(19) DANMARK

(10) **DK/EP 3584252 T3**

(12) Oversættelse af europæisk patentskrift

Patent- og Varemærkestyrelsen

(51) Int.Cl.: ***C 07 K 14/16 (2006.01)***

(45) Oversættelsen bekendtgjort den: **2021-11-15**

(80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2021-08-25**

(86) Europæisk ansøgning nr.: **19186284.6**

(86) Europæisk indleveringsdag: **2016-12-15**

(87) Den europæiske ansøgnings publiceringsdag: **2019-12-25**

(30) Prioritet: **2015-12-15 EP 15200138** **2016-10-17 EP 16194124**

(62) Stamansøgningsnr: **16822410.3**

(84) Designerede stater: **AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR**

(73) Patenthaver: **Janssen Vaccines & Prevention B.V., Archimedesweg 4, 2333 CN Leiden, Holland**

(72) Opfinder: **LANGEDIJK, Johannes, Petrus, Maria, Archimedesweg 4-6, 2333 CN Leiden, Holland**
CALLENDRET, Benoit, C.,S., Archimedesweg 4-6, 2333 CN Leiden, Holland
VAN MANEN, Danielle, Archimedesweg 4-6, 2333 CN Leiden, Holland
KRARUP, Anders, Archimedesweg 4-6, 2333 CN Leiden, Holland
STITZ, Jörn, Archimedesweg 4-6, 2333 CN Leiden, Holland
WEGMANN, Frank, Archimedesweg 4-6, 2333 CN Leiden, Holland
VELLINGA, Jort, Archimedesweg 4-6, 2333 CN Leiden, Holland

(74) Fuldmægtig i Danmark: **RWS Group, Europa House, Chiltern Park, Chiltern Hill, Chalfont St Peter, Bucks SL9 9FG, Storbritannien**

(54) Benævnelse: **HUMAN IMMUNDEFEKTVIRUS-ANTIGENER, -VEKTORER, -SAMMENSÆTNINGER OG FREMGANGSMÅDER TIL ANVENDELSE DERAF**

(56) Fremdragne publikationer:
WO-A1-2010/059732
WO-A1-2014/107744
"RecName: Full=Endogenous retrovirus group K member 9 Env polyprotein {ECO:0000256|SAAS:SAAS00159347}; AltName: Full=Endogenous retrovirus group K member 113 Env polyprotein {ECO:0000256|SAAS:SAAS00159454}; AltName: Full=Endogenous retrovirus group K member 13-1 Env polyprotein {ECO:0000256|SAAS:SAAS", UNIPROT, 1 March 2005 (2005-03-01), XP002757477,
COMPANS R W ET AL: "Recombinant protein gp41 heterologous transmembrane region, SEQ ID 1", GENESEQ,, 15 January 2009 (2009-01-15), XP002757478,
MARASCO W A: "Transmembrane domain peptide, SEQ ID 14", GENESEQ,, 26 January 2006 (2006-01-26), XP002757479,
"GCN4 fusion linker peptide, SEQ ID NO 3", GENESEQ,, 11 January 2007 (2007-01-11), XP002757480,

DESCRIPTION

BACKGROUND OF THE INVENTION

[0001] Human Immunodeficiency Virus (HIV) affects millions of people worldwide, and the prevention of HIV through an efficacious vaccine remains a very high priority, even in an era of widespread antiretroviral treatment. HIV-1 is the most common and pathogenic strain of the virus, with more than 90% of HIV/AIDS cases deriving from infection with HIV-1 group M. The M group is subdivided further into clades or subtypes. An efficacious vaccine ideally would be capable of eliciting both potent cellular responses and broadly neutralizing antibodies capable of neutralizing HIV-1 strains from different clades.

[0002] The high genetic variability of HIV-1 makes the development of a HIV-1 vaccine an unprecedented challenge. In order to improve coverage of potential T-cell epitopes, and improve cellular responses, "mosaic" HIV-1 Gag, Pol and Env antigens, derived from HIV Group Antigen (Gag), Polymerase (Pol), and Envelope (Env) proteins, were described by others and developed in an attempt to provide maximal coverage of potential T-cell epitopes (e.g., Barouch et al, Nat Med 2010, 16: 319-323). The mosaic antigens are similar in length and domain structure to wild-type, naturally occurring HIV-1 antigens.

[0003] For example, mosaic HIV antigens described and used in vaccines include those described in Barouch et al, *supra*, and WO 2010/059732 such as:

1. (a) Gag mosaic antigens including:

   (a)(i) a first mosaic Gag sequence ("mos1Gag") having the amino acid sequence as set forth herein in SEQ ID NO: 1, and

   (a)(ii) a second mosaic Gag sequence ("mos2Gag") having the amino acid sequence as set forth herein in SEQ ID NO: 2;

2. (b) Pol mosaic antigens including:

   (b)(i) a first mosaic Pol sequence ("mos1Pol") having the amino acid sequence as set forth herein in SEQ ID NO: 3, and

   (b)(ii) a second mosaic Pol sequence ("mos2Pol") having the amino acid sequence as set forth herein in SEQ ID NO: 4; and

3. (c) Env mosaic antigens including:

   (c)(i) a first mosaic Env sequence ("mos1Env") having the amino acid sequence as set forth herein in SEQ ID NO: 5, and

   (c)(ii) a second mosaic Env sequence ("mos2Env") having the amino acid sequence as

set forth herein in SEQ ID NO: 6.

[0004] Sequences encoding these antigens have been cloned in vectors, for example, such as recombinant adenoviral vectors, e.g., recombinant adenovirus serotype 26 (rAd26), and these recombinant vectors were previously used as vaccines to generate immune responses to the antigens (see e.g. Barouch et al, *supra*; and WO 2010/059732). For example, the mos1Gag and mos1Pol mosaic antigen sequences are typically combined into a fusion protein of Gag and Pol ("mos1GagPol"), and the coding sequence of which is cloned into a first Ad26 vector ("rAd26.mos1GagPol"); and the mos2Gag and mos2Pol antigen sequences are combined into another fusion protein of Gag and Pol ("mos2GagPol"), and the coding sequence of which is cloned into a second Ad26 vector ("rAd26.mos2GagPol"). Constructs encoding mos1Env and mos2Env are typically cloned into separate Ad26 vectors ("rAd26.mos1Env" and "rAd26.mos2Env", respectively).

[0005] A set of such mosaic antigens as described above gives good global coverage of Group M HIV-1 isolates, where rAd26 vectors encoding mosaic 1 antigen sequences (e.g., rAd26.mos1GagPol and rAd26.mos1Env) favor clade B and CRF01 HIV-1 subtypes, and rAd26 vectors encoding mosaic 2 antigen sequences (e.g., rAd26.mos2GagPol and rAd26.mos2Env) favor clade C strains. Mosaic HIV-1 Gag, Pol, and Env antigens expressed in rAd26 vectors can be used to improve both the breadth and depth of antigen-specific T-lymphocyte responses in rhesus monkeys, without compromising the magnitude of both cellular and humoral responses when compared with consensus or natural sequence HIV-1 antigens (Barouch et al, *supra*; and WO 2010/059732).

[0006] However, upon further development efforts on the vaccine components described above, it was found that rAd26.mos2Env showed non-optimal cell surface expression and immune response in non-human primates, but moreover displayed a hitherto unreported, unexpected and unpredictable non-optimal genetic stability during the manufacturing process as compared to the other rAd26 vectors, such as rAd26.mos1Env. Thus, vaccines containing rAd26.mos2Env may result in non-optimal immune responses against Clade C HIV-1 subtypes, since the mos2Env mosaic antigen favors clade C HIV-1 strains. Accordingly, there is a need for an alternative to the mos2Env antigen in vaccines against HIV that can be used to induce improved immune responses against HIV-1 clade C.

BRIEF SUMMARY OF THE INVENTION

[0007] The invention relates to novel synthetic human immunodeficiency virus (HIV) envelope proteins that have improved cell surface expression and genetic stability as compared to the previously described mos2Env antigen. The invention also relates to compositions and methods of using such novel synthetic HIV envelope proteins and/or coding sequences thereof to induce increased immune responses against HIV-1, particularly HIV-1 clade C, preferably

when used in combination with other HIV antigens.

[0008] In one general aspect, the invention relates to a nucleic acid encoding a synthetic HIV envelope protein comprising the amino acid sequence of SEQ ID NO: 8, or SEQ ID NO:8 having one or more mutations selected from the group consisting of (i) I529P (i.e., a substitution of Ile to Pro at position 529 of SEQ ID NO:8), (ii) K480E (i.e., a substitution of Lys to Glu at position 480 of SEQ ID NO:8), and (iii) a combination of EK479-480RRRR (i.e. a replacement of Glu-Lys at positions 479-480 of SEQ ID NO:8 with four consecutive Arg residues), I529P, A471C (i.e., a substitution of Ala to Cys at position 471 of SEQ ID NO:8) and T575C (i.e., a substitution of Thr to Cys at position 575 of SEQ ID NO:8). In one embodiment, the synthetic HIV envelope protein further comprises a signal sequence, for instance a signal sequence having the amino acid sequence selected from the group consisting of SEQ ID NOs: 9-12. In one embodiment, the signal sequence has the amino acid sequence of SEQ ID NO: 9.

[0009] In certain embodiments, the synthetic HIV envelope protein further comprises a transmembrane domain, preferably a transmembrane domain having the amino acid sequence of SEQ ID NO: 13.

[0010] In certain embodiments, the synthetic HIV envelope protein further comprises a fragment of a cytoplasmic domain, preferably a fragment of a cytoplasmic domain comprising the amino acid sequence of SEQ ID NO: 14, or amino acids 1-4 thereof (i.e., NRVR). In embodiments wherein the synthetic HIV envelope protein further comprises a transmembrane domain and a fragment of a cytoplasmic domain, it is preferred that the protein also comprises the amino acid sequence of SEQ ID NO: 37, which is fused to the carboxyl-terminus (C-terminus) of SEQ ID NO:8 and the amino-terminus (N-terminus) of the transmembrane region.

[0011] In another embodiment, the synthetic HIV envelope protein comprises a trimerization domain, for instance, a trimerization domain comprising the amino acid sequence of SEQ ID NO: 15 (GCN4) or SEQ ID NO:16 (foldon domain). In one preferred embodiment, the trimerization domain comprises the amino acid sequence of SEQ ID NO: 15. Such embodiments with trimerization domains are useful for soluble (i.e. non membrane-bound) synthetic HIV envelope proteins based on the ectodomain sequences provided herein, such as that comprising the amino acid sequence of SEQ ID NO: 8, wherein the trimerization domain is located at the C-terminus of the synthetic HIV envelope protein.

[0012] In yet other embodiments, the synthetic HIV envelope protein comprises SEQ ID NO: 8 with the following mutations: EK479-480RRRR, I529P, A471C and T575C. The introduction of 6 consecutive arginine residues (positions 478 and 481 in the native sequence of SEQ ID NO: 8 already are Arg residues) results in a further optimized furin cleavage site, so that an improved processed (i.e. cleaved) ectodomain is obtained. The three mutations of I529P, A471C and T575C are known as SOSIP mutations, wherefrom the last two mutations result in introduction of a possible disulfide bridge between the newly created cysteine residues. Overall, these mutations result in a soluble, trimerized, synthetic HIV envelope protein, without necessity for a trimerization domain.

[0013] In a preferred embodiment, the invention relates to a nucleic acid encoding a synthetic HIV envelope protein comprising the amino acid sequence of SEQ ID NO: 17, SEQ ID NO: 18, or aa 1-686 of SEQ ID NO: 19. Most preferably the synthetic HIV envelope protein encoded by the nucleic acid comprises or consists of the amino acid sequence of SEQ ID NO: 18.

[0014] In another general aspect, the invention relates to a vector comprising a nucleic acid encoding a synthetic HIV envelope protein according to an embodiment of the invention. In one embodiment, the vector is a viral vector. In a preferred embodiment, the viral vector is an adenoviral vector. In one preferred embodiment, the adenoviral vector is an adenovirus 26 vector.

[0015] Another general aspect of the invention relates to a composition, preferably a vaccine composition, comprising an immunogenically effective amount of a vector according to an embodiment of the invention, and a carrier, wherein the nucleic acid encoding the synthetic HIV envelope protein is operably linked to a promoter sequence. In one embodiment, the composition comprises an adenovirus vector, preferably an adenovirus 26 vector, encoding a synthetic HIV envelope protein comprising the amino acid sequence of SEQ ID NO: 18.

[0016] In another general aspect, the invention relates to a vaccine combination for inducing an immune response against a human immunodeficiency virus (HIV) in a subject in need thereof. The vaccine combination comprises a first composition comprising an immunogenically effective amount of a vector, preferably an adenovirus vector, more preferably an adenovirus 26 vector, encoding a synthetic HIV envelope protein having the amino acid sequence of SEQ ID NO: 18, a second composition comprising an immunogenically effective amount of a second vector, preferably a second adenovirus vector, more preferably a second adenovirus 26 vector, encoding an HIV antigenic polypeptide comprising the amino acid sequence of SEQ ID NO: 5, and optionally at least one additional composition comprising an immunogenically effective amount of at least one selected from the group consisting of a vector encoding an antigenic polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NOs: 1-4, 28 and 29, and a polypeptide comprising an immunogenically effective amount of an isolated HIV antigenic polypeptide, including but not limited to, a polypeptide having residues 30-708 of the amino acid sequence of SEQ ID NO: 7, or a polypeptide having residues 30-724 of SEQ ID NO:36, wherein the first composition, second composition and optional additional composition are present in the same composition or in one or more different compositions.

[0017] Yet another general aspect of the invention relates to methods of inducing an immune response against a human immunodeficiency virus (HIV) in a subject in need thereof, comprising administering to the subject a composition or vaccine combination according to an embodiment of the invention. The invention also relates to methods of inducing an immune response against an HIV comprising priming and boosting the immune response using a composition or a vaccine combination according to an embodiment of the invention.

[0018] Yet a further aspect of the invention relates to a synthetic HIV envelope protein comprising the amino acid sequence of SEQ ID NO: 8, or SEQ ID NO: 8 having one or more mutations selected from the group consisting of (i) I529P, (ii) K480E, (iii) a combination of EK479-480RRRR, I529P, A471C and T575C. In one embodiment, the synthetic HIV envelope protein comprises SEQ ID NO:8 with the mutations of EK479-480RRRR, I529P, A471C and T575C. In another embodiment, the synthetic HIV envelope protein comprises residues 30-704 or 30-711 of the amino acid sequence of SEQ ID NO: 18. In yet another embodiment the synthetic HIV envelope protein comprises residues 30-686 of the amino acid sequence of SEQ ID NO:19.

[0019] Another aspect of the invention relates to a cell, preferably an isolated cell, comprising a vector according to an embodiment of the invention.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

[0020] The foregoing summary, as well as the following detailed description of the invention, will be better understood when read in conjunction with the appended drawings. It should be understood that the invention is not limited to the precise embodiments shown in the drawings.

[0021] In the drawings:

FIGS. 1A-1C are schematic representations of the structure of HIV envelope proteins; FIG. 1A shows a full length HIV envelope protein; FIG. 1B shows the structure of a soluble single chain HIV envelope protein according to an embodiment of the invention in which the transmembrane domain (TM) is replaced with a GCN4 trimerization domain, and the furin cleavage site is mutated (sC4); FIG. 1C shows the structure of a membrane bound HIV envelope protein according to an embodiment of the invention comprising a transmembrane domain and a fragment of a cytoplasmic domain (C4D7);

FIG. 2 shows expression levels of the soluble sC1 envelope protein, which is based on the mos2Env mosaic antigen sequence with an additional C-terminal trimerization domain, and a soluble synthetic HIV envelope protein (sC4) according to an embodiment of the invention; expression was measured by quantitative Western blot using a polyclonal antibody against gp120; plasmids encoding sC1 or sC4 were transiently expressed twice, and each transfection was quantified twice by densitometry; the sC1 protein showed very low expression levels compared to the sC4 synthetic HIV envelope protein, which showed relatively high expression levels;

FIGS. 3A and 3B show the binding of synthetic HIV envelope proteins with monoclonal antibody 17b (mAb17b) in the presence (light gray) and absence (dark gray) of soluble CD4 as determined by ELISA assay; FIG. 3A shows binding of sC1; FIG. 3B shows binding of sC4;

FIG. 4 is an image of a Western blot from a native polyacrylamide gel electrophoresis of the sC1 protein, and the sC4 synthetic HIV envelope protein;

FIG. 5 shows the relative cell surface expression levels of the membrane-bound C1, C1D7, C4 and C4D7 synthetic HIV envelope proteins by FACS analysis of cells expressing these proteins using an anti-gp120 polyclonal antibody (GP120), and by binding to broadly neutralizing antibodies PG9 (PG9) and PG16 (PG16) that are quaternary-structure dependent and preferentially bind to correctly folded Env trimer;

FIG. 6 is a graphical representation of the stability of adenovirus vectors containing sequences encoding synthetic HIV envelope proteins of the invention including full-length C4 (FLC4), C4D7, and sC4 after multiple viral passages; recombinant adenovirus 26 vectors were generated in PER.C6 cells; after the initial 3 passages for transfection and plaque purification, 5 plaques were selected and upscaled for 10 passages in T25 format, resulting in a total viral passage number (vpn) of 13; the stability after vpn 3, 5, 10, and 13 as determined by E1 transgene cassette polymerase chain reaction (PCR) is shown; for example, 3/5 means 3 plaques were stable out of 5 plaques tested, and 5/5 means 5 plaques were stable out of 5 plaques tested; and

FIGS. 7A and 7B show virus neutralization titers against HIV-1 envelope pseudotyped virus particles (EVPs) in a TZM-bl cell-based neutralization assay in rabbits; log10-transformed IC_{50} values of the high-adenoviral vector dosed groups were measured against EVPs VSV-G (negative control) and MW965.26 (Tier 1A clade C) at weeks 1, 8, 14, and 20; each dot represents the log10-transformed IC_{50} value of an individual rabbit, with the group mean indicated by a horizontal line; HD: Highest Dilution tested (upper solid line); LD: Lowest Dilution tested (lower solid line); LOB: limit of background, 95 percentile value of compiled negative samples (dotted line); Log10 IC_{50} values exceeding the LD or HD threshold were set at the corresponding line; a one-way non-parametric comparison with control using the Dunn method for joint ranking was done for each time point; statistically significant differences are indicated in the graphs: * = $P<0.05$, ** = $P<0.01$, and *** = $P<0.001$; FIG. 7A shows the results with VSV-G (negative control); and FIG. 7B shows the results with MW965.26 (Tier 1A clade C).

DETAILED DESCRIPTION OF THE INVENTION

[0022] Discussion of documents, acts, materials, devices, articles or the like which has been included in the present specification is for the purpose of providing context for the invention. Such discussion is not an admission that any or all of these matters form part of the prior art with respect to any inventions disclosed or claimed.

[0023] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this invention pertains. Otherwise, certain terms used herein have the meanings as set forth in the specification. It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise.

[0024] As used herein, "subject" means any animal, preferably a mammal, most preferably a human, to whom will be or has been administered a vector, composition or combination vaccine according to embodiments of the invention. The term "mammal" as used herein, encompasses any mammal. Examples of mammals include, but are not limited to, cows, horses, sheep, pigs, cats, dogs, mice, rats, rabbits, guinea pigs, monkeys, humans, etc., more preferably a human.

[0025] The invention generally relates to synthetic HIV envelope proteins, nucleic acid and vectors encoding the synthetic HIV envelope proteins, and methods of inducing an immune response against HIV with vectors encoding the synthetic HIV envelope proteins or the synthetic HIV envelope proteins, alone or in combination with one or more additional vectors encoding one or more additional HIV antigenic polypeptides and/or in combination with one or more additional isolated HIV antigenic polypeptides.

[0026] Human immunodeficiency virus (HIV) is a member of the genus Lentivirinae, which is part of the family of Retroviridae. Two species of HIV infect humans: HIV-1 and HIV-2. HIV-1 is the most common strain of HIV virus, and is known to be more pathogenic than HIV-2. As used herein, the terms "human immunodeficiency virus" and "HIV" refer, but are not limited to, HIV-1 and HIV-2.

[0027] HIV is categorized into multiple clades with a high degree of genetic divergence. As used herein, the term "HIV clade" or "HIV subtype" refers to related human immunodeficiency viruses classified according to their degree of genetic similarity. There are currently three groups of HIV-1 isolates: M, N and O. Group M (major strains) consists of at least ten clades, A through J. Group O (outer strains) can consist of a similar number of clades. Group N is a new HIV-1 isolate that has not been categorized in either group M or O.

[0028] As used herein, the terms "HIV antigenic polypeptide," "HIV antigenic protein," and "HIV immunogen" refer to a polypeptide capable of inducing an immune response, e.g., a humoral and/or cellular mediated response, against HIV in a subject. The antigenic polypeptide can be a protein of the HIV, a fragment or epitope thereof, or a combination of multiple HIV proteins or portions thereof, that can induce an immune response or produce an immunity, e.g., protective immunity, against the HIV in a subject.

[0029] Preferably, an antigenic polypeptide is capable of raising in a host a protective immune response, e.g., inducing an immune response against a viral disease or infection, and/or producing an immunity in (i.e., vaccinatcs) a subject against a viral disease or infection, that protects the subject against the viral disease or infection. For example, the antigenic polypeptide can comprise a protein or fragments thereof from Simian Immunodeficiency Virus (SIV) or an HIV, such as the HIV or SIV envelope gp160 protein, the HIV or SIV matrix/capsid proteins, and the HIV or SIV *gag, pol* and *env* gene products.

[0030] An HIV antigenic polypeptide can be any HIV-1 or HIV-2 antigen or fragment thereof.

Examples of HIV antigens include, but are not limited to *gag, pol,* and *env* gene products, which encode structural proteins and essential enzymes. *Gag, pol,* and *env* gene products are synthesized as polyproteins, which are further processed into multiple other protein products. The primary protein product of the *gag* gene is the viral structural protein gag polyprotein, which is further processed into MA, CA, SP1, NC, SP2, and P6 protein products. The *pol* gene encodes viral enzymes (Pol, polymerase), and the primary protein product is further processed into RT, RNase H, IN, and PR protein products. The *env* gene encodes structural proteins, specifically glycoproteins of the virion envelope. The primary protein product of the *env* gene is gp160, which is further processed into gp120 and gp41. Other examples of HIV antigens include gene regulatory proteins Tat and Rev; accessory proteins Nef, Vpr, Vif and Vpu; capsid proteins, nucleocapsid proteins, and p24 viral protein.

[0031] In certain embodiments, the HIV antigenic polypeptide comprises an HIV Gag, Env, or Pol antigen, or any antigenic portion or epitope or combination thereof, preferably an HIV-1 Gag, Env, or Pol antigen or any antigenic portion or epitope or combination thereof.

[0032] HIV antigenic polypeptides can also be mosaic HIV antigens. As used herein, "mosaic antigen" refers to a recombinant protein assembled from fragments of natural sequences. Mosaic antigens resemble natural antigens, but are optimized to maximize the coverage of potential T-cell epitopes found in the natural sequences, which improves the breadth and coverage of the immune response. Mosaic HIV antigens for use with the invention are preferably mosaic Gag, Pol, and/or Env antigens, and more preferably a mosaic HIV-1 Gag, Pol, and/or Env antigens. As used herein, "a mosaic HIV Gag, Pol, and/or Env antigen" specifically refers to a mosaic antigen comprising multiple epitopes derived from one or more of the Gag, Pol and/or Env polyprotein sequences of HIV.

[0033] In one embodiment, a mosaic HIV antigen for use with the invention is a mosaic HIV Gag antigen with epitopes derived from the sequences of *gag* gene products (examples are provided in SEQ ID NOs: 1, 2); a mosaic HIV Pol antigen with epitopes derived from the sequences of *pol* gene products (examples are provided in SEQ ID NOs: 3, 4); or a mosaic HIV Env antigen with epitopes derived from the sequences of *env* gene products (examples arc provided in SEQ ID NOs: 5, 6; also the novel antigens of the invention, e.g. in SEQ ID NOs: 8, 17, 18, 19, can be considered mosaic HIV Env antigens). In certain embodiments, a mosaic HIV antigen for use with the invention may comprise a combination of epitopes derived from sequences of *gag, pol,* and/or *env* gene products. Illustrative and non-limiting examples include mosaic Env-Pol antigens with epitopes derived from the sequences of *env* and *pol* gene products; mosaic Gag-Pol antigens with epitopes derived from the sequences of *gag* and *pol* gene products; and mosaic Gag-Env antigens with epitopes derived from the sequences of *gag* and *env* gene products. The sequences of *gag, pol,* and *env* gene products can be derived from one or more clades.

[0034] Examples of mosaic HIV Gag, Pol and/or Env antigens that can be used in the invention include those described in, e.g., US20120076812; Barouch et al., Nat Med 2010, 16:319-323; and Barouch et al., Cell 155:1-9, 2013, all of which are incorporated herein by reference in

their entirety. Preferably, mosaic HIV Gag, Pol, and/or Env antigens for use with the present invention include, but are not limited to, mos1Gag (SEQ ID NO: 1), mos2Gag (SEQ ID NO: 2), mos1Pol (SEQ ID NO: 3), mos2Pol (SEQ ID NO: 4), mos1Env (SEQ ID NO: 5), mos2Env (SEQ ID NO: 6), mos1GagPol (SEQ ID NO: 28), mos2GagPol (SEQ ID NO: 29), and combinations thereof.

[0035] As used herein, each of the terms "HIV envelope protein," "env protein," and "Env" refers to a protein that is expressed on the envelope of an HIV virion and enables an HIV to target and attach to the plasma membrane of HIV infected cells, or a fragment or derivative thereof that can induce an immune response or produce an immunity against the HIV in a subject in need thereof. The HIV *env* gene encodes the precursor protein gp160, which is proteolytically cleaved into the two mature envelope glycoproteins, gp120 and gp41. The cleavage reaction is mediated by a host cell protease, furin, at a sequence highly conserved in retroviral envelope glycoprotein precursors. More specifically, gp160 trimerizes to $(gp\ 160)_3$ and then undergoes cleavage into the two noncovalently associated gp120 and gp41. Viral entry is subsequently mediated by a trimer of gp120/gp41 heterodimers. Gp120 is the receptor binding fragment, and binds to the CD4 receptor on a target cell that has such a receptor, such as, e.g., a T-helper cell. Gp41, which is noncovalently bound to gp120, is the fusion fragment and provides the second step by which HIV enters the cell. Gp41 is originally buried within the viral envelope, but when gp120 binds to a CD4 receptor, gp120 changes its conformation causing gp41 to become exposed, where it can assist in fusion with the host cell. Gp140 is the uncleaved ectodomain of trimeric gp160, i.e., $(gp160)_3$, that has been used as a surrogate for the native state of the clcavcd, viral spike.

[0036] According to embodiments of the invention, an "HIV envelope protein" can be a gp160, gp140, gp120, gp41 protein, combinations, fusions, truncations or derivatives thereof. For example, an "HIV envelope protein" can include a gp120 protein noncovalently associated with a gp41 protein. It can also include a stabilized trimeric gp140 protein that can have or can be modified to include a trimerization domain that stabilizes trimers of gp140. Examples of trimerization domains include, but are not limited to, the T4-fibritin "foldon" trimerization domain; the coiled-coil trimerization domain derived from GCN4; and the catalytic subunit of *E. coli* aspartate transcarbamoylase as a trimer tag. An "HIV envelope protein" can also be a truncated HIV envelope protein including, but not limited to, envelope proteins comprising a C-terminal truncation in the ectodomain (i.e. the domain that extends into the extracellular space), a truncation in the gp41, such as a truncation in the transmembrane domain of gp41, or a truncation in the cytoplasmic domain of gp41. An "HIV envelope protein" can further be a derivative of a naturally occurring HIV envelope protein having sequence mutations, e.g., in the furin cleavage sites, and/or so-called SOSIP mutations.

[0037] Preferably, an "HIV envelope protein" is a "synthetic HIV envelope protein." As used herein, the term "synthetic HIV envelope protein" refers to a non-naturally occurring HIV envelope protein that is optimized to induce an immune response or produce an immunity against one or more naturally occurring HIV strains in a subject in need thereof. Mosaic HIV Env proteins are examples of synthetic HIV Env proteins, and the invention provides novel

synthetic HIV Env antigens, e.g. the ones comprising SEQ ID NOs: 8, 17, 18, or 19.

Synthetic HIV envelope proteins and coding sequences thereof

[0038] Embodiments of the invention relate to novel synthetic HIV envelope proteins and nucleic acid molecules encoding these.

[0039] In one embodiment, the invention relates to a synthetic HIV envelope protein comprising the amino acid sequence of SEQ ID NO: 8, or SEQ ID NO:8 having one or more mutations selected from the group consisting of (i) I529P, (ii) K480E, and (iii) a combination of EK479-480RRRR, I529P, A471C and T575C. SEQ ID NO:8 comprises a synthetic mature gp120 and a synthetic truncated gp41 without the transmembrane region, nor the cytoplasmic domain. SEQ ID NO:8 is a non-naturally occurring sequence comprised of a chimera of sequences from the mos2Env mosaic antigen (SEQ ID NO: 6), and other HIV envelope protein sequences. The sequence of the novel synthetic Env antigen comprising SEQ ID NO:8 is optimized to provide broad coverage and an enhanced T-cell response against HIV clade C (as compared to the mos2Env antigen (SEQ ID NO: 6)). In certain embodiments, further amino acids can be added to SEQ ID NO: 8 or one of its variants defined herein.

[0040] In certain embodiments, the synthetic HIV envelope protein further comprises a signal sequence. The synthetic HIV envelope protein is synthesized with a signal sequence that is cleaved from the nascent polypeptide chain during its transport into the lumen of the endoplasmic reticulum (ER). In principle, any known signal sequence could be used. Preferably an HIV Env signal sequence or a variant thereof is used. Different signal sequences have been used in the art for HIV Env proteins (see e.g. WO 2014/107744). In certain embodiments, the signal sequence comprises SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11 or SEQ ID NO:12. In one preferred embodiment, the signal sequence comprises SEQ ID NO: 9.

[0041] In certain embodiments, the synthetic HIV envelope protein further comprises a transmembrane domain. The transmembrane domain anchors the synthetic HIV envelope protein to the ER membrane, and contributes to membrane assembly and function of the HIV envelope. Preferably, the transmembrane domain comprises SEQ ID NO:13.

[0042] In another embodiment, the synthetic HIV envelope protein comprises a gp41 having a truncated cytoplasmic domain. The gp41 has an unusually long cytoplasmic domain at its carboxyl end, typically about 150 amino acids (Edwards et al., J. Virology, 2002, 76:2683-2691). Truncation of the cytoplasmic domain was reported to induce exposure of conserved regions in the ectodomain of HIV-1 Env protein (*Id.*). The truncated cytoplasmic domain in a synthetic HIV envelope of the invention can range from one to about 140 amino acids, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, or 140 amino acids of a full-length cytoplasmic domain. In certain embodiments the truncated cytoplasmic domain is derived from amino acids 704-862 of SEQ ID NO: 17 (i.e. from the cytoplasmic domain of the C4 molecule of the invention), by truncation after a given amino acid up to the C-

terminus. In a preferred embodiment, the synthetic HIV envelope protein comprises a truncated cytoplasmic domain having 1 to 10 amino acids residues, more preferably 4 to 8 amino acid residues, and most preferably 7 amino acid residues of an HIV gp41 cytoplasmic domain. The cytoplasmic domain or fragment thereof of a synthetic HIV envelope protein is located C-terminal to the extracellular domain (ectodomain), and when the synthetic HIV envelope protein also comprises a transmembrane domain, the cytoplasmic domain or fragment thereof is located C-terminal to the transmembrane domain. See, e.g., FIGS. 1A and 1C. In a particular embodiment, the synthetic HIV envelope protein comprises a gp41 with a truncated cytoplasmic domain having the amino acid sequence of SEQ ID NO:14 or a fragment thereof, such as residues 1-4 thereof (i.e. NRVR). Other truncated cytoplasmic domains have been described and could be used (e.g. Schiernle et al., PNAS 1997; Abrahamyan et al., J Virol 2005).

[0043] In embodiments wherein the synthetic HIV envelope protein further comprises a transmembrane domain and a fragment of a cytoplasmic domain, it is preferred that the protein also comprises the amino acid sequence of SEQ ID NO: 37, which contains residues 655-682 of SEQ ID NO: 18, wherein the amino acid sequence of SEQ ID NO: 37 is fused to the C-terminus of SEQ ID NO: 8 and the N-terminus of the transmembrane domain.

[0044] In a particularly preferred embodiment of the invention, the synthetic HIV envelope protein further comprises a transmembrane domain, such as that having the amino acid sequence of SEQ ID NO:13, and a truncated cytoplasmic domain or a fragment of cytoplasmic domain, such as that having the amino acid sequence of SEQ ID NO: 14 or residues 1-4 of SEQ ID NO:14 (i.e., NRVR). Most preferably, the synthetic HIV envelope protein comprises or consists of the amino acid sequence of SEQ ID NO: 18, with or without the signal sequence (i.e., amino acid resides 1-29 of SEQ ID NO:18).

[0045] In another embodiment, the synthetic HIV envelope protein comprises a trimerization domain that replaces an Env transmembrane region. The trimerization domain increases the stability of an Env trimeric structure. Preferably, the synthetic HIV envelope protein comprises a gp140 polypeptide that is modified to include a trimerization domain that stabilizes trimers of gp140. Examples of trimerization domains include, but are not limited to, the T4-fibritin "foldon" trimerization domain, such as that comprising the amino acid sequence of SEQ ID:16; the coiled-coil trimerization domain derived from GCN4, such as that comprising the amino acid sequence of SEQ ID:15; the catalytic subunit of *E. coli* aspartate transcarbamoylase as a trimer tag; or matrillin-based trimerization motifs. If present, the trimerization domain typically is located C-terminal to the extracellular domain (see FIG. 1B). In certain preferred embodiments where the synthetic HIV envelope protein comprises a trimerization domain, the synthetic HIV envelope protein comprises the amino acid sequence of SEQ ID NO: 19, with or without the signal sequence (i.e., amino acid residues 1-29 of SEQ ID NO:19). These embodiments with trimerization domains are mainly useful for soluble ectodomain variants of the synthetic HIV envelope protein. In certain embodiments of such soluble variants of the invention, it is possible to mutate the furin cleavage site (e.g. mutation of Lys to Glu at position 480 in SEQ ID NO: 8) to inactivate this cleavage site, so that the protein will be a single chain; this combines

well with a trimerization domain, especially with the GCN4 trimerization domain of SEQ ID NO: 19.

[0046] Alternative versions of such soluble ectodomain variants of the synthetic HIV envelope protein without use of trimerization domains are also embodiments of the invention, and can be prepared from SEQ ID NO: 8 by combining mutations that optimize the furin cleavage site (replacing the Gly-Lys dipeptide at positions 479-480 by four Arg residues) as well as so-called SOSIP mutations (I>P mutation at position 529, and introduction of a disulfide bridge between positions 471 and 575 by replacement of the respective Ala and Thr on those positions in SEQ ID NO: 8 each with a Cys residue). This yields a protein having the amino acid sequence of SEQ ID NO: 8 with the following combination of mutations: EK479-480RRRR, I529P, A471C and T575C.

[0047] One possible modification to further increase the trimer content of a synthetic HIV envelope protein of the invention (comprising SEQ ID NO: 8), is modification of Ile to Pro at position 529. This can be effective for both soluble and membrane-bound variants.

Vectors

[0048] Another general aspect of the invention relates to vectors comprising nucleic acid encoding a synthetic HIV envelope protein. According to embodiments of the invention, the vectors can comprise any of the synthetic HIV envelope proteins described herein. In a preferred embodiment of the invention, the vector comprises nucleic acid encoding a synthetic HIV envelope protein comprising the amino acid sequence of SEQ ID NO: 8, SEQ ID NO:17, SEQ ID NO: 18, or SEQ ID NO: 19, and more preferably SEQ ID NO: 18.

[0049] According to embodiments of the invention, the nucleic acid encoding the synthetic HIV envelope protein is operably linked to a promoter, meaning that the nucleic acid is under the control of a promoter. The promoter can be a homologous promoter (i.e., derived from the same genetic source as the vector) or a heterologous promoter (i.e., derived from a different vector or genetic source). Examples of suitable promoters include the cytomegalovirus (CMV) promoter and the Rous Sarcoma virus (RSV) promoter. Preferably, the promoter is located upstream of the nucleic acid within an expression cassette. An exemplary CMV promoter sequence that can be operably linked to nucleic acid encoding the synthetic HIV envelope protein is shown in SEQ ID NO: 24.

[0050] According to embodiments of the invention, a vector can be an expression vector. Expression vectors include, but are not limited to, vectors for recombinant protein expression and vector for delivery of nucleic acid into a subject for expression in a tissue of the subject, such as a viral vector. Examples of viral vectors suitable for use with the invention include, but are not limited to adenoviral vectors, adeno-associated virus vectors, pox virus vectors, MVA vectors, enteric virus vectors, Venezuelan Equine Encephalitis virus vectors, Semliki Forest Virus vectors, Tobacco Mosaic Virus vectors, lentiviral vectors, etc. The vector can also be a

non-viral vector. Examples of non-viral vectors include, but are not limited to plasmids, bacterial artificial chromosomes, yeast artificial chromosomes, bacteriophages, etc.

[0051] In certain embodiments of the invention, the vector is an adenovirus vector. An adenovirus according to the invention belongs to the family of the Adenoviridae, and preferably is one that belongs to the genus Mastadenovirus. It can be a human adenovirus, but also an adenovirus that infects other species, including but not limited to a bovine adenovirus (e.g. bovine adenovirus 3, BAdV3), a canine adenovirus (e.g. CAdV2), a porcine adenovirus (e.g. PAdV3 or 5), or a simian adenovirus (which includes a monkey adenovirus and an ape adenovirus, such as a chimpanzee adenovirus or a gorilla adenovirus). Preferably, the adenovirus is a human adenovirus (HAdV, or AdHu), or a simian adenovirus such as chimpanzee or gorilla adenovirus (ChAd, AdCh, or SAdV). In the invention, a human adenovirus is meant if referred to as Ad without indication of species, e.g. the brief notation "Ad26" means the same as HadV26, which is human adenovirus serotype 26. Also, as used herein, the notation "rAd" means recombinant adenovirus, e.g., "rAd26" refers to recombinant human adenovirus 26.

[0052] Most advanced studies have been performed using human adenoviruses, and human adenoviruses are preferred according to certain aspects of the invention. In certain preferred embodiments, a recombinant adenovirus according to the invention is based upon a human adenovirus. In preferred embodiments, the recombinant adenovirus is based upon a human adenovirus serotype 5, 11, 26, 34, 35, 48, 49, 50, 52, etc. According to a particularly preferred embodiment of the invention, an adenovirus is a human adenovirus of serotype 26. An advantage of this serotypes is a low seroprevalence and/or low pre-existing neutralizing antibody titers in the human population, and experience with use in human subjects in clinical trials.

[0053] Simian adenoviruses generally also have a low seroprevalence and/or low pre-existing neutralizing antibody titers in the human population, and a significant amount of work has been reported using chimpanzee adenovirus vectors (e.g. US6083716; WO 2005/071093; WO 2010/086189; WO 2010085984; Farina et al, 2001, J Virol 75: 11603-13 [13]; Cohen et al, 2002, J Gen Virol 83: 151-55 [69]; Kobinger et al, 2006, Virology 346: 394-401 [70]; Tatsis et al., 2007, Molecular Therapy 15: 608-17 [71]; see also review by Bangari and Mittal, 2006, Vaccine 24: 849-62 [72]; and review by Lasaro and Ertl, 2009, Mol Ther 17: 1333-39 [73]). Hence, in other embodiments, the recombinant adenovirus according to the invention is based upon a simian adenovirus, e.g. a chimpanzee adenovirus. In certain embodiments, the recombinant adenovirus is based upon simian adenovirus type 1, 7, 8, 21, 22, 23, 24, 25, 26, 27.1, 28.1, 29, 30, 31.1, 32, 33, 34, 35.1, 36, 37.2, 39, 40.1, 41.1, 42.1, 43, 44, 45, 46, 48, 49, 50 or SA7P.

[0054] Preferably, the adenovirus vector is a replication deficient recombinant viral vector, such as rAd26, rAd35, rAd48, rAd5HVR48, etc.

[0055] In a preferred embodiment of the invention, the adenoviral vectors comprise capsid

proteins from rare serotypes including Ad26. In the typical embodiment, the vector is an rAd26 virus. An "adenovirus capsid protein" refers to a protein on the capsid of an adenovirus (*e.g.*, Ad26, Ad35, rAd48, rAd5HVR48 vectors) that is involved in determining the serotype and/or tropism of a particular adenovirus. Adenoviral capsid proteins typically include the fiber, penton and/or hexon proteins. As used herein a "capsid protein" for a particular adenovirus, such as an "Ad26 capsid protein" can be, for example, a chimeric capsid protein that includes at least a part of an Ad26 capsid protein. In certain embodiments, the capsid protein is an entire capsid protein of Ad26. In certain embodiments, the hexon, penton and fiber are of Ad26.

[0056] One of ordinary skill in the art will recognize that elements derived from multiple serotypes can be combined in a single recombinant adenovirus vector. Thus, a chimeric adenovirus that combines desirable properties from different serotypes can be produced. Thus, in some embodiments, a chimeric adenovirus of the invention could combine the absence of pre-existing immunity of a first serotype with characteristics such as temperature stability, assembly, anchoring, production yield, redirected or improved infection, stability of the DNA in the target cell, and the like.

[0057] In certain embodiments the recombinant adenovirus vector useful in the invention is derived mainly or entirely from Ad26 (*i.e.,* the vector is rAd26). In some embodiments, the adenovirus is replication deficient, e.g., because it contains a deletion in the E1 region of the genome. For adenoviruses being derived from non-group C adenovirus, such as Ad26 or Ad35, it is typical to exchange the E4-orf6 coding sequence of the adenovirus with the E4-orf6 of an adenovirus of human subgroup C such as Ad5. This allows propagation of such adenoviruses in well-known complementing cell lines that express the E1 genes of Ad5, such as for example 293 cells, PER.C6 cells, and the like (*see, e.g.* Havenga, et al., 2006, J Gen Virol 87: 2135-43; WO 03/104467). However, such adenoviruses will not be capable of replicating in non-complcmcnting cells that do not express the E1 genes of Ad5.

[0058] The preparation of recombinant adenoviral vectors is well known in the art. Preparation of rAd26 vectors is described, for example, in WO 2007/104792 and in Abbink et al., (2007) Virol 81(9): 4654-63. Exemplary genome sequences of Ad26 are found in GenBank Accession EF 153474 and in SEQ ID NO:1 of WO 2007/104792. Examples of vectors useful for the invention for instance include those described in WO2012/082918, the disclosure of which is incorporated herein by reference in its entirety.

[0059] Typically, a vector useful in the invention is produced using a nucleic acid comprising the entire recombinant adenoviral genome (*e.g.,* a plasmid, cosmid, or baculovirus vector). Thus, the invention also provides isolated nucleic acid molecules that encode the adenoviral vectors of the invention. The nucleic acid molecules of the invention can be in the form of RNA or in the form of DNA obtained by cloning or produced synthetically. The DNA can be double-stranded or single-stranded.

[0060] The adenovirus vectors useful in the invention are typically replication deficient. In these embodiments, the virus is rendered replication deficient by deletion or inactivation of regions

critical to replication of the virus, such as the E1 region. The regions can be substantially deleted or inactivated by, for example, inserting a gene of interest, such as a gene encoding a synthetic HIV envelope protein (usually linked to a promoter), or a gene encoding an HIV antigenic polypeptide (usually linked to a promoter) within the region. In some embodiments, the vectors of the invention can contain deletions in other regions, such as the E2, E3 or E4 regions, or insertions of heterologous genes linked to a promoter within one or more of these regions. For E2- and/or E4-mutated adenoviruses, generally E2- and/or E4-complementing cell lines are used to generate recombinant adenoviruses. Mutations in the E3 region of the adenovirus need not be complemented by the cell line, since E3 is not required for replication.

[0061] A packaging cell line is typically used to produce sufficient amounts of adenovirus vectors for use in the invention. A packaging cell is a cell that comprises those genes that have been deleted or inactivated in a replication deficient vector, thus allowing the virus to replicate in the cell. Suitable packaging cell lines for adenoviruses with a deletion in the E1 region include, for example, PER.C6, 911, 293, and E1 A549.

[0062] According to embodiments of the invention, and as noted above, any of the synthetic HIV envelope proteins described herein can be expressed in the vectors of the invention. In view of the degeneracy of the genetic code, the skilled person is well aware that several nucleic acid sequences can be designed that encode the same protein, according to methods entirely routine in the art. The nucleic acid encoding the synthetic HIV envelope protein can optionally be codon-optimized to ensure proper expression in the treated host (*e.g.*, human). Codon-optimization is a technology widely applied in the art. Some non-limiting examples of sequences encoding a synthetic HIV envelope protein of the invention are provided in SEQ ID NOs: 25, 26 and 27. Typically, the nucleic acid encoding the synthetic HIV envelope protein is cloned into the E1 and/or the E3 region of the adenoviral genome.

[0063] In a preferred embodiment of the invention, the vector is an adenovirus vector, and more preferably a rAd26 vector, most preferably a rAd26 vector with at least a deletion in the E1 region of the adenoviral genome, e.g. such as that described in Abbink, J Virol, 2007. 81(9): p. 4654-63, which is incorporated herein by reference.

[0064] The invention also provides cells, preferably isolated cells, comprising any of the vectors described herein. The cells can be used for recombinant protein production, or for the production of viral particles.

[0065] Embodiments of the invention thus also relate to a method of a making a synthetic HIV antigenic polypeptide. The method comprises transfecting a host cell with an expression vector comprising nucleic acid encoding the synthetic HIV antigenic polypeptide operably linked to a promoter, growing the transfected cell under conditions suitable for expression of the synthetic HIV antigenic polypeptide, and isolating the synthetic HIV antigentic polypeptide from the cell. The synthetic HIV antigenic polypeptide can be isolated or collected from the cell by any method known in the art including affinity chromatography, etc. Techniques used for recombinant protein expression will be well known to one of ordinary skill in the art in view of

the present disclosure.

[0066] The invention also includes a method for manufacturing a vector encoding a synthetic HIV antigenic polypeptide of the invention, the method comprising culturing a cell that comprises the vector, to propagate and multiply the vector during said culturing, and isolating the vector that encodes the synthetic HIV antigenic polypeptide of the invention from the cell culture, e.g. from the cells, from the culture medium, or both. The vector may be further purified according to methods known in the art.

[0067] In certain embodiments, the invention provides a vector according to an embodiment of the invention comprising a nucleic acid encoding a synthetic HIV antigenic polypeptide, and in certain exemplary embodiments the nucleic acid has a nucleotide sequence selected from the group consisting of SEQ ID NO: 25, 26 and 27.

Compositions

[0068] In another general aspect, the invention relates to a composition comprising a vector comprising a nucleic acid encoding a synthetic HIV envelope protein and a carrier. According to embodiments of the invention, any of vectors described herein can be included in the composition. Preferably, the vector is a viral vector, more preferably an adenovirus vector, and even more preferably an adenovirus 26 vector. In a preferred embodiment, a composition comprises an adenovirus vector, preferably an adenovirus 26 vector encoding a synthetic HIV envelope protein comprising the amino acid sequence of SEQ ID NO: 8, SEQ ID NO: 18, or SEQ ID NO: 19, and more preferably the amino acid sequence of SEQ ID NO: 18.

[0069] In one aspect, the invention provides a combination vaccine comprising one or more vectors together comprising nucleic acid sequences encoding (i) a synthetic HIV envelope protein comprising the amino acid sequence of SEQ ID NO: 8 (e.g. SEQ ID NO: 18 or 19) and (ii) a second HIV envelope protein comprising the amino acid sequence of SEQ ID NO: 5. The vectors may each be in a separate composition, or be combined in a single composition. Both nucleic acids in the vector(s) are intended to be administered to one subject, which will result in an immune response to HIV that is broader than the immune response that would be obtained upon administration of either vector alone. Both nucleic acid sequences could also be present on one single vector.

[0070] According to embodiments of the invention, a composition comprises an immunogenically effective amount of a vector, such as a viral vector. As used herein, "an immunogenically effective amount" or "immunologically effective amount" means an amount of a composition sufficient to induce a desired immune effect or immune response in a subject in need thereof. In one embodiment, an immunogenically effective amount means an amount sufficient to induce an immune response in a subject in need thereof. In another embodiment, an immunogenically effective amount means an amount sufficient to produce immunity in a subject in need thereof, e.g., provide a protective effect against a disease such as a viral

infection. An immunogenically effective amount can vary depending upon a variety of factors, such as the physical condition of the subject, age, weight, health, etc.; the particular application, whether inducing immune response or providing protective immunity; the specific recombinant vector administered; the immunogen or antigenic polypeptide encoded by the recombinant vector administered; the specific antigenic polypeptide administered; and the particular disease, e.g., viral infection, for which immunity is desired. An immunogenically effective amount can readily be determined by one of ordinary skill in the art in view of the present disclosure.

[0071] As general guidance, an immunogenically effective amount when used with reference to a recombinant viral vector such as an adenviral vector can range from about 10^8 viral particles to about 10^{12} viral particles, for example 10^8, 10^9, 10^{10}, 10^{11}, or 10^{12} viral particles. An immunogenically effective amount can be administered in a single composition, or in multiple compositions, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 compositions (e.g., tablets, capsules or injectables), wherein the administration of the multiple capsules or injections collectively provides a subject with the immunogenically effective amount. In general, when used with reference to a polypeptide, such as an isolated antigenic polypeptide, an immunogenically effective amount can range from, e.g. about 0.3 to about 3000 microgram (µg), e.g. 1-1000 µg, e.g. 10-500 µg, e.g. about 10, 50, 100, 150, 200, 250, 300, 350, 400, 450 or 500 µg. As a non-limiting example, it is possible to combine administration of the vector encoding the synthetic HIV Env antigen of the invention (having SEQ ID NO: 8) with administration of an Env polypeptide, e.g. 250 µg of HIV Clade C Env trimer protein having amino acids 30-708 of SEQ ID NO: 7. It is also possible to administer an immunogenically effective amount to a subject, and subsequently administer another dose of an immunogenically effective amount to the same subject, in a so-called prime-boost regimen. This general concept of a prime-boost regimen is well known to the skill person in the vaccine field. Further booster administrations can optionally be added to the regimen, as needed.

[0072] Compositions of the invention further comprise a carrier. A carrier can include one or more pharmaceutically acceptable excipients such as binders, disintegrants, swelling agents, suspending agents, emulsifying agents, wetting agents, lubricants, flavorants, sweeteners, preservatives, dyes, solubilizers and coatings. The precise nature of the carrier or other material can depend on the route of administration, e.g., intramuscular, subcutaneous, oral, intravenous, cutaneous, intramucosal (e.g., gut), intranasal or intraperitoneal routes. For liquid injectable preparations, for example, suspensions and solutions, suitable carriers and additives include water, glycols, oils, alcohols, preservatives, coloring agents and the like. For solid oral preparations, for example, powders, capsules, caplets, gelcaps and tablets, suitable carriers and additives include starches, sugars, diluents, granulating agents, lubricants, binders, disintegrating agents and the like. For nasal sprays/inhalant mixtures, the aqueous solution/suspension can comprise water, glycols, oils, emollients, stabilizers, wetting agents, preservatives, aromatics, flavors, and the like as suitable carriers and additives.

[0073] Compositions of the invention can be formulated in any matter suitable for administration to a subject to facilitate administration and improve efficacy, including, but not

limited to, oral (enteral) administration and parenteral injections. The parenteral injections include intravenous injection or infusion, intra-arterial injection, subcutaneous injection, intramuscular injection, and intra-articular injection. Compositions of the invention can also be formulated for other routes of administration including transmucosal, ocular, rectal, long acting implantation, sublingual administration, under the tongue, from oral mucosa bypassing the portal circulation, inhalation, or intranasal.

[0074] According to certain embodiments of the invention, a composition comprises an immunogenically effective amount of purified or partially purified adenovirus vector, such as an adenovirus 26 vector, comprising a nucleic acid encoding a synthetic HIV envelope protein of the invention. Said compositions can be formulated as a vaccine (also referred to as an "immunogenic composition") according to methods well known in the art.

[0075] Compositions of the invention can further optionally comprise an adjuvant to enhance immune responses. The terms "adjuvant" and "immune stimulant" are used interchangeably herein, and are defined as one or more substances that cause stimulation of the immune system. In this context, an adjuvant is used to enhance an immune response to the vectors encoding synthetic HIV envelope proteins of the invention and/or HIV antigenic polypeptides used in combination with vectors encoding synthetic HIV envelope proteins of the invention.

[0076] Adjuvants suitable for use with the invention should be ones that are potentially safe, well tolerated and effective in people, such as for instance QS-21, Detox-PC, MPL-SE, MoGM-CSF, TiterMax-G, CRL- 1005, GERBU, TERamide, PSC97B, Adjumer, PG-026, GSK-I, GcMAF, B-alethine, MPC-026, Adjuvax, CpG ODN, Betafectin, aluminum salts (e.g. AdjuPhos), Adjuplex, and MF59. The optimal ratios of each component in the formulation can be determined by techniques well known to those skilled in the art in view of the present disclosure.

[0077] In a preferred embodiment, the adjuvant is an aluminum salt, such as AdjuPhos.

[0078] The preparation and use of immunogenic compositions are well known to those of ordinary skill in the art. Liquid pharmaceutical compositions generally include a liquid carrier such as water, petroleum, animal or vegetable oils, mineral oil or synthetic oil. Physiological saline solution, dextrose or other saccharide solution or glycols such as ethylene glycol, propylene glycol or polyethylene glycol can also be included.

[0079] For instance, recombinant adenovirus vector may be stored in the buffer that is also used for the Adenovirus World Standard (Hoganson et al., 2002, Bioprocessing J 1: 43-8): 20 mM Tris pH 8, 25 mM NaCl, 2.5% glycerol. Another useful adenovirus formulation buffer suitable for administration to humans is 20 mM Tris, 2 mM $MgCl_2$, 25 mM NaCl, sucrose 10% w/v, polysorbate-80 0.02% w/v. Another formulation buffer that is suitable for recombinant adenovirus comprises 10-25 mM citrate buffer pH 5.9-6.2, 4-6% (w/w) hydroxypropyl-beta-cyclodextrin (HBCD), 70-100 mM NaCl, 0.018-0.035% (w/w) polysorbate-80, and optionally 0.3-0.45% (w/w) ethanol. Obviously, many other buffers can be used, and several examples of

suitable formulations for the storage and for pharmaceutical administration of purified vectors are known.

[0080] According to embodiments of the invention, a composition of the invention can be used together with one or more additional vectors encoding one or more additional HIV antigenic polypeptides, and/or one or more isolated HIV antigenic polypeptides. The additional vectors and/or HIV antigenic polypeptides can be present in the same composition comprising a synthetic HIV Env protein of the invention. They can also be present in one or more different compositions that can be used together with a composition comprising a synthetic HIV Env protein of the invention in a vaccine combination. Preferably, the one or more additional vectors are viral vectors, such as adenovirus vectors, and are most preferably adenovirus 26 vectors. The one or more additional vectors can encode any HIV antigenic polypeptide known to those skilled in the art in view of the present disclosure.

[0081] In one embodiment, a composition or a vaccine combination further comprises a second adenovirus vector, preferably an adenovirus 26 vector, encoding a HIV antigenic polypeptide comprising the amino acid sequence of SEQ ID NO: 5. An advantage of such embodiments is increased breadth of the immune response (covering strains from Clades B and C).

[0082] In another embodiment, a composition or a vaccine combination of the invention further comprises an adenovirus vector, preferably an adenovirus 26 vector, encoding an HIV antigenic polypeptide comprising the amino acid sequence of SEQ ID NO: 28 (mos1GagPol).

[0083] In another embodiment, a composition or a vaccine combination of the invention further comprises an adenovirus vector, preferably an adenovirus 26 vector, encoding an HIV antigenic polypeptide comprising the amino acid sequence of SEQ ID NO: 29 (mos2GagPol).

[0084] In a particular embodiment, a composition or a vaccine combination of the invention further comprises a second adenovirus vector, preferably an adenovirus 26 vector, encoding a HIV antigenic polypeptide comprising the amino acid sequence of SEQ ID NO: 5, and one or more additional adenovirus vectors, preferably adenovirus 26 vectors, encoding one or more HIV antigenic polypeptides comprising the amino acid sequence selected from the group consisting of SEQ ID NO: 28 or SEQ ID NO: 29. For example, a composition or a vaccine combination according to an embodiment of the invention can comprise four adenovirus vectors, preferably adenovirus 26 vectors, with a first vector encoding a synthetic HIV envelope protein comprising the amino acid sequence of SEQ ID NO: 8 (e.g. SEQ ID NO: 18); a second vector encoding a HIV antigenic polypeptide comprising the amino acid sequence of SEQ ID NO: 5; a third vector encoding a synthetic HIV envelope protein comprising the amino acid sequence of SEQ ID NO: 28; and a fourth vector encoding a synthetic HIV envelope protein comprising the amino acid sequence of SEQ ID NO: 29.

[0085] In some embodiments, the composition or a vaccine combination further comprises one or more isolated HIV antigenic polypeptides. Any HIV antigenic polypeptide known to those

skilled in the art in view of the present disclosure can be further included in a composition or a vaccine combination of the invention, including, but not limited to an HIV envelope protein (e.g., gp160, gp140, gp120, or gp41), preferably a stabilized trimeric gp140 protein, such as a stabilized clade C or clade A gp140 protein. In a preferred embodiment, the isolated HIV antigenic polypeptide is a stabilized HIV clade C trimeric gp140 protein, such as that comprising residues 30-708 of the amino acid sequence of SEQ ID NO:7 (residues 1-29 of SEQ ID NO:7 are in the signal sequence). An alternative or additional HIV Env polypeptide that could be used in addition to the clade C gp140 protein or alone, is a mosaic Env trimer protein, for instance having an amino acid sequence as disclosed in amino acids 30-724 of SEQ ID NO: 36 (corresponding to SEQ ID NO: 2 of WO 2014/107744, residues 1-29 of SEQ ID NO:36 are in the signal sequence).

[0086] According to a particular embodiment of the invention, an HIV antigenic protein can be a synthetic HIV envelope protein of the invention. Thus, a synthetic envelope protein of the invention can be used in isolated and/or purified form to induce an immune response or provide a protective immunity, etc. against HIV in a subject in need thereof. Any of the synthetic envelope proteins described herein comprising the amino acid sequence of SEQ ID NO: 8 can be used as an HIV antigenic protein in isolated and/or purified form. In a preferred embodiment, when used in isolated form as an HIV antigenic protein, the synthetic envelope protein comprises residues 30-711 of the amino acid sequence of SEQ ID NO: 18 or residues 30-686 of the amino acid sequence of SEQ ID NO: 19, and more preferably residues 30-704 of the amino acid sequence of SEQ ID NO: 18. The isolated HIV antigenic polypeptide can also comprise SEQ ID NO: 8 with the following mutations: EK479-480RRRR, I529P, A471C and T575C.

[0087] Embodiments of the invention also relate to compositions or vaccine combinations comprising an isolated synthetic HIV envelope protein comprising the amino acid sequence of SEQ ID NO: 8. Any of the synthetic HIV envelope proteins described herein can be used. In particular embodiments of the invention, the isolated synthetic HIV envelope protein comprises residues 30-704 or 30-711 of the amino acid sequence of SEQ ID NO:18, residues 30-686 of the amino acid sequence of SEQ ID NO: 19, or the amino acid sequence of SEQ ID NO: 8 with the following mutations: EK479-480RRRR, I529P, A471C and T575C. Such compositions or vaccine combinations can further comprise one or more expression vectors, e.g., adenoviral vectors such as adenovirus 26 vectors, encoding one or more additional HIV antigenic polypeptides, such as the synthetic HIV envelope proteins of the invention, or other HIV antigenic proteins such as those set forth in SEQ ID NOs: 4, 5, 7, 28 or 29, or fragments thereof.

[0088] The invention also relates to a method of producing a composition or a vaccine combination of the invention. According to embodiments of the invention, a method of producing a composition or a combination comprises combining a vector comprising nucleic acid encoding the synthetic HIV envelope protein of the invention with a carrier, and optionally one or more additional vectors encoding one or more additional HIV antigenic polypeptides and/or one or more isolated HIV antigenic polypeptides. One of ordinary skill in the art will be

familiar with conventional techniques used to prepare such compositions.

Vaccine and Vaccine Combinations

[0089] According to embodiments of the invention, a composition can be a vaccine. As used herein, the term "vaccine" refers to a composition comprising an expression vector, preferably a viral vector, encoding a synthetic HIV envelope protein of the invention that can provide protective immunity or a protective immune response to a subject, or to vaccinate a subject. According to embodiments of the invention, upon administration of the composition to a subject, the expression vector expresses the encoded synthetic HIV envelope protein, and the expressed synthetic HIV envelope protein is presented to the immune system of the subject, thereby inducing the required response to produce immunity, or induce an immune response.

[0090] Thus, in another general aspect, the invention provides a vaccine for inducing an immune response against a human immunodeficiency virus (HIV) in a subject. According to embodiments of the invention, the vaccine comprises a composition comprising an immunogenically effective amount of an expression vector encoding a synthetic HIV envelope protein comprising the amino acid sequence of SEQ ID NO: 8, and preferably the amino acid sequence of SEQ ID NO:18. Preferably, the expression vector is a viral vector, more preferably an adenovirus vector, and most preferably an adenovirus 26 vector.

[0091] According to embodiments of the invention, "inducing an immune response" when used with reference to the methods and compositions described herein encompasses providing protective immunity and/or vaccinating a subject against an infection, such as a HIV infection, for prophylactic purposes, as well as causing a desired immune response or effect in a subject in need thereof against an infection, such as a HIV infection, for therapeutic purposes. Preferably, the methods of the invention are for prophylactic purposes, such as for providing protective immunity. The immune response can be a cellular immune response and/or a humoral immune response.

[0092] As used herein, the term "protective immunity" or "protective immune response" means that the vaccinated subject is able to control an infection with the pathogenic agent against which the vaccination was done. Usually, the subject having developed a "protective immune response" develops only mild to moderate clinical symptoms or no symptoms at all. Usually, a subject having a "protective immune response" or "protective immunity" against a certain agent will not die as a result of the infection with said agent.

[0093] According to embodiments of the invention, vaccine compositions can further comprise one or more additional vectors, e.g., viral vectors, such as adenovirus vectors, preferably adenovirus 26 vectors, encoding one or more additional HIV antigenic polypeptides and/or one or more isolated HIV antigenic polypeptides. The synthetic HIV envelope protein, additional vectors and/or one or more isolated HIV antigenic polypeptides can be formulated in the same composition or one or more different compositions in the vaccine.

[0094] The invention also relates to vaccine combinations for priming and boosting an immune response to one or more HIV clades in a subject in need thereof using one or more vectors in combination with an isolated antigenic polypeptide. Thus, in another general aspect, the invention provides a vaccine combination for inducing an immune response against a HIV in a subject. According to embodiments of the invention, the vaccine combination comprises:

1. (i) a first composition comprising an immunogenically effective amount of an expression vector encoding a synthetic HIV envelope protein comprising the amino acid sequence of SEQ ID NO: 8 or SEQ ID NO: 8 having one or more mutations selected from the group consisting of (a) I529P, (b) K480E, and (c) a combination of EK479-480RRRR, I529P, A471C and T575C, and a carrier; and
2. (ii) a second composition comprising an immunogenically effective amount of an isolated HIV antigenic polypeptide and a carrier,

wherein one of the first and second compositions is for priming immunization and the other composition is for boosting immunization.

[0095] According to embodiments of the invention, the vaccine combination optionally further comprises an immunogenically effective amount of one or more additional expression vectors encoding one or more additional HIV antigenic polypeptides. The one or more additional expression vectors can be included in the first composition or the second composition, or the one or more additional expression vectors can be included in one or more additional compositions to be administered together with the first and/or second composition.

[0096] As used herein, the terms "co-delivery", "co-administration" or "administered together with" refers to simultaneous administration of two or more components, such as a viral expression vector and an isolated antigenic polypeptide, or multiple viral expression vectors. "Simultaneous administration" can be administration of the two or more components at least within the same day. When two components are "administered together with," they can be administered in separate compositions sequentially within a short time period, such as 24, 20, 16, 12, 8 or 4 hours, or within 1 hour or less, or they can be administered in a single composition at the same time.

[0097] In particular embodiments of a vaccine combination of the invention, the first composition comprises an adenovirus vector, preferably an adenovirus 26 vector, encoding a synthetic HIV envelope protein comprising the amino acid sequence of SEQ ID NO: 18; and the isolated HIV antigenic polypeptide comprises residues 30-708 of the amino acid sequence of SEQ ID NO: 7 or residues 30-724 of SEQ ID NO: 36. In one particular embodiment, the first composition further comprises an adenovirus vector, preferably an adenovirus 26 vector, encoding a HIV antigenic polypeptide comprising the amino acid sequence of SEQ ID NO: 5. In another particular embodiment, the first composition further comprises one or more additional adenovirus vectors, preferably adenovirus 26 vectors, encoding one or more additional HIV antigenic polypeptides comprising the amino acid sequences selected from the group consisting of SEQ ID NOs: 28 and 29.

[0098] Another general aspect of the invention relates to a kit comprising a vaccine combination according to an embodiment of the invention.

[0099] Other embodiments of the synthetic HIV envelope protein, expression vectors, additional expression vectors, HIV antigenic polypeptides encoded by the expression vectors, and isolated HIV antigenic polypeptide etc. that can be used in the vaccine combinations of the invention are discussed in detail above and in the illustrative examples below.

Method for Inducing Protective Immunity Against HIV Infection

[0100] The invention also relates to a method of inducing an immune response against one or more HIV clades in a subject in need thereof. The methods described herein include methods of priming and boosting an immune response using one or more expression vectors in combination with one or more isolated antigenic polypeptides.

[0101] In one general aspect, a method of inducing an immune response against a human immunodeficiency virus (HIV) in a subject comprises administering to the subject a composition comprising an immunogenically effective amount of an expression vector comprising a nucleic acid encoding a synthetic HIV envelope protein comprising the amino acid sequence of SEQ ID NO: 8. Any of the compositions described herein can be used in a method of inducing an immune response against HIV in a subject. Preferably, the composition comprises an adenovirus vector, preferably an adenovirus 26 vector, comprising a nucleic acid encoding a synthetic HIV envelope protein comprising the amino acid sequence of SEQ ID NO: 18. The composition can further comprise one or more additional vectors encoding one or more additional HIV antigenic polypeptides and/or one or more additional isolated HIV antigenic polypeptides.

[0102] In another general aspect, a method of inducing an immune response against a human immunodeficiency virus (HIV) in a subject comprises:

1. (i) administering to the subject a first composition comprising an immunogenically effective amount of an expression vector encoding a mosaic HIV envelope protein having the amino acid sequence of SEQ ID NO: 8 or SEQ ID NO: 8 having one or more mutations selected from the group consisting of (a) I529P, (b) K480E, and (c) a combination of EK479-480RRRR, I529P, A471C and T575C, and a carrier;
2. (ii) administering to the subject a second composition comprising an immunogenically effective amount of an isolated HIV antigenic polypeptide and a carrier; and
3. (iii) optionally, administering to the subject an immunogenically effective amount of one or more additional expression vectors encoding one or more additional HIV antigenic polypeptides,

wherein steps (i) and (ii) are conducted in either order, with one of the steps for priming

immunization and the other step for boosting immunization. According to embodiments of the invention, the optional, effective amount of the one more additional expression vectors is administered together with the first composition or the second composition. In a preferred embodiment, the optional effective amount of the one or more additional expression vectors is administered together with the first composition.

[0103] In a particular embodiment of a method of inducing an immune response, the first composition comprises an adenovirus vector, preferably an adenovirus 26 vector, encoding a synthetic HIV envelope protein comprising the amino acid sequence of SEQ ID NO: 8 and a second adenovirus vector, preferably an adenovirus 26 vector, encoding a HIV antigenic polypeptide comprising the amino acid sequence of SEQ ID NO: 5; the second composition comprises an isolated HIV antigenic polypeptide having residues 30-708 of the amino acid sequence of SEQ ID NO:7 or residues 30-724 of SEQ ID NO:36; and the one or more additional expression vectors are adenovirus vectors, preferably adenovirus 26 vectors, encoding one or more additional HIV antigenic polypeptides comprising the amino acid sequences selected from the group consisting of SEQ ID NOs: 28 and 29; wherein the first composition is administered to the subject, together with the one or more additional expression vectors, one or more times for priming immunization, and the second composition is administered to the subject one or more times for boosting immunization.

[0104] Administration of the immunogenic compositions comprising the expression vectors and/or antigenic polypeptides is typically intramuscular, intradermal or subcutaneous. However, other modes of administration such as intravenous, rectal, cutaneous, oral, nasal, etc can be envisaged as well. Intramuscular administration of the immunogenic compositions can be achieved by using a needle to inject a suspension of the expression vectors, e.g. adenovirus vectors, and/or antigenic polypeptides. An alternative is the use of a needleless injection device to administer the composition (using, e.g., Biojector™) or a freeze-dried powder containing the vaccine.

[0105] For intramuscular, intravenous, cutaneous or subcutaneous injection, or injection at the site of affliction, the vector will be in the form of a parenterally acceptable aqueous solution which is pyrogen-free and has suitable pH, isotonicity and stability. Likewise, the isolated antigenic polypeptide will be in the form of a parenterally acceptable solution having a suitable pH, isotonicity, and stability. Those of ordinary skill in the art are well able to prepare suitable solutions using, for example, isotonic vehicles such as Sodium Chloride Injection, Ringer's Injection, Lactatcd Ringer's Injection. Preservatives, stabilizers, buffers, antioxidants and/or other additives can be included, as required. A slow-release formulation can also be employed.

[0106] Typically, administration of the vaccine compositions according to embodiments of the invention will have a prophylactic aim to generate an immune response against an HIV antigen before infection or development of symptoms. In other embodiments, the expression vectors, e.g., adenovirus vectors, and/or HIV antigenic polypeptides can be administered for post-exposure prophylactics.

[0107] The immunogenic compositions containing the expression vectors, e.g., adenovirus vectors, and/or antigenic polypeptides are administered to a subject, giving rise to an anti-HIV immune response in the subject. An amount of a composition sufficient to induce a detectable immune response is defined to be an "immunogenically effective dose" or "immunogenically effective amount." In a typical embodiment of the invention, the immune response is a protective immune response.

[0108] The actual amount administered, and rate and time-course of administration, will depend on the nature and severity of what is being treated. Prescription of treatment, e.g., decisions on dosage etc., is within the responsibility of general practitioners and other medical doctors, or in a veterinary context a veterinarian, and typically takes account of the disorder to be treated, the condition of the individual patient, the site of delivery, the method of administration and other factors known to practitioners. Examples of the techniques and protocols mentioned above can be found in Remington's Pharmaceutical Sciences, 16th edition, Osol, A. ed., 1980.

[0109] Following production of adenovirus vectors and optional formulation of such particles into compositions, the vectors can be administered to an individual, particularly a human or other primate. Delivery to a non-human mammal need not be for a therapeutic purpose, but can be for use in an experimental context, for instance in investigation of mechanisms of immune responses to the synthetic HIV envelope protein expressed by the adenovirus vectors of the invention.

[0110] In one embodiment of the disclosed methods, one or more adenovirus vectors encoding one or more HIV antigenic polypeptides are used to prime the immune response. One or more isolated HIV antigenic polypeptides can be used together with the one or more adenovirus vectors for the priming immunization. The priming immunization may be administered only once but can optionally also be administered multiple times, for example, initial priming administration at time 0, followed by another priming administration about 4-14 weeks, e.g. 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 or 14 weeks, or any time in between, after the initial priming administration. One or more isolated HIV antigenic polypeptides optionally together with one or more additional adenovirus or other vectors encoding one or more additional HIV antigenic polypeptides can be used to boost the immune response. A boosting immunization can also be administered once or multiple times, for example, first at about 18-36, e.g. 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35 or 36 weeks, or any time in between, after the initial priming administration, followed by another boosting administration at about 36-52, e.g. 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59 or 60 weeks, or any time in between, after the initial priming administration. The immune response induced by the immunization is monitored.

[0111] Embodiments of the disclosed methods also contemplate shorter prime-boost regimens, meaning that the final boosting immunization is administered about 22-26 weeks after the initial priming administration. The priming immunization can be administered at week 0. The boosting immunization can be administered multiple times, for example, first at about 7-

9 weeks or 11-13 weeks, or at about 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 14 weeks, or any time in between, after the initial priming administration, followed by another boosting administration at about 22-26 weeks, or at about 16, 17, 18, 19, 20, 21, 22, 23, 24 ,25, 26, 27, or 28 weeks, or any time in between, after the initial priming administration. In certain embodiments, one or more isolated HIV antigenic polypeptides is administered together with the one or more adenovirus vectors for the priming and/or boosting immunization.

[0112] It is readily appreciated by those skilled in the art that the regimen for the priming and boosting administrations can be adjusted based on the measured immune responses after the administrations. For example, the boosting compositions are generally administered weeks or months after administration of the priming composition, for example, about 2-3 weeks or 4 weeks, or 8 weeks, or 16 weeks, or 20 weeks, or 24 weeks, or 28 weeks, or 30 weeks or 32 weeks or one to two years after administration of the priming composition.

[0113] According to embodiments of the invention, an adjuvant can be administered together with the isolated HIV antigenic polypeptide as part of the priming and/or boosting immunization. Any adjuvant can be used in view of the present disclosure, and in certain embodiments the adjuvant is an aluminum salt, such as AdjuPhos.

[0114] In a preferred embodiment of the invention, the adenovirus vectors used in the methods disclosed herein include a rAd26 vector. Preferably, an rAd26 vector encoding a synthetic HIV envelope protein comprising the amino acid sequence of SEQ ID NO: 18 or SEQ ID NO: 19, most preferably SEQ ID NO: 18, is used to prime the immune response, alone or in combination with one or more additional rAd26 vectors encoding one or more additional HIV antigenic polypeptides, such as mos1Env having the amino acid sequence of SEQ ID NO:5, and an isolated HIV antigenic polypeptide, such as that comprising residues 30-708 of the amino acid sequence of SEQ ID NO: 7 or residues 30-724 of SEQ ID NO: 36, is used to boost the immune response, or vice versa.

[0115] In one exemplary embodiment, an rAd26 vector encoding a synthetic HIV envelope protein comprising the amino acid sequence of SEQ ID NO: 18 is used to prime the immune response in combination with an rAd26 vector encoding an HIV antigenic polypeptide having the amino acid sequence of SEQ ID NO: 5. One or more additional rAd26 vectors encoding one or more additional HIV antigenic polypeptides having the amino acid sequences selected from the group consisting SEQ ID NOs: 1-4, 28 and 29 can also be administered together with the other rAd26 vectors to prime the immune response. The priming administration in certain embodiments is administered twice before any boosting immunization is administered. An isolated HIV antigenic polypeptide, such as that comprising residues 30-708 of the amino acid sequence of SEQ ID NO: 7 (preferably), or that comprising residues 30-724 of the amino acid sequence of SEQ ID NO:36, or a combination of at least two of such isolated HIV antigenic polypeptides, is then administered to boost the immune response, and is preferably administered more than once. Preferably, an adjuvant is further administered with the isolated HIV antigenic polypeptide in the boosting immunization.

[0116] In a particular embodiment, the immune response is primed by administration of four HIV antigens encoded on adenoviral vectors, preferably rAd26 vectors, the four antigens that are encoded being: (i) a synthetic HIV envelope protein comprising the amino acid sequence of SEQ ID NO: 18, (ii) polypeptide having the amino acid sequence of SEQ ID NO: 5, (iii) polypeptide having the amino acid sequence of SEQ ID NO: 28, and (iv) polypeptide having the amino acid sequence of SEQ ID NO: 29. Each of these four antigens can be encoded on a separate adenoviral vector, preferably a rAd26 vector, administered at a total dose of about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 $x10^{10}$ viral particles (vp), e.g. about $5x10^{10}$ vp (for all vectors together). The vectors may be pre-mixed, e.g. in a 1:1:1:1 ratio. The priming administration may be repeated after the initial priming administration, e.g. at 8, 9, 10, 11, 12, 13, 14, 15 or 16 weeks after the initial priming administration. In this embodiment, an immune response is boosted by administration of the same adenoviral vector vaccine used for the priming administration together with isolated HIV Env gp140 protein, e.g. clade C gp140 protein (comprising residues 30-708 of the amino acid sequence of SEQ ID NO: 7), or mosaic gp140 protein (comprising residues 30-724 of the amino acid sequence of SEQ ID NO:36), or clade C gp140 protein and mosaic gp140 protein, at a total dose of about 50-300 µg protein, e.g. 50, 100, 150, 200, 250, or 300 microgram, or any amount in between, of clade C gp140 protein, or e.g. 50, 100, 150, 200, 250, or 300 microgram, or any amount in between, of mosaic gp140 protein, or e.g. 50, 100, 150, 200, 250, or 300 microgram, or any amount in between, of a combination of clade C gp140 protein and mosaic gp140 protein (e.g. in a 1:1 ratio, either mixed together or separately administered). Preferably the gp140 protein is administered together with adjuvant, e.g. aluminium phosphate. The adenovirus plus gp140 protein administration to boost the immune response may be performed at about 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 or 30 weeks, or at any time in between, after the initial priming administration. The boost administration may be repeated, e.g. at about 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53 or 54 weeks, or any time in between, after the initial priming administration. All administrations according to this embodiment are preferably performed via the intramuscular route.

EXAMPLES

Example 1: Design of HIV envelope antigen sequences

[0117] Several HIV envelope antigen sequences were designed having sequence similarity to the mosaic HIV antigen mos2Env (SEQ ID NO: 6; previously also described in WO 2010/059732). The newly designed, membrane bound, sequences were based on (a combination of) fully natural wild-type sequences from HIV envelope proteins, or a chimera of mos2Env sequence and wild-type HIV envelope protein sequences. In addition to full length envelope protein sequences (see FIG. 1A), sequences having a C-terminal truncation of the cytoplasmic domain were also designed (see, e.g., FIG. 1C). See also e.g., Schiernle et al., PNAS 1997; Abrahamyan et al., J Virol 2005); Edwards et al., J. Virology, 2002, 76:2683-2691.

Soluble variants were also prepared by C-terminal truncation before the transmembrane (TM) region, which was replaced by a trimerization domain, such as a GCN4 trimerization domain (see, e.g., FIG. 1B). These soluble variants were further converted into a single chain variant by mutation of the furin-cleavage site, thus inhibiting the processing of the extracellular domain of the envelope protein into gp120 and gp41 subunits.

[0118] Of the all the constructs generated and tested, constructs based on C4 had the most optimal properties, e.g., good manufacturability, folding, immunogenicity, etc. and these were selected for further studies. A soluble variant of the C4 construct having a GCN4 trimerization domain in place of the transmembrane domain (sC4, FIG. 1B), and a variant comprising a 7-amino acid fragment of the cytoplasmic domain (C4D7, FIG. 1C) were also generated and tested in further studies. The amino acid sequences of C4, sC4, and C4D7 are shown in SEQ ID NOs: 17, 19, and 18, respectively. Sequences encoding these are shown in SEQ ID NOs: 25, 27, and 26, respectively. Construct C1 has an extracellular domain sequence based on the mos2Env sequence (SEQ ID NO: 6). A soluble variant of construct C1 having a GCN4 trimerization domain in place of the transmembrane domain (sC1), and a variant comprising a 7-amino acid fragment of the cytoplasmic domain (C1D7), similar to sC4 and C4D7 as shown in FIGS. 1B and 1C, respectively, were also generated. Construct C1 and its variants were used in further studies for comparison purposes, since these are essentially based on the mos2Env sequence of the prior art. The amino acid sequences of C1, sC1 and C1D7 are shown in SEQ ID NOs: 31, 30, and 32, respectively. Nucleic acid sequences encoding these are shown in SEQ ID NOs: 34, 33, and 35, respectively. Other constructs that were tested were less optimal than the ones based on construct C4, and were not taken into further development.

Example 2: Expression and Folding of Synthetic HIV envelope proteins

[0119] The expression level, folding, and cell-surface expression of synthetic HIV envelope proteins were measured.

Expression Levels

[0120] HEK293F cells were transiently transfected with a plasmid encoding the soluble synthetic HIV envelope proteins sC1 and sC4 as described in Example 1. Expression levels of the soluble protein were measured in the supernatant using quantitative Western blot (QWB). The results are shown in FIG. 2. The low expression levels for sC1 (which essentially corresponds to mos2Env with an added transmembrane domain) are in line with our recent insights for mos2Env. As demonstrated by the results, the sC4 variant of the invention showed significantly higher expression levels than the sC1 variant (control).

Protein Folding

[0121] Protein folding was tested by measuring the binding of soluble synthetic HIV envelope proteins to an antibody (MAb 17b) known to bind the co-receptor binding site of the HIV envelope protein, which is exposed only after binding of CD4, by enzyme-linked immunosorbent assay (ELISA). In particular, binding of purified sC4 was tested for binding to MAb 17b with prior binding of sC4 to CD4, and without prior binding of sC4 to CD4. Purified sC1 was used as a control. Binding of MAb 17b to sC4 without prior CD4 binding to the envelope protein is an indication of partially unfolded or pre-triggered envelope protein (i.e., an unstable Env that adopts the "open" conformation in the absence of CD4 binding). The results of the ELISA assay are shown in FIGS. 3A and 3B.

[0122] As shown in FIG. 3B, sC4 shows strong binding to MAb 17b with prior binding to CD4, but no detectable binding to MAb 17b without prior binding to CD4. In contrast, as shown in FIG. 3A, sC1 showed much lower binding to MAb 17 both with and without prior binding to CD4. The results suggest that sC4 has a correct folding pattern, with no exposure of the co-receptor binding site prior to CD4 binding.

[0123] Protein folding was also analyzed by native polyacrylamide gel electrophoresis (PAGE) of sC1 and sC4 to evaluate the quaternary structure of the soluble protein variants, and possible incorrect disulfide bridge formation between protomers. After electrophoresis on a native gel, protein in the gel was detected by Western blot analysis. As shown by the results in FIG. 4, the majority of sC4 is present in a trimeric state, which is the correct quaternary structure.

[0124] Taken together, the results of the protein folding experiments demonstrate that the sC4 soluble synthetic HIV envelope protein has the desired folding profile, which is improved as compared to the folding profile of the existing mos2Env antigen (represented by sC1).

Cell surface expression

[0125] Cell surface expression of the membrane-bound variants of HIV envelope proteins C1 (full length), C4 (full length, see FIG. 1A), C1D7, and C4D7 was also studied. HEK293T cells were transiently transfected with only eGFP-encoding plasmid (negative control, NC), or with eGFP-encoding plasmid together with an expression construct encoding an HIV envelope protein variant. Two days post-transfection, cells were subjected to fluorescence activated cell sorting (FACS)-analysis upon exposure to several poly- and monoclonal antibodies directed against gp120, and secondary antibodies, and then examined for envelope protein cell-surface expression levels. Quality of the envelope variants was assessed by determining the overall expression levels using an anti-gp120 polyclonal antibody, and by assessing relative binding of the broadly neutralizing antibodies PG9 and PG16, which are quaternary-structure dependent, and preferentially bind to correctly folded envelope trimer.

[0126] The results of the cell surface expression experiments are shown in FIG. 5. The surface expression levels of truncated variants C1D7 and C4D7 as measured using an anti-gp120 antibody, are much higher than the surface expression levels of their full length counterparts, C1 and C4, respectively. This confirms that deletion of 144 residues from the carboxy-terminus of Env increases envelope surface expression levels. The full length C4 construct of the invention also showed improved PG9 and PG16 binding as compared to full length C1, suggesting that the C4 envelope sequence is properly folded (i.e., a trimer) on the cell surface.

[0127] The results also demonstrate that the C1D7 variant, which is essentially Mos2Env with an added transmembrane domain and 7 amino acids of the cytoplasmic domain, can be surface-expressed on HEK293T cells. This is in contrast to the soluble construct in Ad26.mos2Env, which cannot be expressed at detectable levels on the surface when transfected to A549 cells. However, relative binding to PG9 and PG16 is barely detectable above background, suggesting that the C1D7 envelope sequence is poorly folded and is probably not present as an intact trimer on the cell surface.

[0128] Overall, the C4D7 envelope variant has the most optimal antibody binding profile, with higher gp120 expression than its full-length counterpart C4, and with greater than 15-fold increased PG9 and PG16 binding compared to C1 and C1D7 (FIG. 5).

Example 3: Stability of vectors encoding HIV envelope sequences

[0129] Previous work in our laboratories (unpublished) indicated that adenovirus 26 (Ad26) vectors encoding the mos2Env antigen sequence showed relatively high VP/IU ratios (indicating lower quality of adenovirus product batches) and moreover that such vectors displayed stability issues. Accordingly, it was important to test the stability of the synthetic HIV envelope proteins constructs of the invention in an adenovirus background.

[0130] Recombinant Ad26 (rAd26) vectors encoding HIV antigen sequences of the invention C4, C4D7, and sC4 as described above in Example 1 were generated in PER.C6 cells (referred to as "rAd26.C4", "rAd26.C4D7", and "rAd26.sC4", respectively). Vector clones (plaques) were picked and scaled-up for the generation of research batches. A maximum of 5 viral clones (plaques) were scaled-up to T25 format and serially passaged for 10 passages in T25 format (passages 1-3 being the transfection and plaque purification steps, followed by 10 passages in T25 format, resulting in a total of 13 passages). Genetic stability was assessed at viral passage number (vpn) 3, 5, 10 and 13 by an E1 transgene cassette PCR assay, followed by sequencing at vpn 13. The results are shown in FIG. 6.

[0131] The rAd26 vectors encoding full length C4 (rAd26.C4) showed poor growth characteristics, as determined by no full cytopathogenic effect (CPE) in 2-3 days; genetic instability, as determined by deletions of the E1 transgene cassette region; or a combination thereof (FIG. 6). Due to the poor growth characteristics and observed genetic instability, this vector encoding full length C4 was not pursued further.

[0132] In contrast, for the rAd26 vectors encoding C4D7 (rAd26.C4D7) and sC4 (rAd26.sC4), all propagated plaques remained genetically stable during the course of the experiment (FIG. 6). Thus, the novel sC4 and C4D7 constructs outperform the original mos2Env construct with respect to stability in an adenoviral vector background. The genetic stability testing up to vpn 13 represents propagation several passages beyond that used in the industrial scale preparation of the vectors.

Example 4: Expression and *in vivo* antigenicity of HIV envelope sequences in adenovirus vectors

[0133] Expression and antigenicity of rAd26.C4D7 and rAd26.sC4 were assessed separately or in combination with a recombinant Ad26 vector encoding mos1Env (SEQ ID NO: 5) (hereinafter "rAd26.mos1Env") in vector-transduced A549 cells (human cell line) *in vitro* (data not shown). Flow cytometry analysis demonstrated that all antigens were expressed in cell cultures transduced with either $2x10^4$ viral particles (vp) of the single envelope antigens as controls, or with $1x10^4$ vp of the 2 combined Env antigens by adenovirus transduction. All transductions additionally contained single doses ($1x10^4$ vp) of adenovirus vectors encoding mos1GagPol ("rAd26.mos1GagPol") and mos2GagPol ("rAd26.mos2GagPol") (Barouch et al, Nat Med 2010, 16:319-323), so that the assessed vector combinations exhibited the same relative ratios of the different adenoviral vectors as intended for pre-clinical and clinical use. Preferably, the vectors encoding synthetic HIV envelope proteins of the invention are combined with vectors encoding the mos1GagPol and the mos2GagPol antigens for clinical use.

[0134] The combination of rAd26.mos1Env and rAd26.C4D7 yielded a maximal coverage of the assessed epitopes as determined by monoclonal antibody binding. Particularly, the exposure of the PG16 epitope, which was contributed by transformation with Ad26.C4D7 is promising for vaccine use since PG16 represents a broadly neutralizing monoclonal antibody recognizing the V1/V2 loop region of HIV-1 Env (Walker et al, Science. 2009). Hence, the synthetic HIV envelope protein of the invention derived from the C4 sequence increases the breadth of the immune response against the HIV envelope protein compared to the immune response generated by mos1Env only. Vaccine-induced antibody responses directed towards the envelope protein region have been shown to correlate with protection from HIV-1 infection in the RV144 study (Haynes et al, N Engl J Med. 2012), and thus the synthetic HIV envelope protein of the invention is a promising candidate to include in HIV vaccine regimens.

Example 5: Immunogenicity of vectors encoding synthetic HIV envelope proteins

[0135] The synthetic HIV envelope protein sequences of the invention in an Ad26 vector background were tested in rabbits to determine if these constructs were an immunogenic alternative to the rAd26.mos2Env construct.

[0136] The immunogenicity of adenovirus vector encoding mos1Env (rAd26.mos1Env; SEQ ID NO: 5) was tested alone, and in combination with adenovirus vectors encoding synthetic HIV envelope proteins of the invention (rAd26.C4D7 and rAd26.sC4; comprising SEQ ID NO: 8, in particular SEQ ID NOs: 18 and 19, respectively). In all cases, adenovirus 26 vectors encoding mos1GagPol and mos2GagPol antigens (rAd26.mos1GagPol [SEQ ID NO: 28] and rAd26.mos2GagPol [SEQ ID NO: 29], respectively) were also administered. More specifically, the immunogenicity of rAd26.mos1Env alone (trivalent vaccine: rAd26.mos1GagPol, rAd26.mos2GagPol and rAd26.mos1Env) was compared to the immunogenicity of rAd26.mos1Env in combination with one of rAd26.C4D7 or rAd26.sC4 (tetravalent vaccine: administration of either rAd26.mos1GagPol, rAd26.mos2GagPol, rAd26.mos1Env and rAd26.C4D7; or administration of rAd26.mos1GagPol, rAd26.mos2GagPol, rAd26.mos1Env and rAd26.sC4). This comparison of the trivalent vaccine, which lacks any vectors encoding the synthetic HIV envelope proteins of the invention, with the tetravalent vaccine, which contains vectors encoding the synthetic HIV envelope proteins of the invention, allows for a determination of whether the HIV envelope proteins of the invention contribute to the breadth of protection.

[0137] Administration was done in vaccine regimens, wherein these Ad26 vectors were administered at weeks 0 and 6 as a double prime, and a clade C gp140 protein (a trivalent Env gp140 protein having SEQ ID NO: 7 without the signal peptide sequence of residues 1-29, see also WO 2010/042942) at weeks 12 and 18 as a double boost (see e.g. Barouch et al, 2015, Science 349: 320-324). Table 1 describes the vaccine regimens used for the current study. rAd26.Empty refers to a control vector lacking any gene encoding a sequence for an HIV antigenic protein. Each group contained six rabbits.

Table 1: Vaccine regimens tested in immunogenicity study in rabbits

Group	First and second Immunizations			Third and fourth immunizations			N=
	adeno vectors	Dose (vp)	Total dose (vp)	protein boost	Dose (ug)	Adjuvant	
1	rAd26.MoslEnv	$2.5x10^{10}$	$5x10^{10}$	GP140 (clade C)	10	AdjuPhos 250µg	6
	rAd26.Mcs1GagPol	$1.25x10^{10}$					
	rAd26.Mos2Gagpol	$1.25x10^{10}$					
2	rAd26.Mos1Env	$1.25x10^{10}$	$5x10^{10}$	GP140 (clade C)	10	AdjuPhos 250µg	6
	rAd26.C4D7	$1.25x10^{10}$					
	rAd26.Mos1GagPol	$1.25x10^{10}$					
	rAd26.Mos2Gagpol	$1.25x10^{10}$					
	rAd26.MoslEnv	$1.25x10^{10}$					

Group	First and second Immunizations			Third and fourth immunizations			N=
	adeno vectors	Dose (vp)	Total dose (vp)	protein boost	Dose (ug)	Adjuvant	
3	rAd26.sC4	$1.25x10^{10}$	$5x10^{10}$	GP140 (clade C)	10	AdjuPhos 250µg	6
	rAd26.Mos1GagPol	$1.25x10^{10}$					
	rAd26.Mos2Gagpol	$1.25x10^{10}$					
control	rAd26.Empty	$5x10^{10}$	$5x10^{10}$	NA	0	AdjuPhos 250µg	6

[0138] The comparison of the trivalent Ad26 vaccine (lacking the novel Env antigens of the invention) with the tetravalent Ad26 vaccine (which comprises the novel sC4 or C4D7 Env antigens) allows for testing if the novel antigens contribute to breadth of protection. An established TZM-bl cell-based neutralization assay [Montefiori DC. Methods Mol Biol 2009,485:395-405; Sarzotti-Kelsoe M et al., J Immunol Methods 2014,409:131-146] was used to measure neutralizing activity of the vaccine candidates.

[0139] Results are shown in Fig 7, and were statistically analyzed by using the trivalent vaccine (group 1 in Table 1) as control group and comparing to each of the novel quadrivalent vaccines (groups 2 and 3 in Table 1).

[0140] Overall, the novel C4-derived (i.e. encoding Env proteins comprising SEQ ID NO: 8, being an alternative for mos2Env) adeno constructs were immunogenic after two homologous intramuscular immunizations in rabbits.

[0141] Neutralization capacity of rabbit immune sera against Tier 1B pseudoviruses was absent (data not shown), which is not unexpected as it was known that such viruses are more difficult to neutralize.

[0142] Pseudovirus neutralization capacity of rabbit immune sera against a clade B Tier 1A virus was unaffected by the addition of new components (data not shown). This demonstrates that the novel antigen did not negatively interfere with immunogenicity of the existing clade B antigen present in the vaccine (although the new components were directed to clade C, such undesirable interference could not be excluded a priori before it had been tested).

[0143] Pseudovirus neutralization capacity of rabbit immune sera against a clade C Tier 1A virus was significantly enhanced in the quadrivalent novel C4D7 containing adeno (quadrivalent, group 2), compared to trivalent (having only mos1Env) immunization alone (group 1) (Fig 7 panel B). In addition, pseudovirus neutralization capacity of rabbit immune sera against a clade C Tier 1A virus at week 8 was significantly enhanced in the quadrivalent novel sC4 containing adeno (quadrivalent, group 3), compared to trivalent (having only

mos1Env) immunization alone (group 1) (Fig 7 panel B).

[0144] In conclusion, the C4D7 and sC4 constructs encoded in Ad26 were immunogenic and addition thereof expanded the binding- and neutralization capacity of a vaccine that has mos1Env (mainly clade B) as sole Ad26-encoded Env component, towards clade C strains (Fig 7B).

Example 6: Immunogenicity of vaccine regimens including vectors encoding synthetic HIV envelope proteins of the invention

[0145] One further rabbit study assessed the tetravalent vector combination Ad26.Mos4.HIV (consisting of four adenoviral vectors: Ad26.Mos1GagPol [encoding SEQ ID NO: 28], Ad26.Mos2GagPol [encoding SEQ ID NO: 29], Ad26.Mos1Env [encoding SEQ ID NO: 5] and Ad26.Mos2SEnv [the name "C4D7" as used above is also referred to as "Mos2S"; this vector encodes the novel SEQ ID NO: 18 according to the invention], in a 1:1:1:1 mixture at a total dose of $5x10^9$ vp,) applied intramuscularly as double prime immunizations in weeks 0 and 6, in combination with recombinant HIV-1 Env protein boosts using Clade C gp140 [having the sequence of amino acid residues 30-708 of SEQ ID NO: 7], Mosaic gp140 [having the sequence of amino acid residues 30-724 of SEQ ID NO: 36], or a combination of Clade C gp140 and Mosaic gp140, in weeks 13 and 19. These protein boosts were applied intramuscularly at a total dose of 10 or 50 micrograms of protein combined with 250 mcg aluminum phosphate adjuvant formulated on the day of immunization.

[0146] Results indicate that all tested regimens were immunogenic in all animals, inducing high antibody titers and moderate neutralization activity against Tier 1 Env pseudotyped viruses. If Mosaic gp140 was used as vaccine antigen, either alone or in combination with Clade C gp140, Mosaic gp140-specific ELISA titers and Clade B pseudovirus recognition were significantly increased at week 15 in comparison to the reference group boosted with Clade C gp140 alone. The overall effect size of the improvement was moderate, and bigger for the group boosted with the bivalent Clade C gp140 - Mosaic gp140 combination compared to Mosaic gp140 alone. At week 21 of the study, these differences were lost and immune responses measured for the cohorts receiving bivalent Clade C gp140 - Mosaic gp140 boosts or monovalent Clade C gp140 boosts were statistically indistinguishable.

[0147] The bivalent protein regimen showed comparable induction of Clade C ELISA titers and pseudovirus recognition as the Clade C gp140 alone boosted regimen, indicating that the inclusion of the clade B-related immunogen Mosaic gp140 had no negative effect on clade C antigen coverage, whilst significantly enhancing clade B coverage at week 15 of the study.

[0148] The data confirm that the Ad26.Mos2SEnv vector encoding a synthetic Env antigen according to the invention can successfully be used in vaccine regimens.

REFERENCES

[0149]

1. 1. Barouch et al, Nat Med 2010, 16: 319-323
2. 2. WO 2010/059732
3. 3. Schiernle et al., PNAS 94: 8640-8645, 1997
4. 4. Abrahamyan et al., J Virol 79: 106-115, 2005
5. 5. US20120076812
6. 6. Barouch et al., Cell 155:1-9, 2013
7. 7. Havenga, et al., 2006, J Gen Virol 87: 2135-43;
8. 8. WO 03/104467
9. 9. WO 2004/001032
10. 10. WO 2007/104792
11. 11. Abbink et al., (2007) Virol 81(9): 4654-63
12. 12. US Patent No. 7,270,811
13. 13. Vogels et al., (2003) J Virol 77(15): 8263-71
14. 14. WO 00/70071
15. 15. WO2012/082918
16. 16. Walker LM, Phogat SK, Chan-Hui PY, Wagner D, Phung P, Goss JL, et al. Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. Science 2009,326:285-289.
17. 17. Haynes BF, Gilbert PB, McElrath MJ, Zolla-Pazner S, Tomaras GD, Alam SM, et al. Immune-correlates analysis of an HIV-1 vaccine efficacy trial. N Engl J Med 2012,366:1275-1286.
18. 18. Barouch et al. (2015) Science 349: 320-324
19. 19. Montefiori DC. Measuring HIV neutralization in a luciferase reporter gene assay. Methods Mol Biol 2009,485:395-405.
20. 20. Sarzotti-Kelsoe M, Bailer RT, Turk E, Lin CL, Bilska M, Greene KM, et al. Optimization and validation of the TZM-bl assay for standardized assessments of neutralizing antibodies against HIV-1. J Immunol Methods 2014,409:131-146.
21. 21. Edwards et al., J. Virology, 2002, 76:2683-2691.

SEQUENCE LISTING

[0150]

<110> Janssen Vaccines & Prevention B.V.

<120> Human Immunodeficiency Virus Antigens, Vectors, Compositions, and Methods of Use Thereof

<130> 0265 EP P01 PRI

<160> 37

<170> PatentIn version 3.5

<210> 1
<211> 500
<212> PRT
<213> Artificial Sequence

<220>
<223> mos1Gag mosaic antigen sequence

<400> 1

```
Met Gly Ala Arg Ala Ser Val Leu Ser Gly Gly Glu Leu Asp Arg Trp
1               5                  10                  15

Glu Lys Ile Arg Leu Arg Pro Gly Gly Lys Lys Lys Tyr Arg Leu Lys
            20                  25                  30

His Ile Val Trp Ala Ser Arg Glu Leu Glu Arg Phe Ala Val Asn Pro
        35                  40                  45

Gly Leu Leu Glu Thr Ser Glu Gly Cys Arg Gln Ile Leu Gly Gln Leu
    50                  55                  60

Gln Pro Ser Leu Gln Thr Gly Ser Glu Glu Leu Arg Ser Leu Tyr Asn
65                  70                  75                  80

Thr Val Ala Thr Leu Tyr Cys Val His Gln Arg Ile Glu Ile Lys Asp
                85                  90                  95

Thr Lys Glu Ala Leu Glu Lys Ile Glu Glu Glu Gln Asn Lys Ser Lys
            100                 105                 110

Lys Lys Ala Gln Gln Ala Ala Ala Asp Thr Gly Asn Ser Ser Gln Val
        115                 120                 125

Ser Gln Asn Tyr Pro Ile Val Gln Asn Ile Gln Gly Gln Met Val His
    130                 135                 140

Gln Ala Ile Ser Pro Arg Thr Leu Asn Ala Trp Val Lys Val Val Glu
145                 150                 155                 160

Glu Lys Ala Phe Ser Pro Glu Val Ile Pro Met Phe Ser Ala Leu Ser
                165                 170                 175

Glu Gly Ala Thr Pro Gln Asp Leu Asn Thr Met Leu Asn Thr Val Gly
            180                 185                 190

Gly His Gln Ala Ala Met Gln Met Leu Lys Glu Thr Ile Asn Glu Glu
        195                 200                 205
```

```
Ala Ala Glu Trp Asp Arg Val His Pro Val His Ala Gly Pro Ile Ala
    210                 215                 220

Pro Gly Gln Met Arg Glu Pro Arg Gly Ser Asp Ile Ala Gly Thr Thr
225                 230                 235                 240

Ser Thr Leu Gln Glu Gln Ile Gly Trp Met Thr Asn Asn Pro Pro Ile
                245                 250                 255

Pro Val Gly Glu Ile Tyr Lys Arg Trp Ile Ile Leu Gly Leu Asn Lys
            260                 265                 270

Ile Val Arg Met Tyr Ser Pro Val Ser Ile Leu Asp Ile Arg Gln Gly
        275                 280                 285

Pro Lys Glu Pro Phe Arg Asp Tyr Val Asp Arg Phe Tyr Lys Thr Leu
    290                 295                 300

Arg Ala Glu Gln Ala Ser Gln Asp Val Lys Asn Trp Met Thr Glu Thr
305                 310                 315                 320

Leu Leu Val Gln Asn Ala Asn Pro Asp Cys Lys Thr Ile Leu Lys Ala
                325                 330                 335

Leu Gly Pro Ala Ala Thr Leu Glu Glu Met Met Thr Ala Cys Gln Gly
            340                 345                 350

Val Gly Gly Pro Gly His Lys Ala Arg Val Leu Ala Glu Ala Met Ser
        355                 360                 365

Gln Val Thr Asn Ser Ala Thr Ile Met Met Gln Arg Gly Asn Phe Arg
    370                 375                 380

Asn Gln Arg Lys Thr Val Lys Cys Phe Asn Cys Gly Lys Glu Gly His
385                 390                 395                 400

Ile Ala Lys Asn Cys Arg Ala Pro Arg Lys Lys Gly Cys Trp Lys Cys
                405                 410                 415

Gly Lys Glu Gly His Gln Met Lys Asp Cys Thr Glu Arg Gln Ala Asn
            420                 425                 430

Phe Leu Gly Lys Ile Trp Pro Ser Asn Lys Gly Arg Pro Gly Asn Phe
        435                 440                 445

Leu Gln Asn Arg Pro Glu Pro Thr Ala Pro Pro Glu Glu Ser Phe Arg
    450                 455                 460

Phe Gly Glu Glu Thr Thr Thr Pro Ser Gln Lys Gln Glu Pro Ile Asp
465                 470                 475                 480

Lys Glu Met Tyr Pro Leu Ala Ser Leu Lys Ser Leu Phe Gly Asn Asp
                485                 490                 495

Pro Ser Ser Gln
            500
```

<210> 2

<211> 491

<212> PRT

<213> Artificial Sequence

<220>

<223> mos2Gag mosaic antigen sequence

<400> 2

```
Met Gly Ala Arg Ala Ser Ile Leu Arg Gly Gly Lys Leu Asp Lys Trp
1               5                  10                  15

Glu Lys Ile Arg Leu Arg Pro Gly Gly Lys Lys His Tyr Met Leu Lys
            20                  25                  30

His Leu Val Trp Ala Ser Arg Glu Leu Glu Arg Phe Ala Leu Asn Pro
        35                  40                  45

Gly Leu Leu Glu Thr Ser Glu Gly Cys Lys Gln Ile Ile Lys Gln Leu
    50                  55                  60

Gln Pro Ala Leu Gln Thr Gly Thr Glu Glu Leu Arg Ser Leu Phe Asn
65                  70                  75                  80

Thr Val Ala Thr Leu Tyr Cys Val His Ala Glu Ile Glu Val Arg Asp
                85                  90                  95

Thr Lys Glu Ala Leu Asp Lys Ile Glu Glu Glu Gln Asn Lys Ser Gln
            100                 105                 110

Gln Lys Thr Gln Gln Ala Lys Glu Ala Asp Gly Lys Val Ser Gln Asn
        115                 120                 125

Tyr Pro Ile Val Gln Asn Leu Gln Gly Gln Met Val His Gln Pro Ile
    130                 135                 140

Ser Pro Arg Thr Leu Asn Ala Trp Val Lys Val Ile Glu Glu Lys Ala
145                 150                 155                 160

Phe Ser Pro Glu Val Ile Pro Met Phe Thr Ala Leu Ser Glu Gly Ala
                165                 170                 175

Thr Pro Gln Asp Leu Asn Thr Met Leu Asn Thr Val Gly Gly His Gln
            180                 185                 190

Ala Ala Met Gln Met Leu Lys Asp Thr Ile Asn Glu Glu Ala Ala Glu
        195                 200                 205

Trp Asp Arg Leu His Pro Val His Ala Gly Pro Val Ala Pro Gly Gln
    210                 215                 220

Met Arg Glu Pro Arg Gly Ser Asp Ile Ala Gly Thr Thr Ser Asn Leu
225                 230                 235                 240

Gln Glu Gln Ile Ala Trp Met Thr Ser Asn Pro Pro Ile Pro Val Gly
```

```
                245                 250                 255

Asp Ile Tyr Lys Arg Trp Ile Ile Leu Gly Leu Asn Lys Ile Val Arg
            260                 265                 270

Met Tyr Ser Pro Thr Ser Ile Leu Asp Ile Lys Gln Gly Pro Lys Glu
        275                 280                 285

Pro Phe Arg Asp Tyr Val Asp Arg Phe Phe Lys Thr Leu Arg Ala Glu
    290                 295                 300

Gln Ala Thr Gln Asp Val Lys Asn Trp Met Thr Asp Thr Leu Leu Val
305                 310                 315                 320

Gln Asn Ala Asn Pro Asp Cys Lys Thr Ile Leu Arg Ala Leu Gly Pro
                325                 330                 335

Gly Ala Thr Leu Glu Glu Met Met Thr Ala Cys Gln Gly Val Gly Gly
            340                 345                 350

Pro Ser His Lys Ala Arg Val Leu Ala Glu Ala Met Ser Gln Thr Asn
        355                 360                 365

Ser Thr Ile Leu Met Gln Arg Ser Asn Phe Lys Gly Ser Lys Arg Ile
    370                 375                 380

Val Lys Cys Phe Asn Cys Gly Lys Glu Gly His Ile Ala Arg Asn Cys
385                 390                 395                 400

Arg Ala Pro Arg Lys Lys Gly Cys Trp Lys Cys Gly Lys Glu Gly His
                405                 410                 415

Gln Met Lys Asp Cys Thr Glu Arg Gln Ala Asn Phe Leu Gly Lys Ile
            420                 425                 430

Trp Pro Ser His Lys Gly Arg Pro Gly Asn Phe Leu Gln Ser Arg Pro
        435                 440                 445

Glu Pro Thr Ala Pro Pro Ala Glu Ser Phe Arg Phe Glu Glu Thr Thr
    450                 455                 460

Pro Ala Pro Lys Gln Glu Pro Lys Asp Arg Glu Pro Leu Thr Ser Leu
465                 470                 475                 480

Arg Ser Leu Phe Gly Ser Asp Pro Leu Ser Gln
                485                 490
```

<210> 3

<211> 850

<212> PRT

<213> Artificial Sequence

<220>

<223> mos1Pol mosaic antigen sequence

<400> 3

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

Gly Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu Thr Glu Glu Lys
            20                  25                  30

Ile Lys Ala Leu Thr Ala Ile Cys Glu Glu Met Glu Lys Glu Gly Lys
        35                  40                  45

Ile Thr Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe Ala
    50                  55                  60

Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg
65                  70                  75                  80

Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile
                85                  90                  95

Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu Ala
            100                 105                 110

Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Gly Phe Arg Lys
        115                 120                 125

Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly Ile
    130                 135                 140

Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala
145                 150                 155                 160

Ile Phe Gln Cys Ser Met Thr Arg Ile Leu Glu Pro Phe Arg Ala Lys
                165                 170                 175

Asn Pro Glu Ile Val Ile Tyr Gln Tyr Met Ala Ala Leu Tyr Val Gly
            180                 185                 190

Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys Ile Glu Glu Leu Arg
        195                 200                 205

Glu His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His Gln
    210                 215                 220

Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys
225                 230                 235                 240

Trp Thr Val Gln Pro Ile Gln Leu Pro Glu Lys Asp Ser Trp Thr Val
                245                 250                 255

Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile
            260                 265                 270

Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys Leu Leu Arg Gly Ala
        275                 280                 285

Lys Ala Leu Thr Asp Ile Val Pro Leu Thr Glu Glu Ala Glu Leu Glu
    290                 295                 300
```

```
Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr
305                 310                 315                 320

Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly His
                325                 330                 335

Asp Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys
            340                 345                 350

Thr Gly Lys Tyr Ala Lys Met Arg Thr Ala His Thr Asn Asp Val Lys
        355                 360                 365

Gln Leu Thr Glu Ala Val Gln Lys Ile Ala Met Glu Ser Ile Val Ile
    370                 375                 380

Trp Gly Lys Thr Pro Lys Phe Arg Leu Pro Ile Gln Lys Glu Thr Trp
385                 390                 395                 400

Glu Thr Trp Trp Thr Asp Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp
                405                 410                 415

Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu
            420                 425                 430

Lys Asp Pro Ile Ala Gly Val Glu Thr Phe Tyr Val Ala Gly Ala Ala
        435                 440                 445

Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asp Arg Gly
    450                 455                 460

Arg Gln Lys Ile Val Ser Leu Thr Glu Thr Thr Asn Gln Lys Thr Ala
465                 470                 475                 480

Leu Gln Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly Ser Glu Val Asn
                485                 490                 495

Ile Val Thr Ala Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro
            500                 505                 510

Asp Lys Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Gln Leu Ile
        515                 520                 525

Lys Lys Glu Arg Val Tyr Leu Ser Trp Val Pro Ala His Lys Gly Ile
    530                 535                 540

Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ser Gly Ile Arg Lys
545                 550                 555                 560

Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Glu Glu His Glu Lys
                565                 570                 575

Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro
            580                 585                 590

Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Gln Cys Gln Leu Lys
        595                 600                 605
```

```
Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln
    610                 615                 620

Leu Ala Cys Thr His Leu Glu Gly Lys Ile Ile Leu Val Ala Val His
625                 630                 635                 640

Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly
                645                 650                 655

Gln Glu Thr Ala Tyr Phe Ile Leu Lys Leu Ala Gly Arg Trp Pro Val
            660                 665                 670

Lys Val Ile His Thr Ala Asn Gly Ser Asn Phe Thr Ser Ala Ala Val
        675                 680                 685

Lys Ala Ala Cys Trp Trp Ala Gly Ile Gln Gln Glu Phe Gly Ile Pro
    690                 695                 700

Tyr Asn Pro Gln Ser Gln Gly Val Val Ala Ser Met Asn Lys Glu Leu
705                 710                 715                 720

Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr
                725                 730                 735

Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly
            740                 745                 750

Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Ile Asp Ile Ile Ala Thr
        755                 760                 765

Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Ile Lys Ile Gln Asn
    770                 775                 780

Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asp Pro Ile Trp Lys Gly Pro
785                 790                 795                 800

Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn
                805                 810                 815

Ser Asp Ile Lys Val Val Pro Arg Arg Lys Val Lys Ile Ile Lys Asp
            820                 825                 830

Tyr Gly Lys Gln Met Ala Gly Ala Asp Cys Val Ala Gly Arg Gln Asp

        835                 840                 845

Glu Asp
    850
```

<210> 4

<211> 850

<212> PRT

<213> Artificial Sequence

<220>

<223> mos2Pol mosaic antigen sequence

<400> 4

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

Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys
            20                  25                  30

Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys
        35                  40                  45

Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe Ala
    50                  55                  60

Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg
65                  70                  75                  80

Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile
                85                  90                  95

Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu Ala
            100                 105                 110

Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg Lys
        115                 120                 125

Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile
    130                 135                 140

Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala
145                 150                 155                 160

Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln
                165                 170                 175

Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Ala Ala Leu Tyr Val Gly
            180                 185                 190

Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg
        195                 200                 205

Gln His Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His Gln
    210                 215                 220

Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys
225                 230                 235                 240

Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val
                245                 250                 255

Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile
            260                 265                 270

Tyr Ala Gly Ile Lys Val Lys Gln Leu Cys Lys Leu Leu Arg Gly Thr
```

```
Tyr Ala Gly Ile Lys Val Lys Gln Leu Cys Lys Leu Leu Arg Gly Thr
        275                 280                 285

Lys Ala Leu Thr Glu Val Val Pro Leu Thr Glu Glu Ala Glu Leu Glu
    290                 295                 300

Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr
305                 310                 315                 320

Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln
                325                 330                 335

Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys
            340                 345                 350

Thr Gly Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp Val Lys
        355                 360                 365

Gln Leu Thr Glu Ala Val Gln Lys Ile Ala Thr Glu Ser Ile Val Ile
    370                 375                 380

Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp
385                 390                 395                 400

Glu Ala Trp Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp
                405                 410                 415

Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu
            420                 425                 430

Lys Glu Pro Ile Val Gly Ala Glu Thr Phe Tyr Val Ala Gly Ala Ala

        435                 440                 445

Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asp Arg Gly
    450                 455                 460

Arg Gln Lys Val Val Ser Leu Thr Asp Thr Thr Asn Gln Lys Thr Ala
465                 470                 475                 480

Leu Gln Ala Ile His Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn
                485                 490                 495

Ile Val Thr Ala Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro
            500                 505                 510

Asp Lys Ser Glu Ser Glu Leu Val Ser Gln Ile Ile Glu Gln Leu Ile
        515                 520                 525

Lys Lys Glu Lys Val Tyr Leu Ala Trp Val Pro Ala His Lys Gly Ile
    530                 535                 540

Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Arg Gly Ile Arg Lys
545                 550                 555                 560

Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Glu Glu His Glu Lys
                565                 570                 575
```

```
Tyr His Ser Asn Trp Arg Ala Met Ala Ser Glu Phe Asn Leu Pro Pro
            580                 585                 590

Ile Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys
        595                 600                 605

Gly Glu Ala Ile His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln
    610                 615                 620

Leu Ala Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val His
625                 630                 635                 640

Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly
                645                 650                 655

Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val
            660                 665                 670

Lys Thr Ile His Thr Ala Asn Gly Ser Asn Phe Thr Ser Ala Thr Val
        675                 680                 685

Lys Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro
    690                 695                 700

Tyr Asn Pro Gln Ser Gln Gly Val Val Ala Ser Ile Asn Lys Glu Leu
705                 710                 715                 720

Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr
                725                 730                 735

Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly
            740                 745                 750

Ile Gly Glu Tyr Ser Ala Gly Glu Arg Ile Val Asp Ile Ile Ala Ser
        755                 760                 765

Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln Asn
    770                 775                 780

Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asp Pro Leu Trp Lys Gly Pro
785                 790                 795                 800

Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn
                805                 810                 815

Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp
            820                 825                 830

Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp
        835                 840                 845

Glu Asp
    850
```

<210> 5

<211> 685

<212> PRT

<213> Artificial Sequence

<220>

<223> mos1Env mosaic antigen sequence

<400> 5

```
Met Arg Val Thr Gly Ile Arg Lys Asn Tyr Gln His Leu Trp Arg Trp
1               5                  10                  15

Gly Thr Met Leu Leu Gly Ile Leu Met Ile Cys Ser Ala Ala Gly Lys
            20                  25                  30

Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala Thr

        35                  40                  45

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

His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
65                  70                  75                  80

Gln Glu Val Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp Lys
                85                  90                  95

Asn Asn Met Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp
            100                 105                 110

Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
        115                 120                 125

Asn Cys Thr Asp Asp Val Arg Asn Val Thr Asn Asn Ala Thr Asn Thr
    130                 135                 140

Asn Ser Ser Trp Gly Glu Pro Met Glu Lys Gly Glu Ile Lys Asn Cys
145                 150                 155                 160

Ser Phe Asn Ile Thr Thr Ser Ile Arg Asn Lys Val Gln Lys Gln Tyr
                165                 170                 175

Ala Leu Phe Tyr Lys Leu Asp Val Val Pro Ile Asp Asn Asp Ser Asn
            180                 185                 190

Asn Thr Asn Tyr Arg Leu Ile Ser Cys Asn Thr Ser Val Ile Thr Gln
        195                 200                 205

Ala Cys Pro Lys Val Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala
    210                 215                 220

Pro Ala Gly Phe Ala Ile Leu Lys Cys Asn Asp Lys Lys Phe Asn Gly
225                 230                 235                 240
```

```
Thr Gly Pro Cys Thr Asn Val Ser Thr Val Gln Cys Thr His Gly Ile
                245                 250                 255

Arg Pro Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu
            260                 265                 270

Glu Glu Val Val Ile Arg Ser Glu Asn Phe Thr Asn Asn Ala Lys Thr
        275                 280                 285

Ile Met Val Gln Leu Asn Val Ser Val Glu Ile Asn Cys Thr Arg Pro
    290                 295                 300

Asn Asn Asn Thr Arg Lys Ser Ile His Ile Gly Pro Gly Arg Ala Phe
305                 310                 315                 320

Tyr Thr Ala Gly Asp Ile Ile Gly Asp Ile Arg Gln Ala His Cys Asn
                325                 330                 335

Ile Ser Arg Ala Asn Trp Asn Asn Thr Leu Arg Gln Ile Val Glu Lys
            340                 345                 350

Leu Gly Lys Gln Phe Gly Asn Asn Lys Thr Ile Val Phe Asn His Ser
        355                 360                 365

Ser Gly Gly Asp Pro Glu Ile Val Met His Ser Phe Asn Cys Gly Gly
    370                 375                 380

Glu Phe Phe Tyr Cys Asn Ser Thr Lys Leu Phe Asn Ser Thr Trp Thr
385                 390                 395                 400

Trp Asn Asn Ser Thr Trp Asn Asn Thr Lys Arg Ser Asn Asp Thr Glu
                405                 410                 415

Glu His Ile Thr Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Trp
            420                 425                 430

Gln Glu Val Gly Lys Ala Met Tyr Ala Pro Pro Ile Arg Gly Gln Ile
        435                 440                 445

Arg Cys Ser Ser Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly
    450                 455                 460

Asn Asp Thr Ser Gly Thr Glu Ile Phe Arg Pro Gly Gly Gly Asp Met
465                 470                 475                 480

Arg Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Lys Ile
                485                 490                 495

Glu Pro Leu Gly Val Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln
            500                 505                 510

Ser Glu Lys Ser Ala Val Gly Ile Gly Ala Val Phe Leu Gly Phe Leu
        515                 520                 525

Gly Ala Ala Gly Ser Thr Met Gly Ala Ala Ser Met Thr Leu Thr Val
    530                 535                 540
```

```
Gln Ala Arg Leu Leu Leu Ser Gly Ile Val Gln Gln Gln Asn Asn Leu
545                 550                 555                 560

Leu Arg Ala Ile Glu Ala Gln Gln His Leu Leu Gln Leu Thr Val Trp
                565                 570                 575

Gly Ile Lys Gln Leu Gln Ala Arg Val Leu Ala Val Glu Arg Tyr Leu
            580                 585                 590

Lys Asp Gln Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile
        595                 600                 605

Cys Thr Thr Thr Val Pro Trp Asn Ala Ser Trp Ser Asn Lys Ser Leu
    610                 615                 620

Asp Lys Ile Trp Asn Asn Met Thr Trp Met Glu Trp Glu Arg Glu Ile
625                 630                 635                 640

Asn Asn Tyr Thr Ser Leu Ile Tyr Thr Leu Ile Glu Glu Ser Gln Asn
                645                 650                 655

Gln Gln Glu Lys Asn Glu Gln Glu Leu Leu Glu Leu Asp Lys Trp Ala
            660                 665                 670

Ser Leu Trp Asn Trp Phe Asp Ile Ser Asn Trp Leu Trp
        675                 680                 685
```

<210> 6

<211> 684

<212> PRT

<213> Artificial Sequence

<220>

<223> mos2Env mosaic antigen sequence

<400> 6

```
Met Arg Val Arg Gly Ile Gln Arg Asn Trp Pro Gln Trp Trp Ile Trp
1               5                   10                  15

Gly Ile Leu Gly Phe Trp Met Ile Ile Ile Cys Arg Val Met Gly Asn
            20                  25                  30

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

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

His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
65                  70                  75                  80

Gln Glu Met Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp Lys
                85                  90                  95

Asn Asp Met Val Asp Gln Met His Glu Asp Ile Ile Arg Leu Trp Asp
```

```
            100                 105                 110

Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
        115                 120                 125

Glu Cys Arg Asn Val Arg Asn Val Ser Ser Asn Gly Thr Tyr Asn Ile
    130                 135                 140

Ile His Asn Glu Thr Tyr Lys Glu Met Lys Asn Cys Ser Phe Asn Ala
145                 150                 155                 160

Thr Thr Val Val Glu Asp Arg Lys Gln Lys Val His Ala Leu Phe Tyr
                165                 170                 175

Arg Leu Asp Ile Val Pro Leu Asp Glu Asn Asn Ser Ser Glu Lys Ser
            180                 185                 190

Ser Glu Asn Ser Ser Glu Tyr Tyr Arg Leu Ile Asn Cys Asn Thr Ser
        195                 200                 205

Ala Ile Thr Gln Ala Cys Pro Lys Val Ser Phe Asp Pro Ile Pro Ile
    210                 215                 220

His Tyr Cys Ala Pro Ala Gly Tyr Ala Ile Leu Lys Cys Asn Asn Lys
225                 230                 235                 240

Thr Phe Asn Gly Thr Gly Pro Cys Asn Asn Val Ser Thr Val Gln Cys
                245                 250                 255

Thr His Gly Ile Lys Pro Val Val Ser Thr Gln Leu Leu Leu Asn Gly
            260                 265                 270

Ser Leu Ala Glu Glu Glu Ile Ile Ile Arg Ser Glu Asn Leu Thr Asn
        275                 280                 285

Asn Ala Lys Thr Ile Ile Val His Leu Asn Glu Thr Val Asn Ile Thr
    290                 295                 300

Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser Ile Arg Ile Gly Pro
305                 310                 315                 320

Gly Gln Thr Phe Tyr Ala Thr Gly Asp Ile Ile Gly Asp Ile Arg Gln
                325                 330                 335

Ala His Cys Asn Leu Ser Arg Asp Gly Trp Asn Lys Thr Leu Gln Gly
            340                 345                 350

Val Lys Lys Lys Leu Ala Glu His Phe Pro Asn Lys Thr Ile Asn Phe
        355                 360                 365

Thr Ser Ser Ser Gly Gly Asp Leu Glu Ile Thr Thr His Ser Phe Asn
    370                 375                 380

Cys Arg Gly Glu Phe Phe Tyr Cys Asn Thr Ser Gly Leu Phe Asn Gly
385                 390                 395                 400

Thr Tyr Met Pro Asn Gly Thr Asn Ser Asn Ser Ser Ser Asn Ile Thr
                405                 410                 415
```

405 410 415

Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly
420 425 430

Arg Ala Met Tyr Ala Pro Pro Ile Ala Gly Asn Ile Thr Cys Arg Ser
435 440 445

Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly Ser Asn Asn Gly
450 455 460

Val Pro Asn Asp Thr Glu Thr Phe Arg Pro Gly Gly Gly Asp Met Arg
465 470 475 480

Asn Asn Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Glu Val Lys
485 490 495

Pro Leu Gly Val Ala Pro Thr Glu Ala Lys Arg Arg Val Val Glu Ser
500 505 510

Glu Lys Ser Ala Val Gly Ile Gly Ala Val Phe Leu Gly Ile Leu Gly
515 520 525

Ala Ala Gly Ser Thr Met Gly Ala Ala Ser Ile Thr Leu Thr Val Gln
530 535 540

Ala Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Ser Asn Leu Leu
545 550 555 560

Arg Ala Ile Glu Ala Gln Gln His Met Leu Gln Leu Thr Val Trp Gly
565 570 575

Ile Lys Gln Leu Gln Thr Arg Val Leu Ala Ile Glu Arg Tyr Leu Gln
580 585 590

Asp Gln Gln Leu Leu Gly Leu Trp Gly Cys Ser Gly Lys Leu Ile Cys
595 600 605

Thr Thr Ala Val Pro Trp Asn Thr Ser Trp Ser Asn Lys Ser Gln Thr
610 615 620

Asp Ile Trp Asp Asn Met Thr Trp Met Gln Trp Asp Lys Glu Ile Gly
625 630 635 640

Asn Tyr Thr Gly Glu Ile Tyr Arg Leu Leu Glu Glu Ser Gln Asn Gln
645 650 655

Gln Glu Lys Asn Glu Lys Asp Leu Leu Ala Leu Asp Ser Trp Lys Asn
660 665 670

Leu Trp Asn Trp Phe Asp Ile Thr Asn Trp Leu Trp
675 680

<210> 7

<211> 708

<212> PRT

<213> Artificial Sequence

<220>

<223> stabilized clade C gp140 trimer: C97ZA012-gp140-foldon with cleavage mutations

<400> 7

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

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

Leu Trp Val Gly Asn Met Trp Val Thr Val Tyr Tyr Gly Val Pro Val
        35                  40                  45

Trp Thr Asp Ala Lys Thr Thr Leu Phe Cys Ala Ser Asp Thr Lys Ala
    50                  55                  60

Tyr Asp Arg Glu Val His Asn Val Trp Ala Thr His Ala Cys Val Pro
65                  70                  75                  80

Thr Asp Pro Asn Pro Gln Glu Ile Val Leu Glu Asn Val Thr Glu Asn
                85                  90                  95

Phe Asn Met Trp Lys Asn Asp Met Val Asp Gln Met His Glu Asp Ile
            100                 105                 110

Ile Ser Leu Trp Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro
        115                 120                 125

Leu Cys Val Thr Leu His Cys Thr Asn Ala Thr Phe Lys Asn Asn Val
    130                 135                 140

Thr Asn Asp Met Asn Lys Glu Ile Arg Asn Cys Ser Phe Asn Thr Thr
145                 150                 155                 160

Thr Glu Ile Arg Asp Lys Lys Gln Gln Gly Tyr Ala Leu Phe Tyr Arg
                165                 170                 175

Pro Asp Ile Val Leu Leu Lys Glu Asn Arg Asn Asn Ser Asn Asn Ser
            180                 185                 190

Glu Tyr Ile Leu Ile Asn Cys Asn Ala Ser Thr Ile Thr Gln Ala Cys
        195                 200                 205

Pro Lys Val Asn Phe Asp Pro Ile Pro Ile His Tyr Cys Ala Pro Ala
    210                 215                 220

Gly Tyr Ala Ile Leu Lys Cys Asn Asn Lys Thr Phe Ser Gly Lys Gly
225                 230                 235                 240

Pro Cys Asn Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Lys Pro
                245                 250                 255

Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Lys Glu
            260                 265                 270
```

```
Ile Ile Ile Arg Ser Glu Asn Leu Thr Asp Asn Val Lys Thr Ile Ile
        275                 280                 285

Val His Leu Asn Lys Ser Val Glu Ile Val Cys Thr Arg Pro Asn Asn
    290                 295                 300

Asn Thr Arg Lys Ser Met Arg Ile Gly Pro Gly Gln Thr Phe Tyr Ala
305                 310                 315                 320

Thr Gly Asp Ile Ile Gly Asp Ile Arg Gln Ala Tyr Cys Asn Ile Ser
                325                 330                 335

Gly Ser Lys Trp Asn Glu Thr Leu Lys Arg Val Lys Glu Lys Leu Gln

            340                 345                 350

Glu Asn Tyr Asn Asn Asn Lys Thr Ile Lys Phe Ala Pro Ser Ser Gly
        355                 360                 365

Gly Asp Leu Glu Ile Thr Thr His Ser Phe Asn Cys Arg Gly Glu Phe
    370                 375                 380

Phe Tyr Cys Asn Thr Thr Arg Leu Phe Asn Asn Asn Ala Thr Glu Asp
385                 390                 395                 400

Glu Thr Ile Thr Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Trp
                405                 410                 415

Gln Gly Val Gly Arg Ala Met Tyr Ala Pro Pro Ile Ala Gly Asn Ile
            420                 425                 430

Thr Cys Lys Ser Asn Ile Thr Gly Leu Leu Leu Val Arg Asp Gly Gly
        435                 440                 445

Glu Asp Asn Lys Thr Glu Glu Ile Phe Arg Pro Gly Gly Gly Asn Met
    450                 455                 460

Lys Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Ile Glu Leu
465                 470                 475                 480

Lys Pro Leu Gly Ile Ala Pro Thr Gly Ala Lys Glu Arg Val Val Glu
                485                 490                 495

Arg Glu Glu Arg Ala Val Gly Ile Gly Ala Val Phe Leu Gly Phe Leu
            500                 505                 510

Gly Ala Ala Gly Ser Thr Met Gly Ala Ala Ser Leu Thr Leu Thr Val
        515                 520                 525

Gln Ala Arg Gln Leu Leu Ser Ser Ile Val Gln Gln Gln Ser Asn Leu
    530                 535                 540

Leu Arg Ala Ile Glu Ala Gln Gln His Met Leu Gln Leu Thr Val Trp
545                 550                 555                 560
```

```
Gly Ile Lys Gln Leu Gln Thr Arg Val Leu Ala Ile Glu Arg Tyr Leu
                565                 570                 575

Lys Asp Gln Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile
            580                 585                 590

Cys Thr Thr Asn Val Pro Trp Asn Ser Ser Trp Ser Asn Lys Ser Gln
        595                 600                 605

Thr Asp Ile Trp Asn Asn Met Thr Trp Met Glu Trp Asp Arg Glu Ile
    610                 615                 620

Ser Asn Tyr Thr Asp Thr Ile Tyr Arg Leu Leu Glu Asp Ser Gln Thr
625                 630                 635                 640

Gln Gln Glu Lys Asn Glu Lys Asp Leu Leu Ala Leu Asp Ser Trp Lys
                645                 650                 655

Asn Leu Trp Ser Trp Phe Asp Ile Ser Asn Trp Leu Trp Tyr Ile Lys
            660                 665                 670

Ser Arg Ile Glu Gly Arg Gly Ser Gly Gly Tyr Ile Pro Glu Ala Pro
        675                 680                 685

Arg Asp Gly Gln Ala Tyr Val Arg Lys Asp Gly Glu Trp Val Leu Leu
    690                 695                 700

Ser Thr Phe Leu
705
```

<210> 8

<211> 625

<212> PRT

<213> Artificial Sequence

<220>

<223> C4 fragment: gp120-truncated gp41 without signal peptide and transmembrane domain

<400> 8

```
Met Gly Asn Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys
1               5                   10                  15

Asp Ala Lys Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Glu
            20                  25                  30

Lys Glu Val His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp
        35                  40                  45

Pro Asn Pro Gln Glu Ile Val Leu Gly Asn Val Thr Glu Asn Phe Asn
    50                  55                  60

Met Trp Lys Asn Asp Met Val Asp Gln Met His Glu Asp Ile Ile Ser
65                  70                  75                  80

Leu Trp Asp Ala Ser Leu Glu Pro Cys Val Lys Leu Thr Pro Leu Cys
```

```
                85                  90                  95

Val Thr Leu Asn Cys Arg Asn Val Arg Asn Val Ser Ser Asn Gly Thr
            100                 105                 110

Tyr Asn Ile Ile His Asn Glu Thr Tyr Lys Glu Met Lys Asn Cys Ser
        115                 120                 125

Phe Asn Ala Thr Thr Val Val Glu Asp Arg Lys Gln Lys Val His Ala
    130                 135                 140

Leu Phe Tyr Arg Leu Asp Ile Val Pro Leu Asp Glu Asn Asn Ser Ser
145                 150                 155                 160

Glu Lys Ser Ser Glu Asn Ser Ser Glu Tyr Tyr Arg Leu Ile Asn Cys
                165                 170                 175

Asn Thr Ser Ala Ile Thr Gln Ala Cys Pro Lys Val Ser Phe Asp Pro
            180                 185                 190

Ile Pro Ile His Tyr Cys Ala Pro Ala Gly Tyr Ala Ile Leu Lys Cys
        195                 200                 205

Asn Asn Lys Thr Phe Asn Gly Thr Gly Pro Cys Asn Asn Val Ser Thr
    210                 215                 220

Val Gln Cys Thr His Gly Ile Lys Pro Val Val Ser Thr Gln Leu Leu
225                 230                 235                 240

Leu Asn Gly Ser Leu Ala Glu Glu Glu Ile Ile Ile Arg Ser Glu Asn
                245                 250                 255

Leu Thr Asn Asn Ala Lys Thr Ile Ile Val His Leu Asn Glu Thr Val
            260                 265                 270

Asn Ile Thr Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser Ile Arg
        275                 280                 285

Ile Gly Pro Gly Gln Thr Phe Tyr Ala Thr Gly Asp Ile Ile Gly Asp
    290                 295                 300

Ile Arg Gln Ala His Cys Asn Leu Ser Arg Asp Gly Trp Asn Lys Thr
305                 310                 315                 320

Leu Gln Gly Val Lys Lys Lys Leu Ala Glu His Phe Pro Asn Lys Thr
                325                 330                 335

Ile Lys Phe Ala Pro His Ser Gly Gly Asp Leu Glu Ile Thr Thr His
            340                 345                 350

Thr Phe Asn Cys Arg Gly Glu Phe Phe Tyr Cys Asn Thr Ser Asn Leu
        355                 360                 365

Phe Asn Glu Ser Asn Ile Glu Arg Asn Asp Ser Ile Ile Thr Leu Pro
    370                 375                 380

Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly Arg Ala
385                 390                 395                 400
```

385 390 395 400

Ile Tyr Ala Pro Pro Ile Ala Gly Asn Ile Thr Cys Arg Ser Asn Ile
405 410 415

Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly Ser Asn Asn Gly Val Pro
420 425 430

Asn Asp Thr Glu Thr Phe Arg Pro Gly Gly Gly Asp Met Arg Asn Asn
435 440 445

Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Glu Val Lys Pro Leu
450 455 460

Gly Val Ala Pro Thr Glu Ala Lys Arg Arg Val Val Glu Arg Glu Lys
465 470 475 480

Arg Ala Val Gly Ile Gly Ala Val Phe Leu Gly Ile Leu Gly Ala Ala
485 490 495

Gly Ser Thr Met Gly Ala Ala Ser Ile Thr Leu Thr Val Gln Ala Arg
500 505 510

Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Ser Asn Leu Leu Arg Ala
515 520 525

Ile Glu Ala Gln Gln His Met Leu Gln Leu Thr Val Trp Gly Ile Lys
530 535 540

Gln Leu Gln Thr Arg Val Leu Ala Ile Glu Arg Tyr Leu Gln Asp Gln
545 550 555 560

Gln Leu Leu Gly Leu Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr
565 570 575

Ala Val Pro Trp Asn Thr Ser Trp Ser Asn Lys Ser Gln Thr Asp Ile
580 585 590

Trp Asp Asn Met Thr Trp Met Gln Trp Asp Lys Glu Ile Gly Asn Tyr
595 600 605

Thr Gly Glu Ile Tyr Arg Leu Leu Glu Glu Ser Gln Asn Gln Gln Glu
610 615 620

Lys
625

<210> 9

<211> 29

<212> PRT

<213> Artificial Sequence

<220>

<223> signal sequence

<400> 9

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

Ser Ser Leu Gly Phe Trp Met Leu Met Ile Tyr Ser Val
            20                  25
```

<210> 10

<211> 29

<212> PRT

<213> Artificial Sequence

<220>

<223> signal sequence

<400> 10

```
Met Arg Val Thr Gly Ile Arg Lys Asn Tyr Gln His Leu Trp Arg Trp
1               5                   10                  15

Gly Thr Met Leu Leu Gly Ile Leu Met Ile Cys Ser Ala
            20                  25
```

<210> 11

<211> 29

<212> PRT

<213> Artificial Sequence

<220>

<223> signal sequence

<400> 11

```
Met Arg Val Arg Gly Ile Gln Arg Asn Trp Pro Gln Trp Trp Ile Trp
1               5                   10                  15

Gly Ile Leu Gly Phe Trp Met Ile Ile Ile Cys Arg Val
            20                  25
```

<210> 12

<211> 29

<212> PRT

<213> Artificial Sequence

<220>

<223> signal sequence

<400> 12

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

Gly Thr Leu Ile Leu Gly Met Leu Met Ile Cys Ser Ala
            20                  25
```

<210> 13

<211> 22

<212> PRT

<213> Artificial Sequence

<220>

<223> transmembrane domain

<400> 13

```
Ile Phe Ile Met Ile Val Gly Gly Leu Ile Gly Leu Arg Ile Ile Phe
1               5                   10                  15

Ala Val Leu Ser Ile Val
            20
```

<210> 14

<211> 7

<212> PRT

<213> Artificial Sequence

<220>

<223> truncated cytoplasmic region

<400> 14

```
Asn Arg Val Arg Gln Gly Tyr
1               5
```

<210> 15

<211> 32

<212> PRT

<213> Artificial Sequence

<220>

<223> GCN4 trimerization domain

<400> 15

```
Met Lys Gln Ile Glu Asp Lys Ile Glu Glu Ile Leu Ser Lys Ile Tyr
1               5                   10                  15

His Ile Glu Asn Glu Ile Ala Arg Ile Lys Lys Leu Ile Gly Glu Val
            20                  25                  30
```

<210> 16

<211> 30

<212> PRT

<213> Artificial Sequence

<220>

<223> foldon trimerization domain

<400> 16

```
Gly Ser Gly Gly Tyr Ile Pro Glu Ala Pro Arg Asp Gly Gln Ala Tyr
1               5                   10                  15

Val Arg Lys Asp Gly Glu Trp Val Leu Leu Ser Thr Phe Leu
```

```
            20                  25                  30
```

<210> 17

<211> 862

<212> PRT

<213> Artificial Sequence

<220>

<223> C4 sequence

<400> 17

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

Ser Ser Leu Gly Phe Trp Met Leu Met Ile Tyr Ser Val Met Gly Asn
            20                  25                  30

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

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

His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
65                  70                  75                  80

Gln Glu Ile Val Leu Gly Asn Val Thr Glu Asn Phe Asn Met Trp Lys
                85                  90                  95

Asn Asp Met Val Asp Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp

            100                 105                 110

Ala Ser Leu Glu Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
        115                 120                 125

Asn Cys Arg Asn Val Arg Asn Val Ser Ser Asn Gly Thr Tyr Asn Ile
    130                 135                 140

Ile His Asn Glu Thr Tyr Lys Glu Met Lys Asn Cys Ser Phe Asn Ala
145                 150                 155                 160

Thr Thr Val Val Glu Asp Arg Lys Gln Lys Val His Ala Leu Phe Tyr
                165                 170                 175

Arg Leu Asp Ile Val Pro Leu Asp Glu Asn Asn Ser Ser Glu Lys Ser
            180                 185                 190

Ser Glu Asn Ser Ser Glu Tyr Tyr Arg Leu Ile Asn Cys Asn Thr Ser
        195                 200                 205

Ala Ile Thr Gln Ala Cys Pro Lys Val Ser Phe Asp Pro Ile Pro Ile
    210                 215                 220
```

```
His Tyr Cys Ala Pro Ala Gly Tyr Ala Ile Leu Lys Cys Asn Asn Lys
225                 230                 235                 240

Thr Phe Asn Gly Thr Gly Pro Cys Asn Asn Val Ser Thr Val Gln Cys
                245                 250                 255

Thr His Gly Ile Lys Pro Val Val Ser Thr Gln Leu Leu Leu Asn Gly
            260                 265                 270

Ser Leu Ala Glu Glu Glu Ile Ile Ile Arg Ser Glu Asn Leu Thr Asn
        275                 280                 285

Asn Ala Lys Thr Ile Ile Val His Leu Asn Glu Thr Val Asn Ile Thr
    290                 295                 300

Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser Ile Arg Ile Gly Pro
305                 310                 315                 320

Gly Gln Thr Phe Tyr Ala Thr Gly Asp Ile Ile Gly Asp Ile Arg Gln
                325                 330                 335

Ala His Cys Asn Leu Ser Arg Asp Gly Trp Asn Lys Thr Leu Gln Gly
            340                 345                 350

Val Lys Lys Lys Leu Ala Glu His Phe Pro Asn Lys Thr Ile Lys Phe
        355                 360                 365

Ala Pro His Ser Gly Gly Asp Leu Glu Ile Thr Thr His Thr Phe Asn
    370                 375                 380

Cys Arg Gly Glu Phe Phe Tyr Cys Asn Thr Ser Asn Leu Phe Asn Glu
385                 390                 395                 400

Ser Asn Ile Glu Arg Asn Asp Ser Ile Ile Thr Leu Pro Cys Arg Ile
                405                 410                 415

Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly Arg Ala Ile Tyr Ala
            420                 425                 430

Pro Pro Ile Ala Gly Asn Ile Thr Cys Arg Ser Asn Ile Thr Gly Leu
        435                 440                 445

Leu Leu Thr Arg Asp Gly Gly Ser Asn Asn Gly Val Pro Asn Asp Thr
    450                 455                 460

Glu Thr Phe Arg Pro Gly Gly Gly Asp Met Arg Asn Asn Trp Arg Ser
465                 470                 475                 480

Glu Leu Tyr Lys Tyr Lys Val Val Glu Val Lys Pro Leu Gly Val Ala
                485                 490                 495

Pro Thr Glu Ala Lys Arg Arg Val Val Glu Arg Glu Lys Arg Ala Val
            500                 505                 510

Gly Ile Gly Ala Val Phe Leu Gly Ile Leu Gly Ala Ala Gly Ser Thr
        515                 520                 525
```

```
Met Gly Ala Ala Ser Ile Thr Leu Thr Val Gln Ala Arg Gln Leu Leu
    530                 535                 540

Ser Gly Ile Val Gln Gln Gln Ser Asn Leu Leu Arg Ala Ile Glu Ala
545                 550                 555                 560

Gln Gln His Met Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu Gln
                565                 570                 575

Thr Arg Val Leu Ala Ile Glu Arg Tyr Leu Gln Asp Gln Gln Leu Leu
            580                 585                 590

Gly Leu Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr Ala Val Pro
        595                 600                 605

Trp Asn Thr Ser Trp Ser Asn Lys Ser Gln Thr Asp Ile Trp Asp Asn
    610                 615                 620

Met Thr Trp Met Gln Trp Asp Lys Glu Ile Gly Asn Tyr Thr Gly Glu
625                 630                 635                 640

Ile Tyr Arg Leu Leu Glu Glu Ser Gln Asn Gln Gln Glu Lys Asn Glu
                645                 650                 655

Lys Asp Leu Leu Ala Leu Asp Ser Trp Asn Asn Leu Trp Asn Trp Phe
            660                 665                 670

Ser Ile Ser Lys Trp Leu Trp Tyr Ile Lys Ile Phe Ile Met Ile Val
        675                 680                 685

Gly Gly Leu Ile Gly Leu Arg Ile Ile Phe Ala Val Leu Ser Ile Val
    690                 695                 700

Asn Arg Val Arg Gln Gly Tyr Ser Pro Leu Ser Leu Gln Thr Leu Thr
705                 710                 715                 720

Gln Asn Pro Gly Gly Leu Asp Arg Leu Gly Arg Ile Glu Glu Glu Gly
                725                 730                 735

Gly Glu Gln Asp Lys Asp Arg Ser Ile Arg Leu Val Asn Gly Phe Phe
            740                 745                 750

Ala Leu Phe Trp Asp Asp Leu Arg Ser Leu Cys Leu Phe Ser Tyr His
        755                 760                 765

Arg Leu Arg Asp Phe Ile Leu Ile Val Ala Arg Ala Val Glu Leu Leu
    770                 775                 780

Gly Arg Ser Ser Leu Arg Gly Leu Gln Arg Gly Trp Glu Ile Leu Lys
785                 790                 795                 800

Tyr Leu Gly Ser Leu Leu Gln Tyr Trp Gly Leu Glu Leu Lys Lys Ser
                805                 810                 815

Ala Ile Asn Leu Leu Asp Thr Ile Ala Ile Ala Val Ala Glu Gly Thr
            820                 825                 830

- - -- -- -- - -- -- - -- - - -- -- - -
```

```
Asp Arg Ile Ile Glu Leu Ile Gln Arg Ile Cys Arg Ala Ile Cys Asn
        835                 840                 845

Ile Pro Arg Arg Ile Arg Gln Gly Phe Glu Ala Ala Leu Gln
    850                 855                 860
```

<210> 18

<211> 711

<212> PRT

<213> Artificial Sequence

<220>

<223> C4D7 sequence

<400> 18

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

Ser Ser Leu Gly Phe Trp Met Leu Met Ile Tyr Ser Val Met Gly Asn
            20                  25                  30

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

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

His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
65                  70                  75                  80

Gln Glu Ile Val Leu Gly Asn Val Thr Glu Asn Phe Asn Met Trp Lys
                85                  90                  95

Asn Asp Met Val Asp Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp
            100                 105                 110

Ala Ser Leu Glu Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
        115                 120                 125

Asn Cys Arg Asn Val Arg Asn Val Ser Ser Asn Gly Thr Tyr Asn Ile
    130                 135                 140

Ile His Asn Glu Thr Tyr Lys Glu Met Lys Asn Cys Ser Phe Asn Ala
145                 150                 155                 160

Thr Thr Val Val Glu Asp Arg Lys Gln Lys Val His Ala Leu Phe Tyr
                165                 170                 175

Arg Leu Asp Ile Val Pro Leu Asp Glu Asn Asn Ser Ser Glu Lys Ser
            180                 185                 190

Ser Glu Asn Ser Ser Glu Tyr Tyr Arg Leu Ile Asn Cys Asn Thr Ser
        195                 200                 205

Ala Ile Thr Gln Ala Cys Pro Lys Val Ser Phe Asp Pro Ile Pro Ile
    210                 215                 220
```

```
His Tyr Cys Ala Pro Ala Gly Tyr Ala Ile Leu Lys Cys Asn Asn Lys
225                 230                 235                 240

Thr Phe Asn Gly Thr Gly Pro Cys Asn Asn Val Ser Thr Val Gln Cys
                245                 250                 255

Thr His Gly Ile Lys Pro Val Val Ser Thr Gln Leu Leu Leu Asn Gly
            260                 265                 270

Ser Leu Ala Glu Glu Glu Ile Ile Ile Arg Ser Glu Asn Leu Thr Asn
        275                 280                 285

Asn Ala Lys Thr Ile Ile Val His Leu Asn Glu Thr Val Asn Ile Thr
    290                 295                 300

Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser Ile Arg Ile Gly Pro
305                 310                 315                 320

Gly Gln Thr Phe Tyr Ala Thr Gly Asp Ile Ile Gly Asp Ile Arg Gln
                325                 330                 335

Ala His Cys Asn Leu Ser Arg Asp Gly Trp Asn Lys Thr Leu Gln Gly
            340                 345                 350

Val Lys Lys Lys Leu Ala Glu His Phe Pro Asn Lys Thr Ile Lys Phe
        355                 360                 365

Ala Pro His Ser Gly Gly Asp Leu Glu Ile Thr Thr His Thr Phe Asn
    370                 375                 380

Cys Arg Gly Glu Phe Phe Tyr Cys Asn Thr Ser Asn Leu Phe Asn Glu
385                 390                 395                 400

Ser Asn Ile Glu Arg Asn Asp Ser Ile Ile Thr Leu Pro Cys Arg Ile
                405                 410                 415

Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly Arg Ala Ile Tyr Ala
            420                 425                 430

Pro Pro Ile Ala Gly Asn Ile Thr Cys Arg Ser Asn Ile Thr Gly Leu
        435                 440                 445

Leu Leu Thr Arg Asp Gly Gly Ser Asn Asn Gly Val Pro Asn Asp Thr
    450                 455                 460

Glu Thr Phe Arg Pro Gly Gly Gly Asp Met Arg Asn Asn Trp Arg Ser
465                 470                 475                 480

Glu Leu Tyr Lys Tyr Lys Val Val Glu Val Lys Pro Leu Gly Val Ala
                485                 490                 495

Pro Thr Glu Ala Lys Arg Arg Val Val Glu Arg Glu Lys Arg Ala Val
            500                 505                 510

Gly Ile Gly Ala Val Phe Leu Gly Ile Leu Gly Ala Ala Gly Ser Thr
        515                 520                 525
```

```
Met Gly Ala Ala Ser Ile Thr Leu Thr Val Gln Ala Arg Gln Leu Leu
    530                 535                 540

Ser Gly Ile Val Gln Gln Gln Ser Asn Leu Leu Arg Ala Ile Glu Ala
545                 550                 555                 560

Gln Gln His Met Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu Gln
                565                 570                 575

Thr Arg Val Leu Ala Ile Glu Arg Tyr Leu Gln Asp Gln Gln Leu Leu
            580                 585                 590

Gly Leu Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr Ala Val Pro
        595                 600                 605

Trp Asn Thr Ser Trp Ser Asn Lys Ser Gln Thr Asp Ile Trp Asp Asn
    610                 615                 620

Met Thr Trp Met Gln Trp Asp Lys Glu Ile Gly Asn Tyr Thr Gly Glu
625                 630                 635                 640

Ile Tyr Arg Leu Leu Glu Glu Ser Gln Asn Gln Gln Glu Lys Asn Glu
                645                 650                 655

Lys Asp Leu Leu Ala Leu Asp Ser Trp Asn Asn Leu Trp Asn Trp Phe
            660                 665                 670

Ser Ile Ser Lys Trp Leu Trp Tyr Ile Lys Ile Phe Ile Met Ile Val
        675                 680                 685

Gly Gly Leu Ile Gly Leu Arg Ile Ile Phe Ala Val Leu Ser Ile Val
    690                 695                 700

Asn Arg Val Arg Gln Gly Tyr

705                 710
```

<210> 19

<211> 704

<212> PRT

<213> Artificial Sequence

<220>

<223> sC4 sequence

<400> 19

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

Ser Ser Leu Gly Phe Trp Met Leu Met Ile Tyr Ser Val Met Gly Asn
            20                  25                  30

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

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

His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
65                  70                  75                  80

Gln Glu Ile Val Leu Gly Asn Val Thr Glu Asn Phe Asn Met Trp Lys
                85                  90                  95

Asn Asp Met Val Asp Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp
            100                 105                 110

Ala Ser Leu Glu Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
        115                 120                 125

Asn Cys Arg Asn Val Arg Asn Val Ser Ser Asn Gly Thr Tyr Asn Ile
    130                 135                 140

Ile His Asn Glu Thr Tyr Lys Glu Met Lys Asn Cys Ser Phe Asn Ala
145                 150                 155                 160

Thr Thr Val Val Glu Asp Arg Lys Gln Lys Val His Ala Leu Phe Tyr
                165                 170                 175

Arg Leu Asp Ile Val Pro Leu Asp Glu Asn Asn Ser Ser Glu Lys Ser
            180                 185                 190

Ser Glu Asn Ser Ser Glu Tyr Tyr Arg Leu Ile Asn Cys Asn Thr Ser
        195                 200                 205

Ala Ile Thr Gln Ala Cys Pro Lys Val Ser Phe Asp Pro Ile Pro Ile
    210                 215                 220

His Tyr Cys Ala Pro Ala Gly Tyr Ala Ile Leu Lys Cys Asn Asn Lys
225                 230                 235                 240

Thr Phe Asn Gly Thr Gly Pro Cys Asn Asn Val Ser Thr Val Gln Cys
                245                 250                 255

Thr His Gly Ile Lys Pro Val Val Ser Thr Gln Leu Leu Leu Asn Gly
            260                 265                 270

Ser Leu Ala Glu Glu Glu Ile Ile Ile Arg Ser Glu Asn Leu Thr Asn
        275                 280                 285

Asn Ala Lys Thr Ile Ile Val His Leu Asn Glu Thr Val Asn Ile Thr
    290                 295                 300

Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser Ile Arg Ile Gly Pro
305                 310                 315                 320

Gly Gln Thr Phe Tyr Ala Thr Gly Asp Ile Ile Gly Asp Ile Arg Gln
                325                 330                 335
```

```
Ala His Cys Asn Leu Ser Arg Asp Gly Trp Asn Lys Thr Leu Gln Gly
            340                 345                 350

Val Lys Lys Lys Leu Ala Glu His Phe Pro Asn Lys Thr Ile Lys Phe
        355                 360                 365

Ala Pro His Ser Gly Gly Asp Leu Glu Ile Thr Thr His Thr Phe Asn
    370                 375                 380

Cys Arg Gly Glu Phe Phe Tyr Cys Asn Thr Ser Asn Leu Phe Asn Glu
385                 390                 395                 400

Ser Asn Ile Glu Arg Asn Asp Ser Ile Ile Thr Leu Pro Cys Arg Ile
                405                 410                 415

Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly Arg Ala Ile Tyr Ala
            420                 425                 430

Pro Pro Ile Ala Gly Asn Ile Thr Cys Arg Ser Asn Ile Thr Gly Leu
        435                 440                 445

Leu Leu Thr Arg Asp Gly Gly Ser Asn Asn Gly Val Pro Asn Asp Thr

    450                 455                 460

Glu Thr Phe Arg Pro Gly Gly Gly Asp Met Arg Asn Asn Trp Arg Ser
465                 470                 475                 480

Glu Leu Tyr Lys Tyr Lys Val Val Glu Val Lys Pro Leu Gly Val Ala
                485                 490                 495

Pro Thr Glu Ala Lys Arg Arg Val Val Glu Arg Glu Glu Arg Ala Val
            500                 505                 510

Gly Ile Gly Ala Val Phe Leu Gly Ile Leu Gly Ala Ala Gly Ser Thr
        515                 520                 525

Met Gly Ala Ala Ser Ile Thr Leu Thr Val Gln Ala Arg Gln Leu Leu
    530                 535                 540

Ser Gly Ile Val Gln Gln Gln Ser Asn Leu Leu Arg Ala Ile Glu Ala
545                 550                 555                 560

Gln Gln His Met Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu Gln
                565                 570                 575

Thr Arg Val Leu Ala Ile Glu Arg Tyr Leu Gln Asp Gln Gln Leu Leu
            580                 585                 590

Gly Leu Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr Ala Val Pro
        595                 600                 605

Trp Asn Thr Ser Trp Ser Asn Lys Ser Gln Thr Asp Ile Trp Asp Asn
    610                 615                 620

Met Thr Trp Met Gln Trp Asp Lys Glu Ile Gly Asn Tyr Thr Gly Glu
625                 630                 635                 640
```

625 630 635 640

Ile Tyr Arg Leu Leu Glu Glu Ser Gln Asn Gln Gln Glu Lys Met Lys
645 650 655

Gln Ile Glu Asp Lys Ile Glu Glu Ile Leu Ser Lys Ile Tyr His Ile
660 665 670

Glu Asn Glu Ile Ala Arg Ile Lys Lys Leu Ile Gly Glu Val Gly Ser
675 680 685

Gly Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys Arg
690 695 700

<210> 20

<211> 4050

<212> DNA

<213> Artificial Sequence

<220>

<223> nucleotide sequence encoding mos1GagPol

<400> 20

atgggagcca gagccagcgt gctgtccgga ggggagctgg accgctggga gaagatcagg 60

ctgaggcctg gagggaagaa gaagtacagg ctgaagcaca tcgtgtgggc cagcagagag 120

ctggaacggt ttgccgtgaa ccctggcctg ctggaaacca gcgagggctg taggcagatt 180

ctgggacagc tgcagcccag cctgcagaca ggcagcgagg aactgcggag cctgtacaac 240

accgtggcca ccctgtactg cgtgcaccag cggatcgaga tcaaggacac caaagaagcc 300

ctggaaaaga tcgaggaaga gcagaacaag agcaagaaga aagcccagca ggctgccgct 360

gacacaggca acagcagcca ggtgtcccag aactacccca tcgtgcagaa catccaggga 420

cagatggtgc accaggccat cagccctcgg accctgaacg cctgggtgaa ggtggtggag 480

gaaaaggcct tcagccctga ggtgatcccc atgttctctg ccctgagcga gggagccaca 540

ccccaggacc tgaacaccat gctgaacacc gtgggagggc accaggctgc catgcagatg 600

ctgaaagaga caatcaacga ggaagctgcc gagtgggaca gggtccaccc agtgcacgct 660

ggacctatcg ctcctggcca gatgagagag cccagaggca gcgatattgc tggcaccacc 720

tccacactgc aggaacagat cggctggatg accaacaacc ctcccatccc tgtgggagag 780

atctacaagc ggtggatcat tctgggactg aacaagatcg tgcggatgta cagccctgtg 840

agcatcctgg acatcaggca gggacccaaa gagcccttca gggactacgt ggaccggttc 900

tacaagaccc tgagagccga gcaggccagc caggacgtga agaactggat gaccgagaca 960

ctgctggtgc agaacgccaa ccctgactgc aagaccatcc tgaaagccct gggacctgct 1020

gccaccctgg aagagatgat gacagcctgc cagggagtgg gaggacctgg ccacaaggcc 1080

agggtgctgg ccgaggccat gagccaggtg accaactctg ccaccatcat gatgcagaga 1140

ggcaacttcc ggaaccagag aaagaccgtg aagtgcttca actgtggcaa agagggacac 1200

attgccaaga actgcagggc tcccaggaag aaaggctgct ggaagtgcgg aaaagaaggc 1260

gaggagatga aggagtgcaa cgagaggcag gccaagttca tgggcaagat ctggggtagc 1320

caccagatga aggactgcac cgagaggcag gccaacttcc tgggcaagat ctggcctagc 1320

aacaagggca ggcctggcaa cttcctgcag aacagacccg agcccaccgc tcctcccgag 1380

gaaagcttcc ggtttggcga ggaaaccacc acccctagcc agaagcagga acccatcgac 1440

aaagagatgt accctctggc cagcctgaag agcctgttcg gcaacgaccc cagcagccag 1500

atggctccca tcagcccaat cgagacagtg cctgtgaagc tgaagcctgg catggacgga 1560

cccagggtga agcagtggcc tctgaccgag gaaaagatca aagccctgac agccatctgc 1620

gaggaaatgg aaaaagaggg caagatcacc aagatcggac ccgagaaccc ctacaacacc 1680

cctgtgttcg ccatcaagaa gaaagacagc accaagtgga ggaaactggt ggacttcaga 1740

gagctgaaca agcggaccca ggacttctgg gaggtgcagc tgggcatccc tcaccctgct 1800

ggcctgaaga aaaagaaaag cgtgaccgtg ctggctgtgg gagatgccta cttcagcgtg 1860

cctctggacg agggcttccg gaagtacaca gccttcacca tccccagcac caacaacgag 1920

acacctggca tcagatacca gtacaacgtg ctgcctcagg gctggaaagg cagccctgcc 1980

atcttccagt gcagcatgac cagaatcctg gaacccttca gagccaagaa ccctgagatc 2040

gtgatctacc agtatatggc tgccctctac gtgggcagcg acctggaaat cggacagcac 2100

agagccaaaa tcgaagaact ccgcgagcac ctgctgaagt ggggattcac cacccctgac 2160

aagaagcacc agaaagagcc tcccttcctg tggatgggct acgagctgca ccctgacaag 2220

tggaccgtgc agcccatcca gctgccagag aaggactcct ggaccgtgaa cgacatccag 2280

aaactggtcg gcaagctgaa ctgggccagc cagatctacc ctggcatcaa agtcagacag 2340

ctgtgtaagc tgctgagggg agccaaagca ctgaccgaca tcgtgcctct gacagaagaa 2400

gccgagctgg aactggccga gaacagagag atcctgaaag aacccgtgca cggagtgtac 2460

tacgaccccc ccaaggacct gattgccgag atccagaaac agggacacga ccagtggacc 2520

taccagatct atcaggaacc tttcaagaac ctgaaaacag gcaagtacgc caagatgcgg 2580

acagcccaca ccaacgacgt gaagcagctg accgaagccg tgcagaaaat cgccatggaa 2640

agcatcgtga tctggggaaa gacacccaag ttcaggctgc ccatccagaa agagacatgg 2700

gaaacctggt ggaccgacta ctggcaggcc acctggattc ccgagtggga gttcgtgaac 2760

accccacccc tggtgaagct gtggtatcag ctggaaaagg accctatcgc tggcgtggag 2820

acattctacg tggctggagc tgccaacaga gagacaaagc tgggcaaggc tggctacgtg 2880

accgacagag gcagacagaa aatcgtgagc ctgaccgaaa ccaccaacca gaaaacagcc 2940

ctgcaggcca tctatctggc actgcaggac agcggaagcg aggtgaacat cgtgacagcc 3000

agccagtatg ccctgggcat catccaggcc cagcctgaca agagcgagag cgagctggtg 3060

aaccagatca tcgagcagct gatcaagaaa gaacgggtgt acctgagctg ggtgccagcc 3120

cacaagggca tcggagggaa cgagcaggtg gacaagctgg tgtccagcgg aatccggaag 3180

gtgctgttcc tggacggcat cgataaagcc caggaagagc acgagaagta ccacagcaat 3240

tggagagcca tggccagcga cttcaacctg cctcccgtgg tggccaaaga aatcgtggcc 3300

agctgcgacc agtgccagct gaaaggcgag gccatgcacg gacaggtgga ctgctccct 3360

ggcatctggc agctggcatg cacccacctg gaaggcaaga tcattctggt ggccgtgcac 3420

gtggccagcg gatacatcga agccgaagtg atccctgccg agacagggca ggaaacagcc 3480

tacttcatcc tgaagctggc tggcagatgg cctgtgaagg tgatccacac agccaacggc 3540

```
agcaacttca cctctgctgc cgtgaaggct gcctgttggt gggctggcat tcagcaggaa     3600

tttggcatcc cctacaatcc ccagtctcag ggagtggtgg ccagcatgaa caaagagctg     3660

aagaagatca tcggacaggt cagggatcag gccgagcacc tgaaaactgc cgtccagatg     3720

gccgtgttca tccacaactt caagcggaag ggagggatcg gagggtactc tgctggcgag     3780

cggatcatcg acatcattgc caccgatatc cagaccaaag agctgcagaa acagatcatc     3840

aagatccaga acttcagggt gtactacagg gacagcaggg accccatctg gaagggacct     3900

gccaagctgc tgtggaaagg cgaaggagcc gtcgtcatcc aggacaacag cgacatcaag     3960

gtggtgccca gacggaaggt gaaaatcatc aaggactacg gcaaacagat ggctggagcc     4020

gactgtgtcg ctggcaggca ggacgaggac                                      4050
```

<210> 21
<211> 4023
<212> DNA
<213> Artificial Sequence

<220>
<223> nucleotide sequence encoding mos2GagPol

<400> 21

```
atgggagcca gagccagcat cctgcgagga gggaagctgg acaagtggga gaagatcagg       60

ctgaggcctg gagggaagaa acactacatg ctgaagcacc tggtctgggc cagcagagag      120

ctggaacggt ttgccctcaa tcctggcctg ctggaaacca gcgagggctg caagcagatc      180

atcaagcagc tgcagcctgc cctgcagaca ggcaccgagg aactgcggag cctgttcaac      240

accgtggcca ccctgtactg cgtgcatgcc gagatcgaag tgagggacac caaagaagcc      300

ctggacaaga tcgaggaaga gcagaacaag agccagcaga aaacccagca ggccaaagaa      360

gccgacggca aggtctccca gaactacccc atcgtgcaga acctgcaggg acagatggtg      420

caccagccca tcagccctcg gacactgaat gcctgggtga aggtgatcga ggaaaaggcc      480

ttcagccctg aggtgatccc catgttcaca gccctgagcg agggagccac accccaggac      540

ctgaacacca tgctgaacac cgtgggaggg caccaggctg ccatgcagat gctgaaggac      600

accatcaacg aggaagctgc cgagtgggac aggctgcacc ctgtgcacgc tggacctgtg      660

gctcctggcc agatgagaga gcccagaggc agcgatattg ctggcaccac ctccaatctg      720

caggaacaga tcgcctggat gaccagcaac cctcccatcc ctgtgggaga catctacaag      780

cggtggatca tcctgggact gaacaagatc gtgcggatgt acagccctac ctccatcctg      840

gacatcaagc agggacccaa agagcctttc agggactacg tggaccggtt cttcaagacc      900

ctgagagccg agcaggccac ccaggacgtg aagaactgga tgaccgacac cctgctggtg      960

cagaacgcca accctgactg caagaccatc ctgagagccc tgggacctgg agccaccctg     1020

gaagagatga tgacagcctg ccagggagtg ggaggaccct ctcacaaggc tagggtgctg     1080

gccgaggcca tgagccagac caacagcacc atcctgatgc agcggagcaa cttcaagggc     1140

agcaagcgga tcgtgaagtg cttcaactgt ggcaaagagg gacacattgc cagaaactgt     1200

aggggcaccca ggaagaaagg ctgctggaag tgcggaaaag aaggccacca gatgaaggac     1260
```

tgcaccgaga ggcaggccaa cttcctgggc aagatctggc ctagccacaa gggcagacct 1320

ggcaacttcc tgcagagcag acccgagccc accgctcctc cagccgagag cttccggttc 1380

gaggaaacca cccctgctcc caagcaggaa cctaaggaca gagagcctct gaccagcctg 1440

agaagcctgt tcggcagcga ccctctgagc cagatggctc ccatctcccc tatcgagaca 1500

gtgcctgtga agctgaagcc tggcatggac ggacccaagg tgaaacagtg gcctctgacc 1560

gaggaaaaga tcaaagccct ggtggagatc tgtaccgaga tggaaaaaga gggcaagatc 1620

agcaagatcg gacccgagaa cccctacaac acccctatct tcgccatcaa gaagaaagac 1680

agcaccaagt ggaggaaact ggtggacttc agagagctga acaagcggac ccaggacttc 1740

tgggaggtgc agctgggcat ccctcaccct gctggcctga agaaaaagaa aagcgtgacc 1800

gtgctggccg tgggagatgc ctacttcagc gtgcctctgg acgaggactt cagaaagtac 1860

acagccttca ccatccccag catcaacaac gagacacctg gcatcagata ccagtacaac 1920

gtgctgcctc agggatggaa gggctctcct gcaatcttcc agagcagcat gaccaagatc 1980

ctggaaccct tccggaagca gaaccctgac atcgtgatct accagtacat ggcagccctg 2040

tacgtcggca gcgacctgga aatcggacag caccggacca agatcgaaga actcaggcag 2100

cacctgctgc ggtggggatt caccaccccc gacaagaagc accagaaaga gcctcccttc 2160

ctgtggatgg gctacgagct gcacccagac aagtggaccg tgcagcccat cgtgctgcct 2220

gagaaggact cctggaccgt gaacgacatc cagaaactgg tcggcaagct gaactgggcc 2280

agccagatct acgctggcat caaagtgaag cagctgtgta agctcctgag aggcaccaaa 2340

gccctgaccg aggtggtgcc actgacagag gaagccgagc tggaactggc cgagaacaga 2400

gagatcctga aagaacccgt gcacggagtg tactacgacc ccagcaagga cctgattgcc 2460

gagatccaga agcagggaca gggacagtgg acctaccaga tctaccagga acccttcaag 2520

aacctgaaaa caggcaagta cgccaggatg aggggagccc acaccaacga cgtcaaacag 2580

ctgaccgaag ccgtgcagaa gatcgccacc gagagcatcg tgatttgggg aaagacaccc 2640

aagttcaagc tgcccatcca gaaagagaca tgggaggcct ggtggaccga gtactggcag 2700

gccacctgga ttcccgagtg ggagttcgtg aacaccccac ccctggtgaa gctgtggtat 2760

cagctggaaa aagaacccat cgtgggagcc gagacattct acgtggctgg agctgccaac 2820

agagagacaa agctgggcaa ggctggctac gtgaccgaca gaggcaggca gaaagtggtg 2880

tccctgaccg ataccaccaa ccagaaaaca gccctgcagg ccatccacct ggctctgcag 2940

gactctggcc tggaagtgaa catcgtgaca gccagccagt atgccctggg catcattcag 3000

gcacagcctg acaagagcga gagcgagctg gtgtctcaga tcattgagca gctgatcaag 3060

aaagaaaagg tgtacctggc ctgggtgcca gcccacaagg ggatcggagg gaacgagcag 3120

gtggacaagc tggtgtccag gggcatccgg aaggtgctgt ttctggacgg catcgacaaa 3180

gcccaggaag agcacgagaa gtaccacagc aattggagag ccatggccag cgagttcaac 3240

ctgcctccca tcgtggccaa agaaatcgtg gcctcttgcg acaagtgcca gctgaaaggc 3300

gaggccattc acggacaggt ggactgcagc ccaggcatct ggcagctggc ctgcacccac 3360

ctggaaggca aggtgatcct ggtggccgtg cacgtggcct ctggatacat cgaagccgaa 3420

gtgatccctg ccgagacagg ccaggaaaca gcctacttcc tgctgaagct ggctggcagg 3480

tggcctgtga aaaccatcca cacagccaac ggcagcaact tcacctctgc caccgtgaag 3540

gctgcctgtt ggtgggctgg cattaagcag gaatttggca tcccctacaa ccctcagtct 3600

cagggagtgg tggcctccat caacaaagag ctgaagaaga tcatcggaca ggtcagggat 3660

caggccgagc atctgaaaac agccgtccag atggccgtgt tcatccacaa cttcaagcgg 3720

aagggaggga tcggagagta ctctgctggc gagaggatcg tggacattat cgccagcgat 3780

atccagacca aagaactgca gaagcagatc acaaagatcc agaacttcag ggtgtactac 3840

agggacagca gagatccccc gtggaaggga cctgccaagc tgctgtggaa aggcgaagga 3900

gccgtcgtca tccaggacaa cagcgacatc aaggtggtgc ccagacggaa ggccaagatc 3960

atcagagact acggcaaaca gatggctggc gacgactgcg tcgcctctag gcaggacgag 4020

gac 4023

<210> 22
<211> 2055
<212> DNA
<213> Artificial Sequence

<220>
<223> nucleotide sequence encoding moslEnv

<400> 22
atgcgggtga ccggcatccg gaagaactac cagcacctgt ggcggtgggg caccatgctg 60

ctgggcatcc tgatgatttg ctctgccgcc ggaaagctgt gggtgaccgt gtactacggc 120

gtgcccgtgt ggaaagaggc caccaccacc ctgttctgcg ccagcgacgc caaggcctac 180

gacaccgagg tgcacaacgt gtgggccacc cacgcctgcg tgcccaccga ccccaacccc 240

caggaagtgg tcctggaaaa cgtgaccgag aacttcaaca tgtggaagaa caacatggtg 300

gagcagatgc acgaggacat catcagcctg tgggaccaga gcctgaagcc ctgcgtgaag 360

ctgaccccc tgtgcgtgac cctgaactgc accgacgacg tgcggaacgt gaccaacaac 420

gccaccaaca ccaacagcag ctggggcgag cctatggaaa agggcgagat caagaactgc 480

agcttcaaca tcaccacctc catccggaac aaggtgcaga agcagtacgc cctgttctac 540

aagctggacg tggtgcccat cgacaacgac agcaacaaca ccaactaccg gctgatcagc 600

tgcaacacca gcgtgatcac ccaggcctgc cccaaggtgt ccttcgagcc catccccatc 660

cactactgcg cccctgccgg cttcgccatc ctgaagtgca acgacaagaa gttcaacggc 720

accggccct gcaccaacgt gagcaccgtg cagtgcaccc acggcatccg gcccgtggtg 780

tccacccagc tgctgctgaa cggcagcctg gccgaggaag aggtggtgat cagaagcgag 840

aatttcacca acaatgccaa gaccatcatg gtgcagctga acgtgagcgt ggagatcaac 900

tgcacccggc ccaacaacaa cacccggaag agcatccaca tcggccctgg cagggccttc 960

tacacagccg gcgacatcat cggcgacatc cggcaggccc actgcaacat cagccgggcc 1020

aactggaaca acaccctgcg gcagatcgtg gagaagctgg gcaagcagtt cggcaacaac 1080

aagaccatcg tgttcaacca cagcagcggc ggagaccccg agatcgtgat gcacagcttc 1140

aactgtggcg gcgagttctt ctactgcaac agcaccaagc tgttcaacag cacctggacc 1200

```
tggaacaact ccacctggaa taacaccaag cggagcaacg acaccgaaga gcacatcacc     1260
ctgccctgcc ggatcaagca gattatcaat atgtggcagg aggtcggcaa ggccatgtac     1320
gcccctccca tccggggcca gatccggtgc agcagcaaca tcaccggcct gctgctgacc     1380
cgggacggcg gcaacgatac cagcggcacc gagatcttcc ggcctggcgg cggagatatg     1440
cgggacaact ggcggagcga gctgtacaag tacaaggtgg tgaagatcga gcccctgggc     1500
gtggctccca ccaaggccaa gcggcgggtg gtgcagagcg agaagagcgc cgtgggcatc     1560
ggcgccgtgt ttctgggctt cctgggagcc gccggaagca ccatgggagc cgccagcatg     1620
accctgaccg tgcaggcccg gctgctgctg tccggcatcg tgcagcagca gaacaacctg     1680
ctccgggcca tcgaggccca gcagcacctg ctgcagctga ccgtgtgggg catcaagcag     1740
ctgcaggcca gggtgctggc cgtggagaga tacctgaagg atcagcagct cctggggatc     1800
tggggctgca gcggcaagct gatctgcacc accaccgtgc cctggaacgc cagctggtcc     1860
aacaagagcc tggacaagat ctggaacaat atgacctgga tggaatggga gcgcgagatc     1920
aacaattaca ccagcctgat ctacaccctg atcgaggaaa gccagaacca gcaggaaaag     1980
aacgagcagg aactgctgga actggacaag tgggccagcc tgtggaactg gttcgacatc     2040
agcaactggc tgtgg                                                      2055
```

<210> 23
<211> 2052
<212> DNA
<213> Artificial Sequence

<220>
<223> nucleotide sequence encoding mos2Env

<400> 23

```
atgagagtgc ggggcatcca gcggaactgg ccccagtggt ggatctgggg catcctgggc       60
ttttggatga tcatcatctg ccgggtgatg ggcaacctgt gggtgaccgt gtactacggc      120
gtgcccgtgt ggaaagaggc caagaccacc ctgttctgcg ccagcgacgc caaggcctac      180
gagaaagagg tgcacaacgt gtgggccacc cacgcctgcg tgcccaccga ccccaacccc      240
caggaaatgg tcctggaaaa cgtgaccgag aacttcaaca tgtggaagaa cgacatggtg      300
gaccagatgc acgaggacat catccggctg tgggaccaga gcctgaagcc ctgcgtgaag      360
ctgaccccca tgtgcgtgac cctggaatgc cggaacgtga gaaacgtgag cagcaacggc      420
acctacaaca tcatccacaa cgagacctac aaagagatga agaactgcag cttcaacgcc      480
accaccgtgg tggaggaccg gaagcagaag gtgcacgccc tgttctaccg gctggacatc      540
gtgcccctgg acgagaacaa cagcagcgag aagtccagcg agaacagctc cgagtactac      600
cggctgatca actgcaacac cagcgccatc acccaggcct gccccaaggt gtccttcgac      660
cccatcccca tccactactg cgcccctgcc ggctacgcca tcctgaagtg caacaacaag      720
accttcaacg gcaccggccc ctgcaacaac gtgagcaccg tgcagtgcac ccacggcatc      780
aagcccgtgg tgtccaccca gctgctgctg aacggcagcc tggccgagga agagatcatc      840
atccggtccg agaacctgac caacaacgcc aagaccatca tcgtgcacct gaatgagacc      900
```

gtgaacatca cctgcacccg gcccaacaac aacacccgga agagcatccg gatcggccct 960

ggccagacct tttacgccac cggcgacatc atcggcgaca tccggcaggc ccactgcaac 1020

ctgagccggg acggctggaa caagaccctg cagggcgtga agaagaagct ggccgagcac 1080

ttccccaata agaccatcaa cttcaccagc agcagcggcg gagacctgga aatcaccacc 1140

cacagcttca actgcagggg cgagttcttc tactgcaata cctccggcct gttcaatggc 1200

acctacatgc ccaacggcac caacagcaac agcagcagca acatcaccct gccctgccgg 1260

atcaagcaga tcatcaatat gtggcaggag gtcggcaggg ccatgtacgc ccctcccatc 1320

gccggcaata tcacctgccg gtccaacatc accggcctgc tgctgaccag ggacggcggc 1380

agcaacaacg gcgtgcctaa cgacaccgag accttccggc ctggcggcgg agatatgcgg 1440

aacaactggc ggagcgagct gtacaagtac aaggtggtgg aggtgaagcc cctgggcgtg 1500

gctcctaccg aggccaagcg gcgggtggtg gagagcgaga agagcgccgt gggcatcggc 1560

gccgtgtttc tgggcattct gggagccgcc ggaagcacca tgggagccgc cagcatcacc 1620

ctgaccgtgc aggcccggca gctgctgtcc ggcatcgtgc agcagcagag caacctgctg 1680

agagccatcg aggcccagca gcacatgctg cagctgaccg tgtggggcat caagcagctg 1740

cagacccggg tgctggccat cgagagatac ctgcaggatc agcagctcct gggcctgtgg 1800

ggctgcagcg gcaagctgat ctgcaccacc gccgtgccct ggaacaccag ctggtccaac 1860

aagagccaga ccgacatctg ggacaacatg acctggatgc agtgggacaa agagatcggc 1920

aactacaccg gcgagatcta caggctgctg gaagagagcc agaaccagca ggaaaagaac 1980

gagaaggacc tgctggccct ggacagctgg aagaacctgt ggaactggtt cgacatcacc 2040

aactggctgt gg 2052

<210> 24
<211> 829
<212> DNA
<213> Artificial Sequence

<220>
<223> CMV promoter used for expression of antigens in Ad26 vectors

<400> 24
tcaatattgg ccattagcca tattattcat tggttatata gcataaatca atattggcta 60

ttggccattg catacgttgt atccatatca taatatgtac atttatattg gctcatgtcc 120

aacattaccg ccatgttgac attgattatt gactagttat taatagtaat caattacggg 180

gtcattagtt catagcccat atatggagtt ccgcgttaca taacttacgg taaatggccc 240

gcctggctga ccgcccaacg acccccgccc attgacgtca ataatgacgt atgttcccat 300

agtaacgcca atagggactt tccattgacg tcaatgggtg gagtatttac ggtaaactgc 360

ccacttggca gtacatcaag tgtatcatat gccaagtacg cccctattg acgtcaatga 420

cggtaaatgg cccgcctggc attatgccca gtacatgacc ttatgggact ttcctacttg 480

gcagtacatc tacgtattag tcatcgctat taccatggtg atgcggtttt ggcagtacat 540

caatgggcgt ggatagcggt ttgactcacg gggatttcca agtctccacc ccattgacgt 600

caatgggagt ttgttttggc accaaaatca acgggacttt ccaaaatgtc gtaacaactc 660

cgccccattg acgcaaatgg gcggtaggcg tgtacggtgg gaggtctata taagcagagc 720

tcgtttagtg aaccgtcaga tcgcctggag acgccatcca cgctgttttg acctccatag 780

aagacaccgg gaccgatcca gcctccgcgg ccgggaacgg tgcattgga 829

<210> 25
<211> 2586
<212> DNA
<213> Artificial Sequence

<220>
<223> nucleotide sequence encoding C4

<400> 25
atgagagtgc ggggcatgct gagaaactgg cagcagtggt ggatctggtc cagcctgggc 60

ttctggatgc tgatgatcta cagcgtgatg ggcaacctgt gggtcaccgt gtactacggc 120

gtgcccgtgt ggaaggacgc caagaccacc ctgttttgcg cctccgatgc caaggcctac 180

gagaaagagg tgcacaacgt ctgggccacc cacgcctgtg tgcccaccga ccccaatccc 240

caggaaatcg tcctgggcaa cgtgaccgag aacttcaaca tgtggaagaa cgacatggtc 300

gatcagatgc acgaggacat catctccctg tgggacgcct ccctggaacc ctgcgtgaag 360

ctgacccctc tgtgcgtgac cctgaactgc cggaacgtgc gcaacgtgtc cagcaacggc 420

acctacaaca tcatccacaa cgagacatac aaagagatga agaactgcag cttcaacgct 480

accaccgtgg tcgaggaccg gaagcagaag gtgcacgccc tgttctaccg gctggacatc 540

gtgcccctgg acgagaacaa cagcagcgag aagtcctccg agaacagctc cgagtactac 600

agactgatca actgcaacac cagcgccatc acccaggcct gccccaaggt gtccttcgac 660

cctatcccca tccactactg cgcccctgcc ggctacgcca tcctgaagtg caacaacaag 720

accttcaatg gcaccggccc ctgcaacaat gtgtccaccg tgcagtgcac ccacggcatc 780

aagcccgtgg tgtctaccca gctgctgctg aacggcagcc tggccgagga agagatcatt 840

atcagaagcg agaacctgac caacaacgcc aaaaccatca tcgtccacct gaacgaaacc 900

gtgaacatca cctgtacccg gcctaacaac aacacccgga agtccatccg gatcggccct 960

ggccagacct tttacgccac cggcgatatt atcggcgaca tccggcaggc ccactgcaat 1020

ctgagccggg acggctggaa caagacactg cagggcgtca agaagaagct ggccgaacac 1080

ttccctaaca agactatcaa gttcgccct cactctggcg gcgacctgga aatcaccacc 1140

cacaccttca actgtcgggg cgagttcttc tactgcaata cctccaacct gttcaacgag 1200

agcaacatcg agcggaacga cagcatcatc acactgcctt gccggatcaa gcagattatc 1260

aatatgtggc aggaagtggg cagagccatc tacgcccctc caatcgccgg caacatcaca 1320

tgccggtcca atatcaccgg cctgctgctc accagagatg gcggctccaa caatggcgtg 1380

ccaaacgaca ccgagacatt cagacccggc ggaggcgaca tgcggaacaa ttggcggagc 1440

gagctgtaca agtacaaggt ggtggaagtg aagcccctgg gcgtggcccc taccgaggcc 1500

aagagaagag tggtcgaacg cgagaagcgg gccgtgggaa tcggagccgt gtttctggga 1560

atcctgggag ccgctggctc taccatgggc gctgcctcta tcaccctgac agtgcaggcc 1620

agacagctgc tcagcggcat cgtgcagcag cagagcaacc tgctgagagc cattgaggcc 1680

cagcagcaca tgctgcagct gaccgtgtgg ggcattaagc agctccagac acgggtgctg 1740

gccatcgaga gatacctgca ggatcagcag ctcctgggcc tgtggggctg tagcggcaag 1800

ctgatctgta ccaccgccgt gccctggaat acctcttgga gcaacaagag ccagaccgac 1860

atctgggaca acatgacctg gatgcagtgg gacaaagaaa tcggcaacta taccggcgag 1920

atctatagac tgctggaaga gtcccagaac cagcaggaaa agaacgagaa ggacctgctg 1980

gccctggatt cttggaacaa tctgtggaac tggttcagca tctccaagtg gctgtggtac 2040

atcaagatct tcatcatgat cgtgggcggc ctgatcggcc tgcggatcat ctttgccgtg 2100

ctgagcatcg tgaaccgcgt gcggcaggga tacagccctc tgagcctgca gaccctgact 2160

cagaaccctg gcggactgga cagactgggc cggattgagg aagaaggcgg cgagcaggac 2220

aaggatcgga gcatcaggct ggtcaacggc ttcttcgctc tgttttggga cgacctgcgg 2280

agcctgtgcc tgttcagcta ccacagactg cgggacttta tcctgattgt ggccagagcc 2340

gtcgaactgc tggggagaag ctctctgaga ggcctgcagc ggggctggga gattctgaag 2400

tacctgggct ccctgctgca gtactggggc ctggaactga agaagtctgc catcaatctg 2460

ctcgacacaa tcgctattgc cgtggccgaa ggcaccgata gaatcatcga gctgatccag 2520

cggatctgcc gggccatctg caacatcccc agacggatca gacagggctt cgaggccgct 2580

ctgcag 2586

<210> 26
<211> 2133
<212> DNA
<213> Artificial Sequence

<220>
<223> nucleotide sequence encoding C4D7

<400> 26
atgagagtgc ggggcatgct gagaaactgg cagcagtggt ggatctggtc cagcctgggc 60

ttctggatgc tgatgatcta cagcgtgatg ggcaacctgt gggtcaccgt gtactacggc 120

gtgcccgtgt ggaaggacgc caagaccacc ctgttttgcg cctccgatgc caaggcctac 180

gagaaagagg tgcacaacgt ctgggccacc cacgcctgtg tgcccaccga ccccaatccc 240

caggaaatcg tcctgggcaa cgtgaccgag aacttcaaca tgtggaagaa cgacatggtc 300

gatcagatgc acgaggacat catctccctg tgggacgcct ccctggaacc ctgcgtgaag 360

ctgaccccctc tgtgcgtgac cctgaactgc cggaacgtgc gcaacgtgtc cagcaacggc 420

acctacaaca tcatccacaa cgagacatac aaagagatga agaactgcag cttcaacgct 480

accaccgtgg tcgaggaccg gaagcagaag gtgcacgccc tgttctaccg gctggacatc 540

gtgcccctgg acgagaacaa cagcagcgag aagtcctccg agaacagctc cgagtactac 600

agactgatca actgcaacac cagcgccatc acccaggcct gccccaaggt gtccttcgac 660

cctatcccca tccactactg cgcccctgcc ggctacgcca tcctgaagtg caacaacaag 720

accttcaatg gcaccggccc ctgcaacaat gtgtccaccg tgcagtgcac ccacggcatc 780

aagcccgtgg tgtctaccca gctgctgctg aacggcagcc tggccgagga agagatcatt 840

atcagaagcg agaacctgac caacaacgcc aaaaccatca tcgtccacct gaacgaaacc 900

gtgaacatca cctgtacccg gcctaacaac aacacccgga agtccatccg gatcggccct 960

ggccagacct tttacgccac cggcgatatt atcggcgaca tccggcaggc ccactgcaat 1020

ctgagccggg acggctggaa caagacactg cagggcgtca agaagaagct ggccgaacac 1080

ttccctaaca agactatcaa gttcgccccc cactctggcg gcgacctgga aatcaccacc 1140

cacaccttca actgtcgggg cgagttcttc tactgcaata cctccaacct gttcaacgag 1200

agcaacatcg agcggaacga cagcatcatc acactgcctt gccggatcaa gcagattatc 1260

aatatgtggc aggaagtggg cagagccatc tacgcccctc caatcgccgg caacatcaca 1320

tgccggtcca atatcaccgg cctgctgctc accagagatg gcggctccaa caatggcgtg 1380

ccaaacgaca ccgagacatt cagacccggc ggaggcgaca tgcggaacaa ttggcggagc 1440

gagctgtaca agtacaaggt ggtggaagtg aagcccctgg gcgtggcccc taccgaggcc 1500

aagagaagag tggtcgaacg cgagaagcgg gccgtgggaa tcggagccgt gtttctggga 1560

atcctgggag ccgctggctc taccatgggc gctgcctcta tcaccctgac agtgcaggcc 1620

agacagctgc tcagcggcat cgtgcagcag cagagcaacc tgctgagagc cattgaggcc 1680

cagcagcaca tgctgcagct gaccgtgtgg ggcattaagc agctccagac acgggtgctg 1740

gccatcgaga gatacctgca ggatcagcag ctcctgggcc tgtggggctg tagcggcaag 1800

ctgatctgta ccaccgccgt gccctggaat acctcttgga gcaacaagag ccagaccgac 1860

atctgggaca acatgacctg gatgcagtgg gacaaagaaa tcggcaacta taccggcgag 1920

atctatagac tgctggaaga gtcccagaac cagcaggaaa agaacgagaa ggacctgctg 1980

gccctggatt cttggaacaa tctgtggaac tggttcagca tctccaagtg gctgtggtac 2040

atcaagatct tcatcatgat cgtgggcggc ctgatcggcc tgcggatcat ctttgccgtg 2100

ctgagcatcg tgaaccgcgt gcggcagggc tac 2133

<210> 27
<211> 2112
<212> DNA
<213> Artificial Sequence

<220>
<223> nucleotide sequence encoding sC4

<400> 27
atgagagtgc ggggcatgct gagaaactgg cagcagtggt ggatctggtc cagcctgggc 60

ttctggatgc tgatgatcta cagcgtgatg ggcaacctgt gggtcaccgt gtactacggc 120

gtgcccgtgt ggaaggacgc caagaccacc ctgttttgcg cctccgatgc caaggcctac 180

gagaaagagg tgcacaacgt ctgggccacc cacgcctgtg tgcccaccga ccccaatccc 240

caggaaatcg tcctgggcaa cgtgaccgag aacttcaaca tgtggaagaa cgacatggtc 300

gatcagatgc acgaggacat catctccctg tgggacgcct ccctggaacc ctgcgtgaag 360

ctgacccctc tgtgcgtgac cctgaactgc cggaacgtgc gcaacgtgtc cagcaacggc 420

acctacaaca tcatccacaa cgagacatac aaagagatga agaactgcag cttcaacgct 480

accaccgtgg tcgaggaccg gaagcagaag gtgcacgccc tgttctaccg gctggacatc 540

gtgcccctgg acgagaacaa cagcagcgag aagtcctccg agaacagctc cgagtactac 600

agactgatca actgcaacac cagcgccatc acccaggcct gccccaaggt gtccttcgac 660

cctatcccca tccactactg cgcccctgcc ggctacgcca tcctgaagtg caacaacaag 720

accttcaatg gcaccggccc ctgcaacaat gtgtccaccg tgcagtgcac ccacggcatc 780

aagcccgtgg tgtctaccca gctgctgctg aacggcagcc tggccgagga agagatcatt 840

atcagaagcg agaacctgac caacaacgcc aaaaccatca tcgtccacct gaacgaaacc 900

gtgaacatca cctgtacccg gcctaacaac aacacccgga agtccatccg gatcggccct 960

ggccagacct tttacgccac cggcgatatt atcggcgaca tccggcaggc ccactgcaat 1020

ctgagccggg acggctggaa caagacactg cagggcgtca agaagaagct ggccgaacac 1080

ttccctaaca agactatcaa gttcgccct cactctggcg gcgacctgga aatcaccacc 1140

cacaccttca actgtcgggg cgagttcttc tactgcaata cctccaacct gttcaacgag 1200

agcaacatcg agcggaacga cagcatcatc acactgcctt gccggatcaa gcagattatc 1260

aatatgtggc aggaagtggg cagagccatc tacgcccctc caatcgccgg caacatcaca 1320

tgccggtcca atatcaccgg cctgctgctc accagagatg gcggctccaa caatggcgtg 1380

ccaaacgaca ccgagacatt cagacccggc ggaggcgaca tgcggaacaa ttggcggagc 1440

gagctgtaca agtacaaggt ggtggaagtg aagcccctgg gcgtggcccc taccgaggcc 1500

aagagaagag tggtcgaacg cgaggaacgg gccgtgggaa tcggagccgt gtttctggga 1560

atcctgggag ccgctggctc taccatgggc gctgcctcta tcaccctgac agtgcaggcc 1620

agacagctgc tcagcggcat cgtgcagcag cagagcaacc tgctgagagc cattgaggcc 1680

cagcagcaca tgctgcagct gaccgtgtgg ggcattaagc agctccagac acgggtgctg 1740

gccatcgaga gatacctgca ggatcagcag ctcctgggcc tgtggggctg tagcggcaag 1800

ctgatctgta ccaccgccgt gccctggaat acctcttgga gcaacaagag ccagaccgac 1860

atctgggaca acatgacctg gatgcagtgg gacaaagaaa tcggcaacta taccggcgag 1920

atctatagac tgctggaaga gtcccagaac cagcaggaaa agatgaagca gatcgaggac 1980

aagatcgaag agattctgag caagatctac cacatcgaga acgagatcgc ccgcatcaag 2040

aaactgatcg gcgaagtggg atccggcgct cccacaaagg ccaaaagacg ggtggtgcag 2100

cgcgagaaac gc 2112

<210> 28
<211> 1350
<212> PRT
<213> Artificial Sequence

<220>

<223> mos1GagPol mosaic antigen sequence

<400> 28

```
Met Gly Ala Arg Ala Ser Val Leu Ser Gly Gly Glu Leu Asp Arg Trp
1               5                  10                  15

Glu Lys Ile Arg Leu Arg Pro Gly Gly Lys Lys Lys Tyr Arg Leu Lys
            20                  25                  30

His Ile Val Trp Ala Ser Arg Glu Leu Glu Arg Phe Ala Val Asn Pro
        35                  40                  45

Gly Leu Leu Glu Thr Ser Glu Gly Cys Arg Gln Ile Leu Gly Gln Leu
    50                  55                  60

Gln Pro Ser Leu Gln Thr Gly Ser Glu Glu Leu Arg Ser Leu Tyr Asn
65                  70                  75                  80

Thr Val Ala Thr Leu Tyr Cys Val His Gln Arg Ile Glu Ile Lys Asp
                85                  90                  95

Thr Lys Glu Ala Leu Glu Lys Ile Glu Glu Glu Gln Asn Lys Ser Lys
            100                 105                 110

Lys Lys Ala Gln Gln Ala Ala Ala Asp Thr Gly Asn Ser Ser Gln Val
        115                 120                 125

Ser Gln Asn Tyr Pro Ile Val Gln Asn Ile Gln Gly Gln Met Val His
    130                 135                 140

Gln Ala Ile Ser Pro Arg Thr Leu Asn Ala Trp Val Lys Val Val Glu
145                 150                 155                 160

Glu Lys Ala Phe Ser Pro Glu Val Ile Pro Met Phe Ser Ala Leu Ser
                165                 170                 175

Glu Gly Ala Thr Pro Gln Asp Leu Asn Thr Met Leu Asn Thr Val Gly
            180                 185                 190

Gly His Gln Ala Ala Met Gln Met Leu Lys Glu Thr Ile Asn Glu Glu
        195                 200                 205

Ala Ala Glu Trp Asp Arg Val His Pro Val His Ala Gly Pro Ile Ala
    210                 215                 220

Pro Gly Gln Met Arg Glu Pro Arg Gly Ser Asp Ile Ala Gly Thr Thr

225                 230                 235                 240

Ser Thr Leu Gln Glu Gln Ile Gly Trp Met Thr Asn Asn Pro Pro Ile
                245                 250                 255

Pro Val Gly Glu Ile Tyr Lys Arg Trp Ile Ile Leu Gly Leu Asn Lys
            260                 265                 270
```

```
Ile Val Arg Met Tyr Ser Pro Val Ser Ile Leu Asp Ile Arg Gln Gly
        275                 280                 285

Pro Lys Glu Pro Phe Arg Asp Tyr Val Asp Arg Phe Tyr Lys Thr Leu
    290                 295                 300

Arg Ala Glu Gln Ala Ser Gln Asp Val Lys Asn Trp Met Thr Glu Thr
305                 310                 315                 320

Leu Leu Val Gln Asn Ala Asn Pro Asp Cys Lys Thr Ile Leu Lys Ala
                325                 330                 335

Leu Gly Pro Ala Ala Thr Leu Glu Glu Met Met Thr Ala Cys Gln Gly
            340                 345                 350

Val Gly Gly Pro Gly His Lys Ala Arg Val Leu Ala Glu Ala Met Ser
        355                 360                 365

Gln Val Thr Asn Ser Ala Thr Ile Met Met Gln Arg Gly Asn Phe Arg
    370                 375                 380

Asn Gln Arg Lys Thr Val Lys Cys Phe Asn Cys Gly Lys Glu Gly His
385                 390                 395                 400

Ile Ala Lys Asn Cys Arg Ala Pro Arg Lys Lys Gly Cys Trp Lys Cys
                405                 410                 415

Gly Lys Glu Gly His Gln Met Lys Asp Cys Thr Glu Arg Gln Ala Asn
            420                 425                 430

Phe Leu Gly Lys Ile Trp Pro Ser Asn Lys Gly Arg Pro Gly Asn Phe
        435                 440                 445

Leu Gln Asn Arg Pro Glu Pro Thr Ala Pro Pro Glu Glu Ser Phe Arg
    450                 455                 460

Phe Gly Glu Glu Thr Thr Thr Pro Ser Gln Lys Gln Glu Pro Ile Asp
465                 470                 475                 480

Lys Glu Met Tyr Pro Leu Ala Ser Leu Lys Ser Leu Phe Gly Asn Asp
                485                 490                 495

Pro Ser Ser Gln Met Ala Pro Ile Ser Pro Ile Glu Thr Val Pro Val
            500                 505                 510

Lys Leu Lys Pro Gly Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu
        515                 520                 525

Thr Glu Glu Lys Ile Lys Ala Leu Thr Ala Ile Cys Glu Glu Met Glu
    530                 535                 540

Lys Glu Gly Lys Ile Thr Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr
545                 550                 555                 560

Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu
                565                 570                 575
```

565 570 575

Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val
580 585 590

Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val
595 600 605

Thr Val Leu Ala Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu
610 615 620

Gly Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu
625 630 635 640

Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys
645 650 655

Gly Ser Pro Ala Ile Phe Gln Cys Ser Met Thr Arg Ile Leu Glu Pro
660 665 670

Phe Arg Ala Lys Asn Pro Glu Ile Val Ile Tyr Gln Tyr Met Ala Ala
675 680 685

Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys Ile
690 695 700

Glu Glu Leu Arg Glu His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp
705 710 715 720

Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu
725 730 735

His Pro Asp Lys Trp Thr Val Gln Pro Ile Gln Leu Pro Glu Lys Asp
740 745 750

Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp
755 760 765

Ala Ser Gln Ile Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys Leu
770 775 780

Leu Arg Gly Ala Lys Ala Leu Thr Asp Ile Val Pro Leu Thr Glu Glu
785 790 795 800

Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val
805 810 815

His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln
820 825 830

Lys Gln Gly His Asp Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe
835 840 845

Lys Asn Leu Lys Thr Gly Lys Tyr Ala Lys Met Arg Thr Ala His Thr
850 855 860

Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln Lys Ile Ala Met Glu
865 870 875 880

```
Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe Arg Leu Pro Ile Gln
                885                 890                 895

Lys Glu Thr Trp Glu Thr Trp Trp Thr Asp Tyr Trp Gln Ala Thr Trp
            900                 905                 910

Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp
        915                 920                 925

Tyr Gln Leu Glu Lys Asp Pro Ile Ala Gly Val Glu Thr Phe Tyr Val
    930                 935                 940

Ala Gly Ala Ala Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val
945                 950                 955                 960

Thr Asp Arg Gly Arg Gln Lys Ile Val Ser Leu Thr Glu Thr Thr Asn
                965                 970                 975

Gln Lys Thr Ala Leu Gln Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly
            980                 985                 990

Ser Glu Val Asn Ile Val Thr Ala  Ser Gln Tyr Ala Leu  Gly Ile Ile
        995                 1000                1005

Gln Ala  Gln Pro Asp Lys Ser  Glu Ser Glu Leu Val  Asn Gln Ile
    1010                1015                1020

Ile Glu  Gln Leu Ile Lys Lys  Glu Arg Val Tyr Leu  Ser Trp Val
    1025                1030                1035

Pro Ala  His Lys Gly Ile Gly  Gly Asn Glu Gln Val  Asp Lys Leu
    1040                1045                1050

Val Ser  Ser Gly Ile Arg Lys  Val Leu Phe Leu Asp  Gly Ile Asp
    1055                1060                1065

Lys Ala  Gln Glu Glu His Glu  Lys Tyr His Ser Asn  Trp Arg Ala
    1070                1075                1080

Met Ala  Ser Asp Phe Asn Leu  Pro Pro Val Val Ala  Lys Glu Ile
    1085                1090                1095

Val Ala  Ser Cys Asp Gln Cys  Gln Leu Lys Gly Glu  Ala Met His
    1100                1105                1110

Gly Gln  Val Asp Cys Ser Pro  Gly Ile Trp Gln Leu  Ala Cys Thr
    1115                1120                1125

His Leu  Glu Gly Lys Ile Ile  Leu Val Ala Val His  Val Ala Ser
    1130                1135                1140

Gly Tyr  Ile Glu Ala Glu Val  Ile Pro Ala Glu Thr  Gly Gln Glu
    1145                1150                1155

Thr Ala  Tyr Phe Ile Leu Lys  Leu Ala Gly Arg Trp  Pro Val Lys
    1160                1165                1170
```

```
Val Ile  His Thr Ala Asn Gly  Ser Asn Phe Thr Ser  Ala Ala Val
    1175                 1180                 1185

Lys Ala  Ala Cys Trp Trp Ala  Gly Ile Gln Gln Glu  Phe Gly Ile
    1190                 1195                 1200

Pro Tyr  Asn Pro Gln Ser Gln  Gly Val Val Ala Ser  Met Asn Lys
    1205                 1210                 1215

Glu Leu  Lys Lys Ile Ile Gly  Gln Val Arg Asp Gln  Ala Glu His

    1220                 1225                 1230

Leu Lys  Thr Ala Val Gln Met  Ala Val Phe Ile His  Asn Phe Lys
    1235                 1240                 1245

Arg Lys  Gly Gly Ile Gly Gly  Tyr Ser Ala Gly Glu  Arg Ile Ile
    1250                 1255                 1260

Asp Ile  Ile Ala Thr Asp Ile  Gln Thr Lys Glu Leu  Gln Lys Gln
    1265                 1270                 1275

Ile Ile  Lys Ile Gln Asn Phe  Arg Val Tyr Tyr Arg  Asp Ser Arg
    1280                 1285                 1290

Asp Pro  Ile Trp Lys Gly Pro  Ala Lys Leu Leu Trp  Lys Gly Glu
    1295                 1300                 1305

Gly Ala  Val Val Ile Gln Asp  Asn Ser Asp Ile Lys  Val Val Pro
    1310                 1315                 1320

Arg Arg  Lys Val Lys Ile Ile  Lys Asp Tyr Gly Lys  Gln Met Ala
    1325                 1330                 1335

Gly Ala  Asp Cys Val Ala Gly  Arg Gln Asp Glu Asp
    1340                 1345                 1350
```

<210> 29

<211> 1341

<212> PRT

<213> Artificial Sequence

<220>

<223> mos2GagPol mosaic antigen sequence

<400> 29

```
Met Gly Ala Arg Ala Ser Ile Leu Arg Gly Gly Lys Leu Asp Lys Trp
1               5                  10                  15

Glu Lys Ile Arg Leu Arg Pro Gly Gly Lys Lys His Tyr Met Leu Lys
            20                  25                  30

His Leu Val Trp Ala Ser Arg Glu Leu Glu Arg Phe Ala Leu Asn Pro
```

```
        35                  40                  45

Gly Leu Leu Glu Thr Ser Glu Gly Cys Lys Gln Ile Ile Lys Gln Leu
    50                  55                  60

Gln Pro Ala Leu Gln Thr Gly Thr Glu Glu Leu Arg Ser Leu Phe Asn
65                  70                  75                  80

Thr Val Ala Thr Leu Tyr Cys Val His Ala Glu Ile Glu Val Arg Asp
                85                  90                  95

Thr Lys Glu Ala Leu Asp Lys Ile Glu Glu Glu Gln Asn Lys Ser Gln
            100                 105                 110

Gln Lys Thr Gln Gln Ala Lys Glu Ala Asp Gly Lys Val Ser Gln Asn
        115                 120                 125

Tyr Pro Ile Val Gln Asn Leu Gln Gly Gln Met Val His Gln Pro Ile
    130                 135                 140

Ser Pro Arg Thr Leu Asn Ala Trp Val Lys Val Ile Glu Glu Lys Ala
145                 150                 155                 160

Phe Ser Pro Glu Val Ile Pro Met Phe Thr Ala Leu Ser Glu Gly Ala
                165                 170                 175

Thr Pro Gln Asp Leu Asn Thr Met Leu Asn Thr Val Gly Gly His Gln
            180                 185                 190

Ala Ala Met Gln Met Leu Lys Asp Thr Ile Asn Glu Glu Ala Ala Glu
        195                 200                 205

Trp Asp Arg Leu His Pro Val His Ala Gly Pro Val Ala Pro Gly Gln
    210                 215                 220

Met Arg Glu Pro Arg Gly Ser Asp Ile Ala Gly Thr Thr Ser Asn Leu
225                 230                 235                 240

Gln Glu Gln Ile Ala Trp Met Thr Ser Asn Pro Pro Ile Pro Val Gly
                245                 250                 255

Asp Ile Tyr Lys Arg Trp Ile Ile Leu Gly Leu Asn Lys Ile Val Arg
            260                 265                 270

Met Tyr Ser Pro Thr Ser Ile Leu Asp Ile Lys Gln Gly Pro Lys Glu
        275                 280                 285

Pro Phe Arg Asp Tyr Val Asp Arg Phe Phe Lys Thr Leu Arg Ala Glu
    290                 295                 300

Gln Ala Thr Gln Asp Val Lys Asn Trp Met Thr Asp Thr Leu Leu Val
305                 310                 315                 320

Gln Asn Ala Asn Pro Asp Cys Lys Thr Ile Leu Arg Ala Leu Gly Pro

                325                 330                 335
```

```
Gly Ala Thr Leu Glu Glu Met Met Thr Ala Cys Gln Gly Val Gly Gly
            340                 345                 350

Pro Ser His Lys Ala Arg Val Leu Ala Glu Ala Met Ser Gln Thr Asn
        355                 360                 365

Ser Thr Ile Leu Met Gln Arg Ser Asn Phe Lys Gly Ser Lys Arg Ile
    370                 375                 380

Val Lys Cys Phe Asn Cys Gly Lys Glu Gly His Ile Ala Arg Asn Cys
385                 390                 395                 400

Arg Ala Pro Arg Lys Lys Gly Cys Trp Lys Cys Gly Lys Glu Gly His
                405                 410                 415

Gln Met Lys Asp Cys Thr Glu Arg Gln Ala Asn Phe Leu Gly Lys Ile
            420                 425                 430

Trp Pro Ser His Lys Gly Arg Pro Gly Asn Phe Leu Gln Ser Arg Pro
        435                 440                 445

Glu Pro Thr Ala Pro Pro Ala Glu Ser Phe Arg Phe Glu Glu Thr Thr
    450                 455                 460

Pro Ala Pro Lys Gln Glu Pro Lys Asp Arg Glu Pro Leu Thr Ser Leu
465                 470                 475                 480

Arg Ser Leu Phe Gly Ser Asp Pro Leu Ser Gln Met Ala Pro Ile Ser
                485                 490                 495

Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro
            500                 505                 510

Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys Ala Leu Val
        515                 520                 525

Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly
    530                 535                 540

Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe Ala Ile Lys Lys Lys Asp
545                 550                 555                 560

Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu Asn Lys Arg
                565                 570                 575

Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly
            580                 585                 590

Leu Lys Lys Lys Lys Ser Val Thr Val Leu Ala Val Gly Asp Ala Tyr
        595                 600                 605

Phe Ser Val Pro Leu Asp Glu Asp Phe Arg Lys Tyr Thr Ala Phe Thr
    610                 615                 620

Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn
625                 630                 635                 640
```

```
Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser
                645                 650                 655

Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro Asp Ile Val
            660                 665                 670

Ile Tyr Gln Tyr Met Ala Ala Leu Tyr Val Gly Ser Asp Leu Glu Ile
        675                 680                 685

Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg Gln His Leu Leu Arg
    690                 695                 700

Trp Gly Phe Thr Thr Pro Asp Lys Lys His Gln Lys Glu Pro Pro Phe
705                 710                 715                 720

Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro
                725                 730                 735

Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp Ile Gln Lys
            740                 745                 750

Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Ala Gly Ile Lys
        755                 760                 765

Val Lys Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu
    770                 775                 780

Val Val Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg
785                 790                 795                 800

Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys
                805                 810                 815

Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln Gly Gln Trp Thr Tyr
            820                 825                 830

Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala
        835                 840                 845

Arg Met Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu Thr Glu Ala
    850                 855                 860

Val Gln Lys Ile Ala Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro
865                 870                 875                 880

Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Ala Trp Trp Thr
                885                 890                 895

Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe Val Asn Thr
            900                 905                 910

Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val
        915                 920                 925

Gly Ala Glu Thr Phe Tyr Val Ala Gly Ala Ala Asn Arg Glu Thr Lys
    930                 935                 940
```

```
Leu Gly Lys Ala Gly Tyr Val Thr Asp Arg Gly Arg Gln Lys Val Val
945                 950                 955                 960

Ser Leu Thr Asp Thr Thr Asn Gln Lys Thr Ala Leu Gln Ala Ile His
                965                 970                 975

Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn Ile Val Thr Ala Ser
            980                 985                 990

Gln Tyr Ala Leu Gly Ile Ile Gln  Ala Gln Pro Asp Lys  Ser Glu Ser
        995                 1000                1005

Glu Leu  Val Ser Gln Ile Ile  Glu Gln Leu Ile Lys  Lys Glu Lys
    1010                1015                1020

Val Tyr  Leu Ala Trp Val Pro  Ala His Lys Gly Ile  Gly Gly Asn
    1025                1030                1035

Glu Gln  Val Asp Lys Leu Val  Ser Arg Gly Ile Arg  Lys Val Leu
    1040                1045                1050

Phe Leu  Asp Gly Ile Asp Lys  Ala Gln Glu Glu His  Glu Lys Tyr
    1055                1060                1065

His Ser  Asn Trp Arg Ala Met  Ala Ser Glu Phe Asn  Leu Pro Pro
    1070                1075                1080

Ile Val  Ala Lys Glu Ile Val  Ala Ser Cys Asp Lys  Cys Gln Leu
    1085                1090                1095

Lys Gly  Glu Ala Ile His Gly  Gln Val Asp Cys Ser  Pro Gly Ile
    1100                1105                1110

Trp Gln  Leu Ala Cys Thr His  Leu Glu Gly Lys Val  Ile Leu Val
    1115                1120                1125

Ala Val  His Val Ala Ser Gly  Tyr Ile Glu Ala Glu  Val Ile Pro
    1130                1135                1140

Ala Glu  Thr Gly Gln Glu Thr  Ala Tyr Phe Leu Leu  Lys Leu Ala
    1145                1150                1155

Gly Arg  Trp Pro Val Lys Thr  Ile His Thr Ala Asn  Gly Ser Asn
    1160                1165                1170

Phe Thr  Ser Ala Thr Val Lys  Ala Ala Cys Trp Trp  Ala Gly Ile
    1175                1180                1185

Lys Gln  Glu Phe Gly Ile Pro  Tyr Asn Pro Gln Ser  Gln Gly Val
    1190                1195                1200

Val Ala  Ser Ile Asn Lys Glu  Leu Lys Lys Ile Ile  Gly Gln Val
    1205                1210                1215

Arg Asp  Gln Ala Glu His Leu  Lys Thr Ala Val Gln  Met Ala Val
    1220                1225                1230
```

```
Phe Ile  His Asn Phe Lys Arg  Lys Gly Gly Ile Gly  Glu Tyr Ser
    1235                 1240                 1245

Ala Gly  Glu Arg Ile Val Asp  Ile Ile Ala Ser Asp  Ile Gln Thr
    1250                 1255                 1260

Lys Glu  Leu Gln Lys Gln Ile  Thr Lys Ile Gln Asn  Phe Arg Val
    1265                 1270                 1275

Tyr Tyr  Arg Asp Ser Arg Asp  Pro Leu Trp Lys Gly  Pro Ala Lys
    1280                 1285                 1290

Leu Leu  Trp Lys Gly Glu Gly  Ala Val Val Ile Gln  Asp Asn Ser
    1295                 1300                 1305

Asp Ile  Lys Val Val Pro Arg  Arg Lys Ala Lys Ile  Ile Arg Asp

    1310                 1315                 1320

Tyr Gly  Lys Gln Met Ala Gly  Asp Asp Cys Val Ala  Ser Arg Gln
    1325                 1330                 1335

Asp Glu  Asp
    1340
```

<210> 30

<211> 709

<212> PRT

<213> Artificial Sequence

<220>

<223> sC1 sequence

<400> 30

```
Met Arg Val Arg Gly Ile Gln Arg Asn Trp Pro Gln Trp Trp Ile Trp
1               5                  10                  15

Gly Ile Leu Gly Phe Trp Met Ile Ile Ile Cys Arg Val Met Gly Asn
            20                  25                  30

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

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

His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
65                  70                  75                  80

Gln Glu Met Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp Lys
                85                  90                  95

Asn Asp Met Val Asp Gln Met His Glu Asp Ile Ile Arg Leu Trp Asp
            100                 105                 110
```

```
            100                 105                 110

Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
        115                 120                 125

Glu Cys Arg Asn Val Arg Asn Val Ser Ser Asn Gly Thr Tyr Asn Ile
    130                 135                 140

Ile His Asn Glu Thr Tyr Lys Glu Met Lys Asn Cys Ser Phe Asn Ala
145                 150                 155                 160

Thr Thr Val Val Glu Asp Arg Lys Gln Lys Val His Ala Leu Phe Tyr
                165                 170                 175

Arg Leu Asp Ile Val Pro Leu Asp Glu Asn Asn Ser Ser Glu Lys Ser
            180                 185                 190

Ser Glu Asn Ser Ser Glu Tyr Tyr Arg Leu Ile Asn Cys Asn Thr Ser
        195                 200                 205

Ala Ile Thr Gln Ala Cys Pro Lys Val Ser Phe Asp Pro Ile Pro Ile
    210                 215                 220

His Tyr Cys Ala Pro Ala Gly Tyr Ala Ile Leu Lys Cys Asn Asn Lys
225                 230                 235                 240

Thr Phe Asn Gly Thr Gly Pro Cys Asn Asn Val Ser Thr Val Gln Cys
                245                 250                 255

Thr His Gly Ile Lys Pro Val Val Ser Thr Gln Leu Leu Leu Asn Gly
            260                 265                 270

Ser Leu Ala Glu Glu Glu Ile Ile Ile Arg Ser Glu Asn Leu Thr Asn
        275                 280                 285

Asn Ala Lys Thr Ile Ile Val His Leu Asn Glu Thr Val Asn Ile Thr
    290                 295                 300

Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser Ile Arg Ile Gly Pro
305                 310                 315                 320

Gly Gln Thr Phe Tyr Ala Thr Gly Asp Ile Ile Gly Asp Ile Arg Gln
                325                 330                 335

Ala His Cys Asn Leu Ser Arg Asp Gly Trp Asn Lys Thr Leu Gln Gly
            340                 345                 350

Val Lys Lys Lys Leu Ala Glu His Phe Pro Asn Lys Thr Ile Asn Phe
        355                 360                 365

Thr Ser Ser Ser Gly Gly Asp Leu Glu Ile Thr Thr His Ser Phe Asn
    370                 375                 380

Cys Arg Gly Glu Phe Phe Tyr Cys Asn Thr Ser Gly Leu Phe Asn Gly
385                 390                 395                 400

Thr Tyr Met Pro Asn Glu Thr Asn Ser Asn Ser Ser Ser Asn Ile Thr
```

```
                405                 410                 415

Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly

            420                 425                 430

Arg Ala Met Tyr Ala Pro Pro Ile Ala Gly Asn Ile Thr Cys Arg Ser
        435                 440                 445

Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly Ser Asn Asn Gly
    450                 455                 460

Val Pro Asn Asp Thr Glu Thr Phe Arg Pro Gly Gly Gly Asp Met Arg
465                 470                 475                 480

Asn Asn Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Glu Val Lys
                485                 490                 495

Pro Leu Gly Val Ala Pro Thr Glu Ala Lys Arg Arg Val Val Glu Arg
            500                 505                 510

Glu Glu Arg Ala Val Gly Ile Gly Ala Val Phe Leu Gly Ile Leu Gly
        515                 520                 525

Ala Ala Gly Ser Thr Met Gly Ala Ala Ser Ile Thr Leu Thr Val Gln
    530                 535                 540

Ala Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Ser Asn Leu Leu
545                 550                 555                 560

Arg Ala Ile Glu Ala Gln Gln His Met Leu Gln Leu Thr Val Trp Gly
                565                 570                 575

Ile Lys Gln Leu Gln Thr Arg Val Leu Ala Ile Glu Arg Tyr Leu Gln
            580                 585                 590

Asp Gln Gln Leu Leu Gly Leu Trp Gly Cys Ser Gly Lys Leu Ile Cys
        595                 600                 605

Thr Thr Ala Val Pro Trp Asn Thr Ser Trp Ser Asn Lys Ser Gln Thr
    610                 615                 620

Asp Ile Trp Asp Asn Met Thr Trp Met Gln Trp Asp Lys Glu Ile Gly
625                 630                 635                 640

Asn Tyr Thr Gly Glu Ile Tyr Arg Leu Leu Glu Glu Ser Gln Asn Gln
                645                 650                 655

Gln Glu Lys Met Lys Gln Ile Glu Asp Lys Ile Glu Glu Ile Leu Ser
            660                 665                 670

Lys Ile Tyr His Ile Glu Asn Glu Ile Ala Arg Ile Lys Lys Leu Ile
        675                 680                 685

Gly Glu Val Gly Ser Gly Ala Pro Thr Lys Ala Lys Arg Arg Val Val
    690                 695                 700
```

```
Gln Arg Glu Lys Arg
705
```

<210> 31

<211> 867

<212> PRT

<213> Artificial Sequence

<220>

<223> C1 sequence

<400> 31

```
Met Arg Val Arg Gly Ile Gln Arg Asn Trp Pro Gln Trp Trp Ile Trp
1               5                  10                  15

Gly Ile Leu Gly Phe Trp Met Ile Ile Ile Cys Arg Val Met Gly Asn
            20                  25                  30

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

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

His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
65                  70                  75                  80

Gln Glu Met Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp Lys
                85                  90                  95

Asn Asp Met Val Asp Gln Met His Glu Asp Ile Ile Arg Leu Trp Asp
            100                 105                 110

Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
        115                 120                 125

Glu Cys Arg Asn Val Arg Asn Val Ser Ser Asn Gly Thr Tyr Asn Ile
    130                 135                 140

Ile His Asn Glu Thr Tyr Lys Glu Met Lys Asn Cys Ser Phe Asn Ala
145                 150                 155                 160

Thr Thr Val Val Glu Asp Arg Lys Gln Lys Val His Ala Leu Phe Tyr

                165                 170                 175

Arg Leu Asp Ile Val Pro Leu Asp Glu Asn Asn Ser Ser Glu Lys Ser
            180                 185                 190

Ser Glu Asn Ser Ser Glu Tyr Tyr Arg Leu Ile Asn Cys Asn Thr Ser
        195                 200                 205

Ala Ile Thr Gln Ala Cys Pro Lys Val Ser Phe Asp Pro Ile Pro Ile
```

```
    210                 215                 220

His Tyr Cys Ala Pro Ala Gly Tyr Ala Ile Leu Lys Cys Asn Asn Lys
225                 230                 235                 240

Thr Phe Asn Gly Thr Gly Pro Cys Asn Asn Val Ser Thr Val Gln Cys
                245                 250                 255

Thr His Gly Ile Lys Pro Val Val Ser Thr Gln Leu Leu Leu Asn Gly
            260                 265                 270

Ser Leu Ala Glu Glu Glu Ile Ile Ile Arg Ser Glu Asn Leu Thr Asn
        275                 280                 285

Asn Ala Lys Thr Ile Ile Val His Leu Asn Glu Thr Val Asn Ile Thr
    290                 295                 300

Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser Ile Arg Ile Gly Pro
305                 310                 315                 320

Gly Gln Thr Phe Tyr Ala Thr Gly Asp Ile Ile Gly Asp Ile Arg Gln
                325                 330                 335

Ala His Cys Asn Leu Ser Arg Asp Gly Trp Asn Lys Thr Leu Gln Gly
            340                 345                 350

Val Lys Lys Lys Leu Ala Glu His Phe Pro Asn Lys Thr Ile Asn Phe
        355                 360                 365

Thr Ser Ser Ser Gly Gly Asp Leu Glu Ile Thr Thr His Ser Phe Asn
    370                 375                 380

Cys Arg Gly Glu Phe Phe Tyr Cys Asn Thr Ser Gly Leu Phe Asn Gly
385                 390                 395                 400

Thr Tyr Met Pro Asn Gly Thr Asn Ser Asn Ser Ser Ser Asn Ile Thr
                405                 410                 415

Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly
            420                 425                 430

Arg Ala Met Tyr Ala Pro Pro Ile Ala Gly Asn Ile Thr Cys Arg Ser
        435                 440                 445

Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly Ser Asn Asn Gly
    450                 455                 460

Val Pro Asn Asp Thr Glu Thr Phe Arg Pro Gly Gly Gly Asp Met Arg
465                 470                 475                 480

Asn Asn Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Glu Val Lys
                485                 490                 495

Pro Leu Gly Val Ala Pro Thr Glu Ala Lys Arg Arg Val Val Glu Arg
            500                 505                 510

Glu Lys Arg Ala Val Gly Ile Gly Ala Val Phe Leu Gly Ile Leu Gly
```

```
                515                 520                 525

Ala Ala Gly Ser Thr Met Gly Ala Ala Ser Ile Thr Leu Thr Val Gln
    530                 535                 540

Ala Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Ser Asn Leu Leu
545                 550                 555                 560

Arg Ala Ile Glu Ala Gln Gln His Met Leu Gln Leu Thr Val Trp Gly
                565                 570                 575

Ile Lys Gln Leu Gln Thr Arg Val Leu Ala Ile Glu Arg Tyr Leu Gln
            580                 585                 590

Asp Gln Gln Leu Leu Gly Leu Trp Gly Cys Ser Gly Lys Leu Ile Cys
        595                 600                 605

Thr Thr Ala Val Pro Trp Asn Thr Ser Trp Ser Asn Lys Ser Gln Thr
    610                 615                 620

Asp Ile Trp Asp Asn Met Thr Trp Met Gln Trp Asp Lys Glu Ile Gly
625                 630                 635                 640

Asn Tyr Thr Gly Glu Ile Tyr Arg Leu Leu Glu Glu Ser Gln Asn Gln
                645                 650                 655

Gln Glu Lys Asn Glu Lys Asp Leu Leu Ala Leu Asp Ser Trp Lys Asn
            660                 665                 670

Leu Trp Asn Trp Phe Asp Ile Thr Asn Trp Leu Trp Tyr Ile Lys Ile
        675                 680                 685

Phe Ile Met Ile Val Gly Gly Leu Ile Gly Leu Arg Ile Ile Phe Ala
    690                 695                 700

Val Leu Ser Ile Val Asn Arg Val Arg Gln Gly Tyr Ser Pro Leu Ser
705                 710                 715                 720

Leu Gln Thr Leu Thr Gln Asn Pro Gly Gly Leu Asp Arg Leu Gly Arg
                725                 730                 735

Ile Glu Glu Glu Gly Gly Glu Gln Asp Lys Asp Arg Ser Ile Arg Leu
            740                 745                 750

Val Asn Gly Phe Phe Ala Leu Phe Trp Asp Asp Leu Arg Ser Leu Cys
        755                 760                 765

Leu Phe Ser Tyr His Arg Leu Arg Asp Phe Ile Leu Ile Val Ala Arg
    770                 775                 780

Ala Val Glu Leu Leu Gly Arg Ser Ser Leu Arg Gly Leu Gln Arg Gly
785                 790                 795                 800

Trp Glu Ile Leu Lys Tyr Leu Gly Ser Leu Leu Gln Tyr Trp Gly Leu
                805                 810                 815
```

```
Glu Leu Lys Lys Ser Ala Ile Asn Leu Leu Asp Thr Ile Ala Ile Ala
            820                 825                 830

Val Ala Glu Gly Thr Asp Arg Ile Ile Glu Leu Ile Gln Arg Ile Cys
        835                 840                 845

Arg Ala Ile Cys Asn Ile Pro Arg Arg Ile Arg Gln Gly Phe Glu Ala
    850                 855                 860

Ala Leu Gln
865
```

<210> 32

<211> 716

<212> PRT

<213> Artificial Sequence

<220>

<223> C1D7 sequence

<400> 32

```
Met Arg Val Arg Gly Ile Gln Arg Asn Trp Pro Gln Trp Trp Ile Trp
1               5                   10                  15

Gly Ile Leu Gly Phe Trp Met Ile Ile Ile Cys Arg Val Met Gly Asn
            20                  25                  30

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

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

His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
65                  70                  75                  80

Gln Glu Met Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp Lys
                85                  90                  95

Asn Asp Met Val Asp Gln Met His Glu Asp Ile Ile Arg Leu Trp Asp
            100                 105                 110

Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
        115                 120                 125

Glu Cys Arg Asn Val Arg Asn Val Ser Ser Asn Gly Thr Tyr Asn Ile
    130                 135                 140

Ile His Asn Glu Thr Tyr Lys Glu Met Lys Asn Cys Ser Phe Asn Ala
145                 150                 155                 160

Thr Thr Val Val Glu Asp Arg Lys Gln Lys Val His Ala Leu Phe Tyr
                165                 170                 175

Arg Leu Asp Ile Val Pro Leu Asp Glu Asn Asn Ser Ser Glu Lys Ser
            180                 185                 190
```

```
Ser Glu Asn Ser Ser Glu Tyr Tyr Arg Leu Ile Asn Cys Asn Thr Ser
        195                 200                 205

Ala Ile Thr Gln Ala Cys Pro Lys Val Ser Phe Asp Pro Ile Pro Ile
    210                 215                 220

His Tyr Cys Ala Pro Ala Gly Tyr Ala Ile Leu Lys Cys Asn Asn Lys
225                 230                 235                 240

Thr Phe Asn Gly Thr Gly Pro Cys Asn Asn Val Ser Thr Val Gln Cys
                245                 250                 255

Thr His Gly Ile Lys Pro Val Val Ser Thr Gln Leu Leu Leu Asn Gly
            260                 265                 270

Ser Leu Ala Glu Glu Glu Ile Ile Ile Arg Ser Glu Asn Leu Thr Asn
        275                 280                 285

Asn Ala Lys Thr Ile Ile Val His Leu Asn Glu Thr Val Asn Ile Thr
    290                 295                 300

Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser Ile Arg Ile Gly Pro
305                 310                 315                 320

Gly Gln Thr Phe Tyr Ala Thr Gly Asp Ile Ile Gly Asp Ile Arg Gln
                325                 330                 335

Ala His Cys Asn Leu Ser Arg Asp Gly Trp Asn Lys Thr Leu Gln Gly
            340                 345                 350

Val Lys Lys Lys Leu Ala Glu His Phe Pro Asn Lys Thr Ile Asn Phe
        355                 360                 365

Thr Ser Ser Ser Gly Gly Asp Leu Glu Ile Thr Thr His Ser Phe Asn
    370                 375                 380

Cys Arg Gly Glu Phe Phe Tyr Cys Asn Thr Ser Gly Leu Phe Asn Gly
385                 390                 395                 400

Thr Tyr Met Pro Asn Gly Thr Asn Ser Asn Ser Ser Ser Asn Ile Thr
                405                 410                 415

Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly
            420                 425                 430

Arg Ala Met Tyr Ala Pro Pro Ile Ala Gly Asn Ile Thr Cys Arg Ser
        435                 440                 445

Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly Ser Asn Asn Gly
    450                 455                 460

Val Pro Asn Asp Thr Glu Thr Phe Arg Pro Gly Gly Gly Asp Met Arg
465                 470                 475                 480

Asn Asn Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Glu Val Lys
                485                 490                 495
```

```
Pro Leu Gly Val Ala Pro Thr Glu Ala Lys Arg Arg Val Val Glu Arg
            500                 505                 510

Glu Lys Arg Ala Val Gly Ile Gly Ala Val Phe Leu Gly Ile Leu Gly
        515                 520                 525

Ala Ala Gly Ser Thr Met Gly Ala Ala Ser Ile Thr Leu Thr Val Gln
    530                 535                 540

Ala Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Ser Asn Leu Leu
545                 550                 555                 560

Arg Ala Ile Glu Ala Gln Gln His Met Leu Gln Leu Thr Val Trp Gly
                565                 570                 575

Ile Lys Gln Leu Gln Thr Arg Val Leu Ala Ile Glu Arg Tyr Leu Gln
            580                 585                 590

Asp Gln Gln Leu Leu Gly Leu Trp Gly Cys Ser Gly Lys Leu Ile Cys
        595                 600                 605

Thr Thr Ala Val Pro Trp Asn Thr Ser Trp Ser Asn Lys Ser Gln Thr
    610                 615                 620

Asp Ile Trp Asp Asn Met Thr Trp Met Gln Trp Asp Lys Glu Ile Gly
625                 630                 635                 640

Asn Tyr Thr Gly Glu Ile Tyr Arg Leu Leu Glu Glu Ser Gln Asn Gln
                645                 650                 655

Gln Glu Lys Asn Glu Lys Asp Leu Leu Ala Leu Asp Ser Trp Lys Asn
            660                 665                 670

Leu Trp Asn Trp Phe Asp Ile Thr Asn Trp Leu Trp Tyr Ile Lys Ile
        675                 680                 685

Phe Ile Met Ile Val Gly Gly Leu Ile Gly Leu Arg Ile Ile Phe Ala
    690                 695                 700

Val Leu Ser Ile Val Asn Arg Val Arg Gln Gly Tyr
705                 710                 715
```

<210> 33

<211> 2127

<212> DNA

<213> Artificial Sequence

<220>

<223> nucleotide sequence encoding sC1

<400> 33

```
atgagagtgc ggggcattca gagaaactgg ccccagtggt ggatctgggg catcctgggc       60

ttttggatga tcattatctg ccgcgtgatg ggcaacctgt gggtcaccgt gtactacggc      120
```

gtgcccgtgt ggaaagaggc caagaccacc ctgttctgcg ccagcgacgc caaggcctac 180

gagaaagagg tgcacaacgt ctgggccacc cacgcctgtg tgcccaccga ccccaatccc 240

caggaaatgg tcctggaaaa cgtgaccgag aacttcaaca tgtggaagaa cgacatggtg 300

gaccagatgc acgaggacat catccggctg tgggaccaga gcctgaagcc ctgcgtgaag 360

ctgacccctc tgtgcgtgac cctggaatgc cggaacgtgc gcaacgtgtc cagcaacggc 420

acctacaata tcatccacaa cgagacatac aaagagatga agaactgcag cttcaacgct 480

accaccgtgg tcgaggaccg gaagcagaag gtgcacgccc tgttttaccg gctggacatc 540

gtgcccctgg acgagaacaa cagcagcgag aagtcctccg agaacagctc cgagtactac 600

agactgatca actgcaacac cagcgccatc acccaggcct gccccaaggt gtccttcgac 660

cctatcccca tccactactg cgccccctgcc ggctacgcca tcctgaagtg caacaacaag 720

accttcaatg gcaccggccc ctgcaacaat gtgtccaccg tgcagtgcac ccacggcatc 780

aagcccgtgg tgtctaccca gctgctgctg aacggcagcc tggccgagga agagatcatc 840

atcagaagcg agaacctgac caacaacgcc aagacaatca tcgtccacct gaacgaaacc 900

gtgaacatca cctgtacccg gcctaacaac aacacccgga agtccatccg gatcggccct 960

ggccagacct tttacgccac cggcgatatt atcggcgaca tccggcaggc ccactgcaat 1020

ctgagccggg acggctggaa caagacactg cagggcgtca agaagaagct ggccgaacac 1080

ttccccaaca aaaccatcaa cttcaccagc tcctctggcg gcgacctgga aatcaccacc 1140

cacagcttta actgcagagg cgagttcttc tactgcaata cctccggcct gttcaatgga 1200

acctacatgc ccaacgggac caacagcaac tccagcagca atatcaccct gccttgccgg 1260

atcaagcaga ttatcaatat gtggcaggaa gtgggcagag ctatgtacgc ccctccaatc 1320

gccggcaaca tcacatgcag aagcaacatt accggcctgc tgctcaccag ggacggcggc 1380

tctaacaatg gcgtgccaaa cgacaccgag acattcagac ccggcggagg cgacatgcgg 1440

aacaattggc ggagcgagct gtacaagtac aaggtggtgg aagtgaagcc cctgggcgtg 1500

gcccctaccg aagccaagag aagagtggtc gaacgcgagg aacgggccgt gggcattgga 1560

gccgtgtttc tgggaatcct gggagccgct ggcagcacca tgggcgctgc ctctatcaca 1620

ctgacagtgc aggccagaca gctcctgagc ggcatcgtgc agcagcagag caacctgctg 1680

agagccatcg aggcacagca gcacatgctg cagctgaccg tgtggggcat taagcagctc 1740

cagacacggg tgctggccat tgagagatac ctgcaggatc agcagctgct cggcctgtgg 1800

ggctgtagcg gcaagctgat ctgtaccacc gccgtgcctt ggaacacctc ctggtccaac 1860

aagagccaga ccgacatctg ggacaacatg acctggatgc agtgggacaa agaaatcggc 1920

aactataccg gcgagatcta ccgactgctg gaagagtccc agaaccagca ggaaaagatg 1980

aagcagatcg aggacaagat cgaagagatt ctgagcaaaa tctaccacat cgagaacgag 2040

atcgcccgca tcaagaaact gatcggcgaa gtgggatccg gcgctcccac aaaggccaaa 2100

agacgggtgg tgcagcgcga gaaacgc 2127

<210> 34

<211> 2601

<212> DNA

<213> Artificial Sequence

<220>
<223> nucleotide sequence encoding C1

<400> 34

atgagagtgc ggggcattca gagaaactgg ccccagtggt ggatctgggg catcctgggc 60

ttttggatga tcattatctg ccgcgtgatg ggcaacctgt gggtcaccgt gtactacggc 120

gtgcccgtgt ggaaagaggc caagaccacc ctgttctgcg ccagcgacgc caaggcctac 180

gagaaagagg tgcacaacgt ctgggccacc cacgcctgtg tgcccaccga ccccaatccc 240

caggaaatgg tcctggaaaa cgtgaccgag aacttcaaca tgtggaagaa cgacatggtg 300

gaccagatgc acgaggacat catccggctg tgggaccaga gcctgaagcc ctgcgtgaag 360

ctgacccctc tgtgcgtgac cctggaatgc cggaacgtgc gcaacgtgtc cagcaacggc 420

acctacaata tcatccacaa cgagacatac aaagagatga agaactgcag cttcaacgct 480

accaccgtgg tcgaggaccg gaagcagaag gtgcacgccc tgttttaccg gctggacatc 540

gtgcccctgg acgagaacaa cagcagcgag aagtcctccg agaacagctc cgagtactac 600

agactgatca actgcaacac cagcgccatc acccaggcct gccccaaggt gtccttcgac 660

cctatcccca tccactactg cgcccctgcc ggctacgcca tcctgaagtg caacaacaag 720

accttcaatg gcaccggccc ctgcaacaat gtgtccaccg tgcagtgcac ccacggcatc 780

aagcccgtgg tgtctaccca gctgctgctg aacggcagcc tggccgagga agagatcatc 840

atcagaagcg agaacctgac caacaacgcc aagacaatca tcgtccacct gaacgaaacc 900

gtgaacatca cctgtacccg gcctaacaac aacacccgga agtccatccg gatcggccct 960

ggccagacct tttacgccac cggcgatatt atcggcgaca tccggcaggc ccactgcaat 1020

ctgagccggg acggctggaa caagacactg cagggcgtca agaagaagct ggccgaacac 1080

ttccccaaca aaaccatcaa cttcaccagc tcctctggcg gcgacctgga aatcaccacc 1140

cacagcttta actgcagagg cgagttcttc tactgcaata cctccggcct gttcaatgga 1200

acctacatgc ccaacgggac caacagcaac tccagcagca atatcaccct gccttgccgg 1260

atcaagcaga ttatcaatat gtggcaggaa gtgggcagag ctatgtacgc ccctccaatc 1320

gccggcaaca tcacatgcag aagcaacatt accggcctgc tgctcaccag ggacggcggc 1380

tctaacaatg gcgtgccaaa cgacaccgag acattcagac ccggcggagg cgacatgcgg 1440

aacaattggc ggagcgagct gtacaagtac aaggtggtgg aagtgaagcc cctgggcgtg 1500

gcccctaccg aagccaagag aagagtggtc gaacgcgaga agcgggccgt gggcattgga 1560

gccgtgtttc tgggaatcct gggagccgct ggcagcacca tgggcgctgc ctctatcaca 1620

ctgacagtgc aggccagaca gctcctgagc ggcatcgtgc agcagcagag caacctgctg 1680

agagccatcg aggcacagca gcacatgctg cagctgaccg tgtggggcat taagcagctc 1740

cagacacggg tgctggccat tgagagatac ctgcaggatc agcagctgct cggcctgtgg 1800

ggctgtagcg gcaagctgat ctgtaccacc gccgtgcctt ggaacacctc ctggtccaac 1860

aagagccaga ccgacatctg ggacaacatg acctggatgc agtgggacaa agaaatcggc 1920

aactataccg gcgagatcta ccgactgctg gaagagtccc agaaccagca ggaaaagaac 1980

gagaagaacc tgctggccct ggacagctgg aaaaatctgt ggaattggtt cgacatcagc 2040

gagaaggacc tgctggccct ggacagctgg aaaaatctgt ggaattggct cgacatcacc 2040

aactggctgt ggtacatcaa gatcttcatc atgatcgtgg gcggcctgat cggcctgcgg 2100

atcatctttg ccgtgctgag catcgtgaac cgcgtgcggc agggatacag ccctctgagc 2160

ctgcagaccc tgacccagaa tccaggcgga ctggatcggc tgggccggat tgaggaagaa 2220

ggcggcgagc aggacaagga ccgcagcatc agactcgtga acggcttctt cgctctgttt 2280

tgggacgacc tgcggagcct gtgcctgttc tcctaccaca gactgcggga ctttatcctg 2340

attgtggcca gagccgtcga gctgctgggc agatcttctc tgagaggcct gcagcggggc 2400

tgggagattc tgaagtacct gggctccctg ctgcagtatt ggggcctgga actgaagaag 2460

tccgccatca atctgctcga cacaatcgct attgccgtgg ccgaaggcac cgacagaatc 2520

atcgagctga tccagcggat ctgccgggcc atctgcaaca tccccagacg gatcagacag 2580

ggctttgaag ccgccctcca g 2601

<210> 35
<211> 2148
<212> DNA
<213> Artificial Sequence

<220>
<223> nucleotide sequence encoding C1D7

<400> 35
atgagagtgc ggggcattca gagaaactgg ccccagtggt ggatctgggg catcctgggc 60

ttttggatga tcattatctg ccgcgtgatg ggcaacctgt gggtcaccgt gtactacggc 120

gtgcccgtgt ggaaagaggc caagaccacc ctgttctgcg ccagcgacgc caaggcctac 180

gagaaagagg tgcacaacgt ctgggccacc cacgcctgtg tgcccaccga ccccaatccc 240

caggaaatgg tcctggaaaa cgtgaccgag aacttcaaca tgtggaagaa cgacatggtg 300

gaccagatgc acgaggacat catccggctg tgggaccaga gcctgaagcc ctgcgtgaag 360

ctgacccctc tgtgcgtgac cctggaatgc cggaacgtgc gcaacgtgtc cagcaacggc 420

acctacaata tcatccacaa cgagacatac aaagagatga agaactgcag cttcaacgct 480

accaccgtgg tcgaggaccg gaagcagaag gtgcacgccc tgttttaccg gctggacatc 540

gtgcccctgg acgagaacaa cagcagcgag aagtcctccg agaacagctc cgagtactac 600

agactgatca actgcaacac cagcgccatc acccaggcct gccccaaggt gtccttcgac 660

cctatcccca tccactactg cgcccctgcc ggctacgcca tcctgaagtg caacaacaag 720

accttcaatg gcaccggccc ctgcaacaat gtgtccaccg tgcagtgcac ccacggcatc 780

aagcccgtgg tgtctaccca gctgctgctg aacggcagcc tggccgagga agagatcatc 840

atcagaagcg agaacctgac caacaacgcc aagacaatca tcgtccacct gaacgaaacc 900

gtgaacatca cctgtacccg gcctaacaac aacacccgga agtccatccg gatcggccct 960

ggccagacct tttacgccac cggcgatatt atcggcgaca tccggcaggc ccactgcaat 1020

ctgagccggg acggctggaa caagacactg cagggcgtca agaagaagct ggccgaacac 1080

ttccccaaca aaaccatcaa cttcaccagc tcctctggcg gcgacctgga aatcaccacc 1140

```
cacagcttta actgcagagg cgagttcttc tactgcaata cctccggcct gttcaatgga     1200

acctacatgc ccaacgggac caacagcaac tccagcagca atatcaccct gccttgccgg     1260

atcaagcaga ttatcaatat gtggcaggaa gtgggcagag ctatgtacgc ccctccaatc     1320

gccggcaaca tcacatgcag aagcaacatt accggcctgc tgctcaccag ggacggcggc     1380

tctaacaatg gcgtgccaaa cgacaccgag acattcagac ccggcggagg cgacatgcgg     1440

aacaattggc ggagcgagct gtacaagtac aaggtggtgg aagtgaagcc cctgggcgtg     1500

gcccctaccg aagccaagag aagagtggtc gaacgcgaga agcgggccgt gggcattgga     1560

gccgtgtttc tgggaatcct gggagccgct ggcagcacca tgggcgctgc ctctatcaca     1620

ctgacagtgc aggccagaca gctcctgagc ggcatcgtgc agcagcagag caacctgctg     1680

agagccatcg aggcacagca gcacatgctg cagctgaccg tgtggggcat taagcagctc     1740

cagacacggg tgctggccat tgagagatac ctgcaggatc agcagctgct cggcctgtgg     1800

ggctgtagcg gcaagctgat ctgtaccacc gccgtgcctt ggaacacctc ctggtccaac     1860

aagagccaga ccgacatctg ggacaacatg acctggatgc agtgggacaa agaaatcggc     1920

aactataccg gcgagatcta ccgactgctg gaagagtccc agaaccagca ggaaaagaac     1980

gagaaggacc tgctggccct ggacagctgg aaaaatctgt ggaattggtt cgacatcacc     2040

aactggctgt ggtacatcaa gatcttcatc atgatcgtgg gcggcctgat cggcctgcgg     2100

atcatctttg ccgtgctgag catcgtgaac cgcgtgcggc agggctac                  2148
```

<210> 36

<211> 724

<212> PRT

<213> Artificial Sequence

<220>

<223> mosaic Env trimer sequence

<400> 36

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

Gly Thr Leu Ile Leu Gly Met Leu Met Ile Cys Ser Ala Ala Gly Lys
            20                  25                  30

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

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

His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
65                  70                  75                  80

Gln Glu Val Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp Lys
                85                  90                  95

Asn Asn Met Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp
            100                 105                 110
```

```
Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
        115                 120                 125

Asn Cys Thr Asp Asp Val Arg Asn Val Thr Asn Asn Ala Thr Asn Thr
    130                 135                 140

Asn Ser Ser Trp Gly Glu Pro Met Glu Lys Gly Glu Ile Lys Asn Cys
145                 150                 155                 160

Ser Phe Asn Ile Thr Thr Ser Ile Arg Asn Lys Val Gln Lys Gln Tyr
                165                 170                 175

Ala Leu Phe Tyr Lys Leu Asp Val Val Pro Ile Asp Asn Asp Ser Asn
            180                 185                 190

Asn Thr Asn Tyr Arg Leu Ile Ser Cys Asn Thr Ser Val Ile Thr Gln
        195                 200                 205

Ala Cys Pro Lys Val Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala
    210                 215                 220

Pro Ala Gly Phe Ala Ile Leu Lys Cys Asn Asp Lys Lys Phe Asn Gly
225                 230                 235                 240

Thr Gly Pro Cys Thr Asn Val Ser Thr Val Gln Cys Thr His Gly Ile
                245                 250                 255

Arg Pro Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu
            260                 265                 270

Glu Glu Val Val Ile Arg Ser Glu Asn Phe Thr Asn Asn Ala Lys Thr
        275                 280                 285

Ile Met Val Gln Leu Asn Val Ser Val Glu Ile Asn Cys Thr Arg Pro
    290                 295                 300

Asn Asn Asn Thr Arg Lys Ser Ile His Ile Gly Pro Gly Arg Ala Phe
305                 310                 315                 320

Tyr Thr Ala Gly Asp Ile Ile Gly Asp Ile Arg Gln Ala His Cys Asn
                325                 330                 335

Ile Ser Arg Ala Asn Trp Asn Asn Thr Leu Arg Gln Ile Val Glu Lys
            340                 345                 350

Leu Gly Lys Gln Phe Gly Asn Asn Lys Thr Ile Val Phe Asn His Ser
        355                 360                 365

Ser Gly Gly Asp Pro Glu Ile Val Met His Ser Phe Asn Cys Gly Gly
    370                 375                 380

Glu Phe Phe Tyr Cys Asn Ser Thr Lys Leu Phe Asn Ser Thr Trp Thr
385                 390                 395                 400

Trp Asn Asn Ser Thr Trp Asn Asn Thr Lys Arg Ser Asn Asp Thr Glu
                405                 410                 415
```

```
Glu His Ile Thr Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Trp
            420                 425                 430

Gln Glu Val Gly Lys Ala Met Tyr Ala Pro Pro Ile Arg Gly Gln Ile
        435                 440                 445

Arg Cys Ser Ser Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly
    450                 455                 460

Asn Asp Thr Ser Gly Thr Glu Ile Phe Arg Pro Gly Gly Gly Asp Met
465                 470                 475                 480

Arg Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Lys Ile
                485                 490                 495

Glu Pro Leu Gly Val Ala Pro Thr Lys Ala Lys Glu Arg Val Val Gln
            500                 505                 510

Arg Glu Glu Arg Ala Val Gly Ile Gly Ala Val Phe Leu Gly Phe Leu
        515                 520                 525

Gly Ala Ala Gly Ser Thr Met Gly Ala Ala Ser Met Thr Leu Thr Val
    530                 535                 540

Gln Ala Arg Leu Leu Leu Ser Gly Ile Val Gln Gln Gln Asn Asn Leu
545                 550                 555                 560

Leu Arg Ala Ile Glu Ala Gln Gln His Leu Leu Gln Leu Thr Val Trp
                565                 570                 575

Gly Ile Lys Gln Leu Gln Ala Arg Val Leu Ala Val Glu Arg Tyr Leu
            580                 585                 590

Lys Asp Gln Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile
        595                 600                 605

Cys Thr Thr Thr Val Pro Trp Asn Ala Ser Trp Ser Asn Lys Ser Leu
    610                 615                 620

Asp Lys Ile Trp Asn Asn Met Thr Trp Met Glu Trp Glu Arg Glu Ile
625                 630                 635                 640

Asn Asn Tyr Thr Ser Leu Ile Tyr Thr Leu Ile Glu Glu Ser Gln Asn
                645                 650                 655

Gln Gln Glu Lys Asn Glu Gln Glu Leu Leu Glu Leu Asp Lys Trp Ala
            660                 665                 670

Ser Leu Trp Asn Trp Phe Asp Ile Ser Asn Trp Leu Trp Tyr Ile Lys
        675                 680                 685

Ser Arg Ile Glu Gly Arg Gly Ser Gly Gly Tyr Ile Pro Glu Ala Pro
    690                 695                 700

Arg Asp Gly Gln Ala Tyr Val Arg Lys Asp Gly Glu Trp Val Leu Leu
705                 710                 715                 720
```

Ser Thr Phe Leu

<210> 37
<211> 28
<212> PRT
<213> Artificial Sequence

<220>
<223> Residues 655-682 of SEQ ID NO:18

<400> 37

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

Trp Phe Ser Ile Ser Lys Trp Leu Trp Tyr Ile Lys
            20                  25
```

REFERENCES CITED IN THE DESCRIPTION

Cited references

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- WO2010059732A [0003] [0004] [0005] [0117] [0149]
- US20120076812A [0034] [0149]
- WO2014107744A [0040] [0085]
- US6083716A [0053]
- WO2005071093A [0053]
- WO2010086189A [0053]
- WO2010085984A [0053]
- WO03104467A [0057] [0149]
- WO2007104792A [0058] [0058] [0149]
- WO2012082918A [0058] [0149]
- WO2010042942A [0137]

- WO2004001032A [0149]
- US7270811B [0149]
- WO0070071A [0149]

Non-patent literature cited in the description

- **BAROUCH et al.**Nat Med, 2010, vol. 16, 319-323 [0002] [0034] [0133] [0149]
- **BAROUCH et al.**Cell, 2013, vol. 155, 1-9 [0034] [0149]
- **EDWARDS et al.**J. Virology, 2002, vol. 76, 2683-2691 [0042] [0117] [0149]
- **SCHIERNLE et al.**PNAS, 1997, [0042] [0117]
- **ABRAHAMYAN et al.**J Virol, 2005, [0042] [0117]
- **FARINA et al.**J Virol, 2001, vol. 75, 11603-13 [0053]
- **COHEN et al.**J Gen Virol, 2002, vol. 83, 151-55 [0053]
- **KOBINGER et al.**Virology, 2006, vol. 346, 394-401 [0053]
- **TATSIS et al.**Molecular Therapy, 2007, vol. 15, 608-17 [0053]
- **BANGARIMITTAL**Vaccine, 2006, vol. 24, 849-62 [0053]
- **LASAROERTL**Mol Ther, 2009, vol. 17, 1333-39 [0053]
- **HAVENGA et al.**J Gen Virol, 2006, vol. 87, 2135-43 [0057] [0149]
- **ABBINK et al.**Virol, 2007, vol. 81, 94654-63 [0058] [0149]
- **ABBINK**J Virol, 2007, vol. 81, 94654-63 [0063]
- **HOGANSON et al.**Bioprocessing, 2002, vol. 1, 43-8 [0079]
- Remington's Pharmaceutical Sciences19800000 [0108]
- **WALKER et al.**Science, 2009, [0134]
- **HAYNES et al.**N Engl J Med., 2012, [0134]
- **BAROUCH et al.**Science, 2015, vol. 349, 320-324 [0137] [0149]
- **MONTEFIORI DC**Methods Mol Biol, 2009, vol. 485, 395-405 [0138]
- **SARZOTTI-KELSOE M et al.**J Immunol Methods, 2014, vol. 409, 131-146 [0138]
- **SCHIERNLE et al.**PNAS, 1997, vol. 94, 8640-8645 [0149]
- **ABRAHAMYAN et al.**J Virol, 2005, vol. 79, 106-115 [0149]
- **VOGELS et al.**J Virol, 2003, vol. 77, 158263-71 [0149]
- **WALKER LMPHOGAT SKCHAN-HUI PYWAGNER DPHUNG PGOSS JL et al.**Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine targetScience, 2009, vol. 326, 285-289 [0149]
- **HAYNES BFGILBERT PBMCELRATH MJZOLLA-PAZNER STOMARAS GDALAM SM et al.**Immune-correlates analysis of an HIV-1 vaccine efficacy trialN Engl J Med, 2012, vol. 366, 1275-1286 [0149]
- **MONTEFIORI DC**Measuring HIV neutralization in a luciferase reporter gene assayMethods Mol Biol, 2009, vol. 485, 395-405 [0149]
- **SARZOTTI-KELSOE MBAILER RTTURK ELIN CLBILSKA MGREENE KM et al.**Optimization and validation of the TZM-bl assay for standardized assessments of

neutralizing antibodies against HIV-1J Immunol Methods, 2014, vol. 409, 131-146 [0149]

Patentkrav

1. Vektor, som omfatter en nukleinsyre, der koder for et syntetisk HIV-kappeprotein, hvor det syntetiske HIV-kappeprotein omfatter eller består af aminosyresekvensen ifølge SEQ ID NO: 18.

2. Vektor ifølge krav 1, hvor vektoren er en virusvektor.

3. Vektor ifølge krav 1 eller 2, hvor vektoren er en adenovirusvektor.

4. Vektor ifølge krav 3, hvor adenovirusvektoren er en human adenovirusvektor eller en simian-adenovirusvektor.

5. Vektor ifølge krav 3, hvor adenovirusvektoren er en human adenovirus serotype 26 (Ad26)-vektor.

6. Vektor ifølge krav 1 eller 2, hvor vektoren er en adenoassocieret virusvektor, en poxvirusvektor, en MVA-vektor, en enterisk virusvektor, en venezuelansk hesteencefalitisvirusvektor, en Semliki Forest-virusvektor, en tobaksmosaikvirusvektor eller en lentivirusvektor.

7. Vektor ifølge krav 1 eller 2, hvor vektoren er en MVA-vektor.

8. Vektor ifølge krav 1 eller 2, hvor vektoren er et plasmid, et kunstigt kromosom fra bakterie, et kunstigt kromosom fra gær eller en bakteriofagvektor.

9. Sammensætning, som omfatter en immunologisk effektiv mængde af en vektor ifølge et hvilket som helst af kravene 6 eller 7 og et bæremateriale, hvor nukleinsyren, der koder for det syntetiske HIV-kappeprotein er operabelt koblet til en promotorsekvens.

10. Vaccinekombinationsprodukt, som omfatter:

(i) en første sammensætning, der omfatter en immunologisk effektiv mængde af en Ad26-vektor, der koder for et syntetisk HIV-kappeprotein med aminosyresekvensen ifølge SEQ ID NO: 18; og

(ii) en anden sammensætning, der omfatter en immunologisk effektiv mængde af en anden Ad26-vektor, der koder for et HIV-antigent polypeptid, der omfatter aminosyresekvensen ifølge SEQ ID NO: 5,

hvor den første og anden sammensætning er til stede i den samme sammensætning eller i forskellige sammensætninger.

11. Vaccinekombinationsprodukt ifølge krav 10, som yderligere omfatter mindst én ekstra sammensætning, der omfatter:
en immunologisk effektiv mængde af en vektor, der koder for mindst ét antigent polypeptid med aminosyresekvensen valgt fra gruppen, der består af SEQ ID NO: 1-4, 28 og 29,
hvor den første sammensætning, anden sammensætning og ekstra sammensætning er til stede i den samme sammensætning eller i en eller flere forskellige sammensætninger.

12. Vaccinekombinationsprodukt ifølge krav 10 eller 11, som yderligere omfatter mindst én yderligere sammensætning, der omfatter en immunologisk effektiv mængde af et eller flere isolerede HIV-antigene polypeptider,
hvor den første sammensætning, anden sammensætning, ekstra sammensætning og yderligere sammensætning er til stede i den samme sammensætning eller i en eller flere forskellige sammensætninger.

13. Vaccinekombinationsprodukt ifølge krav 12, hvor det ene eller de flere isolerede HIV-antigene polypeptider indbefatter et eller flere HIV-antigene polypeptider valgt fra gruppen, der består af stabiliseret HIV-clade C-trimert gp140-protein, der omfatter resterne 30-708 af aminosyresekvensen ifølge SEQ ID NO: 7, og mosaik-HIV-Env-trimert protein, der omfatter resterne 30-724 af SEQ ID NO: 36.

14. Kit, der omfatter et vaccinekombinationsprodukt ifølge et

hvilket som helst af kravene 10-13.

15. Sammensætning ifølge krav 9 eller vaccinekombinationsprodukt ifølge et hvilket som helst af kravene 10-13 til anvendelse til inducering af et immunrespons mod HIV hos et individ.

16. Sammensætning, som omfatter:
(i) en første adenovirus 26 (Ad26)-vektor, der koder for et syntetisk HIV-kappeprotein, der omfatter aminosyresekvensen ifølge SEQ ID NO: 18;
(ii) en anden Ad26-vektor, der koder for et HIV-antigent polypeptid, der omfatter aminosyresekvensen ifølge SEQ ID NO: 5;
(iii) en tredje Ad26-vektor, der koder for et HIV-antigent polypeptid, der omfatter aminosyresekvensen ifølge SEQ ID NO: 28; og
(iv) en fjerde Ad26-vektor, der koder for et HIV-antigent polypeptid, der omfatter aminosyresekvensen ifølge SEQ ID NO: 29.

17. Sammensætning ifølge krav 16, hvor den første, anden, tredje og fjerde Ad26-vektor er til stede i et viruspartikelforhold på 1:1:1:1.

18. Vaccinekombination, som omfatter en sammensætning ifølge krav 16 og et eller flere isolerede HIV-antigene polypeptider.

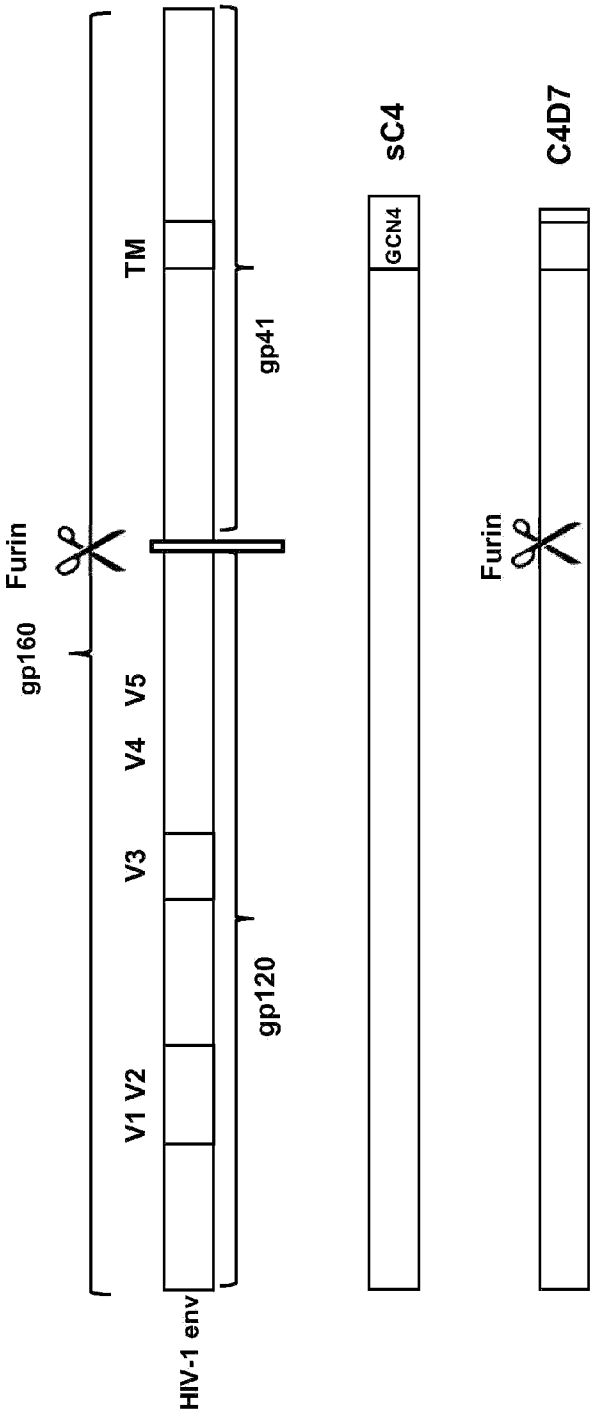
19. Vaccinekombination ifølge krav 18, hvor det ene eller de flere isolerede HIV-antigene polypeptider omfatter et HIV-kappeprotein, for eksempel et stabiliseret HIV-clade C-trimert gp140-protein og/eller et mosaikkappe-trimert protein.

20. Vaccinekombination ifølge krav 19, hvor clade C-trimert gp140-proteinet omfatter resterne 30-708 af aminosyresekvensen ifølge SEQ ID NO: 7, og/eller hvor det mosaikkappe-trimere protein omfatter resterne 30-724 af aminosyresekvensen ifølge SEQ ID NO: 36.

21. Ad26-vektor, som omfatter en nukleinsyre, der koder for et syntetisk HIV-kappeprotein, der omfatter aminosyresekvensen ifølge SEQ ID NO: 18, til anvendelse til inducering af et immunrespons mod HIV hos et individ.

22. Adenovirusvektorvaccine, som omfatter: (i) en første adenovirus 26 (Ad26)-vektor, der koder for et syntetisk HIV-kappeprotein, der omfatter aminosyresekvensen ifølge SEQ ID NO: 18; (ii) en anden Ad26-vektor, der koder for et HIV-antigent polypeptid, der omfatter aminosyresekvensen ifølge SEQ ID NO: 5; (iii) en tredje Ad26-vektor, der koder for et HIV-antigent polypeptid, der omfatter aminosyresekvensen ifølge SEQ ID NO: 28; og (iv) en fjerde Ad26-vektor, der koder for et HIV-antigent polypeptid, der omfatter aminosyresekvensen ifølge SEQ ID NO: 29,
til anvendelse til en primer-administration,
og
isoleret HIV-Env-gp140-protein, for eksempel (a) clade C-gp140-protein, der omfatter resterne 30-708 af SEQ ID NO: 7, eller (b) mosaik-gpl40-protein, der omfatter resterne ifølge SEQ ID NO: 36, eller (c) en kombination af clade C-gpl40-protein og mosaik-gpl40-protein,
til anvendelse sammen med adenovirusvektorvaccinen til en booster-administration, hvor primer- og booster-administrationerne er til fremkaldelse af et immunrespons mod HIV hos et individ.

DRAWINGS



A. B. C. Fig. 1

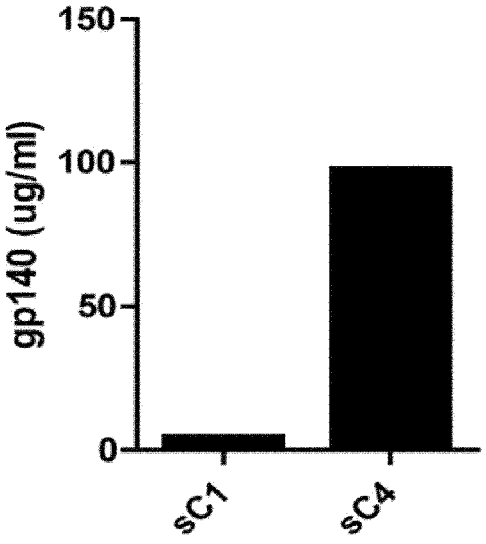
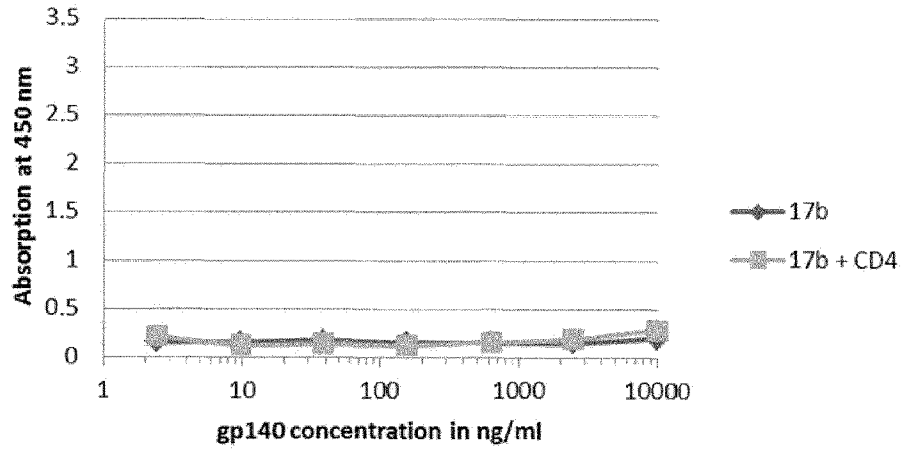


Fig. 2

A.


sC1
Absorption at 450 nm
3.5
3
2.5
2
1.5
1
0.5
0
1
10
100
1000
10000
gp140 concentration in ng/ml
17b
17b + CD4

B.

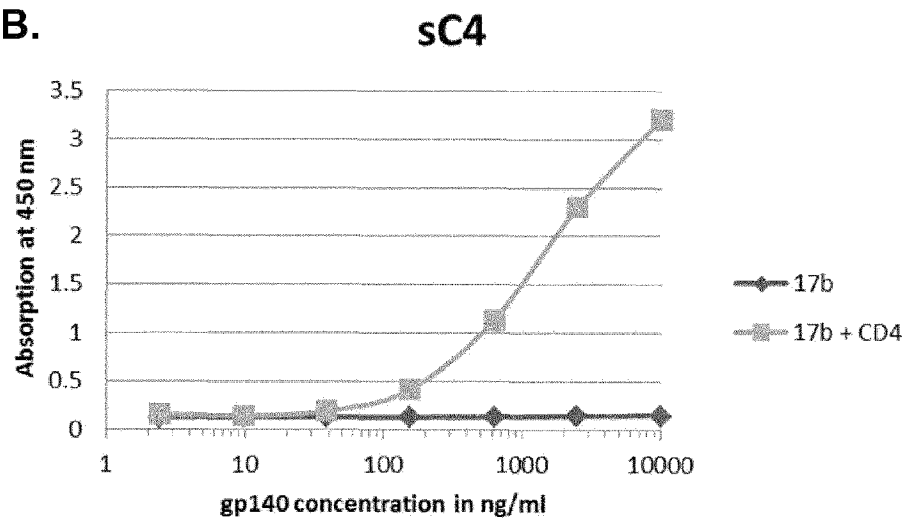

sC4
Absorption at 450 nm
3.5
3
2.5
2
1.5
1
0.5
0
1
10
100
1000
10000
gp140 concentration in ng/ml
17b
17b + CD4

Fig. 3

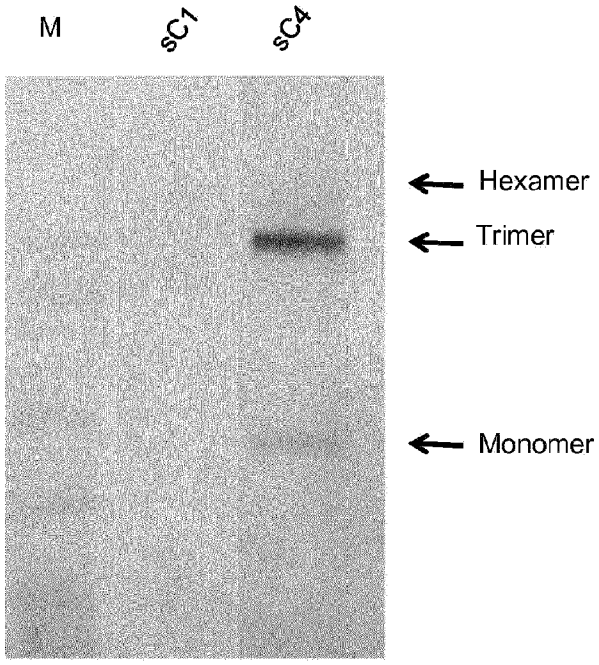


Fig. 4

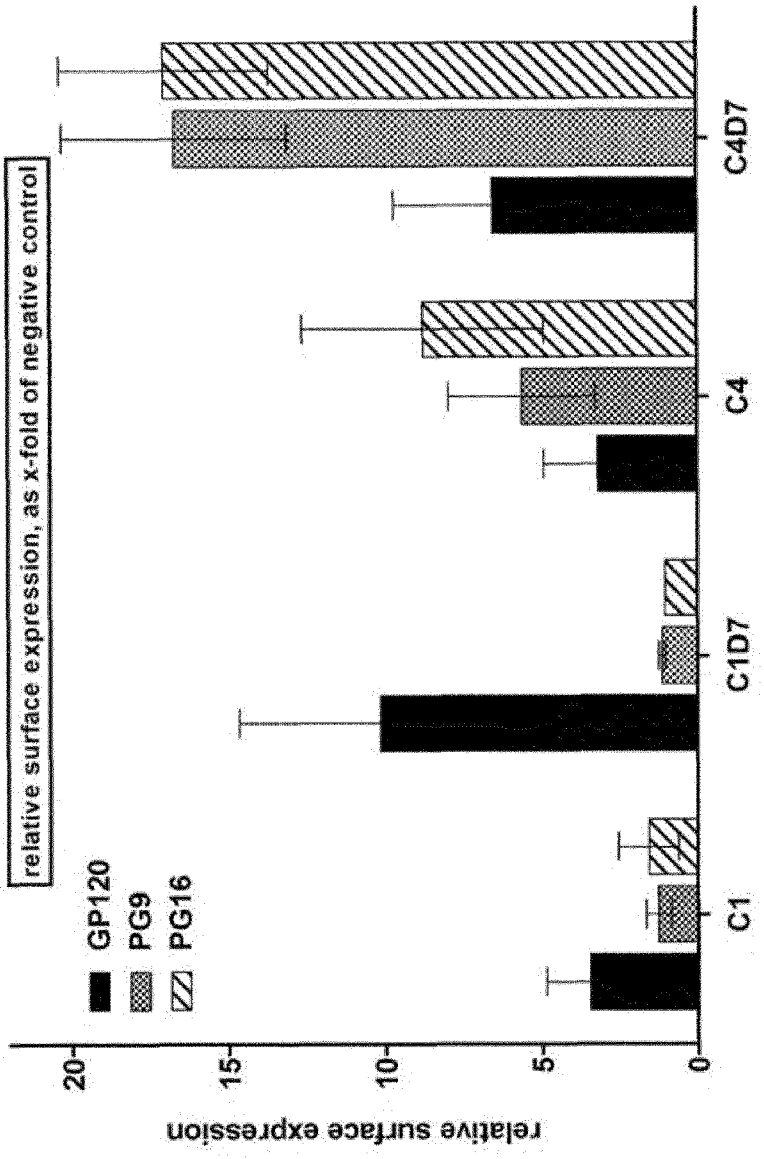

relative surface expression, as x-fold of negative control
GP120
PG9
PG16
relative surface expression
0
5
10
15
20
C1
C1D7
C4
C4D7

Fig. 5

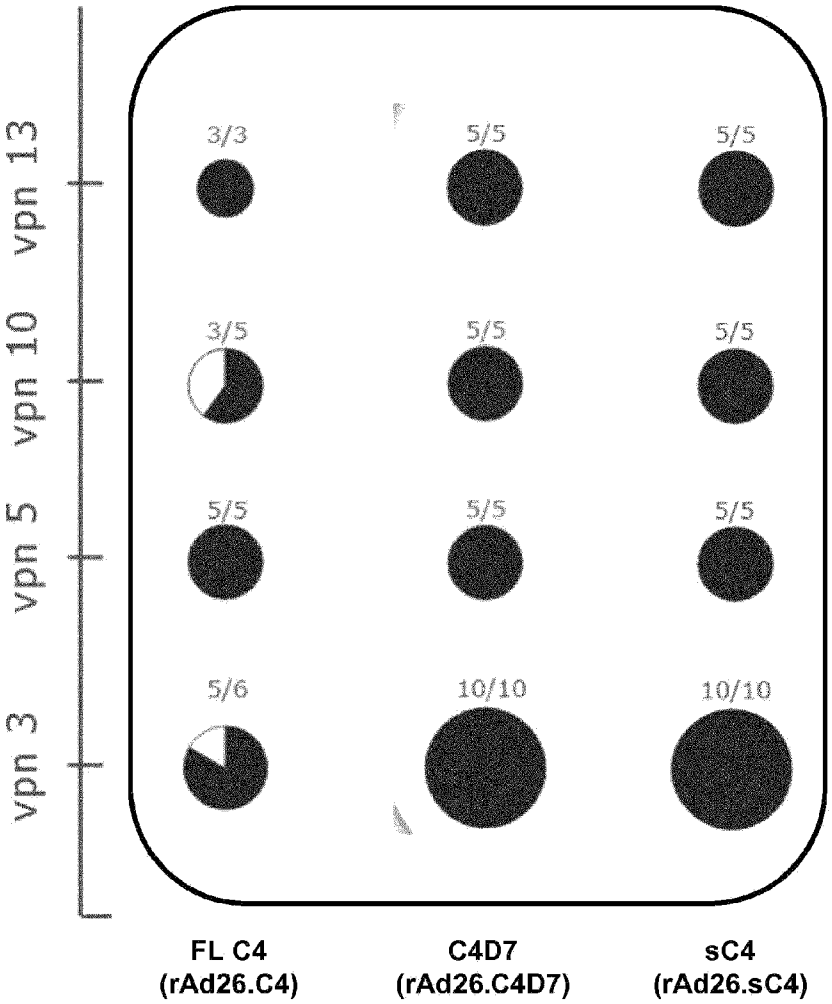
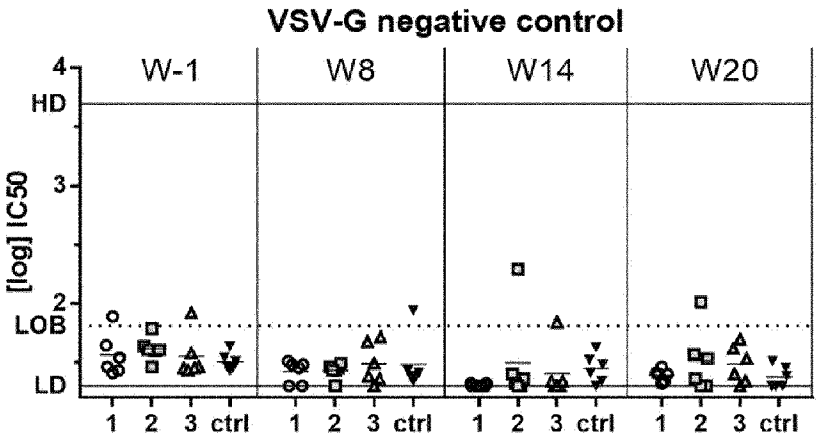
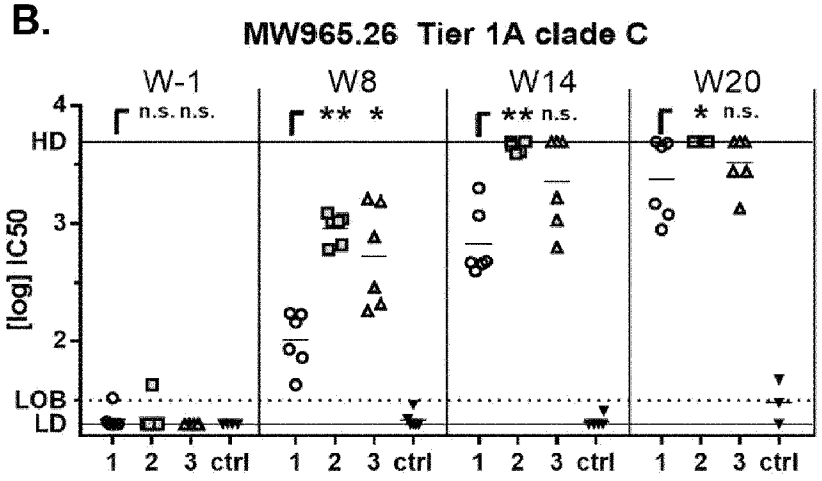

vpn 13
vpn 10
vpn 5
vpn 3
3/3
5/5
5/5
3/5
5/5
5/5
5/5
5/5
5/5
5/6
10/10
10/10
FL C4
(rAd26.C4)
C4D7
(rAd26.C4D7)
sC4
(rAd26.sC4)

Fig. 6

A.



B.



3-valent

4-valent (C4D7)

4-valent (sC4)

control (Ad26.empty)

Fig. 7