

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
12 January 2012 (12.01.2012)

(10) International Publication Number
WO 2012/06180 A1

(51) International Patent Classification:
C12N 15/49 (2006.01) **A61P 31/18** (2006.01)
A61K 39/21 (2006.01)

(21) International Application Number:
PCT/US201 1/042434

(22) International Filing Date:
29 June 2011 (29.06.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/359,738 29 June 2010 (29.06.2010) US
61/443,550 16 February 2011 (16.02.2011) US

(71) Applicants (for all designated States except US): **THE UNITED STATES OF AMERICA, AS REPRESENTED BY THE SECRETARY, DEPARTMENT OF HEALTH AND HUMAN SERVICES** [US/US]; National Institutes of Health, Office of Technology Transfer, 6011 Executive Blvd., Suite 325, MSC 7660, Bethesda, MD 20892-7660 (US). **PURDUE RESEARCH FOUNDATION** [US/US]; 3000 Kent Avenue, West Lafayette, IN 47906 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **NABEL, Gary, J.** [US/US]; 2520 30th Street, NW, Washington, DC 20008 (US). **YANG, Zhi-yong** [US/US]; 10004 Penfold Court, Potomac, MD 20854 (US). **AKAHATA, Wataru** [JP/US]; 11317 Palisades Ct., Kensington, MD 20895 (US). **KANEKIYO, Masaru** [JP/US]; 4803 Wellington Drive, Apt. 104, Chevy Chase, MD 20815 (US). **XU, Ling** [US/US]; 10004 Penfold Court, Potomac, MD 20854 (US). **KWONG, Peter, D.** [US/US]; 1436 T Street NW, Washington, DC 20009 (US). **ZHOU, Tongqing** [US/US]; 18303 Thundercloud Road, Boyds, MD 20841 (US). **BOYINGTON, Jeffrey, C.** [US/US]; 22816 Timber Creek Ln., Clarksburg, MD 20871 (US). **GEORGIEV, Ivelin, S.** [BG/US]; 9309 Kobe Way,

Gaithersburg, MD 20886 (US). **ZHU, Jiang** [US/US]; 19805 Helix Dr. Apt. 217, Ashburn, VA 20147 (US). **ROSSMANN, Michael** [US/US]; 1208 Wiley Dr., West Lafayette, IN 47906 (US).

(74) Agent: **GRISWOLD, Ian, J.**; Klarquist Sparkman, LLP, One World Trade Center, Suite 1600, 121 SW Salmon Street, Portland, OR 97204 (US).

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

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

Declarations under Rule 4.17:

— of inventorship (Rule 4.17(iv))

Published:

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

— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

— with sequence listing part of description (Rule 5.2(a))

(54) Title: HIV IMMUNOGENS

(57) Abstract: Disclosed are HIV immunogens. Also disclosed are nucleic acids encoding these immunogens and methods of producing these antigens. Methods for generating an immune response in a subject are also disclosed. In some embodiments, the method is a method for treating or preventing a human immunodeficiency type 1 (HIV-1) infection in a subject.



WO 2012/06180 A1

HIV IMMUNOGENS

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application No:
5 61/443,550, filed February 16, 2011, and U.S. Provisional Application No:
61/359,738, filed June 29, 2010, which are hereby incorporated herein by reference
in their entirety.

STATEMENT OF JOINT RESEARCH

10 The research work described herein was performed under a Collaboration
Agreement between the U.S. Government and Purdue University.

FIELD

The present disclosure relates to immunogenic polypeptides, and specifically
15 to polypeptides that can provoke an immune response to human immunodeficiency
virus (HIV).

BACKGROUND

Over 30 million people are infected with HIV worldwide, and 2.5 to 3
20 million new infections have been estimated to occur yearly. Although effective
antiretroviral therapies are available, millions succumb to AIDS every year,
especially in sub-Saharan Africa, underscoring the need to develop measure to
prevent the spread of this disease.

The primary immunologic abnormality resulting from infection by HIV is
25 the progressive depletion and functional impairment of T lymphocytes expressing
the CD4 cell surface glycoprotein. The loss of CD4 helper/inducer T cell function
probably underlies the defects in cellular and humoral immunity leading to the
opportunistic infections and malignancies characteristic of the acquired
immunodeficiency syndrome (AIDS) (Lane *et al.*, *Ann. Rev. Immunol.* 3:477, 1985).
30 Studies of HIV-1 infection of fractionated CD4 and CD8 T cells from normal donors
and AIDS patients have revealed that depletion of CD4 T cells results from the

ability of HIV-1 to selectively infect, replicate in, and ultimately destroy this T lymphocyte subset (Klatzmann *et al.*, *Science* 225:59, 1984). The possibility that CD4 itself is an essential component of the cellular receptor for HIV-1 was first indicated by the observation that monoclonal antibodies directed against CD4 block
5 HIV-1 infection and syncytia induction (Dalglish *et al.*, *Nature* 312:767,1984; McDougal *et al.*, *J. Immunol.* 135:3151, 1985). This hypothesis has been confirmed by the demonstration that a molecular complex forms between CD4 and the major envelope glycoprotein of HIV-1 (McDougal *et al.*, *Science* 231:382, 1986).

Over the past few decades the development of agents to combat HIV
10 infections has vastly increased the average lifespan of those infected throughout the world. However, in some instances HIV is able to acquire resistance to antivirals within a few replication cycles.

To combat the ever-changing landscape of HIV resistance to the current therapies, the standard course of action for pharmaceutical companies is to develop
15 an ever-increasing array of small molecule therapeutic agents. As an alternative, vaccines have been developed which stimulate the body to fight an infection by eliciting antibody responses to the target pathogen(s). In some examples, these vaccines are polypeptide epitopes that induce an immune response to pathogens and can be referred to as immunogens. These immunogens can be introduced into a
20 subject where they can elicit an antibody response to specific epitopes of the pathogen. For example, immunogens derived from the envelope protein of HIV have been used to produce an antibody response.

Initial attempts at generating neutralizing antibodies by vaccination with recombinant HIV gp120 protein analogous to some highly effective vaccines have
25 thus far proved unsuccessful in generating protective immunity. For example, in 2009, a large multicenter, double-blind, placebo-controlled clinical study revealed that priming immunization with ALVAC-HIV (a canary pox vector vaccine expressing HIV Env genes) followed by AIDSVAX B/E booster immunizations (recombinant HIV Env gp120 vaccine) can reduce the risk of HIV infection amongst
30 heterosexuals by 31 percent. However, the vaccine only induced short-term protection, and there has been speculation that short-lived antibody responses to

HIV Env protein might neutralize or block HIV mucosal transmission. An effective HIV-1 vaccine will likely need to induce neutralizing antibodies (NAbs) that block HIV-1 entry into human cells. To be effective, vaccine induced antibodies will have to be active against most circulating strains of HIV-1. Thus, there is a need for immunogens that can be used to elicit an immune response to pathogens, such as HIV.

SUMMARY

Disclosed are immunogens that include a human immunodeficiency virus (HIV) gpl20 outer domain polypeptide, which has been substantially modified from the wild type amino acid sequence to focus the immune response to the CD4 binding site of gpl20. In some examples, the gpl20 outer domain polypeptide includes amino acids 213-492 of gpl20, amino acids 213-482 of gpl20, amino acids 252-492 of gpl20, or 252-482 of gpl20. In some examples, the gpl20 outer domain polypeptide is part of a larger peptide, such as a HIV-1 gpl40 or HIV-1 gpl20 peptide.

In some embodiments, one or more of the loops of the outer domain peptide (such as one or more of the VI, V2, V3, V4 or V5 loop) have been altered to enhance or focus immunogenicity and/or increase stability of the immunogen. In some embodiments, the β 20/ β 21 bridging sheet of the gpl20 outer domain is removed or truncated.

In some embodiments, the immunogen includes one or more amino acid substitutions relative to the wild type sequence, for example to add or remove glycosylation sites, and/or to increase protein stability. In some embodiments, the isolated immunogen includes one or more of a foldon domain, a ferritin polypeptide, a hybrid of different ferritin polypeptides, a six-histidine residue tag, a secretion signal sequence and a transmembrane domain. In specific examples, the immunogen includes an amino acid sequence at least 95% identical to the amino acid sequence set forth as any one of SEQ ID NOs : 203, 58-77, 1-57, 79, or 80.

In some embodiments, the immunogen is part of a virus-like particle (VLP), such as a Chikungunya virus VLP. Also disclosed are nucleic acids molecules

encoding the disclosed immunogen. In specific examples, such nucleic acids include the nucleotide sequences set forth as one of SEQ ID NOs: 81-161.

Methods of generating an immune response in a subject are disclosed, as are methods of treating, inhibiting or preventing a HIV-1 infection in a subject. In such
5 methods a subject, such as a human subject, is administered and effective amount of a disclosed antigen.

Methods for detecting or isolating an HIV-1 binding antibody in a subject infected with HIV-1 are disclosed. In such methods, a disclosed immunogen is contacted with an amount of bodily fluid from a subject and the binding of the HIV-
10 1 binding antibody to the immunogen is detected, thereby detecting or isolating the HIV-1 binding antibody in a subject.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGS. 1A and 1B are schematic representations depicting the development
15 of representative outer domain immunogens disclosed herein.

FIG. 2 includes a ribbon diagram of modeled HIV gpl20 OD structures (upper left). The V3 loops are highlighted and circled. Also shown is an amino acid sequence alignment of various V3 modifications (upper right) (GSL; SEQ ID NO: 172) (V3.1; SEQ ID NO: 173), (V3.2; SEQ ID NO: 174), (V3.3; SEQ ID NO: 175),
20 V3.4; SEQ ID NO: 176). The symbol "y" indicates a potential glycosylation site. The graphs (lower left) show antibody binding profiles of V3-modified ODs. The purified parental OD4.2F and V3-modified ODs v3.3 and v3.4 proteins were coated onto plates pre-coated with snowdrop lectin (GNA) and tested for binding abilities to VRC01 and b12 monoclonal antibodies by ELISA. The EC₅₀ values (lower right)
25 were calculated from the curve and represented as mg/ml. The constructs used were A2.6 T383F (OD4.2F), A2.6.2 V3.1 (v3.1), A2.6.2 V3.2 (v3.2), A2.6.2 V3.3 (v3.3) and A2.6.2 V3.4 (v3.4).

FIG. 3 includes a ribbon diagram of modeled HIV gpl20 OD structures (upper left). The V5 loops are circled. Also shown is an amino acid sequence
30 alignment of various V5 modifications (upper right) (wild type; SEQ ID NO: 204) (v5.1; SEQ ID NO: 187) (v5.2; SEQ ID NO: 188) (v5.3; SEQ ID NO: 189). The

symbol "y" indicates a potential glycosylation site. The graphs (lower left) show antibody binding profiles of V5-modified ODs. The purified parental OD4.2F and V5-modified ODs v5.2 and v5.3 proteins were coated onto plates pre-coated with snowdrop lectin (GNA) and tested for binding abilities to VRCOL and b12

- 5 monoclonal antibodies by ELISA. The EC_{50} values (lower right) were calculated from the curve and represented as mg/ml. The constructs used were A2.6 T383F (OD4.2F), A2.6.2 V5.1 (v5.1), A2.6.2 V5.2 (v5.2) and A2.6.2 V5.3 (v5.3).

FIG. 4 includes a ribbon diagram of modeled OD structures (upper left). The V3 and V5 modifications are circled. Brief descriptions of each modification are listed (upper right). The graphs show antibody binding profiles of OD4.2 and OD4.2'. The purified parental OD4.2 and modified OD4.2' proteins were coated onto plates pre-coated with snowdrop lectin (GNA) and tested for binding abilities to VRCOL, bl2 and bl3 monoclonal antibodies by ELISA. The constructs used were A2.3 3ab (OD4.2), A2.6.3 N276D V3.3 V5.2 (OD4.2').

- 15 **FIG. 5** shows a surface representation of modeled OD structures with artificially added glycans (upper). The modeled OD surface is colored in gray, and VRCOL-contacting regions are highlighted. The native glycans and artificially added glycans are modeled. The numbers indicate the positions that have been glycosylated. The front (looking down to CD4BS) and back views are shown. The designed OD variants with various degrees and positions of glycosylation are also listed (bottom). The constructs used were A2.9a-j Glyc(295, 442) (HG.1), A2.9a-j Glyc(295, 442, 479) (HG.2), A2.9a-j Glyc(295, 442, 436*) (HG.3), A2.9a-j Glyc(295, 442, 424*) (HG.4), A2.9a-j Glyc(295, 442, 398) (HG.5), A2.9a-j Glyc(295, 442, 369) (HG.6), A2.9a-j Glyc(295, 442, 273, 377, 348) (HG.7), A2.9a-j Glyc(295, 442, 479, 273, 377, 436*, 398, 348, 369) (HG.8), A2.9a-j Glyc(295, 442, 479, 273, 377, 424*, 398, 348, 369) (HG.9) and A2.9a-j Glyc(295, 442, 479, 273, 377) (HG.10).

FIG. 6 is a series of graphs showing antibody binding profiles of the glycosylation variants of OD4.2'. The purified OD variant proteins were coated onto plates pre-coated with snowdrop lectin (GNA) and tested for binding abilities to VRCOL, bl2 and bl3 monoclonal antibodies by ELISA. The constructs used were

A2.9a-j Glyc(295, 442) (HG.1), A2.9a-j Glyc(295, 442, 479) (HG.2), A2.9a-j Glyc(295, 442, 436*) (HG.3), A2.9a-j Glyc(295, 442, 424*) (HG.4), A2.9a-j Glyc(295, 442, 398) (HG.5), A2.9a-j Glyc(295, 442, 369) (HG.6), A2.9a-j Glyc(295, 442, 273, 377, 348) (HG.7), A2.9a-j Glyc(295, 442, 479, 273, 377, 436*, 398, 348, 369) (HG.8) and A2.9a-j Glyc(295, 442, 479, 273, 377) (HG. 10).

FIG. 7 depicts the identification of the bl2-binding knock out variant HG.3. The antibody binding profile of HG.3 is highlighted with a frame (upper). The binding of VRCOI, bl2 and bl3 monoclonal antibodies are plotted as squares, circles and triangles, respectively. Also shown is a surface representation of modeled HG.3 structure (bottom). The modeled HG.3 surface is colored in gray, and VRCOI-contacting regions are highlighted. The native and artificially added glycans are modeled. The constructs used were A2.9a-j Glyc(295, 442, 436*) (HG.3).

FIG. 8 includes a surface representation of modeled HG.3 structures with artificially added glycans (upper left). The VRCOI-contacting regions are highlighted. The native glycans and artificially added glycans in HG.3 and in HG.3 variants are modeled. The numbers indicated positions that have been glycosylated. Also listed are the designed HG.3 variants with various degrees and positions of glycosylation (upper right). The purified **OD** variant proteins were coated onto plates pre-coated with snowdrop lectin (GNA) and tested for binding abilities to VRCOI, bl2 and bl3 monoclonal antibodies by ELISA (lower panels). The constructs used were A2.9a-j Glyc(295, 442, 436*) (HG.3), A2.HG3 Glyc(479) (HG.3.1), A2.HG3 Glyc(479, 398) (HG.3.2), A2.HG3 Glyc(479, 348) (HG.3.3), A2.HG3 Glyc(479, 377) (HG.3.4) and A2.HG3 Glyc(479, 273) (HG.3.5).

FIG. 9 includes a surface representation of modeled HG3.2 structures with artificially added glycans (upper left). The modeled HG3.2 surface is colored in gray, and VRCOI-contacting regions are highlighted. The native glycans and artificially added glycans are modeled. The HG3.2 structure complex with VRCOI (bottom left), bl2 (bottom center) and bl3 (bottom right) were modeled. The added glycans may prevent binding of antibodies with unfavorable angle (less potent neutralizing antibodies like bl2 and bl3).

FIG. 10 includes schematic representations of the insertion of a representative immunogen into the genome of the CHIK, and the creation of virus like particles presenting the immunogen on the surface. OD4.0 was inserted between amino acid 205 and amino acid 206 E2. Corresponding E2 from Sindbis map were shown in the middle and right panel. The purified CHIK-OD was reconstructed by Cryoelectron microscopy.

FIG. 11 shows antibody titers of sera against RSC3 or delta RSC3 mutant that knockout CD4 binding domain from an immunized rabbit. The rabbits were immunized intramuscularly two times with CHIK-OD. The curve fit was calculated by Prism software.

FIG. 12 shows the results of immunized monkeys with one of the CHIKV-OD VLP. Significant amount of gpl40 binding antibodies were elicited after 3 times of immunization. Interestingly, in 2 out of 3 sera the elicitation of CD4 BS antibodies was seen, again based on the RSC3/ARSC3 ELISA assay.

FIG. 13 shows that a heavily glycosylated OD displayed on CHIK VLP reacts only with VRCO1 and VRCPG04.

FIG. 14 is a flow chart showing the work flow of designing the OD-based vaccine immunogen.

FIG. 15 shows the construction of a ferritin nanoparticle expressing an outer domain immunogen.

FIG. 16 shows the immunogenicity of HIV OD-ferritin nanoparticles.

FIG. 17 shows the induction of CD4 BS antibodies by glycan modified RSC3: Y5.

FIG. 18 is schematic representation depicting the development of representative RSC3 immunogens disclosed herein.

SEQUENCES

The nucleic acid and amino acid sequences disclosed herein are shown using standard letter abbreviations for nucleotide bases, and one letter code for amino acids. Only one strand of each nucleic acid sequence is shown, but the complementary strand is understood as included by any reference to the displayed

strand. The Sequence Listing is submitted as an ASCII text file in the form of the file named Sequence_Listing.txt, which was created on June 29, 2011, is 432,644 bytes, and is incorporated by reference herein.

5 SEQ ID NOs: 1-77, 79, 80, 148, 149 and 203 are the amino acid sequences of exemplary HIV antigens.

 SEQ ID NO: 78 is the amino acid sequence of the HXB2 core.

 SEQ ID NOs: 81-147, 150-161 and 206-215 are the nucleic acid sequences encoding exemplary HIV antigens.

10 SEQ ID NO: 162 is the amino acid sequence of an exemplary modified V1/V2 loop.

 SEQ ID NO: 163 is the amino acid sequence of an exemplary VI wild type loop.

15 SEQ ID NOs: 164-165 are the amino acid sequence of exemplary modified VI loops.

 SEQ ID NOs: 166-167 are the amino acid sequence of exemplary modified V2 loops.

 SEQ ID NOs: 168-181 are the amino acid sequence of exemplary modified V3 loops.

20 SEQ ID NOs: 182-186 are the amino acid sequence of exemplary modified V4 loops.

 SEQ ID NOs: 187-194 and 202 are the amino acid sequence of exemplary modified V5 loops.

 SEQ ID NO: 195 is the amino acid sequence of an exemplary fibrin foldon.

25 SEQ ID NO: 196 is the amino acid sequence of an exemplary CD4 transmembrane domain.

 SEQ ID NO: 197 is the amino acid sequence of an influenza transmembrane neuraminidase domain.

30 SEQ ID NO: 198 is the amino acid sequence of an influenza transmembrane hemagglutinin domain.

 SEQ ID NO: 199 is the amino acid sequence of a CD5 leader sequence.

SEQ ID NO: 200 is the amino acid sequence of an IL-2 secretion signal sequence.

SEQ ID NO: 201 is the amino acid sequence of an exemplary peptide linker.

5 SEQ ID NO: 204 is the amino acid sequence of an exemplary wild type V5 loop.

SEQ ID NO: 205 is the amino acid sequence of an exemplary wild type V3 loop.

DETAILED DESCRIPTION

10 /. ***Summary of Terms***

Unless otherwise noted, technical terms are used according to conventional usage. Definitions of common terms in molecular biology can be found in Benjamin Lewin, *Genes VII*, published by Oxford University Press, 1999; Kendrew *et al.* (eds.), *The Encyclopedia of Molecular Biology*, published by Blackwell Science Ltd., 1994; and Robert A. Meyers (ed.), *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, published by VCH Publishers, Inc., 1995; and other similar references.

20 As used herein, the singular forms "a," "an," and "the," refer to both the singular as well as plural, unless the context clearly indicates otherwise. For example, the term "an antigen" includes single or plural antigens and can be considered equivalent to the phrase "at least one antigen"

As used herein, the term "comprises" means "includes." Thus, "comprising an antigen" means "including an antigen" without excluding other elements.

25 It is further to be understood that any and all base sizes or amino acid sizes, and all molecular weight or molecular mass values, given for nucleic acids or polypeptides are approximate, and are provided for descriptive purposes, unless otherwise indicated. Although many methods and materials similar or equivalent to those described herein can be used, particular suitable methods and materials are described below. In case of conflict, the present specification, including
30 explanations of terms, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

To facilitate review of the various embodiments of the invention, the following explanations of terms are provided:

Adjuvant: A vehicle used to enhance antigenicity. Adjuvants include a suspension of minerals (alum, aluminum hydroxide, or phosphate) on which antigen is adsorbed; or water-in-oil emulsion in which antigen solution is emulsified in mineral oil (Freund incomplete adjuvant), sometimes with the inclusion of killed mycobacteria (Freund's complete adjuvant) to further enhance antigenicity (inhibits degradation of antigen and/or causes influx of macrophages). Immunostimulatory oligonucleotides (such as those including a CpG motif) can also be used as adjuvants (for example see U.S. Patent No. 6,194,388; U.S. Patent No. 6,207,646; U.S. Patent No. 6,214,806; U.S. Patent No. 6,218,371; U.S. Patent No. 6,239,116; U.S. Patent No. 6,339,068; U.S. Patent No. 6,406,705; and U.S. Patent No. 6,429,199). Adjuvants include biological molecules (a "biological adjuvant"), such as costimulatory molecules. Exemplary adjuvants include IL-2, RANTES, GM-CSF, TNF- α , IFN- γ , G-CSF, LFA-3, CD72, B7-1, B7-2, OX-40L and 4-1 BBL. Adjuvants can be used in combination with the disclosed antigens.

Administration: The introduction of a composition into a subject by a chosen route. For example, if the chosen route is intravenous, the composition (such as a disclosed antigen) is administered by introducing the composition into a vein of the subject.

Amino acid substitutions: The replacement of one amino acid in an antigen with a different amino acid. In some examples, an amino acid in an antigen is substituted with an amino acid from a homologous antigen.

Amplification: A technique that increases the number of copies of a nucleic acid molecule (such as an RNA or DNA). An example of amplification is the polymerase chain reaction, in which a biological sample is contacted with a pair of oligonucleotide primers, under conditions that allow for the hybridization of the primers to a nucleic acid template in the sample. The primers are extended under suitable conditions, dissociated from the template, and then re-annealed, extended, and dissociated to amplify the number of copies of the nucleic acid. The product of amplification can be characterized by electrophoresis, restriction endonuclease

cleavage patterns, oligonucleotide hybridization or ligation, and/or nucleic acid sequencing using standard techniques. Other examples of amplification include strand displacement amplification, as disclosed in U.S. Patent No. 5,744,311; transcription-free isothermal amplification, as disclosed in U.S. Patent No.

5 6,033,881; repair chain reaction amplification, as disclosed in WO 90/01069; ligase chain reaction amplification, as disclosed in EP-A-320 308; gap filling ligase chain reaction amplification, as disclosed in U.S. Patent No. 5,427,930; and NASBA™ RNA transcription-free amplification, as disclosed in U.S. Patent No. 6,025,134.

Animal: A living multi-cellular vertebrate or invertebrate organism, a
10 category that includes, for example, mammals. The term mammal includes both human and non-human mammals. Similarly, the term "**subject**" includes both human and veterinary subjects, such as non-human primates. Thus, administration to a subject can include administration to a human subject. Particular examples of veterinary subjects include domesticated animals (such as cats and dogs), livestock
15 (for example, cattle, horses, pigs, sheep, and goats), laboratory animals (for example, mice, rabbits, rats, gerbils, guinea pigs, and non-human primates).

Antibody: A polypeptide substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof, which specifically binds and recognizes an analyte (such as an antigen or immunogen) such as a gpl20, gpl40
20 polypeptide or antigenic fragment thereof, such as a gpl20 outer domain or an resurfaced gpl20, gpl40 polypeptide or antigenic fragment thereof.

Immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon and mu constant region genes, as well as the myriad immunoglobulin variable region genes.

25 Antibodies exist, for example as intact immunoglobulins and as a number of well characterized fragments produced by digestion with various peptidases. For instance, Fabs, Fvs, and single-chain Fvs (SCFvs) that bind to gpl20, or gpl40 would be gpl20-, or gpl40 -specific binding agents. This includes intact immunoglobulins and the variants and portions of them well known in the art, such
30 as Fab' fragments, F(ab)'₂ fragments, single chain Fv proteins ("scFv"), and disulfide stabilized Fv proteins ("dsFv"). A scFv protein is a fusion protein in which a light

chain variable region of an immunoglobulin and a heavy chain variable region of an immunoglobulin are bound by a linker, while in dsFvs, the chains have been mutated to introduce a disulfide bond to stabilize the association of the chains. The term also includes genetically engineered forms such as chimeric antibodies (such as
5 humanized murine antibodies), heteroconjugate antibodies (such as bispecific antibodies). See also, *Pierce Catalog and Handbook*, 1994-1995 (Pierce Chemical Co., Rockford, IL); Kuby, J., *Immunology*, 3rd Ed., W.H. Freeman & Co., New York, 1997.

Antibody fragments are defined as follows: (1) Fab, the fragment which
10 contains a monovalent antigen-binding fragment of an antibody molecule produced by digestion of whole antibody with the enzyme papain to yield an intact light chain and a portion of one heavy chain; (2) Fab', the fragment of an antibody molecule obtained by treating whole antibody with pepsin, followed by reduction, to yield an intact light chain and a portion of the heavy chain; two Fab' fragments are obtained
15 per antibody molecule; (3) (Fab')₂, the fragment of the antibody obtained by treating whole antibody with the enzyme pepsin without subsequent reduction; (4) F(ab')₂, a dimer of two Fab' fragments held together by two disulfide bonds; (5) Fv, a genetically engineered fragment containing the variable region of the light chain and the variable region of the heavy chain expressed as two chains; and (6) single chain
20 antibody ("SCA"), a genetically engineered molecule containing the variable region of the light chain, the variable region of the heavy chain, linked by a suitable polypeptide linker as a genetically fused single chain molecule. The term "antibody," as used herein, also includes antibody fragments either produced by the modification of whole antibodies or those synthesized de novo using recombinant
25 DNA methodologies.

Typically, a naturally occurring immunoglobulin has heavy (H) chains and light (L) chains interconnected by disulfide bonds. There are two types of light chain, lambda (λ) and kappa (κ). There are five main heavy chain classes (or isotypes) which determine the functional activity of an antibody molecule: IgM,
30 IgD, IgG, IgA and IgE.

Each heavy and light chain contains a constant region and a variable region, (the regions are also known as "domains"). In combination, the heavy and the light chain variable regions specifically bind the antigen. Light and heavy chain variable regions contain a "framework" region interrupted by three hypervariable regions, also called "complementarity-determining regions" or "CDRs." The extent of the framework region and CDRs have been defined (see, Kabat *et al.*, *Sequences of Proteins of Immunological Interest*, U.S. Department of Health and Human Services, 1991, which is hereby incorporated by reference). The Kabat database is now maintained online. The sequences of the framework regions of different light or heavy chains are relatively conserved within a species. The framework region of an antibody, that is the combined framework regions of the constituent light and heavy chains, serves to position and align the CDRs in three-dimensional space.

The CDRs are primarily responsible for binding to an epitope of an antigen. The CDRs of each chain are typically referred to as CDR1, CDR2, and CDR3, numbered sequentially starting from the N-terminus, and are also typically identified by the chain in which the particular CDR is located. Thus, a V_H CDR3 is located in the variable domain of the heavy chain of the antibody in which it is found, whereas a V_L CDR1 is the CDR1 from the variable domain of the light chain of the antibody in which it is found. Light chain CDRs are sometimes referred to as CDR L1, CDR L2, and CDR L3. Heavy chain CDRs are sometimes referred to as CDR H1, CDR H2, and CDR H3.

References to " V_H " or "VH" refer to the variable region of an immunoglobulin heavy chain, including that of an Fv, scFv, dsFv or Fab. References to " V_L " or "VL" refer to the variable region of an immunoglobulin light chain, including that of an Fv, scFv, dsFv or Fab.

A "monoclonal antibody" is an antibody produced by a single clone of B-lymphocytes or by a cell into which the light and heavy chain genes of a single antibody have been transfected. Monoclonal antibodies are produced by methods known to those of skill in the art, for instance by making hybrid antibody-forming cells from a fusion of myeloma cells with immune spleen cells. These fused cells

and their progeny are termed "hybridomas." Monoclonal antibodies include humanized monoclonal antibodies.

Antigen: A compound, composition, or substance that can stimulate the production of antibodies or a T cell response in an animal, including compositions that are injected or absorbed into an animal. An antigen reacts with the products of specific humoral or cellular immunity, including those induced by heterologous antigens, such as the disclosed antigens. "Epitope" or "antigenic determinant" refers to the region of an antigen to which B and/or T cells respond. In one embodiment, T cells respond to the epitope, when the epitope is presented in conjunction with an MHC molecule. Epitopes can be formed both from contiguous amino acids or noncontiguous amino acids juxtaposed by tertiary folding of a protein. Epitopes formed from contiguous amino acids are typically retained on exposure to denaturing solvents whereas epitopes formed by tertiary folding are typically lost on treatment with denaturing solvents. An epitope typically includes at least 3, and more usually, at least 5, about 9, or about 8-10 amino acids in a unique spatial conformation. Methods of determining spatial conformation of epitopes include, for example, x-ray crystallography and nuclear magnetic resonance.

Examples of antigens include, but are not limited to, peptides, lipids, polysaccharides, and nucleic acids containing antigenic determinants, such as those recognized by an immune cell. In some examples, antigens include peptides derived from a pathogen of interest. Exemplary pathogens include bacteria, fungi, viruses and parasites. In specific examples, an antigen is derived from HIV, such as a gpl20, gpl40 polypeptide or antigenic fragment thereof, such as an outer domain of gpl20.

CD4: Cluster of differentiation factor 4 polypeptide, a T-cell surface protein that mediates interaction with the MHC class II molecule. CD4 also serves as the primary receptor site for HIV on T-cells during HIV-1 infection.

The known sequence of the CD4 precursor has a hydrophobic signal peptide, an extracellular region of approximately 370 amino acids, a highly hydrophobic stretch with significant identity to the membrane-spanning domain of the class II

MHC beta chain, and a highly charged intracellular sequence of 40 residues (Maddon, *Cell* 42:93, 1985).

The extracellular domain of CD4 consists of four contiguous immunoglobulin-like regions (D1, D2, D3, and D4, see Sakihama *et al.*, *Proc. Natl. Acad. Sci.* 92:6444, 1995; U.S. Patent No. 6,117,655), and amino acids 1 to 183 have been shown to be involved in gp120 binding.

CD4BS antibodies: Antibodies that bind to or substantially overlap the CD4 binding surface of a gp120, or gp140, polypeptide. The antibodies interfere with or prevent CD4 from binding to a gp120 polypeptide.

10 **CD4i antibodies:** Antibodies that bind to a conformation of gp120 induced by CD4 binding.

CD8: Cluster of differentiation factor 8, a T cell surface protein that mediates interaction with the MHC Class I molecule. Cells that express CD8 are often cytotoxic T cells.

15 **Contacting:** Placement in direct physical association; includes both in solid and liquid form. Contacting includes contact between one molecule and another molecule, for example the amino acid on the surface of one polypeptide, such as an antigen, that contact another polypeptide, such as an antibody. Contacting also includes administration, such as administration of a disclosed antigen to a subject by
20 a chosen route.

Degenerate variant and conservative variant: A polynucleotide encoding a polypeptide or an antibody that includes a sequence that is degenerate as a result of the genetic code. For example, a polynucleotide encoding a disclosed antigen or an antibody that specifically binds a disclosed antigen includes a sequence that is
25 degenerate as a result of the genetic code. There are 20 natural amino acids, most of which are specified by more than one codon. Therefore, all degenerate nucleotide sequences are included as long as the amino acid sequence of the antigen or antibody that binds the antigen encoded by the nucleotide sequence is unchanged. Because of the degeneracy of the genetic code, a large number of functionally
30 identical nucleic acids encode any given polypeptide. For instance, the codons CGU, CGC, CGA, CGG, AGA, and AGG all encode the amino acid arginine. Thus, at

every position where an arginine is specified within a protein encoding sequence, the codon can be altered to any of the corresponding codons described without altering the encoded protein. Such nucleic acid variations are "silent variations," which are one species of conservative variations. Each nucleic acid sequence herein
5 that encodes a polypeptide also describes every possible silent variation.

One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule by standard techniques. Accordingly, each "silent variation" of a nucleic acid which encodes a polypeptide is implicit in each
10 described sequence.

One of ordinary skill will recognize that individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids (for instance less than 5%, in some embodiments less than 1%) in an encoded sequence are conservative variations where the alterations result in the
15 substitution of an amino acid with a chemically similar amino acid.

Conservative amino acid substitutions providing functionally similar amino acids are well known in the art. The following six groups each contain amino acids that are conservative substitutions for one another:

- 1) Alanine (A), Serine (S), Threonine (T);
- 20 2) Aspartic acid (D), Glutamic acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

25 Not all residue positions within a protein will tolerate an otherwise "conservative" substitution. For instance, if an amino acid residue is essential for a function of the protein, even an otherwise conservative substitution may disrupt that activity, for example the specific binding of an antibody to a target epitope may be disrupted by a conservative mutation in the target epitope.

Expression: Translation of a nucleic acid into a protein. Proteins may be expressed and remain intracellular, become a component of the cell surface membrane, or be secreted into the extracellular matrix or medium.

Expression Control Sequences: Nucleic acid sequences that regulate the expression of a heterologous nucleic acid sequence to which it is operatively linked. Expression control sequences are operatively linked to a nucleic acid sequence when the expression control sequences control and regulate the transcription and, as appropriate, translation of the nucleic acid sequence. Thus expression control sequences can include appropriate promoters, enhancers, transcription terminators, a start codon (ATG) in front of a protein-encoding gene, splicing signal for introns, maintenance of the correct reading frame of that gene to permit proper translation of mRNA, and stop codons. The term "control sequences" is intended to include, at a minimum, components whose presence can influence expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences. Expression control sequences can include a promoter.

A promoter is a minimal sequence sufficient to direct transcription. Also included are those promoter elements which are sufficient to render promoter-dependent gene expression controllable for cell-type specific, tissue-specific, or inducible by external signals or agents; such elements may be located in the 5' or 3' regions of the gene. Both constitutive and inducible promoters are included (see for example, Bitter *et al*, *Methods in Enzymology* 153:516-544, 1987). For example, when cloning in bacterial systems, inducible promoters such as pL of bacteriophage lambda, plac, ptrp, ptac (ptrp-lac hybrid promoter) and the like may be used. In one embodiment, when cloning in mammalian cell systems, promoters derived from the genome of mammalian cells (such as metallothionein promoter) or from mammalian viruses (such as the retrovirus long terminal repeat; the adenovirus late promoter; the vaccinia virus 7.5K promoter) can be used. Promoters produced by recombinant DNA or synthetic techniques may also be used to provide for transcription of the nucleic acid sequences.

A polynucleotide can be inserted into an expression vector that contains a promoter sequence, which facilitates the efficient transcription of the inserted genetic sequence of the host. The expression vector typically contains an origin of replication, a promoter, as well as specific nucleic acid sequences that allow
5 phenotypic selection of the transformed cells.

Ferritin: A protein that stores iron and releases it in a controlled fashion. The protein is produced by almost all living organisms. Ferritin assembles into a globular protein complex that in some cases consists of 24 protein subunits. In some examples, ferritin is used to form a particle (such as a virus-like particle) presenting
10 antigens on its surface, for example HIV antigens, such as the disclosed gpl20, gpl40 antigens or immunogenic fragments thereof.

Foldon domain: An amino acid sequence that naturally forms a trimeric structure. In some examples, a foldon domain can be included in the amino acid sequence of a disclosed antigen so that the antigen will form a trimer. In one
15 example, a foldon domain is the T4 foldon domain.

Glycoprotein (gp): A protein that contains oligosaccharide chains (glycans) covalently attached to polypeptide side-chains. The carbohydrate is attached to the protein in a cotranslational or posttranslational modification. This process is known as glycosylation. In proteins that have segments extending extracellularly, the
20 extracellular segments are often glycosylated. Glycoproteins are often important integral membrane proteins, where they play a role in cell-cell interactions. In some examples a glycoprotein is an HIV glycoprotein, such as a HIV gpl20, gpl40 or an immunogenic fragment thereof.

Glycosylation site: An amino acid sequence on the surface of a polypeptide, such as a protein, which accommodates the attachment of a glycan. An N-linked glycosylation site is triplet sequence of NXS/T in which N is asparagine, X is any residues except proline, S/T means serine or threonine. A glycan is a polysaccharide or oligosaccharide. Glycan may also be used to refer to the carbohydrate portion of a glycoconjugate, such as a glycoprotein, glycolipid, or a proteoglycan.
25

gpl20: An envelope protein from human immunodeficiency virus (HIV). The envelope protein is initially synthesized as a longer precursor protein of 845-
30

870 amino acids in size, designated gpl60. gpl60 forms a homotrimer and undergoes glycosylation within the Golgi apparatus. It is then cleaved by a cellular protease into gpl20 and gp41. gpl20 contains most of the external, surface-exposed, domains of the envelope glycoprotein complex, and it is gpl20 that binds both to the cellular CD4 receptor and to the cellular chemokine receptors (such as CCR5).

The mature gpl20 wild-type polypeptides have about 500 amino acids in the primary sequence. The gpl20 is heavily N-glycosylated giving rise to an apparent molecular weight of 120 kD. The polypeptide is comprised of five conserved regions (C1-C5) and five regions of high variability (V1-V5). Exemplary sequences of wild-type gpl60 polypeptides are shown on GENBANK®, for example Accession Nos. AAB05604 and AAD12142, which are incorporated herein by reference in their entirety as available on June 29, 2010. Exemplary sequences of gpl20 polypeptides from HIV-1 DU156 are shown on GENBANK®, for example Accession Nos. ABD83635, AAO50350 and AAT91997, which are incorporated herein by reference in their entirety as available on September 27, 2010. Exemplary sequences of gpl20 polypeptides from HIV-1 ZA012 are shown on GENBANK®, for example Accession No. ACF75939, which is incorporated herein by reference in its entirety as available on September 27, 2010.

The gpl20 core has a unique molecular structure, which comprises two domains: an "inner" domain (which faces gp41) and an "outer" domain (which is mostly exposed on the surface of the oligomeric envelope glycoprotein complex). The two gpl20 domains are separated by a "bridging sheet" that is not part of either of these domains. The gpl20 core comprises 25 beta strands, 5 alpha helices, and 10 defined loop segments.

The core gpl20 comprises 25 β -strands, 5 α -helices and 10 defined loop segments. The polypeptide chain of gpl20 is folded into two major domains, plus certain excursions that emanate from this body. The inner domain (inner with respect to the N and C termini) features a two-helix, two-strand bundle with a small five-stranded β -sandwich at its termini-proximal end and a projection at the distal end from which the V1/V2 stem emanates. The outer domain is a stacked double barrel that lies alongside the inner domain so that the outer barrel and inner bundle

axes are approximately parallel. The bridging sheet ($\beta 3$, $\beta 2$, $\beta 21$, $\beta 20$) packs primarily over the inner domain, although some surface residues of the outer domain, such as Phenylalanine 382, reach in to form part of its hydrophobic core.

The numbering used in the gpl20 derived antigens disclosed herein is
5 relative to the HXB2 numbering scheme as set forth in *Numbering Positions in HIV Relative to HXB2CG* Bette Korber *et al*, Human Retroviruses and AIDS 1998: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences. Korber *et al*, Eds. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM, which is incorporated by reference herein in its
10 entirety.

gpl40: An oligomeric form of HIV envelope protein, which contains all of gpl20 and the entire gp41 ectodomain.

Homologous proteins: Proteins from two or more species that have a similar structure and function in the two or more species. For example a gpl20
15 antigen from one species of lentivirus such as HIV-1 is a homologous antigen to a gpl20 antigen from a related species such as HIV-2 or SIV. Homologous proteins share the same protein fold and can be considered structural homologs.

Homologous proteins typically share a high degree of sequence conservation, such as at least 30% at least 40% at least 50%, at least 60%, at least 70%, at least
20 80% or at least 90% sequence conservation. Homologous proteins can share a high degree of sequence identity, such as at least 30% at least 40% at least 50%, at least 60%, at least 70%, at least 80% or at least 90% sequence identity.

Host cells: Cells in which a vector can be propagated and its DNA expressed. The cell may be prokaryotic or eukaryotic. The term also includes any
25 progeny of the subject host cell. It is understood that all progeny may not be identical to the parental cell since there may be mutations that occur during replication. However, such progeny are included when the term "host cell" is used.

Immunogenic polypeptide: A protein or a portion thereof that is capable of inducing an immune response in a mammal, such as a mammal infected or at risk of
30 infection with a pathogen. Administration of an immunogenic polypeptide derived from a pathogen of interest that inducing an immune response. Administration of an

immunogenic polypeptide can lead to protective immunity against a pathogen of interest. In some examples, an immunogenic polypeptide is an antigen that is resurfaced to focus immunogenicity to a target epitope. An "**immunogenic gpl20 or gpl40 polypeptide**" is gpl20, or gpl40 molecule, a resurfaced gpl20, or gpl40 molecule, or a portion thereof, such as a gpl20 outer domain (OD) that is capable of inducing an immune response in a mammal, such as a mammal with or without an HIV infection. Administration of an immunogenic gpl20 or gpl40 polypeptide that induces an immune response can lead to protective immunity against HIV. In some examples, an immunogenic gpl20 or gpl40 polypeptide is a disclosed antigen that is resurfaced to focus immunogenicity to a target epitope.

Immunogenic surface: A surface of a molecule, for example a protein such as gpl20 or gpl40 protein or polypeptide, capable of eliciting an immune response. An immunogenic surface includes the defining features of that surface, for example the three-dimensional shape and the surface charge. In some examples, an immunogenic surface is defined by the amino acids on the surface of a protein or peptide that are in contact with an antibody, such as a neutralizing antibody, when the protein and the antibody are bound together. A target epitope includes an immunogenic surface. Immunogenic surface is synonymous with antigenic surface.

Immune response: A response of a cell of the immune system, such as a B cell, T cell, or monocyte, to a stimulus. In one embodiment, the response is specific for a particular antigen (an "antigen-specific response"). In one embodiment, an immune response is a T cell response, such as a CD4+ response or a CD8+ response. In another embodiment, the response is a B cell response, and results in the production of specific antibodies.

Immunogenic composition: A composition comprising an immunogenic peptide that induces a measurable CTL response against virus expressing the immunogenic peptide, or induces a measurable B cell response (such as production of antibodies) against the immunogenic peptide. In one example, an "immunogenic composition" is composition includes a disclosed antigen derived from a gpl20, or gpl40 peptide or an antigenic fragment thereof, such as a gpl20 outer domain (OD) peptide that induces a measurable CTL response against virus expressing gpl20

polypeptide, or induces a measurable B cell response (such as production of antibodies) against a gpl20, or gpl40 polypeptide. It further refers to isolated nucleic acids encoding an antigen, such as a nucleic acid that can be used to express the antigen (and thus be used to elicit an immune response against this polypeptide).

5 For *in vitro* use, an immunogenic composition may consist of the isolated protein, peptide epitope, or nucleic acid encoding the protein, or peptide epitope. For *in vivo* use, the immunogenic composition will typically include the protein, immunogenic peptide or nucleic acid in pharmaceutically acceptable carriers, and/or other agents. Any particular peptide, such as a disclosed antigen or a nucleic acid
10 encoding the antigen, can be readily tested for its ability to induce a CTL or B cell response by art-recognized assays. Immunogenic compositions can include adjuvants, which are well known to one of skill in the art.

Inhibiting or treating a disease: Inhibiting the full development of a disease or condition, for example, in a subject who is at risk for a disease such as
15 acquired immune deficiency syndrome (AIDS), AIDS related conditions, HIV-1 infection, or combinations thereof. "Treatment" refers to a therapeutic intervention that ameliorates a sign or symptom of a disease or pathological condition after it has begun to develop. The term "ameliorating," with reference to a disease or pathological condition, refers to any observable beneficial effect of the treatment.
20 The beneficial effect can be evidenced, for example, by a delayed onset of clinical symptoms of the disease in a susceptible subject, a reduction in severity of some or all clinical symptoms of the disease, a slower progression of the disease, a reduction in the number of metastases, an improvement in the overall health or well- being of the subject, or by other parameters well known in the art that are specific to the
25 particular disease. A "prophylactic" treatment is a treatment administered to a subject who does not exhibit signs of a disease or exhibits only early signs for the purpose of decreasing the risk of developing pathology.

Isolated: An "isolated" biological component (such as a protein, for example a disclosed antigen or nucleic acid encoding such an antigen) has been substantially
30 separated or purified away from other biological components in which the component naturally occurs, such as other chromosomal and extrachromosomal

DNA, RNA, and proteins. Proteins, peptides and nucleic acids that have been "isolated" include proteins purified by standard purification methods. The term also embraces proteins or peptides prepared by recombinant expression in a host cell as well as chemically synthesized proteins, peptides and nucleic acid molecules.

- 5 Isolated does not require absolute purity, and can include protein, peptide, or nucleic acid molecules that are at least 50% isolated, such as at least 75%, 80%, 90%, 95%, 98%, 99%, or even 99.9% isolated.

Label: A detectable compound or composition that is conjugated directly or indirectly to another molecule to facilitate detection of that molecule. Specific, non-
10 limiting examples of labels include fluorescent tags, enzymatic linkages, and radioactive isotopes. In some examples, a disclosed antigen is labeled with a detectable label. In some examples, label is attached to a disclosed antigen or nucleic acid encoding such an antigen.

Nucleic acid: A polymer composed of nucleotide units (ribonucleotides, deoxyribonucleotides, related naturally occurring structural variants, and synthetic
15 non-naturally occurring analogs thereof) linked via phosphodiester bonds, related naturally occurring structural variants, and synthetic non-naturally occurring analogs thereof. Thus, the term includes nucleotide polymers in which the nucleotides and the linkages between them include non-naturally occurring synthetic analogs, such
20 as, for example and without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-**O**-methyl ribonucleotides, peptide-nucleic acids (PNAs), and the like. Such polynucleotides can be synthesized, for example, using an automated DNA synthesizer. The term "oligonucleotide" typically refers to short polynucleotides, generally no greater than
25 about 50 nucleotides. It will be understood that when a nucleotide sequence is represented by a DNA sequence (*i.e.*, A, T, G, C), this also includes an RNA sequence (*i.e.*, A, U, G, C) in which "U" replaces "T."

"Nucleotide" includes, but is not limited to, a monomer that includes a base linked to a sugar, such as a pyrimidine, purine or synthetic analogs thereof, or a base
30 linked to an amino acid, as in a peptide nucleic acid (PNA). A nucleotide is one

monomer in a polynucleotide. A nucleotide sequence refers to the sequence of bases in a polynucleotide.

Conventional notation is used herein to describe nucleotide sequences: the left-hand end of a single-stranded nucleotide sequence is the 5'-end; the left-hand
5 direction of a double-stranded nucleotide sequence is referred to as the 5'-direction. The direction of 5' to 3' addition of nucleotides to nascent RNA transcripts is referred to as the transcription direction. The DNA strand having the same sequence as an mRNA is referred to as the "coding strand;" sequences on the DNA strand having the same sequence as an mRNA transcribed from that DNA and which are
10 located 5' to the 5'-end of the RNA transcript are referred to as "upstream sequences;" sequences on the DNA strand having the same sequence as the RNA and which are 3' to the 3' end of the coding RNA transcript are referred to as "downstream sequences."

"cDNA" refers to a DNA that is complementary or identical to an mRNA, in
15 either single stranded or double stranded form.

"Encoding" refers to the inherent property of specific sequences of nucleotides in a polynucleotide, such as a gene, a cDNA, or an mRNA, to serve as templates for synthesis of other polymers and macromolecules in biological processes having either a defined sequence of nucleotides (for example, rRNA,
20 tRNA and mRNA) or a defined sequence of amino acids and the biological properties resulting therefrom. Thus, a gene encodes a protein if transcription and translation of mRNA produced by that gene produces the protein in a cell or other biological system. Both the coding strand, the nucleotide sequence of which is identical to the mRNA sequence and is usually provided in sequence listings, and
25 non-coding strand, used as the template for transcription, of a gene or cDNA can be referred to as encoding the protein or other product of that gene or cDNA. Unless otherwise specified, a "nucleotide sequence encoding an amino acid sequence" includes all nucleotide sequences that are degenerate versions of each other and that encode the same amino acid sequence. Nucleotide sequences that encode proteins
30 and RNA may include introns. In some examples, a nucleic acid encodes a disclosed antigen.

"Recombinant nucleic acid" refers to a nucleic acid having nucleotide sequences that are not naturally joined together. This includes nucleic acid vectors comprising an amplified or assembled nucleic acid which can be used to transform a suitable host cell. A host cell that comprises the recombinant nucleic acid is referred to as a "recombinant host cell." The gene is then expressed in the recombinant host cell to produce, such as a "recombinant polypeptide." A recombinant nucleic acid may serve a non-coding function (such as a promoter, origin of replication, ribosome-binding site, etc.) as well.

Operably linked: A first nucleic acid sequence is operably linked with a second nucleic acid sequence when the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences are contiguous and, where necessary to join two protein-coding regions, in the same reading frame.

Peptide modifications: Peptides, such as the HIV immunogens disclosed herein can be modified by a variety of chemical techniques to produce derivatives having essentially the same activity as the unmodified peptides, and optionally having other desirable properties. For example, carboxylic acid groups of the protein, whether carboxyl-terminal or side chain, may be provided in the form of a salt of a pharmaceutically-acceptable cation or esterified to form a C_1 - C_{16} ester, or converted to an amide of formula NR_1R_2 wherein R_1 and R_2 are each independently H or CrC_{16} alkyl, or combined to form a heterocyclic ring, such as a 5- or 6-membered ring. Amino groups of the peptide, whether amino-terminal or side chain, may be in the form of a pharmaceutically-acceptable acid addition salt, such as the HCl , HBr , acetic, benzoic, toluene sulfonic, maleic, tartaric and other organic salts, or may be modified to CrC_{16} alkyl or dialkyl amino or further converted to an amide.

Hydroxyl groups of the peptide side chains can be converted to C_1 - C_{16} alkoxy or to a C_1 - C_{16} ester using well-recognized techniques. Phenyl and phenolic rings of the peptide side chains can be substituted with one or more halogen atoms,

such as F, Cl, Br or I, or with C₁-C₁₆ alkyl, C₁-C₁₆ alkoxy, carboxylic acids and esters thereof, or amides of such carboxylic acids. Methylene groups of the peptide side chains can be extended to homologous C₂-C₄ alkylenes. Thiols can be protected with any one of a number of well-recognized protecting groups, such as acetamide
5 groups. Those skilled in the art will also recognize methods for introducing cyclic structures into the peptides of this disclosure to select and provide conformational constraints to the structure that result in enhanced stability. For example, a C- or N-terminal cysteine can be added to the peptide, so that when oxidized the peptide will contain a disulfide bond, generating a cyclic peptide. Other peptide cyclizing
10 methods include the formation of thioethers and carboxyl- and amino-terminal amides and esters.

Peptide: Any compound composed of amino acids, amino acid analogs, chemically bound together. Peptide as used herein includes oligomers of amino acids, amino acid analog, or small and large peptides, including polypeptides or
15 proteins. Any chain of amino acids, regardless of length or post-translational modification (such as glycosylation or phosphorylation). "Peptide" applies to amino acid polymers to naturally occurring amino acid polymers and non-naturally occurring amino acid polymer as well as in which one or more amino acid residue is a non-natural amino acid, for example a artificial chemical mimetic of a
20 corresponding naturally occurring amino acid. A "residue" refers to an amino acid or amino acid mimetic incorporated in a polypeptide by an amide bond or amide bond mimetic. A peptide has an amino terminal (N-terminal) end and a carboxy terminal (C-terminal) end.

A "**protein**" is a peptide that folds into a specific three-dimensional structure.
25 A protein can include surface-exposed amino acid residues and non-surface-exposed amino acid residues. "**Surface-exposed amino acid residues**" are those amino acids that have some degree of exposure on the surface of the protein, for example such that they can contact the solvent when the protein is in solution. In contrast, non-surface-exposed amino acids are those amino acid residues that are not exposed on the surface
30 of the protein, such that they do not contact solution when the protein is in solution. In

some examples, the non-surface-exposed amino acid residues are part of the protein core.

A "**protein core**" is the interior of a folded protein, which is substantially free of solvent exposure, such as solvent in the form of water molecules in solution.

- 5 Typically, the protein core is predominately composed of hydrophobic or apolar amino acids. In some examples, a protein core may contain charged amino acids, for example aspartic acid, glutamic acid, arginine, and/or lysine. The inclusion of uncompensated charged amino acids (a compensated charged amino can be in the form of a salt bridge) in the protein core can lead to a destabilized protein. That is, a
- 10 protein with a lower T_m than a similar protein without an uncompensated charged amino acid in the protein core. In other examples, a protein core may have a cavity within the protein core. Cavities are essentially voids within a folded protein where amino acids or amino acid side chains are not present. Such cavities can also destabilize a protein relative to a similar protein without a cavity. Thus, when creating
- 15 a stabilized form of a protein, it may be advantageous to substitute amino acid residues within the core in order to fill cavities present in the wild-type protein.

Amino acids in a peptide, polypeptide or protein generally are chemically bound together via amide linkages (CONH). Additionally, amino acids may be bound together by other chemical bonds. For example, linkages for amino acids or

20 amino acid analogs can include $\text{CH}_2\text{NH-}$, $-\text{CH}_2\text{S-}$, $-\text{CH}_2\text{-CH}_2-$, $-\text{CH=CH-}$ (cis and trans), $-\text{COCH}_2-$, $-\text{CH(OH)CH}_2-$, and $-\text{CHH}_2\text{SO-}$ (These and others can be found in Spatola, in Chemistry and Biochemistry of Amino Acids, Peptides, and Proteins, B. Weinstein, eds., Marcel Dekker, New York, p. 267 (1983); Spatola, A. F., Vega Data (March 1983), Vol. 1, Issue 3, Peptide Backbone Modifications (general

25 review); Morley, Trends Pharm Sci pp. 463-468, 1980; Hudson, *et al.*, *Int J Pept ProtRes* 14:177-185, 1979; Spatola *et al.* *Life Sci* 38:1243-1249, 1986; Harm / . *Chem. Soc Perkin Trans.* 1307-314, 1982; Almquist *et al.* *J. Med. Chem.* 23:1392-1398, 1980; Jennings-White *et al.* *Tetrahedron Lett* 23:2533, 1982; Holladay *et al.* *Tetrahedron. Lett* 24:4401-4404, 1983; and Hruby *Life Sci* 31:189-199, 1982.

Pharmaceutical agent or drug: A chemical compound or composition capable of inducing a desired therapeutic or prophylactic effect when properly administered to a subject.

Pharmaceutically acceptable carriers: The pharmaceutically acceptable carriers useful in this disclosure are conventional. *Remington's Pharmaceutical Sciences*, by E. W. Martin, Mack Publishing Co., Easton, PA, 19th Edition (1995), describes compositions and formulations suitable for pharmaceutical delivery of the proteins and other compositions herein disclosed.

In general, the nature of the carrier will depend on the particular mode of administration being employed. For instance, parenteral formulations usually comprise injectable fluids that include pharmaceutically and physiologically acceptable fluids such as water, physiological saline, balanced salt solutions, aqueous dextrose, glycerol or the like as a vehicle. For solid compositions, powder, pill, tablet, or capsule forms, conventional non-toxic solid carriers can include, for example, pharmaceutical grades of mannitol, lactose, starch, or magnesium stearate. In addition to biologically-neutral carriers, pharmaceutical compositions to be administered can contain minor amounts of non-toxic auxiliary substances, such as wetting or emulsifying agents, preservatives, and pH buffering agents and the like, for example sodium acetate or sorbitan monolaurate.

Purified: The term purified does not require absolute purity; rather, it is intended as a relative term. Thus, for example, a purified protein is one in which the protein is more enriched than the protein is in its natural environment within a cell. Preferably, a preparation is purified such that the protein represents at least 50% of the protein content of the preparation.

The resurfaced immunogens disclosed herein, or antibodies that specifically bind the disclosed resurfaced immunogens, can be purified by any of the means known in the art. See for example *Guide to Protein Purification*, ed. Deutscher, *Meth. Enzymol.* 185, Academic Press, San Diego, 1990; and Scopes, *Protein Purification: Principles and Practice*, Springer Verlag, New York, 1982. Substantial purification denotes purification from other proteins or cellular components. A substantially purified protein is at least 60%, 70%, 80%, 90%, 95% or 98% pure.

Thus, in one specific, non-limiting example, a substantially purified protein is 90% free of other proteins or cellular components.

Resurfaced antigen or resurfaced immunogen: A polypeptide immunogen derived from a wild-type antigen in which amino acid residues outside or exterior to a target epitope are mutated in a systematic way to focus the immunogenicity of the antigen to the selected target epitope. In some examples a resurfaced antigen is referred to as an antigenically-cloaked immunogen or antigenically-cloaked antigen.

Sequence identity/similarity: The identity/similarity between two or more nucleic acid sequences, or two or more amino acid sequences, is expressed in terms of the identity or similarity between the sequences. Sequence identity can be measured in terms of percentage identity; the higher the percentage, the more identical the sequences are. Homologs or orthologs of nucleic acid or amino acid sequences possess a relatively high degree of sequence identity/similarity when aligned using standard methods.

Methods of alignment of sequences for comparison are well known in the art. Various programs and alignment algorithms are described in: Smith & Waterman, *Adv. Appl. Math.* 2:482, 1981; Needleman & Wunsch, *J. Mol. Biol.* 48:443, 1970; Pearson & Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444, 1988; Higgins & Sharp, *Gene*, 73:237-44, 1988; Higgins & Sharp, *CABIOS* 5:151-3, 1989; Corpet *et al.*, *Nuc. Acids Res.* 16:10881-90, 1988; Huang *et al.* *Computer Appls. in the Biosciences* 8, 155-65, 1992; and Pearson *et al.*, *Meth. Mol. Bio.* 24:307-31, 1994. Altschul *et al.*, *J. Mol. Biol.* 215:403-10, 1990, presents a detailed consideration of sequence alignment methods and homology calculations.

Once aligned, the number of matches is determined by counting the number of positions where an identical nucleotide or amino acid residue is present in both sequences. The percent sequence identity is determined by dividing the number of matches either by the length of the sequence set forth in the identified sequence, or by an articulated length (such as 100 consecutive nucleotides or amino acid residues from a sequence set forth in an identified sequence), followed by multiplying the resulting value by 100. For example, a peptide sequence that has 1166 matches when aligned with a test sequence having 1554 nucleotides is 75.0 percent identical

to the test sequence ($1166 \div 1554 * 100 = 75.0$). The percent sequence identity value is rounded to the nearest tenth. For example, 75.11, 75.12, 75.13, and 75.14 are rounded down to 75.1, while 75.15, 75.16, 75.17, 75.18, and 75.19 are rounded up to 75.2. The length value will always be an integer.

5 For sequence comparison of nucleic acid sequences and amino acids sequences, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are
10 designated. Default program parameters are used. Methods of alignment of sequences for comparison are well known in the art. Optimal alignment of sequences for comparison can be conducted, for example, by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482, 1981, by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443, 1970, by the
15 search for similarity method of Pearson & Lipman, *Proc. Nat'l Acad. Sci. USA* 85:2444, 1988, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection (see for example, *Current Protocols in Molecular Biology*
20 (Ausubel *et al.*, eds 1995 supplement)). The NCBI Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, *J. Mol. Biol.* 215:403-10, 1990) is available from several sources, including the National Center for Biological Information (NCBI, National Library of Medicine, Building 38A, Room 8N805, Bethesda, MD 20894) and on the Internet, for use in connection with the sequence analysis programs
25 blastp, blastn, blastx, tblastn, and tblastx. Blastn is used to compare nucleic acid sequences, while blastp is used to compare amino acid sequences. Additional information can be found at the NCBI web site.

Another example of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and the BLAST 2.0
30 algorithm, which are described in Altschul *et al.*, *J. Mol. Biol.* 215:403-410, 1990 and Altschul *et al.*, *Nucleic Acids Res.* 25:3389-3402, 1977. Software for performing

BLAST analyses is publicly available through the National Center for Biotechnology Information (World Wide Web address ncbi.nlm.nih.gov). The BLASTN program (for nucleotide sequences) uses as defaults a word length (W) of 11, alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands. The BLASTP program (for amino acid sequences) uses as defaults a word length (W) of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915, 1989).

Another indicia of sequence similarity between two nucleic acids is the ability to hybridize. The more similar are the sequences of the two nucleic acids, the more stringent the conditions at which they will hybridize. The stringency of hybridization conditions are sequence-dependent and are different under different environmental parameters. Thus, hybridization conditions resulting in particular degrees of stringency will vary depending upon the nature of the hybridization method of choice and the composition and length of the hybridizing nucleic acid sequences. Generally, the temperature of hybridization and the ionic strength (especially the Na⁺ and/or Mg⁺⁺ concentration) of the hybridization buffer will determine the stringency of hybridization, though wash times also influence stringency. Generally, stringent conditions are selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Conditions for nucleic acid hybridization and calculation of stringencies can be found, for example, in Sambrook *et al*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001; Tijssen, *Hybridization With Nucleic Acid Probes, Part I: Theory and Nucleic Acid Preparation*, Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Ltd., NY, NY, 1993. and Ausubel *et al. Short Protocols in Molecular Biology*, 4th ed., John Wiley & Sons, Inc., 1999.

"Stringent conditions" encompass conditions under which hybridization will only occur if there is less than 25% mismatch between the hybridization molecule and the target sequence. "Stringent conditions" may be broken down into particular

levels of stringency for more precise definition. Thus, as used herein, "moderate stringency" conditions are those under which molecules with more than 25% sequence mismatch will not hybridize; conditions of "medium stringency" are those under which molecules with more than 15% mismatch will not hybridize, and
5 conditions of "high stringency" are those under which sequences with more than 10% mismatch will not hybridize. Conditions of "very high stringency" are those under which sequences with more than 6% mismatch will not hybridize. In contrast nucleic acids that hybridize under "low stringency conditions include those with much less sequence identity, or with sequence identity over only short subsequences
10 of the nucleic acid.

T Cell: A white blood cell critical to the immune response. T cells include, but are not limited to, CD4⁺ T cells and CD8⁺ T cells. A CD4⁺ T lymphocyte is an immune cell that carries a marker on its surface known as "cluster of differentiation 4" (CD4). These cells, also known as helper T cells, help orchestrate the immune
15 response, including antibody responses as well as killer T cell responses. CD8⁺ T cells carry the "cluster of differentiation 8" (CD8) marker. In one embodiment, a CD8 T cells is a cytotoxic T lymphocytes. In another embodiment, a CD8 cell is a suppressor T cell.

Therapeutic agent: A chemical compound, small molecule, or other
20 composition, such as nucleic acid molecule, capable of inducing a desired therapeutic or prophylactic effect when properly administered to a subject.

Therapeutically effective amount or Effective amount: The amount of agent, such as nucleic acid vaccine or other therapeutic agent, that is sufficient to prevent, treat (including prophylaxis), reduce and/or ameliorate the symptoms
25 and/or underlying causes of any of a disorder or disease, for example to prevent, inhibit, and/or treat HIV. In some embodiments, an "effective amount" is sufficient to reduce or eliminate a symptom of a disease, such as AIDS. For instance, this can be the amount necessary to inhibit viral replication or to measurably alter outward symptoms of the viral infection, such as increase of T cell counts in the case of an
30 HIV-1 infection. In general, this amount will be sufficient to measurably inhibit virus (for example, HIV) replication or infectivity. When administered to a subject,

a dosage will generally be used that will achieve target tissue concentrations (for example, in lymphocytes) that has been shown to achieve in vitro inhibition of viral replication. An "anti-viral agent" or "anti-viral drug" is an agent that specifically inhibits a virus from replicating or infecting cells. Similarly, an "anti-retroviral agent" is an agent that specifically inhibits a retrovirus from replicating or infecting cells.

Transformed: A transformed cell is a cell into which has been introduced a nucleic acid molecule by molecular biology techniques. As used herein, the term transformation encompasses all techniques by which a nucleic acid molecule might be introduced into such a cell, including transfection with viral vectors, transformation with plasmid vectors, and introduction of DNA by electroporation, lipofection, and particle gun acceleration.

Vaccine: A pharmaceutical composition that elicits a prophylactic or therapeutic immune response in a subject. In some cases, the immune response is a protective immune response. Typically, a vaccine elicits an antigen-specific immune response to an antigen of a pathogen, for example a viral pathogen, or to a cellular constituent correlated with a pathological condition. A vaccine may include a polynucleotide (such as a nucleic acid encoding a disclosed antigen), a peptide or polypeptide (such as a disclosed antigen), a virus, a cell or one or more cellular constituents.

Vector: A nucleic acid molecule as introduced into a host cell, thereby producing a transformed host cell. Recombinant DNA vectors are vectors having recombinant DNA. A vector can include nucleic acid sequences that permit it to replicate in a host cell, such as an origin of replication. A vector can also include one or more selectable marker genes and other genetic elements known in the art. Viral vectors are recombinant DNA vectors having at least some nucleic acid sequences derived from one or more viruses.

Virus: A virus consists essentially of a core of nucleic acid surrounded by a protein coat, and has the ability to replicate only inside a living cell. "Viral replication" is the production of additional virus by the occurrence of at least one viral life cycle. A virus may subvert the host cells' normal functions, causing the cell

to behave in a manner determined by the virus. For example, a viral infection may result in a cell producing a cytokine, or responding to a cytokine, when the uninfected cell does not normally do so. In some examples, a virus is a pathogen.

"Retroviruses" are RNA viruses wherein the viral genome is RNA. When a host cell is infected with a retrovirus, the genomic RNA is reverse transcribed into a DNA intermediate which is integrated very efficiently into the chromosomal DNA of infected cells. The integrated DNA intermediate is referred to as a provirus. The term "lentivirus" is used in its conventional sense to describe a genus of viruses containing reverse transcriptase. The lentiviruses include the "immunodeficiency viruses" which include human immunodeficiency virus (HIV) type 1 and type 2 (HIV-1 and HIV-2), simian immunodeficiency virus (SIV), and feline immunodeficiency virus (FIV).

HIV-1 is a retrovirus that causes immunosuppression in humans (HIV disease), and leads to a disease complex known as the acquired immunodeficiency syndrome (AIDS). "HIV disease" refers to a well-recognized constellation of signs and symptoms (including the development of opportunistic infections) in persons who are infected by an HIV virus, as determined by antibody or western blot studies. Laboratory findings associated with this disease are a progressive decline in T cells.

Virus-like particle (VLP): A non-replicating, viral shell, derived from any of several viruses. VLPs are generally composed of one or more viral proteins, such as, but not limited to, those proteins referred to as capsid, coat, shell, surface and/or envelope proteins, or particle-forming polypeptides derived from these proteins. VLPs can form spontaneously upon recombinant expression of the protein in an appropriate expression system. Methods for producing particular VLPs are known in the art. The presence of VLPs following recombinant expression of viral proteins can be detected using conventional techniques known in the art, such as by electron microscopy, biophysical characterization, and the like. See, for example, Baker *et al.* (1991) *Biophys. J.* 60:1445-1456; and Hagensee *et al.* (1994) *J. Virol.* 68:4503-4505. For example, VLPs can be isolated by density gradient centrifugation and/or identified by characteristic density banding. Alternatively, cryoelectron microscopy

can be performed on vitrified aqueous samples of the VLP preparation in question, and images recorded under appropriate exposure conditions.

Wild-type antigen: An antigen that has not been modified by selective mutation to focus that antigenicity of the antigen to a target epitope.

5

//. ***Description of Several Embodiments***

As the human immunodeficiency virus (HIV) pandemic continues to infect millions of people each year, the need for an effective vaccine increases. However the development of such a vaccine has been stymied due to the difficulty in
10 developing an immunogen capable of eliciting broadly neutralizing antibodies. The current disclosure meets these needs.

Because many epitopes on HIV are derived from different and non-contiguous regions of the primary amino acid sequence of the antigen, it has proven difficult to synthetically reproduce these complex epitopes. Thus, an effective
15 immunogen and vaccine will likely need to be derived from a wild type HIV antigen.

One of the major hurdles to the construction of such an antigen is focusing the immune response to regions of HIV proteins which mostly produce broadly neutralizing antibodies. As disclosed herein, the inventors have constructed a series
20 of HIV immunogens based on the HIV glycoprotein gp120 that focus the immune response to the CD4 binding site of gp120 by a using combination of antigen resurfacing and/or paring back the structural elements of gp120 that do not contribute to elicitation of broadly neutralizing antibodies. These appropriately constructed immunogens retain the ability to stimulate the production of neutralizing
25 antibodies directed to a target epitope, while eliminating extraneous epitopes on gp120 that do not contribute to the elicitation of neutralizing epitopes. Such molecules have utility as both potential vaccines for HIV and as diagnostic molecules (for example, to detect and quantify target antibodies in a polyclonal serum response).

30

A. **Antigens**

Provided herein in various embodiments are antigens derived from HIV glycoproteins, which are useful to induce immunogenic responses in vertebrate animals (such as mammals, for example primates, such as humans) to HIV (for
5 example HIV-1 and HIV-2). In some embodiments, the antigen is an HIV antigen, such as a modified gpl20, or gpl40 or an immunogenic fragment thereof. In specific embodiments, the disclosed antigen is HIV-1 gpl20 or an immunogenic fragment thereof, for example, the outer domain (OD) of gpl20.

The disclosed antigens have been substantially resurfaced from the wild type
10 sequence, such that the surface of the antigen has been altered to focus the immune response to a particular feature, or epitope on the surface of gpl20 or gpl40. In some embodiments, the disclosed antigen is a resurfaced gpl20 antigen in which the one or more of the VI, V2, V3, V4 and/or V5 variable loops from gpl20 or gpl40 or an immunogenic fragment, thereof such as a gpl20 outer domain, are removed or
15 truncated. In some examples, the disclosed antigens have been modified to substitute the surface-exposed amino acids located exterior to the target epitope to focus the antigenicity of the antigen to the target epitope. For example, the method can remove non-target epitopes that might interfere with specific binding of an antibody to the target epitope. In some examples, the amino acid substitutions result in the
20 antigen not being bound by antibodies in a polyclonal serum that specifically bind surface-exposed amino acid residues of the wild-type antigen located exterior of the target epitope. In some embodiments, the amino acid substitutions alter antigenicity of the antigen *in vivo* as compared to the wild-type antigen (unsubstituted antigen), but do not introduce additional glycosylation sites as compared to the wild-type
25 antigen. In some embodiments, that antigen is glycosylated. Examples of antigen resurfacing methods are given in PCT Publication No. WO 09/100376, which is specifically incorporated by reference in its entirety.

HIV-I can be classified into four groups: the "major" group M, the "outlier" group O, group N, and group P. Within group M, there are several genetically
30 distinct clades (or subtypes) of HIV-I. The disclosed immunogens can be derived from any subtype of HIV, such as groups M, N, O, or P or clade A, B, C, D, F, G, H,

J or K and the like. The nucleic acid sequence can encode an Env polypeptide from any group or clade of HIV. HIV gpl20 and gpl40 proteins from the different HIV clades, as well as nucleic acid sequences encoding such proteins and methods for the manipulation and insertion of such nucleic acid sequences into vectors, are known
5 (see, e.g., HIV Sequence Compendium, Division of AIDS, National Institute of Allergy and Infectious Diseases (2003); HIV Sequence Database (hiv-web.lanl.gov/content/hiv-db/mainpage.html); Sambrook et al., Molecular Cloning, a Laboratory Manual, 2d edition, Cold Spring Harbor Press, Cold Spring Harbor, N. Y. (1989); Ausubel et al., Current Protocols in Molecular Biology, Greene
10 Publishing Associates and John Wiley & Sons, New York, N. Y. (1994)).

In some examples, the disclosed antigen is a gpl20 from a HIV-1 Clade A virus. In some examples, the HIV-1 Clade A virus is KER2018, such that the disclosed immunogen is a gpl20 or immunogenic fragment thereof from HIV-1 KER2018. In some examples, the disclosed antigen is a gpl20 from a HIV-1 Clade
15 C virus. In some examples, the HIV-1 Clade C virus is DU156, such that the disclosed immunogen is a gpl20 or immunogenic fragment thereof from HIV-1 DU156. In some examples, the HIV-1 Clade C virus is ZA012, such that the disclosed immunogen is a gpl20 or immunogenic fragment thereof from HIV-1 ZA012.

20 In some embodiments, a disclosed antigen includes the outer domain of a gpl20.

In some examples, the outer domain of gpl20 includes amino acid residues 252-482 of gpl20. In some examples, the outer domain of gpl20 includes amino acid residues 213-492 of gpl20. In some examples, the outer domain of gpl20
25 includes amino acid residues 213-482 of gpl20. In some examples, the outer domain of gpl20 includes amino acid residues 252-492 of gpl20. In specific embodiments, the disclosed antigen is HIV-1 gpl20. In other specific embodiments, the disclosed antigen is HIV-1 gpl40.

In some examples, V1/V2 and β 20/ β 21 regions of gpl20 or gpl40 are
30 modified to reduce the immunogenicity or at least alter the immunogenicity of those regions. In some examples, the β 20/ β 21 bridging sheet of the gpl20 or gpl40

antigen or immunogenic fragment thereof is removed by replacing the amino acid residues between 1423 and Y435 with Gly-Gly. In some examples, the β 20/ β 21 bridging sheet of the antigen is removed, for example by replacing residues 422-436 of gpl20 or gpl40 or the outer domain of gpl20 with Gly-Gly. In some examples, residues 128-194 of gpl20 in the VI, V2 loop region are replaced by Gly-Arg-Gly. In some examples, a modified V1/V2 is taken from a core gpl20 previously designed that has improved expression yields known as "new 9c" (see International Patent Publication NO. WO 2007/030518, which is incorporated herein by reference), and includes the insertion of VKLTPLAGATSVITQA (SEQ ID NO: 162) between C119 and C205.

In some examples, the wild type VI loop (residues 131-157) includes the amino acid sequence CTDLRNATNTTSSSWETMEKGEIKNC (SEQ ID NO: 163). In some examples, the VI loop is replaced with the amino acid sequence CTDLRGSGGGSGGGSEIKNC (SEQ ID NO: 164), which in some examples is designated AVI. 1.

In some examples, the wild type V2 loop (residues 157-198) includes the amino acid sequence CSFNITTSIRDKVQKEYALFYKLDVVPIDNDNTSYRLINC (SEQ ID NO: 165). In some examples, the V2 loop is replaced with the amino acid sequence CSFNITTSIRDKVQKEYALFYKLDVVPIGSGGSGSYRLINC (SEQ ID NO: 166), which in some examples is designated V2.1, or the amino acid sequence CSFNITTSIRDKVQKEYALFYKLDVVPIGSGGSGSYRLINC (SEQ ID NO: 167), which in some examples is designated V2.2.

In some embodiments, the V3 loop of a disclosed antigen is mutated and/or truncated as compared to a wild type gpl20, gpl40 or immunogenic fragment thereof. In some examples, the wild-type V3 loop (residues 296-331) includes the amino acid sequence CSRPNNTNRKSIPMGPGRAFYTGTGQIIGDIRQAHC (SEQ ID NO: 168). In some examples, the V3 loop is shortened by 18 amino acids, for example by removing 9 amino acids from each end of the loop, which in certain examples is designated V3-9,9. In some examples, residues 302-323 of gpl20, part of the V3 loop, are replaced with a basic hexapeptide (NTRGRR; SEQ ID NO: 169). In some examples, residues 305-323 of gpl20 in the V3 loop are replaced by Gly-

Arg-Arg. In some examples, the V3 loop is replaced with the m5 loop, having the amino acid sequence CSRPNNNTRGRRGSSGGSHC (SEQ ID NO: 170). In some examples, the V3 loop is replaced with the m4 loop, having the amino acid sequence CSRPNNGGSGSGGSSGGSHC (SEQ ID NO: 171). In some examples, the V3 loop is replaced with Gly-Val-Gly. In some examples, the V3 loop is at least partially replaced with Gly-Ser. In some examples, the V3 loop is at least partially replaced with Gly-Ser-Leu. In some examples, the V3 loop is truncated to an 11-mer. In some examples, the V3 loop is replaced with the amino acid sequence CARPSNNTRGRRGDIRQAYC (SEQ ID NO: 172). In some examples, the V3 loop is replaced with the amino acid sequence CARPSNNTDIRQAYC (SEQ ID NO: 173), which in some examples is designated V3.1. In some examples, the V3 loop is replaced with the amino acid sequence CARPSNNTRQAYC (SEQ ID NO: 174), which in some examples is designated V3.2. In some examples, the V3 loop is replaced with the amino acid sequence CARPSNNTQYC (SEQ ID NO: 175), which is some examples is designated V3.3. In some examples, the V3 loop is replaced with the amino acid sequence CARGSGSGSYC (SEQ ID NO: 176), which in some examples is designated V3.4. In some examples, the V3 loop is replaced with the amino acid sequence CSRPNNNTRGRRGDIRQAH C (SEQ ID NO: 177), which in some examples is designated V3(GSL). In some examples, the V3 loop is replaced with the amino acid sequence CSRPNNNTRRQAH C (SEQ ID NO: 178), which in some examples is designated V3.2. In some examples, the V3 loop is replaced with the amino acid sequence CSRPNNGGSGQAH C (SEQ ID NO: 179), which in some examples is designated V3.2GS. In some examples, the wild-type V3 loop includes the amino acid sequence CTRPNNNTRKSIHIGPGQAFYATGDIIGDIRQAH C (SEQ ID NO: 205). In some examples, the V3 loop is replaced with the amino acid sequence CTRPNNGGSGSGGSSGGSHC (SEQ ID NO: 180), which in some examples is designated V3.4. In some examples, the V3 loop is replaced with the amino acid sequence CTRPNNNTRGRRGSSGGSHC (SEQ ID NO: 181), which in some examples is designated V3.5. In some examples, the V3 loop is replaced with a 15 mer with a native glycan at the tip. In some examples, the V3 loop is replaced with a 5 mer with a slightly shifted glycan at the tip.

In some embodiments, the V4 loop of a disclosed antigen is mutated and/or truncated as compared to a wild type gpl20, gpl40 or immunogenic fragment thereof. In some examples, the V4 loop is at least partially replaced with Gly-Ser. In some examples, 9 amino acids are removed from the V4 loop, which in some examples is designated V4.2.1. In some examples, the V4 loop is shortened to SIWNNGGSGGGSGGGSDTIT (SEQ ID NO: 182), which in some examples is designated V4.GS. In some examples, the V4 loop is replaced with the V4 loop from strain Ker2018 (a clade A strain), which in some examples is designated V4.Ker A. In some examples, one of the two glycan sites is removed from the V4 loop from strain Ker2018, which in some examples is designated V4.Ker A/AG2, or V4.Ker A/AG1. In some examples, both of the two glycan sites are removed from the V4 loop from strain Ker2018r, which in some examples is designated V4.Ker A/AG2, or V4.Ker A/AG12. In some examples, the wild-type V4 loop includes the amino acid sequence STWFNSTWSTKGSNNTEGSDTIT (SEQ ID NO: 183). In some examples, the V4 loop is replaced with the amino acid sequence STWFNGSGSGSGGTIT (SEQ ID NO: 184), which in some examples is designated V4.1. In some examples, the V4 loop is replaced with the amino acid sequence STWFNSTWSTKGSNNTEGSDTIT (SEQ ID NO: 185), which in some examples is designated V4.2. In some examples, the V4 loop is replaced with the amino acid sequence STWFQSGSGSGGTIT (SEQ ID NO: 186), which in some examples is designated V4.3. In some examples, the glycan site at the N terminal end of the V4 is removed, which is designated the V4.7 loop. In some examples, the glycan site at the C terminal end of the V4 is removed, which is designated v4.8. In some examples, the glycan site at both the N and C terminal end is removed, which is designated v4.9. In some examples, the V4 loop is replaced with a Gly-Ser repeat, which is designated V4.x

In some embodiments, the V5 loop of a disclosed antigen is mutated and/or truncated as compared to a wild type gpl20, gpl40 or immunogenic fragment thereof. In some examples, the V5 loop is at least partially replaced with Gly-Ser. In some examples, a wild type V5 loop has the amino acid sequence GGNTGNNSRTC (SEQ ID NO: 202). In some examples, the V5 loop is truncated to a 7-mer. In some

examples, the V5 loop is truncated to a 5-mer. In some examples, the V5 loop is replaced with the amino acid sequence GGNTNRTC (SEQ ID NO: 187), which in some examples is designated V5.1. In some examples, the V5 loop is replaced with the amino acid sequence GGSGSGTC (SEQ ID NO: 188), which in some examples is designated V5.2. In some examples, the V5 loop is replaced with the amino acid sequence GGSGSTC (SEQ ID NO: 189), which in some examples is designated V5.2. In some examples, the V5 loop is truncated to a 16-mer. In some examples, the V5 loop is truncated to NDSGNETFR (SEQ ID NO: 190) for example from KDDNSRDGNETFR (SEQ ID NO: 191), which in some examples is designated V5.2.1. In some examples, the V5 loop is replaced with the amino acid sequence SGGSGQETFR (SEQ ID NO: 192), which in some examples is designated V5.2GS. In some examples, the wild-type V5 loop includes the amino acid sequence GGNDNNESEI (SEQ ID NO: 193). In some examples, the V5 loop is replaced with the amino acid sequence GGGSGSGEI (SEQ ID NO: 194), which in some examples is designated V5.1. In some examples, the V5 loop is replaced with an 8 amino acid Gly-Ser repeat.

In some embodiments, an immunogen is an outer domain of gpl20 and includes residues 252-482 of gpl20. In specific embodiments, the outer domain of gpl20, and thus the immunogen, includes the sequence set forth as SEQ ID NO: 58, or a variant thereof that retains the outer domain fold of gpl20. In some examples, the outer domain of gpl20 includes additional mutations. In specific examples, the immunogen including the outer domain of gpl20 includes one or more of the following mutations made in the context of SEQ ID NO: 203 (OD 1.0): 273N, T283N, T339N, A341T, 360N, 362N, N363Q, P369N, 137IT, 37IN, T373N, T373G, 377N, A379T, F383T, N386Q, D392N, R419N, G421N, G424N, A431T, P437T, N465Q, W479N, N280C and G458, K358C and N465C, and/or V255C and M475C. The cysteine residues can be introduced to stabilize the polypeptide, for example by stabilizing the loops and/or core of the folded polypeptide. When the residue number is not preceded by a residue but is followed by a N it is meant that an asparagine is introduced at that position regardless of the starting residue, for example to insert a glycosylation site at that position.

In specific examples, the immunogen including the outer domain of gpl20 includes one or more of the following mutations made in the context of SEQ ID NO: 1 (A2.3_KER2018.11): 273N, T283N, T339N, A341T, 360N, 362N, N363Q, P369N, 137 IT, 37 IN, T373N, T373G, 377N, A379T, F383T, N386Q, D392N, 5 R419N, G421N, G424N, A431T, P437T, N465Q, W479N, N280C and G458, K358C and N465C, and/or V255C and M475C. The cysteine residues can be introduced to stabilize the polypeptide, for example by stabilizing the loops and/or core of the folded polypeptide. In specific examples, the immunogen including the outer domain of gpl20 includes one or more of the following mutations made in the context of SEQ ID NO: 1: V257N, V272N, N276Q, N276D, N276E, T283V, 10 A297T, D368R, N362T, E363N, P364S, T373N, S375T, F382T, S388A, A388S, K389D, E398N, N406Q, N410Q, K421T, G422V, V427N, V442N, R444T, N478L, S48 IT, and/or E482S.

In specific examples, the immunogen including the outer domain of gpl20 includes one or more of the following mutations made in the context of SEQ ID NO: 78 (HXB2 core): T26N, K55N, A49S, D91T, T92N, D94T, D98N, M100T V103N, K113N, P118N, T207N, R209T, V257N, V272N, N276D, R421N, I423T, T424N, M426T, I434N, A436T, R487N, V489T, and/or R490N. In specific examples, the immunogen including the outer domain of gpl20 includes all of the following mutations made in the context of SEQ ID NO: 74: T26N, V257N, A49S, V103N, 20 V272N. In specific examples, the immunogen including the outer domain of gpl20 includes all of the following mutations made in the context of SEQ ID NO: 74: T84N, A107S, K113N, V244N, V427N, and V442N.

In some examples, the immunogen includes the amino acid sequence set forth as anyone of SEQ ID NOs: 74-76. In some examples, the immunogen includes the amino acid sequence set forth as anyone of SEQ ID NOs: 74-76.

In some examples, glycosylation sites are mutational introduced into the immunogen, for example as show in the table in FIG. 5, in which the numbers indicate the positions of glycosylation. In some examples, glycosylation sites are introduced at one or more of position 295, 442, 479, 272, 377, 436, 424, 398, 348, 30 and/or 369. In some examples glycosylation sites are introduced positions 295 and

442, which is termed HG.1. In some examples glycosylation sites are introduced positions 295, 442, and 479, which is termed HG.2. In some examples glycosylation sites are introduced positions 295, 442, and 436, which is termed HG.3. In some examples glycosylation sites are introduced positions 295, 442, and 424, which is
5 termed HG.4. In some examples glycosylation sites are introduced positions 295, 442, and 398, which is termed HG.5. In some examples glycosylation sites are introduced positions 295, 442, and 369, which is termed HG.6. In some examples glycosylation sites are introduced positions 295, 442, 273, 377, and 348, which is termed HG.7. In some examples glycosylation sites are introduced positions 295,
10 442, 479, 273, 377, 436*, 398, 348, and 369, which is termed HG.8. In some examples glycosylation sites are introduced positions 295, 442, 479, 273, 377, 424*, 398, 348, and 369, which is termed HG.9. In some examples glycosylation sites are introduced positions 295, 442, 479, 273, and 377, which is termed HG.10. In specific examples, the immunogen including the outer domain of gpl20 includes
15 one or more of the following mutations V257N, V272N, N276Q, N276D, N276E, T283V, A297T, D368R, N362T, E363N, P364S, T373N, S375T, F382T, S388A, A388S, K389D, E398N, N406Q, N410Q, K421T, G422V, V427N, V442N, R444T, N478L, S481T, and/or E482S. Mutations that are recited outside of the gpl20 outer domain are made in the context of a larger HIV immunogen, such as a gpl20 or
20 gpl40 immunogen.

In specific examples, the immunogen including the outer domain of gpl20 includes one or more of the following mutations V257N, V272N, N276Q, N276D, N276E, T283V, A297T, D368R, N362T, E363N, P364S, T373N, S375T, F382T, S388A, A388S, K389D, E398N, N406Q, N410Q, K421T, G422V, V427N, V442N,
25 R444T, N478L, S481T, and/or E482S. Mutations that are recited outside of the gpl20 outer domain are made in the context of a larger HIV immunogen, such as a gpl20 or gpl40 immunogen. In specific examples, the immunogen including the outer domain of gpl20 includes all of the following mutations N363Q, K358C-N465C, T283N, N386Q, and N465Q. In specific examples, the immunogen
30 including the outer domain of gpl20 includes all of the following mutations K358C-N465C, T283N, N386Q, and N465Q. In specific examples, the immunogen

including the outer domain of gpl20 includes all of the following mutations V255C-M475C, T283N, D392N, T339N, and N465Q. In specific examples, the immunogen including the outer domain of gpl20 includes all of the following mutations N363Q, V255C-M475C, T283N, N465Q, D392N, T339N, A431T, and T373N. In specific examples, the immunogen including the outer domain of gpl20 includes all of the following mutations N363Q, V255C-M475C, T283N, N465Q, and R419N. In specific examples, the immunogen including the outer domain of gpl20 includes all of the following mutations V255C-M475C, T283N, N465Q, D392N, T339N, G424N, and P437T. In specific examples, the immunogen including the outer domain of gpl20 includes all of the following mutations N276D, T283N, A297T, E398N, V442N, R444T, N478L, W479N, S481T, E482S, N362T, E363N, and P364S. In specific examples, the immunogen including the outer domain of gpl20 includes all of the following mutations K358C-N465C, T283N, and N386Q. In specific examples, the immunogen including the outer domain of gpl20 includes all of the following mutations K358C-N465C, T283N, N386Q, and N465Q. In specific examples, the immunogen including the outer domain of gpl20 includes all of the following mutations F383T, and N280C-G458C, In specific examples, the immunogen including the outer domain of gpl20 includes all of the following mutations F383T, K358C-N465C, In specific examples, the immunogen including the outer domain of gpl20 includes all of the following mutations F383T, V255C-M475C. In specific examples, the immunogen including the outer domain of gpl20 includes all of the following mutations D392N, T339N, and A341T, In specific examples, the immunogen including the outer domain of gpl20 includes all of the following mutations D392N, T339N, A341T, and 386N, In specific examples, the immunogen including the outer domain of gpl20 includes all of the following mutations P369N and 137 IT. In specific examples, the immunogen including the outer domain of gpl20 includes all of the following mutations 42IN, 377N, and 479N. In specific examples, the immunogen including the outer domain of gpl20 includes all of the following mutations 421N, 377N, 479N, and 273N. In specific

examples, the immunogen including the outer domain of gpl20 includes all of the following mutations 424N, 377N, 479N, 273N. In specific examples, the immunogen including the outer domain of gpl20 includes all of the following mutations 424N, 377N, and 479N. In specific examples, the immunogen including the outer domain of gpl20 includes all of the following mutations 421N and 363N. In specific examples, the immunogen including the outer domain of gpl20 includes all of the following mutations 421N, 377N, 479N, and 363N.

In some embodiments the antigen is a multimer, such as a multimer of gpl20, gpl40 or an immunogenic fragment thereof, for example, a dimer, trimer, etc. of gpl20, gpl40 or an immunogenic fragment thereof, such as an outer domain of gpl20. Several exogenous oligomerization motifs have been successfully used to promote stable trimers of soluble recombinant proteins: the GCN4 leucine zipper (Harbury *et al.* 1993 *Science* 262:1401-1407), the trimerization motif from the lung surfactant protein (Hoppe *et al.* 1994 *FEBS Lett* 344:191-195), collagen (McAlinden *et al.* 2003 *J Biol Chem* 278:42200-42207), and the phage T4 fibrin foldon (Miroshnikov *et al.* 1998 *Protein Eng* 11:329-414). Thus, in some examples, the disclosed antigen includes one or more of a foldon domain. In specific examples, the foldon domain is a T4 fibrin foldon domain such as the amino acid sequence GYIPEAPRDGQAYVRKDGWVLLSTF (SEQ ID NO: 195), which adopts a β -propeller conformation, and can fold and trimerize in an autonomous way (Tao *et al.* 1997 *Structure* 5:789-798). In some specific examples, an immunogen including a T4 fibrin foldon domain includes the amino acid sequence set forth as one of SEQ ID NOs: 71-74.

In some embodiments, the disclosed antigen includes a ferritin polypeptide or hybrid of different ferritin polypeptides (for example, to induce multimerization). In specific examples, the ferritin polypeptide is *E. coli* ferritin, human light chain ferritin, bullfrog ferritin or a hybrid thereof, such as *E. coli*-human hybrid ferritin, *E. coli*-bullfrog hybrid ferritin, or human-bullfrog hybrid ferritin. Exemplary amino acid sequences of ferritin polypeptides and nucleic acid sequences encoding ferritin polypeptides for use in the disclosed antigens can be found in GENBANK®, for example at accession numbers ZP_03085328, ZP_06990637, AAA35832,

NP_000137 AAA49532, AAA49525, AAA49524 and AAA49523, which are specifically incorporated by reference herein in their entirety as available June 29, 2010; and in the Protein Data Base (PDB), for example at PDB accession numbers leum, 2ffx and lrcc, which are specifically incorporated by reference herein in their
 5 entirety as available June 29, 2010.

Ferritin is the iron ion storage protein ubiquitously found in almost all living organisms, including bacteria, fungi, and higher plants, and animals. It forms an octahedron consisting of 24 subunits of ~20 kDa protein. In some examples, a disclosed immunogen has been genetically fused to the amino terminus of
 10 engineered ferritin, such as eumrcc (a hybrid *E coli* and human ferritin), with a Ser-Gly linker. When the constructs have been made in HEK 293 Freestyle cells, the fusion proteins are secreted from the cells and self-assembled into octahedral particles. The particles can be purified by a few different chromatography procedures, *e.g.* Mono Q (anion exchange) followed by size exclusion
 15 (SUPEROSE® 6) chromatography.

In some embodiments, a disclosed antigen includes a transmembrane domain, for example to anchor the antigen to the surface of a cell. In some examples, the transmembrane domain is a gp41 transmembrane domain. In some examples, the transmembrane domain is a CD4 transmembrane domain. In specific
 20 examples, a CD4 transmembrane domain is set forth as the amino acid sequence LrVLGGVAGLLLFIGLGI (SEQ ID NO: 196). In some embodiments, a transmembrane domain is an influenza transmembrane domain, such as a neuraminidase (NA) or hemagglutinin (HA) amino acid sequence. In specific examples, the HA sequence includes the amino acid sequence
 25 ILAIYSTVASSLVLLVSLGAISF (SEQ ID NO: 197). In some specific examples, an immunogen including a HA transmembrane domain includes the amino acid sequence set forth as one of SEQ ID NOs: 53-55. In specific examples, the NA sequence includes the amino acid sequence
 30 MNPNQKIITIGSICMVVGIIISLILQIGNIISIWVS (SEQ ID NO: 198). In some specific examples, an immunogen including a HA transmembrane domain includes the amino acid sequence set forth as one of SEQ ID NOs: 56-57.

In some examples a disclosed antigen includes a secretion signal sequence, such as human CD5-derived secretion signal sequence or an IL-2 secretion signal sequence at the N-terminus so that the antigen is secreted from a cell, for example to aid in production and purification of the antigen. In some examples, the secretion
5 signal sequence is a CD5 leader amino acid sequence, such as MPMGSLQPLATLYLLGMLVASVLA (SEQ ID NO. 199). In some specific examples, an immunogen including a CD5 leader amino acid sequence includes the amino acid sequence set forth as one of SEQ ID NOs: 53-55 and 75-76. In other examples, the secretion signal sequence, is an IL-2 secretion signal sequence, such
10 as an murine IL-2 amino acid sequence, for example MQLASCVTLTLLVNSAP (SEQ ID NO: 200). In some specific examples, an immunogen including a murine IL-2 amino acid sequence includes the amino acid sequence set forth as one of SEQ ID NOs: 58-70.

In some examples, the disclosed antigen is a part of a virus-like particle
15 (VLP), such as a CHIKV VLP. In one non-limiting example, the disclosed antigen such as a gpl20, gpl40 or immunogenic fragment thereof, for example a gpl20 outer domain (OD), is inserted between Chikungunya virus (CHIKV) E2 205 amino acid (a.a.) and 206 a.a. on CHIKV VLP (strain 37997). Immunogens are typically presented multimerically (240 molecules per CHIKV VLP particle) to immune cells
20 such as B cells and antigen presenting cells. This results in effectively inducing immune responses against the immunogen, in particular, antibody responses. In specific embodiments, the antigen that is part of a CHIKV VLP includes the amino acid sequence set forth as one of SEQ ID NOs: 79-80.

In some embodiments, a disclosed antigen includes a six-histidine residue
25 tag (for example, to induce oligomerization and/or aid in purification). In some examples the disclosed antigen includes a 3C protease cleavage site, for example so that a 6X His tag or other peptide fragment, such as those described herein can be cleaved from the antigen.

A disclosed antigen can be covalently linked to a carrier, which is an
30 immunogenic macromolecule to which an antigenic molecule can be bound. When bound to a carrier, the bound polypeptide becomes more immunogenic. Carriers are

chosen to increase the immunogenicity of the bound molecule and/or to elicit higher titers of antibodies against the carrier which are diagnostically, analytically, and/or therapeutically beneficial. Covalent linking of a molecule to a carrier can confer enhanced immunogenicity and T cell dependence (see Pozsgay *et al.*, *PNAS* 96:5194-97, 1999; Lee *et al.*, *J. Immunol.* 116:1711-18, 1976; Dintzis *et al.*, *PNAS* 73:3671-75, 1976). Useful carriers include polymeric carriers, which can be natural (for example, polysaccharides, polypeptides or proteins from bacteria or viruses), semi-synthetic or synthetic materials containing one or more functional groups to which a reactant moiety can be attached. Bacterial products and viral proteins (such as hepatitis B surface antigen and core antigen) can also be used as carriers, as well as proteins from higher organisms such as keyhole limpet hemocyanin, horseshoe crab hemocyanin, edestin, mammalian serum albumins, and mammalian immunoglobulins. Additional bacterial products for use as carriers include bacterial wall proteins and other products (for example, streptococcal or staphylococcal cell walls and lipopolysaccharide (LPS)).

In some embodiments, the disclosed antigen includes one or more peptide linkers, for example to attach the gpl20, gpl40 or immunogenic fragment thereof to one or more of a foldon domain, a ferritin polypeptide, a hybrid of different ferritin polypeptides a six-histidine residue tag and a transmembrane domain and the like. Linker peptides are typically a short amino acid sequence providing a flexible linker that permits attachment of an antigenic polypeptide, such as a gpl20 or a gpl40 antigen or an antigenic fragment thereof, without disruption of the structure, aggregation (multimerization) or activity of the self-aggregating polypeptide component. Typically, a linear linking peptide consists of between two and 25 amino acids. Usually, the linear linking peptide is between two and 15 amino acids in length, although in certain circumstances it can be only one, such as a single glycine residue. In one example, the linker polypeptide is two to three amino acids in length, such as a serine and an arginine, or two serine residues and an arginine residue, or two arginine residues and a serine residue, two glycines and a serine, two serines and a glycine or any combination thereof. In specific examples, a peptide linker includes the amino acid sequence set forth as (G₃S)₂G (SEQ ID NO: 201). In

specific examples, a peptide linker includes the amino acid sequence set forth as GSG. In other examples, a linker includes a single G. In other examples, the linear linking peptide can be a short sequence of alternating glycines and prolines, such as the amino acid sequence glycine-proline- glycine-proline. A linking peptide can also
 5 consist of one or more repeats of the sequence glycine-glycine-serine. Alternatively, the linear linking peptide can be somewhat longer, such as the glycine(4)- serine spacer described by Chaudhary *et al*, *Nature* 339:394-397,1989.

The antigens disclosed herein can be chemically synthesized by standard methods, or can be produced recombinantly, for example by expression of the
 10 antigen from a nucleic acid molecule that encodes the antigen (see Section C below). An exemplary process for polypeptide production is described in Lu *et al*, *Federation of European Biochemical Societies Letters*. 429:31-35, 1998. They can also be isolated by methods including preparative chromatography and immunological separations.

15 Exemplary amino acid sequences of gpl20, gpl40 antigens or immunogenic fragments thereof are given below as SEQ ID NOs: 1-77, 79 and 80. In some embodiments, a disclosed gpl20, gpl40 antigen or immunogenic fragment thereof comprises an amino acid sequence that is at least 95% identical to the amino acid sequence set forth as one of SEQ ID NOs: 1-80, such as at least 95% identical, at
 20 least 96% identical, at least 97% identical, at least 98% identical, at least 99% identical, or even 100% identical to the amino acid sequence set forth as one of SEQ ID NOs: 1-77, 79 and 80. In some embodiments, the a disclosed gpl20, gpl40 antigen or immunogenic fragment thereof consists of the amino acid sequence set forth as one of SEQ ID NOs: 1-77, 79 and 80.

25 A2.3_KER2018.11 (SEQ ID NO: 1):
 KPVVSTQLLLNGSLAEKEIRIKSENISDNAKTIIVQLTKPVLINCARPSNNTRG
 RRGDIRQAYCVVNRTQWNTLGQVAIQLRKHWNTTIIFNEPSGGDLEITTHS
 FNCGGFTYCNTSDLFNSTWNIEGTASINGTESNDNITLPCRIKGSGAPPIQGV
 IRCQSNITGILLTRDGGNTGNNRSTNETFRPGGGDMRDNRSE

30 A2.5_KER2018.11 (SEQ ID NO: 2):
 KPVVSTQLLLNGSLAEKEIRIKSENISDCAKTIIVQLTKPVLINCARPSNNTRG
 RRGDIRQAYCVVNRTQWNTLGQVAIQLRKHWNTTIIFNEPSGGDLEITTHS

FNCGGFEFTYCNTSDLFNSTWNIEGTASINGTESNDNITLPCRIKGS GAPPIQGV
IRCQSNITGILLTRDCGNTGNN SRTNETFRPGGGDMRDNRSE

A2.6_KER2018.11 (SEQ ID NO: 3):

5 KPVVSTQLLLNGSLAEKEIRIKSENISDNAKTIIVQLTKPVLINCARPSNNTRG
RRGDIRQAYCVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHS
FNCGGFEFTYCNTSDLFNSTWNIEGTASINGTESNDNITLPCRIKGS GAPPIQGV
IRCQSNITGILLTRDGGNTGNN SRTCETFRPGGGDMRDNRSE.

10 A2.7_KER2018.11 (SEQ ID NO: 4):

KPVCSTQLLLNGSLAEKEIRIKSENISDNAKTIIVQLTKPVLINCARPSNNTRG
RRGDIRQAYCVVNRTQWNDTLGQVAIQLRKHWNTTIIIFNEPSGGDLEITTHS
FNCGGFEFTYCNTSDLFNSTWNIEGTASINGTESNDNITLPCRIKGS GAPPIQGV
IRCQSNITGILLTRDGGNTGNN SRTNETFRPGGGDCRDNRSE.

15

A2.6.1_KER2018.11 (SEQ ID NO: 5):

KPVVSTQLLLNGSLAEKEIRIKSENISDNAKNIIVQLTKPVLINCARPSNNTRG
RRGDIRQAYCVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHS
FNCGGFEFTYCNTSDLFNSTWNIEGTASINGTESNDNITLPCRIKGS GAPPIQGV
20 IRCQSNITGILLTRDGGNTGNN SRTCETFRPGGGDMRDNRSE.

A2.6.2_KER2018.11 (SEQ ID NO: 6):

KPVVSTQLLLNGSLAEKEIRIKSENISDNAKTIIVQLTKPVLINCARPSNNTRG
RRGDIRQAYCVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHS
25 FNCGGFEFTYCNTSDLFNSTWNIEGTASINGTESNDNITLPCRIKGS GAPPIQGV
IRCQSNITGILLTRDGGNTGNN SRTCETFRPGGGDMRDNRSE.

A2.6.3_KER2018.11 (SEQ ID NO: 7):

KPVVSTQLLLNGSLAEKEIRIKSENISDNAKNIIVQLTKPVLINCARPSNNTRG
30 RRGDIRQAYCVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHS
FNCGGFEFTYCNTSDLFNSTWNIEGTASINGTESNDNITLPCRIKGS GAPPIQGV
IRCQSNITGILLTRDGGNTGNN SRTCETFRPGGGDMRDNRSE.

A2.6.5_KER2018.11 (SEQ ID NO: 8):

35 KPVVSTQLLLNGSLAEKEIRIKSEDISDNAKTIIVQLTKPVLINCARPSNNTRG
RRGDIRQAYCVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHS
FNCGGFEFTYCNTSDLFNSTWNIEGTASINGTESNDNITLPCRIKGS GAPPIQGV
IRCQSNITGILLTRDGGNTGNN SRTCETFRPGGGDMRDNRSE.

40 A2.8a_KER2018.11 (SEQ ID NO: 9):

KPVVSTQLLLNGSLAEKEIRIKSENISDNAKTIIVQLTKPVLINCARPSNNTQY
CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGFEFTY
CNTSDLFNSTWNIEGTASINGTESNDNITLPCRIKGS GAPPIQGVIRCQSNITGI
LLTRDGGNTGNN SRTCETFRPGGGDMRDNRSE.

45

- A2.8b_KER2018.11(SEQ ID NO: 10):
KPVVSTQLLLNGSLAEKEIRIKSENISDNAKTIIVQLTKPVLINCARGSGSGSY
CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
CNTSDLFNSTWNIEGTASINGTESNDNITLPCRIKSGAPPIQGVIRCQSNITGI
5 LLTRDGGNTGNNRSTCETFRPGGGDMRDNRSE.
- A2.8j_KER2018.11 (SEQ ID NO: 11):
KPVVSTQLLLNGSLAEKEIRIKSENISDNAKTIIVQLTKPVLINCARPSNNTRG
RRGDIRQAYCVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHS
10 FNCGGEFFYCNTSDLFNSTWNIEGTASINGTESNDNITLPCRIKSGAPPIQGV
IRCQSNITGILLTRDGGSGSGTCETFRPGGGDMRDNRSE.
- A2.8k_KER2018.11 (SEQ ID NO: 12):
KPVVSTQLLLNGSLAEKEIRIKSENISDNAKTIIVQLTKPVLINCARPSNNTRG
15 RRGDIRQAYCVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHS
FNCGGEFFYCNTSDLFNSTWNIEGTASINGTESNDNITLPCRIKSGAPPIQGV
IRCQSNITGILLTRDGGSGSGTCETFRPGGGDMRDNRSE.
- A2.8af_KER2018.11 (SEQ ID NO: 13):
20 KPVVSTQLLLNGSLAEKEIRIKSENISDNAKTIIVQLTKPVLINCARPSNNTQY
CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
CNTSDLFNSTWNIEGTASIDGTESNDNITLPCRIKSGAPPIQGVIRCQSNITGI
LLTRDGGNTGNNRSTCETFRPGGGDMRDNRSE.
- 25 A2.8ag_KER2018.11 (SEQ ID NO: 14):
KPVVSTQLLLNGSLAEKEIRIKSENISDNAKTIIVQLTKPVLINCARPSNNTQY
CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
CNTSDLFNSTWNIEGTASINGTESNDNITLPCRIKSGAPPIQGVIRCQSNITGI
LLTRDGGNTGNNRSTCETFRPGGGDMRDNRSE.
30
- A2.8ah_KER2018.11 (SEQ ID NO: 15):
KPVVSTQLLLNGSLAEKEIRIKSENISDNAKTIIVQLTKPVLINCARPSNNTQY
CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
CNTSDLFNSTWNIEGTASIDGTESNDNITLPCRIKSGAPPIQGVIRCQSNITGI
35 LLTRDGGNTGNNRSTCETFRPGGGDMRDNRSE.
- A2.8ai_KER2018.11 (SEQ ID NO: 16):
KPVVSTQLLLNGSLAEKEIRIKSENISDNAKTIIVQLTKPVLINCARPSNNTQY
CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
40 CNTSDLFNSTWGS GSGSGSDGSGSDITLPCRIKSGAPPIQGVIRCQSNITG
ILLTRDGGNTGNNRSTCETFRPGGGDMRDNRSE.
- A2.8aj_KER2018.11 (SEQ ID NO: 17):
KPVVSTQLLLNGSLAEKEIRIKSENISDNAKTIIVQLTKPVLINCARPSNNTQY
45 CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY

CNTSDLFNSTWNIEGTASINGTESNDNITLPCRIKSGAPPIQGVIRCQSNITGI
LLTRDGGSGSGTCETFRPGGGDMRDNRSE.

A2.8afj_KER2018.11 (SEQ ID NO: 18):

5 KPVVSTQLLLNGSLAEKEIRIKSENISDNAKTIIVQLTKPVLINCARPSNNTQY
CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
CNTSDLFNSTWNIEGTASIDGTESNDNITLPCRIKSGAPPIQGVIRCQSNITGI
LLTRDGGSGSGTCETFRPGGGDMRDNRSE.

10 A2.8agj_KER2018.11 (SEQ ID NO: 19):

KPVVSTQLLLNGSLAEKEIRIKSENISDNAKTIIVQLTKPVLINCARPSNNTQY
CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
CNTSDLFNSTWNIEGTASINGTESNDNITLPCRIKSGAPPIQGVIRCQSNITGI
LLTRDGGSGSGTCETFRPGGGDMRDNRSE.

15

A2.8ahj_KER2018.11 (SEQ ID NO: 20):

KPVVSTQLLLNGSLAEKEIRIKSENISDNAKTIIVQLTKPVLINCARPSNNTQY
CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
CNTSDLFNSTWNIEGTASIDGTESNDNITLPCRIKSGAPPIQGVIRCQSNITGI
20 LLTRDGGSGSGTCETFRPGGGDMRDNRSE.

A2.8aij_KER2018.11 (SEQ ID NO: 21):

KPVVSTQLLLNGSLAEKEIRIKSENISDNAKTIIVQLTKPVLINCARPSNNTQY
CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
25 CNTSDLFNSTWGS GSGSGSDGSGSDITLPCRIKSGAPPIQGVIRCQSNITG
ILLTRDGGSGSGTCETFRPGGGDMRDNRSE.

A2.9a_KER2018.11 (SEQ ID NO: 22):

KPVVSTQLLLNGSLAEKEIRIKSEDISDNAKNIIVQLTKPVLINCARPSNNTQY
30 CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
CNTSDLFNSTWNIEGTASINGTESNDNITLPCRIKSGAPPIQGVIRCQSNITGI
LLTRDGGNTGNNRSTCETFRPGGGDMRDNRSE.

A2.9af_KER2018.11 (SEQ ID NO: 23):

35 KPVVSTQLLLNGSLAEKEIRIKSEDISDNAKNIIVQLTKPVLINCARPSNNTQY
CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
CNTSDLFNSTWNIEGTASIDGTESNDNITLPCRIKSGAPPIQGVIRCQSNITGI
LLTRDGGNTGNNRSTCETFRPGGGDMRDNRSE.

40 A2.9ag_KER2018.11 (SEQ ID NO: 24):

KPVVSTQLLLNGSLAEKEIRIKSEDISDNAKNIIVQLTKPVLINCARPSNNTQY
CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
CNTSDLFNSTWNIEGTASINGTESNDNITLPCRIKSGAPPIQGVIRCQSNITGI
LLTRDGGNTGNNRSTCETFRPGGGDMRDNRSE.

45

A2.9ah_KER2018.11 (SEQ ID NO: 25):

KPVVSTQLLLNGSLAEKEIRIKSEDISDNAKNIIVQLTKPVLINCARPSNNTQY
 CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
 CNTSDLFNSTWNIEGTASIDGTESNDDITLPCRIKGSGAPPIQGVIRCQSNITGI
 5 LLTRDGGNTGNNRSTCETFRPGGGDMRDNRSE.

A2.9ai_KER2018.11 (SEQ ID NO: 26):

KPVVSTQLLLNGSLAEKEIRIKSEDISDNAKNIIVQLTKPVLINCARPSNNTQY
 CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
 10 CNTSDLFNSTWGS GSGSGSDGSGSGSDITLPCRIKGSGAPPIQGVIRCQSNITG
 ILLTRDGGNTGNNRSTCETFRPGGGDMRDNRSE.

A2.9aj_KER2018.11 (SEQ ID NO: 27):

KPVVSTQLLLNGSLAEKEIRIKSEDISDNAKNIIVQLTKPVLINCARPSNNTQY
 15 CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEIT THSFNCGGEFFY
 CNTSDLFNSTWNIEGTASINGTESNDNITLPCRIKGSGAPPIQGVIRCQSNITGI
 LLTRDGGSGSGTCETFRPGGGDMRDNRSE.

A2.9afj_KER2018.11 (SEQ ID NO: 28):

KPVVSTQLLLNGSLAEKEIRIKSEDISDNAKNIIVQLTKPVLINCARPSNNTQY
 20 CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
 CNTSDLFNSTWNIEGTASIDGTESNDNITLPCRIKGSGAPPIQGVIRCQSNITGI
 LLTRDGGSGSGTCETFRPGGGDMRDNRSE.

A2.9agj_KER2018.11 (SEQ ID NO: 29):

KPVVSTQLLLNGSLAEKEIRIKSEDISDNAKNIIVQLTKPVLINCARPSNNTQY
 CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
 CNTSDLFNSTWNIEGTASINGTESNDDITLPCRIKGSGAPPIQGVIRCQSNITGI
 25 LLTRDGGSGSGTCETFRPGGGDMRDNRSE

A2.9ahj_KER2018.11 (SEQ ID NO: 30):

KPVVSTQLLLNGSLAEKEIRIKSEDISDNAKNIIVQLTKPVLINCARPSNNTQY
 CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
 CNTSDLFNSTWNIEGTASIDGTESNDDITLPCRIKGSGAPPIQGVIRCQSNITGI
 30 35 LLTRDGGSGSGTCETFRPGGGDMRDNRSE

A2.9aij_KER2018.11 (SEQ ID NO: 31):

KPVVSTQLLLNGSLAEKEIRIKSEDISDNAKNIIVQLTKPVLINCARPSNNTQY
 CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
 40 CNTSDLFNSTWGS GSGSGSDGSGSGSDITLPCRIKGSGAPPIQGVIRCQSNITG
 ILLTRDGGSGSGTCETFRPGGGDMRDNRSE

A2.6.2 V3.1 (OD4.2F v3.1) (SEQ ID NO: 32):

KPVVSTQLLLNGSLAEKEIRIKSENISDNAKTIIVQLTKPVLINCARPSNNTDIR
 45 QAYCVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGG

EFFYCNTSDLFNSTWNIEGTASINGTESNDNITLPCRIKGSGAPPIQGVIRCQS
NrrGILLTRDGGNTGNNRSTCETFRPGGGDMRDNRSE

A2.6.2 V3.2 (OD4.2F v3.2) (SEQ ID NO: 33):

5 KPVVSTQLLLNGSLAEKEIRIKSENISDNAKTIIVQLTKPVLINCARPSNNTRQ
AYCVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFF
FYCNTSDLFNSTWNIEGTASINGTESNDNITLPCRIKGSGAPPIQGVIRCQSNIT
GILLTRDGGNTGNNRSTCETFRPGGGDMRDNRSE

10 A2.6.2 V5.1 (OD4.2F v5.1) (SEQ ID NO: 34):

KPVVSTQLLLNGSLAEKEIRIKSENISDNAKTIIVQLTKPVLINCARPSNNTRG
RRGDIRQAYCVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHS
FNCGGEFFYCNTSDLFNSTWNIEGTASINGTESNDNITLPCRIKGSGAPPIQGV
IRCQSNITGILLTRDGGNTNRTCETFRPGGGDMRDNRSE

15

A2.HG.1 (OD4.2' HG.1) (SEQ ID NO: 35):

KPVVSTQLLLNGSLAEKEIRIKSEDISDNAKNIIVQLTKPVLINCTRPSNNTQY
CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
CNTSDLFNSTWNIEGTASINGTESNDNITLPCRIKGSGAPPIQGNITCQSNrrGI

20

LLTRDGGSGSGTCETFRPGGGDMRDNRSE

A2.HG.2 (OD4.2' HG.2) (SEQ ID NO: 36):

KPVVSTQLLLNGSLAEKEIRIKSEDISDNAKNIIVQLTKPVLINCTRPSNNTQY
CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
25 CNTSDLFNSTWNIEGTASINGTESNDNITLPCRIKGSGAPPIQGNITCQSNrrGI
LLTRDGGSGSGTCETFRPGGGDMRDNRSE

A2.HG.3 (OD4.2' HG.3) (SEQ ID NO: 37):

KPVVSTQLLLNGSLAEKEIRIKSEDISDNAKNIIVQLTKPVLINCTRPSNNTQY
30 CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
CNTSDLFNSTWNIEGTASINGTESNDNITLPCRINGTIQGNITCQSNITGILLTR
DGGSGSGTCETFRPGGGDMRDNRSE

A2.HG.4 (OD4.2' HG.4) (SEQ ID NO: 38):

KPVVSTQLLLNGSLAEKEIRIKSEDISDNAKNIIVQLTKPVLINCTRPSNNTQY
35 CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
CNTSDLFNSTWNIEGTASINGTESNDNITLPCRIKGSNATVIQGNITCQSNrrGI
LLTRDGGSGSGTCETFRPGGGDMRDNRSE

A2.HG.5 (OD4.2' HG.5) (SEQ ID NO: 39):

KPVVSTQLLLNGSLAEKEIRIKSEDISDNAKNIIVQLTKPVLINCTRPSNNTQY
40 CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
CNTSDLFNSTWNINGTASINGTESNDNITLPCRIKGSGAPPIQGNITCQSNITGI
LLTRDGGSGSGTCETFRPGGGDMRDNRSE

45

A2.HG.6 (OD4.2' HG.6) (SEQ ID NO: 40):

KPVVSTQLLLNGSLAEKEIRIKSEDISDNAKNIIVQLTKPVLINCTRPSNNTQY
 CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDNETTTHSFNCGGEFF
 YCNTSDLFNSTWNIEGTASINGTESNDNITLPCRIGSGAPPIQGNITCQSNIT
 5 GILLTRDGGSGSGTCETFRPGGGDMRDNRSE

A2.HG.7 (OD4.2' HG.7) (SEQ ID NO: 41):

KPVVSTQLLLNGSLAEKEIRINSTDISDNAKNIIVQLTKPVLINCTRPSNNTQY
 CVVNRTQWNDTLGQVAINLTKHWNTCIIFNEPSGGDLEITTHSFNCTGEFFY
 10 CNTSDLFNSTWNIEGTASINGTESNDNITLPCRIGSGAPPIQGNITCQSNITGI
 LLTRDGGSGSGTCETFRPGGGDMRDNRSE

A2.HG.8 (OD4.2' HG.8) (SEQ ID NO: 42):

KPVVSTQLLLNGSLAEKEIRINSTDISDNAKNIIVQLTKPVLINCTRPSNNTQY
 15 CVVNRTQWNDTLGQVAINLTKHWNTCIIFNEPSGGDNETTTHSFNCTGEFF Y
 CNTSDLFNSTWNINGTASINGTESNDNITLPCRINGTIQGNITCQSNITGILLTR
 DGGSGSGTCETFRPGGGDMRDLNRTS

A2.HG.9 (OD4.2' HG.9) (SEQ ID NO: 43):

KPVVSTQLLLNGSLAEKEIRINSTDISDNAKNIIVQLTKPVLINCTRPSNNTQY
 20 CVVNRTQWNDTLGQVAINLTKHWNTCIIFNEPSGGDNETTTHSFNCTGEFFY
 CNTSDLFNSTWNINGTASINGTESNDNITLPCRIGSNATVIQGNITCQSNITG
 ILLTRDGGSGSGTCETFRPGGGDMRDLNRTS

A2.HG.10 (OD4.2' HG.10) (SEQ ID NO: 44):

KPVVSTQLLLNGSLAEKEIRINSTDISDNAKNIIVQLTKPVLINCTRPSNNTQY
 25 CVVNRTQWNDTLGQVAINLTKHWNTCIIFNEPSGGDLEITTHSFNCTGEFFY
 CNTSDLFNSTWNIEGTASINGTESNDNITLPCRIGSGAPPIQGNITCQSNITGI
 LLTRDGGSGSGTCETFRPGGGDMRDLNRTS

A2.HG.3.1 (OD4.2' HG.3.1) (SEQ ID NO: 45):

KPVVSTQLLLNGSLAEKEIRIKSEDISDNAKNIIVQLTKPVLINCTRPSNNTQY
 30 CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
 CNTSDLFNSTWNIEGTASINGTESNDNITLPCRINGTIQGNITCQSNITGILLTR
 35 DGGSGSGTCETFRPGGGDMRDLNRTS

A2.HG.3.2 (OD4.2' HG.3.2) (SEQ ID NO: 46):

KPVVSTQLLLNGSLAEKEIRIKSEDISDNAKNIIVQLTKPVLINCTRPSNNTQY
 40 CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
 CNTSDLFNSTWNINGTASINGTESNDNITLPCRINGTIQGNITCQSNITGILLTR
 DGGSGSGTCETFRPGGGDMRDLNRTS

A2.HG.3.3 (OD4.2' HG.3.3) (SEQ ID NO: 47):

KPVVSTQLLLNGSLAEKEIRIKSEDISDNAKNIIVQLTKPVLINCTRPSNNTQY
 45 CVVNRTQWNDTLGQVAINLTKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY

CNTSDLFNSTWNIEGTASINGTESNDNITLPCRINGTIQGNITCQSNITGILLTR
DGGSGSGTCETFRPGGGDMRDNRSTS

A2.HG.3.4 (OD4.2' HG.3.4) (SEQ ID NO: 48):

5 KPVVSTQLLLNGSLAEKEIRIKSEDISDNAKNIIVQLTKPVLINCTRPSNNTQY
CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCTGEFFY
CNTSDLFNSTWNIEGTASINGTESNDNITLPCRINGTIQGNITCQSNITGILLTR
DGGSGSGTCETFRPGGGDMRDNRSTS

10 A2.HG.3.5 (OD4.2' HG.3.5) (SEQ ID NO: 49):

KPVVSTQLLLNGSLAEKEIRINSTDISDNAKNIIVQLTKPVLINCTRPSNNTQY
CVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFY
CNTSDLFNSTWNIEGTASINGTESNDNITLPCRINGTIQGNITCQSNITGILLTR
DGGSGSGTCETFRPGGGDMRDNRSTS

15

A2.HG.3.6 (OD4.2' HG.3.6) (SEQ ID NO: 50):

KPVVSTQLLLNGSLAEKEIRIKSEDISDNAKNIIVQLTKPVLINCTRPSNNTQY
CVVNRTQWNDTLGQVAIQLRKHWNTCIIFTNSSGGDLEITTHSFNCGGEFFY
CNTSDLFNSTWNIEGTASINGTESNDNITLPCRINGTIQGNITCQSNITGILLTR
DGGSGSGTCETFRPGGGDMRDNRSTS

20

A2.HG.3.2.1 (OD4.2' HG.3.2.1) (SEQ ID NO: 51):

KPVVSTQLLLNGSLAEKEIRIKSEDISDNAKNIIVQLTKPVLINCTRPSNNTQY
CVVNRTQWNDTLGQVAIQLRKHWNTCIIFTNSSGGDLEITTHSFNCGGEFFY
25 CNTSDLFNSTWNINGTASINGTESNDNITLPCRINGTIQGNITCQSNITGILLTR
DGGSGSGTCETFRPGGGDMRDNRSTS

25

A2.6_SG_eumrcc (SEQ ID NO: 52):

KPVVSTQLLLNGSLAEKEIRIKSENISDNAKTIIVQLTKPVLINCARPSNNTRG
30 RRGDIRQAYCVVNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHS
FNCGGEFTYCNTSDLFNSTWNIEGTASINGTESNDNITLPCRIRKSGAPPIQGV
IRCQSNITGILLTRDGGNTGNSRSTCETFRPGGGDMRDNRSESGESQVRQN
FKPEMEEKLNEQMNLELYSSLLYQQMSAWCSYHTFEGAAFLRRHAQEEM
THMQRLFDYLTDTGNLPRINTVESPF AEYSSLDL FQETYKHEQLITQKINEL
35 AHAAMTNQDYPTFNFLQWYVSEQHEEEKLFKSIIDKLSLAGKSGEGLYFIDK
ELSTLD

35

KER2018_OD4.2'_HG3.2_HATM (SEQ ID NO: 53):

MPMGSQPLATLYLLGMLVASVLAKPVVSTQLLLNGSLAEKEIRIKSEDISD
40 NAKNIIVQLTKPVLINCTRPSNNTQYCVVNRTQWNDTLGQVAIQLRKHWNT
CIIFNEPSGGDLEITTHSFNCGGEFFYCNTSDLFNSTWNINGTASINGTESNDN
ITLPCRINGTIQGNITCQSNITGILLTRDGGSGSGTCETFRPGGGDMRDNRSTS
GSGILAIYSTVASSLVLLVSLGAISF

40

45

KER2018_OD4.2'_HG3.2_HATM (SEQ ID NO: 54):
 MPMGSLQPLATLYLLGMLVASVLAKPVVSTQLLLNGSLAEKEIRIKSEDISD
 NAKNIIVQLTKPVLINCTRPSNNTQYCVVNRTQWNDTLGQVAIQLRKHWNT
 CIIFNEPSGGDLEITTHSFNCGGEFFYCNTSDLFNSTWNINGTASINGTESNDN
 5 ITLPCRINGTIQGNITCQSNITGILLTRDGGSGSGTCETFRPGGGDMRDLNRTS
 GSGILAIYSTVASSLVLLVSLGAISF

KER2018_OD4.2'_HG3.2.1_HATM (SEQ ID NO: 55):
 MPMGSLQPLATLYLLGMLVASVLAKPVVSTQLLLNGSLAEKEIRIKSEDISD
 10 NAKNIIVQLTKPVLINCTRPSNNTQYC VVNRTQWNDTLGQVAIQLRKHWNT
 CIIFTNSSGGDLEITTHSFNCGGEFFYCNTSDLFNSTWNINGTASINGTESNDN
 ITLPCRINGTIQGNITCQSNITGILLTRDGGSGSGTCETFRPGGGDMRDLNRTS
 GSGILAIYSTVASSLVLLVSLGAISF

15 KER2018_OD4.2'_HG3.2_NATM (SEQ ID NO: 56):
 MNPNQKIITIGSICMVVGIIISLILQIGNIISIWVSGKPVVSTQLLLNGSLAEKEIR
 IKSEDISDNAKNIIVQLTKPVLINCTRPSNNTQYCVVNRTQWNDTLGQVAIQL
 RKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFYCNTSDLFNSTWNINGTASIN
 GTESNDNITLPCRINGTIQGNITCQSNITGILLTRDGGSGSGTCETFRPGGGDM
 20 RDLNRTS

KER2018_OD4.2'_HG3.2.1_NATM (SEQ ID NO: 57):
 MNPNQKIITIGSICMVVGIIISLILQIGNIISIWVSGKPVVSTQLLLNGSLAEKEIR
 IKSEDISDNAKNIIVQLTKPVLINCTRPSNNTQYCVVNRTQWNDTLGQVAIQL
 25 RKHWNTCIIFTNSSGGDLEITTHSFNCGGEFFYCNTSDLFNSTWNINGTASIN
 GTESNDNITLPCRINGTIQGNITCQSNITGILLTRDGGSGSGTCETFRPGGGDM
 RDLNRTS

R2 OD1.0 (parent sequence) (SEQ ID NO: 58):
 30 MQLASCVTLTLVLLVNSAPRRPVVSTQLLLNGSLAEEEVVIRSENFTNNAKT
 IIVQLREPVKINCSRPNNNTRKSIPMGPGRAFYTGTGQIIGDIRQAHCNISKTNW
 TNALKQVVEKLGEQFNKTIVFTNSSGGNPEIVTHSFNCAGEFFYCNTTQLF
 DSrWNSENG TWNITRGLNNTGRNDTITLPCRQIINRWQEVGKAMYAPPIK
 GNISCSSNITGLLLTRDGGKDDNSRDGNETFRPGGGDMRDNRWRSEGGSHHHH
 35 HH

R2 OD1.0 (parent sequence) (SEQ ID NO: 203):
 RRPVVSTQLLLNGSLAEEEVVIRSENFTNNAKTIVQLREPVKINCSRPNNNT
 RKSIPMGPGRAFYTGTGQIIGDIRQAHCNISKTNWTNALKQVVEKLGEQFNKT
 40 KIVFTNSSGGNPEIVTHSFNCAGEFFYCNTTQLFDSrWNSENG TWNITRGLN
 TGRNDTITLPCRQIINRWQEVGKAMYAPPIKGNISCSSNITGLLLTRDGGK
 DDNSRDGNETFRPGGGDMRDNRWRSE

R2 OD1.1 (SEQ ID NO: 59):
 45 MQLASCVTLTLVLLVNSAPRRPVVSTQLLLNGSLAEEEVVIRSENFTNNAKT
 IIVQLREPVKINCSRPNNNTRKSIPMGPGRAFYTGTGQIIGDIRQAHCNISKTNW

TNALKQVVEKLGEQFNKTKIVFTQSSGGNPEIVTHSFNCAGEFFYCNTTQLF
DSIWNSNGTWNITRGLNNTGRNDTITLPCRIKQIINRWQEVGKAMYAPPIK
GNISCSSNITGLLLTRDGGKDDNSRDGNETFRPGGGDMRDNWRSEGSHHHH
HH

5

R2 OD2.2 (SEQ ID NO: 60):

MQLASCVTLTLVLLVNSAPRRPVVSTQLLLNGLAEEEEVVIRSENFTNNAKT
IIVQLREPVKINCSRPNNNTRGRRGDIRQAHCNISKTNWTNALKQVVEKLGE
QFNKTKIVFTQSSGGDPErVTHSFNCAGEFFYCNTTQLFDSrWNSNGTWNIT
RGLNNTGRNDTITLPCRIKSGAPPIKGNISCSSNITGLLLTRDGGKDDNSRD
GNETFRPGGGDMRDNWRSEGSHHHHHH

10

R2 OD3.2 (SEQ ID NO: 61):

MQLASCVTLTLVLLVNSAPRRPVVSTQLLLNGLAEEEEVVIRSENFTNNAKT
IIVQLREPVKINCSRPNNNTRGRRGDIRQAHCNISKTNWTNALKQVVEKLGE
QFNKTKIVFTQSSGGDPErVTHSFNCAGEFTYCNTTQLFDSrWNSNGTWNIT
RGLNNTGRNDTITLPCRIKSGAPPIKGNISCSSNITGLLLTRDGGKDDNSRD
GNETFRPGGGDMRDNWRSEGSHHHHHH

15

R2 OD4.1 (SEQ ID NO: 62):

MYSMQLASCVTLTLVLLVNSAPRRPVVSTQLLLNGLAEEEEVVIRSENFTNC
AKTIIVQLREPVKINCSRPNNNTRGRRGDIRQAHCNISKTNWTNALKQVVEK
LGEQFNKTKIVFTQSSGGDPEIVTHSFNCAGEFTYCNTTQLFDSrWNSNGT
WNrRGLNNTGRNDTITLPCRIKSGAPPIKGNISCSSNITGLLLTRDCGKDD
NSRDGNETFRPGGGDMRDNWRSEGSLVLFQGP GHHHHHHH

20

25

R2 OD4.2 (SEQ ID NO: 63):

MQLASCVTLTLVLLVNSAPRRPVVSTQLLLNGLAEEEEVVIRSENFTNNAKT
IIVQLREPVKINCSRPNNNTRGRRGDIRQAHCNISKTNWTNALKQVVEKLGE
QFNKTCIVFTQSSGGDPErVTHSFNCAGEFTYCNTTQLFDSnVNSNGTWNr
RGLNNTGRNDTITLPCRIKSGAPPIKGNISCSSNITGLLLTRDGGKDDNSRD
GCETFRPGGGDMRDNWRSEGSHHHHHH

30

R2 OD4.2.2 (SEQ ID NO: 64):

MQLASCVTLTLVLLVNSAPRRPVVSTQLLLNGLAEEEEVVIRSENFTNNAKN
IIVQLREPVKINCSRPNNNTRGRRGDIRQAHCNISKTNWTNALKQVVEKLGE
QFNKTCIVFTQSSGGDPErVTHSFNCAGEFFYCQTTQLFDSnVNSNGTWNr
RGLNNTGRNDTITLPCRIKSGAPPIKGNISCSSNITGLLLTRDGGKDDNSRD
GCETFRPGGGDMRDNWRSEGSLVLFQGP GHHHHHHH

35

40

R2 OD4.2.2/V3.2 (SEQ ID NO: 65):

MQLASCVTLTLVLLVNSAPRRPVVSTQLLLNGLAEEEEVVIRSENFTNNAKN
IIVQLREPVKINCSRPNNNTRRQAHCNISKTNWTNALKQVVEKLGEQFNKTC
IVFTQSSGGDPEIVTHSFNCAGEFFYCQTTQLFDSTWNIETASINGTESNDQI
TLPCRIKSGAPPIKGNISCSSNITGLLLTRDGGSGGSGCETFRPGGGDMRDN
WRSEGSLVLFQGP GHHHHHHH

45

R2 OD4.2.2/V3.2/p (SEQ ID NO: 66):

5 MQLASCVTTLTLVLLVNSAPRRPVVSTQLLLNGLANESVVIRSENFTNNAKN
IIVQLREPVKINCSRPNNNTRRQAHCNISKTNWTNALKQVVEKLGEQFNKTC
IVFTQSSGGDPEIVTHSFNCAGEFFYCQTTQLFDSTWNIEGTASINGTESNDQI
TLPCRIKGSGANATKGNISCSSNITGLLLTRDGGSGGSGCETFRPGGGDMRD
NNRSEGSLEVLFGPGHHHHHH

R2 OD4.3 (SEQ ID NO: 67):

10 MQLASCVTTLTLVLLVNSAPRRPVCSTQLLLNGLAEEEEVVIRSENFTNNAKT
irVQLREPVKINCSRPNNNTRGRRGDIRQAHCNISKTNWTNALKQVVEKLGE
QFNKTKIVFTQSSGGDPErVTHSFNCAGEFTYCNTTQLFDSrWNSENGTDWNIT
RGLNNTGRNDTITLPCRIKGSGAPPIKGNISCSSNITGLLLTRDGGKDDNSRD
GNETFRPGGGDCRDNRSEGSHHHHHH

R2 OD4.3.2 (SEQ ID NO: 68):

15 MQLASCVTTLTLVLLVNSAPRRPVCSTQLLLNGLAEEEEVVIRSENFTNNAKN
irVQLREPVKINCSRPNNNTRGRRGDIRQAHCNISKTNWTNALKQVVEKLGE
QFNKTKIVFTQSSGGDPErVTHSFNCAGEFFYCQTTQLFDSrWNSENGTDWNIT
20 RGLNNTGRNDTITLPCRIKGSGAPPIKGNISCSSNITGLLLTRDGGKDDNSRD
GQETFRPGGGDCRDNRSEGSLEVLFGPGHHHHHH

R2 OD4.3.2/V3.2/p (SEQ ID NO: 69):

25 MQLASCVTTLTLVLLVNSAPRRPVCSTQLLLNGLAEEEEVVIRSENFTNNAKN
IIVQLREPVKINCSRPNNNTRRQAHCNISKTNWTNALKQVVEKLGEQFNKTK
IVFTQSSGGDPEIVTHSFNCAGEFFYCQTTQLFDSTWNIEGTASINGTESNDNI
TLPCRIKGSGAPPIKGNISCSSNITGLLLTRDGGSGGSGQETFRPGGGDCRD
WRSEGSLEVLFGPGHHHHHH

R2 OD4.3.2/V4.2.1 R2/p (SEQ ID NO: 70):

30 MQLASCVTTLTLVLLVNSAPRRPVCSTQLLLNGLANESVVIRSENFTNNAKN
IIVQLREPVKINCSRPNNNTRRQAHCNISKTNWTNALKQVVEKLGEQFNKTK
IVFTQSSGGDPEIVTHSFNCAGEFFYCQTTQLFDSTWNIEGTASINGTESNDNI
TLPCRIKGSGANATKGNISCSSNITGLLLTRDGGSGGSGQETFRPGGGDCRD
35 NNRSEGSLEVLFGPGHHHHHH

R2 gpl40(N363Q)(AV123/GSL)(Ap2021)(mFC)/Foldon/3C/His (SEQ ID NO: 71):

MRVKGIRRNYPHWGWGTMLLGLLMICSATEKLWVTVYYGVPVWKEAT
TTLFCASDAKAYDTEAHNVWATHACVPTDPNPQEVELNVNVTENFNMWKN
40 NMVEQMHEDIISLWDQSLKPCVKLTPLCVGSGSCNTSVrrQACPKISFEPIPIH
YCAPAGFAILKCNDKKFSGKGSCKNVSTVQCTHGIRPVVSTQLLLNGLAEE
EVVIRSENFTNNAKTIIIVQLREPVKINCSRPNNNTRGRRGDIRQAHCNISKTN
WTNALKQVVEKLGEQFNKTKIVFTQSSGGDPErVTHSFNCAGEFFYCNTTQL
FDSIWNSENGTDWNITRGLNNTGRNDTITLPCRIKGSGAPPIKGNISCSSNITGL
45 LLTRDGGKDDNSRDGNETFRPGGGDMRDNRSELYKYKVVKIEPLGVAPT
KAKRRVVQSEESATGLGATYIGYLGAAGSTMGAASVTTLTVQARQLLSGIVQ

QQSNLLRAIEAQQHLLQLTVWGIKQLQARILAVEYLKDQQLLGIWGCSGK
 LICTTTVPWNASWSKNKTLEAIWNNMTWMQWDKEIDNYTSLIYSLIEESQI
 QQEKNEQELLELDKWANLWNWFDISNWLWGSGYIPEAPRDGQAYVRKDG
 EWVLLSTFLGLEVLFGPGHHHHHH

5

R2_gpl40(N363Q))(AV13/GSL)(Ap2021)(mFC)/Foldon/3C/His (SEQ ID NO: 72):
 MRVKGIRRNYPHWWGWGTMLLGLLMICSATEKLWVTVYYGVPVWKEAT
 TTLFCASDAKAYDTEAHNVWATHACVPTDPNPQEVELVNVTFENFNMWKN
 NMVEQMHEDIISLWDQSLKPCVKLTPLCVGSKNCSFNIAISIGDKMQKEYA
 10 LLYKLDIEPIDNDNTSYRLISCNTSVITQACPKISFEPIPIHYCAPAGFAILKCN
 DKKFSGKGSCKNVSTVQCTHGIRPVVSTQLLNGSLAEEEVVIRSENFTNNA
 KTIIVQLREPVKINCSRPNNNTRGRRGDIRQAHCNISKNTNWTNALKQVVEKL
 GEQFNKTKIVFTQSSGGDPEIVTHSFNCAGEFFYCNTTQLFDSIWNSNGTW
 NITRGLNNTGRNDTITLPCRIKSGAPPIKGNISCSSNITGLLLTRDGGKDDNS
 15 RDGNETFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRRVVQSEES
 ATGLGATYIGYLGAAGSTMGAASVTLTVQARQLLSGIVQQSNLLRAIEAQ
 QHLLQLTVWGIKQLQARILAVEYLKDQQLLGIWGCSGKLICTTTVPWNAS
 WSKNKTLEARVNNMTWMQWDKEIDNYTSLIYSLIEESQIQQEKNEQELLEL
 DKWANLWNWFDISNWLWGSGYIPEAPRDGQAYVRKDG
 20 EVVLLSTFLGLEVLFGPGHHHHHH

R2_gpl40(N363Q))(AV123/m5)(Ap2021)(mFC)/Foldon/3C/His (SEQ ID NO: 73):
 MRVKGIRRNYPHWWGWGTMLLGLLMICSATEKLWVTVYYGVPVWKEAT
 TTLFCASDAKAYDTEAHNVWATHACVPTDPNPQEVELVNVTFENFNMWKN
 25 NMVEQMHEDIISLWDQSLKPCVKLTPLCVGSGSCNTSV_{tr}QACPKISFEPIPIH
 YCAPAGFAILKCNDDKKFSGKGSCKNVSTVQCTHGIRPVVSTQLLNGSLAEE
 EVVIRSENFTNNAKTIIVQLREPVKINCSRPNNNTRGRRGSSGSHCNISKTN
 WTNALKQVVEKLGEQFNKTKIVFTQSSGGDPE_rVTHSFNCAGEFFYCNTTQL
 FDSIWNSNGTWNITRGLNNTGRNDTITLPCRIKSGAPPIKGNISCSSNITGL
 30 LLTRDGGKDDNSRDGNETFRPGGGDMRDNRSELYKYKVVKIEPLGVAPT
 KAKRRRVVQSEESATGLGATYIGYLGAAGSTMGAASVTLTVQARQLLSGIVQ
 QQSNLLRAIEAQQHLLQLTVWGIKQLQARILAVEYLKDQQLLGRVGCSSGK
 LICTTTVPWNASWSKNKTLEA_rWNNMTWMQWDKEIDNYTSLIYSLIEESQI
 QQEKNEQELLELDKWANLWNWFDISNWLWGSGYIPEAPRDGQAYVRKDG
 35 EWVLLSTFLGLEVLFGPGHHHHHH

RS3 (SEQ ID NO: 74):
 TTVTVNVTVTFDWCADDMVATMNTAICTLWKTSNDPCTKCPTVRFKPVPIR
 YCAPPGYAILKCNRRDFNGTGPCTNVSVVTCTDGIHPVVSSQLLNGTLADE
 40 KVVIRSCNFSDNAKTIIVQLNTSVEINCTGQGHCNITRAKWNQTLKQIAEKL
 REQFGNNKTIIFRPSSGGDPEIVTHWFNCGGKFFYCNSTQLFNSTWFNSTWS
 TKGSNNTEGSDTITLPCRIR_{Srr}GMVCTVGKMIYAPPVEGVITCSSNITGLLLT
 RDGGNDNNESEIFRPGGGDMRDNRSELYKYRVVRLT

RSC3(Y5) (SEQ ID NO: 75):

MPMGSLQPLATLYLLGMLVASVLATNVTNVTNVTFTDWCADDMVATMNTS
ICTLWKTSNDPCTKCPTVRFKPVPIRYCAPPGYAILKCNNRDFNGTGPCTNV
SNVTCTDGIHPVVSSQLLLNGTLADEKVVIRSCNFSDNAKTIIVQLNTSVEIN
5 CTGQGHHCNITRAKWNQTLKQIAEKLREQFGNNKTIIFRPSSGGDPEIVTHWF
NCGGKFFYCNSTQLFNSTWFNSTWSTKGSNNTEGSDTITLPCRIRSITGMNC
TVGKMIYAPPVEGNITCSSNITGLLLTRDGGNDNNESEIFRPGGGDMRDNR
SELYKYRVVRLTGSHHHHHH.

10 RSC3 Y6.1 His (SEQ ID NO: 76):

MPMGSLQPLATLYLLGMLVASVLATNVTNVTNVTFTDWCADDMVATMNTS
ICTLWNTSNDPCTKCPTVRFKPVPIRYCAPPGYAILKCNNRDFNGTGPCTNV
SNVTCTDGIHPVVSSQLLLNGTLADEKVVIRSCNFSDNAKTIIVQLNTSVEIN
CTGQGHHCNITRAKWNQTLKQIAEKLREQFGNNKTIIFRPSSGGDPEIVTHWF
15 NCGGKFFYCNSTQLFNSTWFNSTWSTKGSNNTEGSDTITLPCRIRSITGMNC
TVGKMIYAPPVEGNITCSSNITGLLLTRDGGNDNNESEIFRP
GGGDMRDNRSELYKYRVVRLTGSHHHHHH

Mosaic 2.1/gpl45(AVLIV2.1V3.5)(ACFI) (SEQ ID NO: 77):

20 LWVTVYYGVPVWKEATTTLFCASDAKAYDTEVHNWATHACVPTDPNPQ
EVVLGNVTENFNMWKNNMVEQMHEDIISLWDESLKPCVKLTPLCVTLNCT
DLRGSGGGSGGGSEIKNCSFNITTSIRDKVQKEYALFYKLDVVPIGGSGGSY
RLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNDKKFNGTGPKKNVST
VQCTHGIRPVVSTQLLNGSLAEEEVVIRSENFTNNAKTIIVQLNESVEINCTR
25 PNNNTRGRRGSSGGSHCNISRAKWNNTLKQIVIKLREQFGNKTIIVFNQSSGG
DPErVMHSFNCGGEFFYCNTTQLFNSTWNVNGTWNGTGSENITLPCRICKIV
NMWQEVGKAMYAPPIRGQIRCSSNITGLLLTRDGGNNNSTNETFRPGGGDM
RDNRSELYKYKVVKIEPLGVAPTKAKLTVQARLLLSGIVQQQNNLLRAIE
AQQHMLQLTVWGIKQLQARVLAVERYLKDQQLLERVNNMTWMEWEREID
30 NYTGLIYTLIEESQNQQEKNEQELLELDKWASLWNWFDITNWLWYIKIFIMI
VGGLIGLRIVFAVLSI

HXB2 core (SEQ ID NO: 78):

EVVLNVVTENFNMWKNDMVEQMHEDIISLWDQSLKPCVKLTPLCVGAGSC
35 NTSVITQACPKVSFEPIPIHYCAPAGFAILKCNNKTFNGTGPCTNVSTVQCTH
GIRPVVSTQLLNGSLAEEEVVIRSVNFTDNAKTIIVQLNTSVEINCTGAGHC
NIARAKWNNTLKQIASKLREQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFY
CNSTQLFNSTWFNSTWSTEGSNNTGSDTITLPCRICKIINMWQKVGKAMY
APPISGQIRCSSNITGLLLTRDGGNSNNESEIFRPGGGDMRDNRSELYKYK
40 VVKIE

CHIKV185_OD4.2'_HG3.2 (SEQ ID NO: 79)

MEFIPTQTFYNRRYQPRPWAPRPTIQVIRPRPRPQRQAGQLAQLISAVNKL
45 MRAVPQQKPRNRNRKNKKQRQKKQAPQNDPKQKKQPPQKKPAQKKKKPGR
RERMCMKIENDCIFEVKHEGKVMGYACLVGDKVMKPAHVKGKTIDNADLA

KLAFKRSSKYDLECAQIPVHMKSDASKFTHEKPEGYYNWHHGAVQYSSGR
 FTIPTGAGKPGDSGRPIFDNKGRVVAIVLGGANEGARTALSVVTWNKDIVTK
 ITPEGAEWSLALPVLCLLANTTFPCSQPPCTPCCYEKEPESTLRMLEDNVM
 RPGYYQLLKASLTCSPHRQRRSTKDNFNVYKATRPYLAHCPDCGEGHSCHS
 5 PIALERIRNEATDGTLKIQVSLQIGIKTDDSHDWTCLRYMDSHTPADAEERAG
 LLVRTSAPCTITGTMGHFILARCPKGETLTVGFTDSRKISHTCTHPFHHEPPVI
 GRERFHSRPQHKGELPCSTYVQSTAATAEEIEVHMPPDTPDRTLMTQQSGKP
 VVSTQLLLNGSLAEKEIRIKSEDISDNAKNIIIVQLTKPVLINCTRPSNNTQYCV
 VNRTQWNDTLGQVAIQLRKHWNTCIIFNEPSGGDLEITTHSFNCGGEFFYCN
 10 TSDLFNSTWNINGTASINGTESNDNITLPCRINGTIQGNITCQSNITGILLTRDG
 GSGSGTCETFRPGGGDMRD LNRTSGSGNVKITVNGQTVRYKCNCGGSNEG
 LTTTDKVINNCKIDQCHAAVTNHKNWQYNSPLVPRNAELGDRKGKIHIPFPL
 ANVTCRVPKARNPTVTYGKNQVTMLLYPDHPTLLSYRNMGQEPNYHEEW
 VTHKKEVTLTVPTEGLEVTWGNNEPYKYWPQMSTNGTAHGHPHEIILYYY
 15 ELYPTMTVVIVSVASFVLLSMVGTAVGMCVCARRRCrrPYELTPGATVPFLL
 SLLCCVRTTKAATYYEAAAYLWNEQQPLFWLQALIPLAALIVLCNCLKLLP
 CCCKTLAFLAVMSIGAHTVSAYEHVTVIPNTVGVPYKTLVNRPGYSPMVLE
 MELQSVTLEPTLSLDYITCEYKTVIPSPYVKCCGTAECKDKSLPDYSCKVFT
 GVYPFMWGGAYCFDAENTQLSEAHVEKSECKTEFASAYRAHTASASAK
 20 LRVLYQGNNrrVAAYANGDHA VTVKDAKFVVGPMSSAWTPFDNKIVVYK
 GDVYNMDYPPFGAGRPGQFGDIQSRTPEKDVYANTQLVLQRPAAGTVHV
 PYSQAPSGFKYWLKERGASLQHTAPFGCQIATNPVRAVNCAVGNIPI SIDIPD
 AAFTRVVDAPSVTDMSCVPACTHSSDFGGVAIIKYTASKKGKCAVHSMTN
 AVTIREADVEVEGNSQLQISFSTALASAEFRVQVCSTQVHCAAACHPPKDH
 25 VNYPASHTTLGVQDISTTAMSWVQKITGGVGLIVAVAALILIVLVCVSFSRH

CHIKV185_OD4.2'_HG3.2.1 (SEQ ID NO: 80)
 MEFIPTQTFYNRRYQPRPWAPRPTIQVIRPRPRPQRQAGQLAQLISAVNKL
 MRAVPQQKPRRNRKNKKQRQKKQAPQNDPKQKKQPPQKKPAQKKKKPGR
 30 RERMCMKIENDCIFEVKHEGKVMGYACLVGDKVMKPAHVKGITDNADLA
 KLAFKRSSKYDLECAQIPVHMKSDASKFTHEKPEGYYNWHHGAVQYSSGR
 FTIPTGAGKPGDSGRPIFDNKGRVVAIVLGGANEGARTALSVVTWNKDIVTK
 ITPEGAEWSLALPVLCLLANTTFPCSQPPCTPCCYEKEPESTLRMLEDNVM
 RPGYYQLLKASLTCSPHRQRRSTKDNFNVYKATRPYLAHCPDCGEGHSCHS
 35 PIALERIRNEATDGTLKIQVSLQIGIKTDDSHDWTCLRYMDSHTPADAEERAG
 LLVRTSAPCTITGTMGHFILARCPKGETLTVGFTDSRKISHTCTHPFHHEPPVI
 GRERFHSRPQHKGELPCSTYVQSTAATAEEIEVHMPPDTPDRTLMTQQSGKP
 VVSTQLLLNGSLAEKEIRIKSEDISDNAKNIIIVQLTKPVLINCTRPSNNTQYCV
 VNRTQWNDTLGQVAIQLRKHWNTCIIFTNSSGGDLEITTHSFNCGGEFFYCN
 40 TSDLFNSTWNINGTASINGTESNDNITLPCRINGTIQGNITCQSNITGILLTRDG
 GSGSGTCETFRPGGGDMRD LNRTSGSGNVKITVNGQTVRYKCNCGGSNEG
 LTTTDKVINNCKIDQCHAAVTNHKNWQYNSPLVPRNAELGDRKGKIHIPFPL
 ANVTCRVPKARNPTVTYGKNQVTMLLYPDHPTLLSYRNMGQEPNYHEEW
 VTHKKEVTLTVPTEGLEVTWGNNEPYKYWPQMSTNGTAHGHPHEIILYYY
 45 ELYPTMTVVIVSVASFVLLSMVGTAVGMCVCARRRCrrPYELTPGATVPFLL
 SLLCCVRTTKAATYYEAAAYLWNEQQPLFWLQALIPLAALIVLCNCLKLLP

CCCKTLAFLAVMSIGAHTVSAYEHVTVIPNTVGVVPYKTLVNRPGYSPMVLE
 MELQSVTLEPTLSLDYITCEYKTVIPSPYVKCCGTAECKDKSLPDYSCKVFT
 GVYPPFMWGGAYCFCDAENTQLSEAHVEKSECKTEFASAYRAHTASASAK
 LRVLYQGNITVAAYANGDHAHTVTKDAKFVVGPMSSAWTPFDNKIVVYK
 5 GDVYNMDYPPFGAGRPGQFGDIQSRTPESKDVYANTQLVLQRPAAAGTVHV
 PYSQAPSGFKYWLKERG ASLQHTAPFGCQIATNP VRAVNCAVGNIPISIDIPD
 AAFTRVVDAPSVTDMSCCEVPACTHSSDFGGVAIIKYTASKKGKCAVHSMNTN
 AVTIREADVEVEGNSQLQISFSTALASAEFRVQVCSTQVHCAAACHPPKDH
 VNYPASHTTLGVQDISTTAMSWVQKITGGVGLIVAVAALILIVVLCVSFSRH

10

CHIKV-OD A2.6 (SEQ ID NO: 148):

MEFIPTQTFYNRRYQPRPWAPRTIQVIRPRPRPQRQAGQLAQLISAVNKL
 MRAPVQQKPRRNRKNKKQRQKKQAPQNDPKQKKQPPQKKPAQKKKKPGR
 RERMCMKIENDCIFEVKHEGKVMGYACLVGDKVMKPAHVKGITDNADLA
 15 KLAFLRSSKYDLECAQIPVHMKSDASKFTHEKPEGYYNWHHGAVQYSGGR
 FTIPTGAGKPGDSGRPIFDNKGRVVAIVLGGANEGARTALSVVTWNKDIVTK
 ITPEGAEESLALPVLCLLANTTFPCSQPPCTPCCYEKEPESTLRMLEDNVM
 RPGYYQLLKASLTCSPHRQRRSTKDNFNVYKATRPYLAHCPDCGEGHSCHS
 PIALERIRNEATDGTLLKIQVSLQIGIKTDDSHDWTCLRYMDSHTPADAEAG
 20 LLVRTSAPCTITGTMGHFILARCPKGETLTVGFTDSRKISHTCTHPFHHEPPVI
 GRERFHSRPHGKELPCSTYVQSTAATAEEIEVHMPPDTPDRTLMTQQSGN
 VKITVNGQTVRYKCNCGSGKPVVSTQLLLNGSLAEKEIRIKSENISDNAKTI
 IVQLTKPVLINCARPSNNTRGRRGDIRQAYCVVNRTQWNDTLGQVAIQLRK
 HWNTCIIFNEPSGGDLEITTHSFNCGGEFTYCNTSDLFNSTWNIEGTASINGTE
 25 SNDNITLPCRIKSGAPPIQGVRQSNITGILLTRDGGNTGNNRSTCETFRPG
 GGDMDRDNWRSESGSNEGLTTTDKVINNCKIDQCHAAVTNHKNWQYNSPL
 VPRNAELGDRKGKIHIPFLANVTCRVPKARNPTVTYGKNQVTMLLYPDHP
 TLLSYRNMGQEPNYHEEWVTHKKEVTLTVPTEGLEVTWGNNEPYKYWPQ
 MSTNGTAHGHHPHEIILYYYELYPTMTVVIVSVASFVLLSMVGTAVGMCVCA
 30 RRRCTrPYELTPGATVPFLLSLLCCVVRTTKAATYYEAAAYLWNEQQPLFWL
 QALIPLAALIVLCNCLKLLPCCCKTLAFLAVMSIGAHTVSAYEHVTVIPNTV
 GVPYKTLVNRPGYSPMVLEMELQSVTLEPTLSLDYITCEYKTVIPSPYVKCC
 GTAECKDKSLPDYSCKVFTGVYPPFMWGGAYCFCDAENTQLSEAHVEKSES
 CKTEFASAYRAHTASASAKLRVLYQGNITVAAYANGDHAHTVTKDAKFVV
 35 GPMSSAWTPFDNKIVVYKGDVYNMDYPPFGAGRPGQFGDIQSRTPESKDVY
 ANTQLVLQRPAAAGTVHVPYSQAPSGFKYWLKERGASLQHT APFGCQIATNP
 VRAVNCAVGNIPISIDIPDAAAFTRVVDAPSVTDMSCCEVPACTHSSDFGGVAII
 KYTASKKGKCAVHSMNTNAVIREADVEVEGNSQLQISFSTALASAEFRVQV
 CSTQVHCAAACHPPKDHVNYPASHTTLGVQDISTTAMSWVQKITGGVGLI
 40 VAVAALILIVVLCVSFSRH

CHIKV-OD4.1 (SEQ ID NO: 149):

MEFIPTQTFYNRRYQPRPWAPRTIQVIRPRPRPQRQAGQLAQLISAVNKL
 MRAPVQQKPRRNRKNKKQRQKKQAPQNDPKQKKQPPQKKPAQKKKKPGR
 45 RERMCMKIENDCIFEVKHEGKVMGYACLVGDKVMKPAHVKGITDNADLA
 KLAFLRSSKYDLECAQIPVHMKSDASKFTHEKPEGYYNWHHGAVQYSGGR

FTIPTGAGKPGDSGRPIFDNKGRVVAIVLGGANEGARTALSVVTWNKDIVTK
 ITPEGAEWSLALPVLCLLANTTFPCSQPPCTPCCYEKEPESTLRMLEDNVM
 RPGYYQLLKASLTCSPHRQRRSTKDNFNVYKATRPYLAHCPDCGEGHSCHS
 PIALERIRNEATDGTLLKIQVSLQIGIKTDDSHDWTCLRYMDSHTPADAERAG
 5 LLVRTSAPCTITGTMGHFILARCPKGETLTVGFTDSRKISHTCTHPFHHEPPVI
 GRERFHSRPQHKGELPCSTYVQSTAATAEEIEVHMPPDTPDRTLMTQQSGN
 VK_{rr}VNGQTVRYKCNCGSGRPVSTQLLNGSLAEEEVVIRSENFTNCAKT
 IIVQLREPVKINCSRPNNTNRGRRGDIRQAHCNISKTNWTNALKQVVEKLGE
 QFNKTKIVFTQSSGGDP_{er}VTHSFNCAGEFTYCNTTQLFDS_rWSENGETWNIT
 10 RGLNNTGRNDTITLPCRIKSGAPPIKGNISCSNITGLLLTRDCGKDDNSRD
 GNETFRPGGGDMRDNRSESGSNEGLTTTDKVINNCKIDQCHAAVTNHK
 NWQYNSPLVPRNAELGDRKGKIHIPPLANVTCRVPKARNPTVTYGKNQVT
 MLLYPDHTLLSYRNMGQEPNYHEEWVTHKKEVTLTVPTEGLEVTWGNNE
 PYKYWPQMSTNGTAHGHPEIILYYYELYPTMTVVIVSVASFVLLSMVGTA
 15 VGMVCVARRRCITPYELTPGATVPFLSLCCVRRTTKAATYYEAAAYLWNE
 QQPLFWLQALIPLAALIVLCNCLKLLPCCCKTLAFLAVMSIGAHTVSAYEHV
 TVIPNTVGVPYKTLVNRPGYSPMVLEMELQSVTLEPTLSLDYITCEYKTVIPS
 PYVKCCGTAECKDKSLPDYSCKVFTGVYPFMWGGAYCFDAENTQLSEAH
 VEKSECKTEFASAYRAHTASASAKLRVLYQGNNITVAAYANGDHAVTVK
 20 DAKFVVGPMSSAWTPFDNKIVVYKGDVYNMDYPPFGAGRPGQFGDIQSRT
 PESKD VYANTQLVLQRP AAGTVHVPYSQAPSGFKYWLKERG ASLQHTAPF
 GCQIATNPVRAVNCAVGNIPISIDIPDAAFRVVDAPSVTDMSCVPACTHSS
 DFGGVAIIKYTASKKGKCAVHSMTNAV TIREADVEVEGNSQLQISFSTALAS
 AEF RVQVCSTQVHCAAACHPPKDHIVNYPASHTTLGVQDISTTAMSWVQKI
 25 TGGVGLIVAVAALILIVVLCVSFSRH

B. Polynucleotides Encoding Antigens

Polynucleotides encoding the antigens disclosed herein are also provided.
 These polynucleotides include DNA, cDNA and RNA sequences which encode the
 30 antigen.

Methods for the manipulation and insertion of the nucleic acids of this
 disclosure into vectors are well known in the art (see for example, Sambrook *et al.*,
Molecular Cloning, a Laboratory Manual, 2d edition, Cold Spring Harbor Press,
 Cold Spring Harbor, N.Y., 1989, and Ausubel *et al.*, *Current Protocols in Molecular*
 35 *Biology*, Greene Publishing Associates and John Wiley & Sons, New York, N.Y.,
 1994).

A nucleic acid encoding an antigen can be cloned or amplified by *in vitro*
 methods, such as the polymerase chain reaction (PCR), the ligase chain reaction
 (LCR), the transcription-based amplification system (TAS), the self-sustained

sequence replication system (3SR) and the Q β replicase amplification system (QB). For example, a polynucleotide encoding the protein can be isolated by polymerase chain reaction of cDNA using primers based on the DNA sequence of the molecule. A wide variety of cloning and *in vitro* amplification methodologies are well known to persons skilled in the art. PCR methods are described in, for example, U.S. Patent No. 4,683,195; Mullis *et al*, *Cold Spring Harbor Symp. Quant. Biol.* 51:263, 1987; and Erlich, ed., *PCR Technology*, (Stockton Press, NY, 1989). Polynucleotides also can be isolated by screening genomic or cDNA libraries with probes selected from the sequences of the desired polynucleotide under stringent hybridization conditions.

The polynucleotides encoding an antigen include a recombinant DNA which is incorporated into a vector into an autonomously replicating plasmid or virus or into the genomic DNA of a prokaryote or eukaryote, or which exists as a separate molecule (such as a cDNA) independent of other sequences. The nucleotides of the invention can be ribonucleotides, deoxyribonucleotides, or modified forms of either nucleotide. The term includes single and double forms of DNA.

DNA sequences encoding the antigen can be expressed *in vitro* by DNA transfer into a suitable host cell. The cell may be prokaryotic or eukaryotic. The term also includes any progeny of the subject host cell. It is understood that all progeny may not be identical to the parental cell since there may be mutations that occur during replication. Methods of stable transfer, meaning that the foreign DNA is continuously maintained in the host, are known in the art.

Polynucleotide sequences encoding antigens can be operatively linked to expression control sequences. An expression control sequence operatively linked to a coding sequence is ligated such that expression of the coding sequence is achieved under conditions compatible with the expression control sequences. The expression control sequences include, but are not limited to, appropriate promoters, enhancers, transcription terminators, a start codon (*i.e.*, ATG) in front of a protein-encoding gene, splicing signal for introns, maintenance of the correct reading frame of that gene to permit proper translation of mRNA, and stop codons.

Hosts can include microbial, yeast, insect and mammalian organisms. Methods of expressing DNA sequences having eukaryotic or viral sequences in

prokaryotes are well known in the art. Non-limiting examples of suitable host cells include bacteria, archaea, insect, fungi (for example, yeast), plant, and animal cells (for example, mammalian cells, such as human). Exemplary cells of use include *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Salmonella*

5 *typhimurium*, SF9 cells, C129 cells, 293 cells, *Neurospora*, and immortalized mammalian myeloid and lymphoid cell lines. Techniques for the propagation of mammalian cells in culture are well-known (see, Jakoby and Pastan (eds), 1979, Cell Culture. Methods in Enzymology, volume 58, Academic Press, Inc., Harcourt Brace Jovanovich, N.Y.). Examples of commonly used mammalian host cell lines
10 are VERO and HeLa cells, CHO cells, and WI38, BHK, and COS cell lines, although cell lines may be used, such as cells designed to provide higher expression desirable glycosylation patterns, or other features.

Transformation of a host cell with recombinant DNA can be carried out by conventional techniques as are well known to those skilled in the art. Where the host
15 is prokaryotic, such as, but not limited to, *E. coli*, competent cells which are capable of DNA uptake can be prepared from cells harvested after exponential growth phase and subsequently treated by the CaCl_2 method using procedures well known in the art. Alternatively, MgCl_2 or RbCl can be used. Transformation can also be performed after forming a protoplast of the host cell if desired, or by electroporation.

20 When the host is a eukaryote, such methods of transfection of DNA as calcium phosphate coprecipitates, conventional mechanical procedures such as microinjection, electroporation, insertion of a plasmid encased in liposomes, or viral vectors can be used. Eukaryotic cells can also be co-transformed with polynucleotide sequences encoding a disclosed antigen, and a second foreign DNA
25 molecule encoding a selectable phenotype, such as the herpes simplex thymidine kinase gene. Another method is to use a eukaryotic viral vector, such as simian virus 40 (SV40) or bovine papilloma virus, to transiently infect or transform eukaryotic cells and express the protein (see for example, *Eukaryotic Viral Vectors*, Cold Spring Harbor Laboratory, Gluzman ed., 1982).

30 A number of viral vectors have been constructed, that can be used to express the disclosed antigens, including polyoma, *i.e.*, SV40 (Madzak *et al*, 1992, *J. Gen.*

Virol, 73:15331536), adenovirus (Berkner, 1992, *Cur. Top. Microbiol. Immunol.*, 158:39-6; Berliner *et al*, 1988, *Bio Techniques*, 6:616-629; Gorziglia *et al*, 1992, *J. Virol*, 66:4407-4412; Quantin *et al*, 1992, *Proc. Natl. Acad. Sci. USA*, 89:2581-2584; Rosenfeld *et al*, 1992, *Cell*, 68:143-155; Wilkinson *et al*, 1992, *Nucl. Acids Res.*, 20:2233-2239; Stratford-Perricaudet *et al*, 1990, *Hum. Gene Ther.*, 1:241-256), vaccinia virus (Mackett *et al*, 1992, *Biotechnology*, 24:495-499), adeno-associated virus (Muzyczka, 1992, *Curr. Top. Microbiol. Immunol.*, 158:91-123; On *et al*, 1990, *Gene*, 89:279-282), herpes viruses including HSV and EBV (Margolskee, 1992, *Curr. Top. Microbiol. Immunol.*, 158:67-90; Johnson *et al*, 1992, *J. Virol*, 66:29522965; Fink *et al*, 1992, *Hum. Gene Ther.* 3:11-19; Breakfield *et al*, 1987, *Mol. Neurobiol.*, 1:337-371; Fresse *et al*, 1990, *Biochem. Pharmacol.*, 40:2189-2199), Sindbis viruses (H. Herweijer *et al*, 1995, *Human Gene Therapy* 6:1161-1167; U.S. Pat. Nos. 5,091,309 and 5,2217,879), alphaviruses (S. Schlesinger, 1993, *Trends Biotechnol.* 11:18-22; I. Frolov *et al*, 1996, *Proc. Natl. Acad. Sci. USA* 93:11371-11377) and retroviruses of avian (Brandyopadhyay *et al*, 1984, *Mol. Cell Biol.*, 4:749-754; Petropoulos *et al*, 1992, *J. Virol*, 66:3391-3397), murine (Miller, 1992, *Curr. Top. Microbiol. Immunol.*, 158:1-24; Miller *et al*, 1985, *Mol. Cell Biol.*, 5:431-437; Sorge *et al*, 1984, *Mol. Cell Biol.*, 4:1730-1737; Mann *et al*, 1985, *J. Virol*, 54:401-407), and human origin (Page *et al*, 1990, *J. Virol*, 64:5370-5276; Buchschalcher *et al*, 1992, *J. Virol*, 66:2731-2739). Baculovirus (*Autographa californica* multinuclear polyhedrosis virus; AcMNPV) vectors are also known in the art, and may be obtained from commercial sources (such as PharMingen, San Diego, Calif.; Protein Sciences Corp., Meriden, Conn.; Stratagene, La Jolla, Calif.).

In some embodiments, a nucleic acid molecule that encodes a disclosed antigen is a nucleic acid given below as any one of SEQ ID NOs: 81-147, 150-161 and 206-215. In some embodiments, a nucleic acid molecule that encodes a disclosed antigen comprises a nucleic acid sequence at least about 95% identical, such as about 95%, about 96%, about 97%, about 98%, about 99% or even 100% identical to the nucleic acid sequence according to one of SEQ ID NOs: 81-147, 150-161 and 206-215. In some embodiments, a nucleic acid molecule that encodes a

disclosed antigen consists of a nucleic acid sequence according to one of SEQ ID NOs: 81-147, 150-161 and 206-215.

5 R2 gpl40(N363Q))(AV123/GSL)(Ap2021)(mFC)/Foldon/3C/His (SEQ ID NO: 116):
 atgcgggtgaagggcatcagacggaactatcagcattgggtggggctggggcaccatgctgctgggactgctgatgatct
 gtagcgccaccgagaaactgtgggtgaccgtgtactacggcgtgcctgtgtggaaagaggccaccaccacctgtttg
 tgcctctgacgccaaggcctatgataccgagggccacaatgtgtgggctactcatgcctgtgtgccaccgatcccaatc
 ctgaggaagtggagctgggtcaacgtgaccgagaactcaacatgtggaagaacaacatggaggagcagatgcacgag
 10 gacatcatcagcctgtgggaccagtctctgaagccttgcgtgaagctgacacctctgtgcgtgggacagcggcagctgca
 acaccagcgtgatcacacaggcctgccccaaagatcagcttcgagcctatcccatccactattgtgccccctgccggttt
 gccatcctgaagtgaacgacaagaagttcagcggcaagggcagctgcaagaacgtgagcaccgtgcagtgtacaca
 cggcatcagacctgtggtgtctacacagctgctgctgaatggctcttgccgaggaagaggtggtgatcagaagcgag
 aatttaccacaacgccaagaccatcatcgtgcagctgagggaaaccgtgaagatcaactgcagccggcccaacaac
 15 aatacccggggcagaagaggagacatcagacagggccactgcaacatcagcaagaccaactggaccaacgcctga
 aacaggtgggtggagaagctgggagcaggttcaacaagaccaagatcgtgttaccagagcagcggcgagatcct
 gagatcgtgaccacagcttcaattgtgccggcgagttcttactgcaataccaccagctgttcgacagcatctggaac
 agcgagaacggcacctggaatatcaccaggggcctgaacaacaccggcaggaacgataccatcacctgcctgcag
 gatcaagggaagcggagcccctccatcaagggaatattagctgcagcagcaacatcacaggactgctgctgacaag
 20 agatggcggaaggacgacaatagcagggacggcaacgagacattcagacctggcgggcggagacatgagggacaa
 ttggcgagcagctgtacaagtacaaggtggtgaagatcgagcccctgggctgggtcccacaaaggccaagagaa
 gagtgggtcagagcgaggagagcgccaccggcctgggcgccacctacatcggtacctgggagccgcccgaagca
 ccatggcgctgcccagcgtgacactgacagtgcaggctagacagctgctgtctggcattgtgcagcagcagagcaatct
 gctgagagccatcgaagcccagcagcacctgctgcagctgacagctgtggggcatcaaacagctgcaggctcgcattct
 25 ggccgtggagagatacctgaaggatcagcagctgctcggaatctggggctgtagcggcaagctgatctgtaccaccac
 cgtgccttggaaatgccagctggtccaagaacaagaccctggaagccatctggaataacatgacctggatgcagtggga
 caaagagatcgacaactacaccagcctgatctacagcctgatcgaggaaagccagatccagcaggaaaagaacgagc
 aggaactgctggaactggacaagtggccaacctgtggaactggttcgacatcagcaactggctgtggggatccggat
 acatccccgaggctccacgcgacggccaggcttacgtgcgcaaggacggcgagtggtgctgctgagcaccttctg
 30 ggcctggaggtgtgttccaggggccaggccaccaccaccaccactga

R2 gpl40(N363Q))(AV13/GSL)(Ap2021)(mFC)/Foldon/3C/His (SEQ ID NO: 117):
 atgcgggtgaagggcatcagacggaactatcagcattgggtggggctggggcaccatgctgctgggactgctgatgatct
 gtagcgccaccgagaaactgtgggtgaccgtgtactacggcgtgcctgtgtggaaagaggccaccaccacctgtttg
 35 tgcctctgacgccaaggcctatgataccgagggccacaatgtgtgggctactcatgcctgtgtgccaccgatcccaatc
 ctgaggaagtggagctgggtcaacgtgaccgagaactcaacatgtggaagaacaacatggaggagcagatgcacgag
 gacatcatcagcctgtgggaccagtctctgaagccttgcgtgaagctgacacctctgtgcgtgggaagcaagaactgca
 gcttcaacattgccacctccatggcgacaagatgcagaaagagtacgccctgctgtacaagctggacatcgagcccat
 cgacaacgacaacaccagctacaggctgatcagctgcaacaccagcgtgatcacaggcctgccccaaagatcagctt
 40 cgagcctatcccatccactattgtgccctgcccgtttgccatcctgaagtgaacgacaagaagttcagcggcaagg
 gcagctgcaagaacgtgagcaccgtgcagtgtacacacggcatcagacctgtggtgtctacacagctgctgctgaatgg
 ctcttgccgaggaagaggtggtgatcagaagcgagaatttaccacaacaacgccaagaccatcatcgtgcagctgag
 ggaacccgtgaagatcaactgcagccggcccaacaatacccggggcagaagaggagacatcagacaggccca
 ctgcaacatcagcaagaccaactggaccaacgcctgaacaggtggtggagaagctgggcgagcagttcaacaaga
 45 ccaagatcgtgttaccagagcagcggcgagatcctgagatcgtgaccacagcttcaattgtccggcgagttcttc

tactgcaataaccaccagctgttcgacagcatctggaacagcgagaacggcacctggaatatcaccaggggcctgaac
 aacaccggcaggaacgataccatcacctgcccctgcaggatcaagggaagcgagccccctccatcaagggaat
 tagctgcagcagcaacatcacaggactgtgtgacaagagatggcggcaaggacgacaatagcagggacggcaac
 gagacattcagacctggcggcgagacatgagggacaattggcggagcgagctgtacaagtacaaggtggtgaagat
 5 cgagccccctgggctggctccacaaaaggccaagagaagagtgtgagagcgaggagagcgccaccggcctggg
 cgccacctacatcggctacctgggagccgcccgaagcaccatgggctgctgagcgtgacactgacagtgcaggcta
 gacagctgctgtctggcattgtgcagcagcagagcaatctgctgagagccatcgaagcccagcagcacctgtgcagc
 tgacagtgtggggcatcaaacagctgcaggctgcattctggcctggagagatacctgaagatcagcagctgtctg
 gaatctggggctgtagcggcaagctgatctgtaccaccacctgcttgggaatgccagctggtccaagaacaagacct
 10 ggaagccatctggaataacatgacctggatgagtgaggacaaagagatcgacaactacaccagcctgatctacagcct
 gatcgaggaaagccagatccagcaggaagaaagcagcaggaactgtggaactggacaagtgggccaacctgtgg
 aactggttcgacatcagcaactggctgtggggatccggatacatccccgaggctccacgcgacggccaggcttacgtg
 cgcaaggacggcgagtggtgctgctgagcaccttctgggctggaggtgctgttccagggcccaggccaccacca
 ccaccacctga
 15 R2 gpl40(N363Q))(AV123/m5)(Ap2021)(mFC)/Foldon/3C/His (SEQ ID NO: 118):
 atgcgggtgaagggcacagacggaactatcagcattggtggggctggggcaccatgctgctgggactgctgatgatct
 gtagcggccaccgagaaactgtgggtgaccgtgtactacggcgtgctgtgtggaagaggccaccaccacctgtttg
 tgctctgacgccaaggcctatgataccgagggccacaatgtgtgggctactcatgcctgtgtgcccaccgatcccaatc
 20 ctgaggaagtggagctggtcaacgtgaccgagaactcaacatgtggaagaacaacatggtggagcagatgcacgag
 gacatcatcagcctgtgggaccagtctgaagccttgcgtgaagctgacacctctgtgctgggcagcggcagctgca
 acaccagcgtgatcacacaggcctgccccaaagatcagcttcgagcctatccccatccactattgtgcccctgcccgttt
 gccatcctgaagtgaacgacaagaagttcagcggcaagggcagctgcaagaacgtgagcaccgtgagtgatcacaca
 cggcatcagacctgtggtgtctacacagctgctgctgaatggctcttggccgaggaagaggtggtgatcagaagcgag
 25 aatttccacaacaacccaagaccatcatcgtgcagctgagggaaaccgtgaagatcaactgcagccggcccaacaac
 aacaccgggggcccggcgggggaagcagcggcgagccactgcaacatcagcaagaccaactggaccaacgccctg
 aaacaggtggtggagaagctgggcgagcagttcaacaagaccaagatcgtgttcaccagagcagcggcgagatcc
 tgagatcgtgaccacagcttcaattgtgcccggcaggttcttactgcaataccaccagctgttcgacagcatctggaa
 cagcgagaacggcacctggaatatcaccaggggcctgaacaacaccggcaggaacgataccatcacctgcccctgca
 30 ggatcaagggaagcggagccccctccatcaagggaatattagctgcagcagcaacatcacaggactgtgtgacaa
 gagatggcggcaaggacgacaatagcagggacggcaacgagacattcagacctggcggcgagacatgagggaca
 attggcggagcgagctgtacaagtacaaggtggtgaagatcgagccccctgggctggctccacaaaaggccaagaga
 agagtgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc
 accatgggctgctgacactgacagtgcaggctagacagctgctgtctggcattgtgcagcagcagcagcagcagc
 35 ctgctgagagccatcgaagcccagcagcacctgtgcagctgacagtgtggggcatcaaacagctgcaggctgcatt
 ctggcctgagagatacctgaaggatcagcagctgctcggaatctggggctgtagcggcaagctgatctgtaccacc
 accgtgccttgggaatgccagctggtccaagaacaagacctggaagccatctggaataacatgacctggatgagtggtg
 gacaaagagatcgacaactacaccagcctgatctacagcctgatcgaggaaagccagatccagcaggaagaaacga
 gcaggaactgctggaactggacaagtgggccaacctgtggaactggttcgacatcagcaactggctgtggggatccgg
 40 atacatccccgaggctccacgcgacggccaggcttacgtgcgaaggacggcgagtggtgctgctgagcaccttct
 gggcctggaggtgctgttccagggcccaggccaccaccaccaccactga

CHIKV-OD A2.6 (SEQ ID NO: 140):

atggagttcatcccgacgcaaaactttctataacagaaggtaccaaccccgaccctgggccccacgcctacaattcaagt
 45 aattagacctagaccagctccacagaggcaggctgggcaactcgcccagctgatctccgcagtcacaaaattgacctg
 cgcgcggtacctcaacagaagcctcgagaaatcggaacaaagaagcaaaaggcagaagaagcagggcgccgcaa

aacgacccaaagcaaaagaagcaaccaccacaaaagaagccggctcaaaagaagaagaaccaggccgtagggag
 agaatgtgcatgaaaattgaaaatgattgcatcttcgaagtcaagcatgaaggcaaagtgatgggctacgcatgcctggt
 gggggataaagtaatgaaaccagcacatgtgaagggaactatcgacaatgccgatctggctaaactggccttaagcgg
 tcgtctaaatacgcattgaatgtgcacagataccgggtgcacatgaagtctgatgcctcgaagttaccacgagaaacc
 5 gaggggtactataactggcatcacggagcagtgagattcaggaggccgggtcactatcccgacgggtgcaggcaag
 ccgggagacagcggcagaccgatcttcgacaacaaaggacgggtgggtggccatcgtcctaggagggggccaacgaag
 gtgcccgcacggccctctccgtgggtgacgtggaacaaagacatcgtcacaaaaattaccctgaggagccgaagagt
 ggagcctcgcctcccggtcttgtgcctgttggaacactacattccctgctctcagccgcttgacaccctgctgct
 acgaaaaggaaaccggaaagcaccttgcgcatgcttgaggacaacgtgatgagaccggatactaccagctactaaaag
 10 catcgtgacttgctctccaccgccaagacgcagtactaaggacaatttaattgtctataaagccacaagaccatatct
 agctcattgtcctgactgcggagaagggtcattcgtgccacagccctatcgattggagcgcacagaaatgaagcaacg
 gacggaacgctgaaaatccagggtctcttgcatgcgggataaagacagatgacagccacgattggaccaagctgcgct
 atatggatagccatacgcacgcggacgcggagcgcgggattgcttgaaggacttcagaccgtgcacgatcaccg
 ggaccatgggacactttattctgcccgatgcccgaaggagagacgctgacagtgggatttacggacgcagaaaga
 15 tcagccacacatgcacacaccgttccatcatgaaccacctgtgataggtaggagaggttccactctgaccacaacat
 ggtaaagagttaccttgacgcacgtacgtgcagagcaccgctgccactgctgaggagatagaggtgcatatgccccca
 gatactcctgaccgcacgctgatgacgcagcagctctggcaacgtgaagatcacagttaatgggcagacgggtgcggtac
 aagtgcaactgcgggtggctccggaagcccgtgggtgagcaccagctgctgctgaacggcagcctggccgagaagg
 agatcaggatcaagagcgagaacatcagcgacaacgccaagaccatcatcgtgcagctgaccaagcccgtgctgatc
 20 aactgcgccaggcccagcaacaacaccaggggagggggcgacatcaggcaggcctactgctggtgaacagg
 acccagtggaacgcacacctgggcccaggtggccatccagctgaggaagcactggaacacctgcatcttcaacgag
 ccagcggcgggcgacctggagatcaccaccacagcttcaactgcggcgggcgagttcacctactgcaacaccagcga
 cctgttcaacagcacctggaacatcaggggcaccgcccagcatcaacggcaccgagagcaacgacaacatcacctgc
 cctgcaggatcaagggcagcggcgccccctccatccaggggcgtgatcaggtgccagagcaacatcaccggcatcctg
 25 ctgaccaggagcggcggaacaccggcaacaacagcaggacctgcgagacctcaggccccggcgggcgacatg
 agggacaactggaggagcgcggggtccggatcaaacgagggactgacaaccacagacaaagtgatcaataactgca
 aaattgatcagtgccatgctgcagtcactaatcacaagaattggcaatacaactcccccttagtcccgcgcaacgctgaac
 tcggggaccgtaaaggaaagatccacatccattcccattggcaaacgtgacttgagagtgccaaaagcaagaaacc
 ctacagtaacttacggaaaaaaccaagtcacatgctgctgtatcctgaccatccgacactctgtcttaccgtaacatggg
 30 acaggaaccaaattaccacgaggagtgggtgacacacaagaaggaggttacctgaccgtgcctactgagggtctgga
 ggtcacttggggcaacaacgaaccatacaagtactggccgcagatgtctacgaacgggtactgctcatggtcaccacat
 gagataatctgtactattatgagctgtaccccactatgactgtatgctcgttggtggcctgctgctgctgctgatgg
 tgggcacagcagtggggaatgtgtgtgtgctgcacggcgagatgcattacacatgatgaattaacaccaggagccactgt
 tcccttctgctcagcctgctatgctgcgtcagaacgaccaaggcgggccacatattacgaggctgcggcatatctatgga
 35 acgaacagcagcccctgttctggtgacggctcttaccgctggccgcttgatcgtcctgtgcaactgtctgaaactctt
 gccatgctgctgaagaccctggcttttttagccgtaatgagcatcggtgccacactgtgagcgcgtacgaacacgtaac
 agtgatcccgaacacgggtgggagtagcgtataagactctgtcaacagaccgggttacagccccatgggtgttgagatg
 gagctacaatcagtcaccttgaaccaacactgtcacttgactacatcacgtgcgagtacaaaactgtcatccccctcccc
 tacgtgaagtgtgtgtgacagcagagtgcaaggacaagacctaccagactacagctgcaaggtctttactggagctca
 40 cccatttatgtggggcgggcctactgcttttcgacgcggaaaatacgaattgagcgaggcacatgtagagaaatctg
 aatcttgcaaaacagagtttgcatcggcctacagagcccacaccgcatcggcgtcggcgaagctccgcgtcctttacca
 aggaacaacattaccgtagctgcctacgctaacggtagcatgcccgtcacagtaaaaggacgccaagtttgctgtgggc
 ccaatgtcctccgctggacaccttttgacaacaaaatcgtgggtgtacaaaggcgacgtctacaacatggactaccacct
 tttggcgaggaagaccaggacaatttgggtgacattcaaagtcgtacaccggaaagtaaaagacgttatgccaacactca
 45 gttggtactacagaggccagcagcaggcacggtacatgtaccatactctcaggcaccatctggcttcaagtattggctga
 aggaacgaggagcatcgctacagcacacggcaccgttcgggtgccagattgcgacaaacccggtaagagctgtaaat

gcgctgtggggaacataccaatttccatcgacataccggatgcggcctttactaggggtgtcgatgcaccctctgtaacgg
 acatgtcatgcgaagtaccagcctgcactcactcctccgactttgggggctgcgccatcatcaaatacacagctagcaag
 aaaggtaaatgtgcagtacattcgtgaccaacgccgttaccattcgagaagccgacgtagaagtagaggggaactccc
 agctgcaaatacttctcaacagccctggcaagcgccgagtttcgctgcaagtgtgtccacacaagtacactgcgca
 5 gccgcatgccaccctccaaaggaccacatagtcattaccagcatcacacaccaccttgggggtccaggatatatcca
 caacggcaatgtcttgggtgcagaagattacgggaggagtaggattaattgtgtgtgtgtgccttaatttaattgtgtg
 ctatgcgtgtcgttagcaggcactaa

CHIKV-0D4.1 (SEQ ID NO: 141):

10 atggagttcatcccgacgcaaactttctataacagaaggtaccaacccccgacctggggccccacgcctacaattcaagt
 aattagacntagaccagtcacagaggcaggctgggcaactcgccagctgatctccgagtcacaaaattgacatg
 cgcgcggtacctaacagaagcctcgcaaaatcgaaaaaacaagaagcaaaaggcagaagaagcaggcgccgcaa
 aacgacccaaagcaaaagaagcaaccaccacaaaagaagccggctcaaaagaagaagaaccaggccgtaggggag
 agaattgtcatgaaaattgaaaatgattgcattctcgaagtcaagcatgaaggcaaaagtgtgggtacgcctgtgt
 15 gggggataaaagtaatgaaaccagcacatgtgaagggaaactatcgacaatgccgatctggctaaactggccttaagcgg
 tcgtctaaatacgtattgaatgtgcacagataccgggtgcacatgaagtctgatgcctcgaagttaccacagagaaacc
 gaggggtactataactggcatcacggagcagtgagtgattcaggaggccgggtcactatcccgacgggtgcaggcaag
 ccgggagacagcggcagaccgatcttcgacaacaaaggacgggtggtggccatcgtcctaggaggggccaacgaag
 gtgcccgcacggccctctccgtggtgacgtggaacaaagacatcgtcacaaaaattacccctgaggggagccgaagagt
 20 ggagcctcgccctcccggtctgtgctgttgcaaacactacattccctgctctcagccgccttgacaccctgtgtgt
 acgaaaaggaaaccggaagcaccttgcgcatgcttgaggacaacgtgatgagacccggatactaccagctactaaaag
 catcgtgactgtctctccaccgcaaaagacgcagtactaaggacaatttaattgtctataaagccacaagaccatatct
 agctcattgtcctgactgcggagaagggcattcgtgccacagccctatcgattggagcgcacagaaatgaagcaacg
 gacggaacgctgaaaatccaggtctcttgcagatcgggataaagacagatgacagccacgattggaccaagctgcgt
 25 atatggatagccatacgccagcggacgcggagcgcgggattgctgtaaggacttcagcaccgtgcacgatcaccg
 ggaccatgggacactttattctcgcccgatgcccgaaaggagagacgctgacagtgaggattacggacagcagaaaga
 tcagccacacatgcacacacccgttccatcatgaaccacctgtgataggtaggagagggttcactctcgaccacaacat
 ggtaaagagttaccttgacgacgtacgtgcagagcaccgctgccactgctgaggagatagaggtgcataatgccccca
 gatactcctgaccgcacgctgatgacgcagcagtgctggaacgtgaagatcacagttaatgggcagacgggtgcggtac
 30 aagtgaactgcgggtggctccggaagacctgtggtgtctacacagctgctgctgaatggctctctggccgaggaagagg
 tggatgacagaagcgagaattcaccaactgcgcaagaccatcatcgtgcagctgagggaacccgtgaagatcaactg
 cagccggcccaacaacaataaccggggcagaagaggagacatcagacaggccactgcaacatcagcaagacca
 ctggaccaacgcctgaaacaggtggtggagaagctgggcgagcagttcaacaagaccaagatcgtgttaccacaga
 gcagcggcgagatcctgagatcgtgaccacagcttaattgtgccggcgagttcacctactgcaataaccaccagct
 35 gttcgacagcatctggaacagcgagaacggcaccttgaatatcaccaggggctgaacaacaccggcaggaacgata
 ccatcaccctgccctgcaggatcaagggaagcggagcccctcccatcaagggaatattagctgcagcagcaacatca
 caggactgctgtgacaagagattgcggcaaggacgacaatagcagggacggcaacgagacattcagacctggcgg
 cggagacatgagggaacattggcggagcggggctccggatcaaacgagggactgacaaccacagacaaagtgatc
 aataactgcaaaattgatcagtgccatgctgcagtcactaatcacaagaattggcaataaactcccccttagtcccgcgca
 40 acgctgaactcggggaccgtaaaggaaagatccacatcccattccattggcaaacgtgacttgacagagtccaaaagc
 aagaaaccctacagtaacttacgaaaaaaacaaagtcacatgctgctgtatcctgacctccgacactctgttaccgt
 aacatgggacaggaaccaaattaccacgaggagtggtgacacacaagaaggaggtaccttgaccgtgctactgag
 ggtctggagggtcacttggggcaacaacgaaccatacaagtaactggccgagatgtctacgaacgggtactgctcatggtc
 acccatgagataatctgtactattatgagctgtacccactatgactgtatgctcattgtgtcgggtggcctcgttcgtctt
 45 gtcgatggtgggcacagcagtggggaatgtgtgtgtgcgcagggcgagatgcattacacatgaattaacaccagga
 gccactgttccctcctgctcagcctgtatgctgcgcagaacgaccaaggcggccacatattacagggtgcggcata

tctatggaacgaacagcagccccctgttctgggtgcaggctcttatcccgctggccgccttgatcgctctgtgcaactgtctg
 aaactcttgccatgctgctgtaagacctggccttttttagccgtaatgagcatcggtgccacactgtgagcgctacgaa
 cacgtaacagtgatcccgaacacgggtgggagtagccgtataagactcttgtcaacagaccgggttacagccccatgggtg
 ggagatggagctacaatcagtcaccttggaaaccaactgtcacttgactacatcacgtgcgagtacaaaactgtcatcc
 5 cctccccgtacgtgaagtgtgtgtgtacagcagagtgcaaggacaagagcctaccagactacagctgcaaggctttac
 tggagtctacccatttatgtggggcggcgcctactgcttttgacgcgcgaaaatacgcaattgagcgaggcacatgtag
 agaaatctgaatcttgcaaacagagtttgcacggcctacagagcccacaccgcatcggtcggtcgaagctccgcgt
 cctttaccaaggaaacaacattaccgtagctgcctacgctaacgggtgaccatgccgtcacagtaaaggacgccaagttg
 tcgtgggcccattgtcctccgctggacaccttttgacaacaaaatcgtggtgtacaaaggcgacgtctacaacatggac
 10 taccaccttttggcgcaggaagaccaggacaatttggtagacattcaaagtcgtacaccggaaagtaaacgctttatgc
 caacactcagttggtactacagaggccagcagcaggcacgggtacatgtaccatactctcaggcaccatctggttcaagt
 attggctgaaggaacgaggagcatcgctacagcacacggcaccgttcggttgccagattgcgacaaaccggtaagag
 ctgtaaattgcgctgtggggaacataccaatttccatcgacataccggatgcggcctttactagggtgtcgatgcacccctc
 tgaacggacatgtatgcgaagtaccagcctgcactcactcctccgactttggggcgctgccatcatcaatacacag
 15 ctgcaagaaaggtaaattgtgcagtacattcgatgaccaacgccgttaccattcgagaagccgacgtagaagtagaggg
 gaactcccagctgcaaatatccttcaacagccctggcaagcgccgagtttcgctgcaagtggtgtccacacaagtac
 actgcgcagccgcatgccaccctcaaaggaccacatagtcattaccagcatcacacaccacccttgggggtccagg
 atatatccacaacggcaatgtcttgggtgcagaagattacgggaggagtaggattaattgttgcgtgtgtgccttaatttta
 atttggtgctatgcgtgtcgttttagcaggcactaa

20

CMV/R RSC3 Y6.1 His (SEQ ID NO: 143):

tcgcgcgtttcgggtgatgacgggtgaaaacctctgacacatgcagctccgggagacgggtcacagcttgtctgtaagcggat
 gccgggagcagacaagcccgtcagggcgctgcagcgggtgttggcggggtgcggggctggcctaactatcgggcatc
 agagcagattgtactgagagtgaccatatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaga
 25 ttggctattggccattgcatacgttgtatccatatcataatgtacatttatattggctcatgtccaacattaccgccattga
 cattgattattgactagttattaatagtaataattacggggctcattagttcatagcccatatatggagttccgcttacataact
 tacggtaaatggccccgctgggtgaccgcccacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaa
 cgccaatagggactttccattgacgtcaatgggtggagattttacggtaaaactgccacttggcagtagacatcaagtgtatca
 30 tatgccaaagtacggccctattgacgtcaatgacggtaaattggccgctggcattatgccagtagacatgacattatggga
 ctttctacttggcagtagacatctacgtattatgcatcgctattaccatgggtgatgcgggtttggcagtagacatcaatgggcgtgg
 atagcgggttgactcacggggatttccaaagtctccacccttgcaggtcaatgggagttgttttggcaccaaaatcaacgg
 gactttccaaaatgtcgtaaacactccgccccattgacgcaaatgggcggttaggcgtgtacgggtgggaggtctatataag
 cagagctcgtttagtaaccgtcagatcgctgtagagacgcatccacgctgttttgacctccatagaagacaccgggacc
 gatccagcctccatcggtcgcacatctctccttcacgcgcccgccttacctgaggccgcatccacgccggttgagtc
 35 gcgttctgcgcctcccgctgtggtgctcctgaactacgtccgctctaggttaagtttagagctcaggtcgagaccg
 ggctttgtccggcgctcccttggagcctacgtacgtcagccggtctccacgctttgcctgacctgcttgcctcaactct
 agttaacgggtggaggagcagtgtagtctgagcagtagtctgtgctgccgcgcgccaccagacataatagctgacagac
 taacagactgttctttccatgggtcttttgcagtcaccgtcgtgcacacgtgtgatcagatatcgggccgctctagaga
 tatcgccaccatgcctatgggatctctgcagccttggccacactgtatctgctgggaatgctggtcgcttctgtgctggcc
 40 acaaacgtgaccgtgaatgtgaccgtgaccttcgattgggtgcgccgatgatattggtggctacaatgaacaccagcatctg
 caccctgtggaacaccagcaacgaccctgcaccaagtgtcctaccgtgcgggttaagcccgtgccatcagatattgtg
 ccctcctggctatgccatcctgaatgcaacaaccgggactttaatggcaccggcccttcacaaaatgtgtccaacgtg
 acctgtacagatggcatccaccctgtggtgtctagtcagctgctgctgaatggcacactggccgatgagaaggtggtgat
 cagaagctgcaacttcagcgacaacgcaagaccatcatcgtgcagctgaacaccagcgtggagatcaattgtacagg
 45 ccaggggcactgcaatatcaccgggccaagtgggaatcagaccctgaagcagatcgccgagaagctgagagagcagt
 tcggcaacaacaagacaatcatcttcaggcctagctctggcgagatcctgagatcgtgaccactggttcaattgcggc

ggcaagttcttactgcaacagcaccagctgttcaacagcacctgggtcaactctacttggagaccaagggcagcaa
 caacaccgagggcagcgataccatcacctgcccctgcaggatcagatctatcacccggcatgaactgcacagtgggcaa
 gatgatctacgcccctcctgtggaaggcaacatcacctgcagcagcaacatcacaggcctgctgctgacaagagatgg
 cggcaacgacaacaacgagagcgagatctttagacctggcggcggagacatgagggacaattggcggagcgagctg
 5 tacaagtacagagtgggtgaggctgaccggatcccatcatcatcatcattagtctggaagggcgaaattgatccagatct
 gctgtgccttctagtgtccagccatctgtgttggcccctccccctgcttcccttgaccctggaaggtgccactccactgt
 cctttcctaataaaatgaggaaattgcatcgattgtctgagtaggtgtcattctattctgggggggtgggggtggggcaggac
 agcaaggggggaggattgggaagacaatagcaggcatgctggggatgcgggtgggctctatgggtaccaggtgtgaa
 gaattgacccgggttctcctggggccagaaaagaagcaggcacatccccttctgtgacacacccctgtccacgcccctggt
 10 tcttagttccagccccactcataggacactcatagctcaggagggtccgccttcaatcccaccgcgtaaagtacttggag
 cggctctcctccctcatcagcccaccaaaccacccagcctccaagagtgggaagaaataaagcaagataggctat
 taagtgcagagggagagaaaatgcctccaacatgtgagggaagtaatgagagaaatcatagaatttaaggccatcatgg
 ccttaatctccgcttcctcgtcactgactcgctgcgctcggtcgttcggctgcggcgagcggatcagctcactcaaaag
 gcggtaatacgggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccag
 15 gaaccgtaaaaaggccggtgctggcgttttccataggctccgccccctgacgagcatcaaaaaatcgacgctca
 agtcagaggtggcgaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctcctg
 ttccgaccctgccgttacccgatacctgtccgcctttctccttcgggaagcgtggcgttttctatagctcacgctgtag
 gtatctcagttcgggtgtaggtcgttcgctcaagctgggctgtgtgcacgaacccccgttcagcccagccgtgcgctt
 atccggtaaactatcgtcttgagtccaacccggtgaagacacgacttatcgccactggcagcagccactggtaacaggatta
 20 gcagagcgaggtatgtaggcgggtgctacagagttctgaagtgggtggcctaactacggctacactagaagaacagtattt
 ggtatctgcgctcgtgtaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaaacaaccaccgctggt
 agcgggtggtttttgttgcaagcagcagattacgcgcagaaaaaaggatctcaagaagatcctttgatctttctacggg
 gtctgacgctcagtggaacgaaaactcacgttaagggttttgggtcatgagattatcaaaaaggatcttcacctagatcctt
 taaataaaaatgaagtttaaatcaatctaaagtatatatgagtaaaacttggctgacagttaccaatgcttaacagtgagg
 25 cacctatctcagcgatctgtctatttcgttcatccatagttgcctgactcggggggggggggcgctgaggtctgcctcgtg
 aagaaggtgttgctgactcataccaggcctgaatcgccccatcatccagccagaaaagtgagggagccacgggtgatga
 gagctttgtttaggtggaccagttggtgattttgaacttttgccttgccacggaacggctcgtgttgctgggaagatgcgtg
 atctgatccttcaactcagcaaaagttcgattattcaacaaagccgcccgtcaagtacgctaattgctctgccagt
 ttacaaccaattaaccaattctgattagaaaaactcatcgagcatcaaatgaaactgcaattattcatatcaggattatcaat
 30 accatattttgaaaaagccgtttctgtaatgaaggagaaaaactcaccgaggcagttccataggtggcaagatcctggtat
 cggctcgcgattccgactcgtccaacatcaatacaacctattatccccctcgtcaaaaataaggttatcaagtgaagaaatc
 accatgagtgacgactgaatccggtgagaatggcaaaagcttatgcatcttccagactgttcaacagggccagccatta
 cgctcgtcatcaaaatcactcgcatacaacaaaccgttattcattcgtgattgcgcctgagcgagacgaaatcgcgatcg
 ctgttaaaaggacaattacaaacaggaatcgaatgcaaccggcgaggaacactgccagcgcatcaacaatattttcac
 35 ctgaatcaggatattcttctaatacctggaatgctgtttccggggatcgagtggtgagtaacctgcatcatcaggagt
 acggataaaatgcttgatgggtcggaagaggcataaattccgtcagccagtttagtctgacctctcatctgtaacatcattg
 gcaacgctacctttgccatgtttcagaaacaactctggcgcacatggggcttccatacaatcgatagattgtcgcacctgatt
 gcccagattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaacgcggcctcgagcaagacgt
 ttcccggtgaatatggctcataacacccctgtattactgtttatgtaagcagacagttttattgttcatgatgatattttatctt
 40 gtgcaatgtaacatcagagattttgagacacaacgtggccttccccccccccattattgaagcatttatcagggttattgt
 ctcatgagcggatacatatttgatgtatttagaaaaataaacaataaggggtccgcgcacattccccgaaaagtgccac
 ctgacgtctaagaaaccattattatcatgacattaacctataaaaataggcgtatcacaggccctttcgtc

CMV8x/R_KER2018_OD4.2'_HG3.2_HATM (SEQ ID NO: 144):

45 tcgcgcgtttcgggtgatgacgggtgaaaacctctgacacatgcagctcccgagacgggtcacagcttctgtgaagcggat
 gccgggagcagacaagcccgctcagggcgcgctcagcgggtgttggcgggtgtcggggctggcttaactatgcggcatc

agagcagattgtactgagagtgcaccatatgCGgtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaga
 ttggctattggccattgcatacgttgatccatatcataatatgtacatttatattggctcatgtccaacattaccgccatgttga
 cattgattattgactagtattataatagtaatacaattacgggaacttccatagcccatatatggagttccgcgttacataacttac
 gggaatttccaaacctggctgaccgcccacgacccccgccattgacgtcaataatgacgtatgtcccatagtaacgc
 5 caatagggaaacttccattgacgtcaatgggtggagattttacggtaactgccacttgggaatttccaaagtgtatcatatgc
 caagtacgccccctattgacgtcaatgacgggaacttccataagcttgacattatgccagttacatgaccttatgggaatttc
 ctacttggcagttacatctacgtattagtcacgtctattaccatgggtgatgcggttttggcagttacatcaatgggcgtggatag
 cggtttgactcacgggaacttccaaagtctccacccattgacgtcaatgggagtttgtttgactcacaaaaatcaacggga
 attcccaaaatgtcgtaaacaactccgccccattgacgcaaatgggcggttaggcgtgtacgggtgggaggtctatataagca
 10 gagctcgttttagtgaaccgtcagatcgctggagacgccatccacgtgtttgacctccatagaagacaccgggaccg
 atccagcctccatcggtcgcacgtctctcttcacgcgccccgccctacctgaggccgccatccacgcgggttgagtc
 gcgttctgcgcctcccgcctgtggtgcctcctgaactgcgtccgcgtctaggttaagttaaagctcaggtcgagaccg
 ggctttgtccggcgctcccttggagcctacatagactcagccggctctccacgtttgctgacctgcttgcctcaactct
 agttaacgggtggaggcagtgtagtctgagcagttactcgttgcgtccgcgcgccaccagacataatagctgacagac
 15 taacagactgttcttccatgggtcttttctgcagtcaccgtcgtcgacacgtgtgatcagatatcgcgccgctctagaca
 ccatgccccatgggcagcctgcagcccctggccaccctgtacctgtgggcatgctggtggctagcgtgctggccaagc
 ccgtggtagcaccagctgctgctgaacggcagcctggccgagaaggagatcaggatcaagagcaggacatcag
 cgacaacccaagaacatcatcgtgcagctgaccaagcccgtgctgatcaactgcaccaggcccagcaacaacccc
 agtactgcgtggtgaacaggaccagtggaacgacaccctgggcccaggtggccatccagctgaggaagcactggaac
 20 acctgcatcatctcaacgagcccagcgcgccgacctggagatcaccacccacagcttcaactgcggcgccgagttc
 TTCtactgcaacaccagcgacctgttcaacagcacctggaacatcaacggcaccgccagcatcaacggcaccgaga
 gcaacgacaacatcacctgccctgcaggatcaacggcaccatccagggaacatcacctgccagagcaacatcacc
 ggcatcctgtgaccagggacggcgccagcgccagcgccacctgcgagacctcaggcccggcgccgagcatg
 agggacctgaacaggaccagcgGATCCGGCATCCTGGCCATCTACAGCACCGTGGCC
 25 AGCAGCCTGGTGCTGCTGGTGAGCCTGGGCGCCATCAGCTTCTAgtacctagc
 atcatcatcatcatcattagtctgaagggcgaattgatccagctgtgccttctagtgtccagccatctgtgttttccccctccc
 ccgtgccttcttgacctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgattgtctgagt
 aggtgtcattctattctgggggggtgggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgc
 tggggatcggggtgggtctatgggtaccaggtgctgaagaattgaccgggttctcctgggcccagaaagaagcaggc
 30 acatccccctctctgtgacacaccctgtccacgcccctgggtcttagttccagccccactcataggacactcatagctcagg
 agggctccgcttcaatcccacccgctaaagtacttggagcgggtctctccctccctcatagcccacaaacaaaccta
 gcctccaagagtgggaagaaataaagcaagataggctattaagtgcagagggagagaaaatgcctccaacatgtgag
 gaagtaatgagagaaatcatagaattttaaggccatgatttaaggccatcatggccttaattctccgcttctcgtcactga
 ctcgctgcgtcggctcgttcgggtcgggcgagcgggtatcagctcactcaaaggcggtaatacgggttatccacagaatca
 35 ggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaacctgaaaaaggccggttgcgtg
 gcgtttttccataggctccgccccctgacgagcatcacaataatcgacgtcaagtgcagaggtggcgaaacccgacag
 gactataaagataccaggcggtttccccctggaagctccctcgtgcgtctcctgttccgacctgcccgttaccggatacc
 tgtccgctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgct
 ccaagctgggctgtgtgcacgaacccccgttcagcccagccgctgcgccttatccggtaactatcgtcttgagttccaac
 40 ccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgct
 acagagttcttgaagtgggtggcctaactacggctacactagaagaacagttatgtgtatctgcgtctgctgaagccagtta
 cttcggaaaaagagttgtagcttctgatccggcaaaacaaaccaccgctggtagcggtggtttttgtttgcaagcagca
 gattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactc
 acgttaagggtattttggtcatgagattatcaaaaaggatcttcacctagatccttttaattaaaaatgaagtttaaatcaatct
 45 aaagtatatatgagtaacttggctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttc
 atccatagttgcctgactcggggggggggggcgctgaggtctgcctcgtgaagaaggtgtgtgactcataccaggcc

tgaatcgccccatcatccagccagaaagtgagggagccacgggtgatgagagctttgttaggtggaccagttgggtgat
 ttgaaacttttgccttccacgggaacgggtctgcgttgcgggaagatgcgtgatctgaccttcaactcagcaaaagttcgatt
 tattcaacaaagccgcccgtcccgtcaagtcagcgaatgctctgccagtgttacaaccaattaaccaattctgattagaaaa
 actcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataccatattttgaaaaagccgttctgtaataga
 5 aggagaaaactcaccgaggcagttccataggtatggcaagatcctggatcgggtctgcgattccgactcgtccaacatcaa
 tacaacctattaatttccccctcgtaaaaaataaggttatcaagtgagaaatcacatgagtgacgactgaatccgggtgagaa
 tggcaaaagcttatgcatttcttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcacatcaacca
 aaccgttattcattcgtgattgcgcctgagcgagacgaaatacgcgatcgtgttaaaaggacaattacaacagggaatc
 gaatgcaaccggcgaggaacactgccagcgcacaaatattttcacctgaatcaggatattcttctaatacttggaat
 10 gctgttttccggggatcgcagtggtgagtaacatgcatcatcaggagtagcgataaaatgcttgatggctcggaagagg
 cataaattccgctcagccagtttagtctgacctctcatctgtaacatcattggcaacgctacctttgccatgttcagaaacaa
 ctctggcgcatcgggcttccatacaatcgatagattgtcgcacctgattgcccagattatcgcgagccatttatacca
 tataaatcagcatccatgttggaattaatcgcgccctcgagcaagacggttcccggtgaatatggctcataacacccctgt
 attactgtttatgaagcagacagttttattgttcatgatgatattttatctgtgcaatgaacatcagagattttgagacaca
 15 acgtggctttccccccccccattattgaagcatttatcagggttattgtctcatgagcggatacatatttgatgtatttaga
 aaaataacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccattattatcatgacatt
 aacctataaaaatagcggtatcacgagggcctttcgtc

CMV8x/R_KER2018_OD4.2'_HG3.2.1_HATM (SEQ ID NO: 145):

20 tcgcgcgtttcgggtgatgacggtgaaaacctctgacacatgcagctcccggagacgggtcacagcttctgttaagcggat
 gccgggagcagacaagcccgtcagggcgctcagcgggtgttggcgggtgtcggggctggcttaactatgcggcatc
 agagcagattgtactgagagtgaccatatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaga
 ttggctattggccattgcatacgttgcataatgtacatttatattggctcatgtccaacattaccgccatgttga
 cattgattattgactagtattataatgaatcaattacgggaacttccatagcccatataggagtccgcgttacataacttac
 25 gggaatttccaaacctggctgaccgcccacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgc
 caataggggaacttccattgacgtcaatgggtggagttttacggtaaaactgccacttgggaatttccaaagtatcatatgc
 caagtacgccccctattgacgtcaatgacgggaacttccataagcttgattatgccagtagacattatgggaatttc
 ctacttggcagtagacatctatgtagtcatcgtattaccatgggtgatgcggttttggcagtagacatcaatgggcgtggatag
 cgggttgactcacgggaacttccaaagtctccacccattgacgtcaatgggagtttgggttactacaaaatcaacggga
 30 attcccaaaatgtcgaactccgccccattgacgcaaatgggcggtaggcgtgtacgggtgggaggtctatataagca
 gagctcgttttagtaaccgtcagatcgctggagacgccatccacgctgtttgacctcatagaagacaccgggaccg
 atccagcctccatcggtcgcacatctctcttcacgcgcccgcgacctgagggcccatccacgcccgttgagtc
 gcgttctgccgctcccgcctgtggtgcctcctgaactgcgtccgcttaggtaagtttaaagctcaggtcgagaccg
 ggcttttgcggcgctccctggagcctacctagactcagccggctctccacgcttgcctgacctgttgcacactct
 35 agttaacgggtggagggcagtgtagtctgagcagtagctgttgcgcgcgcccaccagacataatagctgacagac
 taacagactgttctttccatgggtcttttctgcagtcaccgtcgtcgacacgtgtgatcagatatcgcgccgctctagaca
 ccatgccccatgggcagcctgcagccccggccacccgtgtacctgctgggcatgctggtggctagcgtgctggccaagc
 ccgtggtgagcaccagctgctgctgaacggcagcctggccgagaaggagatcaggatcaagagcgaggacatcag
 cgacaacgccaagaacatcatcgtgcagctgaccaagcccgtgctgatcaactgcaccaggcccagcaacaacccc
 40 agtactgcgtggtgaacaggacccagtggaacgacaccctgggcccaggtggccatccagctgaggaagcactggaac
 acctgcatcatcttccaaacagcagcggcgacctggagatcaccacccacagcttcaactgcggcgccgaggttc
 TTCtactgcaacaccagcgacctgttcaacagcacttgaacatcaacggcaccgcccagcatcaacggcaccgaga
 gcaacgacaacatcacctgccctgcaggatcaacggcaccatccagggcaacatcacctgccagagcaacatcac
 ggcacatcctgctgaccaggagcggcgagcggcagcggcacctgcgagaccttcaggcccggcgggcgacatg
 45 agggacctgaacaggaccagcgGATCCGGCATCCTGGCCATCTACAGCACCGTGGCC
 AGCAGCCTGGTGCTGCTGGTGAGCCTGGGCGCCATCAGCTTCTAgatcctagc

- atcatcatcatcattagctgaagggcgaattgatccagctgtgccttctagttgccagccatctgtgtttgccctccc
ccgtgccttccctgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaatgcatcgattgtctgagt
aggtgtcatttctattctgggggggtgggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgc
tggggatgcggtgggctctatgggtaccaggtgctgaagaattgaccgggtcctcctgggccagaaagaagcaggc
5 acatcccccttctgtgacacaccctgtccacgcccctgggttcttagttccagccccactcataggacactcatagctcagg
aggggtccgccttcaatcccaccgctaagtaacttggagcgggtctctccctccctcatcagcccacaaacaaaccta
gcctcaagagtgaggaaataaagcaagataggctattaagtgcagagggagagaaaaatgcctccaacatgtgag
gaagtaatgagagaatcatagaatttaaggccatgatttaaggccatcatggccttaattctccgcttctcgtcactga
ctcgtcgcgtcgggtcgttgggtcgggtcgggtatcagctcactcaaaggcggtaatacgggtatccacagaatca
10 ggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaaccgtaaaaaggccggttgctg
gcgtttttccatagggtccgccccctgacgagcatcacaataatcgacgtcaagtcagaggtggcgaaaccgcagac
gactataaagataccaggcgtttccccctggaaagctccctcgtgcgctcctcgttccgacctgcccgttaccggatacc
tgtccgcctttctccctcgggaagcgtggcgcttttctcatagctcacgctgtaggtatctcagttcgggtgtaggtcgttcgct
ccaagctgggctgtgtgcacgaacccccgttcagcccagccgctgcgccttatccgtaactatcgttctgagtccaac
15 ccggttaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggggtgct
acagagttctgaagtgggtggcctaactacggctacactagaagaacagtatttggtatctgcgctcgtcgaagccagtta
ccttcgaaaaagagttggtagctcttgatccggcaaaacacaccgctggtagcgggtggtttttgttgcaagcagca
gattacgcgcagaaaaaaggatctcaagaagatcctttgatctttctacggggtctgacgtcagtggaacgaaaactc
acgttaagggtatttggtcatgagattatcaaaaaggatcttcacctagatccttttaataaaaaatgaagtttaataatct
20 aaagtatatatgagtaaaacttggtctgacagttaccaatgcttaacagtgaggcacctatctcagcgatctgtctatttcgttc
atccatagttgcctgactcggggggggggggcgctgaggtctgcctcgtgaagaaggtgttgctgactcataccaggcc
tgaatcgccccatcatccagccagaaagtgaggagccacgggtgatgagagctttgttgtaggtggaccaggttggtgat
tttgaacttttgctttgccacggacgggtcgtgttcgggaagatgcgtgatctgatccttcaactcagcaaaagttcgatt
tattcaaaaagccgctcccggtcaagtcagcgtaatgctctgccagtggtacaaccaattaaccaattctgattagaaaa
25 actcatcgagcatcaaatgaaactgcaattattcatatcaggattatcaataccatattttgaaaaagccgttctgtaatga
aggagaaaaactcaccgaggcagttccataggatggcaagatcctgggtatcggtctgcgattccgactcgtccaacatcaa
tacaacctattaatttccctcgtcaaaaataagggtatcaagtgagaaatcaccatgagtgacgactgaatccggtgagaa
tggcaaaagcttatgcatttcttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcataacca
aaccgttattcattctgattgcgctgagcgagacgaaatacgcgatcgtgttaaaaggacaattacaacaggaatc
30 gaatgcaaccggcgaggaacactgccagcgcatcaacaatattttcacctgaatcaggatattcttctaatacctggaat
gctgttttccggggatcgagtggtgagtaacatgcatcatcaggagtacggataaaatgcttgatggtcggaagagg
cataaattccgtcagccagtttagtctgacctctcatctgtaacatcattggcaacgctacctttgcatgttcagaaaca
ctctggcgcatcggggttcccatacaatcgatagattgtcgcacctgattgcccacattatcgcgagcccatttatacca
tataaatcagcatccatgttggaattaatcgcgccctcgagcaagacgtttccggtgaatatggctcataacacccttgt
35 attactgtttatgaagcagacagttttattgttcatgatgatattttatcttgcaatgaatacagagattttgagacaca
acgtggctttccccccccccattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttaga
aaaataaaacaataggggtccgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccattattatcatgacatt
aacctataaaaataggcgtatcacgagcccttctgc
- 40 CMV8x/R_KER2018_OD4.2'_HG3.2_NATM (SEQ ID NO: 146):
tcgcgcgtttcgggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgctgtaagcggat
gccgggagcagacaagcccgtagggcgctcagcgggtgttgccgggtgtcggggctggcttaactatcgcgcatc
agagcagattgtactgagagtgcaccatatcggtgtgaaataaccgcacagatgcgtaaggagaaaaataccgcatcaga
ttggctattggccattgcatacgttgatccatatacataatgtacatttatattggctcatgtccaacattaccgcatgtga
45 cattgattattgactagtatttaataagtaataacggaacttccatagcccatatatggagtccgcgttacataacttac
gggaatttccaaacctgggtgaccgccaacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgc

caatagggaaactccattgacgtcaatgggtggagtagtttacggtaaactgccacttgggaatttccaagtgtatcatatgc
 caagtacgccccctattgacgtcaatgacgggaacttcataagcttgacattatgccagtagacacattatgggaatttc
 ctacttggcagtagacatctacgtattagtcacgtcattaccatgggtgatgcgggtttggcagtagacatcaatgggcgtggatag
 cgggttgactcacgggaactccaagtctccacccattgacgtcaatgggagttgtttgactcacaaaaatcaacggga
 5 attcccaaaatgtcgtacaactccgccccattgacgcaaatgggcggtaggcgtgtacgggtgggaggtctatataagca
 gagctcgttttagtgaaccgtcagatcgctggagacgccatccacgctgtttgacctcatagaagacaccgggaccg
 atccagcctccatcggtcgcacgtctctcttcacgcgcccggcgccctacctgaggccgcatccacgcccgttgagtc
 gcgttctgccgctcccgctgtggtgctcctgaactgcgtccgctctaggttaagtttaaagctcaggtcgagaccg
 ggctttgtccggcgctcccttggagcctacctagactcagccggctctccacgctttgctgacctgcttgcctcaactct
 10 agttaacgggtggaggcagtgtagtctgagcagtagctgttgcgcgcgcgccaccagacataatagctgacagac
 taacagactgttctttccatgggtctttctgacgtcaccgtcgtcgacacgtgtgatcagatatcgggccgcttagaca
 ccATGAACCCCAACCAGAAGATCATCACCATCGGCAGCATCTGCATGGTG
 GTGGGCATCATCAGCCTGATCCTGCAGATCGGCAACATCATCAGCATCT
 GGGTGTCCGGAaagcccgtggtgagcaccagctgctgctgaacggcagcctggccgagaaggagatcag
 15 gatcaagagcgaggacatcagcgacaacgccaagaacatcatcgtgcagctgaccaagcccgtgctgatcaactgca
 ccaggcccagcaacaacaccagtagctgctggtgaacaggaccagtggaacgacaccctgggcccaggtggccat
 ccagctgaggaagcactggaacacctgcatcttcaacgagcccagcgggcgacactggagatcaccaccaca
 gcttcaactgcggcgaggttcTTCtactgcaacaccagcgacactgttcaacgacactggaacatcaacggcacc
 gccagcatcaacggcaccgagagcaacgacaacatcacctgcccctgcaggatcaacggcaccatccagggcaacat
 20 cacctgccagagcaacatcacccggcatcctgctgaccaggacggcgccgagcgccgagcgccacactgcgagaccttc
 agggccggcgccgagacatgaggacactgaacaggaccagcggtcctagcatcatcatcatcattagctgaa
 gggcgaaatgatccagctgtgcttctagttgccagccatctgtgttgccttccccctgcttcttgccttgccttgaag
 gtgcccactcccactgtccttctataaaaatgaggaaatgcatcgattgtctgagtaggtgtcattctattctggggggtg
 ggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgtgggtagcggtgggctctatgg
 25 gtacccaggtgtgaaagaattgaccgggttctcctgggcccagaaagaagcaggcacatccccttctctgtgacacacc
 ctgtccacgccccctggttcttagttccagccccactcataggacactcatagctcaggagggtcccgcttcaatcccacc
 cgctaaagtacttggagcggtcctcctccctcatcagcccacaaacaaacctagcctccaagagtgggaagaaatt
 aaagcaagataggctattaagtgcagagggagagaaaatgctccaacatgtgaggaaatgagagaaatcataga
 attttaaggccatgatttaaggccatcatggccttaattctccgcttctcgtcactgactcgtcgcctcgttccggtc
 30 gcggcgagcggtatcagctcactcaaaggcggtataacgggtatccacagaatcaggggataacgcaggaaagaacat
 gtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccggttgcgtggttccataggtccgcccccc
 tgacgagcatcacaataatcagctcaagtgcagaggtggcgaaaccgacaggactataagataaccaggcggttcc
 ccctggaagctccctcgtgcgtcctcgttccgacctgcccgttaccggatacctgtccgcttctcccttcgggaagc
 gtggcgcttctcatagctcagctgtaggtatctcagttcggtgtaggtcggtcgtccaagctgggctgtgtgcacgaac
 35 cccccgttcagcccagcgctgccttatccggttaactatcgttgcgtcgaacccggtaagacagacttatcgcca
 ctggcagcagccactggtaacaggattagcagagcgaggtatgtaggggtgtacagagttcttgaagtgggtggccta
 actacggctacactagaagaacagtagtttggtatctgcgtcgtgctgaagccagttaccttcggaaaaagagtggtagct
 cttgatccggcaaaacaccaccgctggttagcggtggtttttgttgcaagcagcagattacgcgcagaaaaaaggat
 ctcaagaagatccttctgatctttctacggggctgacgctcagtggaacgaaaactcacgttaagggttttgggtcatgag
 40 ctgacagttaccaatgcttaacagtgaggcacctatctcagcgatctgtctatttcgttcatcatagttgcctgactcgggg
 gggggggcgctgaggtctgcctcgtgaagaagggtgttgcgtactataccaggcctgaatgccccatcatccagcc
 agaaagtgaggagccacgggtgatgagagcttgtgttaggtggaccagttggtgattttgaacttttgccttccacgga
 acgggtctgcgttgcgggaagatgcgtgatctgacgttcaactcagcaaaagttcgattattcaacaaagcccgctccc
 45 gtcaagtcagcgtaatgctctgccagtggttacaaccaattaaccaattctgattagaaaaactcatcgagcatcaaatgaaa
 ctgcaatttattcatatcaggattatcaataccatatttttgaaaaagccgttctgtaaatgaaggagaaaactcaccgaggca

- gttccataggatggcaagatcctgggtatcggtctgcgattccgactcgtccaacatcaatacaacctattaatttcccctcgt
caaaaataaggttatcaagtgagaaatcacatgagtgacgactgaatccgggtgagaatggcaaaagcttatgcatttcttt
ccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcacatcaacaaaccgttattcattcgtgattgc
gcctgagcgcgagacgaaatacgcgatcgtgtttaaaggacaattacaaacaggaatcgaatgcaaccggcgagga
5 cactgccagcgcacatcaacaatattttcacctgaatcaggatattcttctaatacctggaatgctgttttccgggggagcgcag
tggtgagtaacatgcatcatcaggagtagcgataaaatgcttgatggctcggaagaggcataaattccgtcagccagtta
gtctgacctctcatctgtaacatcattggcaacgctacctttgccatgttcagaaacaactctggcgcatcgggcttccca
tacaatcgatagattgtcgacctgattggccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggga
atthaatcgcgccctcgagcaagacgtttccgttgaaatggctcataacacccctgtattactgtttatgtaagcagaca
10 gttttattgttcatgatgatataattttatcttgtgcaatgtaacatcagagattttgagacacaacgtggctttcccccccccc
cattattgaagcatttatcaggggtattgtctcatgagcggatacatattgaatgtatttagaaaaataaacaataggggttc
cgcgacatttccccgaaaagtccacctgacgtctaagaaaccattattatcatgacattaacctataaaaaatagggcgtat
cacgaggcccttctgc
- 15 CMV8x/R_KER2018_OD4.2'_HG3.2.1_NATM (SEQ ID NO: 147):
tcgcgcgtttcgggtgatgacgggtgaaaacctctgacacatgcagctcccggagacgggtcacagcttgtctgtaagcggat
gccgggagcagacaagcccgtcagggcgctcagcgggtgttggcggggtgtcggggctggcttaactatgcggcatc
agagcagattgtactgagagtgcacatgacgggtgtgaaataccgcacagatgcgtaaggagaaaataccgcacaga
ttggctattggccattgcatacgttgatccatatcataatgtacatttatattggctcatgtccaacattaccgccatgttga
20 cattgattattgactagtattaatagtaataacacgggaacttccatagcccataatggagttccgcgttacataacttac
gggaatttccaaacctggctgaccgccaacgacccccgccattgacgtcaataatgacgtatgttccatagtaacgc
caatagggaacttccattgacgtcaatgggtggagtatttaccggtaaactgccacttgggaatttcaagtgtatcatatgc
caagtacgccccctattgacgtcaatgacgggaacttccataagcttgacattatgccagtagcatgacctatgggaatttc
ctacttggcagtagcatctacgtattatgcatcgtattaccatgggtgatgcgggttttggcagtagcatcaatgggcgtggatag
25 cgggttgactcacgggaacttccaaagtctccacccattgacgtcaatgggagtttgtttgactacaaaaatcaacggga
attcccaaaatgtcgtaaactccgccccattgacgcaaatggcggttaggcgtgtacgggtgggaggtctatataagca
gagctcgttttagtgaaccgtcagatgcctggagacgccatccacgctgtttgacctccatagaagacaccgggaccg
atccagctccatcggtcgcacatctctcttcacgcgcccgcctacctgaggccgcatccacgcccgttgagtc
gcgttctgccgctcccgtgtgtgctcctgaactcgtccgcttaggtaagtttaagctcaggtcgagaccg
30 ggcctttgtccggcgctccctggagcctacctagactcagccggctctccacgcttgctgacctgtgtcactct
agttaacgggtggaggcagtgtagtctgagcagtagctgtgtcgcgcgcgcgccaccagacataatagctgacagac
taacagactgttctttccatgggtcttttgcagtcaccgctcgtcgacacgtgtgatcagatcgcggccgcttagaca
ccATGAACCCCAACCAGAAGATCATCACCATCGGCAGCATCTGCATGGTG
GTGGGCATCATCAGCCTGATCCTGCAGATCGGCAACATCATCAGCATCT
35 GGGTGTCCGGAaagcccgtggtgagcaccagctgctgctgaacggcagcctggccgagaaggagatcag
gatcaagagcgaggacatcagcgacaacgccaagaacatcatcgtgcagctgaccaagcccgtgctgatcaactgca
ccaggcccagcaacaacacccagtagctggtggaacaggaccagtggaacgacacccctgggccaaggtggccat
ccagctgaggaagcactggaacacctgcatcatcttcaccaacagcagcggcgggcagctggagatcaccaccaca
gcttcaactcggcgggcgagttcTTCtactgcaacaccagcgacctgttcaacagcacctggaacatcaacggcacc
40 gccagcatcaacggcaccgagagcaacgacaacatcacctgccctgcaggatcaacggcaccatccagggaacat
cacctgccagagcaacatcacggcatcctgtgaccagggacggcggcagcggcgagcggcacctgcgagaccttc
aggcccgggcgggcgacatgaggacctgaacaggaccagcgatcctagcatcatcatcatcattagctgaa
gggcgaattgatccagctgtgccttctagtgtccagccatctgtttgttccctccccctgccttcttgacctggaag
gtgccactcccactgtccttcttaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggtg
45 ggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcgggtgggctctatgg
gtacccaggtgctgaagaattgacccgggtcctcctgggcccagaagaagcaggcacatcccccttctgtgacacacc

ctgtccacgcccctgggtcttagttccagccccactcataggacactcatagctcaggagggctccgcttcaatcccacc
 cgctaaagtacttggagcgggtctctccctccctcatcagcccacaaacaaacctagcctccaagagtgggaagaaatt
 aaagcaagataggctattaagtgcagaggggagagaaaatgcctccaacatgtgaggaagtaatgagagaaatcataga
 attttaaggccatgatttaaggccatcatggccttaattctccgcttctcgtcactgactcgtcgcgtcggctgctcggct
 5 gcggcgagcgggtatcagctcactcaaaggcggtaatacgggtatccacagaatcaggggataacgcaggaaagaacat
 gtgagcaaaaggccagcaaaaggccaggaaaccgtaaaaaggccggttgctggcggttttccataggctccgcccccc
 tgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcggttcc
 ccctggaagctccctcgtgcgctctcgttccgacctgcccgttaccggatacctgtccgcttttctccttcgggaagc
 gtggcgcttttctatagctcacgctgtaggtatctcagttcgggtgtaggtcgttcgctccaagctgggctgtgtgcacgaac
 10 cccccgttcagcccagccgctgcgccttatccggttaactatcgtctgagtccaacccggttaagacacgacttatcgcca
 ctggcagcagccactggtaacaggattagcagagcgggtatgtaggcgggtgctacagagttctgaagtgggtggccta
 actacggctacactagaagaacagttatgtgctcgtcgtgaagccagttaccttcggaaaaagagttggttagct
 ctgatccggcaaaacaaaccaccgctggtagcgggtggtttttgttgcaagcagcagattacgcgcagaaaaaaaggat
 ctcaagaagatcctttgatctttctacggggtctgacgctcagtggaacgaaaactcacgttaagggttttggctcatgag
 15 attatcaaaaaggatcttcacctagatccttttaataaaaaatgaagtttaaatcaatctaaagtatatatgagtaaaacttgg
 ctgacagttaccaatgcttaatcagtgaggcacctatctcagcgtatctgtctatttcgttcatccatagttgctgactcgggg
 gggggggggcgctgaggtctgcctcgtgaagaagggtgtgctgactcataccaggcctgaatcgccccatcatccagcc
 agaaagtgagggagccacgggtgatgagagcttggtaggtggaccagttgggtgattttgaacttttgccttgccacgga
 acggtctcgttgcgggaagatgcgtgatctgatccttaactcagcaaaagttcgattattcaacaaagccgcccgtccc
 20 gtcaagtcagcgaatgctctgccagtggtacaaccaattaaccaattctgattagaaaaactcatcgagcatcaaatgaaa
 ctgcaatttattcatatcaggattatcaataccatattttgaaaaagccgtttctgtaatgaaggagaaaactcaccgaggca
 gttccataggatggcaagatcctggtatcggtctgcgattccgactcgtccaacatcaataaacctattaatttccctcgt
 caaaaaataaggttatcaagtgagaaatcacctagtgacgactgaatccgggtgagaatggcaaaagcttatgcatttctt
 ccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcacatcaacaaaccgttattcattcgtgattgc
 25 gcctgagcgagacgaaatacgcgatcgtgttaaaaggacaattacaaacaggaatcgaatgcaaccggcgagga
 cactgccagcgcatcaacaatattttcacctgaatcaggatattcttctaatacctggaatgctgttttccggggatcgag
 tggtagtaacatgcatcatcaggagtacggataaaatcgttgatggctggaagaggcataaattccgtcagccagttta
 gtctgacctctcatctgtaacatcattggcaacgctaccttggcatgttcagaaacaactctggcgcatcgggcttccca
 tacaatcgatagattgtcgacactgattgcccgacattatcgcgagccatttataccatataaatcagcatccatgttggga
 30 attaatcgcgccctcgagcaagacgtttccggtgaatatgggtcataacacccctgtattactgtttatgtaagcagaca
 gttttattgttcatgatgatattttatcttgtgcaatgtaacatcagagattttgagacacaacgtggctttccccccccc
 cattattgaagcatttatcagggttattgtctcatgagcggatacatatttgatgtatttagaaaaataaacaataaggggttc
 cgcgcacatttccccgaaaagtgccacctgacgtctaagaacacattattatcatgacattaacctataaaaatagggctat
 cagaggccctttcgtc

35

0D4.1 (SEQ ID NO: 150):

Atgcagctggccagctgtgtgacactgacactgggtgctgctgggaattccgcccctaggagacctgtgtgctctacaca
 gctgctgctgaatggctcttgccgaggaagagggtggtgatcagaagcgagaatttcaccaacaacgccaagaacat
 catcgtgcagctgagggaaacccgtgaagatcaactgcagccggcccaacaacaataccggagacaggccactgca
 40 acatcagcaagaccaactggaacaacacctgaacagggtggtggagaagctgggcgagcagttcaacaagaccaa
 gatcgtgttaccagagcagcggcgagatcctgagatcgtgaacctgaccttcaattgtccggcgagttcttctactg
 caataccaccagctgttcaacagcacctggaatatcaggggcaccgccagcatcaacggaaccgagagcaacgata
 acatcacctgccctgcaggatcaagggaagcggagccccctccatcaagggaatattagctgcagcagcaacatca
 caggactgctgctgacaagagatggcggcagcggcgagcggccaggagacattcagacctggcggcgagact
 45 gcaggggacaattggcggagcaggggatccctggagggtgctgttccaggggccaggccaccaccaccaccactg
 atga

CHIKV-OD4.1 (SEQ ID NO: 151):

5 tcgcgcgtttcgggtgatgacgggtgaaaacctctgacacatgcagctcccggagacgggtcacagcttgctgtaagcggat
 gccggggagcagacaagcccgtagggcgctcagcgggtgtggcggggtgcggggctggcttaactatgcggcatc
 agagcagattgtactgagagtgcaccatatgcgggtgtaataccgcacagatgcgtaaggagaaaataccgcatcaga
 ttggctattggccattgcatacgttgatccatatcataatatgtacattatattggctcatgtccaacattaccgccatgtga
 cattgattattgactagttattaatagtaatacaattacgggaacttccatagcccatatatggagtccgcgttacataacttac
 10 gggaatttccaaacctggctgaccgcccacgacccccgccattgacgtcaataatgacgtatgttccatagtaacgc
 caatagggaacttccattgacgtcaatgggtggagtattacggtaaaactgccacttgggaatttccaaagtgtatcatatgc
 caagtacgccccctattgacgtcaatgacgggaacttccataagcttgacattatgccagtacatgaccttatgggaatttc
 ctacttggcagtagacatctcgtattatgcgtattaccatgggtgatgcgggtttggcagtagacatcaatgggcgtggatag
 cggtttgactcacgggaacttccaaagtctccacccattgacgtcaatgggagtttgggttactacaaaatcaacggga
 15 attcccaaaatgtcgtacaactccgccccattgacgcaaattggcggtaggcgtgtacgggtgggaggtctatataagca
 gagctcgtttagtgaaccgtcagatgcctggagacgccatccacgctgtttgacctccatagaagacaccgggaccg
 atccagcctccatcggtcgcattctccttcacgcgcgcgcgcctacctgaggccgcatccacgcccgttgagtc
 gcgttctgcgcctcccgcctgtggtgctcctgaactgcgtccgcgtctaggttaagtttaagctcaggtcgagaccg
 ggctttgtccggcgctcccttgagcctacatagactcagccggctctccacgctttgcctgacctgctgtcaactct
 20 agttaacgggtggaggcagtgtagctgagcagtagctgttgcgcgcgcgcaccagacataatagctgacagac
 taacagactgttctttccatgggtctttctgcagtcaccgtcgtcgcacggtgtatcagatcgcggccgctctagaca
 ccatgtatagcatgcagctggccagctgtgtgacactgacactgggtgctgctggtgaattccgccccaggagacctgtgt
 gctctacacagctgctgctgaatggctcttgccgaggaagagggtggtgatcagaagcgagaatttcaccaacaacgc
 caagaacatcatcgtgcagctgagggaacccgtgaagatcaactgcagccggccaacaacaatacccgagacagg
 25 cccactgcaacatcagcaagaccaactggaacaacacctgaaacaggtgggtggagaagctgggcgagcagttcaac
 aagaccaagatcgtgttcacccagagcagcggcgagatcctgagatcgtgaacctgaccttcaattgtgccggcgagt
 tcttctactgcaataaccacccagctgttcaacagcacctggaatatcgagggcaccgccagcatcaacggaaccgagag
 caacgataacatcacctgcccctgcaggatcaagggaagcgaggccctccatcaagggaatattagctgcagcag
 caacatcacaggactgctgctgacaagagatggcggcagcggcgagcggccaggagacattcagacctggcggc
 30 ggagactgcagggacaattggcgagcaggggatccctggagggtgctgttccaggggccaggccaccaccaccacc
 accactgatgaagatcttcaaggcgcaattctgcagatgatctgctgtgcttctagttgccagccatctgttgttgcctt
 ccccgctgcttcttgcacctggaagggtgacctccactgtcctttcctaataaaatgaggaaattgcacgcattgtctg
 agtaggtgtcattctattctgggggggtgggggtggggcaggacagcaagggggaggattgggaagacaatagcaggca
 tgctgggggatgcgggtgggctctatgggtacccaggtgctgaagaattgacccgggtcctcctgggccagaaagaagca
 35 ggacatcccccttctgtgacacacctgtccacgcccctggttcttagttccagccccactcataggacactcatagctc
 aggagggtccgccttcaatcccacccgctaagtagtctggagcgggtctcctcctcatcagccccacaaacaaac
 ctgacctcaagagtgggaagaaattaaagcaagataggctattaagtgcagaggagagaaaatgcctccaacatgtg
 aggaagtaatgagagaaatcatagaatttaaggccatgatttaaggccatcatggccttaattctccgcttctcgtcact
 gactcgtcgtcgtcggtcggtcggtgctggcgagcgggtatcagctcactcaaaggcggttaatacgggttatccacagaat
 40 caggggataacgcaggaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccggttgc
 ggctgttttccataggctccgccccctgacgagcatcacaataatcgacgtcaagtcagagggtggcgaaacccgaca
 ggactataagataaccaggcggttccccctggaagctccctcgtgcgtctcctgttccgacctgcccgttaccggatac
 ctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcgggttaggtcggtcg
 ctccaagctgggctgtgtgcacgaacccccgttcagcccagcgtgcgccttatccggtaactatcgtcttgagtcca
 45 acccggttaagacacgacttatccactggcagcagccactggtaacaggattagcagagcgagggtatgtaggcggtg
 ctacagagttctgaagtgggtggcctaactacggctacactagaagaacagtatatttggtatctgcgtctgctgaagccagt

taccttcgaaaaagagttggttagctcttgatccggcaaaacaaccaccgctggttagcggtggtttttgttgcaagcag
 cagattacgcgcagaaaaaaggatctcaagaagatcctttgatctttctacggggtctgacgctcagtggaacgaaaa
 ctacggttaagggttttgggtcatgagattatcaaaaaggatcttcacctagatccttttaataaaaaatgaagtttaaatca
 atctaaagtatatatgagtaaaacttgggtctgacagttaccaatgcttaacagtgaggcacctatctcagcgatctgtctatttc
 5 gttcatccatagttgctgactcggggggggggggcgctgaggtctgcctcgtgaagaaggtgtgctgactcatacca
 ggctgaatcgccccatcatccagccagaaagtgaggagccacgggtgatgagagctttgtgtgaggtggaccaggtg
 gtgattttgaacttttgccttgccacggaacgggtctgcgtgtcggaagatgcgtgatctgatccttcaactcagcaaaagt
 tcgattttcaacaaagccgcccgtcccgtaagtcagcgtaaatgctctgacagtggtacaaccaattaaccaattctgatta
 gaaaaactcatcgagcatcaaatgaaactgcaattttatcatatcaggattatcaataccatattttgaaaaagccgtttctgt
 10 aatgaaggagaaaaactcaccgagggcagttccataggtggcaagatcctgggtatcggtctgcgattccgactcgtccaa
 catcaatacaacctattaatttcccctcgtcaaaaataaggttatcaagtgaagaaatccatgagtgacgactgaatccgg
 tgagaatggcaaaagcttatgcatctttccagactgttcaacaggccagccattacgctcgtcatcaaaatcactcgc
 caaccaaacggttattcattcgtgattgcgctgagcgagacgaaatacgcgatcgtgttaaaaggacaattacaaca
 ggaatcgaatgaaccggcgaggaacactgccagcgcatcaacaatatttcacctgaatcaggatattcttcaataacc
 15 tggaatgctgttttccggggatcgagtggtgagtaacctgcatcatcaggagtacggataaaatgcttgatggctgga
 agaggcataaattccgtcagccagtttagtctgacatctcatctgtaacatcattggcaacgctacctttgccatgtttcaga
 aacaactctggcgcatcgggctcccatacaatcgatagattgtcgacactgattgcccgacattatcgcgagccattat
 acccatataaatcagcatcatgttggaatttaatcgcgccctcgagcaagacgtttccggtgaatatggctcataacacc
 cctgtattactgtttatgtaagcagacagttttattgttcatgatgatataattttatctgtgcaatgtaacatcagagattttgag
 20 acacaactgtggctttccccccccccattattgaagcatttatcagggtattgtctcatgagcggatacatatttgatgta
 tttagaaaaataaacaataggggtccgcgacatttccccgaaaagtgccacctgacgtctaagaaccattattatcat
 gacattaacctataaaaataggcgatcacgagcccttctgct

OD4.3.2Vmut_3G/424N (SEQ ID NO: 152):

25 atggagttcatcccagcgaactttctataacagaaggtaccaacccccgacctggggccccacgcctacaattcaagt
 aattagacctagaccagtcacagaggcagggtgggcaactcgcccagctgatctccgcagtcacaacaaattgacatg
 cgcgcggtacctaacaagaagcctcgagaaatcggaacaaagaagcaaaaggcagaagaagcaggcgccgcaa
 aacgacccaaagcaaaagaagcaaccaccacaaaagaagccggctcaaaagaagaagaaccaggccgtagggag
 agaattgtcatgaaaattgaaaatgattgcatcttcgaagtcaagcatgaaggcaaaagtatgggctacgatgcctggt
 30 gggggataaagtaataaaccagcacatgtgaagggaactatcgacaatgccgatctggctaaactggccttaagcgg
 tcgtctaatacgaatcttgatgtgacagataccggtgacatgaagtctgatgcctcgaagttaccacgagaaacc
 gaggggtactataactggcatcacggagcagtgagtgattcaggaggccggttactatcccagcgggtgcaggcaag
 ccgggagacagcggcagaccgatcttcgacaacaaaggacgggtggtggccatcgtcttaggagggggccaacgaag
 gtgcccgcacggccctcctggtgacgtggaacaaagacatcgtcacaaaattaccctgagggagccgaagagt
 35 ggagcctcgccctcccggtctgtgctgttggaacactacattccctgctctcagccgcttgacacccgtgctgct
 acgaaaaggaaaccggaaaagcaccttgcgcatgcttgaggacaacgtgatgagaccggatactaccagctactaaaag
 catcgtgacttgctctccccaccgccaagacgcagtactaaggacaatttaattgtctataaagccacaagaccatatct
 agctcattgtcctgactcggagaagggcattcgtgccacagccctatcgattggagcgcacagaaatgaagcaacg
 gacggaacgctgaaaatccaggtctcttgcagatcgggataaagacagatgacagccacgattggaccaagctgcgct
 40 atatggatagccatcagccagcggacgcggagcgcgggattgctgttaaggacttcagcacctgacgatcaccg
 ggacctggggacactttattctcgcggatgccgaaaggagagacgctgacagtgaggatttacggacagcagaaaga
 tcagccacacatgcacacacccgttccatcatgaaccacctgtgataggtaggagaggttccactctcgaccacaacat
 ggtaaagagttaccttgacgacgtacgtgcagagcaccgctgccactgctgaggagatagaggtgcatatgccccca
 gatactcctgaccgcacgctgatgacgcagcagcttggaacgtgaagatcacagttaatgggcagacgggtgcggtac
 45 aagtgcaactgcgggtggctccggaagacctgtgtgctctacacagctgctgctgaatggctctctggccgaggaagagg
 tgggtgatcagaagcgagaatttcaccaacaacgccaagaacatcatcgtgcagctgagggaacccgtgaagatcaact

- gcagccggcccaacaacaatacccgagacaggcccactgcaacatcagcaagaccaactggaacaacacctgaa
 acaggtgggtggagaagctgggcgagcagttcaacaagaccaagatcgtgttaccagagcagcgggcgagatcctg
 agatcgtgaccacagcttcaattgtgccggcgagttcttactgcaataccaccagctgttcaacagcacctggaatat
 cgagggcaccgccagcatcaacggaaaccgagagcaacgataacatcacctgcccctgcaggatcaagggaagcaac
 5 gccaccgtgatcaagggaatattagctgcagcagcaacatcacaggactgctgctgacaagagatggcgggcagcgg
 cggcagcggccaggagacattcagacctggcgggcgagactgcaggggacaactggaggagcggggctccggatc
 aaacgagggaactgacaaccacagacaaagtgatcaataactgcaaaattgatcagtccatgctgcagtcactaatcac
 aagaattggcaatacaactcccccttagtcccgcgcaacgctgaactcggggaccgtaaaggaaagatccacatcccatt
 cccattggcaaacgtgacttcagagtgccaaaagcaagaaacctacagtaacttacggaaaaaaccaagtcacatg
 10 ctgctgtatcctgacctccgacactcttgtcttaccgtaacatgggacaggaaccaattaccacgaggagtgggtgac
 acacaagaaggaggttaccttgaccgtgcttactgagggctgagggtcacttggggcaacaacgaaccatacaagtac
 tggccgcagatgtctacgaacgggtactgctcatggtcaccacatgagataatctgtactattatgagctgtacccactat
 gactgtagtcatgtgtcggtggcctcgttctgtctgtcgatgggtgggcacagcagtggaatgtgtgtgtgcgcacg
 gcgcagatgcattacaccatatgaattaacaccaggagccactgttcccttctgctcagcctgctatgctgcgtcagaac
 15 gaccaaggcggccacatattacgaggctgcggcatatctatggaacgaacagcagcccctgttctggttcaggtcttta
 tcccgctggccgcttgatcgtcctgtgcaactgtctgaaactcttgcctgctgctgtaagacctggctttttgacccgta
 atgagcatcggtgcccacactgtgagcgcgtacgaacacgtaacagtgtcccgaacacgggtgggagtaccgtataag
 actcttgtcaacagaccgggttacagccccatggtgttggagatggagctacaatcagtcaccttggaaacacactgtc
 acttgactacatcacgtgcgagtacaaaactgtcatccccctccccgtacgtgaagtgtgtgtggtacagcagagtgaagg
 20 acaagagcctaccagactacagctgcaaggtcttactggagtctacccattatgtggggcgggcgctactgctttgctg
 acgccgaaaatacgcaattgagcagggcacatgtagagaaatctgaatcttgcataacagagtttgcacggcctacag
 agcccacaccgcacgcgtcggcggaagctccgcgtctttaccaaggaaacaacattaccgtagctgcttacgctaac
 ggtgacctgcccgtcacagtaaaaggacgccaagttgtcgtgggcccgaatgtcctccgctggacacctttgacaacaa
 aatcgtgggtgtacaaaggcgacgtctacaacatggactaccaccttttggcgaggaagaccaggacaatttggtgac
 25 attcaaagtcgtacaccgaaagtaaagacgtttatgccaacactcagttggtactacagaggccagcagggcacgg
 tacatgtaccatactctcaggcaccatctggcttcaagtattggctgaaggaaacgaggagcatcgctacagcacacggca
 ccgttcggttgccagattgcgacaaaccggtaagagctgtaaattgcgtgtgggggaacataccaatttccatcgacata
 ccggatgcggcctttactaggggtgtcgatgcaccctctgtaacggacatgtcatgcgaagtaccagcctgactactcc
 tccgactttgggggctgcctcatcatcaatacacagctagcaagaaaggtaaatgtgcagtacattcgatgaccaacgc
 30 cgttaccattcgagaagccgacgtagaagtagaggggaactcccagctgcaaatatccttctcaacagccctggcaagc
 gccgagtttcgctgcaagtgtgtccacacaagtacactgcgcagccgcatgccaccctcaaaggaccacatagtc
 attaccagcatcacacaccaccttgggtccaggatatccacaacggcaatgtcttgggtgcagaagattacggga
 ggagtaggattaattgttctgtgtcgtcctaattttaattgtggtgctatgcgtgtcgttttagcaggactaa
- 35 CHIKV-OD4.3.2Vmut_3G/424N (SEQ ID NO: 153):
 tcgcgcgtttcgggtgatgacggtgaaaaacctctgacacatgcagctcccggagacgggtcacagctgtctgtaagcggat
 gccggggagcagacaagcccgtcagggcgctcagcgggtgttggcggggtgtcggggctggcttaactatgcggcatc
 agagcagattgtactgagagtgcacatatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaga
 ttggctattggccattgcatacgttgatccatatcataatgtacatttatattggctcatgtccaacattaccgccatgttga
 40 cattgattattgactagtattataatagtaataacacgggtcattagttcatagcccatatatggagttccgcgttacataact
 tacggtaaatggccccctggctgaccgcccacgacccccgccattgacgtcaataatgacgtatgtcccatagtaa
 cgccaatagggactttccattgacgtcaatgggtggagattttacggtaaaactgccacttggcagtagatcaagtgtatca
 tatgccaaagtacggccctattgacgtcaatgacggtaaatggcccgctggcattatgccagtagatgacattatggga
 ctttctacttggcagtagatctacgtattagtcacgtattaccatgggtgatgcgggttttggcagtagatcaatgggcgtgg
 45 atagcgggttgactcacggggatttccaagtctccacccattgacgtcaatgggagtttgggttttggcaccaaaatcaacgg
 gactttccaaaatgtcgtacaactccgccccattgacgcaaatgggcggttaggcgtgtacgggtgggaggtctatataag

cagagctcgttttagtgaaccgtcagatcgctggagacgccatccacgctgtttgacctccatagaagacaccgggacc
 gatccagcctccatcggtcgcctctcttcacgcgcccgcctacctgaggccgcatccacgccggttgagt
 cgcgttctgccgctcccgctgtggtgcctcctgaactgcgtccgcccgtctaggttaagtttaagctcaggtcgagacc
 gggcctttgtccggcgctcccttgagcctacctagactcagccggctctccacgctttgacctgcttgcctcaactc
 5 tagttaacggtggaggggcagtgtagtctgagcagctactcgttgctgccgcgcgcgccaccagacataatagctgacaga
 ctaacagactgttctttccatgggtctttctgcagtcaccgtcgtcgacacgtgtgatcagatatcgcgccgctctagac
 accatggagttcatcccagcgaactttctataacagaaggtaccaacccccgacctgggccccacgcccataattca
 agtaattagacctagaccacgtccacagaggcaggtgggcaactcgccagctgatctccgcagtcacaaaattgacc
 atgcgcgcggtacctcaacagaagcctcgcagaaatcggaacaaagaagcaaggcagaagaagcaggcgccg
 10 caaacgacccaaagcaaaagcaaccaccacaaaagaagccggctcaaaagaagaagaaccaggccgtagg
 gagagaatgtgatgaaaattgaaaatgattgcatcttgaagtcaagcatgaaggcaagtgtgggtacgcagtcct
 ggtgggggataaagtaataaaaccagcacatgtgaagggaactatcgacaatgccgatctggctaaactggcctttaag
 cggtcgtctaaatacgtcttgaatgtgcacagataccggtgcacatgaagtctgatgcctcgaagttaccacagagaaa
 cccgaggggtactataactggcatcacggagcagtgagtgatcaggaggccggttactatcccagcgggtgcaggc
 15 aagccgggagacagcggcagaccgatcttcgacaacaaaggacgggtgggtggccatcgtcctaggaggggccaacg
 aaggtgcccgcacggccctctccgtgggtgacgtggaacaaagacatcgtcacaaaaattaccctgaggggagccgaa
 gagtggagcctcgccctcccgggtcttgcctgttggcaaacactacattcccctgctctcagccgcttgcacacccctgc
 tgctacgaaaaggaaaccggaaagcaccttgcgcagcttgaggacaacgtgatgagaccggatactaccagctacta
 aaagcatcgctgacttgctctccccaccgccaagacgcagctactaaggacaattttaatgtctataaagccacaagacc
 20 atatctagctcattgtcctgactgcggagaagggcattcgtgccacagccctatcgattggagcgcacagaaatgaag
 caacggacggaacgctgaaaatccaggtctcttgcagatcgggataaagacagatgacagccacgattggaccaagc
 tgcgctatatggatagccatacgccagcggagcggagcgcgagccgattgcttgaaggacttcagcacccgtgcacgat
 caccgggacatgggacactttattctcgccgatgcccgaaggagagacgctgacagtggtgattacggacagcag
 aaagatcagccacacatgcacacaccgttccatcatgaaccacctgtgataggtaggagaggttccactctcgacca
 25 caacatggtaaagagttaccttgagcagctacgtgcagagcaccgctgccactgctgaggagatagaggtgcatatgc
 cccagatactctgaccgcacgctgatgacgcagcagcttggaacgtgaagatcacagttaatgggcagacgggtgc
 ggtacaagtgaactgcgggtggctccggaagacctgtgtgctctacacagctgctgctgaatggctctctggccgagga
 agaggtggtgatcagaagcgagaatttaccacaacgccaagaacatcatcgtgcagctgaggggaacccgtgaagat
 caactgcagccggcccaacaataaccggagacagggccactgcaacatcagcaagaccaactggaacaacacc
 30 ctgaaacaggtggtggagaagctgggcgagcagttcaacaagaccaagatcgtgttaccacagagcagcggcgag
 atctgagatcgtgaccacagcttcaattgtccggcgaggttcttactgcaataccaccagctgttcaacagcacctg
 gaatatcaggggcaccgcccagcatcaacggaaccgagagcaacgataacatcacctgccctgcaggatcaaggga
 gcaacgccaccgtgatcaagggaatattagctgcagcagcaacatcacaggactgctgctgacaagagatggcggc
 agcggcgagcggccagggagacattcagacctggcgcgagactgcagggacaactggaggagcagagggtc
 35 cggatcaaacgaggggactgacaaccacagacaaagtatcaataactgcaaaattgatcagtgccatgctgcagtcact
 aatcacagaattggcaatacaactcccccttagtcccgcgcaacgctgaactcggggaccgtaaaaggaaagatccaca
 tccattcccattggcaaacgtgacttgcagagtgcacaaagcaagaacccctacagtaacttacggaaaaaaccaagtc
 accatgctgctgtatcctgaccatccgacactctgttaccgtaacatgggacaggaaccaaattaccacgaggagtgg
 gtgacacacaagaaggaggttaccttgacctgacctgaggggtctggaggtcacttggggcaacaacgaaccatac
 40 aagtactggccgcagatgtctacgaacggtactgctcatggtcaccacatgagataatctgtactattatgagctgtacc
 ccactatgactgtagtattgtcgggtggcctcgtcgttctgtcgtgatgggtgggcacagcagtggggaatgtgtgtg
 cgcacggcgagatgcattacacatataaataacaccaggagccactgttcccttctgctcagcctgctatgctgcgt
 cagaacgaccaaggcgccacatattacagggtcgggcatalctatggaacgaacagcagccctgttctggttcag
 gctcttatcccgtggccgcttgatcgtcctgtgcaactgtctgaaactcttgccatgctgctgtaagacctggccttttta
 45 gccgtaatgagcatcgggtgccacactgtgagcgcgtacgaacacgtaacagtgatccgaacacgggtgggagtaccg
 tataagactctgtcaacagaccgggttacagcccatgggtgttgagatggagctacaatcagtcaccttggaaaccaac

actgtcacttgactacatcacgtgagtagacaaaactgtcatccctccccgtacgtgaagtgtgtgtgtacagcagagtg
 caaggacaagagcctaccagactacagctgcaaggcttttactggagctacccatttatgtggggcgggcgctactgctt
 ttgcgacgccgaaaaatcgcaattgagcgggacacatgtagagaaatctgaatcttgcaaaacagagtttgcacggcct
 acagagccccacaccgcatcggtcggtcggaagctccgctctttaccaaggaaacaacattaccgtagctgcctacg
 5 ctaacggtgaccatgccgtcacagtaaaggacgccaagtttgcgtggggcccaatgtcctccgcctggacacctttgac
 aaaaaatcggtgtgtacaaaggcgacgtctacaacatggactaccaccttttggcgaggaagaccaggacaatttg
 gtgacattcaaagtcgtacaccggaagtaaagacgtttatgccaacactcagttgggtactacagaggccagcagcagg
 cacggtacatgtaccatactctcaggcaccatctggctcaagtattggctgaaggaaacgaggagcatcgctacagcaca
 cggcaccgttcggttgccagattgacgacaaacccggtaagagctgtaaattgcgctgtgggggaacataccaatttccatc
 10 gacataccggatgctggcctttactaggggttgcgtatgcaccctctgaacggacatgtcatgcgaagtaccagcctgcact
 cactcctccgactttggggcgctgccatcatcaatacacagctagcaagaaaggtaaattgtcagtagcattcgtatgac
 caacgccgttaccattcgagaagccgacgtagaagtagaggggaactcccagctgcaaatatccttctcaacagccctg
 gcaagcgccgagtttcgctgcaagtgtgctccacacaagtacactgcgcagccgcatgccaccctcaaaggaccac
 atagtcaattaccagcatcacacaccaccttgggggtccaggatataccacaacggcaatgtcttgggtgcagaagatt
 15 acggggaggagtaggattaattgttgcgtgttgccttaattttaattgtgtgtctatgcgtgtctttagcaggcactaatga
 ggatccagatctgctgtccttctagtgtccagccatctgttgttgccttccccgtgccttcttacccttgggaaggtgc
 cactcccactgtcctttcctaataaaatgaggaaattgcacgcattgtctgagtaggtgtcattctattctgggggggtgggg
 tggggcgaggacgaaggggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatgggtac
 ccaggtgctgaagaattgacccggtcctcctgggcccagaaagaagcaggcacatcccccttctgtgacacacccgtc
 20 cacgcccctggttcttagttccagccccactcataggacactcatagctcaggagggtccgccttcaatcccacccgcta
 aagtacttggagcgggtctctccctccctcatagcccacaaaccaaaccctagcctcaagagtgaggaaagaattaaag
 caagataggctattaagtgcagagggagagaaaatgcctccaacatgtgaggaaagtaagtagagaaatcatagaatttta
 aggccatgatttaaggccatcatggccttaattctccgcttctcgtcactgactcgtgcgctcggctcgttcggctgcgg
 cgagcgggtatcagctcactcaaaggcggttaatacgggtatccacagaatcaggggataacgcaggaaagaacatgtga
 25 gcaaaaggccagcaaaaggccaggaaccgtaaaaggccgctgtggcgtttttccataggctccgccccctgac
 gagcatcaaaaaatcgacgctcaagtacaggtggcgaaacccgacaggactataaagatacaggcgctttccccct
 ggaagctccctcgtgcgctcctgttccgacctgcccgttaccggatacctgtccgcctttctccttcgggaagcgtg
 gcgctttctatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgtccaagctgggctgtgtgcagaaaccc
 ccgttcagcccagaccgtgcgcttaccggtaactatcgtcttagtccaacccgtaagacacgacttatcgccactgg
 30 cagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttctgaagtgggtggcctaacta
 cggctacactagaagaacagtatctgcgctcgtctgaagccagttaccttcggaaaaagagttggtagctcttga
 tccggcaaaacaaaccaccgctggtagcgggtgtttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctca
 agaagatcctttgatcttttacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattat
 caaaaaggatcttcactagatccttttaataaaaaatgaagttttaaatcaatctaaagtatatatgagtaaaacttggctga
 35 cagttaccaatgcttaacagtgaggcacctatctcagcgtatctgtctatttcgttcatccatagttgcctgactcggggggg
 gggggcgctgaggtcgtcctcgtgaagaaggtgttgcgtactataccaggcctgaatcgccccatcatccagccagaa
 agtgagggagccacgggtgatgagagcttgggttaggtggaccagttgggtgattttgaacttttgccttggcacggaacgg
 tctgcgttgcgggaagatgcgtgatctgatcctcaactcagcaaaagttcgatttattcaaaaagccgcccgtccgtca
 agtcagcgtaatgctctgccagtgttacaaccaattaaccaattctgattagaaaaactcatcgagcatcaaatgaaactgc
 40 aatttattcatatcaggattatcaataccatattttgaaaaagccgttctgtaatgaaggagaaaactcaccgaggcagttc
 cataggatggcaagatcctggtatcggtctgcgattccgactcgtccaacatcaatacaacctaatttccctcgtcaaa
 aataaggttatcaagtgaagaatcaccatgagtgacgactgaatccggtgagaatggcaaaagcttatgcatttcttccag
 acttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcatcaaccaaaccgttattcattcgtgattgcgctg
 agcgagacgaaatcgcgatcgctgttaaaaggacaattacaacaggaatcgaatgaaccggcgaggaacactg
 45 ccagcgcatcaacaatattttcacctgaatcaggatatttcttaatacctggaatgctgtttcccggggatcgagtggtg
 agtaacatgcatcatcaggagtacggataaaatgcttgatggtcggaagaggcataaattccgtcagccagtttagtctg

accatctcatctgtaacatcattggcaacgctacctttgccatgtttcagaaacaactctggcgcatcgggcttccatacaa
tcgatagattgtcgacactgattgcccacattatcgcgagccattataccatataaatcagcatccatgttggaatttaa
tcgcggcctcgagcaagacgtttcccggtgaatatggctcataacacccctgttactgtttatgtaagcagacagtttat
tggtcatgatgatataattttatcttgcaatgtaacatcagagattttgagacacaacgtggcttccccccccccattatt
5 gaagcatttatcaggggtattgtctcatgagcggatacatattgaaatgtatttagaaaaataaacaataaggggtccgcgc
acatttccccgaaaagtgccacctgacgtctaagaaaccattattatcatgacattaacctataaaaataggcgatcacga
ggccctttcgtc

OD 4.3.2Vmut 3G (SEQ ID NO: 154):

10 atggagttcatcccgacgcaaaactttctataacagaaggtaccaaccccgaccctgggccccacgcctacaattcaagt
aattagacntagaccagtcacagaggcaggctgggcaactcgcccagctgatctccgcagtcacaaaattgaccatg
cgcgcggtacctaacagaagcctcgagaaatcggaacaaagaagcaaaaggcagaagaagcaggcgccgcaa
aacgacccaaagcaaaagaagcaaccaccacaaaagaagcggctcaaaagaagaagaaccaggccgtagggag
agaatgtgcatgaaaattgaaaatgattgcatcttcgaagtcaagcatgaaggcaaaagtgtgggctacgcatgcctgg
15 gggggataaaagtaataaccagcacatgtgaagggaactatcgacaatgccgatctgggctaaactggccttaagcgg
tcgtctaataacgatcttgaatgtgcacagataccgggtgcacatgaagtctgatgcctcgaagttaccacagagaaacc
gaggggtactataactggcatcacggagcagtgacagtattcaggaggccgggtcactatcccagcgggtgcaggcaag
ccgggagacagcggcagaccgatcttcgacaacaaaggacgggtggggccatcgtcctaggagggggccaacgaag
gtgcccgcacggccctctccgtggtagctggaacaaagacatcgtcacaaaaattaccctgagggagccgaagagt
20 ggagcctcgccctcccgggtctgtgctgttggaacactacattcccctgctctcagccgccttgcacaccctgctgct
acgaaaaggaaccggaagcaccttgcgcatgcttgaggacaacgtgatgagaccggatactaccagctactaaaag
catcgtgactgtctctcccaccgcaaaagacgcagtactaaggacaattttaatgtctataaagccacaagaccatatct
agctcattgtcctgactgcggagaagggcattcgtgccacagccctatcgattggagcgcacagaaatgaagcaacg
gacggaacgctgaaaatccaggtctcttgcagatcgggataaagacagatgacagccacgattggaccaagctgcgt
25 atatggatagccatacgcagcggacgcggagcgcggagtgcttgaaggacttcagcaccgtgcacgatcaccg
ggaccatgggacactttattctcgcccgatccccgaaaggagagacgctgacagtgggatttacggacagcagaaaga
tcagccacacatgcacacacccgttccatcatgaaccacctgtgataggtaggagagaggttccactctcgaccacaacat
ggtaaaagagttaccttgacgacgtacgtgcagagcaccgctgccactgctgaggagatagaggtgcatatgccccca
gatactctgaccgcacgctgatgacgcagcagtgctggaacgtgaagatcacagttaatgggcagacgggtgcggtac
30 aagtgcaactgcgggtgggtccggaagacctgtgtgctctacacagctgctgctgaatggctctctggccgaggaagagg
tggtgatcagaagcgagaatttcaccaacaacgccaagaacatcatcgtgcagctgagggaacccgtgaagatcaact
gcagccggccaacaacaatacccgagacagggccactgcaacatcagcaagaccaactggaacaacaccctgaa
acaggtgggtggagaagctgggagcaggttcaacaagaccaagatcgtgttaccagagcagcggcgagatcctg
agatcgtgaccacagcttcaattgtgccggcgagttcttactgcaataaccaccagctgttcaacagcacctggaatat
35 cgagggcaccgcagcatcaacggaaccgagagcaacgataacatcacctgccctgcaggatcaagggaagcggga
gcccctcccatcaagggaatattagctgcagcagcaacatcacaggactgctgctgacaagagatggcggcagcgg
cggcagcggccaggagacattcagacctggcgggcggagactgcagggacaactggaggagcaggggtccggatc
aaacgagggaactgacaaccacagacaaagtatcaataactgcaaaattgatcagtccatgctgcagtcactaatcac
aagaattggcaataaactccccttagtcccgcgcaacgctgaactcggggaccgtaaggaaagatccacatcccatt
40 cccattggcaaacgtgacttgagagtgccaaaagcaagaaacctacagtaacttacggaaaaaaccaagtcacatg
ctgctgtatcctgacctccgacactctgtcttaccgtaacatgggacaggaaccaaattaccacgaggagtgggtgac
acacaagaaggaggttaccttgaccgtgctactgagggctgagggtcacttggggcaacaacgaaccatacaagtac
tggccgcagatgtctacgaacgggtactgctcatgggtcacccacatgagataatctgtactattatgagctgtacccactat
gactgtagtcatgtgtcggtggcctcgttcgtctctgtcgtatgggtgggcacagcagtgggaaatgtgtgtgtgcgcacg
45 gcgcagatgcattacacatgaattaacaccaggagccactgttcccttctgctcagcctgctatgctgcgtcagaac
gaccaaggcggccacatattacgaggctgcggcatatctatggaacgaacagcagccctgttctggttgaggctctta

- tcccgtggccgccttgatcgtcctgtgcaactgtctgaaactcttgccatgctgctgtaagaccctggcttttttagccgta
atgagcatcggtgccacactgtgagcgcgtacgaacacgtaacagtgatcccgaacacgggtgggagtagccgtataag
actcttgtaacagaccgggttacagccccatgggtgtggagatggagctacaatcagtcaccttggaaacacactgtc
acttgactacatcacgtgcgagtacaaaactgtcatcccctccccgtacgtgaagtgtgtggtagacagagtgcaagg
5 acaagagcctaccagactacagctgcaaggctttactggagtgtagccattatgtggggcgccgctactgcttttgcg
acgccgaaaatacgcaattgagcagggcacatgtagagaaatctgaatcttgcaaaacagagtttgcacggcctacag
agcccacaccgcatcggtcggcgaagctccgcgtctttaccaaggaaacaacattaccgtagctgcttacgctaac
ggtagcatgcccgtcacagtaaaggacccaagttgtcgtgggccaatgtctccgcctggacacctttgacaacaa
aatcgtggtgtacaaaggcgacgtctacaacatggactaccaccttttggcgaggaagaccaggacaatttggtgac
10 attcaaaagtctacaccggaagtaaagacgtttatgccaacactcagttggtactacagaggccagcagcaggcacgg
tacatgtaccatactctcaggcaccatctggcttcaagtattggctgaaggaaacgaggagcatcgctacagcacacggca
ccgttcggttgccagattgcgacaaaccggtaagagctgtaaattgcgctgtgggggaacataccaatttccatcgacata
ccggatgcggcctttactagggtgtcgatgcaccctctgtaacggacatgtcatgcgaagtaccagcctgcactactcc
tccgactttggggcgctgccatcatcaatacacagctagcaagaaaggtaaattgtcagtagattcgatgaccaacgc
15 cgttaccattcgagaagccgacgtagaagtagaggggaactcccagctgcaaatatccttctcaacagccctggcaagc
gccgagtttcgctgcaagtgtgctccacacaagtacactgcgcagccgcatgccaccctcaaaggaccacatagtca
attaccagcatcacaccacccttgggtccaggatataccacaacggcaatgtcttgggtgcagaagattacggga
ggagtaggattaattgttctgtgtgccttaatttaattgtgggtgctatgcgtgtcgttttagcaggcactaa
- 20 CHIKV OD 4.3.2Vmut 3G (SEQ ID NO: 155):
tcgcgcgtttcggtgatgacgggtgaaaacctctgacacatgcagctcccgagacgggtcacagcttctgtaagcggat
gccgggagcagacaagcccgctcagggcgctcagcgggtgttggcggtgtcggggctggcttaactatgcggcatc
agagcagattgtactgagagtgacccatatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaga
ttggctattggccattgcatacgttgatccatatcataatatgtacatttatattggctcatgtccaacattaccgcatgtga
25 cattgattattgactagtatttaataagtaatacattacgggtcattagttcatagcccatatatggagttccgcgttacataact
tacggtaaatggccccgctggctgaccgccccacgacccccgccattgacgtcaataatgacgtatgttcccatagtaa
cgccaatagggactttccattgacgtcaatgggtggagattttacggtaaactgccacttggcagtagatcaagtgtatca
tatgccaagtacgccccctattgacgtcaatgacggtaaatggccgcctggcattatgccagtagatgaccttatggga
ctttctacttggcagtagatctacgtattagtcacgtattaccatgggtgatgcggttttggcagtagatcaatgggcgtgg
30 atagcgggttgactcacggggatttccaagtctccacccattgacgtcaatgggagttgttttggcaccaaaatcaacgg
gactttccaaaatgtcgtacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataag
cagagctcgttttagtaaccgtcagatcgcttgagacgccatccacgctgttttgacctccatagaagacaccgggacc
gatccagcctccatcggtcgcacatctctccttcacgcgccccgcccctacctgaggccgcatccacgccgggtgagt
cgcggtctgcgcctcccgcctgtgggtgcctctgaactgcgtccgcgttaggtaagtttaaagctcaggtcgagacc
35 gggcctttgtccggcgctcccttgagcctacgtacagcggctctccacgctttgctgaccctgcttgcctaactc
tagttaacgggtggagggcagtgtagtctgagcagtagctgtgctgcccgcgcgcccaccagacataatagctgacaga
ctaacagactgttctttccatgggtcttttctgagtcaccgtcgtcgacacgtgtgatcagatatcgcgccgctctagac
accatggagttcatcccagcgaactttctataacagaaggtaccaaccccgaccctggggccccacgcccataattca
agtaattagacctagaccagtcacagaggcaggtgggcaactcggccagctgatctccgcagtagcaacaaattgacc
40 atgcgcgcggttacctcaacagaagcctcgagaaatcggaacaaagaagcaaaggcagagaagaagcaggcgccg
caaaacgacccaaagcaaaagaagcaaccaccacaaaagaagccggtcaaaagaagaagaacaggcgctagg
gagagaatgtgatgaaaattgaaatgattgcatcttcgaagtcaagcatgaaggcaagtgatgggctacgcatgcct
gggtgggggataaagtaataaaccagcacatgtgaagggaactatcgacaatgccgatctggctaaactggcctttaag
cggtcgtctaaatacagatcttgatgtgcacagataccggtgcacatgaagtctgatgcctcgaagtttaccacgagaaa
45 cccgaggggtactataactggcatcacggagcagtgtagtattcaggaggccggttactatcccaggggtgcaggc
aagccgggagacagcggcagaccgatcttcgacaacaaaggacgggtgtggccatcgctcaggaggggccaacg

aaggtgcccgcacggccctctccgtgggtgacgtggaacaaagacatcgtcacaaaaattaccctgagggagccgaa
 gagtgagcctcgccctcccggtcttgctgttggtgcaaacactacattcccctgctctcagccgccttgacacccctgc
 tgctacgaaaaggaaccggaaagcaccttgcatgcttgaggacaacgtgatgagaccggatactaccagctacta
 aaagcatcgctgacttgctctccccaccgcaaaagacgcagtactaaggacaattttaatgtctataaagccacaagacc
 5 atatctagctcattgtcctgactgaggagaagggcattcgtgccacagccctatcgattggagcgcacagaaatgaag
 caacggacggaacgctgaaaatccaggtctctttgcagatcgggataaagacagatgacagccacgattggaccaagc
 tgcgctatatggatagccatacggcagcggacgcggagcgcgggattgctgtaaggacttcagaccgtgcacgat
 caccgggacatgggacactttattctgcccgatgcccgaaggagagacgctgacagtgggatttacggacagcag
 aaagatcagccacacatgcacacaccgttccatcatgaaccacgtgtgataggtaggagaggttccactctcgacca
 10 caacatggtaaagagttaccttgagcacgtacgtgcagagcaccgctgccactgctgaggagatagaggtgcatatgc
 cccagatactcctgaccgcacgctgatgacgcagcagcttggaacgtgaagatcacagttaatgggcagacgggtgc
 ggtacaagtgaactgcgggtggctccggaagacgtgtgtctacacagctgctgctgaatggctctctggccgagga
 agaggtgggtgatcagaagcgagaattcaccaacaacgccaagaacatcatcgtgcagctgagggaaaccgtgaagat
 caactgcagccggcccaacaataaccggagacaggcccactgcaacatcagcaagaccaactggaacaacacc
 15 ctgaaacaggtgggtggagaagctgggagcaggtcaacaagaccaagatcgtgttaccagagcagcggcggag
 atcctgagatcgtgaccacagcttcaattgtgcccgcgagttcttctactgcaataccaccagctgttcaacagcacctg
 gaatacagagggcaccgccagcatcaacggaaccgagagcaacgataacatcacctgcctgaggatcaagggaa
 gcggagccccctccatcaagggcaatattagctgcagcagcaacatcacaggactgctgctgacaagagatggcggc
 agcggcggcagcggccaggagacattcagacctggcggcggagactgcagggacaactggaggagcaggggctc
 20 cggatcaaacgagggactgacaaccacagacaaagtgatcaataactgcaaattgatcagtgcctatgctgcagtcact
 aatcacagaattggcaatacaactcccccttagtcccgcgcaacgctgaactcggggaccgtaaaggaaagatccaca
 tccattccattggcaaacgtgacttgagagtgccaaaagcaagaaaccctacagtaacttacggaaaaaccaagtc
 accatgctgctgtatcctgacctccgacactctgtcttaccgtaacatgggacaggaaaccaattaccacgaggagtgg
 gtgacacacaagaaggaggttaccttgaccgtgcctactgagggctggaggctcacttggggcaacaacgaaccatac
 25 aagtactggccgcagatgtctacgaacggtactgctcatggtcaccacatgagataatctgtactattatgagctgtacc
 ccactatgactgtagtcatgtgctgggtggcctcgctgcttctgtcagtggtgggcacagcagtgggaatgtgtgtgtg
 cgcacggcgcagatgcattacaccatgaattaacaccaggagccactgttcccttctgctcagcctgctatgctgcgt
 cagaacgaccaaggcggccacatattacgaggctgcggcatactatggaacgaacagcagcccctgttctggttgag
 gctcttatcccgtggccgccttgatcgtcctgtgcaactgtctgaaactcttccatgctgctgtaagacctggcttttta
 30 gccgtaatgagcatcgggtgccacactgtgagcgcgtacgaacacgtaacagtgatcccgaacacgggtgggagtaccg
 tataagactctgtcaacagaccgggttacagccccatgggtgttgagatggagctacaatcagtcaccttgaaccaac
 actgtcacttgactacatcacgtgcgagtacaaaactgtatccccctcccgtacgtgaagtgtgtgtgtacagcagagt
 caaggacaagagcctaccagactacagctgcaaggcttctactggagtctacccatttatgtggggcggcgccactgctt
 ttgcgacgcccgaataacgcaattgagcgaggcacatgtagagaaatctgaatcttgcaaacagagtttgcatcggcct
 35 acagagcccacaccgcatcggcgtcggcgaagctccgcgtcctttaccaaggaaacaacattaccgtagctgcctacg
 ctaacgggtgacctgcccgtcacagtaaaggacgccaagtttgcgtggggcccaatgtcctccgcctggacaccttttgac
 aaaaaatcgtgggtgtacaaaggcgacgtctacaacatggactaccaccttttggcgagggaagaccaggacaatttg
 gtgacattcaagtcgtacaccggaagtaagacgtttatgccaacactcagttgggtactacagaggccagcagcagg
 cacggtacatgtaccatactctcaggcaccatctggcttcaagtattggctgaaggaaacgaggagcatcgtacagcaca
 40 cggcaccgttcgggtgccagattgcgacaaacccggtaagagctgaaattgcgctgtgggggaacataccaatttccatc
 gacataccggatgcggccttactagggtgtcgatgacccctctgtaaccggacatgtcatgcaagtaccagcctgcact
 cactctccgactttggggcgctgccatcatcaatacacagctagcaagaaaggtaaatgtgcagtacattcgtatgac
 caacgccgttaccattcgagaagccgacgtagaagtagaggggaactcccagctgcaaatatccttctcaacagccctg
 gcaagcggcaggttcgctgcaagtgtgtccacaaagtacactgcgcagccgcatgccaccctcaaaggaccac
 45 atagtcaattaccagcatcacacaccaccttggggtcaggatatatccacaacggcaatgtcttgggtgcagaagatt
 acgggaggagtaggattaattgtgtgtgtgtgccttaattttaattgtgtgtgtatgcgtgtcttagcaggcactaatga

ggatccagatctgctgtgccttctagttgccagccatctgttgttggccctccccgtgccttcttgaccctggaagggtgc
 cactcccactgtcctttcctaataaaatgaggaaattgcatcgattgtctgagtaggtgtcattctattctgggggggtgggg
 tggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcgggtgggctctatgggtac
 ccaggtgctgaagaattgacccgggtcctcctggggccagaaagaagcaggcacatccccttctctgtgacacacctgtc
 5 cagcccctgggttcttagttccagccccactcataggacactcatagctcaggagggtccgccttcaatcccacccgcta
 aagtacttggagcgggtctctccctccctcatcagccccacaaaccaaacctagcctccaagagtgggaagaaattaag
 caagataggctattaagtgcagaggggagagaaaaatgcctccaacatgtgaggaagtaatgagagaaatcatagaatttta
 aggccatgatttaaggccatcatggccttaattctccgcttctcgtcactgactcgtcgtcggctcgttcggctgcgg
 cgagcgggtatcagctcactcaaaggcggtaatacgggtatccacagaatcaggggataacgcaggaaagaacatgtga
 10 gcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcggttttccataggctccgccccctgac
 gagcatcacaataatcgacgtcaagtcagaggtggcgaaacccgacaggactataaagataaccaggcgtttcccct
 ggaagctccctcgtgcgtctcctgttccgaccctgccgttaccggatacctgtccgcctttctccctcgggaagcgtg
 gcgctttctcatagctcacgctgtaggtatctcagttcgggtgtaggtcggtcgtccaagctgggctgtgtgcacgaacccc
 ccgttcagcccagccgtgcgccttatccgtaactatcgtttagtccaacccggtaagacacgacttatcgccactgg
 15 cagcagccactggtaacaggattagcagagcgaggtatgtaggcgggtctacagagttctgaagtgggtggcctaacta
 cggctacactagaagaacagatatttgggtatctgcgctcgtcgaagccagttacctcggaataagagttggtagctctga
 tccggcaaaacacaccgctggtagcgggtgggtttttgttgaagcagcagattacgcgcagaaaaaaaggatctca
 agaagatcctttgatctttctacggggctgacgctcagtggaacgaaaactcacgtaagggttttggatcatgagattat
 caaaaaggatcttcactagatccttttaataaaatgaagtttaaatcaatctaaagtatatatgagtaacttggctga
 20 cagttaccaatgcttaacagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactcggggggg
 gggggcgctgaggtctgcctcgtgaagaagggtgtgctgactcataccaggcctgaatcgccccatcatccagccagaa
 agtgagggagccacgggtgatgagagctttgttaggtggaccagttgggtatttgaactttgctttgccacggaacgg
 tctgcgtgtcgggaagatgcgtgatctgatcctcaactcagcaaaagttcgattattcaacaaagcccgctcccgtca
 agtcagcgtaatgctctgccagtgttacaaccaattaaccaattctgattagaaaaactcatcgagcatcaatgaaactgc
 25 aatttattcatatcaggattatcaataccatattttgaaaaagccgtttctgtaatgaaggagaaaaactcaccgaggcagttc
 cataggatggcaagatcctgggtatcgggtcgtcgattccgactcgtccaacatcaatacaacctattatccctcgtcaaa
 aataaggttatcaagtgagaaatccatagtgacgactgaatccgggtgagaatggcaaaagcttatgatttctttccag
 actgttcaacaggccagccattacgctcgtcatcaaaatcactcgcatcaaccaaaccgttattcattcgtgattgcgctg
 agcgagacgaaatacgcgatcgtgttaaaaggacaattacaacaggaatcgaatgaacggcgaggaacactg
 30 ccagcgcatcaacaataattttcacctgaatcaggatattcttctaatacctggaatgctgtttccggggatcgagtggtg
 agtaacatgcatcatcaggagtacggataaatgcttgatggcgggaagaggcataaattccgtcagccagtttagctg
 accatctcatctgtaacatcattggcaacgctacctttgccatgttcagaaacaactctggcgcatcgggcttccatacaa
 tcgatagattgtcgacactgattgcccagacattatcgcgagcccatttataccatataaatacagcatccatgttggaattaa
 tcgcggtcctgagcaagacgtttccggtgaatatggctcataacacccttgattactgttatgtaagcagacagtttat
 35 tgttcatgatgatatttttctgtgcaatgtaacatcagagattttgagacacaacgtggctttccccccccccattatt
 gaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaataagggttccgcgc
 acatttccccgaaaagtgccactgacgtctaagaaaccattattatcatgacattaacctataaaaataggcgatcacga
 ggccctttcgtc

40 OD4.3.2Vmut_3G/419N-3 CH (SEQ ID NO: 156):

atgcagctggccagctgtgtgacactgacactgggtgctgctgggtgaattccgcccctaggagacctgtgtgctctacaca
 gctgctgctgaatggctctctggccgaggaagaggtggtgatcagaagcgagaatttcaccaacaacgccaagaacat
 catcgtgcagctgagggaaacccgtgaagatcaactgcagccggcccaacaacaataccggagacaggccactgca
 acatcagcaagaccaactggaacaacaccctgaacaggtggtggagaagctgggcgagcagttcaacaagaccaa
 45 gatcgtgttaccagagcagcggcgagatcctgagatcgtgaccacagcttcaattgtccggcgagttcttact
 gcaataaccaccagctgttcaacagcacctggaatatcaggggcaccgccagcatcaacggaaccgagagcaacgat

aacatcacctgcctgcaacatcacctgagcggagccccccatcaagggaatattagctgcagcagcaacatca
caggactgctgctgacaagagatggcggcagcggcggcagcggccaggagacattcagacctggcggcggagact
gcagggacaattggcggagcggggatccctggagggtgctgttccaggggccaggccaccaccaccaccactg
atga

5

CHIKV-OD4.3.2Vmut_3G/419N-3 CH (SEQ ID NO: 157):

tcgcgcgtttcggatgacgggtgaaaacctctgacacatgcagctcccggagacgggtcacagcttctgtaagcggat
gccgggagcagacaagcccgtcagggcgctcagcgggtgttggcgggtgtcggggctggcttaactatgcggcatc
agagcagattgtactgagagtgcaccatatgcgggtgtgaaataccgcacagatgcgtaaggagaaaataccgcacaga
10 ttggctattggcattgcatacgttatccatatcataatatgtacattatattggctcatgtccaacattaccgccatgttga
cattgattattgactagttaataagtaataacacgggaactccatagcccatataggagtccgcgttacataacttac
gggaatttccaaacctggctgaccgcccacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgc
caatagggaacttcattgacgtcaatgggtggagtattacggtaaaactgccacttgggaatttccaagtgtatcatatgc
caagtacccccctattgacgtcaatgacgggaacttcataagcttgattatgccagtacatgaccttatgggaatttc
15 ctacttggcagtagcatctacgtattatgcctgattaccatgggtgatgcgggttttggcagtagcatcaatgggcgtggatag
cgggtttagctacgggaacttccaagctccacccattgacgtcaatgggagtttgggttagctacacaaaatcaacggga
attcccaaatgtcgtacaactccgccccattgacgcaaatgggcggtaggcgtgtacgggtgggaggtctatataagca
gagctcgttttagtaaccgtcagatgcctggagacgccatccacgctgtttgacctccatagaagacaccgggaccg
atccagcctccatcggtcgcacatctctcttcacgcgcccgcgcctacctgaggcccatccacgcccgttgagtc
20 gcgttctgcgcctcccgcctgtgggtgcctcctgaactgcgtccgccttaggtaagttaaagctcaggctcagacccg
ggccttttgcggcgctccctggagcctacatagactcagccggctctccacgcttgcctgacctgctgtcaactct
agttaacgggtggaggcagtgtagtctgagcagtagctgtgctgcgcgcgcgcaccagacataatagctgacagac
taacagactgttctttcatgggtctttctgcagtcaccgctgcgcacgtgtgatcagatacgcggccgcttagaca
ccatgtatagcatgcagctggccagctgtgtgacactgacactgggtgctgctggtgaattccgcccctaggagacctgtgt
25 gctctacacagctgctgctgaatggctctggtggcggaggaagagggtggtgatcagaagcgagaatttccaacaacgc
caagaacatcatcgtgcagctgagggaacccgtgaagatcaactgcagccggcccaacaacaataacccggagacagg
cccactgcaacatcagcaagaccaactggaacaacacctgaaacagggtgggtggagaagctgggcgagcagttcaac
aagaccaagatcgtgttcacccagagcagcggcgagatcctgagatcgtgacctacagcttcaattgtgccggcgag
ttcttactgcaataaccaccagctgttcaacagcacttggaatatcagagggcaccgccagcatcaacggaaccgaga
30 gcaacgataacatcacctgcccctgcaacatcacctgagcggagccccccatcaagggaatattagctgcagca
gcaacatcacaggactgtgctgacaagagatggcggcagcggcgagcggccaggagacattcagacctggcgg
cggagactgcagggacaattggcggagcggggatccctggagggtgctgttccaggggccaggccaccaccaccac
caccactgatgaagatcttaaggcggaattctgcagatgatctgctgtgccttctagtgtccagccatctgttgttcccc
tccccgtgccttcttgacctggaagggtgccactccactgtccttcttaataaaatagggaattgcatgcattgtct
35 gagtaggtgtcattctattctgggggggtgggtggggcaggacagcaagggggaggattgggaagacaatagcaggc
atgctgggggatgcgggtgggtctatgggtaccaggtgctgaagaattgacctgggttctcctggggcagaaagaagca
ggcacatcccccttctgtgacacaccctgtccacgccctgggtcttagttccagccccactcataggacactcatagctc
aggagggtccgccttcaatcccaccgcctaaagtacttgagcgggtctcctcctccatcagcccacaaacaaac
ctagcctccaagagtgggaagaaataaagcaagataggctattaagtgcagagggagagaaaatgcctccaacatgtg
40 aggaagtaatgagagaaatcatagaatttaaggccatgatttaaggccatcatggccttaattctccgcttctcgtcact
gactcgtcgtcgtcggctgttcggctgcggcgagcgggtatcagctcactcaaaggcggaataacgggttatccagaaat
caggggataacgcaggaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaggccggttgc
ggcgtttttccataggctccgccccctgacgagcatcacaataatcgacgtcaagtcagagggtggcgaaacccgaca
ggactataaagataaccaggcgtttccccctggaagctccctcgtgcgtctcctgttccgacctgcccgttaccggatac
45 ctgtccgcctttctccctcgggaagcgtggcgcttctcatagctcacgctgtaggtatctcagttcggttaggtcgttcg
ctccaagctgggtgctgtgcacgaacccccgttcagccgaccgtgcgccttatccggaactatcgtttagtcca

acccggttaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgagggtatgtaggcgggtg
 ctacagagttcttgaagtgggtggcctaactacggctacactagaagaacagtatttgggtatctgcgctctgctgaagccagt
 taccttcggaaaaagagttggtagctcttgatccggcaaaaccaccgctggtagcgggtgggtttttgttgaagcag
 cagattacgcgcagaaaaaaggatctcaagaagatcctttgatctttctacggggctgacgctcagtggaacgaaaa
 5 ctcacgttaagggattttgggtcatgagattatcaaaaaggatcttcacctagatccttttaataaaaaatgaagtttaaatca
 atctaaagtatatatgagtaaacttgggtctgacagttaccaatgcttaacagtgaggcacctatctcagcgatctgtctatttc
 gttcatccatagttgcctgactcggggggggggggcgctgaggtctgcctcgtgaagaaggtgttgctgactcatacca
 ggctgtaatcgccccatcatccagccagaaagtgaggagccacgggtgatgagagcttggtaggtggaccaggttg
 gtgatttgaacttttgccttgccacgggaacgggtctgcgtgtcggaagatgcgtgatctgaccttcaactcagcaaaagt
 10 tcgatttattcaacaagccgcccgtcccgtcaagtcagcgtaatgctctgccagtgttacaaccaattaaccaattctgatta
 gaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataccatattttgaaaaagccgtttctgt
 aatgaaggagaaaaactcaccgaggcagttccataggtggcaagatcctgggtatcgggtctgcgattccgactcgtccaa
 catcaatacaacctattaatttcccctcgcaaaaataaggttatcaagtgagaaatcaccatgagtgacgactgaatccgg
 tgagaatggcaaaagcttatgcatttcttccagactgttcaacaggccagccattacgctcgtcatcaaatcactcgc
 15 caaccaaaccttattcattcgtgattgcgcctgagcgagacgaatacgcgatcgtgttaaaaggacaattacaaca
 ggaatcgaatgaaccggcgaggaacactgccagcgcatcaacaatatttccactgaatcaggatatttcttaatacc
 tggatgctgttttccggggatcgagtggtgagtaacctgcatcatcaggagtagcggataaaatgcttgatggtcgga
 agaggcataaattccgtcagccagtttagtctgacctatcatctgtaacatcattggcaacgctacctttgccatgttcaga
 aacaactctggcgcatcgggcttccatacaatcgatagattgctgcacctgattgccgacattatcgcgagcccattat
 20 acccatataaatcagcatccatgttggaatttaatcgcgccctcgagcaagacgttcccgtgaatatggctcataacacc
 cctgtattactgtttatgtaagcagacagttttattgtcatgatgatattttatcttgcaatgtaacatcagagattttgag
 acacaactgtggctttccccccccccattattgaagcatttatcagggttattgtctcatgagcggatacatattgaatgta
 ttgaaaaataaacaataggggttccgcgcacatttcccgaagtgccacctgacgtctaagaaccattattatcat
 gacattaacctataaaaataggcgtatcacaggcccttctgctc

25

OD4.3Vmut3G-3CH (SEQ ID NO: 158):

Atgcagctggccagctgtgtgacactgacactgggtgctgctgggaattccgcccctaggagacctgtgtgctctacaca
 gctgctgctgaatggctctctggccgaggaagaggtgggtgatcagaagcgagaatttcaccaacaacgcaagaacat
 catcgtgcagctgagggaaaccgtgaagatcaactgcagccggcccaacaacaataaccggagacaggccactgca
 30 acatcagcaagaccaactggaacaacacctgaacaggtgggtggagaagctgggcgagcagttcaacaagaccaa
 gatcgtgttccacagagcagcggcgagatcctgagatcgtgaccacagcttcaattgtgccggcgagttcttact
 gcaataaccacccagctgttcaacagcacctggaatatcaggggcaccgccagcatcaacggaaccgagagcaacgat
 aacatcaccttgcctgcaggatcaagggaagcggagcccctcccataagggaatattagctgcagcagcaacatc
 acaggactgctgctgacaagagatggcggcagcggcgagcggccaggagacattcagacctggcggcgagac
 35 tgcagggacaattggcggagcaggggatccctggaggtgctgttcaggggccaggccaccaccaccaccactg
 atga

CHIKV-OD4.3Vmut3G-3CH (SEQ ID NO: 159):

tcgcgctttcgggtgatgacgggtgaaaacctctgacacatgcagctcccggagacgggtcacagcttgtctgtaagcggat
 40 gccgggagcagacaagcccgtcagggcgctcagcgggtgttggcggggtgcggggctggcttaactatgcggcatc
 agagcagattgtactgagagtgaccatatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaga
 ttggctattggccattgcatacgttgcataatataatgtacatttatattggctcatgtccaacattaccgccatgttga
 cattgattattgactagttattaatagtaatcaattacgggaacttccatagcccatataggagtccgcgttacataacttac
 gggaatttcaaacctggctgaccgccaacgacccccgccattgacgtcaataatgacgtatgttcccatagtaacgc
 45 caatagggaacttccattgacgtcaatgggtggagtattacggtaaaactcccacttgggaatttcaagtgtatcatatgc
 caagtacgccccctattgacgtcaatgacgggaacttccataagcttgattatgccagttacatgaccttatgggaatttc

ctacttggcagtacatctacgtattagtcacgctattaccatgggtgatgcggtttggcagtacatcaatgggcgtggatag
 cggtttgactcacgggaacttccaagtctccacccattgacgtcaatgggagttgtttgactcacaaaaatcaacggga
 attcccaaatgtcgaacaactccgccccattgacgcaaatgggcggtaggcgtgtacgggtgggaggtctatataagca
 gagctcgtttagtgaaccgtcagatcgctggagacgccatccacgctgtttgacctccatagaagacacgggaccg
 5 atccagcctccatcggtcgcacatctctcttcacgcgcccgcgcctacctgaggccgcatccacgcccgttgagtc
 gcgttctgccgctccccgctgtgggtcctcctgaactgcgtccgcgtctaggtaagttaaagctcaggtcgagaccg
 ggctttgtccggcgtcccttggagcctacctagactcagccggctcctccacgctttgcctgacctgcttgcctcaactct
 agttaacgggtggaggggcagtgtagtctgagcagtagctgttgcgtccgcgcgccaccagacataatagctgacagac
 taacagactgttcttccatgggtctttctgcagtcaccgtcgtcgacacgtgtgatcagatatcgcgccgctctagaca
 10 ccatgtatagcatgcagctggccagctgtgtgacactgacactgggtgctgctgggtgaattccgcccctaggagacctgtgt
 gctctacacagctgctgtgaatggctctctggccgaggaagaggtgggtgatcagaagcgagaatttcaccaacaacgc
 caagaacatcatcgtgcagctgagggaacccgtgaagatcaactgcagccggcccaacaacaatacccgagacagg
 cccactgcaacatcagcaagaccaactggaacaacaccctgaaacaggtgggtggagaagctgggcgagcagttcaac
 aagaccaagatcgtgttccaccagagcagcggcgagatcctgagatcgtgaccacagcttcaattgtccggcgag
 15 ttcttactgcaataccaccagctgttcaacagcacctggaatatcaggggcaccgccagcatcaacggaaccgaga
 gcaacgataacatcacctgcctgcaggatcaagggaagcggagcccctccatcaagggaatattagctgcagca
 gcaacatcacaggactgctgtgacaagagatggcgccagcggcgagcggccaggagacattcagacctggcg
 cgggagactgcagggaacattggcgagcaggggatccctggaggtgctgttccaggggccaggccaccaccaccac
 caccactgatgaagatcttcaaggcggaattctgcagatgatctgctgtgccttctagtgtccagccatctgttgttcccc
 20 tccccgtgccttcttgacctggaaggtgccactcccactgtccttcttaataaaatgaggaaattgcatcgattgtct
 gagtaggtgtcattctattctgggggggtgggggtggggcaggacagcaagggggaggattgggaagacaatagcaggc
 atgctggggatgcggtgggtcttatgggtaccaggtgctgaagaattgaccgggtcctcctggggccagaaagaagca
 ggcacatccccttctctgtgacacaccctgtccacgcccctgggttcttagttccagccccactcataggacactcatagctc
 agggagggtccgccttcaatcccaccgctaaagtacttggagcgggtctctcctcctcatcagcccaccaaaccac
 25 ctacgtccaagagtgggaagaaattaaagcaagataggctattaagtgcagaggagagaaaaatgcctccaacatgtg
 aggaagtaatgagagaaatcatagaattttaaggccatgatttaaggccatcatggccttaattctccgcttctcgtcact
 gactcgtcgcgtcggctcgttcggctcggcgagcgggtatcagctcactcaaaggcggtataacgggtatccacagaat
 caggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgctgtgct
 30 ggcgtttttccataggctccgccccctgacgagcatcaaaaaatcgacgctcaagtgcaggtggcgaaacccgaca
 ggactataaagataaccaggcgtttccccctggaagctcctcgtgcgtctcctgttccgacctgcccgttaccggatac
 ctgtccgcttttctccttcgggaagcgtggcgcttttctatagctcacgctgtaggtatctcagttcggtgtaggtcgttcg
 ctcaaagctgggctgtgtgcacgaacccccgttcagcccagcgtgcgccttatccggtaactatcgtcttgagcca
 acccggtgaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcaggtatgtaggcggtg
 35 ctacagagttcttgaagtgggtggcctaactacggctacactagaagaacagatttggtatctgcgctctgctgaagccagt
 taccttcggaaaaagagttggtagctttgatccggcaaaacacaccgctggtagcgggtgtttttgtttgcaagcag
 cagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacgggggtctgacgctcagtggaacgaaaa
 ctacgtaagggtatttggatgagattatcaaaaaggatcttcacctagatccttttaataaaaaatgaagttttaaatca
 atctaaagtatatatgagtaaacttgggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctattc
 40 gttcatccatagttgcctgactcggggggggggggcgctgagggtcgcctcgtgaagaagggtgttgcctgactcatacca
 ggctgaatcgccccatcatccagccagaaagtggaggagccacgggtgatgagagcttgtttaggtggaccaggtg
 gtgattttgaacttttgccttggcacggaacggctcgttgcgttgcgggaagatgcgtgatctgatccttcaactcagcaaaagt
 tcgatttattcaaaaagccgccgtccgtcaagtgcgtaagtctcgtccaggtttacaaccaattaaccaattctgatta
 gaaaaactcatcgagcatcaaatgaaactgaatttattcatatcaggattatcaataccatattttgaaaaagccgtttctgt
 45 aatgaaggagaaaaactcaccgaggcagttccataggatggcaagatcctggatcggtctgcgattccgactcgtccaa
 catcaatacaacctattaatttcccctcgtcaaaaataagggtatcaagtgaagaatcaccatgagtgacgactgaatccgg
 tgagaatggcaaaagcttatgcatttcttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcac

caaccaaaccgttattcattcgtgattgcgcctgagcgagacgaatacgcgatcgtgttaaaaggacaattacaaaca
 ggaatcgaatgcaaccggcgaggaacactgccagcgcatcaacaatattttcacctgaatcaggatattcttctaatacc
 tggaatgctgtttccggggatcgagtggtgagtaacatgcatcatcaggagtagggataaaatgcttgatggcgga
 agaggcataaattccgctcagccagtttagtctgacatctcatctgaacatcattggcaacgctacctttgccatgttcaga
 5 aacaactctggcgcatcgggcttccatacaatcgatagattgtcgcacctgattgcccacattatcgcgagcccattat
 acccatataaatcagcatccatgttggaatttaatcgcgccctcgagcaagacgttcccgttgaatatggctcataacacc
 cttgtattactgtttatgtaagcagacagttttattgtcatgatgatattttatcttgcaatgtaacatcagagattttgag
 acacaacgtggctttccccccccccattattgaagcatttatcagggttattgtctcatgagcggatacatattgaatgta
 tttagaaaaataacaaataggggtccgcgcacattccccgaaaagtccacctgacgtctaagaaccattattatcat
 10 gacattaacctataaaaataggcgatcacgagcccttcgtc

CMV/R_CHIKV185_OD4.2'_HG3.2 (SEQ ID NO: 160):

tcgcgcgtttcgggtgatgacggtgaaaacctctgacacatgcagctcccggagacgggtcacagcttgtctgtaagcggat
 gccgggagcagacaagcccgtcagggcgctcagcgggtgttgccgggtgtcggggctggcttaactatgcggcatc
 15 agagcagattgtactgagagtgacccatattcggtgtgaaataaccgcacagatgcgtaaggagaaaataaccgcatcaga
 ttggctattggccattgcatacgttgatccatatacaatatgtacatttatattggctcatgtccaacattaccgccatgttga
 cattgattattgactagttattaatagtaatacaattacggggcttagttcatagcccatatatggagtccgcgttacataact
 tacggtaaatggcccgcctggctgacgcgccaacgacccccgccattgacgtcaataatgacgtatgttcccatagtaa
 cgccaataggagctttccattgacgtcaatgggtggagattttacggtaaacctgcccacttgccagtagatcaagtgtatca
 20 tatgccaaagtagccccctattgacgtcaatgacggtaaatggccgcctggcattatgccagtagatgaccttatggga
 ctttctacttggcagtagatctacgtattagtcacgtattaccatgggtgatgcgggtttggcagtagatcaatgggcgtgg
 atagcgggttgactcacggggatttccaagtctccacccattgacgtcaatgggagttgttttggcaccaaaatcaacgg
 gactttccaaaatgtcgaataactccgcccattgacgcaaatgggcggtaggcgtgtacgggtgggaggtctatataag
 cagagctcgttttagtaaccgtcagatcgctggagacgccatccacgtgtttgacctccatagaagacacggggacc
 25 gatccagcctccatcggctcgcacatctctcttcacgcgcccgccctacctgaggcccatccacgccggttagt
 cgctgttctcgcctcccgcctgtggtgcctctgaactgcgtccgccttaggttaagttaaaagctcaggctcagagacc
 gggcctttgtccggcgctcccttgagcctacctagactcagccggctctccacgcttgcctgacctgcttgcctaactc
 tagttaacgggtggagggcagtgtagtctgagcagtagctgttgcgtccgcgcgccaccagacataatagctgacaga
 ctaacagactgttctttccatgggtcttttgcagtcaccgtcgtcgacagctgtgatcagatatcgggccgctctagac
 30 accatggaggtcatcccgacgcaaatcttataacagaaggtaccaacccgacctgggccccacgccctacaattca
 agtaattagacctagaccacgtccacagaggcaggctgggcaactcggcagctgatctccgagtagcaaaaattgacc
 atgcgcgcgttacctcaacagaagcctcgagaaatcggaacaaagaagcaaggcagaagaagcaggcgccg
 caaacgacccaaagcaaaagaagcaaccaccacaaaagaagccggctcaaaagaagaagaaccaggccgtagg
 gagagaatgtgcatgaaaattgaaaatgattgcatcttcgaagtcaagcatgaaggcaaaagtgtgggctacgcatgcct
 35 ggtgggggataaagtaataaaccagcacatgtgaagggaactatcgacaatccgatctggctaaactggcctttaag
 cggctgtctaaatacagatcttgaatgtgcacagataccggtgcacatgaagtctgatgcctgaagtttaccacagagaaa
 cccgaggggtactataactggcatcacggagcagtgacgtattcaggaggccgggttactatcccagcgggtgcaggc
 aagccgggagacagcggcagaccgatcttcgacaacaaaggacgggtggtggccatcgtcctaggaggggccaacg
 aaggtgcccgcacggccctctccgtggtgacgttggaacaaagacatcgtcacaaaaattaccctgagggagccgaa
 40 gagtgaggacctcgcctcccgttctgtgcctgttggcaaacactacattcccctgctctcagccgcttgcacaccctgc
 tgctacgaaaaggaaaccgaaagcaccttgcgatgcttgaggacaacgtgatgagaccggatactaccagctacta
 aaagcatcgtgacttgcctccccaccgcaaaagacgcagtactaaggacaatttaattgtctataaagccacaagacc
 atatctagctcattgtcctgactgcggagaagggcattcgtccacagccctatcgattggagcgcacatcagaatgaag
 caacggacggaacgctgaaaatccaggtctctttgcagatcgggataaagacagatgacagccacgattggaccaagc
 45 tgcgtatatggatagccatagccagcggacgcggagcagccgattgcttgaaggacttcagcaccgtgcacgat
 caccgggacatgggacactttattctcggccgatgccgaaaggagagacgtgacagtgaggatttaccggacagcag

- 93 -

ggcggaataacgggtatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggcca
 ggaaccgtaaaaaggccgctgtgctggcggttttccataggtccgccccctgacgagcatcaaaaatcgacgctc
 aagtcagaggtggcgaaacccgacaggactataaagataaccaggcggtttcccctggaagctccctcgtgcgctcct
 5 gttccgacctgcccgttaccggatacctgtccgcctttctccctcgggaagcgtggcgctttctcatagctcacgctgta
 ggatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccgttcagcccagccgctgcgcc
 ttatccggaactatcgtcttgagccaacccggaagacacgacttatcgccactggcagcagccactggtaacaggatt
 agcagagcgaggtatgtaggcggtgctacagagttctgaagtgggtggcctaactacggctacactagaagaacagtatt
 tggatctgcgctcgtgtaagccagttaccttcggaaaaagagttggtagctcttgatccggcaacaaaccaccgctgg
 tagcgggtggtttttgtttgcaagcagcagattacgcgcagaaaaaaggatctcaagaagatcctttgatctttctacgg
 10 ggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatcct
 tttaaattaaaaatgaagttttaaatacaatctaaagtatatatgagtaaaacttggctgacagttaccaatgcttaacagttag
 gcacctatctcagcgaatctgtctatttcgttcatccatagttgcctgactcggggggggggggcgctgaggtctgcctcgt
 gaagaagggtgtgctgactcataccaggcctgaatcgccccatcatccagccagaaaagttagggagccacgggtgatg
 agagctttgttaggtggaccagttggtgattttgaacttttgccttgccacggaacggtctgcgttgcgggaagatgcgt
 15 gatctgatccttcaactcagcaaaagttcgattattcaaaaagcccgctcccgtaagtcagcgaatgctctgccagt
 gttacaaccaattaaccaattctgattagaaaaactcatcgagcatcaaatgaaactgcaattattcatatcaggattatcaa
 taccatattttgaaaaagccgtttctgtaatgaaggagaaaaactcaccgaggcagttccataggatggcaagatcctggta
 tcggctgcgattccgactcgtccaacatcaatacaacctattaatttcccctcgtcaaaaataagggtatcaagttagaaaat
 caccatgagtgacgactgaatccggtgagaatggcaaaagcttatgcatctttccagacttggtcaacaggccagccatt
 20 acgctcgtcatcaaaactcctgcacaaacacggtattcattcgtgattgctgcctgagcagagacgaaatacgcgac
 gctgttaaaaggacaattacaacaggaatcgaatgaaccggcgaggaacactgccagcgcacatcaacaattttca
 cctgaatcaggatattcttaatacctggatgctgtttcccggggagtcagtggtgagtaacatgcatcatcaggag
 tacggataaaatgcttgatggtcggaagaggcataaattccgtcagccagtttagtctgacctctcatctgtaacatcattg
 gcaacgctacctttgccatgttcagaaacaactctggcgcatcggttccatacaatcgatagattgtcgcacctgatt
 25 gcccagacattatcgcgagcccatttatcccatataaatcagcatccatgttggaattaatcgcgccctcgagcaagacgt
 tccccgtgaatatggctcataacaccccttgattactgttatgtaagcagacagttttattgttcatgatgatattttatctt
 gtgcaatgtaacatcagagattttgagacacaacgtggcttccccccccccattattgaagcatttatcagggttattgt
 ctcatgagcggatacatatttgatgtatttagaaaaataaacaataaggggtccgcgcacattccccgaaaagtgccac
 ctgacgtctaagaaaccattattatcatgacattaacctataaaaataggcgatcacgaggcccttctgc

30 CMV/R_CHIKV185_OD4.2'_HG3.2.1 (SEQ ID NO: 161):
 tcgcgctttcgggtgatgacgggtgaaaacctctgacacatgcagctcccggagacgggtcacagcttctgtaagcggat
 gccgggagcagacaagcccgtcagggcgctcagcgggtgttgccgggtgtcggggctggcttaactatgcggcatc
 35 agagcagattgtactgagagtgaccatattgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaga
 ttggctattggccattgcatacgttgatccatataatgtacatttatattggctcatgtccaacattaccgccatgttga
 cattgattattgactagttattaatagtaataattacggggctcattagttcatagcccatatatggagttccgcgttacataact
 tacggtaaatggcccgcctggctgaccgccaacgacccccgccattgacgtcaataatgacgtatgttcccatagtaa
 cgccaatagggactttccattgacgtcaatgggtggagtatttacggtaaaactgccacttggcagtacatcaagtgtatca
 40 tatgccaaagtacgccccctattgacgtcaatgacggtaaatggccgcctggcattatgccagctacatgaccttatggga
 ctttctacttggcagctacatctacgtattatgcatcgtattaccatgggtgatgcggttttggcagctacatcaatgggcgtgg
 atagcgggttgactcacggggatttccaagtctccacccattgacgtcaatgggagttgttttggcaccaaaatcaacgg
 gactttccaaaatgtcgaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataag
 cagagctcgttttagtaaccgtcagatcgctggagacgccatccacgctgtttgacctccatagaagacaccgggacc
 gatccagcctccatcggctcgcactctctcttcacgcgcccgcgccctacctgaggcccatccacgccgggttagt
 45 cgcgttctccgctcccgcctgtggtgctcctgaactgcgtccgcgtctaggttaagtttaaagctcaggtcgagacc
 gggcctttgtccggcgctcccttgagcctacctagactcagccggtctccacgcttgcctgacctgctgtcctaact

tagttaacgggtggaggcagtgtagtctgagcagtactcgttgctgccgcgcgcgccaccagacataatagctgacaga
 ctaacagactgttctttccatgggtctttctgcagtcaccgtcgtcgacacgtgtgatcagatatcgcgccgctctagac
 accatggagttcatcccagcgaactttctataacagaaggtaccaaccccaccctgggccccacgcctacaattca
 agtaattagacctagaccacgtccacagaggcaggctgggcaactcgccagctgatctccgcagtcaacaaattgacc
 5 atgcgcgcgggtacctcaacagaagcctcgcagaaatcggaaaaaaagaagcaaaaggcagaagaagcaggcgccg
 caaacgacccaaagcaaaagaaagcaaccaccacaaaagaagccggctcaaaagaagaagaaccaggccgtagg
 gagagaatgtgcatgaaaattgaaaatgattgcatcttcgaagtcaagcatgaaggcaaaagtgtgggtctacgcatgcct
 ggtgggggataaagtaatgaaaccagcacatgtgaagggaactatcgacaatgccgatctggctaaactggcctttaag
 cggtcgtctaaatacgtcttgatgtgcacagataccgggtgcacatgaagtctgatgcctcgaagttatcccacgagaaa
 10 cccgaggggtactataactggcatcacggagcagtgacgtattcaggaggccgggtcactatcccaggggtgcaggc
 aagccgggagacagcggcagaccgatcttcgacaacaaaggacgggtgggtggccatcgtcctaggagggggccaacg
 aagggtgccgcacggccctctccgtgggtgacgtggaacaaagacatcgtcacaacaaattacccctgagggagccgaa
 gaggtaggcctcgcctcccgggtcttgctgctgtggcaaacactacattcccctgctctcagccgccttgcacaccctgc
 tgctacgaaaaggaaccggaaagcaccttgcgcatgcttgaggacaacgtgatgagacccggatactaccagctacta
 15 aaagcatcgtgacttgctctccccaccgccaagacgcagtactaaggacaattttaatgtctataaagccacaagacc
 atatctagctcattgtcctgactgcggagaagggtccttgccacagccctatcgattggagcgcacagaaatgaag
 caacggacggaaacgctgaaaatccaggtctcttgcagatcgggataaagacagatgacagccacgattggaccaagc
 tgcgctatatggatagccatacgcagcggacgcggagcgcgggattgcttgtaaggacttcagaccgtgcacgat
 caccgggacatgggacactttattctgcccgatgcccgaaggagagacgctgacagtggtgattacggacagcag
 20 aaagatcagccacacatgcacacaccgttccatcatgaaccctgtgataggtaggagaggttccactctcgacca
 caacatggtaaagagttaccttgacgcacgtacgtgcagagcaccgctgccactgctgaggagatagaggtgcatatgc
 cccagatactctgaccgcacgctgatgacgcagcagtcgggaaagcccgtggtgagcaccagctgctgtgaaac
 ggcagcctggccgagaaggagatcaggatcaagagcgcaggacatcagcgacaacgccaagaacatcatcgtgcagc
 tgaccaagcccgtgctgatcaactgcaccaggcccagcaacaacaccagtagtgcgtggtgaacaggaccagtggtg
 25 aacgacacctgggcccaggtggccatccagctgaggaagcactggaacacctgcatcatcttcaccaacagcagcgg
 cggcgacctggagatcaccacccacagcttcaactgcggcgaggagttcTTCtactgcaacaccagcgacctgttca
 acagcacttggaaacatcaacggcaccgcccagcatcaacggcaccgagagcaacgacaacatcaccttgccttcgag
 gatcaacggcaccatccagggaacatcacctgccagagcaacatcaccggcatcctgctgaccaggagcggcgga
 gcggcagcggcacctgcgagaccttcaggcccggcgggcgacatgagggacctgaacaggaccagcggctccg
 30 gaaacgtgaagatcacagttaatgggcagacgggtgcggtacaagtgcactgcggtggctcaaacgagggactgaca
 accacagacaaaagtgatcaataactgaaaattgatcagtgccatgctgcagtcactaatcacaagaattggcaataca
 ctccccttagtcccgcgcaacgctgaactcggggaccgttaaaggaaagatccacatcccattcccattggcaaacgtga
 cttgcagagtgccaaaagcaagaaaccctacagtaacttacggaaaaaaccaagtcacatgctgctgtatcctgacat
 ccgacactcttgtcttaccgtaacatgggacaggaaccaattaccacgaggagtggggtgacacacaagaaggaggtta
 35 ccttgaccgtgcctactgaggggtctggaggtcacttggggcaacaacgaaccatacaagtactggccgcagatgtctac
 gaacgggtactgctcatggtcaccacatgagataatctgtactattatgagctgtacccactatgactgtagtattgtgc
 ggtggcctcgttgcgttctgtcgtatgggtgggcacagcagtggggaatgtgtgtgctgcgcacggcgagatgcattacac
 catatgaattaacaccaggagccactgttccctcctgctcagcctgctatgctgctcagaacgaccaaggcggccaca
 tattacgaggtcgcggcatatctatggaacgaacagcagccctgttctggttgacggctcttatcccgtggccgcttg
 40 atcgtcctgtgcaactgtctgaaactcttgccatgctgctgtaagaccctggcttttttagccgtaatgagcatcggtgccca
 cactgtgagcgcgtacgaacacgtaacagtgatcccgaacacgggtgggagtaccgtataagactcttgtcaacagaccg
 ggttacagccccatgggtgttgagatggagctacaatcagtcaccttgaaccaacactgtcacttgactacatcacgtgc
 gagtacaaaactgtcatcccctcccgtacgtgaagtgtgtgtgtgtacagcagagtgaaggacaagagcctaccagact
 acagctgcaaggtctttactggagtctaccattatgtggggcgccgctactgcttttgacgcgcccgaataacgcaatt
 45 gagcagggcacatgtagagaaatctgaatctgcaaacagagtttgcatcgccctacagagcccacaccgcacgcggc
 gtcggcggaagctccgcgtcctttaccaaggaaacaacattaccgtagctgcctacgctaaccggtagcatgcctcaca

gtaaaggacgccaagtttgcgtggggcccaatgtcctccgctggacacctttgacaacaaaatcgtgggtgacaaagg
 cgacgtctacaacatggactaccaccttttggcgaggaagaccaggacaatttggtagacattcaaagtcgtacaccgg
 aaagtaaagacgtttatgccaacactcagttgggtactacagaggccagcagcaggcacgggtacatgtaccatactctcag
 gcacatctggcttcaagtattggctgaagggaacgaggagcatcgctacagcacacggcaccgttcgggttgccagattg
 5 cgacaaacccggtaagagctgtaaattgcgtgtggggaacataccaatttccatcgacataccggatgcggcctttact
 aggggtgtcgatgcaccctctgaacggacatgtcatgcgaagtaccagcctgcactcactcctccgactttggggggcgt
 cgccatcatcaaatacacagctagcaagaaaggtaaattgtgcagtacattcgtatgaccaacgccgttaccattcgagaag
 ccgacgtagaagtagaggggaactcccagctgcaaatactcttcaacagccctggcaagcgccgagtttcgctgca
 agtgtgctccacacaagtacactgcgcagccgatgccaccctccaaaggaccacatagtaattaccagcatcacac
 10 accacccttgggggtccaggatatatccacaacggcaatgtcttgggtgcagaagattacgggaggagtaggattaattgtt
 gctgttgccttaatttaattgtgggtctatgcgtgtcgtttagcaggcactaatgaggatccagatctgctgtgccttcta
 gttgccagccatctgttgtttgccctccccgtgccttcttgacacctggaagggtgccactcccactgtcctttctaataaa
 atgaggaaattgcatcgattgtctgagtaggtgtcattctattctgggggggtgggggtggggcaggacagcaagggggga
 ggattgggaagacaatagcaggcatgctggggatgcgggtgggctctatgggtaccagggtgctgaagaattgacccgg
 15 ttctctggggccagaaagaagcaggcacatccccttctgtgacacaccctgtccacgcccctgggttctagtccagc
 cccactcataggacactcatagctcaggaggggtccgccttcaatcccaccgctaagttgtagcggtctctccct
 ccctcatcagcccaccaaaccacctaagcctccaagagtgggaagaaattaaagcaagataggctattaagtgcagag
 ggagagaaaaatgctccaacatgtgaggaagtaatgagagaaatcatagaatttaaggccatgatttaaggccatcatg
 gccttaattctccgcttctcgtcactgactcgtgcgtcggtcgttcggctgcggcgagcggtatcagctcactcaaa
 20 ggcggtataacggttatccacagaatcaggggataacgcaggaagaacatgtgagcaaaaggccagcaaaaggcca
 ggaaaccgtaaaaaggccgctgtgctggcggttttccataggctccgccccctgacgagcatcacaaaaatcgacgctc
 aagttagaggtggcgaaacccgacaggactataaagataccaggcggttccccctggaagctccctcgtgcgtctcct
 gttccgaccctgcccgttaccggatacctgtccgcctttctccctcgggaagcgtggcgctttctcatagctcacgctgta
 ggatctcagttcgggtgtaggtcgttcgctcaagctgggctgtgtgcacgaacccccgttcagcccagccgctgcgcc
 25 ttatccggtaatatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggatt
 agcagagcgagggtatgtaggggtgctacagagttctgaagtgggtggcctaactacggctacactagaagaacagatt
 tggtatctgcgtctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaacaaaccaccgctgg
 tagcgggtgggtttttgttgcaagcagcagattacgcgcagaaaaaaggatctcaagaagatcctttgatctttctacgg
 ggtctgacgctcagtggaacgaaaactcacgttaagggttttggctcatgagattatcaaaaaggatcttcacctagatcct
 30 tttaaataaaaatgaagtttaaatcaatctaaagtatatatgagtaaaacttggctgacagttaccaatgcttaacagttag
 gcacctatctcagcagatctgtctatttctcatccatagttgcctgactcggggggggggggcgctgaggtctgcctcgt
 gaagaagggtgtgctgactcataccaggcctgaatcgccccatcatccagccagaaaagttagggagccacgggtgatg
 agagctttgtgtaggtggaccagtgtggtgatttgaactttgctttgccacggaacgggtctgcgtgtcggggaagatgcgt
 gatctgatccttcaactcagcaaaagtctgatttattcaaaaagccgctcccgtaagtcagcgaatgctctgccagt
 35 gttacaaccaattaaccaattctgattagaaaaactcatcgagcatcaaatgaaactgaatttattcatatcaggattatcaa
 taccatatttttgaaaaagccgtttctgtaatgaaggagaaaaactcaccgaggcagttccataggtggcaagatcctggta
 tcgggtctgcgattccgactcgtccaacatcaatacaacctattatccctcgtcaaaaataaggttatcaagtgagaaat
 caccatgagtgacgactgaatccgggtgagaatggcaaaaagcttatgcatttcttccagacttgttcaacaggccagccatt
 acgctcgtcatcaaaatcactcgcatcaacaaacgggtatttctcgtgattgcgcctgagcgagacgaaatacgcgatc
 40 gctgttaaaaggacaattacaacagggaatcgaatgcaaccggcgaggaacactgccagcgcatcaacaatattttca
 cctgaatcaggatattcttaatacctggaatgctgtttccggggatcgagtggtgagtaaccatgcatcatcaggag
 tacggataaaatgcttgatggctggaagaggcataaattccgtcagccagtttagtctgaccatctcatctgtaacatcattg
 gcaacgctacctttgccatgtttcagaacaactctggcgcatcggggttccatacaatcgatagattgtcgacactgatt
 gcccgaattatcgcgagcccattatacccatataaatcagcatccatgttggaatttaacgcggcctcgagcaagacgt
 45 tccccgtgaatatggctcataacacccctgtattactgtttatgtaagcagacagttttattgtcatgatgatattttatctt
 gtgcaatgtaacatcagagattttgagacacaacgtggctttccccccccccattattgaagcatttatcagggttattgt

ctcatgagcggatacatatttgaatgtatttagaaaaataaacaatataggggtccgcgcacatttccccgaaaagtgccac
ctgacgtctaagaaccattattatcatgacattaacctataaaaaataggcgtatcacgaggcccttcgtc

C. Therapeutic Methods and Pharmaceutical Compositions

5 The antigens as disclosed herein, or a nucleic acid molecule encoding the disclosed immunogen, can be administered to a subject in order to generate an immune response to a pathogen, such as HIV.

In exemplary applications, compositions are administered to a subject suffering from HIV infection or at risk of becoming infected from HIV. In other
10 applications, the immunogens disclosed herein can be administered prophylactically, for example as part of an immunization regimen. The antigen is administered in an amount sufficient to raise an immune response against HIV.

The immunogen is administered in an amount sufficient to raise an immune response against HIV virus. Administration induces a sufficient immune response to
15 treat the pathogenic infection, for example, to inhibit the infection and/or reduce the signs and/or symptoms of the infection. Amounts effective for this use will depend upon the severity of the disease, the general state of the subject's health, and the robustness of the subject's immune system. A therapeutically effective amount of the antigen is that which provides either subjective relief of a symptom(s) or an
20 objectively identifiable improvement as noted by the clinician or other qualified observer.

An antigen can be administered by any means known to one of skill in the art (see Banga, A., "Parenteral Controlled Delivery of Therapeutic Peptides and Proteins," in *Therapeutic Peptides and Proteins*, Technomic Publishing Co., Inc.,
25 Lancaster, PA, 1995) either locally or systemically, such as by intramuscular, subcutaneous, or intravenous injection, but even oral, nasal, or anal administration is contemplated. In one embodiment, administration is by subcutaneous or intramuscular injection. To extend the time during which the disclosed antigen is available to stimulate a response, the antigen can be provided as an implant, an oily
30 injection, or as a particulate system. The particulate system can be a microparticle, a microcapsule, a microsphere, a nanocapsule, or similar particle. (see, e.g., Banga, *supra*). A particulate carrier based on a synthetic polymer has been shown to act as

an adjuvant to enhance the immune response, in addition to providing a controlled release. Aluminum salts can also be used as adjuvants to produce an immune response.

Optionally, one or more cytokines, such as interleukin (IL)-2, IL-6, IL-12, IL-15, RANTES, granulocyte macrophage colony stimulating factor (GM-CSF), tumor necrosis factor (TNF) - α , interferon (IFN)- α or IFN- γ , one or more growth factors, such as GM-CSF or G-CSF, one or more costimulatory molecules, such as ICAM-1, LFA-3, CD72, B7-1, B7-2, or other B7 related molecules; one or more molecules such as OX-40L or 41 BBL, or combinations of these molecules, can be used as biological adjuvants (see, for example, Salgaller *et al.*, 1998, *J. Surg. Oncol.* 68(2):122-38; Lotze *et al.*, 2000, *Cancer J Sci. Am.* 6(Suppl 1):S61-6; Cao *et al.*, 1998, *Stem Cells* 16(Suppl 1):251-60; Kuiper *et al.*, 2000, *Adv. Exp. Med. Biol.* 465:381-90). These molecules can be administered systemically (or locally) to the host. In several examples, IL-2, RANTES, GM-CSF, TNF- α , IFN- γ , G-CSF, LFA-3, CD72, B7-1, B7-2, B7-1 B7-2, OX-40L, 41 BBL and ICAM-1 are administered.

A number of means for inducing cellular responses, both *in vitro* and *in vivo*, are known. Lipids have been identified as agents capable of assisting in priming CTL *in vivo* against various antigens. For example, as described in U.S. Patent No. 5,662,907, palmitic acid residues can be attached to the alpha and epsilon amino groups of a lysine residue and then linked (for example, via one or more linking residues, such as glycine, glycine-glycine, serine, serine-serine, or the like) to an immunogenic peptide. The lipidated peptide can then be injected directly in a micellar form, incorporated in a liposome, or emulsified in an adjuvant. As another example, *E. coli* lipoproteins, such as tripalmitoyl-S-glycerylcysteinylserine can be used to prime tumor specific CTL when covalently attached to an appropriate peptide (see, Deres *et al.*, *Nature* 342:561, 1989).

A pharmaceutical composition including an isolated immunogen is provided. In some embodiments, the immunogen is mixed with an adjuvant containing two or more of a stabilizing detergent, a micelle-forming agent, and an oil. Suitable stabilizing detergents, micelle-forming agents, and oils are detailed in U.S. Patent No. 5,585,103; U.S. Patent No. 5,709,860; U.S. Patent No. 5,270,202; and U.S.

Patent No. 5,695,770. A stabilizing detergent is any detergent that allows the components of the emulsion to remain as a stable emulsion. Such detergents include polysorbate, 80 (TWEEN) (Sorbitan-mono-9-octadecenoate-poly(oxy-1,2-ethanediyl; manufactured by ICI Americas, Wilmington, DE), TWEEN 40™, 5 TWEEN 20™, TWEEN 60™, ZWITTERGENT™ 3-12, TEEPOL HB7™, and SPAN 85™. These detergents are usually provided in an amount of approximately 0.05 to 0.5%, such as at about 0.2%. A micelle forming agent is an agent which is able to stabilize the emulsion formed with the other components such that a micelle-like structure is formed. Such agents generally cause some irritation at the site of 10 injection in order to recruit macrophages to enhance the cellular response. Examples of such agents include polymer surfactants described by BASF Wyandotte publications, *e.g.*, Schmolka, *J. Am. Oil. Chem. Soc.* 54: 110, 1977, and Hunter *et al.*, *J. Immunol* 129:1244, 1981, PLURONIC™ L62LF, L101, and L64, PEG1000, and TETRONIC™ 1501, 150R1, 701, 901, 1301, and 130R1. The chemical structures of 15 such agents are well known in the art. In one embodiment, the agent is chosen to have a hydrophile-lipophile balance (HLB) of between 0 and 2, as defined by Hunter and Bennett, *J. Immun.* 133:3167, 1984. The agent can be provided in an effective amount, for example between 0.5 and 10%, or in an amount between 1.25 and 5%.

The oil included in the composition is chosen to promote the retention of the 20 antigen in oil-in-water emulsion, for example to provide a vehicle for the desired antigen, and preferably has a melting temperature of less than 65°C such that emulsion is formed either at room temperature (about 20°C to 25°C), or once the temperature of the emulsion is brought down to room temperature. Examples of such oils include squalene, Squalane, EICOSANE™, tetratetracontane, glycerol, and 25 peanut oil or other vegetable oils. In one specific, non-limiting example, the oil is provided in an amount between 1 and 10%, or between 2.5 and 5%. The oil should be both biodegradable and biocompatible so that the body can break down the oil over time, and so that no adverse effects, such as granulomas, are evident upon use of the oil.

30 In one embodiment, the adjuvant is a mixture of stabilizing detergents, micelle-forming agent, and oil available under the name PROVAX® (IDEC

Pharmaceuticals, San Diego, CA). An adjuvant can also be an immunostimulatory nucleic acid, such as a nucleic acid including a CpG motif, or a biological adjuvant (see above).

Controlled release parenteral formulations can be made as implants, oily
5 injections, or as particulate systems. For a broad overview of protein delivery systems, see Banga, *Therapeutic Peptides and Proteins: Formulation, Processing, and Delivery Systems*, Technomic Publishing Company, Inc., Lancaster, PA, 1995. Particulate systems include microspheres, microparticles, microcapsules, nanocapsules, nanospheres, and nanoparticles. Microcapsules contain the therapeutic
10 protein as a central core. In microspheres, the therapeutic agent is dispersed throughout the particle. Particles, microspheres, and microcapsules smaller than about 1 μm are generally referred to as nanoparticles, nanospheres, and nanocapsules, respectively. Capillaries have a diameter of approximately 5 μm so that only nanoparticles are administered intravenously. Microparticles are typically
15 around 100 μm in diameter and are administered subcutaneously or intramuscularly (see Kreuter, *Colloidal Drug Delivery Systems*, J. Kreuter, ed., Marcel Dekker, Inc., New York, NY, pp. 219-342, 1994; Tice & Tabibi, *Treatise on Controlled Drug Delivery*, A. Kydonieus, ed., Marcel Dekker, Inc. New York, NY, pp. 315-339, 1992).

20 Polymers can be used for ion-controlled release. Various degradable and nondegradable polymeric matrices for use in controlled drug delivery are known in the art (Langer, *Accounts Chem. Res.* 26:537, 1993). For example, the block copolymer, polaxamer 407 exists as a viscous yet mobile liquid at low temperatures but forms a semisolid gel at body temperature. It has shown to be an effective
25 vehicle for formulation and sustained delivery of recombinant interleukin-2 and urease (Johnston *et al.*, *Pharm. Res.* 9:425, 1992; and Pec, *J. Parent. Sci. Tech.* 44(2):58, 1990). Alternatively, hydroxyapatite has been used as a microcarrier for controlled release of proteins (Ijntema *et al.*, *Int. J. Pharm.* 112:215, 1994). In yet another aspect, liposomes are used for controlled release as well as drug targeting of
30 the lipid-capsulated drug (Betageri *et al.*, *Liposome Drug Delivery Systems*, Technomic Publishing Co., Inc., Lancaster, PA, 1993). Numerous additional

systems for controlled delivery of therapeutic proteins are known (*e.g.*, U.S. Patent No. 5,055,303; U.S. Patent No. 5,188,837; U.S. Patent No. 4,235,871; U.S. Patent No. 4,501,728; U.S. Patent No. 4,837,028; U.S. Patent No. 4,957,735; and U.S. Patent No. 5,019,369; U.S. Patent No. 5,055,303; U.S. Patent No. 5,514,670; U.S. Patent No. 5,413,797; U.S. Patent No. 5,268,164; U.S. Patent No. 5,004,697; U.S. Patent No. 4,902,505; U.S. Patent No. 5,506,206; U.S. Patent No. 5,271,961; U.S. Patent No. 5,254,342; and U.S. Patent No. 5,534,496).

In another embodiment, a pharmaceutical composition includes a nucleic acid encoding a disclosed immunogen. A therapeutically effective amount of the nucleic acid can be administered to a subject in order to generate an immune response. In one specific, non-limiting example, a therapeutically effective amount of a nucleic acid encoding a disclosed gpl20, gpl40 antigen, or immunogenic fragment thereof is administered to a subject to treat or prevent or inhibit HIV infection.

Optionally, one or more cytokines, such as IL-2, IL-6, IL-12, RANTES, GM-CSF, TNF- α , or IFN- γ , one or more growth factors, such as GM-CSF or G-CSF, one or more costimulatory molecules, such as ICAM-1, LFA-3, CD72, B7-1, B7-2, or other B7 related molecules; one or more molecules such as OX-40L or 41 BBL, or combinations of these molecules, can be used as biological adjuvants (see, for example, Salgaller *et al.*, 1998, *J. Surg. Oncol.* 68(2): 122-38; Lotze *et al.*, 2000, *Cancer J Sci. Am.* 6(Suppl 1):S61-6; Cao *et al.*, 1998, *Stem Cells* 16(Suppl 1):251-60; Kuiper *et al.*, 2000, *Adv. Exp. Med. Biol.* 465:381-90). These molecules can be administered systemically to the host. It should be noted that these molecules can be co-administered via insertion of a nucleic acid encoding the molecules into a vector, for example, a recombinant pox vector (see, for example, U.S. Pat. No. 6,045,802). In various embodiments, the nucleic acid encoding the biological adjuvant can be cloned into same vector as the disclosed antigen coding sequence, or the nucleic acid can be cloned into one or more separate vectors for co-administration. In addition, nonspecific immunomodulating factors such as Bacillus Calmette-Guerin (BCG) and levamisole can be co-administered.

One approach to administration of nucleic acids is direct immunization with plasmid DNA, such as with a mammalian expression plasmid. As described above, the nucleotide sequence encoding the disclosed antigen can be placed under the control of a promoter to increase expression of the molecule.

5 Immunization by nucleic acid constructs is well known in the art and taught, for example, in U.S. Patent No. 5,643,578 (which describes methods of immunizing vertebrates by introducing DNA encoding a desired antigen to elicit a cell-mediated or a humoral response), and U.S. Patent No. 5,593,972 and U.S. Patent No. 5,817,637 (which describe operably linking a nucleic acid sequence encoding an
10 antigen to regulatory sequences enabling expression). U.S. Patent No. 5,880,103 describes several methods of delivery of nucleic acids encoding immunogenic peptides or other antigens to an organism. The methods include liposomal delivery of the nucleic acids (or of the synthetic peptides themselves), and immune-stimulating constructs, or ISCOMS™, negatively charged cage-like structures of 30-
15 40 nm in size formed spontaneously on mixing cholesterol and Quil A™ (saponin). Protective immunity has been generated in a variety of experimental models of infection, including toxoplasmosis and Epstein-Barr virus-induced tumors, using ISCOMS™ as the delivery vehicle for antigens (Mowat and Donachie, *Immunol. Today* 12:383, 1991). Doses of antigen as low as 1 µg encapsulated in ISCOMS™
20 have been found to produce Class I mediated CTL responses (Takahashi *et al*, *Nature* 344:873, 1990).

In another approach to using nucleic acids for immunization, a disclosed antigen can also be expressed by attenuated viral hosts or vectors or bacterial vectors. Recombinant vaccinia virus, adeno-associated virus (AAV), herpes virus,
25 retrovirus, cytomegalo virus or other viral vectors can be used to express the peptide or protein, thereby eliciting a CTL response. For example, vaccinia vectors and methods useful in immunization protocols are described in U.S. Patent No. 4,722,848. BCG (Bacillus Calmette Guerin) provides another vector for expression of the peptides (see Stover, *Nature* 351:456-460, 1991).

30 Simultaneous production of an immunostimulatory molecule and the disclosed antigen enhances the generation of specific effectors. Without being bound

by theory, dependent upon the specific immunostimulatory molecules, different mechanisms might be responsible for the enhanced immunogenicity: augmentation of help signal (IL-2), recruitment of professional APC (GM-CSF), increase in CTL frequency (IL-2), effect on antigen processing pathway and MHC expression (IFN γ and TNF α) and the like. For example, IL-2, IL-6, interferon, tumor necrosis factor, or a nucleic acid encoding these molecules, can be administered in conjunction with a disclosed antigen, or a nucleic acid encoding a disclosed antigen. The co-expression of a disclosed antigen together with at least one immunostimulatory molecule can be effective in an animal model to show anti-pathogen effects.

10 In one embodiment, a nucleic acid encoding a disclosed antigen is introduced directly into cells. For example, the nucleic acid can be loaded onto gold microspheres by standard methods and introduced into the skin by a device such as Bio-Rad's HELIOS™ Gene Gun. The nucleic acids can be "naked," consisting of plasmids under control of a strong promoter. Typically, the DNA is injected into muscle, although it can also be injected directly into other sites, including tissues in proximity to metastases. Dosages for injection are usually around 0.5 μ g/kg to about 50 mg/kg, and typically are about 0.005 mg/kg to about 5 mg/kg (see, *e.g.*, U.S. Patent No. 5,589,466).

20 In one specific, non-limiting example, a pharmaceutical composition for intravenous administration would include about 0.1 μ g to 10 mg of a disclosed antigen per subject per day. Dosages from 0.1 up to about 100 mg per subject per day can be used, particularly if the agent is administered to a secluded site and not into the circulatory or lymph system, such as into a body cavity or into a lumen of an organ. Actual methods for preparing administrable compositions will be known or apparent to those skilled in the art and are described in more detail in such publications as *Remington's Pharmaceutical Sciences*, 19th Ed., Mack Publishing Company, Easton, Pennsylvania, 1995.

25 Single or multiple administrations of the compositions are administered depending on the dosage and frequency as required and tolerated by the subject. In one embodiment, the dosage is administered once as a bolus, but in another embodiment can be applied periodically until a therapeutic result is achieved.

Generally, the dose is sufficient to treat or ameliorate symptoms or signs of disease without producing unacceptable toxicity to the subject. Systemic or local administration can be utilized.

The immunogenic compositions of this disclosure can be employed to
5 generate antibodies that recognize the antigens disclosed herein and the antigen from which the disclosed antigen was derived. The methods include administering to a subject an immunogenic composition including a disclosed antigen or administering to the subject a polynucleotide encoding a disclosed antigen to generate antibodies that recognize the disclosed antigen. The subject employed in this embodiment is
10 one typically employed for antibody production. Mammals, such as, rodents, rabbits, goats, sheep, etc., are preferred.

The antibodies generated can be either polyclonal or monoclonal antibodies. Polyclonal antibodies are raised by injecting (for example subcutaneous or intramuscular injection) antigenic polypeptides into a suitable animal (for example,
15 a mouse or a rabbit). The antibodies are then obtained from blood samples taken from the animal. The techniques used to produce polyclonal antibodies are extensively described in the literature. Polyclonal antibodies produced by the subjects can be further purified, for example, by binding to and elution from a matrix that is bound with the polypeptide against which the antibodies were raised.
20 Those of skill in the art will know of various standard techniques for purification and/or concentration of polyclonal, as well as monoclonal, antibodies. Monoclonal antibodies can also be generated using techniques known in the art.

Any of the disclosed immunogens and nucleic acid molecules encoding such immunogens can be used to elicit an immune response (immunogenic compositions)
25 to gpl20 such as to a gpl20 expressing virus, for example to reduce HIV-1 infection or a symptom of HIV-1 infection. Following administration of a therapeutically effective amount of the disclosed therapeutic compositions, the subject can be monitored for HIV-1 infection, symptoms associated with HIV-1 infection, or both. Disclosed herein are methods of administering the therapeutic molecules disclosed
30 herein (such as gpl20, gpl40 antigens or immunogenic fragments thereof and nucleic acids encoding such antigens) to reduce HIV-1 infection.

Immunogenic compositions can be administered for therapeutic treatments. In therapeutic applications, a therapeutically effective amount of the immunogenic composition is administered to a subject suffering from a disease, such as HIV-1 infection or AIDS.

5 In therapeutic applications, a therapeutically effective amount of the composition is administered to a subject prior to or following exposure to or infection by HIV. When administered prior to exposure, the therapeutic application can be referred to as a prophylactic administration (such as in the form of a vaccine). Single or multiple administrations of the compositions are administered depending
10 on the dosage and frequency as required and tolerated by the subject. In one embodiment, the dosage is administered once as a bolus, but in another embodiment can be applied periodically until a therapeutic result, such as a protective immune response, is achieved. Generally, the dose is sufficient to treat or ameliorate symptoms or signs of disease without producing unacceptable toxicity to the subject.
15 Systemic or local administration can be utilized.

 It may be advantageous to administer the immunogenic compositions disclosed herein with other agents such as proteins, peptides, antibodies, and other antiviral agents, such as anti-HIV agents. Examples of such anti-HIV therapeutic agents include nucleoside reverse transcriptase inhibitors, such as abacavir, AZT,
20 didanosine, emtricitabine, lamivudine, stavudine, tenofovir, zalcitabine, zidovudine, and the like, non-nucleoside reverse transcriptase inhibitors, such as delavirdine, efavirenz, nevirapine, protease inhibitors such as amprenavir, atazanavir, indinavir, lopinavir, nelfinavir, osamprenavir, ritonavir, saquinavir, tipranavir, and the like, and fusion protein inhibitors such as enfuvirtide and the like. In certain
25 embodiments, immunogenic compositions are administered concurrently with other anti-HIV therapeutic agents. In some examples, the disclosed antigens are administered with T-helper cells, such as exogenous T-helper cells. Exemplary methods for the producing and administering T-helper cells can be found in International Patent Publication WO 03/020904, which is incorporated herein by
30 reference.

In certain embodiments, the immunogenic compositions are administered sequentially with other anti-HIV therapeutic agents, such as before or after the other agent. One of ordinary skill in the art would know that sequential administration can mean immediately following or after an appropriate period of time, such as hours,
5 days, weeks, months, or even years later.

The gpl20, gpl40 antigens or immunogenic fragments thereof and nucleic acids encoding these antigens can be used in a multistep immunization regime. In some examples, the regime includes administering to a subject a therapeutically effective amount of a first gpl20, gpl40 antigens or immunogenic fragments thereof
10 as disclosed herein (the prime) and boosting the immunogenic response with one or more additional gpl20, gpl40 antigens or immunogenic fragments thereof after an appropriate period of time. The method of eliciting such an immune reaction is what is known as "prime-boost." In this method, the antibody response to the selected immunogenic surface is focused by giving the subject's immune system a chance to
15 "see" the antigenic surface in multiple contexts. In other words, the use of multiple gpl20, gpl40 antigens or immunogenic fragments thereof with an antigenic surface in common selects for antibodies that bind the antigen's surface in common. Different dosages can be used in a series of sequential inoculations. Thus, a practitioner may administer a relatively large dose in a primary inoculation and then
20 boost with relatively smaller doses of the boost. The immune response against the gpl20, gpl40 antigens or immunogenic fragments thereof can be generated by one or more inoculations of a subject with a disclosed immunogenic composition.

In some examples, the gpl20, gpl40 antigens or immunogenic fragments thereof and nucleic acids encoding these antigens can be administered in "prime-
25 boost" immunization regimes with stabilized gpl40 trimer (see for example Yang *et al. J Virol.* 76(9):4634-42, 2002), and/or stabilized gpl20 polypeptides (such as those described in WO 07/030518). In some examples of this method, gpl20, gpl40 antigens or immunogenic fragments thereof is initially administered to a subject and at periodic times thereafter stabilized gpl40 trimer boosts are administered. In other
30 examples of this method stabilized gpl40 trimer is initially administered to a subject and at periodic times thereafter one or gpl20, gpl40 antigens or immunogenic

fragments thereof are administered. Examples of stabilized gpl40 or gpl20 trimers can be found for example in U.S. Patent No. 6,911,205 which is incorporated herein in its entirety.

One can also use cocktails containing a variety of different HIV strains to
5 prime and boost with trimers from a variety of different HIV strains or with trimers that are a mixture of multiple HIV strains. For example, the first prime could be with a gpl20, gpl40 antigens or immunogenic fragments thereof from one primary HIV isolate, with subsequent boosts different primary isolates.

In one embodiment, a suitable immunization regimen includes at least three
10 separate inoculations with one or more immunogenic compositions of the invention, with a second inoculation being administered more than about two, about three to eight, or about four, weeks following the first inoculation. Generally, the third inoculation is administered several months after the second inoculation, and in specific embodiments, more than about five months after the first inoculation, more
15 than about six months to about two years after the first inoculation, or about eight months to about one year after the first inoculation. Periodic inoculations beyond the third are also desirable to enhance the subject's "immune memory." The adequacy of the vaccination parameters chosen, *e.g.*, formulation, dose, regimen and the like, can be determined by taking aliquots of serum from the subject and assaying antibody
20 titers during the course of the immunization program. Alternatively, the T cell populations can be monitored by conventional methods. In addition, the clinical condition of the subject can be monitored for the desired effect, *e.g.*, prevention of HIV-1 infection or progression to AIDS, improvement in disease state (*e.g.*, reduction in viral load), or reduction in transmission frequency to an uninfected
25 partner. If such monitoring indicates that vaccination is sub-optimal, the subject can be boosted with an additional dose of immunogenic composition, and the vaccination parameters can be modified in a fashion expected to potentiate the immune response. Thus, for example, the dose of the chimeric non-HIV polypeptide or polynucleotide and/or adjuvant can be increased or the route of administration can
30 be changed.

It is contemplated that there can be several boosts, and that each boost can be a different gpl20, gpl40 antigen or immunogenic fragment thereof. It is also contemplated that in some examples that the boost may be the same disclosed antigen as another boost, or the prime.

5 The prime can be administered as a single dose or multiple doses, for example two doses, three doses, four doses, five doses, six doses or more can be administered to a subject over days, weeks or months. The boost can be administered as a single dose or multiple doses, for example two to six doses, or more can be administered to a subject over a day, a week or months. Multiple boosts
10 can also be given, such one to five, or more. Different dosages can be used in a series of sequential inoculations. For example a relatively large dose in a primary inoculation and then a boost with relatively smaller doses. The immune response against the selected antigenic surface can be generated by one or more inoculations of a subject with an immunogenic composition disclosed herein.

15

D. Immunodiagnostic Reagents and Kits

In addition to the therapeutic methods provided above, any of the disclosed antigens disclosed herein can be utilized to produce antigen specific immunodiagnostic reagents, for example, for serosurveillance. Immunodiagnostic
20 reagents can be designed from any of the antigenic polypeptide described herein. For example, in the case of the gpl20 and/or gpl40 antigens, the presence of serum antibodies to HIV is monitored using the isolated gpl20 and/or gpl40 antigens disclosed herein, such as to detect an HIV infection and/or the presence of antibodies that specifically bind the gpl20 and/or gpl40 antigens.

25 Generally, the method includes contacting a sample from a subject, such as, but not limited to a blood, serum, plasma, urine or sputum sample from the subject with one or more of the disclosed antigens disclosed herein (including a polymeric form thereof) and detecting binding of antibodies in the sample to the disclosed antigens. The binding can be detected by any means known to one of skill in the art,
30 including the use of labeled secondary antibodies that specifically bind the

antibodies from the sample. Labels include radiolabels, enzymatic labels, and fluorescent labels.

Any such immunodiagnostic reagents can be provided as components of a kit. Optionally, such a kit includes additional components including packaging,
5 instructions and various other reagents, such as buffers, substrates, antibodies or ligands, such as control antibodies or ligands, and detection reagents

Methods are further provided for a diagnostic assay to monitor HIV-1 induced disease in a subject and/or to monitor the response of the subject to immunization by an HIV vaccine. By "HIV-1 induced disease" is intended any
10 disease caused, directly or indirectly, by HIV. An example of an HIV-1 induced disease is acquired immunodeficiency syndrome (AIDS). The method includes contacting a disclosed gpl20 and/or gpl40 antigen with a sample of bodily fluid from the subject, and detecting binding of antibodies in the sample to the disclosed antigens. In addition, the detection of the HIV-1 binding antibody also allows the
15 response of the subject to immunization by a HIV vaccine to be monitored. In still other embodiments, the titer of the HIV-1 binding antibodies is determined. The binding can be detected by any means known to one of skill in the art, including the use of labeled secondary antibodies that specifically bind the antibodies from the sample. Labels include radiolabels, enzymatic labels, and fluorescent labels. In other
20 embodiments, a disclosed gpl20 and/or gpl40 antigen is used to isolate antibodies present in a subject or biological sample obtained from a subject.

EXAMPLES

Example 1

25 This example describes exemplary procedures for the production of antigens and immunization of animals with the antigens.

In some examples nucleic acid molecules encoding the disclosed antigens are cloned into expression vector CMV/R. Expression vectors are then transfected into 293F cells using 293Fectin (Invitrogen, Carlsbad, CA). Five days after transfection,
30 cell culture supernatant is harvested and concentrated/buffer-exchanged to 500mM NaCl/50 mM Tris pH8.0. The protein initially is purified using HiTrap IMAC HP

Column (GE, Piscataway, NJ), and subsequent gel-filtration using SUPERDEX™ 200 (GE). In some examples the 6x His tag is cleaved off using 3C protease (Novagen, Madison, WI).

For vaccinations with the disclosed antigens 3-4 months old rabbits (NZW)(Covance, Princeton, NJ) are immunized using the Sigma Adjuvant System (Sigma, St. Louis, MO) according to manufacture's protocol. Specifically, three rabbits in each group are vaccinated with 50 µg of protein in 300µl PBS emulsified with 300µl of adjuvant intramuscularly (both legs, 300µl each leg) for example at week 0, 4, 8, 12, 16. Sera are collected for example at week 6 (Post-1), 10 (Post-2), 14 (Post-3), and 18 (Post-4), and subsequently analyzed for their neutralization activities against a panel of HIV-1 strains, and the profile of antibodies that mediate the neutralization.

The disclosed antigens are also tested for antigenic profiling using well-characterized human monoclonal antibodies. For example:

- 1) potent neutralizing CD4 binding site (CD4BS) antibody (IgG1b12, the b12 neutralizing antibody);
- 2) the non-neutralizing CD4 binding site antibodies (IgG1b13 and F105,);
- 3) the non-neutralizing CD4 induced (CD4i) antibodies (17b, 48d, m6); and
- 4) the broad neutralizing antibody 2G12 which binds to glycan motifs on the outer domain, serving as a control for structural integrity of the resurfaced protein.

In some examples, the disclosed antigens are used to coat ELISA plates (4µg/ml in PBS). An amount of the different antibodies is added to the wells, and incubated at room temperature for one hour. The plates are washed six times with PBS+0.05% TWEEN®20, and followed by incubation with HRP conjugated goat anti-human IgG (1:5000) for another hour. The plates are washed again six times with PBS+0.05% TWEEN®20 and developed by adding OPN substrate for 30 minutes at room temperature. The results are evaluated by plotting the readout at OD490 vs. the antibody concentrations.

The antigens are also used to probe for rabbit anti-sera for existence of CD4BS antibodies in the anti-sera.

Example 2

Treatment of HIV in a Subject

This example describes exemplary methods for treating or inhibiting an HIV
5 infection in a subject, such as a human subject by administration of one or more of
the antigens disclosed herein. Although particular methods, dosages and modes of
administrations are provided, one skilled in the art will appreciate that variations can
be made without substantially affecting the treatment.

HIV, such as HIV type 1 (HIV-1) or HIV type 2 (HIV-2), is treated by
10 administering a therapeutically effective amount of a disclosed antigen that induces
an immune response to HIV, for example by inducing an immune response, such as
a neutralizing antibody response to a protein present on the surface of HIV, for
example a gpl20 peptide.

Briefly, the method includes screening subjects to determine if they have
15 HIV, such as HIV-1 or HIV-2. Subjects having HIV are selected for further
treatment. In one example, subjects are selected who have increased levels of HIV
antibodies in their blood, as detected with an enzyme-linked immunosorbent assay,
Western blot, immunofluorescence assay or nucleic acid testing, including viral
RNA or proviral DNA amplification methods. In one example, half of the subjects
20 follow the established protocol for treatment of HIV (such as a highly active
antiretroviral therapy). The other half follow the established protocol for treatment
of HIV (such as treatment with highly active antiretroviral compounds) in
combination with administration of the agents including a therapeutically effective
amount of a disclosed gpl20 antigen that induces an immune response to HIV. In
25 another example, half of the subjects follow the established protocol for treatment of
HIV (such as a highly active antiretroviral therapy). The other subjects receive a
therapeutically effective amount of a disclosed resurfaced gpl20 antigen that
induces an immune response to HIV, such as a neutralizing antibody response.

Screening subjects

In particular examples, the subject is first screened to determine if the subject has HIV. Examples of methods that can be used to screen for HIV include measuring a subject's CD4+ T cell count and the level of HIV in serum blood levels.

5 In some examples, HIV testing consists of initial screening with an enzyme-linked immunosorbent assay (ELISA) to detect antibodies to HIV, such as to HIV-1. Specimens with a nonreactive result from the initial ELISA are considered HIV-negative unless new exposure to an infected partner or partner of unknown HIV status has occurred. Specimens with a reactive ELISA result are retested in
10 duplicate. If the result of either duplicate test is reactive, the specimen is reported as repeatedly reactive and undergoes confirmatory testing with a more specific supplemental test (for example, Western blot or an immunofluorescence assay (IFA)). Specimens that are repeatedly reactive by ELISA and positive by IFA or reactive by Western blot are considered HIV-positive and indicative of HIV
15 infection. Specimens that are repeatedly ELISA-reactive occasionally provide an indeterminate Western blot result, which may be either an incomplete antibody response to HIV in an infected person or nonspecific reactions in an uninfected person. IFA can be used to confirm infection in these ambiguous cases. In some instances, a second specimen will be collected more than a month later and retested
20 for subjects with indeterminate Western blot results. In additional examples, nucleic acid testing (for example, viral RNA or proviral DNA amplification method) can also help diagnosis in certain situations.

 The detection of HIV in a subject's blood is indicative that the subject has HIV and is a candidate for receiving the therapeutic compositions disclosed herein.
25 Moreover, detection of a CD4+ T cell count below 350 per microliter, such as 200 cells per microliter, is also indicative that the subject is likely to have HIV.

 Pre-screening is not required prior to administration of the therapeutic compositions disclosed herein.

Pre-treatment of subjects

In particular examples, the subject is treated prior to diagnosis of AIDS with the administration of a therapeutically effective amount of a disclosed antigen that induces an immune response to HIV. In some examples, the subject is treated with
5 an established protocol for treatment of AIDS (such as a highly active antiretroviral therapy) prior to treatment with the administration of a therapeutic agent that includes one or more of the disclosed antigens that induces an immune response to HIV. However, such pre-treatment is not always required and can be determined by a skilled clinician.

10

Administration of therapeutic compositions

Following selection, a therapeutic effective dose of a therapeutically effective amount of a disclosed antigen that induces an immune response to HIV is administered to the subject (such as an adult human or a newborn infant either at
15 risk for contracting HIV or known to be infected with HIV). Additional agents, such as anti-viral agents, can also be administered to the subject simultaneously or prior to or following administration of the disclosed agents. Administration can be achieved by any method known in the art, such as oral administration, inhalation, intravenous, intramuscular, intraperitoneal or subcutaneous.

20 The amount of the composition administered to prevent, reduce, inhibit, and/or treat HIV or a condition associated with it depends on the subject being treated, the severity of the disorder and the manner of administration of the therapeutic composition. Ideally, a therapeutically effective amount of an agent is the amount sufficient to prevent, reduce, and/or inhibit, and/or treat the condition
25 (for example, HIV) in a subject without causing a substantial cytotoxic effect in the subject. An effective amount can be readily determined by one skilled in the art, for example using routine trials establishing dose response curves. In addition, particular exemplary dosages are provided above. The therapeutic compositions can be administered in a single dose delivery, *via* continuous delivery over an extended
30 time period, in a repeated administration protocol (for example, by a daily, weekly or monthly repeated administration protocol). In one example, a therapeutically

effective amount of a disclosed antigen that induces an immune response to HIV is administered intravenously to a human. As such, these compositions may be formulated with an inert diluent or with a pharmaceutically acceptable carrier. Therapeutic compositions can be taken long term (for example over a period of
5 months or years).

Assessment

Following the administration of one or more therapies, subjects having HIV (for example, HIV-1 or HIV-2) can be monitored for reductions in HIV levels,
10 increases in a subjects CD4+ T cell count or reductions in one or more clinical symptoms associated with HIV infection. In particular examples, subjects are analyzed one or more times, starting 7 days following treatment. Subjects can be monitored using any method known in the art. For example, biological samples from the subject, including blood, can be obtained and alterations in HIV or CD4+ T cell
15 levels evaluated.

Additional treatments

In particular examples, if subjects are stable or have a minor, mixed or partial response to treatment, they can be re-treated after re-evaluation with the same
20 schedule and preparation of agents that they previously received for the desired amount of time, including the duration of a subject's lifetime. A partial response is a reduction, such as at least a 10%, at least 20%, at least 30%, at least 40%, at least 50% or at least 70% reduction of HIV viral load, HIV replication or combination thereof. A partial response may also be an increase in CD4+ T cell count such as at
25 least 350 T cells per microliter.

Example 3

Treatment of Subjects

This example describes methods that can be used to treat a subject that has or
30 is at risk of having an infection from HIV that can be treated by eliciting an immune response, such as a neutralizing antibody response to the pathogen of interest. In

particular examples, the method includes screening a subject having, thought to have or at risk of having a HIV infection. Subjects of an unknown infection status can be examined to determine if they have an infection, for example using serological tests, physical examination, enzyme-linked immunosorbent assay (ELISA), radiological screening or other diagnostic technique known to those of skill in the art. In some examples, subjects are screened to identify a HIV infection, with a serological test, or with a nucleic acid probe specific for a HIV. Subjects found to (or known to) have a HIV infection can be administered a disclosed antigen that can elicit an antibody response to HIV. Subjects may also be selected who are at risk of developing HIV for example, subjects exposed to HIV.

Subjects selected for treatment can be administered a therapeutic amount of the disclosed antigen. The disclosed antigen can be administered at doses of 1 µg/kg body weight to about 1 mg/kg body weight per dose, such as 1 µg/kg body weight - 100 µg/kg body weight per dose, 100 µg/kg body weight - 500 µg/kg body weight per dose, or 500 µg/kg body weight - 1000 µg/kg body weight per dose. However, the particular dose can be determined by a skilled clinician. The disclosed antigen can be administered in one or several doses, for example continuously, daily, weekly, or monthly. When administered sequentially the time separating the administration of the disclosed antigen can be seconds, minutes, hours, days, or even weeks.

The mode of administration can be any used in the art. The amount of agent administered to the subject can be determined by a clinician, and may depend on the particular subject treated. Specific exemplary amounts are provided herein (but the disclosure is not limited to such doses).

25

Example 4

Identification of Immunogenic Fragments of gp120

This example describes the selection of immunogenic fragments of the disclosed antigens.

A nucleic acid molecule encoding a disclosed antigen is expressed in a host using standard techniques (see Sambrook *et al*, Molecular Cloning; A Laboratory

Manual, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. 1989). Preferable gpl20 antigen fragment is expressed such that the antigen can be isolated or purified in sufficient quantity. The antigens that are expressed are analyzed by various techniques known in the art, such as immunoblot, and ELISA, and for binding to
5 known neutralizing antibodies of HIV, for example the bl2 antibody.

To determine the antigenic antigen fragments, subjects such as mice, rabbits or other suitable subjects are immunized with resurfaced gpl20 antigen fragments. Sera from such immunized subjects are tested for antibody activity for example by ELISA with the expressed polypeptide. They are also tested in a CD4 binding assay,
10 for example by qualitative biacore, and the binding of neutralizing antibodies, for example, by using the bl2 antibody. Thus antigenic fragments of gpl20 are selected to achieve broadly reactive neutralizing antibody responses.

Example 5

15 HIV Immunogens

The class of immunogens disclosed herein is aimed at eliciting anti-HIV-1 neutralization antibodies with special emphasis on induction of neutralization antibodies targeting the CD4 binding site (BS), the primary receptor for HIV-1 entry. The outer domain (OD) of HIV-1 envelope gpl20 is an engineered subunit
20 protein containing residues 252 to 482. The V3 and β 20-21 loops are truncated and replaced by short linkers. The OD retains a functional CD4-binding site (CD4BS) which represents a potential target for vaccine design. The same region is also targeted by potent anti-HIV neutralization antibodies VRCOL. The ODs described in this example retain the antigenic area for eliciting potent anti-CD4BS antibody, but
25 with much more selective exposure of this region to the immune system. The ODs described herein contain various modifications in the one or more of the VI, V2, V3, V4 and V5 loops for improved protein conformation and stability and various additions of glycosylation sites for covering exposed immunodominant sites other than CD4BS.

Biosynthesis of outer domain proteins

Genes encoding the outer domain of HIV-1 gp120 were synthesized by PCR-based accurate synthesis (PAS). All genes were ligated with human CD5 secretion signal sequence and cloned into CMV/R expression vector. For protein production, the plasmids were transfected to 293-F cells by 293fectin (Invitrogen) and the supernatants were harvested 4 days later. The OD proteins were purified from the supernatants by cobalt- (TALON, Clontech) or nickel-coupled resin (ProPur IMAC, Nunc), or by covalently cross-linked bl2-sepharose 6 affinity column followed by gel filtration on Superdex 200 16/60 (GE Healthcare).

Enzyme-linked immunosorbent assay (ELISA)

ELISA was conducted with Maxisorp plates (Nunc). The plates were first coated with snowdrop lectin (GNA, Sigma) and blocked with 1% BSA in PBS. After washing the wells, the plates were incubated with the purified proteins (200 ng/well). The plates were then washed and incubated with serial dilutions of monoclonal antibodies. After washing the wells, the bound antibodies were detected by HRP-labeled anti-human IgG (Southern Biotech). The wells were developed by the substrates (OPD, Sigma), stopped by sulfuric acid and the absorbance was read at 490 nm. The curve fits and the EC₅₀ values were calculated by using Prism 5 software (Graphpad Software).

While this disclosure has been described with an emphasis upon particular embodiments, it will be obvious to those of ordinary skill in the art that variations of the particular embodiments may be used and it is intended that the disclosure may be practiced otherwise than as specifically described herein. Features, characteristics, compounds, chemical moieties or examples described in conjunction with a particular aspect, embodiment or example of the invention are to be understood to be applicable to any other aspect, embodiment or example of the invention. Accordingly, this disclosure includes all modifications encompassed within the spirit and scope of the disclosure as defined by the following claims.

We Claim:

1. An isolated immunogen, comprising a human immunodeficiency virus (HIV) gpl20 outer domain polypeptide, wherein the gpl20 outer domain polypeptide, comprises amino acids 252-482 of gpl20, and wherein the outer domain comprises the amino acid sequence according to SEQ ID NO: 203, SEQ ID NO: 1 or SEQ ID NO: 74 and wherein the immunogen optionally further comprises one or more mutations selected from the following groups, a-h:
 - a: wherein the $\beta 20/\beta 21$ bridging sheet of the gpl20 outer domain is removed or truncated by replacing the amino acid residues between residue number 423 and residue number 435 in gpl20 with Gly-Gly or by replacing amino acid residues 421 through 436 of gpl20 with Gly-Gly;
 - b: wherein the VI and V2 loops of the gpl20 outer domain are removed or truncated by replacing amino acid residues 128 through 194 of gpl20 with GAG or by replacing the amino acid residues between residue number C119 and residue number C205 in gpl20 with VKLTPLAGATSVITQA (SEQ ID NO: 162);
 - c: wherein the VI loop of the gpl20 outer domain is removed or truncated by replacing amino acid residues 131-157 with the amino acid sequence CTDLRGSGGGSGGGSEIKNC (SEQ ID NO: 164);
 - d: wherein the V2 loop of the gp120 outer domain is removed or truncated by replacing amino acid residues 157-198 in gpl20 with the amino acid sequence CSFNITTSIRDKVQKEYALFYKLDVVPIDNDNTSYRLINC (SEQ ID NO: 165), the amino acid sequence CSFNITTSIRDKVQKEYALFYKLDVVPISGGSGGSYRLINC (SEQ ID NO: 166), or the amino acid sequence CSFNITTSIRDKVQKEYALFYKLDVVPISGGSGGSYRLINC (SEQ ID NO: 167);
 - e: wherein the V3 loop of the gpl20 outer domain is removed or truncated by replacing amino acid residues 302-323 of gpl20 with the amino acid sequence NTRGRR (SEQ ID NO: 169), by replacing amino acid residues 302-323 of gpl20, with the amino acid sequence GRR, by replacing amino acid residues 296-

- 331 of gpl20 with the amino acid sequence CSRPNNTNRGRRGSSGGS HC (SEQ ID NO: 170), by replacing amino acid residues 296-331 of gpl20 with the amino acid sequence CSRPNNGGSGSGGSSGGS HC (SEQ ID NO: 171), by replacing amino acid residues 296-331 of gpl20 with the amino acid sequence
- 5 CARPSNNTRGRRGDIRQAYC (SEQ ID NO: 172), by replacing amino acid residues 296-331 of gpl20 with the amino acid sequence CARPSNNTDIRQAYC (SEQ ID NO: 173), by replacing amino acid residues 296-331 of gpl20 with the amino acid sequence CARPSNNTRQAYC (SEQ ID NO: 174), by replacing amino acid residues 296-331 of gpl20 with the amino acid sequence CARPSNNTQYC
- 10 (SEQ ID NO: 175), by replacing amino acid residues 296-331 of gpl20 with the amino acid sequence CARGSGSGSYC (SEQ ID NO: 176), by replacing amino acid residues 296-331 of gpl20 with the amino acid sequence CSRPNNTNRGRRGDIRQAHC (SEQ ID NO: 177), by replacing amino acid residues 296-331 of gpl20 with the amino acid sequence CSRPNNTNRQAHC
- 15 (SEQ ID NO: 178), by replacing amino acid residues 296-331 of gpl20 with the amino acid sequence CSRPNNGGSGQAHC (SEQ ID NO: 179), by replacing amino acid residues 296-331 of gpl20 with the amino acid sequence CTRPNNGGSGSGGSSGGS HC (SEQ ID NO: 180), or by replacing amino acid residues 296-331 of gpl20 with the amino acid sequence
- 20 CTRPNNTNRGRRGSSGGS HC (SEQ ID NO: 181);
- f: wherein the V4 loop of the gpl20 outer domain is removed or truncated by replacing the V4 loop with the amino acid sequence SrWNNGGGSGGGSGGGSDTIT (SEQ ID NO: 182), the amino acid sequence STWFNGSGSGSGGTIT (SEQ ID NO: 184), the amino acid sequence
- 25 STWFNSTWSTKGSNNTGSDTIT (SEQ ID NO: 185), the amino acid sequence STWFQGS GSGSGGTIT (SEQ ID NO: 186), by at least partially replacing the V4 loop with Gly-Ser, by truncating the V4 loop by removing 9 amino acids from the V4 loop, or replacing the V4 loop with a V4 loop from HIV strain Ker2018;
- g: wherein the V5 loop of the gpl20 outer domain is removed or
- 30 truncated by replacing the V5 loop with the amino acid sequence GGNTNR TC (SEQ ID NO: 187), the amino acid sequence GGSGSGTC (SEQ ID NO: 188), the

amino acid sequence GGSGSTC (SEQ ID NO: 189), the amino acid sequence SGGSGQETFR (SEQ ID NO: 192), the amino acid sequence GGGSGSGEI (SEQ ID NO: 194), by truncating the V5 loop to NDSDGNETFR (SEQ ID NO: 190) or by replacing the V5 loop with an 8 amino acid Gly-Ser repeat; and

5 h: wherein the immunogen comprises one or more amino acid substitutions corresponding to: 273N, T283N, T339N, A341T, 360N, 362N, N363Q, P369N, 137 IT, 37 IN, T373N, T373G, 377N, A379T, F383T, N386Q, D392N, R419N, G421N, G424N, A431T, P437T, N465Q, W479N, N280C and G458, K358C and N465C, V255C and M475C, V257N, V272N, N276Q, N276D, N276E,
10 T283V, A297T, D368R, N362T, E363N, P364S, T373N, S375T, F382T, S388A, A388S, K389D, E398N, N406Q, N410Q, K421T, G422V, V427N, V442N, R444T, N478L, S481T, E482S, V257N, V272N, N276D, A388S, K421S, R421N, I423T, T424N, M426T, I434N, A436T, R487N, V489T, and/or R490N.

15 2. The isolated immunogen of claim 1, wherein the gpl20 outer domain polypeptide comprises or consists of amino acids 213-492 of gpl20, amino acids 213-482 of gpl20, amino acids 252-492 of gpl20, or 252-482 of gpl20.

20 3. The isolated immunogen of any one of claims 1 or 2, wherein the immunogen comprises or consists of amino acids 213-492 of gpl20, amino acids 213-482 of gpl20, amino acids 252-492 of gpl20, or 252-482 of gpl20.

 4. The isolated immunogen of any one of claims 1 or 2, wherein the immunogen comprises or consists of HIV-1 gpl40.

25

 5. The isolated immunogen of any one of claims 1 or 2, wherein the immunogen comprises or consists of HIV-1 gpl20.

30 6. The isolated immunogen of any one of claims 4 or 5, wherein the immunogen comprises one or more amino acid substitutions corresponding to:

T26N, K55N, A49S, D91T, T92N, D94T, D98N, M100T V103N, K113N, P118N, T207N, R209T.

7. The isolated immunogen of any one of claims 1-6, wherein the
 5 β 20/ β 21 bridging sheet of the gpl20 outer domain is removed or truncated by replacing the amino acid residues between residue number 423 and residue number 435 in gpl20 with Gly-Gly or by replacing amino acid residues 421 through 436 of gpl20 with Gly-Gly.

10 8. The isolated immunogen of any one of claims 1-7, wherein the VI and V2 loops of the gpl20 outer domain are removed or truncated by replacing amino acid residues 128 through 194 of gpl20 with GAG or by replacing the amino acid residues between residue number C119 and residue number C205 in gpl20 with VKLTPLAGATSVITQA (SEQ ID NO: 162).

15 9. The isolated immunogen of any one of claims 1-7, wherein the wherein the VI loop of the gpl20 outer domain is removed or truncated by replacing amino acid residues 131-157 with the amino acid sequence CTDLRGSGGGSGGGSEIKNC (SEQ ID NO: 164).

20 10. The isolated immunogen of any one of claims 1-7 or 9, wherein the V2 loop of the gpl20 outer domain is removed or truncated by replacing amino acid residues 157-198 in gpl20 with the amino acid sequence CSFNITTSIRDKVQKEYALFYKLDVVPIDNDNTSYRLINC (SEQ ID NO: 165),
 25 the amino acid sequence CSFNITTSIRDKVQKEYALFYKLDVVPIGGSGGSYRLINC (SEQ ID NO: 166), or the amino acid sequence CSFNITTSIRDKVQKEYALFYKLDVVPIGSGGGSGSYRLINC (SEQ ID NO: 167).

30

11. The isolated immunogen of any one of claims 1-10, wherein the V3 loop of the gpl20 outer domain is removed or truncated by replacing amino acid residues 302-323 of gpl20 with the amino acid sequence NTRGRR (SEQ ID NO: 169), by replacing amino acid residues 302-323 of gpl20, with the amino acid sequence GRR, by replacing amino acid residues 296-331 of gpl20 with the amino acid sequence CSRPNNNTRGRRGSSGGSHC (SEQ ID NO: 170), by replacing amino acid residues 296-331 of gpl20 with the amino acid sequence CSRPNNGGSGSGGSSGGSHC (SEQ ID NO: 171), by replacing amino acid residues 296-331 of gpl20 with the amino acid sequence CARPSNNTRGRRGDIRQAYC (SEQ ID NO: 172), by replacing amino acid residues 296-331 of gpl20 with the amino acid sequence CARPSNNTDIRQAYC (SEQ ID NO: 173), by replacing amino acid residues 296-331 of gpl20 with the amino acid sequence CARPSNNTRQAYC (SEQ ID NO: 174), by replacing amino acid residues 296-331 of gpl20 with the amino acid sequence CARPSNNTQYC (SEQ ID NO: 175), by replacing amino acid residues 296-331 of gpl20 with the amino acid sequence CARGSGSGSYC (SEQ ID NO: 176), by replacing amino acid residues 296-331 of gpl20 with the amino acid sequence CSRPNNNTRGRRGDIRQAHC (SEQ ID NO: 177), by replacing amino acid residues 296-331 of gpl20 with the amino acid sequence CSRPNNNTRRQAHC (SEQ ID NO: 178), by replacing amino acid residues 296-331 of gpl20 with the amino acid sequence CSRPNNGGSGQAHC (SEQ ID NO: 179), by replacing amino acid residues 296-331 of gpl20 with the amino acid sequence CTRPNNGGSGSGGSSGGSHC (SEQ ID NO: 180), or by replacing amino acid residues 296-331 of gpl20 with the amino acid sequence CTRPNNNTRGRRGSSGGSHC (SEQ ID NO: 181).

12. The isolated immunogen of any one of claims 1-11, wherein the V4 loop of the gpl20 outer domain is removed or truncated by replacing the V4 loop with the amino acid sequence SIWNNGGGSGGGSGGGSDTIT (SEQ ID NO: 182), the amino acid sequence STWFNGSGSGGSGTIT (SEQ ID NO: 184), the amino acid sequence STWFNSTWSTKGSNNTEGSDTIT (SEQ ID NO: 185), the amino

acid sequence STWFQSGSGSGSGTIT (SEQ ID NO: 186), by at least partially replacing the V4 loop with Gly-Ser, by truncating the V4 loop by removing 9 amino acids from the V4 loop, or replacing the V4 loop with a V4 loop from HIV strain Ker2018.

5

13. The isolated immunogen of any one of claims 1-12, wherein the V5 loop of the gp120 outer domain is removed or truncated by replacing the V5 loop with the amino acid sequence GGNTNRTC (SEQ ID NO: 187), the amino acid sequence GGS GSGTC (SEQ ID NO: 188), the amino acid sequence GGS GSTC (SEQ ID NO: 189), the amino acid sequence SGGSGQETFR (SEQ ID NO: 192), the amino acid sequence GGGSGSGEI (SEQ ID NO: 194), by truncating the V5 loop to NDS DGNETFR (SEQ ID NO: 190) or by replacing the V5 loop with a 8 amino acid Gly-Ser repeat.

14. The isolated immunogen of any one of claims 1-13, wherein the immunogen comprises one or more amino acid substitutions corresponding to: 273N, T283N, T339N, A341T, 360N, 362N, N363Q, P369N, 137 IT, 37 IN, T373N, T373G, 377N, A379T, F383T, N386Q, D392N, R419N, G421N, G424N, A431T, P437T, N465Q, W479N, N280C and G458, K358C and N465C, V255C and M475C, V257N, V272N, N276Q, N276D, N276E, T283V, A297T, D368R, N362T, E363N, P364S, T373N, S375T, F382T, S388A, A388S, K389D, E398N, N406Q, N410Q, K421T, G422V, V427N, V442N, R444T, N478L, S481T, E482S, T26N, K55N, A49S, D91T, T92N, D94T, D98N, M100T V103N, K113N, P118N, T207N, R209T, V257N, V272N, N276D, A388S, K421S, R421N, I423T, T424N, M426T, I434N, A436T, R487N, V489T, and/or R490N.

15. The isolated immunogen of any one of claims 1-14, wherein the immunogen comprises one or more of a foldon domain, a ferritin polypeptide, a hybrid of different ferritin polypeptides, a six-histidine residue tag, a secretion signal sequence and a transmembrane domain .

16. The isolated immunogen of claim 15, wherein the immunogen comprises a T4 fibrin foldon comprising the amino acid sequence GYIPEAPRDGQAYVRKDGEWVLLSTF (SEQ ID NO: 195).

5 17. The isolated immunogen of any one of claims 15 or 16, wherein the immunogen comprises a gp41 transmembrane domain, a CD4 transmembrane domain or an influenza transmembrane domain

10 18. The isolated immunogen of claim 17, wherein the CD4 transmembrane domain comprises the amino acid sequence LIVLGGVAGLLLFGLGI (SEQ ID NO: 196).

15 19. The isolated immunogen of claim 17, wherein the influenza transmembrane domain comprises a neuraminidase (NA) transmembrane domain comprising the amino acid sequence MNPNQKIITIGSICMVVGIIISLILQIGNIISIWVS (SEQ ID NO: 198) or a hemagglutinin (HA) transmembrane domain comprising the amino acid sequence ILAIYSTVASSLVLLVSLGAISF (SEQ ID NO: 197).

20 20. The isolated immunogen of any one of claims 15-19, wherein secretion signal sequence is a CD5 secretion signal sequence comprising the amino acid sequence MPMGSLQPLATLYLLGMLVASVLA (SEQ ID NO: 199) or a IL-2 secretion signal sequence comprising the amino acid sequence MQLASCVTLTLVLLVNSAP (SEQ ID NO: 200).

25 21. The isolated immunogen of any one of claims 1-20, wherein the immunogen comprises an amino acid sequence at least 95% identical to the amino acid sequence set forth as any one of SEQ ID NOs: 203, 58-77, 1-57, 79, or 80 148, or 149.

30

22. A virus-like particle (VLP) comprising the isolated immunogen of any one of claims 1-21.

23. The VLP of claim 22, wherein the VLP comprises a Chikungunya virus peptide covalently linked to the immunogen.

24. An isolated nucleic acid molecule encoding the antigen of any one of claims 1-21 or a fusion protein comprising a Chikungunya virus peptide covalently linked to the immunogen of any one of claims 22 and 23.

25. The isolated nucleic acid molecule of claim 24, wherein the nucleotide sequence comprises the nucleic acid sequence set forth as one of SEQ ID NOs: 81-147, 150-161 and 206-215.

26. The isolated nucleic acid molecule of any one of claims 24 or 25 operably linked to a promoter.

27. A vector comprising the nucleic acid molecule of claim 26.

28. An isolated host cell comprising the vector of claim 27.

29. A pharmaceutical composition comprising an effective amount of the isolated immunogen of any one of claims 1-21, an isolated nucleic acid molecule encoding the isolated immunogen of any one of claims 1-21, or a VLP comprising the isolated immunogen of any one of claims 22 or 23, and a pharmaceutically acceptable carrier.

30. A method for generating an immune response in a subject, comprising administering to the subject an effective amount of the pharmaceutical composition of claim 29, thereby generating the immune response.

31. A method for treating or preventing a human immunodeficiency type 1 (HIV-1) infection in a subject, comprising administering to the subject a therapeutically effective amount of a first isolated immunogen, wherein the first isolated immunogen comprises the isolated immunogen of any one of claims 1-21,
5 an isolated nucleic acid molecule encoding the isolated immunogen, or a VLP comprising the first isolated immunogen, and wherein the first isolated immunogen comprises a target epitope, thereby treating the subject or preventing infection of the subject.

10 32. The method of claim 31, further comprising administering a therapeutically effective amount of at least one additional isolated immunogen of any one of claims 1-20, an isolated nucleic acid molecule encoding the isolated immunogen, or a VLP comprising the first isolated immunogen, wherein a target epitope of the additional immunogen is identical to the first immunogen, and
15 wherein surface-exposed amino acid residues located exterior to the target epitope of the first immunogen are not identical to surface-exposed amino acid residues located exterior to the target epitope on the additional isolated immunogen.

33. The method of any one of claims 31-32, further comprising
20 administering a therapeutically effective amount of a polypeptide comprising:
a) a monomeric or trimeric gpl40 polypeptide;
b) an monomeric or trimeric wild-type gpl20 polypeptide;
c) a wild-type outer domain gpl20 polypeptide;
d) a nucleic acid molecule expressing the polypeptide of a-c; or
25 e) any combination of a-d, above.

34. The method of any one of claims 31-33, further comprising administering to the subject a therapeutically effective amount of an anti-viral agent.

30 35. A method for detecting or isolating an HIV-1 binding antibody in a subject infected with HIV-1, comprising:

- providing the immunogen of any one of claims 1-21;
 - contacting the immunogenic composition with an amount of bodily fluid from the subject; and
 - detecting binding of the HIV-1 binding antibody to the immunogen, thereby
- 5 detecting or isolating the HIV-1 binding antibody in a subject.

FIG. 1A
Development of Fourth Generation Ods (I)

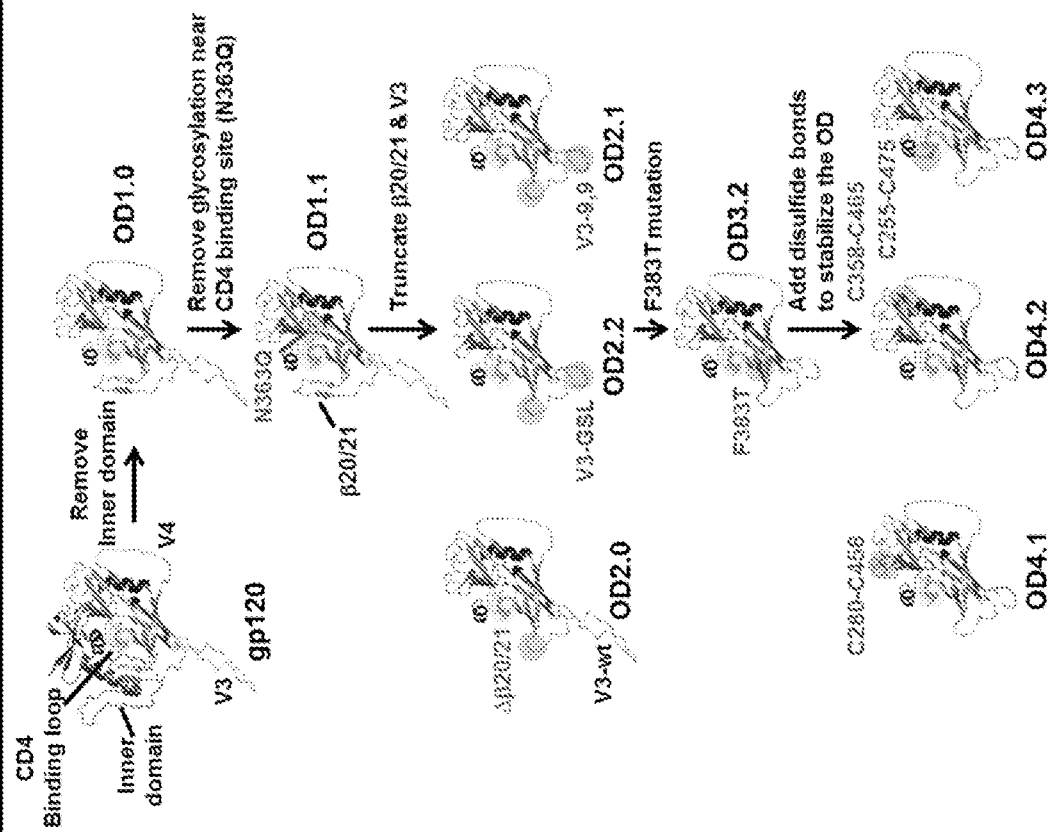
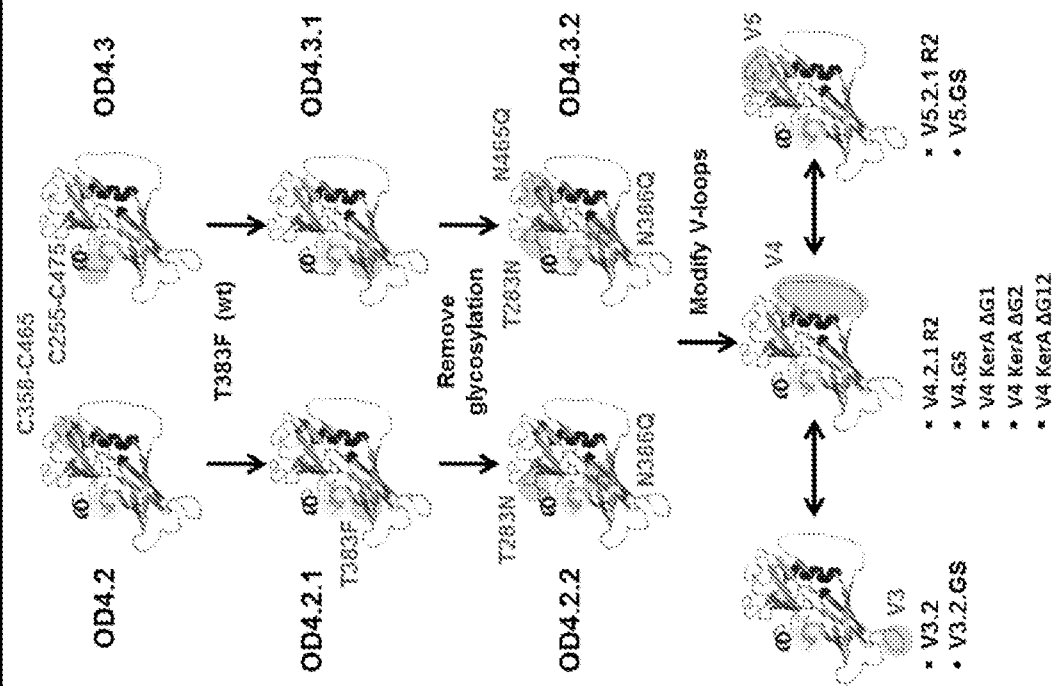


FIG. 1B
Development of Fourth Generation Ods (II)

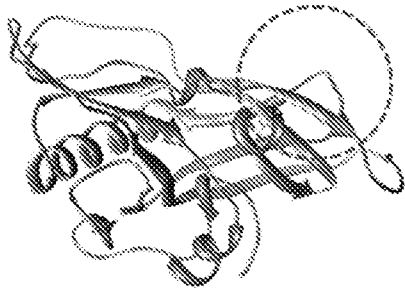
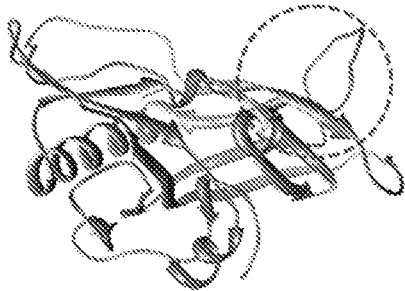


V3 loop modification

FIG. 2

OD4.2F (V3-GSL)

v3.3

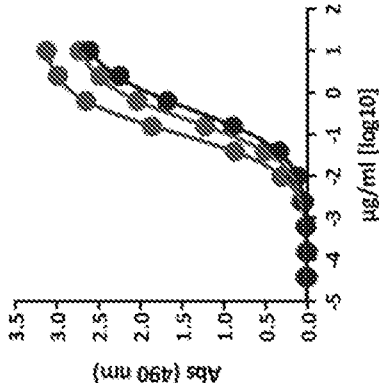


V3 loop sequence

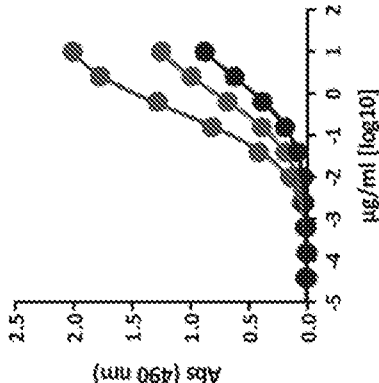
	ψ
GSL:	CARPSNNTGRRGDIRQAYC
v3.1:	CARPSNNT-----DIRQAYC
v3.2:	CARPSNNT-----RQAYC
v3.3:	CARPSNNT-----Q-YC
v3.4:	CAR-----GSGSGS---YC

● OD4.2F
● v3.3
● v3.4

VRC01



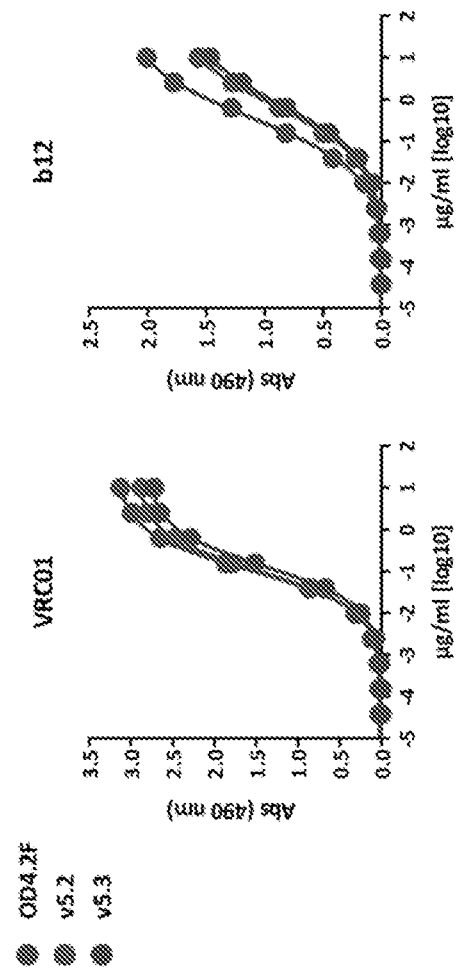
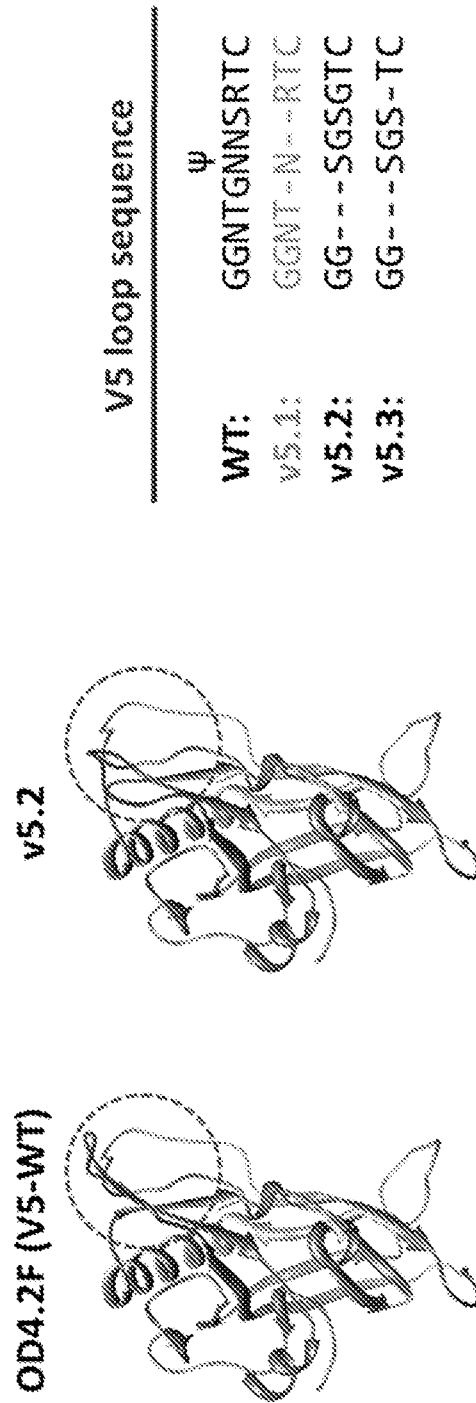
b12



	EC50: VRC01	EC50: b12
OD4.2F	0.1068	0.3592
v3.3	0.2095	0.9199
v3.4	0.3846	2.0680

V5 loop modification

FIG. 3



	EC50: VRC01	EC50: b12
OD4.2F	0.1068	0.3592
v5.2	0.1004	0.7041
v5.3	0.1299	0.7129

Updated OD4.2'

FIG. 4

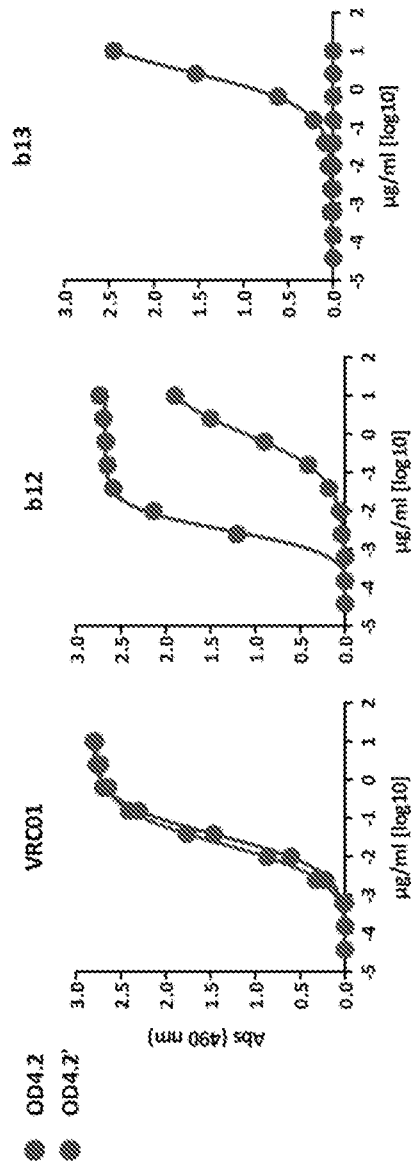
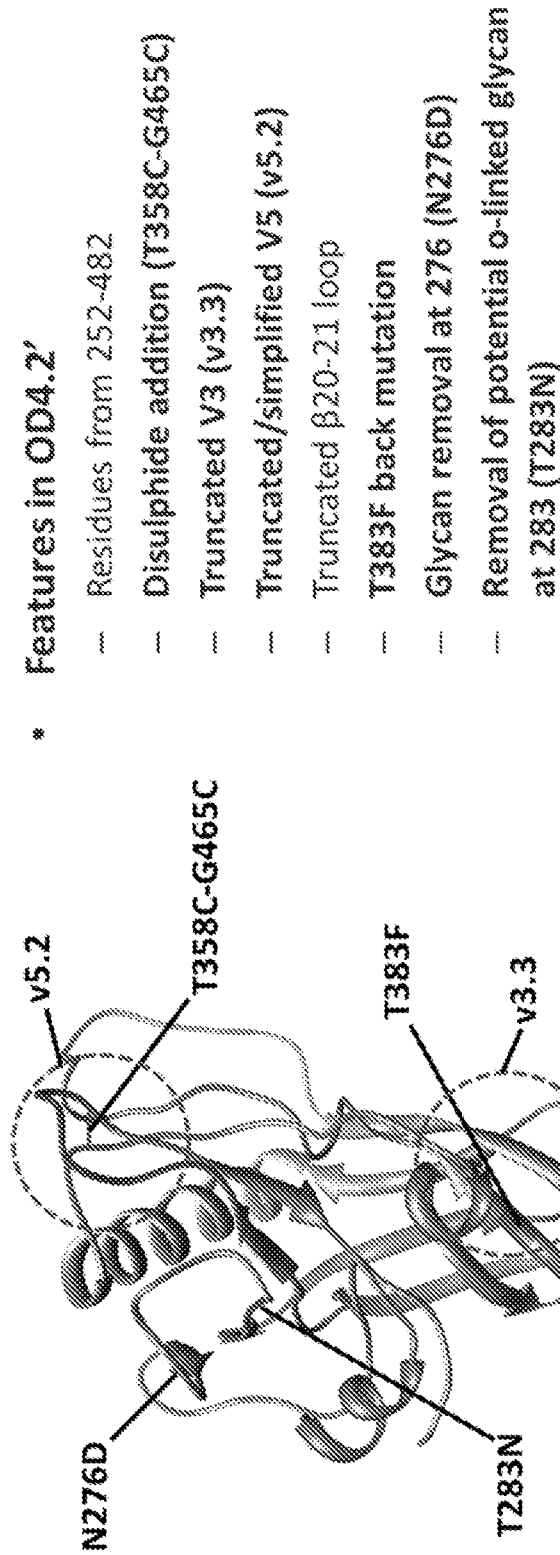
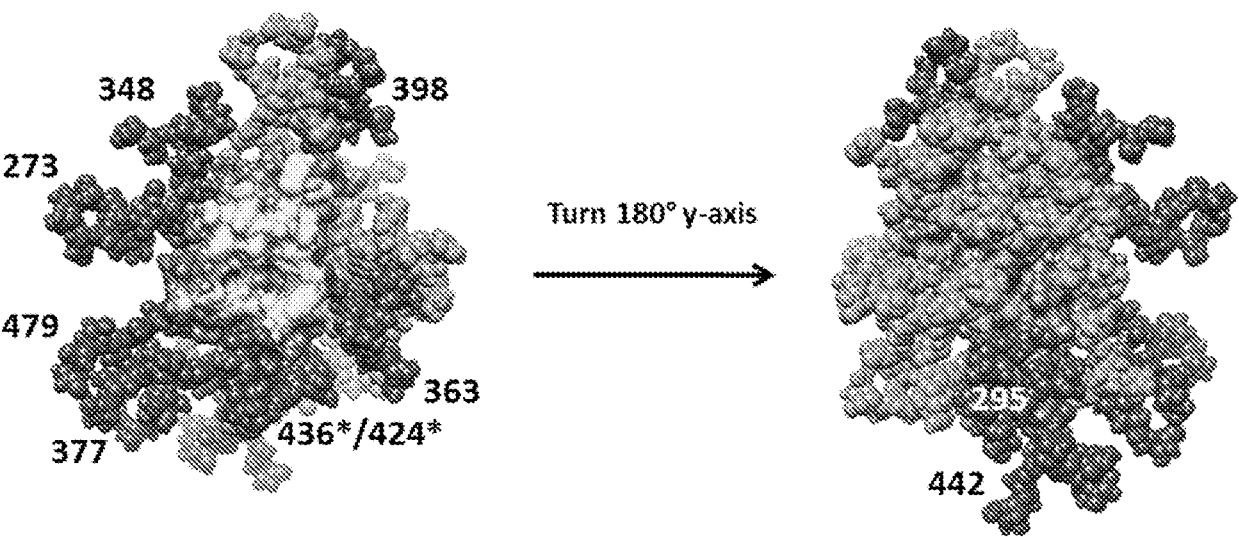


FIG. 5

Strategy for glycan masking

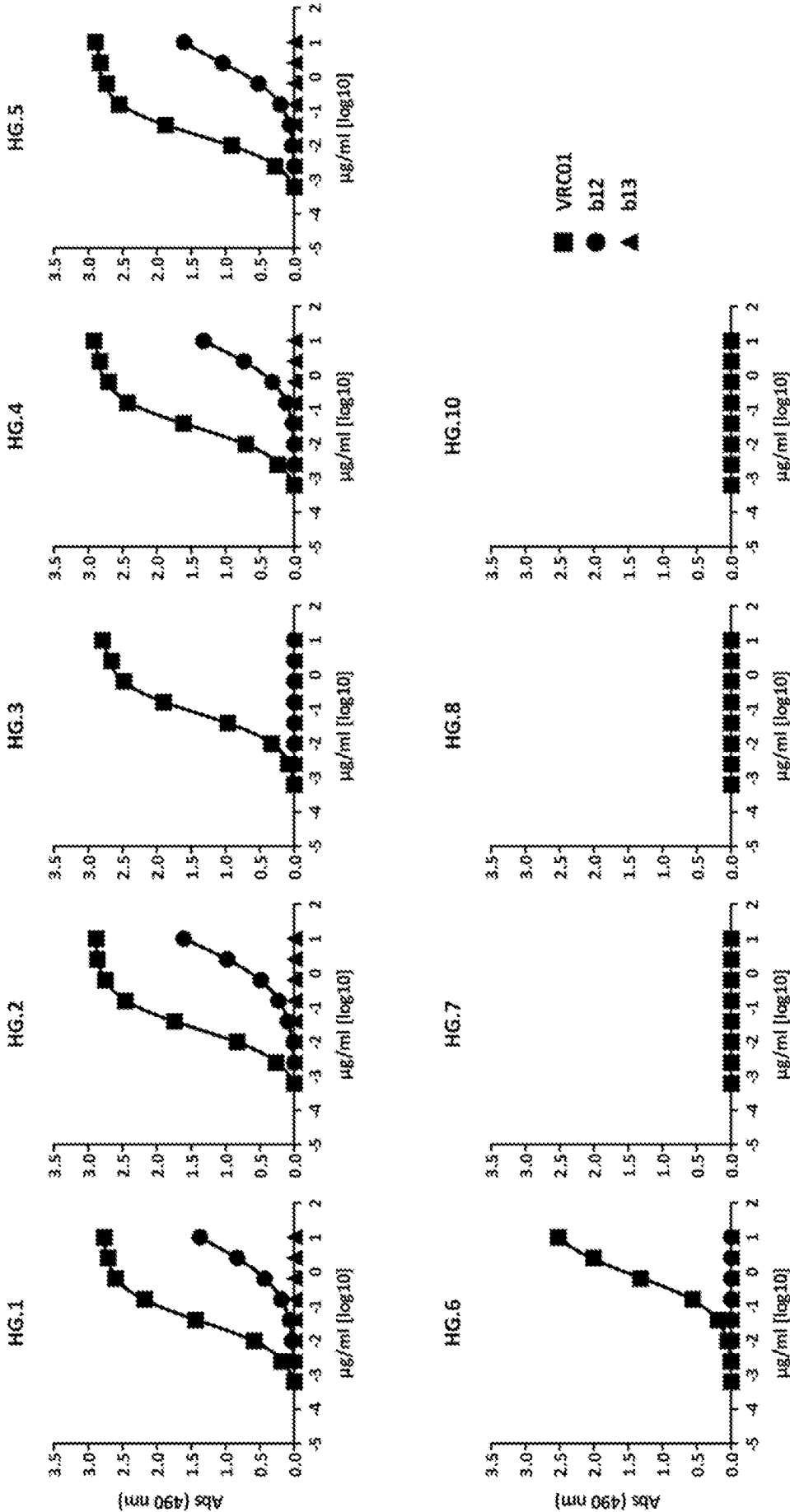


	@295	@442	@479	@273	@377	@436 *	@424 *	@398	@348	@363
HG.1	X	X								
HG.2	X	X	X							
HG.3	X	X				X				
HG.4	X	X					X			
HG.5	X	X						X		
HG.6	X	X								X
HG.7	X	X		X	X				X	
HG.8	X	X	X	X	X	X		X	X	X
HG.9	X	X	X	X	X		X	X	X	X
* HG.10	X	X	X	X	X				X	

* Residue numbering was altered due to the deletion of β 20-21 loop.

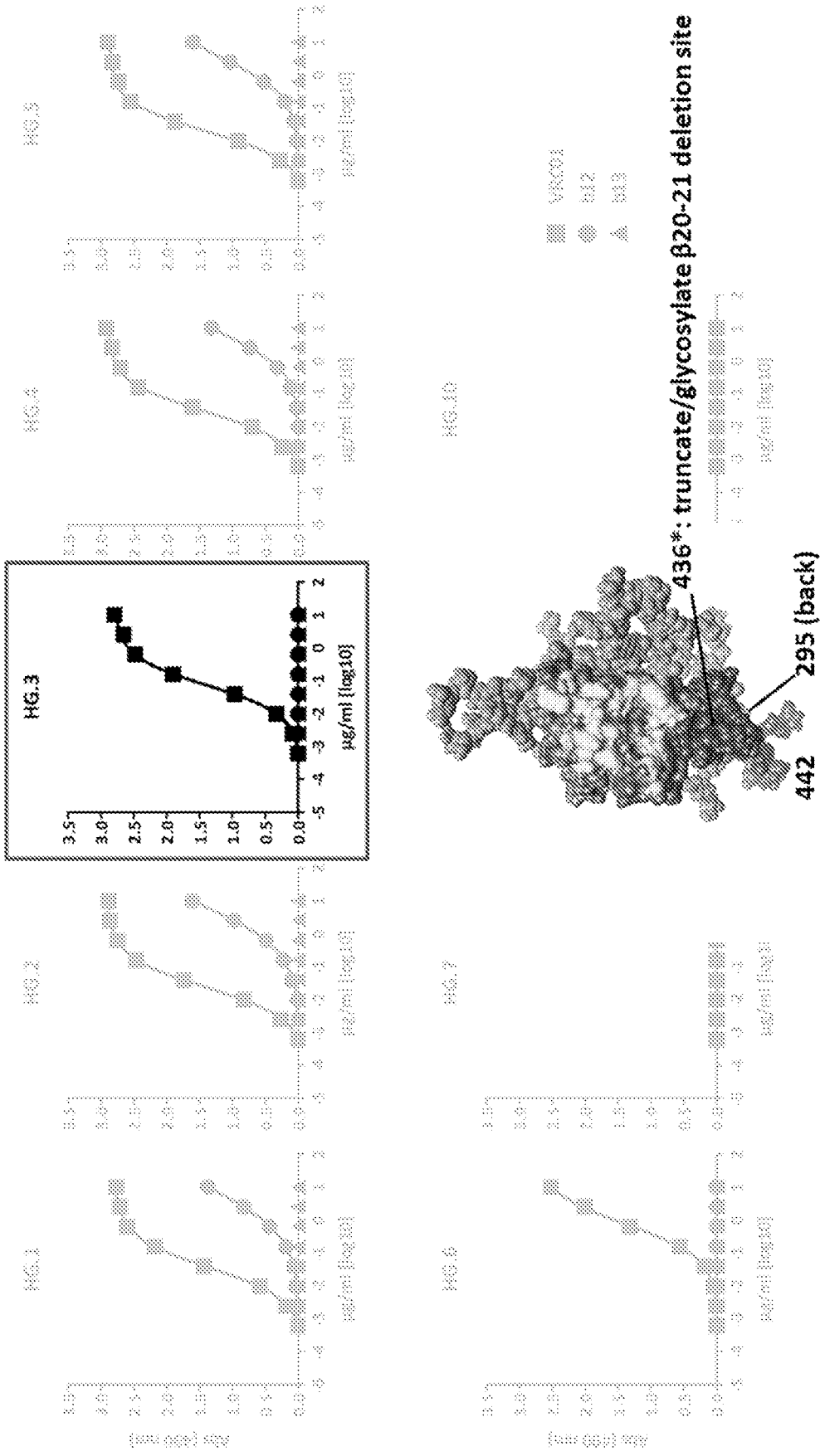
Identification of a CD4BS immunogen that binds to VRC01 but does not bind to b12

FIG. 6



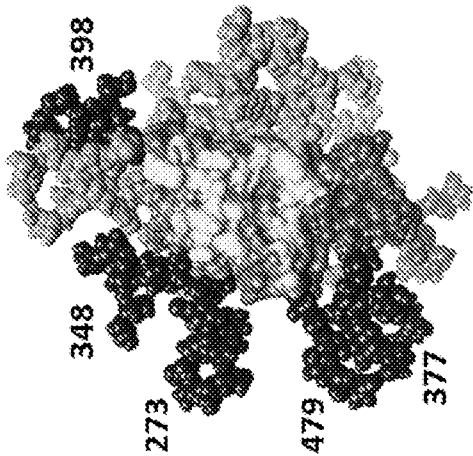
Identification of a CD4BS immunogen that binds to VRC01 but does not bind to b12

FIG. 7



Adding more glycans onto HG.3

FIG. 8



	@295	@442	@479	@273	@377	@436	@395	@346
HG.3	X	X				X		
HG.3.1	X	X	X			X		
HG.3.2	X	X	X			X	X	
HG.3.3	X	X	X			X		X
HG.3.4	X	X	X		X	X		
HG3.5	X	X	X	X		X		

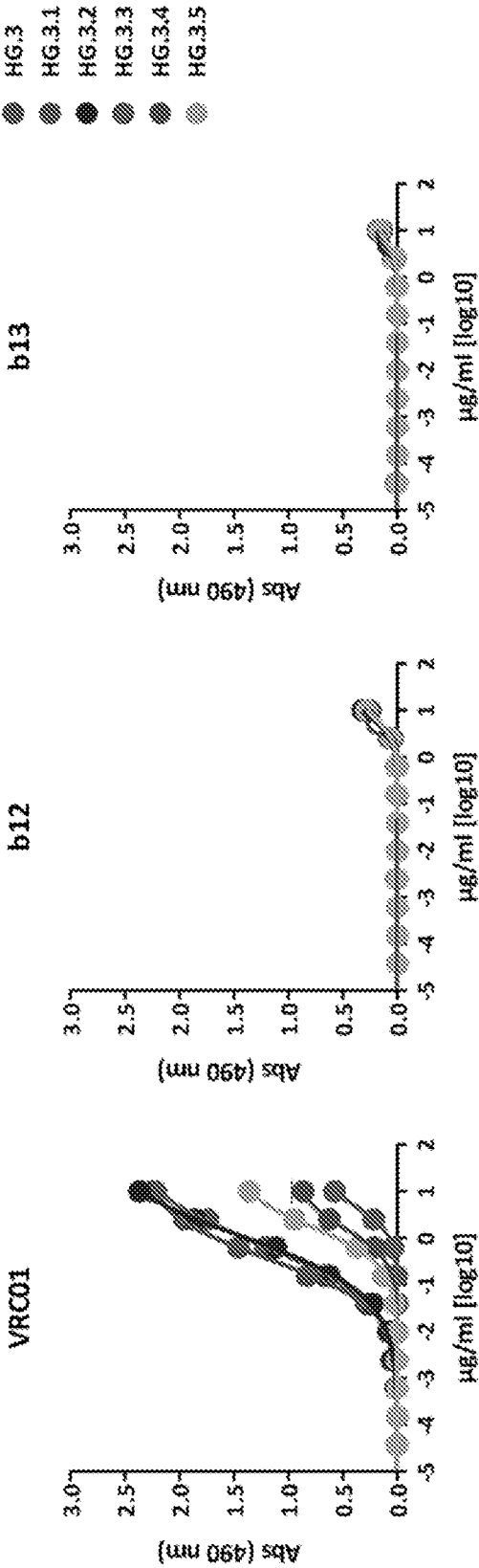


FIG. 9
Glycan masking may prevent binding of unfavorable antibodies

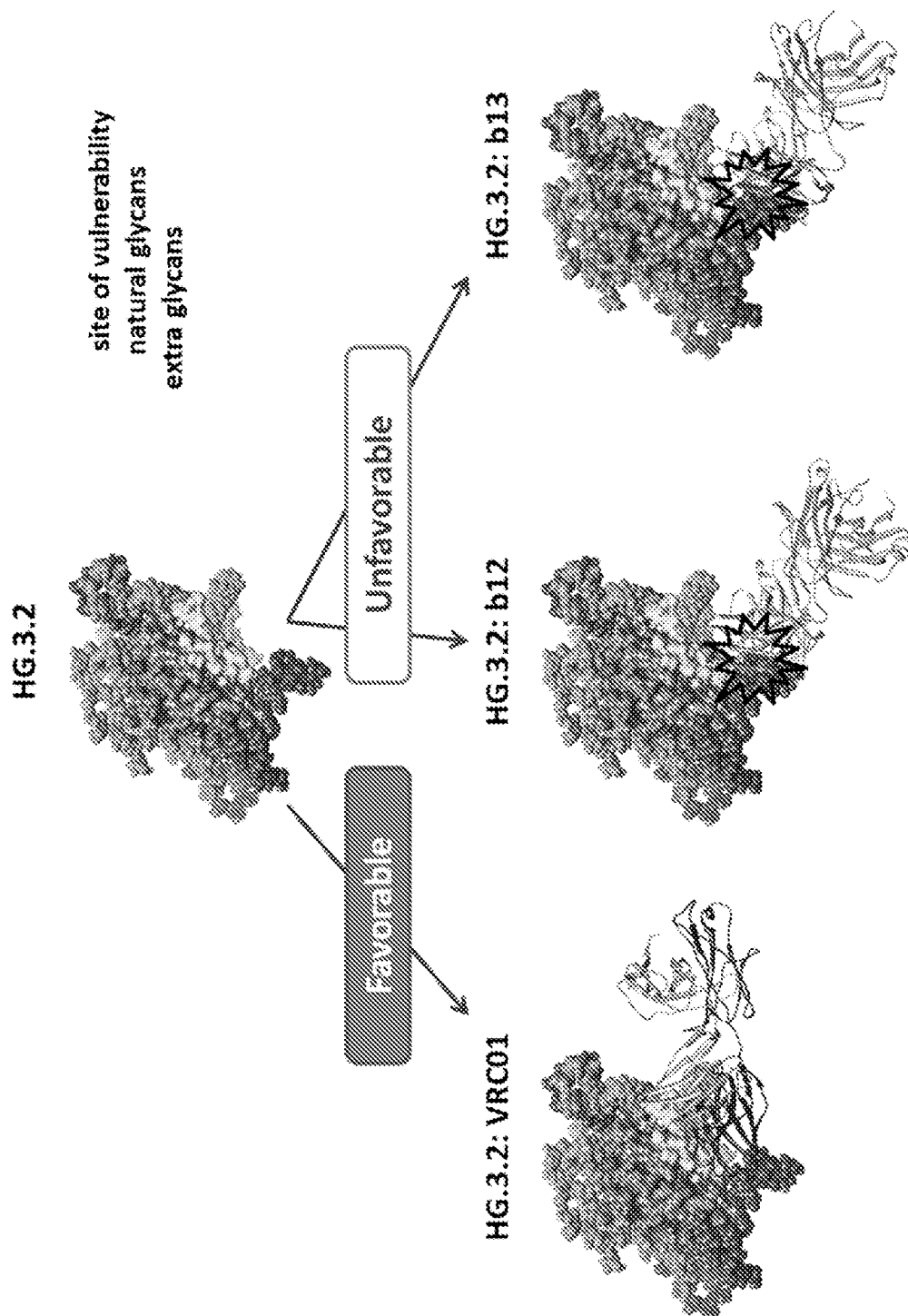
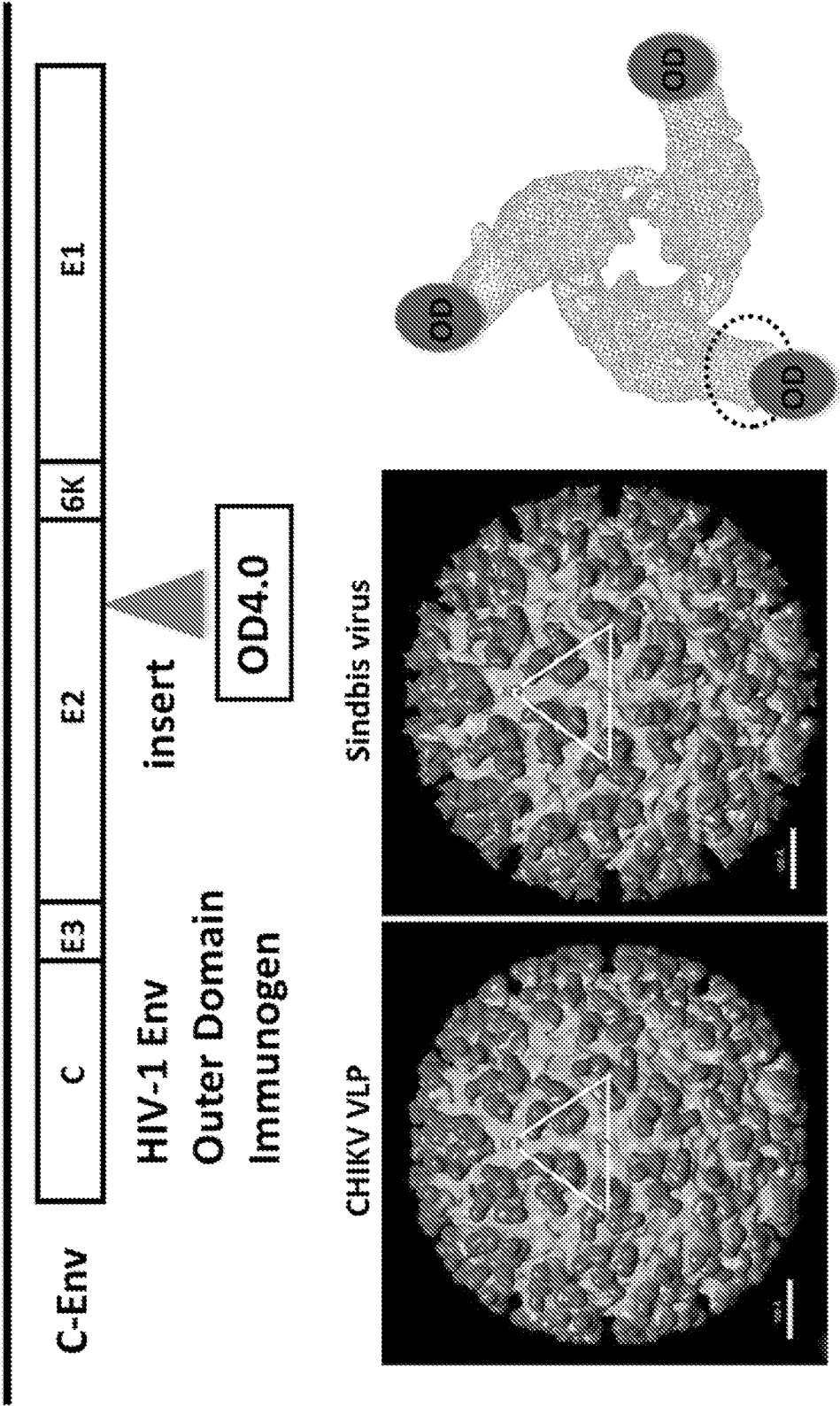


FIG. 10
Development of CHIKV VLPs
to Present HIV-1 Antigens



Akahata *et al.*, Nat Med, 2010

FIG. 11

CHIKV-OD Induced CD4-Binding Site Antibodies

Immunogen: CHIKV-ODm7V3.2V4Ker

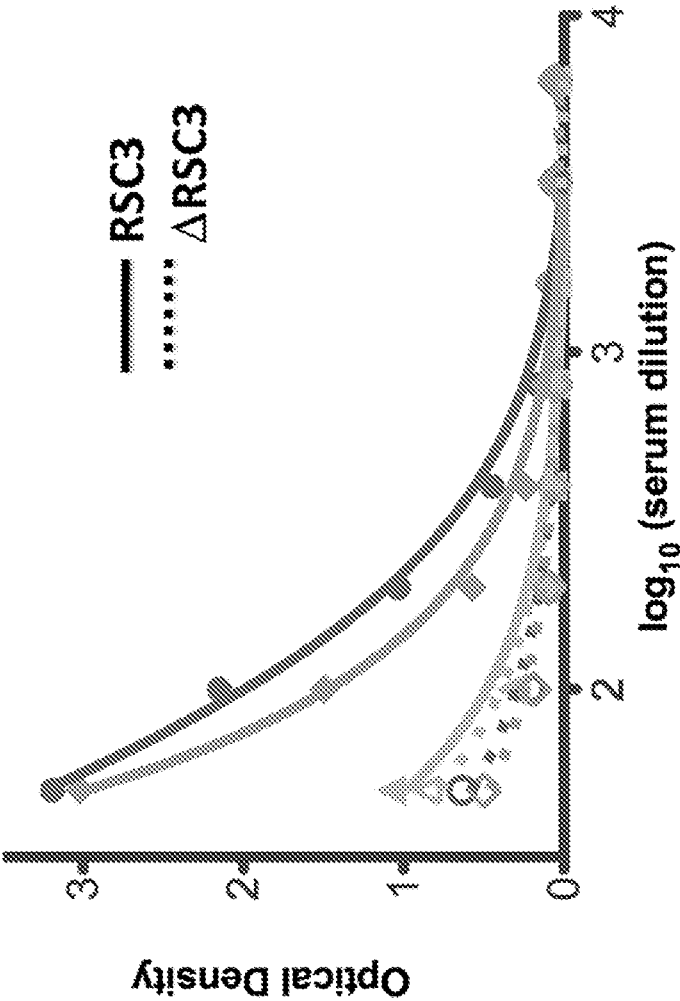


FIG. 12
Elicitation of CD4 BS Antibodies in NHP by
CHIKV-OD VLP

RSC3 vs. ΔRSC3

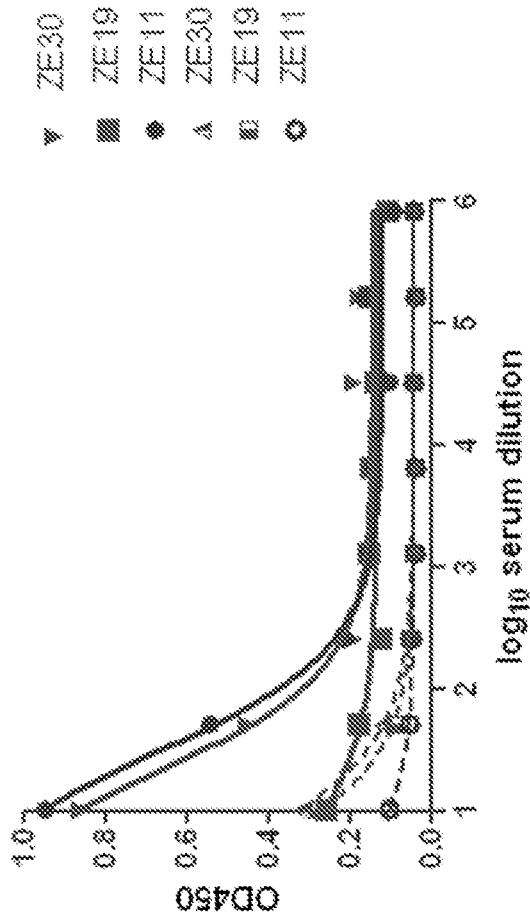


FIG. 13
Heavily glycosylated OD displayed on CHIKVLP
reacts only with VRC01 and VRCPG04

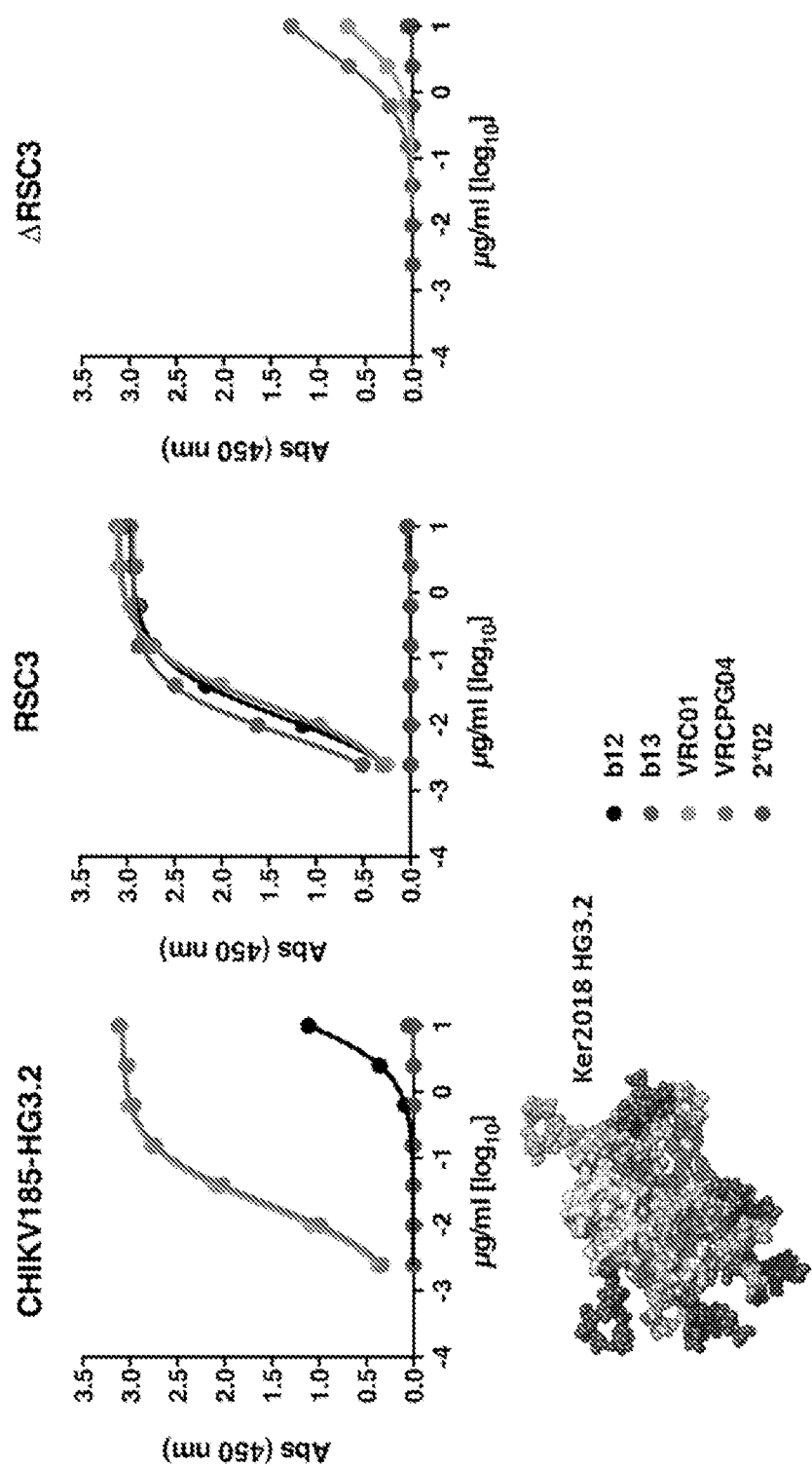


FIG. 14
Summary

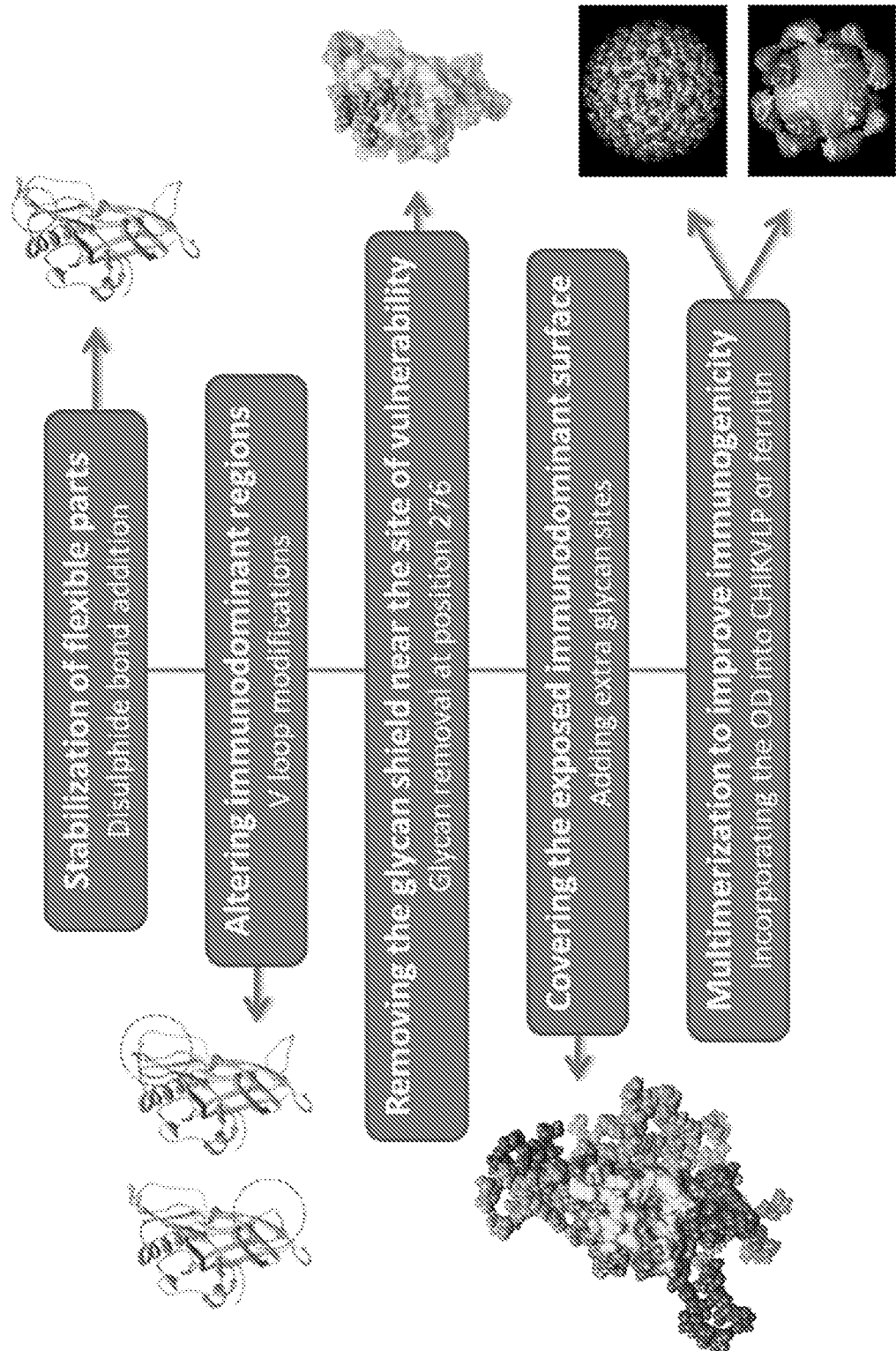


FIG. 15
Ferritin Nanoparticle Presentation of HIV OD

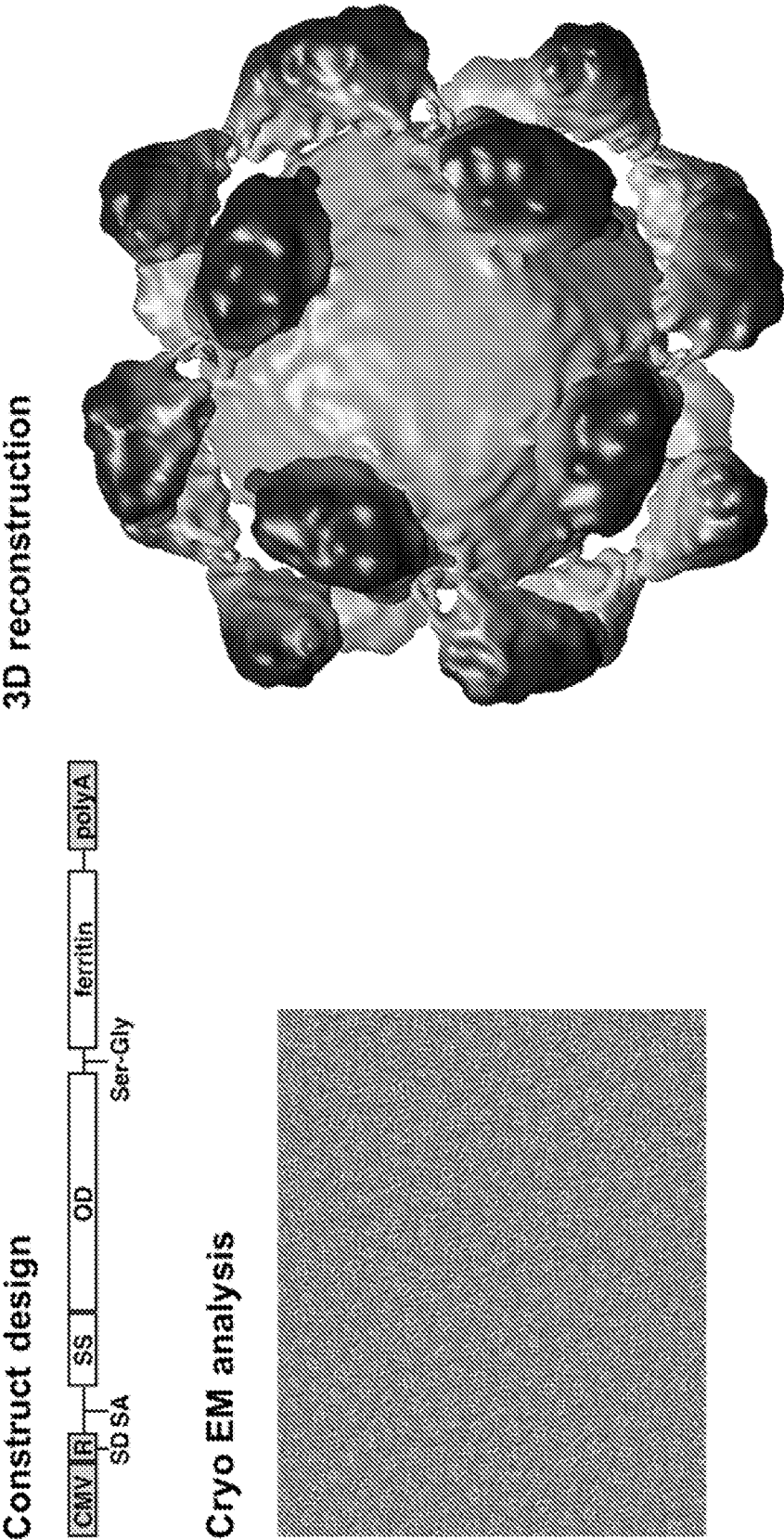


FIG. 16

Immunogenicity of HIV OD-ferritin nanoparticles

Rabbit experiments (n=3)

Experimental timeline diagram showing immunization at Day 0 and bleeding at Day 28 and Day 38.

Immunogen	Adjuvant	Dose
OD	Ribi	50 µg
OD-ferritin	Ribi	25 µg
OD-ferritin	no	25 µg
Irrelevant-ferritin	no	25 µg

Anti-OD IgG responses

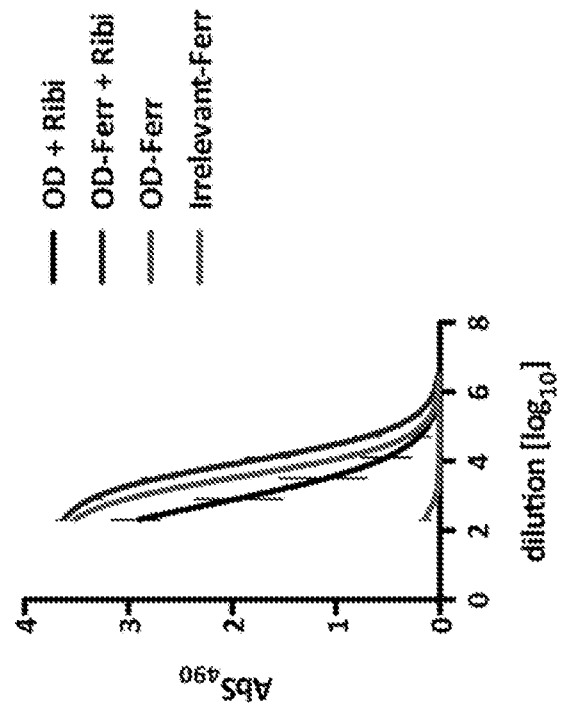


FIG. 17
Induction of CD4 BS Antibodies by
Glycan Modified RSC3: Y5

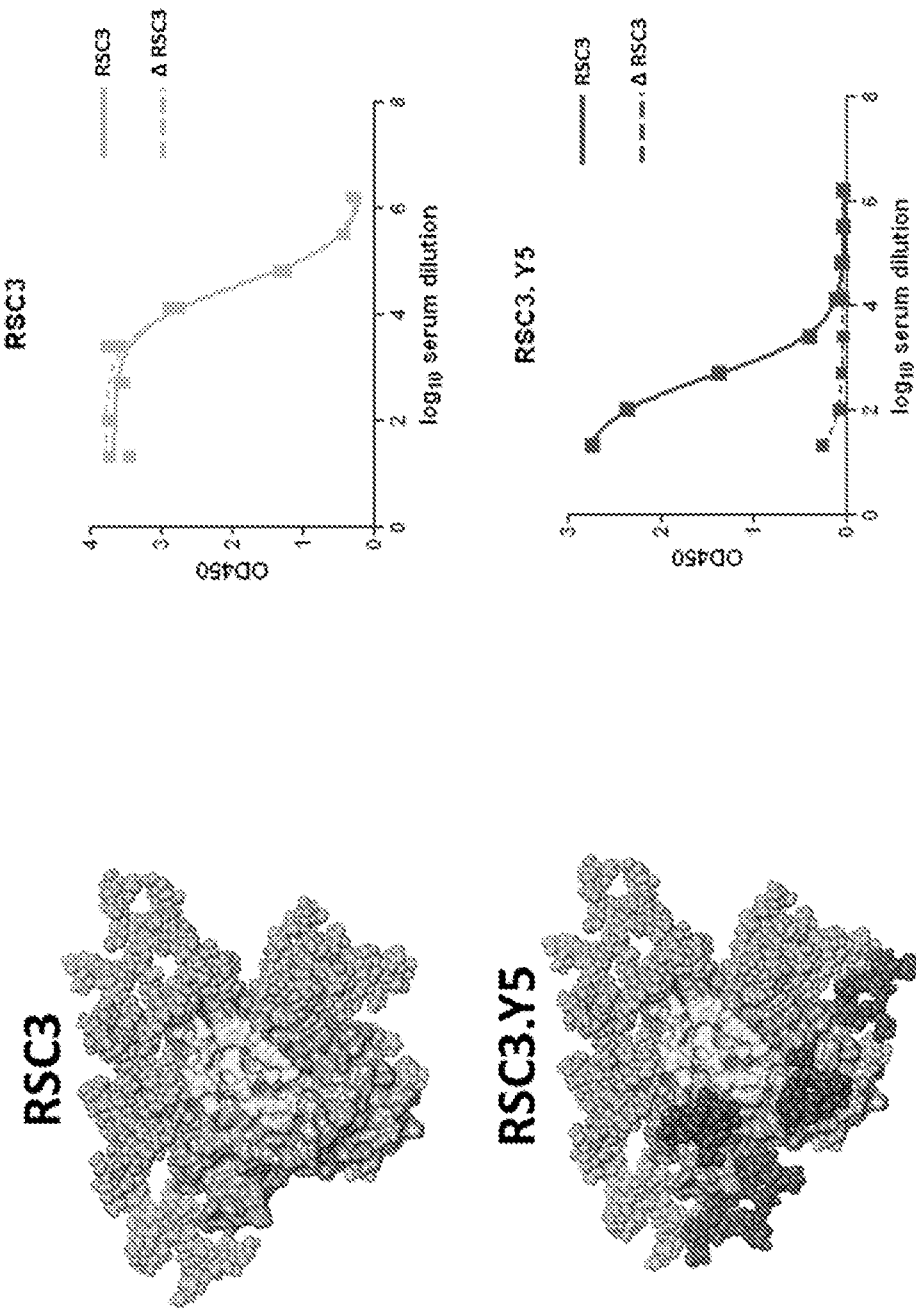
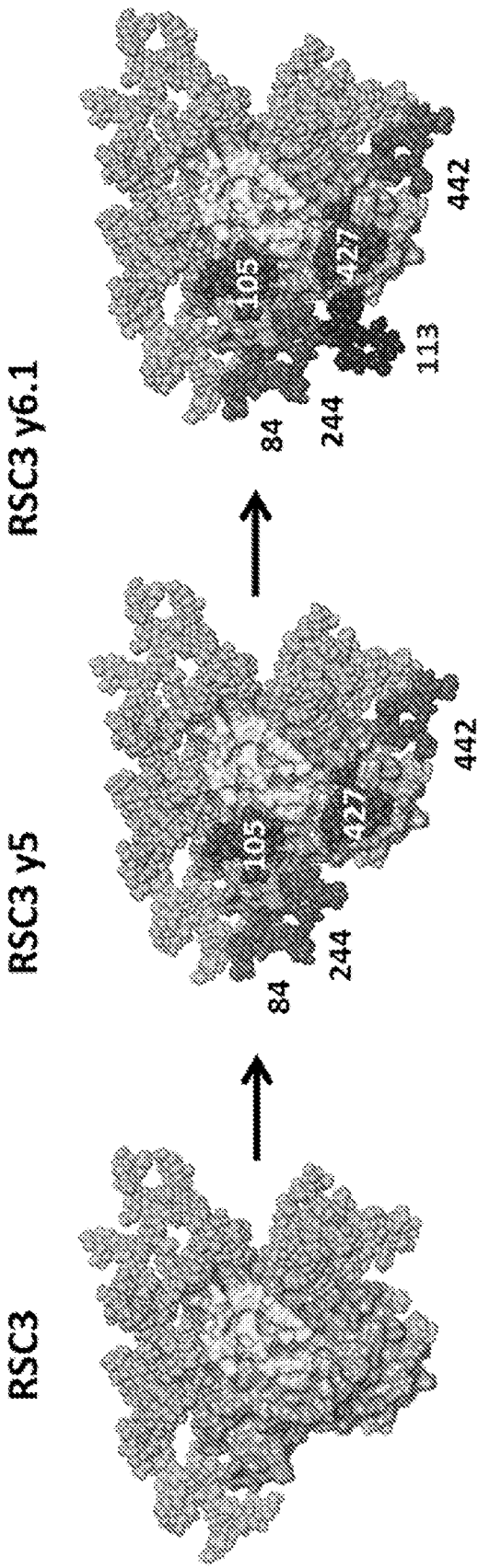


FIG. 18

Glycan masking of exposed regions on RSC3



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US20 11/042434

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item I.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing filed or furnished:
- a. (means)
- ☐ on paper
- ☒ in electronic form
- b. (time)
- ☐ in the international application as filed
- ☒ together with the international application in electronic form
- ☐ subsequently to this Authority for the purposes of search
2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.

PCT /US2011/042434

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

See Supplemental Box I

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☒ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US201 1/042434

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl.

C12N 15/49 (2006.0 1)

A61K 39/21 (2006.0 1)

A61P 31/18 (2006.0 1)

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

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

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

WPI, EPODOC & keywords: HIV, gpl20, immunogen, mutation, substitution and other like terms

Medline, Caplus, Biosis & Keywords: HIV, gpl20, immunogen, mutation, substitution and other like terms

GenomeQuest: Sequence search for SEQ ID NO: 1-161, 203 and 206-215

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2009/100376 A2 (THE UNITED STATES OF AMERICA, AS REPRESENTED BY THE SECRETARY, DEPARTMENT OF HEALTH & HUMAN SERVICES et al.) 13 August 2009 See Page 3, lines 28-30, Page 43, line 10 to Page 44, line 15, Page 53, lines 1-15, Page 54, Pages 60-61, Pages 71, 72, and 75-79 and Examples 1, 3 and 4 SEQ ID NO: 3 (Page 55) comprises a sequence identical to present SEQ ID NO: 74 and shares > 95% identity with present SEQ ID NO: 75 and 76 SEQ ID NO: 12 (Pages 63-64) shares >95% identity with present SEQ ID NO: 129 and 210	1-5, 7-11, 15-18, 21, 22, 24-35
A	US 6858712 B1 (CHANG et al.) 22 February 2005 See Column 4, line 66 to Column 5, line 3 and SEQ ID NO: 11 comprises a sequence sharing > 90% identity to present SEQ ID NO: 203	



Further documents are listed in the continuation of Box C



See patent family annex

* Special categories of cited documents:	
"A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"E" earlier application or patent but published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

24 October 2011

Date of mailing of the international search report

4 NOVEMBER 2011

Name and mailing address of the ISA/AU

AUSTRALIAN PATENT OFFICE
PO BOX 200, WODEN ACT 2606, AUSTRALIA
E-mail address: pct@ipaaustralia.gov.au
Facsimile No. +61 2 6283 7999

Authorized officer

ESTHER NG

AUSTRALIAN PATENT OFFICE
(ISO 9001 Quality Certified Service)

Telephone No : +61 2 6283 3129

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US201 1/042434

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	BROWN, B.K. et al., 'Biologic and Genetic Characterization of a Panel of 60 Human Immunodeficiency Virus Type 1 Isolates, Representing Clades A, B, C, D, CRF01_AE, and CRF02_AG, for the Development and Assessment of Candidate Vaccines', Journal of Virology, 2005, vol. 79, No. 10, pages 6089-6101 and GenBank Accession No. AAW72235 and AY736810 31 August 2005 See Abstract and Figure 2	

Supplemental Box I

(To be used when the space in any of Boxes I to IV is not sufficient)

Continuation of Box No. III:

The international application does not comply with the requirements of unity of invention because it does not relate to one invention or to a group of inventions so linked as to form a single general inventive concept.

PCT Rule 13.2 states that where a group of inventions is claimed in one and the same international application, the requirement of unity of invention referred to in Rule 13.1 shall be fulfilled only where there is a technical relationship among those inventions involving one or more of the same corresponding special technical features. Rule 13.2 states that the expression 'special technical features' shall mean those technical features that define features that define a contribution which each of the claimed inventions, considered as a whole, makes over the prior art.

In assessing whether there is more than one invention claimed, I have given consideration to those features which can be considered to potentially distinguish the claimed combination of features from the prior art. Where different claims have different distinguishing features they define different inventions.

The International Searching Authority has found that there are 3 different inventions as follows:

Invention 1

Claims 1-35 (all in part as they relate to SEQ ID NO: 203) are directed to an isolated immunogen including a human immunodeficiency virus (HTV) outer domain polypeptide which comprises SEQ ID NO: 203, further modifications in one or more of the β 20-21, VI, V2, V3, V4 and V5 loops of SEQ ID NO: 203 or SEQ ID NO: 203 with one or more amino acid substitutions at designated positions and its use in a method for treating/preventing HTV-1 infection or detecting/isolating HTV-1 binding antibody in a subject.

An isolated immunogen comprising SEQ ID NO: 203 with or without modifications in one or more of the β 20-21, VI, V2, V3, V4 and V5 loops or at designated positions within said sequence and its use in treating/preventing HIV-1 infection or detecting/isolating HTV-1 binding antibody in a subject is considered to define a first distinguishing feature.

Invention 2

Claims 1-35 (all in part as they relate to SEQ ID NO: 1) are directed to an isolated immunogen including a human immunodeficiency virus (HTV) outer domain polypeptide which comprises SEQ ID NO: 1, further modifications in one or more of the β 20-21, VI, V2, V3, V4 and V5 loops of SEQ ID NO: 1 or SEQ ID NO: 1 with one or more amino acid substitutions at designated positions and its use in a method for treating/preventing HTV-1 infection or detecting/isolating HTV-1 binding antibody in a subject.

An isolated immunogen comprising SEQ ID NO: 1 with or without modifications in one or more of the β 20-21, VI, V2, V3, V4 and V5 loops or at designated positions within said sequence and its use in treating/preventing HTV-1 infection or detecting/isolating HTV-1 binding antibody in a subject is considered to define a second distinguishing feature.

Invention 3

Claims 1-35 (all in part as they relate to SEQ ID NO: 74) are directed to an isolated immunogen including a human immunodeficiency virus (HTV) outer domain polypeptide which comprises SEQ ID NO: 74, further modifications in one or more of the β 20-21, VI, V2, V3, V4 and V5 loops of SEQ ID NO: 74 or SEQ ID NO: 74 with one or more amino acid substitutions at designated positions and its use in a method for treating/preventing HTV-1 infection or detecting/isolating HTV-1 binding antibody in a subject.

An isolated immunogen comprising SEQ ID NO: 74 with or without modifications in one or more of the β 20-21, VI, V2, V3, V4 and V5 loops or at designated positions within said sequence and its use in treating/preventing HTV-1 infection or detecting/isolating HTV-1 binding antibody in a subject is considered to define a third distinguishing feature.

.../Continued in Supplemental Box II

Supplemental Box II

(To be used when the space in any of Boxes I to VIII is not sufficient)

Continuation of Supplemental Box No I:

The only feature common to all of the claims is that the immunogen comprises modifications within the outer domain of the HTV envelope gpl20 protein. However this feature does not make a contribution over the prior art because it is disclosed in the following document:

WO 2009/100376 A2 (THE UNITED STATES OF AMERICA, AS REPRESENTED BY THE SECRETARY, DEPARTMENT OF HEALTH AND HUMAN SERVICES et al.) 13 August 2009

WO 2009/100376 discloses antigens (including HTV gpl20) which can be used to treat or prevent HTV-1 infection in a subject (Page 3 to Page 4, first paragraph). WO 2009/100376 further discloses that said antigen can include the outer domain of a gpl20 polypeptide that is appropriately mutated to provide antigenic cloaking (Page 53, line 30 to Page 54, line 23). These mutations include modifications to the $\beta 20/\beta 21$ bridging sheet, VI, V2 and/or V3 variable loops (Page 54, lines 1- 21). Moreover, WO 2009/100376 exemplifies an amino acid sequence (identified as SEQ ID NO: 3) that encodes said antigenically-cloaked gpl20 antigen and comprises present SEQ ID NO: 74 (Page 55, lines 9-14).

Thus in light of this disclosure, the common feature cannot constitute a special technical feature within the meaning of PCT Rule 13.2, second sentence, since it makes no contribution over the prior art. As a result of the common feature not satisfying the requirement for being a special technical feature, it follows that it cannot provide the necessary technical relationship between the identified inventions. Therefore the claims do not satisfy the requirement of unity of invention *a posteriori*. Thus each HTV gpl20 outer domain polypeptide (SEQ ID NO: 203, SEQ ID NO: 1 and SEQ ID NO: 74) and its use in generating immunogens for therapeutic or detection methods for HTV infection is considered to be a separate invention.

Accordingly, because the claims do not define inventions which share a special technical feature or single inventive concept, there is a lack of unity.

Information on patent family members

PCT/US2011/042434

Patent Document Cited in Search Report		Patent Family Member					
WO	2009100376	NONE					
US	6858712	AU	47765/85	EP	0173529	EP	0185444
		EP	0552850	IL	76082	JP	2003038190
		JP	7149796	JP	62026300	JP	S61501431
		JP	2004224798	NZ	213155	US	6001977
		US	6610476	US	7815916	US	2005095581
		WO	8601535				

END OF ANNEX