

**NOVEL DRUGGABLE REGIONS IN THE HUMAN CYTOMEGALOVIRUS GLYCOPROTEIN B
POLYPEPTIDE AND METHODS OF USE THEREOF**

RELATED APPLICATIONS

5 This application claims priority to U.S. Provisional Application No. 63/153,164 filed February 24, 2021 and U.S. Provisional Application No. 63/306,669 filed February 4, 2022. The entire content of each of the foregoing applications is herein incorporated by reference in its entirety.

REFERENCE TO SEQUENCE LISTING

10 This application is being filed electronically via EFS-Web and includes an electronically submitted sequence listing in .txt format. The .txt file contains a sequence listing entitled "PC72715_Feb2022_ST25.txt" created on February 3, 2022 and having a size of 1,374 KB. The sequence listing contained in this .txt file is part of the specification and is incorporated herein by
15 reference in its entirety.

FIELD OF THE INVENTION

 The present invention relates to novel druggable regions in the human cytomegalovirus (HCMV) glycoprotein B (gB) polypeptide and methods of using same, e.g. for drug discovery.

BACKGROUND OF THE INVENTION

 Human cytomegalovirus (HCMV) is a double stranded DNA virus of the β -herpesvirus family. HCMV is the leading cause of congenital and neonatal hearing loss resulting from vertical virus transmission following infection or reactivation of latent virus in pregnant women.
25 In addition, HCMV is a common opportunistic pathogen affecting immunosuppressed patients, such as solid organ and stem cell transplant patients, AIDS patients, etc. Though development of a vaccine against HCMV has been listed as a top priority by the Institute of Medicine, none has been licensed to date.

 The HCMV genome encodes several envelope glycoproteins, one of which is
30 glycoprotein B (gB). Glycoprotein B is a fusogen that is required for virus entry into cells and an important target for neutralizing antibody (nAb) responses to infection. HCMV vaccines that incorporate gB subunit antigens have been under development.

 Because fusion is a key step in viral infectivity, identification of druggable regions within the gB protein will further identification and development of therapeutics that can specifically
35 inhibit fusion and, therefore, inhibit viral infection by HCMV.

SUMMARY OF THE INVENTION

HCMV gB protein structures in both its pre-fusion and post-fusion conformations have been solved as described in detail below, thereby providing information about the structure of the polypeptide, and druggable regions, domains and the like contained therein, all of which
5 may be used in rational-based drug design efforts.

Accordingly, the present invention provides in part novel druggable regions in the HCMV gB protein. The interaction of a compound with such regions, or the modulation of the activity of such regions with a compound, could inhibit viral fusion and hence viral infectivity. In one aspect, the present invention provides methods of screening compounds against these
10 druggable regions in order to discover a candidate therapeutic for HCMV infection caused by HCMV having gB protein such as, for example, a viral fusion inhibitor.

Furthermore, this invention provides a method for identifying a candidate therapeutic for treating or preventing a disease caused by HCMV infection comprising contacting a HCMV having a glycoprotein B (gB) polypeptide which comprises a druggable region with a compound,
15 wherein binding of said compound indicates a candidate therapeutic. Compounds may in certain embodiments be selected from the following classes of compounds: proteins, peptides, polypeptides, peptidomimetics, antibodies, nucleic acids, and small molecules, or may be selected from a library of compounds. Such a library may be generated by combinatorial synthetic methods. Binding may be assayed either *in vitro* or *in vivo*. In certain embodiments of
20 this method, the protein is HCMV gB protein and comprises at least one residue, preferably three residues, from a druggable region of HCMV gB protein. Such druggable regions also may be utilized in the structure determination, drug screening, drug design, and other methods described and claimed herein.

In one embodiment, the druggable region comprises residues K130-A135, D216-W233,
25 R258-K260, A267-V273, R327-D329, W349-E350, V480-K518 and N676-Y690 of SEQ ID NO:1. These residues form the binding pocket for Domain V of HCMV gB protein in the postfusion conformation.

In yet another embodiment, a druggable region comprises a fusion loop or a portion thereof of HCMV gB protein in the postfusion conformation.

30 In another aspect, the present invention provides methods for identifying a candidate therapeutic for treating or preventing an infection caused by HCMV having gB protein. In certain embodiments, such methods comprise contacting the gB protein which comprises a druggable region with a compound, wherein the modulation of the activity of said gB protein indicates a candidate therapeutic. In other embodiments, such methods comprise contacting the gB protein
35 which comprises a druggable region with a compound, wherein the preclusion of movement or interaction of said druggable region indicates a candidate therapeutic. In still other embodiments, the modulation of the function or activity of said gB protein involves precluding the completion of the post-fusion conformational change. In yet another embodiment, the

modulation of the function or activity of said gB protein involves interfering with the first stage of the conformational change. In another embodiment, such method comprises contacting the gB protein which comprises a druggable region with a compound, wherein the inhibition of fusion of said virus indicates a candidate therapeutic. In yet another embodiment, the method comprises
5 contacting the gB protein which comprises a druggable region with a compound, wherein the inhibition of viral infectivity of said HCMV indicates a candidate therapeutic. In still another embodiment, the reduction of at least one symptom of a disease caused by HCMV infection in a subject indicates a candidate therapeutic.

In a further embodiment, the invention provides a method for identifying a candidate
10 therapeutic for disease caused by infection with HCMV having gB protein comprising contacting the gB protein which comprises a druggable region with a compound, wherein the compound prevents Domain V of the gB protein or a fragment thereof from binding in its binding pocket. In one embodiment, the binding pocket for Domain V of the gB protein comprises residues K130-A135, D216-W233, R258-K260, A267-V273, R327-D329, W349-E350, V480-K518 and N676-
15 Y690 of SEQ ID NO:1.

In another aspect, all of the information learned and described herein about the gB protein may be used in methods of designing modulators of its biological activity. In one embodiment, a method for designing a modulator for the prevention or treatment of a disease caused by infection with HCMV having gB protein, comprises: (a) providing a three-dimensional
20 structure for gB protein; (b) identifying a potential modulator for the prevention or treatment of disease caused by infection with HCMV having gB protein by reference to the three-dimensional structure; (c) contacting the gB protein with the potential modulator; and (d) assaying the activity of the gB protein or determining the viability of the virus having said gB protein after contact with the modulator, wherein a change in the activity of the protein or the
25 viability of the virus indicates that the modulator may be useful for prevention or treatment of a virus-related disease or disorder. In certain embodiments, the potential modulator is identified by reference to the three-dimensional structure of HCMV having gB protein. In other embodiments, the potential modulator is identified by reference to the three-dimensional structure of a druggable region of the gB protein or a fragment thereof.

30 In a further aspect, the present invention provides modulators (in certain embodiments, inhibitors) of HCMV infection, as well as pharmaceutical compositions and kits comprising the same. Such modulators may in certain embodiments interact with a druggable region of the invention. In still another aspect, the present invention is directed toward a modulator that is a fragment (or homolog of such fragment or mimetic of such fragment) of Domain V of a HCMV
35 gB protein and competes for that druggable region. Modulators of any of the above-described druggable regions may be used alone or in complementary approaches to treat or prevent infection by HCMV.

Finally, the present invention provides a druggable region of HCMV gB comprising residues K130-A135, D216-W233, R258-K260, A267-V273, R327-D329, W349-E350, V480-K518 and N676-Y690 of SEQ ID NO: 1. In one aspect, the residues of the druggable region are in a postfusion conformation.

- 5 The embodiments and practices of the present invention, other embodiments, and their features and characteristics, will be apparent from the description, figures and claims that follow, with all of the claims hereby being incorporated by this reference into this Summary.

BRIEF DESCRIPTION OF THE DRAWINGS

- 10 **FIG. 1A - 1B** describe two-dimensional (2D) class averages of gB conformers. **FIG. 1A** depicts the 2D projections from a postfusion gB structure. Projection images of an electron cryomicroscopy structure of postfusion gB bound with antibody Fabs are shown. **FIG. 1B** depicts 2D class averages. Two-dimensional class averages from electron cryomicroscopy images obtained from a preparation of gB extracted from CMV virions after treatment with a fusion inhibitor and a cross-linker and binding of an antibody fragment are shown on the right. Class averaged images that do not resemble any of the reference postfusion gB two dimensional projections are identified by circles.

- FIG. 2** describes glycoprotein B amino acids included in the prefusion and postfusion gB-Fab complex models from our electron cryomicroscopy structures. The amino acids that can be modeled in the electron cryomicroscopy density maps are highlighted with the domain codes (Domain I (*italics only*, i.e., upper sequence (prefusion) residues 133-344; lower sequence (post-fusion) residues 133-344); Domain II (**bold and underlined**, i.e., upper sequence (prefusion) residues 121-132 and 345-436; lower sequence (post-fusion) residues 121-132 and 345-439); Domain III (**bold only**, i.e., upper sequence (prefusion) residues 86-120 and 483-550; lower sequence (post-fusion) residues 86-120 and 474-550); Domain IV (*italics and underlined*, i.e., upper sequence (prefusion) residues 551-641; lower sequence (post-fusion) residues 551-641); Domain V (*italics and bold*, i.e., upper sequence (prefusion) residues 642-724; lower sequence (post-fusion) residues 642-697); MPR (underline only, i.e., upper sequence (prefusion) residues 25-7507; lower sequence (post-fusion) no residues); TM (*italics, bold, and underlined*, i.e., upper sequence (prefusion) residues 751-769; lower sequence (post-fusion) no residues)). The upper and lower sequences are for the prefusion and postfusion structure models, respectively.

- 35 **FIG. 3A - 3B** depict the fitting of models into the density maps. The models of inhibitor compound stabilized prefusion (**FIG. 3A**) and postfusion gB conformation (**FIG. 3B**) are fitted into the light gray density maps. gB components are dark gray, and SM5-1 fab components are

black. Approximate position of the virus envelope as determined by the position of the TM region in the prefusion structure is indicated by black horizontal lines.

FIG. 4A - 4B depict a comparison of the structures of gB in two conformations. The gB stabilized prefusion structure (**FIG. 4A**) and postfusion structure (**FIG. 4B**) are shown with one protomer to indicate the domains: I, II, III, IV, V, MPR and TM. The vertical black dashed line extending from the top of the prefusion structure represents residues missing from the model due to a less defined density map. The overall dimensions of the buildable ectodomain parts of the structure are indicated by the dashed line rectangles. The arrows indicate the direction pointed by the C-termini of the central 3-helix bundle in domain III of each conformation. The 115Å dimension on the prefusion structure (**FIG. 4A**) indicates the height of the modeled part of the ectodomain.

FIG. 5A - 5D: **FIG. 5A** depicts the location of fusion inhibitor compound N-{4-[(1S)-1-[3,5-bis(trifluoromethyl)phenyl]ethyl]carbamothioyl}amino]phenyl}-1,3-thiazole-4-carboxamide in the prefusion gB model is shown in black. The chemical structure of the compound is shown in **FIG. 5D**. **FIG. 5B:** A close view of the electron density around the compound (grey transparent surface). Nearby amino acid residues are shown and domains are labeled. **FIG. 5C:** The interacting residues around the compound are shown.

FIG. 6A - 6C depict a model of structural rearrangements of gB during membrane fusion. FL (and asterisks) – fusion loop. DI – domain 1. DII – domain 2. DV – domain 5. TM – transmembrane region. The solid lines depict membranes: host cell and viral membranes. **FIG. 6A** (prefusion) depicts a prefusion conformation; **FIG. 6B** (Extended intermediate) depicts an extended intermediate conformation; **FIG. 6C** (postfusion) depicts a postfusion conformation.

FIG. 7A - 7B depict an exemplary disulfide bond mutation to stabilize gB in a prefusion conformation. The locations of the residues participating in the disulfide bond are depicted as gray spheres in a prefusion conformation (**FIG. 7A**) and postfusion conformation (**FIG. 7B**).

FIG. 8 depicts information from Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) file: 5CXF, Crystal structure of the extracellular domain of glycoprotein B from Human Cytomegalovirus, from Human cytomegalovirus (strain AD169), deposited 2015-07-28; DOI: 10.2210/pdb5CXF/ pdb.

Unit Cell:

Length (Å) Angle (°)
a = 92.183 α = 90.00
b = 133.930 β = 90.00

Length (Å) Angle (°)
c = 295.376 γ = 90.00

FIG. 9 depicts sequences of gB from clinical and laboratory-adapted HCMV strains (SEQ ID NO: 110 - SEQ ID NO: 111). Additional sequences may be found in an amino acid sequence alignment of gB from clinical and laboratory-adapted HCMV found in S4 Fig., from Burke et al.,
5 PLoS Pathog. 2015 Oct 20;11(10): e1005227. According to Burke et al. sixty HCMV gB sequences from clinical and laboratory-adapted strains, downloaded from NCBI's RefSeq data base, were aligned and analyzed using ClustalW2 and ESPript 3.x. Identical residues are shown as white text on red background, and similar residues are highlighted in yellow in S4 Fig. of Burke et al., said S4 Fig. and the description thereof is incorporated herein by reference in its
10 entirety.

FIG. 10 depicts the amino acid sequences for SEQ ID NOs: 1-43 and SEQ ID NOs: 47-106.

FIG. 11 depicts the dose-dependent IgG responses in both gB1666 and wild type gB (Towne)
15 immunized mice. The graph shows that 10 out of 10 mice immunized with wild type gB DNA, and 9 out of 10 mice immunized with gB1666 DNA generated detectable anti-gB IgG titers. Mean ± SD, LLOQ = 25.

FIG. 12 depicts the structural model of engineered gB1666 (light gray, structure code:P-GB-
20 002) is overlaid with the structural model of wild type HCMV gB (dark gray, structure code:P-GB-001). The new structure allows modeling of additional residues, 437-448 and 478-482 at the membrane distal end of the molecule and 770-779 in the transmembrane domain.

FIG. 13 depicts an example of the combinations of additional mutation combinations on the
25 pSB1666 background that could further stabilize prefusion gB. Disulfide bond at M371, W506 could link domains II and III. Disulfide bonds at N524, M684 and F541, E681 could link domains IV and V. Mutations of negatively charged patches at E686 to hydrophobic residues could further stabilize gB in prefusion conformation. Each domain is identified. Abbreviations: membrane proximal region (MPR), and transmembrane domain (TMD).
30

FIG. 14 depicts a SDS-PAGE documenting the expression and purification of recombinant gB2459 protein. The pSB2459 expression plasmid was transiently transfected into Expi293F cells. The cell pellets were harvested 68 hours after transfection, and the glycoprotein product gB2459 was purified in 25 mM HEPES pH 7.5, 250 mM NaCl, 0.02% n-Dodecyl β-D-maltoside
35 (DDM), 0.002% cholesteryl hemi-succinate (CHS) through a series of processes of solubilization, affinity and size exclusion chromatography. This figure shows the purified protein

analyzed by stain-free 4-20% SDS-PAGE under reducing conditions. The smearing of the protein band is consistent with gB2459 being heavily glycosylated. Lane M: protein marker; Lane 1: gB2457; and Lane 2: gB2459.

5 **FIG. 15** depicts the construct pSB2459 which contains N524C and M684C mutations on the pSB1666 background. The protein product gB2459 was purified through affinity tags without the presence of any fusion inhibitors. There are prefusion classes observed in the 2D class averaged images (two classes with obvious prefusion features are indicated with the numbers 1 and 2). In addition, the prefusion gB2459 is stable over a period of a few days. Sample solution
10 of gB2459 was stored at 4°C, aliquots of the sample were obtained at day 1 and day 7 to prepare the negative stained grids. An images dataset was collected and processed on these two grids. For each dataset, the particle populations in the prefusion and postfusion 2D classes were counted: the ratio between prefusion and postfusion conformation was 5:1 for the sample at day 1 and 3:1 at day 7.

15

FIG. 16A - 16D depict the design of soluble, detergent-free gB ectodomains. **FIG. 16A** shows gB ectodomain (1-707) with MPR, TM and CT regions removed. **FIG. 16B** shows gB ectodomain stabilized with additional cysteine mutations, e.g. D703C and P704C, in Domain V.

20 **FIG. 16C** shows gB ectodomain fused to a C-terminal GCN4 trimerization motif. **FIG. 16D** shows gB ectodomain fused to a C-terminal T4 fibrin foldon domain. *Legend:* Domain I (residues 134-344) - dark gray 3D volume structure; Domain II (residues 121-133 and 345-436) - light gray 3D volume structure; Domain III (residues 97-111, 475-539 and 640-648) - top center light gray vertical coils; Domain V (residues 649-707) - bottom internal dark gray coils
25 (see arrow **FIG. 16B**); and rectangle – trimerization location.

FIG. 17 depicts gel-filtration profiles of purified gB ectodomains, gB2264-gB2269, analyzed by Superose 6 Increase 10/300 in 20mM HEPES pH 7.5, 250mM NaCl.

30 **FIG. 18A – 18B** depict a Negative stain EM image (**FIG. 18A**) and representative 2D class averages (**FIG. 18B**) from negative stain EM of recombinant gB2555 without bound fusion inhibitor, which show monodispersed gB proteins are suitable to use as a framework to add more stabilizing mutations towards a prefusion form of gB in the absence of inhibitor and detergents.

35

FIG. 19A - 19B depict a Negative stain EM image (**FIG. 19A**) and representative 2D class averages (**FIG. 19B**) from negative stain EM of recombinant gB2556 without bound fusion inhibitor, which show monodispersed gB proteins are suitable to use as a framework to add

more stabilizing mutations towards a prefusion form of gB in the absence of inhibitor and detergents.

FIG. 20A – 20B depict the space-filled model of HCMV (Towne strain) domain V (residues I642-V697 (SEQ ID NO: 281)) in prefusion (**FIG. 20A**) and postfusion (**FIG. 20B**) structures. Hydrophobic residues are labeled.

FIG. 21A – 21C depict (**FIG. 21A**) HCMV gB (Towne Strain) domain V (light grey) and its binding pocket (dark grey) in postfusion conformation; (**FIG. 21B**) the space-filled model of the pocket without domain V is shown; and (**FIG. 21C**) ribbon representation of the same structure as in **FIG. 21A** with certain residues labeled.

SEQUENCE IDENTIFIERS

SEQ ID NO: 1 sets forth an amino acid sequence derived from a native HCMV gB (strain Towne).

SEQ ID NO: 2 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: Q98C, G271C.

SEQ ID NO: 3 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: Q98C, I653C.

SEQ ID NO: 4 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: G99C, A267C.

SEQ ID NO: 5 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: T100C, A267C.

SEQ ID NO: 6 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: T100C, S269C.

SEQ ID NO: 7 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: T100C, L651C.

SEQ ID NO: 8 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: D217C, F584C.

SEQ ID NO: 9 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: Y218C, A585C.

SEQ ID NO: 10 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: S219C, D654C.

SEQ ID NO: 11 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: N220C, D652C.

SEQ ID NO: 12 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: T221C, D652C.

- SEQ ID NO: 13 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: W240C, G718C.
- SEQ ID NO: 14 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: Y242C, K710C.
- 5 SEQ ID NO: 15 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: Y242C, D714C.
- SEQ ID NO: 16 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: S269C, I653C.
- SEQ ID NO: 17 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: G271C, P614C.
- 10 SEQ ID NO: 18 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: S367C, L499C.
- SEQ ID NO: 19 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: T372C, W506C.
- 15 SEQ ID NO: 20 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: F541C, Q669C.
- SEQ ID NO: 21 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: L548C, A650C.
- SEQ ID NO: 22 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: A549C, I653C.
- 20 SEQ ID NO: 23 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: S550C, D652C.
- SEQ ID NO: 24 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: G604C, F661C.
- 25 SEQ ID NO: 25 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: N605C, E665C.
- SEQ ID NO: 26 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: R607C, S675C.
- SEQ ID NO: 27 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: T608C, D679C.
- 30 SEQ ID NO: 28 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: E609C, F678C.
- SEQ ID NO: 29 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: R673C, S674C.
- 35 SEQ ID NO: 30 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: N676C, V677C.
- SEQ ID NO: 31 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: L680C, E681C.

- SEQ ID NO: 32 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: I683C, M684C.
- SEQ ID NO: 33 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: F687C, N688C.
- 5 SEQ ID NO: 34 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: Y690C, K691C.
- SEQ ID NO: 35 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: K695C, K724C.
- SEQ ID NO: 36 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following
- 10 mutations are included: T746C, F747C.
- SEQ ID NO: 37 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutations are included: K749C, N750C.
- SEQ ID NO: 38 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutation is included: K670L.
- 15 SEQ ID NO: 39 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutation is included: K670F.
- SEQ ID NO: 40 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutation is included: R673L.
- SEQ ID NO: 41 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following
- 20 mutation is included: R673F.
- SEQ ID NO: 42 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutation is included: K691L.
- SEQ ID NO: 43 sets forth the amino acid sequence of SEQ ID NO: 1, wherein the following mutation is included: K691F.
- 25 SEQ ID NO: 44 sets forth the amino acid sequence for a native HCMV gB (AD169; PDB: 5CXF) that folds into a postfusion conformation when expressed.
- SEQ ID NO: 45 sets forth the amino acid sequence for an HCMV gB variant (gB705) that folds into a postfusion conformation when expressed.
- SEQ ID NO: 46 sets forth the amino acid sequence for a native HCMV gB (Merlin strain) that
- 30 folds into a postfusion conformation when expressed.
- SEQ ID NO: 47 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: M96C and D660C.
- SEQ ID NO: 48 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: Q98C and N658C.
- 35 SEQ ID NO: 49 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: T100C and R258C.
- SEQ ID NO: 50 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: T100C and L656C.

- SEQ ID NO: 51 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: T100C and N658C.
- SEQ ID NO: 52 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: I117C and T406C.
- 5 SEQ ID NO: 53 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: I117C and S407C.
- SEQ ID NO: 54 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: Y153C and L712C.
- SEQ ID NO: 55 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: L162C and M716C.
- 10 SEQ ID NO: 56 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: D217C and S587C.
- SEQ ID NO: 57 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: D217C and Y589C.
- 15 SEQ ID NO: 58 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: S219C and F584C.
- SEQ ID NO: 59 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: S219C and A585C.
- SEQ ID NO: 60 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: S219C and N586C.
- 20 SEQ ID NO: 61 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: N220C and T659C.
- SEQ ID NO: 62 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: S223C and T659C.
- 25 SEQ ID NO: 63 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: W240C and A732A.
- SEQ ID NO: 64 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: W240C and G735C.
- SEQ ID NO: 65 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: Y242C and V728C.
- 30 SEQ ID NO: 66 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: Y242C and G731C.
- SEQ ID NO: 67 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: R258C and L656C.
- 35 SEQ ID NO: 68 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: S269C and L656C.
- SEQ ID NO: 69 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: S269C and N658C.

- SEQ ID NO: 70 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: D272C and P614C.
- SEQ ID NO: 71 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: V273C and V629C.
- 5 SEQ ID NO: 72 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: W349C and A650C.
- SEQ ID NO: 73 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: S367C and A500C.
- SEQ ID NO: 74 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: S367C and A503C.
- 10 SEQ ID NO: 75 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: K370C and Q501C.
- SEQ ID NO: 76 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: K522C and I683C.
- 15 SEQ ID NO: 77 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: I523C and I683C.
- SEQ ID NO: 78 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: I523C and M684C.
- SEQ ID NO: 79 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: N524C and M684C.
- 20 SEQ ID NO: 80 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: P525C and E681C.
- SEQ ID NO: 81 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: R540C and L680C.
- 25 SEQ ID NO: 82 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: F541C and L680C.
- SEQ ID NO: 83 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: L548C and P655C.
- SEQ ID NO: 84 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: A549C and N658C.
- 30 SEQ ID NO: 85 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: S550C and P655C.
- SEQ ID NO: 86 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: S550C and E657C.
- 35 SEQ ID NO: 87 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: Q591C and S668C.
- SEQ ID NO: 88 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: L603C and Y667C.

- SEQ ID NO: 89 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: G604C and L672C.
- SEQ ID NO: 90 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: R607C and N688C.
- 5 SEQ ID NO: 91 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: T608C and Q692C.
- SEQ ID NO: 92 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: E609C and K691C.
- SEQ ID NO: 93 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are
- 10 included: E610C and S674C.
- SEQ ID NO: 94 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: E610C and S675C.
- SEQ ID NO: 95 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: Q612C and V663C.
- 15 SEQ ID NO: 96 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: V737C and F755C.
- SEQ ID NO: 97 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are included: V741C and A754C.
- SEQ ID NO: 98 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutations are
- 20 included: V741C and F755C.
- SEQ ID NO: 99 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutation is included: D679S.
- SEQ ID NO: 100 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutation is included: D679N.
- 25 SEQ ID NO: 101 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutation is included: E682S.
- SEQ ID NO: 102 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutation is included: E682Q.
- SEQ ID NO: 103 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutation is
- 30 included: E686S.
- SEQ ID NO: 104 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutation is included: E686Q.
- SEQ ID NO: 105 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutation is included: N118P.
- 35 SEQ ID NO: 106 sets forth the amino acid of SEQ ID NO: 1, wherein the following mutation is included: D646P.
- SEQ ID NO: 107 sets forth the amino acid sequence for >5CXF:A|PDBID|CHAIN|SEQUENCE, from **FIG. 8**.

SEQ ID NO: 108 sets forth the amino acid sequence for >5CXF:B|PDBID|CHAIN|SEQUENCE, from **FIG. 8**.

SEQ ID NO: 109 sets forth the amino acid sequence for >5CXF:C|PDBID|CHAIN|SEQUENCE, from **FIG. 8**.

- 5 SEQ ID NO: 110 sets forth the amino acid sequence for a gB polypeptide from HAN13 gij|242345614|gb|GQ221973.1|:81988-84705 Human herpesvirus 5 strain HAN13, complete genome reverse complement, referenced in the description for **FIG. 9**.

SEQ ID NOs: 111 sets forth the amino acid sequence for a gB polypeptide from VR1814 gij|270355759|gb|GU179289.1|:81925-84642 Human herpesvirus 5 strain VR1814, complete genome reverse complement, referenced in the description for **FIG. 9**.

- 10 SEQ ID NOs: 112-140 sets forth the amino acid sequence for a gB polypeptide from various CMV gB strains described in **FIG. 9**.

SEQ ID NO: 141 – SEQ ID NO: 210 set forth a polynucleotide sequence encoding a polypeptide derived from HCMV, such as for example, gH, gL, UL128, UL130, UL131, gB or pp65.

- 15 SEQ ID NO: 211 – SEQ ID NO: 223 set forth an amino acid sequence for a polypeptide derived from HCMV, such as for example, gH, gL, UL128, UL130, UL131, gB or pp65.

SEQ ID NO: 224 sets forth an amino acid sequence for a polypeptide derived from HCMV.

SEQ ID NO: 225 – SEQ ID NO: 254 set forth a polynucleotide sequence encoding a

- 20 polypeptide derived from HCMV.

SEQ ID NO: 255 – SEQ ID NO: 259 set forth the C-Term fusion sequences of various gB ectodomain proteins set forth in **Table 9**.

SEQ ID NO: 260 – SEQ ID NO: 280 set forth the amino acid sequences of various fusion inhibitory peptides.

- 25 SEQ ID NO: 281 sets forth the amino acid sequence of Domain V of the gB polypeptide (Towne).

SEQ ID NO: 282 – SEQ ID NO: 284 set forth the amino acid sequences of interacting regions of HCMV gB protein as shown in **FIG. 20A – 20B**.

SEQ ID NO: 285 – SEQ ID NO: 289 set forth the amino acid residues that form the binding

- 30 pocket for Domain V of HCMV gB protein (in addition to R258-K260, R327-D329, and W349-E350 of SEQ ID NO: 1).

DETAILED DESCRIPTION

As described herein, the inventors elucidated a three-dimensional structure of a HCMV glycoprotein B (gB) polypeptide in a conformation that differs from the postfusion conformation and which we refer to as a prefusion conformation. Mutations to stabilize the polypeptide in a prefusion conformation were also discovered. The structures may be used to generate HCMV neutralizing antibody responses greater than those achieved with prior HCMV gB-based immunogens. The polypeptides described herein, and the nucleic acids that encode the polypeptides, may be used, for example, as potential immunogens in a vaccine against HCMV and as diagnostic tools, among other uses.

The inventors further discovered mutations that can be introduced into a cytomegalovirus (CMV) gB polypeptide, which can, among other things, greatly facilitate the production and subsequent purification of a gB antigen stabilized in the prefusion conformation; significantly improve the efficiency of production of a gB polypeptide in the prefusion conformation; alter the antigenicity of a gB polypeptide, as compared to the wild-type gB polypeptide; facilitate a focused immune response to prefusion gB; and reduce and/or eliminate steric occlusion of neutralizing epitopes of gB.

Because fusion is a key step in viral infectivity, identification of druggable regions within the gB protein will further development of therapeutics that can specifically inhibit viral infection by HCMV.

A. Native HCMV gB

Native HCMV gB is synthesized as a 906 or 907 amino acid polypeptide (depending upon the strain of CMV) that undergoes extensive posttranslational modification, including glycosylation at *N*- and *O*-linked sites and cleavage by ubiquitous cellular endoproteases into amino- and carboxy-terminal fragments. The *N*- and *C*-terminal fragments of gB, gp116 and gp55, respectively, are covalently connected by disulfide bonds, and the mature, glycosylated gB assumes a trimeric configuration. The gB polypeptide contains a large ectodomain (which is cleaved into gp116 and the ectodomain of gp55), a transmembrane domain (TM), and the intraviral (or cytoplasmic) domain (cytodomain).

Native HCMV gBs from various strains are known. For example, at least sixty HCMV gB sequences from clinical and laboratory-adapted strains are available from NCBI's RefSeq database. See also **FIG. 9**.

Accordingly, the term "CMV gB" polypeptide or "HCMV gB" polypeptide as used herein is to be understood as the native HCMV gB polypeptide from any human HCMV strain (not limited to the Towne strain). The actual residue position number may need to be adjusted for gBs from other human CMV strains depending on the actual sequence alignment.

HCMV gB is encoded by the UL55 gene of HCMV genome. It is an envelope glycoprotein that mediates the fusion of the HCMV viral membrane with a host cell membrane.

The protein undergoes a series of conformational changes from a prefusion to a postfusion form. The crystal structure of gB in its postfusion form is available (PDB accession code 5CXF), and the prefusion conformation is described hereinbelow.

5 B. Conformations

A HCMV gB postfusion conformation refers to a structural conformation adopted by HCMV gB subsequent to the fusion of the virus envelope with the host cellular membrane. The native HCMV gB may also assume the postfusion conformation outside the context of a fusion event, for example, under stress conditions such as exposure to heat, extraction from a
 10 membrane, expression as an ectodomain or storage. More specifically, the gB postfusion conformation is described, for example, in Burke et al., Crystal Structure of the Human Cytomegalovirus Glycoprotein B. *PLoS Pathog.* 2015 Oct 20;11(10): e1005227. See also, Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB): 5CXF, Crystal structure of the extracellular domain of glycoprotein B from Human Cytomegalovirus,
 15 from Human cytomegalovirus (strain AD169), deposited 2015-07-28; DOI: 10.2210/pdb5CXF/pdb; and **FIG. 9**. A sequence of a protein that when expressed, can fold into a postfusion conformation, is provided as SEQ ID NO: 44. Another example of a protein that when expressed folds into a postfusion conformation is provided as SEQ ID NO: 45. The postfusion conformation is about 165 Å tall and 65 Å wide.

As used herein, a “prefusion conformation” refers to a structural conformation adopted by the polypeptide that differs from the HCMV gB postfusion conformation at least in terms of molecular dimensions or three-dimensional coordinates. The prefusion conformation refers to a structural conformation adopted by HCMV gB prior to triggering of the fusogenic event that leads to transition of gB to the postfusion conformation. Isolating HCMV gB in a stable
 20 prefusion conformation may be useful in informing and directing development of improved vaccines and immunogenic compositions to address the important public health problem of cytomegalovirus infections. In some embodiments, a prefusion conformation includes a conformation that can bind to a prefusion-specific antibody. In some embodiments, a prefusion conformation includes a conformation that is characterized by coordinates set forth in **Table 1A**
 30 described in WO2021/260510 which is hereby incorporated by reference in its entirety. In some embodiments, the polypeptide is characterized by structure coordinates comprising a root mean square deviation (RMSD) of conserved residue backbone atoms when superimposed on backbone atoms described by structural coordinates of **Table 1A** described in WO2021/260510 which is hereby incorporated by reference in its entirety. In some embodiments, a prefusion
 35 conformation includes a conformation that is characterized by coordinates set forth in **Table 1B** described in WO2021/260510 which is hereby incorporated by reference in its entirety. In some embodiments, the polypeptide is characterized by structure coordinates comprising a root mean square deviation (RMSD) of conserved residue backbone atoms when superimposed on

backbone atoms described by structural coordinates of **Table 1B** described in WO2021/260510 which is hereby incorporated by reference in its entirety. In some embodiments, a polypeptide having a HCMV gB prefusion conformation refers to a polypeptide that includes a trimeric helix bundle, centered on the three-fold axis of the trimer and comprising residues L479 to K522 of each protomer, wherein the direction of the bundle from N-terminal to C-terminal along the three-fold axis (shown by the arrows in **FIG. 4A & FIG. 4B**) is towards the point on the three-fold axis intersected by the plane defined by residue W240 of each protomer, which is in a fusion loop near the tip of each Domain I of the trimer. In some embodiments, the helix bundle comprises the residues between L479 and K522, according to the numbering of SEQ ID NO: 1.

C. Mutants of Wild-type HCMV gB

The present invention includes polypeptides that comprise amino acid mutations relative to the amino acid sequence of the corresponding wild-type HCMV gB. The amino acid mutations include amino acid substitutions, deletions, or additions relative to a wild-type HCMV gB. Accordingly, the polypeptides are mutants of wild-type HCMV gBs.

In some embodiments, the polypeptides possess certain beneficial characteristics, such as being immunogenic. In some embodiments, the polypeptides possess increased immunogenic properties or improved stability in the prefusion conformation, as compared to the corresponding wild-type HCMV gB. Stability refers to the degree to which a transition of the HCMV gB conformation from prefusion to postfusion is hindered or prevented. In still other embodiments, the present disclosure provides polypeptides that display one or more introduced mutations as described herein, which may also result in improved stability in the prefusion conformation. The introduced amino acid mutations in the HCMV gB include amino acid substitutions, deletions, or additions. In some embodiments, the only mutations in the amino acid sequences of the mutants are amino acid substitutions relative to a wild-type HCMV gB.

Several modes of stabilizing the polypeptide conformation include amino acid substitutions that introduce disulfide bonds, introduce electrostatic mutations, fill cavities, alter the packing of residues, introduce N-linked glycosylation sites, and combinations thereof, as compared to a native HCMV gB.

In one aspect, the invention relates to a polypeptide that exhibits a conformation that is not the postfusion conformation. That is, the polypeptide exhibits a prefusion conformation as described above and does not exhibit a postfusion conformation. See, for example, the prefusion conformation illustrated in **FIG. 3A**, as compared to the postfusion conformation illustrated in **FIG. 3B**; **FIG. 4A**, as compared to the postfusion conformation illustrated in **FIG. 4B**; and **FIG. 6A**, as compared to the postfusion conformation illustrated in **FIG. 6C**. In some embodiments, the polypeptide is characterized by structure coordinates comprising a root mean square deviation (RMSD) of conserved residue backbone atoms when superimposed on backbone atoms described by structural coordinates of **Table 1A** described in WO2021/260510

which is hereby incorporated by reference in its entirety. In some embodiments, the polypeptide is characterized by structure coordinates comprising a root mean square deviation (RMSD) of conserved residue backbone atoms when superimposed on backbone atoms described by structural coordinates of **Table 1B** described in WO2021/260510 which is hereby
5 incorporated by reference in its entirety.

In some embodiments, the polypeptides are isolated, i.e., separated from HCMV gB polypeptides having a postfusion conformation. Thus, the polypeptide may be, for example, at least 80% isolated, at least 90%, 95%, 98%, 99%, or even 99.9% isolated from HCMV gB polypeptides in a postfusion conformation. In one aspect, the invention relates to a polypeptide
10 that specifically binds to an HCMV gB prefusion-specific antibody.

It will be understood that a homogeneous population of polypeptides in a particular conformation can include variations (such as polypeptide modification variations, e.g., glycosylation state), that do not alter the conformational state of the polypeptide. In several embodiments, the population of polypeptides remains homogeneous over time. For example, in
15 some embodiments, the polypeptide, when dissolved in aqueous solution, forms a population of polypeptides stabilized in the prefusion conformation for at least 12 hours, such as at least 24 hours, at least 48 hours, at least one week, at least two weeks, or more.

Without being bound by theory, the polypeptides disclosed herein are believed to facilitate a stabilized prefusion conformation of an HCMV gB polypeptide. The polypeptides
20 include at least one mutation as compared to a corresponding native HCMV gB polypeptide. A person of ordinary skill in the art will appreciate that the polypeptides are useful to elicit immune responses in mammals to CMV.

The native HCMV gB is conserved among the HCMV entry glycoproteins and is required for entry into all cell types. In view of the substantial conservation of HCMV gB sequences, the
25 amino acid positions amongst different native HCMV gB sequences may be compared to identify corresponding HCMV gB amino acid positions among different HCMV strains. Thus, the conservation of native HCMV gB sequences across strains allows use of a reference HCMV gB sequence for comparison of amino acids at particular positions in the HCMV gB polypeptide. Accordingly, unless expressly indicated otherwise, the polypeptide amino acid positions
30 provided herein refer to the reference sequence of the HCMV gB polypeptide set forth in SEQ ID NO: 1.

However, it should be noted that different native HCMV gB sequences may have different numbering systems from SEQ ID NO: 1, for example, there may be additional amino acid residues added or removed as compared to SEQ ID NO: 1 in a native HCMV gB sequence
35 derived from a strain other than Towne. As such, it is to be understood that when specific amino acid residues are referred to by their number, the description is not limited to only amino acids located at precisely that numbered position when counting from the beginning of a given amino acid sequence, but rather that the equivalent or corresponding amino acid residue in any

and all HCMV gB sequences is intended even if that residue is not at the same precise numbered position, for example if the HCMV sequence is shorter (e.g., a fragment) or longer than SEQ ID NO: 1, or has insertions or deletions as compared to SEQ ID NO: 1.

In some embodiments, the polypeptide is full-length, wherein the polypeptide includes the same number of amino acid residues as the mature full-length wild-type HCMV gB. In some embodiments, the polypeptide is a fragment, wherein the polypeptide includes less than the total number of amino acid residues as the mature full-length wild-type HCMV gB. In some embodiments, the truncated gB polypeptide includes only the ectodomain sequence.

10 C.1. Cysteine (C) Substitutions

In some embodiments, the polypeptide includes cysteine substitutions that are introduced, as compared to a native HCMV gB. In some embodiments, the polypeptide includes any one of 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 cysteine substitutions. Without being bound by theory or mechanism, the cysteine substitutions described herein are believed to facilitate stability of the polypeptide in a conformation that is not the HCMV gB postfusion conformation. The introduced cysteine substitutions may be introduced by protein engineering, for example, by including one or more substituted cysteine residues that form a disulfide bond. In several embodiments, the amino acid positions of the cysteines are within a sufficiently close distance for formation of a disulfide bond in the prefusion, and not postfusion, conformation of the HCMV gB.

The cysteine residues that form a disulfide bond can be introduced into native HCMV gB sequence by two or more amino acid substitutions. For example, in some embodiments, two cysteine residues are introduced into a native HCMV gB sequence to form a disulfide bond.

In some embodiments, the polypeptide includes a recombinant HCMV gB stabilized in a prefusion conformation by a disulfide bond between cysteines that are introduced into a pair of amino acid positions that are close to each other in the prefusion conformation and more distant in the postfusion conformation.

Exemplary cysteine substitutions as compared to a native HCMV gB include any mutation selected from **Table 2**, the numbering of which based on the numbering of SEQ ID NO: 1.

Table 2. Exemplary cysteine pairs for disulfide bond stabilization

Row	(i) Mutant ID	(ii) HCMV gB residue pairs for cysteine substitution, according to the numbering of SEQ ID NO: 1	(iii) Substitutions corresponding to SEQ ID NO: 1	(iv) Exemplary sequence that includes the mutations is set forth in:
1	gB-001; pSB01582	98 and 271	Q98C and G271C	SEQ ID NO: 2

2	gB-002	98 and 653	Q98C and I653C	SEQ ID NO: 3
3	gB-003; pSB01579	99 and 267	G99C and A267C	SEQ ID NO: 4
4	gB-004; pSB01580	100 and 267	T100C and A267C	SEQ ID NO: 5
5	gB-005; pSB01581	100 and 269	T100C and S269C	SEQ ID NO: 6
6	gB-006	100 and 651	T100C and L651C	SEQ ID NO: 7
7	gB-007	217 and 584	D217C and F584C	SEQ ID NO: 8
8	gB-008	218 and 585	Y218C and A585C	SEQ ID NO: 9
9	gB-009	219 and 654	S219C and D654C	SEQ ID NO: 10
10	gB-010	220 and 652	N220C and D652C	SEQ ID NO: 11
11	gB-011	221 and 652	T221C and D652C	SEQ ID NO: 12
12	gB-012	240 and 718	W240C and G718C	SEQ ID NO: 13
13	gB-013	242 and 710	Y242C and K710C	SEQ ID NO: 14
14	gB-014	242 and 714	Y242C and D714C	SEQ ID NO: 15
15	gB-015	269 and 653	S269C and I653C	SEQ ID NO: 16
16	gB-016	271 and 614	G271C and P614C	SEQ ID NO: 17
17	gB-017	367 and 499	S367C and L499C	SEQ ID NO: 18
18	gB-018	372 and 506	T372C and W506C	SEQ ID NO: 19
19	gB-019	541 and 669	F541C and Q669C	SEQ ID NO: 20
20	gB-020	548 and 650	L548C and A650C	SEQ ID NO: 21
21	gB-021	549 and 653	A549C and I653C	SEQ ID NO: 22
22	gB-022	550 and 652	S550C and D652C	SEQ ID NO: 23
23	gB-023	604 and 661	G604C and F661C	SEQ ID NO: 24
24	gB-024	605 and 665	N605C and E665C	SEQ ID NO: 25
25	gB-025	607 and 675	R607C and S675C	SEQ ID NO: 26
26	gB-026	608 and 679	T608C and D679C	SEQ ID NO: 27
27	gB-027	609 and 678	E609C and F678C	SEQ ID NO: 28
28	gB-028	673 and 674	R673C and S674C	SEQ ID NO: 29
29	gB-029	676 and 677	N676C and V677C	SEQ ID NO: 30
30	gB-030	680 and 681	L680C and E681C	SEQ ID NO: 31
31	gB-031	683 and 684	I683C and M684C	SEQ ID NO: 32
32	gB-032	687 and 688	F687C and N688C	SEQ ID NO: 33
33	gB-033	690 and 691	Y690C and K691C	SEQ ID NO: 34
34	gB-034	695 and 724	K695C and K724C	SEQ ID NO: 35
35	gB-035	746 and 747	T746C and F747C	SEQ ID NO: 36

36	gB-036	749 and 750	K749C and N750C	SEQ ID NO: 37
37	gB-043; pSB01656	96 and 660	M96C and D660C	SEQ ID NO: 47
38	gB-044; pSB01657	98 and 658	Q98C and N658C	SEQ ID NO: 48
39	gB-045; pSB01658	100 and 258	T100C and R258C	SEQ ID NO: 49
40	gB-046; pSB01659	100 and 656	T100C and L656C	SEQ ID NO: 50
41	gB-047; pSB01660	100 and 658	T100C and N658C	SEQ ID NO: 51
42	gB-048; pSB01661	117 and 406	I117C and T406C	SEQ ID NO: 52
43	gB-049; pSB01662	117 and 407	I117C and S407C	SEQ ID NO: 53
44	gB-050; pSB01663	153 and 712	Y153C and L712C	SEQ ID NO: 54
45	gB-051; pSB01664	162 and 716	L162C and M716C	SEQ ID NO: 55
46	gB-052; pSB01665	217 and 587	D217C and S587C	SEQ ID NO: 56
47	gB-053; pSB01666	217 and 589	D217C and Y589C	SEQ ID NO: 57
48	gB-054; pSB01667	219 and 584	S219C and F584C	SEQ ID NO: 58
49	gB-055; pSB01668	219 and 585	S219C and A585C	SEQ ID NO: 59
50	gB-056; pSB01669	219 and 586	S219C and N586C	SEQ ID NO: 60
51	gB-057; pSB01670	220 and 659	N220C and T659C	SEQ ID NO: 61
52	gB-058; pSB01671	223 and 659	S223C and T659C	SEQ ID NO: 62
53	gB-059; pSB01672	240 and 732	W240C and A732C	SEQ ID NO: 63
54	gB-060; pSB01673	240 and 735	W240C and G735C	SEQ ID NO: 64

55	gB-061; pSB01674	242 and 728	Y242C and V728C	SEQ ID NO: 65
56	gB-062; pSB01675	242 and 731	Y242C and G731C	SEQ ID NO: 66
57	gB-063	258 and 656	R258C and L656C	SEQ ID NO: 67
58	gB-064	269 and 656	S269C and L656C	SEQ ID NO: 68
59	gB-065	269 and 658	S269C and N658C	SEQ ID NO: 69
60	gB-066; pSB01679	272 and 614	D272C and P614C	SEQ ID NO: 70
61	gB-067; pSB01680	273 and 629	V273C and V629C	SEQ ID NO: 71
62	gB-068	349 and 650	W349C and A650C	SEQ ID NO: 72
63	gB-069; pSB01682	367 and 500	S367C and A500C	SEQ ID NO: 73
64	gB-070; pSB01683	367 and 503	S367C and A503C	SEQ ID NO: 74
65	gB-071; pSB01684	370 and 501	K370C and Q501C	SEQ ID NO: 75
66	gB-072; pSB01685	522 and 683	K522C and I683C	SEQ ID NO: 76
67	gB-073; pSB01686	523 and 683	I523C and I683C	SEQ ID NO: 77
68	gB-074; pSB01687	523 and 684	I523C and M684C	SEQ ID NO: 78
69	gB-075	524 and 684	N524C and M684C	SEQ ID NO: 79
70	gB-076	525 and 681	P525C and E681C	SEQ ID NO: 80
71	gB-077	540 and 680	R540C and L680C	SEQ ID NO: 81
72	gB-078	541 and 680	F541C and L680C	SEQ ID NO: 82
73	gB-079	548 and 655	L548C and P655C	SEQ ID NO: 83
74	gB-080	549 and 658	A549C and N658C	SEQ ID NO: 84
75	gB-081	550 and 655	S550C and P655C	SEQ ID NO: 85
76	gB-082	550 and 657	S550C and E657C	SEQ ID NO: 86
77	gB-083	591 and 668	Q591C and S668C	SEQ ID NO: 87

78	gB-084; pSB01697	603 and 667	L603C and Y667C	SEQ ID NO: 88
79	gB-085; pSB01698	604 and 672	G604C and L672C	SEQ ID NO: 89
80	gB-086; pSB01699	607 and 688	R607C and N688C	SEQ ID NO: 90
81	gB-087; pSB01700	608 and 692	T608C and Q692C	SEQ ID NO: 91
82	gB-088; pSB01701	609 and 691	E609C and K691C	SEQ ID NO: 92
83	gB-089; pSB01702	610 and 674	E610C and S674C	SEQ ID NO: 93
84	gB-090; pSB01703	610 and 675	E610C and S675C	SEQ ID NO: 94
85	gB-091; pSB01704	612 and 663	Q612C and V663C	SEQ ID NO: 95
86	gB-092; pSB01705	737 and 755	V737C and F755C	SEQ ID NO: 96
87	gB-093; pSB01706	741 and 754	V741C and A754C	SEQ ID NO: 97
88	gB-094; pSB01707	741 and 755	V741C and F755C	SEQ ID NO: 98
89		217 and 589	D217C and Y589C	
90		356 and 500	I356C and A500C	
91		367 and 500	S367C and A500C	
92		367 and 503	S367C and A503C	
93		371 and 505	M371C and A505C	
94		371 and W506	M371C and W506C	
95		374 and 503	T374C and A503C	
96		160 and 708	Y160C and Y708C	
97		162 and 716	L162C and M716C	
98		524 and 684	N524C and M684C	
99		99 and 267	G99C and A267C	

100		100 and 258	T100C and R258C	
101		221 and 657	T221C and E657C	
102		223 and 659	S223C and T659C	
103		541 and 681	F541C and E681C	
104		603 and 667	L603C and Y667C	
105		605 and 670	N605C and K670C	
106		607 and 688	R607C and N688C	
107		609 and 691	E609C and K691C	

In some embodiments, the polypeptide includes one or more (such as 2, 3, 4, 5, 6, 7, 8, 9 or 10) cysteine substitutions at any one of the positions listed in one or more of rows 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, or 36 of column (ii) of **Table 2**, wherein the resulting polypeptide does not exhibit an HCMV postfusion conformation.

In some embodiments, the polypeptide includes one or more (such as 2, 3, 4, 5, 6, 7, 8, 9 or 10) cysteine substitutions at any one of the positions listed in one or more of rows 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, or 88 of column (ii) of, wherein the resulting polypeptide does not exhibit an HCMV postfusion conformation.

In a preferred embodiment, the polypeptide includes cysteine substitutions at positions 98 and 653 (listed in row 2, column (ii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In another preferred embodiment, the polypeptide includes cysteine substitutions at positions 100 and 269 (listed in row 5, column (ii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a further preferred embodiment, the polypeptide includes cysteine substitutions at positions 217 and 584 (listed in row 7, column (ii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes cysteine substitutions at positions 242 and 710 (listed in row 13, column (ii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In another preferred embodiment, the polypeptide includes cysteine substitutions at positions 242 and 714 (listed in row 14, column (ii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a further preferred embodiment, the

- polypeptide includes cysteine substitutions at positions 367 and 499 (listed in row 17, column (ii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes cysteine substitutions at positions 372 and 506 (listed in row 18, column (ii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In another preferred embodiment, the polypeptide includes cysteine substitutions at positions 550 and 652 (listed in row 22, column (ii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a further preferred embodiment, the polypeptide includes cysteine substitutions at positions 608 and 679 (listed in row 26, column (ii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes cysteine substitutions at positions 695 and 724 (listed in row 34, column (ii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB.
- 15 In some embodiments, the polypeptide includes one or more (such as 2, 3, 4, 5, 6, 7, 8, 9 or 10) disulfide bonds between pairs of cysteine residues substituted at any one of the pairs of positions listed in one or more of rows 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 20 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, and 88 of column (ii) of **Table 2**.
- In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 98 and 653 (listed in row 2, column (ii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In another preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 100 and 269 (listed in row 5, column (ii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a further preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 217 and 30 584 (listed in row 7, column (ii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 242 and 710 (listed in row 13, column (ii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In another preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 242 and 714 (listed in row 14, column (ii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a further preferred embodiment, the polypeptide includes a disulfide bond between a

pair of cysteine residues substituted at positions 367 and 499 (listed in row 17, column (ii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 372 and 506 (listed in row 18, column (ii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In another preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 550 and 652 (listed in row 22, column (ii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a further preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 608 and 679 (listed in row 26, column (ii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 695 and 724 (listed in row 34, column (ii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB.

In further embodiments, the polypeptide includes one or more (such as 2, 3, 4, 5, 6, 7, 8, 9 or 10) disulfide bonds between pairs of cysteine residues that are introduced by cysteine amino acid substitutions at any one of the pairs of positions listed in one or more of rows 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, or 36 of column (iii) of **Table 2**, wherein the polypeptide does not exhibit an HCMV postfusion conformation.

In further embodiments, the polypeptide includes one or more (such as 2, 3, 4, 5, 6, 7, 8, 9 or 10) disulfide bonds between pairs of cysteine residues that are introduced by cysteine amino acid substitutions at any one of the pairs of positions listed in one or more of rows 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, or 88 of column (iii) of **Table 2**, wherein the polypeptide does not exhibit an HCMV postfusion conformation.

In a preferred embodiment, the polypeptide includes cysteine substitutions at Q98C and I653C (listed in row 2, column (iii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In another preferred embodiment, the polypeptide includes cysteine substitutions at T100C and S269C (listed in row 5, column (iii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a further preferred embodiment, the polypeptide includes cysteine substitutions at D217C and F584C (listed in row 7, column (iii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-

type HCMV gB. In a preferred embodiment, the polypeptide includes cysteine substitutions at Y242C and K710C (listed in row 13, column (iii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In another preferred embodiment, the polypeptide includes cysteine substitutions at Y242C and D714C (listed in row 14, column (iii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a further preferred embodiment, the polypeptide includes cysteine substitutions at S367C and L499C (listed in row 17, column (iii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes cysteine substitutions at T372C and W506C (listed in row 18, column (iii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In another preferred embodiment, the polypeptide includes cysteine substitutions at S550C and D652C (listed in row 22, column (iii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a further preferred embodiment, the polypeptide includes cysteine substitutions at T608C and D679C (listed in row 26, column (iii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes cysteine substitutions at K695C and K724C (listed in row 34, column (iii) of **Table 2**) according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB.

In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 96 and 660 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In another preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 98 and 658 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a further preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 100 and 258 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 100 and 656 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In another preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 100 and 658 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a further preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 117 and 406 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues

substituted at positions 117 and 407 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 153 and 712 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 162 and 716 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 217 and 587 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 217 and 589 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 219 and 584 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 219 and 585 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 219 and 586 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 220 and 659 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 223 and 659 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 240 and 732 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 240 and 735 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 242 and 728 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at

positions 242 and 731 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 258 and 656 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 269 and 656 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 269 and 658 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 272 and 614 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 273 and 629 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 349 and 650 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 367 and 500 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 367 and 503 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 370 and 501 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 522 and 683 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 523 and 683 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 523 and 684 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 524 and 684

according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 525 and 681 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 540 and 680 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 541 and 680 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 548 and 655 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 549 and 658 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 550 and 655 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 550 and 657 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 591 and 668 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 603 and 667 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 604 and 672 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 607 and 688 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 608 and 692 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 609 and 691 according to the numbering of

SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 610 and 674 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the

5 polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 610 and 675 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 612 and 663 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-

10 type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 737 and 755 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 741 and 754 according to the numbering of SEQ ID NO: 1, relative to

15 the amino acid sequence of the wild-type HCMV gB. In a preferred embodiment, the polypeptide includes a disulfide bond between a pair of cysteine residues substituted at positions 741 and 755 according to the numbering of SEQ ID NO: 1, relative to the amino acid sequence of the wild-type HCMV gB.

In some embodiments, the polypeptide includes a combination of two or more of the

20 disulfide bonds between cysteine residues listed in **Table 2**.

In some embodiments, the polypeptide includes an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to any sequence selected from: SEQ ID NO: 2; SEQ ID NO: 3; SEQ ID NO: 4; SEQ ID NO: 5; SEQ ID NO: 6; SEQ ID NO: 7; SEQ ID NO: 8; SEQ ID

25 NO: 9; SEQ ID NO: 10; SEQ ID NO: 11; SEQ ID NO: 12; SEQ ID NO: 13; SEQ ID NO: 14; SEQ ID NO: 15; SEQ ID NO: 16; SEQ ID NO: 17; SEQ ID NO: 18; SEQ ID NO: 19; SEQ ID NO: 20; SEQ ID NO: 21; SEQ ID NO: 22; SEQ ID NO: 23; SEQ ID NO: 24; SEQ ID NO: 25; SEQ ID NO: 26; SEQ ID NO: 27; SEQ ID NO: 28; SEQ ID NO: 29; SEQ ID NO: 30; SEQ ID NO: 31; SEQ ID NO: 32; SEQ ID NO: 33; SEQ ID NO: 34; SEQ ID NO: 35; SEQ ID NO: 36; and SEQ ID NO: 37.

30 In some embodiments, the polypeptide includes an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to any sequence selected from: SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO:

35 59, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70, SEQ ID NO: 71, SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 75, SEQ ID NO:

76, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 81, SEQ ID NO: 82, SEQ ID NO: 83, SEQ ID NO: 84, SEQ ID NO: 85, SEQ ID NO: 86, SEQ ID NO: 87, SEQ ID NO: 88, SEQ ID NO: 89, SEQ ID NO: 90, SEQ ID NO: 91, SEQ ID NO: 92, SEQ ID NO: 93, SEQ ID NO: 94, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 97, and SEQ ID NO: 98.

5 In some embodiments, the polypeptide includes an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, preferably 99%, or 100% identity to any sequence selected from SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, and SEQ ID NO: 60.

10 In some embodiments, the polypeptide includes an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, preferably 99%, or 100% identity to any sequence selected from SEQ ID NO: 51, SEQ ID NO: 73, SEQ ID NO: 70, and SEQ ID NO: 78

15 In some embodiments, the composition preferably does not include a polypeptide having the sequence set forth in any one of SEQ ID NO: 59, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 71, SEQ ID NO: 52, SEQ ID NO: 96, and SEQ ID NO: 50.

In additional embodiments, the polypeptide includes the amino acid sequence as set forth in any one of the SEQ ID NOs listed in column (iv) of **Table 2**. That is, an exemplary polypeptide includes a polypeptide having the amino acid sequence selected from any one of: SEQ ID NO: 2; SEQ ID NO: 3; SEQ ID NO: 4; SEQ ID NO: 5; SEQ ID NO: 6; SEQ ID NO: 7; SEQ ID NO: 8; SEQ ID NO: 9; SEQ ID NO: 10; SEQ ID NO: 11; SEQ ID NO: 12; SEQ ID NO: 13; SEQ ID NO: 14; SEQ ID NO: 15; SEQ ID NO: 16; SEQ ID NO: 17; SEQ ID NO: 18; SEQ ID NO: 19; SEQ ID NO: 20; SEQ ID NO: 21; SEQ ID NO: 22; SEQ ID NO: 23; SEQ ID NO: 24; SEQ ID NO: 25; SEQ ID NO: 26; SEQ ID NO: 27; SEQ ID NO: 28; SEQ ID NO: 29; SEQ ID NO: 30; SEQ ID NO: 31; SEQ ID NO: 32; SEQ ID NO: 33; SEQ ID NO: 34; SEQ ID NO: 35; SEQ ID NO: 36; and SEQ ID NO: 37. In some embodiments, the polypeptide has the amino acid sequence selected from any one of SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70, SEQ ID NO: 71, SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 75, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 81, SEQ ID NO: 82, SEQ ID NO: 83, SEQ ID NO: 84, SEQ ID NO: 85, SEQ ID NO: 86, SEQ ID NO: 87, SEQ ID NO: 88, SEQ ID NO: 89, SEQ ID NO: 90, SEQ ID NO: 91, SEQ ID NO: 92, SEQ ID NO: 93, SEQ ID NO: 94, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 97, and SEQ ID NO: 98.

In a preferred embodiment, the polypeptide includes the amino acid sequence as set forth in any one of SEQ ID NO: 3; SEQ ID NO: 6; SEQ ID NO: 8; SEQ ID NO: 14; SEQ ID NO: 15; SEQ ID NO: 18; SEQ ID NO: 19; SEQ ID NO: 23; SEQ ID NO: 27; and SEQ ID NO: 35.

In some embodiments, amino acids can be inserted (or deleted) from the native HCMV gB sequence to adjust the alignment of residues in the polypeptide structure, such that particular residue pairs are within a sufficiently close distance to form a disulfide bond in the prefusion, but not postfusion, conformation. In several such embodiments, the polypeptide

5 includes a disulfide bond between cysteine residues located at any of the pairs of positions listed in one or more of rows 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, or 88 of column (ii) of **Table 2**,

10 in addition to including at least one amino acid insertion.

In some embodiments, the polypeptide includes a phenylalanine substitution as compared to a native HCMV gB. In some embodiments, the polypeptide includes a leucine substitution as compared to a native HCMV gB. In some embodiments, the polypeptide may be stabilized by amino acid mutations (such as, for example, phenylalanine (F) and leucine (L)

15 substitutions) that decrease ionic repulsion between residues that are proximate to each other in the folded structure of the polypeptide, as compared to a HCMV gB polypeptide in postfusion conformation. In some embodiments, the polypeptide may be stabilized by amino acid mutations that increase ionic attraction between residues that are proximate to each other in the folded structure of the polypeptide, as compared to a HCMV gB in postfusion conformation.

20 Exemplary mutations include any mutation selected from **Table 3**, according to the numbering of SEQ ID NO: 1 as compared to a native HCMV gB.

Table 3. Exemplary Phenylalanine (F) and Leucine (L) Substitutions

Row	(i) Mutant ID	(ii) Mutated residue position, according to the numbering of SEQ ID NO: 1	(iii) Substitutions corresponding to SEQ ID NO: 1	(iv) Exemplary sequence that includes the mutation is set forth in:
1	gB-037	670	K670L	SEQ ID NO: 38
2	gB-038	670	K670F	SEQ ID NO: 39
3	gB-039	673	R673L	SEQ ID NO: 40
4	gB-040	673	R673F	SEQ ID NO: 41
5	gB-041	691	K691L	SEQ ID NO: 42
6	gB-042	691	K691F	SEQ ID NO: 43
7		686	E686L	
8		354	R354F	
9		573	R573F	
10		101 and 260	D101L and K260L	

5

Table 4. Further exemplary substitutions

Row	(i) Mutant ID	(ii) Mutated residue position, according to the numbering of SEQ ID NO: 1	(iii) Substitutions corresponding to SEQ ID NO: 1	(iv) Exemplary sequence that includes the mutation is set forth in:
1	gB-095	679	D679S	SEQ ID NO: 99
2	gB-096	679	D679N	SEQ ID NO: 100
3	gB-097	682	E682S	SEQ ID NO: 101
4	gB-098	682	E682Q	SEQ ID NO: 102
5	gB-099	686	E686S	SEQ ID NO: 103
6	gB-100	686	E686Q	SEQ ID NO: 104
7	gB-101	118	N118P	SEQ ID NO: 105
8	gB-102	646	D646P	SEQ ID NO: 106
9		686	E686L	
10		686	E686I	
11		679	D679A	
12		354	R354F	
13		573	R573F	
14		101 and 260	D101L and K260L	
15		217 and 589	D217C and Y589C	
16		217, 589, 703, and 704	D217C, Y589C, D703C, and P704C	
17		217, 589, 696, and 697	D217C, Y589C, Y696C, and V697C	
18		217 and 589	D217C and Y589C	
19		217 and 589	D217C and Y589C	
20		217 and 589	D217C and Y589C	
21		217 and 589	D217C and Y589C	
22		217, 371, 506, and 589	D217C, M371C, W506C, and Y589C	
23		217, 524, 589, and 684	D217C, N524C, Y589C, M684C	
24		217, 524, 589, 684, 703, and 704	D217C, N524C, Y589C, M684C, D703C, and P704C	

In some embodiments, the polypeptide includes one or more (such as 2, 3, 4, 5, 6, 7, 8, 9 or 10) residues substituted at any one of the positions listed in one or more of rows 1, 2, 3, 4, 5, or 6 of column (ii) of **Table 3**, wherein the polypeptide does not exhibit an HCMV postfusion conformation.

In some embodiments, the polypeptide includes one or more (such as 2, 3, 4, 5, 6, 7, 8, 9 or 10) residues substituted at any one of the positions listed in one or more of rows 1, 2, 3, 4, 5, 6, 7, and 8 of column (ii) of **Table 4**, wherein the polypeptide does not exhibit an HCMV postfusion conformation.

In some embodiments, the polypeptide includes a mutation at position 670 (listed in rows 1 and 2, column (ii) of **Table 3**) according to the numbering of SEQ ID NO: 1. In some embodiments, the polypeptide includes a mutation at position 673 (listed in rows 3 and 4, column (ii) of **Table 3**) according to the numbering of SEQ ID NO: 1. In some embodiments, the polypeptide includes a mutation at position 691 (listed in rows 5 and 6, column (ii) of **Table 3**) according to the numbering of SEQ ID NO: 1.

In some embodiments, the polypeptide includes a mutation at position 670 according to the numbering of SEQ ID NO: 1. In some embodiments, the polypeptide includes a mutation at position 682 according to the numbering of SEQ ID NO: 1. In some embodiments, the polypeptide includes a mutation at position 686 according to the numbering of SEQ ID NO: 1. In some embodiments, the polypeptide includes a mutation at position 118 according to the numbering of SEQ ID NO: 1. In some embodiments, the polypeptide includes a mutation at position 646 according to the numbering of SEQ ID NO: 1.

In further embodiments, the polypeptide includes an electrostatic mutation that is introduced by substitutions at any one of the positions listed in one or more of rows 1, 2, 3, 4, 5, or 6 of column (iii) of **Table 3**, wherein the polypeptide does not exhibit an HCMV postfusion conformation.

In a preferred embodiment, the polypeptide includes a substitution K670L (listed in row 1, column (iii) of **Table 3**) according to the numbering of SEQ ID NO: 1. In another preferred embodiment, the polypeptide includes a substitution K670F (listed in row 2, column (iii) of **Table 3**) according to the numbering of SEQ ID NO: 1. In a further preferred embodiment, the polypeptide includes a substitution R673L (listed in row 3, column (iii) of **Table 3**) according to the numbering of SEQ ID NO: 1. In a preferred embodiment, the polypeptide includes a substitution R673F (listed in row 4, column (iii) of **Table 3**) according to the numbering of SEQ ID NO: 1. In another preferred embodiment, the polypeptide includes a substitution K691L (listed in row 5, column (iii) of **Table 3**) according to the numbering of SEQ ID NO: 1. In a further preferred embodiment, the polypeptide includes a substitution K691F (listed in row 6, column (iii) of **Table 3**) according to the numbering of SEQ ID NO: 1.

In some embodiments, the polypeptide includes a combination of two or more of the phenylalanine (F) and leucine (L) substitutions listed in **Table 3**.

In a preferred embodiment, the polypeptide includes a substitution D679S according to the numbering of SEQ ID NO: 1. In another preferred embodiment, the polypeptide includes a substitution D679N according to the numbering of SEQ ID NO: 1. In another preferred embodiment, the polypeptide includes a substitution E682S according to the numbering of SEQ ID NO: 1. In another preferred embodiment, the polypeptide includes a substitution E682Q according to the numbering of SEQ ID NO: 1. In another preferred embodiment, the polypeptide includes a substitution E686S according to the numbering of SEQ ID NO: 1. In another preferred embodiment, the polypeptide includes a substitution E686Q according to the numbering of SEQ ID NO: 1. In another preferred embodiment, the polypeptide includes a substitution N118P according to the numbering of SEQ ID NO: 1. In another preferred embodiment, the polypeptide includes a substitution D646P according to the numbering of SEQ ID NO: 1.

In some embodiments, the polypeptide includes a combination of two or more of the phenylalanine (F) and leucine (L) substitutions listed in **Table 4**. In some embodiments, the polypeptide includes an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to any sequence selected from: SEQ ID NO: 38; SEQ ID NO: 39; SEQ ID NO: 40; SEQ ID NO: 41; SEQ ID NO: 42; and SEQ ID NO: 43.

In some embodiments, the polypeptide includes an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to any sequence selected from: SEQ ID NO: 99; SEQ ID NO: 100; SEQ ID NO: 102; SEQ ID NO: 103; SEQ ID NO: 104; SEQ ID NO: 105, and SEQ ID NO: 106.

In additional embodiments, the polypeptide includes the amino acid sequence as set forth in any one of the SEQ ID NOs listed in column (iv) of **Table 3**. That is, an exemplary polypeptide includes a polypeptide having the amino acid sequence selected from any one of: SEQ ID NO: 38; SEQ ID NO: 39; SEQ ID NO: 40; SEQ ID NO: 41; SEQ ID NO: 42; and SEQ ID NO: 43. In some embodiments, the polypeptide has the amino acid sequence selected from any one of: SEQ ID NO: 99; SEQ ID NO: 100; SEQ ID NO: 102; SEQ ID NO: 103; SEQ ID NO: 104; SEQ ID NO: 105, and SEQ ID NO: 106.

In some embodiments, amino acids can be inserted (or deleted) from the native HCMV gB sequence to adjust the alignment of residues in the polypeptide structure, such that particular residue pairs are within a sufficiently close distance to form a desired electrostatic interaction in the prefusion, but not postfusion, conformation. In several such embodiments, the polypeptide includes a desired electrostatic interaction at any of the positions listed in one or

more of rows 1, 2, 3, 4, 5, or 6 of column (ii) of **Table 3**, wherein the polypeptide does not exhibit an HCMV postfusion conformation.

5

C.2. Further Embodiments of the polypeptide

In some embodiments, the polypeptide does not include a mutation at any one of the following amino acid positions: 280, 281, 283, 284, 285, 286, 290, 292, 295, 297, 298, 299, or any combinations thereof, according to the numbering of reference sequence SEQ ID NO: 46.

10 In some exemplary embodiments, the polypeptide does not include a substitution of any one of the following residues, according to the numbering of reference sequence SEQ ID NO: 46: Y280; N281; T283; N284; R285; N286; F290; E292; N293; F297; F298; I299; F298; and any combinations thereof. Without being bound by theory or mechanism, residues important for neutralizing antibodies may include Y280/N284 and Y280/N293/D295. Accordingly, in a
15 preferred embodiment, the polypeptide does not include mutations at Y280, N293, N284, and D295, as compared to reference sequence SEQ ID NO: 46.

In some embodiments, the polypeptide does not include a mutation at any one of the following amino acid positions: R562, P577, S587, Y588, G592, G595, L601/H605, C610, L612, P613, Y625, Y627, F632, and K633, and any combinations thereof, according to the numbering
20 of reference sequence SEQ ID NO: 44. In some embodiments, the polypeptide does not include any one of the following amino acid mutations: R562C, P577L, S587L, Y588C, G592S, G595D, L601P/H605N, C610Y, L612F, P613Y, Y625C, Y627C, F632L, and K633T, or any combinations thereof, according to the numbering of reference sequence SEQ ID NO: 44. Without being bound by theory or mechanism, P577 and Y627 are believed to be located next
25 to each other within the domain IV core while C610 participates in a conserved disulfide bond. Thus, all three residues may help maintain the position of domain IV in the prefusion structure and, therefore, the stability of entire antigenic site AD-1. Moreover, without being bound by theory or mechanism, F632 and G595 are believed to be exposed on the surface of the prefusion form of gB. Accordingly, in a preferred embodiment, the polypeptide does not include
30 a mutation at P577, Y627, C610, F632, and G595, or any combinations thereof, according to the numbering of reference sequence SEQ ID NO: 44.

C.3. Cavity filling mutations

In still other embodiments, the polypeptide includes amino acid mutations that are one
35 or more cavity filling mutations. Examples of amino acids that may be replaced with the goal of cavity filling include small aliphatic (e.g. Gly, Ala, and Val) or small polar amino acids (e.g. Ser and Thr) and amino acids that are buried in the pre-fusion conformation, but exposed to solvent

in the post-fusion conformation. Examples of the replacement amino acids include large aliphatic amino acids (Ile, Leu and Met) or large aromatic amino acids (His, Phe, Tyr and Trp).

C.4. Combination of Mutations

5 In another aspect, the present invention relates to a polypeptide that includes a combination of two or more different types of mutations selected from engineered disulfide bond mutations, cavity filling mutations, and electrostatic mutations, each as described herein above. In some embodiments, the polypeptide includes at least one disulfide bond mutation and at least electrostatic mutation. More specifically, in some embodiments, the polypeptide includes
10 at least one cysteine substitution and at least one phenylalanine substitution. In some embodiments, the polypeptide includes at least one cysteine substitution and at least one leucine substitution.

In some further embodiments, the polypeptide includes at least one mutation selected from any one of the mutations in **Table 2** and at least one mutation selected from any one of the
15 mutations in **Table 3**. In some further embodiments, the polypeptide includes at least one mutation selected from any one of the mutations in **Table 2** and at least one mutation selected from any one of the mutations in **Table 4**. In some further embodiments, the polypeptide includes at least one mutation selected from any one of the mutations in **Table 3** and at least one mutation selected from any one of the mutations in **Table 4**.

20

D. Druggable Regions of HCMV gB polypeptide

The present invention provides, in part, novel druggable regions in the HCMV gB protein. The interaction of a drug with such regions, or the modulation of the activity of such regions with a drug, could inhibit viral fusion and hence viral infectivity. In one aspect, the
25 present invention provides methods of screening compounds against these druggable regions in order to discover a candidate therapeutic for HCMV infection caused by HCMV having gB protein, for example such as a small molecule viral fusion inhibitor. In one embodiment, a method for identifying a candidate therapeutic for a HCMV infection caused by HCMV having gB protein comprises contacting a gB protein which comprises a druggable region with a
30 compound, wherein binding of said compound indicates a candidate therapeutic. Compounds may in certain embodiments be selected from the following classes: peptides, polypeptides, peptidomimetics, or small molecules, or may be selected from a library of compounds. Such a library may be generated by combinatorial synthetic methods. Binding may be assayed either in vitro or in vivo. In certain embodiments of this method, the protein is HCMV gB protein and
35 comprises at least one residue, preferably three residues, from a druggable region of HCMV gB protein. Such druggable regions also may be utilized in the structure determination, drug screening, drug design, and other methods described and claimed herein.

The term “druggable region”, when used in reference to a polypeptide, nucleic acid, complex and the like, refers to a region of HCMV gB protein which is a target or is a likely target for binding an agent that reduces or inhibits viral infectivity. For a polypeptide, a druggable region generally refers to a region wherein several amino acids of a polypeptide would be

5 capable of interacting with an agent. For a polypeptide or complex thereof, exemplary druggable regions include binding pockets and sites, interfaces between domains of a polypeptide or complex, surface grooves or contours or surfaces of a polypeptide or complex which are capable of participating in interactions with another molecule, such as a cell membrane.

10 A druggable region may be described and characterized in a number of ways. For example, a druggable region may be characterized by some or all of the amino acids that make up the region, or the backbone atoms thereof, or the side chain atoms thereof (optionally with or without Ca atoms). Alternatively, a druggable region may be characterized by comparison to other regions on the same or other molecules. For example, the term “affinity region” refers to a

15 druggable region on a molecule (such as a HCMV gB protein) that is present in several other molecules, in so much as the structures of the same affinity regions are sufficiently the same so that they are expected to bind the same or related structural analogs. An example of an affinity region is an ATP-binding site of a protein kinase that is found in several protein kinases (whether or not of the same origin). The term “selectivity-region” refers to a druggable region of

20 a molecule that may not be found on other molecules, in so much as the structures of different selectivity regions are sufficiently different so that they are not expected to bind the same or related structural analogs. An exemplary selectivity region is a catalytic domain of a protein kinase that exhibits specificity for one substrate. In certain instances, a single modulator may bind to the same affinity region across a number of proteins that have a substantially similar

25 biological function, whereas the same modulator may bind to only one selectivity region of one of those proteins.

Continuing with examples of different druggable regions, the term “undesired region” refers to a druggable region of a molecule that upon interacting with another molecule results in an undesirable effect. For example, a binding site that oxidizes the interacting molecule (such

30 as P-450 activity) and thereby results in increased toxicity for the oxidized molecule may be deemed a “undesired region”. Other examples of potential undesired regions include regions that upon interaction with a drug decrease the membrane permeability of the drug, increase the excretion of the drug, or increase the blood brain transport of the drug. It may be the case that, in certain circumstances, an undesired region will no longer be deemed an undesired region

35 because the affect of the region will be favorable, e.g., a drug intended to treat a brain condition would benefit from interacting with a region that resulted in increased blood brain transport, whereas the same region could be deemed undesirable for drugs that were not intended to be delivered to the brain.

When used in reference to a druggable region, the "selectivity" or "specificity" of a molecule such as a modulator to a druggable region may be used to describe the binding between the molecule and a druggable region. For example, the selectivity of a modulator with respect to a druggable region may be expressed by comparison to another modulator, using the
5 respective values of K_d (i.e., the dissociation constants for each modulator-druggable region complex) or, in cases where a biological effect is observed below the K_d , the ratio of the respective EC_{50} 's (i.e., the concentrations that produce 50% of the maximum response for the modulator interacting with each druggable region).

The term "modulation", when used in reference to a functional property or biological
10 activity or process (e.g., enzyme activity or receptor binding), refers to the capacity to either up regulate (e.g., activate or stimulate), down regulate (e.g., inhibit or suppress) or otherwise change a quality of such property, activity or process. In certain instances, such regulation may be contingent on the occurrence of a specific event, such as activation of a signal transduction pathway, and/or may be manifest only in particular cell types.

The term "modulator" refers to a peptide, polypeptide, nucleic acid, macromolecule,
15 complex, molecule, small molecule, compound, species or the like (naturally-occurring or non-naturally-occurring), or an extract made from biological materials such as bacteria, plants, fungi, or animal cells or tissues, that may be capable of causing modulation. Modulators may be evaluated for potential activity as modulators or activators (directly or indirectly) of a functional
20 property, biological activity or process, or combination of them, (e.g., agonist, partial antagonist, partial agonist, inverse agonist, antagonist, anti-microbial agents, modulators of microbial infection or proliferation, and the like) by inclusion in assays. In such assays, many modulators may be screened at one time. The activity of a modulator may be known, unknown or partially known. The term "inhibitor" refers to a peptide, polypeptide, nucleic acid, macromolecule,
25 complex, molecule, small molecule, compound, species or the like (naturally-occurring or non-naturally-occurring), or an extract made from biological materials such as bacteria, plants, fungi, or animal cells or tissues, that may be capable of down-regulating or suppressing a functional property or biological activity or process.

The term "motif" refers to an amino acid sequence that is commonly found in a protein of
30 a particular structure or function. Typically, a consensus sequence is defined to represent a particular motif. The consensus sequence need not be strictly defined and may contain positions of variability, degeneracy, variability of length, etc. The consensus sequence may be used to search a database to identify other proteins that may have a similar structure or function due to the presence of the motif in its amino acid sequence. For example, on-line databases
35 may be searched with a consensus sequence in order to identify other proteins containing a particular motif. Various search algorithms and/or programs may be used, including FASTA, BLAST or ENTREZ. FASTA and BLAST are available as a part of the GCG sequence analysis package (University of Wisconsin, Madison, Wis.). ENTREZ is available through the National

Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Md.

The term "small molecule" refers to a compound, which has a molecular weight of less than about 5 kD, less than about 2.5 kD, less than about 1.5 kD, or less than about 0.9 kD.

- 5 Small molecules may be, for example, nucleic acids, peptides, polypeptides, peptide nucleic acids, peptidomimetics, carbohydrates, lipids or other organic (carbon containing) or inorganic molecules. Many pharmaceutical companies have extensive libraries of chemical and/or biological mixtures, often fungal, bacterial, or algal extracts, which can be screened with any of the assays of the invention. The term "small organic molecule" refers to a small molecule that is
- 10 often identified as being an organic or medicinal compound, and does not include molecules that are exclusively nucleic acids, peptides or polypeptides.

- In one embodiment, the druggable region comprises residues K130-A135, D216-W233, R258-K260, A267-V273, R327-D329, W349-E350, V480-K518 and N676-Y690 of SEQ ID NO:1, which residues form the binding pocket for Domain V of HCMV gB protein in the
- 15 postfusion conformation.

In yet another embodiment, a druggable region comprises a fusion loop or a portion thereof.

- In another aspect, the present invention is directed towards methods for identifying a candidate therapeutic for a disease caused by HCMV having gB protein. In certain
- 20 embodiments, such methods comprise contacting the gB protein which comprises a druggable region with a compound, wherein the modulation of the activity of said gB protein indicates a candidate therapeutic. In other embodiments, such methods comprise contacting the gB protein which comprises a druggable region with a compound, wherein the preclusion of the movement or interaction of said druggable region indicates a candidate therapeutic. In still other
- 25 embodiments, the modulation of the function or activity of said gB protein involves precluding the completion of the post-fusion conformational change. In yet another embodiment, the modulation of the function or activity of said gB protein involves interfering with the first stage of the conformational change. In another embodiment, a method for identifying a candidate therapeutic for a disease caused by infection with HCMV having gB protein comprises
- 30 contacting the gB protein which comprises a druggable region with a compound, wherein the inhibition of fusion in said virus indicates a candidate therapeutic. In yet another embodiment, a method for identifying a candidate therapeutic for a disease caused by infection with HCMV having gB protein, comprising contacting the gB protein which comprises a druggable region with a compound, wherein the inhibition of viral infectivity of said HCMV indicates a candidate
- 35 therapeutic. In still another embodiment, a method for identifying a candidate therapeutic for disease caused by infection with HCMV having gB protein comprises contacting the gB protein which comprises a druggable region with a compound, wherein the reduction of at least one symptom of said disease in a subject indicates a candidate therapeutic.

In a further embodiment, the invention provides a method for identifying a candidate therapeutic for disease caused by infection with HCMV having gB protein comprising contacting the gB protein which comprises a druggable region with a compound, wherein the compound prevents Domain V of the gB protein or a fragment thereof from binding in its binding pocket. In one embodiment, the binding pocket for Domain V of the gB protein comprises residues K130-A135, D216-W233, R258-K260, A267-V273, R327-D329, W349-E350, V480-K518 and N676-Y690 of SEQ ID NO:1.

In another aspect, all of the information learned and described herein about the gB protein may be used in methods of designing modulators of one or more of their biological activities. In one embodiment, a method for designing a modulator for the prevention or treatment of a disease caused by infection with HCMV having gB protein, comprises: (a) providing a three-dimensional structure for gB protein; (b) identifying a potential modulator for the prevention or treatment of disease caused by HCMV having gB protein by reference to the three-dimensional structure; (c) contacting the gB protein with the potential modulator; and (d) assaying the activity of the gB protein or determining the viability of the virus having said gB protein after contact with the modulator, wherein a change in the activity of the polypeptide or the viability of the virus indicates that the modulator may be useful for prevention or treatment of a virus-related disease or disorder. In certain embodiments, the potential modulator is identified by reference to the three-dimensional structure of HCMV having gB protein. In other embodiments, the potential modulator is identified by reference to the three-dimensional structure comprising a druggable region or fragment of the gB protein.

In yet another aspect, the present invention provides modulators (in certain embodiments, inhibitors) of gB protein activity, as well as pharmaceutical compositions and kits comprising the same. Such modulators may in certain embodiments interact with a druggable region of the invention. In still another aspect, the present invention is directed toward a modulator that is a fragment of (or homolog of such fragment or mimetic of such fragment) the druggable region of a HCMV gB protein and competes with that druggable region. Modulators of any of the above-described druggable regions may be used alone or in complementary approaches to treat infection by HCMV.

Finally, the present invention is directed toward methods of identifying and designing modulators which bind with, interact with, or modulate the function or activity of an active or binding site of a HCMV gB polypeptide.

As used herein, the term "compound" shall include, but not be limited to, a modulator, an inhibitor, a therapeutic, a therapeutic drug, a prophylactic, a drug, or an agent. As used herein, the term "candidate therapeutic" (also known as a "candidate agent" or "test agent") shall include, but not be limited to, a compound, detergents, proteins, peptides, peptidomimetics, antibodies, nucleic acids, small molecules, cytokines, or hormones.

Set forth below are aspects of drug discovery useful for the practice of this invention by one of skill in the art.

D.1. Druggable Regions

5 In one embodiment, the druggable region comprises residues K130-A135, D216-W233, R258-K260, A267-V273, R327-D329, W349-E350, V480-K518 and N676-Y690 of SEQ ID NO:1 or a fragment thereof. These residues form the binding pocket for Domain V of the gB protein or a portion thereof in the postfusion conformation. As used herein, the terms “binding pocket” and “binding groove” shall have the same meaning and are interchangeable.

10 In another embodiment, a druggable region comprises a fusion loop or a portion thereof.

In yet another aspect, the present invention is directed toward methods of identifying and designing modulators which bind with, interact with, or modulate the function or activity of an active or binding site of a HCMV gB protein.

15 **D.2. Modulators, Modulator Design and Screening Using the Subject Druggable Regions**

In one aspect, the present invention provides methods of screening for potential modulators of the subject druggable regions, as well as methods of designing such modulators. Modulators to HCMV gB polypeptides of the invention and other structurally related molecules, and complexes containing the same, may be identified and developed as set forth below and otherwise using techniques and methods known to those of skill in the art. The modulators of the invention may be employed, for instance, to inhibit and treat disease caused by HCMV.

In one aspect, the present invention is directed towards a modulator that interacts with the subject druggable regions so as to reduce the activity of the HCMV. Such modulators may in certain embodiments interact with a druggable region of the virus. In certain embodiments, a modulator interacts with the binding pocket for Domain V of the gB polypeptide so as to preclude Domain V from binding in the pocket, thereby modulating the activity of the HCMV. In still another aspect, the present invention is directed toward a modulator that is a fragment of Domain V of the HCMV gB protein (or homolog of such fragment or mimetic of such fragment) and competes with the druggable region, i.e. the binding pocket for Domain V. In one aspect, modulators comprise fragments of Domain V which are selected from the group consisting of fragments having residues (Towne strain) M648-K700, M648-V697, S647-V697, S647-V663, I653-V697, I653-Q692, I653-L680, I653-S675, I653-Y667, R662-V697, R662-Q692, R662-L680, R662-S675, L664-F678, S668-V697, S668-Q692, S668-V677, D679-V697, L680-V697, L680-Q692, and R693-K700 of SEQ ID NO: 1. More specifically, the modulators are fragments of SEQ ID NO:1 having residues M648-K700 (SEQ ID NO: 260), M648-V697 (SEQ ID NO: 261), S647-V697 (SEQ ID NO: 262), S647-V663 (SEQ ID NO: 263), I653-V697 (SEQ ID NO: 264), I653-Q692 (SEQ ID NO: 265), I653-L680 (SEQ ID NO: 266), I653-S675 (SEQ ID NO: 267), I653-Y667 (SEQ ID NO: 268), R662-V697 (SEQ ID NO: 269), R662-Q692 (SEQ ID NO: 270), R662-L680 (SEQ ID NO: 271), L664-F678 (SEQ ID NO: 272), S668-V697 (SEQ ID NO: 273), S668-Q692 (SEQ ID NO: 274), S668-V677 (SEQ ID NO: 275), D679-V697 (SEQ ID NO: 276), L680-V697 (SEQ ID NO: 277), L680-Q692 (SEQ ID NO: 278), and R693-K700 (SEQ ID NO: 279) of SEQ ID NO: 1.

267), I653-Y667 (SEQ ID NO: 268), R662-V697 (SEQ ID NO: 269), R662-Q692 (SEQ ID NO: 270), R662-L680 (SEQ ID NO: 271), R662-S675 (SEQ ID NO: 272), L664-F678 (SEQ ID NO: 273), S668-V697 (SEQ ID NO: 274), S668-Q692 (SEQ ID NO: 275), S668-V677 (SEQ ID NO: 276), D679-V697 (SEQ ID NO: 277), L680-V697 (SEQ ID NO: 278), L680-Q692 (SEQ ID NO: 279), and R693-K700 (SEQ ID NO: 280).

Modulators of any of the above-described druggable regions may be used alone or in complementary approaches to treat or prevent HCMV infections.

A variety of methods for inhibiting the growth or infectivity of HCMV using the modulators are contemplated by the present invention. For example, exemplary methods involve contacting a HCMV gB protein with a modulator thought or shown to be effective against such pathogen.

For example, in one aspect, the present invention contemplates a method for treating a subject suffering from an infection of HCMV comprising administering to the subject an amount of a modulator effective to modulate the expression and/or activity of HCMV. The present invention further contemplates a method for treating a subject suffering from an HCMV infection, comprising administering to the subject having the infection a therapeutically effective amount of a molecule identified using one of the methods of the present invention.

In another embodiment, the present invention contemplates a method for preventing infection of HCMV in a subject by stimulating an immune response to HCMV. An immune response can be stimulated by administering to the subject an amount of a modulator effective to modulate the expression and/or activity of HCMV. In some embodiments the immune response induced is a protective immune response, i.e., the response reduces the risk or severity of or clinical consequences of a CMV infection. Stimulating a protective immune response is particularly desirable in some populations particularly at risk from CMV infection and disease. For example, at-risk populations include solid organ transplant (SOT) patients, bone marrow transplant patients, and hematopoietic stem cell transplant (HSCT) patients. Modulators can be administered to a transplant donor pre-transplant, or a transplant recipient pre- and/or post-transplant. Because vertical transmission from mother to child is a common source of infecting infants, administering modulators to a woman who is pregnant or can become pregnant is particularly useful.

In certain instances, the subject is a mammal. In some embodiments, the mammal is a human. In some embodiments, the human is a child, such as an infant. In some other embodiments, the human is female, including an adolescent female, a female of childbearing age, a female who is planning pregnancy, a pregnant female, and females who recently gave birth. In some embodiments, the human is a transplant patient.

For CMVs of nonhuman mammals, such as rhesus monkeys, guinea pigs, or mice, homologous sequences and their mimetics are expected to have the same properties and be used for experimental, prophylactic, or therapeutic purposes in those nonhuman mammals.

Because CMVs are highly species specific, construction of animal CMV (including, but not limited to, mouse CMV, guinea pig CMV, and rhesus CMV) homologues of the HCMV domain V and testing them for binding and inhibition of activity of the animal CMV homologues of gB and for the inhibition of infection of the animals by MCMV, gpCMV, and rhCMV are also contemplated in this invention.

In another embodiment, modulators or inhibitors of HCMV, or biological complexes containing them, may be used in the manufacture of a medicament for any number of uses, including, for example, preventing or treating any disease or other treatable condition of a subject (including humans and animals), and particularly a disease caused by HCMV infection.

(i) Modulator Design

A number of techniques can be used to screen, identify, select and design chemical entities capable of associating with a HCMV gB polypeptide, structurally homologous molecules, and other molecules. Knowledge of the structure for a HCMV gB polypeptide, determined in accordance with the methods described herein, permits the design and/or identification of molecules and/or other modulators which have a shape complementary to the conformation of a HCMV gB polypeptide, or more particularly, a druggable region thereof. It is understood that such techniques and methods may use, in addition to the exact structural coordinates and other information for a HCMV gB polypeptide, structural equivalents thereof described above (including, for example, those structural coordinates that are derived from the structural coordinates of amino acids contained in a druggable region as described above).

The term "chemical entity," as used herein, refers to chemical compounds, complexes of two or more chemical compounds, and fragments of such compounds or complexes. In certain instances, it is desirable to use chemical entities exhibiting a wide range of structural and functional diversity, such as compounds exhibiting different shapes (e.g., flat aromatic rings(s), puckered aliphatic rings(s), straight and branched chain aliphatics with single, double, or triple bonds) and diverse functional groups (e.g., carboxylic acids, esters, ethers, amines, aldehydes, ketones, and various heterocyclic rings).

In one aspect, the method of drug design generally includes computationally evaluating the potential of a selected chemical entity to associate with any of the molecules or complexes of the present invention (or portions thereof). For example, this method may include the steps of (a) employing computational means to perform a fitting operation between the selected chemical entity and a druggable region of the molecule or complex; and (b) analyzing the results of said fitting operation to quantify the association between the chemical entity and the druggable region.

A chemical entity may be examined either through visual inspection or through the use of computer modeling using a docking program such as GRAM, DOCK, or AUTODOCK (Dunbrack et al., *Folding & Design*, 2:27-42 (1997)). This procedure can include computer fitting

of chemical entities to a target to ascertain how well the shape and the chemical structure of each chemical entity will complement or interfere with the structure of a HCMV gB polypeptide (Bugg et al., Scientific American, December: 92-98 (1993); West et al., TIPS, 16:67-74 (1995)). Computer programs may also be employed to estimate the attraction, repulsion, and steric hindrance of the chemical entity to a druggable region, for example. Generally, the tighter the fit (e.g., the lower the steric hindrance, and/or the greater the attractive force) the more potent the chemical entity will be because these properties are consistent with a tighter binding constant. Furthermore, the more specificity in the design of a chemical entity the more likely that the chemical entity will not interfere with related proteins, which may minimize potential side-effects due to unwanted interactions.

A variety of computational methods for molecular design, in which the steric and electronic properties of druggable regions are used to guide the design of chemical entities, are known: Cohen et al. (1990) J. Med. Cam. 33: 883-894; Kuntz et al. (1982) J. Mol. Biol. 161: 269-288; DesJarlais (1988) J. Med. Cam. 31: 722-729; Bartlett et al. (1989) Spec. Publ., Roy. Soc. Chem. 78: 182-196; Goodford et al. (1985) J. Med. Cam. 28: 849-857; and DesJarlais et al. J. Med. Cam. 29: 2149-2153. Directed methods generally fall into two categories: (1) design by analogy in which 3-D structures of known chemical entities (such as from a crystallographic database) are docked to the druggable region and scored for goodness-of-fit; and (2) de novo design, in which the chemical entity is constructed piece-wise in the druggable region. The chemical entity may be screened as part of a library or a database of molecules. Databases which may be used include ACD (Molecular Designs Limited), NCI (National Cancer Institute), CCDC (Cambridge Crystallographic Data Center), CAST (Chemical Abstract Service), Derwent (Derwent Information Limited), Maybridge (Maybridge Chemical Company Ltd), Aldrich (Aldrich Chemical Company), DOCK (University of California in San Francisco), and the Directory of Natural Products (Chapman & Hall). Computer programs such as CONCORD (Tripos Associates) or DB-Converter (Molecular Simulations Limited) can be used to convert a data set represented in two dimensions to one represented in three dimensions.

Chemical entities may be tested for their capacity to fit spatially with a druggable region or other portion of a target protein. As used herein, the term "fits spatially" means that the three-dimensional structure of the chemical entity is accommodated geometrically by a druggable region. A favorable geometric fit occurs when the surface area of the chemical entity is in close proximity with the surface area of the druggable region without forming unfavorable interactions. A favorable complementary interaction occurs where the chemical entity interacts by hydrophobic, aromatic, ionic, dipolar, or hydrogen donating and accepting forces. Unfavorable interactions may be steric hindrance between atoms in the chemical entity and atoms in the druggable region.

If a model of the present invention is a computer model, the chemical entities may be positioned in a druggable region through computational docking. If, on the other hand, the

model of the present invention is a structural model, the chemical entities may be positioned in the druggable region by, for example, manual docking. As used herein the term “docking” refers to a process of placing a chemical entity in close proximity with a druggable region, or a process of finding low energy conformations of a chemical entity/druggable region complex.

5 In an illustrative embodiment, the design of potential modulator begins from the general perspective of shape complimentary for the druggable region of a HCMV gB polypeptide, and a search algorithm is employed which is capable of scanning a database of small molecules of known three-dimensional structure for chemical entities which fit geometrically with the target druggable region. Most algorithms of this type provide a method for finding a wide assortment of
10 chemical entities that are complementary to the shape of a druggable region of a HCMV gB polypeptide. Each of a set of chemical entities from a particular database, such as the Cambridge Crystallographic Data Bank (CCDB) (Allen et al. (1973) J. Chem. Doc. 13: 119), is individually docked to the druggable region of a HCMV gB polypeptide in a number of geometrically permissible orientations with use of a docking algorithm. In certain embodiments,
15 a set of computer algorithms called DOCK, can be used to characterize the shape of invaginations and grooves that form the active sites and recognition surfaces of the druggable region (Kuntz et al. (1982) J. Mol. Biol. 161: 269-288). The program can also search a database of small molecules for templates whose shapes are complementary to particular binding sites of a HCMV gB polypeptide (DesJarlais et al. (1988) J Med Chem 31: 722-729).

20 The orientations are evaluated for goodness-of-fit and the best are kept for further examination using molecular mechanics programs, such as AMBER or CHARMM. Such algorithms have previously proven successful in finding a variety of chemical entities that are complementary in shape to a druggable region.

Goodford (1985, J Med Chem 28:849-857) and Boobbyer et al. (1989, J Med Chem
25 32:1083-1094) have produced a computer program (GRID) which seeks to determine regions or high affinity for different chemical groups (termed probes) of the druggable region. GRID hence provides a tool for suggesting modifications to known chemical entities that might enhance binding. It may be anticipated that some of the sites discerned by GRID as regions of high affinity correspond to “pharmacophoric patterns” determined interferentially from a series of
30 known ligands. As used herein, a “pharmacophoric pattern” is a geometric arrangement of features of chemical entities that is believed to be important for binding. Attempts have been made to use pharmacophoric patterns as a search screen for novel ligands (Jakes et al. (1987) J Mol Graph 5:41-48; Brint et al. (1987) J Graph 5:49-56; Jakes et al. (1986) J Mol Graph 4:12-20).

35 Yet a further embodiment of the present invention utilizes a computer algorithm such as CLIX which searches such databases as CCDB for chemical entities which can be oriented with the druggable region in a way that is both sterically acceptable and has a high likelihood of achieving favorable chemical interactions between the chemical entity and the surrounding

amino acid residues. The method is based on characterizing the region in terms of an ensemble of favorable binding positions for different chemical groups and then searching for orientations of the chemical entities that cause maximum spatial coincidence of individual candidate chemical groups with members of the ensemble. The algorithmic details of CLIX is described in

5 Lawrence et al. (1992) Proteins 12:31-41.

In this way, the efficiency with which a chemical entity may bind to or interfere with a druggable region may be tested and optimized by computational evaluation. For example, for a favorable association with a druggable region, a chemical entity must preferably demonstrate a relatively small difference in energy between its bound and free states (i.e., a small deformation energy of binding). Thus, certain, more desirable chemical entities will be designed with a
10 deformation energy of binding of not greater than about 10 kcal/mole, and more preferably, not greater than 7 kcal/mole. Chemical entities may interact with a druggable region in more than one conformation that is similar in overall binding energy. In those cases, the deformation energy of binding is taken to be the difference between the energy of the free entity and the
15 average energy of the conformations observed when the chemical entity binds to the target.

In this way, the present invention provides computer-assisted methods for identifying or designing a potential modulator of the activity of a HCMV gB polypeptide including: supplying a computer modeling application with a set of structure coordinates of a molecule or complex, the molecule or complex including at least a portion of a druggable region from a HCMV gB
20 polypeptide; supplying the computer modeling application with a set of structure coordinates of a chemical entity; and determining whether the chemical entity is expected to bind to the molecule or complex, wherein binding to the molecule or complex is indicative of potential modulation of the activity of a HCMV gB polypeptide.

In another aspect, the present invention provides a computer-assisted method for
25 identifying or designing a potential modulator to a HCMV gB polypeptide, supplying a computer modeling application with a set of structure coordinates of a molecule or complex, the molecule or complex including at least a portion of a druggable region of a HCMV gB polypeptide; supplying the computer modeling application with a set of structure coordinates for a chemical entity; evaluating the potential binding interactions between the chemical entity and active site
30 of the molecule or molecular complex; structurally modifying the chemical entity to yield a set of structure coordinates for a modified chemical entity, and determining whether the modified chemical entity is expected to bind to the molecule or complex, wherein binding to the molecule or complex is indicative of potential modulation of the HCMV gB polypeptide.

In one embodiment, a potential modulator can be obtained by screening a peptide or
35 other compound or chemical library (Scott and Smith, Science, 249:386-390 (1990); Cwirla et al., Proc. Natl. Acad. Sci., 87:6378-6382 (1990); Devlin et al., Science, 249:404-406 (1990)). A potential modulator selected in this manner could then be systematically modified by computer modeling programs until one or more promising potential drugs are identified. Such analysis has

been shown to be effective in the development of HIV protease modulators (Lam et al., Science 263:380-384 (1994); Wlodawer et al., Ann. Rev. Biochem. 62:543-585 (1993); Appelt, Perspectives in Drug Discovery and Design 1:23-48 (1993); Erickson, Perspectives in Drug Discovery and Design 1:109-128 (1993)). Alternatively, a potential modulator may be selected
5 from a library of chemicals such as those that can be licensed from third parties, such as chemical and pharmaceutical companies. A third alternative is to synthesize the potential modulator *de novo*.

For example, in certain embodiments, the present invention provides a method for making a potential modulator for a HCMV gB polypeptide, the method including synthesizing a
10 chemical entity or a molecule containing the chemical entity to yield a potential modulator of a HCMV gB polypeptide, the chemical entity having been identified during a computer-assisted process including supplying a computer modeling application with a set of structure coordinates of a molecule or complex, the molecule or complex including at least one druggable region from a HCMV gB polypeptide; supplying the computer modeling application with a set of structure
15 coordinates of a chemical entity; and determining whether the chemical entity is expected to bind to the molecule or complex at the active site, e.g. druggable region, wherein binding to the molecule or complex is indicative of potential modulation. This method may further include the steps of evaluating the potential binding interactions between the chemical entity and the active site, e.g. druggable region, of the molecule or molecular complex and structurally modifying the
20 chemical entity to yield a set of structure coordinates for a modified chemical entity, which steps may be repeated one or more times.

Once a potential modulator is identified, it can then be tested in any standard assay for the macromolecule depending of course on the macromolecule, including in high throughput assays. Further refinements to the structure of the modulator will generally be necessary and
25 can be made by the successive iterations of any and/or all of the steps provided by the particular screening assay, in particular further structural analysis by e.g., 15N NMR relaxation rate determinations or x-ray crystallography with the modulator bound to a HCMV gB polypeptide. These studies may be performed in conjunction with biochemical assays.

Once identified, a potential modulator may be used as a model structure, and analogs to
30 the compound can be obtained. The analogs are then screened for their ability to bind to a HCMV gB polypeptide. An analog of the potential modulator might be chosen as a modulator when it binds to a HCMV gB polypeptide with a higher binding affinity than the predecessor modulator.

In a related approach, iterative drug design is used to identify modulators of a target
35 protein. Iterative drug design is a method for optimizing associations between a protein and a modulator by determining and evaluating the three-dimensional structures of successive sets of protein/modulator complexes. In iterative drug design, crystals of a series of protein/modulator complexes are obtained and then the three-dimensional structures of each complex is solved.

Such an approach provides insight into the association between the proteins and modulators of each complex. For example, this approach may be accomplished by selecting modulators with modulatory activity, obtaining crystals of this new protein/modulator complex, solving the three-dimensional structure of the complex, and comparing the associations between the new
5 protein/modulator complex and previously solved protein/modulator complexes. By observing how changes in the modulator affected the protein/modulator associations, these associations may be optimized.

In addition to designing and/or identifying a chemical entity to associate with a druggable region, as described above, the same techniques and methods may be used to design and/or
10 identify chemical entities that either associate, or do not associate, with affinity regions, selectivity regions or undesired regions of protein targets. By such methods, selectivity for one or a few targets, or alternatively for multiple targets, from the same species or from multiple species, can be achieved.

For example, a chemical entity may be designed and/or identified for which the binding
15 energy for one druggable region, e.g., an affinity region or selectivity region, is more favorable than that for another region, e.g., an undesired region, by about 20%, 30%, 50% to about 60% or more. It may be the case that the difference is observed between (a) more than two regions, (b) between different regions (selectivity, affinity or undesirable) from the same target, (c) between regions of different targets, (d) between regions of homologs from different species, or
20 (e) between other combinations. Alternatively, the comparison may be made by reference to the K_d , usually the apparent K_d , of said chemical entity with the two or more regions in question.

In another aspect, prospective modulators are screened for binding to two nearby druggable regions on a target protein. For example, a modulator that binds a first region of a target polypeptide does not bind a second nearby region. Binding to the second region can be
25 determined by monitoring changes in a different set of amide chemical shifts in either the original screen or a second screen conducted in the presence of a modulator (or potential modulator) for the first region. From an analysis of the chemical shift changes, the approximate location of a potential modulator for the second region is identified. Optimization of the second modulator for binding to the region is then carried out by screening structurally related
30 compounds (e.g., analogs as described above). When modulators for the first region and the second region are identified, their location and orientation in the ternary complex can be determined experimentally. On the basis of this structural information, a linked compound, e.g., a consolidated modulator, is synthesized in which the modulator for the first region and the modulator for the second region are linked. In certain embodiments, the two modulators are
35 covalently linked to form a consolidated modulator. This consolidated modulator may be tested to determine if it has a higher binding affinity for the target than either of the two individual modulators. A consolidated modulator is selected as a modulator when it has a higher binding affinity for the target than either of the two modulators. Larger consolidated modulators can be

constructed in an analogous manner, e.g., linking three modulators which bind to three nearby regions on the target to form a multilinked consolidated modulator that has an even higher affinity for the target than the linked modulator. In this example, it is assumed that is desirable to have the modulator bind to all the druggable regions. However, it may be the case that binding to certain of the druggable regions is not desirable, so that the same techniques may be used to identify modulators and consolidated modulators that show increased specificity based on binding to at least one but not all druggable regions of a target.

The present invention provides a number of methods that use drug design as described above. For example, in one aspect, the present invention contemplates a method for designing a candidate therapeutic for screening for modulators of a HCMV gB polypeptide, the method comprising: (a) determining the three dimensional structure of a crystallized HCMV gB polypeptide or a fragment thereof; and (b) designing a candidate therapeutic based on the three dimensional structure of the crystallized polypeptide or fragment.

(ii) Modulator Libraries

The synthesis and screening of combinatorial libraries is a validated strategy for the identification and study of organic molecules of interest. According to the present invention, the synthesis of libraries containing molecules that bind, interact with, or modulate the activity/function of a subject druggable region may be performed using established combinatorial methods for solution phase, solid phase, or a combination of solution phase and solid phase synthesis techniques. The synthesis of combinatorial libraries is well known in the art and has been reviewed (see, e.g., "Combinatorial Chemistry", Chemical and Engineering News, Feb. 24, 1997, p. 43; Thompson et al., Chem. Rev. (1996) 96:555). Many libraries are commercially available. One of ordinary skill in the art will realize that the choice of method for any particular embodiment will depend upon the specific number of molecules to be synthesized, the specific reaction chemistry, and the availability of specific instrumentation, such as robotic instrumentation for the preparation and analysis of the inventive libraries. In certain embodiments, the reactions to be performed to generate the libraries are selected for their ability to proceed in high yield, and in a stereoselective and regioselective fashion, if applicable.

In one aspect of the present invention, the inventive libraries are generated using a solution phase technique. Traditional advantages of solution phase techniques for the synthesis of combinatorial libraries include the availability of a much wider range of reactions, and the relative ease with which products may be characterized, and ready identification of library members, as discussed below. For example, in certain embodiments, for the generation of a solution phase combinatorial library, a parallel synthesis technique is utilized, in which all of the products are assembled separately in their own reaction vessels. In a particular parallel synthesis procedure, a microtitre plate containing n rows and m columns of tiny wells which are

capable of holding a few milliliters of the solvent in which the reaction will occur, is utilized. It is possible to then use n variants of reactant A, such as a ligand, and m variants of reactant B, such as a second ligand, to obtain $n \times m$ variants, in $n \times m$ wells. One of ordinary skill in the art will realize that this particular procedure is most useful when smaller libraries are desired, and the specific wells may provide a ready means to identify the library members in a particular well.

In other embodiments of the present invention, a solid phase synthesis technique is utilized. Solid phase techniques allow reactions to be driven to completion because excess reagents may be utilized and the unreacted reagent washed away. Solid phase synthesis also allows the use a technique called "split and pool", in addition to the parallel synthesis technique, developed by Furka. See, e.g., Furka et al., *Abstr. 14th Int. Congr. Biochem.*, (Prague, Czechoslovakia) (1988) 5:47; Furka et al., *Int. J. Pept. Protein Res.* (1991) 37:487; Sebestyen et al., *Bioorg. Med. Chem. Lett.* (1993) 3:413. In this technique, a mixture of related molecules may be made in the same reaction vessel, thus substantially reducing the number of containers required for the synthesis of very large libraries, such as those containing as many as or more than one million library members. As an example, the solid support with the starting material attached may be divided into n vessels, where n represents the number species of reagent A to be reacted with the such starting material. After reaction, the contents from n vessels are combined and then split into m vessels, where m represents the number of species of reagent B to be reacted with the now modified starting materials. This procedure is repeated until the desired number of reagents is reacted with the starting materials to yield the inventive library.

The use of solid phase techniques in the present invention may also include the use of a specific encoding technique. Specific encoding techniques have been reviewed by Czarnik in *Current Opinion in Chemical Biology* (1997) 1:60. One of ordinary skill in the art will also realize that if smaller solid phase libraries are generated in specific reaction wells, such as 96 well plates, or on plastic pins, the reaction history of these library members may also be identified by their spatial coordinates in the particular plate, and thus are spatially encoded. In other embodiments, an encoding technique involves the use of a particular "identifying agent" attached to the solid support, which enables the determination of the structure of a specific library member without reference to its spatial coordinates. Examples of such encoding techniques include, but are not limited to, spatial encoding techniques, graphical encoding techniques, including the "tea bag" method, chemical encoding methods, and spectrophotometric encoding methods. One of ordinary skill in the art will realize that the particular encoding method to be used in the present invention must be selected based upon the number of library members desired, and the reaction chemistry employed.

In certain embodiments, molecules of the present invention may be prepared using solid support chemistry known in the art. For example, polypeptides having up to twenty amino acids or more may be generated using standard solid phase technology on commercially available equipment (such as Advanced Chemtech multiple organic synthesizers). In certain

embodiments, a starting material or later reactant may be attached to the solid phase, through a linking unit, or directly, and subsequently used in the synthesis of desired molecules. The choice of linkage will depend upon the reactivity of the molecules and the solid support units and the stability of these linkages. Direct attachment to the solid support via a linker molecule
5 may be useful if it is desired not to detach the library member from the solid support. For example, for direct on-bead analysis of biological activity, a stronger interaction between the library member and the solid support may be desirable. Alternatively, the use of a linking reagent may be useful if more facile cleavage of the inventive library members from the solid support is desired.

10 In regard to automation of the present subject methods, a variety of instrumentation may be used to allow for the facile and efficient preparation of chemical libraries of the present invention, and methods of assaying members of such libraries. In general, automation, as used in reference to the synthesis and preparation of the subject chemical libraries, involves having instrumentation complete one or more of the operative steps that must be repeated a multitude
15 of times because a library instead of a single molecule is being prepared. Examples of automation include, without limitation, having instrumentation complete the addition of reagents, the mixing and reaction of them, filtering of reaction mixtures, washing of solids with solvents, removal and addition of solvents, and the like. Automation may be applied to any steps in a reaction scheme, including those to prepare, purify and assay molecules for use in the
20 compositions of the present invention.

There is a range of automation possible. For example, the synthesis of the subject libraries may be wholly automated or only partially automated. If wholly automated, the subject library may be prepared by the instrumentation without any human intervention after initiating the synthetic process, other than refilling reagent bottles or monitoring or programming the
25 instrumentation as necessary. Although synthesis of a subject library may be wholly automated, it may be necessary for there to be human intervention for purification, identification, or the like of the library members.

In contrast, partial automation of the synthesis of a subject library involves some robotic assistance with the physical steps of the reaction schema that gives rise to the library, such as
30 mixing, stirring, filtering and the like, but still requires some human intervention other than just refilling reagent bottles or monitoring or programming the instrumentation. This type of robotic automation is distinguished from assistance provided by convention organic synthetic and biological techniques because in partial automation, instrumentation still completes one or more of the steps of any schema that is required to be completed a multitude of times because a
35 library of molecules is being prepared.

In certain embodiments, the subject library may be prepared in multiple reaction vessels (e.g., microtitre plates and the like), and the identity of particular members of the library may be determined by the location of each vessel. In other embodiments, the subject library may be

synthesized in solution, and by the use of deconvolution techniques, the identity of particular members may be determined.

In one aspect of the invention, the subject screening method may be carried out utilizing immobilized libraries. In certain embodiments, the immobilized library will have the ability to bind to a microorganism as described above. The choice of a suitable support will be routine to the skilled artisan. Important criteria may include that the reactivity of the support not interfere with the reactions required to prepare the library. Insoluble polymeric supports include functionalized polymers based on polystyrene, polystyrene/divinylbenzene copolymers, and the like, including any of the particles described herein. It will be understood that the polymeric support may be coated, grafted or otherwise bonded to other solid supports.

In another embodiment, the polymeric support may be provided by reversibly soluble polymers. Such polymeric supports include functionalized polymers based on polyvinyl alcohol or polyethylene glycol (PEG). A soluble support may be made insoluble (e.g., may be made to precipitate) by addition of a suitable inert nonsolvent. One advantage of reactions performed using soluble polymeric supports is that reactions in solution may be more rapid, higher yielding, and more complete than reactions that are performed on insoluble polymeric supports.

Once the synthesis of either a desired solution phase or solid support bound template has been completed, the template is then available for further reaction to yield the desired solution phase or solid support bound structure. The use of solid support bound templates enables the use of more rapid split and pool techniques.

Characterization of the library members may be performed using standard analytical techniques, such as mass spectrometry, Nuclear Magnetic Resonance Spectroscopy, including ¹⁹⁵Pt and ¹H NMR, chromatography (e.g., liquid etc.) and infra-red spectroscopy. One of ordinary skill in the art will realize that the selection of a particular analytical technique will depend upon whether the inventive library members are in the solution phase or on the solid phase. In addition to such characterization, the library member may be synthesized separately to allow for more ready identification.

(iii) In Vitro Assays

Any form of HCMV gB polypeptide, e.g. a full-length polypeptide or a fragment thereof comprising the target druggable region, may be used to assess the activity of candidate therapeutics in *in vitro* assays. In one embodiment of such an assay, agents are identified which modulate the biological activity of a druggable region, the protein-protein interaction of interest or formation of a protein complex involving a subject druggable region. In another embodiment of such an assay, agents are identified which bind or interact with subject druggable region. In certain embodiments, the test agent is a small organic molecule. The candidate agents may be selected, for example, from the following classes of compounds: detergents, proteins, peptides, peptidomimetics, antibodies, small molecules, cytokines, or hormones. In some embodiments,

the candidate therapeutic (also known as a “candidate agent” or “test agent”) may be in a library of compounds. These libraries may be generated using combinatorial synthetic methods as described above. In certain embodiments of the present invention, the ability of said candidate therapeutics to bind a target druggable region may be evaluated by an *in vitro* assay. In either
5 embodiment, discussed in the next section, the binding assay may also be *in vivo*.

 The invention also provides a method of screening multiple compounds to identify those which modulate the action of peptides or polypeptides of the invention, or polynucleotides encoding the same. The method of screening may involve high-throughput techniques. For example, to screen for modulators, a synthetic reaction mix, a coellular compartment, such as a
10 membrane, cell envelope or cell wall, or a preparation of any thereof, a whole cell or tissue, or even a whole organism, comprising a HCMV gB polypeptide and a labeled substrate or ligand of such polypeptide is incubated in the absence or the presence of a candidate therapeutic that may be a modulator of a HCMV gB polypeptide. The ability of the candidate therapeutic to modulate a HCMV gB polypeptide is reflected in decreased binding of the labeled ligand or
15 decreased production of product from such substrate. Detection of the rate or level of production of product from substrate may be enhanced by using a reporter system. Reporter systems that may be useful in this regard include but are not limited to colorimetric labeled substrate converted into product, a reporter gene that is responsive to changes in a nucleic acid of the invention or polypeptide activity, and binding assays known in the art. For example, in
20 one embodiment, the modulator or inhibitor peptide is conjugated to FITC and the affinity of the FITC-labelled peptide for the target binding pocket is measured using a fluorescence polarization assay. The specificity of the modulator or inhibitor for the target binding pocket may be shown by using a biotinylated derivative of the modulator or inhibitor using the “pull down” method described in Schmidt et al. ((2010) PLoS Pathog 6(4): e1000851). To detect the
25 modulator or inhibitor association with virions, incubate biotinyl-labelled modulator or inhibitor with virion and add streptavidin resin to detect by immunoblotting. To further confirm modulator or inhibitor association with virions, use pyrene labelled virions in the presence of modulator or inhibitor and check excimer intensity. In a preferred embodiment, fluorescence polarization and FRET are methods of detecting binding of the domain V peptides and inhibition of that binding.
30 In a further embodiment, surface plasmon resonance is used to screen for binding of the modulator or inhibitor to the druggable region.

 Another example of an assay for a modulator of a HCMV gB polypeptide is a competitive assay that combines a HCMV gB peptide or polypeptide and a potential modulator with molecules that bind to a HCMV gB polypeptide, recombinant molecules that bind to a
35 HCMV gB polypeptide, natural substrates or ligands, or substrate or ligand mimetics, under appropriate conditions for a competitive inhibition assay. Peptides or polypeptides of the invention can be labeled, such as by radioactivity or a colorimetric compound, such that the number of molecules of a HCMV gB peptide or polypeptide bound to a binding molecule or

converted to product can be determined accurately to assess the effectiveness of the potential modulator. In one aspect, the HCMV gB peptides are fragments of SEQ ID NO:1 having residues M648-K700 (SEQ ID NO: 260), M648-V697 (SEQ ID NO: 261), S647-V697 (SEQ ID NO: 262), S647-V663 (SEQ ID NO: 263), I653-V697 (SEQ ID NO: 264), I653-Q692 (SEQ ID NO: 265), I653-L680 (SEQ ID NO: 266), I653-S675 (SEQ ID NO: 267), I653-Y667 (SEQ ID NO: 268), R662-V697 (SEQ ID NO: 269), R662-Q692 (SEQ ID NO: 270), R662-L680 (SEQ ID NO: 271), R662-S675 (SEQ ID NO: 272), L664-F678 (SEQ ID NO: 273), S668-V697 (SEQ ID NO: 274), S668-Q692 (SEQ ID NO: 275), S668-V677 (SEQ ID NO: 276), D679-V697 (SEQ ID NO: 277), L680-V697 (SEQ ID NO: 278), L680-Q692 (SEQ ID NO: 279), and R693-K700 (SEQ ID NO: 280).

A number of methods for identifying a molecule which modulates the activity of a polypeptide are known in the art. For example, in one such method, a HCMV gB polypeptide is contacted with a test compound, and the activity of the HCMV gB polypeptide in the presence of the test compound is determined, wherein a change in the activity of the HCMV gB polypeptide is indicative that the test compound modulates the activity of the HCMV gB polypeptide. In certain instances, the test compound agonizes the activity of the HCMV gB polypeptide, and in other instances, the test compound antagonizes the activity of the HCMV gB polypeptide.

In another example, a compound which modulates HCMV gB polypeptide dependent growth or infectivity of HCMV may be identified by (a) contacting a HCMV gB polypeptide with a test compound; and (b) determining the activity of the polypeptide in the presence of the test compound, wherein a change in the activity of the polypeptide is indicative that the test compound may modulate the growth or infectivity of HCMV.

In certain of the subject assays, to evaluate the results using the subject compositions, comparisons may be made to known molecules, such as one with a known binding affinity for the target. For example, a known molecule and a new molecule of interest may be assayed. The result of the assay for the subject complex will be of a type and of a magnitude that may be compared to result for the known molecule. To the extent that the subject complex exhibits a type of response in the assay that is quantifiably different from that of the known molecule then the result for such complex in the assay would be deemed a positive or negative result. In certain assays, the magnitude of the response may be expressed as a percentage response with the known molecule result, e.g. 100% of the known result if they are the same.

As those skilled in the art will understand, based on the present description, binding assays may be used to detect agents that bind a polypeptide. Cell-free assays may be used to identify molecules that are capable of interacting with a polypeptide. In a preferred embodiment, cell-free assays for identifying such molecules are comprised essentially of a reaction mixture containing a target and a test molecule or a library of test molecules. A test molecule may be, e.g., a derivative of a known binding partner of the target, e.g., a biologically inactive peptide, or a small molecule. Agents to be tested for their ability to bind may be produced, for example, by

bacteria, yeast or other organisms (e.g. natural products), produced chemically (e.g. small molecules, including peptidomimetics), or produced recombinantly. In certain embodiments, the test molecule is selected from the group consisting of lipids, carbohydrates, peptides, peptidomimetics, peptide-nucleic acids (PNAs), proteins (including antibodies), small molecules, 5 natural products, aptamers and oligonucleotides. In other embodiments of the invention, the binding assays are not cell-free. In a preferred embodiment, such assays for identifying molecules that bind a target comprise a reaction mixture containing a target microorganism and a test molecule or a library of test molecules.

In many candidate screening programs which test libraries of molecules and natural 10 extracts, high throughput assays are desirable in order to maximize the number of molecules surveyed in a given period of time. Assays of the present invention which are performed in cell-free systems, such as may be derived with purified or semi-purified proteins or with lysates, are often preferred as "primary" screens in that they may be generated to permit rapid development and relatively easy detection of binding between a target and a test molecule. Moreover, the 15 effects of cellular toxicity and/or bioavailability of the test molecule may be generally ignored in the in vitro system, the assay instead being focused primarily on the ability of the molecule to bind the target. Accordingly, potential binding molecules may be detected in a cell-free assay generated by constitution of functional interactions of interest in a cell lysate. In an alternate format, the assay may be derived as a reconstituted protein mixture which, as described below, 20 offers a number of benefits over lysate-based assays.

In one aspect, the present invention provides assays that may be used to screen for molecules that bind HCMV gB polypeptide druggable regions. In an exemplary binding assay, the molecule of interest is contacted with a mixture generated from target cell surface polypeptides. Detection and quantification of expected binding to a target polypeptide provides 25 a means for determining the molecule's efficacy at binding the target. The efficacy of the molecule may be assessed by generating dose response curves from data obtained using various concentrations of the test molecule. Moreover, a control assay may also be performed to provide a baseline for comparison. In the control assay, the formation of complexes is quantitated in the absence of the test molecule.

30 Complex formation between a molecule and a target HCMV gB polypeptide or microorganism containing a HCMV gB polypeptide may be detected by a variety of techniques, many of which are effectively described above. For instance, modulation in the formation of complexes may be quantitated using, for example, detectably labeled proteins (e.g. radiolabeled, fluorescently labeled, or enzymatically labeled), by immunoassay, or by 35 chromatographic detection.

Accordingly, one exemplary screening assay of the present invention includes the steps of contacting a HCMV gB polypeptide or functional fragment thereof with a test molecule or library of test molecules and detecting the formation of complexes. For detection purposes, for

example, the molecule may be labeled with a specific marker and the test molecule or library of test molecules labeled with a different marker. Interaction of a test molecule with a polypeptide or fragment thereof may then be detected by determining the level of the two labels after an incubation step and a washing step. The presence of two labels after the washing step is
5 indicative of an interaction. Such an assay may also be modified to work with a whole target cell.

An interaction between a HCMV gB polypeptide target and a molecule may also be identified by using real-time BIA (Biomolecular Interaction Analysis, Pharmacia Biosensor AB) which detects surface plasmon resonance (SPR), an optical phenomenon. Detection depends
10 on changes in the mass concentration of macromolecules at the biospecific interface, and does not require any labeling of interactants. In one embodiment, a library of test molecules may be immobilized on a sensor surface, e.g., which forms one wall of a micro-flow cell. A solution containing the target is then flowed continuously over the sensor surface. A change in the resonance angle as shown on a signal recording, indicates that an interaction has occurred.
15 This technique is further described, e.g., in BIAtechnology Handbook by Pharmacia.

In a preferred embodiment, it will be desirable to immobilize the target to facilitate separation of complexes from uncomplexed forms, as well as to accommodate automation of the assay. Binding of polypeptide to a test molecule may be accomplished in any vessel suitable for containing the reactants. Examples include microtitre plates, test tubes, and micro-
20 centrifuge tubes. In one embodiment, a fusion protein may be provided which adds a domain that allows the target to be bound to a matrix. For example, glutathione-S-transferase/polypeptide (GST/polypeptide) fusion proteins may be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, Mo.) or glutathione derivatized microtitre plates, which are then combined with a labeled test molecule (e.g., S35 labeled, P33 labeled, and the
25 like, and the mixture incubated under conditions conducive to complex formation, e.g. at physiological conditions for salt and pH, though slightly more stringent conditions may be desired. Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly (e.g. beads placed in scintillant), or in the supernatant after the complexes are subsequently dissociated. Alternatively, the complexes
30 may be dissociated from the matrix, separated by SDS-PAGE, and the level of polypeptide or binding partner found in the bead fraction quantitated from the gel using standard electrophoretic techniques such as described in the appended examples. The above techniques could also be modified in which the test molecule is immobilized, and the labeled target is incubated with the immobilized test molecules. In one embodiment of the invention, the
35 test molecules are immobilized, optionally via a linker, to a particle of the invention, e.g. to create the ultimate composition.

Other techniques for immobilizing targets or molecules on matrices may be used in the subject assays. For instance, a target or molecule may be immobilized utilizing conjugation of

biotin and streptavidin. For instance, biotinylated polypeptide molecules may be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with a target or molecule may be derivatized to the wells of the plate, and the target or molecule trapped in the wells by antibody conjugation. As above, preparations of test molecules are incubated in the polypeptide presenting wells of the plate, and the amount of complex trapped in the well may be quantitated. Exemplary methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the complex, or which are reactive with one of the complex components; as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with a target or molecule, either intrinsic or extrinsic activity. In an instance of the latter, the enzyme may be chemically conjugated or provided as a fusion protein with the target or molecule. To illustrate, a target polypeptide may be chemically cross-linked or genetically fused with horseradish peroxidase, and the amount of polypeptide trapped in a complex with a molecule may be assessed with a chromogenic substrate of the enzyme, e.g. 3,3'-diamino-benzidine tetrahydrochloride or 4-chloro-1-naphthol. Likewise, a fusion protein comprising the polypeptide and glutathione-S-transferase may be provided, and complex formation quantitated by detecting the GST activity using 1-chloro-2,4-dinitrobenzene (Habig et al (1974) J Biol Chem 249:7130).

For processes that rely on immunodetection for quantitating one of the components trapped in a complex, antibodies against a component, such as anti-polypeptide antibodies, may be used. Alternatively, the component to be detected in the complex may be "epitope tagged" in the form of a fusion protein which includes, in addition to the polypeptide sequence, a second polypeptide for which antibodies are readily available (e.g. from commercial sources). For instance, the GST fusion proteins, described above may also be used for quantification of binding using antibodies against the GST moiety. Other useful epitope tags include myc-epitopes (e.g., see Ellison et al. (1991) J Biol Chem 266:21150-21157) which includes a 10-residue sequence from c-myc, as well as the pFLAG system (International Biotechnologies, Inc.) or the pEZZ-protein A system (Pharmacia, NJ).

In certain *in vitro* embodiments of the present assay, the solution containing the target comprises a reconstituted protein mixture of at least semi-purified proteins. By semi-purified, it is meant that the components utilized in the reconstituted mixture have been previously separated from other cellular or viral proteins. For instance, in contrast to cell lysates, a target protein is present in the mixture to at least 50% purity relative to all other proteins in the mixture, and more preferably are present at 90-95% purity. In certain embodiments of the subject method, the reconstituted protein mixture is derived by mixing highly purified proteins such that the reconstituted mixture substantially lacks other proteins (such as of cellular or viral origin)

which might interfere with or otherwise alter the ability to measure binding activity. In one embodiment, the use of reconstituted protein mixtures allows more careful control of the target:molecule interaction conditions.

5 In still other embodiments of the present invention, variations of viral fusion or viral infectivity assays may be utilized in order to determine the ability of a test molecule to prevent a virus expressing HCMV gB polypeptide from binding to, fusing with, or infecting cells. If fusion, binding, or infecting is prevented, then the molecule or composition may be useful as a therapeutic agent.

All of the screening methods may be accomplished by using a variety of assay formats. In light of the present disclosure, those not expressly described herein will nevertheless be known and comprehended by one of ordinary skill in the art. Assay formats which approximate such conditions as formation of protein complexes or protein-nucleic acid complexes, and enzymatic activity may be generated in many different forms, as those skilled in the art will appreciate based on the present description and include but are not limited to assays based on cell-free systems, e.g. purified proteins or cell lysates, as well as cell-based assays which utilize intact cells. Assaying binding resulting from a given target:molecule interaction may be accomplished in any vessel suitable for containing the reactants. Examples include microtitre plates, test tubes, and micro-centrifuge tubes. Any of the assays may be provided in kit format and may be automated. Many of the following particularized assays rely on general principles, such as blockage or prevention of fusion, that may apply to other particular assays.

(iv) *In Vivo* Assays

Animal models of viral infection and/or disease may be used as an *in vivo* assay for evaluating the effectiveness of a potential drug target in treating or preventing HCMV infection. A number of suitable animal models are described briefly below, however, these models are only examples and modifications, or completely different animal models, may be used in accord with the methods of the invention. Animal models may be developed by methods known in the art, for example, by infecting an animal with HCMV, or by genetically engineering an animal to be predisposed to such infection (see, e.g., Maidji, E. et al. Impaired Surfactant Production by Alveolar Epithelial Cells in a SCID-hu Lung Mouse Model of Congenital Human Cytomegalovirus Infection. J Virol. 2012 Dec; 86(23): 12795–12805; Crawford, L.B. et al. Humanized Mouse Models of Human Cytomegalovirus Infection. Curr Opin Virol. 2015 Aug; 13: 86–92).

Further, viral infectivity assays may be used as *in vivo* assays to assess the effectiveness of a potential drug target in treating or preventing HCMV infection. For example, such as competitive, asymmetric reverse transcriptase-mediated PCR (RT-PCR) assays and flow cytometric assays that measure viral antigen, and plaque assays, may be used to assess the effectiveness of a potential drug target.

Still further, cell-cell fusion assays may be used as *in vivo* assays to assess the effectiveness of a potential drug target in treating or preventing HCMV infection.

A variety of other *in vivo* models are available and may be used when appropriate for specific pathogens or specific test agents.

5 It is also relevant to note that the species of animal used for an infection model, and the specific genetic make-up of that animal, may contribute to the effective evaluation of the effects of a particular test agent. For example, immuno-incompetent animals may, in some instances, be preferable to immuno-competent animals. For example, the action of a competent immune system may, to some degree, mask the effects of the test agent as compared to a similar
10 infection in an immuno-incompetent animal. In addition, many opportunistic infections, in fact, occur in immuno-compromised patients, so modeling an infection in a similar immunological environment is appropriate.

E. Pharmaceutical Compositions

15 Pharmaceutical compositions of this invention include any modulator identified according to the present invention, or a pharmaceutically acceptable salt thereof, and a pharmaceutically acceptable carrier, adjuvant, or vehicle. The term "pharmaceutically acceptable carrier" refers to a carrier(s) that is "acceptable" in the sense of being compatible with the other ingredients of a composition and not deleterious to the recipient thereof.

20 Methods of making and using such pharmaceutical compositions are also included in the invention. The pharmaceutical compositions of the invention can be administered orally, parenterally, by inhalation spray, topically, rectally, nasally, buccally, vaginally, or via an implanted reservoir. The term parenteral as used herein includes subcutaneous, intracutaneous, intravenous, intramuscular, intra articular, intrasynovial, intrasternal, intrathecal,
25 intralesional, and intracranial injection or infusion techniques.

Dosage levels of between about 0.01 and about 100 mg/kg body weight per day, preferably between about 0.5 and about 75 mg/kg body weight per day of the modulators described herein are useful for the prevention and treatment of disease and conditions caused by HCMV infection, including diseases and conditions mediated by pathogenic species of origin
30 for the polypeptides of the invention. The amount of active ingredient that may be combined with the carrier materials to produce a single dosage form will vary depending upon the host treated and the particular mode of administration. A typical preparation will contain from about 5% to about 95% active compound (w/w). Alternatively, such preparations contain from about 20% to about 80% active compound.

35

F. Kits

The present invention provides kits for treating or preventing HCMV infections. For example, a kit may comprise compositions comprising compounds identified herein as

modulators of HCMV gB polypeptide. The compositions may be pharmaceutical compositions comprising a pharmaceutically acceptable excipient. In other embodiments involving kits, this invention contemplates a kit including compositions of the present invention, and optionally instructions for their use. Kit components may be packaged for either manual or partially or
5 wholly automated practice of the foregoing methods. Such kits may have a variety of uses, including, for example, imaging, diagnosis, therapy, and other applications.

G. Preparation of the polypeptide

The polypeptides described herein may be prepared by routine methods known in the
10 art, such as by expression in a recombinant host system using a suitable vector. Suitable recombinant host cells include, for example, insect cells, mammalian cells, avian cells, bacteria, and yeast cells. Examples of suitable insect cells include, for example, Sf9 cells, Sf21 cells, Tn5 cells, Schneider S2 cells, and HIGH FIVE cells (a clonal isolate derived from the parental Trichoplusia ni BTI- TN-5B1-4 cell line). Examples of suitable mammalian cells include Chinese
15 hamster ovary (CHO) cells, human embryonic kidney cells (HEK293 or Expi 293 cells, typically transformed by sheared adenovirus type 5 DNA), NIH-3T3 cells, 293-T cells, Vero cells, and HeLa cells. Suitable avian cells include, for example, chicken embryonic stem cells (e.g., EBx.RTM. cells), chicken embryonic fibroblasts, chicken embryonic germ cells, quail fibroblasts (e.g. ELL-O), and duck cells. Suitable insect cell expression systems, such as baculovirus-
20 vectored systems, are known to those of skill in the art. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from. Avian cell expression systems are also known to those of skill in the art. Similarly, bacterial and mammalian cell expression systems are also known in the art.

A number of suitable vectors for expression of recombinant proteins in insect or
25 mammalian cells are well-known and conventional in the art. Suitable vectors can contain a number of components, including, but not limited to one or more of the following: an origin of replication; a selectable marker gene; one or more expression control elements, such as a transcriptional control element (e.g., a promoter, an enhancer, a terminator), and/or one or more translation signals; and a signal sequence or leader sequence for targeting to the secretory
30 pathway in a selected host cell (e.g., of mammalian origin or from a heterologous mammalian or non-mammalian species). For example, for expression in insect cells a suitable baculovirus expression vector, such as PFASTBAC, is used to produce recombinant baculovirus particles. The baculovirus particles are amplified and used to infect insect cells to express recombinant protein. For expression in mammalian cells, a vector that will drive expression of the construct
35 in the desired mammalian host cell (e.g., Chinese hamster ovary cells) is used.

The polypeptide can be purified using any suitable methods. For example, methods for purifying a polypeptide by immunoaffinity chromatography are known in the art. Suitable methods for purifying desired polypeptides including precipitation and various types of

chromatography, such as hydrophobic interaction, ion exchange, affinity, chelating and size exclusion are known in the art. Suitable purification schemes can be created using two or more of these or other suitable methods. If desired, the polypeptide may include a "tag" that facilitates purification, such as an epitope tag or a histidine tag. Such tagged polypeptides can be purified, for example from conditioned media, by chelating chromatography or affinity chromatography.

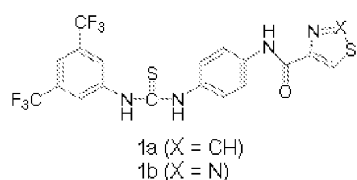
1. Nucleic Acids Encoding polypeptides

In another aspect, the invention relates to nucleic acid molecules that encode a polypeptide described herein. These nucleic acid molecules include DNA, cDNA, and RNA sequences. The nucleic acid molecule can be incorporated into a vector, such as an expression vector.

In some embodiments, the nucleic acid includes a self-replicating RNA molecule. In some embodiments, the nucleic acid includes a modified RNA molecule. In another aspect, the invention relates to a composition including a nucleic acid according to any one of the embodiments described herein.

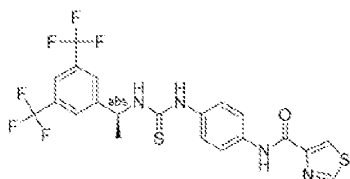
2. Compound-stabilized Polypeptide

The inventors discovered a polypeptide stabilized in a prefusion conformation that can be identified by, for example, the binding of a bis(aryl)thiourea compound to an HCMV gB as described in WO2021/260510 which is hereby incorporated by reference in its entirety. Bis(aryl)thiourea compounds, as exemplified by structures 1a,b (Formula I), are highly potent and specific inhibitors of CMV. In one aspect, the HCMV gB polypeptide is capable of binding to a bis(aryl)thiourea compound. In preferred embodiments, the compound does not bind to a postfusion conformation of the HCMV gB polypeptide.

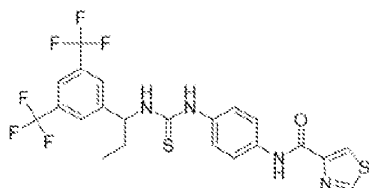


(Formula I)

In a preferred embodiment, the compound is a bis(aryl)thiourea thiozole analog thereof. Most preferably, in some embodiments, the compound is N-{4-[(1S)-1-[3,5-bis(trifluoromethyl)phenyl]ethyl]carbamothioyl}amino]phenyl]-1,3-thiazole-4-carboxamide, having the following structure:



In another embodiment, the compound has the following structure:



5

In several embodiments, the polypeptide includes an HCMV gB prefusion epitope, which is not present in a native HCMV gB postfusion conformation.

In some embodiments, at least about 90% of the polypeptides (such as at least about 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% of the polypeptides in the homogeneous population are bound by a bis(aryl)thiourea compound (e.g., such as a thiazole analog of bis(aryl)thiourea compounds, more preferably N-{4-[(1S)-1-[3,5-bis(trifluoromethyl)phenyl] ethyl]carbamoethyl} amino]phenyl}-1,3-thiazole-4-carboxamide).

In some embodiments, the polypeptide that can bind to the bis(aryl)thiourea compound does not have a postfusion conformation. Rather, the polypeptide has a prefusion conformation, such as an HCMV gB prefusion conformation.

In another embodiment, the polypeptide can be at least 80% isolated, at least 90%, 95%, 98%, 99%, or preferably 99.9% isolated from HCMV gB polypeptides that are not specifically bound by a bis(aryl)thiourea compound.

20 3. Compositions Including a Polypeptide and Methods of Use Thereof

The invention relates to compositions and methods of using the modulators described herein. For example, the modulator of the invention can be delivered directly as a component of an immunogenic composition. Alternatively, if the modulator comprises amino acids, the nucleic acids that encode the amino acid sequence of the invention can be administered to produce the peptide, polypeptide or immunogenic fragment *in vivo*. Certain preferred embodiments, such as protein formulations, recombinant nucleic acids (e.g., DNA, RNA, self-replicating RNA, or any variation thereof) and viral vectors (e.g., live, single-round, non-replicative assembled virions, or otherwise virus-like particles, or alphavirus VRP) that contain

sequences encoding polypeptides are further described herein and may be included in the composition.

In one aspect, the invention provides an immunogenic composition comprising the modulator described herein. The immunogenic composition can include additional CMV proteins, such as gO, gH, gL, pUL128, pUL130, pUL131, pp65, an immunogenic fragment thereof, or a combination thereof. For example, the modulator can be combined with CMV pentameric complex comprising: gH or a pentamer-forming fragment thereof, gL or a pentamer-forming fragment thereof, pUL128 or a pentamer-forming fragment thereof, pUL130 or a pentamer-forming fragment thereof, and pUL131 or a pentamer-forming fragment thereof. The modulator of the invention can also be combined with CMV trimeric complex comprising: gH or a trimer-forming fragment thereof, gL or a trimer-forming fragment thereof, and gO or a trimer-forming fragment thereof.

In another aspect, the invention relates to a composition including a polynucleotide that may elicit an immune response in a mammal. The polynucleotide encodes at least one polypeptide of interest, e.g., an antigen. Antigens disclosed herein may be wild type (i.e., derived from the infectious agent) or preferably modified (e.g., engineered, designed or artificial). The nucleic acid molecules described herein, specifically polynucleotides, in some embodiments, encode one or more peptides or polypeptides of interest. Such peptides or polypeptides may serve as an antigen or antigenic molecule. The term "nucleic acid" includes any compound that includes a polymer of nucleotides. These polymers are referred to as "polynucleotides." Exemplary nucleic acids or polynucleotides of the invention include, but are not limited to, ribonucleic acids (RNAs), including mRNA, and deoxyribonucleic acids (DNAs).

In some embodiments, the composition includes DNA encoding a polypeptide or fragment thereof described herein. In some embodiments, the composition includes RNA encoding a polypeptide or fragment thereof described herein. In some embodiments, the composition includes an mRNA polynucleotide encoding a polypeptide or fragment thereof described herein. Such compositions may produce the appropriate protein conformation upon translation.

In one aspect, the invention relates to a composition that includes at least one polynucleotide encoding a polypeptide including at least one amino acid mutation relative to the amino acid sequence of the wild-type HCMV gB.

In one aspect, the invention relates to a composition that includes at least one DNA polynucleotide encoding a polypeptide including at least one amino acid mutation relative to the amino acid sequence of the wild-type HCMV gB.

In one aspect, the invention relates to a composition that includes at least one RNA polynucleotide encoding a polypeptide including at least one amino acid mutation relative to the amino acid sequence of the wild-type HCMV gB.

In some embodiments, the invention relates to a composition that includes at least one polynucleotide encoding at least one HCMV gB polypeptide or an immunogenic fragment or epitope thereof.

5 In some embodiments, the composition includes at least one polynucleotide encoding two or more antigenic polypeptides or an immunogenic fragment or epitope thereof. In some embodiments, the composition includes two or more polynucleotides encoding two or more antigenic polypeptides or immunogenic fragments or epitopes thereof. The one or more antigenic polypeptides may be encoded on a single polynucleotide or may be encoded individually on multiple (e.g., two or more) polynucleotides.

10 In another aspect, the invention relates to a composition that includes (a) a polynucleotide encoding a polypeptide including at least one introduced amino acid mutation relative to the amino acid sequence of the wild-type HCMV glycoprotein B (gB); and (b) a polynucleotide encoding an additional polypeptide.

In another aspect, the invention relates to a composition that includes (a) a
15 polynucleotide encoding a polypeptide including at least one introduced amino acid mutation relative to the amino acid sequence of the wild-type HCMV glycoprotein B (gB); and (b) a polynucleotide encoding an additional polypeptide, preferably an HCMV antigenic polypeptide. The additional polypeptide may be selected from HCMV gH, gL, gB, gO, gN, and gM and an immunogenic fragment or epitope thereof. In some embodiments, the additional polypeptide is
20 HCMV pp65. In some embodiments, the additional polypeptide may be selected from gH, gL, gO, gM, gN, UL128, UL130, and UL131A, and fragments thereof. In some embodiments, the additional polypeptide is HCMV gH polypeptide. In some embodiments, the additional polypeptide is an HCMV gL polypeptide. In some embodiments, the additional polypeptide is an HCMV gB polypeptide. In some embodiments, the additional polypeptide is an HCMV gO
25 polypeptide. In some embodiments, the additional polypeptide is an HCMV gN polypeptide. In some embodiments, the additional polypeptide is an HCMV gM polypeptide. In some embodiments, the additional polypeptide is a variant gH polypeptide, a variant gL polypeptide, or a variant gB polypeptide. In some embodiments, the variant HCMV gH, gL, or gB polypeptide is a truncated polypeptide lacking one or more of the following domain sequences: (1) the
30 hydrophobic membrane proximal domain, (2) the transmembrane domain, and (3) the cytoplasmic domain. In some embodiments, the truncated HCMV gH, gL, or gB polypeptide lacks the hydrophobic membrane proximal domain, the transmembrane domain, and the cytoplasmic domain. In some embodiments, the truncated HCMV gH, gL, or gB polypeptide includes only the ectodomain sequence. In some embodiments, an antigenic polypeptide is an
35 HCMV protein selected from UL83, UL123, UL128, UL130 and UL131A or an immunogenic fragment or epitope thereof. In some embodiments, the antigenic polypeptide is an HCMV UL83 polypeptide. In some embodiments, the antigenic polypeptide is an HCMV UL123 polypeptide. In some embodiments, the antigenic polypeptide is an HCMV UL128 polypeptide. In some

embodiments, the antigenic polypeptide is an HCMV UL130 polypeptide. In some embodiments, the antigenic polypeptide is an HCMV UL131 polypeptide.

In another aspect, the invention relates to a composition that includes (a) a polynucleotide encoding a polypeptide including at least one introduced amino acid mutation relative to the amino acid sequence of the wild-type HCMV glycoprotein B (gB); and (b) a polynucleotide encoding an additional polypeptide having any one of the amino acid sequences SEQ ID NOs: 211-223. In another aspect, the invention relates to a composition that includes (a) a polynucleotide encoding a polypeptide including at least one introduced amino acid mutation relative to the amino acid sequence of the wild-type HCMV glycoprotein B (gB); and (b) a polynucleotide having any one of the sequences selected from SEQ ID NOs: 141-210. In another aspect, the invention relates to a composition that includes (a) a polynucleotide encoding a polypeptide including at least one introduced amino acid mutation relative to the amino acid sequence of the wild-type HCMV glycoprotein B (gB); and (b) an additional polypeptide having any one of the amino acid sequences selected from SEQ ID NOs: 211-223. In some embodiments, the polynucleotide encoding the additional polypeptide includes at least one nucleic acid sequence selected from any of SEQ ID NOs: 224-254. In some embodiments, the polynucleotide encoding the additional polypeptide includes at least one nucleic acid sequence selected from any of SEQ ID NOs: 141-147. In some embodiments, the polynucleotide encoding the additional polypeptide has at least one sequence selected from any of SEQ ID NOs: 220-223.

In some embodiments, the antigenic polypeptide includes two or more HCMV proteins, fragments, or epitopes thereof. In some embodiments, the antigenic polypeptide includes two or more glycoproteins, fragments, or epitopes thereof. In some embodiments, the antigenic polypeptide includes at least one HCMV polypeptide, fragment or epitope thereof and at least one other HCMV protein, fragment or epitope thereof. In some embodiments, the two or more HCMV polypeptides are encoded by a single RNA polynucleotide. In some embodiments, the two or more HCMV polypeptides are encoded by two or more RNA polynucleotides, for example, each HCMV polypeptide is encoded by a separate RNA polynucleotide. In some embodiments, the two or more HCMV polypeptides can be any combination of HCMV gH, gL, gB, gO, gN, and gM polypeptides or immunogenic fragments or epitopes thereof. In some embodiments, the two or more glycoproteins includes pp65 or immunogenic fragments or epitopes thereof; and any combination of HCMV gH, gL, gB, gO, gN, and gM polypeptides or immunogenic fragments or epitopes thereof. In some embodiments, the two or more glycoproteins can be any combination of HCMV gB and one or more HCMV polypeptides selected from gH, gL, gO, gN, and gM polypeptides or immunogenic fragments or epitopes thereof. In some embodiments, the two or more glycoproteins can be any combination of HCMV gH and one or more HCMV polypeptides selected from gL, gO, gN, and gM polypeptides or immunogenic fragments or epitopes thereof. In some embodiments, the two or more

glycoproteins can be any combination of HCMV gL and one or more HCMV polypeptides selected from gB, gH, gO, gN, and gM polypeptides or immunogenic fragments or epitopes thereof. In some embodiments, the two or more HCMV polypeptides are gB and gH. In some embodiments, the two or more HCMV polypeptides are gB and gL. In some embodiments, the two or more HCMV polypeptides are gH and gL. In some embodiments, the two or more HCMV polypeptides are gB, gL, and gH. In some embodiments, the two or more HCMV proteins can be any combination of HCMV UL83, UL123, UL128, UL130, and UL131A polypeptides or immunogenic fragments or epitopes thereof. In some embodiments, the two or more HCMV polypeptides are UL123 and UL130. In some embodiments, the two or more HCMV polypeptides are UL123 and 131 A. In some embodiments, the two or more HCMV polypeptides are UL130 and 131 A. In some embodiments, the two or more HCMV polypeptides are UL 128, UL130 and 131 A. In some embodiments, the two or more HCMV proteins can be any combination of HCMV gB, gH, gL, gO, gM, gN, UL83, UL123, UL128, UL130, and UL131A polypeptides or immunogenic fragments or epitopes thereof. In some embodiments, the two or more glycoproteins can be any combination of HCMV gH and one or more HCMV polypeptides selected from gL, UL128, UL130, and UL131A polypeptides or immunogenic fragments or epitopes thereof. In some embodiments, the two or more glycoproteins can be any combination of HCMV gL and one or more HCMV polypeptides selected from gH, UL128, UL130, and UL131A polypeptides or immunogenic fragments or epitopes thereof. In some embodiments, the two or more HCMV polypeptides are gL, gH, UL 128, UL130 and 131 A. In any of these embodiments in which the composition includes two or more HCMV proteins, the HCMV gH may be a variant gH, such as any of the variant HCMV gH glycoproteins disclosed herein, for example, any of the variant HCMV gH disclosed herein. In any of these embodiments in which the composition includes two or more HCMV proteins, the HCMV gB may be a variant gB, such as any of the variant HCMV gB glycoproteins disclosed herein, for example, any of the variant HCMV gB disclosed herein. In any of these embodiments in which the composition includes two or more HCMV gL proteins, the HCMV gL may be a variant gL, such as any of the variant HCMV gL glycoproteins disclosed herein, for example, any of the variant HCMV gL disclosed herein.

In certain embodiments in which the composition includes two or more RNA polynucleotides encoding two or more HCMV antigenic polypeptides or an immunogenic fragment or epitope thereof (either encoded by a single RNA polynucleotide or encoded by two or more RNA polynucleotides, for example, each protein encoded by a separate RNA polynucleotide), the two or more HCMV proteins are a variant gB, for example, any of the variant gB polypeptides disclosed herein, and an HCMV protein selected from gH, gL, gO, gM, gN, UL128, UL130, and UL131 polypeptides or immunogenic fragments or epitopes thereof. In some embodiments, the two or more HCMV proteins are a variant gH, for example, any of the variant gH polypeptides disclosed herein, and an HCMV protein selected from gH, gL, gO, gM,

gN, UL128, UL130, and UL131A polypeptides or immunogenic fragments or epitopes thereof. In some embodiments, the two or more HCMV proteins are a variant gH, for example, any of the variant gH polypeptides disclosed herein, and an HCMV protein selected from gH, gL, gO, gM, gN, UL128, UL130, and UL131 polypeptides or immunogenic fragments or epitopes thereof. In some embodiments in which the variant HCMV proteins are variant HCMV gB, variant HCMV gL, and variant HCMV gH, the variant HCMV polypeptide is a truncated polypeptide selected from the following truncated polypeptides: lacks the hydrophobic membrane proximal domain; lacks the transmembrane domain; lacks the cytoplasmic domain; lacks two or more of the hydrophobic membrane proximal, transmembrane, and cytoplasmic domains; and includes only the ectodomain. In some embodiments, the composition includes multimeric RNA polynucleotides encoding at least one HCMV antigenic polypeptide or an immunogenic fragment or epitope thereof. In some embodiments, the composition includes at least one RNA polynucleotide encoding at least one HCMV antigenic polypeptide or an immunogenic fragment or epitope thereof, wherein the 5'UTR of the RNA polynucleotide includes a patterned UTR. In some embodiments, the patterned UTR has a repeating or alternating pattern, such as ABABAB or AABBAABBAABB or ABCABCABC or variants thereof repeated once, twice, or more than 3 times. In these patterns, each letter, A, B, or C represent a different UTR at the nucleotide level. In some embodiments, the 5' UTR of the RNA polynucleotide (e.g., a first nucleic acid) has regions of complementarity with a UTR of another RNA polynucleotide (a second nucleic acid). For example, UTR nucleotide sequences of two polynucleotides sought to be joined (e.g., in a multimeric molecule) can be modified to include a region of complementarity such that the two UTRs hybridize to form a multimeric molecule. In some embodiments, the 5' UTR of an RNA polynucleotide encoding an HCMV antigenic polypeptide is modified to allow the formation of a multimeric sequence. In some embodiments, the 5' UTR of an RNA polynucleotide encoding an HCMV protein selected from UL128, UL130, UL131 is modified to allow the formation of a multimeric sequence. In some embodiments, the 5' UTR of an RNA polynucleotide encoding an HCMV polypeptide is modified to allow the formation of a multimeric sequence. In some embodiments, the 5' UTR of an RNA polynucleotide encoding an HCMV polypeptide selected from gH, gL, gB, gO, gM, and gN is modified to allow the formation of a multimeric sequence. In any of these embodiments, the multimer may be a dimer, a trimer, pentamer, hexamer, heptamer, octamer nonamer, or decamer. Thus, in some embodiments, the 5' UTR of an RNA polynucleotide encoding an HCMV protein selected from gH, gL, gB, gO, gM, gN, UL128, UL130, and UL131 is modified to allow the formation of a dimer. In some embodiments, the 5' UTR of an RNA polynucleotide encoding an HCMV protein selected from gH, gL, gB, gO, gM, gN, UL128, UL130, and UL131A is modified to allow the formation of a trimer. In some embodiments, the 5' UTR of an RNA polynucleotide encoding an HCMV protein selected from gH, gL, gB, gO, gM, gN, UL128, UL130, and UL131 is modified to allow the formation of a pentamer. In some embodiments, the

composition includes at least one RNA polynucleotide having a single open reading frame encoding two or more (for example, two, three, four, five, or more) HCMV antigenic polypeptides or an immunogenic fragment or epitope thereof. In some embodiments, the composition includes at least one RNA polynucleotide having more than one open reading

5 frame, for example, two, three, four, five or more open reading frames encoding two, three, four, five or more HCMV antigenic polypeptides. In either of these embodiments, the at least one RNA polynucleotide may encode two or more HCMV antigenic polypeptides selected from gH, gB, gL, gO, gM, gN, UL83, UL123, UL128, UL130, UL131A, and fragments or epitopes thereof. In some embodiments, the at least one RNA polynucleotide encodes UL83 and UL123. In

10 some embodiments, the at least one RNA polynucleotide encodes gH and gL. In some embodiments, the at least one RNA polynucleotide encodes UL128, UL130, and UL131. In some embodiments, the at least one RNA polynucleotide encodes gH, gL, UL128, UL130, and UL131. In some embodiments, in which the at least one RNA polynucleotide has a single open reading frame encoding two or more (for example, two, three, four, five, or more) HCMV

15 antigenic polypeptides, the RNA polynucleotide further comprises additional sequence, for example, a linker sequence or a sequence that aids in the processing of the HCMV RNA transcripts or polypeptides, for example a cleavage site sequence. In some embodiments, the additional sequence may be a protease sequence, such as a furin sequence. In some embodiments, the additional sequence may be self-cleaving 2A peptide, such as a P2A, E2A,

20 F2A, and T2A sequence. In some embodiments, the linker sequences and cleavage site sequences are interspersed between the sequences encoding HCMV polypeptides.

In some embodiments, at least one RNA polynucleotide includes any nucleic acid sequence selected from any one of nucleic acid sequences disclosed herein, or homologs thereof having at least 80% (e.g., 85%, 90%, 95%, 98%, 99%) identity with a nucleic acid

25 sequence disclosed herein. In some embodiments, the open reading frame is encoded is codon-optimized. Some embodiments include a composition that includes at least one RNA polynucleotide encoding at least one HCMV antigenic polypeptide or an immunogenic fragment thereof and at least one 5' terminal cap. In some embodiments, a 5' terminal cap is 7mG(5')ppp(5')NimpNp.

30 In some embodiments, the at least one polynucleotide includes a nucleic acid sequence selected from any one of SEQ ID NOs: 141-210. In some embodiments, the at least one polynucleotide encodes a polypeptide having at least 90% identity to any one of the amino acid sequences of SEQ ID NOs: 211-223. In some preferred embodiments, the composition does not include a polypeptide having the amino acid sequence SEQ ID NO: 216. In some preferred

35 embodiments, the composition does not include a polynucleotide encoding the amino acid sequence SEQ ID NO: 216. In some preferred embodiments, the composition does not include a polynucleotide having the sequence SEQ ID NO: 152.

In some embodiments, the composition includes at least one polynucleotide, wherein the at least one polynucleotide has at least one chemical modification. In some embodiments, the at least one polynucleotide further includes a second chemical modification. Preferably, the polynucleotide is RAN. In some embodiments, the at least one polynucleotide having at least one chemical modification has a 5' terminal cap. In some embodiments, the at least one chemical modification is selected from pseudouridine, N1-methylpseudouridine, N1-ethylpseudouridine, N1-ethylpseudouridine, 2-thiouridine, 4'-thiouridine, 5-methylcytosine, 2-thio-l-methyl-1-deaza-pseudouridine, 2-thio-l-methyl-pseudouridine, 2-thio-5-aza-uridine, 2-thio-dihydropseudouridine, 2-thio-dihydrouridine, 2-thio-pseudouridine, 4-methoxy-2-thio-pseudouridine, 4-methoxy-pseudouridine, 4-thio-l-methyl-pseudouridine, 4-thio-pseudouridine, 5-aza-uridine, dihydropseudouridine, 5-methoxyuridine and 2'-O-methyl uridine. In some embodiments, the composition includes at least one polynucleotide, wherein at least 80% (e.g., 85%, 90%, 95%, 98%, 99%, 100%) of the uracil in the open reading frame has a chemical modification, optionally wherein the composition is formulated in a lipid nanoparticle. In some embodiments, 100% of the uracil in the open reading frame has a chemical modification. In some embodiments, a chemical modification is in the 5-position of the uracil. In some embodiments, a chemical modification is a N1-methyl pseudouridine.

In some embodiments, the additional polypeptides or immunogenic fragments encoded by the polynucleotide (e.g., in an mRNA composition) are selected from gB, gH, gL, gO, gM, gN, UL83, UL123, UL128, UL130, UL131A, pp65 and IE1 antigens.

In some embodiments, a first composition and a second composition are administered to the mammal. In some embodiments, a first composition includes a polynucleotide encoding a polypeptide including at least one introduced amino acid mutation relative to the amino acid sequence of the wild-type HCMV gB; and a second composition includes a polynucleotide encoding HCMV pp65 or an antigenic fragment or epitope thereof. In some embodiments, a first composition includes a polynucleotide encoding a polypeptide including at least one introduced amino acid mutation relative to the amino acid sequence of the wild-type HCMV gB; and a second composition includes a polynucleotide encoding at least one polynucleotide encoding an additional polypeptide selected from HCMV gH, gL, UL128, UL130, and UL131, or antigenic fragments or epitopes thereof.

In another aspect, the invention relates to methods of inducing an immune response in a mammal, including administering to the mammal a composition in an amount effective to induce an immune response, wherein the composition includes a polynucleotide encoding a modulator of the wild-type HCMV gB.

In some embodiments, the immune response includes a T cell response or a B cell response. In some embodiments, the immune response includes a T cell response and a B cell response. In some embodiments, the method involves a single administration of the composition. In some embodiments, a method further includes administering to the subject a

booster dose of the composition. The composition including a polynucleotide disclosed herein may be formulated in an effective amount to produce an antigen specific immune response in a mammal.

The immunogenic composition may include an adjuvant. Exemplary adjuvants to enhance effectiveness of the composition include: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc.; (2) oil-in-water emulsion formulations (with or without other specific adjuvants such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59 (PCT Publ. No. WO 90/14837), containing 5% Squalene, 0.5% TWEEN 80, and 0.5% Span 85 formulated into submicron particles using a microfluidizer, (b) SAF, containing 10% Squalene, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) RIBI™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, Mont.) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL+CWS (DETOX™); (3) saponin adjuvants, such as QS-21, STIMULON™ (Cambridge Bioscience, Worcester, Mass.), which may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes), ALFQ; (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (IL-1, IL-2, etc.), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc.; and (6) other substances that act as adjuvants to enhance the effectiveness of the composition. In a preferred embodiment, the adjuvant is a saponin adjuvant, namely comprising QS-21. In some embodiments, the composition does not include an adjuvant. In some embodiments, the composition further includes a lipid nanoparticle. In some embodiments, the composition is formulated in a nanoparticle. In some embodiments, the composition further includes a cationic or polycationic compounds, including protamine or other cationic peptides or proteins, such as poly-L-lysine (PLL).

Each of the immunogenic compositions discussed herein may be used alone or in combination with one or more other antigens, the latter either from the same viral pathogen or from another pathogenic source or sources. These compositions may be used for prophylactic (to prevent infection) or therapeutic (to treat disease after infection) purposes.

In one embodiment, the composition may include a "pharmaceutically acceptable carrier," which includes any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as adjuvants. Furthermore, the antigen may

be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, and etc. pathogens.

In one embodiment, the composition includes a diluent, such as water, saline, glycerol, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles.

The compositions described herein may include an immunologically effective amount of the polypeptide or polynucleotide, as well as any other of the above-mentioned components, as needed. By "immunologically effective amount," it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for eliciting an immune response. The immune response elicited may be sufficient, for example, for treatment and/or prevention and/or reduction in incidence of illness, infection or disease. This amount varies depending upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated (e.g., nonhuman primate, primate, etc.), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.

The composition may be administered parenterally, e.g., by injection, either subcutaneously or intramuscularly. In some embodiments, the composition is administered to the mammal by intradermal or intramuscular injection. Additional formulations suitable for other modes of administration include oral and pulmonary formulations, nasal formulations, suppositories, and transdermal applications. Oral formulations may be preferred for certain viral proteins. Dosage treatment may be a single dose schedule or a multiple dose schedule. The immunogenic composition may be administered in conjunction with other immunoregulatory agents.

In another aspect, the invention provides a method of eliciting an immune response against cytomegalovirus, comprising administering to a subject in need thereof an immunologically effective amount of the polypeptide and/or an immunogenic composition described herein, which comprises the proteins, DNA molecules, RNA molecules (e.g., self-replicating RNA molecules), or VRPs as described above. In certain embodiments, the immune response comprises the production of neutralizing antibodies against CMV.

The immune response can comprise a humoral immune response, a cell-mediated immune response, or both. In some embodiments an immune response is induced against each delivered CMV protein. A cell-mediated immune response can comprise a Helper T-cell (Th) response, a CD8+ cytotoxic T-cell (CTL) response, or both. In some embodiments the immune response comprises a humoral immune response, and the antibodies are neutralizing antibodies.

Neutralizing antibodies block viral infection of cells. CMV infects epithelial cells and also fibroblast cells. In some embodiments the immune response reduces or prevents infection of both cell types. Neutralizing antibody responses can be complement-dependent or complement-independent. In some embodiments the neutralizing antibody response is complement-independent. In some embodiments the neutralizing antibody response is cross-neutralizing; i.e., an antibody generated against an administered composition neutralizes a CMV virus of a strain other than the strain used in the composition.

The polypeptide and/or immunogenic composition described herein may also elicit an effective immune response to reduce the likelihood of a CMV infection of a non-infected mammal, or to reduce symptoms in an infected mammal, e.g., reduce the number of outbreaks, CMV shedding, and risk of spreading the virus to other mammals.

In one aspect, the invention relates to a method for reducing CMV viral shedding in a mammal. In some embodiments, the invention relates to a method for reducing CMV viral shedding in urine in a mammal. In some embodiments, the invention relates to a method for reducing CMV viral shedding in saliva in a mammal. In another aspect, the invention relates to a method for reducing CMV viral titers in a mammal. In one aspect, the invention relates to a method for reducing CMV nucleic acids in serum in a mammal. The term "viral shedding" is used herein according to its plain ordinary meaning in medicine and virology and refers to the production and release of virus from an infected cell. In some embodiments, the virus is released from a cell of a mammal. In some embodiments, virus is released into the environment from an infected mammal. In some embodiments the virus is released from a cell within a mammal.

In one aspect, the invention relates to a method for reducing CMV viral shedding in a mammal. The method includes administering the modified CMV gB polypeptide and/or immunogenic composition described herein to the mammal that is infected with or is at risk of a CMV infection. In one embodiment, the reduction in CMV viral shedding in a mammal is as compared to the viral shedding in mammals that were not administered the modified CMV gB. In another embodiment, the reduction in CMV viral shedding in a mammal is as compared to the viral shedding following an administration of a CMV pentamer alone or following an administration of a CMV pentamer in the absence of the polypeptide.

In some embodiments, the mammal is a human. In some embodiments, the human is a child, such as an infant. In some other embodiments, the human is female, including an adolescent female, a female of childbearing age, a female who is planning pregnancy, a pregnant female, and females who recently gave birth. In some embodiments, the human is a transplant patient.

In one embodiment, the challenge cytomegalovirus strain is a human CMV strain. In one embodiment, the challenge cytomegalovirus strain is homologous to the CMV strain from which the polypeptide is derived. In another embodiment, the challenge cytomegalovirus strain

is homologous to the CMV strain VR1814. In another embodiment, the challenge cytomegalovirus strain is homologous to the CMV strain Towne.

In one embodiment, the challenge cytomegalovirus strain is a human CMV strain that is heterologous to the CMV strain from which the modified CMV gB polypeptide is derived. In another embodiment, the challenge cytomegalovirus strain is a human CMV strain that is heterologous to the VR1814 CMV strain. In another embodiment, the challenge cytomegalovirus strain is the VR1814 CMV strain. In another embodiment, the challenge cytomegalovirus strain is a human CMV strain that is heterologous to the CMV strain Towne. In another embodiment, the challenge cytomegalovirus strain is the CMV strain Towne.

In another embodiment, the challenge cytomegalovirus strain is a rhesus CMV strain homologous to the macacine herpesvirus 3 isolate 21252 CMV strain. In another embodiment, the challenge cytomegalovirus strain is the macacine herpesvirus 3 isolate 21252 CMV strain.

A useful measure of antibody potency in the art is "50% neutralization titer." Another useful measure of antibody potency is any one of the following: a "60% neutralization titer"; a "70% neutralization titer"; a "80% neutralization titer"; and a "90% neutralization titer." To determine, for example, a 50% neutralizing titer, serum from immunized animals is diluted to assess how dilute serum can be yet retain the ability to block entry of 50% of infectious viruses into cells. For example, a titer of 700 means that serum retained the ability to neutralize 50% of infectious virus after being diluted 700-fold. Thus, higher titers indicate more potent neutralizing antibody responses. In some embodiments, this titer is in a range having a lower limit of about 200, about 400, about 600, about 800, about 1000, about 1500, about 2000, about 2500, about 3000, about 3500, about 4000, about 4500, about 5000, about 5500, about 6000, about 6500, or about 7000. The 50%, 60%, 70%, 80%, or 90% neutralization titer range can have an upper limit of about 400, about 600, about 800, about 1000, about 1500, about 2000, about 2500, about 3000, about 3500, about 4000, about 4500, about 5000, about 5500, about 6000, about 6500, about 7000, about 8000, about 9000, about 10000, about 11000, about 12000, about 13000, about 14000, about 15000, about 16000, about 17000, about 18000, about 19000, about 20000, about 21000, about 22000, about 23000, about 24000, about 25000, about 26000, about 27000, about 28000, about 29000, or about 30000. For example, the 50% neutralization titer can be about 3000 to about 6500. "About" means plus or minus 10% of the recited value. Neutralization titer can be measured as described in the specific examples, below.

An immune response can be stimulated by administering proteins, DNA molecules, RNA molecules (e.g., self-replicating RNA molecules or nucleoside modified RNA molecules), or VRPs to an individual, typically a mammal, including a human. In some embodiments the immune response induced is a protective immune response, i.e., the response reduces the risk or severity of or clinical consequences of a CMV infection. Stimulating a protective immune response is particularly desirable in some populations particularly at risk from CMV infection

and disease. For example, at-risk populations include solid organ transplant (SOT) patients, bone marrow transplant patients, and hematopoietic stem cell transplant (HSCT) patients. VRPs can be administered to a transplant donor pre-transplant, or a transplant recipient pre- and/or post-transplant. Because vertical transmission from mother to child is a common source of infecting infants, administering VRPs to a woman who is pregnant or can become pregnant is particularly useful.

Any suitable route of administration can be used. For example, a composition can be administered intramuscularly, intraperitoneally, subcutaneously, or transdermally. Some embodiments will be administered through an intra-mucosal route such as intra-orally, intra-nasally, intra-vaginally, and intra-rectally. Compositions can be administered according to any suitable schedule.

Also provided herein is a method of inhibiting cytomegalovirus entry into a cell, comprising contacting the cell with the composition described herein.

In one aspect, the invention relates to compositions that include a modulator described above. In another aspect, the invention relates to compositions that include a nucleic acid molecule or vector encoding such modulator. In a further aspect, the invention relates to compositions that include a modulator described above and a nucleic acid molecule or vector encoding such modulator.

In some embodiments, the composition is an immunogenic composition capable of eliciting an immune response against CMV in a subject. In some particular embodiments, the immunogenic composition is a pharmaceutical composition, which includes a modulator provided by the present disclosure and a pharmaceutically acceptable carrier. In still other embodiments, the pharmaceutical composition is a vaccine.

In some embodiments, a composition, such as an immunogenic composition or a vaccine, includes two or more different modulators described above. The two or more different modulators may include the same introduced amino acid mutations but may be derived from gB from different HCMV strains or subtypes. In another embodiment, the two or more different modulators may include amino acid mutations, as compared to a native HCMV gB, that differ from one another.

In preferred embodiments, the modulator is soluble in aqueous solution. In some embodiments, the modulator is soluble in a solution that lacks detergent.

4. Antibodies and Diagnostic Uses

The modulators described above may be used to produce antibodies, both polyclonal and monoclonal. If polyclonal antibodies are desired, a selected mammal (e.g., mouse, rabbit, goat, guinea pig, horse, etc.) is immunized with an immunogenic modulator. Serum from the immunized animal is collected and treated according to known procedures. If serum containing polyclonal antibodies to a CMV epitope contains antibodies to other antigens, the polyclonal

antibodies can be purified by immunoaffinity chromatography. Techniques for producing and processing polyclonal antisera are known in the art.

Monoclonal antibodies directed against CMV epitopes can also be readily produced by one skilled in the art. The general methodology for making monoclonal antibodies by hybridomas is known. Immortal antibody-producing cell lines can be created by cell fusion, and also by other techniques such as direct transformation of B lymphocytes with oncogenic DNA, or transfection with Epstein-Barr virus. Panels of monoclonal antibodies produced against CMV epitopes can be screened for various properties; i.e., for isotype, epitope affinity, etc.

Antibodies, both monoclonal and polyclonal, which are directed against CMV epitopes are particularly useful in diagnosis, and those which are neutralizing are useful in passive immunotherapy. Monoclonal antibodies, in particular, may be used to raise anti-idiotypic antibodies.

Both the modulators which react immunologically with serum containing CMV antibodies, and the antibodies raised against these modulators, may be useful in immunoassays to detect the presence of CMV antibodies, or the presence of the virus, in biological samples, including for example, blood or serum samples. Design of the immunoassays is subject to a great deal of variation, and a variety of these are known in the art. For example, the immunoassay may utilize the polypeptide having the sequence set forth in any one of SEQ ID NOs: 2-43.

Alternatively, the immunoassay may use a combination of viral antigens derived from the polypeptides described herein. It may use, for example, a monoclonal antibody directed towards at least one polypeptide described herein, a combination of monoclonal antibodies directed towards the polypeptides described herein, monoclonal antibodies directed towards different viral antigens, polyclonal antibodies directed towards the polypeptides described herein, or polyclonal antibodies directed towards different viral antigens. Protocols may be based, for example, upon competition, or direct reaction, or may be sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed by packaging the appropriate materials, including the polypeptides of the invention containing CMV epitopes or antibodies directed against epitopes in suitable containers, along with the remaining reagents and materials required for the conduct of the assay, as well as a suitable set of assay instructions.

The polynucleotide probes can also be packaged into diagnostic kits. Diagnostic kits include the probe DNA, which may be labeled; alternatively, the probe DNA may be unlabeled

and the ingredients for labeling may be included in the kit. The kit may also contain other suitably packaged reagents and materials needed for the particular hybridization protocol, for example, standards, as well as instructions for conducting the test.

Some embodiments of the present disclosure provide a HCMV vaccine that includes at least one ribonucleic acid (RNA) polynucleotide having an open reading frame encoding at least one HCMV antigenic polypeptide or an immunogenic fragment thereof and at least one 5' terminal cap. In some embodiments, a 5' terminal cap is 7mG(5')ppp(5')NlmpNp.

Some embodiments of the present disclosure provide a HCMV vaccine that includes at least one ribonucleic acid (RNA) polynucleotide having an open reading frame encoding at least one HCMV antigenic polypeptide or an immunogenic fragment thereof, wherein the at least one ribonucleic acid (RNA) polynucleotide has at least one chemical modification. In some embodiments, the at least one ribonucleic acid (RNA) polynucleotide further comprises a second chemical modification. In some embodiments, the at least one ribonucleic acid (RNA) polynucleotide having at least one chemical modification has a 5' terminal cap. In some embodiments, the at least one chemical modification is selected from pseudouridine, N1-methylpseudouridine, N1-ethylpseudouridine, 2-thiouridine, 4'-thiouridine, 5-methylcytosine, 2-thio-1-methyl-1-deaza-pseudouridine, 2-thio-1-methyl-pseudouridine, 2-thio-5-aza-uridine, 2-thio-dihydropseudouridine, 2-thio-dihydrouridine, 2-thio-pseudouridine, 4-methoxy-2-thio-pseudouridine, 4-methoxy-pseudouridine, 4-thio-1-methyl-pseudouridine, 4-thio-pseudouridine, 5-aza-uridine, dihydropseudouridine, 5-methoxyuridine and 2'-O-methyl uridine.

In some embodiments, the chemical modification is selected from the group consisting of pseudouridine, N1-methylpseudouridine, N1-ethylpseudouridine, 2-thiouridine, 4'-thiouridine, 5-methylcytosine, 2-thio-1-methyl-1-deaza-pseudouridine, 2-thio-1-methyl-pseudouridine, 2-thio-5-aza-uridine, 2-thio-dihydropseudouridine, 2-thio-dihydrouridine, 2-thio-pseudouridine, 4-methoxy-2-thio-pseudouridine, 4-methoxy-pseudouridine, 4-thio-1-methyl-pseudouridine, 4-thio-pseudouridine, 5-aza-uridine, dihydropseudouridine, 5-methoxyuridine, and 2'-O-methyl uridine.

Some embodiments of the present disclosure provide a HCMV vaccine that includes at least one ribonucleic acid (RNA) polynucleotide having an open reading frame encoding at least one HCMV antigenic polypeptide or an immunogenic fragment thereof, wherein at least 80% (e.g., 85%, 90%, 95%, 98%, 99%, 100%) of the uracil in the open reading frame have a chemical modification, optionally wherein the vaccine is formulated in a lipid nanoparticle. In some embodiments, 100% of the uracil in the open reading frame have a chemical modification. In some embodiments, a chemical modification is in the 5-position of the uracil. In some embodiments, a chemical modification is a N1-methyl pseudouridine.

Some embodiments of the present disclosure provide a HCMV vaccine that is formulated within a cationic lipid nanoparticle, also referred to herein as ionizable cationic lipid nanoparticles, ionizable lipid nanoparticles and lipid nanoparticles, which are used

interchangeably. In some embodiments, the lipid nanoparticle comprises a cationic lipid, a PEG-modified lipid, a sterol and a non-cationic lipid. In some embodiments, the cationic lipid is an ionizable cationic lipid and the non-cationic lipid is a neutral lipid, and the sterol is a cholesterol. In some embodiments, the cationic lipid is selected from the group consisting of 2,2-dilinoleyl-4-dimethylaminoethyl-[1,3]-dioxolane (DLin-KC2-DMA), dilinoleyl-methyl-4-dimethylaminobutyrate (DLin-MC3-DMA), and di((Z)-non-2-en-1-yl) 9-((4-(dimethylamino)butanoyl)oxy)heptadecanedioate (L319). In some embodiments, the lipid nanoparticle has a molar ratio of about 20-60% cationic lipid, about 5-25% non-cationic lipid, about 25-55% sterol, and about 0.5-15% PEG-modified lipid. In some embodiments, the nanoparticle has a polydispersity value of less than 0.4. In some embodiments, the nanoparticle has a net neutral charge at a neutral pH. In some embodiments, the nanoparticle has a mean diameter of 50-200 nm.

In some embodiments, 80% of the uracil in the open reading frame have a chemical modification. In some embodiments, 100% of the uracil in the open reading frame have a chemical modification. In some embodiments, the chemical modification is in the 5-position of the uracil. In some embodiments, the chemical modification is N1-methylpseudouridine, N1-ethylpseudouridine. In some embodiments, the vaccine is formulated within a lipid nanoparticle. In some embodiments, the lipid nanoparticle comprises a cationic lipid, a PEG-modified lipid, a sterol and a non-cationic lipid. In some embodiments, the cationic lipid is an ionizable cationic lipid and the non-cationic lipid is a neutral lipid, and the sterol is a cholesterol. In some embodiments, the cationic lipid is selected from the group consisting of 2,2-dilinoleyl-4-dimethylaminoethyl-[1,3]-dioxolane (DLin-KC2-DMA), dilinoleyl-methyl-4-dimethylaminobutyrate (DLin-MC3-DMA), and di((Z)-non-2-en-1-yl) 9-((4-(dimethylamino)butanoyl)oxy)heptadecanedioate (L319).

Some embodiments of the present disclosure provide methods of inducing an antigen specific immune response in a subject, comprising administering to the subject a HCMV RNA vaccine in an amount effective to produce an antigen specific immune response. In some embodiments, an antigen specific immune response comprises a T cell response or a B cell response. In some embodiments, an antigen specific immune response comprises a T cell response and a B cell response. In some embodiments, a method of producing an antigen specific immune response involves a single administration of the vaccine. In some embodiments, a method further includes administering to the subject a booster dose of the vaccine. In some embodiments, a vaccine is administered to the subject by intradermal or intramuscular injection.

Also provided herein are HCMV RNA vaccines for use in a method of inducing an antigen specific immune response in a subject, the method comprising administering the vaccine to the subject in an amount effective to produce an antigen specific immune response.

Further provided herein are uses of HCMV RNA vaccines in the manufacture of a medicament for use in a method of inducing an antigen specific immune response in a subject, the method comprising administering the vaccine to the subject in an amount effective to produce an antigen specific immune response.

- 5 Further provided herein are methods of preventing or treating HCMV infection comprising administering to a subject the vaccine of the present disclosure. The HCMV vaccine disclosed herein may be formulated in an effective amount to produce an antigen specific immune response in a subject.

The term "polypeptide variant" refers to molecules which differ in their amino acid
10 sequence from a native or reference sequence. The amino acid sequence variants may possess substitutions, deletions, and/or insertions at certain positions within the amino acid sequence, as compared to a native or reference sequence. Ordinarily, variants possess at least 50% identity to a native or reference sequence. In some embodiments, variants share at least 80%, or at least 90% identity with a native or reference sequence.

- 15 In some embodiments "variant mimics" are provided. As used herein, the term "variant mimic" is one which contains at least one amino acid that would mimic an activated sequence. For example, glutamate may serve as a mimic for phosphoro-threonine and/or phosphoro-serine. Alternatively, variant mimics may result in deactivation or in an inactivated product containing the mimic, for example, phenylalanine may act as an inactivating substitution for
20 tyrosine; or alanine may act as an inactivating substitution for serine. "Orthologs" refers to genes in different species that evolved from a common ancestral gene by speciation. Normally, orthologs retain the same function in the course of evolution. Identification of orthologs is critical for reliable prediction of gene function in newly sequenced genomes. "Analog" is meant to include polypeptide variants which differ by one or more amino acid alterations, for example,
25 substitutions, additions or deletions of amino acid residues that still maintain one or more of the properties of the parent or starting polypeptide.

- The present disclosure provides several types of compositions that are polynucleotide or polypeptide based, including variants and derivatives. These include, for example, substitutional, insertional, deletion and covalent variants and derivatives. The term "derivative"
30 is used synonymously with the term "variant" but generally refers to a molecule that has been modified and/or changed in any way relative to a reference molecule or starting molecule.

- As such, polynucleotides encoding peptides or polypeptides containing substitutions, insertions and/or additions, deletions and covalent modifications with respect to reference sequences, in particular the polypeptide sequences disclosed herein, are included within the
35 scope of this disclosure. For example, sequence tags or amino acids, such as one or more lysines, can be added to peptide sequences (e.g., at the N-terminal or C-terminal ends). Sequence tags can be used for peptide detection, purification or localization. Lysines can be used to increase peptide solubility or to allow for biotinylation. Alternatively, amino acid residues

located at the carboxy and amino terminal regions of the amino acid sequence of a peptide or protein may optionally be deleted providing for truncated sequences. Certain amino acids (e.g., C-terminal or N-terminal residues) may alternatively be deleted depending on the use of the sequence, as for example, expression of the sequence as part of a larger sequence which is soluble, or linked to a solid support. "Substitutional variants" when referring to polypeptides are those that have at least one amino acid residue in a native or starting sequence removed and a different amino acid inserted in its place at the same position. Substitutions may be single, where only one amino acid in the molecule has been substituted, or they may be multiple, where two or more amino acids have been substituted in the same molecule.

10

EXAMPLES

The invention is further described by the following illustrative examples. The examples do not limit the invention in any way. They merely serve to clarify the invention.

15

EXAMPLE 1: Isolation and Purification of Crosslinked and Native HCMV gB (Towne strain) with Fusion Inhibitor

During the sample preparation the HCMV fusion inhibitor (compound 28 described in Bloom et al., *Bioorganic & Medicinal Chemistry Letters* 14 (2004) 3401–3406; see also **FIG. 5D**) was added to each step during the virus concentration, processing, extraction and purification to inhibit conversion of gB to the postfusion form.

Following crosslinking of the proteins on the virion surface with bis(sulfosuccinimidyl) glutarate (BS²G) and extraction of gB from the virion with detergent, the SM5-1 His/Strep-tagged Fab (Potzsch et al., *PLoS pathogens* 7(8):e1002172, 2011) was added to assist in purification and identification of gB by electron cryomicroscopy. The Fab-gB complexes were purified by an affinity column.

These extracted and purified proteins were then analyzed by electron cryomicroscopy for the presence of prefusion gB and used to solve the structure of a prefusion form.

EXAMPLE 2: Electron microscopy

Graphene oxide film-supported electron microscopy grids were prepared. The gB sample solutions were vitrified using a Vitrobot (ThermoFisher). The frozen grids were transferred to a FEI Titan Krios transmission electron microscope that operates at 300kV. Target positions were set up in the SerialEM program, and high magnification (18000X) images were automatically collected with the program using a K2 direct detector camera (Gatan) using super resolution movie mode. The unbinned pixel size was 0.638 Å and the beam intensity was ~8e/unbin pixel/s. The total electron dose on the sample for each movie was ~40e/Å². A total of 7,771 movies, each with 28 frames, was collected in three sessions.

Image processing

Drift correction was done using the MotionCor2 program (Zheng S et al. *Nature Methods* 14, 331–332 (2017)), and the final micrographs were binned 2X and averaged from all frames. Contrast transfer function parameters were calculated with Gctf (Kai Zhang, *Journal of Structural Biology* 193(1), 1-12 (2016)). For particle picking, the published structure of HCMV gB in postfusion conformation (PDB:5CXF) was used to generate a 30Å density map using pdb2mrc (EMAN) (Ludtke, S. et al. *Journal of Structural Biology* 128(1), 82-97 (1999)). Projection images from this density map was generated with project3d (EMAN) (FIG. 1) and used as a template for the automatic particle picking using Gautomatch program (Urnavicius L, et al. *Science* 347(6229):1441-1446 (2015)). Relion v2.1-beta (Scheres, S.H. *Journal of Structural Biology* 180(3): 519-530 (2012)) was used to extract the resulting ~1.9 million particles and to carry out all subsequent image processing steps, including 2D classification, 3D classification, auto-refinement and post-processing. The 2D classes were put into three groups based on the image features: the first group consisted of the 2D classes that showed features that resemble the crystallographically determined postfusion gB structure (>50%); the second group contained 2D classes with well resolved protein features that do not resemble the structural features from postfusion gB (<10%); the third group contained 2D classes that did not contain clearly defined protein (~40%) (FIG. 1). The first and second groups were further processed with 3D classification, auto refinement and post processing procedures with Relion. Following this processing, a ~3.5 Å resolution electron density map showing the postfusion conformation structure was reconstructed from the first group; a ~3.6 Å resolution electron density map showing a prefusion conformation structure was reconstructed from the second group. Based on these density maps and the known HCMV gB amino acid sequence (Towne strain P13201, SEQ ID NO:1), atomic models were built with the Coot program (Emsley P. et al *Acta Crystallogr D Biol Crystallogr* 66(Pt 4): 486-501 (2010)) for the prefusion and postfusion conformation structures. The postfusion gB crystal structure (PDB accession code 5CXF) and a crystal structure of a complex between the SM5-1 fab and gB domain II (PDB accession code 4OT1) were used as initial models for both structures. For the postfusion structure model, small adjustment was enough to obtain a good fit to the electron density. For the prefusion conformation model, domains I, II, III and IV from the reference PDB model could be docked as rigid bodies into the electron density map as a starting point. Then, adjustments of individual residues were made for optimal fitting. The model for domains V, MPR and TM were built *de novo*. The models were iteratively refined with the Phenix.real_space_refine tool (Afonine PV et al. *Acta Crystallogr D Struct Biol* 74(Pt 6): 531-544 (2018)) followed by local manual adjusting for several rounds.

Results

Sample screening by cryoEM

The prefusion conformation of gB is unstable, with a propensity to rearrange to the postfusion state, including during sample handling. Therefore, the samples studied contained a mixture of gB conformers, complicating structure determination. In addition, there was no pre-existing reliable information on the arrangement of domains or the unique structural features of prefusion gB. We used direct visualization by electron microscopy and image processing to screen different sample preparation conditions. Image sorting by 2D and 3D classification permits multiple structures to be determined from heterogeneous samples. However, it requires a large data set so that enough particles for each structure can be combined to produce a class average with good signal. This was especially the case for the gB samples because prefusion gB was a small population in the mixtures. Therefore, we collected ~1,000 movies for each condition, and decided whether to pursue image processing with more data from the same sample or switch to another at the 2D classification stage. The structure of antibody Fab-bound postfusion conformation gB was readily obtained from many datasets. The projection images from these Fab-bound postfusion conformation structures were used as a reference to avoid selecting images for the prefusion image reconstruction. We selected any good class average with protein features that did not resemble any of the postfusion gB projection images for further image processing. We screened dozens of conditions for sample preparation with this strategy and eventually found a sample that produced some alternative 2D classes as a minor species in the particle populations (**FIG. 1B**, circled). Then a total of 7,771 movies were collected from that sample and used for determination of a prefusion gB structure. Projection images of the antibody Fab-bound postfusion gB structure are shown in **FIG. 1A**. The 2D class averages from the dataset collected are shown in **FIG. 1B**. Some classes that do not resemble any of the postfusion gB reference 2D projections are circled.

Obtaining a prefusion conformation structure

Approximately 1.9 million raw particle images were automatically selected from the data set. After 2D classification, the images were grouped into a postfusion class (55% of the particle population) and a prefusion class (10% of the particle population). The two groups were further processed in 3D with C3 symmetry applied to yield a density map of SM5-1 Fab-bound postfusion gB at 3.5 Å resolution and a density map of SM5-1 Fab-bound prefusion gB at 3.6 Å resolution.

The X-ray crystallography-based models of the SM5-1 Fab and of the ectodomain of postfusion gB were fit to the postfusion density map with rigid body docking. Except for the constant domain of the Fab (which is likely too flexible to produce strong electron density), the density map of the postfusion gB-Fab complex and the model agreed well with each other (**FIG. 3A**). The membrane proximal region, transmembrane region and cytoplasmic domain were not resolved in our final postfusion gB density map, suggesting that these regions of postfusion gB are flexible either intrinsically or through detergent solubilization in the sample preparations (**FIG. 2**, lower line). The interaction of the Fab and DII of postfusion gB in the electron

cryomicroscopy-based model agrees well with the previously determined crystal structure of the complex (PDB accession code 4OT1).

To build a prefusion gB model, guided by the known Fab binding position, domains I, II, III and part of domain IV from the postfusion gB crystal structure were docked into the density map of the prefusion gB-Fab complex individually and individual residues were manually adjusted as necessary for optimal fit of the electron density. The rest of the prefusion gB structure was built *de novo*. The amino acids of gB that were modeled in the prefusion structure are indicated in **FIG. 2**, the top line. The model of the prefusion gB-Fab complex fits most parts of the prefusion density map, and the presence of Fab density confirms the identity of gB in the novel structure (**FIG. 3B**).

The coordinates and structure factors for the model of the prefusion gB associated with the present Example are provided in **Table 1A** described in WO2021/260510 which is hereby incorporated by reference in its entirety.

The structure of gB in a prefusion conformation and comparison to postfusion gB

The electron density for the complex of prefusion gB and the SM5-1 Fab allowed the building of a prefusion gB model that includes the gB ectodomain, membrane proximal region (MPR - a helical region that is oriented parallel to the viral membrane), and single span transmembrane helix (TM) (**FIG. 3B** and **FIG. 4B**). The MPR and TM regions were not resolved in the structural data for postfusion gB or included in postfusion gB models.

The overall dimensions of prefusion and postfusion gB are different (**FIG. 4A** vs. **FIG. 4B**). The postfusion gB trimer ectodomain has a rod shape, with an approximate height of 165 Å (the distance between planes formed by proline 570 of each protomer at the membrane distal end and tryptophan 240 of the each protomer at the membrane proximal end; **FIG. 4A**). It has a width of approximately 65Å (the distance between alanine 315 on adjacent protomers). The structures described here were derived from gB of HCMV strain Towne. Although there is some natural variations of gB amino acid sequence, the overall postfusion structure of Towne gB is almost identical to the postfusion structure of gB from the strain AD169 (PDB accession code 5CXF). Thus, the description of the postfusion gB structure applies to both strains with measurements from equivalent amino acids from sequence alignments.

The prefusion gB trimer has a more squat shape than the postfusion gB trimer (**FIG. 4A** vs. **FIG. 4B**). The distance between the plane formed by W240 of each protomer and the most membrane distal modeled residue in the prefusion structure, Q483, is roughly 115Å. The prefusion model is 95Å in width (measured by the distance between any two A315 from different protomers).

The individual subunit structures of domains I, II, III and IV are similar in the prefusion and postfusion conformations. However, the overall arrangement of these domains is very different in the two conformations (**FIGs. 4A-4B** and **FIGs. 6A-6C**). In the prefusion conformation, the fusion loops at the tip of DI and the C-termini of the central helix bundle in

domain III all point in the same direction, toward the virion envelope, as identified by the position of the TM region (**FIG. 4A** and **FIG. 6A**). In contrast, in the postfusion conformation, the fusion loops and the C-termini of the central helix bundle point in opposite directions (**FIG. 4B** and **FIG. 6C**).

5 In the prefusion structure, the hydrophobic residues in the fusion loops (residues Y155, I156, H157 and W240, L241) are in close proximity to the MPR and are likely surrounded by detergents (**FIG. 4A** and **FIG. 6A**).

In the transition from prefusion to post fusion, domain II shifts from a position mid-way up the domain III central coiled-coil to a position at the membrane proximate end of the coiled-coil and near end of domain I opposite the fusion loops (**FIG. 4A** and **FIG. 4B**).

10 The structure of DIII (**FIGs. 4A-4B** and **FIGs. 6A-6C**) is very similar in the prefusion and postfusion conformations. The central helix in both conformations spans from L479 to P525, indicating a minimal rearrangement during the prefusion to postfusion transition. However, the other domains change their positions relative to the central helix of domain III, so that, as noted above, the direction of the DIII helix bundle (from N-terminal to C-terminal) points away from the fusion loops towards the distal end of the trimer in the postfusion conformation and toward the viral membrane, in the same direction as the fusion loops in the prefusion conformation.

In the prefusion structure, domain IV (**FIG. 4A** and **FIG. 6A**) is buried at the interface between domain I on the exterior of the trimer and domains III and V at the center of the trimer. In contrast, in the postfusion structure, domain IV forms a highly exposed "crown" at the membrane-distal tip of the trimer.

Domain V has different structures in prefusion gB (**FIG. 4A** and **FIG. 6A**) and postfusion gB (**FIG. 4B** and **FIG. 6C**). In prefusion gB, the N-terminal half of the domain (about residues 642-660) is sandwiched between domain I and domain IV of an adjacent protomer and is sequestered from solvent. The region between residue 683-704 of domain V forms a trimeric helix bundle with its counterpart in other protomers. This helix bundle is cuddled mostly inside of the pocket of the "crown" formed by domain IV. There is an additional short helix (approximately residues 710-719) linking the helix bundle from domain V to the MPR region. In contrast, in the postfusion conformation (**FIG. 4B** and **FIG. 6C**), domain V is solvent exposed and extends along the outside of domain III helix bundle and the groove formed by the interface between domain I from adjacent protomers.

Comparison of the prefusion and postfusion gB structures suggests a progression of conformational changes that is familiar from other well-studied fusion proteins (Harrison, S.C. *Virology* 0:498–507 (2015)). The comparison provides confidence that the structure described in this invention is, in fact, in a prefusion conformation. In the prefusion state (**FIG. 6A**), the fusion loops of domain I are buried by interaction with the MPR and potentially with the viral membrane. In the prefusion structure of the distant gB homolog, the vesicular stomatitis virus G glycoprotein, the fusion loops also point toward the viral membrane (also the anticipated

position of an MPR region, which is not seen in that structure) (Roche et al. *Science* 315:843-8 (2007)).

Based on analogy to other fusion proteins, it is likely that rearrangement proceeds with lengthening of the central helix as part of a transition to a proposed extended intermediate between the prefusion and postfusion states (**FIG. 6B**). In the proposed extended intermediate state, the TM region would still be anchored in the viral membrane, and the fusion loops, now extended far from the viral membrane at the tips of a rotated and translocated domain I, would interact with a cellular membrane. The transition from the proposed extended intermediate to the postfusion conformation would involve a fold-back so that the transmembrane region and the fusion loops are again in proximity to each other at the same end of the molecule, this time both interacting with the fused viral and cellular membrane (**FIG. 6C**).

We speculate that, in prefusion gB, there may be dynamic changes in the length of the central helix, with the prefusion structure we have determined representing a “snapshot” of a “breathing” molecule, locked into the conformation we see in the electron density by the fusion inhibitor and by the cross-linking agent used to prepare the sample studied by electron cryomicroscopy.

Stabilizing factors for the observed prefusion conformation

After modeling the gB amino acids into the electron density map, a region of density that was not filled by amino acid residues remained between the MPR, domain V, and the tip of domain I that contains the fusion loops (**FIG. 5A**). The size and shape of the unfilled density fits the chemical structure of the HCMV fusion inhibitor, N-{4-[(1S)-1-[3,5-bis(trifluoromethyl)phenyl]ethyl]carbamothioyl}amino]phenyl]-1,3-thiazole-4-carboxamide (**FIG. 5D**), which had been present throughout the production of the sample studied by electron cryomicroscopy (**FIG. 5B**). The compound adopted a pose with a kink between the trifluoromethyl phenyl moiety and the rest of the compound. The thiazole forms contacts with hydrophobic residues of L712, A738 and Y153, Y155 from an adjacent protomer. The phenyl is surrounded in a hydrophobic environment formed by residues of L715, the aliphatic hydrocarbon of D714 from domain V, G734 and I 730 from MPR, and F752 from the TM domain of an adjacent protomer. The trifluoromethyl phenyl resides in a hydrophobic environment near the hinge between MPR and TM helices from another protomer. It may act as a hook to prevent the outward movement of MPR and TM domains. In addition to the interaction coordinated by the inhibitor compound, the W240, Y242 from other fusion loop are forming van der waals interactions with the hydrophobic patch from the MPR region and L715 in domain V respectively. (**FIG. 5C**). These specific interactions around the fusion inhibitor would be expected to hold domain I, domain V, and the MPR together and restrict movements among domain I, domain V, and the MPR during the fusion process (**FIGs. 6A-6C**).

The effects of cross linking on the stability of the prefusion conformation were also tested. During the sample preparation steps, BS²G cross linking reagents either were or were

not added. In the absence of the cross linker, the ratio of particles in prefusion versus postfusion conformations was 1:100, while the ratio was 1:4 in the sample that had been cross linked by the BS²G reagent. The cross linker was not identified in the electron density.

5 The prefusion structure of CMV gB and color versions of the prefusion and postfusion structures set forth in the Figures described herein may also be found in Liu et al. *Science Advances* 7(10): eabf3178 (2021), which is hereby incorporated by reference herein in its entirety.

EXAMPLE 3: Expression and purification of gB1666

10 For the production of gB1666, the PSB1666 construct was transiently transfected into Expi293F cells. The cell pellets were harvested 96 hours after transfection. The PSB1666 protein was purified in 25mM HEPES pH 7.5, 250mM NaCl, 0.02% DDM, 0.002% CHS, 3μg/ml WAY-174865 (inhibitor, see FIG. 5D) through a series of processes of solubilization, affinity column and size exclusion chromatography. The protein was analyzed on SDS-PAGE and by
15 EM with negative staining to ensure at least 50% of the proteins displaying prefusion conformation. The PSB1666 protein is expressed efficiently in transfection of Expi293F cells and 1L expression would generate ~0.1mg of purified PSB1666 in high quality.

The polypeptide gB1666 (PSB1666) (SEQ ID NO: 57) includes a mutation in Domains I and IV. The polypeptide includes the following mutations, D217C and Y589C, relative to the
20 corresponding wild-type gB (Towne) set forth in SEQ ID NO: 1.

EXAMPLE 4: DNA-expressed gB1666 is immunogenic in Balb/c mice

One of the proposed stabilized full length prefusion gB constructs, gB1666 (SEQ ID NO: 57), has been shown by EM to have an increased proportion of molecules in the prefusion
25 conformation relative to wild type gB of the Towne strain after purification from transfected mammalian cells in the presence of a fusion inhibitor (WAY-174865; see FIG. 5D). To assess whether this molecule can elicit immune responses *in vivo*, the DNA sequence corresponding to gB1666 and wild type gB were cloned into an in-house mammalian expression vector. Ten Balb/c mice were electroporated with 100 ug of DNA encoding gB1666 twice at a three-week
30 interval (D0 and D21). An additional 10 mice were electroporated by the same protocol with DNA encoding wild type gB, and a third group was electroporated with a placebo, consisting of phosphate-buffered saline. Serum samples were collected at Day 28. ELISA was performed against recombinant gB protein produced from mammalian cells, based on the wild type sequence of Towne strain but with the transmembrane domain removed (Sino Biologicals) to
35 determine the anti-gB IgG responses according to a standard protocol. Ten out of ten animals from the wild type gB DNA immunized mice and nine of ten gB1666 DNA immunized mice generated detectable anti-gB IgG titers (FIG. 11, showing mean ± SD, LLOQ = 25). The study demonstrates that gB1666 is immunogenic in Balb/c mice.

EXAMPLE 5: Immunogenicity study of stabilized prefusion gB1666 protein**Immunogenicity study of gB1666 in mice.**

- To evaluate the antibody response in mice, the following immunization scheme will be followed. At week 8, mice will be exsanguinated and the neutralization titers from the immunized animal serum will be determined and compared with those immunized with gB705 (postfuion) and/or gB wild type proteins.

Table 5. Mouse immunogenicity study design with gB1666 protein

Group	No. of Mice	Immunogen	Adjuvant	Route	Dosing Schedule
1	10	gB705 (postfusion) (1.25 mcg/0.2ml)	-	0.2ml/SC	Weeks 0, 3, 6
2	10	gB705 (postfusion) (0.25 mcg/0.2ml)	-	0.2ml/SC	Weeks 0, 3, 6
3	10	gB1666 (in inhibitor-containing buffer) (1.25 mcg/0.2ml)	-	0.2ml/SC	Weeks 0, 3, 6
4	10	gB1666 (in inhibitor-containing buffer) (0.25 mcg/0.2ml)	-	0.2ml/SC	Weeks 0, 3, 6
5	10	gB wt (in inhibitor-containing buffer) (1.25 mcg/0.2ml)	-	0.2ml/SC	Weeks 0, 3, 6
6	10	gB wt (in inhibitor-containing buffer) (0.25 mcg/0.2ml)	-	0.2ml/SC	Weeks 0, 3, 6
7	5	Buffer (+Inhibitor)	-	0.2ml/SC	Weeks 0, 3, 6
8	5	Buffer only	-	0.2ml/SC	Weeks 0, 3, 6

10

EXAMPLE 6

- In Example 2, we disclosed the electron cryomicroscopy (cryoEM) structure of prefusion human cytomegalovirus (HCMV) strain Towne glycoprotein B (gB) in complex with an antibody fragment. The gB used for structure determination was obtained by adding a small molecule fusion inhibitor, WAY-174865, to a fermentation of authentic HCMV in mammalian cell culture and maintaining the presence of the inhibitor throughout production and analysis of gB;

purifying the virus; treating the virus with a chemical cross linker, bis(sulfosuccinimidyl) glutarate (BS²G; 7.7Å spacer arm); extracting gB from the virus with detergent; binding gB on the virion with an affinity tagged antibody fragment; and purifying the gB by affinity and sizing columns. We also disclosed the use of the prefusion gB cryoEM structure to engineer

5 mutations that stabilize gB in the prefusion state. Specifically, we disclosed the recombinant gB protein gB1666, in which two residues are mutated to cysteine (D217C, Y589C). The resulting formation of an engineered disulfide bond between C217 and C589 increases the conformational stability of the recombinant gB in the prefusion state. gB1666 maintained prefusion structural features when it was expressed in Expi293F cells and purified in the

10 presence of a fusion inhibitor, compound WAY-174865. In the absence of the inhibitor, gB1666 tends to undergo a conformational change and lose its prefusion structural state. Loss of prefusion conformational stability in the absence of inhibitor is not a desirable characteristic for use of the recombinant glycoprotein as an antigen for immunization. Even if gB1666 were formulated with the inhibitor, there is a risk that, upon injection into a person or animal, the

15 dilution of the inhibitor in vivo would lead to its dissociation from gB1666 and the loss of prefusion conformation of gB1666. Thus, it is desired that HCMV gB be stabilized sufficiently in the prefusion conformation to remain in the prefusion state in the absence of WAY-174865. It is also preferable that a prefusion gB immunogen includes a soluble ectodomain to improve manufacturability, improve solubility, improve homogeneity, and reduce or eliminate the need

20 for formulation with a detergent or other excipient to prevent aggregation or precipitation mediated by the gB transmembrane region.

We now report the invention, through a structure-based engineering approach, of new mutations in HCMV gB that confer these improved characteristics for use of prefusion gB as an immunogen. First, we determined the structure by cryoEM of gB1666, which was solubilized by

25 anchoring in nanodiscs and stabilized in the prefusion conformation by the presence of WAY-174865 (**FIG. 12**). Most of the new structure of the recombinant, D217C and Y589C mutant gB is similar to the structure of the virion-derived, chemically cross-linked and antibody fragment bound HCMV Towne prefusion gB that we determined previously, but there are subtle differences between the two structures in certain local regions. The difference in the structures

30 could reflect several differences in the preparations: first, the presence of the engineered disulfide bond in gB1666, which should restrict the breathing motion of the glycoprotein; second, the anchoring of gB1666 in a nanodisc, which provides a more natural local lipid environment for the transmembrane domain than the detergents used to extract and maintain gB in solution for the previous structure determination; third, the absence of chemical cross-

35 linking of gB1666; fourth, the higher resolution of the new structure at 3.3Å, compared to the 3.6Å resolution of the previous structure, allowing more accurate modeling of amino acid side chains.

Based on the new structural information, we designed additional stabilizing mutations on the background of the full length gB construct pSB1666 (**Table 6** and **Table 7**). We hypothesized that adding these additional mutations on the pSB1666 background would further stabilize the gB in a prefusion state (**FIG. 13**). For example, cysteine mutations at residues

5 M371 and W506 may introduce a disulfide bond between domains II and III; cysteine mutations at the pairs of (F541, E681) and (N524, M684) may introduce disulfide bonds between domains IV and V; mutations of residues E686, D679 to hydrophobic residues could remove a locally destabilizing same charge repulsion patch and increase protein stability. Recombinant glycoproteins with a selection of the new, added mutations were expressed and purified in the

10 absence of fusion inhibitor and without chemical crosslinking. The electrophoretic mobility of the expressed glycoproteins by SDS-PAGE showed the expected apparent molecular weight and heterogeneity consistent with glycosylation (**FIG. 14**). The samples were stored at 4 °C, and aliquots were taken for negatively stained electron microscopy analysis on day 1 and day 7. In the 2D class averaged images, triangular shape features that resemble “top views” of the

15 prefusion conformation of the gB were apparent. The ratio of particles in the population belonging to prefusion and postfusion classes were 5:1 on day 1 and 3:1 at day 7 (**FIG. 15**).

Based on the new structural information, we designed several soluble, detergent-free gB ectodomains (**Table 8**) with prefusion-stabilizing mutations as illustrated in **FIGs. 16A-16D**. The purified ectodomain of HCMV gB, residues 1-707, formed rosette-like aggregates, in which gB

20 proteins associated through their exposed fusion loops. To eliminate aggregation and increase protein secretion to the condition media, we replaced four exposed hydrophobic residues within the fusion loops with the corresponding more hydrophilic amino acids from herpes simplex virus-1 (HSV-1) gB, e.g. Y1H (155-157)→GHR, W240→A. We also mutated the exposed Cys246 to Ser (C246→S) to prevent formation of spurious disulfide bonds. To further stabilize

25 the prefusion trimeric state of the antigen, we either introduced cysteine residues capable of forming inter-protomer disulfide bonds or appended C-terminal trimerization motifs, e.g. GCN4 or foldon from T4-bacteriophage fibrin. Disulfide mutations, e.g. D217C-Y589C, M317C-W506C, N524C-M684C, were further introduced to lock the proteins into the prefusion state. Recombinant glycoproteins were expressed, secreted to the conditioned media and purified in

30 the absence of fusion inhibitor and without chemical detergent. Notably, the recombinant variant, fused to a GCN4 trimerization motif, showed optimal size-exclusion chromatography profile (**FIG. 17**). The negatively stained electron microscopy showed recombinant proteins, gB2555 and gB2556, as monodispersed proteins in the absence of inhibitor and detergents. Expected gB protein features are observed in the 2D class averaged images (**FIG. 18** and **FIG.**

35 **19**). These results confirm that these engineered constructs are suitable to be used as a framework to add more stabilizing mutations towards a prefusion form of gB in the absence of inhibitor and detergents if needed.

The coordinates and structural factors for the model of the prefusion gB associated with the present Example are provided in **Table 1B** described in WO2021/260510 which is hereby incorporated by reference in its entirety.

5 **Table 6. Exemplary cysteine pair mutations for disulfide bond stabilization.**

Row	Mutations
1	D217C, Y589C
2	I356C, A 500C
3	S367C, A500C
4	S367C, A503C
5	M371C, A505C
6	M371C, W506C
7	T374C, A503C
8	Y160C, Y708C
9	L162C, M716C
10	N524C, M684C
11	G99C, A267C
12	T100C, R258C
13	T221C, E657C
14	S223C, T659C
15	F541C, E681C
16	L603C, Y667C
17	N605C, K670C
18	R607C, N688C
19	E609C, K691C

Table 7. Exemplary charge mutations for stabilization.

Row	Mutations
1	E686L
2	E686I
3	D679A
4	R354F
5	R573F
6	D101L, K260L

Table 8. Construct mutations. The constructs were made from the designs for purpose of testing the presence of prefusion gB in the purified recombinant protein preparation in different conditions.

Construct	Mutation 1	Mutation 2	Mutation 3	Mutation 4
pSB1688	N524C	M684C		
pSB2455	M371C	W506C		
pSB2456	F541C	E681C		
pSB2457	D217C	Y589C	M371C	W506C
pSB2459	D217C	Y589C	N524C	M684C
pSB1666	D217C	Y589C		

5

Table 9. Exemplary Soluble, detergent-free gB ectodomain proteins

Code	gB*	Mutations	C-Term Fusion Sequence
gB707	gB (1-707)	D217C, Y589C	
gB2264	gB (1-707)	D217C, Y589C, D703C, P704C	
gB2265	gB (1-707)	D217C, Y589C, Y696C, V697C	
gB2266	gB (1-702)	D217C, Y589C	RIKQIEDKIEEILSKQYHIENEIARIKKLIG (SEQ ID NO: 255)
gB2267	gB (1-702)	D217C, Y589C	KIEEILSKQYHIENEIARIKKLIG (SEQ ID NO: 256)
gB2268	gB (1-702)	D217C, Y589C	KIYHIENEIARIKKLIG (SEQ ID NO: 257)
gB2269	gB (1-703)	D217C, Y589C	LSAIGGYIPEAPRDGQAYVRKDGWVLLS TFL (SEQ ID NO: 258)
gB2555	gB (1-702)	D217C, M371C, W506C, Y589C	KIEEILSKQYHIENEIARIKKLIG (SEQ ID NO: 259)
gB2556	gB (1-702)	D217C, N524C, Y589C, M684C	KIEEILSKQYHIENEIARIKKLIG (SEQ ID NO: 259)
gB2557	gB (1-707)	D217C, N524C, Y589C, M684C, D703C, P704C	

* Mutations, including Y1H (155-157)→GHR, W240→A and C246→S, had been incorporated in gB to decrease aggregation and increase protein secretion.

10 EXAMPLE 7: Determining Druggable Regions of HCMV gB polypeptide

The prefusion and postfusion structures of the HCMV gB polypeptide are described in the Examples set forth hereinabove. When HCMV gB polypeptide rearranges to the postfusion structure, long beta-sheet rich globular domains are formed tipped by fusion loops that come together in a trimer, leaving grooves between them. A long, extended coil/alpha-helical structure from closer to the membrane anchor of the protein tracks up those grooves to bring

the membrane anchor and fusion loops together, completing the fusion reaction that effects cell entry of the virus.

FIG. 20A and 20B depict the space-filled model of HCMV (Towne strain) domain V (residues I642-V697 (SEQ ID NO: 281)) in prefusion (**FIG. 20A**) and postfusion (**FIG. 20B**) structures. Hydrophobic residues are labeled. In prefusion conformation, these hydrophobic residues are disordered, but they rearrange to an amphipathic extended conformation to complement with the hydrophobic surface of other parts of the postfusion structure. For example, as shown in **FIG. 20B**, the region between I642—F661 (SEQ ID NO: 282) interacts with the surface formed by Q485—L520 (SEQ ID NO: 283) from the other two protomers; the region between L672—V697 (SEQ ID NO: 284) interacts with the same region from the other two protomers. This structure arrangement is necessary to drive the fusion between the virus and host cell membranes.

This conformational change creates a druggable region in HCMV gB polypeptide domain V up the groove between domains I and II in the post-fusion form.

FIG. 21A – 21C show (**FIG. 21A**) the HCMV gB (Towne Strain) domain V (light grey) and its binding pocket (dark grey) in postfusion conformation and (**FIG. 21B**) the the space-filled model of the pocket without domain V is shown. **FIG. 21C** shows a ribbon representation of the same structure as in **FIG. 21A** with certain residues labeled.

In one embodiment, peptides corresponding to parts of the long extended coil/a-helix that tracks up the grooves are added to the virus to compete with their authentic counterparts for occupancy of the grooves, preventing completion of the rearrangement, thereby blocking membrane fusion, virus entry, and virus infectivity. In another embodiment, blockade of binding of the peptides into the groove may be used to screen a library of small molecules, leading to the discovery of small molecule CMV fusion inhibitors.

This invention provides peptides comprising residues (Towne strain) of SEQ ID NO:1 selected from the group consisting of M648-K700 (SEQ ID NO: 260), M648-V697 (SEQ ID NO: 261), S647-V697 (SEQ ID NO: 262), S647-V663 (SEQ ID NO: 263), I653-V697 (SEQ ID NO: 264), I653-Q692 (SEQ ID NO: 265), I653-L680 (SEQ ID NO: 266), I653-S675 (SEQ ID NO: 267), I653-Y667 (SEQ ID NO: 268), R662-V697 (SEQ ID NO: 269), R662-Q692 (SEQ ID NO: 270), R662-L680 (SEQ ID NO: 271), R662-S675 (SEQ ID NO: 272), L664-F678 (SEQ ID NO: 273), S668-V697 (SEQ ID NO: 274), S668-Q692 (SEQ ID NO: 275), S668-V677 (SEQ ID NO: 276), D679-V697 (SEQ ID NO: 277), L680-V697 (SEQ ID NO: 278), L680-Q692 (SEQ ID NO: 279), and R693-K700 (SEQ ID NO: 280).

Verification Experiments for Peptide Activity

The above-described peptides (see also **Table 10**) are synthesized and tested in a virus cell infection inhibitory activity. The peptides are expected to have a potency to reduce 50% infection (IC50) at a concentration around 10uM or lower, comparable to peptides with similar

functions in dengue virus (A. G. Schmidt, P. L. Yang, S. C. Harrison, Peptide inhibitors of dengue-virus entry target a late-stage fusion intermediate. *PLoS Pathog* 6, e1000851 (2010)). The engagement of selected peptides in the druggable region of gB may be verified with binding assays. These include a measurement of the binding affinity by SPR to show the dissociation constant (K_d) at around 10uM or lower and cryoEM structure to show the density for the peptide in the druggable region of gB. A druggable region of HCMV gB comprises residues K130-A135, D216-W233, R258-K260, A267-V273, R327-D329, W349-E350, V480-K518 and N676-Y690 of SEQ ID NO: 1. The residues of the druggable region are in a postfusion conformation.

These methods are well known to those of skill in the art and described in more detail below:

1. Virus infection inhibition assay

HCMV infection is tested in MRC-5 fibroblast cells and ARPE-19 epithelial cells by a microneutralization assay. Briefly, cells are seeded in 96 well plates overnight at 37°C, 5% CO₂. Serially diluted peptides are combined with HCMV Towne strain or VR1814 at approximately 500 pfu/well and allowed to incubate for one hour at 37°C. The mixture is then transferred to MRC-5 or ARPE-19 cells and allowed to infect for 24 hours at 37°C, 5% CO₂. After fixing with methanol, HCMV-infected cells are detected by immunostaining using a mouse monoclonal anti-HCMV immediate early protein IE1 antibody (developed in-house). A secondary antibody, Alexa Flour 488 labeled Goat anti-mouse IgG (H+L) (Molecular Probes) antibody is then used for counting infected cells using a CTL-ImmunoSpot Analyzer. Titers that inhibit 50% of the infectivity of the samples (IC₅₀) are determined by interpolation using a 4-parameter nonlinear regression.

2. Binding affinity by SPR

For measuring the binding of HCMV gB protein and peptides, the SPR CM5 sensor chip surface is prepared with 1xHBSP+ (0.01 M Hepes pH 7.4, 0.15 M NaCl, 0.05% v/v Surfactant P20) running buffer. All flow cells are activated with 0.4 M EDC + 0.1 M NHS, and amine-coupled with Anti-Avi pAb in 10 mM Acetate pH 4.5. The surface will then be blocked with 0.1 M EDA in 0.2 M borate buffer pH 8.5, and then regenerated three times with 75 mM phosphoric acid.

Multicycle kinetic assays are performed at room temperature with 1xHBSP+ supplemented with 1 g/L BSA as running buffer. For capture, purified HCMV gB proteins are diluted to 0.2 ug/mL and loaded onto flow-cell 2 of each channel. The designed gB peptides will then be injected at a concentration of 100 uM for the initial binding screen. If positive binding peptides are identified, then the positive peptides are injected at concentrations of 0, 5, 10, 20,

50, and 100 μ M for measuring the binding kinetics. At the end of every cycle, the surface is regenerated three times with 75 mM phosphate.

3. Binding to the druggable site by cryoEM

5 A purified gB protein that lack domain V, MPR, TM and cytoplasmic domains (roughly ends at residue 642 of SEQ ID NO: 1) and in its postfusion form is used in this experiment. The protein is incubated with the designed peptide at room temperature for 4 hours with about two-fold molar excess of the peptide. The sample is then subject to cryo grids preparation and imaging in cryo conditions. The reconstructed structure is obtained by popular image
10 processing software such as CryoSparc (A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296 (2017)) or Relion (S. H. Scheres, RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519-530 (2012)). The specific druggable site binding is verified by the presence of the peptide density in the corresponding
15 location as in the postfusion structure.

Unlike other viruses, CMV causes lifelong infection. Accordingly, these CMV modulators or inhibitors may be used to treat subjects suffering from CMV infection or to prevent infection in transplant patients or patients with immunodeficiency. In another embodiment, the modulators or inhibitors may be used to treat newborns with CMV infection to prevent or reduce the
20 manifestations of congenital CMV. Treatment of prevention in newborns may also comprise administering to pregnant women.

Table 10

SEQ ID NO:	Description	Sequence
260	M648-K700 of SEQ ID NO:1	MIALDIDPLENTDFRVLELYSQKELRSSNVFDLEEIMREFNSYKQRVKYVEDK
261	M648-V697 of SEQ ID NO:1	MIALDIDPLENTDFRVLELYSQKELRSSNVFDLEEIMREFNSYKQRVKYV
262	S647-V697 of SEQ ID NO:1	SMIALDIDPLENTDFRVLELYSQKELRSSNVFDLEEIMREFNSYKQRVKYV
263	S647-V663 of SEQ ID NO:1	SMIALDIDPLENTDFRV
264	I653-V697 of SEQ ID NO:1	IDPLENTDFRVLELYSQKELRSSNVFDLEEIMREFNSYKQRVKYV
265	I653-Q692 of SEQ ID NO:1	IDPLENTDFRVLELYSQKELRSSNVFDLEEIMREFNSYKQ
266	I653-L680 of SEQ ID NO:1	IDPLENTDFRVLELYSQKELRSSNVFDL
267	I653-S675 of SEQ ID NO:1	IDPLENTDFRVLELYSQKELRSS
268	I653-Y667 of SEQ ID NO:1	IDPLENTDFRVLELY

269	R662-V697 of SEQ ID NO:1	RVLELYSQKELRSSNVFDLEEIMREFNSYKQRVKYV
270	R662-Q692 of SEQ ID NO:1	RVLELYSQKELRSSNVFDLEEIMREFNSYKQ
271	R662-L680 of SEQ ID NO:1	RVLELYSQKELRSSNVFDL
272	R662-S675 of SEQ ID NO:1	RVLELYSQKELRSS
273	L664-F678 of SEQ ID NO:1	LELYSQKELRSSNVF
274	S668-V697 of SEQ ID NO:1	SQKELRSSNVFDLEEIMREFNSYKQRVKYV
275	S668-Q692 of SEQ ID NO:1	SQKELRSSNVFDLEEIMREFNSYKQ
276	S668-V677 of SEQ ID NO:1	SQKELRSSNV
277	D679-V697 of SEQ ID NO:1	DLEEIMREFNSYKQRVKYV
278	L680-V697 of SEQ ID NO:1	LEEIMREFNSYKQRVKYV
279	L680-Q692 of SEQ ID NO:1	LEEIMREFNSYKQ
280	R693-K700 of SEQ ID NO:1	RVKYVEDK
281	I642-V697 of SEQ ID NO:1	ISTVDSMIALDIDPLENTDFRVLELYSQKELRSSNVFDLEEIM REFNSYKQRVKYV
282	I642-F661 of SEQ ID NO:1	ISTVDSMIALDIDPLENTDF
283	Q485-L520 of SEQ ID NO:1	QFTYDTLRGYINRALAQIAEAWCVDQRRRTLEVFKEL
284	L672-V697 of SEQ ID NO:1	LRSSNVFDLEEIMREFNSYKQRVKYV
285	K130-A135 of SEQ ID NO:1	KRNIVA
286	D216-W233 of SEQ ID NO:1	DDYSNTHSTRYVTVKDQW
	R258-K260 of SEQ ID NO:1	RSK
287	A267-V273 of SEQ ID NO:1	ATSTGDV
	R327-D329 of SEQ ID NO:1	RAD
	W349-E350 of SEQ ID NO:1	WE
288	V480-K518 of SEQ ID NO:1	VYAQLQFTYDTLRGYINRALAQIAEAWCVDQRRRTLEVFK
289	N676-Y690 of SEQ ID NO:1	NVFDLEEIMREFNSY

Embodiments of the present invention are set out in the following numbered clauses:

- 5 C1. A method for identifying a candidate therapeutic for treating or preventing a disease caused by a human cytomegalovirus (HCMV) infection comprising contacting a HCMV having a glycoprotein B (gB) polypeptide which comprises a druggable region with a compound, wherein binding of said compound indicates a candidate therapeutic.
- C2. A method for identifying a candidate therapeutic for treating or preventing a disease caused by HCMV infection, comprising contacting a HCMV gB polypeptide comprising a druggable region with a compound, wherein the modulation of the function or activity of said gB polypeptide indicates a candidate therapeutic.
- 10 C3. The method of clause C2, wherein said modulation of the activity of said gB polypeptide involves precluding the binding of domain V to the domain V binding groove in the postfusion conformation.
- C4. A method for identifying a candidate therapeutic for treating or preventing a disease caused by infection with HCMV having a gB polypeptide, comprising contacting the gB polypeptide comprising a druggable region with a compound, wherein the inhibition of fusion of said virus indicates a candidate therapeutic.
- 15 C5. A method for identifying a candidate therapeutic for a disease caused by infection with HCMV having a gB polypeptide, comprising contacting the gB polypeptide comprising a druggable region with a compound, wherein the inhibition of viral infectivity of said virus indicates a candidate therapeutic.
- 20 C6. A method for identifying a candidate therapeutic for a disease caused by infection with HCMV having a gB polypeptide, comprising contacting the gB polypeptide comprising a druggable region with a compound, wherein the reduction of at least one symptom of said disease in a subject indicates a candidate therapeutic.
- 25 C7. The method of any one of clauses C1-C6, wherein said compound is selected from the following classes of compounds: proteins, peptides, polypeptides, peptidomimetics, antibodies, nucleic acids, and small molecules.
- C8. The method of clause C7, wherein binding is determined using an *in vitro* assay.
- C9. The method of clause C7, wherein binding is determined using an *in vivo* assay.
- 30 C10. The method of clause C7, wherein the druggable region comprises at least one residue from residues K130-A135, D216-W233, R258-K260, A267-V273, R327-D329, W349-E350, V480-K518 or N676-Y690 of SEQ ID NO: 1.
- C11. The method of clause C10, wherein the residue of the druggable region is in a postfusion conformation.
- 35 C12. The method of any one of clauses C1-C11, wherein the compound is a peptide comprising residues (Towne strain) of SEQ ID NO:1 selected from the group consisting of M648-K700 (SEQ ID NO: 260), M648-V697 (SEQ ID NO: 261), S647-V697 (SEQ ID NO: 262), S647-V663 (SEQ ID NO: 263), I653-V697 (SEQ ID NO: 264), I653-Q692 (SEQ ID

- NO: 265), I653-L680 (SEQ ID NO: 266), I653-S675 (SEQ ID NO: 267), I653-Y667 (SEQ ID NO: 268), R662-V697 (SEQ ID NO: 269), R662-Q692 (SEQ ID NO: 270), R662-L680 (SEQ ID NO: 271), R662-S675 (SEQ ID NO: 272), L664-F678 (SEQ ID NO: 273), S668-V697 (SEQ ID NO: 274), S668-Q692 (SEQ ID NO: 275), S668-V677 (SEQ ID NO: 276),
 5 D679-V697 (SEQ ID NO: 277), L680-V697 (SEQ ID NO: 278), L680-Q692 (SEQ ID NO: 279), and R693-K700 (SEQ ID NO: 280).
- C13. The method of any one of clauses C1-C11, wherein said compound is in a library of compounds.
- C14. The method of clause C13, wherein said library is generated using combinatorial
 10 synthetic methods.
- C15. A candidate therapeutic, wherein said candidate therapeutic is a modulator of HCMV activity which interacts with domain V region of glycoprotein B (gB) of HCMV.
- C16. A candidate therapeutic, wherein said candidate therapeutic is a modulator of HCMV activity which precludes the movement of domain V of glycoprotein B (gB) of HCMV.
- 15 C17. A candidate therapeutic, wherein said candidate therapeutic is a modulator of HCMV activity which precludes completion of the conformational change by interacting with at least one residue from the domain V residues at the trimer interface formed by any subunit in the postfusion trimer.
- C18. A candidate therapeutic, wherein the candidate therapeutic is an inhibitor of HCMV
 20 activity comprising a polypeptide sequence with at least 80% homology to SEQ ID NO: 1.
- C19. The candidate therapeutic of clauses C15-C18, wherein the candidate inhibitor is a peptide comprising residues of SEQ ID NO:1 (Towne strain) selected from the group consisting of M648-K700 (SEQ ID NO: 260), M648-V697 (SEQ ID NO: 261), S647-V697 (SEQ ID NO: 262), S647-V663 (SEQ ID NO: 263), I653-V697 (SEQ ID NO: 264), I653-Q692 (SEQ ID NO: 265), I653-L680 (SEQ ID NO: 266), I653-S675 (SEQ ID NO: 267),
 25 I653-Y667 (SEQ ID NO: 268), R662-V697 (SEQ ID NO: 269), R662-Q692 (SEQ ID NO: 270), R662-L680 (SEQ ID NO: 271), R662-S675 (SEQ ID NO: 272), L664-F678 (SEQ ID NO: 273), S668-V697 (SEQ ID NO: 274), S668-Q692 (SEQ ID NO: 275), S668-V677 (SEQ ID NO: 276), D679-V697 (SEQ ID NO: 277), L680-V697 (SEQ ID NO: 278), L680-Q692 (SEQ ID NO: 279), and R693-K700 (SEQ ID NO: 280).
 30
- C20. The candidate therapeutic of clauses C15-C19, wherein the candidate therapeutic is a nucleic acid.
- C21. A pharmaceutical composition comprising a candidate therapeutic of any of clauses C15-C20.
- 35 C22. A method of treating a subject having a disease or disorder associated with HCMV infection comprising administering to said subject a pharmaceutical composition of clause C21.

- C23. A method of preventing a disease or disorder associated with HCMV infection in a subject comprising administering to said subject a pharmaceutical composition of clause C21.
- 5 C24. A kit for treating or preventing a disease or disorder associated with HCMV infection, comprising a pharmaceutical composition of clause C21 and optionally instructions for use.
- C25. A druggable region of HCMV gB comprising residues K130-A135, D216-W233, R258-K260, A267-V273, R327-D329, W349-E350, V480-K518 and N676-Y690 of SEQ ID NO: 1.
- 10 C26. The druggable region of of HCMV gB of clause C25, wherein the residues of the druggable region are in a postfusion conformation.

CLAIMS

1. A method for identifying a candidate therapeutic for treating or preventing a disease caused by a human cytomegalovirus (HCMV) infection comprising contacting a HCMV having a glycoprotein B (gB) polypeptide which comprises a druggable region with a compound, wherein binding of said compound indicates a candidate therapeutic.
2. A method for identifying a candidate therapeutic for treating or preventing a disease caused by HCMV infection, comprising contacting a HCMV gB polypeptide comprising a druggable region with a compound, wherein the modulation of the function or activity of said gB polypeptide indicates a candidate therapeutic.
3. The method of claim 2, wherein said modulation of the activity of said gB polypeptide involves precluding the binding of domain V to the domain V binding groove in the postfusion conformation.
4. A method for identifying a candidate therapeutic for treating or preventing a disease caused by infection with HCMV having a gB polypeptide, comprising contacting the gB polypeptide comprising a druggable region with a compound, wherein the inhibition of fusion of said virus indicates a candidate therapeutic.
5. A method for identifying a candidate therapeutic for a disease caused by infection with HCMV having a gB polypeptide, comprising contacting the gB polypeptide comprising a druggable region with a compound, wherein the inhibition of viral infectivity of said virus indicates a candidate therapeutic.
6. A method for identifying a candidate therapeutic for a disease caused by infection with HCMV having a gB polypeptide, comprising contacting the gB polypeptide comprising a druggable region with a compound, wherein the reduction of at least one symptom of said disease in a subject indicates a candidate therapeutic.
7. The method of any one of claims 1-6, wherein said compound is selected from the following classes of compounds: proteins, peptides, polypeptides, peptidomimetics, antibodies, nucleic acids, and small molecules.
8. The method of claim 7, wherein binding is determined using an *in vitro* assay.
9. The method of claim 7, wherein binding is determined using an *in vivo* assay.
10. The method of claim 7, wherein the druggable region comprises at least one residue from residues K130-A135, D216-W233, R258-K260, A267-V273, R327-D329, W349-E350, V480-K518 or N676-Y690 of SEQ ID NO: 1.
11. The method of claim 10, wherein the residue of the druggable region is in a postfusion conformation.
12. The method of any one of claims 1-11, wherein the compound is a peptide comprising residues (Towne strain) of SEQ ID NO:1 selected from the group consisting of M648-K700 (SEQ ID NO: 260), M648-V697 (SEQ ID NO: 261), S647-V697 (SEQ ID NO: 262), S647-V663 (SEQ ID NO: 263), I653-V697 (SEQ ID NO: 264), I653-Q692 (SEQ ID NO: 265),

- I653-L680 (SEQ ID NO: 266), I653-S675 (SEQ ID NO: 267), I653-Y667 (SEQ ID NO: 268), R662-V697 (SEQ ID NO: 269), R662-Q692 (SEQ ID NO: 270), R662-L680 (SEQ ID NO: 271), R662-S675 (SEQ ID NO: 272), L664-F678 (SEQ ID NO: 273), S668-V697 (SEQ ID NO: 274), S668-Q692 (SEQ ID NO: 275), S668-V677 (SEQ ID NO: 276), D679-V697 (SEQ ID NO: 277), L680-V697 (SEQ ID NO: 278), L680-Q692 (SEQ ID NO: 279), and R693-K700 (SEQ ID NO: 280).
13. The method of any one of claims 1-11, wherein said compound is in a library of compounds.
 14. The method of claim 13, wherein said library is generated using combinatorial synthetic methods.
 15. A candidate therapeutic, wherein said candidate therapeutic is a modulator of HCMV activity which interacts with domain V region of glycoprotein B (gB) of HCMV.
 16. A candidate therapeutic, wherein said candidate therapeutic is a modulator of HCMV activity which precludes the movement of domain V of glycoprotein B (gB) of HCMV.
 17. A candidate therapeutic, wherein said candidate therapeutic is a modulator of HCMV activity which precludes completion of the conformational change by interacting with at least one residue from the domain V residues at the trimer interface formed by any subunit in the postfusion trimer.
 18. A candidate therapeutic, wherein the candidate therapeutic is an inhibitor of HCMV activity comprising a polypeptide sequence with at least 80% homology to SEQ ID NO: 1.
 19. The candidate therapeutic of claims 15-18, wherein the candidate inhibitor is a peptide comprising residues of SEQ ID NO:1 (Towne strain) selected from the group consisting of M648-K700 (SEQ ID NO: 260), M648-V697 (SEQ ID NO: 261), S647-V697 (SEQ ID NO: 262), S647-V663 (SEQ ID NO: 263), I653-V697 (SEQ ID NO: 264), I653-Q692 (SEQ ID NO: 265), I653-L680 (SEQ ID NO: 266), I653-S675 (SEQ ID NO: 267), I653-Y667 (SEQ ID NO: 268), R662-V697 (SEQ ID NO: 269), R662-Q692 (SEQ ID NO: 270), R662-L680 (SEQ ID NO: 271), R662-S675 (SEQ ID NO: 272), L664-F678 (SEQ ID NO: 273), S668-V697 (SEQ ID NO: 274), S668-Q692 (SEQ ID NO: 275), S668-V677 (SEQ ID NO: 276), D679-V697 (SEQ ID NO: 277), L680-V697 (SEQ ID NO: 278), L680-Q692 (SEQ ID NO: 279), and R693-K700 (SEQ ID NO: 280).
 20. The candidate therapeutic of claims 15-19, wherein the candidate therapeutic is a nucleic acid.
 21. A pharmaceutical composition comprising a candidate therapeutic of any of claims 15-20.
 22. A method of treating a subject having a disease or disorder associated with HCMV infection comprising administering to said subject a pharmaceutical composition of claim 21.

23. A method of preventing a disease or disorder associated with HCMV infection in a subject comprising administering to said subject a pharmaceutical composition of claim 21.
24. A kit for treating or preventing a disease or disorder associated with HCMV infection, comprising a pharmaceutical composition of claim 21 and optionally instructions for use.
25. A druggable region of HCMV gB comprising residues K130-A135, D216-W233, R258-K260, A267-V273, R327-D329, W349-E350, V480-K518 and N676-Y690 of SEQ ID NO: 1.
26. The druggable region of of HCMV gB of claim 25, wherein the residues of the druggable region are in a postfusion conformation.