The nucleic acids and vectors of the invention are useful for providing genetic vaccines, i.e., vaccines for delivering the nucleic acids

encoding the antigens of the invention to a subject, such as a human, such that the antigens are then expressed in the subject to elicit an immune response.

Immunogenic Compositions

The term "immunogenic protein or peptide" as used herein also includes peptides and polypeptides that are immunologically active in the sense that once administered to the host, it is able to evoke an immune response of the humoral and/or cellular type directed against the protein. Preferably the protein fragment is such that it has substantially the same immunological activity as the total protein. Thus, a protein fragment according to the invention comprises at least one epitope or antigenic determinant. The term epitope relates to a protein site able to induce an immune reaction of the humoral type (B cells) and/or cellular type (T cells).

The term "immunogenic protein or peptide" further contemplates deletions, additions and substitutions to the sequence, so long as the polypeptide functions to produce an immunological response as defined herein.

The term "epitope" refers to the site on an antigen or hapten to which specific B cells and/or T cells respond. The term is also used interchangeably with "antigenic determinant" or "antigenic determinant site". Antibodies that recognize the same epitope can be identified in a simple immunoassay showing the ability of one antibody to block the binding of another antibody to a target antigen.

An "immunological response" to a composition or vaccine is the development in the host of a cellular and/or antibody-mediated immune response to a composition or vaccine of interest. Usually, an "immunological response" includes but is not limited to one or more of the following effects: the production of antibodies, B cells, helper T cells, suppressor-T cells, and/or cytotoxic T cells and/or $\gamma\delta$ T cells, directed specifically to an antigen or antigens included in the composition or vaccine of interest. Preferably, the host will display either a therapeutic or protective immunological response such that resistance to new infection will be enhanced and/or the clinical severity of the disease reduced. Such protection will be demonstrated by either a reduction or lack of symptoms

normally displayed by an infected host, a quicker recovery time and/or a lowered viral titer in the infected host.

Generation of an immunological response may involve antigen presenting cells (APCs). APCs may be "professional" antigen presenting cells or may be another cell that may be induced to present antigen to T cells. APCs include dendritic cells (DCs) such as interdigitating DCs or follicular DCs, Langerhans cells, PBMCs, macrophages, B-lymphocytes, or other cell types such as epithelial cells, fibroblasts or endothelial cells, activated or engineered by transfection to express a MHC molecule (Class I or II) on their surfaces. APCs also include hybridomas, lymphomas, and synthetic APCs such as lipid membranes. Precursors of APCs include CD34+ cells, monocytes, fibroblasts and endothelial cells. Cytokine genes which may promote immune potentiation include IL-2, IL-12, IFN- γ , TNF- α , IL-18, etc. Such proteins include MHC molecules (Class I or Class II), CD80, CD86, or CD40. Examples of T cells include helper T cells (CD4+) and CD8+ cells.

The terms "immunogenic" protein or polypeptide as used herein also refers to an amino acid sequence which elicits an immunological response as described above. An "immunogenic" protein or polypeptide, as used herein, includes the full-length sequence of the protein, analogs thereof, or immunogenic fragments thereof. By "immunogenic fragment" is meant a fragment of a protein which includes one or more epitopes and thus elicits the immunological response described above. Such fragments can be identified using any number of epitope mapping techniques, well known in the art. See, e.g., Epitope Mapping Protocols in Methods in Molecular Biology, Vol. 66 (Glenn E. Morris, Ed., 1996) Humana Press, Totowa, N.J. For example, linear epitopes may be determined by e.g., concurrently synthesizing large numbers of peptides on solid supports, the peptides corresponding to portions of the protein molecule, and reacting the peptides with antibodies while the peptides are still attached to the supports. Such techniques are known in the art and described in, e.g., U.S. Pat. No. 4,708,871; Geysen et al. (1984) Proc. Natl. Acad. Sci. USA 81:3998-4002; Geysen et al. (1986) Molec. Immunol. 23:709-715, all incorporated herein by reference in their entireties. Similarly,

conformational epitopes are readily identified by determining spatial conformation of amino acids such as by, e.g., x-ray crystallography and 2-dimensional nuclear magnetic resonance. See, e.g., *Epitope Mapping Protocols*, *supra*.

Synthetic antigens are also included within the definition, for example, polypeptides, flanking epitopes, and other recombinant or synthetically derived antigens. See, e.g., Bergmann et al. (1993) *Eur. J. Immunol.* 23:2777-2781; Bergmann et al. (1996) *J. Immunol.* 157:3242-3249; Suhrbier, A. (1997) *Immunol. and Cell Biol.* 75:402-408; Gardner et al. (1998) 12th World AIDS Conference, Geneva, Switzerland, Jun. 28-Jul. 3, 1998. Immunogenic fragments, for purposes of the present invention, will usually include at least about 3 amino acids, at least about 5 amino acids, at least about 10-15 amino acids, or at least about 25 or more amino acids, of the molecule. There is no critical upper limit to the length of the fragment, which could comprise nearly the full-length of the protein sequence, or even a fusion protein comprising at least one epitope of the protein.

As mentioned earlier, epitope determination procedures, such as, generating overlapping peptide libraries (Hemmer B. et al., *Immunology Today*, 1998, 19 (4), 163-168), Pepscan (Geysen et al., (1984) *Proc. Nat. Acad. Sci. USA*, 81, 3998-4002; Geysen et al., (1985) *Proc. Nat. Acad. Sci. USA*, 82, 178-182; Van der Zee R. et al., (1989) *Eur. J. Immunol.*, 19, 43-47; Geysen H.M., (1990) *Southeast Asian J. Trop. Med. Public Health*, 21, 523-533; Multipin.RTM. Peptide Synthesis Kits de Chiron) and algorithms (De Groot A. et al., (1999) *Nature Biotechnology*, 17, 533-561), and in PCT Application Serial No. PCT/US2004/022605 all of which are incorporated herein by reference in their entirety, can be used in the practice of the invention, without undue experimentation. Other documents cited and incorporated herein may also be consulted for methods for determining epitopes of an immunogen or antigen and thus nucleic acid molecules that encode such epitopes.

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.

The term "antibody" includes intact molecules as well as fragments thereof, such as Fab, F(ab')₂, Fv and scFv which are capable of binding the epitope determinant. These antibody fragments retain some ability to selectively bind with its antigen or receptor and include, for example:

(i) Fab, the fragment which contains a monovalent antigen-binding fragment of an antibody molecule can be produced by digestion of whole antibody with the enzyme papain to yield an intact light chain and a portion of one heavy chain;

(ii) Fab', the fragment of an antibody molecule can be obtained by treating whole antibody with pepsin, followed by reduction, to yield an intact light chain and a portion of the heavy chain; two Fab' fragments are obtained per antibody molecule;

(iii) F(ab')₂, the fragment of the antibody that can be obtained by treating whole antibody with the enzyme pepsin without subsequent reduction; F(ab')₂ is a dimer of two Fab' fragments held together by two disulfide bonds;

(iv) scFv, including a genetically engineered fragment containing the variable region of a heavy and a light chain as a fused single chain molecule.

General methods of making these fragments are known in the art. (See for example, Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, New York (1988), which is incorporated herein by reference).

A "neutralizing antibody" is one that can neutralize the ability of that pathogen to initiate and/or perpetuate an infection in a host and/or in target cells *in vitro*. A neutralizing antibody may inhibit the entry of HIV-1 virus with a neutralization index >1.5 or >2.0. Broad and potent neutralizing antibodies may neutralize greater than about 50% of HIV-1 viruses (from diverse clades and different strains within a clade) in a neutralization assay. The inhibitory concentration of the monoclonal antibody may be less than about 25 mg/ml to neutralize about 50% of the input virus in the neutralization assay.

An “isolated antibody” or “non-naturally occurring antibody” is one that has been separated and/or recovered from a component of its natural environment. Contaminant components of its natural environment are materials that would interfere with diagnostic or therapeutic uses for the antibody, and may include enzymes, hormones, and other proteinaceous or nonproteinaceous solutes. In preferred embodiments, the antibody is purified: (1) to greater than 95% by weight of antibody as determined by the Lowry method, and most preferably more than 99% by weight; (2) to a degree sufficient to obtain at least 15 residues of N-terminal or internal amino acid sequence by use of a spinning cup sequenator; or (3) to homogeneity by SDS-PAGE under reducing or non-reducing conditions using Coomassie blue or, preferably, silver stain. Isolated antibody includes the antibody *in situ* within recombinant cells since at least one component of the antibody's natural environment will not be present. Ordinarily, however, isolated antibody will be prepared by at least one purification step.

The term “monoclonal antibody” as used herein refers to an antibody obtained from a population of substantially homogeneous antibodies, *i.e.*, the individual antibodies comprising the population are identical except for possible naturally occurring mutations that may be present in minor amounts. Monoclonal antibodies are highly specific, being directed against a single antigenic site. Furthermore, in contrast to polyclonal antibody preparations that include different antibodies directed against different determinants (epitopes), each monoclonal antibody is directed against a single determinant on the antigen. In addition to their specificity, the monoclonal antibodies are advantageous in that they may be synthesized uncontaminated by other antibodies. The modifier “monoclonal” is not to be construed as requiring production of the antibody by any particular method. For example, the monoclonal antibodies useful in the present invention may be prepared by the hybridoma methodology first described by Kohler et al., *Nature*, 256:495 (1975), or may be made using recombinant DNA methods in bacterial, eukaryotic animal or plant cells (see, *e.g.*, U.S. Pat. No. 4,816,567). The “monoclonal antibodies” may also be isolated from phage antibody libraries using the

techniques described in Clackson *et al.*, Nature, 352:624-628 (1991) and Marks *et al.*, J. Mol. Biol., 222:581-597 (1991), for example.

An “antibody fragment” comprises a portion of an intact antibody, preferably the antigen binding or variable region of the intact antibody. Examples of antibody fragments include Fab, Fab', F(ab')₂, and Fv fragments; diabodies; linear antibodies (see U.S. Pat. No. 5,641,870; Zapata *et al.*, Protein Eng. 8(10): 1057-1062 [1995]); single-chain antibody molecules; and multispecific antibodies formed from antibody fragments.

It should be understood that the proteins, including the antibodies 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, cysteine, serine threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified as aromatic amino acids. It is reasonably predictable that an isolated or non-naturally occurring 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.

According to the invention, in certain embodiments, administration of a vaccine of the invention can be combined with other vaccinations within the framework of vaccination programs, in the form of immunization or vaccination kits or methods, or in the form of multivalent immunogenic compositions and multivalent vaccines, e.g.,

comprising at least one vaccine component against a target pathogenic agent, such as HIV, and at least one vaccine component against at least one other pathogenic agent. This also includes the expression by the same expression vector of genes of at least two pathogenic agents.

The invention thus also relates to a multivalent or "cocktail" immunogenic composition or a multivalent or "cocktail" vaccine against a target pathogenic agent, such as HIV, and against at least one other pathogen of the target species, using the same *in vivo* expression vector containing and expressing at least one polynucleotide of the target pathogenic agent, such as HIV, according to the invention and at least one polynucleotide expressing an immunogen of another pathogen

As discussed herein, these multivalent compositions or vaccines can also comprise a pharmaceutically acceptable carrier or vehicle or excipient, and optionally an adjuvant.

The immunogenic compositions or vaccines as discussed herein can also be combined with at least one conventional vaccine (e.g., inactivated, live attenuated, or subunit) directed against the same pathogen or at least one other pathogen of the species to which the composition or vaccine is directed. The immunogenic compositions or vaccines discussed herein can be administered prior to or after the conventional vaccine, e.g., in a "prime-boost" regimen.

Formulations

The compositions of the invention can include any pharmaceutically acceptable carrier known in the art.

To facilitate the administration of a vaccine of the invention, the vaccine can be formulated into suitable pharmaceutical compositions. Generally, such compositions include the active ingredient (e.g., a DNA vaccine) and a pharmacologically acceptable carrier. Such compositions can be suitable for delivery of the active ingredient to a patient for medical application, and can be manufactured in a manner that is itself known, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes.

Pharmaceutical compositions for use in accordance with the present invention can be formulated in a conventional manner using one or more pharmacologically or physiologically acceptable carriers comprising excipients, as well as optional auxiliaries, which facilitate processing of the active compounds into preparations, which can be used pharmaceutically. Proper formulation is dependent upon the route of administration chosen. Thus, for injection, the active ingredient can be formulated in aqueous solutions, preferably in physiologically compatible buffers. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art. For oral administration, the active ingredient can be combined with carriers suitable for inclusion into tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like. For administration by inhalation, the active ingredient is conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant. The active ingredient can be formulated for parenteral administration by injection, e.g., by bolus injection or continuous infusion. Such compositions can take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and can contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Other pharmacological excipients are known in the art.

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 of the invention, e.g., a DNA vaccine, 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/caprate), glyceryl tri(caprylate/caprate) 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 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 Publishing Company, (ed.) 1980).

The term “adjuvant” encompasses vaccine adjuvants. A vaccine adjuvant is a component that potentiates the immune responses to an antigen and/or modulates it towards the desired immune responses. *See* The European Medicines Agency (EMA) Evaluation of Medicines for Human Use, Guideline on Adjuvants in Vaccines, (2005), page 6. Examples of suitable adjuvants include mineral salts, such as aluminum hydroxide and aluminum or calcium phosphate gels; oil emulsions and surfactant based formulations, e.g., MF59 (microfluidized detergent stabilized oil-in-water emulsion), QS21 (purified saponin), AS02 [SBAS2] (oil-in-water + MPL + QS-21), Montanide ISA-51 and ISA-720 (stabilized water-in-oil emulsion); particulate adjuvants, e.g., virosomes (unilamellar liposomal vehicles incorporating influenza hemagglutinin), AS04 ([SBAS4] Al salt with MPL), ISCOMS (structured complex of saponins and lipids), polylactide co-glycolide (PLG); microbial derivatives (natural and synthetic), e.g., monophosphoryl lipid A (MPL), Detox (MPL + *M. Phlei* cell wall skeleton), AGP [RC-529] (synthetic acylated monosaccharide), DC_Chol (lipoidal immunostimulators able to self-organize into liposomes), OM-174 (lipid A derivative), CpG motifs (synthetic oligonucleotides containing immunostimulatory CpG motifs), modified LT and CT (genetically modified bacterial toxins to provide non-toxic adjuvant effects); endogenous human immunomodulators, e.g., hGM-CSF or hIL-12 (cytokines that can be administered either as protein or plasmid encoded), Immudaptin (C3d tandem array); and inert vehicles, such as gold particles. *Id.*

Adjuvants that enhance the effectiveness of the vaccine may also be added to the formulation. Further to the above, 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.* 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 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 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 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, 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 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 oil in water emulsion, which is especially appropriate for viral vectors, can be based on: light liquid paraffin oil (European pharmacopoeia type), isoprenoid oil such as

squalane, squalene, oil resulting from the oligomerization of alkenes, e.g. isobutene or decene, esters of acids or alcohols having a straight-chain alkyl group, such as vegetable oils, ethyl oleate, propylene glycol, di(caprylate/caprates), glycerol tri(caprylate/caprates) and propylene glycol dioleate, or esters of branched, fatty alcohols or acids, especially isostearic acid esters. The oil is used in combination with emulsifiers to form an emulsion. The emulsifiers may be nonionic surfactants, such as: esters of on the one hand sorbitan, mannide (e.g., anhydromannitol oleate), glycerol, polyglycerol or propylene glycol and on the other hand oleic, isostearic, ricinoleic or hydroxystearic acids, said esters being optionally ethoxylated, polyoxypropylene-polyoxyethylene copolymer blocks, such as Pluronic, e.g., L121.

For maleic anhydride-alkenyl derivative copolymers, EMA (Monsanto) may be used, which are straight-chain or crosslinked ethylene-maleic anhydride copolymers and they are, for example, crosslinked by divinyl ether. Reference is also made to J. Fields et al., *Nature* 186: 778-780, Jun. 4, 1960. With regard to structure, the acrylic or methacrylic acid polymers and EMA are preferably formed by basic units having the following formula in which: R1 and R2, which can be the same or different, represent H or CH₃, x=0 or 1, preferably x=1, y=1 or 2, with x+y=2. For EMA, x=0 and y=2 and for carbomers x=y=1. These polymers are soluble in water or physiological salt solution (20 g/l NaCl) and the pH can be adjusted to 7.3 to 7.4, e.g., by soda (NaOH), to provide the adjuvant solution in which the expression vector(s) can be incorporated.

A further instance of an adjuvant is a compound chosen from the polymers of acrylic or methacrylic acid and the copolymers of maleic anhydride and alkenyl derivative. Advantageous adjuvant compounds are the polymers of acrylic or methacrylic acid which are cross-linked, especially with polyalkenyl ethers of sugars or polyalcohols. These compounds are known by the term carbomer (*Pharmeuropa* Vol. 8, No. 2, June 1996). Persons skilled in the art can also refer to U.S. Patent No. 2,909,462 (incorporated herein by reference) which describes such acrylic polymers cross-linked with a polyhydroxylated compound having at least 3 hydroxyl groups, preferably not more than 8, the hydrogen atoms of at least three hydroxyls being replaced by

unsaturated aliphatic radicals having at least 2 carbon atoms. The preferred radicals are those containing from 2 to 4 carbon atoms, e.g. vinyls, allyls and other ethylenically unsaturated groups. The unsaturated radicals may themselves contain other substituents, such as methyl. The products sold under the name CARBOPOL® (BF Goodrich, Ohio, USA) are particularly appropriate. They are cross-linked with an allyl sucrose or with allyl pentaerythritol. Among them, there may be mentioned CARBOPOL® 974P, 934P and 971P. Among the copolymers of maleic anhydride and alkenyl derivative, the copolymers EMA® (Monsanto) which are copolymers of maleic anhydride and ethylene, linear or cross-linked, for example cross-linked with divinyl ether, are preferred. Reference may be made to J. Fields et al., *Nature*, 186 : 778-780, 4 June 1960, incorporated herein by reference.

The term "liposome" as used herein encompasses smectic mesophases, which may comprise either phospholipid or nonphospholipid smectic mesophases. See, for example, "smectic mesophase" in Small, D.M., in "The Physical Chemistry of Lipids, From Alkanes to Phospholipids" Handbook of Lipid Research, Vol, 4, Plenum, NY, 1986, pp. 49-50, which states that "[w]hen a given molecule is heated, instead of melting directly into an isotropic liquid, it may instead pass through intermediate states called mesophases or liquid crystals, characterized by residual order in some directions but by lack of order in others....In general, the molecules of liquid crystals are somewhat longer than they are wide and have a polar or aromatic part somewhere along the length of the molecule. The molecular shape and the polar-polar, or aromatic, interaction permit the molecules to align in partially ordered arrays....These structures characteristically occur in molecules that possess a polar group at one end. Liquid crystals with long-range order in the direction of the long axis of the molecule are called smectic, layered, or lamellar liquid crystals....In the smectic states the molecules may be in single or double layers, normal or tilted to the plane of the layer, and with frozen or melted aliphatic chains." See also Figs. 3-4 of Small.

Advantageously, the immunogenic compositions and vaccines according to the invention comprise an effective quantity to elicit an immunological response and/or a

protective immunological response of one or more expression vectors and/or polypeptides as discussed herein; and, an effective quantity can be determined from this disclosure, including the documents incorporated herein, and the knowledge in the art, without undue experimentation. The immunogenic compositions can be designed to introduce the 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 formulation can be prepared using appropriate macromolecules (for 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, 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) 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.

Administration

Suitable dosages of the antigens, nucleic acids and expression vectors of the invention (collectively, the immunogens) in an 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 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.

Assays for assessing the cellular response to HIV vaccines of the instant invention include intracellular staining (e.g., flow cytometry) and ELISPOT (an enzyme-linked immunosorbent assay format), which allow detecting and counting cells producing cytokines (e.g., TNF α and IFN- γ) in response to antigens. For example, isolation of splenocytes or peripheral blood monocyte cells (PBMCs) from animals or human patients followed by in vitro challenge with an HIV epitope such as 2F5 or 4E10, and finally testing by ELISPOT and/or intracellular cytokine staining (ICS), can determine the potential for a cell-mediated immune response in vaccine recipients. Flow cytometry using tetramers (e.g., molecules consisting of four copies of a given class I molecule bound to their cognate peptide and alkaline phosphatase) allows the enumeration of antigen-specific T cells (e.g., detection of T cells that recognize specific peptides bound to major histocompatibility complex (MHC) class I molecules). A standard chromium release assay can be used to assess cytotoxicity. To assess a cell-mediated immune response to a DNA vaccine, the more traditional approaches of measuring T cell proliferation in response to antigen and CTL-mediated killing of autologous cells expressing HIV epitopes can also be used.

ELISA assays and Western blots can be used to assess humoral immune responses. In particular, ELISA and Western blots can be used to assess antibody binding, antibody neutralizing capability, antibody-mediated fusion inhibition, and antibody-dependent cytotoxicity.

An MT-2 assay can be performed to measure neutralizing antibody responses. Antibody-mediated neutralization can be measured in an MT-2 cell-killing assay as described previously (Montefiori et al., 1988, *J. Clin. Microbiol.*, 26:231-237). The inhibition of the formation of syncytia by the sera shows the activity of neutralizing antibodies present within the sera, induced by vaccination. Briefly, vaccinated test and control sera can be exposed to virally infected cells (e.g., MT-2 T cell line). Neutralization can be measured by staining viable cells (e.g., with Finter's neutral red when cytopathic effects in control wells are about >70% but less than 100%). Percentage protection can be determined by calculating the difference in absorption (A_{540}) between test wells (cells+virus) and dividing this result by the difference in absorption between cell control wells (cells only) and virus control wells (virus only). Neutralizing titers are then expressed as the reciprocal of the plasma dilution required to protect at least 50% of cells from virus-induced killing.

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, 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 infection or to prevent or attenuate the progression of AIDS in a subject already infected with HIV. When provided therapeutically, the immunogenic compositions can serve to ameliorate and treat AIDS symptoms and are 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.

Suitable doses of nucleic acid compositions for humans can range from 1 $\mu\text{g/kg}$ to 1 mg/kg of total nucleic acid, e.g., from 5 $\mu\text{g/kg}$ -500 mg/kg of total DNA, 10 $\mu\text{g/kg}$ -250 $\mu\text{g/kg}$ of total DNA, or 10 $\mu\text{g/kg}$ -170 $\mu\text{g/kg}$ of total DNA. In one embodiment, a human subject (18-50 years of age, 45-75 kg) is administered 1.2 mg-7.2 mg of DNA. DNA vaccines can be administered multiple times, e.g., between two-six times, e.g., three

times. In a particular embodiment, 100 µg of a DNA composition is administered to a human subject at 0, 4, and 12 weeks (100 µg per administration).

An example of range for an immunogenic amount of protein composition is 5 µg/kg-500 µg/kg, e.g., 10-100 µg/kg of total protein, with adjuvant. In one embodiment, a dose of 325 µg of a protein composition is administered to a human (18-55 years of age, 45-75 kg).

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 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; 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). Additional methods of delivery of DNA to animal tissue include electroporation, jet injection, sonoporation, microneedle-assisted delivery, etc. Alternatively, delivery routes can be oral, intranasal or by any other suitable route. Delivery 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. In certain embodiments, 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 and up to 6 months or more. With DNA tattooing, the interval is typically only 3 days (e.g., 0, 3, and 6 days). 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. 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).

The immunogenic compositions of the invention can be administered alone, or can be co-administered, or sequentially administered, with other immunogens and/or 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. For example, in some embodiments, an HIV Env protein of the instant invention is administered in a viral vector, such as an MVA, which also comprises genes encoding one or more other HIV proteins, such as, e.g., gag and pol. Again, the ingredients and manner (e.g., 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.

In certain embodiments, the immunogenic compositions of the invention are administered to a mammal. In further embodiments, the mammal is a human, a non-human primate, a dog, a rabbit, a guinea pig, or a mouse.

Those of ordinary skill in the art can easily make a determination of the proper

dosage of a protein subunit and/or DNA vaccine. Generally, certain factors will impact the dosage that is administered; although the proper dosage is such that, in one context, in embodiments where a DNA vaccine is administered, the exogenous gene is expressed and the gene product is produced in the particular cell of the mammal. Preferably, the dosage is sufficient to have a therapeutic and/or prophylactic effect on the animal.

Combination Therapies

The methods of treating subjects infected with HIV with the compositions of the instant invention can include combination therapies, in which other HIV treatments are administered. For example, a subject undergoing HIV Env protein subunit vaccination according to the instant invention can be administered anti-retroviral drugs individually, or in combination, for example, with various combinations of nucleoside reverse transcriptase inhibitors, non-nucleoside reverse transcriptase inhibitors, and HIV protease inhibitors.

Nucleoside reverse transcriptase inhibitors include, e.g., zidovudine (AZT); didanosine (ddI); zalcitabine (ddC); stavudine (d4T); lamivudine (3TC); abacavir (1592U89); adefovir dipivoxil [bis(POM)-PMEA]; lobucavir (BMS-180194); and lodenosine (FddA), 9-(2,3-dideoxy-2-fluoro-b-D-threo-pentofuranosyl)adenine.

Non-nucleoside reverse transcriptase inhibitors include nevirapine (BI-RG-587); delaviradine (BHAP, U-90152); and efavirenz (DMP-266).

Protease inhibitors include saquinavir (Ro 31-8959); ritonavir (ABT-538); indinavir (MK-639); nelfinavir (AG-1343) available under the VIRACEPT™ tradename from Agouron Pharmaceuticals, Inc.; amprenavir (141W94), a non-peptide protease inhibitor, tradename AGENERASE™; and lasinavir (BMS-234475).

Kits

The compositions of the instant invention and their methods of use are ideally suited for preparation of kits. HIV Env nucleic acid and/or protein may be provided in containers that can be in any form, e.g., lyophilized, or in solution (e.g., a distilled water or buffered solution), etc. In the kits of the invention, a set of instructions will typically be included.

The kits can include one or more other elements including: other reagents, e.g., a diluent, devices or other materials for preparing the composition for administration; pharmaceutically acceptable carriers; and devices or other materials for administration to a subject. Instructions for use can include instructions for therapeutic application including suggested dosages and/or modes of administration, e.g., in a human subject, as described herein.

The kit can further contain at least one additional reagent, such as a diagnostic or therapeutic agent, e.g., a diagnostic agent to monitor a response to immune response to the compositions in the subject, or an additional therapeutic agent as described herein.

In one embodiment, the kit includes a vial (or other suitable container) containing one or more recombinant HIV Env proteins of the instant invention. In certain embodiments, the kit further includes an adjuvant and an excipient. The adjuvant and the excipient are formulated with the protein, and can be included in the formulation or packaged separately within the kit.

The invention will now be further described by way of the following non-limiting examples.

EXAMPLE 1

Ba-L gp140 DC 4E10

The HIV-1 Ba-L gp160 gene, minus the signal peptide was codon optimized and synthesized at Geneart. The gp120/gp41 cleavage site was altered to prevent cleavage: Arg 501 and Arg 509 were changed to Serines. The gp140 DC 4E10 sequence was amplified by PCR and inserted into the pJWIREs expression plasmid in frame with the tPa signal. The amino terminus begins with E(30) and the carboxyl terminus ends at ...WLWYIK(681) (SEQ ID NO: 45) with an additional KKK added to help solubility. The pJWIREs includes the following:

Expression of the inserted gene is driven by the CMV promotor and Bovine Growth Hormone (BGH) Poly A. It has puromycin acetyl transferase gene linked to the inserted gene through IRES sequence for Puromycin resistance.

The resulting construct is referred to as pJW Ba-L gp140 DC 4E10 Puro (FIGURE 1).

Transfection studies were performed using HEK 293 cells using lipofectamine 2000. IP western blot of the conditioned media using 4E10 and 2F5 Human MPER antibodies (Polymun Scientific GmbH, Klosterneuburg, Austria) showed reactivity with the gp140 DC 4E10 as shown in Figure 2.

See also Figures 3-4 for the nucleic acid and protein sequences of the Clade B sequence, Ba-L gp140 DC 4E10.

Figures 5-6 depict the nucleic acid and protein sequences for a second modified Clade B sequence, Ba-L gp145.

EXAMPLE 2

Clade D gp140 Methods

Cell line development and molecular cloning

Chinese Hamster Ovary (CHO) cell lines stably expressing the extracellular domains of gp160 (gp140) for 3 HIV-1 clade D isolates were developed. The goal was to develop cell lines that secrete high levels of gp140 that can be purified and be used in HIV-1 vaccine development. The isolates chosen for gp140 expression are A07412 (parent sequence, GenBank Accession No. AF484477; see also GenBank Accession No. AY736828), 57128 (parent sequence, GenBank Accession No. AF484502; see also GenBank Accession No. AY736829) and 57140 (parent sequence, GenBank Accession No. AF484511). In order to maximize expression in this system, the gp140 codes were created synthetically at GENEART. Through this process, the genes were codon optimized for expression in human cells by designing the genes using codons that correspond to the most abundant tRNAs present in human cells. Human codon optimization is ideal for any DNA vaccine component to be used in humans, but is also

quite effective in yielding high levels of expression in CHO. In addition to codon optimization, the synthetic genes were also designed to eliminate various cis-acting elements that can reduce transcription/translation efficiency (such as splice sites, poly A sites, adenine-rich elements, the Rev Responsive Elements (RRE), and other mRNA secondary structures) as well as other motifs (such as GC-rich stretches, internal TATA boxes, Qui site) that may destabilize mRNA.

For each isolate, the gp140 genes were mutated at the primary and secondary gp120/gp41 cleavage sites using a PCR-based process. This was done to prevent gp120/gp41 cleavage, resulting in stable gp140 molecules upon secretion. In addition, the native signal peptide for each was removed so that the efficient Tissue Plasminogen Activator (tPA) signal in the expression vector can be used as the secretory signal. The gp140 codes each have a stop codon inserted just prior to the Transmembrane (TM) Domain to prevent the gp140 from being bound to the cell membrane upon secretion. The gp140 genes were ligated into the mammalian expression vector pJWTCDE-N at the NheI and EcoRI sites for stable expression in CHO cells.

A07412

The amino acid sequence of the amino terminus is SL(30)WVT.. (SEQ ID NO: 46), and the carboxyl terminus is ...FSITK(673)-Stop (SEQ ID NO: 47). The amino terminal serine is a foreign residue from the NheI cloning site at the end of the tPa signal. The gp120/gp41 cleavage site was altered from RAKRRVVEREKR(507) (SEQ ID NO: 48) to RAKSRVVEREKS (SEQ ID NO: 49). See also Figure 51, SEQ ID NO: 3.

57140

The amino acid sequence of the amino terminus is SL(33)WVT.. (SEQ ID NO: 46), and the carboxyl terminus is ...FSISN(673)-Stop (SEQ ID NO: 50). The amino terminal serine is a foreign residue from the NheI cloning site at the end of the tPa signal. The gp120/gp41 cleavage site was altered from RAKRRVVEREKR(507) (SEQ ID NO: 48) to RAKSRVVEREKS (SEQ ID NO: 49). See also Figure 51, SEQ ID NO: 4.

57128

The amino acid sequence of the amino terminus is SL(33)WVT.. (SEQ ID NO: 46), and the carboxyl terminus is ...FSITK(671)-Stop (SEQ ID NO: 47). The amino terminal serine is a foreign residue from the NheI cloning site at the end of the tPa signal. The gp120/gp41 cleavage site was altered from KARRRVVEREKR(507) (SEQ ID NO: 51) to KARSRVVEREKS (SEQ ID NO: 52). See also Figure 51, SEQ ID NO: 5.

The pJWTCDE-N contains the following elements for efficient expression of foreign genes in CHO cells (Figure 7):

- (1) Transcription of the gp140 genes are driven by the CMV promoter/intron A and Bovine Growth Hormone (BGH) Poly A
- (2) Neomycin Phosphotransferase II (NPT II) gene, driven by the SV40 promoter and a synthetic poly A for the selection of stably transfected cells under G418 Sulfate selection.
- (3) Dihydrofolate Reductase (DHFR) gene, driven by a partially crippled SV40 promoter and SV40 poly A for fairly weak expression of DHFR in transfected Cells. This facilitates selection in nucleoside-free media as well as inducing gene amplification of the foreign DNA through treatment with the DHFR inhibitor Methotrexate (MTX). This gene amplification can greatly increase the level gp140 production along with the increase in DHFR production needed to sustain life in MTX-containing media.

After cloning the expression vectors for each, sequencing of the inserted gp140 genes are performed to ensure proper construction.

To establish stable CHO cell lines secreting gp140, CHO cells deficient in DHFR

(CHO-dhfr-) were transfected using Lipofectamine 2000 (Gibco). Transfected cells were analyzed in a gp120 antigen capture assay and in radioimmunoprecipitation (RIP) with HIV-1 (+) human serum to detect the presence and quality of gp140 production. Transfected cells were plated into 96-well plates for selection in alpha MEM with 10% dialyzed Fetal Bovine Serum and 550 µg/ml G418 Sulfate. Surviving cells

were screened with a gp120 antigen capture assay, and fairly high gp140 producers were selected for expansion and single-cell cloning.

The gp120 antigen capture assay is an ELISA-based assay for the detection and quantification of gp120 protein. Molecules, such as gp140 and the later described gp140 DC 4E10 and gp145 proteins, which contain the gp120 sequence, can also be detected with this assay. The microtiter wells in a 96-well plate are coated with two murine monoclonal antibodies that react with unique epitopes on HIV-1 gp120. When gp120 standard solutions or tissue culture test samples are added to the wells, an immune complex forms with the plate-bound antibodies and the gp120 in solution. Unbound materials are then thoroughly washed away. The conjugate solution, containing peroxidase-conjugated human anti-gp120 polyclonal antibodies is then added. The conjugated antibodies complex with other epitopes on the captured gp120. After washing away the unbound conjugate solution, the peroxidase substrate is added. The enzyme-substrate reaction results in the substrate's blue color change. Upon adding the stop solution (2 N sulfuric acid), the blue changes to yellow, which can be quantitatively measured by reading the absorbance at 450nm. The amount of gp120 in the gp120 standards and test samples is relative to the absorbance. The concentration of gp120 in a test sample can be calculated based on the standard curve.

True clones were compared to find a few strong producers using a gp120 antigen capture assay and RIP. The best were treated with 0.02 μ M MTX to facilitate gene amplification of the foreign DNA. Once cells were able to grow at normal rates in MTX, they were cloned again to find higher producers. Cell lines were cloned by limiting dilution, analyzed for optimum expression and adapted for growth in the protein-free media HyQPFCHO Liquid Soy (HyClone; Logan, Utah).

Protein Purification

Conditioned media from CHO cultures were harvested by centrifugation and concentrated with tangential flow 100 kDa molecular weight cutoff filters to about 2L. Media was buffered with phosphate buffer, and pH was adjusted to 7.2. Sodium chloride

concentration was adjusted to 300 mM, and media was filtered with 0.22 micron filter. Media was passed through a GNL agarose (Vector Laboratories; Burlingame, CA) column, and Env proteins were eluted in PBS containing 500 mM methyl α -D-mannopoyranoside. Media was passed on the GNL agarose column 3-4 additional times to remove all of the Env protein from the conditioned media. Additional procedures were performed to further purify the Env proteins. The sodium chloride concentration of the GNL agarose eluates were adjusted to 212 mM and passed through a column of Q-sepharose (Amersham Biosciences; Piscataway, NJ). The high molecular weight impurities bind Q-sepharose, but Env does not under these conditions. To disrupt any abnormal multimers formed through air oxidation, the Q sepharose treated Env proteins were concentrated to about 3 ml and treated with 50 mM DTT for 15 hours at 4°C, followed by 1 hour at 21°C. DTT treated preparations were then run on a Superdex 200 26/60 (Amersham Biosciences) gel filtration column to remove additional high and low molecular weight impurities, as well as to reduce the amount of Env breakdown products. The column was run at 0.5 ml/min. in PBS containing 1 mM DTT. Fractions containing the purest Env proteins, as analyzed on SDS-PAGE, were pooled. Proteins were then buffer exchanged on 10 ml PD-10 columns (Amersham Biosciences) equilibrated with PBS. Finally, proteins were filtered with 0.22 μ m filters, aliquoted and stored at -70°C.

EXAMPLE 3

Introduction

Plasmid DNA constructs expressing the Env proteins of four subtype C isolates isolated from patients at the acute and early seroconversion stages of infection were developed and tested to downselect the best candidate for gp145 expression. Isolate C06980v0c22 was selected, stable cell lines expressing C06980v0c22 gp145 were developed, and research cell banks were produced. Purified gp145 protein was produced and supplied for study in preclinical immunogenicity studies.

As described above, Applicants collaborated on the development of a subtype D HIV-1 subunit vaccine. Sequences for 4 subtype D HIV-1 isolates were provided and

several gp140 and gp120-secreting CHO cell lines were prepared. Cell lines were adapted to serum-free media and the Env proteins purified for preclinical immunogenicity studies. Small animals were immunized and gp140 and gp120-specific serum antibody binding titers evaluated by ELISA and neutralizing antibody titers against the homologous primary isolate evaluated using the pseudotype assay. While all gp120 and gp140 proteins were immunogenic, none elicited detectable neutralizing antibody against the homologous pseudotyped isolate.

HIV Env subunit vaccine efforts were pursued using subtype C Env sequences with the goal of eliciting more potent and broadly neutralizing antibody responses. As discussed above, a DNA construct had been developed encoding modifications of the HIV-1 Ba-L (Subtype B) Env. This construct coded for a truncated gp160 molecule referred to as gp145. This gp145 protein includes a modified tissue plasminogen activator (t-Pa) signal peptide upstream of a cleavage deficient gp160 that is truncated at the end of the membrane proximal external region (MPER). At the C terminus, three additional lysine residues were included, theoretically to increase the hydrophilicity of the C tail in order to present potentially neutralizing MPER epitopes to the immune system. Unlike the previous gp140 molecules discussed above, this gp145 molecule reacted to the neutralizing anti-MPER huMAb 4E10 in ELISA and western blot.

Subtype C is known to be the most common international subtype, and since preliminary data suggested subtype C infections may induce the most broadly cross-reactive HIV-1 neutralizing response in natural infection, subtype C sequences were investigated. A gp145 construct was proposed to be created using a primary CCR5-dependant subtype C Env sequence. Stable CHO cell lines expressing this Env protein were developed.

Envelope Downselection

The env sequences from 4 subtype C strains were provided for codon optimization and synthesis. Transient expression studies were performed to select which isolate would be used for further gp145 development.

An electronic copy was provided of four South African subtype C R5 HIV-1 envelope sequences from three acute (C06838v1c48, C06980v0c22 and C3728v2c6) and one early seroconverted (C06980v1c3) HIV-1 infections. Two of the sequences are from the same individual, one during acute infection (C06980v0c22) and the other after seroconversion (C06980v1c3).

In order to maximize expression in Chinese Hamster Ovary (CHO) cells, the env genes were synthesized incorporating *C. griseus* (Chinese Hamster) codon bias by Geneart AG (Regensburg, Germany). To further optimize expression, cis-acting motifs that can reduce translational efficiency were eliminated (e.g., internal TATA boxes, chi-sites, ribosomal entry sites, RNA secondary structure, repeat sequences, etc.). Two versions of each env gene were synthesized: a) gp160, full-length gp160 minus the native signal peptide and b) gp160 DC, full-length gp160 minus the native signal peptide and with mutations in the gp120/gp41 primary and secondary cleavage sites to prevent protease cleavage.

The translations of the four gp160 (WT) and gp160 DC (cleavage mutant) genes are compared (Figure 35). Molecules are shown as the sequences following the t-Pa signal peptide cleavage in the expressed proteins. The shaded regions are areas of variability. The boxed region highlights the gp120/gp41 cleavage sites; arginine to serine mutations in the gp160DC genes prevent the proteolytic cleavage. The gp160 genes were cloned into pSWTIPK3, a proprietary mammalian expression plasmid (Advanced BioSciences Laboratories, Inc.), at the NheI and EcoRI sites, in frame with the tPa signal peptide (Figure 8). The native leader sequences are replaced by the tPa signal peptide, which provides a more efficient secretion signal, enhancing gp160 production and transport to the cell membrane. The expression plasmids contain the Cytomegalovirus (CMV) promoter to control expression. The plasmids were expanded and purified from transformed *Escherichia coli*, and the gp160 coding regions were sequenced to confirm sequence identity.

Chinese hamster ovary cells (CHO-K1) and Human embryonic kidney cells (HEK293; clone 293H) cells were transfected and analyzed for gp160 production in

western blot and antigen capture ELISA. Based on production quality and quantity of gp160 molecules, a decision was made which isolate will be used to develop CHO cell lines producing the gp145 protein. Cells were transfected with the four gp160 and gp160DC plasmid constructs using lipofection (Lipofectamine 2000; Invitrogen, Carlsbad, CA). Cultures transfected with HIV-1Ba-L gp145, HIV-1Ba-L gp160 or HIV-1 subtype C gp140 expression constructs served as positive controls. Naïve CHO-K1 and HEK293 served as negative controls. Media and cell lysates were harvested 48 hours post-transfection for analysis.

Media and cell lysate samples from CHO-K1 transfections were evaluated for gp41/gp120/gp160 content via IP western blot using an HIV-1 positive human serum for immunoprecipitation and an HIV-1Ba-L gp160 immunized rabbit's serum for detection. No Env expression was detected in the media for any construct except the Ba-L gp145 control (data not shown). If gp160 is processed into gp41 and gp120, the gp120 could shed into the media; the amount of shed gp120 is below the assay detection limit. From the cell lysates, Env expression is evident with each of the subtype C gp160 and gp160DC constructs (Figure 19). Each construct produces gp160; each isolate runs at a different size, likely due to different glycosylation patterns. Only pSWC06980v0c22 gp160 shows gp160 processing into gp120 and gp41.

Because expression levels in CHO-K1 were quite low, transfection and analysis using HEK293 cells was performed to further evaluate each construct. Cell lysate samples from HEK293 transfections were evaluated for gp41/gp160 content via IP western blot using huMAb to gp41 (4E10) for immunoprecipitation and HIV-1Ba-L gp160 immunized rabbit's serum for detection. Env expression is strongly evident with each of the subtype C gp160 and gp160DC constructs (Figure 20). Each construct produces gp160; each isolate runs at a different size, likely due to different glycosylation patterns. Again, only pSWC06980v0c22 gp160 shows gp160 processing into gp120 and gp41.

To quantify gp120/gp160 production levels, media and cell lysates from transfected CHO-K1 and HEK293 were analyzed in an HIV-1 gp120 antigen capture

assay (Table 1). In the CHO-K1 transfections, Env expression was detected with each construct except C06838v1c48 gp160DC. Expression levels were quite low overall, compared to the Ba-L gp145 control. This was due to very low expression levels and difficulty in the detection of the Env proteins from these isolates. Level of the Ba-L gp145 expression was also quite low due to the low efficiency in transfecting CHO cells. The highest producing subtype C constructs were the C06980v0c22 and C3728v2c6 gp160s. It should be noted that concentration values are based on relative reactivities to a subtype C gp120 standard from a different isolate. Exact concentrations may differ than as reported due to possible differences in each isolates' affinities to the antibodies used in the assays.

Table 1: HIV-1 gp120 Antigen Capture Assay of CHO-K1 cells transfected with HIV-1 subtype C gp160 and gp160 DC expression plasmids. Quantities of gp120 and gp160 are detected at 48 hrs post-transfection in media and cell lysates.

Construct	gp120/gp160 (ng/ml)	
	Media	Cell Lysate
pSWC06838v1c48 gp160	0.6	0.0
pSWC06980v1c3 gp160	1.0	0.7
pSWC06980v0c22 gp160	3.2	5.7
pSWC3728v2c6 gp160	7.2	6.7
pSWC06838v1c48 gp160DC	0.0	0.0
pSWC06980v1c3 gp160DC	0.4	0.7
pSWC06980v0c22 gp160DC	1.6	1.2
pSWC3728v2c6 gp160DC	2.5	3.4
(-) Control	0.0	0.0
(+) Control: Ba-L gp145	25.0	3.0

Since detection of Env proteins was so weak in CHO-K1, analysis of HEK293 transfections was performed to verify results. In HEK293, Env expression was detected using antigen capture with each construct (Table 2). In HEK293, 06838v1c48 gp160DC was now detected, but very weakly. Expression levels were much higher than with the

CHO-K1. The highest producing subtype C constructs were again the C06980v0c22 and C3728v2c6 gp160s.

Table 2: HIV-1 gp120 Antigen Capture Assay of HEK293 cells transfected with HIV-1 subtype C gp160 and gp160 DC expression plasmids. Quantities of gp120 and gp160 are detected at 48 hrs post-transfection in media and cell lysates.

Construct	Cell Type	gp120/gp160 (ng/ml)	
		Media	Cell Lysate
pSWC06838v1c48 gp160	293H	0.0	0.9
pSWC06980v1c3 gp160	293H	28.2	11.6
pSWC06980v0c22 gp160	293H	94.1	94.6
pSWC3728v2c6 gp160	293H	198.2	87.7
pSWC06838v1c48 gp160DC	293H	0.0	0.8
pSWC06980v1c3 gp160DC	293H	2.7	8.8
pSWC06980v0c22 gp160DC	293H	13.9	37.7
pSWC3728v2c6 gp160DC	293H	59.3	92.1
(-) Control	293H	0.0	0.0
(+) Control: Ba-L gp145	293H	237.0	61.0
(+) Control: Ba-L gp160	293H	79.0	62.0
(-) Control	CHO-K1	0.0	0.0
(+) Control: Ba-L gp145	CHO-K1	41.0	3.0
(+) Control: Ba-L gp160	CHO-K1	21.0	29.0

Further evaluation of gp160 production was performed using two gp160 antigen capture assays utilizing the human monoclonal antibodies 4E10 (Table 3) and 2F5 (Table 4). The gp160 from each isolate reacted strongly with the 4E10-based assay. Even the C06838v1c48 isolate reacted strongly, indicating that the gp120 assay gives artificially low results for this isolate. The gp160 from each isolate reacted with the 2F5-based assay, although weaker than the 4E10-based assay. The C06980v1c22 isolate reacted the strongest. The weaker 2F5 reactivity is explained by the fact that the 2F5 epitope is quite different in the subtype C isolates from that in the subtype B isolates from which the antibody was developed. Without being bound to theory, Applicants think reactivity to 2F5 may possibly be through interactions with gp160-bound lipids rather than the amino acid backbone. These assays demonstrate that the MPER is exposed on each construct.

Table 3: HIV-1 gp160 Antigen Capture Assay using of HEK293 cells transfected with HIV-1 subtype C gp160 and gp160 DC expression plasmids. 4E10 huMAb to gp41 MPER is used as the capture antibody. Relative quantities of gp160 are detected at 48 hrs post-transfection in media and cell lysates.

Construct	OD 450 nm	
	Media	Cell Lysate
pSWC06838v1c48 gp160	0.146	2.655
pSWC06980v1c3 gp160	0.926	2.718
pSWC06980v0c22 gp160	2.534	2.736
pSWC3728v2c6 gp160	0.589	2.779
pSWC06838v1c48 gp160DC	0.162	2.703
pSWC06980v1c3 gp160DC	1.105	2.588
pSWC06980v0c22 gp160DC	2.608	2.616
pSWC3728v2c6 gp160DC	0.681	2.652
(-) Control	0.123	0.156
(+) Control: Ba-L gp145	2.417	2.553

Table 4: HIV-1 gp160 Antigen Capture Assay using of HEK293 cells transfected with HIV-1 subtype C gp160 and gp160 DC expression plasmids. 2F5 huMAb to gp41 MPER is used as the capture antibody. Relative quantities of gp160 are detected at 48 hrs post-transfection in media and cell lysates.

Construct	OD 450 nm	
	Media	Cell Lysate
pSWC06838v1c48 gp160	0.146	0.530
pSWC06980v1c3 gp160	0.200	0.399
pSWC06980v0c22 gp160	0.344	1.458
pSWC3728v2c6 gp160	0.198	0.781
pSWC06838v1c48 gp160DC	0.155	0.506
pSWC06980v1c3 gp160DC	0.179	0.446
pSWC06980v0c22 gp160DC	0.298	1.210
pSWC3728v2c6 gp160DC	0.219	0.829
(-) Control	0.135	0.150
(+) Control: Ba-L gp145	2.268	2.638

Applicants concluded that the isolate C06980v0c22 would be used to establish CHO cell lines producing gp145. The decision to use this isolate was based on several factors:

- Its relatively strong expression, as compared to isolates C06838v1c48 and C06980v1c2
- Processing of gp160 into gp120 and gp41 was most pronounced with this isolate
- This strain was isolated early from the patient during acute infection
- Its strong reactivity with the MPER antibody 4E10
- This virus was infection competent

CHO-K1/C06980v0c22 gp145 cell line development

The C06980v0c22 gp145 DNA expression plasmid was constructed and used to establish stably transfected CHO-K1 cells producing gp145. These cell lines were adapted for growth in protein-free media, and cell banks were established. The clone H-73-9-2-8 was selected for gp145 protein production.

The C06980v0c22 gp145 DNA construct was developed by modifying the gp160DC gene using a PCR-based technique. The gp145 gene is composed of residues N30, directly downstream of the native signal peptide cleavage site, through K676 just prior to the transmembrane domain. Following K676, the gp145 terminates with three additional lysines (Figure 36 and Figure 37). These are included to theoretically increase the hydrophilicity of the C terminus, thus increasing exposure of the MPER for presentation to the immune system. Upon signal peptide cleavage, it is predicted that a foreign serine from the modified t-Pa signal will be present at the amino terminus. The gp145 gene was ligated at the NheI and EcoRI sites in the mammalian expression plasmid pSWTIPK3 and named pSWC06980v0c22 gp145 (Figure 21). The plasmid was amplified in the E. coli (Invitrogen) strain DH5 α and purified using an Endofree Plasmid Maxi kit (Qiagen, Valencia, CA). The plasmids were analyzed by restriction digest, and

the sequence of the gp145 coding region was confirmed by DNA bidirectional sequencing. The plasmid contains the puromycin acetyl-transferase gene for selection of stable colonies under puromycin selection. It is driven by an internal ribosomal entry site (IRES) and the CMV promoter. This facilitates a high level of expression of the gp145 gene by linking its expression with that of the puromycin resistance marker.

A summary of important features of the mammalian expression vector can be described as follows:

- The vector contains an antibiotic resistance gene that can be used as a selectable marker in bacteria during construction. Therapeutic products derived from such vectors should avoid use of penicillin or related antibiotics during their construction. Therefore, kanamycin is used instead of ampicillin.
- The gp145 gene to be expressed is codon optimized for enhanced expression of the product. The gene is synthesized using CHO codon bias, using codons that correspond to the most abundant tRNAs present in CHO cells. The synthetic gene is also designed to eliminate any cis-acting elements that can reduce transcription/translation efficiency as well as other motifs that may destabilize mRNA.
- The gp145 gene is introduced into the vector in frame with a modified t-Pa signal peptide to allow for efficient transport to the cell membrane.
- The gene is expressed under control of a strong promoter and efficient poly-A signal. The powerful CMV promoter and the efficient Bovine Growth Hormone (BGH) poly-A are used.
- For the selection of stable protein-expressing cell clones, the vector contains a selectable marker: puromycin acetyl-transferase gene for puromycin resistance driven by an internal ribosomal entry site (IRES) and the CMV promoter.

In developing cell lines, care was taken to perform tasks, keep records and use materials that would be acceptable with the FDA should the need arise to use these cell lines in the clinical setting. The CHO-K1 (cat# CCL.61) cells were obtained from ATCC (Manassas, VA). To establish stable CHO-K1 cell lines, the preferred method of transfection is electroporation. A benefit of this method is in its avoidance of uncharacterized animal-derived components. In addition, animal derived products were avoided unless necessary. Recombinant trypsin was used instead of porcine trypsin and the fetal bovine serum (FBS) was well defined from a New Zealand source to reduce the chances of BSE contamination. FBS was irradiated, heat inactivated and sterile filtered.

CHO-K1 cells were separately electroporated with supercoiled and linearized pSWC06980v0c22 gp145 DNA (linearized with the single cutter: NruI). Both forms of DNA were used, as both have their benefits and drawbacks when establishing cell lines. Supercoiled DNA typically transfects with higher efficiency, which may be beneficial, as CHO-K1 cells transfect poorly. Linear DNA transfects with less efficiency, but incorporates into host genome with better efficiency than supercoiled DNA. Briefly, 5×10^6 CHO-K1 cells were suspended in 0.5 ml electroporation buffer (BioRad, Hercules, CA), mixed with 100 μ l Electroporation Buffer containing 100 μ g plasmid DNA in 0.4 cm electrode cuvettes. Cells were pulsed using a Gene Pulser apparatus (BioRad) at 350V with 125 μ FD, set on ice for 30 minutes, pooled and cultured in 5.5 ml complete F-12 K medium (F12-K (Invitrogen, Inc.), containing 10 % heat inactivated FBS (Hyclone Laboratories, Logan UT), 10 μ g/ml gentamicin (Invitrogen, Inc.) and 2 mM l-glutamine (Quality Biologicals, Inc., Columbia, MD)).

Forty-eight hours post-electroporation, conditioned media and cell lysate samples were taken for analyses in a gp120 antigen capture assay and IP western blot. The gp120 antigen capture assay results confirm the secretion of gp145 into the conditioned media. Production was quite low at only four and nine ng/ml from the linear and supercoiled DNA, respectively. Production from the supercoiled DNA is higher than that from the linearized DNA, as expected. Env proteins from Media and cell lysate samples were analyzed in IP western blot (Figure 26). The presence of gp145 in the conditioned media

from both the supercoiled and linearized DNA electroporated cells is evident at about 140 kDa, as expected. A slightly lower molecular weight species of gp145 is evident in the cell lysates, as expected. This likely represents incompletely processed gp145.

Forty-eight hours post-electroporation, cells were plated at about 4500 cells/well in 96-well plates for selection in cF12-K. After 24 hours, cells were put under puromycin selection in cF12K containing 10 µg/ml puromycin (Sigma-Aldrich, St. Louis, MO). Media was changed twice a week until puromycin-resistant colonies reached about 50% confluency. Conditioned media was analyzed for gp145 production using the gp120 antigen capture assay. Twenty cultures with the highest production were expanded, frozen and cloned by limiting dilution to isolate true clones that stably express gp145. The best production levels of uncloned cultures reached 2 µg/ml.

From the original 20 cultures initially selected and cloned, >100 clones were analyzed by gp120 antigen capture assay. Based on gp145 production levels, 17 cell lines, representing 12 of the original cultures were determined to be potential candidates for gp145 production. Frozen stocks for each cell line were made. Studies to compare production levels of each of the selected clones were performed. Briefly, cells were seeded in a 24-well plate at 1x10⁵ cells in 1 ml tissue culture media and incubated at 37°C for 64 hours. Media was harvested and analyzed by gp120 antigen capture assay and I.P. western blot. Based on their antigen capture and I.P. western blot results, clones H-73-9, H-84-1 and H-94-10 were selected for adaptation to protein-free media required for protein production. A five vial cell bank was frozen for each. Each clone produced between 1 and 2 mg gp145/L. In I.P. western blot, each of the selected clones has strong gp145 reactivity at about 140 kDa as expected.

It was observed that CHO-K1 cells electroporated with supercoiled DNA yielded higher transient gp145 production than those cells electroporated with linearized DNA: 9 ng/ml versus 4 ng/ml, respectively. However, both supercoiled and linearized DNA yielded about 100 stable cell lines following puromycin selection. Interestingly, 16 of the 17 best gp145 producing cell lines were derived from the linearized DNA. This supports

Applicants' prediction that supercoiled DNA is more efficiently taken up by cells, but linearized DNA is more efficiently integrated and results in higher protein yields in stable cell lines.

The three selected cell lines were adapted for growth in protein-free media. As adaptation for growth in protein-free media can be difficult for certain clones, three were selected to increase the likelihood that an adaptable clone is selected. In addition, each clone was adapted to three different media (PowerCHO-1 CD, PowerCHO-2 CD and PowerCHO-3 CD, Lonza, Walkersville, MD), each containing 5 µg/ml puromycin and 4 mM l-glutamine. After several passages in protein-free media, clone H-73-9 grown in PowerCHO-1 CD, PowerCHO-2 CD and PowerCHO-3 CD adapted and were named H-73-9-1, H-73-9-2, and H-73-9-3, respectively. Clone H-84-1 grown in PowerCHO-2 CD also adapted well and was named H-84-1-2. A two vial cell bank was frozen for each.

The four adapted cultures were again cloned by limiting dilution and the best producing clone for each culture was identified by the gp120 antigen capture assay. Of these, two cultures were identified as being the best producers; H-73-9-2-8 and H-73-9-3-9. Two vial cell banks were frozen for each, and cultures were expanded to about 500 ml for small-scale protein purification. Conditioned media was harvested, buffered with 20mM Tris, pH 8, 0.5% Triton-X-100 and 500 mM sodium chloride. Buffered media was run through 2 ml columns of Galanthus nivalis lectin (GNL) agarose (Vector Laboratories, Inc., Burlingame, CA). The columns were washed with Tris, pH 8, 0.5% Triton-X-100 and 500mM sodium chloride, followed by PBS. The bound gp145 was eluted with 400mM methyl α-D-mannopyranoside. Purified gp145 was analyzed in SDS-PAGE and western blot (Figure 27 and Figure 28). Both clones produce an approximately 145 kDa protein that reacts well in western blot. Under nonreducing conditions, some dimer and high order multimers are evident, also. Both cell lines produced >1.2 mg/L according to Comassie Plus Protein Assay (Pierce). Clone H-73-9-2-8 had better growth characteristics (slightly healthier and faster growth), thus was selected gp145 production.

A 10 vial research cell bank (RCB) for H-73-9-2-8 (lot 4/17/08) was made and stored in liquid nitrogen freezer. 2X10⁶ cells were frozen in 10 vials of 1 ml protein-free freezing medium (7.5% DMSO (Sigma-Aldrich), 50% fresh growth media, 42.5% Profreeze CDM (Lonza)). The genomic DNA was isolated from 5x10⁶ cells using Qiamp Blood Mini Kit (Qiagen), and the integrated gp145 gene region was amplified by PCR and sequenced in both directions. There was a 100% sequence match in the gp145 coding region. Two weeks after cell banking, cells were tested for mycoplasma contamination using MycoAlert Mycoplasma detection Kit (Lonza) and were found to be negative. One vial was thawed and put into culture to test for viability. After 3 days of culture, cells were 77% viable and tested positive for gp145 production in gp120 antigen capture assay. Culture was tested for bacterial contamination, and showed no bacterial growth in inoculated SOC broth after incubation at 37° C for 24 hours.

A culture grown from the H-73-9-2-8 RCB lot 4/17/08 was used to make s 25 vial RCB (lot F1144) and stored in liquid nitrogen freezer. 10X10⁶ cells were frozen in 25 vials of 1 ml protein-free freezing medium (7.5% DMSO (Sigma-Aldrich), 50% PowerCHO-2 CD, 42.5% Profreeze CDM (Lonza)). At the time of cell banking, the culture was tested for mycoplasma contamination and was found to be negative. The culture was tested for bacterial and fungal contamination, and showed no bacterial growth in thioglycollate broth or fungal growth in soybean-Casein digest broth. One vial from the cell bank was thawed and cultured in the media previously described to check for viability and to confirm production of gp145 in these cells. Upon thawing, viability was acceptable at 87%. Growth characteristics were as expected and gp145 production was confirmed by antigen capture assay.

C06980v0c22 gp145 Protein production

Three lots of gp145 were produced and delivered for further studies (Table 5).

Table 5: HIV-1_{C06980v0c22} gp145 Lots provided

Lot Number	Concentration (mg/ml)	Total Delivered (mg)
112009	1.1	7
120710A	1.0	18
120710B	0.975	5

Lot 112009

H-73-9-2-8 culture was expanded to 3 L in PowerCHO-2 CD supplemented with 4 mM l-glutamine and 5 µg/ml puromycin using roller bottles. The conditioned media was clarified by centrifugation. Media samples from the 3 L harvest were analyzed in antigen capture. Results predicted nearly 4 mg gp145/L media in gross. Quality and yield was determined to be acceptable for production using the 3L harvest. The gp145 protein was purified as described below and as outlined in Figure 29.

Harvested media was concentrated at room temperature using a 0.1 m² Pellicon filtration unit with molecular weight cut off of 30 kDa. The system was flushed with 1.0 M NaOH followed by WFI water then by 1X PBS buffer. The CHO cell culture supernatant (3 L) was introduced and the system was operated in a recirculation mode. The concentration was performed at a permeate flux of –1 L per hr/ 0.1 m² and a cross flow of 0.5L/min. At the end of concentration, the sample was concentrated to 200 mL. The concentrated cell culture supernatant was stored at –70 °C until further processing.

The gp145 was purified using Lectin Affinity Chromatography. Concentrated media was adjusted to 500 mM sodium chloride and was run over 25 ml column of GNL agarose. The column was washed with PBS and the bound gp145 was eluted with 500 mM methyl α-D-mannopyranoside. The purification was performed at a flow rate of 10 ml/min. The eluate (100 mL) was concentrated and diafiltered into PBS using a tangential flow filtration. A 50 cm² Pellicon filtration unit with molecular weight cut off of 30 kDa was used for this step. The system was flushed with 1.0 M NaOH, followed

by WFI water then by PBS. The eluate was then introduced, and the system was operated in recirculation mode. The ultra filtration was performed at a permeate flux of 2.0 mL/min., and the cross flow was ~ 40 ml/min. The volume of the GNL-Eluate was reduced to 8 mL in PBS.

Protein content was estimated by Bradford assay and found to be 1.1 mg/mL. Endotoxin in the purified gp145 was estimated using colorimetric LAL assay (Lonza) and found to be 31.8 EU/mg protein. SDS-PAGE analysis in reduced and non-reduced condition showed the molecular weight of ~145 kD for the purified protein.

As can be seen from SE-HPLC analysis (Figure 30), lot 112009 gp145 was eluted in four multimeric forms, referred to as A, B, C, and D. Based on the mobility of protein standards, the apparent molecular weight of each gp145 species is calculated (Table 6). The major peak calculates as >669 kDa (estimated to about 895 kDa), corresponding to form A. Two shoulders are evident at > 669 (estimated to about 680 kDa) and 571 kDa, corresponding to forms B and C, respectively. A fourth, but minor peak of 417 kDa corresponds to a form D.

Table 6: Retention time and molecular weight of protein standards and purified gp145 (lot 112009) multimer species

Protein	Retention Time (Min)	Molecular Weight (kDa)	Multimeric form
Thyroglobulin	11.81	669	N/A
Ferritin	13.74	440	N/A
Catalase	16.18	232	N/A
Bovine Gamma Globulin	16.52	150	N/A
Bovine Serum Albumin	16.75	66	N/A
gp145 peak 1	11.15	Approx. 895	A
gp145 peak 2	11.77	Approx. 680	B
gp145 peak 3	12.25	571	C
gp145 peak 4	14.64	417	D

The purity of the gp145 was 96.1 % by SDS-PAGE followed by laser densitometry.

Lot 120710A

H-73-9-2-8 culture was expanded to 11 L in PowerCHO-2 CD supplemented with 4 mM l-glutamine and 5 µg/ml puromycin using roller bottles. The conditioned media was clarified by centrifugation. Media samples from the 11 L harvest were analyzed in antigen capture. Results predict nearly 8 mg gp145/L media in gross. Quality and yield was determined to be acceptable for production using the 11 L harvest. The gp145 protein was purified as described below and as outlined in Figure 31.

Harvested media was concentrated at room temperature using a 0.1 m² Pellicon filtration unit with molecular weight cut off of 30 kDa operating in a recirculation mode as described for lot 112009. The 11L conditioned media was concentrated to 1L.

The concentrated conditioned media was buffered with 20 mM Tris, pH8, 500 mM sodium chloride and 0.5 % Triton-X-100, and then clarified with 0.22µm filter. Conditioned media was passed over 20 ml GNL-Agarose resin at 4° C at about 1 ml/min. The resin was washed with 20 mM Tris, pH8, 500 mM sodium chloride and 0.5% Triton-X-100 buffer, and then equilibrated with PBS. The gp145 was eluted in PBS containing 0.5M Methyl- α -D manopyranoside. GNL-Eluate (88 mL) was concentrated to 20 ml with 50 kDa MWCO filter. 10 ml was set aside for use in preparing lot 120710B. The remaining 10 ml was run on PD10 buffer exchange resin into PBS. Eluted material was sterile filtered with 0.22µm filter, aliquoted and stored at -70° C. Lot 120710A final product has a volume of 22ml.

Protein content was estimated by Bradford assay and found to be 1.0 mg/mL. Endotoxin in the purified gp145 was estimated using colorimetric LAL assay and found to be <0.313 EU/mg protein. SDS-PAGE analysis in reduced and non-reduced condition shows the molecular weight of 142 kD for the purified protein. Under non-reducing conditions, multimers are also evident. This represents multimers held together with disulfide bonds.

As can be seen from SE-HPLC analysis (Figure 32), gp145 lot 120710A was eluted in four multimeric forms as with lot 112009. Based on the mobility of protein standards, the molecular weight of each gp145 species is calculated (Table 7). The major peak calculates as 666 kDa, corresponding to a form B. A second peak and a shoulder calculate to >669 kDa (estimated to about 845 kDa) and 584 kDa, corresponding to forms A and C, respectively. A fourth, but minor peak of 411 kDa corresponds to a form D.

Table 7: Retention time and molecular weight of protein standards and purified HIV-1C06980v0c22 gp145 lot 120710A multimer species

Protein	Retention Time (Min)	Molecular Weight (kDa)	Multimeric form
Thyroglobulin	11.813	669	N/A
Ferritin	13.677	440	N/A
Catalase	16.114	232	N/A
Bovine Gamma Globulin	16.400	150	N/A
Bovine Serum Albumin	16.737	66	N/A
gp145 peak 1	11.226	Approx. 845	A
gp145 peak 2	11.824	666	B
gp145 peak 3	12.217	584	C
gp145 peak 4	14.343	411	D

The purity of HIV-1C_{06980v0c22} gp145 lot 120710A was 94.2 % by SDS-PAGE followed by laser densitometry.

Lot 120710B

Lot 120710B is made from the same gp145 eluted during lectin affinity chromatography as lot 120710A. For lot 120710B, an additional step for the purpose of reducing intermolecular disulfide bonds is employed. The rationale for this is based on the observation that the previous lot of gp145 (lot 112009) is in the form of high order multimers. Without being bound to theory, Applicants believe that many of these multimers are due to oxidation, resulting in intermolecular disulfide bridges. Reduction of these bonds is attempted to produce protein in the form of lower order multimers, preferably trimer.

The gp145 protein was purified as described below and as outlined in Figure 33.

10 ml of the concentrated GNL eluate described for lot 120710A had been set aside for use in preparing lot 120710B. This 10 ml was treated with 50 mM DTT at 37° C for 30 minutes, then run on PD10 buffer exchange resin into PBS. Eluted material was sterile filtered with 0.22µm filter, aliquoted and stored at -70° C. Lot 120710B final product has a volume of 22ml.

Protein content was estimated by Bradford assay and found to be 0.975 mg/mL. Endotoxin in the purified gp145 was estimated using colorimetric LAL assay and found to be <0.321 EU/mg protein. SDS-PAGE analysis in reduced and non-reduced condition shows the molecular weight of 143 kDa for the purified protein. Under non-reducing conditions, only trace amount of multimers are also evident. This represents multimers held together with disulfide bonds. Treatment with DTT reduced many of these bonds compared to the non-DTT treated lot 120710A.

Western blot shows the major band at about 143 kDa under reducing and non-reducing conditions. Several multimeric forms of gp145 are evident under non-reducing conditions, but fewer than seen with the non-DTT treated lot 120710A. Under reducing conditions, these multimers have mainly been reduced to monomer.

As can be seen from SE-HPLC analysis (Figure 34), gp145 was eluted in four multimeric forms as with lots 120710A and 112009. Based on the mobility of protein standards, the molecular weight of each gp145 species is calculated (Table 8). The major peak calculates as 665 kDa, corresponding to a form B. A second peak and a shoulder calculate to >669 kDa (estimated to about 844 kDa) and 572 kDa, corresponding to forms A and C, respectively. A fourth, but minor peak of 412 kDa corresponds to form D.

Table 8: Retention time and molecular weight of protein standards and purified HIV-1C06980v0c22 gp145 lot 120710B multimer species.

Protein	Retention Time (Min)	Molecular Weight (kDa)	Multimeric form
Thyroglobulin	11.813	669	N/A
Ferritin	13.677	440	N/A
Catalase	16.114	232	N/A
Bovine Gamma Globulin	16.400	150	N/A
Bovine Serum Albumin	16.737	66	N/A
gp145 peak 1	11.23	Approx. 844	A
gp145 peak 2	11.83	665	B
gp145 peak 3	12.289	572	C
gp145 peak 4	14.327	412	D

The purity of the gp145 is 94.2 % by SDS-PAGE followed by laser densitometry.

DTT reduction of intermolecular disulfide bonds had some effect on the multimeric form of gp145. In SDS-PAGE, it is clear that most intermolecular disulfide bonds were reduced with DTT treatment, if compared to non-DTT treated. SE-HPLC shows only a modest increase in form C, perhaps due to a modest reduction of form A.

Further investigation of the nature of the multimeric forms is included below in Example 7.

EXAMPLE 4

Induction of Neutralizing Antibodies to HIV-1 by Immunization with CHO-Expressed Recombinant gp145 Derived from Acute Clade C HIV-1

Animals: 24 New Zealand White female rabbits, 1.8-2 kg

The rabbits are divided into 6 groups of 4 animals each. The individual animals are identified by cage cards and ear tags.

Group	Antigen	Adjuvant/Vehicle
1	gp145	Alhydrogel
2	gp145	Liposome formulation 1 containing lipid A, preformed and mixed with gp145
3	gp145	Liposome formulation 1 containing lipid A with encapsulated gp145
4	gp145	Liposome formulation 2 containing lipid A and PIP with encapsulated gp145
5	None	Liposome formulation 1 containing lipid A
6	mper23, clade B	Liposome formulation 2 containing lipid A and PIP with encapsulated mper23

Antigen: 25 µg/rabbit/dose

gp145 as described above expressed in CHO cell (acute clade C, C0698v0c22) in PBS, pH 7.4

mper23 (NK-4): LELDKWASLWNWFDITNWLWYIK (SEQ ID NO: 53), (HBX2 variant; Swiss-Prot accession number P04578 except that the N at position 674 was replaced with D.)

Adjuvants:

Alhydrogel (0.6 mg Al³⁺ /dose) formulated at 0.6 mg Al³⁺ in 0.125 ml of PBS, pH 7.4; Mixed with equal volume of antigen

Liposome formulation 1 – DMPC:cholesterol:DMPG (9:7.5:1); 50 mM phospholipids containing 100 µg of lipid A/0.25 ml dose; PBS, pH 7.4. DPMC refers to dimyristoyl phosphotidylcholine, and DMPG refers to dimyristoyl phosphotidylglycerol.

Liposome formulation 2 – DMPC:cholesterol:PIP (1:1.5:1); 50 mM phospholipids containing 100 µg of lipid A/0.25 ml dose; PBS, pH 7.4. PIP refers to phosphotidyl inositol-4-phosphate.

Bleeding: The animals are bled at weeks -2, 0, 4, 8, and 10 from an ear artery using a 20-24 gauge butterfly catheter. Approximately, 5 ml of blood is taken during each phlebotomy. The blood is incubated at room temperature for 2-3 hr and then refrigerated overnight at 4C prior to centrifugation to remove the serum from the clot. The serum is aliquoted: 1 x 1ml and 3 x 0.5 ml in plastic vials and frozen at -80C.

At week 12, the rabbits are terminally bled by cardiac puncture after anesthesia with atamine/Xylanine with a 60 cc syringe and an 18 gauge needle. The serum is aliquoted : 5 x 1 ml and 5 x 5ml in plastic vials and frozen at -80 °C.

Immunization: Weeks 0, 4 and 8 by the intramuscular route in alternating caudal thigh muscles. Inject 0.25 ml with a 23-27 gauge needle.

Schedule:

Week	Procedure
	Receive rabbits
Week -2	Release from quarantine
Week -2	Pre-bleed
Week 0	Pre-bleed
Week 0	Immunize IM, 0.25 ml
Week 4	Bleed
Week 4	Immunize IM, 0.25 ml
Week 8	Bleed
Week 8	Immunize IM, 0.25 ml
Week 10	Bleed
Week 12	Terminal Bleed

See also Figure 41. Results are presented in Figures 42-48, 52.

EXAMPLE 5

The following methods detailed below for the mouse studies incorporate the immunogenicity/antigenicity methods used in the rabbit studies depicted in the Figures (see Figures 41-48, 52), except rabbits received 25 ug of gp145 or placebo in liposomes as described.

Gp145 Mouse Immunogenicity Study: Antigenicity/Immunogenicity Methods

Antigens

The gp145 protein was produced from an envelope sequence isolated from an acute, subtype C infected individual from Tanzania. The entire ecto-domain of the protein is present, including the MPER of gp41. The protein was designed to include two mutations (R508S, R511S) in the gp120/gp41 cleavage site to prevent protease cleavage and a multi-lysine C-terminal to facilitate production and MPER epitope presentation. The protein was produced in CHO cells, purified by lectin affinity chromatography and is present as a mix of multimers as described in Example 7. The gp145 protein contained the following MPER epitope sequence: ALDSWNNLWNWFDIS (SEQ ID NO: 23).

Liposome preparation

Antigens, experimental M13 phage or gp145 protein, were encapsulated in liposome prior to immunization. Liposomes composed of dimyristoyl phosphatidylcholine, dimyristoyl phosphatidylglycerol and cholesterol in molar ratios of 1.8:0.2:1.5 were prepared by dispersion of lyophilized mixtures of lipids at a phospholipid concentration of 50 mM in Dulbecco's PBS with 0.4 g/ml lipid A, either lacking or containing antigen. Liposomes were washed twice in sterile saline to remove the unencapsulated antigen.

Animal Immunizations

Forty female BALB/C mice, 25 g each, were immunized under a protocol approved by the Institutional Laboratory Animal Care and Use Committee. Animals were divided into eight groups of five animals each (Table 9). Mice were immunized intramuscularly four times in alternating caudal thigh muscles at two or three week intervals with 5×10^{11} phage or 10 μ g gp145 protein each per dose. Blood was collected at two-week intervals starting two weeks prior to the first immunization ending when the animals were euthanized. Blood was incubated at room temperature for 2-3 h, refrigerated overnight at 4°C then centrifuged. Serum was collected and stored at -20°C.

Two weeks after the last boost (week 10) the mice will be euthanized. Blood, spleens, lymph nodes, bone marrows, and livers were obtained and processed from naïve and immunized mice.

Group#	Immunogen	Immunization (Weeks)	Bleeds (Weeks) 150ul/mouse/bleed	Euthanasia (collect blood, spleens, lymph nodes, bone marrow and livers)
1	M13-12D4	0, 3, 6, 8	-2, 0, 2, 4, 6, 8, 10	Week 10
2	M13-12B7	0, 3, 6, 8	-2, 0, 2, 4, 6, 8, 10	Week 10
3	M13-all 5	0, 3, 6, 8	-2, 0, 2, 4, 6, 8, 10	Week 10
4	gp145/M13-all 5	0, 3, 6, 8	-2, 0, 2, 4, 6, 8, 10	Week 10
5	gp145	0, 3, 6, 8	-2, 0, 2, 4, 6, 8, 10	Week 10
6	M13-no insert	0, 3, 6, 8	-2, 0, 2, 4, 6, 8, 10	Week 10
7	Naïve	0, 3, 6, 8	-2, 0, 2, 4, 6, 8, 10	Week 10

Table 9. Mouse immunization plan.

IFN γ -release ELISPOT (Enzyme-linked Immunosorbent Spot) assay

Spleen cells secreting IFN γ were analyzed by ELISPOT. Ninety-six-well nitrocellulose-backed MultiScreen-IP sterile plates (Millipore) were coated overnight at 4°C with 10 μ g/ml of anti-gamma interferon (IFN γ) (PBL Interferon Source) in sterile PBS. The wells were blocked with sterile PBS containing 0.5% bovine serum albumin for 30 min at 37°C and washed with PBS containing 0.025% Tween 20 (wash solution) followed by sterile RPMI-1640 complete medium. Single cell suspensions were prepared from the mouse spleens of each group (five mice/group). Cells (2×10^6 /well) were plated on anti-IFN γ -coated plates and incubated for 18 h at 37°C in a humidified CO₂ incubator. Cells were incubated with 5 μ g/ml acute C gp145 (HIV-1 C06980, Advanced Bioscience Laboratories), gp140 (HIV-1 IIIB, Advanced Bioscience Laboratories), yeast-derived gp41 (Meridian Biosciences) or 10 μ g/ml cathepsin degraded, yeast-derived gp41 or no protein. Plates were washed with wash solution followed by distilled water and overlaid with 0.125 μ g/ml of biotinylated anti-IFN γ (clone XMG 1.2; BD PharMingen) and incubated at room temperature for 2h. The plates were then washed and incubated with a 1:1,000 dilution of avidin-conjugated alkaline phosphatase (Vector Laboratories) for 2 h at room temperature. The plates were washed, and bound IFN γ was detected by the

addition of 5-bromo-4-chloro-3-indolylphosphate (BCIP)/nitroblue tetrazolium (NBT) (Kirkegaard and Perry Labs). The plates were washed with water, and the individual spots were visualized and counted the next day using a stereo binocular microscope. The average number of spots/number of cells plated was plotted.

Antigen presentation and detection of cytokines from T-cells by flow cytometry

Cells from spleens or lymph nodes from the different groups of mice were stimulated with 5 µg/ml acute C gp145 (HIV-1 C06980, Advanced Bioscience Laboratories), gp140 (HIV-1 IIIB, Advanced Bioscience Laboratories), yeast-derived gp41 (Meridian Biosciences) or 10 µg/ml cathepsin degraded, yeast-derived gp41 or ConA as the positive control for 22 h at 37°C. The cells were incubated with the above-mentioned antigens for 2 h before the addition of brefeldin A (1 mg/ml, Sigma-Aldrich) and monensin (0.07 mg/ml, BD Pharmingen). Cells were incubated for an additional 20 h. Cells were analyzed on an LSR II (BD Immunocytometry Systems) flow cytometer and 500,000 events were collected using FACSDiva software (BD Immunocytometry Systems). Dead cells were excluded using a viability marker and B-cells were excluded. The CD3⁺ CD4⁺ and the CD3⁺ CD8⁺ T-cells were gated and analyzed for the expression of IL-2, TNF-α, IFN-γ and CD107a. The data were analyzed using FlowJo software (Tree Star). Percent positively stained cells per antigen are shown for each group. The black bar represents a two-fold range above the control response, M13 – no insert.

Antigen-specific serum IgG ELISA

Antigen specific IgG titers were determined by binding ELISA titrations using gp145 and gp41 as targets. Antigens were diluted to 0.25 µg/ml in PBS (pH 7.4), 100 µl per well was added to 96-well microtiter Immunol 2 polystyrene plates. Plates were incubated overnight at 4°C then washed three times with 300 µl 0.1% PBST (PBS containing 0.1% Tween-20). Serum was titered in 2-fold serial dilutions starting at 1:50 dilution in serum diluent (0.1% PBST containing 5% non-fat milk), and 100 µl each dilution was added to the plate. Plates were incubated at 37°C for 1 h then washed three times with wash buffer. HRP-labeled anti-mouse IgG antibody diluted to 1:16,000 in

serum diluent was added, 100 μ l/well. Plates were incubated for 1 h at 37°C then washed three times with wash buffer. TMB (100 μ l, Kirkegaard and Perry Labs) was added, incubated for 30 min at 37°C and the reaction stopped by adding 100 μ l of 1 M phosphoric acid. Plates were read on a spectrophotometer at 410 nm, 570 nm reference filter. Antigen binding titer was determined by calculating the concentration at which binding was detectable above three times background. Two independent assays were performed and the results were averaged.

Surface Plasmon resonance (SPR) measurements by Biacore

SPR measurements were conducted with a Biacore T200 using CM5 chips. Peptides were immobilized to the chip surface using the Biacore amine coupling kit (Biacore, AB). All immobilization steps used a flow rate of 10 μ l/min and were performed at 25°C. The peptide loading buffer was 20 mM sodium acetate, pH 4.2. The immobilization wizard packaged within the T200 control software was used to immobilize 14700 resonance units (RU) of 10 μ M scrambled MPER peptide and 20500 RU of MPER peptide to their respective flow cells. Both peptides had a 10 min contact time during immobilization. The serum samples were diluted 1:50 in Tris buffered saline, pH 7.4 and passed over the chip surface at 30 μ l/min for 3 min followed by a 5 min dissociation period. At the end of the 5 min period, a 75 μ g/mL solution of sheep anti-mouse IgG(Fc) antibody (The Binding Site) was passed over the flow cells for 2 min at a flow rate of 10 μ l/min. After a 70 s dissociation period, the chip surface was regenerated using a 30 second pulse of 50 mM HCl, a 30 second pulse of 100 mM EDTA in 20 mM Tris, pH 7.4, and 30 second pulse of 50% acetic acid followed by a 1 minute injection of Tris-buffered saline, pH 7.4. Non specific binding was subtracted and data analysis was performed using the BIAevaluation 4.1 software. The reported response units for the IgG specific values are the difference between the average value of a 5 second window taken 60 seconds after the end of the anti-IgG injection and the average value of a 5 second window taken 10 seconds before the beginning of the anti-IgG injection.

Pseudovirus neutralization assay

TZM-bl cells were used as assay targets to determine HIV-1 neutralization. BnAb or plasma were titrated in 4-fold serial dilutions starting at 25 µg/ml or 1:20 dilution respectively, in growth medium [DMEM with 100 U/ml penicillin, 100 µg/ml streptomycin, 2 mM L-glutamine (Quality Biologics Inc.), and 15% fetal calf serum (Gemini Bio-Products)] and 25 µl added in duplicate to a 96-well flat-bottom black plate. Pseudovirus, diluted in growth medium to a dilution optimized to yield ~150,000 relative luminescence units (RLU), was added in equal volume to each well. The samples were incubated at 37°C in a humidified 5% CO₂ incubator for 1 h. All incubations were under these conditions. TZMbl cells were resuspended at 2×10^5 cells/ml in growth medium containing 60 µg/ml DEAE-dextran (Sigma), 50 µl was added to each well. Each plate included wells with cells and pseudovirus (virus control) or cells alone (background control). Plates were incubated for 48 h, and then 100 µl/well of reconstituted Brite Lite Plus (Perkin Elmer) was added. RLU values were measured using a Victor 2 luminometer (Perkin-Elmer). The percent inhibition due to the presence of the antibody was calculated by comparing RLU values from wells containing antibody to well with virus control. Two independent assays were performed and the results were averaged.

PBMC neutralization assay

PBMC, collected from HIV-negative donors and cryopreserved, were used as assay targets to determine HIV-1 neutralization. This assay uses replication-competent HIV-1 infectious molecular clones (IMC) containing a Renilla reniformis luciferase (LucR)-expressing HIV-1 reporter gene; viral production is measured with a luminometer (Edmonds, TG et al. Virology 408:1-13 (2010)). Sera were titrated in 4-fold serial dilutions starting at 1:20 dilution in IL-2 growth medium [RPMI-1600 with 100 U/ml penicillin, 100 µg/ml streptomycin, 2 mM L-glutamine (Quality Biologics Inc.), 15% fetal calf serum (Gemini Bio-Products), and 20 U/ml recombinant interleukin-2 (Roche Diagnostics)] and 25 µl was added in duplicate to a 96-well round-bottom plate. IMC, diluted in IL-2 growth medium to a dilution optimized to yield ~50,000 RLU, was added in equal volume to each well. The samples were incubated at 37°C in a humidified 5%

CO₂ incubator for 1 h. All incubations were under these conditions. PHA/IL-2 stimulated PBMC were resuspended at 2×10^6 cells/ml in IL-2 growth medium then 50 μ l was added to each well. Each plate included wells with cells and IMC (virus control) or cells alone (background control). Plates were incubated for 24 h, 100 μ l of growth medium was added to each well and then plates were incubated for an additional 72 h. Renilla Luciferase Assay System (Promega) was used to quantify luciferase production. Lysis buffer, 50 μ l/well, was added and two freeze/thaw cycles were performed, 20 μ l/well was transferred to a black, flat-bottom plate and RLU in each well were measured immediately after injection of 100 μ l substrate. The percent inhibition due to the presence of the antibody was calculated by comparing RLU values from wells containing antibody to well with virus control. Two independent assays were performed and the results were averaged.

.....g.....city of the five M13-displayed 4E10 epitopes capable of inhibiting neutralization was evaluated *in vivo*. Thirty-five female BALB/C mice, seven groups of five animals each, were vaccinated with a single M13-displayed epitope, all five M13-displayed epitopes or all five M13-displayed epitopes in combination with HIV-1 gp145 envelope protein (Table 9). The gp145 envelope protein, from an acute clade C HIV-1 infection, has been shown to elicit neutralizing antibodies in rabbits.

Analysis of elicited cellular immune response

Cellular immune responses elicited by vaccination were assessed by INF γ -release ELISPOT and intracellular cytokine staining (ICS) assays in both the spleen and lymph node. In these assays HIV-specific responses were measured after stimulation with HIV-1 envelope proteins: gp145, gp140, gp41 or cathepsin degraded gp41. A response two-fold greater than the control group, mice immunized with M13 – no insert, was considered a positive response (Figure 53 and 54). ICS data was analyzed to determine CD3+CD4+ or CD3+CD8+ T-cell specific responses.

In the INF γ -release ELISPOT assay, a single response was observed in all groups in both the lymph node and the spleen; the M13-all 5 immunized group, stimulated with gp140 in the lymph node and the gp145 immunized group, stimulated with gp145 in the spleen (Figure 53). Background responses to gp140 were high in the splenic T-cells. IL-2 responses were observed by ICS for all groups against several HIV-1 envelope antigens; TNF α , CD107a and INF γ responses were not detected. Positive IL-2 responses were more frequent in the lymph node than in the spleen, 85% and 48% positive responses, respectively, but were lower in magnitude, 3.9-fold and 4.5-fold above control, respectively. Positive IL-2 responses were more frequent in the CD4+ T-cell compartment than in the CD8+ T-cell. 73% and 60% positive responses, respectively, and were higher in magnitude, 4.5-fold and 3.8-fold above control, respectively (Figure 54). Mice immunized with liposomes only did not have HIV-1 specific cellular responses.

Analysis of elicited antibody responses

Humoral immune responses were analyzed by IgG binding ELISA, Biacore and by neutralization assays. Binding titers against gp145 and gp140 were determined for all groups (Figure 55). Animals immunized with gp145 or gp145/M13-all 5 produced antibodies with high titers gp145, average of 512000 and 409600 respectively, and gp41, average of 30400 and 43200 respectively; both groups had the highest titers to the gp145 immunizing protein. The other groups did not have detectable binding titers in this assay with the exception of

d a weak binding titer to gp145, average 1200. Biacore was used to -specific IgG binding to MPER peptide in pooled serum; no binding was observed (data not shown).

Neutralization assays were performed using both TZMbl and PBMC as assay targets.

- 5 Sera were titrated against two neutralization-sensitive HIV-1 strains in both assay platforms and ID50 values were calculated (Figure 56). Animals immunized with gp145/M13-all 5 had the highest neutralization titers in both the TZMbl and PBMC assays, a 2.1- and 1.9-fold increase respectively over the gp145-immunized group. Animals immunized with a single M13-displayed MPER epitope, M13-12D4 and M13-12B7, or multiple M13-displayed
- 10 MPER epitopes (M13-all 5), also produced HIV-neutralizing antibodies. All sera were screened against the HIV-2/MPER chimera and a nonspecific viral control, MuLV, no neutralization was observed for either of these viruses.

EXAMPLE 6

α 4 β 7 Blocking Assay

- 15 Materials:

Media : 10% FCS / RPMI / Lglut / PenStrep

Cells : RPMI8866

Table 10 - Reagents:

20

Reagent	Clone#	Mfg	Cat#	Lot#	Working Dil.	Comments
Normal mouse IgG	-	Invitrogen	10400C	645253A	10%	
Normal human IgG	-	GenScript	A01006	A108810	10%	Make lyoph. stock to 1mg/mL H2O
Anti- α 4blocking MAb	HP2/1	Beckman Coulter	IM0764	21	2ug/well	make lyoph. stock to 0.5mg/mL H2O
Anti- β 7-FITC	FIB504	BioLegend	321212	B142196	1:10	
Neutravidin-PE	-	Invitrogen	A2660	866787	1:400	
gp145-AcuteC-biotin	-	V.Polonis	-	LotB	0.5ug/well	biotinylated 2X by RF
Anti-Env mAbs x 14	See Table 12 for list of mAbs tested					rcvd from S. Zolla-Pazner

Prepare α 4 β 7 Binding Buffer

- prepare 1M MnCl₂ fresh from powder (1g MnCl₂·4H₂O (from Arthos Lab - MW 197.9) + 5mL dH₂O
- 25 - prepare binding buffer per table below
- sterile filter solution and store at 4°C

TABLE 11

Reagent	Mfr	Cat#	[Final] (mM or %)	[Initial] (M or %)	mL for 500 mL
HEPES buffer	Gibco	15830	10	1	5
NaCl	Sigma	56546	150	5	15
MnCl ₂	(commercially available)		1	1	0.5
CaCl ₂	Fluka	21115	0.1	1	0.05
BSA	Sigma	A9576	0.5	30	8.3
NaN ₃	Aldrich	438456	0.09	10	4.5
dH ₂ O	Quality Bio	351029101	-	-	466.6

Harvest cells

- 5 - collect non-adherent cells and transfer to a 50ml tube
- pellet cells at 200xg for 10' and pour off supernatant
- combine all cells into 10mL media, resuspend vigorously to break up clumps and count
- adjust volume to 1.0x10⁶ cells/mL media
- 10 - aliquot 100uL (100K) cells per well into assay plates (96-well U-bottom polypropylene)
- pellet at 200xg for 10'
- wash cells 2X with binding buffer

15 Table 12 - Samples

16	Sample ID	Specificity	[IgG] (ug/uL)	[uL for 10ug]	[uL Bndg Bfr]	Total [uL]
A	1361.100.10.10.1.1	V2	1.52	6.6	33.4	40
B	1393A.100.10.1.1	V2	2.99	3.3	36.7	40
C	1357A.100.10.1	V2	0.84	11.9	28.1	40
D	2158.100.100.10.1.1	V2	5.28	1.9	38.1	40
E	697-30D.10.1.1	V2	6.44	1.6	38.4	40
F	830A.10.10.1.1	V2	0.45	22.2	17.8	40
G	2297.100.1.1	V2	3.01	3.3	36.7	40
H	447-52D.10.1.1	V3	4.36	2.3	37.7	40
I	1006-15D.100.1	V3	2.40	4.2	35.8	40
J	3869.100.10.1.1	V3	5.27	1.9	38.1	40
K	729-30D.10.10.1.1	CD4bs	2.15	4.7	35.3	40
L	1331-160E.100.10.1	CD4bs	1.54	6.5	33.5	40
M	1570D.10.1.1	CD4bs	3.86	2.6	37.4	40
N	1418(16).10.1	b19	4.75	2.1	37.9	40

Preincubate Protein + IgG

- add binding buffer to wells according to plate layout and sample calculations
- 20 - add IgG to the appropriate sample wells according to the sample calculations

binding buffer alone to wells for UNTREATED, POS CTRL and NO IgG

- make up stock of protein(s) at 0.025ug/uL in binding buffer

Table 13

Protein	[stock] ug/uL	# Wells	uL per well	Total Vol	[Final] ug/uL	uL Protein	uL Buffer
gp145-AcuteC Lot B2	1.0	20	20	400	0.025	10.0	390

- add 20uL binding buffer to UNTREATED wells
- add 20uL protein to sample wells, POS CTRL and NO IgG CTRL wells (= 0.5ug protein/well)
- incubate plate for 60min at 37°C

10

Binding Assay

- prepare blocking buffer (10% mouse IgG, 10% human IgG in binding buffer)
- add 50uL / well blocking buffer
- add 4uL (2ug) anti- $\alpha 4$ blocking mAb to the POS CTRL well
- incubate on ice x 10min (do not wash off)
- transfer 50uL protein/IgG complexes to assay plate per layout
- incubate on ice x 30min
- wash 2X with binding buffer

20 Staining

- prepare staining cocktail

Table 14:

uL per well	# wells	Total Vol	uL $\beta 7$ -FITC (1:10)	uL NA-PE (1:400)	uL Buffer
50	20	1000	100	2.5	898

- add 50uL binding buffer to unstained wells
- add 50uL staining cocktail to NO IgG CTRL, POS CTRL and all sample wells
- incubate at 4°C for 30min

1 binding buffer

- wash with 100 μ FA x 30-60min at 4°C
- spin cells down and resuspend in 150uL binding buffer
- store plates at 4°C until ready to read

5 Compensation Bead Preparation

- prepare 2 wells of compensation beads, wash beads with staining buffer 2X
- resuspend beads in 45uL staining buffer
- 5uL CD4-FITC to one well and 5uL CD4-PE to the other and mix well
- incubate at 4°C for 30min

- 10
- wash 2X with staining buffer
 - resuspend in 200uL staining buffer

A second α 4B7 binding inhibition assay is a flow cytometry based assay which can be performed on the RMPI8866 cell line, but is primarily performed on isolated CD4+ and

- 15 CD8+ T cells cultured to express the active form of the α 4 β 7 heterodimer. A model for this assay is shown in Figure 57. This assay can be performed as a binding assay to test variants of the HIV-1 env derived from selected acute sequences. The expanded utility of the assay includes the test for functional blocking of the env/ α 4 β 7 interaction by monoclonal antibodies or purified serum IgG against either the env protein or the integrin. Selected HIV-1 env
- 20 protein is generated and biotinylated and pre-incubated with antibody, followed by incubation with the α 4 β 7 expressing cells. Binding is detected by addition of neutravidin-PE and the presence of the α 4 β 7 is confirmed by staining with the non-blocking mAb conjugated to FITC. When testing an anti-integrin antibody, the cells are pre-incubated with the antibody prior to addition of the biotinylated env protein, and detection proceeds as described.
- 25 Applicants have adapted this method for use with whole env protein as well as biotinylated linear and cyclic peptides. In certain embodiments, the assay is developed to be used with the IMC and VLP constructs.

- One of the important features of this assay is the ability to use primary T cells expressing the active form of α 4 β 7. To generate these cells, Applicants isolate CD4+ and
- 30 CD8+ T cells from PBMC by magnetic bead separation. A negative selection protocol is used so the resulting cells are “untouched”, purified and bead-free. Following isolation, cells

lays in the presence of anti-CD3/anti-CD28, IL-2 and retinoic acid to
ssion of $\alpha 4\beta 7$ (Figure 58).

Preliminary experiments during development of this assay were conducted to
determine the binding kinetics and overall utility of the assay with a variety of HIV-1
5 envelope reagents. Recombinant CRF01_AEgp120 and an acute subtype C gp145 (as
described above) were biotinylated and bound to $\alpha 4\beta 7$ expressing CD4+ or CD8+ T cells
(Figure 59, left and center panels). A biotinylated cyclic peptide containing the V2 loop of
HIV-1 Env derived from CRF01-AE also bound both CD4+ $\alpha 4\beta 7$ + and CD8+ $\alpha 4\beta 7$ + cells.
Similar binding was also seen with the RPMI8866 cell line (data not shown). There was no
10 binding detected with a clade B MN derived gp120 or with a cyclic V2 peptide containing a
mutation in the apex of the loop (data not shown).

Initial blocking studies were conducted using human anti-V2 monoclonal antibodies
(kindly provided by S. Zolla-Pazner) and the CRF01_AE-derived gp120 or cyclic V2 peptide.
Both V2-reactive monoclonal antibodies tested, 697-30D and 2158, blocked binding of Env
15 to CD4+ $\alpha 4\beta 7$ + and CD8+ $\alpha 4\beta 7$ + cells (Figure 60). As a positive control, cells were pre-
incubated with the anti- $\alpha 4$ blocking antibody HP2/1 prior to addition of protein or peptide.
Applicants proceeded with these experiments, testing a panel of overlapping linear peptides
derived from the V2 loop to delineate the amino acid residues required for this interaction
(data not shown).

20 Methodology

Preparation of $\alpha 4\beta 7$ T Lymphocytes. Cryopreserved PBMC are thawed in complete
media and CD4+ or CD8+ T cells are isolated by magnetic bead negative selection. Cells are
cultured in the presence of anti-CD3/anti-CD28, IL-2 and retinoic acid for at least 5 days.
Polychromatic flow cytometry is used to monitor phenotype, cell viability and expression of
25 active form of $\alpha 4\beta 7$. For some assays, the human B cell lymphoma line RPMI8866 will be
used as it highly expresses active form of $\alpha 4\beta 7$.

$\alpha 4\beta 7$ Binding/Blocking Assay. Cells expressing $\alpha 4\beta 7$ are incubated with 2-5 μ g
biotinylated V2 peptides or HIV-1 envelope glycoprotein for 30 minutes. Following a wash
to remove unbound peptide/protein, cells are stained with neutravidin-PE and binding is
30 assessed by flow cytometry. For blocking studies, antibodies are pre-incubated with either
the $\alpha 4\beta 7$ expressing cells or with the HIV-1 envelope protein, as appropriate, for 30 minutes
prior to addition.

Synthesis of HIV-1 envelope proteins. Acute envelope sequences are selected from
subjects in the RV217 acute infection study for synthesis. Sequences are submitted to

don optimization and cloning into mammalian expression vectors.

ed in CHO cells or HEK293 cells, which provide different glycosylation patterns that may be important for binding assays. Following expression, a portion of each protein is biotinylated for use in $\alpha 4\beta 7$ binding/blocking assays.

5 Synthesis of biotinylated V2 peptides. Peptides designed by Dr. Tim Cardozo (New York University) are synthesized and biotinylated by Genemed Synthesis Inc. These peptides have been kindly provided to us by Dr. Cardozo.

EXAMPLE 7

10 Summary

 Data from gel filtration supported the presence of a mixture of different multimer species. However, it was not certain how the globular nature, hydrophobic regions and heavy glycosylation of gp145 affect the resolution of the different multimeric forms in this assay. Therefore, it was difficult to conclude what species are present and in what proportion. In
15 addition, poor resolution in this assay made it difficult to determine the relative quantity of each form. To further analyze the oligomeric forms present in the purified lots of C06980v0c22 gp145, Blue Native PAGE (BN PAGE), and EGS crosslinking, SDS-PAGE was performed using purified proteins. In addition, separation of oligomeric forms was attempted using gel filtration chromatography.

20

BN PAGE

 The multimeric composition of purified HIV-1C06980v0c22 gp145 lots 112009, 120710A and 120710B were analyzed using Blue Native PAGE. Purified HIV-1C06980v0c22 gp145 was run on a 4-16% Novex Bis-Tris polyacrylamide gel using
25 Invitrogen's Native PAGE system (Figure 61). For comparison, Ba-L gp145 and the three clade D gp140 proteins (A07412, 57128, and 57140) described above were also run. Laser densitometry analysis run on the BLUE Native PAGE predicts a mixture of multimeric forms for each protein (Table 15).

30

ser densitometry prediction of gp140 and gp145 multimer composition

Clade	Protein	Lot	Apparent Molecular Weight (kDa)	Multimeric Species	Percent composition
B	Ba-L gp145	061308	>800* 691 538 278	A B C D	40.3 % 21.5 % 35.1 % 3.1 %
D	A07412 gp140	3-31-05	>800* 733 574	A B C	30.5 % 32 % 37.5 %
D	57128 gp140	4-13-05	>800* 674 566	A B C	21.4 % 44.5 % 34 %
D	57140 gp140	4-20-05	>800* 773 659	A B C	27.1 % 40.6 % 32.2 %
C	C06980v0c22 gp145	112009	>800* 767 624 302/244**	A B C D	12.8% 52.7% 27.3% 7.2%
C	C06980v0c22 gp145	120710A	>800* 751 621 308/239**	A B C D	11.4 % 61 % 22% 5.6 %
C	C06980v0c22 gp145	120710B (DTT treated)	>800* 770 638 307/239**	A B C D	15.1 % 59.5 % 19.6 % 5.9%

* Extremely diffuse band spanning several hundred kDa; difficult to state molecular weight. More than one species may be present in these fractions.

5 ** Doublet with incomplete separation may represent 2 species including some breakdown products.

From Blue Native PAGE, it is clear that the gp140 and gp145 proteins exist as a mixture of various multimers. However, it is not abundantly clear what species are present. If one assumes that monomer is about 140 kDa, then the expected molecular weight of trimer, tetramer, pentamer and hexamer are about 420 kDa, 560 kDa, 700 kDa, and 840 kDa, respectively. However, due to their globular form, hydrophobic regions, and heavy glycosylation, it is suspected that the proteins may not behave in this manner. Accordingly, for this Example, Applicants designated each multimer species as A, B, C or D, with A being the most complex and D the least complex. All C06980v0c22 gp145 lots behave similarly with B as the predominant form along with significant amounts of C and D. DTT treatment

ars to have no significant affect on the multimer composition. A similar
 clade D 57140 and 57128 gp140s. The clade D A07412 gp140 is similar,
 but the C form is slightly greater in quantity than the B form. For the Ba-L gp145, the
 predominant species are found in a broad, diffuse band corresponding to a likely mixture of
 5 high order multimers, classified as the A species. A distinct major C population and a more
 minor B population are also evident.

EGS Crosslinking/SDS-PAGE

C06980v0c22 gp145 gp145 has been further characterized by SDS-PAGE of proteins
 10 crosslinked with ethylene glycol bis(succinimidylsuccinate) (EGS) to further characterize the
 multimeric forms. The data suggests that trimers predominate, but dimers and trace amounts
 of monomer and higher order multimers are also present. Purified HIV-1_{C06980v0c22} gp145 (lot
 120710A) was crosslinked with 0.2, 1, 5 and 12.5 mM EGS and resolved on a 3-8%
 NuPAGE Tris Acetate polyacrylamide gel (Invitrogen) under reducing and non-reducing
 15 conditions and stained with coomassie blue (Figure 62). Laser densitometry analysis was
 used to estimate the molecular weights of each gp145 species. When treated with 0.2 mM
 EGS, gp145 crosslinking is not complete, and three species are evident at 334, 232 and 139
 kDa. These correspond well to the predicted molecular weights of trimeric, dimeric and
 monomeric forms. As EGS concentrations are increased to 5 and 12.5 mM, crosslinking is
 20 completed, revealing that trimer is the predominate species. A major dimer species is also
 present, but monomer makes up only a trace of the total protein. Under nonreducing
 conditions, a faint band corresponding to a higher order multimer is also evident with the
 fully crosslinked samples. This reveals that some higher order multimers exist that are held
 together with disulfide bridges. Multimeric species A, B and C presumably correspond to the
 25 higher order multimer, trimer and dimer species as resolved with the EGS crosslinked
 protein, respectively. Based on Blue Native PAGE results, the major multimer species B and
 C had apparent molecular weights of 751 and 621 kDa, respectively. These apparent
 molecular weights were too high to conclude with confidence that they represent trimer and
 dimer. However, using EGS crosslinking, the apparent molecular weights are more in line
 30 with the major forms being trimer or dimer. EGS crosslinking SDS-PAGE is a method used
 by J.P. Nkolola, et. al. (1) to describe recombinant HIV-1 92UG037.8 gp140 produced in the
 baculovirus system as trimer. These proteins run at a similar molecular weight as the
 C06980v0c22 gp145 using a similar EGS crosslinking procedure. The apparent
 discrepancies in the predicted molecular weights of the gp145 oligomers observed with BN-

PAGE and EGS crosslinking SDS-PAGE may be due to how the charge, globular nature, hydrophobic regions and heavy glycosylation of gp145 affect the resolution of the different multimeric forms in these assays. It would seem that BN-PAGE is capable of resolving oligomeric forms in the native state. However, the molecular weights cannot be determined due to proportionate differences in the mobility of gp145 and the molecular weight markers. In EGS-crosslinking SDS-PAGE, oligomers are covalently bound together, but denatured as with a regular SDS-PAGE. Under these conditions, gp145 migrates relative to the standards based on its apparent molecular weight much as it would if not crosslinked.

Dimer and Trimer Purification

Gel filtration chromatography has been investigated as to whether it can be used to isolate the various oligomeric species of gp145. Successful isolation of the various forms would allow for the potential investigation of the antigenicity or immunogenicity of each form.

Gel filtration chromatography using Superose 6 was performed. It is thought that Superose 6 would have potential for separating the high molecular weight gp145 species because of its high molecular weight range; optimal separation of proteins is 5 to 500 kDa. Separation of large proteins often proves to be difficult due to relatively poor resolution of gel filtration resins for high molecular weight proteins, such as the gp145 oligomers.

Superose 6 shows some promise for separating the different oligomers.

On an analytical Superose 6 PC 3.2/30 (GE Healthcare) column, 25 µl containing 200 µg HIV-1_{C06980v0c22} gp145 lot 120710A was loaded at 0.05 ml/min in PBS, pH 7.2. 50 µl fractions were collected and analyzed by EGS crosslinking/SDS-PAGE (Figure 63).

Although no fractions contained pure trimer or dimer, some enrichment of trimer was evident in certain fractions. Higher order multimers are evident in some fractions. Different column size and conditions could have better resolution and potentially separate out trimer from the higher order multimers and dimers. In further embodiments, optimization of separation conditions may be performed using a Superose 6 column.

References

Nkolola, J.P., et al. 2010. Breadth of neutralizing antibodies elicited by stable, homogeneous clade A and clade C HIV-1 gp140 envelope trimers in guinea pigs. *J. Virol.* 84: 3270-3279.

* * *

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

- 5 Each patent, patent application, and publication cited or described in the present application is hereby incorporated by reference in its entirety as if each individual patent, patent application, or publication was specifically and individually indicated to be incorporated by reference.

10

WHAT IS CLAIMED IS:

1. An isolated peptide comprising a truncated HIV Env protein, wherein the HIV Env
5 protein is mutated in the native gp120/gp41 cleavage site to prevent protease cleavage,
comprises the MPER of gp41, and is truncated prior to the transmembrane domain.
2. The peptide of claim 1, wherein the HIV Env protein comprises about 1-10
hydrophilic amino acids at its C-terminus.
- 10 3. The peptide of claim 2, wherein the about 1-10 hydrophilic amino acids are three
lysines.
4. The peptide of claim 1, wherein the MPER of gp41 comprises the 4E10 epitope.
- 15 5. The peptide of claim 4, wherein the MPER of gp41 comprises the amino acid
sequence: LWYIK (SEQ ID NO: 24) at its C-terminus.
6. The peptide of claim 5, wherein the HIV Env protein comprises about 1-10 non-native
20 hydrophilic amino acids C-terminal to and contiguous with the LWYIK (SEQ ID NO: 24)
amino acid sequence.
7. The peptide of claim 1, wherein the HIV Env protein is derived from an HIV-1 strain
classified in a group selected the group consisting of: M, O, N, and P.
- 25 8. The peptide of claim 7, wherein the HIV Env protein is derived from an HIV-1 group
M strain.
9. The peptide of claim 8, wherein the HIV-1 group M strain is a subtype (clade)
30 selected from the group consisting of: A, B, C, D, F, G, H, J, and K.
10. The peptide of claim 9, wherein the subtype (clade) is clade B.
11. The peptide of claim 9, wherein the subtype (clade) is clade D.

12. The peptide of claim 9, wherein the subtype (clade) is clade C.
13. The peptide of claim 12, comprising an amino acid sequence having 85 % or greater
5 identity to the amino acid sequence depicted in SEQ ID NO: 1.
14. The peptide of claim 13, wherein the peptide comprises an amino acid sequence having 90 % or greater identity to the amino acid sequence depicted in SEQ ID NO: 1.
- 10 15. The peptide of claim 14, wherein the peptide comprises an amino acid sequence having 95 % or greater identity to the amino acid sequence depicted in SEQ ID NO: 1.
16. The peptide of claim 15, wherein the peptide comprises an amino acid sequence having 98 % or greater identity to the amino acid sequence depicted in SEQ ID NO: 1.
15
17. The peptide of claim 16, wherein the peptide comprises an amino acid sequence having 99 % or greater identity to the amino acid sequence depicted in SEQ ID NO: 1.
18. The peptide of claim 17, wherein the peptide comprises the amino acid sequence
20 depicted in SEQ ID NO: 1.
19. An isolated nucleic acid comprising a nucleic acid sequence encoding the amino acid sequence of any of claims 13-18.
- 25 20. The nucleic acid of claim 19, wherein the nucleic acid sequence encodes the amino acid sequence depicted in SEQ ID NO: 1.
21. A vector comprising the nucleic acid of claim 19.
- 30 22. A host cell comprising the vector of claim 21.
23. The host cell of claim 22, wherein the host cell is a CHO cell.

24. A method of making a peptide comprising an amino acid sequence having at least 85 % or greater identity to the amino acid sequence depicted in SEQ ID NO: 1, comprising culturing the host cell of claim 22 or claim 23 under conditions suitable for protein expression and isolating the peptide.

5

25. A composition comprising the peptide of any of claims 13-18 and a pharmaceutically acceptable carrier.

26. A method of generating antibodies against HIV in a mammal, comprising administering the composition of claim 25 to the mammal.

10

27. The method of claim 26, wherein the composition further comprises an adjuvant.

28. A method of conferring immunity against HIV in a mammal, comprising administering the composition of claim 25 to the mammal.

15

29. The method of claim 28, wherein the composition further comprises an adjuvant.

30. The method of claim 28, comprising administering the composition to the mammal by injection.

20

31. The method of any of claims 26-30, wherein the mammal is selected from the group consisting of: a human, a non-human primate, a dog, a rabbit, a guinea pig, and a mouse.

32. A subunit vaccine comprising the peptide of any of claims 1-3.

25

33. A subunit vaccine comprising the peptide of any of claims 13-18.

34. A nucleic acid vaccine comprising the nucleic acid of claim 19.

30

35. An isolated peptide comprising an amino acid sequence having 90 % or greater identity to the amino acid sequence depicted in SEQ ID NO: 3, SEQ ID NO: 4, or SEQ ID NO: 5.

36. The peptide of claim 35, wherein the peptide comprises an amino acid sequence having 98 % or greater identity to an amino acid sequence depicted in SEQ ID NO: 3, SEQ ID NO: 4, or SEQ ID NO: 5.

5 37. The peptide of claim 36, wherein the peptide comprises an amino acid sequence having 99 % or greater identity to an amino acid sequence depicted in SEQ ID NO: 3, SEQ ID NO: 4, or SEQ ID NO: 5.

38. The peptide of claim 37, wherein the peptide comprises the amino acid sequence depicted in SEQ ID NO: 3, SEQ ID NO: 4, or SEQ ID NO: 5.

10

39. An isolated nucleic acid sequence comprising a nucleic acid sequence encoding the amino acid sequence of any of claims 35-38.

15 40. The nucleic acid of claim 39, wherein the nucleic acid sequence encodes the amino acid sequence depicted in SEQ ID NO: 3, SEQ ID NO: 4, or SEQ ID NO: 5.

41. A kit comprising (a) a composition comprising the peptide of any of claims 1-3 and a pharmaceutically acceptable carrier and (b) instructions for administration of the composition to a mammal.

20

42. A kit comprising (a) the composition of claim 25 and (b) instructions for administration of the composition to a mammal.

25 43. A kit comprising (a) a composition comprising the nucleic acid of claim 19 and a pharmaceutically acceptable carrier and (b) instructions for administration of the composition to a mammal.

44. The peptide of any of claims 7-12, wherein the HIV-1 strain is isolated from an individual with an acute HIV-1 infection.

30

45. The peptide of any of claims 7-12, wherein the HIV-1 strain is isolated from an individual with a chronic HIV-1 infection.

46. The peptide of claim 1, wherein the HIV Env protein is mutated in the leader sequence.

5 47. The peptide of claim 46, wherein the native signal peptide is replaced with a tPA signal peptide.

48. The peptide of claim 47, wherein the tPA signal peptide comprises a sequence selected from the group consisting of: SEQ ID NO: 21 and SEQ ID NO: 22.

10

49. An isolated peptide comprising an amino acid sequence having 90 % or greater identity to the amino acid sequence depicted in SEQ ID NO: 7 or SEQ ID NO: 9.

15 50. The peptide of claim 49, wherein the peptide comprises an amino acid sequence having 98 % or greater identity to an amino acid sequence depicted in SEQ ID NO: 7 or SEQ ID NO: 9.

20 51. The peptide of claim 50, wherein the peptide comprises an amino acid sequence having 99 % or greater identity to an amino acid sequence depicted in SEQ ID NO: 7 or SEQ ID NO: 9.

52. The peptide of claim 51, wherein the peptide comprises the amino acid sequence depicted in SEQ ID NO: 7 or SEQ ID NO: 9.

25 53. An isolated nucleic acid sequence comprising a nucleic acid sequence encoding the amino acid sequence of any of claims 49-52.

54. The nucleic acid of claim 53, wherein the nucleic acid sequence is selected from the group consisting of: SEQ ID NO: 3 and SEQ ID NO: 5.

30

55. The nucleic acid of claim 19, wherein the isolated nucleic acid comprises the nucleic acid sequence depicted in SEQ ID NO: 20.

56. The method of claim 27 or claim 29, wherein the adjuvant comprises a liposome formulation.

57. The method of claim 56, wherein the liposome formulation comprises one or more of:
5 dimyristoyl phosphatidylcholine, dimyristoyl phosphatidylglycerol, cholesterol, and phospholipid.

58. The method of claim 57, wherein the liposome formulation comprises phospholipid A.

10

59. The method of claim 26, wherein the antibodies generated in the mammal are antibodies that compete with the peptide comprising the truncated HIV Env protein for binding integrin $\alpha 4\beta 7$.

15 60. The isolated peptide of claim 1, wherein the peptide binds integrin $\alpha 4\beta 7$.

20

FIG. 1

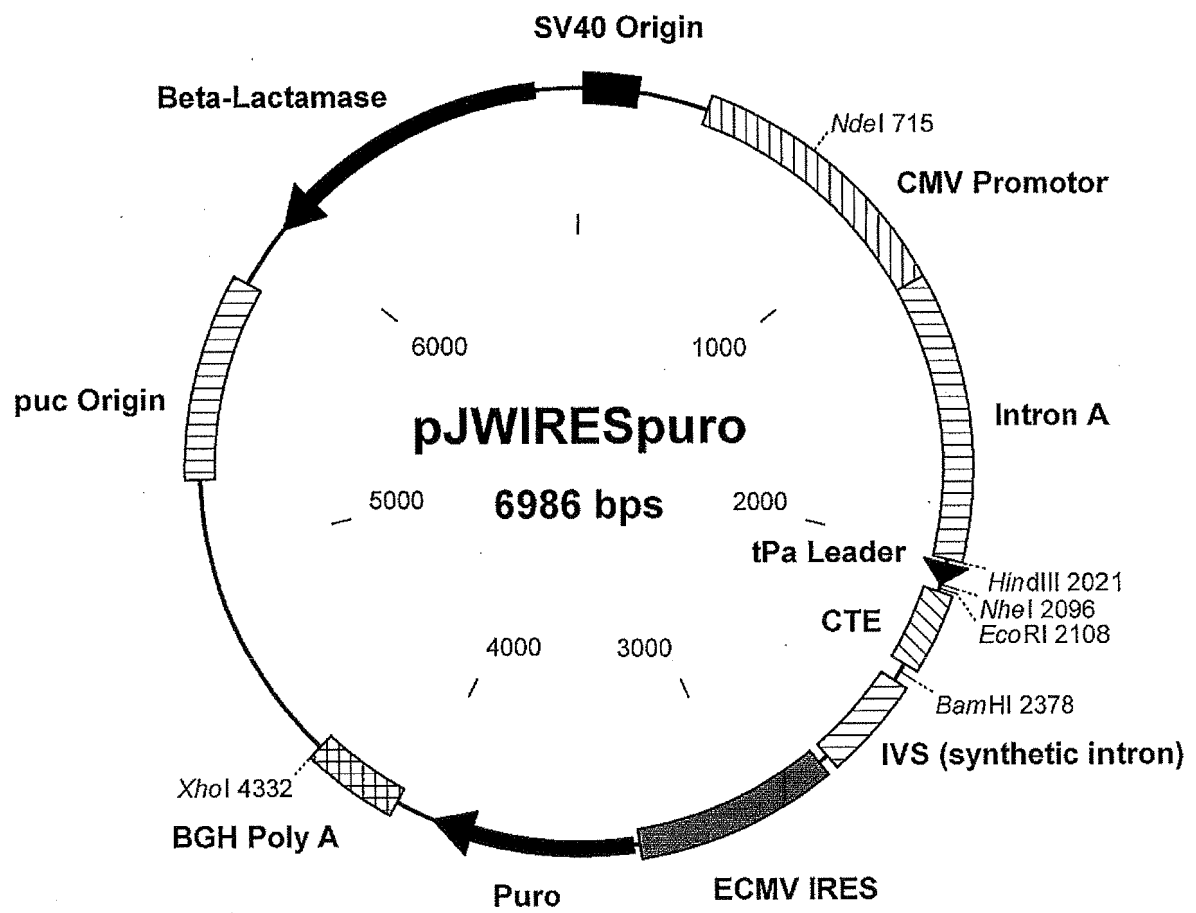


FIGURE 2

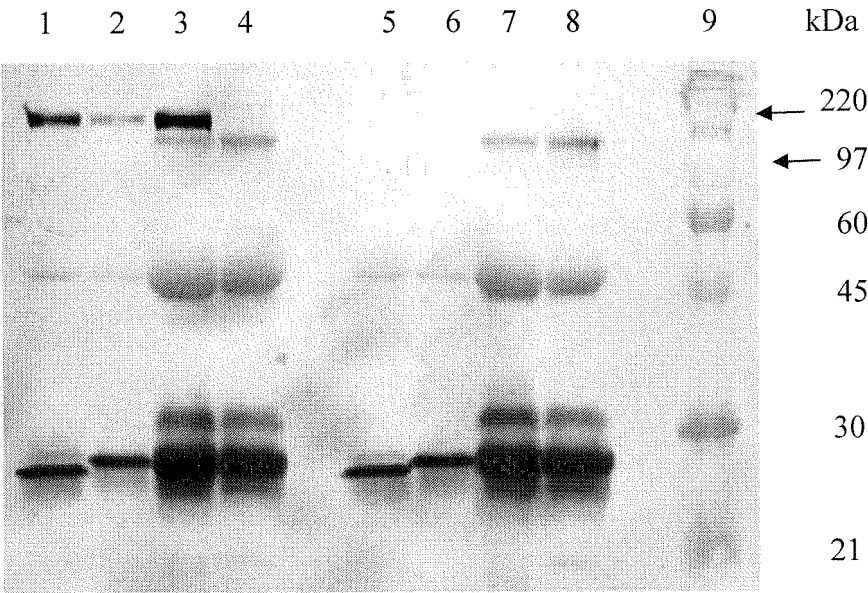


FIGURE 3

Nucleic acid sequence of HIV-1 Ba-L gp140 DC 4E10 (SEQ ID NO: 6)

ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTT
TCGGCTAGC GAGGAAAAGCTGTGGGTGACCGTGTACTACGGCGTGCCCGTGTGGAAAGAG
GCCACCACCACCTGTTCTGCGCCAGCGACGCCAAGGCCTACGACACCGAGGTGCACAAC
GTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCAAGGAAGTGGAGCTGGAA
AACGTGACCGAGAACTTCAACATGTGGAAGAACAACATGGTGGAGCAGATGCACGAGGAC
ATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGTGCGTG
ACCCTGAACTGCACCGACCTGCGGAACGCCACCAACGGCAACGACACCAACACCACCTCC
AGCAGCCGGGAGATGATGGGCGGAGGGGAGATGAAGAACTGCAGCTTCAAGATCACCACC
AACATCCGGGGCAAGGTGCAGAAAGAGTACGCCCTGTTCTACGAGCTGGACATCGTGCCC
ATCGACAACAACAGCAACAACCGGTACAGGCTGATCAGCTGCAACACCAGCGTGATCACC
CAGGCCTGCCCCAAGATCAGCTTCGAGCCCATCCCCATCCACTACTGCGCCCCCTGCCGGC
TTCGCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCAAGGGCCCCTGCAGCAACGTG
AGCACCGTGCAGTGCACCCACGGCATCCGGCCCCGTGGTGTCCACCCAGCTGCTGCTGAAC
GGCAGCCTGGCCGAGGAAGAGGTGGTGATCAGAAGCGAGAACTTCGCCGACAACGCCAAG
ACCATCATCGTGCACTGAACGAGAGCGTGGAGATCAACTGCACCCGGCCCAACAACAAC
ACCCGGAAGAGCATCCACATCGGCCCTGGCAGGGCCCTGTACACCACCGGCGAGATCATC
GGCGACATCCGGCAGGCCCACTGCAACCTGAGCCGGGCCAAGTGGAACGACACCCTGAAC
AAGATCGTGATCAAGCTGCGGGAGCAGTTCGGCAACAAGACCATCGTGTTCAAGCACAGC
AGCGGCGGAGACCCGAGATCGTGACCCACAGCTTCAACTGTGGCGGCGAGTTCTTCTAC
TGCAACAGCACCCAGCTGTTCAACAGCACCTGGAATGTGACCGAGGAAAGCAACAACACC
GTGGAGAACAACACCATCACCTGCCCTGCCGGATCAAGCAGATCATCAATATGTGGCAG
AAAGTGGGCAGAGCCATGTACGCCCTCCCATCCGGGGCCAGATCCGGTGCAGCAGCAAC
ATCACCGGCCTGCTGCTGACCCGGGACGGCGGACCCGAGGCCAACAAGACCGAGGTGTTT

FIGURE 3 (cont.)

CGGCCTGGCGGCGGAGATATGCGGGACAACTGGCGGAGCGAGCTGTACAAGTACAAGGTG
GTGAAGATCGAGCCCCTGGGCGTGGCCCCACCAAGGCCAAGAGCCGGGTGGTGCAGCGG
GAGAAGAGCGCCGTGGGCATCGGCGCCGTGTTTCTGGGCTTCCTGGGAGCCGCCGGAAGC
ACCATGGGAGCCGCCAGCATGACCCTGACCGTGCAGGCCCGGCTGCTGCTGTCCGGCATC
GTGCAGCAGCAGAACAACCTGCTGCGGGCCATTGAGGCACAGCAGCATCTGCTGCAGCTG
ACCGTGTGGGGCATTAAAGCAGCTGCAGGCCAGGGTGTGGCCGTGGAGAGATACCTGCGG
GATCAGCAGCTGCTGGGGATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCACCGCCGTG
CCCTGGAACGCCAGCTGGTCCAACAAGAGCCTGAACAAAATCTGGGACAACATGACCTGG
ATGGAATGGGACCGCGAGATCAACAACCTACACCAGCATCATCTACAGCCTGATCGAGGAA
AGCCAGAACCAGCAGGAAAAGAACGAGCAGGAACTGCTGGAAGTGGACAAGTGGGCCAGC
CTGTGGAAGTGGTTCGACATCACCGAGTGGCTGTGGTACATCAAGAAGAAGAAGTGA

FIGURE 4

Amino acid sequence of HIV-1 Ba-L gp140 DC 4E10 (SEQ ID NO: 7)

MDAMKRGLCVLLCGAVFVSAS EEKLWVTVYYGVPVWKEATTLFCASDAKAYDTEVHN
VWATHACVPTDPNPQEVELENTENFNMWKNNMVEQMHEDIISLWDQSLKPCVKLTPLCV
TLNCTDLRNATNGNDTNTTSSSREMMGGGEMKNCSFKITTNIRGKVQKEYALFYELDIVP
IDNNSNNRYRLISCNSTVITQACPKISFEPIPIHYCAPAGFAILKCKDKKFNGKGPCSNV
STVQCTHGIRPVVSTQLLNGSLAEEEVIRSENFADNAKTIIVQLNESVEINCTRPNNN
TRKSIHIGPGRALYTTGEIIGDIRQAHCNLSRAKWNDTLNKIVIKLREQFGNKTIVFKHS
SGGDPEIVTHSFNCGGEFFYCNSTQLFNSTWNVTEESNNTVENNTITLPCRIKQIINMWQ
KVGRAMYAPPIRGQIRCSSNITGLLLTRDGGPEANKTEVFRPGGGDMRDNRSELYKYKV
VKIEPLGVAPTAKASRVVQREKSAVGIGAVFLGFLGAAGSTMGAASMTLTVQARLLLSGI
VQQQNNLLRAIEAQQHLLQLTVWGIKQLQARVLAVERYLRDQQLGIWGC SGKLICTTAV
PWNASWSNKS LNKIWDNMTWMEWDREINNYTSIIYSLIEESQNQQEKNEQEELLELDK WAS
LWNWFDITEWLWYIKKKK

FIGURE 5

Nucleic acid sequence of HIV-1 Ba-L gp145 (SEQ ID NO: 8)

ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTT
ACTACCACAGAGGCTAGC GAGGAAAAGCTGTGGGTGACCGTGTAACGGCGTGCCCGTG
TGGAAGAGGGCCACCACCACCTGTTCTGCGCCAGCGACGCCAAGGCCTACGACACCGAG
GTGCACAACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAGGAAGTG
GAGCTGGAAAACGTGACCGAGAACTTCAACATGTGGAAGAACAACATGGTGGAGCAGATG
CACGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTGACCCCC
CTGTGCGTGACCCTGAACTGCACCGACCTGCGGAACGCCACCAACGGCAACGACACCAAC
ACCACCTCCAGCAGCCGGGAGATGATGGGCGGAGGGGAGATGAAGAACTGCAGCTTCAAG
ATCACCACCAACATCCGGGGCAAGGTGCAGAAAGAGTACGCCCTGTTCTACGAGCTGGAC
ATCGTGCCCATCGACAACAACAGCAACAACCGGTACAGGCTGATCAGCTGCAACACCAGC
GTGATCACCCAGGCCTGCCCCAAGATCAGCTTCGAGCCCATCCCCATCCACTACTGCGCC
CCTGCCGGCTTCGCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCAAGGGCCCCTGC
AGCAACGTGAGCACCGTGCAGTGCACCCACGGCATCCGGCCCCGTGGTGTCCACCCAGCTG
CTGCTGAACGGCAGCCTGGCCGAGGAAGAGGTGGTGATCAGAAGCGAGAACTTCGCCGAC
AACGCCAAGACCATCATCGTGCACTGAACGAGAGCGTGGAGATCAACTGCACCCGGCCC
AACAACAACACCCGGAAGAGCATCCACATCGGCCCTGGCAGGGCCCTGTACACCACCGGC
GAGATCATCGGCGACATCCGGCAGGCCCACTGCAACCTGAGCCGGGCCAAGTGGAACGAC
ACCCTGAACAAGATCGTGATCAAGCTGCGGGAGCAGTTCGGCAACAAGACCATCGTGTTT
AAGCACAGCAGCGGCGGAGACCCGAGATCGTGACCCACAGCTTCAACTGTGGCGGCGAG
TTCTTCTACTGCAACAGCACCCAGCTGTTCAACAGCACCTGGAATGTGACCGAGGAAAGC
AACAACACCGTGGAGAACAACACCATCACCTGCCCTGCCGGATCAAGCAGATCATCAAT
ATGTGGCAGAAAGTGGGCAGAGCCATGTACGCCCTCCCATCCGGGGCCAGATCCGGTGC
AGCAGCAACATACCGGCCTGCTGCTGACCCGGGACGGCGGACCCGAGGCCAACAAGACC

FIGURE 5 (cont.)

GAGGTGTTCCGGCCTGGCGGCGGAGATATGCGGGACAACCTGGCGGAGCGAGCTGTACAAG
TACAAGGTGGTGAAGATCGAGCCCCTGGGCGTGGCCCCACCAAGGCCAAGAGCCGGGTG
GTGCAGCGGGAGAAGAGCGCCGTGGGCATCGGCGCCGTGTTTCTGGGCTTCCTGGGAGCC
GCCGGAAGCACCATGGGAGCCGCCAGCATGACCCTGACCGTGCAGGCCCGGCTGCTGCTG
TCCGGCATCGTGCAGCAGCAGAACAACCTGCTGCGGGCCATTGAGGCACAGCAGCATCTG
CTGCAGCTGACCGTGTGGGGCATTAAAGCAGCTGCAGGCCAGGGTGCTGGCCGTGGAGAGA
TACCTGCGGGATCAGCAGCTGCTGGGGATCTGGGGCTGCAGCGGCAAGCTGATCTGCACC
ACCGCCGTGCCCTGGAACGCCAGCTGGTCCAACAAGAGCCTGAACAAAATCTGGGACAAC
ATGACCTGGATGGAATGGGACCGCGAGATCAACAACCTACACCAGCATCATCTACAGCCTG
ATCGAGGAAAGCCAGAACCAGCAGGAAAAGAACGAGCAGGAACTGCTGGAACCTGGACAAG
TGGGCCAGCCTGTGGAACCTGGTTCGACATCACCGAGTGGCTGTGGTACATCAAGAAGAAG
AAGTGA

FIGURE 6

Amino acid sequence of HIV-1 Ba-L gp145 (SEQ ID NO: 9):

MDAMKRGLCVLLLCGAVFVTTTEAS EEKLWVTVYYGVPVWKEATTTLCASDAKAYDTE
VHNVWATHACVPTDPNPQEVELENTENFNMWKNNMVEQMHEDIISLWDQSLKPCVKLTP
LCVTLNCTDLRNATNGNDTNTTSSSREMMGGGEMKNCSFKITTNIRGKVQKEYALFYELD
IVPIDNNSNNRYRLSCNTSVITQACPKISFEPIPIHYCAPAGFAILKCKDKKFNGKGPC
SNVSTVQCTHGIRPVVSTQLLNGSLAEEEVVIRSENFADNAKTIIVQLNESVEINCTRP
NNNTRKSIHIGPGRALYTTGEIIGDIRQAHCNLSRAKWNDTLNKIVIKLREQFGNKTIVF
KHSSGGDPEIVTHSFNCGGEFFYCNSTQLFNSTWNVTEESNNTVENNTITLPCRIKQIIN
MWQKVGRAMYAPPIRGQIRCSSNITGLLLTRDGGPEANKTEVFRPGGGDMRDNWRSELYK
YKVVKIEPLGVAPTAKSRVVQREKSAVGIGAVFLGFLGAAGSTMGAASMTLTVQARLLL
SGIVQQQNNLLRAIEAQHLLQLTVWGIKQLQARVLAVERYLRDQQLGIWGCSGKLICT
TAVPWNASWSNKSINKIWDNMTWMEWDREINNYTSIIYSLIEESQNQQEKNEQELLELDK
WASLWNWFDITEWLWYIKKKK

FIGURE 7

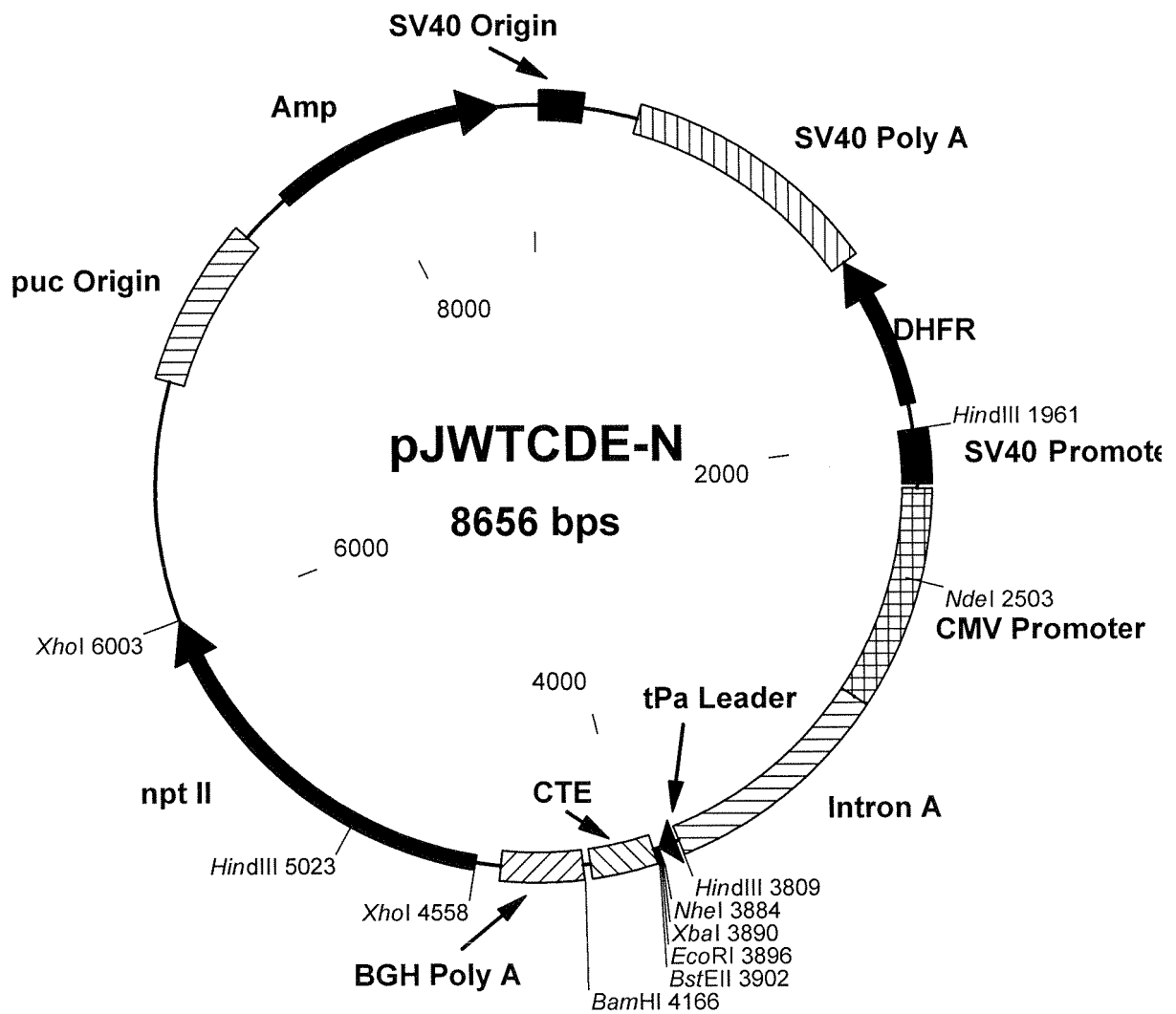


FIGURE 8

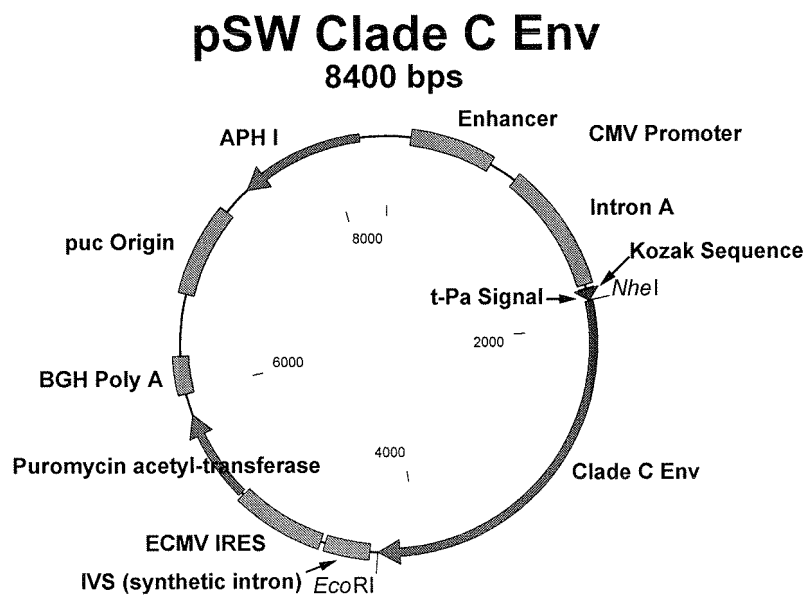


FIGURE 9

HIV-1 C3728v2c6 gp160 Codon Optimized (SEQ ID NO: 10):

tPa signal is highlighted

ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTT
ACTACCACAGAGGCTAGC AACCTGTGGGTGACCGTGTAACGCGTGCCCGTGTGGAAG
GAGGCCAAGACCACCCTGTTCTGCGCCAGCGACGCCAAGGCCTACGAGAAGGAGGTGCAC
AACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAGGAGATGGAGCTG
CCCAACGTGACCGAGAACTTCAACATGTGGAAGAACGACATGGTGGACCAGATGCACGAG
GACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGTGC
GTGACCCTGAACTGCAACAAGGTGAACAACAACCGCAGCATCGACAACAACAGCACCGAG
GAGGAGATGAAGAACTGCACCTTCAACACCACCACCGAGATCAGGGACAAGAAGAGGACC
CAGCAGGCCCTGTTCTACAAGCTGGACATCGTGCCCTGGGCAACAGCAACGAGAGCTAC
AGGCTGATCAACTGCAACACCAGCACCTGACCCAGGCCTGCCCCAAGGTGACCTTCGAC
CCCATCCCCATCCACTACTGCGCCCCAGCCGGCTACGCCATCCTGAAGTGCAAGGACGAG
AAGTTCAACGGCACCGGCCCTGCAACAACGTGTCCACCGTGAGTGACCCACGGCATC
AAGCCCGTGGTGTCCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGAAGGACATCATC
ATCCGCAGCGAGAACCTGACCAACAACGTGAAGACCATCATGGTGCACCTGAACGAGAGC
GTGGAGATCAACTGTACCAGGCCCAACAACAACACCAGGAAGAGCATCAGGATCGGCCCCA
GGCCAGACCTTCTACGCCACCGGCGGCATCATCGGCAACATCAGGCAGGCCTACTGCACC
ATCAGCAAGAACAAGTGGAACACCACCCTGGAGAGAGTGGCTACCAAGCTGAAGGAGTAC
TTCAAGAACACCACCATCCAGTTCGCCCCCACTCTGGCGGCGACCTGGAGATCACCACC
CACAGCTTCAACTGCAGGGGCGAGTTCTTCTACTGCAACACAAGCCAGCTGTTCAACGGG
ACCAGCACCGGCTTCAGCAACAAGAGCACCGGCAACGAGACATTCACCCTGCCCTGCAGA
ATCAAGCAGATCATCAACATGTGGCAGGAAGTGGGCAGGGCCATGTACGCCCCACCAATC

FIGURE 9 (cont.)

GCCGGGAACATCACCTGCGTGAGCAACATCACCGGCCTGCTGCTGACCAGGGACGGCGGG
GACAACAACACAAGAACCAGACATTCCGGCCAGGCGGCGGAGACATGAGGGACAACCTGG
CGGAGCGAGCTGTACAAGTACAAGGTGGTGGAGGTGAAGCCCCTGGGCGTGGCCCCAAGC
GAAGCCAGGCGGAGAGTGGTGGAAACGGGAGAAGAGGGCCGTGGGCCTGGGCGCCGTCTTT
CTGGGATTCTGGGGGCTGCCGGATCCACCATGGGAGCCGCCAGCATCACCTGACCGTG
CAGGCTAGGCAGCTGCTGTCCGGAATCGTGCAGCAGCAGTCCAACCTGCTGAGGGCTATC
GAGGCTCAGCAGCATATGCTCCAGCTGACCGTGTGGGGCATCAAGCAGCTGCAGACCAAG
GTGCTGGCCATCGAGAGATATCTGCAGGACCAGCAGCTGCTCGGCATCTGGGGCTGCAGC
GGCAAGCTGATCTGCACCACCAACGTGCCCTGGAACAGCAGCTGGTCCAACAAGTCCCTG
AAGGCCATCTGGGACAACATGACCTGGATGCAGTGGGACAGAGAGATCAGCAACTACACC
AACACCATCTACAGGCTGCTGGAGGACAGCCACATCCAGCAGGAGAAGAACGAGAAGGAC
CTGCTGGCCCTGAACAGCTGGAACAACCTGTGGAGCTGGTTCAACATCACCAACTGGCTG
TGGTACATCCGGATCTTCATCATGATCGTGGGCGGCCTGATCGGCCTGAGGATCATCTTC
GCCATCCTGTCCATCGTGAACAGAGTGCGCCAGGGCTACAGCCCACTGAGCTTCCAGACC
CTGACACCCAACCAGCGGGGGCCAGACAGACTGGGCAGGATCGAGGAGGAGGGCGGCGAG
CAGGACAAGGACAGATCCATCAGGCTGGTGAACGGCTTCCTGGCCCTGGCCTGGGACGAC
CTGAGAAGCCTGTGCCTGTTCAGCTACCACAGGCTGAGGGACTTCATCCTGATCGCCACC
AGGGCCGTGGAACCTGCTGGGGCACTCCAGCCTGAGGGGCCTGCAGAGGGGCTGGGAGGCC
CTGAAGTACCTGGGCTCCCTGGGCCAGTATTGGGGCCAGGAGCTGAAGAAGAGCGCCATC
AGACTGCTGGACATCACCGCCATCGCCGTGGCCGAGGGCACCGACAGGATCATCGAGTTC
ATCCAGAGGATCTGCAGGGCCATCCACAACGTGCCCCAACAGGATCAGGCAGGGCTTCGAG
GCCGCCCTGCTGTGATGA

FIGURE 10

HIV-1 C3728v2c6 gp160 DC codon optimized Nucleic acid seq (SEQ ID NO: 11):

tPa signal is highlighted

The gp120/gp41 cleavage site is mutated to prevent cleavage

ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTT
ACTACCACAGAGGCTAGC AACCTGTGGGTGACCGTGTACTACGGCGTGCCCGTGTGGAAG
GAGGCCAAGACCACCCTGTTCTGCGCCAGCGACGCCAAGGCCTACGAGAAGGAGGTGCAC
AACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAGGAGATGGAGCTG
CCCAACGTGACCGAGAACTTCAACATGTGGAAGAACGACATGGTGGACCAGATGCACGAG
GACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGTGC
GTGACCCTGAACTGCAACAAGGTGAACAACAACCGCAGCATCGACAACAACAGCACCGAG
GAGGAGATGAAGAACTGCACCTTCAACACCACCACCGAGATCAGGGACAAGAAGAGGACC
CAGCAGGCCCTGTTCTACAAGCTGGACATCGTGCCCTGGGCAACAGCAACGAGAGCTAC
AGGCTGATCAACTGCAACACCAGCACCTGACCCAGGCCTGCCCCAAGGTGACCTTCGAC
CCCATCCCCATCCACTACTGCGCCCCAGCCGGCTACGCCATCCTGAAGTGCAAGGACGAG
AAGTTCAACGGCACCGGCCCTGCAACAACGTGTCCACCGTGCAAGTGACCCACGGCATC
AAGCCCGTGGTGTCCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGAAGGACATCATC
ATCCGCAGCGAGAACCTGACCAACAACGTGAAGACCATCATGGTGCACCTGAACGAGAGC
GTGGAGATCAACTGTACCAGGCCCAACAACAACACCAGGAAGAGCATCAGGATCGGCCCCA
GGCCAGACCTTCTACGCCACCGGCGGCATCATCGGCAACATCAGGCAGGCCTACTGCACC
ATCAGCAAGAACAAGTGGAACACCACCCTGGAGAGAGTGGCTACCAAGCTGAAGGAGTAC
TTCAAGAACACCACCATCCAGTTCGCCCCCACTCTGGCGGCGACCTGGAGATCACCACC
CACAGCTTCAACTGCAGGGGCGAGTTCTTCTACTGCAACACAAGCCAGCTGTTCAACGGG
ACCAGCACCGGCTTCAGCAACAAGAGCACCGGCAACGAGACATTACCCCTGCCCTGCAGA
ATCAAGCAGATCATCAACATGTGGCAGGAAGTGGGCAGGGCCATGTACGCCCCACCAATC

FIGURE 10 (cont.)

GCCGGGAACATCACCTGCGTGAGCAACATCACCGGCCTGCTGCTGACCAGGGACGGCGGG
GACAACAACACAAGAACCAGACATTCCGGCCAGGCGGCGGAGACATGAGGGACAACCTGG
CGGAGCGAGCTGTACAAGTACAAGGTGGTGGAGGTGAAGCCCCTGGGCGTGGCCCCAAGC
GAAGCCAGGAGCAGAGTGGTGGAAACGGGAGAAGAGCGCCGTGGGCCTGGGCGCCGTCTTT
CTGGGATTCTGGGGGCTGCCGGATCCACCATGGGAGCCGCCAGCATCACCTGACCGTG
CAGGCTAGGCAGCTGCTGTCCGGAATCGTGCAGCAGCAGTCCAACCTGCTGAGGGCTATC
GAGGCTCAGCAGCATATGCTCCAGCTGACCGTGTGGGGCATCAAGCAGCTGCAGACCAAG
GTGCTGGCCATCGAGAGATATCTGCAGGACCAGCAGCTGCTCGGCATCTGGGGCTGCAGC
GGCAAGCTGATCTGCACCACCAACGTGCCCTGGAACAGCAGCTGGTCCAACAAGTCCCTG
AAGGCCATCTGGGACAACATGACCTGGATGCAGTGGGACAGAGAGATCAGCAACTACACC
AACACCATCTACAGGCTGCTGGAGGACAGCCACATCCAGCAGGAGAAGAACGAGAAGGAC
CTGCTGGCCCTGAACAGCTGGAACAACCTGTGGAGCTGGTTCAACATCACCAACTGGCTG
TGGTACATCCGGATCTTCATCATGATCGTGGGCGGCCTGATCGGCCTGAGGATCATCTTC
GCCATCCTGTCCATCGTGAACAGAGTGCGCCAGGGCTACAGCCCCTGAGCTTCCAGACC
CTGACACCCAACCAGCGGGGGCCAGACAGACTGGGCAGGATCGAGGAGGAGGGCGGCGAG
CAGGACAAGGACAGATCCATCAGGCTGGTGAACGGCTTCCTGGCCCTGGCCTGGGACGAC
CTGAGAAGCCTGTGCCTGTTCAGCTACCACAGGCTGAGGGACTTCATCCTGATCGCCACC
AGGGCCGTGGAACCTGCTGGGGCACTCCAGCCTGAGGGGCCTGCAGAGGGGCTGGGAGGCC
CTGAAGTACCTGGGCTCCCTGGGCCAGTATTGGGGCCAGGAGCTGAAGAAGAGCGCCATC
AGACTGCTGGACATCACCGCCATCGCCGTGGCCGAGGGCACCGACAGGATCATCGAGTTC
ATCCAGAGGATCTGCAGGGCCATCCACAACGTGCCCAACAGGATCAGGCAGGGCTTCGAG
GCCGCCCTGCTGTGA

FIGURE 11

HIV-1 C06838v1c48 gp160 Codon optimized Nucleic acid seq. (SEQ ID NO: 12):

tPa signal is highlighted

```
ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTT
ACTACCACAGAGGCTAGC AACCTGTGGGTGACCGTGTACTACGGCGTGCCCGTGTGGAGA
GAGGCCAAGACCACCTGTTCTGCGCCAGCGACGCCAAGGCCTACGAAAAGGAGGTGCAC
AACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCGACCCCGAGGAGATGGTGCTG
AAGAACGTGACCGAGAACTTCAACATGTGGAAGAACGACATGGTGGAGCAGATGCACGAG
GACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGTGC
GTGACCTGACCTGCACCAACGTGAACGCCACCGACAACGTGACCTACAAGGAGAAGATG
GAGGGCGAGATCAAGAACTGCAGCTTCAACATCACCACCGAGATCAGGGACAAGAAGAGG
AAGGTCCACGCCCTGTTCTACAGGCTGGACGTGGTGCAGCTGAACAACAGCAACGAGTAC
ATCCTGATCAACTGCAACACCAGCGCCATCACCCAGGCCTGCCCCAAGGTGTCCTTCGAC
CCCATCCCCATCCACTACTGCGCCCCAGCCGGCTACGCCATCCTGAAGTGCAACGACAAG
CCCTTCAACGGCACCCGGCCCATGCCACAACGTGTCTACCGTGCAGTGCACCCACGGCATC
AAGCCCGTGATCAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGAAGGAGGTGATC
ATCCGCTTCGAGAACCTGACCGACAACGCCAAGACCATCATCGTGCAGCTGAACCAGAGC
ATCGAGATCAAGTGCATCAGGCCCAACAACAACACCAGAGAGAGCATCAGGATCGGCCCA
GGCCAGGCCTTCTACGCCACCAGGGACATCATCGGCGACATCAGAAGGGCCTACTGCACC
ATCGAGACAGAGAGGTGGAAGGAGACACTGGGCAACGTACAGAGAAGCTGAAGGAGTAC
TTCCCCAACCAACATCAGCTTTGCTCCCTCCTCTGGCGGGGACCTGGAGATCACCACC
CACAGCTTCAACTGCAGGGGCGAGTTCTTCTACAGAAACACCAGCAAGCTGTTCAACAAC
AACGACACCGAGAACAACTGACCATCACCCCTGCCCTGCAGAATCAAGCAGATCGTGAAC
ATGTGGCAGGGCGTGGGCAGGGCCATGTACGCCCCACCCATCGAGGGCAACATCACCTGC
AAGTCCAACATCACCGGCCTGCTGCTGACCAGAGACGGCGGCAACGGCACAGAGAGGAAC
```

FIGURE 11 (cont.)

GAGACATTCAGGCCAGGCGGCGGAGACATGAGGAACAACTGGCGGAGCGAGCTGTACAAG
TACAAGGTGGTGGAGATCAAGCCCCTGGGGATCGCCCCAACCGGCGCCAAGAGAAGAGTG
GTGGAGAGAGAGAAGAGGGCCGTGCGCATGGGCGCCGTGTTTCTGGGCTTCCTGGGAGCC
GCCGGAAGCACCATGGGAGCCGCCAGCCTGACCCTGACAGTGCAGGCCAGACAGGTGCTG
TCCGGGATCGTGCAGCAGCAGAGCAACCTGCTGAGAGCTATCGAGGCTCAGCAGCACATG
CTGCAGCTGACCGTGTGGGGCATCAAGCAGCTGCAGACCAGGGTGCTGGCCATCGAGAGA
TACCTGAAAGATCAGCAGCTGCTCGGCCTCTGGGGCTGCAGCGGCAAGCTGATCTGCACC
ACAGCCGTGCCATGGAATAGCAGCTGGTCCAACAAGAGCGAGATCGACATCTGGAACAAC
ATGACCTGGATGCAGTGGGACAGAGAGATCAGCAACTACACCAACACCATCTACAGACTG
CTGGAGGACAGCCAGACCCAGCAGGAGAAGAACGAGAAGGACCTGCTGGCCCTGGACAGC
TGGAAGAACCTGTGGAACCTGGTTTAACATCACCAAGTGGCTGTGGTACATCAAGATCTTC
ATCATCATCGTGGGCGGCCTGATCGGCCTGAGGATCATCTTCGCCGTGCTGTCCATCGTG
AACAGAGTGCGCCAGGGGTACAGCCCACTGAGCTTCCAGACCCTGATCCCCAGCCCCAGA
GGCCCCGACAAGCTGGGCAGGATCGAGGAGGAGGGCGGCGAGCAGGACAAGGGCAGATCC
GTGAGGCTGGTGAACGGCTTCCTGGCCCTGGCCTGGGACGACCTGAGAAGCCTGTGCCTG
TTCAGCTATCACCAGCTGAGGGACTTCATCCTGATCGTGGCCAGAGCCGTGGGCCTGCTG
GGCAGGTCCAGCCTGAGGGGCCTGCAGAGAGGCTGGGAGATCCTGAAGTACCTGGGCGGG
CTGGTGCAGTACTGGGGCCTGGAGCTGAAGAAGAGCGCCGTGAGCCTGTTCGACACAATC
GCCATCGCCGTGACCGAGGGCACCGACAGGATCATCGAGCTGATCCAGAGGTCCTGCAGA
GCCATCAGGAACGTGCCACCAAGGATCAGGCAGGGCTTCGAGGCCGCCCTGCAGTGA

FIGURE 12

HIV-1 C06838v1c48 gp160 DC codon optimized Nucleic acid seq (SEQ ID NO: 13)

tPa signal is highlighted

The gp120/gp41 cleavage site is mutated to prevent cleavage

ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTT
ACTACCACAGAGGCTAGC AACCTGTGGGTGACCGTGTAACGGCGTGCCCGTGTGGAGA
GAGGCCAAGACCACCCTGTTCTGCGCCAGCGACGCCAAGGCCTACGAAAAGGAGGTGCAC
AACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCGACCCCGAGGAGATGGTGCTG
AAGAACGTGACCGAGAACTTCAACATGTGGAAGAACGACATGGTGGAGCAGATGCACGAG
GACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTGACCCCTGTGC
GTGACCCTGACCTGCACCAACGTGAACGCCACCGACAACGTGACCTACAAGGAGAAGATG
GAGGGCGAGATCAAGAACTGCAGCTTCAACATCACCACCGAGATCAGGGACAAGAAGAGG
AAGGTCCACGCCCTGTTCTACAGGCTGGACGTGGTGCAGCTGAACAACAGCAACGAGTAC
ATCCTGATCAACTGCAACACCAGCGCCATCACCAGGCCTGCCCCAAGGTGTCCTTCGAC
CCCATCCCCATCCACTACTGCGCCCCAGCCGGCTACGCCATCCTGAAGTGCAACGACAAG
CCCTTCAACGGCACCGGCCCATGCCACAACGTGTCTACCGTGCAGTGCACCCACGGCATC
AAGCCCGTGATCAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGAAGGAGGTGATC
ATCCGCTTCGAGAACCTGACCGACAACGCCAAGACCATCATCGTGCAGCTGAACCAGAGC
ATCGAGATCAAGTGCATCAGGCCCAACAACAACACCAGAGAGAGCATCAGGATCGGCCCCA
GGCCAGGCCTTCTACGCCACCAGGGACATCATCGGCGACATCAGAAGGGCCTACTGCACC
ATCGAGACAGAGAGGTGGAAGGAGACACTGGGCAACGTACAGAGAAGCTGAAGGAGTAC
TTCCCCAACCAACATCAGCTTTGCTCCCTCCTCTGGCGGGGACCTGGAGATCACCACC
CACAGCTTCAACTGCAGGGGCGAGTTCTTCTACAGAAACACCAGCAAGCTGTTCAACAAC
AACGACACCGAGAACAACCTGACCATCACCTGCCCTGCAGAATCAAGCAGATCGTGAAC

FIGURE 12 (cont.)

ATGTGGCAGGGCGTGGGCAGGGCCATGTACGCCCCACCCATCGAGGGCAACATCACCTGC
AAGTCCAACATCACCGGCCTGCTGCTGACCAGAGACGGCGGCAACGGCACAGAGAGGAAC
GAGACATTCAGGCCAGGCGGCGGAGACATGAGGAACAACCTGGCGGAGCGAGCTGTACAAG
TACAAGGTGGTGGAGATCAAGCCCCTGGGGATCGCCCCAACCGGCGCCAAGAGCAGAGTG
GTGGAGAGAGAGAAGAGCGCCGTCGGCATGGGCGCCGTGTTTCTGGGCTTCCTGGGAGCC
GCCGGAAGCACCATGGGAGCCGCCAGCCTGACCCTGACAGTGCAGGCCAGACAGGTGCTG
TCCGGGATCGTGCAGCAGCAGAGCAACCTGCTGAGAGCTATCGAGGCTCAGCAGCACATG
CTGCAGCTGACCGTGTGGGGCATCAAGCAGCTGCAGACCAGGGTGCTGGCCATCGAGAGA
TACCTGAAAGATCAGCAGCTGCTCGGCCTCTGGGGCTGCAGCGGCAAGCTGATCTGCACC
ACAGCCGTGCCATGGAATAGCAGCTGGTCCAACAAGAGCGAGATCGACATCTGGAACAAC
ATGACCTGGATGCAGTGGGACAGAGAGATCAGCAACTACACCAACACCATCTACAGACTG
CTGGAGGACAGCCAGACCCAGCAGGAGAAGAACGAGAAGGACCTGCTGGCCCTGGACAGC
TGGAAGAACCTGTGGAACCTGGTTTAACATCACCAAGTGGCTGTGGTACATCAAGATCTTC
ATCATCATCGTGGGCGGCCTGATCGGCCTGAGGATCATCTTCGCCGTGCTGTCCATCGTG
AACAGAGTGCGCCAGGGGTACAGCCCCACTGAGCTTCCAGACCCTGATCCCCAGCCCCAGA
GGCCCCGACAAGCTGGGCAGGATCGAGGAGGAGGGCGGCGAGCAGGACAAGGGCAGATCC
GTGAGGCTGGTGAACGGCTTCCTGGCCCTGGCCTGGGACGACCTGAGAAGCCTGTGCCTG
TTCAGCTATCACCAGCTGAGGGACTTCATCCTGATCGTGGCCAGAGCCGTGGGCCTGCTG
GGCAGGTCCAGCCTGAGGGGCCTGCAGAGAGGCTGGGAGATCCTGAAGTACCTGGGCGGG
CTGGTGCAGTACTGGGGCCTGGAGCTGAAGAAGAGCGCCGTGAGCCTGTTCGACACAATC
GCCATCGCCGTGACCGAGGGCACCGACAGGATCATCGAGCTGATCCAGAGGTCCTGCAGA
GCCATCAGGAACGTGCCACCAAGGATCAGGCAGGGCTTCGAGGCCGCCCTGCAGTGA

FIGURE 13

HIV-1 C06980v1c3 gp160 codon optimized nucleic acid seq. (SEQ ID NO: 14):

tPa signal is highlighted

ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTT
ACTACCACAGAGGCTAGC AACCTGTGGGTGACCGTGTAACGGCGTGCCCGTGTGGAGA
GAGGCCAAGACCACCCTGTTCTGCGCCAGCGACGCCAAGGCCTACGAGAGGGAGGTGCAC
AACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCGACCCACAGGAAATCTTCTG
GGCAAGAACGTGACCGAGAAGTTCAACATGTGGGAGAACTACATGGTGGACCAGATGCAC
GAGGACATCATCAGCCTGTGGGACCAGAGCCTGCAGCCCTGCGTGAAGCTGACCCCCCTG
TGCATCACCTGAACTGCACCGACTTCACCGCCCAACGGCAGCACCGTGACGACAAC
AACGCCACCGCCAACAGCACCAACGAGATCAAGAACTGCAGCTTCAACATCATCTCCGAG
CTGAGGGACAAGAGGAAGAAGGAGAACGCCCTGTTCAACAACCTGGACATCGTGACAGCTG
GACGGCAACTCCAGCCTGTACAGACTGATCAACTGCAACACCAGCATCATCAAGCAGGCC
TGTCCCAAGATCAGCTTCGACCCCATCCCCATCCACTACTGCGCCCCAGCCGGCTTCGTG
ATCCTGACCTGCAACAACGAGACATTCAACGGCACCGGCCCATGCAACAACGTGTCCGCT
GTGCAGTGTACCCACGGGATCAAGCCCGTGGTGTCCACCCAGCTGCTGCTGAACGGCAGC
CTGGCCAAGGGCGAGATCATGATCAGAAGCGAGAACATCACCAACAACGTGAAGACCATC
ATCGTGCACTTCAACAAGAGCGTGGAGATCGTGTGCACCAGGCCCAACAACAACACCAGG
AAGAGCATCAGGATCGGCCAGGCCAGACCTTCTACGCCACCGGCGACATCATCGGCGAC
ATCCGGCAGGCCTACTGCAGCATCAACGAGAGCAACTGGAACATCACCTGCAGAGGGTG
TCCAAGAAGCTGGCCGAGCACTTCCCCAACAGAACCATCCAGTTCGAGAGCCCCTCTGGC
GGCGACCTGGAGATCACCATGCACAGCTTCAACTGCAGGGGCGAGTTCTTCTACTGCAAC
ACATCCAAGCTGTTCAAGAGCACCTACCACCCCAACGGCACCTACAACCTGAACGGCACA
AACAGCACCTGATCATCCCATGCAGGATCAAGCAGATCATCAACATGTGGCAGGGCGTG

FIGURE 13 (cont.)

GGCAAGGCCATCTACGCCAGCCCAATCGCCGGCAGCATCACCTGCAGAAGCAACATCACC
GGCCTGCTGCTGACCAGGGACGGCGGCGACACCAACGACACCGAGATCTTCAGGCCCACC
GGCGGAGACATGAGGGACAACCTGGCGGAGCGAGCTGTACAAGTACAAGATCGTGGAGATC
AAGCCCCTGGGCGTGGCTCCAACCGAGGCCAAGAGAAGGGTGGTGAAGAGAGAGAAGAGG
GCCGTCACCATCGGCGCCGTGTTTCTGGGCTTCCTGGGAGCCGCCGAAGCACCATGGGA
GCCGCCTCCATCACCTGACCGTGCAGGCCAGGCAGCTGCTGTCCGGGATCGTGCAGCAG
CAGAGCAACCTGCTGCGGGCCATCGAAGCTCAGCAGCACATGCTGCAGCTGGCAGTCTGG
GGCATCAAGCAGCTGCAGGCCAGAGTGCTGGCCATCGAGAGATACCTGAAGGATCAGCAG
CTGCTCGGCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCACCGCCGTGCCCTGGAAC
AACAGCTGGTCCAACAGGACCCAGGACGAGATCTGGAAGAACCTGACCTGGATGGAGTGG
GACAGAGAGATCAGCAACTACACCAACACAATCTACGAGCTGCTGGAGGTGTCCCAGAGC
CAGCAGGAGAGGAACGAGAAGGACCTGCTGGCCCTGGACAGCTGGAACAACCTGTGGAAC
TGGTTCGACATCAGCAACTGGCTGTGGTACATCAAGATCTTCATCATGATCGTGGGCGGC
CTGATCGGCCTGAGGATCATCTTCGCCGTGCTGTCCATCGTGAACAGAGTGCGGCAGGGC
TACAGCCCCCTGAGCTTCCAGACCCTGATCCCCAACCCAGAGAGAGCCCGACAGGCCCCGC
AGGATCGAGGAGGAGGGCGGCGAGCAGGACAAGGATAGATCTATCAGACTGGTGTCCGGC
TTCCTGGCCCTGGCCTGGGACGACCTGAGAAGCCTGTGCATCTTCCTGTACCACCACCTG
AGGGACCTGATCCTGATCGCCGCTAGAGCCACAGAGCTGCTGGGGAGGTCCAGCCTGAGG
GGCCTGCAGAGAGGCTGGGAGGCCCTGAAGTACCTGGGCAGCCTGGTGCAGTACTGGGGC
CTGGAGATCAAGAAGAGCGCCATCAACCTGCTGGACACAATCGCCATCGCCGTGGCCGAG
GGCACCGACAGGATCATCGAGATCGTGCAGAGGGCCTGTAGGGCCGTGCTGAACATCCTG
AGAAGGATCCGGCAGGGCCTGGAGGCTGCCCTGCAGTGA

FIGURE 14

HIV-1 C06980v1c3 gp160 DC codon optimized Nucleic acid seq (SEQ ID NO: 15):

tPa signal is highlighted

The gp120/gp41 cleavage site is mutated to prevent cleavage

ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTT
ACTACCACAGAGGCTAGC AACCTGTGGGTGACCGTGTACTACGGCGTGCCCGTGTGGAGA
GAGGCCAAGACCACCTGTTCTGCGCCAGCGACGCCAAGGCCTACGAGAGGGAGGTGCAC
AACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCGACCCACAGGAAATCTTTCTG
GGCAAGAACGTGACCGAGAAGTTCAACATGTGGGAGAACTACATGGTGGACCAGATGCAC
GAGGACATCATCAGCCTGTGGGACCAGAGCCTGCAGCCCTGCGTGAAGCTGACCCCCCTG
TGCATCACCTGAACTGCACCGACTTCACCGCCCAACGGCAGCACCGTGTACGACAAC
AACGCCACCGCCAACAGCACCAACGAGATCAAGAACTGCAGCTTCAACATCATCTCCGAG
CTGAGGGACAAGAGGAAGAAGGAGAACGCCCTGTTCAACAACCTGGACATCGTGCAGCTG
GACGGCAACTCCAGCCTGTACAGACTGATCAACTGCAACACCAGCATCATCAAGCAGGCC
TGTCCAAGATCAGCTTCGACCCCATCCCCATCCACTACTGCGCCCCAGCCGGCTTCGTG
ATCCTGACCTGCAACAACGAGACATTCAACGGCACCGGCCCATGCAACAACGTGTCCGCT
GTGCAGTGTACCCACGGGATCAAGCCCGTGGTGTCCACCCAGCTGCTGCTGAACGGCAGC
CTGGCCAAGGGCGAGATCATGATCAGAAGCGAGAACATCACCAACAACGTGAAGACCATC
ATCGTGCACTTCAACAAGAGCGTGGAGATCGTGTGCACCAGGCCCAACAACAACACCAGG
AAGAGCATCAGGATCGGCCCAGGCCAGACCTTCTACGCCACCGGCGACATCATCGGCGAC
ATCCGGCAGGCCTACTGCAGCATCAACGAGAGCAACTGGAACATCACCTGCAGAGGGTG
TCCAAGAAGCTGGCCGAGCACTTCCCCAACAGAACCATCCAGTTCGAGAGCCCCCTCTGGC
GGCGACCTGGAGATCACCATGCACAGCTTCAACTGCAGGGGCGAGTTCTTCTACTGCAAC
ACATCCAAGCTGTTCAAGAGCACCTACCACCCCAACGGCACCTACAACCTGAACGGCACA

FIGURE 14 (cont.)

AACAGCACCTGATCATCCCATGCAGGATCAAGCAGATCATCAACATGTGGCAGGGCGTG
GGCAAGGCCATCTACGCCAGCCCAATCGCCGGCAGCATCACCTGCAGAAGCAACATCACC
GGCCTGCTGCTGACCAGGGACGGCGGCGACACCAACGACACCGAGATCTTCAGGCCCACC
GGCGGAGACATGAGGGACAACCTGGCGGAGCGAGCTGTACAAGTACAAGATCGTGGAGATC
AAGCCCCTGGGCGTGGCTCCAACCGAGGCCAAGAGCAGGGTGGTGAAGAGCGAGAAGAGC
GCCGTCACCATCGGCGCCGTGTTTCTGGGCTTCTGGGAGCCGCCGAAGCACCATGGGA
GCCGCCTCCATCACCTGACCGTGCAGGCCAGGCAGCTGCTGTCCGGGATCGTGCAGCAG
CAGAGCAACCTGCTGCGGGCCATCGAAGCTCAGCAGCACATGCTGCAGCTGGCAGTCTGG
GGCATCAAGCAGCTGCAGGCCAGAGTGCTGGCCATCGAGAGATACCTGAAGGATCAGCAG
CTGCTCGGCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCACCGCCGTGCCCTGGAAC
AACAGCTGGTCCAACAGGACCCAGGACGAGATCTGGAAGAACCTGACCTGGATGGAGTGG
GACAGAGAGATCAGCAACTACACCAACACAATCTACGAGCTGCTGGAGGTGTCCCAGAGC
CAGCAGGAGAGGAACGAGAAGGACCTGCTGGCCCTGGACAGCTGGAACAACCTGTGGAAC
TGGTTCGACATCAGCAACTGGCTGTGGTACATCAAGATCTTCATCATGATCGTGGGCGGC
CTGATCGGCCTGAGGATCATCTTCGCCGTGCTGTCCATCGTGAACAGAGTGCGGCAGGGC
TACAGCCCCCTGAGCTTCCAGACCCTGATCCCCAACAGAGAGAGCCCGACAGGCCCGGC
AGGATCGAGGAGGAGGGCGGCGAGCAGGACAAGGATAGATCTATCAGACTGGTGTCCGGC
TTCCTGGCCCTGGCCTGGGACGACCTGAGAAGCCTGTGCATCTTCCTGTACCACCACCTG
AGGGACCTGATCCTGATCGCCGCTAGAGCCACAGAGCTGCTGGGGAGGTCCAGCCTGAGG
GGCCTGCAGAGAGGCTGGGAGGCCCTGAAGTACCTGGGCAGCCTGGTGCAGTACTGGGGC
CTGGAGATCAAGAAGAGCGCCATCAACCTGCTGGACACAATCGCCATCGCCGTGGCCGAG
GGCACCGACAGGATCATCGAGATCGTGCAGAGGGCCTGTAGGGCCGTGCTGAACATCCTG
AGAAGGATCCGGCAGGGCCTGGAGGCTGCCCTGCAGTGATG

FIGURE 15

HIV-1 C06980v0c22 gp160 codon optimized nucleic acid seq (SEQ ID NO: 16):

tPa signal is highlighted

ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGCTGTGGAGCAGTCTTCGTT
ACTACCACAGAGGCTAGC AACCTGTGGGTGACCGTGTACTACGGCGTGCCCGTGTGGAGA
GAGGCCAAGACCACCCTGTTCTGCGCCAGCGACGCCAAGGCCTACGAGCGGGAGGTGCAC
AACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCGACCCTCAGGAAATCTTCCTG
GGCAAGAACGTGACCGAGAAGTTCAACATGTGGAAGAACTACATGGTGGACCAGATGCAC
GAGGACATCATCAGCCTGTGGGACCAGAGCCTGCAGCCCTGCGTGAAGCTGACCCCCCTG
TGCATCACCTGAACTGCACCGACGTGACCGCCCACAACGGCAGCACCGTGTACGACAAC
AACGCCACCGTGGTGAACAGCACCAACGAGATCAAGAACTGCAGCTTCAACATCACCACC
GAGCTGAGGGACAAGAGGAAGAAGGAGCACGCCCTGTTCAACAACCTGGACATCGTGCAG
CTGGACGGCAACAGCTCCCTGTACAGACTGATCAACTGCAACACCAGCATCATCAAGCAG
GCCTGCCCCAAGATCAGCTTCGACCCCATCCCCATCCACTACTGCGCCCCAGCCGGCTTC
GTGATCCTGAAGTGCAACAACGAGACATTCAACGGCACCGGCCCTGTAACAACGTGTCC
GCTGTGCAGTGCACCCACGGCATCAAGCCCGTGGTGTCCACCCAGCTGCTGCTGAACGGC
AGCCTGGCCAAGGGCGAGATCATGATCAGAAGCGAGAACATCACCGACAACGTGAAGACC
ATCATCGTGCACCTGAACAACAGCGTGGAGATCGTGTGCACCAGGCCCAACAACAACACC
AGGAAGAGCATCAGGATCGGCCCAGGCCAGACCTTCTACGCCACCGGCGACATCATCGGC
GACATCCGGCAGGCCTACTGCAGCATCAACGAGAGCAACTGGAACGCCACCCTGCAGAGG
GTGTCCAAGAAGCTGGCCGAGCACTTCCCCAACAAGACCATCCAGTTCAAGAGCCCCTCT
GGCGGCGACCTGGAGATCACCATGCACAGCTTCAACTGCAGGGGCGAGTTCTTCTACTGC
AACACATCCAAGCTGTTTAACGGGACCTACTACCCAAACGGCACATACTACCCAAACGGG
ACCAACAGCACCTGATCATCCCATGCAGGATCAAGCAGATCATCAACATGTGGCAGGGC
GTGGGCAAGGCCATCTACGCCAGCCCAATCGCCGGCAACATCACCTGCCGGTCCAACATC

FIGURE 15 (cont.)

ACCGGCCTGCTGCTGACCAGGGACGGCGGCGACACCAACGACACCGAGATCTTCAGGCCA
GCCGGCGGAGACATGAGGGACAACCTGGCGGAGCGAGCTGTACAAGTACAAGATCGTGGAG
ATCAAGCCCCTGGGCGTGGCTCCAACCGAGGCCAAGAGAAGGGTGGTGAAGAGAGAGAAG
AGGGCCGTCAACATCGGCGCCGTGTTTCTGGGCTTCCTGGGAGCCGCCGAAGCACCATG
GGAGCCGCCTCCATCACCTGACCGTGCAGGCCAGGCAGCTGCTGTCCGGGATCGTGCAG
CAGCAGAGCAACCTGCTGAGAGCCATTGAGGCTCAGCAGCACATGCTGCAGCTGACAGTG
TGGGGCATCAAGCAGCTGCAGGCCAGAGTGCTGGCCATCGAGAGATACCTGAAGGATCAG
CAGCTGCTCGGCATCTGGGGCTGCAGCGCAAGCTGATCTGCACCACCGCCGTGCCCTGG
AACAACCTCTGGTCCAACAGGACCCAGGACGAGATCTGGAAGAACCTGACCTGGATGGAG
TGGGACAGAGAGATCAGCAACTACACCAACACCATCTACGAGCTGCTGGAGGTGTCCAG
AGCCAGCAGGAGAGGAACGAGAAGGACCTGCTGGCCCTGGATAGCTGGAACAACCTGTGG
AACTGGTTCGACATCAGCAACTGGCTGTGGTACATCAAGATCTTCATCATGATCGTGGGC
GGCCTGATCGGCCTGAGGATCATCTTCGCCGTGCTGTCCATCGTGAACAGAGTGCGGCAG
GGCTACAGCCCCCTGAGCTTCAGACCCTGATCCCCAACCCAGAGAGAGCCCGACAGGCC
GGCAGGATCGAGGAGGAGGGCGGCGAGCAGGACAAGGACAGATCCATCAGGCTGGTGTCC
GGCTTCCTGGCCCTGGCCTGGGACGACCTGAGAAGCCTGTGCATCTTCCTGTACCACCAC
CTGAGGGACTTCATCCTGATCGCCGCCAGGGCCACAGAGCTGCTGGGGAGGTCCAGCCTG
AGGGGCCTGCAGAGAGGCTGGGAGGCCCTGAAGTACCTGGGCAGCCTGGTGCAGTACTGG
GGCCTGGAGATCAAGAAGAGCGCCATCAACCTGCTGGACACAATCGCCATCGCCGTGGCC
GAGGGCACCGACAGGATCATCGAGATCGTGCAGAGGGCCTGTAGGGCCGTGCTGAACATC
CCCAGAAGGATCAGACAGGGGCTGGAGGCTGCCCTGCAGTGA

FIGURE 16

HIV-1 C06980v0c22 gp160 DC codon optimized Nucleic acid seq (SEQ ID NO: 17):

tPa signal is highlighted

The gp120/gp41 cleavage site is mutated to prevent cleavage

ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTT
ACTACCACAGAGGCTAGC AACCTGTGGGTGACCGTGTACTACGGCGTGCCCGTGTGGAGA
GAGGCCAAGACCACCTGTTCTGCGCCAGCGACGCCAAGGCCTACGAGCGGGAGGTGCAC
AACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCGACCCTCAGGAAATCTTCCTG
GGCAAGAACGTGACCGAGAAGTTCAACATGTGGAAGAACTACATGGTGGACCAGATGCAC
GAGGACATCATCAGCCTGTGGGACCAGAGCCTGCAGCCCTGCGTGAAGCTGACCCCCCTG
TGCATCACCTGAACTGCACCGACGTGACCGCCCACAACGGCAGCACCGTGTACGACAAC
AACGCCACCGTGGTGAACAGCACCAACGAGATCAAGAACTGCAGCTTCAACATCACCACC
GAGCTGAGGGACAAGAGGAAGAAGGAGCACGCCCTGTTCAACAACCTGGACATCGTGCAG
CTGGACGGCAACAGCTCCCTGTACAGACTGATCAACTGCAACACCAGCATCATCAAGCAG
GCCTGCCCCAAGATCAGCTTCGACCCCATCCCCATCCACTACTGCGCCCCAGCCGGCTTC
GTGATCCTGAAGTGCAACAACGAGACATTCAACGGCACCGGCCCTGTAACAACGTGTCC
GCTGTGCAGTGCACCCACGGCATCAAGCCCCTGGTGTCCACCCAGCTGCTGCTGAACGGC
AGCCTGGCCAAGGGCGAGATCATGATCAGAAGCGAGAACATCACCGACAACGTGAAGACC
ATCATCGTGACCTGAACAACAGCGTGGAGATCGTGTGCACCAGGCCCAACAACAACACC
AGGAAGAGCATCAGGATCGGCCCAGGCCAGACCTTCTACGCCACCGGCGACATCATCGGC
GACATCCGGCAGGCCTACTGCAGCATCAACGAGAGCAACTGGAACGCCACCCTGCAGAGG
GTGTCCAAGAAGCTGGCCGAGCACTTCCCCAACAAGACCATCCAGTTCAAGAGCCCCTCT
GGCGGCGACCTGGAGATCACCATGCACAGCTTCAACTGCAGGGGCGAGTTCTTCTACTGC

FIGURE 16 (cont.)

AACACATCCAAGCTGTTTAACGGGACCTACTACCCAAACGGCACATACTACCCAAACGGG
ACCAACAGCACCTGATCATCCCATGCAGGATCAAGCAGATCATCAACATGTGGCAGGGC
GTGGGCAAGGCCATCTACGCCAGCCCAATCGCCGGCAACATCACCTGCCGGTCCAACATC
ACCGGCCTGCTGCTGACCAGGGACGGCGGCGACACCAACGACACCGAGATCTTCAGGCCA
GCCGGCGGAGACATGAGGGACAACCTGGCGGAGCGAGCTGTACAAGTACAAGATCGTGGAG
ATCAAGCCCCTGGGCGTGGCTCCAACCGAGGCCAAGAGCAGGGTGGTGAAGAGCGAGAAG
AGCGCCGTCACCATCGGCGCCGTGTTTCTGGGCTTCCTGGGAGCCGCCGAAGCACCATG
GGAGCCGCCTCCATCACCTGACCGTGCAGGCCAGGCAGCTGCTGTCCGGGATCGTGCAG
CAGCAGAGCAACCTGCTGAGAGCCATTGAGGCTCAGCAGCACATGCTGCAGCTGACAGTG
TGGGGCATCAAGCAGCTGCAGGCCAGAGTGTGGCCATCGAGAGATACCTGAAGGATCAG
CAGCTGCTCGGCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCACCGCCGTGCCCTGG
AACAACCTCTGGTCCAACAGGACCCAGGACGAGATCTGGAAGAACCTGACCTGGATGGAG
TGGGACAGAGAGATCAGCAACTACACCAACACCATCTACGAGCTGCTGGAGGTGTCCCAG
AGCCAGCAGGAGAGGAACGAGAAGGACCTGCTGGCCCTGGATAGCTGGAACAACCTGTGG
AACTGGTTCGACATCAGCAACTGGCTGTGGTACATCAAGATCTTCATCATGATCGTGGGC
GGCCTGATCGGCCTGAGGATCATCTTCGCCGTGCTGTCCATCGTGAACAGAGTGCGGCAG
GGCTACAGCCCCCTGAGCTTCCAGACCCTGATCCCCAACCAGAGAGAGCCCGACAGGCCC
GGCAGGATCGAGGAGGAGGGCGGCGAGCAGGACAAGGACAGATCCATCAGGCTGGTGTCC
GGCTTCCTGGCCCTGGCCTGGGACGACCTGAGAAGCCTGTGCATCTTCCTGTACCACCAC
CTGAGGGACTTCATCCTGATCGCCGCCAGGGCCACAGAGCTGCTGGGGAGGTCCAGCCTG
AGGGGCCTGCAGAGAGGCTGGGAGGCCCTGAAGTACCTGGGCAGCCTGGTGCAGTACTGG
GGCCTGGAGATCAAGAAGAGCGCCATCAACCTGCTGGACACAATCGCCATCGCCGTGGCC
GAGGGCACCGACAGGATCATCGAGATCGTGCAGAGGGCCTGTAGGGCCGTGCTGAACATC
CCCAGAAGGATCAGACAGGGGCTGGAGGCTGCCCTGCAGTGA

FIG 17

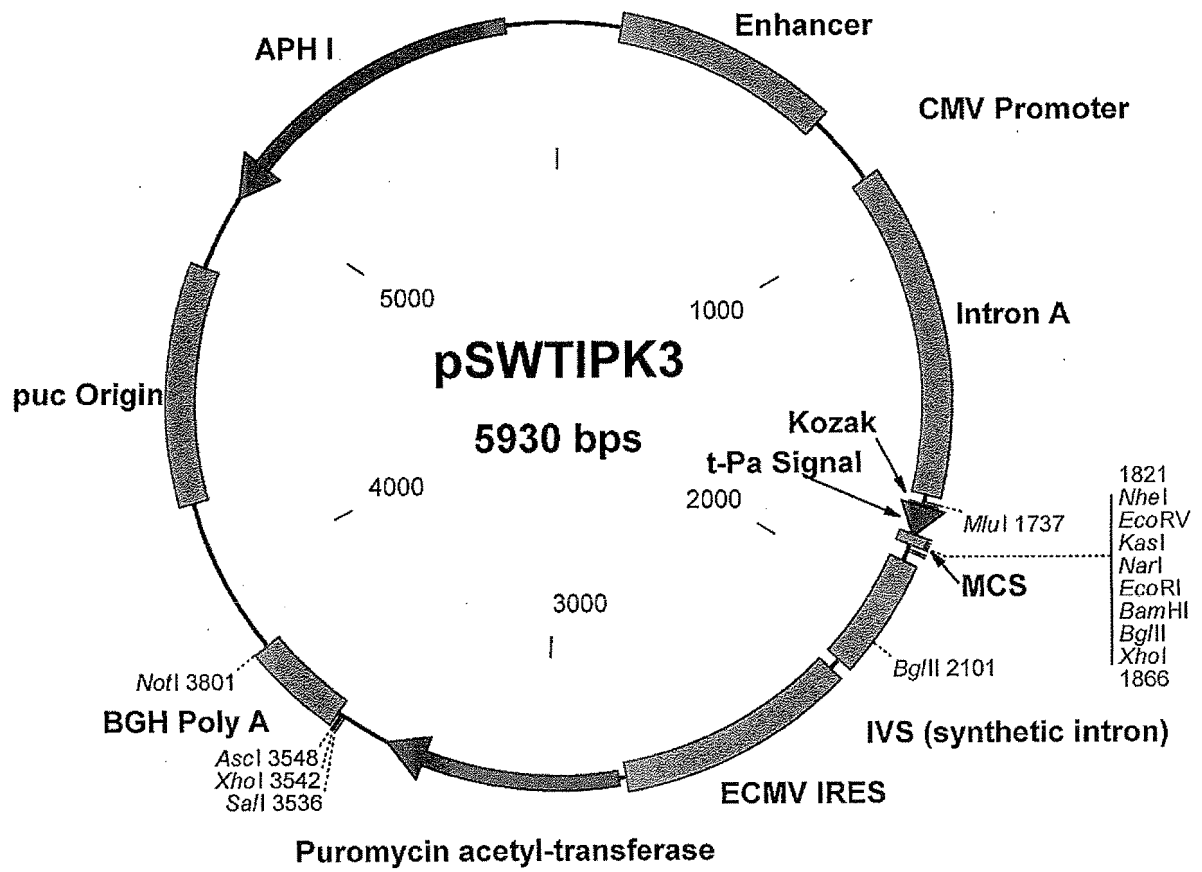


FIGURE 18

pSWTIPK3 sequence (SEQ ID NO: 18):

TGGCACATGGCCAATGCATATCGATCTATACATTGAATCAATATTGGCAATTAGCCATAT
TAGTCATTGGTTATATAGCATAAATCAATATTGGCTATTGGCCATTGCATACGTTGTATC
TATATCATAATATGTACATTTATATTGGCTCATGTCCAATATGACCGCCATGTTGACATT
GATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATA
TGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACC
CCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC
ATTGACGTCAATGGGTGGAGTATTTACGGTAACTGCCCACTTGGCAGTACATCAAGTGT
ATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATT
ATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCA
TCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTG
ACTCACGGGGATTTCGAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACC
AAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCG
GTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCG
CCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGCC
TCCGCGGCCGGGAACGGTGCATTGGAACGCGGATTCCCCGTGCCAAGAGTGACGTAAGTA
CCGCCTATAGACTCTATAGGCACACCCCTTTGGCTCTTATGCATGCTATACTGTTTTTGG
CTTGGGGCCTATACACCCCGCTTCCTTATGCTATAGGTGATGGTATAGCTTAGCCTATA
GGTGTGGGTATTGACCATTATTGACCACTCCCCTATTGGTGACGATACTTCCATTACT
AATCCATAACATGGCTCTTTGCCACAACCTATCTCTATTGGCTATATGCCAATACTCTGTC
CTTCAGAGACTGACACGGACTCTGTATTTTTACAGGATGGGGTCCCATTATTATTTACA
AATTCACATATACAACAACGCCGTCCCCCGTGCCCGCAGTTTTTTATTAAACATAGCGTGG
GATCTCCACGCGAATCTCGGGTACGTGTTCCGGACATGGGCTCTTCTCCGGTAGCGGCGG

FIGURE 18 (cont. - 1)

AGCTTCCACATCCGAGCCCTGGTCCCATGCCTCCAGCGGCTCATGGTCGCTCGGCAGCTC
CTTGCTCCTAACAGTGGAGGCCAGACTTAGGCACAGCACAATGCCCACCACCACAGTGT
GCCGCACAAGGCCGTGGCGGTAGGGTATGTGTCTGAAAATGAGCTCGGAGATTGGGCTCG
CACCGCTGACGCAGATGGAAGACTTAAGGCAGCGGCAGAAGAAGATGCAGGCAGCTGAGT
TGTTGTATTCTGATAAGAGTCAGAGGTAACCTCCCGTTGCGGTGCTGTAAACGGTGGAGGG
CAGTGTAGTCTGAGCAGTACTCGTTGCTGCCGCGCGGCCACCAGACATAATAGCTGACA
GACTAACAGACTGTTCTTTCCATGGGTCTTTTCTGCAGTCACCGTCCATTAAATACGC
GTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAG
TCTTCGTTACTACCACAGAGGCTAGCGGGGATATCGGCGCCGAATTCGGATCCACTAGTA
GATCTCTCGAGAACGGCCGCCAGTGTGCTGGAATTAATTCGCTGTCTGCGAGGGCCAGCT
GTTGGGGTGAGTACTCCCTCTCAAAAGCGGGCATGACTTCTGCGCTAAGATTGTCAGTTT
CCAAAACGAGGAGGATTTGATATTCACCTGGCCCGCGGTGATGCCTTTGAGGGTGGCCG
CGTCCATCTGGTCAGAAAAGACAATCTTTTTGTTGTCAAGCTTGAGGTGTGGCAGGCTTG
AGATCTGGCCATACACTTGAGTGACAATGACATCCACTTTGCCTTTCTCTCCACAGGTGT
CCACTCCCAGGTCCAACTGCAGGTCGAGCATGCATCTAGGGCGGCCAATTCGCCCCCTCT
CCCTCCCCCCCCCTAACGTTACTGGCCGAAGCCGCTTGAATAAGGCCGGTGTGCGTTT
GTCTATATGTTATTTTCCACCATATTGCCGTCTTTTGGCAATGTGAGGGCCCGGAAACCT
GGCCCTGTCTTCTTGACGAGCATTCTAGGGGTCTTTCCCCTCTCGCCAAAGGAATGCAA
GGTCTGTTGAATGTCGTGAAGGAAGCAGTTCCTCTGGAAGCTTCTTGAAGACAAACAACG
TCTGTAGCGACCCCTTGCAGGCAGCGGAACCCCCACCTGGCGACAGGTGCCTCTGCGGC
CAAAAGCCACGTGTATAAGATACACCTGCAAAGGCGGCACAACCCAGTGCCACGTTGTG
AGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCTCCTCAAGCGTATTCAACAAGGGGCTG
AAGGATGCCCAGAAGGTACCCATTGTATGGGATCTGATCTGGGGCCTCGGTGCACATGC
TTTACATGTGTTTAGTCGAGGTTAAAAAACGTCTAGGCCCCCGAACCACGGGGACGTG

FIGURE 18 (cont. - 2)

GTTTTCTTTGAAAAACACGATGATAAGCTTGCCACAACCCACAAGGAGACGACCTTCCA
TGACCGAGTACAAGCCCACGGTGCGCCTCGCCACCCGCGACGACGTCCCCGGGCCGTAC
GCACCCTCGCCGCCGCGTTCGCCGACTACCCGCCACGCGCCACACCGTGGACCCGGACC
GCCACATCGAGCGGGTCACCGAGCTGCAAGAACTCTTCCTACGCGCGTCTGGGCTCGACA
TCGGCAAGGTGTGGGTGCGGACGACGGCGCTGCGGTGGCGGTCTGGACCACGCCGGAGA
GCGTCGAAGCGGGGGCGGTGTTGCCGAGATCGGCCCGCGCATGGCCGAGTTGAGCGGTT
CCCGGCTGGCCGCGCAGCAACAGATGGAAGGCTCCTGGCTCCGCACCGGCCCAAGGAGC
CCGCGTGTTCTTGCCACCGTCGGCGTCTCGCCGACCACCAGGGCAAGGGTCTGGGCA
GCGCCGTCGTGCTCCCCGGAGTGGAGGCGGCCGAGCGCGCCGGGTGCCCGCCTTCCTGG
AGACCTCCGCGCCCCGCAACCTCCCCTTCTACGAGCGGCTCGGCTTACCGTCACCGCCG
ACGTCGAGGTGCCCAGGACCGCGCACCTGGTGCATGACCCGCAAGCCCGGTGCCTGAC
GCCCCCCCCACGACCCGCGCGCCCGACCGAAAGGAGCGCACGACCCCATGGCTCCGACC
GAAGCCACCCGGGGCGGGCCCCGCCGACCCGCGACCCGCCCCGAGGCCACCGACGTCGA
CCTCGAGGGCGCGCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTG
TTTGCCCCTCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCT
AATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTG
GGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATG
CGGTGGGCTCTATGGCTTCTGCGGCCGCGAGCTTGGCGTAATCATGGTCATAGCTGTTCC
TGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTG
TAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCC
CGCTTTCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGG
GAGAGGCGGTTTGCGTATTGGGCGCTCTCCGCTTCCTCGCTCACTGACTCGCTGCGCTC
GGTCGTTCCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCAC

FIGURE 18 (cont. - 3)

AGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAA
CCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCA
CAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGC
GTTTCCCCCTGGAAGCTCCCTCGTGCCTCTCTGTTCCGACCCTGCCGCTTACCGGATA
CCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTA
TCTCAGTTCGGTGTAGGTCGTTCCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTCA
GCCCAGCCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGA
CTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGG
TGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGG
TATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGG
CAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTGCAAGCAGCAGATTACGCGCAG
AAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAA
CGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGAT
CCTTTTAAATTAATAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTC
TGACAGTTAGAAAACTCATCGAGCATCAAATGAACTGCAATTTATTCATATCAGGATT
ATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCA
GTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAAT
ACAACCTATTAATTTCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGT
GACGACTGAATCCGGTGAGAATGGCAAAAGCTTATGCATTTCTTTCCAGACTTGTTCAAC
AGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAACCGTTATTCATTG
TGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTAAAAGGACAATTACAAACAGG
AATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTACCTGAATC
AGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCA

FIGURE 18 (cont. - 4)

TGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATCCGTCAG
CCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTT
CAGAAACAACCTCTGGCGCATCGGGCTTCCCATAACAATCGATAGATTGTCGCACCTGATTG
CCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAA
TCGCGGCCTTGAGCAAGACGTTTCCCGTTGAATATGGCTCATACTCTTCCTTTTCAATA
TTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTA
GAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC

FIGURE 19

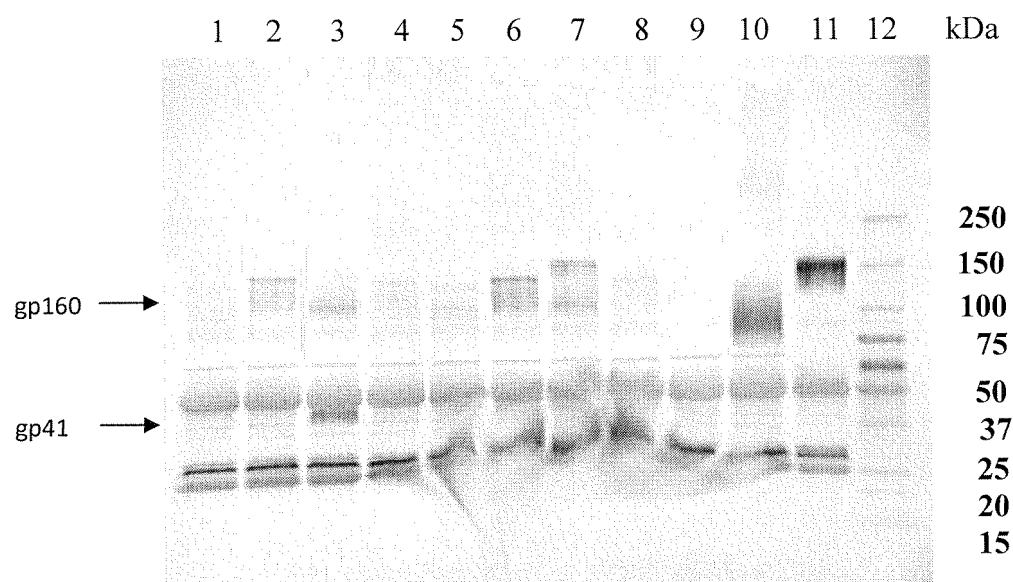


FIGURE 20

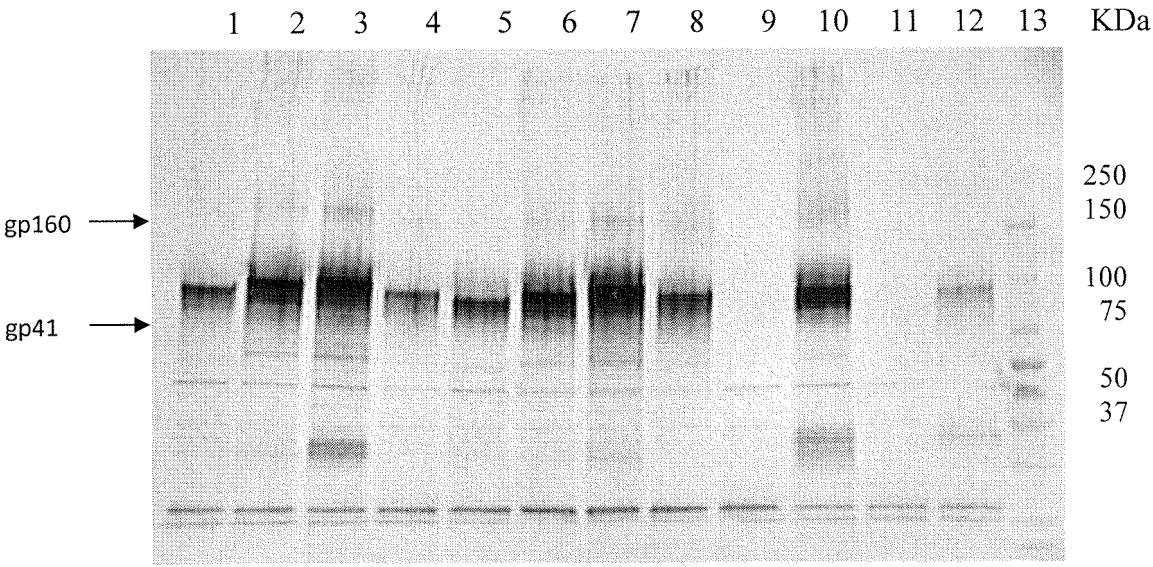


FIGURE 21

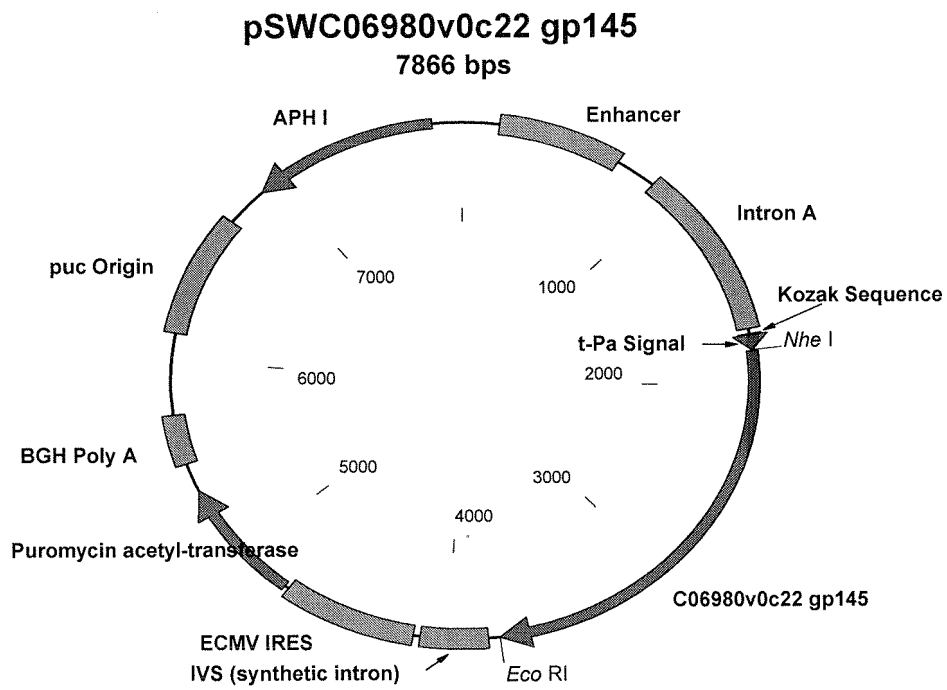


FIGURE 22

pSWC06980v0c22 gp145 sequence (SEQ ID NO: 19):

TGGCACATGGCCAATGCATATCGATCTATACATTGAATCAATATTGGCAATTAGCCATAT
TAGTCATTGGTTATATAGCATAAATCAATATTGGCTATTGGCCATTGCATACGTTGTATC
TATATCATAATATGTACATTTATATTGGCTCATGTCCAATATGACCGCCATGTTGACATT
GATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATA
TGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACC
CCCGCCCATGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC
ATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGT
ATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATT
ATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCA
TCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTG
ACTCACGGGGATTTCGAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACC
AAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCG
GTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCG
CCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGCC
TCCGCGGCCGGGAACGGTGCATTGGAACGCGGATTCCCCGTGCCAAGAGTGACGTAAGTA
CCGCCTATAGACTCTATAGGCACACCCCTTTGGCTCTTATGCATGCTATACTGTTTTTGG
CTTGGGGCCTATACACCCCGCTTCCTTATGCTATAGGTGATGGTATAGCTTAGCCTATA
GGTGTGGGTATTGACCATTATTGACCACTCCCCTATTGGTGACGATACTTCCATTACT
AATCCATAACATGGCTCTTTGCCACAACCTATCTCTATTGGCTATATGCCAATACTCTGTC
CTTCAGAGACTGACACGGACTCTGTATTTTTACAGGATGGGGTCCCATTATTATTACAA
AATTCACATATACAACAACGCCGTCCCCCGTGCCCGCAGTTTTTTATTAAACATAGCGTGG
GATCTCCACGCGAATCTCGGGTACGTGTTCCGGACATGGGCTCTTCTCCGGTAGCGGCGG

FIGURE 22 (cont.-1)

AGCTTCCACATCCGAGCCCTGGTCCCATGCCTCCAGCGGCTCATGGTCGCTCGGCAGCTC
CTTGCTCCTAACAGTGGAGGCCAGACTTAGGCACAGCACAATGCCCACCACCACAGTGT
GCCGCACAAGGCCGTGGCGGTAGGGTATGTGTCTGAAAATGAGCTCGGAGATTGGGCTCG
CACCGCTGACGCAGATGGAAGACTTAAGGCAGCGGCAGAAGAAGATGCAGGCAGCTGAGT
TGTTGTATTCTGATAAGAGTCAGAGGTAACCCCGTTGCGGTGCTGTTAACGGTGGAGGG
CAGTGTAGTCTGAGCAGTACTCGTTGCTGCCGCGCGGCCACCAGACATAATAGCTGACA
GACTAACAGACTGTTCTTTCCATGGGTCTTTTCTGCAGTCACCGTCCATTTAAATACGC
GTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAG
TCTTCGTTACTACCACAGAGGCTAGCAACCTGTGGGTGACCGTGTACTACGGCGTGCCCG
TGTGGAGAGAGGCCAAGACCACCCTGTTCTGCGCCAGCGACGCCAAGGCCTACGAGCGGG
AGGTGCACAACGTGTGGGGCCACCCACGCCTGCGTGCCACCGACCCCGACCCCTCAGGAAA
TCTTCCTGGGCAAGAACGTGACCGAGAAGTTCAACATGTGGAAGAACTACATGGTGGACC
AGATGCACGAGGACATCATCAGCCTGTGGGACCAGAGCCTGCAGCCCTGCGTGAAGCTGA
CCCCCTGTGCATCACCTGAACTGCACCGACGTGACCGCCCACAACGGCAGCACCGTGT
ACGACAACAACGCCACCGTGGTGAACAGCACCAACGAGATCAAGAACTGCAGCTTCAACA
TCACCACCGAGCTGAGGGACAAGAGGAAGAAGGAGCACGCCCTGTTCAACAACCTGGACA
TCGTGCAGCTGGACGGCAACAGCTCCCTGTACAGACTGATCAACTGCAACACCAGCATCA
TCAAGCAGGCCTGCCCAAGATCAGCTTCGACCCCATCCCCATCCACTACTGCGCCCCAG
CCGGCTTCGTGATCCTGAAGTGCAACAACGAGACATTCAACGGCACCGGCCCTGTAACA
ACGTGTCCGCTGTGCAGTGCACCCACGGCATCAAGCCCGTGGTGTCCACCCAGCTGCTGC
TGAACGGCAGCCTGGCCAAGGGCGAGATCATGATCAGAAGCGAGAACATCACCGACAACG
TGAAGACCATCATCGTGCACCTGAACAACAGCGTGGAGATCGTGTGCACCAGGCCCAACA
ACAACACCAGGAAGAGCATCAGGATCGGCCAGGCCAGACCTTCTACGCCACCGGCGACA
TCATCGGCGACATCCGGCAGGCCTACTGCAGCATCAACGAGAGCAACTGGAACGCCACCC

FIGURE 22 (cont.-2)

TGCAGAGGGTGTCCAAGAAGCTGGCCGAGCACTTCCCCAACAAAGACCATCCAGTTCAAGA
GCCCCCTCTGGCGGCGACCTGGAGATCACCATGCACAGCTTCAACTGCAGGGGCGAGTTCT
TCTACTGCAACACATCCAAGCTGTTTAACGGGACCTACTACCCAAACGGCACATACTACC
CAAACGGGACCAACAGCACCTGATCATCCCATGCAGGATCAAGCAGATCATCAACATGT
GGCAGGGCGTGGGCAAGGCCATCTACGCCAGCCCAATCGCCGGCAACATCACCTGCCGGT
CCAACATCACCGGCCTGCTGCTGACCAGGGACGGCGGCGACACCAACGACACCGAGATCT
TCAGGCCAGCCGGCGGAGACATGAGGGACAACCTGGCGGAGCGAGCTGTACAAGTACAAGA
TCGTGGAGATCAAGCCCCTGGGCGTGGCTCCAACCGAGGCCAAGAGCAGGGTGGTGAAGA
GCGAGAAGAGCGCCGTCACCATCGGCGCCGTGTTTCTGGGCTTCCTGGGAGCCGCCGGAA
GCACCATGGGAGCCGCCTCCATCACCTGACCGTGCAGGCCAGGCAGCTGCTGTCCGGGA
TCGTGCAGCAGCAGAGCAACCTGCTGAGAGCCATTGAGGCTCAGCAGCACATGCTGCAGC
TGACAGTGTGGGGCATCAAGCAGCTGCAGGCCAGAGTGTGGCCATCGAGAGATACCTGA
AGGATCAGCAGCTGCTCGGCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCACCGCCG
TGCCCTGGAACAACTCCTGGTCCAACAGGACCCAGGACGAGATCTGGAAGAACCTGACCT
GGATGGAGTGGGACAGAGAGATCAGCAACTACACCAACACCATCTACGAGCTGCTGGAGG
TGTCCCAGAGCCAGCAGGAGAGGAACGAGAAGGACCTGCTGGCCCTGGATAGCTGGAACA
ACCTGTGGAACCTGGTTCGACATCAGCAACTGGCTGTGGTACATCAAGAAGAAGAAGTGAA
TTCGGATCCACTAGTAGATCTCTCGAGAACGGCCGCCAGTGTGCTGGAATTAATTCGCTG
TCTGCGAGGGCCAGCTGTTGGGGTGAGTACTCCCTCTCAAAAGCGGGCATGACTTCTGCG
CTAAGATTGTCAGTTTCAAAAACGAGGAGGATTTGATATTCACCTGGCCCGCGGTGATG
CCTTTGAGGGTGGCCGCGTCCATCTGGTCAGAAAAGACAATCTTTTTGTTGTCAAGCTTG
AGGTGTGGCAGGCTTGAGATCTGGCCATACACTTGAGTGACAATGACATCCACTTTGCCT
TTCTCTCCACAGGTGTCCACTCCCAGGTCCAACCTGCAGGTCGAGCATGCATCTAGGGCGG

FIGURE 22 (cont.-3)

CCAATTCCGCCCCTCTCCCTCCCCCCCCCTAACGTTACTGGCCGAAGCCGCTTGGAATA
AGGCCGGTGTGCGTTTGTCTATATGTTATTTCCACCATATTGCCGTCTTTTGGCAATGT
GAGGGCCCGGAAACCTGGCCCTGTCTTCTTGACGAGCATTCTAGGGGTCTTTCCCCTCT
CGCCAAAGGAATGCAAGGTCTGTTGAATGTCGTGAAGGAAGCAGTTCCTCTGGAAGCTTC
TTGAAGACAAACAACGTCTGTAGCGACCCTTTGCAGGCAGCGGAACCCCCACCTGGCGA
CAGGTGCCTCTGCGGCCAAAAGCCACGTGTATAAGATACACCTGCAAAGCGGCACAACC
CCAGTGCCACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCTCCTCAAGCGT
ATTCAACAAGGGGCTGAAGGATGCCAGAAGGTACCCCATTTGTATGGGATCTGATCTGGG
GCCTCGGTGCACATGCTTTACATGTGTTTAGTCGAGGTTAAAAAACGTCTAGGCCCCC
GAACCACGGGGACGTGGTTTTCTTTGAAAAACACGATGATAAGCTTGCCACAACCCACA
AGGAGACGACCTTCCATGACCGAGTACAAGCCACGGTGCGCCTCGCCACCCGCGACGAC
GTCCCCCGGGCCGTACGCACCCTCGCCGCCGCGTTCCGCCGACTACCCGCCACGCGCCAC
ACCGTGGAACCCGGACCGCCACATCGAGCGGGTCACCGAGCTGCAAGAACTCTTCCTCACG
CGCGTCGGGCTCGACATCGGCAAGGTGTGGGTGCGGGACGACGGCGCTGCGGTGGCGGTG
TGGACCACGCCGAGAGCGTGAAGCGGGGGCGGTGTTCCGCCGAGATCGGCCCCGCGCATG
GCCGAGTTGAGCGGTTCCCGGTGGCCGCGCAGCAACAGATGGAAGGCCTCCTGGCTCCG
CACCGGCCCAAGGAGCCCGCGTGGTTCCTGGCCACCGTCGGCGTCTCGCCCGACCACCAG
GGCAAGGGTCTGGGCAGCGCCGTCGTGCTCCCCGGAGTGAGGGCGGCCGAGCGCGCCGGG
GTGCCCCGCTTCTGAGACCTCCGCGCCCCGCAACCTCCCCTTCTACGAGCGGCTCGGC
TTCACCGTCACCGCCGACGTCGAGGTGCCCCAAGGACCGCGCACCTGGTGCATGACCCGC
AAGCCCGGTGCCTGACGCCCCCCCCACGACCCGCAGCGCCCGACCGAAAGGAGCGCACGA
CCCCATGGCTCCGACCGAAGCCACCCGGGGCGGCCCCGCCGACCCGCACCCGCCCCGA
GGCCACCGACGTCGACCTCGAGGGCGCGCCGCTGATCAGCCTCGACTGTGCCTTCTAGT

FIGURE 22 (cont.-4)

TGCCAGCCATCTGTTGTTTGCCCCCTCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACT
CCCCTGTCTTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTTCAT
TCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGC
AGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGCGGCCGCAGCTTGGCGTAATCAT
GGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAG
CCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTG
CGTTGCGCTCACTGCCCCGCTTTCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAA
TCGGCCAACGCGCGGGGAGAGGCGGTTTTCGTATTGGGCGCTCTTCCGCTTCCTCGCTCA
CTGACTCGCTGCGCTCGGTCGTTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGG
TAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCC
AGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTTCATAGGCTCCGCC
CCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGAC
TATAAAGATAACAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCC
TGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATA
GCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTTCGCTCCAAGCTGGGCTGTGTGC
ACGAACCCCCCGTTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCA
ACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAG
CGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTA
GAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTG
GTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTGTTGCAAGC
AGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGT
CTGACGCTCAGTGGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAA
GGATCTTCACCTAGATCCTTTTAAATTAATAAATGAAGTTTTAAATCAATCTAAAGTATAT

FIGURE 22 (cont.-5)

ATGAGTAACTTGGTCTGACAGTTAGAAAACTCATCGAGCATCAAATGAAACTGCAATT
TATTCATATCAGGATTATCAATACCATATTTTGA AAAAGCCGTTTCTGTAATGAAGGAG
AAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGA
CTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTG
AGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGCTTATGCATTTCTT
TCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCA
AACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTAAAAAG
GACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAA
TATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCG
CAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAG
GCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGC
TACCTTTGCCATGTTTCAGAAACAACCTCTGGCGCATCGGGCTTCCCATACAATCGATAGA
TTGTCGCACCTGATTGCCCCGACATTATCGCGAGCCCATTATACCCATATAAATCAGCAT
CCATGTTGGAATTTAATCGCGGCCTTGAGCAAGACGTTTCCCGTTGAATATGGCTCATAC
TCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACA
TATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAG
TGCCAC

FIGURE 23

HIV-1 C06980v0c22 gp145 codon optimized nucleic acid seq. (SEQ ID NO: 20):

tPa signal is highlighted

ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTT
ACTACCACAGAGGCTAGC AACCTGTGGGTGACCGTGACTACGGCGTGCCCGTGTGGAGA
GAGGCCAAGACCACCTGTTCTGCGCCAGCGACGCCAAGGCCTACGAGCGGGAGGTGCAC
AACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCGACCCTCAGGAAATCTTCCTG
GGCAAGAACGTGACCGAGAAGTTCAACATGTGGAAGAACTACATGGTGGACCAGATGCAC
GAGGACATCATCAGCCTGTGGGACCAGAGCCTGCAGCCCTGCGTGAAGCTGACCCCCCTG
TGCATCACCTGAACTGCACCGACGTGACCGCCCACAACGGCAGCACCGTGTACGACAAC
AACGCCACCGTGGTGAACAGCACCAACGAGATCAAGAACTGCAGCTTCAACATCACCACC
GAGCTGAGGGACAAGAGGAAGAAGGAGCACGCCCTGTTCAACAACCTGGACATCGTGCAG
CTGGACGGCAACAGCTCCCTGTACAGACTGATCAACTGCAACACCAGCATCATCAAGCAG
GCCTGCCCCAAGATCAGCTTCGACCCCATCCCCATCCACTACTGCGCCCCAGCCGGCTTC
GTGATCCTGAAGTGCAACAACGAGACATTCAACGGCACCGGCCCTGTAACAACGTGTCC
GCTGTGCAGTGCACCCACGGCATCAAGCCCGTGGTGTCCACCCAGCTGCTGCTGAACGGC
AGCCTGGCCAAGGGCGAGATCATGATCAGAAGCGAGAACATCACCGACAACGTGAAGACC
ATCATCGTGACCTGAACAACAGCGTGGAGATCGTGTGCACCAGGCCCAACAACAACACC
AGGAAGAGCATCAGGATCGGCCCAGGCCAGACCTTCTACGCCACCGGCGACATCATCGGC
GACATCCGGCAGGCCTACTGCAGCATCAACGAGAGCAACTGGAACGCCACCCTGCAGAGG
GTGTCCAAGAAGCTGGCCGAGCACTTCCCCAACAAGACCATCCAGTTCAAGAGCCCCTCT
GGCGGCGACCTGGAGATCACCATGCACAGCTTCAACTGCAGGGGCGAGTTCTTCTACTGC
AACACATCCAAGCTGTTTAACGGGACCTACTACCCAAACGGCACATACTACCCAAACGGG
ACCAACAGCACCCCTGATCATCCCATGCAGGATCAAGCAGATCATCAACATGTGGCAGGGC

FIGURE 23 (cont.)

GTGGGCAAGGCCATCTACGCCAGCCCAATCGCCGGCAACATCACCTGCCGGTCCAACATC
ACCGGCCTGCTGCTGACCAGGGACGGCGGGGACACCAACGACACCGAGATCTTCAGGCCA
GCCGGCGGAGACATGAGGGACAACCTGGCGGAGCGAGCTGTACAAGTACAAGATCGTGGAG
ATCAAGCCCCTGGGCGTGGCTCCAACCGAGGCCAAGAGCAGGGTGGTGAAGAGCGAGAAG
AGCGCCGTCACCATCGGCGCCGTGTTTCTGGGCTTCCTGGGAGCCGCCGGAAGCACCATG
GGAGCCGCCTCCATCACCTGACCGTGCAGGCCAGGCAGCTGCTGTCCGGGATCGTGCAG
CAGCAGAGCAACCTGCTGAGAGCCATTGAGGCTCAGCAGCACATGCTGCAGCTGACAGTG
TGGGGCATCAAGCAGCTGCAGGCCAGAGTGCTGGCCATCGAGAGATACCTGAAGGATCAG
CAGCTGCTCGGCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCACCGCCGTGCCCTGG
AACAACCTCTGGTCCAACAGGACCCAGGACGAGATCTGGAAGAACCTGACCTGGATGGAG
TGGGACAGAGAGATCAGCAACTACACCAACACCATCTACGAGCTGCTGGAGGTGTCCCAG
AGCCAGCAGGAGAGGAACGAGAAGGACCTGCTGGCCCTGGATAGCTGGAACAACCTGTGG
AACTGGTTCGACATCAGCAACTGGCTGTGGTACATCAAGAAGAAGAAGTGA

FIGURE 24

HIV-1 C06980v0c22 gp145 protein (SEQ ID NO: 32):

tPa signal is highlighted

MDAMKRGGLCCVLLCGAVFVTTTEAS NLWVTVYYGVPVWREAKTTLFCASDAKAYEREVH
NVWATHACVPTDPDPQEIFLGKNVTEKFNMWKNYMVDQMHEDIISLWDQSLQPCVKLTPL
CITLNCTDVTAHNGSTVYDNNATVVNSTNEIKNCSFNITTELDRKRKKEHALFNNLDIVQ
LDGNSSLYRLINCNTSIIKQACPKISFDPIPIHYCAPAGFVILKCNNETFNGTGPCNNVS
AVQCTHGIKPVVSTQLLLNGSLAKGEIMIRSENITDNVKTIIIVHLNNSVEIVCTRPNNNT
RKSIRIGPGQTFYATGDIIGDIRQAYCSINESNWNATLQRVSKKLAEHFPNKTIQFKSPS
GGDLEITMHSFNCRGEFFYCNTSKLFNGTYYPNGTYYPNGTNSTLIIPCRIKQIINMWQG
VGKAIYASPIAGNITCRSNITGLLLTRDGGDTNDTEIFRPAGGDMRDNWRSELYKYKIVE
IKPLGVAPTEAKSRVVKSEKSAVTIGAVFLGFLGAAGSTMGAASITLTVQARQLLSGIVQ
QQSNLLRAIEAQQHMLQLTVWGIKQLQARVLAIERYLKDQQLGIWGC SGKLICTTAVPW
NNSWSNRTQDEIWKNLTWMEWDREISNYTNTIYELLEVSQSQQERNEKDLLALDSWNNLW
NWFDISNWLWYIKKKK

FIGURE 25

(A) tPa signal (SEQ ID NOS: 33 and 21)

Highlighted regions are from the Nhe1 cloning site. This is not part of a native tPa gene, but is incorporated for cloning purposes. Cleavage occurs between the A and S.

Nucleic acid

ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTT

TCG **GCTAGC** (SEQ ID NO: 33)

Protein

MDAMKRGLCCVLLCGAVFVS **AS** (SEQ ID NO: 21)

(B) tPa signal (SEQ ID NOS: 34 and 22)

Highlighted regions are from the Nhe1 cloning site. This is not part of a native tPa gene, but is incorporated for cloning purposes. Cleavage occurs between the A and S.

Nucleic acid

ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTT

ACTACCACAGAG **GCTAGC** (SEQ ID NO: 34)

Protein

MDAMKRGLCCVLLCGAVFVTTE **AS** (SEQ ID NO: 22)

FIGURE 26

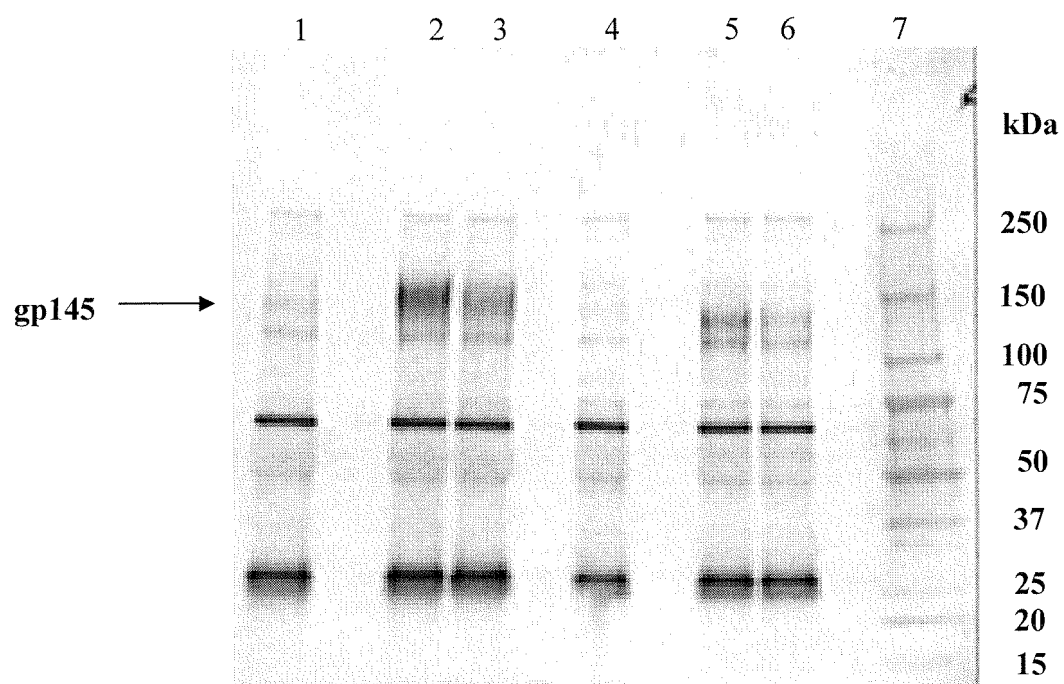


FIGURE 27

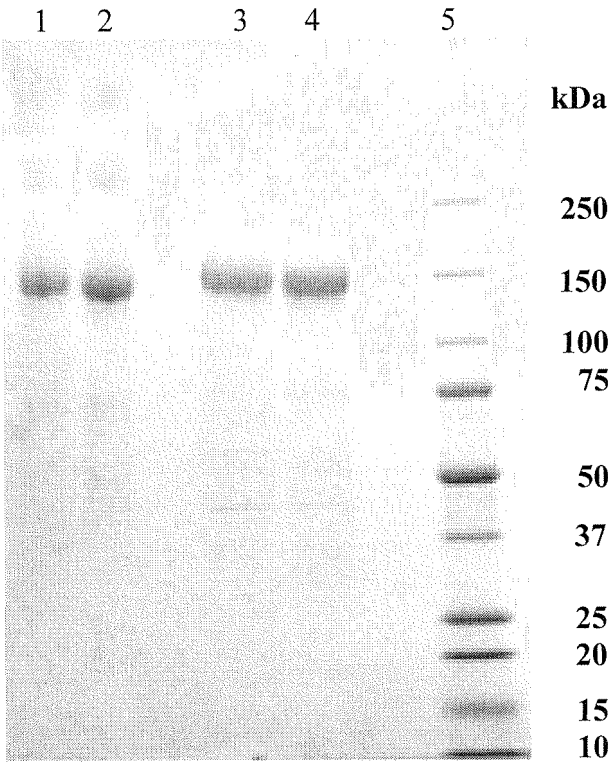


FIGURE 28

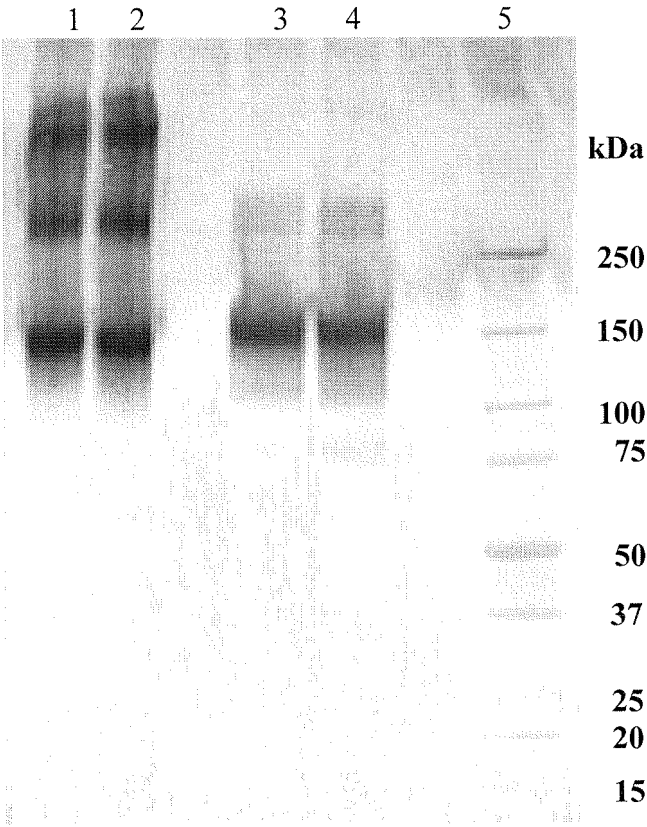


FIGURE 29

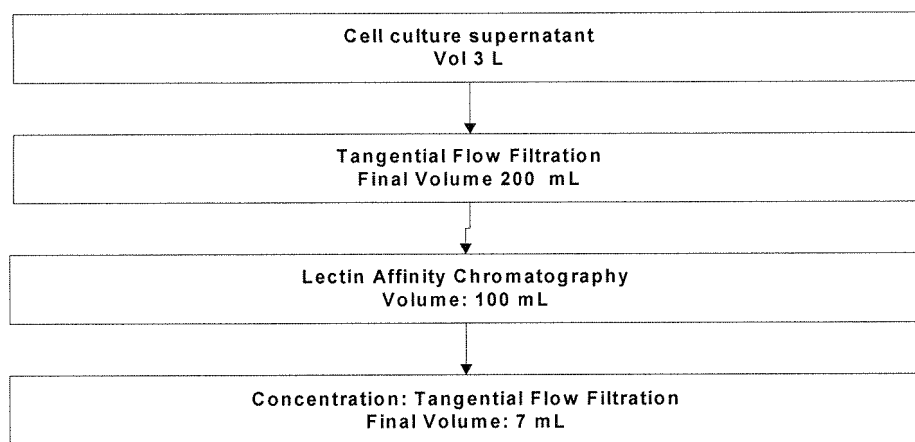


FIGURE 30

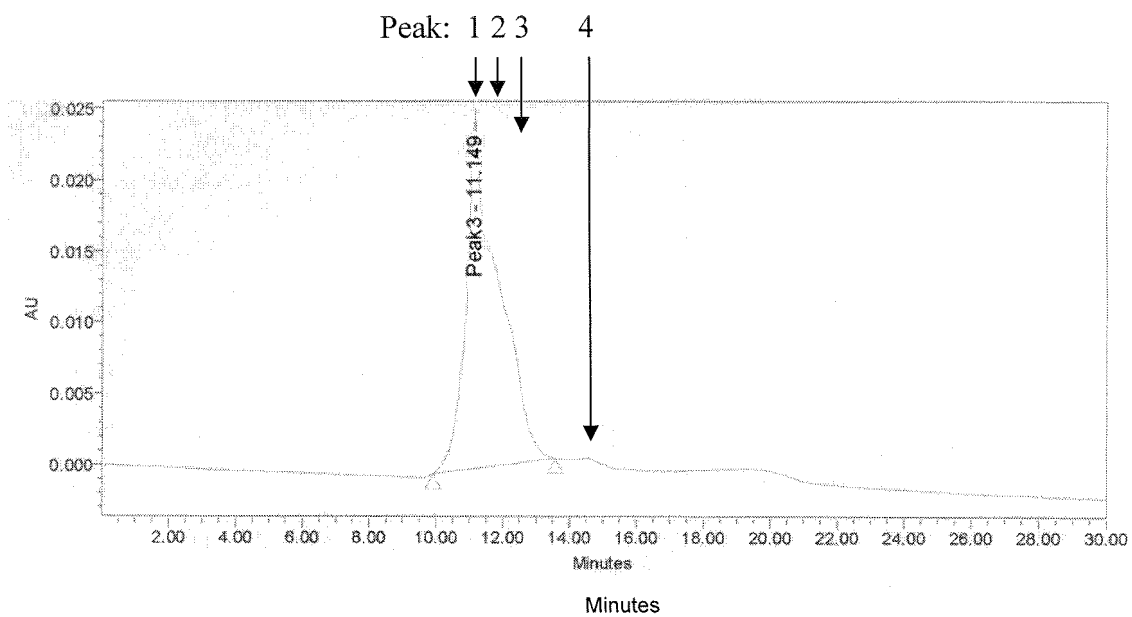


FIGURE 31

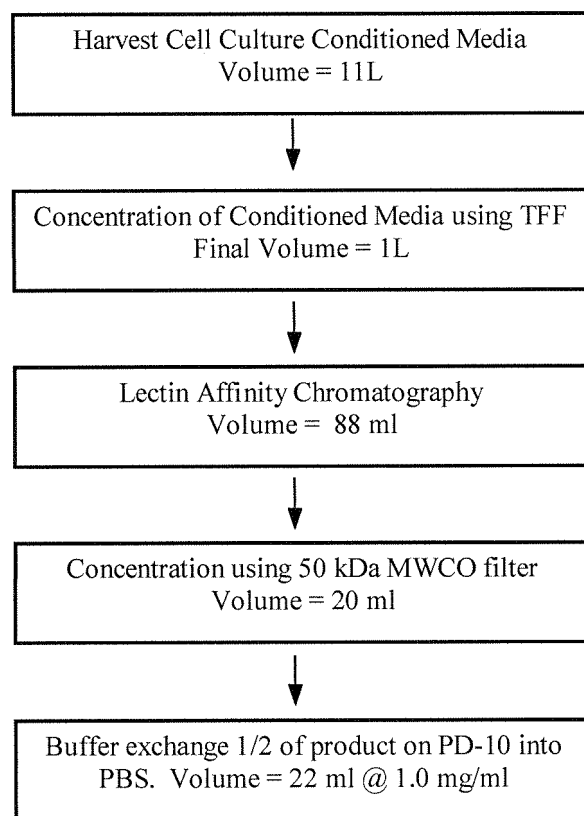


FIGURE 32

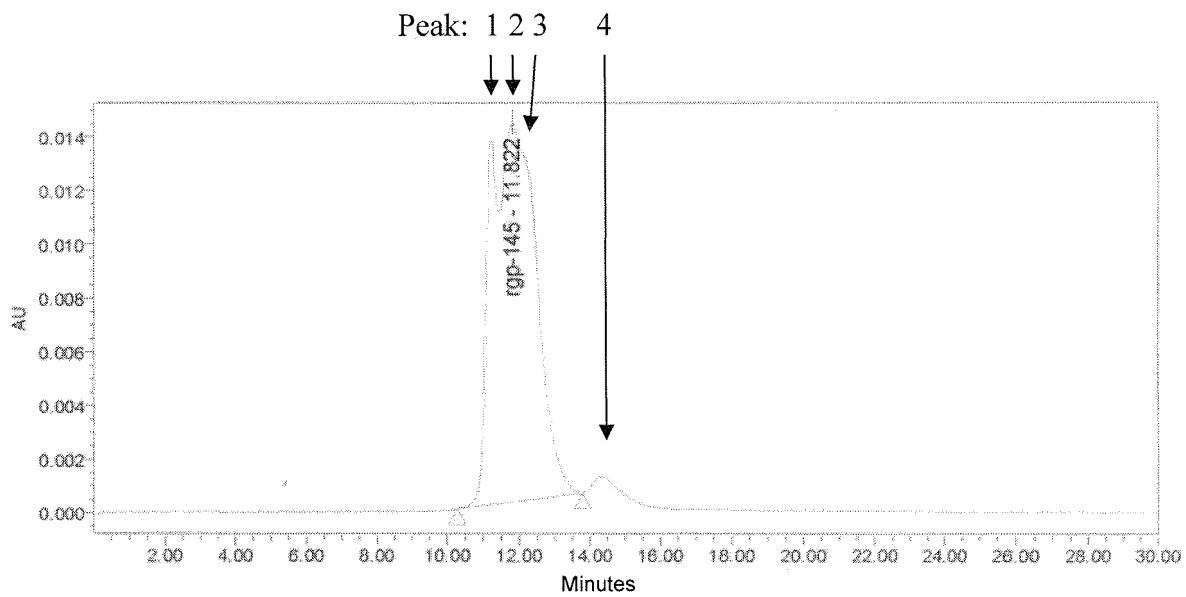


FIGURE 33

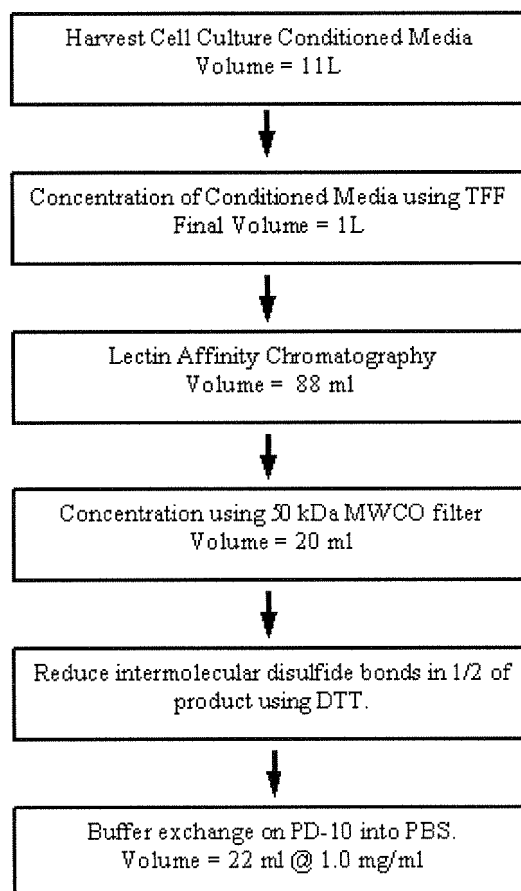


FIGURE 34

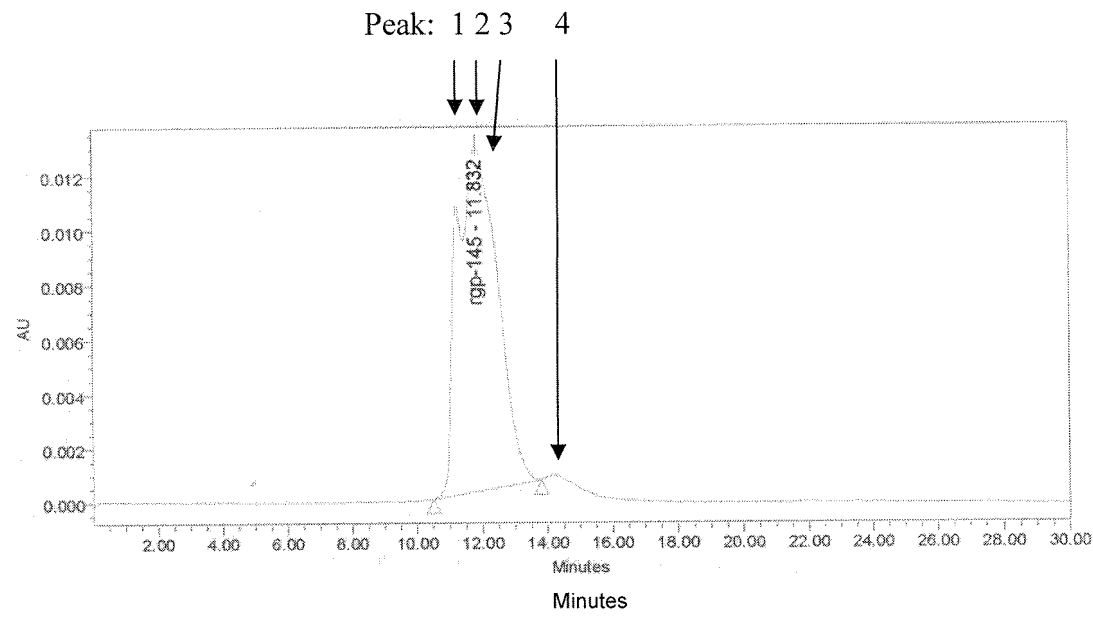


FIGURE 35

C3728v2c6 WT	SNLWVTVYYGVPVWKEAKTTLFCASDAKAYEKEVHNVWATHACVPTDPNP
C3728v2c6 DC	SNLWVTVYYGVPVWKEAKTTLFCASDAKAYEKEVHNVWATHACVPTDPNP
C06838v1c48 WT	SNLWVTVYYGVPVWREAKTTLFCASDAKAYEKEVHNVWATHACVPTDPDP
C06838v1c48 DC	SNLWVTVYYGVPVWREAKTTLFCASDAKAYEKEVHNVWATHACVPTDPDP
C06980v1c3 WT	SNLWVTVYYGVPVWREAKTTLFCASDAKAYEREVHNVWATHACVPTDPDP
C06980v1c3 DC	SNLWVTVYYGVPVWREAKTTLFCASDAKAYEREVHNVWATHACVPTDPDP
C06980v0c22 WT	SNLWVTVYYGVPVWREAKTTLFCASDAKAYEREVHNVWATHACVPTDPDP
C06980v0c22 DC	SNLWVTVYYGVPVWREAKTTLFCASDAKAYEREVHNVWATHACVPTDPDP
C3728v2c6 WT	QEMEL-PNVTENFNMWKNDMVDQMHEDIISLWDQSLKPCVKLTPLCVTLN
C3728v2c6 DC	QEMEL-PNVTENFNMWKNDMVDQMHEDIISLWDQSLKPCVKLTPLCVTLN
C06838v1c48 WT	QEMVL-KNVTENFNMWKNDMVEQMHEDIISLWDQSLKPCVKLTPLCVTLT
C06838v1c48 DC	QEMVL-KNVTENFNMWKNDMVEQMHEDIISLWDQSLKPCVKLTPLCVTLT
C06980v1c3 WT	QEIFLGKNVTEKFNMWENYMVDQMHEDIISLWDQSLQPCVKLTPLCITLN
C06980v1c3 DC	QEIFLGKNVTEKFNMWENYMVDQMHEDIISLWDQSLQPCVKLTPLCITLN
C06980v0c22 WT	QEIFLGKNVTEKFNMWKNYMVDQMHEDIISLWDQSLQPCVKLTPLCITLN
C06980v0c22 DC	QEIFLGKNVTEKFNMWKNYMVDQMHEDIISLWDQSLQPCVKLTPLCITLN
C3728v2c6 WT	CNKVNNN-----RSID-NNSTEEEMKNCTFNTTTEIRDKKRTQQALFYK
C3728v2c6 DC	CNKVNNN-----RSID-NNSTEEEMKNCTFNTTTEIRDKKRTQQALFYK
C06838v1c48 WT	CTNVNAT-----DNVTYKEKMEGEIKNCSFNITTEIRDKKRKVHALFYR
C06838v1c48 DC	CTNVNAT-----DNVTYKEKMEGEIKNCSFNITTEIRDKKRKVHALFYR
C06980v1c3 WT	CTDFTA HNGSTVYDNNATA-NSTNEIKNCSFNIISELDRDKRKKEHALFNN
C06980v1c3 DC	CTDFTA HNGSTVYDNNATA-NSTNEIKNCSFNIISELDRDKRKKEHALFNN
C06980v0c22 WT	CTDFTA HNGSTVYDNNATVVNSTNEIKNCSFNITTELRDKRKKEHALFNN
C06980v0c22 DC	CTDFTA HNGSTVYDNNATVVNSTNEIKNCSFNITTELRDKRKKEHALFNN
C3728v2c6 WT	LDIVPLGNSNESYRLINCNTSTLTQACPKVTFDPIPIHYCAPAGYAILKC
C3728v2c6 DC	LDIVPLGNSNESYRLINCNTSTLTQACPKVTFDPIPIHYCAPAGYAILKC
C06838v1c48 WT	LDVVQLNNSNE-YILINCNTSAITQACPKVSFDPIPIHYCAPAGYAILKC
C06838v1c48 DC	LDVVQLNNSNE-YILINCNTSAITQACPKVSFDPIPIHYCAPAGYAILKC
C06980v1c3 WT	LDIVQLDGNSSLYRLINCNTSIIKQACPKISFDPIPIHYCAPAGFVILTC
C06980v1c3 DC	LDIVQLDGNSSLYRLINCNTSIIKQACPKISFDPIPIHYCAPAGFVILTC
C06980v0c22 WT	LDIVQLDGNSSLYRLINCNTSIIKQACPKISFDPIPIHYCAPAGFVILKC
C06980v0c22 DC	LDIVQLDGNSSLYRLINCNTSIIKQACPKISFDPIPIHYCAPAGFVILKC
C3728v2c6 WT	KDEKFNGTGPCNNVSTVQCTHGIKPVVSTQLLLNGSLAEKDIIIRSENLT
C3728v2c6 DC	KDEKFNGTGPCNNVSTVQCTHGIKPVVSTQLLLNGSLAEKDIIIRSENLT
C06838v1c48 WT	NDKPFNGTGPCNNVSTVQCTHGIKPVISTQLLLNGSLAEKEVIIRFENLT
C06838v1c48 DC	NDKPFNGTGPCNNVSTVQCTHGIKPVISTQLLLNGSLAEKEVIIRFENLT
C06980v1c3 WT	NNETFNGTGPCNNVSAVQCTHGIKPVVSTQLLLNGSLAKGEIMIRSENIT
C06980v1c3 DC	NNETFNGTGPCNNVSAVQCTHGIKPVVSTQLLLNGSLAKGEIMIRSENIT
C06980v0c22 WT	NNETFNGTGPCNNVSAVQCTHGIKPVVSTQLLLNGSLAKGEIMIRSENIT
C06980v0c22 DC	NNETFNGTGPCNNVSAVQCTHGIKPVVSTQLLLNGSLAKGEIMIRSENIT
C3728v2c6 WT	NNVKTIMVHLNLSVEINCTRPNNNTRKSIRIGPGQTFYATGGIIGNIRQA
C3728v2c6 DC	NNVKTIMVHLNLSVEINCTRPNNNTRKSIRIGPGQTFYATGGIIGNIRQA
C06838v1c48 WT	DNAKTIIVQLNQSIETKIRPNNNTRESIRIGPGQAFYATRDIIGDIRRA
C06838v1c48 DC	DNAKTIIVQLNQSIETKIRPNNNTRESIRIGPGQAFYATRDIIGDIRRA
C06980v1c3 WT	NNVKTIIVHENKSVEIVCTRPNNNTRKSIRIGPGQTFYATGDIIGDIRQA
C06980v1c3 DC	NNVKTIIVHENKSVEIVCTRPNNNTRKSIRIGPGQTFYATGDIIGDIRQA
C06980v0c22 WT	DNVKTIIVHLNLSVEIVCTRPNNNTRKSIRIGPGQTFYATGDIIGDIRQA
C06980v0c22 DC	DNVKTIIVHLNLSVEIVCTRPNNNTRKSIRIGPGQTFYATGDIIGDIRQA

FIGURE 35 (cont.-1)

C3728v2c6 WT	YCTISKNNKWNITLERVATKLKEYFKNTTIQFAPHSGGDLEITTHSFNCRG
C3728v2c6 DC	YCTISKNNKWNITLERVATKLKEYFKNTTIQFAPHSGGDLEITTHSFNCRG
C06838v1c48 WT	YCTIETERWKETLGNVTEKLKEYFPNTNISFAPSSGGDLEITTHSFNCRG
C06838v1c48 DC	YCTIETERWKETLGNVTEKLKEYFPNTNISFAPSSGGDLEITTHSFNCRG
C06980v1c3 WT	YCSINESNWNITLQRVSKKLAEHFPNRTIQFESPSGGDLEITMHSFNCRG
C06980v1c3 DC	YCSINESNWNITLQRVSKKLAEHFPNRTIQFESPSGGDLEITMHSFNCRG
C06980v0c22 WT	YCSINESNWNATLQRVSKKLAEHFPNKTIQFKSPSGGDLEITMHSFNCRG
C06980v0c22 DC	YCSINESNWNATLQRVSKKLAEHFPNKTIQFKSPSGGDLEITMHSFNCRG
C3728v2c6 WT	EFFYCNTSQLFNGT--STGFSNKSTGNETFTLPCRIKQIINMWQEVGRAM
C3728v2c6 DC	EFFYCNTSQLFNGT--STGFSNKSTGNETFTLPCRIKQIINMWQEVGRAM
C06838v1c48 WT	EFFYRNTSKLFN-----NNDTENNLTITLPCRIKQIINMWQGVGRAM
C06838v1c48 DC	EFFYRNTSKLFN-----NNDTENNLTITLPCRIKQIINMWQGVGRAM
C06980v1c3 WT	EFFYCNTSKLKFSTYHPNGTYNLTNGTNSTLIIPCRIKQIINMWQGVGKAI
C06980v1c3 DC	EFFYCNTSKLKFSTYHPNGTYNLTNGTNSTLIIPCRIKQIINMWQGVGKAI
C06980v0c22 WT	EFFYCNTSKLFNGTYYPNGTYYPNGTNSTLIIPCRIKQIINMWQGVGKAI
C06980v0c22 DC	EFFYCNTSKLFNGTYYPNGTYYPNGTNSTLIIPCRIKQIINMWQGVGKAI
C3728v2c6 WT	YAPPIAGNITCVSNITGLLLTRDGGDNNTRTETFRPGGDMRDNRSELY
C3728v2c6 DC	YAPPIAGNITCVSNITGLLLTRDGGDNNTRTETFRPGGDMRDNRSELY
C06838v1c48 WT	YAPPIEGNITCKSNITGLLLTRDGGNGTERNETFRPGGDMRNNNRSELY
C06838v1c48 DC	YAPPIEGNITCKSNITGLLLTRDGGNGTERNETFRPGGDMRNNNRSELY
C06980v1c3 WT	YASPIAGSITCRSNITGLLLTRDGGDTND-TEIFRPTGGDMRDNRSELY
C06980v1c3 DC	YASPIAGSITCRSNITGLLLTRDGGDTND-TEIFRPTGGDMRDNRSELY
C06980v0c22 WT	YASPIAGNITCRSNITGLLLTRDGGDTND-TEIFRPAGGDMRDNRSELY
C06980v0c22 DC	YASPIAGNITCRSNITGLLLTRDGGDTND-TEIFRPAGGDMRDNRSELY
C3728v2c6 WT	KYKVVVEVKPLGVAPSEARRRVVEREKRAVGLGAVFLGFLGAAGSTMGAAS
C3728v2c6 DC	KYKVVVEVKPLGVAPSEARRRVVEREKSAVGLGAVFLGFLGAAGSTMGAAS
C06838v1c48 WT	KYKVVVEIKPLGIAPTGAARRRVVEREKRAVGMGAVFLGFLGAAGSTMGAAS
C06838v1c48 DC	KYKVVVEIKPLGIAPTGAARRRVVEREKSAVGMGAVFLGFLGAAGSTMGAAS
C06980v1c3 WT	KYKIVEIKPLGVAPTEAKRRVVKREKRAVTIGAVFLGFLGAAGSTMGAAS
C06980v1c3 DC	KYKIVEIKPLGVAPTEAKSRVVKSEKSAVTIGAVFLGFLGAAGSTMGAAS
C06980v0c22 WT	KYKIVEIKPLGVAPTEAKRRVVKREKRAVTIGAVFLGFLGAAGSTMGAAS
C06980v0c22 DC	KYKIVEIKPLGVAPTEAKSRVVKSEKSAVTIGAVFLGFLGAAGSTMGAAS
C3728v2c6 WT	ITLTVQARQLLSGIVQQQSNLLRAIEAQQHMLQLTVWGIKQLQTKVLAIE
C3728v2c6 DC	ITLTVQARQLLSGIVQQQSNLLRAIEAQQHMLQLTVWGIKQLQTKVLAIE
C06838v1c48 WT	ITLTVQARQVLSGIVQQQSNLLRAIEAQQHMLQLTVWGIKQLQTRVLAIE
C06838v1c48 DC	ITLTVQARQVLSGIVQQQSNLLRAIEAQQHMLQLTVWGIKQLQTRVLAIE
C06980v1c3 WT	ITLTVQARQLLSGIVQQQSNLLRAIEAQQHMLQLAVWGIKQLQARVLAIE
C06980v1c3 DC	ITLTVQARQLLSGIVQQQSNLLRAIEAQQHMLQLAVWGIKQLQARVLAIE
C06980v0c22 WT	ITLTVQARQLLSGIVQQQSNLLRAIEAQQHMLQLTVWGIKQLQARVLAIE
C06980v0c22 DC	ITLTVQARQLLSGIVQQQSNLLRAIEAQQHMLQLTVWGIKQLQARVLAIE
C3728v2c6 WT	RYLQDQQLLGIWGCSGKLICTTNVPWNSSWSNKSILKAIWNNMTWMQWDRE
C3728v2c6 DC	RYLQDQQLLGIWGCSGKLICTTNVPWNSSWSNKSILKAIWNNMTWMQWDRE
C06838v1c48 WT	RYLKDQQLLGLWGCSGKLICTTAVPWNSSWSNKSILKAIWNNMTWMQWDRE
C06838v1c48 DC	RYLKDQQLLGLWGCSGKLICTTAVPWNSSWSNKSILKAIWNNMTWMQWDRE
C06980v1c3 WT	RYLKDQQLLGIWGCSGKLICTTAVPWNSSWSNRTQDEIWKNTLWMEWDRE
C06980v1c3 DC	RYLKDQQLLGIWGCSGKLICTTAVPWNSSWSNRTQDEIWKNTLWMEWDRE
C06980v0c22 WT	RYLKDQQLLGIWGCSGKLICTTAVPWNSSWSNRTQDEIWKNTLWMEWDRE
C06980v0c22 DC	RYLKDQQLLGIWGCSGKLICTTAVPWNSSWSNRTQDEIWKNTLWMEWDRE

FIGURE 35 (cont.-2)

C3728v2c6 WT	ISNYTNTIYRLLED SHIQQEKNEKD LLALNSWNNLWSW FNITN NWLWYIRI
C3728v2c6 DC	ISNYTNTIYRLLED SHIQQEKNEKD LLALNSWNNLWSW FNITN NWLWYIRI
C06838v1c48 WT	ISNYTNTIYRLLED SQTQQEKNEKD LLALDSWKNLWNWF NITKWLWYIKI
C06838v1C48 DC	ISNYTNTIYRLLED SQTQQEKNEKD LLALDSWKNLWNWF NITKWLWYIKI
C06980v1c3 WT	ISNYTNTIYELLEV SQSQQERNEKD LLALDSWNNLWNWF DISN NWLWYIKI
C06980v1c3 DC	ISNYTNTIYELLEV SQSQQERNEKD LLALDSWNNLWNWF DISN NWLWYIKI
C06980v0c22 WT	ISNYTNTIYELLEV SQSQQERNEKD LLALDSWNNLWNWF DISN NWLWYIKI
C06980v0c22 DC	ISNYTNTIYELLEV SQSQQERNEKD LLALDSWNNLWNWF DISN NWLWYIKI
C3728v2c6 WT	FIMIVGGLIGLR IIFAILSIVNRVRQGYSP LSFQTL IPNQ RGPDR LGR IE
C3728v2c6 DC	FIMIVGGLIGLR IIFAILSIVNRVRQGYSP LSFQTL IPNQ RGPDR LGR IE
C06838v1c48 WT	FIIIVGGLIGLR IIFAVLSIVNRVRQGYSP LSFQTL IPSP RGPD KLGR IE
C06838v1C48 DC	FIIIVGGLIGLR IIFAVLSIVNRVRQGYSP LSFQTL IPSP RGPD KLGR IE
C06980v1c3 WT	FIMIVGGLIGLR IIFAVLSIVNRVRQGYSP LSFQTL IPNQ REPDR PGRI E
C06980v1c3 DC	FIMIVGGLIGLR IIFAVLSIVNRVRQGYSP LSFQTL IPNQ REPDR PGRI E
C06980v0c22 WT	FIMIVGGLIGLR IIFAVLSIVNRVRQGYSP LSFQTL IPNQ REPDR PGRI E
C06980v0c22 DC	FIMIVGGLIGLR IIFAVLSIVNRVRQGYSP LSFQTL IPNQ REPDR PGRI E
C3728v2c6 WT	EEGGEQDKDRSIRLV NGFLALAWDDL RLSLC IFS YHRLRDFILIA TRAVEL
C3728v2c6 DC	EEGGEQDKDRSIRLV NGFLALAWDDL RLSLC IFS YHRLRDFILIA TRAVEL
C06838v1c48 WT	EEGGEQDKGRSV RLVNGFLALAWDDL RLSLC IFS YHQLRDFILIVARAV GL
C06838v1C48 DC	EEGGEQDKGRSV RLVNGFLALAWDDL RLSLC IFS YHQLRDFILIVARAV GL
C06980v1c3 WT	EEGGEQDKDRSIRLV SGFLALAWDDL RLSLC IF YHHLRDIILIAARATE L
C06980v1c3 DC	EEGGEQDKDRSIRLV SGFLALAWDDL RLSLC IF YHHLRDIILIAARATE L
C06980v0c22 WT	EEGGEQDKDRSIRLV SGFLALAWDDL RLSLC IF YHHLRDFILIAARATE L
C06980v0c22 DC	EEGGEQDKDRSIRLV SGFLALAWDDL RLSLC IF YHHLRDFILIAARATE L
C3728v2c6 WT	LGHSSLRGLQRGWEALKYL GSLGQYWGQEL KKSAIRLLDITAI AVAEGTD
C3728v2c6 DC	LGHSSLRGLQRGWEALKYL GSLGQYWGQEL KKSAIRLLDITAI AVAEGTD
C06838v1c48 WT	LGRSSLRGLQRGWEI LKYLGG LVQYWGLEI KKSAV SLFDTIAIAVTEG TD
C06838v1C48 DC	LGRSSLRGLQRGWEI LKYLGG LVQYWGLEI KKSAV SLFDTIAIAVTEG TD
C06980v1c3 WT	LGRSSLRGLQRGWEALKYL GSLVQYWGLEI KKSAINLLDTIAIAVAEG TD
C06980v1c3 DC	LGRSSLRGLQRGWEALKYL GSLVQYWGLEI KKSAINLLDTIAIAVAEG TD
C06980v0c22 WT	LGRSSLRGLQRGWEALKYL GSLVQYWGLEI KKSAINLLDTIAIAVAEG TD
C06980v0c22 DC	LGRSSLRGLQRGWEALKYL GSLVQYWGLEI KKSAINLLDTIAIAVAEG TD
C3728v2c6 WT	RIIEFIQRI CRAIHNV PNRIRQGF EAAAL * (SEQ ID NO. 35)
C3728v2c6 DC	RIIEFIQRI CRAIHNV PNRIRQGF EAAAL * (SEQ ID NO. 36)
C06838v1c48 WT	RIIELIQRSCRAIR NVP TRIRQGF EAAAL * (SEQ ID NO. 37)
C06838v1C48 DC	RIIELIQRSCRAIR NVP TRIRQGF EAAAL * (SEQ ID NO. 38)
C06980v1c3 WT	RIIEIVQACRAV INIL RRIRQGLEAALQ* (SEQ ID NO. 39)
C06980v1c3 DC	RIIEIVQACRAV INIL RRIRQGLEAALQ* (SEQ ID NO. 40)
C06980v0c22 WT	RIIEIVQACRAV INIP RRIRQGLEAALQ* (SEQ ID NO. 41)
C06980v0c22 DC	RIIEIVQACRAV INIP RRIRQGLEAALQ* (SEQ ID NO. 42)

FIGURE 36

C06980v0c22 gp145 amino acid sequence (SEQ ID NO: 32):

1 mdamkrglcc vlllcgavfv ttteasnlwv tvyygvpvwr eakttlfcas dakayerevh
t-Pa Signal

61 nvwathacvp tdpdpqeifl gknvtekfnm wknymvdqmh ediislwdqs lqpcvkltp1

121 citlnctdvt ahngstvydn natvvnstne ikncsfnitt elrdkrkkeh alfnnldivq

181 ldgnsslyrl incntsiikq acpkisfdpi pihycapagf vilkcnnetf ngtgpcnnvs

241 avqcthgikp vvstqlllng slakgeimir senitdnvkt iivhltnsve ivctrpnnt

301 rksirigpgq tfyatgdiig dirqaycsin eswnnatlqr vskklaehfp nkti qfksp

361 ggdleitmhs fncrgeffyc ntsklfngty ypngtyypng tnstliipcr ikqiinmwqg

421 vgkaiyaspi agnitrnsni tgllltrdgg dtndteifrp aggdmrnwr selykykive

481 ikplgvapte aksrvvksek savtigavfl gflgaagstm gaasitltvq arqllsgivq
Cleavage site R to S mutation

541 qgsnllraie aqghmlqltv wgikqlgarv laierylkdq qllgiwgcsq klicttavpw

601 nnswnrtqd eiwknltwme wdreisnytn tiyellevsq sqqernekdl laldswnnlw

661 nwfdisnwlw yikkkk
Triple lysine

FIGURE 37

C06980v0c22 gp145 nucleotide sequence (SEQ ID NO: 20) and translation (SEQ ID NO: 32):

```

1  atggatgcaa tgaagagagg gctctgctgt gtgctgctgc tgtgtggagc agtcttcggt
   >>.....t-Pa leader.....>
   m d a m k r g l c c v l l l c g a v f v

       NheI
       -+----
61  actaccacag aggctagcaa cctgtgggtg accgtgtact acggcgtgcc cgtgtggaga
   >...t-Pa leader..>>
   t t t e a s
       >>.....gp145.....>
       s n l w v t v y y g v p v w r

121  gaggccaaga ccaccctggt ctgcgccagc gacgccaagg cctacgagcg ggaggtgcac
   >.....gp145.....>
   e a k t t l f c a s d a k a y e r e v h

181  aacgtgtggg ccaccacgc ctgcgtgcc accgaccccg accctcagga aatcttcctg
   >.....gp145.....>
   n v w a t h a c v p t d p d p q e i f l

241  ggcaagaacg tgaccgagaa gttcaacatg tggaagaact acatggtgga ccagatgcac
   >.....gp145.....>
   g k n v t e k f n m w k n y m v d q m h

301  gaggacatca tcagcctgtg ggaccagagc ctgcagccct gcgtgaagct gacccccctg
   >.....gp145.....>
   e d i i s l w d q s l q p c v k l t p l

361  tgcatcacc tgaactgcac cgacgtgacc gccacaacg gcagcaccgt gtacgacaac
   >.....gp145.....>
   c i t l n c t d v t a h n g s t v y d n

421  aacgccaccg tggatgaacag caccaacgag atcaagaact gcagcttcaa catcaccacc
   >.....gp145.....>
   n a t v v n s t n e i k n c s f n i t t

481  gagctgaggg acaagaggaa gaaggagcac gccctgttca acaacctgga catcgtgcag
   >.....gp145.....>
   e l r d k r k k e h a l f n n l d i v q

541  ctggacggca acagctccct gtacagactg atcaactgca acaccagcat catcaagcag
   >.....gp145.....>
   l d g n s s l y r l i n c n t s i i k q

601  gcctgcccc aagatcagctt cgaccccatc cccatccact actgcgcccc agccggttc
   >.....gp145.....>
   a c p k i s f d p i p i h y c a p a g f

661  gtgatcctga agtgcaacaa cgagacattc aacggcaccg gccctgttaa caacgtgtcc
   >.....gp145.....>
   v i l k c n n e t f n g t g p c n n v s

```

FIGURE 37 (cont.-1)

```

721 gctgtgcagt gcacccacgg catcaagccc gtggtgtcca cccagctgct gctgaacggc
>.....gp145.....>
a v q c t h g i k p v v s t q l l l n g

781 agcctggcca agggcgagat catgatcaga agcgagaaca tcaccgacaa cgtgaagacc
>.....gp145.....>
s l a k g e i m i r s e n i t d n v k t

841 atcatcgtgc acctgaacaa cagcgtggag atcgtgtgca ccaggcccaa caacaacacc
>.....gp145.....>
i i v h l n n s v e i v c t r p n n n t

901 aggaagagca tcaggatcgg cccaggccag accttctacg ccaccggcga catcatcggc
>.....gp145.....>
r k s i r i g p g q t f y a t g d i i g

961 gacatccggc aggcctactg cagcatcaac gagagcaact ggaacgccac cctgcagagg
>.....gp145.....>
d i r q a y c s i n e s n w n a t l q r

1021 gtgtccaaga agctggccga gcacttcccc aacaagacca tccagttcaa gagccccctt
>.....gp145.....>
v s k k l a e h f p n k t i q f k s p s

1081 ggcggcgacc tggagatcac catgcacagc ttcaactgca gggcgagtt cttctactgc
>.....gp145.....>
g g d l e i t m h s f n c r g e f f y c

1141 aacacatcca agctgtttaa cgggacctac tacccaaacg gcacatacta cccaaacggg
>.....gp145.....>
n t s k l f n g t y y p n g t y y p n g

1201 accaacagca cctgatcat cccatgcagg atcaagcaga tcatcaacat gtggcagggc
>.....gp145.....>
t n s t l i i p c r i k q i i n m w q g

1261 gtgggcaagg ccatctacgc cagcccaatc gccggcaaca tcacctgccg gtccaacatc
>.....gp145.....>
v g k a i y a s p i a g n i t c r s n i

1321 accggcctgc tgctgaccag ggacggcggc gacaccaacg acaccgagat cttcaggcca
>.....gp145.....>
t g l l l t r d g g d t n d t e i f r p

1381 gccggcggag acatgaggga caactggcgg agcgagctgt acaagtacaa gatcgtggag
>.....gp145.....>
a g g d m r d n w r s e l y k y k i v e

1441 atcaagcccc tgggcgtggc tccaaccgag gccaagagca ggggtgtgaa gagcgagaag
>.....gp145.....>
i k p l g v a p t e a k s r v v k s e k

1501 agcgccgtca ccatcggcgc cgtgtttctg ggcttctctg gagccgcccg aagcaccatg
>.....gp145.....>
s a v t i g a v f l g f l g a a g s t m

```

Cleavage site R to S mutations

Cleavage site R to S mutation

FIGURE 37 (cont.-2)

```

1561 ggagccgcct ccatcacctt gaccgtgcag gccaggcagc tgctgtccgg gatcgtgcag
>.....gp145.....>
   g a a s i t l t v q a r q l l s g i v q

1621 cagcagagca acctgctgag agccattgag gctcagcagc acatgctgca gctgacagtg
>.....gp145.....>
   q q s n l l r a i e a q q h m l q l t v

1681 tggggcatca agcagctgca ggccagagtg ctggccatcg agagatacct gaaggatcag
>.....gp145.....>
   w g i k q l q a r v l a i e r y l k d q

1741 cagctgctcg gcatctgggg ctgcagcggc aagctgatct gcaccaccgc cgtgccctgg
>.....gp145.....>
   q l l g i w g c s g k l i c t t a v p w

1801 aacaactcct ggtccaacag gaccaggac gagatctgga agaacctgac ctggatggag
>.....gp145.....>
   n n s w s n r t q d e i w k n l t w m e

1861 tgggacagag agatcagcaa ctacaccaac accatctacg agctgctgga ggtgtcccag
>.....gp145.....>
   w d r e i s n y t n t i y e l l e v s q

1921 agccagcagg agaggaacga gaaggacctg ctggccctgg atagctggaa caacctgtgg
>.....gp145.....>
   s q q e r n e k d l l a l d s w n n l w

1981 aactggttcg acatcagcaa ctggctgtgg tacatcaaga agaagaagtg aattc
>.....gp145.....>>
   n w f d i s n w l w y i k k k k -

```

EcoRI

---+---

Lysine terminus

FIGURE 38

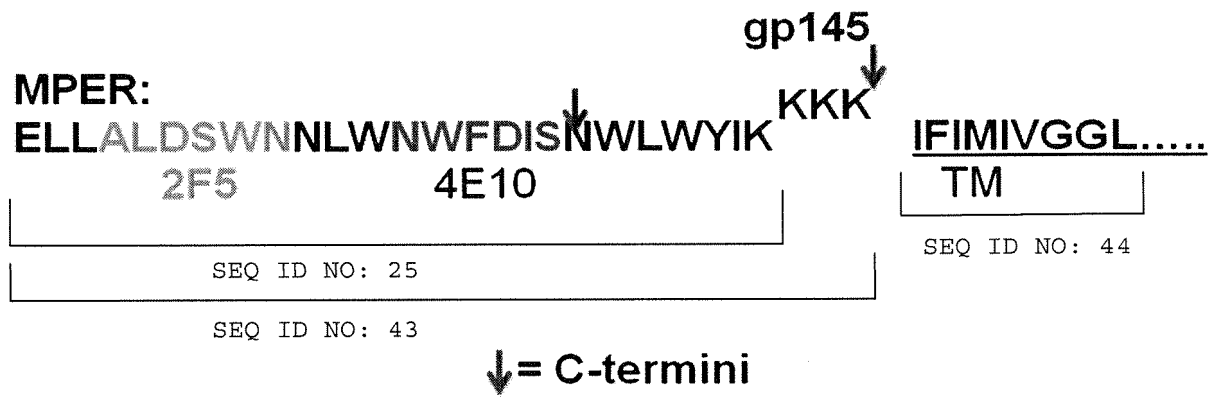


FIGURE 39

Antigen Binding Titers (ug/ml)								
Envelope	Antigen	Clade	Year	2F5	4E10	2G12	b12	VRC01
C06980	gp145	C	2008	>25	1.0	>25	20.0	0.01
57128	gp140	D	1998	>25	>25	0.1	6.7	>25
A07412	gp120	D	1999	>25	>25	>25	>25	0.1
	gp140			>25	>25	22.9	>25	4.3
57140	gp120	D	1998	>25	>25	20.0	>25	0.02
	gp140			>25	>25	>25	>25	0.02

>25
10-25
2-10
<2

FIGURE 40

Clade/CoR.	Envelope	PV Neutralization Sensitivity (ug/ml)				
		2F5	4E10	2G12	b12	VRC01
C/R5	C06980	>25	4.7	>25	>25	0.7
D/R5	57128	>25	11.7	10.5	24.4	>25
D/R5	A07412	1.4	2.2	>25	3.0	0.3
C/R5	20635-4	>25	0.3	>25	0.1	NT
C/R5	93/MW/965	>25	10.3	>25	3.2	NT
C/R5	PBL 288	>25	4.5	>25	>25	NT
D/R5	E13613M4	3.2	12.8	>25	1.1	NT
D/R5	NKU3006	2.1	2.7	>25	>25	NT
D/R5	A03349M1	>25	2.4	>25	1.8	NT

>25
10-25
2-10
<2

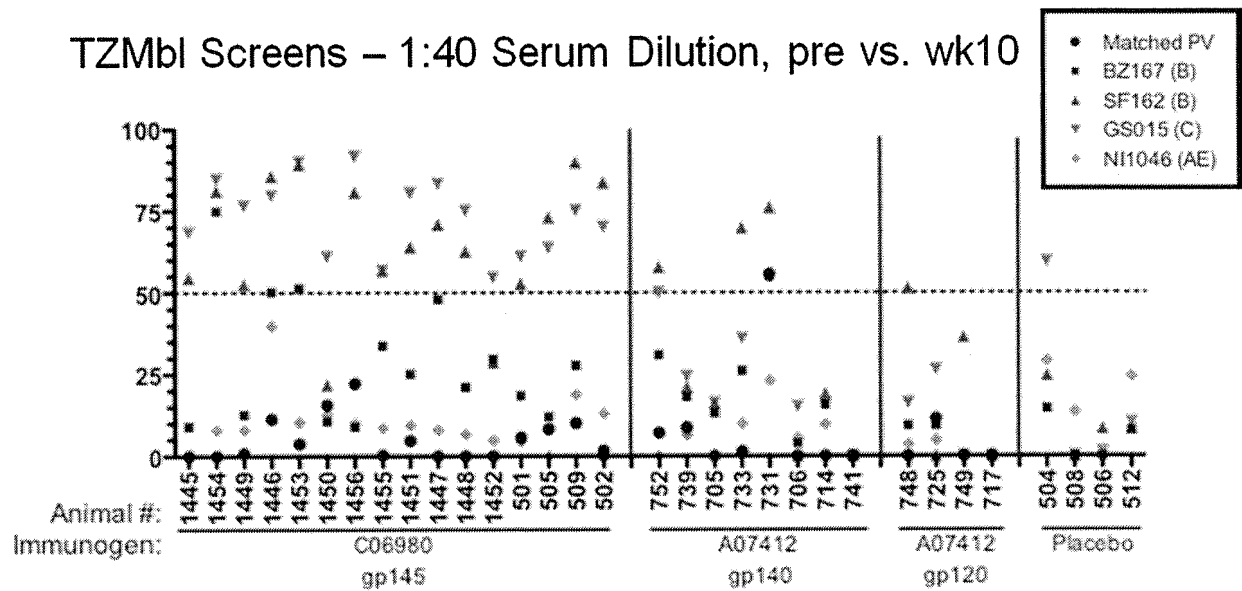
FIGURE 41

6 Groups of 4 New Zealand White females

Group	Antigen	Adjuvant/Vehicle
1	gp145	Alhydrogel
2	gp145	Liposome formulation 1 containing lipid A, preformed and mixed with gp145
3	gp145	Liposome formulation 1 containing lipid A with encapsulated gp145
4	gp145	Liposome formulation 2 containing lipid A and PIP with encapsulated gp145
5	None	Liposome formulation 1 containing lipid A
6	mper23, clade B	Liposome formulation 2 containing lipid A and PIP with encapsulated mper23

- 25 ug dose at weeks 0, 4 and 8
- Bleeds at weeks -2, 0, 4, 8, 10, 12 (terminal)
- Binding titers of 10^4 - 10^5 to:
6980 gp145, CN54 gp140, ZA 1157 gp120

Figure 42



Sera do not neutralize MuLV or the HIV-2/MPER chimera (G Shaw)

FIGURE 43

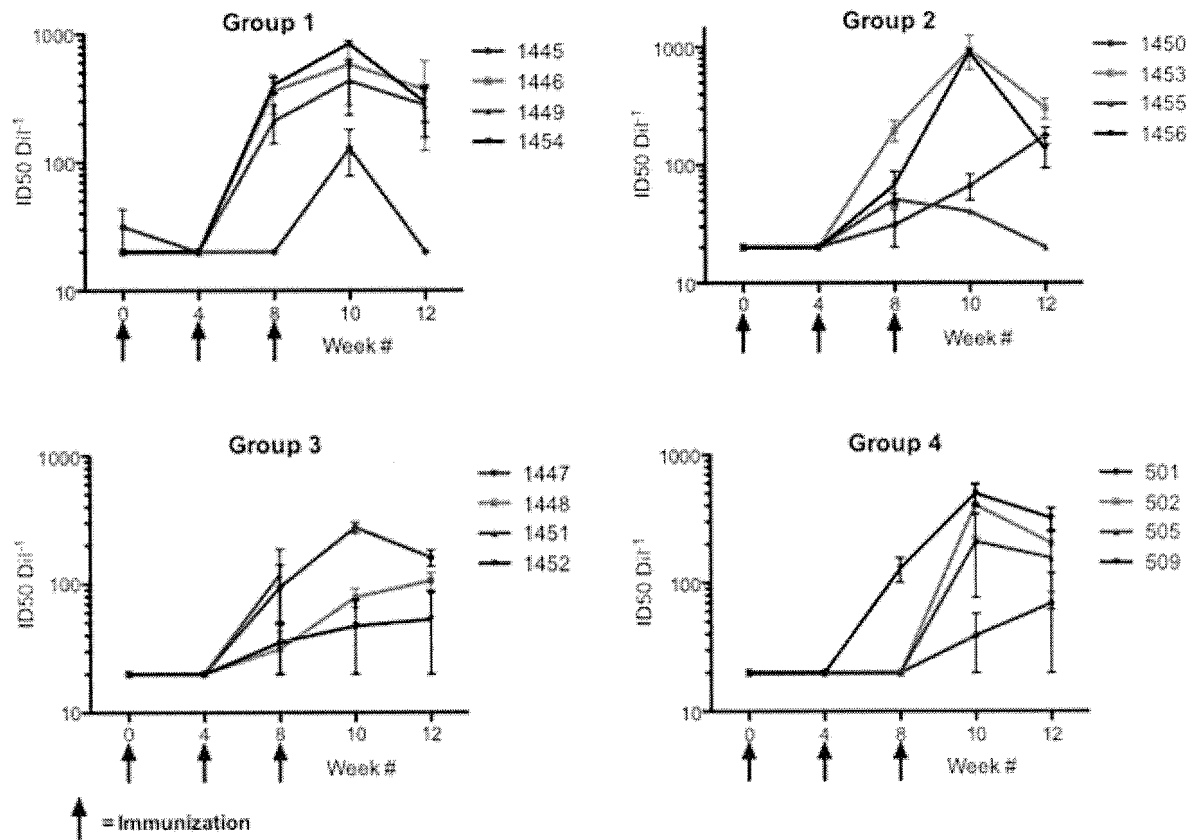


FIGURE 44

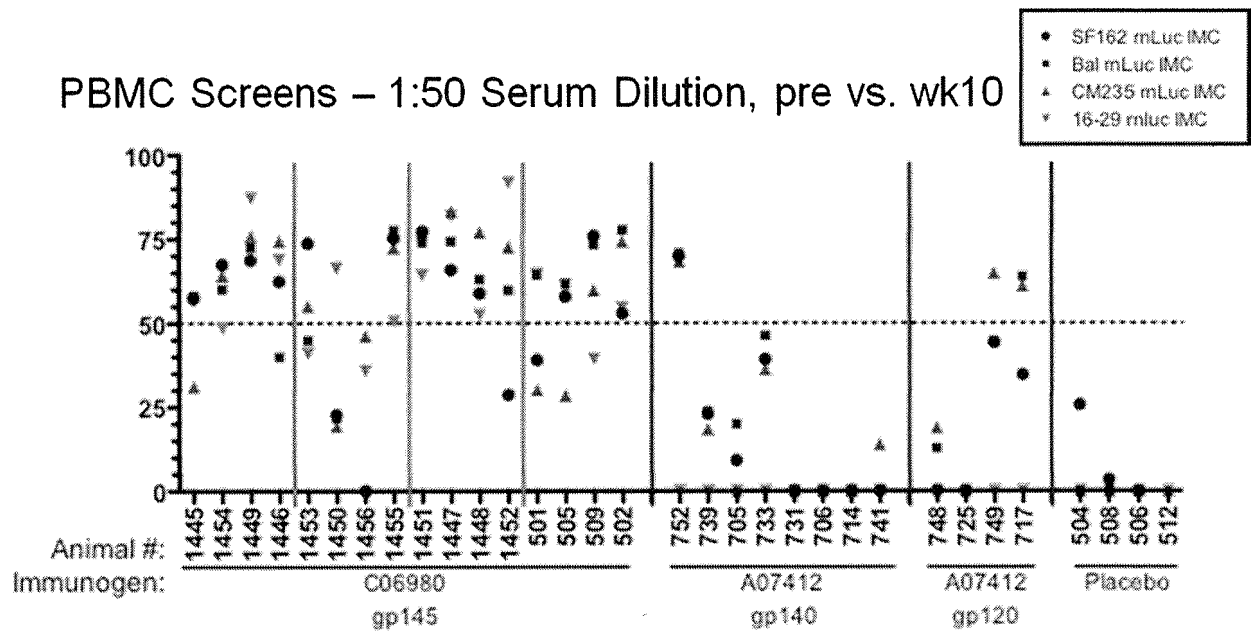


FIGURE 45

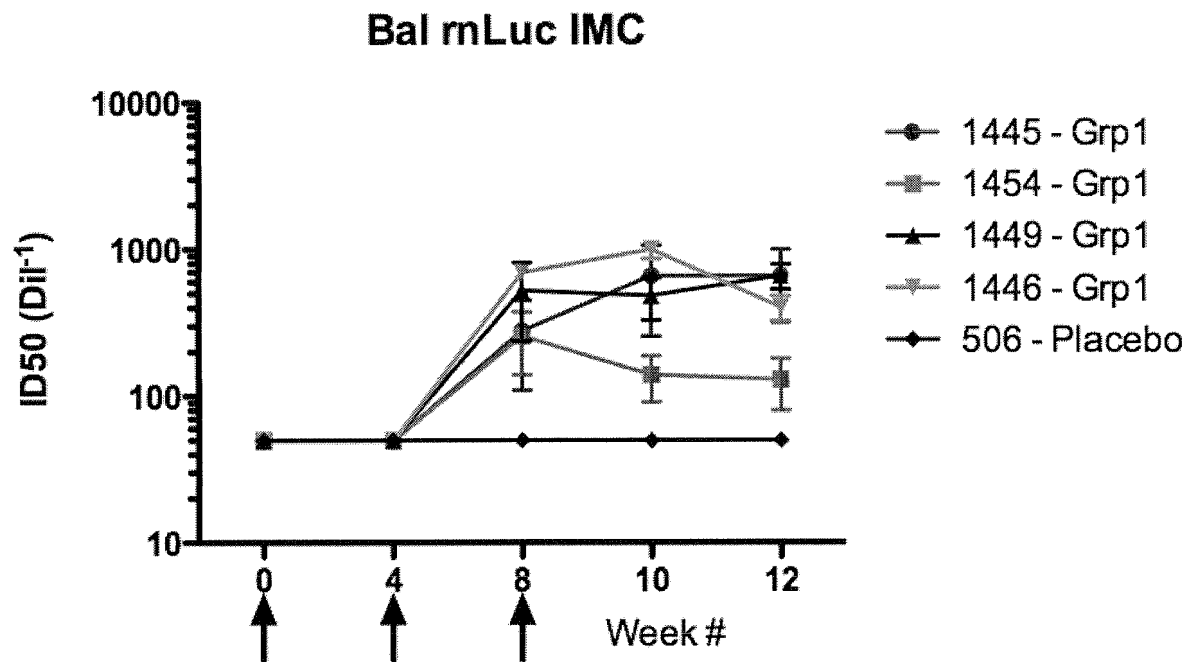
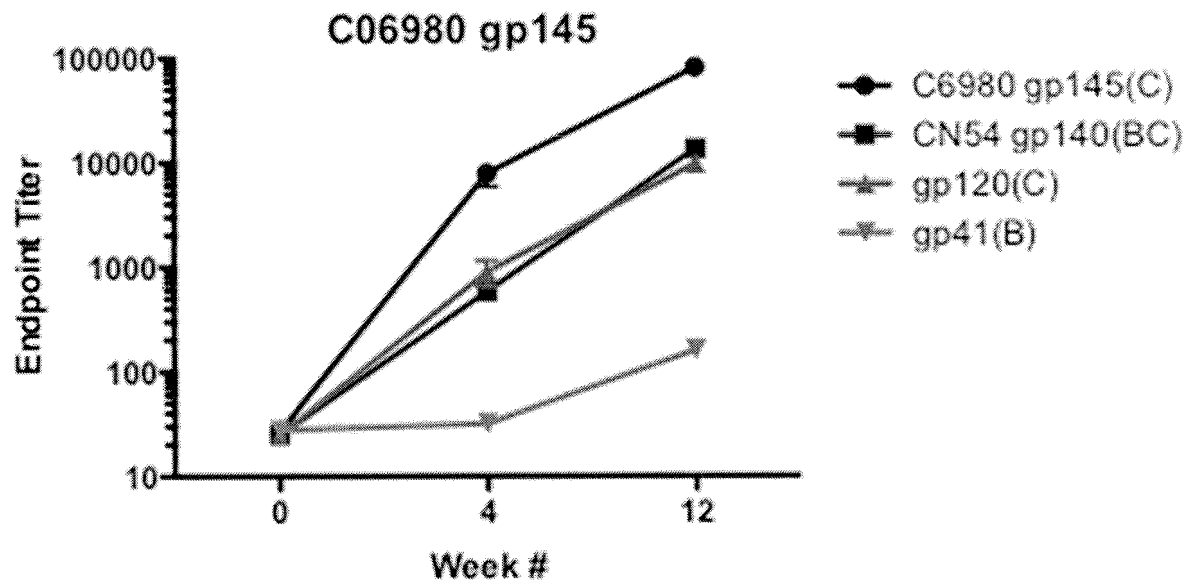


FIGURE 46



Geometric mean titers, all groups (SEM)

(C gp120 = ZA.1197MB, S Africa)

FIGURE 47

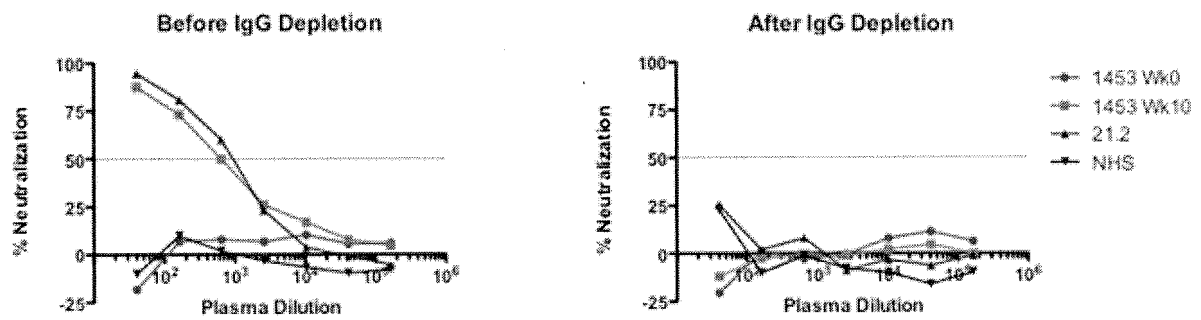


FIGURE 48

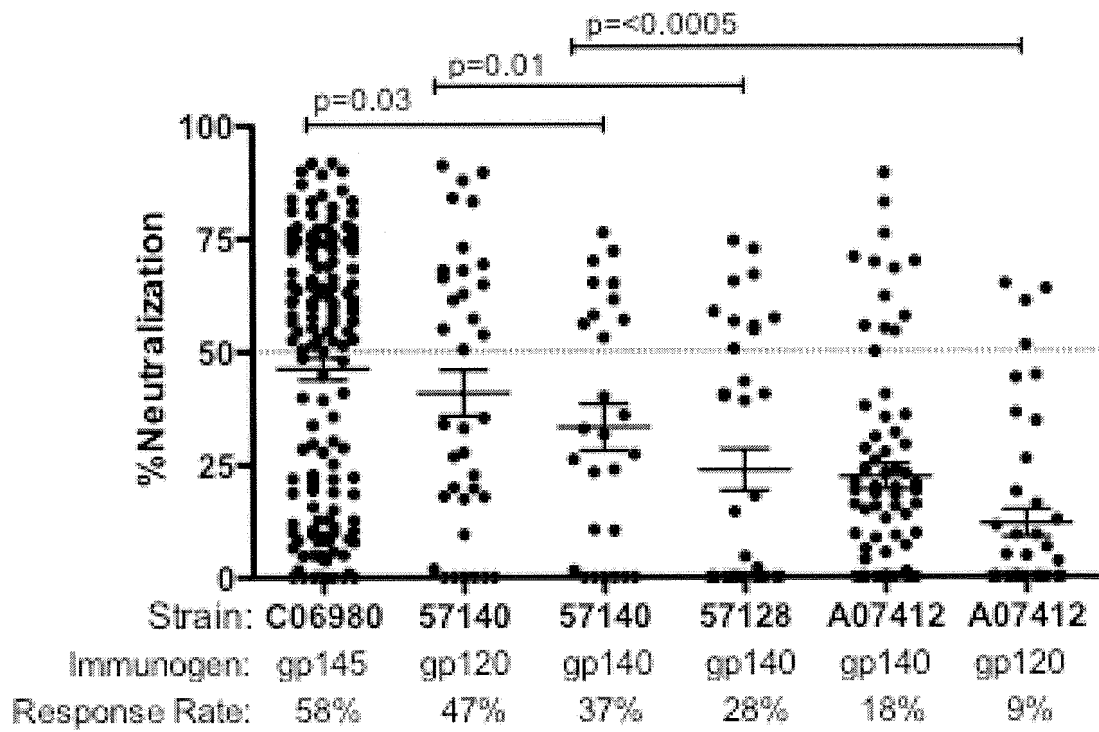
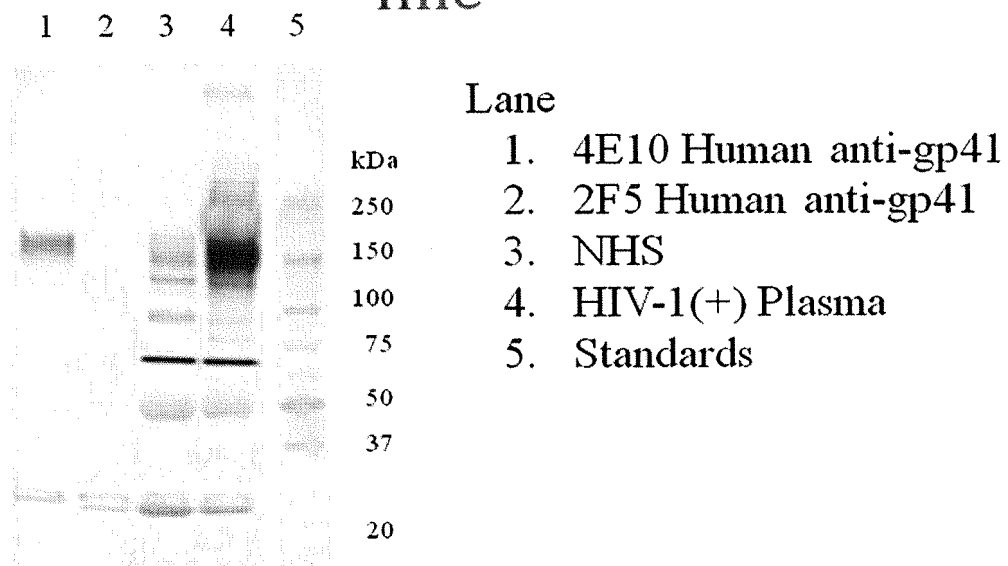


FIGURE 49

I.P. Western Blot of Protein-free media adapted CHO C06980v0c22 gp145 cell line



The gp145 is precipitated from the conditioned media using human antibodies, resolved on 4-15% SDS-PAGE, transferred to PVDF and detected using rabbit antisera to gp120 and gp160.

FIGURE 50

% Sequence Identity:						Potential N-glycosylation sites
	ConC	C06980	A07412	57140	57128	
ConC	100	NA	NA	NA	NA	29
C06980	82.9	100	NA	NA	NA	31
A07412	77.2	71.4	100	NA	NA	29
57140	75.7	71	78.9	100	NA	28
57128	75.7	70.2	77.8	77.8	100	28

FIGURE 51

C06980 SNLKVTVYYGVVWZAKTTLPCASDAKAYEREVHNVWATHACVPTDPDPOEIFLGKNVTEKFNHWNKYNHVDQNHEDIISLWDQSLO
 ConC G.....K.....K.....N.....N.....D.....K.....
 A07412 =S.....K..T.....HKA..A..I.....N.R.MI..=...N.....N..E.....K
 57140 =S.....K..T.....S..A.A..I.....N.R..K.E=...N.....N..E.....K
 57128 =S.....K..T.....KA..A..I.....N....V..=...N.....D..E.....E...K

C06980 PCVXLTPLCITLNCCTOVARNGSTVYONNATVYNSTNEIXNCSFNITTELDRKRKXEHALFNNLDIVQLDGN-----SSLYRLINC
 ConCV....RN..THATNWT..E-----E.....E..VY...YR...P..NE.....E.....
 A07412V.....TYWNGTLQGNHETKGN--RSD.NT.....I...K.Q.T...YK...V.P.EDKDSNKTNYSS.....
 57140VI.D.N--NTNYVNASYVNAT..TEZT.M.....--KQ.V..F.DR..V.PI.K.-----NKS.....
 57128V.....--NAKIEQNVTV.G-----MR.....N....K..K.Q.Y...YK...V..I.NSST---NTD.....

C06980 NTSIIKQACPKISFDPIPIHYCAPAGFVILKCNZTFNGTGPCNNVSAVQCTHGKIPVVSSTQLLNGSLAXGEIMIRSENITONVKT
 ConC ...A.T.....V.....YA.....K.....T.....E.....I.....L.N.A..
 A07412 ...VVT.....VT.E.....A.....K.....K..T.....R.....E.....I.....N.A..
 57140 ...A.T.....T.E.....A.....DKK.....K..T.....E.....I.....L.N.A..
 57128 ...A.T.....T.E.....YA.....K....M...K..T.....R.....E.....V.....L.N.A.I

C06980 IIVHLNNSVEIVCTRPNNTRKSIIRIGPGQTIFYATGDIIGDIRQAYCSINESMNATLQRVSKLAENFP-NKTIQPKSPSGODLEI
 ConCE.....H.N.S.DK..K.....K.....K.EPS.....
 A07412 ...Q..E..T.D.I.....AL.T..~...N...H.NVSKVK.GRM.K..AE..KOLLNQT.N.T..PS...P..
 57140 ...Q..E..P.T.I.....Q.THL...RALFT..-N.V.....H.N.SGEK..K...Q.A...RGL.NGST.MI..PS...P.V
 57128 ...Q..K....N....S....GVH....AI.S..Q.....K.H.N.SRKE..S...Q.T...GSL.N-TTK.I.NAS...P..

C06980 TMHSFNCRGZFFYCNTSKLFNGTYYPNGTY-----YPNGTNSTLIIPCRKIQIINWQGVGKAIYASPIAGNITCRSNITGLLT
 ConC .T.....S..NN.....TNS...ITL.....E...R.N..PP.....K.....T
 A07412 .T....G.....G....SLLNEQ-----PNETS.D.ITLQ.....M..PP...P.S.S.....T
 57140 EH....G.....G...NSVWS.D.SSSND---TSSSS...ITL.....L..E...M..PP...I.K.S.....Q
 57128 .T....N.....AG...-STWNRINSENINSKWNKTEDVNITLQ.....T.....M..PPVS.I.R.S.....T

C06980 RDGGOTND---TEIFRPAGGDMRONNRSELYXYKIVEIKPLGVAPTEAXSRVVKSEKSAYTIGAVFLGFLGAAGSTWGAASITLV
 ConCKK.T=---G.....V.....K.....E..R..G.....
 A07412N.GN--DS.....G.....V.R.E.M.L...R.....ER...IGL..M.....L...
 57140NNDSTNTN...G.....V.RLE...I...R.....ER...IGL..M.....SM...
 57128GADN-NRQN...G.....V.R.E...I...K.R...ER...IGL..L.....T..P...VSM...

C06980 QARQLLSGIYQQQSNNLLRAIEAQQHMLQLTVNGIKQLQARYLAIZRYLXDQQLLGITWGC SGKLICTTAVPNWNSNSNR7QOEIWKNL
 ConCT.....S....KS.ED..D.M
 A07412N.....L.....M.S.....H...T...ST...SVE...N.M
 57140 ...L.....N.....L.....V.S..R.....RH...S...S...KSL.K..N.M
 57128 ...V.....N.....L.....I.V.S.....H....N....S...KSL.Y...M

C06980 TWMZMDREISNYTNTIYELLEVSQSQQERNEXDLLALDSWNNLWNPDISNNLWYIKXXK (SEQ ID NO. 1)
 ConC ...Q.....D...R...D..N...K.....K.....T..... (SEQ ID NO. 2)
 A07412 ...Q.E...E...GL..T.I.E..T...K..QE..Q..K.AS....S.TK (SEQ ID NO. 3)
 57140EK..E...GL..S.I.E..T.....QE..Q..QWAS....S... (SEQ ID NO. 4)
 57128EK..D...EL..S.I...I...K..QE..K...AS....S.TK (SEQ ID NO. 5)

FIGURE 52

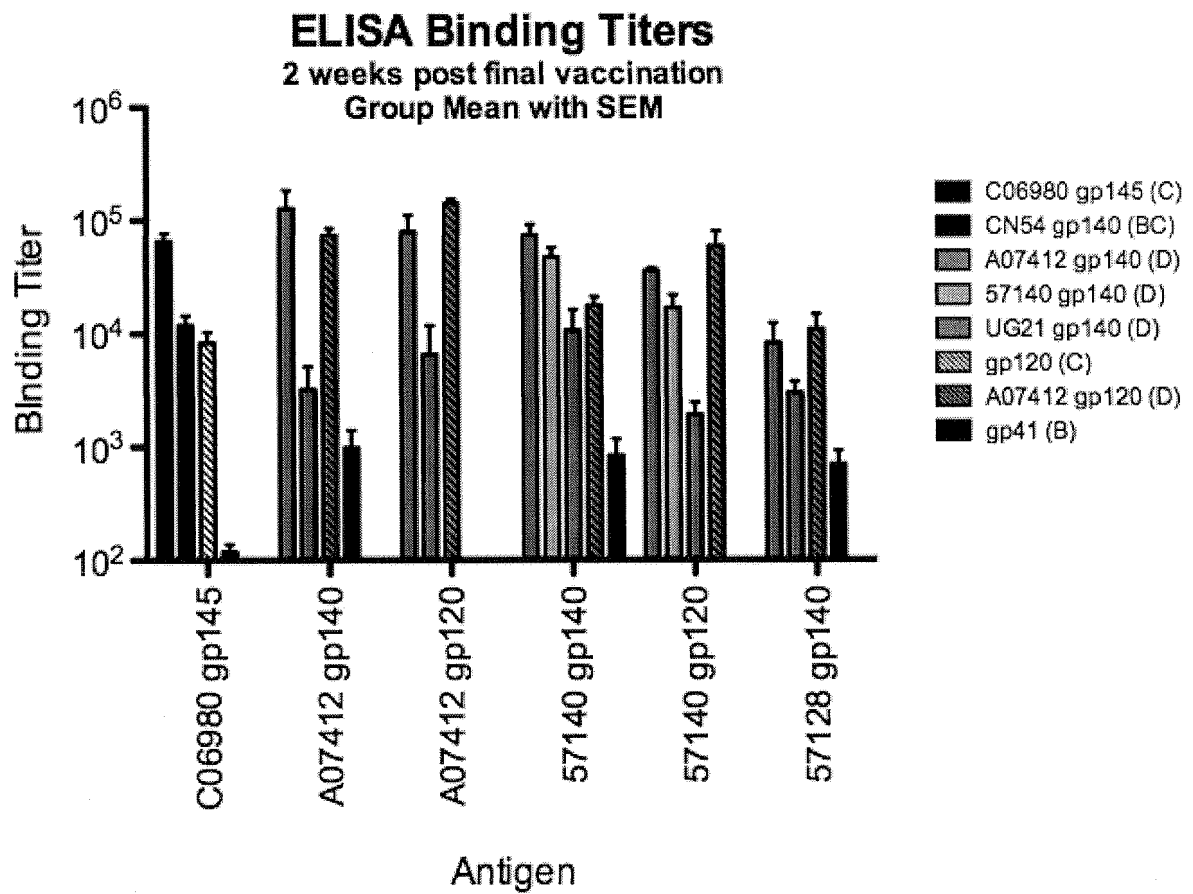


FIGURE 53

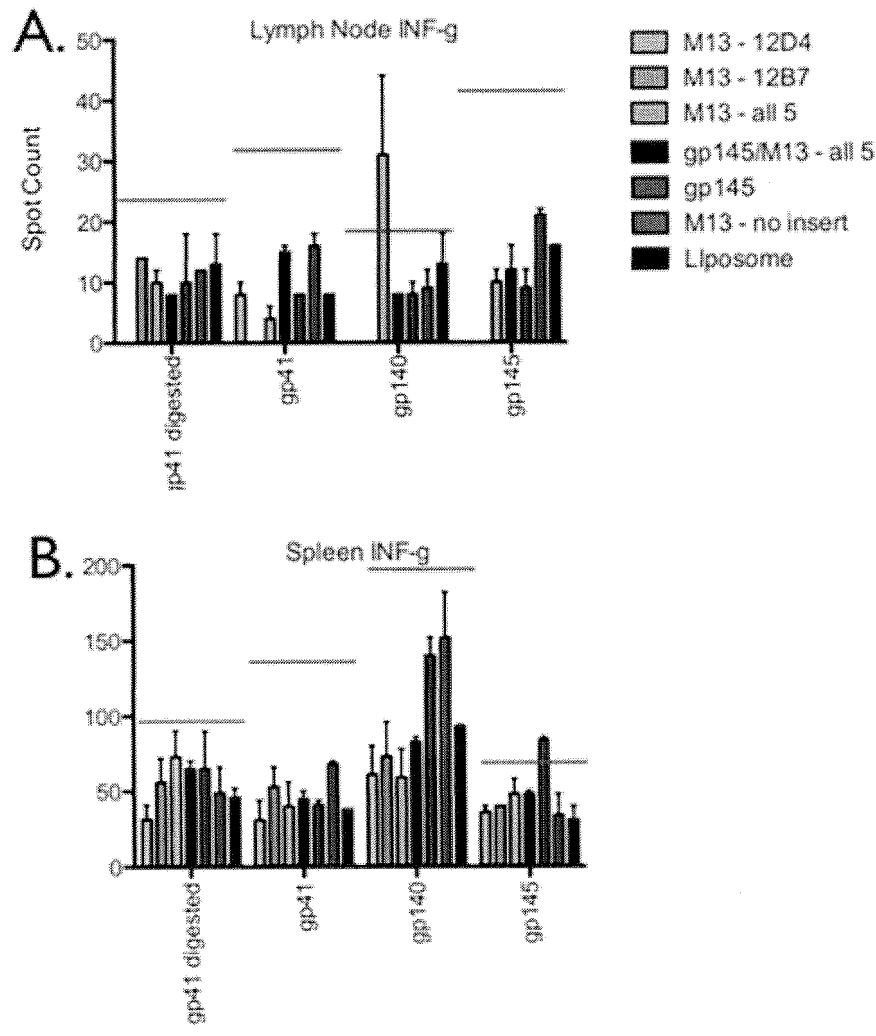


FIGURE 54

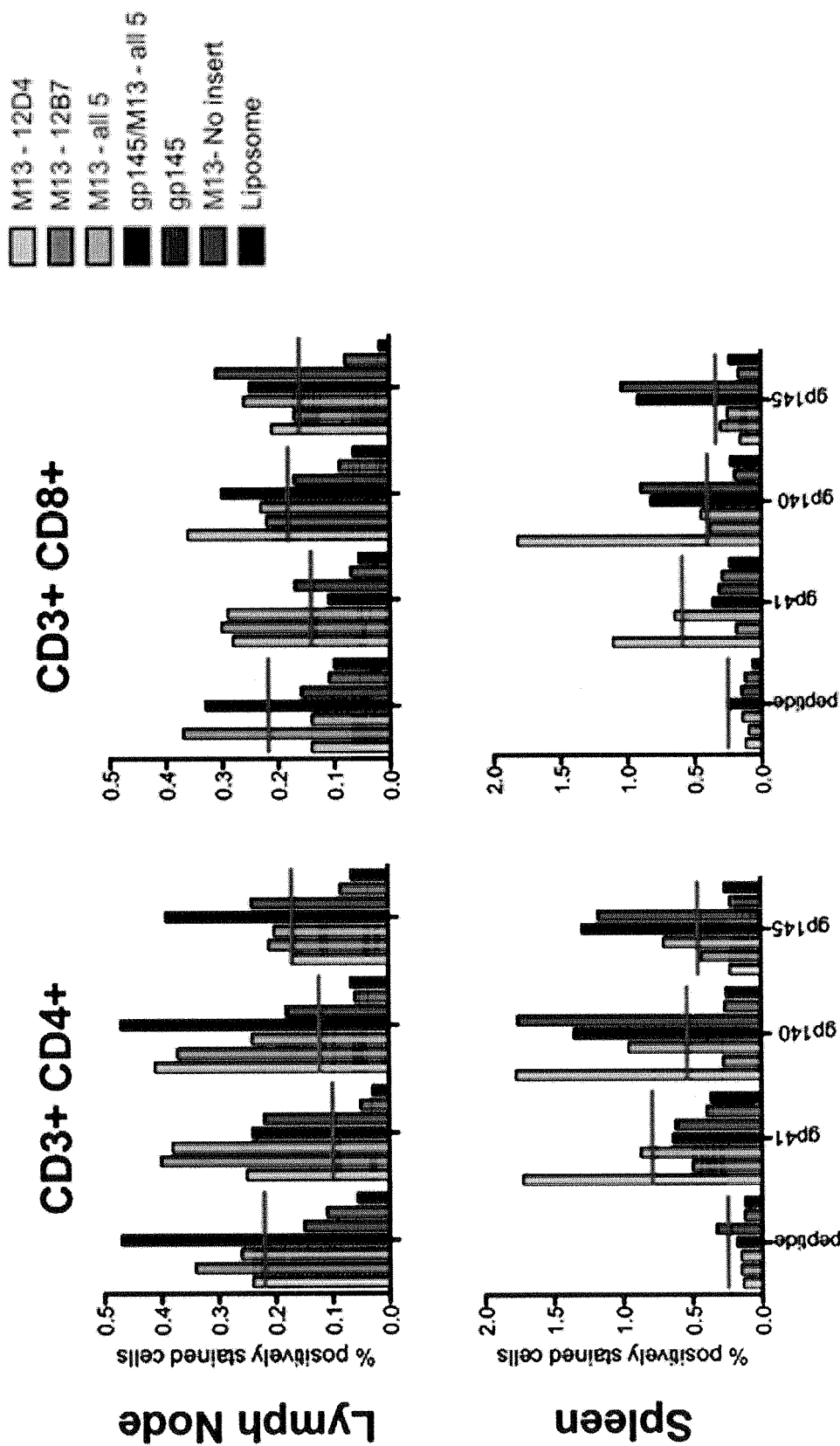


FIGURE 55

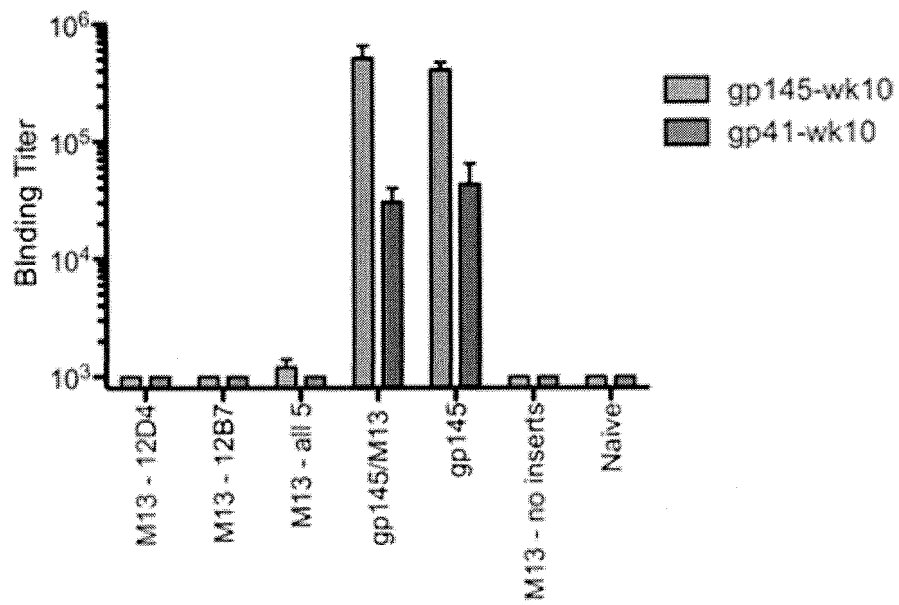


FIGURE 56

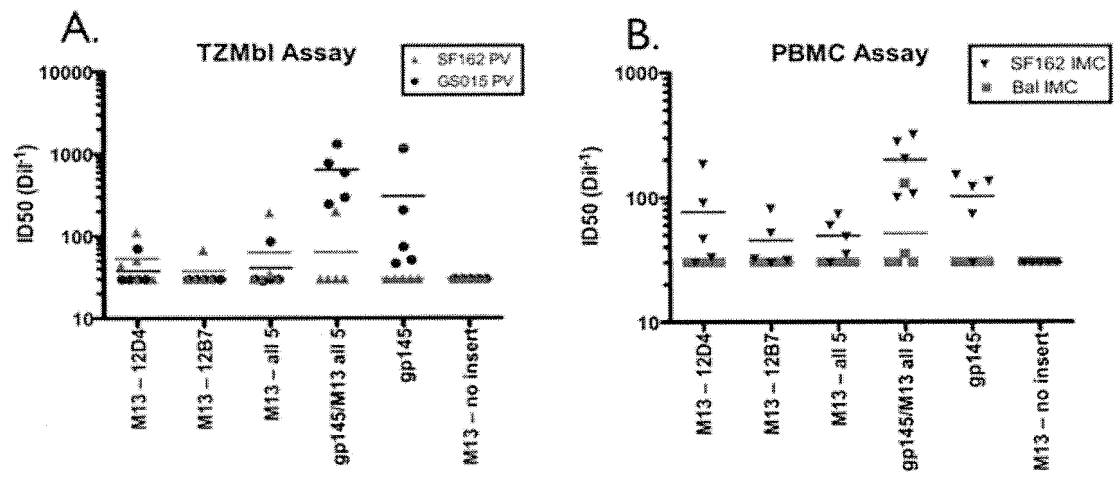


FIGURE 57

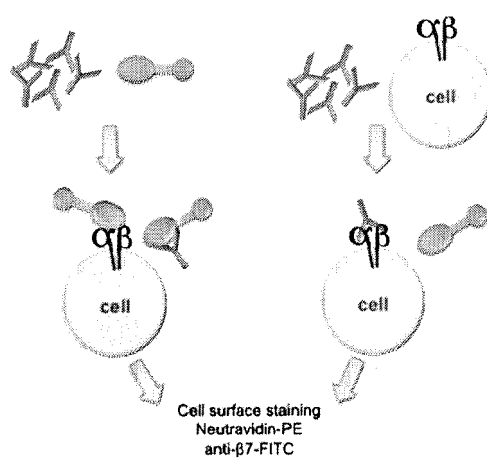


FIGURE 58

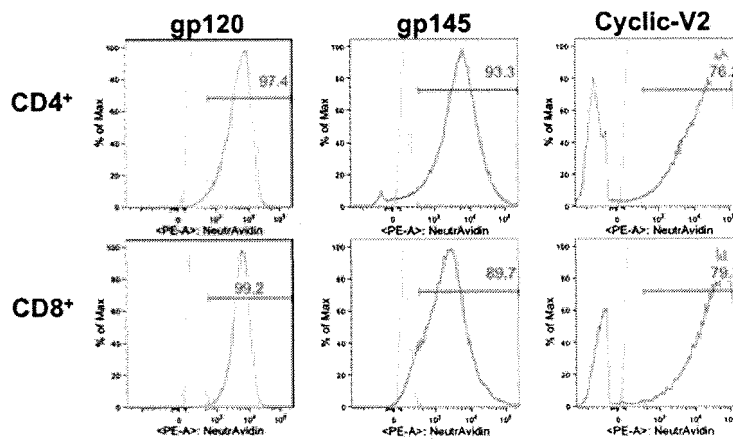


FIGURE 59

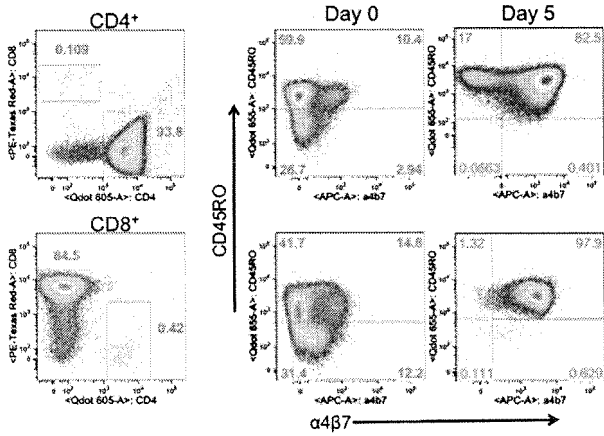


FIGURE 60

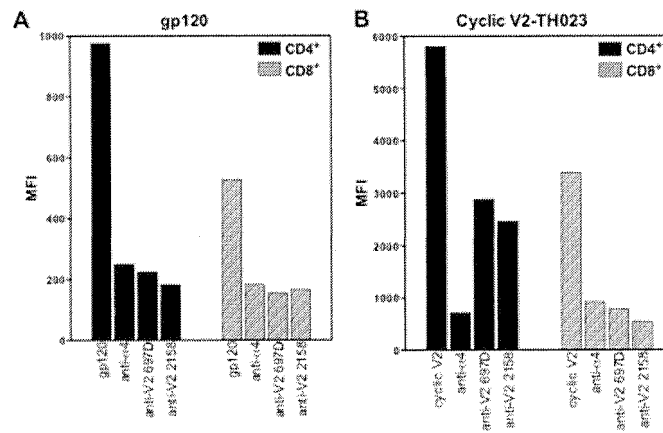


FIGURE 61

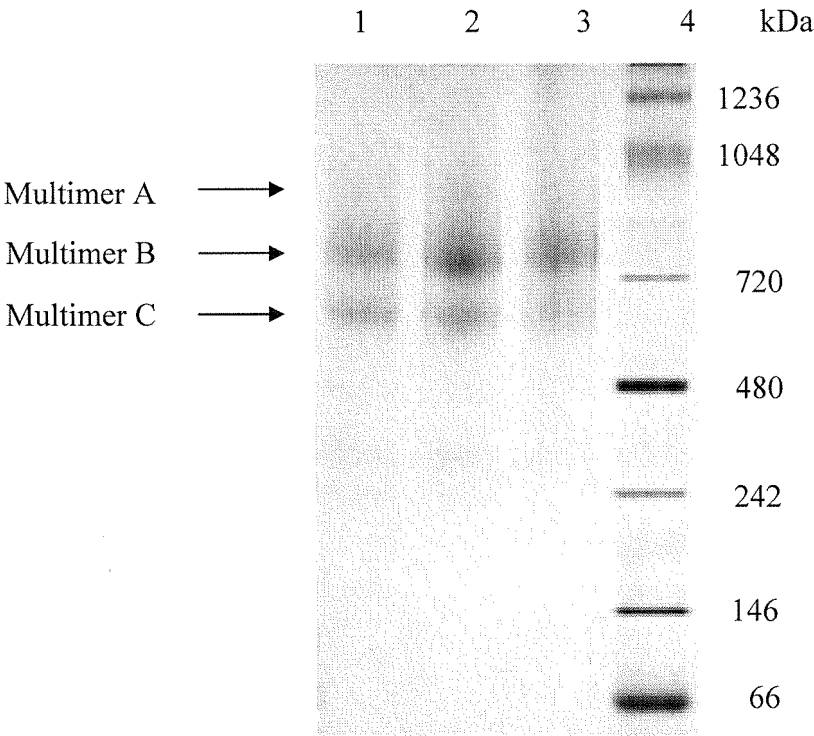


FIGURE 62

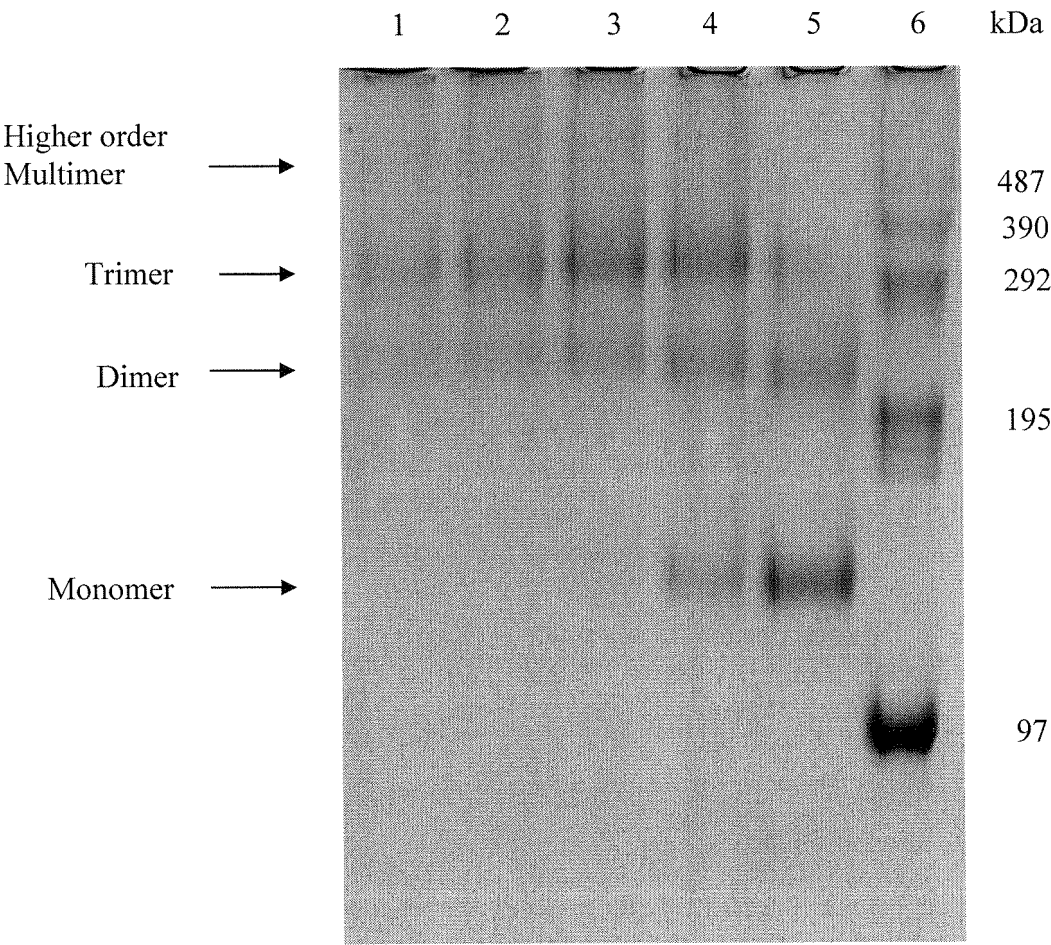
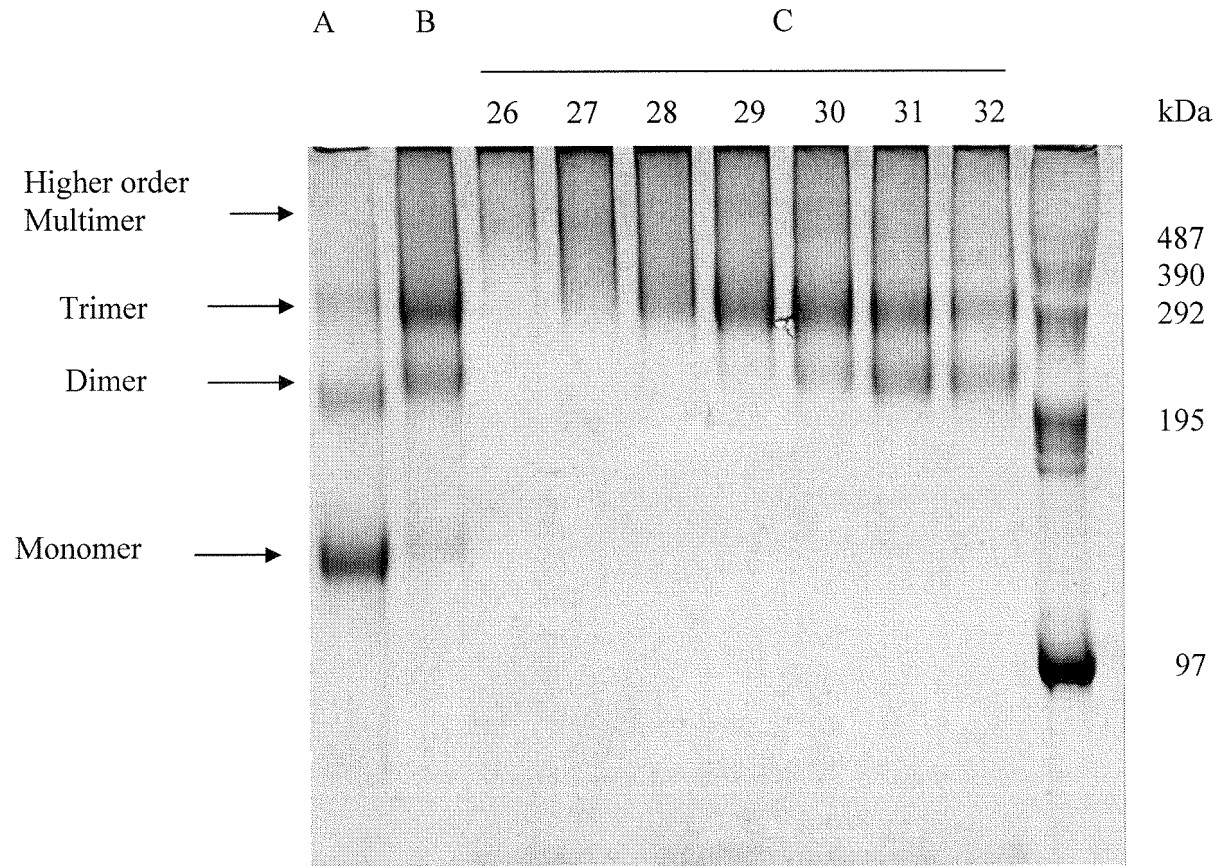


FIGURE 63



INTERNATIONAL SEARCH REPORT

International application No
PCT/US2012/035026

A. CLASSIFICATION OF SUBJECT MATTER
INV. A61K39/21 C07K14/16
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
A61K C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data, EMBL, BIOSIS, CHEM ABS Data, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	W0 03/022869 A2 (PROGENICS PHARM INC) 20 March 2003 (2003-03-20) page 7, line 35 - page 8, line 27 figure 3	1,7-12, 32,42, 44-48
X	----- W0 92/19742 A1 (TRANSGENE SA) 12 November 1992 (1992-11-12) abstract page 3, line 33 - page 4, line 30 ----- -/--	1,4,5, 7-12,32, 41,42



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

24 August 2012

Date of mailing of the international search report

03/09/2012

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Mandl, Birgit

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2012/035026

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BINLEY J M ET AL: "A RECOMBINANT HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 ENVELOPE GLYCOPROTEIN COMPLEX STABILIZED BY AN INTERMOLECULAR DISULFIDE BOND BETWEEN THE GP120 AND GP41 SUBUNITS IS AN ANTIGENIC MIMIC OF THE TRIMERIC VIRION-ASSOCIATED STRUCTURE", JOURNAL OF VIROLOGY, vol. 74, no. 2, 1 January 2000 (2000-01-01), pages 627-643, XP002939087, ISSN: 0022-538X, DOI: 10.1128/JVI.74.2.627-643.2000 the whole document	1
A	----- US 2011/064760 A1 (CHO MICHAEL W [US] ET AL) 17 March 2011 (2011-03-17) abstract	1-60
A	----- MONTERO MARINIEVE ET AL: "The membrane-proximal external region of the human immunodeficiency virus type 1 envelope: dominant site of antibody neutralization and target for vaccine design", MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS, AMERICAN SOCIETY FOR MICROBIOLOGY, US, vol. 72, no. 1, 1 March 2008 (2008-03-01), pages 54-84, tabl, XP002565542, ISSN: 1092-2172, DOI: 10.1128/MMBR.00020-07 the whole document table 3 -----	1-60

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2012/035026

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 03022869	A2	20-03-2003	AT 519500 T 15-08-2011
		AU 2002335709 B2	27-11-2008
		CA 2459426 A1	20-03-2003
		EP 1461079 A2	29-09-2004
		JP 2005502350 A	27-01-2005
		US 2005089526 A1	28-04-2005
		WO 03022869 A2	20-03-2003
-----	-----	-----	-----
WO 9219742	A1	12-11-1992	AT 149203 T 15-03-1997
		CA 2086509 A1	03-11-1992
		DE 69217614 D1	03-04-1997
		DE 69217614 T2	10-07-1997
		DK 0541753 T3	04-08-1997
		EP 0541753 A1	19-05-1993
		ES 2099257 T3	16-05-1997
		FR 2676071 A1	06-11-1992
		GR 3023070 T3	30-07-1997
		JP 3453611 B2	06-10-2003
		JP H05508324 A	25-11-1993
		SG 44828 A1	19-12-1997
		US 6261799 B1	17-07-2001
		US 6284248 B1	04-09-2001
		WO 9219742 A1	12-11-1992
-----	-----	-----	-----
US 2011064760	A1	17-03-2011	NONE
-----	-----	-----	-----

1. 包含截短的 HIV 包膜蛋白的分离的肽,其中所述 HIV 包膜蛋白在天然 gp120/gp41 切割位点处被突变以阻止蛋白酶切割,所述 HIV 包膜蛋白包含 gp41 的 MPER,并在跨膜区前被截短。

2. 权利要求 1 的肽,其中所述 HIV 包膜蛋白在其 C 端包含大约 1-10 个亲水氨基酸。

3. 权利要求 2 的肽,其中所述大约 1-10 个亲水氨基酸是 3 个赖氨酸。

4. 权利要求 1 的肽,其中所述 gp41 的 MPER 包含 4E10 表位。

5. 权利要求 4 的肽,其中所述 gp41 的 MPER 在其 C 端包含氨基酸序列:LWYIK (SEQ ID NO:24)。

6. 权利要求 5 的肽,其中所述 HIV 包膜蛋白包含位于 LWYIK 氨基酸序列 (SEQ ID NO:24) C 端的并与 LWYIK 氨基酸序列 (SEQ ID NO:24) 连续的大约 1-10 个非天然亲水性氨基酸。

7. 权利要求 1 的肽,其中所述 HIV 包膜蛋白衍生自 HIV-1 毒株,所述 HIV-1 毒株归类于选自 M、O、N、P 的组。

8. 权利要求 7 的肽,其中所述 HIV 包膜蛋白衍生自 HIV-1M 组毒株。

9. 权利要求 8 的肽,其中所述 HIV-1M 组毒株选自 A、B、C、D、F、G、H、J 和 K 亚型(进化枝)。

10. 权利要求 9 的肽,其中所述亚型(进化枝)为 B 进化枝。

11. 权利要求 9 的肽,其中所述亚型(进化枝)为 D 进化枝。

12. 权利要求 9 的肽,其中所述亚型(进化枝)为 C 进化枝。

13. 权利要求 12 的肽,其包含与 SEQ ID NO:1 所示氨基酸序列具有 85% 或更高的相同性的氨基酸序列。

14. 权利要求 13 的肽,其中所述肽包含与 SEQ ID NO:1 所示氨基酸序列具有 90% 或更高的相同性的氨基酸序列。

15. 权利要求 14 的肽,其中所述肽包含与 SEQ ID NO:1 所示氨基酸序列具有 95% 或更高的相同性的氨基酸序列。

16. 权利要求 15 的肽,其中所述肽包含与 SEQ ID NO:1 所示氨基酸序列具有 98% 或更高的相同性的氨基酸序列。

17. 权利要求 16 中的肽,其中所述肽包含与 SEQ ID NO:1 所示氨基酸序列具有 99% 或更高的相同性的氨基酸序列。

18. 权利要求 17 的肽,其中所述肽包含 SEQ ID NO:1 所示的氨基酸序列。

19. 分离的核酸,其包含编码权利要求 13-18 中任一项的氨基酸序列的核酸序列。

20. 权利要求 19 的核酸,其中所述核酸序列编码 SEQ ID NO:1 所示的氨基酸序列。

21. 载体,其包含权利要求 19 的核酸。

22. 宿主细胞,其包含权利要求 21 的载体。

23. 权利要求 22 的宿主细胞,其中所述宿主细胞是 CHO 细胞。

24. 制备包含与 SEQ ID NO:1 所示氨基酸序列具有 85% 或更高的相同性的氨基酸序列的肽的方法,所述方法包括在适于蛋白表达的条件下培养权利要求 22 或权利要求 23 的宿主细胞,和分离所述肽。

25. 组合物,其包含权利要求 13-18 中任一项的肽及药用可接受的载体。

26. 在哺乳动物中产生针对 HIV 的抗体的方法,其包括给所述哺乳动物施用权利要求 25 的组合物。

27. 权利要求 26 的方法,其中所述组合物进一步包含佐剂。

28. 在哺乳动物中赋予针对 HIV 的免疫力的方法,其包括给所述哺乳动物施用权利要求 25 的组合物。

29. 权利要求 28 的方法,其中所述组合物进一步包含佐剂。

30. 权利要求 28 的方法,包括通过注射给所述哺乳动物施用所述组合物。

31. 权利要求 26-30 中任一项的方法,其中所述哺乳动物选自人类、非人灵长类、犬、兔、豚鼠和小鼠。

32. 亚单位疫苗,其包含权利要求 1-3 中任一项的肽。

33. 亚单位疫苗,其包含权利要求 13-18 中任一项的肽。

34. 核酸疫苗,其包含权利要求 19 的核酸。

35. 分离的肽,其包含与 SEQ ID NO:3、SEQ ID NO:4 或 SEQ ID NO:5 所示氨基酸序列具有 90% 或更高的相同性的氨基酸序列。

36. 权利要求 35 的肽,其中所述肽包含与 SEQ ID NO:3、SEQ ID NO:4 或 SEQ ID NO:5 所示氨基酸序列具有 98% 或更高的相同性的氨基酸序列。

37. 权利要求 36 的肽,其中所述肽包含与 SEQ ID NO:3、SEQ ID NO:4 或 SEQ ID NO:5 所示氨基酸序列具有 99% 或更高的相同性的氨基酸序列。

38. 权利要求 37 的肽,其中所述肽包含 SEQ ID NO:3、SEQ ID NO:4 或 SEQ ID NO:5 所示的氨基酸序列。

39. 分离的核酸序列,其包含编码权利要求 35-38 中任一项的氨基酸序列的核酸序列。

40. 权利要求 39 的核酸,其中所述核酸序列编码 SEQ ID NO:3、SEQ ID NO:4 或 SEQ ID NO:5 所示的氨基酸序列。

41. 试剂盒,其包含 (a) 包含权利要求 1-3 中任一项的肽和药用可接受的载体的组合物,和 (b) 将所述组合物施用至哺乳动物的说明。

42. 试剂盒,其包含 (a) 权利要求 25 的组合物,和 (b) 将所述组合物施用至哺乳动物的说明。

43. 试剂盒,其包含 (a) 包含权利要求 19 的核酸和药用可接受的载体的组合物,和 (b) 将所述组合物施用至哺乳动物的说明。

44. 权利要求 7-12 中任一项的肽,其中所述 HIV-1 毒株分离自具有急性 HIV-1 感染的个体。

45. 权利要求 7-12 中任一项的肽,其中所述 HIV-1 毒株分离自具有慢性 HIV-1 感染的个体。

46. 权利要求 1 的肽,其中所述 HIV 包膜蛋白在前导序列处被突变。

47. 权利要求 46 的肽,其中天然信号肽被 tPA 信号肽替换。

48. 权利要求 47 中的肽,其中所述 tPA 信号肽包含选自 SEQ ID NO:21 和 SEQ ID NO:22 的序列。

49. 分离的肽,其包含与 SEQ ID NO:7 或 SEQ ID NO:9 所示氨基酸序列具有 90% 或更高的相同性的氨基酸序列。

50. 权利要求 49 的肽,其中所述肽包含与 SEQ ID NO:7 或 SEQ ID NO:9 所示氨基酸序列具有 98% 或更高的相同性的氨基酸序列。

51. 权利要求 50 的肽,其中所述肽包含与 SEQ ID NO:7 或 SEQ ID NO:9 所示氨基酸序列具有 99% 或更高的相同性的氨基酸序列。

52. 权利要求 51 的肽,其中所述肽包含 SEQ ID NO:7 或 SEQ ID NO:9 所示的氨基酸序列。

53. 分离的核酸序列,其包含编码权利要求 49-52 中任一项的氨基酸序列的核酸序列。

54. 权利要求 53 的核酸,其中所述核酸序列选自 SEQ ID NO:3 和 SEQ ID NO:5。

55. 权利要求 19 的核酸,其中所述分离的核酸包含 SEQ ID NO:20 所示的核酸序列。

56. 权利要求 27 或 29 的方法,其中所述佐剂包含脂质体制剂。

57. 权利要求 56 的方法,其中所述脂质体制剂包含以下的一种或多种:二肉豆蔻磷脂酰胆碱、二肉豆蔻磷脂酰甘油、胆固醇和磷脂。

58. 权利要求 57 的方法,其中所述脂质体制剂包含磷脂 A。

59. 权利要求 26 的方法,其中所述在哺乳动物中产生的抗体为与包含截短的 HIV 包膜蛋白的肽竞争结合整联蛋白 $\alpha 4 \beta 7$ 的抗体。

60. 权利要求 1 的分离的肽,其中所述肽结合整联蛋白 $\alpha 4 \beta 7$ 。

截短的 HIV 包膜蛋白 (ENV)、其相关方法和组合物

[0001] 联邦资金

[0002] 本发明部分由政府支持,合作协议号码为 W81XWH-07-2-0067。联邦政府在本发明中享有相应权利。

[0003] 相关申请的交叉引用

[0004] 本申请要求 2011 年 4 月 25 日提交的美国临时专利申请系列号 61/478,857 的优先权权益,其通过引用并入本文。

发明领域

[0005] 本发明大体上涉及新的 HIV 包膜蛋白和其相关方法及组合物。更具体地,本发明涉及涉及用于制备、生产和施用分离的新 HIV 包膜核酸和蛋白质序列(其适于,例如在某些实施方案中,作为 HIV 疫苗)的方法及组合物。

[0006] 发明背景

[0007] AIDS 或获得性免疫缺陷综合征,由人类免疫缺陷病毒(HIV)引起,具有以下几种临床特征,包括:消耗综合征、中枢神经系统退化和导致机会致病菌感染和恶性肿瘤的深度免疫抑制(profound immunosuppression)。HIV 是动物反转录病毒慢病毒(lentivirus)属的成员,该属包括羊的羊髓鞘脱落病毒和牛、猫科及猴免疫缺陷病毒。迄今,人们已鉴定出两种紧密相关的 HIV(HIV) 类型,称为 HIV-1 和 HIV-2,其中 HIV-1 是 AIDS 最常见的致病原因。然而,HIV-2 尽管在基因组结构和抗原性上不同,却导致相似的临床症状。

[0008] 在全球流行的占主要地位的 HIV-1 的形式被称为 HIV-1 的主要组。一共有三种 HIV-1 组,主要组(M组)、异常组(outlier group, O组)和非 M/非 O组(N组)。还有 P 组。M组进一步被分为九种不同的遗传亚型,这些亚型通常称作进化枝(clade)和循环重组体型(CRF)。HIV-1 的 M 组的亚型/进化枝标记为 A、B、C、D、F、G、H、J 和 K。B 进化枝在美国最为普遍,而 C 进化枝在世界范围内最为普遍。CRF01_AE 或之前的 E 进化枝和 CRF02_AG 在 HIV-1 的传染中是最为普遍的亚型间重组株。遗传亚型和亚型间重组形式的地理分布持续改变,现有的数据提供的是不完全的估计。

[0009] 感染性 HIV 颗粒由两条相同的 RNA 链组成,每条链大约 9.2kb 长,包装在一个病毒蛋白质核心中。该核心结构由来自宿主细胞膜的磷脂双层包膜围绕,其中所述磷脂双层包膜还包括病毒编码的膜蛋白(Abbas 等人,Cellular and Molecular Immunology,第 4 版,W.B. Saunders Company,2000,454 页)。HIV 基因组具有反转录病毒家族的特征性的 5' -LTR-Gag-Pol-Env-LTR-3' 组织结构。在病毒基因组的每一端的长末端重复(LTR)充当来自宿主的转录调控蛋白的结合位点,并调控病毒对宿主基因组的整合、病毒基因表达和病毒复制。

[0010] HIV 基因组编码几种结构蛋白。gag 基因编码核壳核心和基质的结构蛋白。pol 基因编码病毒复制过程所需的反转录酶(RT)、整合酶(IN)和病毒蛋白酶(PR)。tat 基因编码病毒转录物延伸所需的蛋白。rev 基因编码促进不完全剪接和未剪接的病毒 RNA 出核转运的蛋白。vif 基因产物增强病毒颗粒的感染性。vpr 基因产物促进病毒 DNA 的入核转运

和调控 G2 细胞周期停滞。vpu 和 nef 基因编码的蛋白质下调宿主细胞 CD4 的表达,同时增强病毒从受感染细胞的释放。env 基因编码病毒包膜糖蛋白,该蛋白可翻译为 160 千道尔顿 (kDa) 的前体 gp160,并被细胞蛋白酶切割生成外部 120-kDa 的包膜糖蛋白 (gp120) 和跨膜 41-kDa 的包膜糖蛋白 gp41,其在细胞感染中是必需的。(Abbas 等人, *Cellular and Molecular Immunology*, 4th edition, W. B. Saunders Company, 2000, pp. 454-456)。gp140 是 Env 糖蛋白的修饰形式,包含外部 120-kDa 包膜糖蛋白部分和 Env 的 gp41 部分的胞外部分,并具有 gp120 和 gp41 两者的特征。nef 基因在灵长类慢病毒中是保守的,并且是感染后最早转录的基因之一。在体外研究中,描述了若干种功能,包括下调 CD4 和 I 类 MHC 的表达、改变 T- 细胞信号转导和激活、以及增强病毒感染性。

[0011] HIV 感染起始于病毒颗粒上的 gp120 与靶细胞的细胞膜上的 CD4 和趋化因子受体分子 (例如 CXCR4, CCR5) 的结合,所述靶细胞诸如 CD4⁺T- 细胞、巨噬细胞和树突细胞。结合的病毒与靶细胞融合,并反转录 RNA 基因组。获得的病毒 DNA 整合进入细胞基因组,并指导产生新的病毒 RNA,并由此产生病毒蛋白质和新的病毒体。这些病毒体从受感染细胞膜上出芽并在其他细胞建立增殖性感染。这个过程也杀死最初受感染的细胞。HIV 也可以间接杀死细胞,这是由于未感染的 T- 细胞上的 CD4 受体具有对表达于受感染细胞表面上的 gp120 的强亲和性。在这些状况中,这些未感染细胞经由 CD4 受体与 gp120 的相互作用结合在感染细胞上并融合为合胞体,所述合胞体不能存活。CD4⁺T- 淋巴细胞对免疫防卫反应非常重要,CD4⁺T- 淋巴细胞的破坏是进行性免疫功能障碍的主要原因,而进行性免疫功能障碍是 AIDS 疾病发展的标志。CD4⁺T 细胞的缺失严重损害人体与大多数外来侵入物作战的能力,尤其剧烈的影响人体对病毒、真菌、寄生虫、某些细菌包括分支杆菌的防御。

[0012] 对 Env 糖蛋白的研究表明该病毒具有很多有效的保护性机制而具有很少弱点 (Wyatt & Sodroski, *Science*. 1998 Jun 19 ; 280 (5371) :1884-8)。为了与靶细胞融合,HIV-1 使用包含 gp120 与 gp41 亚基的三聚体 Env 复合物 (Burton 等人, *Nat Immunol*. 2004 Mar ; 5 (3) :233-6)。与 CD4 受体和共受体 (通常是 CCR5 或 CXCR4) 的接合触发了 Env 复合物的感染潜力。中和抗体或者通过与病毒体表面上成熟的三聚体结合并阻止最初的受体接合事件,或者通过病毒附着后的结合抑制融合过程。(Parren & Burton, *Adv Immunol*. 2001 ; 77 : 195-262)。在后一种情况下,中和抗体可能与因受体结合而增强或触发暴露的表位结合。尽管中和抗体具有以上潜在的抗病毒作用,目前已知 HIV-1 已经进化出多种机制保护自身免受抗体的结合 (Johnson & Desrosiers, *Annu Rev Med*. 2002 ; 53 :499-518)。

[0013] HIV-1 实验疫苗在人体上和 / 或非人类的灵长类动物的大部分实验建议一种成功的疫苗应包含可以引起广谱中和抗体 (bNab) 和强的细胞介导免疫的免疫原。HIV-1 包膜糖蛋白 (Env) 是与病毒进入有关的主要病毒蛋白,也是中和抗体的主要靶点,但是由于免疫逃避策略和包膜蛋白基因剧烈的序列变异性,产生 bNab 是令人畏缩的任务。(Phogat S, Wyatt R. *Curr Pharm Des*. 2007, 13 :213-27 ; Phogat S, 等人 *J Intern Med*. 2007 262 : 26-43, Karlsson Hedestam GB, 等人 *Nat Rev Microbiol*. 2008, 6 :143-55)

[0014] 开发有效的疫苗来预防 HIV 感染或中和 HIV 感染是困难的。引发广谱的和有效的中和抗体的能力在开发 HIV-1 疫苗中是主要的挑战。主要的目标是研制 HIV 疫苗,其可以有效地引发特异性的抗病毒中和抗体以及引起细胞介导的免疫应答从而预防感染和控制 HIV 的传播,并且对跨越不同遗传进化枝的病毒具有相当广泛的反应性。HIV 高度的遗传

多样性对疫苗的研制而言是难以解决的问题。

[0015] 在本申请中引用和标识的任何文献并不代表这样的文献是本发明可获得的现有技术。

[0016] 发明概述

[0017] 在某些实施方案中,本申请提供了包含截短的 HIV 包膜蛋白的分离的肽,其中所述 HIV 包膜蛋白在天然的 gp120/gp41 切割位点处被突变以阻止蛋白酶切割,包含 gp41 的 MPER,且在跨膜区前被截断。

[0018] 在某些实施方案中,HIV 包膜蛋白在其 C 端包含大约 1 到 10 个亲水氨基酸。在某些实施方案中,所述大约 1 到 10 个亲水氨基酸是 3 个赖氨酸。

[0019] 在某些实施方案中,gp41 的 MPER 包含 4E10 表位。在某些实施方案中,gp41 的 MPER 在其 C 端包含氨基酸序列:LWYIK (SEQ ID NO:24)。在进一步的实施方案中,HIV 包膜蛋白包含在 LWYIK 氨基酸序列 (SEQ ID NO:24) C 端的并与 LWYIK 氨基酸序列 (SEQ ID NO:24) 连续的大约 1-10 个非天然亲水性氨基酸。在某些实施方案中,HIV 包膜蛋白与整联蛋白 $\alpha 4 \beta 7$ 结合。

[0020] 在某些实施方案中,HIV 包膜蛋白衍生自 HIV-1 毒株,该 HIV 毒株归类于选自 M、O、N、P 的组。在某些实施方案中,所述 HIV-1 毒株分离自急性 HIV-1 感染的个体。在另外的实施方案中,所述 HIV-1 毒株分离自慢性 HIV-1 感染的个体。在某些实施方案中,HIV 包膜蛋白衍生自 HIV-1 的 M 组毒株。在更进一步的实施方案中,HIV-1M 组毒株是选自 A、B、C、D、F、G、H、J 和 K 的亚型(进化枝)。在某些具体的实施方案中,所述亚型(进化枝)是 B 进化枝。在另一个实施方案中,所述亚型(进化枝)是 D 进化枝。然而另外的实施方案中,所述亚型是 C 进化枝。

[0021] 在某些实施方案中,HIV 包膜蛋白包含与 SEQ ID NO:1 所示氨基酸序列具有 85% 或更高的相同性的氨基酸序列。

[0022] 在更进一步的实施方案中,HIV 包膜蛋白包含与 SEQ ID NO:1 所示氨基酸序列具有 90% 或更高的相同性的氨基酸序列。在另外一些实施方案中,肽包含与 SEQ ID NO:1 所示氨基酸序列具有 95% 或更高的相同性的氨基酸序列。在某些实施方案中,肽包含与 SEQ ID NO:1 所示氨基酸序列具有 98% 或更高的相同性的氨基酸序列。在另外一些实施方案中,肽包含与 SEQ ID NO:1 所示氨基酸序列具有 99% 或更高的相同性的氨基酸序列。在一具体的实施方案中,肽包含与 SEQ ID NO:1 所示的氨基酸序列。

[0023] 在某些实施方案中,本申请涉及分离的核酸,所述核酸包含编码与 SEQ ID NO:1 所示氨基酸序列具有 85%、90%、95%、98%、99% 或更高相同性的氨基酸序列的核酸序列。在某些实施方案中,所述核酸序列编码 SEQ ID NO:1 所示的氨基酸序列。在一具体的实施方案中,分离的核酸包含 SEQ ID NO:20 所示的核酸序列。

[0024] 在另外一些实施方案中,本申请涉及载体,其包含编码与 SEQ ID NO:1 所示氨基酸序列具有 85%、90%、95%、98%、99% 或更高相同性的氨基酸序列的核酸。在某些实施方案中,申请涉及包含载体的宿主细胞,所述载体包含编码与 SEQ ID NO:1 所示氨基酸序列具有 85%、90%、95%、98%、99% 或更高相同性的氨基酸序列的核酸。在一具体的实施方案中,宿主细胞是 CHO 细胞。

[0025] 在其他的实施方案中,本申请涉及制备肽的方法,所述肽包含与 SEQ ID NO:1 所示

氨基酸序列具有至少 85% 的相同性的氨基酸序列,所述方法包括在适于蛋白表达的条件下培养包含载体的宿主细胞和分离所述肽,所述载体包含编码与 SEQ ID NO :1 所示氨基酸序列具有 85%、90%、95%、98%、99% 或更高相同性的氨基酸序列的核酸。

[0026] 在某些实施方案中,本申请提供了组合物,其包含分离的 HIV 包膜蛋白和药用可接受的载体,所述分离的 HIV 包膜蛋白例如包含与 SEQ ID NO :1 所示氨基酸序列具有 85%、90%、95%、98%、99% 或更高的相同性的氨基酸序列的分离的 HIV 包膜蛋白。

[0027] 在仍另外一些实施方案中,本申请涉及在哺乳动物中生成针对 HIV 的抗体的方法,包括给所述哺乳动物施用包含分离的 HIV 包膜蛋白和药用可接受的载体的组合物,所述分离的 HIV 包膜蛋白例如包含与 SEQ ID NO :1 所示氨基酸序列具有 85%、90%、95%、98%、99% 或更高的相同性的氨基酸序列的分离的 HIV 包膜蛋白。在某些实施方案中,该组合物进一步包含佐剂。在某些实施方案中,佐剂包括脂质体制剂。在进一步的实施方案中,脂质体制剂包含以下的一种或多种:二肉豆蔻磷脂酰胆碱、二肉豆蔻磷脂酰甘油、胆固醇和磷脂。在一具体的实施方案中,脂质体制剂包含磷脂 A。在某些实施方案中,在哺乳动物中生成的抗体是与包含截短的 HIV 包膜蛋白的肽竞争结合整合蛋白 $\alpha 4 \beta 7$ 的抗体。

[0028] 在某些实施方案中,本申请涉及在哺乳动物赋予针对 HIV 的免疫力的方法,包括给所述哺乳动物施用包含分离的 HIV 包膜蛋白和药用可接受的载体的组合物,所述分离的 HIV 包膜蛋白例如包含与 SEQ ID NO :1 所示氨基酸序列具有 85%、90%、95%、98%、99% 或更高的相同性的氨基酸序列的分离的 HIV 包膜蛋白。在某些实施方案中,该组合物进一步包含佐剂。在某些实施方案中,佐剂包含脂质体制剂。在进一步的实施方案中,脂质体制剂包含以下的一种或多种:二肉豆蔻磷脂酰胆碱、二肉豆蔻磷脂酰甘油、胆固醇和磷脂。在一具体实施方案中,脂质体制剂包含磷脂 A。在进一步的实施方案中,所述方法包括通过注射给哺乳动物施用所述组合物。

[0029] 例如,可以使用本发明的组合物的哺乳动物的实例包括人类、非人灵长类、犬、兔、豚鼠和小鼠。

[0030] 然而在另一些实施方案中,本申请涉及亚单位疫苗,其包含本发明的 HIV 包膜蛋白,例如包含截短的 HIV 包膜蛋白的分离的肽,其中所述 HIV 包膜蛋白在天然的 gp120/gp41 切割位点处被突变以阻止蛋白酶切割,包含 gp41 的 MPER,且在跨膜区前被截短。在某些实施方案中,HIV 包膜蛋白在其 C 端含有大约 1 到 10 个亲水氨基酸。在某些实施方案中,所述大约 1-10 个亲水性氨基酸是三个赖氨酸。在某些实施方案中,亚单位疫苗包含分离的 HIV 包膜蛋白,例如包含与 SEQ ID NO :1 所示氨基酸序列具有 85%、90%、95%、98%、99% 或更高的相同性的氨基酸序列的分离的 HIV 包膜蛋白。

[0031] 在某些实施方案中,本申请涉及包含分离的核酸的核酸疫苗,所述分离的核酸包含编码与 SEQ ID NO :1 所示氨基酸序列具有 85%、90%、95%、98%、99% 或更高的相同性的氨基酸序列的核酸序列。在某些实施方案中,所述核酸序列编码 SEQ ID NO :1 所示氨基酸序列。

[0032] 在仍其他的实施方案中,本申请涉及分离的肽,所述分离的肽包含与 SEQ ID NO :3、SEQ ID NO :4 或 SEQ ID NO :5 所示氨基酸序列具有 90% 或更高的相同性的氨基酸序列。在某些实施方案中,所述肽包含与 SEQ ID NO :3、SEQ ID NO :4 或 SEQ ID NO :5 所示氨基酸

序列具有 98% 或更高的相同性的氨基酸序列。在进一步的实施方案中,所述肽包含与 SEQ ID NO :3、SEQ ID NO :4 或 SEQ ID NO :5 所示氨基酸序列具有 99% 或更高的相同性的氨基酸序列。在具体的实施方案中,所述肽包含 SEQ ID NO :3 或 SEQ ID NO :4 或 SEQ ID NO :5 所示氨基酸序列。

[0033] 在另一个实施方案中,本申请涉及分离的核酸序列,其包含编码与 SEQ ID NO :3、SEQ ID NO :4 或 SEQ ID NO :5 所示氨基酸序列具有 85%、90%、95%、98%、99% 或更高的相同性的氨基酸序列的核酸序列。在某些实施方案中,所述核酸序列编码 SEQ ID NO :3 或 SEQ ID NO :4 或 SEQ ID NO :5 所示氨基酸序列。

[0034] 在某些实施方案中,本申请涉及试剂盒,其包含 (a) 包含分离的肽和药物学可接受的载体的组合物,所述分离的肽包含截短的 HIV 包膜蛋白,其中所述 HIV 包膜蛋白在天然的 gp120/gp41 切割位点处被突变以阻止蛋白酶切割,包含 gp41 的 MPER,并在跨膜区前被截短,及 (b) 给哺乳动物施用所述组合物的说明。在某些实施方案中,HIV 包膜蛋白在其 C 端包含大约 1-10 个亲水氨基酸。在某些实施方案中,所述大约 1-10 个亲水性氨基酸是三个赖氨酸。

[0035] 在某些实施方案中,本申请涉及试剂盒,其包含 (a) 包含分离的 HIV 包膜蛋白的组合物,所述分离的 HIV 包膜蛋白例如包含与 SEQ ID NO :1 所示氨基酸序列具有 85%、90%、95%、98%、99% 或更高的相同性的氨基酸序列的分离的 HIV 包膜蛋白,以及 (b) 给哺乳动物施用所述组合物的说明。

[0036] 在另一实施方案中,本申请涉及试剂盒,其包含 (a) 包含分离的核酸和药用可接受的载体的组合物,所述分离的核酸包含编码与 SEQ ID NO :1 所示氨基酸序列具有 85%、90%、95%、98%、99% 或更高的相同性的氨基酸序列的核酸序列,及 (b) 给哺乳动物施用所述组合物的说明。在某些实施方案中,所述核酸序列编码 SEQ ID NO :1 所示氨基酸序列。

[0037] 在某些实施方案中,本申请涉及包含截短的 HIV 包膜蛋白的分离的肽,其中所述 HIV 包膜蛋白在天然 gp120/gp41 切割位点处被突变以阻止蛋白酶切割,包含 gp41 的 MPER,并在跨膜区前被截短,其中所述 HIV 包膜蛋白在前导序列处被突变。在某些实施方案中,天然信号肽被 tPA 信号肽替换。在某些实施方案中,tPA 信号肽包含选自 SEQ ID NO :21 和 SEQ ID NO :22 的序列。

[0038] 在仍其他的实施方案中,本申请提供了分离的肽,其包含与 SEQ ID NO :7 或 SEQ ID NO :9 所示氨基酸序列具有 90% 或更高的相同性的氨基酸序列。在进一步的实施方案中,所述肽包含与 SEQ ID NO :7 或 SEQ ID NO :9 所示氨基酸序列具有 98% 或更高的相同性的氨基酸序列。在再进一步的实施方案中,所述肽包含与 SEQ ID NO :7 或 SEQ ID NO :9 所示氨基酸序列具有 99% 或更高的相同性的氨基酸序列。在一具体的实施方案中,所述肽包含 SEQ ID NO :7 或 SEQ ID NO :9 所示的氨基酸序列。

[0039] 在某些实施方案中,本申请涉及分离的核酸序列,其包含编码与 SEQ ID NO :7 或 SEQ ID NO :9 所示氨基酸序列具有 85%、90%、95%、98%、99% 或更高的相同性的氨基酸序列的核酸序列。在某些实施方案中,所述分离的核酸序列包含 SEQ ID NO :3 或 SEQ ID NO :5 所示核酸序列。

[0040] 这些实施方案以及其他的实施方案将在下述发明详述中揭示和阐述。

[0041] 附图简述

[0042] 图 1 为载体 pJWIRESpuro 的图示。

[0043] 图 2293/pJW Ba-L gp140DC4E10Puro 转染免疫共沉淀蛋白质印迹。pJW Ba-L gp140DC4E10Puro 转染的 293 细胞和未处理细胞的条件培养基与人类 MPER 区域 (2F5and4E10) 单克隆抗体、HIV-1(+) 人血清和正常人血清免疫共沉淀。沉淀的蛋白通过聚丙烯酰胺凝胶电泳 SDS-PAGE 分开并转移至 PVDF。Ba-L gp140DC4E10 蛋白采用 gp41 (M25) 的 MoAb 接着山羊抗 - 小鼠 IgG AP 缀合物, 及 BCIP/NBT 底物检测。用 4E10 (泳道 1)、2F5 (泳道 2) 和 HIV-1(+) 人血清 (泳道 3) 检测到 Ba-L gp140DC4E10, 用正常人血清没有检测到 (泳道 4)。对应 Ba-L gp140DC4E10 的条带在未处理样品中用 4E10 (泳道 5)、2F5 (泳道 6) 和 HIV-1(+) 人血清 (泳道 7) 或正常人血清 (泳道 8) 均没有检测到。

[0044] 图 3 示出 HIV-1Ba-L gp140DC4E10 核酸序列 (SEQ ID NO :6)。突出显示的是 tPa 信号。

[0045] 图 4 示出 Ba-L gp140DC4E10 蛋白氨基酸序列 (SEQ ID NO :7)。突出显示的是 tPa 信号。

[0046] 图 5 示出 HIV-1Ba-L gp145 核酸序列 (SEQ ID NO :8)。突出显示的是 tPa 信号肽。该序列与图 3 所示 gp145 区相同, 只是 tPa 信号序列不同。

[0047] 图 6 示出 HIV-1Ba-L gp145 蛋白氨基酸序列 (SEQ ID NO :9)。突出显示的是 tPa 信号。

[0048] 图 7 哺乳动物表达载体 pJWTCDE-N。

[0049] 图 8HIV-1C 亚型 gp160 表达质粒。HIV-1gp160 基因连接进入 pSWTIPK3 质粒的 *NheI* 和 *EcoRI* 的限制性内切酶位点处, 与 t-Pa 信号肽读码框融合。

[0050] 图 9 示出 C 进化枝 C3728v2c6gp160 的核酸序列 (SEQ ID NO :10)。HIV-1C3728v2c6gp160 的核酸序列被密码子优化。突出显示的是 tPa 信号。

[0051] 图 10 示出 C 进化枝 C3728v2c6gp160 的核酸序列 (SEQ ID NO :11)。HIV-1C3728v2c6gp160DC 的核酸序列被密码子优化。突出显示的是 tPa 信号。gp120/gp41 切割位点已被突变以阻止切割。

[0052] 图 11 示出 C 进化枝 C06838v1c48gp160 的核酸序列 (SEQ ID NO :12)。HIV-1C06838v1c48gp160 的核酸序列被密码子优化。突出显示的是 tPa 信号。

[0053] 图 12 示出 C 进化枝 C06838v1c48gp160DC 的核酸序列 (SEQ ID NO :13)。HIV-1C06838v1c48gp160DC 的核酸序列被密码子优化。突出显示的是 tPa 信号。gp120/gp41 切割位点已被突变以阻止切割。

[0054] 图 13 示出 C 进化枝 C06980v1c3gp160 的核酸序列 (SEQ ID NO :14)。HIV-1C06980v1c3gp160 的核酸序列被密码子优化。突出显示的是 tPa 信号。

[0055] 图 14 示出 C 进化枝 C06838v1c48gp160DC 的核酸序列 (SEQ ID NO :15)。HIV-1C06980v1c3gp160DC 的核酸序列被密码子优化。突出显示的是 tPa 信号。gp120/gp41 切割位点已被突变以阻止切割。

[0056] 图 15C 进化枝 C06980v0c22gp160 的核酸序列 (SEQ ID NO :16)。HIV-1C06980v0c22gp160 的核酸序列被密码子优化。突出显示的是 tPa 信号。

[0057] 图 16C 进化枝 C06980v0c22gp160DC 的核酸序列 (SEQ ID NO :17)。HIV-1C06980v0c22gp160DC 的核酸序列被密码子优化。突出显示的是 tPa 信号。gp120/gp41

切割位点已被突变以阻止切割。

[0058] 图 17 为载体 pSWTIPK3 的图示。

[0059] 图 18 示出 pSWTIPK3 载体的核酸序列 (SEQ ID NO :18)。

[0060] 图 19 用 HIV-1C 亚型 gp160 和 gp160DC 的表达质粒转染的 CHO-K1 细胞 IP 蛋白质印迹。使用 HIV-1(+) 人血清从转染后 48 小时的细胞裂解物免疫共沉淀 Env 蛋白,于 4-15% SDS-PAGE 凝胶上分开并转移至 PVDF。使用 HIV-1B 亚型 gp160 的兔抗体检测来自下述构建体的蛋白 :C06838v1c48gp160(泳道 1)、C06980v1c3gp160(泳道 2)、C06980v0c22gp160(泳道 3)、C3728v2c6gp160(泳道 4)、C06838v1c48gp160DC(泳道 5)、C06980v1c3gp160DC(泳道 6)、C06980v0c22gp160DC(泳道 7)、C3728v2c6gp160DC(泳道 8)、空白 CHO-K1(-) 对照(泳道 9)、Ba-L gp145(+) 对照(泳道 10) 和 C 亚型 96ZM651gp140(+) 对照(泳道 11)。分子量蛋白标记位于泳道 12。

[0061] 图 20 转染了 HIV-1C 亚型 gp160 和 gp160DC 的表达质粒的 HEK293 细胞的免疫共沉淀蛋白质印迹。利用针对 gp41(4E10) 的 huMAb 从转染 48 小时后的细胞裂解物免疫共沉淀 Env 蛋白,于 4-15% SDS-PAGE 凝胶上分开,并转移至 PVDF。使用 HIV-1B 亚型 gp160 的兔抗体检测来自下述构建体的蛋白 :C06838v1c48gp160(泳道 1)、C06980v1c3gp160(泳道 2)、C06980v0c22gp160(泳道 3)、C3728v2c6gp160(泳道 4)、C06838v1c48gp160DC(泳道 5)、C06980v1c3gp160DC(泳道 6)、C06980v0c22gp160DC(泳道 7)、C3728v2c6gp160DC(泳道 8)、空白 HEK293(-) 对照(泳道 9)、Ba-L gp160(+) 对照(泳道 10) 空白 CHO-K1(-) 对照(泳道 11) 和 CHO-K1/Ba-L gp160(+) 对照(泳道 12)。分子量蛋白标记位于泳道 13。

[0062] 图 21HIV-1C06980v0c22gp145 表达质粒。HIV-1gp145 基因连接进入 pSWTIPK3 质粒的 NheI 和 EcoRI 的限制性内切酶位点处,与 t-Pa 信号肽读码框融合。

[0063] 图 22 示出 pSWC06980v0c22gp145 的核酸序列 (SEQ ID NO :19)

[0064] 图 23HIV-1C06980v0c22gp145 的密码子优化的核酸序列 (SEQ ID NO :20)。突出显示的是 tPa 信号。

[0065] 图 24 是 HIV-1C06980v0c22gp145 的蛋白质序列 (SEQ ID NO :32)。突出显示的是 tPa 信号。

[0066] 图 25A(SEQ ID NO :21) 和 B(SEQ ID NO :22) 示出本发明的包膜蛋白采用的 tPA 序列。

[0067] 图 26 转染了 pSWC06980v0c22gp145 的 CHO-K1 细胞免疫共沉淀蛋白质印迹。用 HIV-1(+) 人血清从转染 48 小时后的条件培养基与裂解物中免疫共沉淀包膜蛋白,于 4-15% SDS-PAGE 凝胶上分开并转移至 PVDF。利用 HIV-1B 亚型 gp160 和 C 亚型 gp120 的兔抗体检测 gp145 :未处理 CHO-K1(-) 对照培养基(泳道 1)、超螺旋 pSWC06980v0c22gp145 培养基(泳道 2)、线性化 pSWC06980v0c22gp145 培养基(泳道 3)、未处理 CHO-K1(-) 对照细胞裂解物(泳道 4)、超螺旋 pSWC06980v0c22gp145 细胞裂解物(泳道 5)、线性化 pSWC06980v0c22gp145 细胞裂解物(泳道 6)。分子量蛋白标记位于泳道 7。

[0068] 图 27 从 CHO 细胞系 H-73-9-2-8 和 H-73-9-3-9 的条件培养基中纯化的 C06980v0c22gp145 的 4-15% SDS-PAGE。在还原和非还原条件下分开 5 μ g 蛋白质,采用考马斯蓝 R250 染色 :H-73-9-2-8 非还原(泳道 1)、H-73-9-3-9 非还原(泳道 2)、H-73-9-2-8 还原(泳道 3) 和 H-73-9-3-9 还原(泳道 4)。分子量蛋白标记位于泳道 5。

[0069] 图 28 从 CHO 细胞系 H-73-9-2-8 和 H-73-9-3-9 的条件培养基中纯化得到的 C06980v0c22gp145 的蛋白质印迹。在还原和非还原条件下采用 4-15 % SDS-PAGE 分开 0.5 μ g 蛋白质,转移至 PVDF,并用 HIV-1 的血清检测 :H-73-9-2-8 非还原 (泳道 1)、H-73-9-3-9 非还原 (泳道 2)、H-73-9-2-8 还原 (泳道 3) 和 H-73-9-3-9 还原 (泳道 4)。分子量蛋白标记位于泳道 5。

[0070] 图 29gp145 的下游纯化方法流程图。

[0071] 图 30 纯化的重组 HIV-1C06980v0c22gp145 的 SE-HPLC 分析 (批次 112009)。于 1X PBS 制备 1 : 10 稀释的纯化蛋白,将 20 μ L 上样于 TSK-GEL3000SWXL 柱 (TOSOH BIOSEP)。该柱使用 1X PBS 等度洗脱,流速 1.0mL/min,结果鉴定出 4 种 gp145。

[0072] 图 31gp145 的下游纯化方法流程图

[0073] 图 32 纯化的重组 HIV-1C06980v0c22gp145 批次 120710A 的 SE-HPLC 分析。于 1X PBS 制备 1 : 10 稀释的纯化蛋白,取 20 μ L 上样于 TSK- GEL3000SWXL 柱 (TOSOH BIOSEP)。该柱使用 1X PBS 等度洗脱,流速 1.0mL/min,结果鉴定出 4 种 gp145 物质。

[0074] 图 33gp145 的下游纯化方法流程图

[0075] 图 34 纯化的重组 HIV-1C06980v0c22gp145 批次 120710B 的 SE-HPLC 分析。于 1X PBS 制备 1 : 10 稀释的纯化蛋白,取 20 μ L 上样于 TSK-GEL3000SWXL 柱 (TOSOH BIOSEP)。该柱使用 1X PBS 等度洗脱,流速 1.0mL/min,结果鉴定出 4 种 gp145。

[0076] 图 35HIV-1C 亚型包膜蛋白序列比对。框出的区域代表 gp41/gp120 切割结构域。氨基端的氨基酸是来自 tPa 信号肽末端的 NheI 克隆位点的丝氨酸。

[0077] 图 36C06980v0c22gp145 氨基酸序列 (SEQ ID NO :32)。tPA 信号、切割位点突变,和 C 端的三个赖氨酸用方框表明。

[0078] 图 37C06980v0c22gp145 核苷酸序列 (SEQ ID NO :22) 和翻译 (SEQ ID NO :32)。在核酸序列和翻译间示出 tPA 前导序列。方框区标示的是切割位点突变和末端赖氨酸重复。

[0079] 图 38 示出依照本发明的 HIV 包膜蛋白的 C 末端残基 (SEQ ID NO :25 和 43)。

[0080] 图 39 抗原性 :通过 ELISA 测定 4E10 和 VRC01 与 C06980gp145 的结合。

[0081] 图 40 中和 :C06980PV 对 4E10 和 VRC01mAb 敏感。

[0082] 图 41 示出兔 C 进化枝 gp145 的研究设计。

[0083] 图 42gp145C 进化枝免疫的兔血清中和来自 B 和 C 进化枝的 Tier1 假病毒。

[0084] 图 43 免疫接种后生成针对 HIV-1C 进化枝 Tier1 假病毒的中和抗体。

[0085] 图 44 在使用 gp145 免疫的兔血清的 PBMC 测定中 B、C 和 AE IMC 跨进化枝中和。

[0086] 图 45 免疫接种后生成针对 HIV-1B 进化枝 BaL IMC 的中和抗体。

[0087] 图 46gp145 免疫的兔血清与 C 进化枝包膜蛋白结合。

[0088] 图 47GS015 的中和由 IgG 介导 (TZMb1 中和测定)。

[0089] 图 48 免疫原中和值的复合图。

[0090] 图 49 示出适应无蛋白培养基的 CHO C06980v0c22gp145 细胞的 I.P. 蛋白印迹结果。使用人类抗体从条件培养基中沉淀 gp145,于 4-15 % SDS-PAGE 分开,转移至 PVDF,并使用 gp120 和 gp160 抗血清检测。

[0091] 图 50 同源性及糖基化位点。

[0092] 图 51 密码子优化后的 D 进化枝 gp140 和 C 进化枝 gp145 的氨基酸序列。

- [0093] 图 52 经免疫的兔血清结合不同的包膜蛋白。
- [0094] 图 53 在 HIV-1 抗原刺激后 A) 淋巴结、B) 脾中的 IFN γ ELISPOT, 显示为斑点计数。
- [0095] 图 54 细胞内染色检测 IL-2 表达。
- [0096] 图 55 所有组对 HIV-1 包膜蛋白 gp145 和 gp140 的 ELISA 结合滴度。
- [0097] 图 56 所有组在两种测定平台上的中和结果, A) TZMbl 和 B) PBMC。
- [0098] 图 57 基于流式细胞术的 $\alpha 4 \beta 7$ 结合和抑制测定。
- [0099] 图 58 在初始 (primary) T 细胞上 $\alpha 4 \beta 7$ 表达的诱导。从 PBMC 经磁珠分离出的 CD4⁺ (上图) 和 CD8⁺ (下图) T 细胞和抗 -CD3/CD28、IL-2 和视黄酸培养 5 天。培养初始 CD4⁺ (上图) 和 CD8⁺ (下图) T 细胞以表达结合重组 gp120 或 gp145 蛋白、或包含包膜蛋白 V2 环区的环肽的 $\alpha 4 \beta 7$ 。示出了结合的蛋白或肽 (蓝色矩形) 与非蛋白 neutravidin-PE 对照 (绿色矩形)。gp145 图显示出对于 $\alpha 4 \beta 7$ 与 gp145 的结合, CD8⁺ 细胞中 89.7% 阳性, 而 CD4⁺ T 细胞中 93.3% 阳性。
- [0100] 图 59 HIV-1 包膜蛋白与表达 $\alpha 4 \beta 7$ 的 T 细胞的结合。培养初始 CD4⁺ (上图) 和 CD8⁺ (下图) T 细胞使之表达结合重组 gp120 或 gp145 蛋白质、或包含包膜蛋白 V2 环区的环肽的 $\alpha 4 \beta 7$ 。示出结合的蛋白或肽 (蓝色矩形) 与非蛋白 neutravidin-PE 对照 (绿色矩形)。
- [0101] 图 60 封闭 V2 和 $\alpha 4 \beta 7$ 间的相互作用。如所描述, 培养初始分离的 T 细胞以表达 $\alpha 4 \beta 7$ 。抗 -V2 单克隆抗体在与细胞结合前与生物素化的 gp120 或环状 -V2 肽预培养。抗 - $\alpha 4$ 单克隆抗体在加入蛋白质前预先与细胞结合作为正对照。
- [0102] 图 61 C06980v0c22gp145 蛋白质的 Blue Native PAGE。5 μ g 下述蛋白使用 Invitrogen 的 Native PAGE 系统在 4-16% Novex Bis-Tris gel 中分开: 批次 112009 (泳道 1)、批次 120710A (泳道 2) 和批次 120710B (泳道 3)。分子量蛋白标记位于泳道 4。对每一组 gp145, 有 3 种多聚体物质 (A、B 和 C) 都是很明显的。多聚体 B 占主要地位。
- [0103] 图 62 在非还原性条件下 EGS 交联的 C06980v0c22gp145 (批次 120710A) 的 SDS-PAGE。5 μ g gp145 用 12.5、5、1、0.2 和 0mM EGS 处理, 在非还原条件下用 3-8% NuPAGE Tris 醋酸聚丙烯酰胺凝胶电泳分开: 12.5mM EGS (泳道 1)、5mM EGS (泳道 2)、1mM EGS (泳道 3)、0.2mM EGS (泳道 4) 和 0mM EGS (泳道 5)。EGS 交联的磷酸化酶 B 位于泳道 6, 作为分子量蛋白质标记。
- [0104] 图 63 在 Superose6 上纯化的 EGS 交联的 C06980v0c22gp145 (批次 120710A) 的 SDS-PAGE。从柱上装载并洗脱的 10 μ g gp145 与 5mM EGS 交联。交联与未交联的 gp145 在 3-8% NuPAGE Tris Acetate 聚丙烯酰胺凝胶电泳中分开: 5 μ g 未交联的 gp145 柱 (A), 10 μ g 交联的 gp145 柱 (B) 和 10 μ g 交联的洗脱片段 26-32 (C)。EGS 与磷酸化酶 B 交联后位于泳道 6, 作为分子量蛋白质标记 (D)。
- [0105] 发明详述
- [0106] 很多候选的 HIV 疫苗不能与人类天然的中和抗体相互作用。如本文描述的, 申请人证实了 HIV-1 包膜蛋白可以被修饰为结合广谱反应性的抗体。因此, 本申请提供了涉及新 HIV 包膜蛋白的方法和相关组合物。
- [0107] 本发明中新的 HIV 包膜蛋白含有 HIV 包膜蛋白的整个胞外结构域, 包括 gp41 的膜近端外部区域 (MPER)。gp41 蛋白由三个主要结构域组成, 即胞外结构域、跨膜结构域和胞

质尾区。胞外结构域由融合肽、N 端七个重复、C 端七个重复和 MPER 组成。

[0108] gp41 的 MPER 一般含有 gp41 胞外结构域 C 端最后 24-28 个氨基酸。MPER 是 HIV 包膜蛋白一个高度保守的区域, 含有广谱中和人类单克隆抗体的表位, 特别是 2F5、Z13 和 4E10 单克隆抗体的表位。

[0109] 如本文所描述, 本发明的 HIV 包膜蛋白是截短的 HIV 包膜蛋白, 其在天然 gp120/gp41 切割位点处突变以阻止蛋白酶切割, 包含 gp41 的 MPER, 并在跨膜区前被截短。本发明的 HIV 包膜蛋白可能含有任何天然的 HIV 包膜蛋白 gp41 的 MPER 区。

[0110] 在某些实施方案中, 本发明的 HIV 包膜蛋白包含 MPER 序列, 其含有氨基酸序列 ALDSWNNLWNWFDIS (SEQ ID NO :23)。在某些实施方案中, 本发明的 HIV 包膜蛋白包含 MPER 序列, 其含有氨基酸序列 LWYIK (SEQ ID NO :24)。在某些实施方案中, 所述 MPER 序列包含氨基酸序列 ELLALDSWNNLWNWFDISNWLWYIK (SEQ ID NO :25)。在另一些实施方案中, 所述 MPER 序列包含氨基酸序列 DLLALDSWKNLWNWFDITNWLWYIK (SEQ ID NO :26)。

[0111] 一般来说, 本发明的 HIV 包膜蛋白可以含有 2F5、Z13 和 4E10 中至少一种单克隆抗体的表位。2F5 表位的实例包括本文公开的 ALDSWN (SEQ ID NO :27)、ELDKWA (SEQ ID NO :28) 和 EKNEQELLELDKWASLW (SEQ ID NO :29) (参见例如, Montero, M, 等人, Microbiology and Molecular Biology Reviews (2008) 72 (1) :54-84 及其中引用的参考文献)。4E10 表位的实例包括本文公开的 NWFDIS (SEQ ID NO :30) 和 NWFDIT (SEQ ID NO :31) (同上)。

[0112] 本发明的 HIV 包膜蛋白缺少跨膜结构域和胞质尾区, 但是包含 gp41 整个胞外结构域。在某些实施方案中, 胞外结构域被修饰为在 C 端含有大约 1-10 个亲水氨基酸。在某些实施方案中, 这些亲水氨基酸残基一般被添加至在本发明的截短的 HIV 包膜蛋白的胞外结构域以通过使其更加亲水而增加该区域的暴露。在其中 HIV 包膜蛋白在 C 端含有大约 1-10 个亲水氨基酸的实施方案中, 所述亲水氨基酸一般与天然 MPER 序列的最后的氨基酸残基是连续的。因此, 这 1-10 个亲水性氨基酸一般包含 HIV 包膜蛋白 C 端最后的氨基酸残基。在某些实施方案中, 所述 1-10 个亲水性氨基酸是一个或多个赖氨酸残基。

[0113] 在某些实施方案中, 本发明的 HIV 包膜蛋白衍生自从急性 HIV 感染的个体中分离得到的 HIV 毒株。在另一些实施方案中, 所述 HIV 感染是慢性的。在某些实施方案中, HIV 包膜蛋白衍生自归类于 M、O、N 或 P 组的 HIV-1 菌株。在一具体的实施方案中, HIV 包膜蛋白衍生自 HIV-1M 组毒株。在进一步的实施方案中, HIV-1M 组毒株是选自 A、B、C、D、F、G、H、J 和 K 及杂合体的亚型 (进化枝)。在进一步的实施方案中, HIV 包膜蛋白衍生自 HIV-1M 组菌株, 所述 M 组毒株为 B 进化枝、C 进化枝或 D 进化枝毒株。在某些实施方案中, B、C 或 D 进化枝毒株分离自急性感染的个体。在另一些实施方案中, B、C 或 D 进化枝毒株分离自慢性感染的个体。可以衍生出本发明的包膜蛋白的适宜的亲本 HIV 毒株的实例包括序列示于 GenBank 登记号 AF484477、AF484511 和 AF484502 的 HIV-1D 进化枝和序列示于 GenBank 登记号 HM215344 和 HM215345 的 HIV-1C 进化枝。

[0114] 在某些实施方案中, 本文描述的 HIV 包膜蛋白以不同形式作为免疫原使用, 用作 HIV 疫苗的组分来引发针对例如 HIV-1 的 bNab。HIV 包膜蛋白的不同形式可以用于初始免疫 (作为表达所述蛋白的 DNA/ 载体或蛋白) 和 / 或用作加强免疫 (作为蛋白)。例如, 在某些实施方案中, 本发明的 HIV 包膜蛋白作为 DNA 疫苗被施用至哺乳动物, 随后施用蛋白加强免疫。在进一步的实施方案中, HIV 包膜蛋白作为质粒中的核酸而施用, 随后在病毒载体

中施用（如，作为 MVA 中的核酸），随后作为蛋白质施用。在某些实施方案中，本发明的 HIV 包膜蛋白也可以通过与病毒颗粒（如 Qbeta、豇豆花叶病毒、CRM、HPV、HBsAg 等）交联而用作微粒免疫原。

[0115] 在某些实施方案中，本发明的 HIV 包膜蛋白被用作试剂，其用于筛选新的广谱中和抗体，和 / 或依照免疫研究，绘制具有广谱中和血清活性的人类血清和动物血清的图谱。在另一些实施方案中，本发明的 HIV 包膜蛋白被用于筛选与广谱中和抗体竞争结合的小分子。经鉴定的小分子可以被用作免疫原或抗病毒化合物。

[0116] 如本文所描述的，申请人生成了具有独特序列的重组包膜蛋白，其中申请人修饰了前导序列、gp120/gp41 切割位点，加入亲水性氨基酸尾部，并且于跨膜结构域前终止该序列使其具有 gp41 的完整胞外结构域。因为经过密码子优化，DNA 序列也是独特的。

[0117] 在另一些有益的实施方案中，HIV 包膜蛋白基本上具有与图 4、6、24、35、36 和 / 或 37 所示 HIV 包膜蛋白序列相似的序列。在另一个更为有益的实施方案中，HIV 包膜蛋白具有与图 38 所示 MPER 序列基本相似的 MPER 序列。

[0118] 在一个更为有益的实施方案中，本发明的 HIV 包膜蛋白与 SEQ ID NO :1 或图 3-6、23、24 和 35-37 中所示的任一 HIV 包膜蛋白序列具有约 75%、约 76%、约 77%、约 78%、约 79%、约 80%、约 81%、约 82%、约 83%、约 84%、约 85%、约 86%、约 87%、约 88%、约 89%、约 90%、约 91%、约 92%、约 93%、约 94%、约 95%、约 96%、约 97%、约 98%、约 99% 或约 100% 的序列相同性。

[0119] 在一实施方案中，本发明的 HIV 包膜蛋白可以被用于筛选和鉴定新的广谱中和抗体的试剂。用于筛选中和抗体的测定是本领域公知。中和测定方法之前已被报道过 (Binley JM, 等人, (2004) Comprehensive Cross-Clade Neutralization Analysis of a Panel of Anti-Human Immunodeficiency Virus Type1 Monoclonal Antibodies. J. Virol. 78 :13232-13252)。可以通过将至少两个分开编码本发明的可溶性包膜蛋白 cDNA 和其余 HIV 基因组的质粒共转染细胞而产生假病毒。在 HIV 基因组编码载体中，可以用萤火虫萤光素酶基因取代包膜蛋白基因。包含假病毒转染子上清可以与 B 细胞上清或和单克隆或多克隆（血清）抗体共温育过夜，所述 B 细胞上清来自经感染的供体初始外周血单核细胞 (PBMC) 的激活。可以将用 CD4 及 CCR5 和 CXCR4 共受体稳定转染并表达 CD4 及 CCR5 和 CXCR4 的细胞加入混合物中，37℃ 孵育 3 天。感染的细胞可以通过荧光测定定量。

[0120] 在某些实施方案中，为了筛选广谱中和抗体，可以通过将包膜蛋白的 C 端与酶连接构建包膜-酶融合蛋白。可以生成包含所述融合蛋白和野生型和 / 或可溶性包膜糖蛋白的病毒颗粒，并用于感染患者血清中存在的靶细胞。在如此感染的细胞中测量的酶活性是由野生型病毒包膜蛋白介导的病毒结合和进入靶细胞的量度。可以用于生成所述融合蛋白的酶的实例，包含但不限于萤光素酶、细菌或胎盘碱性磷酸酶、β-半乳糖苷酶和荧光蛋白如绿色荧光蛋白，或毒素。一般而言，也可以在 96 孔板上完成该测定。在血清存在情况下酶活性的降低表明血清中含有中和抗体。

[0121] 如本文所用的，下述术语：“药物”、“物质 (agent)”和“化合物”包括任何物质的组合物或混合物，其可以体内或体外证实提供药理学作用。这包括小分子、抗体、微生物 (microbiological)、疫苗、维生素和其他有益的物质。如本文所用的，这些术语进一步还包含了所有在病人产生局部或全身性生理学或药理学活性的物质。

[0122] 核酸、蛋白质和重组技术

[0123] 除非另有说明,本发明采用了普遍的化学、分子生物学、微生物学、DNA 重组和免疫学实验技术,其属于本领域普通技术人员的能力范围。这些在参考文献中有阐述。例如, J. Sambrook, E. F. Fritsch 和 T. Maniatis, 1989, *Molecular Cloning: A Laboratory Manual*, 第二版, Books 1-3, Cold Spring Harbor Laboratory Press; Ausubel, F. M. 等人 (1995 and periodic supplements; *Current Protocols in Molecular Biology*, ch. 9, 13, and 16, John Wiley & Sons, New York, N. Y.); B. Roe, J. Crabtree 和 A. Kahn, 1996, *DNA Isolation and Sequencing: Essential Techniques*, John Wiley & Sons; M. J. Gait (编辑), 1984, *Oligonucleotide Synthesis: A Practical Approach*, Irl Press; D. M. J. Lilley 和 J. E. Dahlberg, 1992, *Methods of Enzymology: DNA Structure Part A: Synthesis and Physical Analysis of DNA Methods in Enzymology*, Academic Press, 以上每一个文本通过引用并入本文。

[0124] 术语“核酸”包括 DNA、RNA (即 mRNA、tRNA)、异源双链体和能够编码多肽的合成分子,并包括本领域普通技术人员将承认能够取代天然存在的核苷酸和其主链的所有类似物和主链取代物例如 PNA。核酸可以是单链或双链,且可以化学修饰。术语“核酸”和“多核苷酸”可互换使用。因为基因密码具有简并性,多于一种密码子可以用于编码一种特定氨基酸,本发明的组合物和方法中包括了编码特定氨基酸序列的核苷酸序列。

[0125] 除非另有说明,核酸的书写从左至右为 5' 至 3' 方向;氨基酸序列的书写从左至右为氨基至羧基方向。

[0126] 本文术语“蛋白质”、“肽”、“多肽”和“氨基酸序列”可互换使用,指的是任意长度的氨基酸残基聚合物。该聚合物可以是线性的或分支的,其可以包含修饰的氨基酸或氨基酸类似物,且其可以被除氨基酸的外化学部分打断。该术语还包括天然地或人为干预地修饰的氨基酸聚合物;例如形成二硫键、糖基化、酯化、乙酰化、磷酸化或任何其他操作与修饰,如与标签或具有生物活性的组分缀合。本文采用的是常规的单字母或三字母氨基酸残基编码。

[0127] 如本文所用的,“合成的”分子是由体外化学合成或酶促合成产生,而不是由生物体产生。

[0128] 如本文所用的,术语“表达”指的是基于基因的核酸序列产生多肽的过程。该过程包含转录和翻译两者。

[0129] “基因”指的是编码多肽或 RNA 的 DNA 节段 (segment)。

[0130] “分离的”多核苷酸或多肽指该多核苷酸或多肽基本上没有在其天然环境与之联合的物质。“基本上没有”意味着要求至少 50%,有利地至少 70%,更有利地至少 80%,甚至更有利地至少 90% 的这些物质没有了。

[0131] “分离的”核酸分子是指从整个生物体中隔开的单独的分离的核酸分子,该分子在自然状态下可以在所述生物体找到;或是指整体或部分缺乏通常自然状态下与之相连的序列的核酸分子;或是自然状态下存在的序列,但是在其含有与之连接的异源序列。

[0132] “天然的”蛋白质或多肽指的所述蛋白质天然存在的来源中分离得到的蛋白质或多肽。“重组”多肽指的是由重组 DNA 技术生产的多肽,例如从转换了编码所需多肽的外源 DNA 构建体的细胞产生的多肽。“合成的”多肽包括通过化学合成制备的多肽及上面所述的

合成的抗原。

[0133] “同系物”指的是与目标氨基酸序列和目标核苷酸序列具有某些程度相同性的实体。如本文所用的,术语“同系物”包括了关于结构和 / 或功能的相同性,例如获得的核苷酸序列的表达产物具有目标氨基酸序列的酶活性。关于序列相同性,优选具有至少 70%、75%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98% 或甚至 99% 的序列相同性。这些术语也包括序列的等位基因变异。术语“同系物”可以用于指物种形成事件分离的基因间的关系,或用于指基因复制事件分离的基因间的关系。

[0134] 相对序列相同性可以通过商业可获得的计算机程序来确定,其采用任何适宜的算法计算出两个或多个序列间的相同性百分比,例如采用缺省参数。此类计算机程序的典型的例子是 CLUSTAL。更有利地,应用 BLAST 算法,参数设置为缺省值。在 National Center for Biotechnology Information (NCBI) 网站上有 BLAST 算法的详细介绍。

[0135] 本文所提供的肽的同系物一般和这样的肽具有结构相似性。多肽的同系物包含一个或多个保守的氨基酸取代,其可以选自这些氨基酸所属的类别的同一或不同成员。

[0136] 在一实施方案中,所述序列还可能具有氨基酸残基的缺失、插入或取代,其造成沉默变化并导致产生功能上等同的物质。考虑的氨基酸取代可以基于残基相似的极性、电荷、可溶性、疏水性、亲水性和 / 或两亲性性质进行,只要保持所述物质的二级结合活性。例如,带负电荷的氨基酸包括天冬氨酸和谷氨酸,带正电荷的氨基酸包括赖氨酸和精氨酸,具有不带电的极性头部具有相似亲水性值的氨基酸包括亮氨酸、异亮氨酸、缬氨酸、甘氨酸、丙氨酸、天冬酰胺、谷氨酰胺、丝氨酸、苏氨酸、苯丙氨酸和酪氨酸。

[0137] 本发明也包括可能存在的保守性取代(置换和取代均在本文用于指已存在的氨基酸残基被可选残基交换),例如相似对相似的取代,如碱性取代碱性、酸性取代酸性、极性取代极性等等。也可以存在非保守性取代,例如从一类残基取代为另一类残基或包括采用非天然氨基酸,例如鸟氨酸(下文表示为 Z)、双氨基丁酸鸟氨酸(下文表示为 B)、正亮氨酸鸟氨酸(下文表示为 O)、吡啶丙氨酸、噻吩丙氨酸、萘基丙氨酸、苯基甘氨酸。保守性取代可以在例如碱性氨基酸(精氨酸、赖氨酸和组氨酸)、酸性氨基酸(谷氨酸和天冬氨酸)、脂肪族氨基酸(丙氨酸、缬氨酸、亮氨酸和异亮氨酸)、极性氨基酸(谷氨酰胺、天冬酰胺、丝氨酸和苏氨酸)、芳香族氨基酸(苯丙氨酸、色氨酸和酪氨酸)、羟基氨基酸(丝氨酸和苏氨酸)、大氨基酸(苯丙氨酸和色氨酸)以及小氨基酸(甘氨酸和丙氨酸)组内进行。

[0138] DNA 扩增的很多方法是本领域公知,并且其中的任一种方法均可使用,例如参见 Sambrook 等人, *Molecular Cloning; A Laboratory Manual* 2d ed. (1989)。例如,感兴趣的 DNA 片段可采用聚合酶链式反应进行扩增,或采用某种其他循环聚合酶介导的扩增反应。

[0139] 扩增的 DNA 区可以采用本领域公知的任一方法测序。有利地,核酸测序还可以采用自动的方法进行(由 Meldrum 所作综述, *Genome Res.* 9. 2000 ;10(9):1288-303, 这些内容以其整体通过引用并入本文),例如采用 Beckman CEQ8000 Genetic Analysis System (Beckman Coulter Instruments, Inc.) 进行测序。核酸测序的方法包括但不限于局限于自动化荧光片段分析(参见例如 Watts & MacBeath, *Methods Mol Biol.* 2001 ;167 :153-70 和 MacBeath 等人 *Methods Mol Biol.* 2001 ;167 :119-52),毛细管电泳法(参见例如 Bosserhoff 等人, *Comb Chem High Throughput Screen.* 12. 2000 ;3(6):455-66), DNA

测序芯片（参见例如 Jain, Pharmacogenomics. 8. 2000 ;1(3) :289-307), 质谱法（参见例如 Yates, Trends Genet. January 2000 ;16(1) :5-8), 焦磷酸测序（参见例如 Ronaghi, Genome Res. January 2001 ;11(1) :3-11), 超薄层凝胶电泳（参见例如 Guttman&Ronai, Electrophoresis. 12. 2000 ;21(18) :3952-64), 上述公开以它们的整体通过引用并入本文。测序也可以由任何商业公司完成。例如这些公司包括但不限于 University of Georgia Molecular Genetics Instrumentation Facility (Athens, Ga.) 或 SeqWright DNA Technologies Services (Houston, Tex.)。

[0140] 可以使用本领域公知的 DNA 扩增的任一种方法, 例如聚合酶链式反应 (PCR)、连接酶链式反应 (LCR) (Barany, F., Proc. Natl. Acad. Sci. (U. S. A.) 88 :189-193 (1991))、链置换测定 (SDA) 或寡核苷酸连接测定 (“OLA”) (Landegren, U. 等人, Science 241 :1077-1080 (1988))。Nickerson, D. A. 等人阐述了一种组合了 PCR 和 OLA 的核酸测定方法 (Nickerson, D. A. 等人, Proc. Natl. Acad. Sci. (U. S. A.) 87 :8923-8927 (1990))。另外一些已知的核酸扩增程序也可以使用, 例如基于转录的扩增系统 (Malek, L. T. 等人, U. S. Pat. No. 5,130,238, Davey, C. 等人, European Patent Application 329,822, Schuster 等人, U. S. Pat. No. 5,169,766, Miller, H. I. 等人 PCT Application W089/06700, Kwoh, D. 等人, Proc. Natl. Acad. Sci. (U. S. A.) 86 :1173 (1989), Gingeras, T. R. 等人 PCT Application W088/10315) 或等温扩增方法 (Walker, G. T. 等人, Proc. Natl. Acad. Sci. (U. S. A.) 89 :392-396 (1992))。

[0141] 为了进行依据本发明的循环酶介导的扩增反应, 引物杂交并退火至靶 DNA 的相对链, 而后提升温度使热稳定的 DNA 聚合酶延伸引物并因此复制出跨越两引物之间的特定 DNA 节段。而后该反应进行热循环, 使得每次循环过后代表两引物间的序列的 DNA 的量加倍, 最终实现基因 DNA 序列（如果存在）的特异性扩增。

[0142] 在本发明中可以使用多种聚合酶。对于热循环反应, 所述聚合酶为热稳定聚合酶, 例如 Taq、KlenTaq、Stoffel Fragment、Deep Vent、Tth、Pfu、Vent 和 UItma, 每一种均易于从商业途径获得。对于非热循环的反应, 以及在某些热循环的反应中, 聚合酶通常是本领域常用的众多聚合酶中的一种, 例如 DNA pol I、Klenow 片段、T7DNA 聚合酶和 T4DNA 聚合酶。这些聚合酶的使用说明可以很容易的从产品说明上或普通分子生物学指导上找到。

[0143] 一般来说, 引物与目标 DNA 序列的退火大约在 37-55°C 进行 2 分钟, 聚合酶（例如 Taq 聚合酶）对引物序列的延伸在三磷酸核苷酸存在下大约 70-75°C 进行 3 分钟, 而变性步骤在大约 90-95°C 进行约 1 分钟以释放已结合的引物。然而, 这些参数可以变化, 对本领域普通技术人员而言可以很容易知晓如何调整反应时间和温度参数以获得所需的结果。例如, 循环可以短至 10、8、6、5、4.5、4、2、1、0.5 分钟或更短。

[0144] 此外, 还可以采用“两种温度”技术, 即退火和延伸步在同样的温度下进行, 一般在约 60-65°C 之间, 从而减少每次扩增循环的长度, 导致较短的测定时间。

[0145] 一般来说, 如本文所描述的反应重复直至生成的产物达到可检测的量。通常这样的可检测的产物的量在大约 10ng 至大约 100ng 之间, 不过更大量的产物, 如 200ng、500ng、1mg 甚至更多当然也可以被检测到。如果浓缩, 可检测的产物的量可以从大约 0.01pmol、0.1pmol、1pmol、10pmol 或更多。因此, 进行的反应循环的数量可以变化, 进行越多次的循环可以得到越多的扩增产物。在某些实施方案中, 反应包含 2、5、10、15、20、30、40、50 或更多

个循环。

[0146] 例如,可以使用 25-50 μ l 样品进行 PCR 反应,所述样品包含大约 0.01 至 1.0ng 模板扩增序列,大约 10 至 100pmol 的每种引物,大约 1.5 单位 Taq DNA 聚合酶 (Promega Corp.),大约 0.2mM dATP,大约 0.2mM dCTP,大约 0.2mM dGATP,大约 0.2mM dTTP,大约 15mM MgCl₂,大约 10mM Tris-HCl (pH9.0),大约 50mM KCl,大约 1 μ g/ml 明胶和大约 10 μ l/ml Triton X-100 (Saiki,1988)。

[0147] 本领域技术人员能够知道在循环聚合酶介导的扩增反应中可用的核苷酸种类。一般来说,核苷酸应至少部分由脱氧核苷酸三磷酸 (dNTP) 组成,这在商业上很容易获得。dNTP 的优化使用参数是本领域技术人员已知的并在在参考文献中描述过。而且,本领域技术人员也知道大量的核苷酸衍生物并且可以在这样的反应中应用。这样的衍生物包括荧光标记的核苷酸,允许如下文所述检测包括如此标记的核苷酸的产物。同时在此类中还包括允许测序包括此类核苷酸的核酸的核苷酸,例如链终止核苷酸、双脱氧核苷酸和硼酸盐核苷酸酶抗性核苷酸。包含了最经常用于这些 DNA 测序方法中的试剂的商品试剂盒可以获得并广泛使用。另外一些核苷酸类物,包括具有溴代、碘代或其他修饰基团的核苷酸,所述修饰基团影响获得的核酸的很多性质,包括其抗原性、其可复制性、其解链温度、其结合特性等等。此外,某些核苷酸包含反应性侧基,例如巯基、氨基、N-羟基 琥珀酸亚胺基,其允许进一步修饰包含它们的核酸。

[0148] 术语“寡核苷酸”指包含两个或更多个脱氧核糖核苷酸的分子,优选多于 3 个。它的精确大小取决于很多因素,其反过来依赖于寡核苷酸的最终的功能和用途。在此采用的术语“引物”指的是寡核苷酸,无论是天然存在的(如由限制性内切酶消化产物中纯化得到)还是合成产生的,当置于诱导引物延伸产物(与核酸链互补)合成的条件下可以作为合成起始点,即,在存在核苷酸和诱导剂(如 DNA 聚合酶)且在适宜的温度和 pH 条件下。引物既可以是单链的,也可以是双链的,并且必须足够长,使得在诱导剂存在的条件下可以引发所需延伸产物的合成。引物的准确长度依赖于多种因素,包括温度、引物来源和使用的方法。在某些实施方案中,可以被用作引物在酶介导的循环扩增反应如 PCR 反应中扩增基因特定的核酸序列的寡核苷酸由寡核苷酸片段组成。这样的片段应足够长以使其特异性退火或杂交至核酸样品。所述序列一般来说长大约 8 到 14 个核苷酸,但是可以更长。对某些实施方案,更长的序列,例如大约 14- 大约 50 核苷酸是有利的。

[0149] 在需要扩增 DNA 片段的实施方案中,考虑具有来自基因序列的大约 8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23 或 24 个核苷酸的连续延伸 (stretch)。

[0150] 如本文所用,“杂交”指的是核酸一条链与其互补链碱基配对过程,如点杂交技术和 PCR 技术中的那样。

[0151] 无论使用哪种探针序列或杂交方法,本领域普通技术人员可以很容易确定适宜的杂交条件,如杂交温度和化学条件。这些杂交方法是本领域公知的。例如,对于需要高选择性的应用,本领域普通技术人员一般需要采用相对严格的杂交反应条件,如,本领域普通技术人员会选用相对低盐和/或高温的条件,如提供大约 0.02M 至大约 0.10M NaCl,温度大约 50°C 至大约 70°C。这些高严格性条件容忍很少的(如果有的话)在探针和模板或目标链之间的错配。通常希望通过添加增加量的甲酰胺来提供更严格的条件。杂交反应条件的其他变化在本领域公知。(例如可以参见 Sambrook 等人, *Molecular Cloning; A Laboratory*

Manual 2d ed. (1989))。

[0152] 杂交条件基于核酸结合复合物的解链温度,例如Berger和Kimmel(1987,Guide to Molecular Cloning Techniques,Methods in Enzymology,Vol 152,Academic Press,San Diego CA)所教导的那样,并赋予下面所解释的明确的“严格性”。

[0153] 通常最高严格性出现在大约 $T_m - 5^{\circ}\text{C}$ (低于探针 $T_m 5^{\circ}\text{C}$);高严格性在低于 T_m 大约 5°C 至 10°C ;中等严格性在低于 T_m 大约 10°C 至 20°C ,而低严格性在低于 T_m 大约 20°C 至 25°C 。如本领域普通技术人员将会理解,最高严格性杂交可以用于检测或鉴定核苷酸序列的相同性;而中等或低严格性杂交可以用于检测或鉴定相似或相关的多核苷酸序列。

[0154] 一方面,本发明采用在严格条件下可杂交于另一核苷酸序列的核苷酸序列(例如 65°C 和 $0.1\times\text{SSC}$ 条件 { $1\times\text{SSC}=0.15\text{M NaCl},0.015\text{M}$ 柠檬酸钠 $\text{pH}7.0$)。当核苷酸序列为双链时,双链体的两条链,单独地或组合,均可以被本发明所采用。当核苷酸序列为单链时,应当理解该核苷酸序列的互补序列也在本发明的涵盖范围内。

[0155] 杂交的严格性指其中多核苷酸的杂交体是稳定的条件。这些条件对本领域普通技术人员是显而易见的。本领域普通技术人员公知,杂交体的稳定性反映于杂交体的解链温度 (T_m),序列相同性每下降 1%,杂交体的解链温度降低大约 $1-1.5^{\circ}\text{C}$ 。通常杂交体的稳定性是钠离子浓度和温度的函数。一般来说,杂交反应在较高严格性条件下进行,随后用不同严格性进行洗涤。

[0156] 如本文所用的,高严格性包括允许那些在 1M Na^+ 在 $65-68^{\circ}\text{C}$ 形成稳定杂交体的核酸序列杂交的条件。可以提供高严格性条件,例如,通过在包含 $6\times\text{SSC}$ 、 $5\times\text{Denhardt' s}$ 、 1% SDS(十二烷基硫酸钠)、 0.1Na^+ 焦磷酸盐和 0.1mg/ml 变性的鲑精 DNA 作为非特异性竞争物的水溶液中进行杂交。杂交之后,可以在若干步骤完成高严格性洗涤,最后洗涤在杂交温度下在 $0.2-0.1\times\text{SSC}$ 、 0.1% SDS 溶液中进行大约 30 分钟。

[0157] 了解这些条件可以使用多种缓冲液(例如基于甲酰胺的缓冲液)和温度来修改和复制。 Denhardt' s 溶液和 SSC 是本领域普通技术人员熟知的其他适宜的杂交缓冲液(参见例如 Sambrook 等人,eds.(1989)Molecular Cloning:A Laboratory Manual,Cold Spring Harbor Laboratory Press,New York 或 Ausubel 等人,eds.(1990)Current Protocols in Molecular Biology,John Wiley&Sons,Inc.)。一般来说凭经验确定最优化的杂交条件,因为杂交对的长度和 GC 含量也起到作用。

[0158] 与本文所公开的引物和探针序列不同的核酸分子也在本发明范围内。与这些序列互补,或如本文所述在标准或严格条件下与所述序列杂交的核酸序列,以及类似物或衍生物也应在本发明范围内。有利地,这些变体与本文所描述的序列只有少数的核苷酸的不同,例如 1 个、2 个或 3 个核苷酸的不同。

[0159] 相应于本文所描述的序列的天然等位基因变体、同源物(即衍生自其他物种的核酸)或其它相关序列(如旁系同源物(paralog))可以基于它们与本文所述的核酸的同源性而分离,例如利用已知序列的全部或部分作探针通过进行标准或严格的杂交反应而分离。这些核酸杂交和克隆的方法本领域公知。

[0160] 类似的,本发明的方法检测到的核酸分子可能只包含所述特定序列的片段。本文提供的片段定义为至少 6 个(连续的)核酸的序列,长度足以允许与核酸探针或引物的特异性杂交,并且至多是少于全长序列的一些部分。片段可以衍生自选择的核酸序列的任何

连续部分。衍生物和类似物可能是全长或不是全长,如果衍生物或类似物含有修饰的核酸或氨基酸,如下所述。

[0161] 本发明的核酸的衍生物、类似物、同源物和变体包括但不限于,包含与本发明的核酸基本上同源的区段的分子,在不同的实施方案中,这些分子与同样大小的核酸序列或与对齐的序列比较(比对采用本领域公知的计算机同源性程序进行)时有至少大约 70%、80%、85%、90%、95%、96%、97%、98%或甚至 99%的序列相同性。

[0162] 出于本发明的目的,序列相同性或同源性通过比较序列确定,所述序列被比对以最大化重叠和相同性同时最小化序列缺口。具体地,序列相同性可以使用多种数学算法中的任一种确定。用于比较两个序列的数学算法的非限制实例为 Karlin&Altschul, Proc. Natl. Acad. Sci. USA1990 ;872264-2268 的算法,由 Karlin&Altschul 修改, Proc. Natl. Acad. Sci. USA1993 ;90 :5873-5877。

[0163] 另一个用于比较序列的数学算法的实例是 Myers&Miller, CABIOS1988 ;4 :11-17 的算法。这样的算法被并入 ALIGN 程序(2.0 版本),其是 GCG 序列比对软件包的一部分。当运用 ALIGN 程序进行氨基酸序列比较时,可以使用 PAM120 权重残基表、12 分的缺口长度罚分和 4 分的缺口罚分。另一种鉴定局部序列相似性和比对的有用的算法是 FASTA 算法,如 Pearson&Lipman, Proc. Natl. Acad. Sci. USA1988 ;85 :2444-2448 中所描述的。

[0164] 依据本发明使用的有益的算法是 WU-BLAST(Washington University BLAST) 版本 2.0 软件。针对若干 UNIX 平台的 WU-BLAST 版本 2.0 可执行程序可以从 <ftp://blast.wustl.edu/blast/executables> 处下载。该程序基于 WU-BLAST 版本 1.4, WU-BLAST 版本 1.4 又是基于公众域 NCBI-BLAST 版本 1.4(Altschul & Gish, 1996, Local alignment statistics; Doolittle ed., Methods in Enzymology266 :460-480 ;Altschul 等人, Journal of Molecular Biology1990215 :403-410 ;Gish&States, 1993 ;Nature Genetics3 :266-272 ;Karlin&Altschul, 1993 ;Proc. Natl. Acad. Sci. USA90 :5873-5877 ;以上全部通过引用并入本文)。

[0165] 在套件内的全部搜索程序中,缺口比对程序整合入数据库搜索本身。如果需要,缺口形成(gapping)可以被关闭。长度为 1 的缺口的缺省罚分(Q)对于蛋白质和 BLASTP 为 Q=9,对于 BLASTN 为 Q=10,但是可以改变为任何整数。延伸缺口的每残基缺省罚分(R)对蛋白质和 BLASTP 为 R=2,对 BLASTN 为 R=10,但是可以改变为任何整数。可以使用 Q 和 R 的任何组合,以对齐序列使得最大化重叠和相同性同时最小化序列缺口。缺省的氨基酸比较矩阵是 BLOSUM62,但是其他氨基酸比较矩阵例如 PAM 也可以使用。

[0166] 可选地或额外地,术语“同源性”或“相同性”,例如,针对核苷酸或氨基酸序列,可以表示两条序列间同源性的定量量度。序列同源性百分比可以用 $(N_{\text{ref}} - N_{\text{dif}}) * 100 / N_{\text{ref}}$ 计算。这里 N_{dif} 是两个对齐的序列间不同残基的总的数量,而 N_{ref} 是其中一条序列的残基数量。因此, DNA 序列 AGTCAGTC 与序列 AATCAATC 具有 75%的序列相同性 ($N_{\text{ref}}=8$; $N_{\text{dif}}=2$)。“同源性”或“相同性”可以指具有相同核苷酸或氨基酸的位置的数量除以两个序列中的较短序列中核苷酸或氨基酸的数量,其中两序列的比对可以依据 Wilbur 和 Lipman 算法确定(Wilbur&Lipman, Proc Natl Acad Sci USA1983 ;80 :726,通过引用并入本文),例如,使用 20 个核苷酸的窗口大小、4 个核苷酸字长、缺口罚分为 4、以及计算机辅助分析,并且包含比对结果的序列数据解释都可以使用商业上易得的程序很方便的执行(如

Intelligenetics. TM. Suite, Intelligenetics Inc. CA)。当述及 RNA 序列和 DNA 序列相似或具有一定程度相同性或同源性时,认为 DNA 序列中的胸腺嘧啶 (T) 等同 RNA 序列中的尿嘧啶 (U)。因此, RNA 序列也在本发明的范围内,且通过将 DNA 序列中的胸腺嘧啶 (T) 考虑等同于 RNA 序列中的尿嘧啶 (U),可以衍生自 DNA 序列。无需过度的实验,本领域技术人员可以借助很多其他的程序或参考文献来计算百分比同源性。

[0167] 至于密码子优化,本发明的核酸分子具有编码本发明的抗原的核苷酸序列,所述核苷酸序列可以被设计成采用在产生抗原的对象的基因中使用的密码子。本领域技术人员公知这些方法和这些方法的选择。此外,还有若干公司会优化密码子序列,例如 Geneart (TUgeneart(dot)comUT)。因此,可以容易地对本发明的核苷酸序列进行密码子优化。

[0168] 如本文所用,术语“探针”指的是分子(如,寡核苷酸,无论是天然存在的由限制酶切消化纯化得到还是合成产生的,重组的还是 PCR 扩增的),其能够与另一感兴趣的分子(如另一寡核苷酸)杂交。当探针是寡核苷酸时,它们可以是单链的或双链的。探针可用于检测、鉴定和分离特定靶(如基因序列)。如本文所述的,设想本发明使用的探针可以采用标签标记以使其在任何检测系统中都可检测,所述系统包括但不限于酶系统(如 ELISA 以及基于酶的组织化学测定)、荧光系统、放射性系统和发光系统。

[0169] 如本文所述的引物和探针可以很容易制备出来,例如通过化学方法直接合成片段或通过引入选定的序列至重组载体进行重组生产。制备载体或重组体或质粒用于在体内或体外扩增片段的方法可以是任何所需的方法,例如以下文献中或其中引用的文献中描述的方法或类似方法:U. S. Pat. Nos. 4,603,112;4,769,330;4,394,448;4,722,848;4,745,051;4,769,331;4,945,050;5,494,807;5,514,375;5,744,140;5,744,141;5,756,103;5,762,938;5,766,599;5,990,091;5,174,993;5,505,941;5,338,683;5,494,807;5,591,639;5,589,466;5,677,178;5,591,439;5,552,143;5,580,859;6,130,066;6,004,777;6,130,066;6,497,883;6,464,984;6,451,770;6,391,314;6,387,376;6,376,473;6,368,603;6,348,196;6,306,400;6,228,846;6,221,362;6,217,883;6,207,166;6,207,165;6,159,477;6,153,199;6,090,393;6,074,649;6,045,803;6,033,670;6,485,729;6,103,526;6,224,882;6,312,682;6,348,450 和 6,312,683;U. S. 专利申请系列号 920,197,1986 年 10 月 16 日提交;W090/01543;W091/11525;W094/16716;W096/39491;W098/33510;EP265785;EP0370573;Andreansky 等人., Proc. Natl. Acad. Sci. USA1996;93:11313-11318;Ballay 等人., EMBO J. 1993;4:3861-65;Felgner 等人., J. Biol. Chem. 1994;269:2550-2561;Frolov 等人, Proc. Natl. Acad. Sci. USA1996;93:11371-11377;Graham, Tibtech1990;8:85-87;Grunhaus 等人, Sem. Virol. 1992;3:237-52;Ju 等人, Diabetologia1998;41:736-739;Kitson 等人, J. Virol. 1991;65:3068-3075;McClements 等人, Proc. Natl. Acad. Sci. USA1996;93:11414-11420;Moss, Proc. Natl. Acad. Sci. USA1996;93:11341-11348;Paoletti, Proc. Natl. Acad. Sci. USA1996;93:11349-11353;Pennock 等人, Mol. Cell. Biol. 1984;4:399-406;Richardson (Ed), Methods in Molecular Biology1995;39, "Baculovirus Expression Protocols," Humana Press Inc.;Smith 等人 (1983) Mol. Cell. Biol. 1983;3:2156-2165;Robertson 等人, Proc. Natl. Acad. Sci. USA1996;93:11334-11340;Robinson

等人, *Sem. Immunol.* 1997 ;9 :271 ; 和 Roizman, *Proc. Natl. Acad. Sci. USA* 1996 ;93 : 11307-11312。探针设计的策略如 W095/11995、EP717,113 和 W097/29212 所述。

[0170] 本发明进一步考虑了直接和间接标记技术。例如,直接标记将荧光染料直接掺入与阵列连接的探针杂交的核苷酸序列(例如在标记的核苷酸或 PCR 引物存在的情况下,通过酶促合成将染料掺入核苷酸序列)。直接标记方案产生强杂交信号,通常使用具有相似化学结构和特性的荧光染料家族,并且很容易实现。在某些包括直接标记的核酸的实施方案中,酞菁或 Alexa 类似物用于多重荧光比较阵列分析。在另一些实施方案中,间接标记方案可以用于在与微阵列探针杂交前或杂交后将表位掺入核酸。一种或多种染色步骤和试剂被用于标记杂交复合物(如,结合表位的荧光分子,由此借助于染料分子与杂交物质的表位的缀合而提供荧光信号)。

[0171] 用作本发明的探针的寡聚核苷酸序列可以用可检测的部分标记。多种标记部分为本领域公知。所述部分可以是,例如,放射性标记(如 ^3H 、 ^{125}I 、 ^{35}S 、 ^{14}C 、 ^{32}P 等)、可检测的酶(如辣根过氧化物酶(HRP)、碱性磷酸酶等)、荧光染料(如异硫氰酸荧光素、得克萨斯红、罗丹明、Cy3、Cy5、Bodipy、Bodipy Far Red、荧光黄、Bodipy630/650-X、Bodipy R6G-X 和 5-CR6G 等等)、比色法标记如胶体金或有色玻璃或塑料(如聚苯乙烯、聚丙烯、乳胶等)、珠或可以形成可检测信号(如比色法、荧光、化学发光法或电化学发光(ECL)信号)的任何其他部分。

[0172] 探针可以用可检测部分直接或间接标记,或合成探针以掺入可检测部分。在一实施方案中,可检测标记在酶介导的循环扩增反应中至少一个循环期间被掺入核酸中。例如,在酶介导的循环扩增反应过程中可以用聚合酶掺入荧光核苷酸。或者,可以在核酸引物或探针合成期间掺入荧光核苷酸。为了用荧光染料标记寡核苷酸,可以使用众所周知的标签标记方法的其中一种(*Nature Biotechnology*, 14, 303-308, 1996 ; *Applied and Environmental Microbiology*, 63, 1143-1147, 1997 ; *Nucleic Acids Research*, 24, 4532-4535, 1996)。有益的探针是用荧光染料在 3' 或 5' 端标记并含有 G 或 C 作为标记末端处的碱基的探针。如果 5' 端被标记而 3' 端未被标记,3' 核糖或脱氧核糖位置的 C 原子上的 OH 基团可以用磷酸基团等修饰,尽管在这方面没有任何强加的限制。

[0173] 可以应用分光光度、光化学、生物化学、免疫化学、电学、光学或化学手段检测这样的标记。检测装置和方法可以包括但不限于光学成像、电子成像、使用 CCD 照相机成像、集成的光学成像和质谱。此外,可以定量与靶点结合的标记的或未标记的探针的数量。这样的定量包括统计学分析。在另一些实施方案中,可以通过一致和不一致位点间的导电性差异,通过猝灭、通过荧光干扰分析或通过供体和受体分子之间的电子转运进行检测。

[0174] 在另外一些实施方案中,可以通过在 PCR 或杂交反应中杂交复合物的分子间的能量传递进行检测,例如通过荧光能量转移(FET)或荧光共振能量转移(FRET)。在 FET 和 FRET 方法中,一或多种核酸探针用荧光分子标记,其中一种可以作为能量供体,另一种作为能量受体分子。它们有时被分别称作报道分子和猝灭分子。供体分子被特定波长的光激发,因此其通常显示荧光发射波长。受体分子也可以在该波长激发,从而通过多种依赖距离的能量转移机制接受供体分子的发射能量。一般来说,当受体分子与供体分子接近时(如在同一个分子上或相邻分子上),受体分子可以接受供体分子的发射能量。FET 和 FRET 为本领域公知。参见例如:U. S. Pat. Nos. 5668648、5707804、5728528、5853992 和 5869255(描述

了 FRET 染料), Tyagi 等人 Nature Biotech. vol. 14, p303-8(1996), 和 Tyagi 等人, Nature Biotech. vol. 16, p49-53(1998) (描述 FET 分子信标), 和 Mergny 等人 Nucleic Acid Res. vol. 22, p920-928, (1994) 和 Wolf 等人 PNAS vol. 85, p8790-94(1988) (FET 和 FRET 一般说明和方法), 这些各自通过引用并入本文。

[0175] 本发明的核苷酸序列可以插入载体中。术语“载体”由本领域普通技术人员广泛了解和使用。如本文所述, 术语“载体”的使用与本领域普通技术人员的了解相同。例如, 由本领域普通技术人员通常使用的术语“载体”指的是一种媒介物 (vehicle), 其允许或便于核酸分子从一种环境到另一种的转移, 或者可以允许或便于对核酸分子的操作。

[0176] 举例来说, 载体是复制子 (replicon), 例如质粒、噬菌体或粘粒, 可以与另一 DNA 节段连接以实现所连接的节段的复制。“复制子”可以是任何遗传元件, 其功能在于作为体内 DNA 复制的自发单元, 即可以在其自身控制下复制。“复制起点”指的是参与 DNA 合成的那些 DNA 序列。“表达控制序列”指的是 DNA 序列, 其控制和调节另一 DNA 序列的转录和翻译。当 RNA 聚合酶将编码序列转录为 mRNA, 所述 RNA 随后被翻译为编码序列编码的蛋白质时, 所述编码序列在细胞内是“可操作地连接的”和在转录和翻译控制序列的控制之下的。

[0177] 一般来说, 包含促进插入的 DNA 片段的有效转录和翻译的启动子序列的表达载体与宿主结合使用。表达载体一般来说包含复制起点、启动子、终止子以及能够在转化的细胞提供表型选择的特定基因。当多核苷酸编码多蛋白片段时, 有利地, 在载体中将起始密码子 (ATG) 置于读码框 5' 端, 将终止密码子置于 3' 端。可能存在控制表达的其他元件, 例如增强子序列、稳定序列和允许蛋白质分泌的信号序列。转化的宿主可以依据本领域公知的方法发酵或培养以达到最佳的细胞生长。

[0178] 依据本发明, 可以进行使用能够表达本发明免疫原的任何载体。在某些实施方案中, 本发明的免疫原可以在体外 (如使用无细胞的表达系统) 和 / 或在体外生长的培养细胞中。对于这样的应用, 可使用任何允许在体外和 / 或在培养的细胞内表达免疫原的载体。

[0179] DNA “编码序列”是双链 DNA 序列, 当置于适合的调控序列的控制下时, 在体内转录并翻译成多肽。编码序列的边界通过 5' (氨基) 端的起始密码子和 3' (羧基) 端的翻译终止密码子来确定。编码序列可以包含但不限于原核生物序列、来自真核生物 mRNA 的 cDNA、真核生物 (例如哺乳动物) 的基因组 DNA、和甚至是合成的 DNA 序列。多腺苷酸化信号和转录终止序列通常位于编码序列的 3' 端。“cDNA”指的是拷贝 -DNA 或互补 -DNA, 是 mRNA 转录物的反转录反应的产物。

[0180] 转录和翻译控制序列是 DNA 调控序列, 例如启动子、增强子、核糖体结合位点、上游调控域、多腺苷酸化信号、终止子等, 其在宿主细胞中提供编码序列的表达。“顺式作用元件”是核苷酸序列, 也被称作为“共有序列”或“基序”, 其与其他能够上调或下调特定基因座的表达的蛋白质相互作用。“信号序列”也可以包含于编码序列中。该序列编码信号肽, 在多肽的 N-端, 与宿主细胞联系并引导多肽进入适当的细胞位置。可以发现信号序列与多种原核生物和真核生物天然蛋白质相关。只要需要的基因可以转录和翻译, 在重组载体中并不总是需要存在所有的这些控制序列。

[0181] “启动子序列”是 DNA 调控区, 可以在细胞内与 RNA 聚合酶结合并起始下游 (3' 方向) 编码序列的转录。一般来说, 启动子序列的边界在其 3' 端限于转录起始位点, 向上游 (5' 方向) 延伸, 直至包含最少量的起始达到高于背景的可以检测水平的转录所必需的碱

基或元件。在启动子区内是转录起始位点,以及负责结合 RNA 聚合酶的蛋白质结合结构域(共有序列)。真核生物的启动子通常但并不是总是含有“TATA”盒和“CAT”盒。原核生物的启动子除了 -10 和 -35 共有序列外还包含 Shine-Dalgarno 序列。

[0182] 如本文所用的,术语“限制性核酸内切酶”和“限制性酶”指的是用来在特定核苷酸序列处或附近切开双链 DNA 的酶。

[0183] “重组 DNA 技术”指的是将两个异源的 DNA 分子联合起来的技术,通常是不同生物体的 DNA 分子在体外连接的结果。重组 DNA 分子通常由遗传工程实验生产。其同义词包括“基因剪接”、“分子克隆”和“遗传工程”。这些操作的产物获得“重组体”或“重组分子”。

[0184] 当外源的或异源的 DNA 被引入细胞内部时,细胞被这样的 DNA “转化”或“转染”。转化的 DNA 可能整合也可能不整合(即共价连接)进细胞基因组。例如在原核动物、酵母菌和哺乳动物细胞中,转化的 DNA 可以保持在独立的游离型(episomal)元件如载体或质粒上。对于真核细胞,稳定转化的细胞是这样的细胞,其中转化的 DNA 整合进入染色体使得在染色体复制时被子细胞遗传。通过真核细胞建立含转化 DNA 的子细胞群组成的细胞系或克隆的能力证实该稳定性。“克隆”是从单个细胞或祖先通过有丝分裂衍生出的细胞群。“细胞系”是原代细胞的克隆,其能在体外稳定生长很多代。被转化进外源 DNA 的生物体,例如植物或动物,称为“转基因的”。

[0185] 如本文所用的,术语“宿主”的意义不仅包括原核生物,也包括真核生物,例如酵母菌、植物和动物细胞。原核生物宿主可以包括大肠杆菌、*S. typhimurium*、粘质沙雷菌(*Serratia marcescens*)和枯草芽孢杆菌(*Bacillus subtilis*)。真核生物宿主包括酵母如巴斯德毕赤氏酵母(*Pichia pastoris*)、哺乳动物细胞和昆虫细胞和植物细胞,例如拟南芥(*Arabidopsis thaliana*)和烟草(*Tobaccum nicotiana*)。很多哺乳动物细胞系是本领域公知的,包括可从 American Type Culture Collection(ATCC)获得的永生细胞系,例如但不限于中国仓鼠卵巢(CHO)细胞、HeLa 细胞、幼仓鼠肾(BHK)细胞、猴肾细胞(COS)、人肝细胞癌细胞(如 Hep G2)、Madin-Darby 牛肾细胞(“MDBK”)及其他。类似的,细菌宿主包括,大肠杆菌、枯草芽孢杆菌(*Bacillus subtilis*)和链球菌属(*Streptococcus* spp.),将可应用于本发明的表达载体。可用于本发明的酵母宿主包括尤其是酿酒酵母(*Saccharomyces cerevisiae*)、白色假丝酵母(*Candida albicans*)、麦芽糖假丝酵母(*Candida maltosa*)、多形汉森酵母(*Hansenula polymorpha*)、脆壁克鲁维酵母(*Kluyveromyces fragilis*)、乳克鲁维酵母(*Kluyveromyces lactis*)、*Pichia guilliermondii*、巴斯德毕赤氏酵母(*Pichia pastoris*)、裂殖酵母菌丝(*Schizosaccharomyces pombe*)和子囊菌酵母(*Yarrowia lipolytica*)。在本发明中可用的插入宿主包括但不限于草地贪夜蛾(*Spodoptera frugiperda*)细胞。

[0186] DNA 构建体的“异源”区域是较大 DNA 分子中的可鉴别的 DNA 节段,所述节段没有发现天然与所述较大分子相关。因此,当异源区编码哺乳动物基因,所述基因通常的侧翼 DNA 不是来源生物体基因组中哺乳动物基因组 DNA 侧翼的 DNA。在另一个实例中,编码序列是构建体,其中所述编码序列自身并不天然存在(如 cDNA,而基因组编码序列中含有内含子;或合成的序列,其具有不同于天然基因的密码子)。等位基因变异或天然发生的突变事件不足以给出如本文定义的 DNA 异源区。例如,多核苷酸可以通过遗传工程技术置入不同来源的质粒或载体中,就是异源多核苷酸。从其天然编码序列上移走并可操作地连接到除所述

天然序列外的编码序列上的启动子是异源启动子。

[0187] 如本文所用的,用于基因或多肽的“片段”或“部分”,通常长度将是至少 10 个残基,更典型地为至少 20 个残基,且优选至少 30 个(如 50 个)残基,但是要短于整个完整的序列。这些基因片段可以通过本领域技术人员公知的方法生成,例如通过 DNA 重组技术产生(使用编码所需基因的载体),或通过化学方法合成。

[0188] 制备和/或施用载体或重组体或质粒的方法以体内或体外表达本发明的基因的基因产物的方法可以是任何所需的方法,例如以下文献描述的方法或类似方法:美国专利号 4,603,112;4,769,330;4,394,448;4,722,848;4,745,051;4,769,331;4,945,050;5,494,807;5,514,375;5,744,140;5,744,141;5,756,103;5,762,938;5,766,599;5,990,091;5,174,993;5,505,941;5,338,683;5,494,807;5,591,639;5,589,466;5,677,178;5,591,439;5,552,143;5,580,859;6,130,066;6,004,777;6,130,066;6,497,883;6,464,984;6,451,770;6,391,314;6,387,376;6,376,473;6,368,603;6,348,196;6,306,400;6,228,846;6,221,362;6,217,883;6,207,166;6,207,165;6,159,477;6,153,199;6,090,393;6,074,649;6,045,803;6,033,670;6,485,729;6,103,526;6,224,882;6,312,682;6,348,450 和 6,312,683;美国专利申请系列号 920,197,1986 年 10 月 16 日提交;W090/01543;W091/11525;W094/16716;W096/39491;W098/33510;EP265785;EP0370573;Andreansky 等人,Proc.Natl.Acad.Sci.USA1996;93:11313-11318;Ballay 等人,EMBO J.1993;4:3861-65;Felgner 等人,J.Biol.Chem.1994;269:2550-2561;Frolov 等人,Proc.Natl.Acad.Sci.USA1996;93:11371-11377;Graham, Tibtech1990;8:85-87;Grunhaus 等人,Sem.Virol.1992;3:237-52;Ju 等人,Diabetologia1998;41:736-739;Kitson 等人,J.Virol.1991;65:3068-3075;McClements 等人,Proc.Natl.Acad.Sci.USA1996;93:11414-11420;Moss, Proc.Natl.Acad.Sci.USA1996;93:11341-11348;Paoletti, Proc.Natl.Acad.Sci.USA1996;93:11349-11353;Pennock 等人,Mol.Cell.Biol.1984;4:399-406;Richardson(Ed),Methods m Molecular Biology 1995;39,“Baculovirus Expression Protocols,”Humana Press Inc.;Smith 等人(1983)Mol.Cell.Biol.1983;3:2156-2165;Robertson 等人,Proc.Natl.Acad.Sci.USA1996;93:11334-11340;Robinson 等人 Sem.Immunol.1997;9:271;and Roizman,Proc.Natl.Acad.Sci.USA1996;93:11307-11312。

[0189] 本发明也提供了包含本发明的载体的转化的宿主细胞。在某些实施方案中,载体通过转染、电穿孔或感染导入细胞中。本发明也提供了一种制备转化的细胞(该转化的细胞可以表达本发明的免疫原)的方法,包括用表达载体(如 DNA 疫苗)转染、电穿孔或感染细胞,从而产生被感染的生产细胞,并在足以在所述宿主细胞中表达所述免疫原的生物学条件下维持细胞生长。

[0190] 依据本发明的另一实施方案,表达载体是用于在适当的细胞系统中体外表达蛋白质的表达载体。表达的蛋白质可以在蛋白质分泌之后或不在分泌之后(如果蛋白质不分泌,一般会出现细胞裂解或进行细胞裂解)从培养物上清中收获,任选地使用浓缩方法(如超滤)进行浓缩和/或使用纯化手段(例如亲和层析、离子交换层析或凝胶过滤型层析方法)纯化。

[0191] 本领域技术人员了解宿主细胞的培养条件依据特定的基因而不同,因此有时必需

通过常规实验来确定含有载体的宿主细胞的最佳培养条件。“宿主细胞”指被遗传学改变的原核或真核细胞,或可以通过施用外源多核苷酸(例如重组的质粒或载体)而在遗传学改变的原核或真核细胞。当提及被遗传学改变的细胞时,该术语既指最初被改变的细胞,也指其后代细胞。

[0192] 含有所需序列的多核苷酸可以插入适宜的克隆或表达载体中,而所述载体进而被引入适宜的宿主细胞中以进行复制和扩增。多核苷酸可以采用本领域公知的任何方法引入宿主细胞中。可以通过的众多适宜的方法中的任何一将包含感兴趣的多核苷酸的载体引入宿主细胞,包括直接摄取、胞吞作用、转染、f⁻接合、电穿孔、使用氯化钙、氯化铷、磷酸钙、二乙氨基乙基葡聚糖或其他物质的转染、微粒轰击、脂转染(lipofection)和感染(其中载体具有感染性,例如反转录病毒载体)。引入载体或多核苷酸的选择通常依赖于宿主细胞的特征。

[0193] 对于免疫原需要在体内表达的应用,例如当本发明的免疫原用于 DNA 或含 DNA 的疫苗时,可以使用任何能够表达本发明的免疫原且体内使用安全的载体。在优选的实施方案中,所用的载体在人类、哺乳动物和/或实验动物中使用安全。

[0194] 一般应该选择本发明使用的载体使得其包含适合的基因调控区(例如启动子和增强子),使得可以表达本发明的免疫原。

[0195] 例如,当目的是体外或在培养的细胞中,或者在为了生产编码这些免疫原的蛋白质的任何原核或真核系统中表达本发明的免疫原时,可以基于所述应用使用任何适合的载体。例如,可使用质粒、病毒载体、细菌载体、原生动物载体、昆虫载体、杆状病毒表达载体、酵母载体、哺乳动物细胞载体等等。技术人员可以考虑载体的性质和在具体环境下表达免疫原的需要,选择适合的载体。

[0196] 当目的是在一个对象体内表达本发明的免疫原时,例如为了产生针对 HIV 抗原的免疫应答和/或针对 HIV 的保护性或治疗性免疫应答,应当选择适合于在该对象上表达,并且体内使用安全的表达载体。例如,在某些实施方案中,需要在实验动物表达本发明的免疫原,例如用于本发明的 HIV 免疫原性组合物和疫苗的临床前实验。在另一些实施方案中,需要在人类对象表达本发明的免疫原,例如用于本发明的 HIV 免疫原性组合物和疫苗进行临床实验和实际临床应用。可以使用任何适合于这些用途的载体,技术人员有能力挑选适合的载体。在某些实施方案中,优选用于这些体内应用的载体被减弱以防止载体在对象内的扩增。例如,如果使用质粒载体,首选那些缺乏在对象中起作用的复制起点的载体,以增强在对象内使用的安全性。如果使用病毒载体,优选地其在对象中是减毒的或是复制缺陷的,也是为了增强在对象内使用的安全性。

[0197] 适于作为疫苗施用的任何载体可以在本发明中采用。在本发明的某些实施方案中,使用适合作为 DNA 疫苗使用的载体,例如 pVAX 和 pcDNA 载体(Invitrogen)。

[0198] 在本发明的另一些实施方案中,使用病毒载体。病毒表达载体是本领域技术人员公知,例如包括腺病毒(如腺病毒亚型 Ad5、Ad11、Ad26、Ad35、Ad48 和 Ad49)、腺伴随病毒(AAV)、甲病毒、反转录病毒和痘病毒(包括鸟痘病毒、减毒痘病毒和痘苗病毒如修饰的牛痘病毒(MVA))。在某些实施方案中,本发明的疫苗包含选自 Ad5、Ad11、Ad26、Ad35、Ad48 和 Ad49 的腺病毒。这样的病毒当作为表达载体使用时,其在选定的对象(如人)中天然不致病或经修饰使其在选定的对象中不致病。例如,复制缺陷腺病毒和甲病毒是公知的且可

用做基因递送载体。

[0199] 表达之后,可以使用任何本领域已知的技术分离和 / 或纯化或浓缩本发明的抗原。例如,可以使用阴离子或阳离子交换层析、磷酸纤维素层析、疏水相互作用层析、亲和层析、免疫亲和层析、羟磷灰石层析、凝集素层析、分子筛层析、等电点聚焦、凝胶电泳或任何其他适合的方法或方法的组合。

[0200] 在某些实施方案中,体内施用本发明的核苷酸序列和 / 或抗原,例如当目的是在对象产生免疫应答时。“对象”在本发明背景下可以是任何动物。例如,在某些实施方案中,需要在实验动物表达本发明的抗原,例如用于本发明的 HIV 免疫原性组合物和疫苗的临床前实验。在另一些实施方案中,需要在人类对象表达本发明的免疫原,例如用于本发明的 HIV 免疫原性组合物和疫苗进行临床实验和实际临床应用。在某些实施方案中,对象是人类,例如感染了 HIV 的人或具有感染 HIV 风险的人。

[0201] 对于这样的体内应用,本发明的核苷酸序列和 / 或抗原,优选地作为免疫原性组合物的组分施用,所述免疫原性组合物包括与药用可接受的载体混合的本发明的核苷酸序列和 / 或抗原。本发明的免疫原性组合物可用于刺激针对 HIV 的免疫应答且可以用作针对 HIV 的预防性或治疗性疫苗的一种或多种组分,用于预防、改善和治疗 HIV。本发明的核酸和载体可用于提供遗传疫苗,即用于将编码本发明抗原的核酸递送至对象(如人)的疫苗,使得所述抗原然后在对象内表达并引起免疫应答。

[0202] 免疫原性组合物

[0203] 如本文所用的术语“免疫原性蛋白质或肽”也包括免疫学活性的肽或多肽,从该意义上说,一旦给宿主施用,可以诱发针对该蛋白的体液和 / 或细胞类型的免疫应答。优选地所述蛋白质片段与总蛋白具有基本上相同的免疫学活性。因此,依据本发明的蛋白质片段含有至少一种表位或抗原决定簇。术语表位指的是能够诱导体液免疫应答(B细胞)或细胞免疫应答(T细胞)的蛋白位点。

[0204] 术语“免疫原性蛋白质或肽”还考虑对序列的缺失、添加和取代,只要所述多肽行使其诱导如本文所述的免疫应答的功能。

[0205] 术语“表位”指的是在抗原或半抗原上的 B 细胞和 / 或 T 细胞特异性应答的位点。该术语也可以与“抗原决定簇”或“抗原决定簇位点”互换使用。识别相同表位的抗体可以用简单的免疫测定鉴定,所述免疫测定显示一种抗体阻断另一种抗体与靶抗原结合的能力。

[0206] 对组合物或疫苗的“免疫应答”指的是在宿主中出现的对感兴趣的组合物或疫苗产生的细胞和 / 或抗体介导的免疫应答。通常,“免疫应答”包含但不限于以下一种或几种效应:生产特异性针对抗原或包含于感兴趣的组合物或疫苗中的抗原的抗体、B 细胞、辅助 T 细胞、抑制性 T 细胞、和 / 或细胞毒性 T 细胞和 / 或 $\gamma \delta$ T 细胞。优选地,宿主将会展示出治疗性或保护性免疫应答,使得对新的感染的抵抗力将得到增强和 / 或疾病的临床症状得到减轻。在受感染的宿主中通常呈现的症状的减少或消失,以及在受感染的宿主中较快的恢复时间和较低的病毒滴度都证明这样的保护。

[0207] 免疫应答的产生可能与抗原呈递细胞(APC)有关。抗原呈递细胞(APC)可能是“专业”的抗原呈递细胞或另外的可被诱导为呈递抗原至 T 细胞的细胞。抗原呈递细胞(APC)包括树突细胞(DC)如指状突树突细胞或滤泡状树突细胞、郎格汉斯细胞、PBMC、巨噬细胞、B 淋巴细胞或其它的细胞类型如上皮细胞、成纤维细胞或内皮细胞,这些细胞通过转染在其

表面表达 I 类或 II 类 MHC 分子从而活化或工程化。抗原呈递细胞 (APC) 也包括杂交瘤、淋巴瘤和合成的 APC 例如脂质膜。抗原呈递细胞 (APC) 的前体包括 CD34+ 细胞、单核细胞、成纤维细胞和内皮细胞。可以促进免疫增强作用的细胞因子基因包括 IL-2、IL-12、IFN- γ 、TNF- α 、IL-18 等。这样的蛋白质包括 MHC 分子 (I 类或 II 类)、CD80、CD86 或 CD40。T 细胞的实例包括辅助 T 细胞 (CD4+) 和 CD8+ 细胞。

[0208] 如本文所用的“免疫原性”蛋白质或多肽也指的是引起如上所述的免疫应答的氨基酸序列。如本文所用的“免疫原性”蛋白质或多肽包括蛋白质的全长序列、其类似物或其具有免疫原性片段。“免疫原性片段”意思是包含一个或多个表位因此可以引起如上所述的免疫应答的蛋白质片段。这样的片段可以使用多种本领域公知的表位定位技术进行鉴定。参见如 Epitope Mapping Protocols in Methods in Molecular Biology, Vol. 66 (Glenn E. Morris, Ed., 1996) Humana Press, Totowa, N. J. 例如, 可以通过如下方式确定线性表位: 在固体支持物上同时合成大量肽, 所述肽相应于蛋白质分子的一部分, 并使所述肽在保持与支持物相连的状况下与抗体反应。这样的技术本领域公知, 如 U. S. Pat. No. 4, 708, 871; Geysen 等人 (1984) Proc. Natl. Acad. Sci. USA 81:3998-4002; Geysen 等人 (1986) Molec. Immunol. 23:709-715 所述, 所有内容以其整体通过引用并入本文。类似的, 构象决定簇易于通过测定氨基酸的空间构象的方法进行鉴定, 例如通过 x-射线晶体学和二维核磁共振鉴定。参见例如 Epitope Mapping Protocols, 同上。

[0209] 合成抗原也包含在本定义范围内, 例如, 多表位、侧翼表位和其他重组的或合成的衍生抗原。参见例如 Bergmann 等人 (1993) Eur. J. Immunol. 23:2777-2781; Bergmann 等人 (1996) J. Immunol. 157:3242-3249; Suhrbier, A. (1997) Immunol. and Cell Biol. 75:402-408; Gardner 等人 (1998) 12th World AIDS Conference, Geneva, Switzerland, Jun. 28-Jul. 3, 1998。为了本发明的用途, 免疫原性片段通常包含至少大约 3 个氨基酸, 至少大约 5 个氨基酸, 至少大约 10-15 个氨基酸或至少大约 25 个或更多个氨基酸。所述片段的长度并没有严格的上限, 可以包含几乎蛋白质序列全长, 或甚至是包含所述蛋白质至少一个表位的融合蛋白。

[0210] 正如前面提到的, 表位确定方法无需过度的实验即可用于本发明的实践, 例如生成重叠肽文库 (Hemmer B. 等人, Immunology Today, 1998, 19(4), 163-168)、肽扫描技术 (Geysen 等人, 1984 Proc. Nat. Acad. Sci. USA, 81, 3998-4002; Geysen 等人, 1985 Proc. Nat. Acad. Sci. USA, 82, 178-182; Van der Zee R. 等人, 1989 Eur. J. Immunol., 19, 43-47; Geysen H. M., 1990 Southeast Asian J. Trop. Med. Public Health, 21, 523-533; Multipin. RTM. Peptide Synthesis Kits de Chiron) 和算法 (De Groot A. 等人, 1999 Nature Biotechnology, 17, 533-561), 以及在 PCT 申请系列号 PCT/US2004/022605, 所有这些通过引用以其整体并入本文。也可以参考其他在此引用和并入的文献查询测定免疫原或抗原的表位的方法以及测定编码该表位的核酸分子的方法。

[0211] 如本文所用的, 术语“抗原”或“免疫原”可交换使用, 指的是物质, 通常为蛋白质, 其能够在对象中诱导免疫应答。本术语也可以指具有免疫活性的蛋白质, 一旦施用至对象 (包括向对象直接施用蛋白质或施用编码该蛋白质的核酸序列或载体) 能够诱发针对此蛋白质的体液和 / 或细胞类型免疫应答。

[0212] 术语“抗体”包括完整的分子以及其片段, 例如 Fab、F(ab')₂、Fv 和 scFv, 其可以

与抗原决定簇结合。这些抗体片段保留对抗原或受体的选择结合的一些能力,并且包括,例如:

[0213] (i) Fab, 抗体片段, 包含抗体分子的单价抗原结合片段, 可以通过整个抗体经木瓜蛋白酶消化而产生, 获得完整的轻链以及重链的一部分。

[0214] (ii) Fab', 抗体分子的片段, 可以通过整个抗体经胃蛋白酶处理, 随后还原得到, 获得完整的轻链以及重链的一部分, 从每个抗体分子可以得到 2 个 Fab' 片段。

[0215] (iii) F(ab')₂, 抗体分子的片段, 可以通过整个抗体经胃蛋白酶处理后不经还原过程得到, F(ab')₂ 是两个 Fab' 片段通过二硫键结合在一起而成的二聚体。

[0216] (iv) scFv, 包括遗传工程化的片段, 包含重链和轻链的可变区, 作为融合的单链分子。

[0217] 制备这些片段的一般方法本领域公知。(参见例如 Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, New York (1988), 通过引用并入本文)

[0218] “中和抗体”, 是能够中和病原体在宿主和 / 或体外的靶细胞中起始和 / 或持续感染的能力的抗体。中和抗体可以以中和指数 > 1.5 或 > 2.0 抑制 HIV-1 的进入。广谱的和有效的中和抗体在中和测定实验中可以中和超过大约 50% 的 HIV-1 的病毒 (来源于不同进化枝或同一进化枝的不同毒株)。在中和测定实验中, 单克隆抗体中和大约 50% 的输入病毒的抑制浓度可以少于大约 25mg/ml。

[0219] “分离的抗体”或“非天然存在的抗体”是从其天然环境的组分中分离的和 / 或回收的抗体。天然环境中的污染成分是干扰抗体用于诊断或治疗的物质, 可能包括酶、激素和其他蛋白质的或非蛋白质溶解物。在优选的实施方案中, 抗体被纯化至: (1) 通过 Lowry 法测量的按重量计多于 95% 的抗体, 且最优选按重量计多于 99%; (2) 足以使用旋杯式测序仪获得 N-端或内部氨基酸序列中至少 15 个残基的程度; 或 (3) 通过 SDS-PAGE 在还原和非还原条件下使用考马斯蓝染色或更优选采用银染法鉴定的均质性。分离的抗体包括在重组细胞中的原位抗体, 因为抗体天然环境的组分中至少一种不存在。一般来说, 尽管如此, 分离抗体将通过至少一个纯化步骤制备。

[0220] 如本文所用的术语“单克隆抗体”指的是从基本上均质的抗体群中获得的抗体, 即组成该抗体群的个体抗体除了自然状态下可能少量出现的突变外是相同的。单克隆抗体是高度特异的, 针对单个抗原位点。此外, 与包含针对不同决定簇 (表位) 的不同抗体的多克隆抗体制备物不同, 每一单克隆抗体针对抗原上的单个决定簇。除了它们的特异性外, 单克隆抗体优于多克隆抗体之处在于可以合成, 不受其他抗体污染。修饰词“单克隆”并不是限制抗体的生产需要通过任何特定的方法。例如, 在本发明中可使用的单克隆抗体可以通过杂交瘤方法制备, 该方法首先见于 Kohler 等人, *Nature*, 256:495 (1975); 或可以通过在细菌、真核动物或植物细胞中施用重组 DNA 的方法得到 (参见例如 U. S. Pat. No. 4, 816, 567)。“单克隆抗体”也可以从噬菌体抗体库中分离得到, 使用的技术参见例如 Clackson 等人, *Nature*, 352:624-628 (1991) 和 Marks 等人, *J. Mol. Biol.*, 222:581-597 (1991)。

[0221] “抗体片段”包含完整抗体的一部分, 优选地包含完整抗体的抗原结合区或可变区。抗体片段的例子包括 Fab、Fab'、F(ab')₂ 和 Fv 片段、双抗体、线性抗体 (参见 U. S. Pat. No. 5, 641, 870; Zapata 等人, *Protein Eng.* 8(10):1057-1062 [1995])、单链抗体

分子和由抗体片段形成的多特异性抗体。

[0222] 需要了解的是,蛋白质,包括本发明的抗体,可以与本文图解和描述的确切序列有区别。因此本发明考虑对所示序列的缺失、添加和取代,只要序列的功能与本发明的方法一致。在这方面中,特别优选的取代在自然界中通常是保守的,即发生在氨基酸家族内部的置换。例如,氨基酸通常被分为4个家族:(1)酸性——天冬氨酸和谷氨酸盐;(2)碱性——赖氨酸、精氨酸和组氨酸;(3)非极性——丙氨酸、缬氨酸、亮氨酸、异亮氨酸、脯氨酸、苯丙氨酸、甲硫氨酸、色氨酸;和(4)不带电极性——甘氨酸、天冬酰胺、谷氨酰胺、半胱氨酸、丝氨酸、苏氨酸、酪氨酸。苯丙氨酸、色氨酸和酪氨酸有时被归类到芳香族的氨基酸。可以合理推测分离的或非天然存在的以下情况:亮氨酸被异亮氨酸或缬氨酸取代,或反之亦然;天冬氨酸被谷氨酸盐取代,或反之亦然;苏氨酸被丝氨酸取代,或反之亦然;或氨基酸与结构类似的氨基酸的相似保守性置换,对蛋白质生理活性没有明显影响。因此,基本上具有与图解或描述的序列相同的氨基酸序列,但是具有较少的且不在实质上影响蛋白质的免疫原性的氨基酸置换的蛋白质在本发明的范围内。

[0223] 依据本发明,在某些实施方案中,在疫苗接种程序框架内本发明中疫苗的施用可以与其他疫苗接种组合,以免疫或疫苗接种试剂盒或方法的形式,或以多价的免疫原性组合和多价疫苗的形式,如包含至少一种针对靶病原体(如HIV)的疫苗组分,和至少一种针对另一种病原体的疫苗组分。这也包括了同一表达载体表达至少两种病原体的基因。

[0224] 本发明因此也关于多价的或“鸡尾酒式的”免疫原性组合,或多价的或“鸡尾酒式的”,针对靶病原体(如HIV),也针对至少另一种靶物种的疫苗,使用同一体内表达载体,其包含和表达至少一种本发明的靶病原体(如HIV)的多核苷酸和至少一种表达另一病原体的免疫原的多核苷酸。

[0225] 如本文所述,这些多价体组合或疫苗也可以包含药用可接受的载体或媒介物或赋形剂,以及任选存在的佐剂。

[0226] 如本文所述的这些免疫原性组合和疫苗,也可以与至少一种常规疫苗(如灭活的、活减毒的或亚单位疫苗)组合,所述常规疫苗针对同一种病原体或所述免疫原性组合和疫苗针对的物种的另一种病原体。如本文所述的这些免疫原性组合和疫苗,可以在常规疫苗施用前或后施用,例如采用一种“初始-加强”方案。

[0227] 配制物

[0228] 本发明的组合可以包括任意本领域已知的药用可接受的载体。

[0229] 为便于本发明的疫苗的施用,疫苗可以配制成成为合适的药物组合。一般来说,这些组合包含活性成分(如DNA疫苗)和药用可接受的载体。用于医学应用,这些组合物适宜递送活性成分至患者,可以按自身已知的方式制备,如借由常规的混合、溶解、粒化、包裹糖衣、研磨、乳化、包囊化、包埋(entrapping)或冻干方法。

[0230] 用于本发明的药物组合可以按常规方式使用一种或多种药理学的或生理学可接受的载体制备,包含赋形剂、及任选的辅剂,其有利于将活性化合物加工成制备物,所述制备物可以在药物学上使用。适当的配制物与选择的施用途径有关。因此,对于注射,活性成分可以配制于水溶液中,优选配制于生理学相容的缓冲液。对于跨粘膜施用,在配制物中采用对于待穿透的屏障为合适的渗透剂。这些渗透剂本领域公知。对于口服施用,活性成分可以与适合包含于片剂、丸剂、糖衣剂、胶囊、液体、凝胶、糖浆剂、膏剂、悬液等的载体组

合。对于通过吸入施用,活性成分方便地以来自加压的包装或喷雾器的气溶胶喷雾的形式递送,使用适合的推进剂。活性成分可以配制用于通过注射肠道外使用,如通过推注或持续输注。这样的组合物可以采用置于油性或水性媒介物中的悬液、溶液、乳状液的形式,并可以包含配方剂(formulatory agent)如悬浮剂、稳定剂和/或分散剂。其他药学赋形剂为本领域公知。

[0231] 本发明的组合物可以是可注射的悬液、溶液、喷雾剂、冻干粉剂、糖浆剂、酏剂等。可以使用任何适当形式的组合物。为了制备这样的组合物,具有所需的纯度的本发明的核酸或载体,与一种或多种药用可接受的载体和/或赋形剂混合。载体和赋形剂必须是可接受的,意思是与组合物中其他成分相容。可接受的载体、赋形剂或稳定剂在施用的剂量和浓度下对受体无毒,并包括但不限于水、盐水、磷酸盐缓冲盐水、右旋糖、甘油、乙醇或这些的组合,缓冲液诸如磷酸盐、柠檬酸盐和其他有机酸;抗氧化剂包括抗坏血酸和甲硫氨酸;防腐剂(诸如十八基二甲基苄基氯化铵、氯化六甲双铵、苯扎氯铵、苄索氯铵、苯酚丁基或苯甲基乙醇、烷基对羟苯甲酸如甲基或对羟基苯甲酸丙酯、邻苯二酚、间苯二酚、环己醇、3-戊醇和间甲酚);低分子量多肽(少于大约10个残基);蛋白质,诸如血清白蛋白、明胶、或免疫球蛋白类、亲水聚合物如聚乙烯吡咯烷酮;氨基酸诸如甘氨酸、谷氨酰胺、天冬酰胺、组氨酸、精氨酸或赖氨酸;单糖类、二糖类和其他碳水化合物包括葡萄糖、甘露糖或糊精;螯合剂如EDTA;糖类如蔗糖、甘露醇、海藻糖或山梨醇;形成盐的抗衡离子如钠;金属复合物(如Zn-蛋白质复合物);和/或非离子型表面活性剂如TWEEN™、PLURONICS™或聚乙二醇(PEG)。

[0232] 本发明中免疫原性或免疫学组合物,如DNA疫苗,也可以被制成水包油乳剂。举例来说,水包油乳剂基于液态石蜡(European Pharmacopea type)、类异戊二烯油,如角鲨烷、角鲨烯、EICOSANEP™或四十四烷,获得自烃聚合的油,如异丁烯或癸烯;包含线性烷基的酸或醇的酯,如植物油、油酸乙酯、丙二醇二(辛酸盐/癸酸盐)、甘油基三(辛酸盐/癸酸盐)或丙二醇二油酸酯;分支脂肪酸或醇的酯,如异硬脂酸酯。油有利地与乳化剂组合以形成乳状液。乳化剂可以是非离子型表面活性剂,如山梨聚糖、二缩甘露醇(如脱水甘露醇油酸)、甘油、聚合甘油、丙二醇和油酸、异硬脂酸、蓖麻油酸或羟硬脂酸的酯,其任选地乙氧基化,以及聚羟丙基-聚氧乙烯共聚物嵌段,例如PLURONIC®产物如L121。佐剂可以是乳化剂、微团形成剂和油的混合物,如商业上可得的产品,名称为PROVAX®(IDEC Pharmaceuticals, San Diego, CA)。

[0233] 本发明的免疫原性组合物包含额外的物质,例如润湿剂或乳化剂、缓冲剂、或佐剂,用于增强疫苗的效果(Remington's Pharmaceutical Sciences, 18th edition, Mack Publishing Company, (ed.) 1980)。

[0234] 术语“佐剂”包含疫苗佐剂。疫苗佐剂指可以赋予针对抗原的免疫应答和/或调节其倾向所需的免疫应答的组分。参见The European Medicines Agency(EMA) Evaluation of Medicines for Human Use, Guideline on Adjuvants in Vaccines, (2005), page 6。合适的佐剂的实例包括矿物盐(例如氢氧化铝和磷酸钙或磷酸铝凝胶)、油乳乳状液和基于表面活性剂的配制物,如MF59(微液滴化的去污剂稳定化的水包油乳状液)、QS21(纯化的皂苷)、AS02[SBAS2](水包油+MPL+QS-21)、Montanide ISA-51和ISA-720(稳定化的水包油乳状液),微粒佐剂例如病毒体(并入流感病毒血凝素单层的脂质体媒介物)、AS04([SBAS4]

加入 MPL 的 A1 盐)、ISCOMS(结构化的皂苷和脂质复合物)、多乳酸化合物共-乙交酯(PLG)、微生物衍生物(天然的和合成的)如单磷酸类脂 A(MPL)、Detox(MPL+M. Phlei 细胞壁支架)、AGP[RC-529](合成的乙酰化单糖)、DC_Chol(可自组织成脂质体的脂质免疫刺激物)、OM-174(脂质 A 衍生物)、CpG 基序(包含免疫刺激性 CpG 基序的合成寡核苷酸)、改良的 LT 和 CT(遗传学修饰的细菌毒素,提供无毒佐剂功效)、内源的人类免疫调节剂如 hGM-CSF 或 hIL-12(细胞因子,可以以蛋白质或编码质粒的形式施用)、Immudaptin(C3d 串联阵列)和插入载体如金颗粒。同上。

[0235] 增强疫苗效果的佐剂也可以加入所述配制物中。除上面所述外,佐剂包括但不限于矿物盐(如 $\text{AlK}(\text{SOB}_{4\text{B}})\text{B}_{2\text{B}}$ 、 $\text{AlNa}(\text{SOB}_{4\text{B}})\text{B}_{2\text{B}}$ 、 $\text{AlNH}(\text{SOB}_{4\text{B}})\text{B}_{2\text{B}}$ 、二氧化硅、明矾、 $\text{Al}(\text{OH})\text{B}_{3\text{B}}$ 、 $\text{CaB}_{3\text{B}}(\text{POB}_{4\text{B}})\text{B}_2$ 、高岭土或碳)、含有或不含免疫刺激复合物的聚核苷酸类(ISCOM)(如 CpG 寡核苷酸,如 Chuang, T. H. et al, (2002) J. Leuk. Biol. 71(3):538-44, Ahmad-Nejad, P. et al(2002) Eur. J. Immunol. 32(7):1958-68 所述;poly IC 或 polyAU 酸、含有或不含 CpG 的聚精氨酸(也为本领域公知为 IC31;参见 Schellack, C. 等人(2003) Proceedings of the 34th Annual Meeting of the German Society of Immunology; Lingnau, K. et al(2002) Vaccine 20(29-30):3498-508), JuvaVaxTM(U. S. Patent No. 6,693,086), 某种天然物质(如来自结核分枝杆菌(*Mycobacterium tuberculosis*)的 D 蜡、在 *Corynebacterium parvum*、百日咳博德特菌(*Bordetella pertussis*)或布鲁杆菌属成员中发现的物质)、鞭毛蛋白(Toll 样受体 5 配体,参见 McSorley, S. J. et al(2002) J. Immunol. 169(7):3914-9), 皂苷类如 QS21、QS17 和 QS7(U. S. Patent Nos. 5,057,540;5,650,398;6,524,584;6,645,495), 单磷酸脂质 A, 特别地 3'-脱-O-酰化单磷酸脂质 A(3D-MPL), 咪喹莫特(也为本领域公知,称为 IQM 以 AldaraTM 名商业上可得;U. S. Patent Nos. 4,689,338;5,238,944; Zuber, A. K. et al(2004) 22(13-14):1791-8) 和 CCR5 抑制剂 CMPD167(见 Veazey, R. S. et al(2003) J. Exp. Med. 198:1551-1562)。

[0236] 氢氧化铝或磷酸铝通常以磷酸缓冲盐溶液中的 0.05-0.1% 溶液使用。也可以使用其他佐剂,特别是 DNA 疫苗中将霍乱毒素用作佐剂,特别是 CTA1-DD/ISCOM(参见 Mowat, A. M. et al(2001) J. Immunol. 167(6):3398-405)、聚偶磷氮(Allcock, H. R. (1998) App. Organometallic Chem. 12(10-11):659-666;Payne, L. G. et al(1995) Pharm. Biotechnol. 6:473-93)、细胞因子例如但不限于 IL-2、IL-4、GM-CSF、IL-12、IGF-1、IFN- α 、IFN- β 和 IFN- γ (Boyer 等人, (2002) J. Liposome Res. 121:137-142; WO01/095919)、免疫调节的蛋白质例如 CD40L(ADX40;参见,例如 WO03/063899)、天然杀伤细胞 CD1 配体(也公知为 CRONY 或 α -半乳糖神经酰胺;参见 Green, T. D. et al, (2003) J. Virol. 77(3):2046-2055)、免疫刺激融合蛋白例如 IL-2 融合免疫球蛋白 Fc 片段(Barouch 等人, Science 290:486-492, 2000) 和共刺激因子 B7.1 和 B7.2 分子(Boyer)。所有这些都可以以蛋白质或 DNA 的形式施用,可以在同一编码本发明中抗原的表达载体,或在单独的表达载体。

[0237] 水包油乳状液特别适合于病毒载体,可以基于液态石蜡(欧洲药典型),类异戊二烯油例如角鲨烷、角鲨烯,从烯烃类例如异丁烯或癸烯聚合获得的油、具有直链的烷基的酸或醇的酯,例如植物油、油酸乙酯、丙二醇、二(辛酸盐/癸酸盐)、甘油基三(辛酸盐/癸酸盐)或丙二醇二油酸酯,或分支的脂肪族醇或酸的酯,特别是异硬脂酸酯。油与乳化

剂组合使用形成乳状液。乳化剂可以是非离子型表面活性剂,如一方面是山梨聚糖、二缩甘露醇(如脱水甘露醇油酸)、甘油、聚合甘油或丙二醇的酯,另一方面是油酸、异硬脂酸、蓖麻油酸或羟硬脂酸的酯,所述酯任选地乙氧基化,聚羟丙基-聚氧乙烯共聚物嵌段,例如 Pluronic,如 L121。

[0238] 对于马来酸酐-烯基衍生共聚物,可以使用 EMA (Monsanto),其是直链或与乙烯-马来酸酐共聚物交联,且它们例如通过二乙烯醚交联,也参考 J.Fields 等人, Nature 186 :778-780, Jun. 4, 1960。关于结构,聚丙烯酸或聚甲基丙烯酸聚合物和 EMA 优选地通过基本单元形成,基本单元具有下式,其中:R1 和 R2,可以相同或不同,代表 H 或 CH₃, x=0 或 1,优选 x=1, y=1 或 2, x+y=2。对于 EMA, x=0 和 y=2。对于卡波姆, x=y=1。这些聚合物可溶于水或生理盐溶液(20g/1 NaCl)且 pH 可以调整至 7.3 到 7.4,例如通过碳酸钠(NaOH)调节,以提供可以掺入表达载体的佐剂溶液。

[0239] 佐剂的进一步的例子是选自丙烯酸或甲基丙烯酸的聚合物或马来酸酐和烯基衍生物的共聚物的化合物。有益的佐剂化合物为丙烯酸或甲基丙烯酸聚合物,其是交联的,尤其具有糖或多元醇的聚烯基酯。这些化合物被称作卡波姆 (Phameuropa Vol. 8, No. 2, June 1996)。本领域技术人员也可以参照 U. S. Patent No. 2, 909, 462(通过引用并入本文),其详细描述这样的与多羟基化合物交联的丙烯酸聚合物,所述多羟基化合物具有至少 3 个羟基,优选不多于 8 个羟基,至少有三个羟基上的氢原子被含有至少两个碳原子的不饱和脂肪基取代。优选的基团包含 2 至 4 个碳原子,例如乙烯基、烯丙基和其他乙烯基不饱和基团。不饱和基团自身可以含有其他取代基,如甲基。以名称 CARBOPOL® (BF Goodrich, Ohio, USA) 出售的产品是最适合的。它们与烯丙基蔗糖或烯丙基季戊四醇交联。其中,可以提及 CARBOPOL® 974P、934P 和 971P。马来酸酐和烯基衍生物的共聚物之中,优选共聚物 EMA® (Monsanto),其是马来酸酐和乙烯的共聚物,线性的或交联的,例如与二乙烯醚交联。可以参考 J.Fields 等人, Nature, 186 :778-780, 4 June 1960, 通过引用并入本文。

[0240] 如本文所用的术语“脂质体”包括近晶中间相,可以包含磷脂近晶中间相,或非磷脂近晶中间相。见,例如,文献 Small, D. M., in "The Physical Chemistry of Lipids, From Alkanes to Phospholipids" 脂质研究手册, Vol. 4, Plenum, NY, 1986, pp. 49-50 中的“近晶中间相”,其中阐述了“加热给定的分子而不是熔化使之直接成为等向性的流体,经过称为中间相或液晶的居间态,具有在某些方向上残基有序而另一些方向上残基序缺乏的特征。一般而言,液晶分子稍微长于其宽度并且在分子长度的某处具有极性或芳香族部分。分子的形状和两极或芳香族的相互作用允许分子排列为局部有序排列。这些结构特异的出现在那些在一端具有极性基团的分子中。液晶的分子长轴方向上长程有序被称为近晶的或分层的或薄片状的液晶。在近晶状态下,分子可能是单层或双层排布,分子层的平面可能是正常的或倾斜的,具有冷冻的或熔化的芳香族链”。也参见 Small 的图 3-4。

[0241] 有利地,依据本发明的免疫原性组合物和疫苗包含有效量的本文所述的一或多种表达载体和/或多肽以引起免疫应答和/或保护性免疫应答;并且,有效量可以由本文公开的内容,包括本文引用的文献,和本领域的知识确定,无需过度的实验。该免疫原性组合物可以设计为将抗原、核酸或表达载体引入所需的作用部位,并以适当的可控的速度释放。制备控释配制物的方法本领域公知。例如控释制备物可以通过使用聚合物复合或吸收所述免疫原和/或免疫原性组合物而产生。控释配制物可以使用适当的大分子(例如,多

酯类、多氨基酸类、聚乙烯、吡咯烷酮、乙烯醋酸乙烯酯、甲基纤维素、羧甲基纤维素或硫酸鱼精蛋白) 制备, 这些大分子可以用来提供所需的控制释放特征或分布。另一种通过控释制备物控制作用持续时间的可能方法是将活性成分掺入至聚合物材料的颗粒中, 所述聚合物材料如聚酯类、聚氨基酸类、水凝胶、聚乳酸、聚乙醇酸、这些酸的共聚物、或乙烯乙酸酯共聚物。或者, 除了使活性成分并入聚合物颗粒, 可能将这些物质包埋进例如通过凝聚法或界面聚合法制备的微胶囊中, 例如在胶体药物递送系统(例如, 脂质体、白蛋白微球、微乳液、纳米微粒和纳米胶囊) 或在大乳状液系统中分别使用羟甲基纤维素或明胶微胶囊和聚-(甲基异丁烯酸) 微胶囊。这些技术在以下文献中公开: *New Trends and Developments in Vaccines*, Voller 等人 (eds.), University Park Press, Baltimore, Md., 1978 and *Remington's Pharmaceutical Sciences*, 16th edition.

[0242] 施用

[0243] 本领域技术人员可以容易地确定本发明的免疫原性组合物中的本发明的抗原、核酸和表达载体(统称免疫原)的适宜剂量。例如, 免疫原的剂量可以依据给药途径和对象大小不同发生变化。本领域技术人员可以确定适宜的剂量, 例如通过测量对象(如实验用动物)的免疫应答, 使用常规免疫学技术, 并且如果合适则调整剂量。此类测量对象的免疫应答的技术包含但不限于铬释放测定、四聚体结合测定、IFN- γ 酶联免疫斑点测定、IL-2 酶联免疫斑点测定、胞内细胞因子测定和其他免疫学检测方法如 Ed Harlow 和 David Lane 所写的“*Antibodies: A Laboratory Manual*”中所详细阐述的技术。

[0244] 在本发明中, 对 HIV 疫苗的细胞应答的测定方法包括细胞内染色(如流式细胞术)和 ELISPOT(酶联免疫斑点测定), 这些方法可以检测对抗原产生应答而产生细胞因子的细胞, 并对其计数。例如, 从动物或人类病人中分离脾细胞或外周血单核细胞(PBMC), 随后采用 HIV 表位如 2F5 或 4E10 体外攻击, 最后利用 ELISPOT 和 / 或胞内细胞染色(ICS)测定, 最终能确定在疫苗接受者中细胞介导的免疫应答潜能。使用四聚体的流式细胞术(如, 由 4 个拷贝的与其同源肽结合的 I 类分子和碱性磷酸酶组成的分子)可以对抗原特异性 T 细胞计数(如检测识别特定肽的 T 细胞, 该特定肽与主要组织相容性复合物(MHC) I 分子结合)。可采用标准铬释放测定评估细胞毒性。为了评估细胞介导的对 DNA 疫苗的免疫应答, 也可采用更常规方法, 包括测量对抗原响应的 T 细胞增殖、表达 HIV 表位的自体细胞的 CTL-介导的杀灭。

[0245] ELISA 测定和蛋白质印迹可用于评估体液免疫应答。特别地, ELISA 和蛋白质印迹可用于评估抗体的结合、抗体的中和能力、抗体介导的融合抑制和抗体依赖的细胞毒性。

[0246] 可以进行 MT-2 测定测量中和抗体应答。抗体介导的中和采用如先前所述的 MT-2 细胞-杀灭测定的方法测量(Montefiori 等人, 1988, *J. Clin. Microbiol.*, 26:231-237)。血清对合胞体的形成的已知显示血清中存在中和抗体(由疫苗接种诱导)活性。简要说来, 将疫苗接种测试和对照血清可以暴露于病毒感染的细胞(如 MT-2T 细胞系)。可以通过活细胞染色测量中和反应(如, 采用 Finter's 中性红, 当对照孔中细胞病变的效应为大约 > 70% 但少于 100% 时)。通过计算 A_{540} 条件下样本孔(细胞 + 病毒)和细胞对照孔(只含细胞)及病毒对照孔(只含病毒)之间吸光度的差异来检测保护的百分比。中和滴度然后表达为保护至少 50% 的细胞免于被病毒诱导杀死所需的血浆稀释度的倒数。

[0247] 当预防性提供时, 本发明的免疫原性组合物理论上可以施用于下述对象: 在 HIV

感染前、表现出 HIV 感染、由于 AIDS 造成的任何症状出现前,特别是高风险对象。免疫原性组合物的预防性施用可以给对象提供针对 HIV 感染的保护性免疫,或阻止或减弱已感染 HIV 的对象的 AIDS 发病进程。当治疗性提供时,所述免疫原性组合物适用于改善或治疗 AIDS 综合症,并且在感染后越早施用效果越好,优选地在任何 AIDS 症状出现前使用最佳,但是在症状出现时和出现后使用亦可。

[0248] 人用核酸组合物的适宜剂量可以是范围从 $1\mu\text{g/kg}$ 至 1mg/kg 的总核酸,例如 $5\mu\text{g/kg}$ – 500mg/kg 的总 DNA、 $10\mu\text{g/kg}$ – $250\mu\text{g/kg}$ 的总 DNA 或 $10\mu\text{g/kg}$ – $170\mu\text{g/kg}$ 的总 DNA。在某些实施方案中,人类对象(18–50 岁,45–75kg)施用 1.2mg–7.2mg 的 DNA。可以多次施用 DNA 疫苗,如 2–6 次,如 3 次。在一具体的实施方案中,人类对象在 0、4 和 12 周时分别施用 $100\mu\text{gDNA}$ 组合物(每次施用 $100\mu\text{g}$)。

[0249] 蛋白质组合物免疫原性量的范围的实例是 $5\mu\text{g/kg}$ – $500\mu\text{g/kg}$,例如含佐剂的 10 – $100\mu\text{g/kg}$ 的总蛋白质。在一实施方案中,给人类对象(18–55 岁,45–75kg)施用的蛋白质组合物剂量为 $325\mu\text{g}$ 。

[0250] 可以采用任何适当的递送方法施用免疫原组合物,包括但不限于肌肉注射、静脉注射、真皮内、粘膜或局部递送。这些技术本领域技术人员公知。递送方式中更具体的例子是肌肉注射、皮内注射和皮下注射。尽管如此,递送并不局限于注射方法。此外,还可以通过阳离子脂质体(Watanabe 等人, (1994) Mol. Reprod. Dev. 38 :268–274 ;and W096/20013)向动物组织中递送 DNA,或是直接注射裸露 DNA 至动物肌肉组织(Robinson 等人, (1993) Vaccine11 :957–960 ;Hoffman 等人, (1994) Vaccine12 :1529–1533 ;Xiang 等人, (1994) Virology199 :132–140 ;Webster 等人, (1994) Vaccine12 :1495–1498 ;Davis 等人, (1994) Vaccine12 :1503– 1509 ;Davis 等人, (1993) Hum. Mol. Gen. 2 :1847–1851),或使用基因枪技术皮内注射 DNA(Johnston 等人, (1994) Meth. Cell Biol. 43 :353–365)。将 DNA 递送至动物组织的方法还包括电穿孔、快速注射、声致穿孔、显微操作针介导的递送等等。递送路径可选择使用口服、鼻内摄取或任何其他适当的途径。递送也通过粘膜表面完成,如肛门、阴道或口腔粘膜。

[0251] 动物(包括人)的免疫程序表(或方案)是众所周知的,并且针对具体对象和免疫原性组合物可以很容易确定。因此,可以对同一个对象一次或多次免疫施用所述免疫原。在某些实施方案中,每次免疫原组合物施用之间有设定的时间间隔。尽管这个时间间隔因对象不同有很大的变化,但是基本上范围是 10 天至几个星期,通常为 2、4、6、或 8 周。对人类来说,此间隔一般从 2 至 6 周一直到 6 个月或更多。对于 DNA 穿刺(tattooing),此间隔一般只有 3 天(如 0、3 和 6 天时施用)。免疫方案基本上包含 1 到 6 次免疫原组合物的施用,但是可能含少至 1 或 2 或 4 次施用。诱导免疫应答的方法也包含和免疫原一起施用佐剂。在一些实例中,初次免疫方案补充一年或两年或其他长间隔(5–10 年一次)的加强免疫。

[0252] 本方法中也包括了多种初始–加强方案。在这些方法中,一次或多次初始免疫后接着进行一次或多次加强免疫。对每次施用可以采用同样的或不同的实际免疫原性组合物,并且免疫原性组合物的类型(如包含蛋白质或表达载体)、途径和免疫原配制物也可以有多种变化。例如,如果将表达载体应用于初始和加强步骤中,可以是相同的或不同的类型(如 DNA 或细菌或病毒表达载体)。

[0253] 本发明的免疫原组合物可以被单独施用,或与其他免疫原和 / 或免疫原性组合物共同施用或依次连续施用,如和其他免疫的、抗原或疫苗或治疗用组合物共同施用或依次连续施用,从而提供了本发明的多价的或“鸡尾酒式的”或组合的组合物及使用它们的方法。例如,在某些实施方案中,本发明的 HIV 包膜蛋白在病毒载体中施用,例如 MVA,其中还包含编码一个或多个 HIV 蛋白质的基因,例如 gag 和 pol。另一方面,施用的成分和方式,包括剂量都需要考虑以下各种因素而确定,包括特定对象的年龄、性别、体重、种类和病症以及施用途径。

[0254] 在某些实施方案中,本发明的免疫原性组合物施用至哺乳动物。进一步详细方案中,哺乳动物可以是人类、非人灵长类、犬、兔、豚鼠或小鼠。

[0255] 本领域普通技术人员可以很容易确定每种蛋白质亚基和 / 或 DNA 疫苗的合适剂量。一般来说,某些因素可以影响施用剂量,尽管如此,在施用 DNA 疫苗的具体方案中,适合的剂量就是,外源基因被表达且在哺乳动物特定细胞中产生基因产物。优选地,所述剂量足以对动物具有治疗和 / 或预防效果。

[0256] 组合治疗

[0257] 感染 HIV 的对象使用本发明的组合物治疗的方法包括组合治疗,在组合治疗中也执行其他 HIV 治疗方法。例如,进行 HIV 包膜蛋白亚单位疫苗注射的对象,可以依照本发明单独注射或混合注射抗反转录病毒的药物,例如核苷反转录酶抑制剂、非核苷反转录酶抑制剂和 HIV 蛋白酶抑制剂的多种组合。

[0258] 核苷反转录酶抑制剂包括:例如齐多夫定 (AZT)、去羟肌苷 (ddI)、扎昔他宾 (ddC)、司他夫定 (d4T)、拉米夫定 (3TC)、阿巴卡韦 (1592U89)、阿德福韦二匹伏酯 [bis (POM)-PMEA]、洛布卡韦 (BMS-180194) 和洛德腺苷 (FddA)、9-(2,3-双脱氧-2-氟代-β-D-苏型-戊呋喃)腺嘌呤。

[0259] 非核苷反转录酶抑制剂包括:奈韦拉平 (BI-RG-587)、地拉韦啉 (BHAP, U-90152) 和依法韦仑 (DMP-266)。

[0260] 蛋白酶抑制剂包括:沙奎那韦 (Ro31-8959)、利托纳韦 (ABT-538)、英地纳韦 (MK-639)、奈夫那韦 (AG-1343) (商业名称 VIRACEPT™,从 Agouron Pharmaceuticals, Inc 可得)、氨普那韦 (141W94) (非肽类蛋白酶抑制剂,商业名称 AGENERASE™) 和拉西那韦 (BMS-234475)。

[0261] 试剂盒

[0262] 本发明的组合物及其使用方法理想地适于制备试剂盒。容器内提供了 HIV 包膜蛋白的核酸和 / 或蛋白质,可能以任何形式存在,如冻干状态的或溶解状态 (如溶解于蒸馏水或缓冲溶液中) 的等等。在本发明的试剂盒中,基本上包括了一组说明。

[0263] 本试剂盒包含一种或多种其他元件,包括:其他试剂 (如稀释剂)、制备用于施用的组合物的装置或其它物质、可接受药用载体、给对象施用 的装置或其它物质。使用说明如本文所述包括治疗应用说明,包含建议用量和 / 或施用方式,如施用至人类的方式。

[0264] 试剂盒还含有至少一种额外的试剂,例如诊断剂或治疗剂,比如监控组合物在对象内的免疫应答的诊断剂或如本文所述的额外添加的治疗剂。

[0265] 在一种实施方案中,试剂盒包括了一个小瓶 (或其他适合的容器),内装本发明的一种或多种重组的 HIV 包膜蛋白。在某些实施方案中,试剂盒进一步包含佐剂和赋形剂。佐

剂和赋形剂和蛋白质一起配制,且可以包含在配制物中或单独地包装在试剂盒中。

[0266] 通过以下几个非限制性的实施例进一步描述本发明。

[0267] 实施例 1

[0268] Ba-L gp140DC4E10

[0269] 在 Genart 人工合成 HIV-1Ba-L gp160 基因,去掉了信号肽并进行了密码子优化。gp120/gp41 切割位点已被改变以阻止蛋白酶对其水解。Arg501 和 Arg509 被变为丝氨酸。gp140DC4E10 序列经 PCR 扩增,并插入进表达载体 pJWIREs,与 tPa 信号读码框融合。氨基端起始于 E(30),羧基端终止于...WLWYIK(681)(SEQ ID NO:45),并额外添加了 KKK 提高溶解性。pJWIREs 包含下述内容:

[0270] 插入基因的表达由 CMV 启动子驱动,由牛生长激素 (BGH)Poly A 终止。嘌呤霉素乙酰基转移酶基因与插入基因通过 IRES 序列相连,已获得嘌呤霉素抗性。

[0271] 构建的载体被称为 pJW Ba-L gp140DC4E10Puro(图 1)。

[0272] 使用阳离子脂质体 2000 对 HEK293 细胞进行转染。使用 4E10 和 2F5 人 MPER 抗体 (Polymun Scientific GmbH,Klosterneuburg,奥地利)对条件培养基进行免疫共沉淀蛋白质印迹。结果如图 2 所示,与 gp140DC4E10 反应。

[0273] B 进化枝 Ba-L gp140DC4E10 核酸和蛋白质序列参见图 3-4。

[0274] 图 5-6 展示了 B 进化枝另一个修饰后序列 Ba-L gp145 的核酸和蛋白质序列。

[0275] 实施例 2

[0276] D 进化枝 gp140 方法

[0277] 细胞系开发和分子克隆

[0278] 开发了 CHO 细胞系,可稳定表达外源 3 种 D 进化枝 HIV-1 分离物的 gp160(gp140)蛋白质结构域。建立该细胞系的目的是分泌高水平的 gp140 蛋白质,从而可以被纯化用于制作 HIV-1 的疫苗。选作用于表达 gp140 的分离物是 A07412(亲本序列,GenBank 登记号 AF484477 参见 GenBank 登记号 AY736828)、57128(亲本序列,GenBank 登记号 AF484502 参见 GenBank 登记号 AY736829)和 57140(亲本序列,GenBank 登记号 AF484511)。为了在本系统中最大化表达,gp140 密码子在 GENEART 人工合成。通过这个过程,即设计的基因使用人类细胞中最为丰富的 tRNAs 的相关密码子,以使得基因在人类细胞内表达密码子最优化。人类密码子最优化是任何应用于人类的 DNA 疫苗的理想选择,但也是非常有效的在 CHO 中获得更高表达产量的方法。除密码子最优化外,人工合成的基因也在设计中去除了多种顺式作用元件和其他使 mRNA 不稳定的基序(例如 GC-富含区、TATA 框、Qui 位点)。去除的那些顺式作用元件可以降低转录和翻译效率(例如剪接位点、poly A 位点、腺嘌呤富含元件、Rev 作用元件(RRE)和其他 mRNA 二级结构)。

[0279] 对每一分离物,采用基于 PCR 的过程,对 gp140 的初级和二级 gp120/gp41 切割位点进行突变。这样阻止了 gp120/gp41 位点处切割,可以产生稳定的 gp140 分泌蛋白质。此外,移除了每一个天然信号肽,这样才能将高效信号肽组织型纤溶酶原激活物(tPA)信号构建到表达载体中作为分泌信号。每个 gp140 序列在跨膜(TM)结构域前插入了终止密码子,阻止 gp140 在分泌过程中结合于细胞膜上。gp140 基因连接到哺乳动物表达载体 pJWTCDE-N 的限制性酶切位点 NheI 和 EcoRI 处,使其可以在 CHO 细胞中稳定表达。

[0280] A07412

[0281] 氨基酸的氨基端序列为 SL(30)WVT(SEQ ID NO:46), 羧基端序列为 ...FSITK(673)-终止子 (SEQ ID NO:47)。氨基端丝氨酸是 tPa 信号末端 NheI 克隆位点处的外来残基。gp120/gp41 切割位点被从 RAKRRVVEREKR(507) (SEQ ID NO:48) 改变成 RAKSRVVEREKS (SEQ ID NO:49), 参见图 51, SEQ ID NO:3。

[0282] 57140

[0283] 氨基酸的氨基端序列为 SL(33)WVT.. (SEQ ID NO:46), 羧基端序列为 ...FSISN(673)-终止子 (SEQ ID NO:50)。氨基端丝氨酸是 tPa 信号末端 NheI 克隆位点处的外来残基。gp120/gp41 切割位点被从 RAKRRVVEREKR(507) (SEQ ID NO:48) 改变成 RAKSRVVEREKS (SEQ ID NO:49), 参见图 51, SEQ ID NO:4。

[0284] 57128

[0285] 氨基酸的氨基端序列为 SL(33)WVT.. (SEQ ID NO:46), 羧基端序列为 ...FSITK(671)-终止子 (SEQ ID NO:47)。氨基端丝氨酸是 tPa 信号末端 NheI 克隆位点处的外来残基。gp120/gp41 切割位点被从 KARRRVVEREKR(507) (SEQ ID NO:51) 改变成 KARSRVVEREKS (SEQ ID NO:52), 参见图 51, SEQ ID NO:5。

[0286] pJWTCDE-N 载体包含下述元件, 使外源基因在 CHO 细胞中高效表达:

[0287] (1) gp140 基因转录由 CMV 启动子/内含子 A 驱动, 牛生长激素 (BGH) Poly A 终止。

[0288] (2) 卡那霉素激酶 II (NPT II) 基因, 由 SV40 启动子驱动, 合成的 poly A 终止, 在硫酸 G418 选择下挑选出稳定转化的细胞。

[0289] (3) 二氢叶酸还原酶 (DHFR) 基因, 由特定缺陷的 SV40 启动子和 SV40poly A 控制, 在转化细胞中非常弱的表达 DHFR。这使得在无核苷酸培养基上的筛选更易进行, 同时可以用 DHFR 抑制子甲氨蝶呤 (MTX) 处理诱导外源基因的扩增。基因扩增可以很大程度的提高 gp140 的产量, 同时提高在含 MTX 培养基上生存所需的 DHFR 的产量。

[0290] 每一个表达载体克隆后, 需要进行插入 gp140 基因的测序, 确保载体构建的正确性。

[0291] 建立稳定分泌 gp140 的 CHO 细胞系, 缺乏 DHFR 的 CHO 细胞

[0292] 使用脂质体 2000 转染试剂 (Gibco) 转化 (CHO-dhfr-)。转化细胞采用 gp120 抗原捕获实验进行分析, 利用与 HIV-1(+) 人血清放射免疫沉淀反应 (RIP) 检测 gp140 产物是否存在及含量。将转化细胞置于 96 孔板中, 用于在 10% 透析胎牛血清和 550 μ g/ml G418 Sulfate 中进行 alpha MEM 筛选。活细胞采用 gp120 抗原捕获筛选, 挑选可以相当大量生产 gp140 的细胞用于扩大培养和单细胞克隆。

[0293] gp120 抗原捕获实验是基于 ELISA 的实验, 用于 gp120 的检测和定量。gp140、后续将要介绍的 gp140DC4E10 和 gp145 蛋白质分子, 均含有 gp120 序列, 也可以用本方法检测。在 96 孔板上的微量滴定, 涂上两种鼠类单克隆抗体, 与 HIV-1gp120 上的特异性表位反应。将 gp120 标准溶液或组织培养检测样品加入样品槽, 与板结合的抗体与溶液中的 gp120 形成了免疫复合物。然后充分洗脱未结合的物质。然后加入含有过氧化物酶结合的人抗-gp120 多克隆抗体的结合溶液。结合抗体复合物捕获 gp120 上的其他表位。洗脱未结合的结合溶液, 加入过氧化物酶底物。酶与底物反应导致出现底物变蓝的结果。紧接着加入终止溶液 (2N 硫酸), 蓝色变为黄色, 可以通过在 450nm 下测量吸光度定量测量。在 gp120 标准和检测样品中 gp120 的含量与吸光度正相关。在检测样品中的 gp120 浓度可基于标准

曲线进行计算。

[0294] 使用 gp120 抗原捕获和 RIP 对各克隆进行比较,以找到一个新的 gp120 强生产者。其中最好的用 $0.02 \mu\text{M}$ MTX 处理,以促进外源 DNA 的扩增。一旦细胞可以在 MTX 中以正常的速率生长,可以对它们进行克隆和再次筛选产量更高的生产者。经有限稀释克隆得到细胞系,分析适合在无蛋白培养基 HyQPFCHO Liquid Soy (HyClone ;Logan, Utah) 中生长和表达的最适条件。

[0295] 蛋白质纯化

[0296] 通过离心从 CHO 培养物中收集条件培养基,并浓缩至大约 2L,浓缩的方法是切向流过 100kDa 分子量筛截。培养基中含有磷酸盐缓冲液,pH 调节至 7.2。调节氯化钠浓度至 300mM,培养基用 0.22 微米的滤筛过滤。培养基通过 GNL 琼脂糖柱,用含 500mM 甲基 α -D-吡喃甘露糖苷的 PBS 溶液洗脱包膜蛋白。培养基再过 3-4 次 GNL 琼脂糖柱,从条件培养基中消除所有的包膜蛋白。而后进行下一步过程以进一步纯化包膜蛋白。调整 GNL 琼脂糖洗脱物的氯化钠浓度至 212mM 并通过含有 Q- 琼脂糖 (Amersham Biosciences ;Piscataway,NJ) 的吸附柱。高分子量的杂志与 Q- 琼脂糖结合,而包膜蛋白在此条件下却不结合。为了使任何经空气氧化形成的不正常的多聚体断裂,用包膜蛋白处理的 Q 琼脂糖浓缩至大约 3ml,并用 50mM DTT 在 4°C 条件下处理 15 小时,而后在 21°C 处理 1 小时。而后将 DTT 处理过的制备物通过 Superdex20026/60 (Amersham Biosciences) 凝胶过滤柱移除另外的高或低分子量的杂志,同时也减少包膜蛋白分解产物量。含 1mM DTT 的 PBS 以 $0.5\text{ml}/\text{min}$ 的速度通过吸附柱。包含纯净包膜蛋白的片段上样进行 SDS-PAGE 分析。然后蛋白质在用 PBS 平衡的 10ml PD-10 吸附柱 (Amersham Biosciences) 上更换缓冲液。最后,蛋白质用 $0.22 \mu\text{m}$ 过滤器过滤,等分并保存于 -70°C 。

[0297] 实施例 3

[0298] 前言

[0299] 构建和检测了表达 4 种 C 亚型分离物包膜蛋白的 DNA 质粒载体,从中挑选表达 gp145 蛋白质的最好的候选,这些包膜蛋白 DNA 是从急性感染和早期血清抗体转化的感染病人身上分离得到。挑选出 C06980v0c22 分离物,得到表达 C06980v0c22gp145 蛋白质的稳定细胞系,并生产出用于研究的细胞库。生产纯化的 gp145 蛋白质,并用于临床免疫学研究。

[0300] 如上所描述的,申请人共同进行研制开发了 D 亚型 HIV-1 的亚单位疫苗。提供了 4 种 D 亚型 HIV-1 分离物的序列,制备了几个 gp140 和 gp120- 分泌 CHO 细胞系。细胞系适合于无血清培养基培养,纯化的包膜蛋白可用于临床免疫学研究。对小动物进行免疫,采用 ELISA 和中和抗体滴度检测了抗同源原始分离物的 gp140 和 gp120 特异性的血清抗体结合滴度,同源原始分离物采用假病毒测定检测。尽管所有的 gp120 和 gp140 蛋白质均具有免疫原性,但是对抗同源假病毒分离物的中和抗体没有被检测到。

[0301] 采用 C 亚型包膜蛋白序列从事 HIV 包膜蛋白亚单位疫苗的研制,目的是诱发更有效的宽广的中和抗体反应。如上所讨论的,构建了编码 HIV-1Ba-L (B 亚型) 包膜蛋白修饰的 DNA 载体。该载体编码了一段截短的糖蛋白 gp160,被命名为 gp145。gp145 蛋白质包含了修饰过的组织型纤溶酶原 激活物 (t-Pa) 信号肽,该信号肽位于缺乏切割位点的 gp160 的上游,且 gp160 在细胞膜近段外侧区域 (MPER) 末端截短。在 C 端包含了 3 个额外的赖氨

酸残基,理论上可以提高 C 端尾部的亲水性,向免疫系统暴露出潜在的 MPER 中和表位。与上面阐述的 gp140 不同, gp145 分子在 ELISA 和蛋白质印迹实验中对抗 -MPER huMAb4E10 起中和反应。

[0302] C 亚型是已知国际上最普遍的亚型,因为初步资料显示出 C 亚型可以在天然感染中诱导最广泛的与 HIV-1 的中和交叉响应,对 C 亚型的序列进行了研究。计划构建一种 gp145 载体,使用原始的 CCR5- 依赖的 C 亚型包膜蛋白序列。开发出了表达包膜蛋白的稳定 CHO 细胞系。

[0303] 包膜蛋白选择 (downselection)

[0304] 提供了来自 4 种 C 亚型毒株中的 env 序列,用于密码子优化和合成。进行瞬时表达研究挑选可用于后续 gp145 生产的分离物。

[0305] 提供了 4 种南非 C 亚型 R5HIV-1 包膜蛋白序列的电子拷贝,这些序列取自 3 个急性感染 HIV-1 的个体 (C06838v1c48, C06980v0c22andC3728v2c6) 和一个早期血清转化感染 HIV-1 (C06980v1c3) 的个体。这些序列中的两个来自同一个个体,一个来自急性感染期 (C06980v0c22),另一个来自血清抗体转换后 (C06980v1c3)。

[0306] 为了最大化的在中国仓鼠卵巢细胞 (CHO) 中表达,包膜蛋白基因合成时通过 Genearth AG (Regensburg, Germany) 结合了中国仓鼠 (*C. griseus*) 的密码子偏爱。为了进一步最大化表达,去除了那些会降低翻译效率的顺式作用元件 (如内部的 TATA 盒、chi- 位点、核糖体进入位点、RNA 二级结构、重复序列等)。每种包膜蛋白基因合成了两个版本:a) gp160,全长 gp160 去掉天然信号肽;b) gp160DC,全长 gp160 去掉天然信号肽在 gp120/gp41 初级和二级切割位点处突变以阻止蛋白酶降解。

[0307] 4 种 gp160 (野生型) 和 gp160DC (切割位点突变型) 基因的表达对比见图 35。在表达蛋白序列中蛋白序列连接在 t-Pa 信号肽切割位点后。阴影区域为可变区。方框标出的区域标示出了 gp120/gp41 切割位点:在 gp160DC 基因中精氨酸至苏氨酸点突变阻止溶蛋白性裂解。gp160 基因克隆进入哺乳动物专利表达载体 pSWTIPK3 (Advanced BioSciences Laboratories, Inc.) 的 NheI 和 EcoRI 位点处,与 tPa 信号肽读码框融合 (图 8)。天然前导序列被 tPa 信号肽取代,这提供了更有效的分泌信号,增强了 gp160 的生产和运输至细胞膜。表达质粒包含了巨细胞病毒 (CMV) 启动子来控制表达。质粒转化至大肠杆菌 (*Escherichia coli*) 中进行扩增和纯化, gp160 编码区测序以验证其序列相同性。

[0308] 转染中国仓鼠卵巢细胞 (CHO-K1) 和人胚肾细胞 (HEK293 ;克隆 293H) 以进行 gp160 生产,并用蛋白质印迹和抗原捕获 ELISA 进行分析。基于 gp160 分子的生产数量和质量确定采用哪个单独分离物建立 CHO 细胞系用于生产 gp145 蛋白质。用 4 种 gp160 和 gp160DC 质粒载体采用脂转染的办法转染细胞 (Lipofectamine2000 ;Invitrogen, Carlsbad, CA)。培养物用 HIV-1Ba-L gp145、HIV-1Ba-L gp160 或 HIV-1C 亚型 gp140 表达载体转染,作为正对照。天然未转染的 CHO-K1 和 HEK293 作为负对照。培养基和细胞裂解物在转染后 48 小时收集以用作分析。

[0309] 转染的 CHO-K1 中的培养基和细胞裂解物样品利用免疫共沉淀蛋白质印迹的方法对 gp41/gp120/gp160 含量进行了评估,使用人类 HIV-1 感染阳性血清进行免疫沉淀,使用免疫了家兔血清的 HIV-1Ba-L gp160 进行检测。任何载体的培养基中都没有检测到包膜蛋白表达,除了 Ba-L gp145 对照 (数据未显示)。如果处理 gp160 使之成为 gp41 和 gp120,

gp120 会流出进入培养基,脱落的 gp120 的量在含量测定检出限以下。在细胞裂解物中,包膜蛋白表达在每一个 C 亚型 gp160 和 gp160DC 载体中都存在。每种载体生产 gp160 蛋白质,每一个分离物生产的都具有不同的大小,可能是由于不同糖基化图谱造成的。只有 pSWC06980v0c22gp160 载体显示出 gp160 被加工成 gp120 和 gp41。

[0310] 由于在 CHO-K1 中的表达水平非常低,进一步对每种载体定量研究时使用 HEK293 进行转染和分析。从转染的 HEK293 细胞裂解物样品中利用免疫共沉淀蛋白质印迹测定 gp41/gp160 含量,使用 huMAb 至 gp41 (4E10) 进行免疫沉淀,兔血清免疫的 HIV-1Ba-L gp160 用于检测。包膜蛋白很强的表达在每一个 C 亚型 gp160 和 gp160DC 载体中。每种载体生产 gp160 蛋白质,每一个分离群生产的都具有不同的大小,可能是由于不同糖基化图谱造成的。再一次,只有 pSWC06980v0c22gp160 载体显示出 gp160 被加工成 gp120 和 gp41。

[0311] 为了定量 gp120/gp160 产物的水平,从转染的 CHO-K1 和 HEK293 中 取得培养基和细胞裂解物,利用 HIV-1gp120 抗原捕获实验进行分析(表 1)。在转染的 CHO-K1 中,在每一个载体中检测到包膜蛋白表达,除了 C06838v1c48gp160DC。与 Ba-L gp145 对照对比,所有的表达水平都很低。这是由于非常低的表达水平和很难从这些分离物中检测包膜蛋白造成的。由于转染 CHO 细胞效率很低造成 Ba-L gp145 的表达水平也很低。生产 C 亚型最多的载体是 C06980v0c22 和 C3728v2c6gp160s。值得注意的是浓度值基于从不同分离物中得到的 C 亚型 gp120 的相关反应性。确切的浓度可能会与报道浓度有差别,这是由于实验所用抗体在每一个分离物中的亲和性可能存在差异。

[0312] 表 1 转化了 HIV-1C 亚型 gp160 和 gp160DC 的表达质粒的 CHO-K1 细胞 HIV-1gp120 抗原捕获测定。转染后 48 小时培养基和细胞裂解物检测到大量的 gp120 和 gp160 蛋白质。

[0313]

构建体	gp120/gp160 (ng/ml)	
	培养基	细胞裂解物
pSWC06838v1c48 gp160	0.6	0.0
pSWC06980v1c3 gp160	1.0	0.7
pSWC06980v0c22 gp160	3.2	5.7
pSWC3728v2c6 gp160	7.2	6.7
pSWC06838v1c48 gp160DC	0.0	0.0
pSWC06980v1c3 gp160DC	0.4	0.7
pSWC06980v0c22 gp160DC	1.6	1.2
pSWC3728v2c6 gp160DC	2.5	3.4
(-) 对照	0.0	0.0
(+) 对照 : Ba-L gp145	25.0	3.0

[0314] 由于在 CHO-K1 中检测到的包膜蛋白非常弱,采用 HEK293 转染分析来检验结果。在 HEK293 中,包膜蛋白的表达采用抗原捕获(利用表 2 中所示载体)的方法被检测到。在 HEK293 中,06838v1c48gp160DC 现在被检测到,但是信号很弱。表达水平明显高于 CHO-K1 中。生产 C 亚型最多的载体是 C06980v0c22 和 C3728v2c6gp160s。

[0315] 表 2 转化了 HIV-1C 亚型 gp160 和 gp160DC 的表达质粒的 HEK293 细胞 HIV-1gp120 抗原捕获测定。转染后 48 小时培养基和细胞裂解物检测到大量的 gp120 和 gp160 蛋白质。

[0316]

构建体	细胞类型	gp120/gp160 (ng/ml)	
		培养基	细胞裂解物
pSWC06838v1c 48 gp160	293H	0.0	0.9
pSWC06980v1c 3 gp160	293H	28.2	11.6
pSWC06980v0c 22 gp160	293H	94.1	94.6
pSWC3728v2c6 gp160	293H	198.2	87.7
pSWC06838v1c 48 gp160DC	293H	0.0	0.8
pSWC06980v1c 3 gp160DC	293H	2.7	8.8
pSWC06980v0c 22 gp160DC	293H	13.9	37.7
pSWC3728v2c6 gp160DC	293H	59.3	92.1
(-) 对照	293H	0.0	0.0
(+) 对照 : Ba-L gp145	293H	237.0	61.0
(+) 对照 : Ba-L gp160	293H	79.0	62.0
(-) 对照	CHO-K1	0.0	0.0
(+) 对照 : Ba-L gp145	CHO-K1	41.0	3.0
(+) 对照 : Ba-L gp160	CHO-K1	21.0	29.0

[0317] gp160 生产的进一步测定采用两种糖蛋白 gp160 抗原捕获实验来进行, 该实验中应用了人类单克隆抗体 4E10 (表 3) 和 2F5 (表 4)。每个分离物的 gp160 与基于 4E10 的含量测定产生很强的反应。甚至分离物 C06838v1c48 反应也很强, 这暗示出 gp120 含量测定对本分离物给出了人工低的结果。每个分离物的 gp160 与基于 2F5 的含量测定也产生反应, 尽管比基于 4E10 的含量测定反应弱。C06980v1c22 分离物反应是最强的。2F5 反应较弱是由于合成抗体的 B 亚型分离物的 2F5 表位与 C 亚型分离物显著不同。不受已知理论限制, 申请人认为 2F5 的反应活性可能出自与 gp160 结合的脂类而非氨基酸骨架的相互作用。这些含量测定表明在每一个载体中 MPER 都被展示出来。

[0318] 表 3 转化了 HIV-1C 亚型 gp160 和 gp160DC 的表达质粒的 HEK293 细胞 HIV-1gp160 抗原捕获测定。4E10huMAb 连接 gp41MPER 被用于捕获抗体。转染后 48 小时培养基和细胞裂解物检测到大量的 gp160 蛋白质。

[0319]

构建体	OD 450 nm	
	培养基	细胞裂解物
pSWC06838v1c48 gp160	0.146	2.655
pSWC06980v1c3 gp160	0.926	2.718
pSWC06980v0c22 gp160	2.534	2.736
pSWC3728v2c6 gp160	0.589	2.779
pSWC06838v1c48 gp160DC	0.162	2.703
pSWC06980v1c3 gp160DC	1.105	2.588
pSWC06980v0c22 gp160DC	2.608	2.616
pSWC3728v2c6 gp160DC	0.681	2.652
(-) 对照	0.123	0.156
(+) 对照 : Ba-L gp145	2.417	2.553

[0320] 表 4 转化了 HIV-1C 亚型 gp160 和 gp160DC 的表达质粒的 HEK293 细胞 HIV-1gp160 抗原捕获测定。2F5huMAb 连接 gp41MPER 被用于捕获抗体。转染后 48 小时培养基和细胞裂

解物检测到大量的 gp160 蛋白质。

[0321]

构建体	OD 450 nm	
	培养基	细胞裂解物
pSWC06838v1c48 gp 160	0.146	0.530
pSWC06980v1c3 gp160	0.200	0.399
pSWC06980v0c22 gp 160	0.344	1.458
pSWC3728v2c6 gp160	0.198	0.781
pSWC06838v1c48 gp 160DC	0.155	0.506
pSWC06980v1c3 gp160DC	0.179	0.446
pSWC06980v0c22 gp 160DC	0.298	1.210
pSWC3728v2c6 gp160DC	0.219	0.829
(-) 对照	0.135	0.150
(+) 对照 : Ba-L gp145	2.268	2.638

[0322] 申请人推断分离物 C06980v0c22 被用于建立 CHO 细胞系,用来生产 gp145。确定采用本分离物基于以下几个因素:

[0323] ●与 C06838v1c48 和 C06980v1c2 分离物相比,具有非常高的表达

[0324] ●采用此分离物将 gp160 加工为 gp120 和 gp41 的效果最显著。

[0325] ●该毒株可以在急性感染的早期分离得到。

[0326] ●与 MPER 抗体 4E10 具有强反应活性。

[0327] ●该病毒是有感染能力的。

[0328] CHO-K1/C06980v0c22gp145 细胞系的形成

[0329] 构建 C06980v0c22gp145DNA 表达质粒并用于建立稳定转染的 CHO-K1 细胞,用于生产 gp145。这些细胞系适合于在无蛋白培养基中生长,同时建立细胞库。挑选出 H-73-9-2-8 克隆用于生产 gp145 蛋白质。

[0330] C06980v0c22gp145DNA 载体通过基于 PCR 的技术修饰 gp160DC 基因而构建出。gp145 由 N30 残基组成,位于天然信号肽切割位点的直接下游,穿过刚好位于跨膜区前的 K676。接着 K676 后是三个额外添加的赖氨酸,gp145 终止于此(图 36 和图 37)。这些包括理论上提高了 C 末端的亲水性,因此提高了免疫系统中 MPER 的暴露性。在信号肽切割位点之上,预测有一个来源于修饰过的 t-Pa 信号的丝氨酸,这位于氨基末端。gp145 基因被连接在哺乳动物表达载体 pSWTIPK3 和被命名为 pSWC06980v0c22gp145 的 NheI and EcoRI 限制性内切酶位点。该质粒在大肠杆菌 DH5 α 菌株 (Invitrogen) 中扩增,使用 Endofree Plasmid Maxi 试剂盒进行纯化 (Qiagen, Valencia, CA)。采用限制酶切消化分析质粒, gp145 序列的编码区通过 DNA 双向测序来确认。质粒包含嘌呤霉素乙酰基转移酶基因,用于使用嘌呤霉素筛选稳定转化菌落。受到内部核糖体进入位点 (IRES) 和 CMV 启动子的引导。通过将 gp145 表达与嘌呤霉素抗性标记结合,推进了 gp145 基因表达水平升高。

[0331] 哺乳动物表达载体重要特征摘要如下所述:

[0332] ●载体包含抗生素抗性基因,可被用于构建载体过程中细菌内的选择标记。从这些载体上衍生的治疗产物在其构建过程中应避免使用青霉素或相关抗生素。因此采用卡那霉素而非氨苄青霉素。

[0333] ● gp145 基因表达密码子最优化,从而增强产物的表达。使用 CHO 密码子偏爱合成基因,使用那些在 CHO 细胞中最为丰富的 tRNAs 相关的密码子。合成的基因中设计消除一切顺式作用元件,除了可以稳定 mRNA 的其他基序,从而降低转录翻译效率。

[0334] ● gp145 基因引入载体符合修饰的 t-Pa 信号肽读码框,从而可以有效的运至细胞膜。

[0335] ● 基因在强启动子和有效的多聚腺核苷酸 poly-A 的控制下表达。采用强大的 CMV 启动子和高效的 Bovine Growth Hormone (BGH) poly-A。

[0336] ● 为了筛选稳定蛋白表达的细胞克隆,该载体包含一个选择标记。引起嘌呤霉素抗性的嘌呤霉素乙酰基转移酶基因受到内部核糖体进入位点 (IRES) 和 CMV 启动子的引导

[0337] 在建立细胞系的过程中,由于临床环境使用这些细胞系的需要,需要小心进行工作,记录档案,使用 FDA 可接受的物质。CHO-K1 (cat#CCL 61) 细胞从 ATCC (Manassas, VA) 得到。为建立稳定 CHO-K1 细胞系,最优化的转化方法为电穿孔。该方法的益处是避免了非特征化的动物衍生元件。此外,动物衍生产物除非必要需要避免。为了降低 BSE 污染的机会重组胰蛋白酶取代猪胰蛋白酶,胎儿牛血清 (FBS) 明确定义其来源为新西兰。FBS 受辐射、热灭活并无菌过滤。

[0338] 超螺旋和线性的 pSWC06980v0c22gp145 (线性载体具有单酶切位点 NruI) 分别独立电穿孔转化 DNACHO-K1 细胞。两种形式的 DNA 均被采用,因为两种形式在建立细胞系时具有各自的缺点和优势。超螺旋 DNA 转化一般效率较高,这点是有益的,因为 CHO-K1 细胞转化很难。线性 DNA 转化效率较低,但是结合进入宿主基因组的效率比超螺旋 DNA 高。简要说,5x10⁶ CHO-K1 细胞在 0.5ml 电穿孔转化缓冲液中悬浮 (BioRad, Hercules, CA), 100 μ l 电穿孔转化缓冲液与 100 μ g 质粒 DNA 在 0.4cm 电极比色杯中混合,采用 Gene Pulser 脉冲设备 (BioRad) 在 350V 电压下 125 μ FD 对细胞进行脉冲,冰上放 30 分钟,混合并培养于 5.5ml F-12K 完全培养基 (Invitrogen, Inc.) 中,其中包含了 10% 热灭活的 FBS (Hyclone Laboratories, Logan UT), 10 μ g/ml 庆大霉素 (Invitrogen, Inc.) 和 2mM 谷氨酰胺 (Quality Biologicals, Inc., Columbia, MD)。

[0339] 转化后 48 小时,取得条件培养基和细胞裂解物样品,进行 gp120 抗原捕获测定分析和免疫共沉淀蛋白质印迹。gp120 抗原捕获测定结果证实 gp145 被分泌至条件培养基。产量很低,线性和超螺旋 DNA 分别只有 4 和 9ng/ml。与预期相同,超螺旋 DNA 的产量比线性 DNA 的产量高。从培养基和细胞裂解物中取得的包膜蛋白样品 IP 蛋白印迹分析 (图 26)。与预期相同,超螺旋和线性 DNA 电转化细胞的条件培养基中均明显存在 gp145 蛋白,约 140KDa。一种稍微低分子量的 gp145 种类与预期相同,在细胞裂解物中出现。这可能不完全代表处理的 gp145。

[0340] 电穿孔转化 48 小时后的细胞在 96 孔板中放置了大约 4500 个细胞,以用于筛选 cF12-K。24 小时后,将细胞放入含 10 μ g/ml 嘌呤霉素的 cF12K 中进行筛选 (Sigma-Aldrich, St. Louis, MO), 一周更换两次培养基,直到嘌呤霉素抗性菌落达到大约 50% 融合。应用 gp120 抗原捕获实验分析条件培养基中 gp145 产物。具有最多产物的 20 种细胞系培养物扩大培养、冷冻保存和克隆,通过有限稀释分离真正稳定表达 gp145 的克隆。非细胞系培养物最佳产物量达到 2 μ g/ml。

[0341] 从最初的 20 种培养物中经初始选择和克隆,可以得到 > 100 个经 gp120 抗原捕获

实验分析的克隆。基于 gp145 的产量水平,确定了代表初始 12 中培养物的 17 个细胞系作为生产 gp145 的潜在候选。每一种细胞系均被冷冻储存。对每一种挑选出的克隆进行产物水平研究。简要来讲,细胞被播种于 24 孔板中,在 1×10^5 个细胞使用 1ml 组织培养基,并于 37°C 孵化培养 64 小时。收集培养基并对其进行 gp120 抗原捕获实验分析及免疫共沉淀蛋白质印迹。基于抗原捕获和免疫共沉淀蛋白质印迹结果,克隆 H-73-9、H-84-1 和 H-94-10 被选出适合进行无蛋白培养基培养生产蛋白质。每种冷冻保存了 5 个细胞库。每种克隆生产 1 至 2mg gp145/L。在免疫共沉淀蛋白质印迹中,每种选出的克隆具有很强的 gp145 反应活性,与预期一致,大约在 140kDa 左右。

[0342] 观察发现,CHO-K1 细胞电穿孔转化超螺旋 DNA 比线性 DNA 获得更高产量的 gp145,产量分别为 9ng/ml 对 4ng/ml。尽管如此,经嘌呤霉素筛选,超螺旋和线性 DNA 均获得大约 100 个稳定细胞系。有意思的是,17 个最佳 gp145 生产细胞系中有 16 个都是从线性 DNA 中衍生的。这一点支持了申请人的预期,即超螺旋 DNA 可以更有效的被细胞所吸收,但是线性 DNA 可以更有效的整合从而在稳定细胞系中得到更高的蛋白质产量。

[0343] 三种选出的细胞系均适合在无蛋白培养基中生长。由于对于某些克隆而言,适应在无蛋白培养基中生长具有难度,选择 3 个可以增加获得一个适应强的克隆的可能。此外,每种克隆在 3 中不同培养基上 (PowerCHO-1CD、PowerCHO-2CD 和 PowerCHO-3CD, Lonza, Walkersville, MD), 培养,每种均包含 $5 \mu\text{g/ml}$ 嘌呤霉素和 4mM 谷氨酰胺。在无蛋白培养基中培养几个阶段后,生长于 PowerCHO-1CD、PowerCHO-2CD 和 PowerCHO-3CD 上的克隆 H-73-9 适应良好,并分别被命名为 H-73-9-1、H-73-9-2 和 H-73-9-3。生长于 PowerCHO-2CD 上的克隆 H-84-1 也适应良好,并被命名为 H-84-1-2。每种冷冻保存两管细胞库。

[0344] 以上 4 种适应培养物经有限稀释再次克隆,并对每种培养物经 gp120 抗原捕获测定鉴定出其中最好的生产克隆。在这些培养物中,有两个被鉴定为最好的生产者: H-73-9-2-8 和 H-73-9-3-9。每种冷冻保存两管细胞库,细胞 500ml 扩大培养以进行小规模蛋白纯化。收集条件培养基,采用 20mM Tris, pH8, 0.5% Triton-X-100 和 500mM 氯化钠缓冲液。缓冲培养基通过 2ml 柱 Galanthus nivalis lectin (GNL) 琼脂糖 (Vector Laboratories, Inc., Burlingame, CA)。该柱采用加入了 Tris pH8, 0.5% Triton-X-100 和 500mM 氯化钠的缓冲液洗脱,而后用 PBS 洗脱。结合的 gp145 经 400mM 甲基 α -D-吡喃甘露糖苷洗脱。纯化的 gp145 蛋白质应用 SDS-PAGE 和蛋白印迹进行分析 (图 27 和图 28)。两种克隆均生产大约 145kDa 的蛋白质,该蛋白质在蛋白质印迹中很好的反应。在非还原性条件下,也可以得到一些二聚体和更高阶的多聚体。两种细胞系依照 Comassie Plus Protein Assay (Pierce) 分析,产量 $> 1.2\text{mg/L}$ 。克隆 H-73-9-2-8 具有更好的生长性质 (更健康生长更快),因此被选择生产 gp145 蛋白。

[0345] 制备了 10 管 H-73-9-2-8 (批次 4/17/08) 研究细胞库 (RCB),并在液氮中冷冻储存。2X106 细胞冷冻在 10 管 1ml 无蛋白冷冻培养基中。该培养基包含: 7.5% 二甲亚砜 DMSO (Sigma-Aldrich)、50% 新鲜生长培养基、42.5% Profreeze CDM (Lonza)。采用试剂盒 Qiamap Blood Mini Kit (Qiagen) 从 5×10^6 细胞中分离基因组 DNA,利用 PCR 扩增 gp145 基因整合区,并双向测序鉴定。gp145 基因编码区具有 100% 序列匹配。细胞库收集两周后,对细胞进行支原体污染检测 (使用 MycoAlert Mycoplasma detection Kit (Lonza) 试剂盒),并发现结果为阴性。将一管解冻置于培养物中检测其活性。培养 3 天后,77% 细胞是

有活力的,并在 gp120 抗原捕获测定中被检测可以生产 gp145 蛋白。培养物进行细菌污染检测,接种了 SOC 培养基后 37℃培养 24 小时显示无细菌生长。

[0346] H-73-9-2-8RCB 批次 4/17/08 培养物生长,用于制作 25 管 RCB(批次 F1144) 并被液氮冷冻保存。10X106 细胞冷冻在 25 管 1ml 无蛋白冷冻培养基中。该培养基包含 :7.5% 二甲亚砜DMSO(Sigma-Aldrich)、50%新鲜生长培养基、42.5%Proffeeze CDM(Lonza)。建立细胞库时,培养物进行支原体污染检测并发现结果为阴性。培养物进行细菌和真菌污染检测,结果显示在巯基乙酸盐液体培养基中无细菌生长,或大豆酪蛋白消化培养基中无真菌生长。从细胞库中取出一管解冻并在如前所述培养基中培养,检测其活性以正式 gp145 在这些细胞中生产。经过溶解,具有 87%活性,可接受。生长特征与预期相同,且 gp145 生产经抗原捕获测定证实。

[0347] C06980v0c22gp145 蛋白质生产

[0348] 生产了 3 批 gp145 蛋白质,并传递以用于后续研究(表 5)。

[0349] 表 5 提供的 HIV-1. C06980v0c22gp145 批次

[0350]

批号	浓度 (mg/ml)	总量 (mg)
112009	1.1	7
120710A	1.0	18
120710B	0.975	5

[0351] 批次 112009

[0352] H-73-9-2-8 培养物扩大培养至 3L,使用摇瓶培养,采用 PowerCHO-2CD 培养基添加 4mM 谷氨酰胺和 5 μ g/ml 嘌呤霉素。条件培养基经离心净化。3L 收集物中取得培养基样品采用抗原捕获进行分析。结果预示了大体总计产生大致每升培养基 4mg gp145。使用 3L 收集物生产的产量和质量经鉴定可以接受。gp145 蛋白质采用如下所描述的方法纯化,并在图 29 中概述出。

[0353] 收获的培养基在室温下采用 0.1m2Pellicon 过滤部件过滤浓缩,分子量被截断至 30kDa。该系统先后应用 1.0M NaOH、WFI 水和 1X PBS 缓冲液冲洗。引入 CHO 细胞培养悬浮物 (3L),并采用在循环模式操作系统。渗透流速 1L 每 hr/0.1m2 及横向流速 0.5L/min 进行浓缩。浓缩结束后,样品被浓缩至 200mL。浓缩的细胞培养悬浮物于 -70℃保存至后续处理。

[0354] 使用凝集素亲和层析Lectin Affinity Chromatography纯化 gp145 蛋白质。调整浓缩培养基使之适于 500mM 氯化钠,而后通过 GNL 凝胶的 25ml 柱。柱使用 PBS 冲洗,采用 500mM 甲基 α -D- 吡喃甘露糖苷洗脱 gp145 蛋白质。10ml/min 的流速下进行纯化,100mL 洗脱物采用正切流动过滤进行浓缩和透析过滤至 PBS 中。在此步骤中,采用 50cm2Pellicon 过滤部件过滤浓缩,分子量被截断至 30kDa 该系统先后应用 1.0M NaOH、WFI 水和 PBS 缓冲液冲洗。而后引入洗脱物,并在循环模式下操作系统。进行环戊氯噻嗪过滤 :渗透流速 2.0mL/min 及横向流速约 40ml/min。GNL-Eluate 的容量被缩减至溶于 8mLPBS。

[0355] 依据 Bradford 含量测定进行蛋白质含量评估,蛋白质含量为 1.1mg/mL。纯化的 gp145 中的内毒素采用色度法 LAL 含量测定 (Lonza),其含量为每毫克蛋白质 31.8EU。在还

原和非还原性条件下进行 SDS-PAGE 分析显示纯化蛋白质分子量约为 145kD。

[0356] 由 SE-HPLC 分析可以看出 (图 30), 批次 112009gp145 以涉及 A、B、C 和 D 四种复合物的形式被洗脱。基于蛋白质链的流动性, 计算出每一种 gp145 的表观分子量 (表 6)。主峰值计算出在 > 669kDa (估计大约为 895kDa) 处, 与 A 形式符合。主峰双肩明显位于 > 669kDa (估计大约为 680kDa) 和 571kDa 处, 分别与 B 和 C 形式符合。第 4 种但不是主峰位于 417kDa 处, 符合 D 形式的大小。

[0357] 表 6 标准和纯化的 gp145 蛋白质 (批次 112009) 多聚体种类的分子量和保留时间。

[0358]

蛋白质	保留时间(Min)	分子量(kDa)	多聚体形式
甲状球蛋白	11.81	669	无或 A
铁蛋白质	13.74	440	无或 A
过氧化氢酶	16.18	232	无或 A
牛 γ 球蛋白	16.52	150	无或 A
牛血清白蛋白	16.75	66	无或 A
gp145 峰 1	11.15	大约 895	A
gp145 峰 2	11.77	大约 680	B
gp145 峰 3	12.25	571	C
gp145 峰 4	14.64	417	D

[0359] 经 SDS-PAGE 纯化后 gp145 蛋白质激光光密度测定法测出纯度为 96.1%。

[0360] 批次 120710A

[0361] H-73-9-2-8 培养物扩大培养至 11L, 使用摇瓶培养, 采用 PowerCHO-2CD 培养基添加 4mM 谷氨酰胺和 5 μ g/ml 嘌呤霉素。条件培养基经离心净化。11L 收集物中取得培养基样品采用抗原捕获进行分析。结果预示了大体总计产生每升培养基 8mg gp145。使用 11L 收集物生产的产量和质量经鉴定可以接受。gp145 蛋白质采用如下所描述的方法纯化, 并在图 31 中概述出。

[0362] 收获的培养基在室温下采用 0.1m2Pellicon 过滤部件过滤浓缩, 分子量被截断至 30kDa。如批次 112009 中所描述的操作再循环模式。11L 条件培养基被浓缩至 1L。

[0363] 浓缩的条件培养基的缓冲液包含 20mM Tris pH8、500mM 氯化钠和 0.5% Triton-X-100, 然后经 0.22 μ m 过滤纯化。条件培养基于 4°C, 以大约 1ml/min 流速通过 20ml GNL- 凝胶树脂。树脂用 20mM Tris pH8、500mM 氯化钠和 0.5% Triton-X-100 缓冲液冲洗, 而后用 PBS 平衡。gp145 蛋白质在含有 0.5M 甲基 - α -D-吡喃甘露糖苷的 PBS 中洗脱。使用 50kDaMWCO 过滤使 88mLGNL- 洗脱物被浓缩为 20ml。留出 10ml 在准备批次 120710B 中使用。剩余的 10ml 在 PD10 缓冲液中经离子交换树脂进入 PBS 缓冲液。洗脱的物质经 0.22 μ m 滤网过滤消毒, 等分成分并储存于 -70°C。批次 120710A 终产物有 22ml 的量。

[0364] 依据 Bradford 含量测定进行蛋白质含量评估, 蛋白质含量为 1.0mg/mL。纯化的 gp145 中的内毒素采用色度法 LAL 含量测定, 其含量为每毫克蛋白质 < 313EU。在还原和非还原性条件下进行 SDS-PAGE 分析显示纯化蛋白质分子量约为 142kD。在非还原性条件下, 多聚体也是存在的。这代表多聚体结合于二硫键上。

[0365] 由 SE-HPLC 分析可以看出 (图 32), 批次 120710A gp145 以四种与批次 112009 复

合物相同的形式被洗脱。基于蛋白质链的流动性,计算出每一种 gp145 的表观分子量(表 7)。主峰值计算出在 666kDa 处,与 B 形式符合。次峰和一肩分别位于 > 669kDa(估计大约为 845kDa)和 584kDa 处,分别与 A 和 C 形式符合。第 4 种但不是主峰位于 411kDa 处,符合 D 形式的大小。

[0366] 表 7 蛋白质标准和纯化的分子量及保留时间

[0367] HIV-1C06980v0c22gp145 批次 120710A 多聚体种类

[0368]

蛋白质	保留时间 (Min)	分子量(kDa)	多聚体形式
甲状腺蛋白	11.813	669	无或 A
铁蛋白质	13.677	440	无或 A
过氧化氢酶	16.114	232	无或 A
牛 γ 球蛋白	16.400	150	无或 A
牛血清白蛋白	16.737	66	无或 A
gp145 峰 1	11.226	大约 845	A
gp145 峰 2	11.824	666	B
gp145 峰 3	12.217	584	C
gp145 峰 4	14.343	411	D

[0369]

[0370] 经 SDS-PAGE 纯化后 gp145 蛋白质激光光密度测定法测出 HIV-1C_{06980v0c22}gp145 批次 120710A 纯度为 94.2%。

[0371] 批次 120, 710B

[0372] 批次 120710B 与批次 120710A 从同一洗脱的 gp145 蛋白质中制备而来, gp145 蛋白质采用凝集素亲和层析方法洗脱得到。为减少分子间的二硫键的目的,制备批次 120710B 的步骤多出一部。基础理论是基于对先前得到的 gp145(批次 112009)的观察发现该蛋白主要表现为高阶聚合体的形式。无已有的理论支持,申请人相信很多多聚体产生归因于氧化作用出现了分子间二硫键的结果。减少这些二硫键将有助于生产低阶聚合体形式甚至三聚体的蛋白质,

[0373] gp145 蛋白质采用如下所描述的方法纯化,并在图 33 中概述出。

[0374] 已经留出批次 120710A 中描述的 10ml 浓缩的 GNL 洗脱物以便用于准备批次 120710B。这 10ml 采用 50mM DTT 在 37°C 处理 30minutes,然后在 PD10 缓冲液中过离子交换树脂进入 PBS 缓冲液。洗脱的物质经 0.22 μ m 滤网过滤消毒,等分成分并储存于 -70°C。批次 120, 710B 终产物有 22ml 的量。

[0375] 依据 Bradford 含量测定进行蛋白质含量评估,蛋白质含量为 0.975mg/mL。纯化的 gp145 中的内毒素采用色度法 LAL 含量测定,其含量为每毫克蛋白质 < 0.321EU。在还原和非还原性条件下进行 SDS-PAGE 分析显示纯化蛋白质分子量约为 143kD。在非还原性条件下,仅仅是痕量的多聚体也是存在的。这代表多聚体结合于二硫键上。与未用 DTT 处理的批次 120710A 比较,DTT 处理后减少了很多这类化学键。

[0376] 蛋白质印迹显示出在还原性和非还原性条件下主条带在大约 143kDa 处。几种复合受体形式的 gp145 很明显出现于非还原性条件下,但是却极少出现在未用 DTT 处理的批次 120710A 中。在还原性条件下,这些多聚体主要降解至单体。

[0377] 由 SE-HPLC 分析可以看出(图 34), gp145 以与批次 120710A 和 112009 复合物相

同的四种形式被洗脱。基于蛋白质链的流动性,计算出每一种 gp145 的表观分子量(表 8)。主峰值计算出在 665kDa 处,与 B 形式符合。次峰和一肩分别位于 > 669kDa(估计大约为 844kDa) 和 572kDa 处,分别与 A 和 C 形式符合。第 4 种但不是主峰位于 412kDa 处,符合 D 形式的大小。

[0378] 表 8 蛋白质标准和纯化的分子量及保留时间

[0379] HIV-1C06980v0c22gp145 批次 120, 710B 多聚体种类

[0380]

蛋白质	保留时间 (Min)	分子量 (kDa)	多聚体形式
甲状腺球蛋白	11.813	669	无或 A
铁蛋白质	13.677	440	无或 A
过氧化氢酶	16.114	232	无或 A
牛 γ 球蛋白	16.400	150	无或 A
牛血清白蛋白	16.737	66	无或 A
gp145 峰 1	11.23	大约 844	A
gp145 峰 2	11.83	665	B
gp145 峰 3	12.289	572	C
gp145 峰 4	14.327	412	D

[0381] 经 SDS-PAGE 纯化后 gp145 蛋白质激光光密度测定法测出纯度为 94.2%。

[0382] DTT 处理减少了分子间二硫键对 gp145 多聚体的存在形式有某些影响。在 SDS-PAGE 中,很明显与未被 DTT 处理的进行对照,DTT 处理破坏了多数分子间二硫键。SE-HPLC 显示 C 型数量只有小幅的上升,这可能由于 A 形式数量小幅下降。

[0383] 进一步对复合形式的性质进行研究探索结果包含于下述实施例 7 中。

[0384] 实施例 4

[0385] 通过与 gp145 蛋白质的免疫进行中和抗体的诱导, gp145 蛋白质衍生于急性 HIV-1C 进化枝,是由 CHO 表达的重组 gp145 蛋白质。

[0386] 动物:24 只新西兰白色雌兔,1.8-2kg

[0387] 兔被分为 6 组每组 4 只。每组的动物个体都被笼子牌和耳标区分。

[0388]

组别	抗原	佐剂或媒介物
1	gp145 蛋白质	铝胶
2	gp145 蛋白质	包含脂质 A 的脂质体制剂物与 gp145 蛋白质混合完成
3	gp145 蛋白质	脂质体制剂物 1, 包含脂质 A, 包裹 gp145 蛋白质
4	gp145 蛋白质	脂质体制剂物 2, 包含脂质 A 和 PIP, 包裹 gp145 蛋白质
5	无	脂质体制剂物 1, 包含脂质 A
6	mper23, B 进化枝	脂质体制剂物 2, 包含脂质 A 和 PIP, 包裹 mper23 蛋白质

[0389] 抗原 25 μ g/兔/剂

[0390] 如上所示的 gp145 蛋白质在 CHO 细胞中的表达(急性感染组 C, C0698v0c22),培养条件是 PBS, pH7.4。

[0391] mper23 (NK-4) :LELDKWASLWNWFDITNWLWYIK (SEQ ID NO :53), (HBX2 变 量 ; Swiss-Prot 登录号 P04578, 除了 674 位 D 氨基酸被 N 取代)

[0392] 佐剂 :

[0393] 铝胶 (0.6mg Al3+ / 剂) 配方 :0.125ml PBS 中 0.6mg Al3+, pH7.4 ; 与等量的抗原混合

[0394] 脂质体制剂物 1 :DMPC : 胆固醇 : DMPG (9 : 7.5 : 1) ; 50mM 磷脂, 含 100 μ g lipid A / 0.25ml 剂 ; PBS, pH7.4。DMPC 指的是二肉豆蔻磷脂酰胆碱, DMPG 指的是二肉豆蔻磷脂酰甘油。

[0395] 脂质体制剂物 2 :DMPC : 胆固醇 : PIP (1 : 1.5 : 1) ; 50mM 磷脂, 含 100 μ g lipid A / 0.25ml 剂 ; PBS, pH7.4。PIP 指的是磷脂酰肌醇 -4- 磷酸。

[0396] 放血 : 在 -2、0、4、8 和 10 周时使用 20-24 规格蝶形导管于动物耳动脉处取血。从每一个放血处取大约 5ml 血。血液于室温保存 2-3 小时而后冷却至 4℃ 过夜, 而后离心移走血清。血清被等分成 : 1x1ml 和 3x0.5ml 于塑料管中在 -80℃ 冷冻保存。 .

[0397] 在第 12 周时, 采用 60cc 皮下注射器和 18 计量注射针用 atamine/Xylanine 对家兔进行麻醉, 而后通过心脏穿刺最后一次放血。血清被等分成 : 5x1ml 和 5x5ml 于塑料管中在 -80℃ 冷冻保存。

[0398] 免疫法 : 第 0、4 和 8 周时交替进行尾部大腿肌肉注射。采用 23-27 计量注射针注射 0.25ml。

[0399] 时间表 :

[0400]

周数	步骤过程
	收到家兔
-2 周	隔离检疫后释放
-2 周	预放血
0 周	预放血
0 周	肌肉注射免疫 , 0.25 ml
4 周	放血
4 周	肌肉注射免疫 , 0.25 ml
8 周	放血
8 周	肌肉注射免疫 , 0.25 ml
10 周	放血
12 周	最后一次放血

[0401] 参见图 41, 结果展示在图 42-48、52 中。

[0402] 实施例 5

[0403] 下述对小鼠研究的方法结合了在家兔研究中使用的免疫原性、抗原性方法, 如图 41-48, 52 所示, 除了家兔收获了 25ug 的 gp145 蛋白质或所述脂质体内的安慰剂。

[0404] Gp145 小鼠免疫原性研究 : 抗原性、免疫原性方法

[0405] 抗原

[0406] 从急性 C 亚型 Tanzania 感染个体中分离得到的包膜蛋白序列生产 gp145 蛋白质。蛋白的整个保外结构与被保留, 包括 gp41 的 MPER 区。蛋白被设计出含有两个突变点 (R508S, R511S), 在 gp120/gp41 切割位点处突变用以阻止蛋白酶的切割, 在 C 末端多赖氨酸

突变可以有助于 MPER 表位呈现及蛋白质生产。蛋白质在 CHO 细胞中生产,通过凝集素亲和层析纯化,并以一种如实施例 7 所述的多聚物混合物的形式出现。gp145 蛋白质含有下述 MPER 表位序列:ALDSWNNLWNWFDIS(SEQ ID NO:23)。

[0407] 脂质体制备:

[0408] 在免疫前抗原即实验用 M13 噬菌体或 gp145 蛋白质均预先被脂质体包被。脂质体由二肉豆蔻磷脂酰胆碱、二肉豆蔻磷脂酰甘油和胆固醇按摩尔比值 1.8 : 0.2 : 1.5 组成,由冻干的脂质合剂于磷脂浓度 50mM 含 0.4g/ml 脂类 A Dulbecco's PBS 中通过色散制备出来,含有或缺乏抗原的两种脂质体均需制备。脂质体采用无菌盐水冲洗两次以移除未包被的抗原。

[0409] 动物免疫:

[0410] 40 只雌性 BALB/C 小鼠,每只 25g,按照 Institutional Laboratory Animal Care and Use Committee 中所批准的方案进行免疫。动物被分为 8 组,每组 5 只(表 9)。小鼠采用在尾侧大腿肌肉肌肉注射 4 次的方法进行免疫,肌肉注射以 2 至 3 周为周期,每次采用 10 μ g gp145 蛋白质剂量或 5×10^{11} 噬菌体进行。血液每两周采集一次,在首次免疫前两周进行第一次采集,并在动物安乐死时进行最后一次采集。血液在室温下保存 2-3 小时,而后冷却至 4°C 过夜,然后离心。收集血清,并于 -20°C 保存。最后一次注射(第 10 周)两周后小鼠被施行安乐死。从未处理和免疫后的小鼠中取得血液、脾脏、淋巴结、骨髓和肝脏。

[0411]

组 #	免疫原	免疫(周)	放血(周) 150ul/小鼠 / 放血	安乐死(收集血液、脾、 淋巴结、骨髓和肝)
1	M13-12D4	0, 3, 6, 8	-2, 0, 2, 4, 6, 8, 10	10周
2	M13-12B7	0, 3, 6, 8	-2, 0, 2, 4, 6, 8, 10	10周
3	M13-all 5	0, 3, 6, 8	-2, 0, 2, 4, 6, 8, 10	10周
4	gp145/M13-all 5	0, 3, 6, 8	-2, 0, 2, 4, 6, 8, 10	10周
5	gp145	0, 3, 6, 8	-2, 0, 2, 4, 6, 8, 10	10周
6	M13- 无插入	0, 3, 6, 8	-2, 0, 2, 4, 6, 8, 10	10周
7	未处理	0, 3, 6, 8	-2, 0, 2, 4, 6, 8, 10	10周

[0412] 表 9 小鼠免疫计划

[0413] IFN γ - 释放 ELISPOT (酶联免疫斑点测定) 实验

[0414] 脾细胞分泌的 IFN γ 通过 ELISPOT 分析。在无菌 PBS 中加入 10 μ g/ml 抗 -gamma 干扰素 (IFN γ) (PBL Interferon Source) 置于 96 孔硝化纤维素 - 支持的多重筛选 -IP 灭菌板 (微孔) 包膜后于 4°C 过夜。这些孔用含有 0.5% 牛血清蛋白的无菌 PBS 在 37°C 封闭 30 分钟,而后用含有 0.025% Tween20 的洗脱液的 PBS 洗脱,随后用无菌 RPMI-1640 完全培养基洗脱。利用每组的小鼠 (每组 5 只小鼠) 脾脏制备单细胞悬浮液。细胞 (每孔含 2×10^6) 置于抗 -IFN γ - 包被的板上在加入 CO₂ 的容器中 37°C 温育培养 18 小时。在 5 μ g/ml 急性感染 C gp145 (HIV-1C06980, Advanced Bioscience Laboratories)、gp140 (HIV-1IIIIB, Advanced Bioscience Laboratories)、酵母衍生的 gp41 (Meridian Biosciences) 或 10 μ g/ml 组织蛋白酶讲解酵母衍生的 gp41 或无蛋白质中培养细胞。培养板采用洗脱液洗脱,随后用蒸馏水洗脱,用 0.125 μ g/ml 生物素化的抗 -IFN γ (XMG1.2 克隆 ;BD PharMingen) 覆盖并于室温下温育培养 2h。而后培养板洗脱,并采用 1 : 1,000 稀释的抗生素蛋白偶联的碱性磷酸酶在室温下培养 2 小时。培养板洗脱,通过添加 5- 溴 -4- 氯 -3- 吡啶 - 磷酸 (BCIP) / 硝基蓝

四唑 (NBT) (Kirkegaard and Perry Labs) 检测结合的 IFN γ 。培养板用水洗脱,第二天使用立体双目显微镜观察独立斑并计数。描绘出培养板上每个号码的斑点细胞数量平均数。

[0415] 通过流式细胞术对 T 细胞中细胞因子的抗原呈递和检测

[0416] 从不同组的小鼠的脾脏或淋巴结中取得的细胞用 $5 \mu\text{g/ml}$ 急性感染 Cgpl45 (HIV-1C06980, Advanced Bioscience Laboratories)、gp140 (HIV-1IIIIB, Advanced Bioscience Laboratories)、酵母衍生的 gp41 (Meridian Biosciences) 或 $10 \mu\text{g/ml}$ 组织蛋白酶降解的酵母衍生 gp41 或作为正对照的 ConA 刺激,于 37°C 培养 22 小时。细胞在上述提到的抗原中温育 2 小时,而后加入布雷菲德菌素 A (1mg/ml , Sigma-Aldrich) 和莫能菌素 (0.07mg/ml , BD Pharmingen)。细胞继续培养 20 小时。在 LSR II (BD Immunocytometry Systems) 流式细胞仪上分析细胞,使用 FACSDiva 软件 (BD Immunocytometry Systems) 收集了 500,000 个事件。使用多种生存活性标记将死细胞和 B 细胞排除在外。CD3+CD4+ 和 CD3+CD8+T- 细胞门控分析其中 IL-2、TNF- α 、IFN- γ 和 CD107a 的表达。使用 FlowJo 软件 (Tree Star) 分析数据。展示出每组中每个抗原染色的阳性细胞百分比。黑色条纹代表对照组响应 2 倍范围之上, M13 未插入。

[0417] 抗原特异性血清 IgG ELISA

[0418] 抗原特异性 IgG 滴定依赖 ELISA 滴定,使用 gp145 和 gp41 作为靶点。抗原在 PBS (pH7.4) 中被稀释至 $0.25 \mu\text{g/ml}$, 96 孔微量滴定聚苯乙烯板上每孔加入 100ul。培养板在 4°C 温育过夜,然后用 300ul 0.1% PBST (含有 0.1% Tween-20 的 PBS) 洗脱 3 次。血清采用 1 : 50 连续稀释 2 次,而后在血清稀释液 (包含 5% 脱脂牛奶的 0.1% PBST) 中进行血清的滴定。培养板在 37°C 温育培养 1 小时,然后用洗脱液洗脱 3 次。HRP- 标记的抗 - 小鼠免疫球蛋白抗体在血清稀释液中以 1 : 16,000 稀释每孔加入 100ul。培养板在 37°C 温育培养 1 小时,然后用洗脱液洗脱 3 次。加入 TMB (100ul, Kirkegaard and Perry Labs) 37°C 温育培养 30 分钟,加入 100ul 1M 磷酸终止反应。培养板在分光光度计中在 410nm、570nm 基准下过滤读出。抗原结合滴定依赖浓度计算,在此过程中本底结合被检测到超过三次。进行两次独立的测定,取结果的平均值。

[0419] 使用 Biacore 进行表面等离子激元 (SPR) 测量

[0420] 使用 CM5 芯片用 Biacore T200 进行表面等离子激元测量。使用 Biacore 胺偶联试剂盒 (Biacore, AB) 将肽固定于芯片表面。所有免疫步骤在 25°C , $10 \mu\text{l/min}$ 流速下进行,肽加样缓冲液为 20mM 醋酸钠, pH4.2。固定术指导使用 T200 控制软件在各自的独立细胞中固定 14700 个共振单位 (RU) 的 10uM 乱序 MPER 肽和 20500 个 RU 的 MPER 肽。两种肽在免疫固定中均具有 10 分钟接触时间。血清样本在三羟甲基氨基甲烷缓冲盐水中以 1 : 50 稀释,然后以 $30 \mu\text{l/min}$ 通过芯片表面 3 分钟,随后是 5 分钟解离时间。在 5 分钟解离时间的最后,绵羊抗 - 小鼠免疫球蛋白 (Fc) 抗体 (结合位点) 的溶液 $75 \mu\text{g/mL}$ 以 $10 \mu\text{l/min}$ 流速通过流式细胞表面 2 分钟。70 秒解离时间后,使用 50mM HCl 30 秒脉冲、100mM EDTA 加入 20mM Tris, pH7.430 秒脉冲、50% 乙酸 30 秒脉冲重建芯片表面,随后进行 1 分钟注射 Tris- 缓冲盐水, pH7.4。使用 BIAevaluation4.1 软件进行数据分析减去非特异性结合。已报道的免疫球蛋白特异性反应单位比值在以下两种情况下具有差异:即在抗免疫球蛋白注射最后 5 秒窗口取 60 秒平均值与在抗免疫球蛋白注射前 5 秒窗口取 10 秒平均值间具有差异。

[0421] 假病毒中和分析

[0422] TZM-bl 细胞用作测定确定 HIV-1 中和的靶点。BnAb 或血浆分别在 25 μ g/ml 或 1 : 20 连续稀释 4 次的系列稀释液中滴定,采用生长培养基 [DMEM 中含 100U/ml 青霉素、100 μ g/ml 链霉素、2mM L-谷氨酰胺 (Quality Biologics Inc.)、和 15%胎牛血清 (Gemini Bio-Products)] 在黑色平底 96 孔板中加入 25 μ l 双份样品。假病毒在生长培养基中稀释至得到约 150,000 相关发光单位 (RLU),每孔中加入等量上述假病毒。样品在 37°C 加入 5% CO₂ 的容器中温育培养 1 小时,所有的培养都是在这样的条件下进行。TZMbl 细胞在 2×10^5 细胞 /ml 的生长培养基中重悬,该培养基含有 60 μ g/ml DEAE-dextran (Sigma),每孔加入 50 μ l。每板的孔内含有细胞和假病毒 (病毒对照) 或仅含有细胞 (背景对照)。培养板温育培养 48 小时,然后每孔加入 100 μ l 再生 Brite Lite Plus (Perkin Elmer)。使用 Victor2 luminometer (Perkin-Elmer) 测量 RLU 值。测量由于抗体存在出现的抑制百分比,通过比较含抗体孔内 RLU 值与病毒对照孔内的 RLU 值进行该测量。进行两次独立的测定,取结果的平均值。

[0423] PBMC 中和测定

[0424] 从 HIV 未感染者供体中收集并冷藏保存的 PBMC 用作确定 HIV-1 中和的测定靶点。含量测定使用具有复制活性的 HIV-1 感染的分子克隆 (IMC),该克隆包含了表达海紫罗兰属萤光素酶 (LucR)-HIV-1 融合蛋白的报告基因。病毒生产量用分光光度计测定 (Edmonds, TG 等人 Virology 408 : 1-13 (2010))。血清采用 IL-2 生长培养基 1 : 20 连续稀释 4 次,该培养基含有以下成分:[RPMI-1600 含有 100U/ml 青霉素、100 μ g/ml 链霉素、2mM L-谷氨酰胺 (Quality Biologics Inc.)、15%胎牛血清 (Gemini Bio-Products) 和 20U/ml 重组白细胞介素 -2 (Roche Diagnostics)],而后取出 25 μ l 作为副本加入圆形底 96 孔板中。IMC 在生长培养基中稀释至得到约 50,000RLU,每孔中加入等量上述 IMC。样品在 37°C 加入 5% CO₂ 的容器中温育培养 1 小时,所有的培养都是在这样的条件下进行。PHA/IL-2 受激的 PBMC 在 2×10^6 细胞 /ml 的 IL-2 生长培养基中重悬,然后每孔加入 50 μ l。每板的孔内含有细胞和 IMC (病毒对照) 或仅含有细胞 (背景对照)。培养板温育培养 24 小时,每孔中加入 100 μ l 生长培养基,然后再温育培养 72 小时。使用 Renilla Luciferase Assay System (Promega) 定量荧光产生量。每孔加入 Lysis 缓冲液 50 μ l,进行两次冷冻解冻循环,而后转移至黑色平底 96 孔板中,每孔加入 20 μ l,100 μ l 底物注射后立即测量每孔中的 RLU。测量由于抗体存在出现的抑制百分比,通过比较含抗体孔内 RLU 值与病毒对照孔内的 RLU 值进行该测量。进行两次独立的测定,取结果的平均值。

[0425] 结果

[0426] 5 种 M13- 展示的 4E10 表位的免疫原性能够抑制体外定量的中和反应。35 只雌性 BALB/C 小鼠分成 7 组,每组 5 只,接种了单个 M13 展示的表位,其中包含了所有 5 种 M13- 展示的表位或所有 5 种结合了 HIV-1gp145 包膜蛋白的 M13- 展示的表位 (表 9)。从急性感染 HIV-1C 进化枝的个体中取得的 gp145 包膜蛋白引出了家兔的中和抗体。

[0427] 引起的细胞免疫应答分析

[0428] 由接种疫苗引起的脾脏和淋巴结中的细胞免疫应答通过 INF γ - 释放酶 联免疫斑点测定 (ELISPOT) 和胞内细胞因子染色 (ICS) 测定评估。在这些测定中, HIV 特异性反应使用 HIV-1 包膜蛋白刺激后进行了测量,包括 gp145、gp140、gp41 或组织蛋白酶讲解的 gp41。小鼠免疫应答是对照组的两倍,其中小鼠与无插入的 M13 的免疫被认定为正对照 (图

53and54)。ICS 数据分析指出 CD3+CD4⁺ 或 CD3+CD8⁺T⁺ 细胞特异反应。

[0429] 在 INF γ - 释放酶联免疫斑点测定中,在淋巴结和脾脏的所有组中均观察到单一的反应,其中 M13- 全部 5 种免疫群,在淋巴结中受 gp140 刺激, gp145 免疫群在脾脏中受 gp145 刺激。(图 53) 在脾脏 T⁺ 细胞中对 gp140 的背景反应很高。通过 ICS 观察 IL-2 所有组对几种 HIV-1 包膜蛋白抗原的免疫应答, TNF α 、CD107a 和 INF γ 反应没有被检测到。阳性 IL-2 反应在淋巴结中比脾脏中更常发生,分别为 85%和 48%阳性反应,但是分别低于对照组 3.9- 折和 4.5- 折。阳性 IL-2 反应在 CD4⁺T⁺ 细胞中比 CD8⁺T⁺ 细胞中更常发生,分别为 73%和 60%阳性反应,且分别高于对照组 4.5- 折和 3.8- 折(图 54)。小鼠脂质体免疫并不仅仅具有 HIV-1 特异细胞免疫应答。

[0430] 引起的抗体应答分析

[0431] 通过免疫球蛋白结合的 ELISA 及中和含量测定分析体液免疫应答。检测了所有组针对 gp145 和 gp140 的结合滴度(图 55)。动物与 gp145 或 gp145/M13- 产生的全部 5 种抗体与高浓度 gp145 发生免疫应答,平均值分别在 512000 和 409600,而与 gp41 反应的均值分别在 30400 和 43200。两种均具有对 gp145 免疫蛋白质最高的滴度。另一些组并不具有在本含量测定中可检测的结合滴度,M13-a115 除外,它具有很弱的对 gp145 的结合滴度,均值 1200。应用 Biacore 展示在混合血清中表位 - 特异的免疫球蛋白与 MPER 肽的结合,没有检测到二者结合(数据没有展示)。

[0432] 中和含量测定使用 TZMb1 和 PBMC 作为含量测定靶点进行测定。血清与两种中和 - 敏感的 HIV-1 毒株在两种含量测定平台上滴定,测出半感染剂量值(图 56)。在 TZMb1 和 PBMC 中,动物与 gp145/M13-a115 免疫均具有最高的中和滴度,分别比 gp145- 免疫组高出 2.1- 和 1.9- 倍。动物与单个 M13- 展示的 MPER 表位、M13-12D4 和 M13-12B7 或多个的 M13- 展示的 MPER 表位(M13-a115) 免疫,也产生 HIV 中和抗体。所有血清都是由 HIV-2/MPER 嵌合体及非特异性病毒对照 MuLV 中筛选得到,所有这些病毒中均未观察到中和反应。

[0433] 实施例 6

[0434] α 4 β 7 封闭分析

[0435] 材料

[0436] 反应介质 :10% FCS/RPMI/Lglut/PenStrep

[0437] 细胞 :RPMI8866

[0438] 表 10 试剂

[0439]

试剂	克隆#	Mfg	Cat #	批次 #	工作稀释	注释
正常小鼠免疫球蛋白	-	Invitrogen	10400C	645253A	10%	
正常人类免疫球蛋白	-	GenScript	A01006	A108810	10%	溶于 1mg/mL H ₂ O 保存
抗- $\alpha 4$ 封闭 MAb	HP2/1	Beckman Coulter	IM0764	21	2ug/孔	溶于 0.5mg/mL H ₂ O 保存
抗 $\beta 7$ -FITC	FIB504	BioLegend	321212	B142196	1:10	
Neutravidin-PE	-	Invitrogen	A2660	866787	1:400	
gp145- 急性感染 C-生物素-	-	V.Polonis	-	LotB	0.5ug/孔	通过 RF 生物素化 2X
抗-包膜蛋白 mAbs x 14	参见表 12 的测试的 mAb 列表					从 S. Zolla-Pazner 处获得

[0440] 制备 $\alpha 4 \beta 7$ 结合缓冲液

[0441] - 使用 MnCl₂ 粉末新鲜制备 1M MnCl₂ (1g MnCl₂-4H₂O (从 Arthos Lab-MW197.9) + 5mL dH₂O)

[0442] - 制备下面表中的结合缓冲液

[0443] - 过滤灭菌并储存于 4°C

[0444] 表 11

[0445]

试剂	Mfr	Cat #	[最终](mM 或 %)	[初始](M or %)	mL 每 500 mL
HEPES 缓冲液	Gibco	15830	10	1	5
NaCl	Sigma	56546	150	5	15
MnCl ₂	(可商购获得)		1	1	0.5
CaCl ₂	Fluka	21115	0.1	1	0.05
BSA	Sigma	A9576	0.5	30	8.3
NaN ₃	Aldrich	438456	0.09	10	4.5
dH ₂ O	Quality Bio	351029101	-	-	466.6

[0446] 采集细胞

[0447] - 采集非贴壁细胞并转移至 50mL 管中

[0448] - 细胞颗粒在 200xg 离心 10 分钟, 倒出上清液。

[0449] - 将所有细胞结合进 10mL 培养基中, 剧烈混悬以打破小簇并计数。

[0450] - 调整容量至 1mL 培养基中 1.0x10⁶ 细胞。

[0451] - 等分成每孔 100uL (100K) 细胞置于 96 孔板 (96-well U-bottom polypropylene) 中。

[0452] - 细胞颗粒在 200xg 离心 10 分钟

[0453] - 使用 2X 结合缓冲液洗脱细胞。

[0454] 表 12 样品

[0455]

#	样品ID	Spicificity	[IgG] ug/uL	uL for 10ug	uL Bndg Bfr	总 uL
A	1361.100.10.10.1.1	V2	1.52	6.6	33.4	40
B	1393A.100.10.1.1	V2	2.99	3.3	36.7	40
C	1357A.100.10.1	V2	0.84	11.9	28.1	40
D	2158.100.100.10.1.1	V2	5.28	1.9	38.1	40
E	697-30D.10.1.1	V2	6.44	1.6	38.4	40
F	830A.10.10.1.1	V2	0.45	22.2	17.8	40
G	2297.100.1.1	V2	3.01	3.3	36.7	40
H	447-52D.10.1.1	V3	4.36	2.3	37.7	40
I	1006-15D.100.1	V3	2.40	4.2	35.8	40
J	3869.100.10.1.1	V3	5.27	1.9	38.1	40
K	729-30D.10.10.1.1	CD4bs	2.15	4.7	35.3	40
L	1331-160E.100.10.1	CD4bs	1.54	6.5	33.5	40
M	1570D.10.1.1	CD4bs	3.86	2.6	37.4	40
N	1418(16).10.1	b19	4.75	2.1	37.9	40

[0456] 加入免疫球蛋白预培养蛋白质

[0457] - 向 96 孔板的孔中加入结合缓冲液并进行样本计算

[0458] - 依据样品计算的结果向适当的样品槽中加入免疫球蛋白

[0459] - 向 UNTREATED、POS CTRL 和 NO IgG CTRL 样本槽中只加入 40uL 结合缓冲液

[0460] - 将储存的蛋白质加入在结合缓冲液,使之达到 0.025ug/uL。

[0461] 表 13

[0462]

蛋白质	[储液] ug/uL	# 孔	uL 每孔	总体积	[最终] ug/uL	uL 蛋白质	uL 缓冲液
gp145-AcuteC Lot B2	1.0	20	20	400	0.025	10.0	390

[0463] - 向 UNTREATED 样本槽中加入 20uL 结合缓冲液。

[0464] - 向 POS CTRL 和 NO IgG CTRL 样本槽中加入 20uL 蛋白质 (每槽=0.5ug 蛋白质)。

[0465] - 培养板在 37°C 温育培养 60 分钟。

[0466] 结合实验

[0467] - 准备封闭缓冲液 (10% 小鼠免疫球蛋白、10% 人类免疫球蛋白溶于结合缓冲液中)

[0468] - 每孔中加入 50uL 封闭缓冲液

[0469] - 向 POS CTRL 样本槽中加入 4uL (2ug) 抗-α 4 封闭 mAb。

[0470] - 冰上放置 10 分钟 (不要洗脱)

[0471] - 每套装置转移蛋白质 / 免疫球蛋白复合物至培养板上。

[0472] - 冰上放置 30 分钟

[0473] - 使用 2X 结合缓冲液洗脱

[0474] 染色

[0475] - 制备染色混合试剂

[0476] 表 14

[0477]

uL 每孔	# 孔	总体积	uL β7-FITC (1:10)	uL NA-PE (1:400)	uL 缓冲液
50	20	1000	100	2.5	898

[0478] - 向未染色样本槽中加入 50uL 结合缓冲液。

[0479] - 向 NO IgG CTRL、POS CTRL 和所有样本槽中加入 50uL 染色混合试剂

[0480] -4℃温育 30min

[0481] - 使用 2X 结合缓冲液洗脱

[0482] - 使用 4% PFA x 在 4℃固定 30-60min。

[0483] - 向下旋转细胞并在 150uL 结合缓冲液中悬浮细胞。

[0484] -4℃保存培养板直至可以读板。

[0485] 补偿珠的制备

[0486] - 准备两管补偿珠,用 2X 染色缓冲液清洗珠子。

[0487] - 在 45uL 染色缓冲液中悬浮珠子

[0488] - 一管中加入 5uL CD4-FITC,另一管中加入 5uL CD4-PE,混合管。

[0489] -4℃温育 30min

[0490] - 使用 2X 染色缓冲液洗脱

[0491] -200uL 染色缓冲液中重悬

[0492] 在细胞系 RPMI8866 中进行基于流式细胞术的 $\alpha 4 \beta 7$ 结合抑制配位测定实验。但是最初在分离的 CD4+ 和 CD8+T 细胞上进行实验,该细胞培养可表达具有活性的 $\alpha 4 \beta 7$ 异源二聚体形式。图 57 展示了本实验的一个模型。本实验可以作为结合实验用来检测从不同的急性感染个体中选取的 HIV-1 的衍生序列的变异。本实验的扩大应用包括检测 env/ $\alpha 4 \beta 7$ 相互作用的功能封闭,这种封闭作用由单克隆抗体或针对包膜蛋白或整联蛋白免疫的纯化的免疫球蛋白血清引起。生产挑选出的 HIV-1 包膜蛋白,并生物素化,与抗体混合培养,而后与 $\alpha 4 \beta 7$ 表达细胞共培养。添加非蛋白抗生素 PE 检测结合,与非封闭的 mAb 偶联的 FITC 染色证实 $\alpha 4 \beta 7$ 的存在。检测抗整联蛋白抗体时,细胞与抗体预培养,而后再加入生物素化的包膜蛋白,如描述观察。申请人发展出这种方法用于整个包膜蛋白也用于生物素化的线性或环状肽。在某些实施方案中,该实验被发展用于 IMC 和 VLP 载体。

[0493] 该实验的最重要的特征为能够使用最初的 T 细胞表达 $\alpha 4 \beta 7$ 的活性形式。为生成这种细胞,申请人从 PBMC 中利用磁性珠分离 CD4+ 和 CD8+T 细胞。应用了阴性选择方案以使得到的细胞为“未受影响的”、纯化的且无小珠的。紧接着的分离,细胞在抗 -CD3/ 抗 -CD28、IL-2 和维 A 酸存在的情况下培养 5 天,诱导 $\alpha 4 \beta 7$ 的表面表达 (图 58)。

[0494] 本实验过程中的预实验是用来指导计算结合动力学并运用许多种 HIV-1 包膜蛋白试剂的全部实验程序。重组的 CRF01_AEgp120 和急性感染 C 亚型 gp145 (如上所述) 生物素化并结合于表达 $\alpha 4 \beta 7$ 的 CD4+ 或 CD8+T 细胞 (图 59, 1 左面和中间长条)。生物素化的环状肽包含了从 CRF01-AE 衍生出的 HIV-1 包膜蛋白的 V2 环,也被结合在 CD4+ $\alpha 4 \beta 7$ + 和 CD8+ $\alpha 4 \beta 7$ + 细胞上。在 RPMI8866 细胞系中也可见类似结合 (数据未显示)。与 B 进化枝 MN 衍生的 gp120 或与在环顶部突变的环状 V2 肽间检测不到结合。

[0495] 最初封闭研究使用人类抗 V2 单克隆抗体 (由 S. Zolla-Pazner 好心提供)、CRF01_AE 衍生的 gp120 或环状 V2 肽。具有反应活性的 V2 单克隆抗体 697-30D 和 2158 被检测,封闭了包膜蛋白与 CD4+ $\alpha 4 \beta 7$ + 和 CD8+ $\alpha 4 \beta 7$ + 细胞的结合。细胞与抗 - $\alpha 4$ 封闭抗体 HP2/1 预培养,而后再加入蛋白质或肽,作为正对照。申请人进行这些实验检测一组从 V2 环中衍生出的重叠线性肽,描绘出了这种相互作用所需要的氨基酸残基。(数据 未显不)

[0496] 方法学:

[0497] $\alpha 4 \beta 7$ T 淋巴细胞的制备:在完全培养基中解冻冷藏保存的 PBMC,通过磁性珠阴

性选择分离 CD4⁺ 或 CD8⁺T 细胞。细胞在抗 -CD3/CD28, IL-2 和类维生素 A 存在条件下培养至少 5 天。应用多色流式细胞术监控表现型、细胞生存和活性形式 $\alpha 4 \beta 7$ 的表达。在某些实验中,人类 B 细胞淋巴瘤细胞系 RPMI8866 用于大量表达活性形式 $\alpha 4 \beta 7$ 。

[0498] $\alpha 4 \beta 7$ 结合封闭实验细胞培养于 2-5 μ g 生物素化的 V2 肽或 HIV-1 包膜糖蛋白中培养 30 分钟,表达 $\alpha 4 \beta 7$ 。随后进行洗脱以移除未结合的肽或蛋白质,用非蛋白抗生素 PE 得到细胞,并通过流式细胞仪评估其结合。封闭研究中,抗体与 $\alpha 4 \beta 7$ 表达细胞或 HIV-1 包膜蛋白预培养,最好在加入试剂前培养 30 分钟。

[0499] HIV-1 包膜蛋白的合成急性的包膜蛋白序列从 RV217 急性感染研究中选出合成。提交序列至 GeneArt, Inc, 进行密码子优化并克隆进入哺乳动物表达载体。蛋白质在 CHO 细胞或 HEK293 细胞中表达,提供了不同的糖基化模式,对于结合实验可能非常重要。下一步表达,每种蛋白质的一部分进行生物素化,以用于 $\alpha 4 \beta 7$ 结合封闭实验。

[0500] 生物素化的 V2 肽的合成。由 Dr. Tim Cardozo (纽约大学) 设计的肽通过 Genemed Synthesis Inc. 合成并生物素化。这些肽由 Dr. Cardozo 馈赠。

[0501] 实施例 7

[0502] 概述

[0503] 凝胶过滤中的数据支持不同多聚体种类混合物的存在。尽管如此,目前还不确定在本实验中球状性质、疏水区和 gp145 重糖基化如何影响不同多聚体形式的分辨。因此,对种类存在的比例下结论是很困难的。并且,在本实验中的弱分辨率使得判定每种形式相关含量非常困难。对在 C06980v0c22gp145 中、Blue Native PAGE (BN PAGE) 中和 EGS 交联中纯化的许多低聚物的进一步分析,使用纯化蛋白质进行。而且,尝试应用凝胶过滤色谱进行不同低聚物形式的分离。

[0504] BN PAGE

[0505] 应用 Blue Native PAGE 分析纯化的 HIV-1C06980v0c22gp145 蛋白质 批次 s112009、120710A 和 120710B 复合受体的组成。纯化的 HIV-1C06980v0c22gp145 蛋白质在 4-16% Novex Bis-Tris 上进行聚丙烯酰胺凝胶电泳,使用了 Invitrogen's Native PAGE 系统 (图 61)。作为对照,如前所述的 Ba-L gp145 和三种 D 进化枝 gp140 蛋白质 (A07412, 57128, and 57140) 同时进行电泳。激光光密度测定分析在 BLUE Native PAGE 上进行电泳预示着每种蛋白质均为多种复合受体形式的混合物 (表 15)。

[0506] 表 15 激光光密度测定预测 gp140 和 gp145 多聚体组成

[0507]

进化枝	蛋白质	批号	相 对 分 子 量 (kDa)	多聚体受体种类	组成百分比
B	Ba-L gp145	061308	>800* 691 538 278	A B C D	40.3 % 21.5 % 35.1 % 3.1 %
D	A07412 gp140	3-31-05	>800* 733 574	A B C	30.5 % 32 % 37.5 %
D	57128 gp140	4-13-05	>800* 674 566	A B C	21.4 % 44.5 % 34 %
D	57140 gp140	4-20-05	>800* 773 659	A B C	27.1 % 40.6 % 32.2 %
C	C06980v0c22 gp145	112009	>800* 767 624 302/244**	A B C D	12.8% 52.7% 27.3% 7.2%
C	C06980v0c22 gp145	120710A	>800* 751 621 308/239**	A B C D	11.4 % 61 % 22% 5.6 %
C	C06980v0c22 gp145	120710B (DTT treated)	>800* 770 638 307/239**	A B C D	15.1 % 59.5 % 19.6 % 5.9%

[0508] * 极端扩散的条带跨越了几百 kDa, 难以确定分子量。在这些片段中可能存在不止一种。

[0509] ** 分隔不全的双联体可能代表了 2 种包括某些分解产物。

[0510] Blue Native PAGE 实验清楚显示 gp140 和 gp145 蛋白质存在于多种多聚体混合物中。尽管如此, 并不是非常清楚到底是哪种存在。一种设想, 单体大约 140kDa, 期望的三聚体、四聚体、五聚体和六聚体分子量分别约为 420kDa、560kDa、700kDa 和 840kDa。尽管如此, 由于它们具有球状形式、疏水区和重糖基化, 蛋白质疑似可能并没有表现出这种方式。相应的, 例如, 申请人将每一种多聚体指定为 A、B、C 或 D, 其中 A 是最 多的复合物, D 是最少的复合物。所有的 C06980v0c22gp145 批号具有类似表现 :B 是占优势地位的形式, 同时存在有大量的 C 和 D。批次 120710B 经 DTT 处理后在多聚体组成上没有表现出明显影响。在 D 进化枝 57140 和 57128 的 gp140 蛋白质也看到了类似的趋势。D 进化枝 A07412gp140 蛋白质是类似的, 但是其中 C 型在数量上稍多于 B 型。对于 Ba-Lgp145, 最占优势的种类是一条宽的扩散的条带, 与高阶多聚体混合物类似, 归类于 A 种。分离的主体 C 群和较少的 B 群也都明显存在。

[0511] EGS 交联 /SDS-PAGE

[0512] C06980v0c22gp145 进一步运用 SDS-PAGE 进行性质鉴定, 其中蛋白质与乙二醇二琥珀酰亚胺琥珀酸盐 (EGS) 交联用于进一步的复合受体特征鉴定。数据显示出三聚体最占优势, 但是二聚体、痕量的单体以及高阶多聚体也是存在的。纯化的 HIV-1_{C06980v0c22}gp145 (批次 120710A) 与 0.2、1.5 和 12.5mM EGS 交联, 在还原与非还原条件下 3-8% NuPAGE Tris 乙酸盐聚丙烯酰胺凝胶电泳中 (Invitrogen) 展开, 并被考马斯亮蓝染色 (图 62)。使用激光光密度测定分析估计每一种 gp145 的分子量。当采用 0.2mM EGS 处理时, gp145 交联不

完全,出现 3 种大小:334、232 和 139kDa。相应的样本槽中为三聚体、二聚体和单体形式的预测分子量。EGS 浓度上升至 5 和 12.5mM 时,交联完全,三聚体是占优势的种类。主要的二聚体形式也存在,但是单体只有总蛋白质的痕量凑数。在非还原性条件下,与高阶多聚物相关的模糊条带出现,这也是完全交联样品的证据。这揭示了一些高阶多聚物的存在,它们是由二硫键结合在一起。复合受体种类 A、B 和 C 推测起来与多聚体有关,分离开的三聚体和二聚体种类是分别与 EGS 交联的蛋白质。基于 Blue Native PAGE 的结果,主要多聚体种类为 B 和 C,其表观分子量分别为 751 和 621kDa。这些表观分子量非常高以至于不能下它们代表三聚体和二聚体的结论。尽管如此,应用 EGS 交联,表观分子量更多的与二聚体或三聚体形式一致。EGS 交联 SDS-PAGE 方法是由 J. P. Nkolola, 等人提出的,用以描述在杆状病毒群系统中生产的重组 HIV-192UG037.8gp140 蛋白质以三聚体形式存在。利用类似的 EGS 交联程序 C06980v0c22gp145 蛋白质在电泳中出现类似的分子量。在 BN-PAGE 和 EGS 交联 SDS-PAGE 中观察到的与预测的 gp145 低聚体的分子量间的显著偏差,可能是由于电荷、球状性质、疏水区、gp145 蛋白质重糖基化影响了本实验中不同多聚体形式的溶解度。BN-PAGE 能够分辨自然状态下的低聚体形式。尽管如此,不能依据 gp145 和分子量标记之间流动性的成比例差异推测出分子量。在 EGS-交联 SDS-PAGE 中,低聚体共价结合,但是在一般的 SDS-PAGE 中变性。在这样的条件下,假设当 gp145 没有交联时,其移动与基于表观分子量的标准高度相关。

[0513] 二聚体和三聚体纯化

[0514] 采用凝胶过滤层析研究关于是否可以用于分离 gp145 多种不同种类低聚物。多种不同形式的成功分离将为可能的研究提供每种形式的抗原性和免疫原性。

[0515] 应用 Superose6 进行凝胶过滤层析。目前认为 Superose6 因其高分子量范围,具有分离高分子量的 gp145 种类的潜能,最佳的分离蛋白大小为 5 至 500kDa。因为高分子量蛋白质在凝胶过滤树脂中的低溶解性,大蛋白的分离通常证明是很难的,例如 gp145 低聚物。Superose6 显示出分离不同低聚物的可能。

[0516] 在分析用 Superose6PC3.2/30(GE Healthcare)柱中,25 μ l 中含有 200 μ g HIV-1_{C06980v0c22}gp145 批次 120710A 溶于 0.05ml/min PBS, pH7.2 中。利用 EGS 交联/SDS-PAGE 收集和分析 50 μ l 片段(图 63)。尽管没有片段含有纯净的二聚体或三聚体,一些三聚体的富集物确实存在于某些片段中。高阶多聚体存在于某些片段中。不同柱大小和条件能够具有更好的分辨率,并且具有将三聚体从高阶多聚体和二聚体中分离出来的能力。在进一步的实施方案中,可能利用 Superose6 进行分离条件的优化。

[0517] 参考文献

[0518] Nkolola, J. P., 等人. 2010. Breadth of neutralizing antibodies elicited by stable, homogeneous clade A and clade C HIV-1gp140envelope trimers in guinea pigs. J. Virol. 84 :3270-3279.

[0519] ***

[0520] 已经详细描述本发明的实施方案,应当理解在以上段落中阐述的本发明并不仅仅局限于上文描述的具体细节,由于很多明显的变化也是可能的,其并不背离本发明的精神和范围。

[0521] 在本申请引用或描述的每项专利、专利申请和出版物通过引用并入本文,就如同

每一个专利、专利申请或出版物被明确和单独地指明通过引用并入本文。

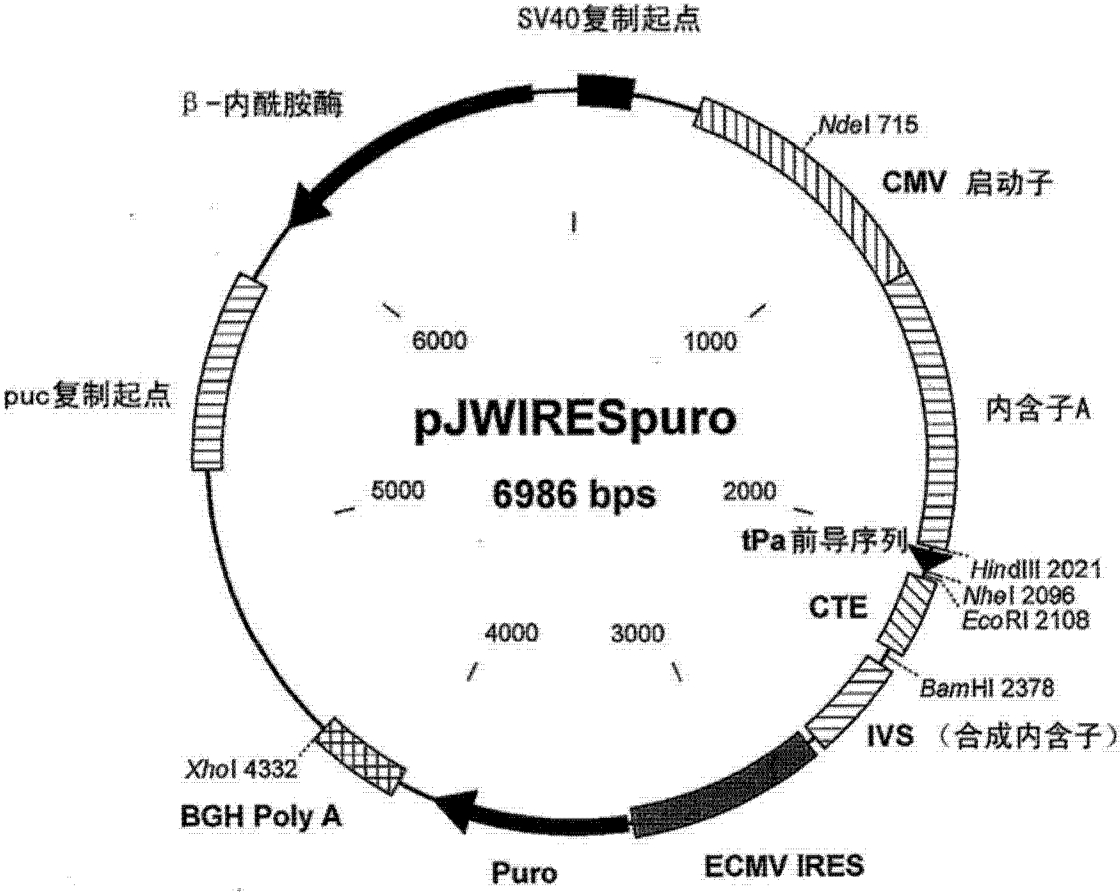


图 1

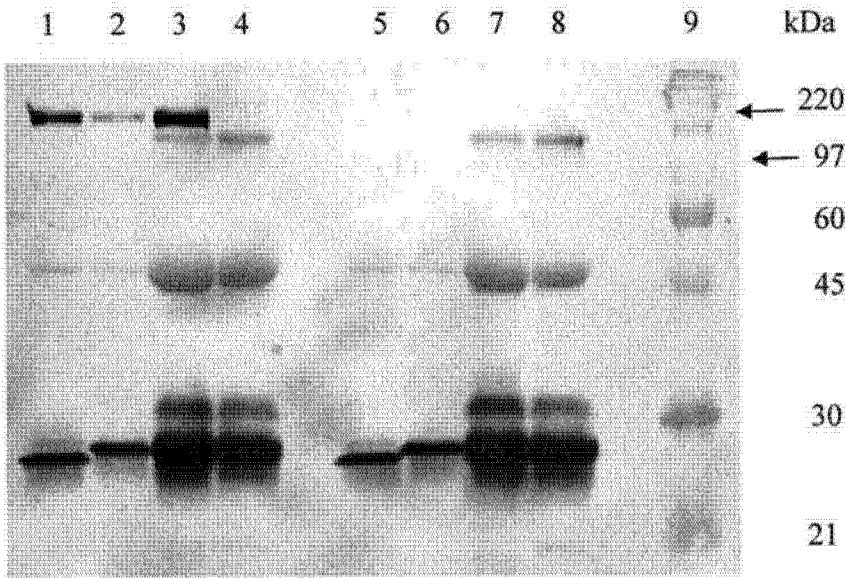


图 2

HIV-1 Ba-L gp140 DC 4E10核苷酸序列(SEQ ID NO:6)

ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTT
TCGGCTAGC GAGGAAAAGCTGTGGGTGACCGTGACTACGGCGTGCCCGTGTGGAAAGAG
GCCACCACCACCCTGTTCTGCGCCAGCGACGCCAAGGCCTACGACACCGAGGTGCACAAC
GTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAGGAAGTGGAGCTGGAA
AACGTGACCGAGAACTTCAACATGTGGAAGAACAACATGGTGGAGCAGATGCACGAGGAC
ATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGTGCGTG
ACCCTGAACTGCACCGACCTGCGGAACGCCACCAACGGCAACGACACCAACACCACCTCC
AGCAGCCGGGAGATGATGGGCGGAGGGGAGATGAAGAACTGCAGCTTCAAGATCACCACC
AACATCCGGGGCAAGGTGCAGAAAGAGTACGCCCTGTTCTACGAGCTGGACATCGTGCCC
ATCGACAACAACAGCAACAACCGGTACAGGCTGATCAGCTGCAACACCAGCGTGATCACC
CAGGCCTGCCCCAAGATCAGCTTCGAGCCCATCCCCATCCACTACTGCGCCCCTGCCGGC
TTCGCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCAAGGGCCCCTGCAGCAACGTG
AGCACCGTGAGTGACCCACGGCATCCGGCCCCGTGGTGTCCACCCAGCTGCTGCTGAAC
GGCAGCCTGGCCGAGGAAGAGGTGGTGATCAGAAGCGAGAACTTCGCCGACAACGCCAAG
ACCATCATCGTGAGCTGAACGAGAGCGTGGAGATCAACTGCACCCGGCCCCAACAACAAC
ACCCGGAAGAGCATCCACATCGGCCCTGGCAGGGCCCTGTACACCACCGGCGAGATCATC
GGCGACATCCGGCAGGCCCACTGCAACCTGAGCCGGGCCAAGTGAACGACACCCTGAAC
AAGATCGTGATCAAGCTGCGGGAGCAGTTCGGCAACAAGACCATCGTGTTCAAGCACAGC
AGCGGCGGAGACCCCGAGATCGTGACCCACAGCTTCAACTGTGGCGGCGAGTTCTTCTAC
TGCAACAGCACCCAGCTGTTCAACAGCACCTGGAATGTGACCGAGGAAAGCAACAACACC
GTGGAGAACAACACCATCACCTGCCCTGCCGGATCAAGCAGATCATCAATATGTGGCAG
AAAGTGGGCAGAGCCATGTACGCCCTCCCATCCGGGGCCAGATCCGGTGAGCAGCAAC
ATCACCGGCCTGCTGCTGACCCGGGACGGCGGACCCGAGGCCAACAAGACCGAGGTGTTT

图 3

CGGCCTGGCGGCGGAGATATGCGGGACAACCTGGCGGAGCGAGCTGTACAAGTACAAGGTG
GTGAAGATCGAGCCCCTGGGCGTGGCCCCACCAAGGCCAAGAGCCGGGTGGTGCAGCGG
GAGAAGAGCGCCGTGGGCATCGGCGCCGTGTTTCTGGGCTTCTGGGAGCCGCCGGAAGC
ACCATGGGAGCCGCCAGCATGACCCTGACCGTGCAGGCCCGGCTGCTGCTGTCCGGCATC
GTGCAGCAGCAGAACAACCTGCTGCGGGCCATTGAGGCACAGCAGCATCTGCTGCAGCTG
ACCGTGTGGGGCATTAAAGCAGCTGCAGGCCAGGGTGCTGGCCGTGGAGAGATACCTGCGG
GATCAGCAGCTGCTGGGGATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCACCGCCGTG
CCCTGGAACGCCAGCTGGTCCAACAAGAGCCTGAACAAAATCTGGGACAACATGACCTGG
ATGGAATGGGACCGCGAGATCAACAACCTACACCAGCATCATCTACAGCCTGATCGAGGAA
AGCCAGAACCAGCAGGAAAAGAACGAGCAGGAACTGCTGGAAGTGGACAAGTGGGCCAGC
CTGTGGAAGTGGTTCGACATCACCGAGTGGCTGTGGTACATCAAGAAGAAGAAGTGA

图 3 续

HIV-1 Ba-L gp140 DC 4E10氨基酸序列(SEQ ID NO:7)

MDAMKRGGLCCVLLCGAVFVSAS EEKLWVTVYGVVPVWKEATTTLCASDAKAYDTEVHN
VWATHACVPTDPNPQEVELENTENFNMWKNNMVEQMHEDIISLWDQSLKPCVKLTPLCV
TLNCTDLRNATNGNDTNTTSSSREMMGGGEMKNCSFKITTNIRGKVQKEYALFYELDVP
IDNNSNNRYRLISCNTSVITQACPKISFEPIPIHYCAPAGFAILKCKDKKFNGKGPCSNV
STVQCTHGIRPVVSTQLLNGSLAEEEVVIRSENFADNAKTIIVQLNESVEINCTRPNNN
TRKSIHIGPGRALYTTGEIIGDIRQAHCNLSRAKWNDTLNKIVIKLREQFGNKTIVFKHS
SGGDPEIVTHSFNCGGEFFYCNSTQLFNSTWNVTEESNNTVENNTITLPCRIKQIINMWQ
KVGRAMYAPPIRGQIRCSSNITGLLLTRDGGPEANKTEVFRPGGGDMRDNWRSELYKYKV
VKIEPLGVAPTKAKSRVVQREKSAVGIGAVFLGFLGAAGSTMGAASMTLTVQARLLLSGI
VQQQNNLLRAIEAQQHLLQLTVWGIKQLQARVLAVERYLRDQQLGIWGC SGKLICTTAV
PWNASWSNKS LNKIWDNMTWMEWDREINNYTSIYSLIEESQNQQEKNEQELLELDK WAS
LWNWFDITEWLWYIKKKK

图 4

HIV-1 Ba-L gp145核苷酸序列(SEQ ID NO:8)

ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTT
ACTACCACAGAGGCTAGC GAGGAAAAGCTGTGGGTGACCGTGTACTACGGCGTGCCCGTG
TGGAAGAGGCCACCACCACCCTGTTCTGCGCCAGCGACGCCAAGGCCTACGACACCGAG
GTGCACAACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAGGAAGTG
GAGCTGGAAAACGTGACCGAGAACTTCAACATGTGGAAGAACAACATGGTGGAGCAGATG
CACGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTGACCCCC
CTGTGCGTGACCCTGAACTGCACCGACCTGCGGAACGCCACCAACGGCAACGACACCAAC
ACCACCTCCAGCAGCCGGGAGATGATGGGCGGAGGGGAGATGAAGAACTGCAGCTTCAAG
ATCACCACCAACATCCGGGGCAAGGTGCAGAAAGAGTACGCCCTGTTCTACGAGCTGGAC
ATCGTGCCCATCGACAACAACAGCAACAACCGGTACAGGCTGATCAGCTGCAACACCAGC
GTGATCACCAGGCCTGCCCAAGATCAGCTTCGAGCCCATCCCCATCCACTACTGCGCC
CCTGCCGGCTTCGCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCAAGGGCCCCCTGC
AGCAACGTGAGCACCGTGCAGTGCACCCACGGCATCCGGCCCCGTGGTGTCCACCCAGCTG
CTGCTGAACGGCAGCCTGGCCGAGGAAGAGGTGGTGATCAGAAGCGAGAACTTCGCCGAC
AACGCCAAGACCATCATCGTGCAGCTGAACGAGAGCGTGGAGATCAACTGCACCCGGCCC
AACAACAACACCCGGAAGAGCATCCACATCGGCCCTGGCAGGGCCCTGTACACCACCGGC
GAGATCATCGGCGACATCCGGCAGGCCCACTGCAACCTGAGCCGGGCCAAGTGGAACGAC
ACCCTGAACAAGATCGTGATCAAGCTGCGGGAGCAGTTCGGCAACAAGACCATCGTGTTT
AAGCACAGCAGCGGCGGAGACCCCGAGATCGTGACCCACAGCTTCAACTGTGGCGGCGAG
TTCTTCTACTGCAACAGCACCCAGCTGTTCAACAGCACCTGGAATGTGACCGAGGAAAGC
AACAACACCGTGGAGAACAACACCATCACCTGCCCTGCCGGATCAAGCAGATCATCAAT
ATGTGGCAGAAAAGTGGGCAGAGCCATGTACGCCCTCCCATCCGGGGCCAGATCCGGTGC
AGCAGCAACATCACCGGCCTGCTGCTGACCCGGGACGGCGGACCCGAGGCCAACAAGACC

图 5

GAGGTGTTCCGGCCTGGCGGCGGAGATATGCGGGACAACCTGGCGGAGCGAGCTGTACAAG
TACAAGGTGGTGAAGATCGAGCCCCTGGGCGTGGCCCCACCAAGGCCAAGAGCCGGGTG
GTGCAGCGGGAGAAGAGCGCCGTGGGCATCGGCGCCGTGTTTCTGGGCTTCCTGGGAGCC
GCCGGAAGCACCATGGGAGCCGCCAGCATGACCCTGACCGTGCAGGCCCGGCTGCTGCTG
TCCGGCATCGTGCAGCAGCAGAACAACCTGCTGCGGGCCATTGAGGCACAGCAGCATCTG
CTGCAGCTGACCGTGTGGGGCATTAAAGCAGCTGCAGGCCAGGGTGCTGGCCGTGGAGAGA
TACCTGCGGGATCAGCAGCTGCTGGGGATCTGGGGCTGCAGCGGCAAGCTGATCTGCACC
ACCGCCGTGCCCTGGAACGCCAGCTGGTCCAACAAGAGCCTGAACAAAATCTGGGACAAC
ATGACCTGGATGGAATGGGACCGCGAGATCAACAACCTACACCAGCATCATCTACAGCCTG
ATCGAGGAAAGCCAGAACCAGCAGGAAAAGAACGAGCAGGAACTGCTGGAACTGGACAAG
TGGGCCAGCCTGTGGAACCTGGTTCGACATCACCGAGTGGCTGTGGTACATCAAGAAGAAG
AAGTGA

图 5 续

HIV-1 Ba-L gp145氨基酸序列 (SEQ ID NO:9)

MDAMKRGLCVLLLCGAVFVTTTEAS EEKLWVTVYYGVPVWKEATTTLCASDAKAYDTE
VHNVWATHACVPTDPNPQEVELENTNFNMWKNMVEQMHEDIISLWDQSLKPCVKLTP
LCVTLNCTDLRNATNGNDTNTTSSSREMMGGGEMKNCSFKITTNIRGKVQKEYALFYELD
IVPIDNNSNNRYRLISCNTSVITQACPKISFEPIPIHYCAPAGFAILKCKDKKFNGKGPC
SNVSTVQCTHGIRPVVSTQLLNGSLAEEVVIRSENFADNAKTIIVQLNESVEINCTRP
NNNTRKSIHIGPGRALYTTGEIIGDIRQAHCNLSRAKWNDTLNKIVIKLREQFGNKTIFF
KHSSGGDPEIVTHSFNCGGEFFYCNSTQLFNSTWNVTEESNNTVENNTITLPCRIKQIIN
MWQKVGRAMYAPPIRGQIRCSSNITGLLLTRDGGPEANKTEVFRPGGGDMRDNRSELYK
YKVVKIEPLGVAPTKAKSRVVQREKSAVGIGAVFLGFLGAAGSTMGAASMTLTVQARLLL
SGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARVLAVERYLRDQQLGIWGCSGKLICT
TAVPWNASWSNKS LNKIWDNMTWMEWDREINNYTSIIYSLIEESQNQQEKNEQELLELDK
WASLWNWFDITEWLWYIKKKK

图 6

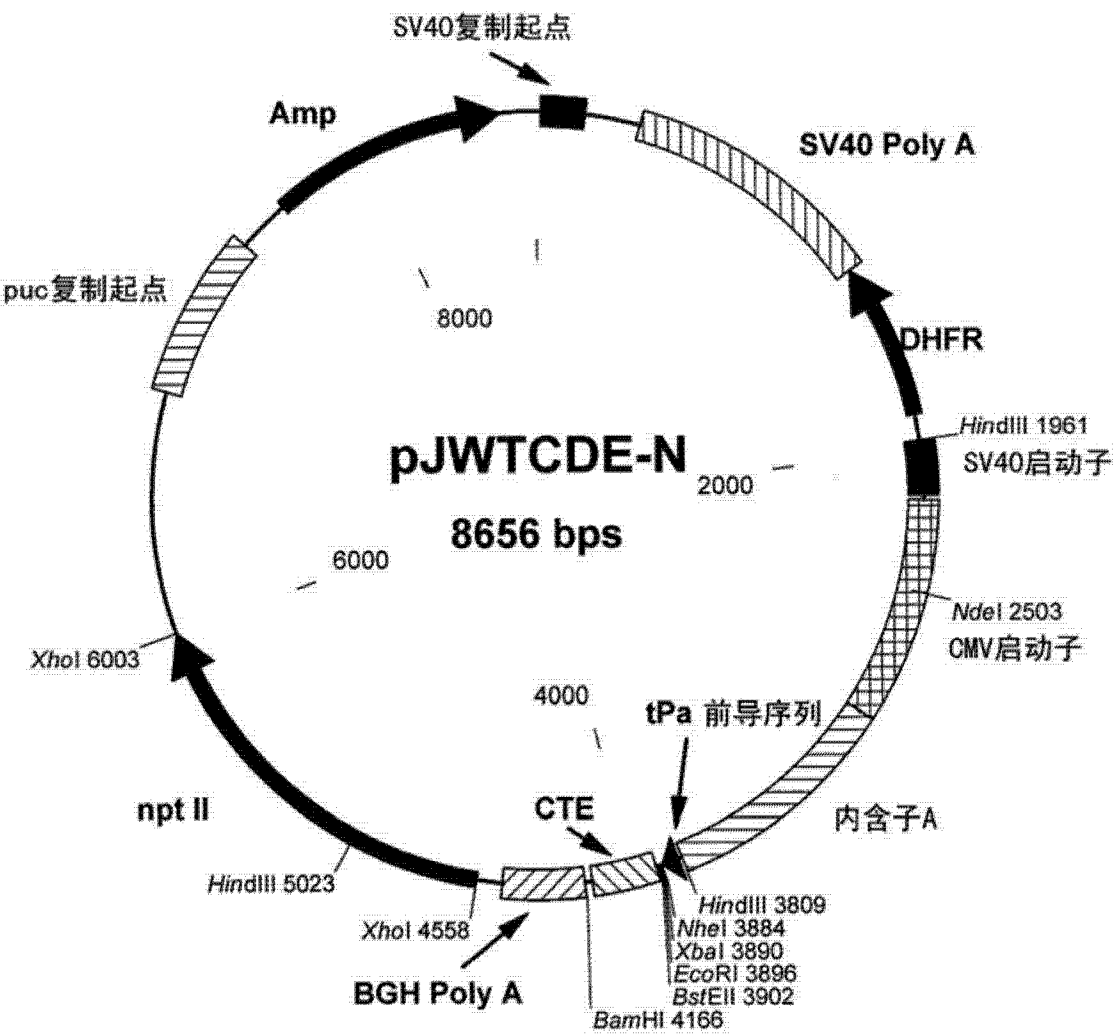


图 7

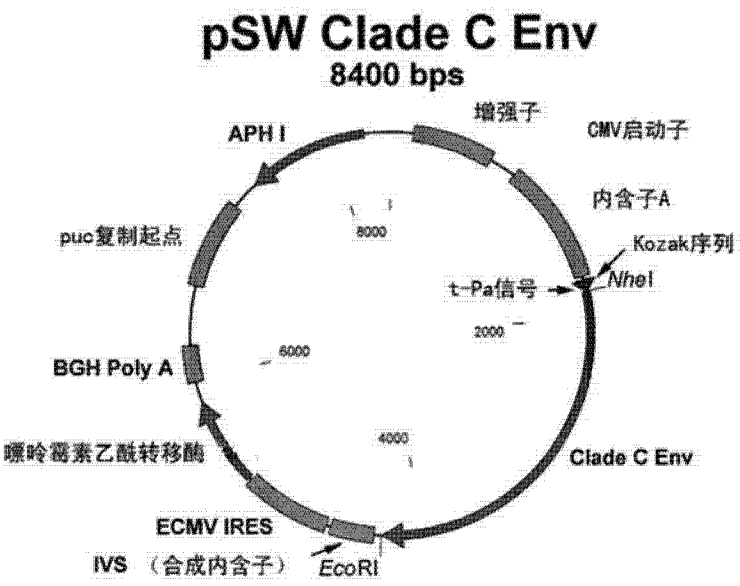


图 8

HIV-1 03728v2c6 gp160密码子优化 (SEQ ID NO:10):

突出显示tPa信号

ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTT
ACTACCACAGAGGCTAGC AACCTGTGGGTGACCGTGTAACGGCGTGCCCGTGGAAG
GAGGCCAAGACCACCCTGTTCTGCGCCAGCGACGCCAAGGCCTACGAGAAGGAGGTGCAC
AACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCAAGGAGATGGAGCTG
CCCAACGTGACCGAGAACTTCAACATGTGGAAGAACGACATGGTGGACCAGATGCACGAG
GACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGTGC
GTGACCCTGAACTGCAACAAGGTGAACAACAACCGCAGCATCGACAACAACAGCACCGAG
GAGGAGATGAAGAACTGCACCTTCAACACCACCACCGAGATCAGGGACAAGAAGAGGACC
CAGCAGGCCCTGTTCTACAAGCTGGACATCGTGCCCTGGGCAACAGCAACGAGAGCTAC
AGGCTGATCAACTGCAACACCAGCACCTGACCCAGGCCTGCCCCAAGGTGACCTTCGAC
CCCATCCCCATCCACTACTGCGCCCCAGCCGGCTACGCCATCCTGAAGTGCAAGGACGAG
AAGTTCAACGGCACCGGCCCTGCAACAACGTGTCCACCGTGCAAGTGCAAGGACGAG
AAGCCCGTGGTGTCCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGAAGGACATCATC
ATCCGCAGCGAGAACCTGACCAACAACGTGAAGACCATCATGGTGCACCTGAACGAGAGC
GTGGAGATCAACTGTACCAGGCCCAACAACAACACCAGGAAGAGCATCAGGATCGGCCCA
GGCCAGACCTTCTACGCCACCGGCGGCATCATCGGCAACATCAGGCAGGCCTACTGCACC
ATCAGCAAGAACAAGTGGAACACCACCCTGGAGAGAGTGGCTACCAAGCTGAAGGAGTAC
TTCAAGAACACCACCATCCAGTTCGCCCCCACTCTGGCGGCGACCTGGAGATCACCACC
CACAGCTTCAACTGCAGGGGCGAGTTCTTCTACTGCAACACAAGCCAGCTGTTCAACGGG
ACCAGCACCGGCTTCAGCAACAAGAGCACCGGCAACGAGACATTACCCTGCCCTGCAGA
ATCAAGCAGATCATCAACATGTGGCAGGAAGTGGGCAGGGCCATGTACGCCCCACCAATC

图 9

GCCGGGAACATCACCTGCGTGAGCAACATCACCGGCCTGCTGCTGACCAGGGACGGCGGG
GACAACAACACAAGAACCGAGACATTCCGGCCAGGCGGCGGAGACATGAGGGACAACCTGG
CGGAGCGAGCTGTACAAGTACAAGGTGGTGGAGGTGAAGCCCCTGGGCGTGCCCCAAGC
GAAGCCAGGCGGAGAGTGGTGGAACGGGAGAAGAGGGCCGTGGGCCTGGGCGCCGTCTTT
CTGGGATTCTGGGGGCTGCCGGATCCACCATGGGAGCCGCCAGCATCACCTGACCGTG
CAGGCTAGGCAGCTGCTGTCCGGAATCGTGCAGCAGCAGTCCAACCTGCTGAGGGCTATC
GAGGCTCAGCAGCATATGCTCCAGCTGACCGTGTGGGGCATCAAGCAGCTGCAGACCAAG
GTGCTGGCCATCGAGAGATATCTGCAGGACCAGCAGCTGCTCGGCATCTGGGGCTGCAGC
GGCAAGCTGATCTGCACCACCAACGTGCCCTGGAACAGCAGCTGGTCCAACAAGTCCCTG
AAGGCCATCTGGGACAACATGACCTGGATGCAGTGGGACAGAGAGATCAGCAACTACACC
AACACCATCTACAGGCTGCTGGAGGACAGCCACATCCAGCAGGAGAAGAACGAGAAGGAC
CTGCTGGCCCTGAACAGCTGGAACAACCTGTGGAGCTGGTTCAACATCACCAACTGGCTG
TGGTACATCCGGATCTTCATCATGATCGTGGGCGGCCTGATCGGCCTGAGGATCATCTTC
GCCATCCTGTCCATCGTGAACAGAGTGCGCCAGGGCTACAGCCCACTGAGCTTCCAGACC
CTGACACCCAACCAGCGGGGCCAGACAGACTGGGCAGGATCGAGGAGGAGGGCGGCGAG
CAGGACAAGGACAGATCCATCAGGCTGGTGAACGGCTTCCTGGCCCTGGCCTGGGACGAC
CTGAGAAGCCTGTGCCTGTTTCAGCTACCACAGGCTGAGGGACTTCATCCTGATCGCCACC
AGGGCCGTGGAACCTGCTGGGGCACTCCAGCCTGAGGGGCCTGCAGAGGGGCTGGGAGGCC
CTGAAGTACCTGGGCTCCCTGGGCCAGTATTGGGGCCAGGAGCTGAAGAAGAGCGCCATC
AGACTGCTGGACATCACCGCCATCGCCGTGGCCGAGGGCACCGACAGGATCATCGAGTTC
ATCCAGAGGATCTGCAGGGCCATCCACAACGTGCCCAACAGGATCAGGCAGGGCTTCGAG
GCCGCCCTGCTGTGATGA

图 9 续

HIV-1 C3728v2c6 gp160 DC密码子优化核苷酸序列 (SEQ ID NO:11):

突出显示tPa信号

gp120/gp140切割位点突变以阻止切割

```
ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTT
ACTACCACAGAGGCTAGC AACCTGTGGGTGACCGTGTACTACGGCGTGCCCGTGTGGAAG
GAGGCCAAGACCACCCTGTTCTGCGCCAGCGACGCCAAGGCCTACGAGAAGGAGGTGCAC
AACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCAGGAGATGGAGCTG
CCCAACGTGACCGAGAACTTCAACATGTGGAAGAACGACATGGTGGACCAGATGCACGAG
GACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGTGC
GTGACCCTGAACTGCAACAAGGTGAACAACAACCGCAGCATCGACAACAACAGCACCGAG
GAGGAGATGAAGAACTGCACCTTCAACACCACCACCGAGATCAGGGACAAGAAGAGGACC
CAGCAGGCCCTGTTCTACAAGCTGGACATCGTGCCCTGGGCAACAGCAACGAGAGCTAC
AGGCTGATCAACTGCAACACCAGCACCTGACCCAGGCCTGCCCCAAGGTGACCTTCGAC
CCCATCCCCATCCACTACTGCGCCCCAGCCGGCTACGCCATCCTGAAGTGCAAGGACGAG
AAGTTCAACGGCACCGGCCCTGCAACAACGTGTCCACCGTGCAGTGCACCCACGGCATC
AAGCCCGTGGTGTCCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGAAGGACATCATC
ATCCGCAGCGAGAACCTGACCAACAACGTGAAGACCATCATGGTGCACCTGAACGAGAGC
GTGGAGATCAACTGTACCAGGCCCAACAACAACACCAGGAAGAGCATCAGGATCGGCCCA
GGCCAGACCTTCTACGCCACCGGCGGCATCATCGGCAACATCAGGCAGGCCTACTGCACC
ATCAGCAAGAACAAGTGGAACACCACCCTGGAGAGAGTGGCTACCAAGCTGAAGGAGTAC
TTCAAGAACACCACCATCCAGTTCGCCCCCACTCTGGCGGCGACCTGGAGATCACCACC
CACAGCTTCAACTGCAGGGGCGAGTTCTTCTACTGCAACACAAGCCAGCTGTTCAACGGG
ACCAGCACCGGCTTCAGCAACAAGAGCACCGGCAACGAGACATTACCCCTGCCCTGCAGA
ATCAAGCAGATCATCAACATGTGGCAGGAAGTGGGCAGGGCCATGTACGCCCCACCAATC
```

图 10

GCCGGGAACATCACCTGCGTGAGCAACATCACCGGCCTGCTGCTGACCAGGGACGGCGGG
GACAACAACACAAGAACCGAGACATTCCGGCCAGGCGGCGGAGACATGAGGGACAAGTGG
CGGAGCGAGCTGTACAAGTACAAGGTGGTGGAGGTGAAGCCCCTGGGCGTGGCCCCAAGC
GAAGCCAGGAGCAGAGTGGTGGAAACGGGAGAAGAGCGCCGTGGGCCTGGGCGCCGTCTTT
CTGGGATTCTGGGGGCTGCCGGATCCACCATGGGAGCCGCCAGCATCACCTGACCGTG
CAGGCTAGGCAGCTGCTGTCCGGAATCGTGCAGCAGCAGTCCAACCTGCTGAGGGCTATC
GAGGCTCAGCAGCATATGCTCCAGCTGACCGTGTGGGGCATCAAGCAGCTGCAGACCAAG
GTGCTGGCCATCGAGAGATATCTGCAGGACCAGCAGCTGCTCGGCATCTGGGGCTGCAGC
GGCAAGCTGATCTGCACCACCAACGTGCCCTGGAACAGCAGCTGGTCCAACAAGTCCCTG
AAGGCCATCTGGGACAACATGACCTGGATGCAGTGGGACAGAGAGATCAGCAACTACACC
AACACCATCTACAGGCTGCTGGAGGACAGCCACATCCAGCAGGAGAAGAACGAGAAGGAC
CTGCTGGCCCTGAACAGCTGGAACAACCTGTGGAGCTGGTTCAACATCACCAACTGGCTG
TGGTACATCCGGATCTTCATCATGATCGTGGGCGGCCTGATCGGCCTGAGGATCATCTTC
GCCATCCTGTCCATCGTGAACAGAGTGCGCCAGGGCTACAGCCCACTGAGCTTCCAGACC
CTGACACCCAACCAGCGGGGCCAGACAGACTGGGCAGGATCGAGGAGGAGGGCGGCGAG
CAGGACAAGGACAGATCCATCAGGCTGGTGAACGGCTTCTGGCCCTGGCCTGGGACGAC
CTGAGAAGCCTGTGCCTGTTCAAGCTACCACAGGCTGAGGGACTTCATCCTGATCGCCACC
AGGGCCGTGGAAGTCTGGGGCACTCCAGCCTGAGGGGCCTGCAGAGGGGCTGGGAGGCC
CTGAAGTACCTGGGCTCCCTGGGCCAGTATTGGGGCCAGGAGCTGAAGAAGAGCGCCATC
AGACTGCTGGACATCACCGCCATCGCCGTGGCCGAGGGCACCGACAGGATCATCGAGTTC
ATCCAGAGGATCTGCAGGGCCATCCACAACGTGCCCAACAGGATCAGGCAGGGCTTCGAG
GCCGCCCTGCTGTGA

图 10 续

HIV-1 C06838v1c48 gp160密码子优化核苷酸序列 (SEQ ID NO:12):

突出显示tPa信号

ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTT
ACTACCACAGAGGCTAGC AACCTGTGGGTGACCGTGTACTACGGCGTGCCCGTGTGGAGA
GAGGCCAAGACCACCCTGTTCTGCGCCAGCGACGCCAAGGCCTACGAAAAGGAGGTGCAC
AACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCGACCCCGAGGAGATGGTGCTG
AAGAACGTGACCGAGAACTTCAACATGTGGAAGAACGACATGGTGGAGCAGATGCACGAG
GACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGTGC
GTGACCCTGACCTGCACCAACGTGAACGCCACCGACAACGTGACCTACAAGGAGAAGATG
GAGGGCGAGATCAAGAACTGCAGCTTCAACATCACCACCGAGATCAGGGACAAGAAGAGG
AAGGTCCACGCCCTGTTCTACAGGCTGGACGTGGTGCAGCTGAACAACAGCAACGAGTAC
ATCCTGATCAACTGCAACACCAGCGCCATCACCAGGCCTGCCCCAAGGTGTCCTTCGAC
CCCATCCCCATCCACTACTGCGCCCCAGCCGGCTACGCCATCCTGAAGTGCAACGACAAG
CCCTTCAACGGCACCGGCCCATGCCACAACGTGTCTACCGTGCACTGCACCCACGGCATC
AAGCCCGTGATCAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGAAGGAGGTGATC
ATCCGCTTCGAGAACCTGACCGACAACGCCAAGACCATCATCGTGCACTGAACCAGAGC
ATCGAGATCAAGTGCATCAGGCCCAACAACAACACCAGAGAGAGCATCAGGATCGGCCCCA
GGCCAGGCCTTCTACGCCACCAGGGACATCATCGGCGACATCAGAAGGGCCTACTGCACC
ATCGAGACAGAGAGGTGGAAGGAGACACTGGGCAACGTACAGAGAAGCTGAAGGAGTAC
TTCCCCAACACCAACATCAGCTTTGCTCCCTCCTCTGGCGGGGACCTGGAGATCACCACC
CACAGCTTCAACTGCAGGGGCGAGTTCTTCTACAGAAACACCAGCAAGCTGTTCAACAAC
AACGACACCGAGAACAACCTGACCATCACCTGCCCTGCAGAATCAAGCAGATCGTGAAC
ATGTGGCAGGGCGTGGGCAGGGCCATGTACGCCCCACCATCGAGGGCAACATCACCTGC
AAGTCCAACATCACCGGCCTGCTGCTGACCAGAGACGGCGGCAACGGCACAGAGAGGAAC

图 11

GAGACATTCAGGCCAGGCGGCGGAGACATGAGGAACAACCTGGCGGAGCGAGCTGTACAAG
TACAAGGTGGTGGAGATCAAGCCCCTGGGGATCGCCCCAACCGGCGCCAAGAGAAGAGTG
GTGGAGAGAGAGAAGAGGGCCGTCGGCATGGGCGCCGTGTTTCTGGGCTTCCTGGGAGCC
GCCGGAAGCACCATGGGAGCCGCCAGCCTGACCCTGACAGTGCAGGCCAGACAGGTGCTG
TCCGGGATCGTGCAGCAGCAGAGCAACCTGCTGAGAGCTATCGAGGCTCAGCAGCACATG
CTGCAGCTGACCGTGTGGGGCATCAAGCAGCTGCAGACCAGGGTGCTGGCCATCGAGAGA
TACCTGAAAGATCAGCAGCTGCTCGGCCTCTGGGGCTGCAGCGGCAAGCTGATCTGCACC
ACAGCCGTGCCATGGAATAGCAGCTGGTCCAACAAGAGCGAGATCGACATCTGGAACAAC
ATGACCTGGATGCAGTGGGACAGAGAGATCAGCAACTACACCAACACCATCTACAGACTG
CTGGAGGACAGCCAGACCCAGCAGGAGAAGAACGAGAAGGACCTGCTGGCCCTGGACAGC
TGGAAGAACCTGTGGAACCTGGTTTAACATCACCAAGTGGCTGTGGTACATCAAGATCTTC
ATCATCATCGTGGGCGGCCTGATCGGCCTGAGGATCATCTTCGCCGTGCTGTCCATCGTG
AACAGAGTGCGCCAGGGGTACAGCCCACTGAGCTTCAGACCCTGATCCCCAGCCCCAGA
GGCCCCGACAAGCTGGGCAGGATCGAGGAGGAGGGCGGCGAGCAGGACAAGGGCAGATCC
GTGAGGCTGGTGAACGGCTTCCTGGCCCTGGCCTGGGACGACCTGAGAAGCCTGTGCCTG
TTCAGCTATCACCAGCTGAGGGACTTCATCCTGATCGTGGCCAGAGCCGTGGGCCTGCTG
GGCAGGTCCAGCCTGAGGGGCCTGCAGAGAGGCTGGGAGATCCTGAAGTACCTGGGCGGG
CTGGTGCACTACTGGGGCCTGGAGCTGAAGAAGAGCGCCGTGAGCCTGTTCGACACAATC
GCCATCGCCGTGACCGAGGGCACCGACAGGATCATCGAGCTGATCCAGAGGTCCTGCAGA
GCCATCAGGAACGTGCCCACCAGGATCAGGCAGGGCTTCGAGGCCGCCCTGCAGTGA

图 11 续

HIV-1 C06838v1c48 gp160 DC密码子优化核苷酸序列 (SEQ ID NO:13):

突出显示tPa信号

gp120/gp140切割位点突变以阻止切割

ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTT
ACTACCACAGAGGCTAGC AACCTGTGGGTGACCGTGACTACGGCGTGCCCGTGTGGAGA
GAGGCCAAGACCACCCTGTTCTGCGCCAGCGACGCCAAGGCCTACGAAAAGGAGGTGCAC
AACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCGACCCCGAGGAGATGGTGCTG
AAGAACGTGACCGAGAACTTCAACATGTGGAAGAACGACATGGTGGAGCAGATGCACGAG
GACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGTGC
GTGACCCTGACCTGCACCAACGTGAACGCCACCGACAACGTGACCTACAAGGAGAAGATG
GAGGGCGAGATCAAGAACTGCAGCTTCAACATCACCACCGAGATCAGGGACAAGAAGAGG
AAGGTCCACGCCCTGTTCTACAGGCTGGACGTGGTGCAGCTGAACAACAGCAACGAGTAC
ATCCTGATCAACTGCAACACCAGCGCCATCACCCAGGCCTGCCCCAAGGTGTCCTTCGAC
CCCATCCCCATCCACTACTGCGCCCCAGCCGGCTACGCCATCCTGAAGTGCAACGACAAG
CCCTTCAACGGCACCGGCCCATGCCACAACGTGTCTACCGTGCAGTGCACCCACGGCATC
AAGCCCGTGATCAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGAAGGAGGTGATC
ATCCGCTTCGAGAACCTGACCGACAACGCCAAGACCATCATCGTGCAGCTGAACCAGAGC
ATCGAGATCAAGTGCATCAGGCCCAACAACAACACCAGAGAGAGCATCAGGATCGGCCCA
GGCCAGGCCTTCTACGCCACCAGGGACATCATCGGCGACATCAGAAGGGCCTACTGCACC
ATCGAGACAGAGAGGTGGAAGGAGACACTGGGCAACGTCAAGAGAAAGCTGAAGGAGTAC
TTCCCCAACACCAACATCAGCTTTGCTCCCTCCTCTGGCGGGGACCTGGAGATCACCACC
CACAGCTTCAACTGCAGGGGCGAGTTCTTCTACAGAAACACCAGCAAGCTGTTCAACAAC
AACGACACCGAGAACAACCTGACCATCACCTGCCCTGCAGAATCAAGCAGATCGTGAAC

图 12

ATGTGGCAGGGCGTGGGCAGGGCCATGTACGCCCCACCCATCGAGGGCAACATCACCTGC
AAGTCCAACATCACCGCCTGCTGCTGACCAGAGACGGCGGCAACGGCACAGAGAGGAAC
GAGACATTCAGGCCAGGCGGCGGAGACATGAGGAACAACCTGGCGGAGCGAGCTGTACAAG
TACAAGGTGGTGGAGATCAAGCCCCCTGGGGATCGCCCCAACCGGCGCCAAGAGCAGAGTG
GTGGAGAGAGAGAAGAGCGCCGTCGGCATGGGCGCCGTGTTTCTGGGCTTCCTGGGAGCC
GCCGGAAGCACCATGGGAGCCGCCAGCCTGACCCTGACAGTGCAGGCCAGACAGGTGCTG
TCCGGGATCGTGCAGCAGCAGAGCAACCTGCTGAGAGCTATCGAGGCTCAGCAGCACATG
CTGCAGCTGACCGTGTGGGGCATCAAGCAGCTGCAGACCAGGGTGCTGGCCATCGAGAGA
TACCTGAAAGATCAGCAGCTGCTCGGCCTCTGGGGCTGCAGCGGCAAGCTGATCTGCACC
ACAGCCGTGCCATGGAATAGCAGCTGGTCCAACAAGAGCGAGATCGACATCTGGAACAAC
ATGACCTGGATGCAGTGGGACAGAGAGATCAGCAACTACACCAACACCATCTACAGACTG
CTGGAGGACAGCCAGACCCAGCAGGAGAAGAACGAGAAGGACCTGCTGGCCCTGGACAGC
TGAAGAACCTGTGGAACCTGGTTTAACATCACCAAGTGGCTGTGGTACATCAAGATCTTC
ATCATCATCGTGGGCGGCCTGATCGGCCTGAGGATCATCTTCGCCGTGCTGTCCATCGTG
AACAGAGTGCGCCAGGGGTACAGCCCACTGAGCTTCAGACCCTGATCCCCAGCCCCAGA
GGCCCCGACAAGCTGGGCAGGATCGAGGAGGAGGGCGGCGAGCAGGACAAGGGCAGATCC
GTGAGGCTGGTGAACGGCTTCCTGGCCCTGGCCTGGGACGACCTGAGAAGCCTGTGCCTG
TTCAGCTATCACCAGCTGAGGGACTTCATCCTGATCGTGGCCAGAGCCGTGGGCCTGCTG
GGCAGGTCCAGCCTGAGGGGCCTGCAGAGAGGCTGGGAGATCCTGAAGTACCTGGGCGGG
CTGGTGCAGTACTGGGGCCTGGAGCTGAAGAAGAGCGCCGTGAGCCTGTTCGACACAATC
GCCATCGCCGTGACCGAGGGCACCGACAGGATCATCGAGCTGATCCAGAGGTCCTGCAGA
GCCATCAGGAACGTGCCCACCAGGATCAGGCAGGGCTTCGAGGCCGCCCTGCAGTGA

图 12 续

HIV-1 C06980v1c3 gp160密码子优化核苷酸序列 (SEQ ID NO: 14):

突出显示tPa信号

ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTT
ACTACCACAGAGGCTAGC AACCTGTGGGTGACCGTGTAACGCGTGCCCGTGTGGAGA
GAGGCCAAGACCACCCTGTTCTGCGCCAGCGACGCCAAGGCCTACGAGAGGGAGGTGCAC
AACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCGACCCACAGGAAATCTTTCTG
GGCAAGAACGTGACCGAGAAGTTCAACATGTGGGAGAACTACATGGTGGACCAGATGCAC
GAGGACATCATCAGCCTGTGGGACCAGAGCCTGCAGCCCTGCGTGAAGCTGACCCCCCTG
TGCATCACCCCTGAACTGCACCGACTTCACCGCCCAACGGCAGCACCGTGTACGACAAC
AACGCCACCGCCAACAGCACCAACGAGATCAAGAACTGCAGCTTCAACATCATCTCCGAG
CTGAGGGACAAGAGGAAGAAGGAGAACGCCCTGTTCAACAACCTGGACATCGTGCAGCTG
GACGGCAACTCCAGCCTGTACAGACTGATCAACTGCAACACCAGCATCATCAAGCAGGCC
TGTCCAAGATCAGCTTCGACCCCATCCCCATCCACTACTGCGCCCCAGCCGGCTTCGTG
ATCCTGACCTGCAACAACGAGACATTCAACGGCACCGGCCCATGCAACAACGTGTCCGCT
GTGCAGTGTACCCACGGGATCAAGCCCGTGGTGTCCACCCAGCTGCTGCTGAACGGCAGC
CTGGCCAAGGGCGAGATCATGATCAGAAGCGAGAACATCACCAACAACGTGAAGACCATC
ATCGTGCACCTTCAACAAGAGCGTGGAGATCGTGTGCACCAGGCCCAACAACAACACCAGG
AAGAGCATCAGGATCGGCCAGGCCAGACCTTCTACGCCACCGGCGACATCATCGGCGAC
ATCCGGCAGGCCTACTGCAGCATCAACGAGAGCAACTGGAACATCACCTGCAGAGGGTG
TCCAAGAAGCTGGCCGAGCACTTCCCCAACAGAACCATCCAGTTCGAGAGCCCCCTCTGGC
GGCGACCTGGAGATCACCATGCACAGCTTCAACTGCAGGGGCGAGTTCTTCTACTGCAAC
ACATCCAAGCTGTTCAAGAGCACCTACCACCCCAACGGCACCTACAACCTGAACGGCACA
AACAGCACCCCTGATCATCCCATGCAGGATCAAGCAGATCATCAACATGTGGCAGGGCGTG

图 13

GGCAAGGCCATCTACGCCAGCCCAATCGCCGGCAGCATCACCTGCAGAAGCAACATCACC
GGCCTGCTGCTGACCAGGGACGGCGGCGACACCAACGACACCGAGATCTTCAGGCCCACC
GGCGGAGACATGAGGGACAACCTGGCGGAGCGAGCTGTACAAGTACAAGATCGTGGAGATC
AAGCCCCTGGGCGTGGCTCCAACCGAGGCCAAGAGAAGGGTGGTGAAGAGAGAGAAGAGG
GCCGTCACCATCGGCGCCGTGTTTCTGGGCTTCCTGGGAGCCGCCGGAAGCACCATGGGA
GCCGCCTCCATCACCTGACCGTGACGGCCAGGCAGCTGCTGTCCGGGATCGTGCAGCAG
CAGAGCAACCTGCTGCGGGCCATCGAAGCTCAGCAGCACATGCTGCAGCTGGCAGTCTGG
GGCATCAAGCAGCTGCAGGCCAGAGTGCTGGCCATCGAGAGATACCTGAAGGATCAGCAG
CTGCTCGGCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCACCGCCGTGCCCTGGAAC
AACAGCTGGTCCAACAGGACCCAGGACGAGATCTGGAAGAACCTGACCTGGATGGAGTGG
GACAGAGAGATCAGCAACTACACCAACACAATCTACGAGCTGCTGGAGGTGTCCCAGAGC
CAGCAGGAGAGGAACGAGAAGGACCTGCTGGCCCTGGACAGCTGGAACAACCTGTGGAAC
TGGTTCGACATCAGCAACTGGCTGTGGTACATCAAGATCTTCATCATGATCGTGGGCGGC
CTGATCGGCCTGAGGATCATCTTCGCCGTGCTGTCCATCGTGAACAGAGTGCGGCAGGGC
TACAGCCCCCTGAGCTTCCAGACCCTGATCCCCAACCAGAGAGAGCCCGACAGGCCCGGC
AGGATCGAGGAGGAGGGCGGCGAGCAGGACAAGGATAGATCTATCAGACTGGTGTCCGGC
TTCCTGGCCCTGGCCTGGGACGACCTGAGAAGCCTGTGCATCTTCCTGTACCACCACCTG
AGGGACCTGATCCTGATCGCCGCTAGAGCCACAGAGCTGCTGGGGAGGTCCAGCCTGAGG
GGCCTGCAGAGAGGCTGGGAGGCCCTGAAGTACCTGGGCAGCCTGGTGCAGTACTGGGGC
CTGGAGATCAAGAAGAGCGCCATCAACCTGCTGGACACAATCGCCATCGCCGTGGCCGAG
GGCACCGACAGGATCATCGAGATCGTGCAGAGGGCCTGTAGGGCCGTGCTGAACATCCTG
AGAAGGATCCGGCAGGGCCTGGAGGCTGCCCTGCAGTGA

图 13 续

HIV-1 C066980v1c3 gp160 DC密码子优化核苷酸序列 (SEQ ID NO:15):

突出显示tPa信号

gp120/gp140切割位点突变以阻止切割

```
ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTT
ACTACCACAGAGGCTAGC AACCTGTGGGTGACCGTGTACTACGGCGTGCCCGTGTGGAGA
GAGGCCAAGACCACCCTGTTCTGCGCCAGCGACGCCAAGGCCTACGAGAGGGAGGTGCAC
AACGTGTGGGCCACCCACGCCTGCGTGCCCAACCGACCCCGACCCACAGGAAATCTTTCTG
GGCAAGAACGTGACCGAGAAGTTCAACATGTGGGAGAACTACATGGTGGACCAGATGCAC
GAGGACATCATCAGCCTGTGGGACCAGAGCCTGCAGCCCTGCGTGAAGCTGACCCCCCTG
TGCATCACCTGAACTGCACCGACTTCACCGCCACAACGGCAGCACCGTGTACGACAAC
AACGCCACCGCCAACAGCACCAACGAGATCAAGAACTGCAGCTTCAACATCATCTCCGAG
CTGAGGGACAAGAGGAAGAAGGAGAACGCCCTGTTCAACAACCTGGACATCGTGCAGCTG
GACGGCAACTCCAGCCTGTACAGACTGATCAACTGCAACACCAGCATCATCAAGCAGGCC
TGTCCAAGATCAGCTTCGACCCCATCCCCATCCACTACTGCGCCCCAGCCGGCTTCGTG
ATCCTGACCTGCAACAACGAGACATTCAACGGCACCGGCCCATGCAACAACGTGTCCGCT
GTGCAGTGTAACCCACGGGATCAAGCCCGTGGTGTCCACCCAGCTGCTGCTGAACGGCAGC
CTGGCCAAGGGCGAGATCATGATCAGAAGCGAGAACATCACCAACAACGTGAAGACCATC
ATCGTGCACTTCAACAAGAGCGTGGAGATCGTGTGCACCAGGCCCAACAACAACACCAGG
AAGAGCATCAGGATCGGCCAGGCCAGACCTTCTACGCCACCGGCGACATCATCGGCGAC
ATCCGGCAGGCCTACTGCAGCATCAACGAGAGCAACTGGAACATCACCTGCAGAGGGTG
TCCAAGAAGCTGGCCGAGCACTTCCCCAACAGAACCATCCAGTTCGAGAGCCCCTCTGGC
GGCGACCTGGAGATCACCATGCACAGCTTCAACTGCAGGGGCGAGTTCTTCTACTGCAAC
ACATCCAAGCTGTTCAAGAGCACCTACCACCCCAACGGCACCTACAACCTGAACGGCACA
```

图 14

AACAGCACCTGATCATCCCATGCAGGATCAAGCAGATCATCAACATGTGGCAGGGCGTG
GGCAAGGCCATCTACGCCAGCCCAATCGCCGGCAGCATCACCTGCAGAAGCAACATCACC
GGCCTGCTGCTGACCAGGGACGGCGGCACACCAACGACACCGAGATCTTCAGGCCACC
GGCGGAGACATGAGGGACAACCTGGCGGAGCGAGCTGTACAAGTACAAGATCGTGGAGATC
AAGCCCCTGGGCGTGGCTCCAACCGAGGCCAAGAGCAGGGTGGTGAAGAGCGAGAAGAGC
GCCGTCACCATCGGCGCCGTGTTTCTGGGCTTCCTGGGAGCCGCCGGAAGCACCATGGGA
GCCGCTCCATCACCTGACCGTGCAGGCCAGGCAGCTGCTGTCCGGGATCGTGCAGCAG
CAGAGCAACCTGCTGCGGGCCATCGAAGCTCAGCAGCACATGCTGCAGCTGGCAGTCTGG
GGCATCAAGCAGCTGCAGGCCAGAGTGCTGGCCATCGAGAGATACCTGAAGGATCAGCAG
CTGCTCGGCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCACCGCCGTGCCCTGGAAC
AACAGCTGGTCCAACAGGACCCAGGACGAGATCTGGAAGAACCTGACCTGGATGGAGTGG
GACAGAGAGATCAGCAACTACACCAACACAATCTACGAGCTGCTGGAGGTGTCCAGAGC
CAGCAGGAGAGGAACGAGAAGGACCTGCTGGCCCTGGACAGCTGGAACAACCTGTGGAAC
TGGTTCGACATCAGCAACTGGCTGTGGTACATCAAGATCTTCATCATGATCGTGGGCGGC
CTGATCGGCCTGAGGATCATCTTCGCCGTGCTGTCCATCGTGAACAGAGTGCGGCAGGGC
TACAGCCCCCTGAGCTTCCAGACCCTGATCCCCAACCCAGAGAGAGCCCGACAGGCCCGGC
AGGATCGAGGAGGAGGGCGGCGAGCAGGACAAGGATAGATCTATCAGACTGGTGTCCGGC
TTCCTGGCCCTGGCCTGGGACGACCTGAGAAGCCTGTGCATCTTCCTGTACCACCACCTG
AGGGACCTGATCCTGATCGCCGCTAGAGCCACAGAGCTGCTGGGGAGGTCCAGCCTGAGG
GGCCTGCAGAGAGGCTGGGAGGCCCTGAAGTACCTGGGCAGCCTGGTGCAGTACTGGGGC
CTGGAGATCAAGAAGAGCGCCATCAACCTGCTGGACACAATCGCCATCGCCGTGGCCGAG
GGCACCGACAGGATCATCGAGATCGTGCAGAGGGCCTGTAGGGCCGTGCTGAACATCCTG
AGAAGGATCCGGCAGGGCCTGGAGGCTGCCCTGCAGTGATG

图 14 续

HIV-1 C06980v0c22 gp160密码子优化核苷酸序列 (SEQ ID NO:16):

突出显示tPa信号

ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTT
ACTACCACAGAGGCTAGC AACCTGTGGGTGACCGTGTACTACGGCGTGCCCGTGTGGAGA
GAGGCCAAGACCACCCTGTTCTGCGCCAGCGACGCCAAGGCCTACGAGCGGGAGGTGCAC
AACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCGACCCTCAGGAAATCTTCCTG
GGCAAGAACGTGACCGAGAAGTTCAACATGTGGAAGAACTACATGGTGGACCAGATGCAC
GAGGACATCATCAGCCTGTGGGACCAGAGCCTGCAGCCCTGCGTGAAGCTGACCCCCCTG
TGCATCACCCCTGAACTGCACCGACGTGACCGCCACAACGGCAGCACCGTGTACGACAAC
AACGCCACCGTGGTGAACAGCACCAACGAGATCAAGAACTGCAGCTTCAACATCACCAAC
GAGCTGAGGGACAAGAGGAAGAAGGAGCACGCCCTGTTCAACAACCTGGACATCGTGCAG
CTGGACGGCAACAGCTCCCTGTACAGACTGATCAACTGCAACACCAGCATCATCAAGCAG
GCCTGCCCCAAGATCAGCTTCGACCCCATCCCATCCACTACTGCGCCCCAGCCGGCTTC
GTGATCCTGAAGTGCAACAACGAGACATTCAACGGCACCGGCCCTGTAACAACGTGTCC
GCTGTGCAGTGCACCCACGGCATCAAGCCCGTGGTGTCCACCCAGCTGCTGCTGAACGGC
AGCCTGGCCAAGGGCGAGATCATGATCAGAAGCGAGAACATCACCGACAACGTGAAGACC
ATCATCGTGACCTGAACAACAGCGTGGAGATCGTGTGCACCAGGCCCAACAACAACACC
AGGAAGAGCATCAGGATCGGCCAGGCCAGACCTTCTACGCCACCGGCGACATCATCGGC
GACATCCGGCAGGCCTACTGCAGCATCAACGAGAGCAACTGGAACGCCACCCTGCAGAGG
GTGTCCAAGAAGCTGGCCGAGCACTTCCCCAACAAGACCATCCAGTTCAAGAGCCCCCTCT
GGCGGCGACCTGGAGATCACCATGCACAGCTTCAACTGCAGGGGCGAGTTCTTCTACTGC
AACACATCCAAGCTGTTTAACGGGACCTACTACCCAAACGGCACATACTACCCAAACGGG
ACCAACAGCACCCCTGATCATCCCATGCAGGATCAAGCAGATCATCAACATGTGGCAGGGC
GTGGGCAAGGCCATCTACGCCAGCCCAATCGCCGGCAACATCACCTGCCGGTCCAACATC

图 15

ACCGGCCTGCTGCTGACCAGGGACGGCGGCGACACCAACGACACCGAGATCTTCAGGCCA
GCCGGCGGAGACATGAGGGACAACCTGGCGGAGCGAGCTGTACAAGTACAAGATCGTGGAG
ATCAAGCCCCTGGGCGTGGCTCCAACCGAGGCCAAGAGAAGGGTGGTGAAGAGAGAGAAG
AGGGCCGTCACCATCGGCGCCGTGTTTCTGGGCTTCCTGGGAGCCGCCGAAGCACCATG
GGAGCCGCCTCCATCACCTGACCGTGCAGGCCAGGCAGCTGCTGTCCGGGATCGTGCAG
CAGCAGAGCAACCTGCTGAGAGCCATTGAGGCTCAGCAGCACATGCTGCAGCTGACAGTG
TGGGGCATCAAGCAGCTGCAGGCCAGAGTGCTGGCCATCGAGAGATACCTGAAGGATCAG
CAGCTGCTCGGCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCACCGCCGTGCCCTGG
AACAACTCCTGGTCCAACAGGACCCAGGACGAGATCTGGAAGAACCTGACCTGGATGGAG
TGGGACAGAGAGATCAGCAACTACACCAACACCATCTACGAGCTGCTGGAGGTGTCCAG
AGCCAGCAGGAGAGGAACGAGAAGGACCTGCTGGCCCTGGATAGCTGGAACAACCTGTGG
AACTGGTTCGACATCAGCAACTGGCTGTGGTACATCAAGATCTTCATCATGATCGTGGGC
GGCCTGATCGGCCTGAGGATCATCTTCGCCGTGCTGTCCATCGTGAACAGAGTGCGGCAG
GGCTACAGCCCCCTGAGCTTCCAGACCCTGATCCCCAACCAGAGAGAGCCCGACAGGCCC
GGCAGGATCGAGGAGGAGGGCGGCGAGCAGGACAAGGACAGATCCATCAGGCTGGTGTCC
GGCTTCCTGGCCCTGGCCTGGGACGACCTGAGAAGCCTGTGCATCTTCCTGTACCACCAC
CTGAGGGACTTCATCCTGATCGCCGCCAGGGCCACAGAGCTGCTGGGGAGGTCCAGCCTG
AGGGGCCTGCAGAGAGGCTGGGAGGCCCTGAAGTACCTGGGCAGCCTGGTGCAGTACTGG
GGCCTGGAGATCAAGAAGAGCGCCATCAACCTGCTGGACACAATCGCCATCGCCGTGGCC
GAGGGCACCGACAGGATCATCGAGATCGTGCAGAGGGCCTGTAGGGCCGTGCTGAACATC
CCCAGAAGGATCAGACAGGGGCTGGAGGCTGCCCTGCAGTGA

图 15 续

HIV-1 C06980v0c22 DC gp160密码子优化核苷酸序列 (SEQ ID NO:17):

突出显示tPa信号

gp120/gp140切割位点突变以阻止切割

```
ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTT
ACTACCACAGAGGCTAGC AACCTGTGGGTGACCGTGTA CTACGGCGTGCCCGTGTGGAGA
GAGGCCAAGACCACCCTGTTCTGCGCCAGCGACGCCAAGGCCTACGAGCGGGAGGTGCAC
AACGTGTGGGCCACCCACGCCTGCGTGCCCAACCGACCCCGACCCTCAGGAAATCTTCCTG
GGCAAGAACGTGACCGAGAAGTTCAACATGTGGAAGAACTACATGGTGGACCAGATGCAC
GAGGACATCATCAGCCTGTGGGACCAGAGCCTGCAGCCCTGCGTGAAGCTGACCCCCCTG
TGCATCACCCCTGAACTGCACCGACGTGACCGCCACAACGGCAGCACCGTGTACGACAAC
AACGCCACCGTGGTGAACAGCACCAACGAGATCAAGAACTGCAGCTTCAACATCACCAAC
GAGCTGAGGGACAAGAGGAAGAAGGAGCACGCCCTGTTCAACAACCTGGACATCGTGCAG
CTGGACGGCAACAGCTCCCTGTACAGACTGATCAACTGCAACACCAGCATCATCAAGCAG
GCCTGCCCCAAGATCAGCTTCGACCCCATCCCCATCCACTACTGCGCCCCAGCCGGCTTC
GTGATCCTGAAGTGCAACAACGAGACATTCAACGGCACCGGCCCTGTAAACAACGTGTCC
GCTGTGCAGTGCACCACGGCATCAAGCCCGTGGTGTCCACCCAGCTGCTGCTGAACGGC
AGCCTGGCCAAGGGCGAGATCATGATCAGAAGCGAGAACATCACCGACAACGTGAAGACC
ATCATCGTGACCTGAACAACAGCGTGGAGATCGTGTGCACCAGGCCCAACAACAACACC
AGGAAGAGCATCAGGATCGGCCAGGCCAGACCTTCTACGCCACCGGCGACATCATCGGC
GACATCCGGCAGGCCTACTGCAGCATCAACGAGAGCAACTGGAACGCCACCCTGCAGAGG
GTGTCCAAGAAGCTGGCCGAGCACTTCCCCAACAAGACCATCCAGTTCAAGAGCCCCCTCT
GGCGGCGACCTGGAGATCACCATGCACAGCTTCAACTGCAGGGGCGAGTTCTTCTACTGC
```

图 16

AACACATCCAAGCTGTTTAACGGGACCTACTACCCAAACGGCACATACTACCCAAACGGG
ACCAACAGCACCCCTGATCATCCCATGCAGGATCAAGCAGATCATCAACATGTGGCAGGGC
GTGGGCAAGGCCATCTACGCCAGCCCAATCGCCGGCAACATCACCTGCCGGTCCAACATC
ACCGGCCTGCTGCTGACCAGGGACGGCGGCGACACCAACGACACCGAGATCTTCAGGCCA
GCCGGCGGAGACATGAGGGACAACCTGGCGGAGCGAGCTGTACAAGTACAAGATCGTGGAG
ATCAAGCCCCTGGGCGTGGCTCCAACCGAGGCCAAGAGCAGGGTGGTGAAGAGCGAGAAG
AGCGCCGTCACCATCGGCGCCGTGTTTCTGGGCTTCCTGGGAGCCGCCGGAAGCACCATG
GGAGCCGCCTCCATCACCCCTGACCGTGCAGGCCAGGCAGCTGCTGTCCGGGATCGTGCAG
CAGCAGAGCAACCTGCTGAGAGCCATTGAGGCTCAGCAGCACATGCTGCAGCTGACAGTG
TGGGGCATCAAGCAGCTGCAGGCCAGAGTGCTGGCCATCGAGAGATACCTGAAGGATCAG
CAGCTGCTCGGCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCACCGCCGTGCCCTGG
AACAACTCCTGGTCCAACAGGACCCAGGACGAGATCTGGAAGAACCTGACCTGGATGGAG
TGGGACAGAGAGATCAGCAACTACACCAACACCATCTACGAGCTGCTGGAGGTGTCCCAG
AGCCAGCAGGAGAGGAACGAGAAGGACCTGCTGGCCCTGGATAGCTGGAACAACCTGTGG
AACTGGTTCGACATCAGCAACTGGCTGTGGTACATCAAGATCTTCATCATGATCGTGGGC
GGCCTGATCGGCCTGAGGATCATCTTCGCCGTGCTGTCCATCGTGAACAGAGTGCGGCAG
GGCTACAGCCCCCTGAGCTTCCAGACCCTGATCCCCAACCAGAGAGAGCCCCGACAGGCCC
GGCAGGATCGAGGAGGAGGGCGGCGAGCAGGACAAGGACAGATCCATCAGGCTGGTGTCC
GGCTTCCTGGCCCTGGCCTGGGACGACCTGAGAAGCCTGTGCATCTTCCTGTACCACCAC
CTGAGGGACTTCATCCTGATCGCCGCCAGGGCCACAGAGCTGCTGGGGAGGTCCAGCCTG
AGGGGCCTGCAGAGAGGCTGGGAGGCCCTGAAGTACCTGGGCAGCCTGGTGCAGTACTGG
GGCCTGGAGATCAAGAAGAGCGCCATCAACCTGCTGGACACAATCGCCATCGCCGTGGCC
GAGGGCACCGACAGGATCATCGAGATCGTGAGAGGGCCTGTAGGGCCGTGCTGAACATC
CCCAGAAGGATCAGACAGGGGCTGGAGGCTGCCCTGCAGTGA

图 16 续

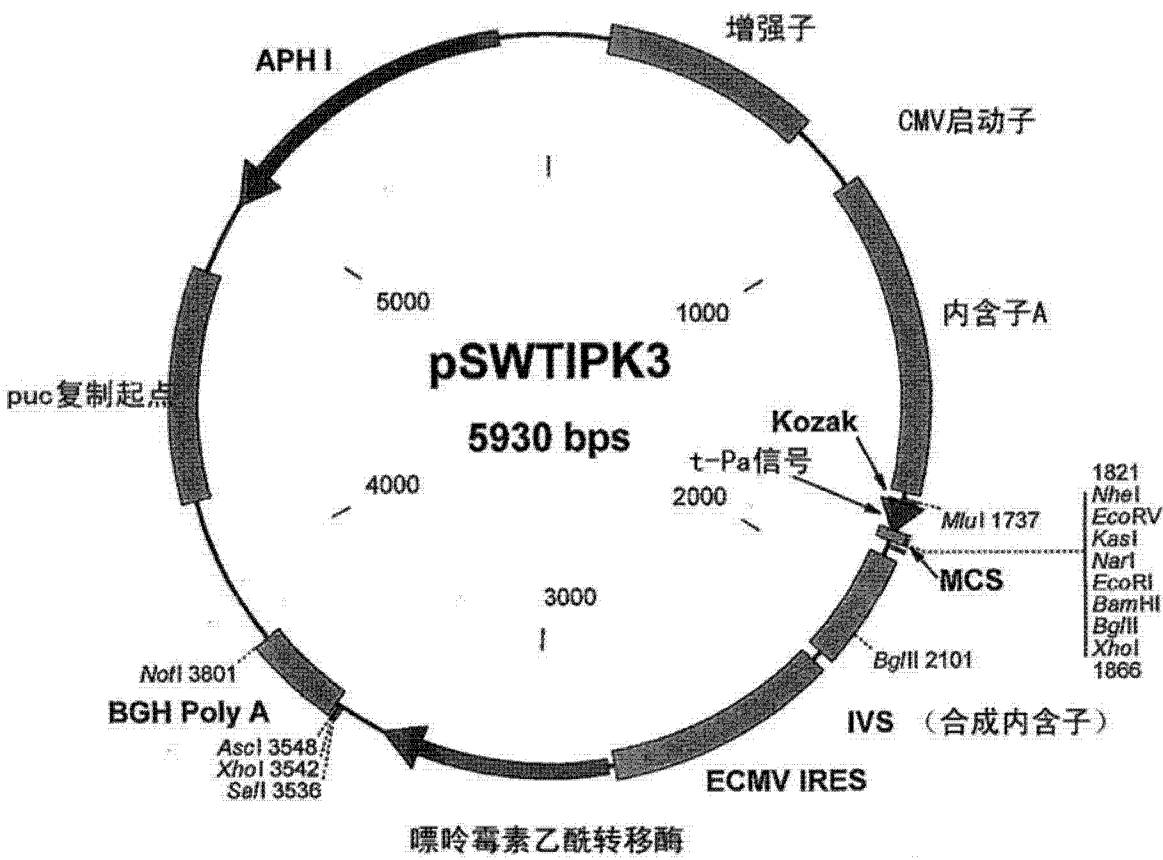


图 17

pSWTIPK3序列 (SEQ ID NO:18):

TGGCACATGGCCAATGCATATCGATCTATACATTGAATCAATATTGGCAATTAGCCATAT
TAGTCATTGGTTATATAGCATAAAATCAATATTGGCTATTGGCCATTGCATACGTTGTATC
TATATCATAATATGTACATTTATATTGGCTCATGTCCAATATGACCGCCATGTTGACATT
GATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATA
TGGAGTTCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACC
CCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC
ATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGT
ATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATT
ATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCA
TCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTTG
ACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACC
AAAATCAACGGGACTTTCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCG
GTAGGCGTGACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCG
CCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGCC
TCCGCGGCCGGGAACGGTGCAATTGGAACGCGGATTCCCCGTGCCAAGAGTGACGTAAGTA
CCGCCTATAGACTCTATAGGCACACCCCTTTGGCTCTTATGCATGCTATACTGTTTTTG
CTTGGGGCCTATACACCCCGCTTCCTTATGCTATAGGTGATGGTATAGCTTAGCCTATA
GGTGTGGGTTATTGACCATTATTGACCACTCCCCTATTGGTGACGATACTTTCATTACT
AATCCATAACATGGCTCTTTGCCACAACATCTCTATTGGCTATATGCCAATACTCTGTC
CTTCAGAGACTGACACGGACTCTGTATTTTACAGGATGGGGTCCCATTTATTATTACA
AATTCACATATACAACAACGCCGTCCCCGTGCCCGCAGTTTTTATTAAACATAGCGTGG
GATCTCCACGCGAATCTCGGGTACGTGTTCCGGACATGGGCTCTTCTCCGGTAGCGGCGG

图 18

AGCTTCCACATCCGAGCCCTGGTCCCATGCCTCCAGCGGCTCATGGTCGCTCGGCAGCTC
CTTGCTCCTAACAGTGGAGGCCAGACTTAGGCACAGCACAATGCCACCACCACCAGTGT
GCCGCACAAGGCCGTGGCGGTAGGGTATGTGTCTGAAAATGAGCTCGGAGATTGGGCTCG
CACCGCTGACGCAGATGGAAGACTTAAGGCAGCGGCAGAAGAAGATGCAGGCAGCTGAGT
TGTTGTATTCTGATAAGAGTCAGAGGTAACCTCCCGTTGCGGTGCTGTTAACGGTGGAGGG
CAGTGTAGTCTGAGCAGTACTCGTTGCTGCCGCGCGGCCACCAGACATAATAGCTGACA
GACTAACAGACTGTTCTTTCCATGGGTCTTTTCTGCAGTCACCGTCCATTTAAATACGC
GTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAG
TCTTCGTTACTACCACAGAGGCTAGCGGGGATATCGGCGCCGAATTCGGATCCACTAGTA
GATCTCTCGAGAACGGCCGCCAGTGTGCTGGAATTAATTCGCTGTCTGCGAGGGCCAGCT
GTTGGGGTGAGTACTCCCTCTCAAAGCGGGCATGACTTCTGCGCTAAGATTGTCAGTTT
CCAAAAACGAGGAGGATTTGATATTCACCTGGCCCCGCGGTGATGCCTTTGAGGGTGGCCG
CGTCCATCTGGTCAGAAAAGACAATCTTTTTGTTGTCAAGCTTGAGGTGTGGCAGGCTTG
AGATCTGGCCATACACTTGAGTGACAATGACATCCACTTTGCCTTTCTCTCCACAGGTGT
CCACTCCCAGGTCCAAGTGCAGGTGAGCATGCATCTAGGGCGGCCAATTCGCCCCCTCT
CCCTCCCCCCCCCTAACGTTACTGGCCGAAGCCGCTTGGAATAAGGCCGGTGTGCGTTT
GTCTATATGTTATTTCCACCATATTGCCGTCTTTTGGAATGTGAGGGCCCCGAAACCT
GGCCCTGTCTTCTTGACGAGCATTCTAGGGGTCTTCCCCTCTCGCCAAAGGAATGCAA
GGTCTGTTGAATGTCGTGAAGGAAGCAGTTCCTCTGGAAGCTTCTTGAAGACAAACAACG
TCTGTAGCGACCTTTGCAGGCAGCGGAACCCCCACCTGGCGACAGGTGCCTCTGCGGC
CAAAAGCCACGTGTATAAGATACACCTGCAAAGGCGGCACAACCCCAGTGCCACGTTGTG
AGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCTCCTCAAGCGTATTCAACAAGGGGCTG
AAGGATGCCCGAGAAGGTACCCCATTTGTATGGGATCTGATCTGGGGCCTCGGTGCACATGC
TTTACATGTGTTTAGTCGAGGTTAAAAAACGTCTAGGCCCCCGAACCACGGGGACGTG

图 18 续-1

GTTTTCTTTGAAAAACACGATGATAAGCTTGCCACAACCCACAAGGAGACGACCTTCCA
TGACCGAGTACAAGCCACGGTGCGCCTCGCCACCCGCGACGACGTCCCCGGGCCGTAC
GCACCCTCGCCGCCGCGTTGCGCGACTACCCCGCCACGCGCCACACCGTGGACCCGGACC
GCCACATCGAGCGGGTCACCGAGCTGCAAGAACTCTTCTCACGCGCGTCGGGCTCGACA
TCGGCAAGGTGTGGGTCGCGGACGACGGCGCTGCGGTGGCGGTCTGGACCACGCCGGAGA
GCGTCGAAGCGGGGGCGGTGTTGCGCGAGATCGGCCCCGCGCATGGCCGAGTTGAGCGGTT
CCCGGCTGGCCGCGCAGCAACAGATGGAAGGCCTCTGGCTCCGCACCGGCCCAAGGAGC
CCGCGTGGTTCTGGCCACCGTCGGCGTCTGCCCCGACCACCAGGGCAAGGGTCTGGGCA
GCGCCGTCGTGCTCCCCGGAGTGGAGGCGGCCGAGCGCGCCGGGGTGCCCGCCTTCTGG
AGACCTCCGCGCCCCGCAACCTCCCCTTCTACGAGCGGCTCGGCTTCACCGTCACCGCCG
ACGTCGAGGTGCCCCAAGGACCGCGCACCTGGTGCATGACCCGCAAGCCCGGTGCCTGAC
GCCCCCCCCACGACCCGCAGCGCCCGACCGAAAGGAGCGCACGACCCCATGGCTCCGACC
GAAGCCACCCGGGGCGGGCCCCGCGACCCCGACCCGCCCCGAGGCCACCGACGTCGA
CCTCGAGGGCGCGCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTG
TTTGCCCTCCCCCGTGCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCT
AATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTG
GGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATG
CGGTGGGCTCTATGGCTTCTGCGGCCGAGCTTGCGGTAATCATGGTCATAGCTGTTTCC
TGTGTGAAATTGTTATCCGCTCACAATCCACACAACATACGAGCCGGAAGCATAAAGTG
TAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCC
CGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGG
GAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTC
GGTCGTTGCGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCAC

图 18 续 -2

AGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAA
CCGTAAAAAGGCCGCGTTGCTGGCGTTTTTTCATAGGCTCCGCCCCCTGACGAGCATCA
CAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGC
GTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATA
CCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTA
TCTCAGTTCGGTGTAGGTCGTTGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTCA
GCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGA
CTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGG
TGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGG
TATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGG
CAAACAAACCACCGCTGGTAGCGGTGGTTTTTTGTTTGCAAGCAGCAGATTACGCGCAG
AAAAAAGGATCTCAAGAAGATCCTTTGATCTTTCTACGGGGTCTGACGCTCAGTGGAA
CGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGAT
CCTTTTAAATTAATAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTC
TGACAGTTAGAAAACTCATCGAGCATCAAATGAACTGCAATTTATTCATATCAGGATT
ATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCA
GTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAAT
ACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGT
GACGACTGAATCCGGTGAGAATGGCAAAAGCTTATGCATTTCTTCCAGACTTGTTCAAC
AGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCTG
TGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTAAAAGGACAATTACAAACAGG
AATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTACCTGAATC
AGGATATTCTTCTAATACCTGGAATGCTGTTTTCCGGGGATCGCAGTGGTGAGTAACCA

图 18 续 -3

TGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATCCGTCAG
CCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTT
CAGAAACAACCTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTG
CCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAA
TCGCGGCCTTGAGCAAGACGTTTCCCGTTGAATATGGCTCATACTCTTCCTTTTCAATA
TTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTGAATGTATTTA
GAAAAATAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC

图 18 续 -4

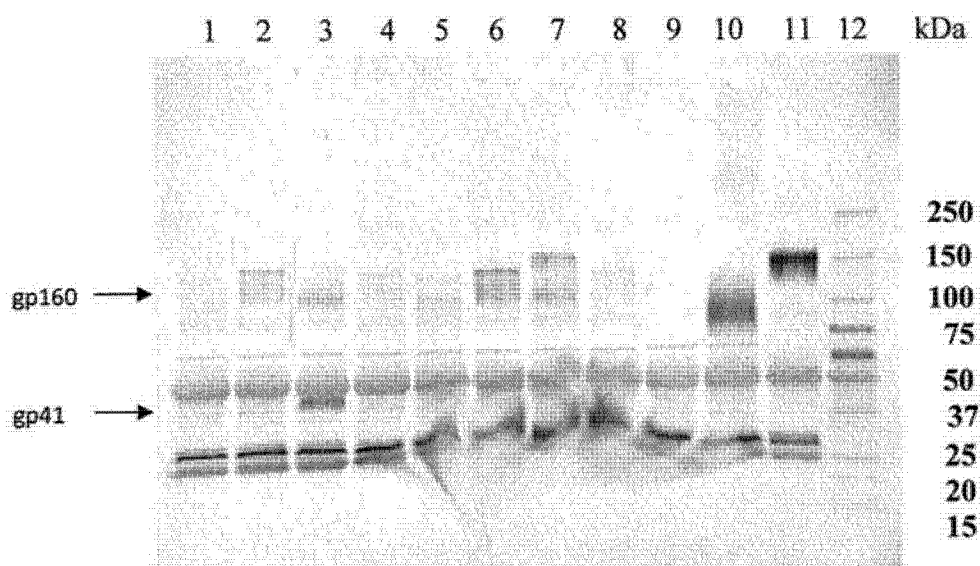


图 19

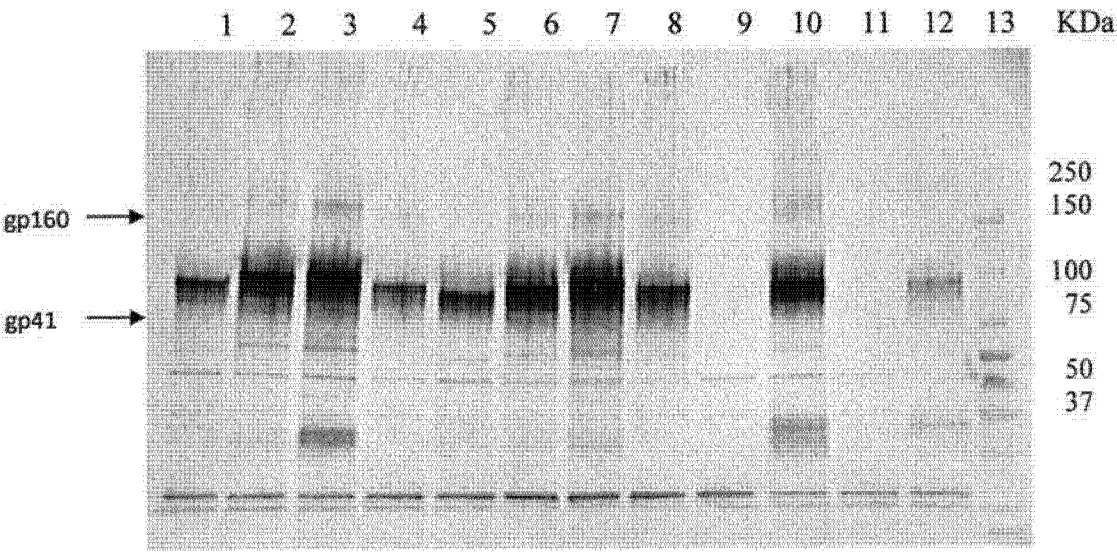


图 20

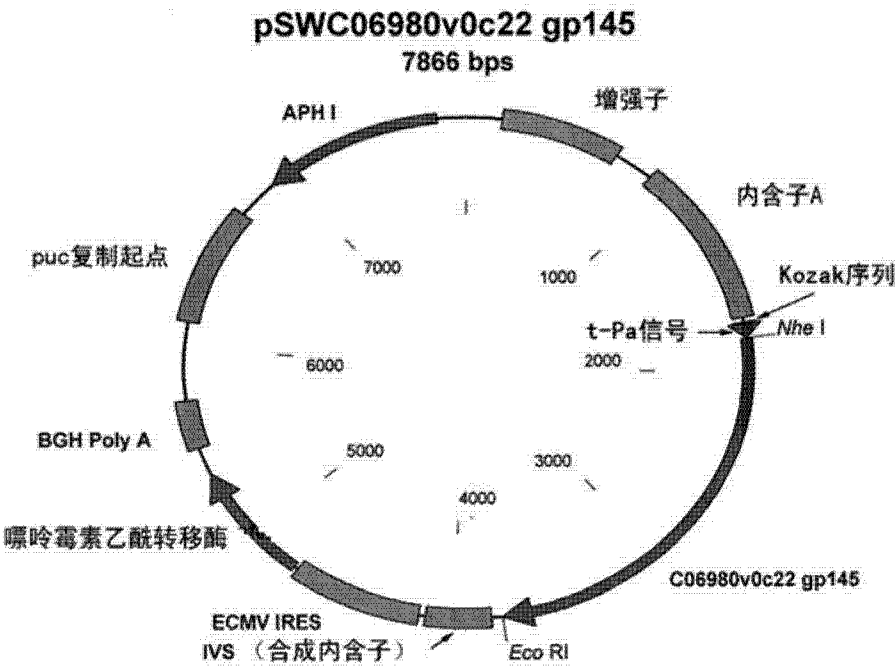


图 21

pSWC06980v0c22 gp145序列 (SEQ ID NO: 19)

TGGCACATGGCCAATGCATATCGATCTATACATTGAATCAATATTGGCAATTAGCCATAT
TAGTCATTGGTTATATAGCATAAATCAATATTGGCTATTGGCCATTGCATACGTTGTATC
TATATCATAATATGTACATTTATATTGGCTCATGTCCAATATGACCGCCATGTTGACATT
GATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATA
TGGAGTTCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACC
CCCGCCCATGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC
ATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGT
ATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATT
ATGCCCAGTACATGACCTTACGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCA
TCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTG
ACTCACGGGGATTTCCAAGTCTCCACCCCATGACGTCAATGGGAGTTTGTTTTGGCACC
AAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCG
GTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCG
CCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGCC
TCCGCGGCCGGAACGGTGCATTGGAACGCGGATTCCCCGTGCCAAGAGTGACGTAAGTA
CCGCTATAGACTCTATAGGCACACCCCTTTGGCTCTTATGCATGCTATACTGTTTTTG
CTTGGGGCCTATACACCCCGCTTCCTTATGCTATAGGTGATGGTATAGCTTAGCCTATA
GGTGTGGGTATTGACCATTATTGACCACTCCCCTATTGGTGACGATACTTTCCATTACT
AATCCATAACATGGCTCTTTGCCACAACCTATCTCTATTGGCTATATGCCAATACTCTGTC
CTTCAGAGACTGACACGGACTCTGTATTTTACAGGATGGGGTCCCATTTATTATTACA
AATTCACATATACAACAACGCCGTCCCCCGTGCCCGCAGTTTTTTATTAACATAGCGTGG
GATCTCCACGCGAATCTCGGGTACGTGTTCCGGACATGGGCTCTTCTCCGGTAGCGGCGG

图 22

AGCTTCCACATCCGAGCCCTGGTCCCATGCCTCCAGCGGCTCATGGTCGCTCGGCAGCTC
CTTGCTCCTAACAGTGGAGGCCAGACTTAGGCACAGCACAATGCCACCACCACCAGTGT
GCCGCACAAGGCCGTGGCGGTAGGGTATGTGTCTGAAAATGAGCTCGGAGATTGGGCTCG
CACCGCTGACGCAGATGGAAGACTTAAGGCAGCGGCAGAAGAAGATGCAGGCAGCTGAGT
TGTTGTATTCTGATAAGAGTCAGAGGTAACCTCCCGTTGCGGTGCTGTTAACGGTGGAGGG
CAGTGTAGTCTGAGCAGTACTCGTTGCTGCCGCGCGGCCACCAGACATAATAGCTGACA
GACTAACAGACTGTTCTTTCCATGGGTCTTTTCTGCAGTCACCGTCCATTAAATACGC
GTGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAG
TCTTCGTTACTACCACAGAGGCTAGCAACCTGTGGGTGACCGTGTAACGGCGTGCCCG
TGTGGAGAGAGGCCAAGACCACCCTGTTCTGCGCCAGCGACGCCAAGGCCTACGAGCGGG
AGGTGCACAACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCGACCCTCAGGAAA
TCTTCCTGGGCAAGAACGTGACCGAGAAGTTCAACATGTGGAAGAACTACATGGTGGACC
AGATGCACGAGGACATCATCAGCCTGTGGGACCAGAGCCTGCAGCCCTGCGTGAAGCTGA
CCCCCTGTGCATCACCTGAACTGCACCGACGTGACCGCCCAACGGCAGCACCGTGT
ACGACAACAACGCCACCGTGGTGAACAGCACCAACGAGATCAAGAACTGCAGCTTCAACA
TCACCACCGAGCTGAGGGACAAGAGGAAGAAGGAGCACGCCCTGTTCAACAACCTGGACA
TCGTGCAGCTGGACGGCAACAGCTCCCTGTACAGACTGATCAACTGCAACACCAGCATCA
TCAAGCAGGCCTGCCCCAAGATCAGCTTCGACCCCATCCCATCCACTACTGCGCCCCAG
CCGGCTTCGTGATCCTGAAGTGCAACAACGAGACATTCAACGGCACCGGCCCTGTAACA
ACGTGTCCGCTGTGCAGTGACCCACGGCATCAAGCCGTGGTGTCCACCCAGCTGCTGC
TGAACGGCAGCCTGGCCAAGGGCGAGATCATGATCAGAAGCGAGAACATCACCGACAACG
TGAAGACCATCATCGTGACCTGAACAACAGCGTGGAGATCGTGTGCACCAGGCCCAACA
ACAACACCAGGAAGAGCATCAGGATCGGCCAGGCCAGACCTTCTACGCCACCGGCGACA
TCATCGGCGACATCCGGCAGGCCTACTGCAGCATCAACGAGAGCAACTGGAACGCCACCC

图 22 续 -1

TGCAGAGGGTGTCCAAGAAGCTGGCCGAGCACTTCCCCAACAAGACCATCCAGTTCAAGA
GCCCCCTCTGGCGGCGACCTGGAGATCACCATGCACAGCTTCAACTGCAGGGGCGAGTTCT
TCTACTGCAACACATCCAAGCTGTTTAACGGGACCTACTACCCAAACGGCACATACTACC
CAAACGGGACCAACAGCACCCCTGATCATCCCATGCAGGATCAAGCAGATCATCAACATGT
GGCAGGGCGTGGGCAAGGCCATCTACGCCAGCCCAATCGCCGGCAACATCACCTGCCGGT
CCAACATCACCGGCCTGCTGCTGACCAGGGACGGCGGCGACACCAACGACACCGAGATCT
TCAGGCCAGCCGGCGGAGACATGAGGGACAACCTGGCGGAGCGAGCTGTACAAGTACAAGA
TCGTGGAGATCAAGCCCCTGGGCGTGGCTCCAACCGAGGCCAAGAGCAGGGTGTTGAAGA
GCGAGAAGAGCGCCGTCACCATCGGCGCCGTGTTTCTGGGCTTCTGGGAGCCGCCGAA
GCACCATGGGAGCCGCCTCCATCACCTGACCGTGCAGGCCAGGCAGCTGCTGTCCGGGA
TCGTGCAGCAGCAGAGCAACCTGCTGAGAGCCATTGAGGCTCAGCAGCACATGCTGCAGC
TGACAGTGTGGGGCATCAAGCAGCTGCAGGCCAGAGTGCTGGCCATCGAGAGATACCTGA
AGGATCAGCAGCTGCTCGGCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCACCGCCG
TGCCCTGGAACAACTCCTGGTCCAACAGGACCCAGGACGAGATCTGGAAGAACCTGACCT
GGATGGAGTGGGACAGAGAGATCAGCAACTACACCAACACCATCTACGAGCTGCTGGAGG
TGTCCCAGAGCCAGCAGGAGAGGAACGAGAAGGACCTGCTGGCCCTGGATAGCTGGAACA
ACCTGTGGAACCTGGTTCGACATCAGCAACTGGCTGTGGTACATCAAGAAGAAGAAGTGAA
TTCGGATCCACTAGTAGATCTCTCGAGAACGGCCGCCAGTGTGCTGGAATTAATTCGCTG
TCTGCGAGGGCCAGCTGTTGGGGTGAGTACTCCCTCTCAAAGCGGGCATGACTTCTGCG
CTAAGATTGTCAGTTTCAAAAAACGAGGAGGATTTGATATTCACCTGGCCCCGCGGTGATG
CCTTTGAGGGTGGCCGCGTCCATCTGGTCAGAAAAGACAATCTTTTGTGTCAAGCTTG
AGGTGTGGCAGGCTTGAGATCTGGCCATACACTTGAGTGACAATGACATCCACTTTGCCT
TTCTCTCCACAGGTGTCCACTCCCAGGTCCAACCTGCAGGTCGAGCATGCATCTAGGGCGG

图 22 续 -2

CCAATTCGCCCCTCTCCCTCCCCCCCCCTAACGTTACTGGCCGAAGCCGCTTGGAATA
AGGCCGGTGTGCGTTTGTCTATATGTTATTTCCACCATATTGCCGTCTTTTGGCAATGT
GAGGGCCCGGAAACCTGGCCCTGTCTTCTTGACGAGCATTCTAGGGGTCTTCCCCTCT
CGCCAAAGGAATGCAAGGTCTGTTGAATGTCGTGAAGGAAGCAGTTCCTCTGGAAGCTTC
TTGAAGACAAACAACGTCTGTAGCGACCCTTTCAGGCAGCGGAACCCCCACCTGGCGA
CAGGTGCCTCTGCGGCCAAAAGCCACGTGTATAAGATACACCTGCAAAGGCGGCACAACC
CCAGTGCCACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCTCCTCAAGCGT
ATTCAACAAGGGGCTGAAGGATGCCAGAAGGTACCCATTGTATGGGATCTGATCTGGG
GCCTCGGTGCACATGCTTTACATGTGTTTAGTCGAGGTTAAAAAACGTCTAGGCCCCC
GAACCACGGGGACGTGGTTTTCTTTGAAAAACACGATGATAAGCTTGCCACAACCCACA
AGGAGACGACCTTCATGACCGAGTACAAGCCACGGTGCGCCTCGCCACCCGCGACGAC
GTCCCCGGGGCGTACGCACCCTCGCCGCCGCGTTCGCCGACTACCCCGCCACGCGCCAC
ACCGTGGACCCGGACCGCCACATCGAGCGGGTCACCGAGCTGCAAGAACTTTCCTCACG
CGCGTCGGGCTCGACATCGGCAAGGTGTGGGTCGCGGACGACGGCGCTGCGGTGGCGGTC
TGGACCACGCCGGAGAGCGTCGAAGCGGGGGCGGTGTTGCGCGAGATCGGCCCCGCGCATG
GCCGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGAAGGCCTCCTGGCTCCG
CACCGGCCCAAGGAGCCCGCGTGGTTCCTGGCCACCGTCGGCGTCTCGCCCGACCACCAG
GGCAAGGGTCTGGGCAGCGCCGTCGTGCTCCCGGAGTGGAAGGCGGCCGAGCGCGCCGGG
GTGCCCCGCTTCTGGAGACCTCCGCGCCCCGCAACCTCCCTTCTACGAGCGGCTCGGC
TTCACCGTCACCGCCGACGTCGAGGTGCCGAAGGACCGCGCACCTGGTGCATGACCCGC
AAGCCCGGTGCCTGACGCCCCGCCACGACCCGCGAGCGCCCGACCGAAAGGAGCGCACGA
CCCCATGGCTCCGACCGAAGCCACCCGGGGCGGCCCGCCGACCCCGCACCCGCCCCGA
GGCCACCGACGTCGACCTCGAGGGCGCGCCGCTGATCAGCCTCGACTGTGCCTTCTAGT

图 22 续 -3

TGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACT
CCCCTGTCTTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCAT
TCTATTCTGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGC
AGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGCGGCCGACGCTTGGCGTAATCAT
GGTCATAGCTGTTTCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAG
CCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTG
CGTTGCGCTCACTGCCCCGCTTTCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAA
TCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTCCGCTTCTCGCTCA
CTGACTCGCTGCGCTCGGTCGTTGCGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGG
TAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCC
AGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCC
CCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGAC
TATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCC
TGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATA
GCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTGCTCCAAGCTGGGCTGTGTGC
ACGAACCCCCGTTAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCA
ACCCGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAG
CGAGGTATGTAGGCGGTGCTACAGAGTTCCTGAAGTGGTGGCCTAACTACGGCTACACTA
GAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTG
GTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTGTTTGCAAGC
AGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGT
CTGACGCTCAGTGGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAA
GGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATAT

图 22 续 -4

ATGAGTAAACTTGGTCTGACAGTTAGAAAACTCATCGAGCATCAAATGAACTGCAATT
TATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAG
AAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGA
CTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAAATAAGGTTATCAAGTG
AGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGCTTATGCATTTCTT
TCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCA
AACCGTTATTCATTTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAG
GACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAA
TATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCG
CAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAG
GCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGC
TACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGA
TTGTCGCACCTGATTGCCCAGATTATCGCGAGCCATTTATACCCATATAAATCAGCAT
CCATGTTGGAATTTAATCGCGGCCTTGAGCAAGACGTTTCCCGTTGAATATGGCTCATAC
TCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACA
TATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCGCGCACATTTCCCCGAAAAG
TGCCAC

图 22 续 -5

HIV-1 C06980v0c22 gp145密码子优化核苷酸序列 (SEQ ID NO:20):

突出显示tPa信号

ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTT
ACTACCACAGAGGCTAGC AACCTGTGGGTGACCGTGACTACGGCGTGCCCGTGTGGAGA
GAGGCCAAGACCACCCTGTTCTGCGCCAGCGACGCCAAGGCCTACGAGCGGGAGGTGCAC
AACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCGACCCTCAGGAAATCTTCCTG
GGCAAGAACGTGACCGAGAAGTTCAACATGTGGAAGAACTACATGGTGGACCAGATGCAC
GAGGACATCATCAGCCTGTGGGACCAGAGCCTGCAGCCCTGCGTGAAGCTGACCCCCCTG
TGCATCACCCCTGAACTGCACCGACGTGACCGCCACAACGGCAGCACCGTGTACGACAAC
AACGCCACCGTGGTGAACAGCACCAACGAGATCAAGAACTGCAGCTTCAACATCACCACC
GAGCTGAGGGACAAGAGGAAGAAGGAGCACGCCCTGTTCAACAACCTGGACATCGTGCAG
CTGGACGGCAACAGCTCCCTGTACAGACTGATCAACTGCAACACCAGCATCATCAAGCAG
GCCTGCCCCAAGATCAGCTTCGACCCCATCCCCATCCACTACTGCGCCCCAGCCGGCTTC
GTGATCCTGAAGTGCAACAACGAGACATTCAACGGCACCGGCCCTGTAACAACGTGTCC
GCTGTGCAGTGCACCCACGGCATCAAGCCCGTGGTGTCCACCCAGCTGCTGCTGAACGGC
AGCCTGGCCAAGGGCGAGATCATGATCAGAAGCGAGAACATCACCGACAACGTGAAGACC
ATCATCGTGACCTGAACAACAGCGTGGAGATCGTGTGCACCAGGCCCAACAACAACACC
AGGAAGAGCATCAGGATCGGCCAGGCCAGACCTTCTACGCCACCGGCGACATCATCGGC
GACATCCGGCAGGCCTACTGCAGCATCAACGAGAGCAACTGGAACGCCACCCTGCAGAGG
GTGTCCAAGAAGCTGGCCGAGCACTTCCCCAACAAGACCATCCAGTTCAAGAGCCCCTCT
GGCGGCGACCTGGAGATCACCATGCACAGCTTCAACTGCAGGGGCGAGTTCTTCTACTGC
AACACATCCAAGCTGTTTAACGGGACCTACTACCCAAACGGCACATACTACCCAAACGGG
ACCAACAGCACCCCTGATCATCCCATGCAGGATCAAGCAGATCATCAACATGTGGCAGGGC

图 23

GTGGGCAAGGCCATCTACGCCAGCCCAATCGCCGGCAACATCACCTGCCGGTCCAACATC
ACCGGCCTGCTGCTGACCAGGGACGGCGGCGACACCAACGACACCGAGATCTTCAGGCCA
GCCGGCGGAGACATGAGGGACAACCTGGCGGAGCGAGCTGTACAAGTACAAGATCGTGGAG
ATCAAGCCCCCTGGGCGTGGCTCCAACCGAGGCCAAGAGCAGGGTGGTGAAGAGCGAGAAG
AGCGCCGTCACCATCGGCGCCGTGTTTCTGGGCTTCCTGGGAGCCGCCGGAAGCACCATG
GGAGCCGCTCCATCACCTGACCGTGCAGGCCAGGCAGCTGCTGTCCGGGATCGTGCAG
CAGCAGAGCAACCTGCTGAGAGCCATTGAGGCTCAGCAGCACATGCTGCAGCTGACAGTG
TGGGGCATCAAGCAGCTGCAGGCCAGAGTGCTGGCCATCGAGAGATACCTGAAGGATCAG
CAGCTGCTCGGCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCACCGCCGTGCCCTGG
AACAACTCCTGGTCCAACAGGACCCAGGACGAGATCTGGAAGAACCTGACCTGGATGGAG
TGGGACAGAGAGATCAGCAACTACCAACACCATCTACGAGCTGCTGGAGGTGTCCCAG
AGCCAGCAGGAGAGGAACGAGAAGGACCTGCTGGCCCTGGATAGCTGGAACAACCTGTGG
AACTGGTTTCGACATCAGCAACTGGCTGTGGTACATCAAGAAGAAGAAGTGA

图 23 续

HIV-1 C06980v0c22 gp145蛋白质序列 (SEQ ID NO:32):

突出显示tPa信号

MDAMKRG LCCVLLCGAVFVTTTEAS NLWVTVYYGVPVWREAKTTLFCASDAKAYEREVH
NVWATHACVPTDPDPQEIFLGKNVTEKFNMWKNYMVDQMHEDIISLWDQSLQPCVKLTPL
CITLNCTDVTAHNGSTVYDNNATVVNSTNEIKNCSFNITTELDRKRKKEHALFNNLDIVQ
LDGNSSLYRLINCNTSIIKQACPKISFDPIPIHYCAPAGFVILKCNNETFNGTGPCNNVS
AVQCTHG IKPVVSTQLLNGSLAKGEIMIRSENITDNVKTIIVHLNNSVEIVCTRPNNNT
RKSIRIGPGQTFYATGDIIGDIRQAYCSINESNWNATLQRVSKKLAEHFPNKTIQFKSPS
GGDLEITMHSFNCRGEFFYCNTSKLFNGTYYPNGTYYPNGTNSTLIIPCRIKQIINMWQG
VGKAIYASPIAGNITCRSNITGLLLTRDGGDTNDTEIFRPAGGDMRDNWRSELYKYKIVE
IKPLGVAPTEAKSRVVKSEKSAVTIGAVFLGFLGAAGSTMGAASITLTVQARQLLSGIVQ
QQSNLLRAIEAQQHMLQLTVWGIKQLQARVLAERYLKDQQLLGIWGC SGKLICTTAVPW
NNSWSNRTQDEIWKNLTWMEWDREISNYTNTIYELLEVSQSQQERNEKDLLALDSWNNLW
NWFDISNWLWYIKKKK

图 24

(A) tPa信号 (SEQ ID NOS: 33和21)

突出显示的区域来自NheI克隆位点。这并非天然tPa基因的一部分，而是为克隆的目的而整合的。切割位点在A和S氨基酸之间。

核酸

ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTT

TCG GCTAGC (SEQ ID NO: 33)

蛋白质

MDAMKRGLCCVLLCGAVFVS AS (SEQ ID NO: 21)

(B) 突出显示的区域来自tPa信号 (SEQ ID NOS: 34和22)。

这并非天然tPa基因的一部分，而是为克隆的目的而整合的。切割位点在A和S之间。

核酸

ATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTT

ACTACCACAGAG GCTAGC (SEQ ID NO: 34)

蛋白质

MDAMKRGLCCVLLCGAVFVTTE AS (SEQ ID NO: 22)

图 25

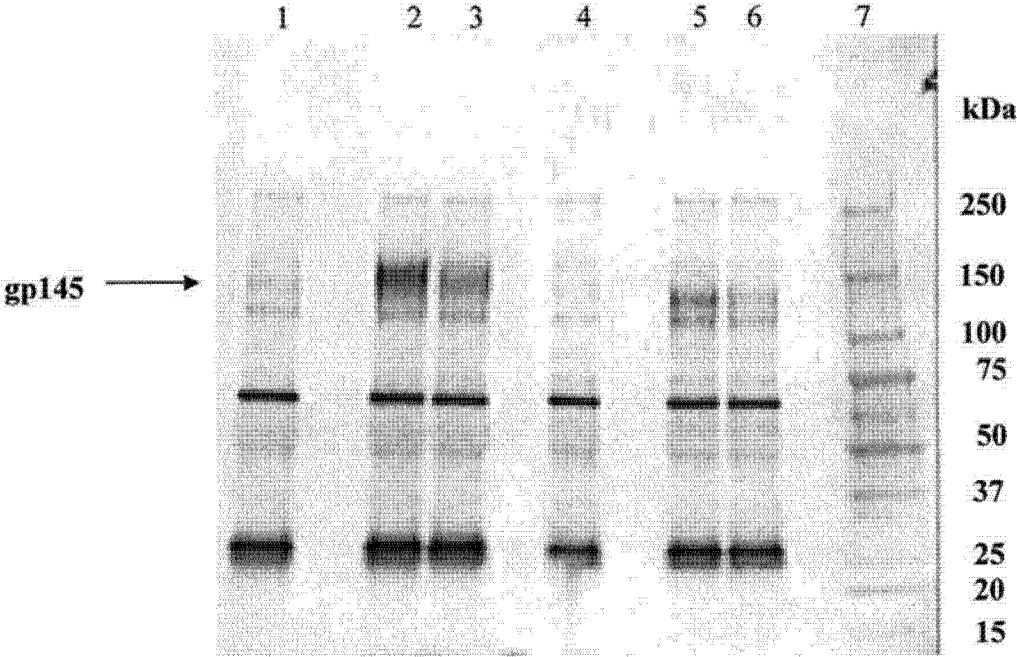


图 26

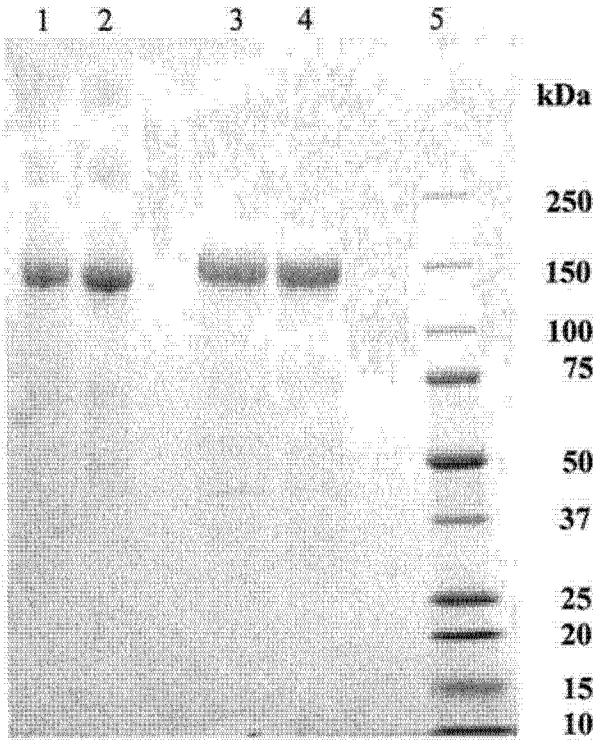


图 27

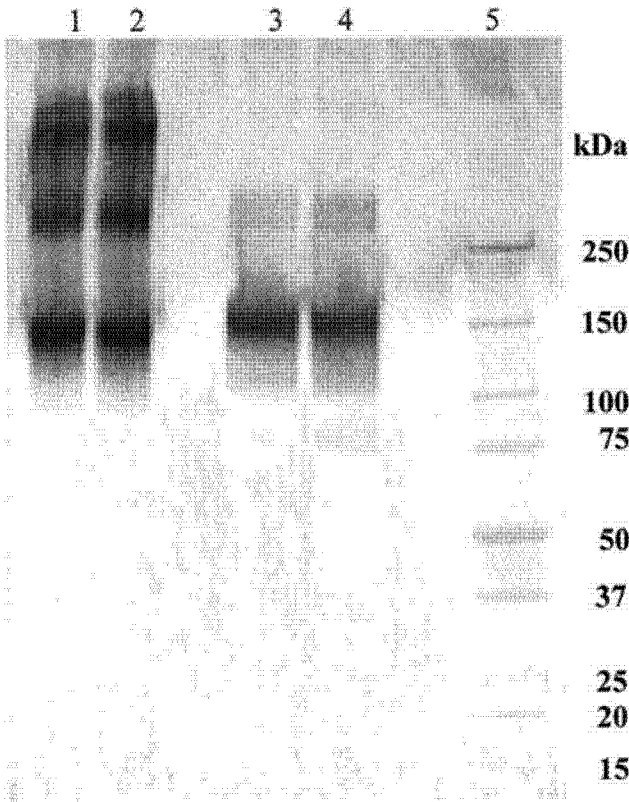


图 28

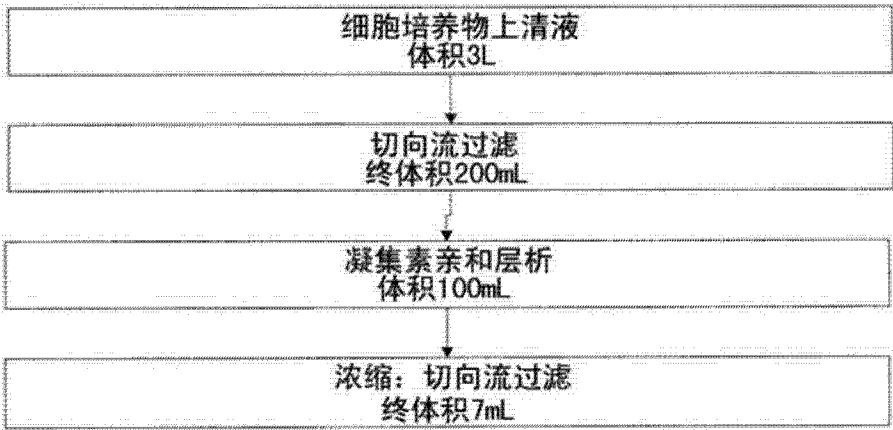


图 29

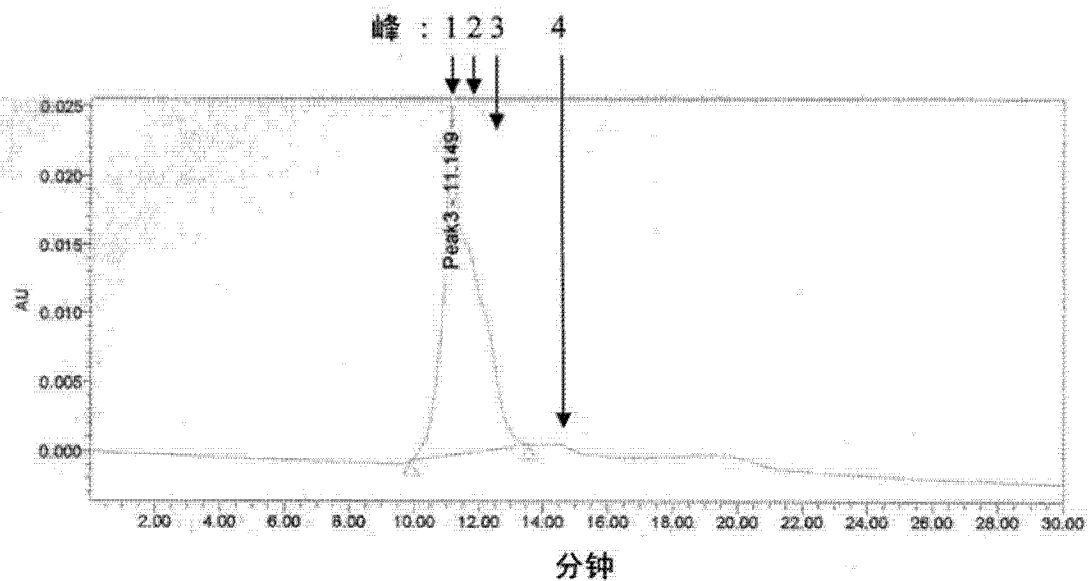


图 30

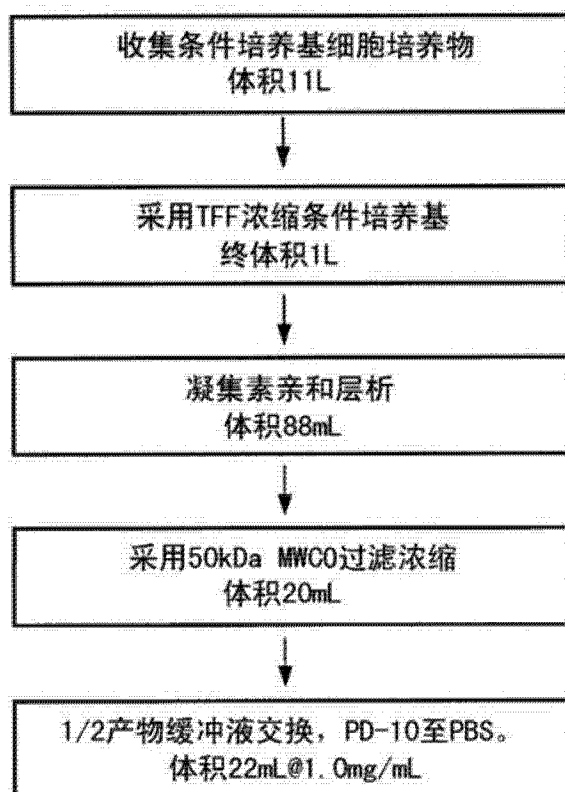


图 31

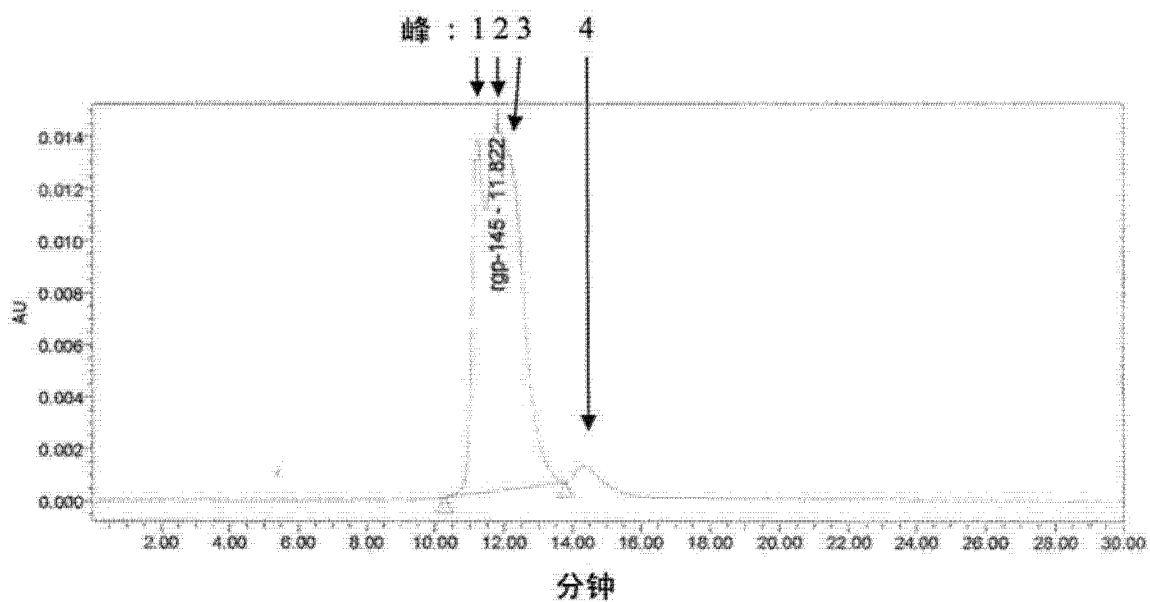


图 32

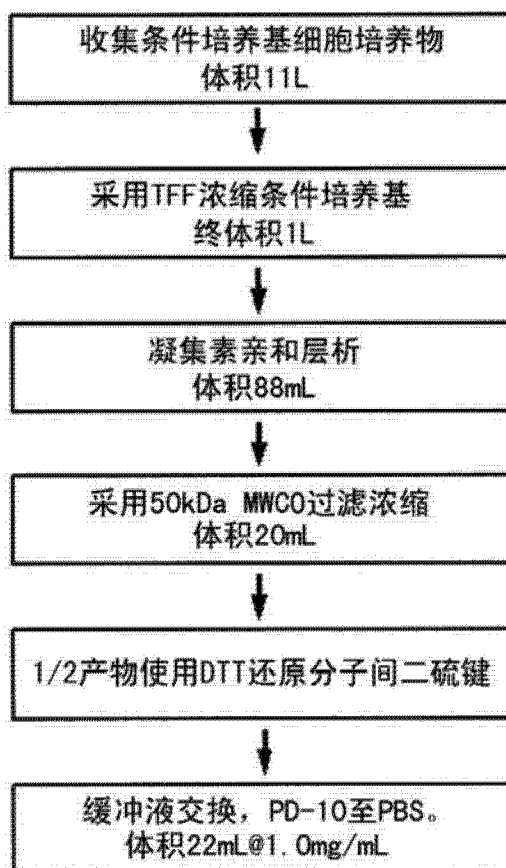


图 33

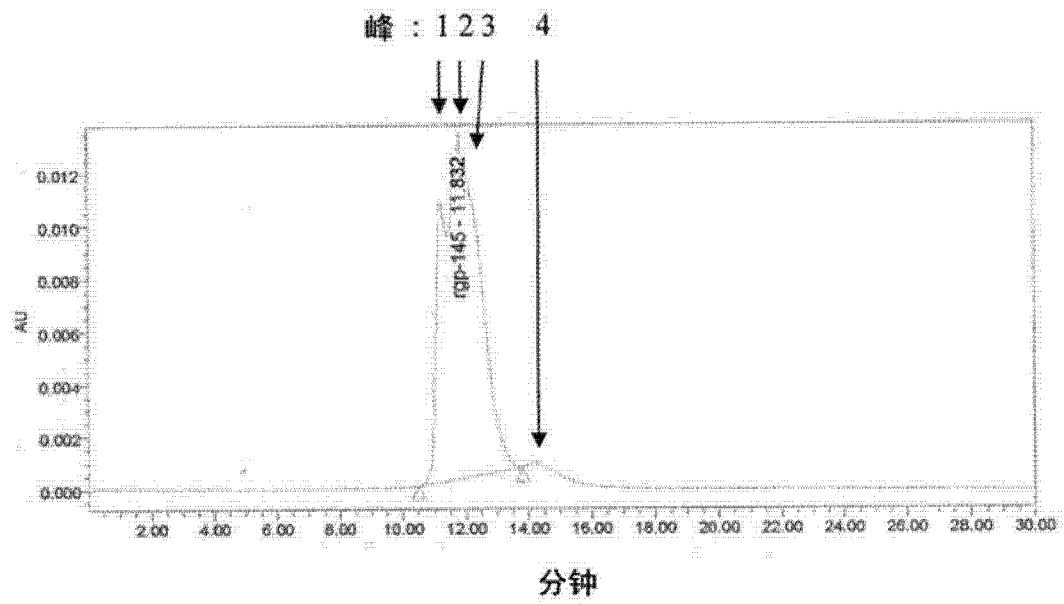


图 34

C3728v2c6 WT	SNLWVTVYYGVPVWKEAKTTLFCASDAKAYEKEVHNVWATHACVPTDPNP
C3728v2c6 DC	SNLWVTVYYGVPVWKEAKTTLFCASDAKAYEKEVHNVWATHACVPTDPNP
C06838v1c48 WT	SNLWVTVYYGVPVWREAKTTLFCASDAKAYEKEVHNVWATHACVPTDPDP
C06838v1c48 DC	SNLWVTVYYGVPVWREAKTTLFCASDAKAYEKEVHNVWATHACVPTDPDP
C06980v1c3 WT	SNLWVTVYYGVPVWREAKTTLFCASDAKAYEKEVHNVWATHACVPTDPDP
C06980v1c3 DC	SNLWVTVYYGVPVWREAKTTLFCASDAKAYEKEVHNVWATHACVPTDPDP
C06980v0c22 WT	SNLWVTVYYGVPVWREAKTTLFCASDAKAYEKEVHNVWATHACVPTDPDP
C06980v0c22 DC	SNLWVTVYYGVPVWREAKTTLFCASDAKAYEKEVHNVWATHACVPTDPDP
C3728v2c6 WT	QEMEL-PNVTFENFMWKNMVDQMHEIIISLWDQSLKPCVKLTPLCVTLN
C3728v2c6 DC	QEMEL-PNVTFENFMWKNMVDQMHEIIISLWDQSLKPCVKLTPLCVTLN
C06838v1c48 WT	QEMVL-KNVTFENFMWKNMVEQMHEIIISLWDQSLKPCVKLTPLCVTLT
C06838v1c48 DC	QEMVL-KNVTFENFMWKNMVEQMHEIIISLWDQSLKPCVKLTPLCVTLT
C06980v1c3 WT	QEIFLGKNVTEKFNMWENYMVDQMHEIIISLWDQSLQPCVKLTPLCITLN
C06980v1c3 DC	QEIFLGKNVTEKFNMWENYMVDQMHEIIISLWDQSLQPCVKLTPLCITLN
C06980v0c22 WT	QEIFLGKNVTEKFNMWKNYMVDQMHEIIISLWDQSLQPCVKLTPLCITLN
C06980v0c22 DC	QEIFLGKNVTEKFNMWKNYMVDQMHEIIISLWDQSLQPCVKLTPLCITLN
C3728v2c6 WT	CNKVNN-----RSID-NNSTEEEMKNCTFNITTEIRDKKRTQOALFYK
C3728v2c6 DC	CNKVNN-----RSID-NNSTEEEMKNCTFNITTEIRDKKRTQOALFYK
C06838v1c48 WT	CTNVNAT-----DNVTYKEKMEGEIKNCSFNITTEIRDKKRKVHALFYR
C06838v1c48 DC	CTNVNAT-----DNVTYKEKMEGEIKNCSFNITTEIRDKKRKVHALFYR
C06980v1c3 WT	CTDETAHNGSTVYDNNATA-NSTNEIKNCSFNIISELRDKRKKENALFNN
C06980v1c3 DC	CTDETAHNGSTVYDNNATA-NSTNEIKNCSFNIISELRDKRKKENALFNN
C06980v0c22 WT	CTDVTAHNGSTVYDNNATVVNSTNEIKNCSFNITTELRDKRKKKEHALFNN
C06980v0c22 DC	CTDVTAHNGSTVYDNNATVVNSTNEIKNCSFNITTELRDKRKKKEHALFNN
C3728v2c6 WT	LDIVFLGNSNESYRLINCNTSTLTQACPKVTFDPIPIHYCAPAGYAILKC
C3728v2c6 DC	LDIVFLGNSNESYRLINCNTSTLTQACPKVTFDPIPIHYCAPAGYAILKC
C06838v1c48 WT	LDVVQLNNSNE-YILINCNTSAITQACPKVSFDPIPIHYCAPAGYAILKC
C06838v1c48 DC	LDVVQLNNSNE-YILINCNTSAITQACPKVSFDPIPIHYCAPAGYAILKC
C06980v1c3 WT	LDIVQLDGNSSLYRLINCNTSIIKQACPKISFDPIPIHYCAPAGFVILTC
C06980v1c3 DC	LDIVQLDGNSSLYRLINCNTSIIKQACPKISFDPIPIHYCAPAGFVILTC
C06980v0c22 WT	LDIVQLDGNSSLYRLINCNTSIIKQACPKISFDPIPIHYCAPAGFVILKC
C06980v0c22 DC	LDIVQLDGNSSLYRLINCNTSIIKQACPKISFDPIPIHYCAPAGFVILKC
C3728v2c6 WT	KDEKFNGTGPCNNVSTVQCTHGKIPVVSTQLLLNGLAEKDIIIRSENLT
C3728v2c6 DC	KDEKFNGTGPCNNVSTVQCTHGKIPVVSTQLLLNGLAEKDIIIRSENLT
C06838v1c48 WT	NDKPFNGTGPCNNVSTVQCTHGKIPVISTQLLLNGLAEKEVIIRFENLT
C06838v1c48 DC	NDKPFNGTGPCNNVSTVQCTHGKIPVISTQLLLNGLAEKEVIIRFENLT
C06980v1c3 WT	NNETFNGTGPCNNVSAVQCTHGKIPVVSTQLLLNGLAKGEIMIRSENIT
C06980v1c3 DC	NNETFNGTGPCNNVSAVQCTHGKIPVVSTQLLLNGLAKGEIMIRSENIT
C06980v0c22 WT	NNETFNGTGPCNNVSAVQCTHGKIPVVSTQLLLNGLAKGEIMIRSENIT
C06980v0c22 DC	NNETFNGTGPCNNVSAVQCTHGKIPVVSTQLLLNGLAKGEIMIRSENIT
C3728v2c6 WT	NNVKTIMVHLNESVEINCTRPNNNTRKSIRIGPGQTFYATGGIIGNIROA
C3728v2c6 DC	NNVKTIMVHLNESVEINCTRPNNNTRKSIRIGPGQTFYATGGIIGNIROA
C06838v1c48 WT	DNAKTIIVQLNQSTIEIKCIRPNNNTRESIRIGPGQAFYATRDIIGDIRRA
C06838v1c48 DC	DNAKTIIVQLNQSTIEIKCIRPNNNTRESIRIGPGQAFYATRDIIGDIRRA
C06980v1c3 WT	NNVKTIIIVHENKSVEIVCTRPNNNTRKSIRIGPGQTFYATGDIIGDIRQA
C06980v1c3 DC	NNVKTIIIVHENKSVEIVCTRPNNNTRKSIRIGPGQTFYATGDIIGDIRQA
C06980v0c22 WT	DNVKTIIVHLNNSVEIVCTRPNNNTRKSIRIGPGQTFYATGDIIGDIRQA
C06980v0c22 DC	DNVKTIIVHLNNSVEIVCTRPNNNTRKSIRIGPGQTFYATGDIIGDIRQA

图 35

C3728v2c6 WT	YCTISKKNWNTTLERVATKLKEYFKNTTIQFAPHSGGDLEITTHSFNCRG
C3728v2c6 DC	YCTISKKNWNTTLERVATKLKEYFKNTTIQFAPHSGGDLEITTHSFNCRG
C06838v1c48 WT	YCTIETERWKETLGNVTEKLKEYFPNTNISFAPSSGGDLEITTHSFNCRG
C06838v1c48 DC	YCTIETERWKETLGNVTEKLKEYFPNTNISFAPSSGGDLEITTHSFNCRG
C06980v1c3 WT	YCSINESNWNITLQRVSKKLAHFNPNTIQFESPSGGDLEITMHSFNCRG
C06980v1c3 DC	YCSINESNWNITLQRVSKKLAHFNPNTIQFESPSGGDLEITMHSFNCRG
C06980v0c22 WT	YCSINESNWNATLQRVSKKLAHFNPNTIQFKSPSGGDLEITMHSFNCRG
C06980v0c22 DC	YCSINESNWNATLQRVSKKLAHFNPNTIQFKSPSGGDLEITMHSFNCRG
C3728v2c6 WT	EFFYCNTSOLFNGT--STGFSNKSTGNETFTLPCRIKQIINMWQEVGRAM
C3728v2c6 DC	EFFYCNTSOLFNGT--STGFSNKSTGNETFTLPCRIKQIINMWQEVGRAM
C06838v1c48 WT	EFFYRNTSKLFN-----NNDTENNLITLPCRIKQIVNMWQGVGRAM
C06838v1c48 DC	EFFYRNTSKLFN-----NNDTENNLITLPCRIKQIVNMWQGVGRAM
C06980v1c3 WT	EFFYCNTSKLFKSTYHPNGTYNLNGTNSTLIIPCRIKQIINMWQGVGKAI
C06980v1c3 DC	EFFYCNTSKLFKSTYHPNGTYNLNGTNSTLIIPCRIKQIINMWQGVGKAI
C06980v0c22 WT	EFFYCNTSKLFNGTYYPNGTYYPNGTNSTLIIPCRIKQIINMWQGVGKAI
C06980v0c22 DC	EFFYCNTSKLFNGTYYPNGTYYPNGTNSTLIIPCRIKQIINMWQGVGKAI
C3728v2c6 WT	YAPPIAGNITCVSNITGLLLTRDGGDNNTTETIFRPGGGDMRDNRSELY
C3728v2c6 DC	YAPPIAGNITCVSNITGLLLTRDGGDNNTTETIFRPGGGDMRDNRSELY
C06838v1c48 WT	YAPPIEGNITCKSNITGLLLTRDGGNGTERNETFRPGGGDMRNNNRSELY
C06838v1c48 DC	YAPPIEGNITCKSNITGLLLTRDGGNGTERNETFRPGGGDMRNNNRSELY
C06980v1c3 WT	YASPIAGSITCRSNITGLLLTRDGGDTND-TEIFRPTGGDMRDNRSELY
C06980v1c3 DC	YASPIAGSITCRSNITGLLLTRDGGDTND-TEIFRPTGGDMRDNRSELY
C06980v0c22 WT	YASPIAGNITCRSNITGLLLTRDGGDTND-TEIFRPAGGDMRDNRSELY
C06980v0c22 DC	YASPIAGNITCRSNITGLLLTRDGGDTND-TEIFRPAGGDMRDNRSELY
C3728v2c6 WT	KYKVVEVKPLGVAPSEARRRVVEREKRAVGLGAVFLGFLGAAGSTMGAAS
C3728v2c6 DC	KYKVVEVKPLGVAPSEARRRVVEREKRAVGLGAVFLGFLGAAGSTMGAAS
C06838v1c48 WT	KYKVVEIKPLGIAPTGAARRRVVEREKRAVGMGAVFLGFLGAAGSTMGAAS
C06838v1c48 DC	KYKVVEIKPLGIAPTGAARRRVVEREKRAVGMGAVFLGFLGAAGSTMGAAS
C06980v1c3 WT	KYKIVEIKPLGVAPTEAKRRVVKREKRAVTIGAVFLGFLGAAGSTMGAAS
C06980v1c3 DC	KYKIVEIKPLGVAPTEAKRRVVKSEKSAVTIGAVFLGFLGAAGSTMGAAS
C06980v0c22 WT	KYKIVEIKPLGVAPTEAKRRVVKREKRAVTIGAVFLGFLGAAGSTMGAAS
C06980v0c22 DC	KYKIVEIKPLGVAPTEAKRRVVKSEKSAVTIGAVFLGFLGAAGSTMGAAS
C3728v2c6 WT	ITLTVQARQLLSGIVQQQSNLLRAIEAQQHMLQLTVWGKIQLOTKVLAIE
C3728v2c6 DC	ITLTVQARQLLSGIVQQQSNLLRAIEAQQHMLQLTVWGKIQLOTKVLAIE
C06838v1c48 WT	ITLTVQARQVLSGIVQQQSNLLRAIEAQQHMLQLTVWGKIQLOTRVLAIE
C06838v1c48 DC	ITLTVQARQVLSGIVQQQSNLLRAIEAQQHMLQLTVWGKIQLOTRVLAIE
C06980v1c3 WT	ITLTVQARQLLSGIVQQQSNLLRAIEAQQHMLQLAVWGKIQLOARVLAIE
C06980v1c3 DC	ITLTVQARQLLSGIVQQQSNLLRAIEAQQHMLQLAVWGKIQLOARVLAIE
C06980v0c22 WT	ITLTVQARQLLSGIVQQQSNLLRAIEAQQHMLQLTVWGKIQLOARVLAIE
C06980v0c22 DC	ITLTVQARQLLSGIVQQQSNLLRAIEAQQHMLQLTVWGKIQLOARVLAIE
C3728v2c6 WT	RYLQDQQLLGIWGCSGKLICTTNVPWNSSWSNKSILKAIWDNMNTWMQWDRE
C3728v2c6 DC	RYLQDQQLLGIWGCSGKLICTTNVPWNSSWSNKSILKAIWDNMNTWMQWDRE
C06838v1c48 WT	RYLKDQQLLGLWGCSGKLICTTAVPWNSSWSNKSEIDIWNNMTWMQWDRE
C06838v1c48 DC	RYLKDQQLLGLWGCSGKLICTTAVPWNSSWSNKSEIDIWNNMTWMQWDRE
C06980v1c3 WT	RYLKDQQLLGIWGCSGKLICTTAVPWNSSWSNRTODEIWKNLTWMEWDRE
C06980v1c3 DC	RYLKDQQLLGIWGCSGKLICTTAVPWNSSWSNRTODEIWKNLTWMEWDRE
C06980v0c22 WT	RYLKDQQLLGIWGCSGKLICTTAVPWNSSWSNRTODEIWKNLTWMEWDRE
C06980v0c22 DC	RYLKDQQLLGIWGCSGKLICTTAVPWNSSWSNRTODEIWKNLTWMEWDRE

图 35 续 -1

C3728v2c6 WT	ISNYTNTIYRLLEDSDHIQOEKNEKDLLALNSWNNLWSWPNITNWLWYIRI
C3728v2c6 DC	ISNYTNTIYRLLEDSDHIQOEKNEKDLLALNSWNNLWSWPNITNWLWYIRI
C06838v1c48 WT	ISNYTNTIYRLLEDSDHIQOEKNEKDLLALDSWKNLWNWPNITKWLWYIKI
C06838v1c48 DC	ISNYTNTIYRLLEDSDHIQOEKNEKDLLALDSWKNLWNWPNITKWLWYIKI
C06980v1c3 WT	ISNYTNTIYELLEVSSQSQQERNEKDLLALDSWNNLWNWPNISNWLWYIKI
C06980v1c3 DC	ISNYTNTIYELLEVSSQSQQERNEKDLLALDSWNNLWNWPNISNWLWYIKI
C06980v0c22 WT	ISNYTNTIYELLEVSSQSQQERNEKDLLALDSWNNLWNWPNISNWLWYIKI
C06980v0c22 DC	ISNYTNTIYELLEVSSQSQQERNEKDLLALDSWNNLWNWPNISNWLWYIKI
C3728v2c6 WT	FIMIVGGLIGLRRIIFAVLSIVNRVRQGYSPLSFQTLTPNQRPDRIGRIE
C3728v2c6 DC	FIMIVGGLIGLRRIIFAVLSIVNRVRQGYSPLSFQTLTPNQRPDRIGRIE
C06838v1c48 WT	FIIVGGLIGLRRIIFAVLSIVNRVRQGYSPLSFQTLIPSPRGPDKLGRIE
C06838v1c48 DC	FIIVGGLIGLRRIIFAVLSIVNRVRQGYSPLSFQTLIPSPRGPDKLGRIE
C06980v1c3 WT	FIMIVGGLIGLRRIIFAVLSIVNRVRQGYSPLSFQTLIPNQREPDRPGRIE
C06980v1c3 DC	FIMIVGGLIGLRRIIFAVLSIVNRVRQGYSPLSFQTLIPNQREPDRPGRIE
C06980v0c22 WT	FIMIVGGLIGLRRIIFAVLSIVNRVRQGYSPLSFQTLIPNQREPDRPGRIE
C06980v0c22 DC	FIMIVGGLIGLRRIIFAVLSIVNRVRQGYSPLSFQTLIPNQREPDRPGRIE
C3728v2c6 WT	EEGGEQDKDRSIRLVNGFLALAWDDLRLSLCLFSYHRLRDFILIAARATEL
C3728v2c6 DC	EEGGEQDKDRSIRLVNGFLALAWDDLRLSLCLFSYHRLRDFILIAARATEL
C06838v1c48 WT	EEGGEQDKGRSVRLVNGFLALAWDDLRLSLCLFSYHQLRDFILIVARAVGL
C06838v1c48 DC	EEGGEQDKGRSVRLVNGFLALAWDDLRLSLCLFSYHQLRDFILIVARAVGL
C06980v1c3 WT	EEGGEQDKDRSIRLVSGFLALAWDDLRLSLCIFLYYHHLRDLILIAARATEL
C06980v1c3 DC	EEGGEQDKDRSIRLVSGFLALAWDDLRLSLCIFLYYHHLRDLILIAARATEL
C06980v0c22 WT	EEGGEQDKDRSIRLVSGFLALAWDDLRLSLCIFLYYHHLRDFILIAARATEL
C06980v0c22 DC	EEGGEQDKDRSIRLVSGFLALAWDDLRLSLCIFLYYHHLRDFILIAARATEL
C3728v2c6 WT	LGHSSLRGLQRGWEALKYLGSLGQYWGOELKKSAILRLDITAIABAEGTD
C3728v2c6 DC	LGHSSLRGLQRGWEALKYLGSLGQYWGOELKKSAILRLDITAIABAEGTD
C06838v1c48 WT	LGRSSLRGLQRGWEILKYLGGLVQYWGLELKKSAYSLEDTIAIAVTEGTD
C06838v1c48 DC	LGRSSLRGLQRGWEILKYLGGLVQYWGLELKKSAYSLEDTIAIAVTEGTD
C06980v1c3 WT	LGRSSLRGLQRGWEALKYLGSLVQYWGLEIKKSAINLLDTIAIABAEGTD
C06980v1c3 DC	LGRSSLRGLQRGWEALKYLGSLVQYWGLEIKKSAINLLDTIAIABAEGTD
C06980v0c22 WT	LGRSSLRGLQRGWEALKYLGSLVQYWGLEIKKSAINLLDTIAIABAEGTD
C06980v0c22 DC	LGRSSLRGLQRGWEALKYLGSLVQYWGLEIKKSAINLLDTIAIABAEGTD
C3728v2c6 WT	RIIEFIQRI CRAIHNVPNRIRQGFEAALL* (SEQ ID NO. 35)
C3728v2c6 DC	RIIEFIQRI CRAIHNVPNRIRQGFEAALL* (SEQ ID NO. 36)
C06838v1c48 WT	RIIEFIQRI CRAIRNVPTIRIRQGFEAALQ* (SEQ ID NO. 37)
C06838v1c48 DC	RIIEFIQRI CRAIRNVPTIRIRQGFEAALQ* (SEQ ID NO. 38)
C06980v1c3 WT	RIIEIVQACRAVINILRRIRQGLEAALQ* (SEQ ID NO. 39)
C06980v1c3 DC	RIIEIVQACRAVINILRRIRQGLEAALQ* (SEQ ID NO. 40)
C06980v0c22 WT	RIIEIVQACRAVINIPRRIRQGLEAALQ* (SEQ ID NO. 41)
C06980v0c22 DC	RIIEIVQACRAVINIPRRIRQGLEAALQ* (SEQ ID NO. 42)

图 35 续 -2

C06980v0c22 gp145氨基酸序列 (SEQ ID NO:32):

```

1  mdamkrglcc vlllcgavfv ttteasnlw tvyygvpvwr eakttlfcas dakayerevh
    tPa信号
61  nvwathacvp tdpdpqeifl gknvtekfnm wknymvdqmh ediislwdqs lqpcvkltp1
121 citlnctdvt ahngstvydn natvvnstne ikncsfnitt elrdkrkkeh alfnnldivq
181 ldgnsslyrl incntsiikq acpkisfdpi pihycapagf vilkennetf ngtgpcnnvs
241 avqcthgikp vvstqlllng slakgeimir senitdnvkt iivhltnsve ivctrpnntt
301 rksirigpgq tfyatgdiig dirqaycsin eswnatlqr vskklaehfp nkti qfksp
361 ggdleitmhs fncrgeffyc ntsklfngty ypngtyypng tnstliipcr ikqiinmwgg
421 vgkaiyaspi agnitorsni tgl1l1trdgg dtndteifrp aggdmrnwr selykykive
481 ikplgvapte aksrvvksek savtigavfl gflgaagstm gaasitltvq arqllsgivq
    切割位点R至S突变
541 qqsnullraie agqhmlqltv wgikqlqarv laierylkdq qilgiwgcsq klicttavpw
601 nnswnrtqd eiwknltwme wdreisnytn tiyellevsq sqqernekdl laidswnnlw
661 nwfdisnwlw yikkkk
    3个赖氨酸

```

图 36

C06980v0c22 gp145核苷酸序列 (SEQ ID NO:20)和翻译序列 (SEQ ID NO:32):

```

1  atggatgcaa tgaagagagg gctctgctgt gtgctgctgc tgtgtggagc agtcttcgtt
   >>.....tPa前导序列.....>
   m d a m k r g l c c v l l l c g a v f v

       NheI
       --+----
61  actaccacag aggctagcaa cctgtgggtg accgtgtact acggcgtgcc cgtgtggaga
   >...tPa前导序列..>>
   t t t e a s
       >>.....gp145.....>
       s n l w v t v y y g v p v w r

121 gaggccaaga ccaccctgtt ctgcgccagc gacgccaagg cctacgagcg ggaggtgcac
   >.....gp145.....>
   e a k t t l f o a s d a k a y e r e v h

181 aacgtgtggg ccaccacagc ctgcgtgccc accgaccccg accctcagga aatcttcctg
   >.....gp145.....>
   n v w a t h a c v p t d p d p q e i f l

241 ggcaagaacg tgaccgagaa gttcaacatg tggaagaact acatggtgga ccagatgcac
   >.....gp145.....>
   g k n v t e k f n m w k n y m v d q m h

301 gaggacatca tcagcctgtg ggaccagagc ctgcagccct gcgtgaagct gacccccctg
   >.....gp145.....>
   e d i i s l w d q s l q p c v k l t p l

361 tgcacacccc tgaactgcac cgacgtgacc gccacacaac gcagcacctg gtacgacaac
   >.....gp145.....>
   c i t l n c t d v t a h n g s t v y d n

421 aacgccaccg tggatgaacg caccaacgag atcaagaact gcagcttcaa catcaccacc
   >.....gp145.....>
   n a t v v n s t n e i k n c s f n i t t

481 gagctgaggg acaagaggaa gaaggagcac gccctgttca acaacctgga catcgtgcag
   >.....gp145.....>
   e l r d k r k k e h a l f n n l d i v q

541 ctggacggca acagctccct gtacagactg atcaactgca acaccagcat catcaagcag
   >.....gp145.....>
   l d g n s s l y r l i n c n t s i i k q

601 gcttgcccca agatcagctt cgaccccatc cccatccact actgogcccc agccggcttc
   >.....gp145.....>
   a c p k i s f d p l p i h y c a p a g f

661 gtgatcctga agtgcaacaa cgagacattc aacggcaccg gccctgttaa caacgtgtcc
   >.....gp145.....>
   v i l k c n n e t f n g t g p c n n v s

```

图 37

```

721 gctgtgcagt gcacccacgg catcaagccc gtgggtgtcca cccagctgct gctgaacggc
>.....gpl45.....>
a v q c t h g i k p v v s t q l l l n g

781 agcctggcca agggcgagat catgatcaga agcgagaaca tcaccgacaa cgtgaagacc
>.....gpl45.....>
s l a k g e i m i r s e n i t d n v k t

841 atcatcgtgc acctgaacaa cagcgtggag atcgtgtgca ccaggcccaa caacaacacc
>.....gpl45.....>
l i v h l n n s v e i v c t r p n n n t

901 aggaagagca tcaggatcgg cccaggccag accttctacg ccaccggcga catcatcggc
>.....gpl45.....>
r k s i r i g p g q t f y a t g d i i g

961 gacatccggc aggcctactg cagcatcaac gagagcaact ggaacgccac cctgcagagg
>.....gpl45.....>
d i r q a y c s i n e s n w n a t l q r

1021 gtgtccaaga agctggccga gcaattcccc aacaagacca tccagttcaa gagcccctct
>.....gpl45.....>
v s k k l a e h f p n k t i q f k s p s

1081 ggcgggcgacc tggagatcac catgcacagc ttcaactgca gggggcgagtt cttctactgc
>.....gpl45.....>
g g d l e i t m h s f n c r g e f f y c

1141 aacacatcca agctgttttaa cgggacctac tacccaaacg gcacatacta cccaaacggg
>.....gpl45.....>
n t s k l f n g t y y p n g t y y p n g

1201 accaacagca ccctgatcat cccatgcagg atcaagcaga tcatcaacat gtggcagggc
>.....gpl45.....>
t n s t l i i p c r i k q i i n m w q g

1261 gtgggcaagg ccattctacg cagcccaatc gccggcaaca tcacctgccg gtccaacatc
>.....gpl45.....>
v g k a i y a s p i a g n i t c r s n i

1321 accggcctgc tgctgaccag ggacggcggc gacaccaacg acaccgagat cttcaggcca
>.....gpl45.....>
t g l l l t r d g g d t n d t e i f r p

1381 gccggcgagg acatgagggg caactggcgg agcgagctgt acaagtacaa gatcgtggag
>.....gpl45.....>
a g g d m r d n w r s e l y k y k i v e

1441 atcaagcccc tgggcgtggc tccaaccgag gccaagagca ggggtggtgaa gagcgagaag
>.....gpl45.....>
i k p l g v a p t e a k s r v v k s e k
          切割位点R至S突变

1501 agcgccgtca ccacggcgcc cgtgtttctg ggcttctctg gagccgcggg aagcaccatg
>.....gpl45.....>
s a v t i g a v f l g f l g a a g s t m
          切割位点R至S突变

```

图 37 续 -1

```

1561 ggagccgcct ccataccct gaccgtgcag gccaggcagc tctgtccgg gatcgtgcag
>.....gp145.....>
   g a a s i t l t v q a r q l l s g i v q

1621 cagcagagca acctgctgag agccattgag gctcagcagc acatgctgca gctgacagtg
>.....gp145.....>
   q q s n l l r a i e a q q h m l q l t v

1681 tggggcatca agcagctgca ggccagagtg ctggccatcg agagatacct gaaggatcag
>.....gp145.....>
   w g i k q l q a r v l a i e r y l k d q

1741 cagctgctcg gcatactgggg ctgcagcggc aagctgatct gcaccaccgc cgtgccctgg
>.....gp145.....>
   q l l g i w q c s g k l i c t t a v p w

1801 aacaactcct ggtccaacag gacccaggac gagatctgga agaactgac ctggatggag
>.....gp145.....>
   n n s w s n r t q d e i w k n l t w m e

1861 tgggacagag agatcagcaa ctacaccaac accatctacg agctgctgga ggtgtcccag
>.....gp145.....>
   w d r e i s n y t n t i y e l l e v s q

1921 agccagcagg agaggaacga gaaggacctg ctggccctgg atagctggaa caacctgtgg
>.....gp145.....>
   s q q e r n e k d l l a l d s w n n l w

                                EcoRI
                                -----
1981 aactggttcg acatcagcaa ctggctgtgg tacatcaaga agaagaagtg aattc
>.....gp145.....>>>
   n w f d i s n w l w y i k k k k -
                                赖氨酸末端

```

图 37 续 -2

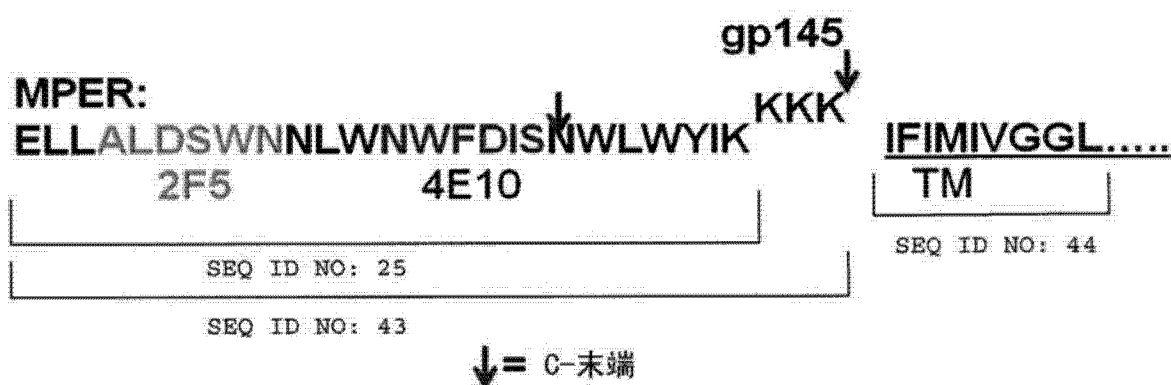


图 38

包膜	抗原	进化枝	年份	抗原结合滴度 (ug/ml)				
				2F5	4E10	2G12	b12	VRC01
C06980	gp145	C	2008	>25	10	>25	20.0	0.0
57128	gp140	D	1998	>25	>25	0.1	6.7	>25
A07412	gp120	D	1999	>25	>25	>25	>25	0.1
	gp140			>25	>25	22.9	>25	4.3
57140	gp120	D	1998	>25	>25	20.0	>25	0.0
	gp140			>25	>25	>25	>25	0.0

>25

10-25

2-10

<2

图 39

进化枝 / CoR.	包膜	PV 中和敏感性 (ug/ml)				
		2F5	4E10	2G12	b12	VRC01
C/R5	C06980	>25	4.7	>25	>25	0.7
D/R5	57128	>25	11.7	10.5	24.4	>25
D/R5	A07412	3.3	2.2	>25	3.0	0.3
C/R5	20635-4	>25	8.3	>25	3.1	NT
C/R5	93/MW/965	>25	10.3	>25	3.2	NT
C/R5	PBL 288	>25	4.5	>25	>25	NT
D/R5	E13613M4	3.2	12.8	>25		NT
D/R5	NKU3006	2.1	2.7	>25	>25	NT
D/R5	A03349M1	>25	2.4	>25		NT

>25

10-25

2-10

<2

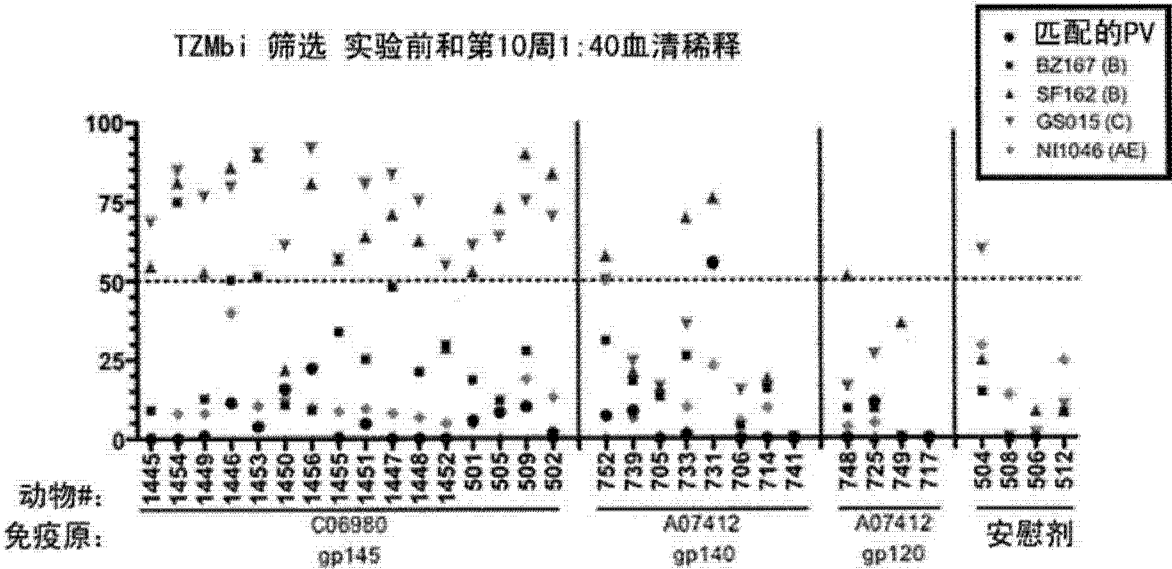
图 40

6组新西兰雌兔，每组4只

组别	抗原	佐剂/媒介物
1	gp145蛋白质	铝胶
2	gp145蛋白质	包含脂质A的脂质体制剂物与gp145蛋白质混合完成
3	gp145蛋白质	脂质体制剂物1，包含脂质A，包裹145蛋白质
4	gp145蛋白质	脂质体制剂物2，包含脂质 A和PIP，包裹gp145蛋白质
5	无	脂质体制剂物1，包含脂质A
6	mper23，B进化枝	脂质体制剂物2，包含脂质 A和PIP，包裹mper23蛋白质

- 在0、4和8周时加入25ug剂量
- 在-2、0、4、8、10、12（终止）周时放血
- 对6980 gp145、CN54 gp140、ZA1157 gp120的结合滴度 10^4 - 10^5

图 41



血清与MuLV和HIV-2/MPER嵌合体 (G Shaw)

图 42

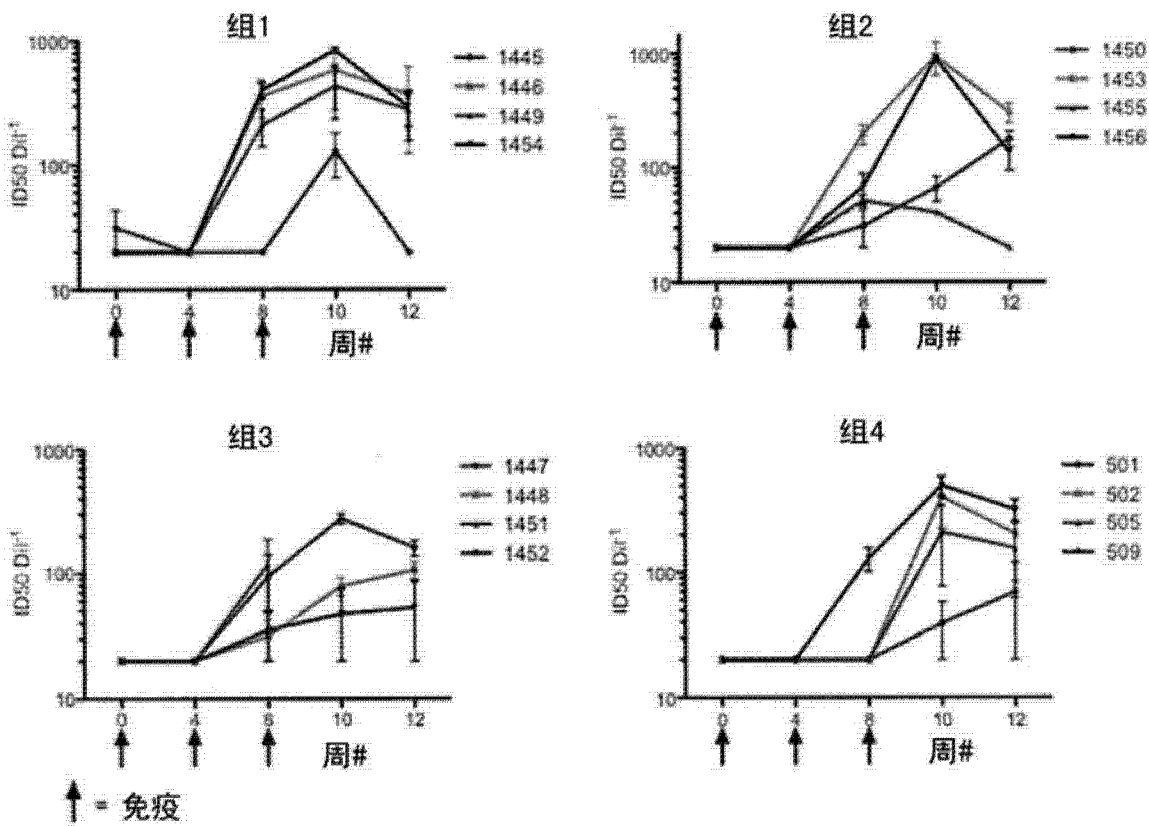


图 43

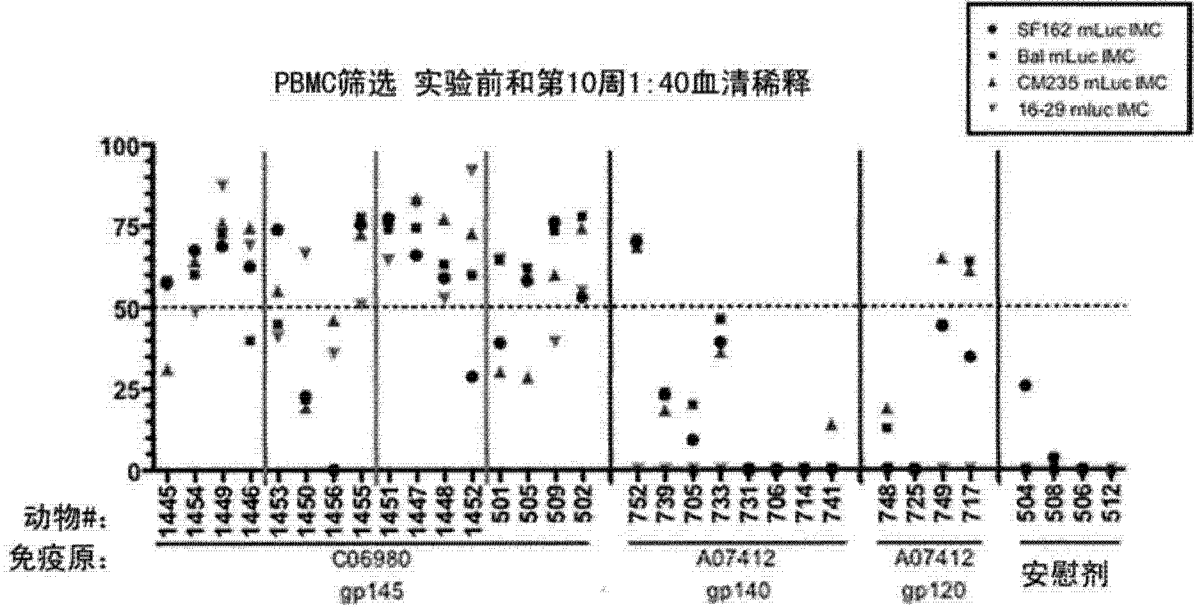


图 44

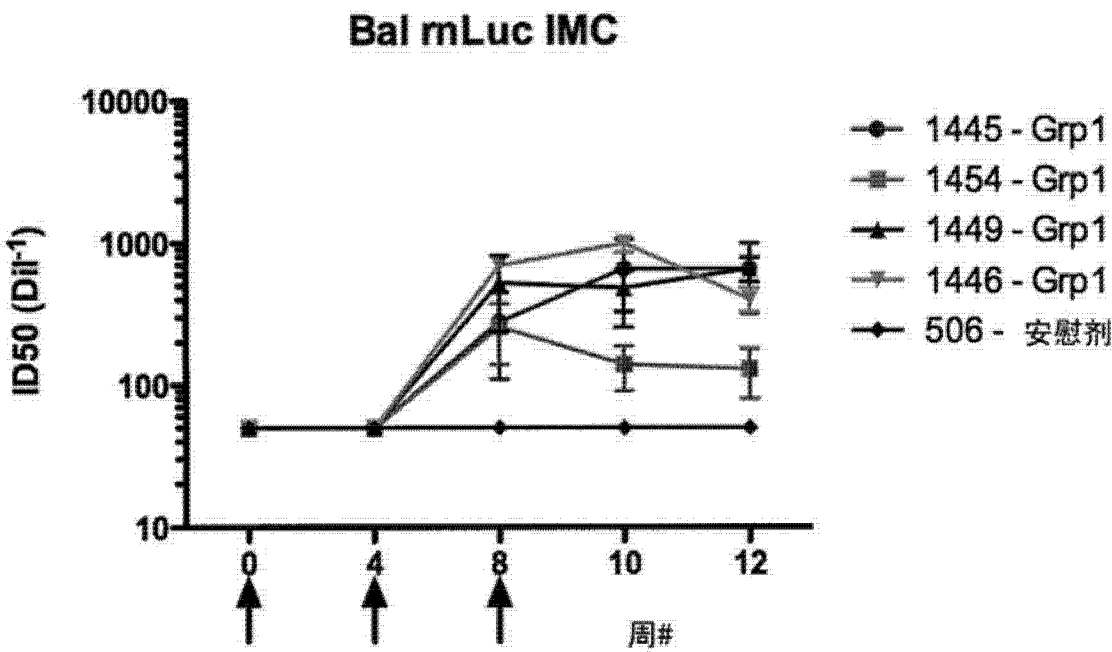
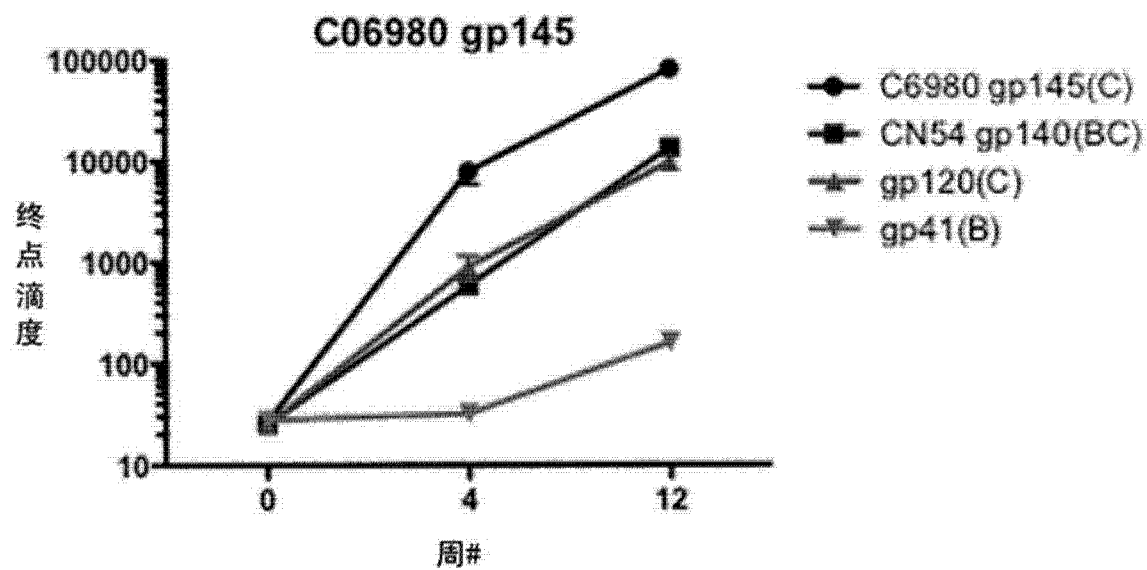


图 45



所有组的几何平均滴度 (SEM)
(C gp120 = ZA.1197MB, S Africa)

图 46

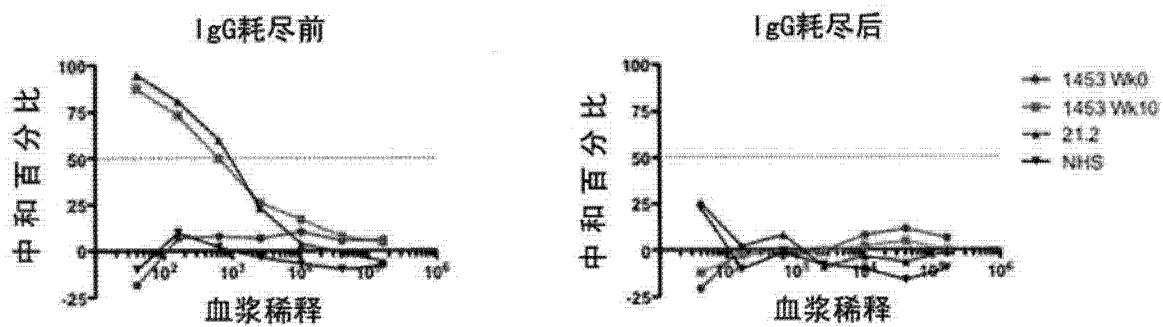


图 47

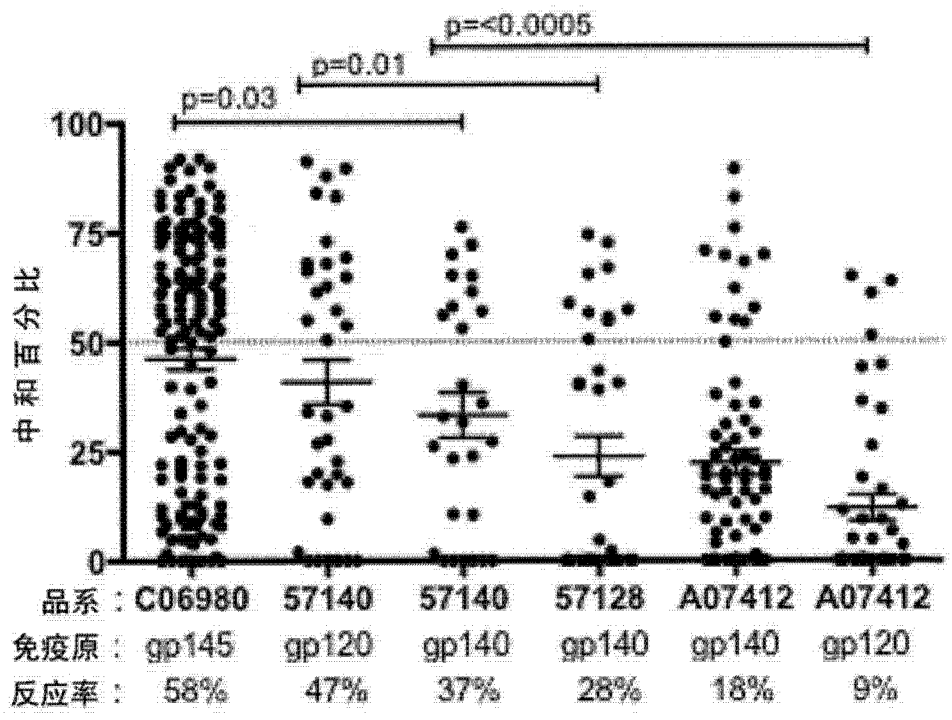
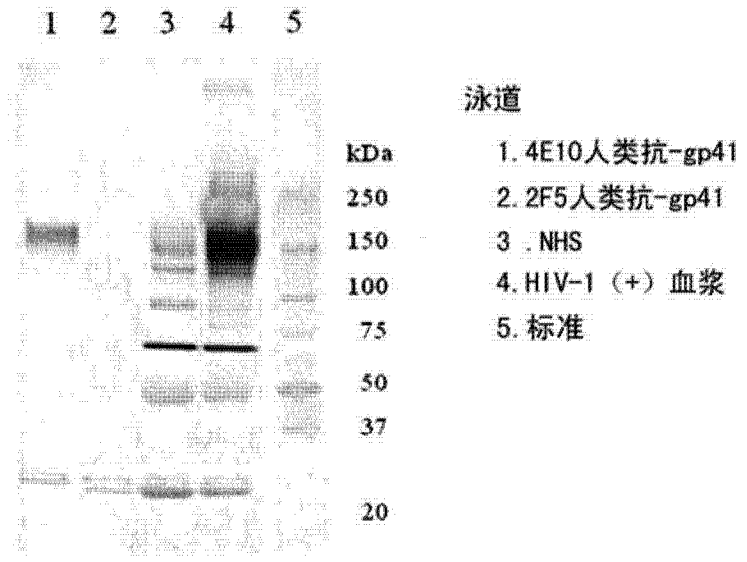


图 48

无蛋白培养基中培养的CHO C06980v0c22 gp145细胞系免疫共沉淀蛋白质印迹分析



使用人抗体将gp145蛋白质从条件培养基中沉淀出来，于4-15%SDS-PAGE上电泳展开，转移至PVDF，使用兔抗gp120和gp160抗血清对其进行检测。

图 49

序列相似性百分比:						潜在N-糖基化位点
	ConC	C06980	A07412	S7140	S7128	
ConC	100	NA	NA	NA	NA	29
C06980	82.9	100	NA	NA	NA	31
A07412	77.2	71.4	100	NA	NA	29
S7140	75.7	71	78.9	100	NA	28
S7128	75.7	70.2	77.8	77.8	100	28

图 50

122

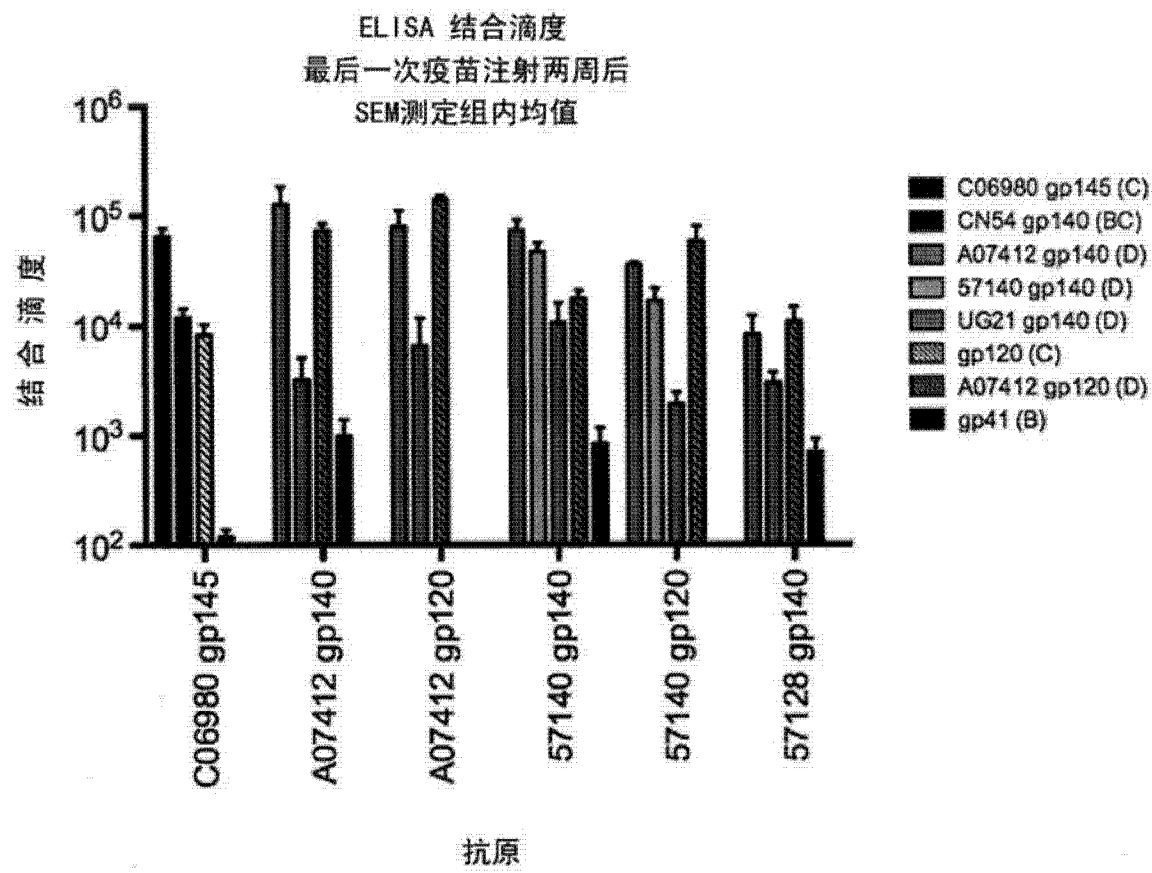


图 52

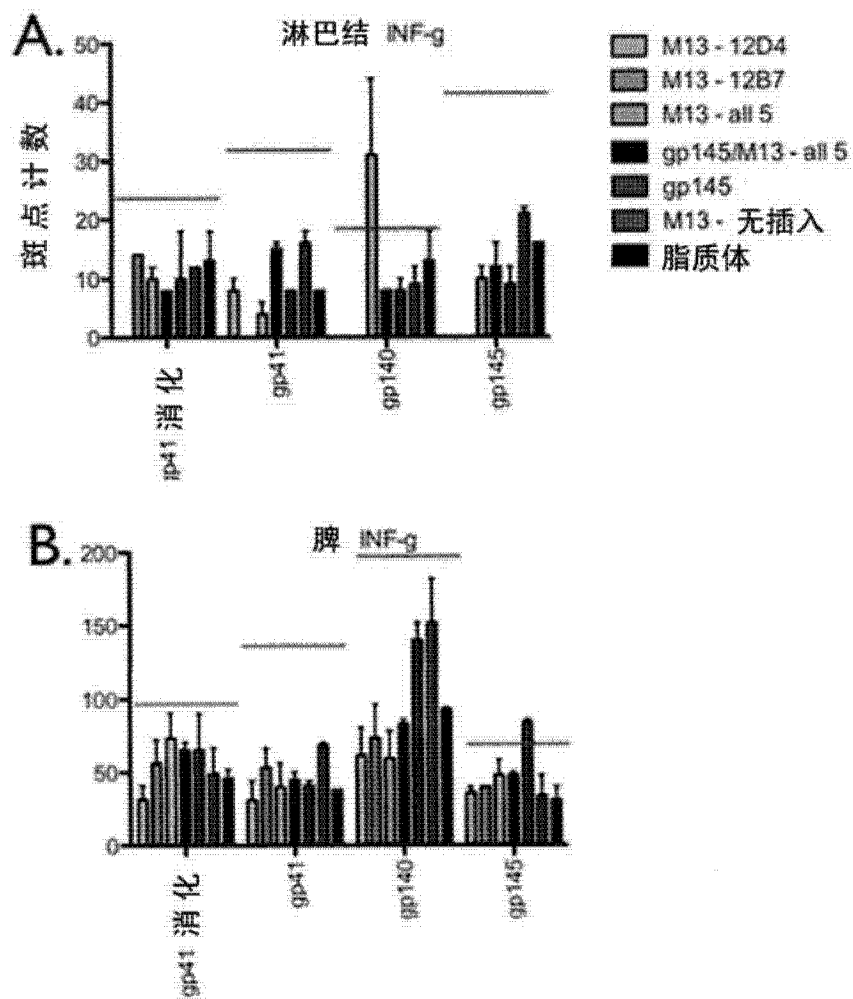


图 53

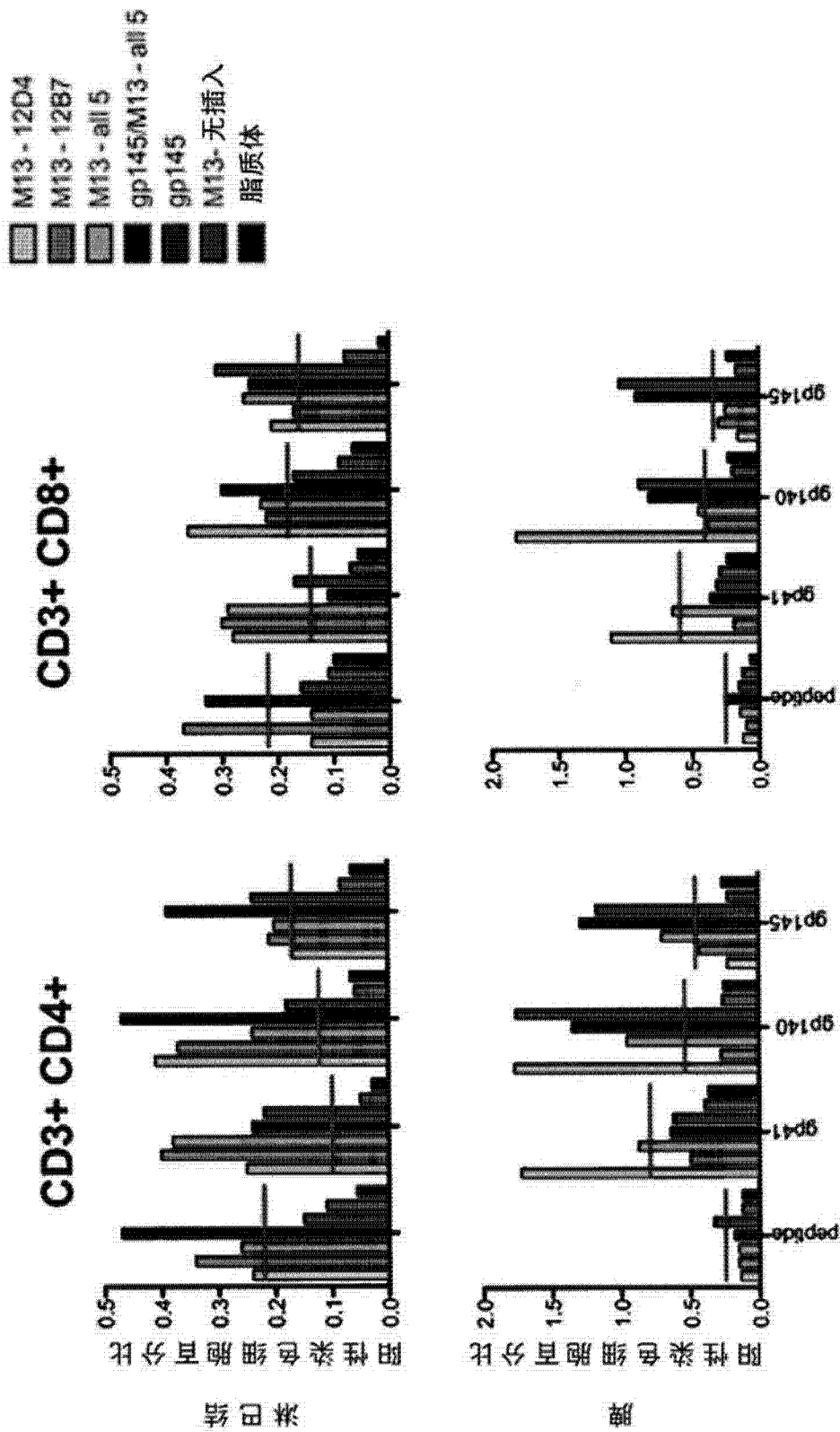


图 54

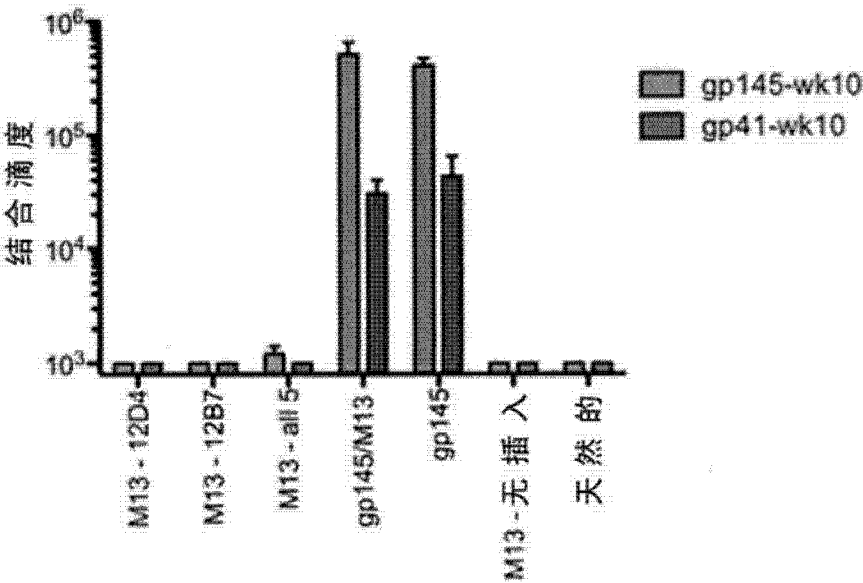


图 55

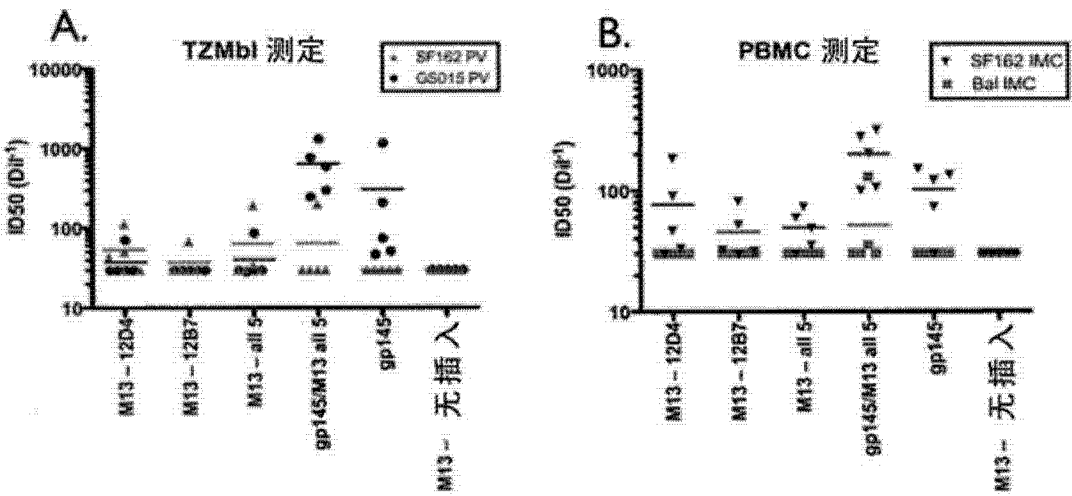


图 56

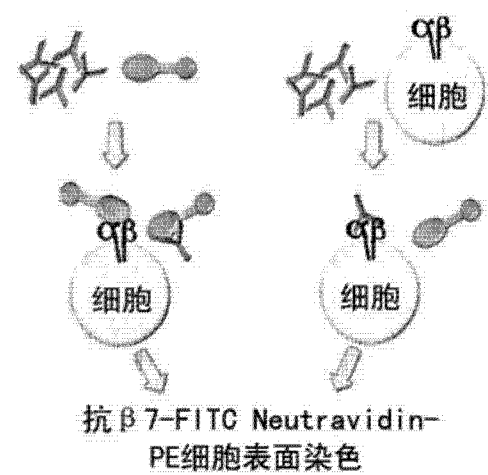


图 57

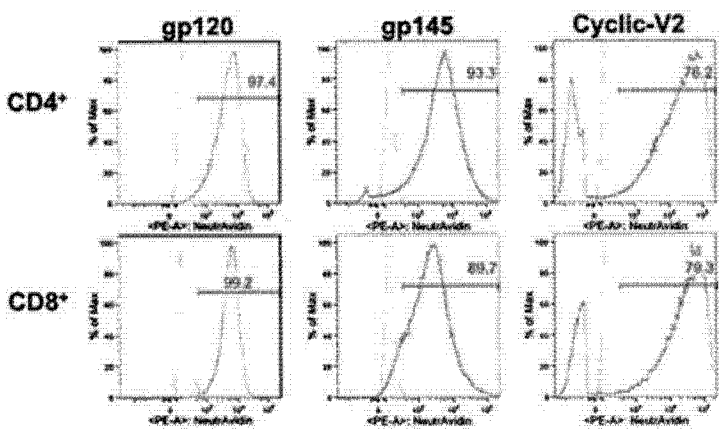


图 58

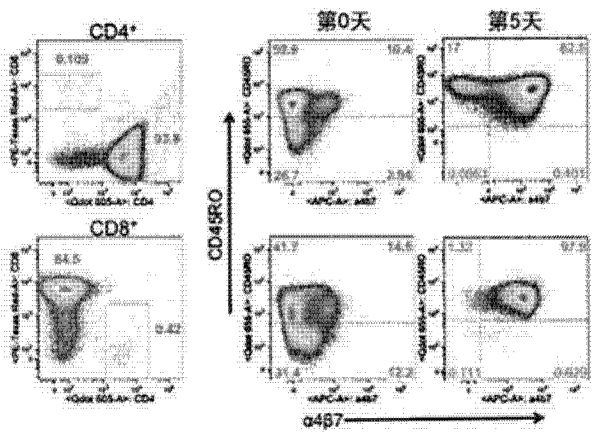


图 59

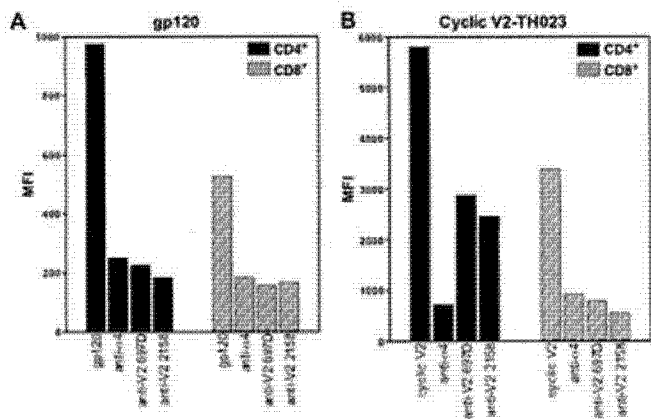


图 60

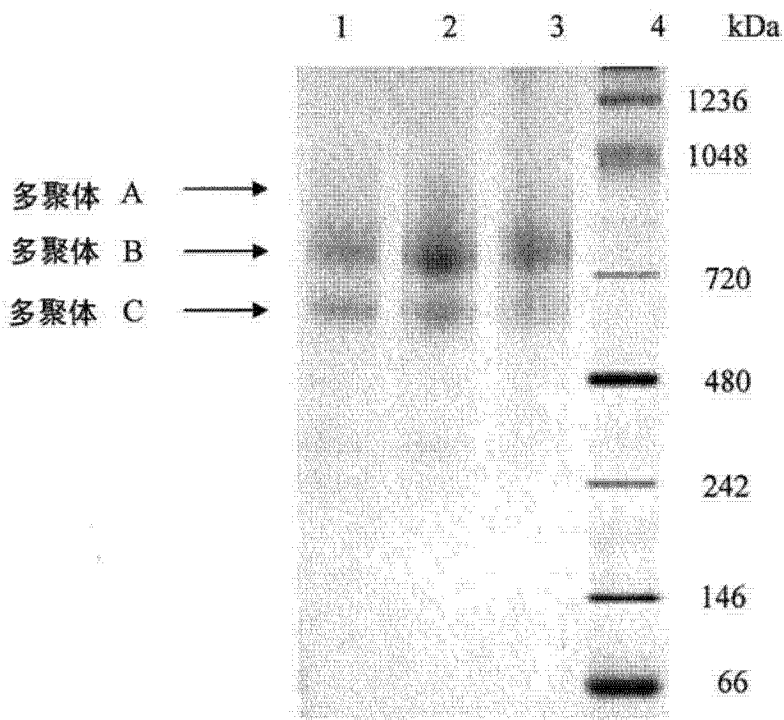


图 61

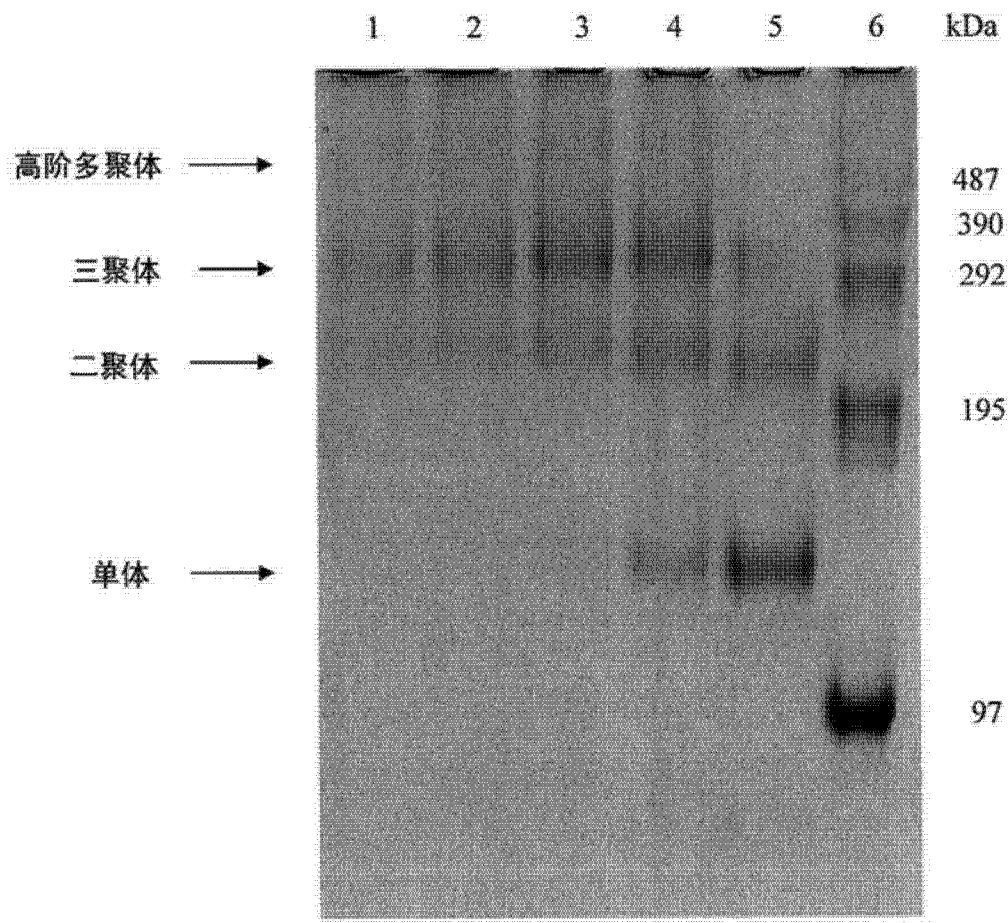


图 62

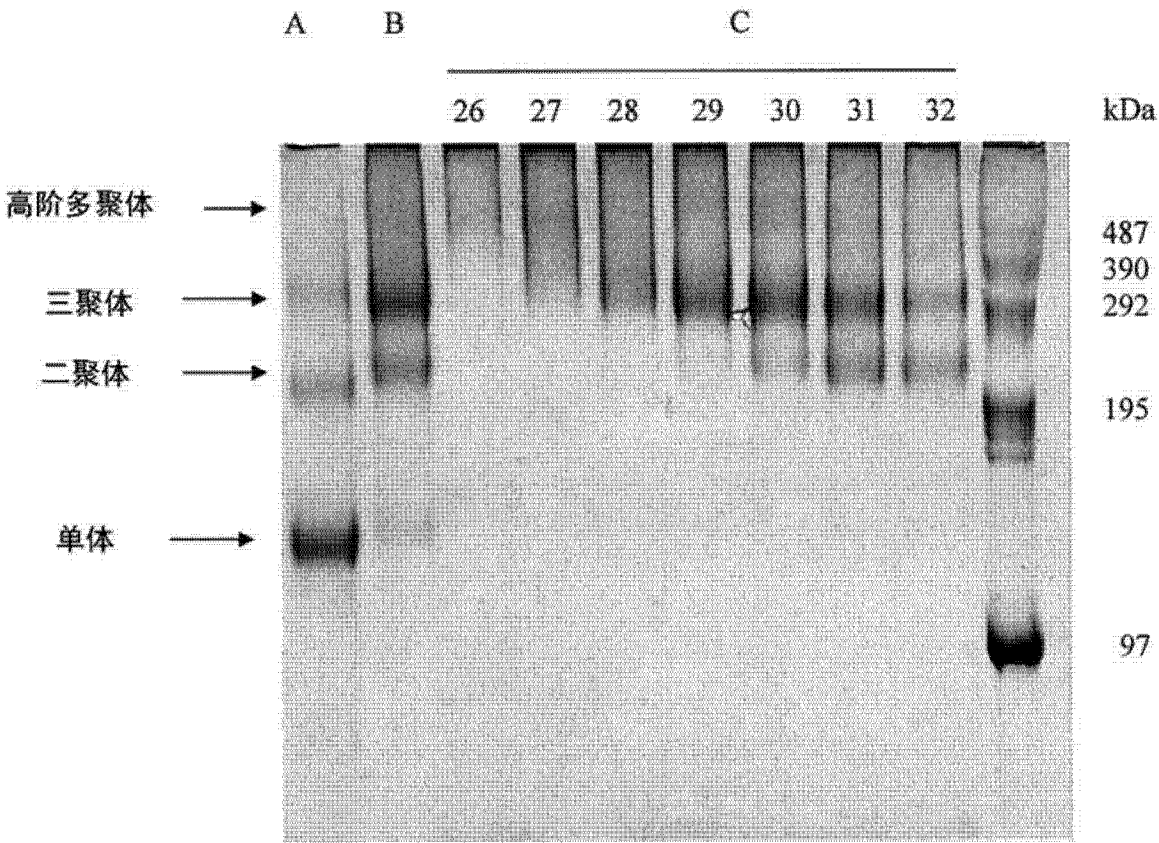


图 63

Abstract

The instant application provides methods and related compositions pertaining to novel HIV envelope proteins. In some embodiments, the invention relates to methods and compositions for the preparation, production, and administration of isolated novel HIV envelope nucleic acid and protein sequences suitable, for example, as vaccines against HIV.