

COMPLEXES OF CYTOMEGALOVIRUS PROTEINS

This application claims the benefit of U.S. Provisional Application No. 61/668,975, filed July 6, 2012 and U.S. Provisional Application No. 61/770,257 filed February 27, 2013, the complete contents of each of the foregoing applications are hereby incorporated herein by reference for all purposes.

TECHNICAL FIELD

This invention is in the field of vaccination against human cytomegalovirus (HCMV), and in particular the isolation of purified complexes comprising gH, gL and at least one more HCMV glycoprotein, preferably the trimeric gH/gL/gO complex or the pentameric gH/gL/pUL128/pUL130/pUL131A complex, and their subsequent use in vaccines.

BACKGROUND ART

Cytomegalovirus (CMV) is a genus of virus that belongs to the viral family known as *Herpesviridae* or herpesviruses. The species that infects humans is commonly known as HCMV or human herpesvirus-5 (HHV-5). Within *Herpesviridae*, HCMV belongs to the *Betaherpesvirinae* subfamily, which also includes cytomegaloviruses from other mammals.

Although they may be found throughout the body, HCMV infections are frequently associated with the salivary glands. HCMV infects between 50% and 80% of adults in the United States (40% worldwide), as indicated by the presence of antibodies in much of the general population. HCMV infection is typically unnoticed in healthy people, but can be life-threatening for the immunocompromised, such as HIV-infected persons, organ transplant recipients, or new born infants (Mocarski, Shenk and Pass 2006). HCMV is the virus most frequently transmitted to a developing fetus. After infection, HCMV has an ability to remain latent within the body for the lifetime of the host, with occasional reactivations from latency.

HCMV seems to have a large impact on immune parameters in later life and may contribute to increased morbidity and eventual mortality (Simanek, et al. (2011)).

To date, the genomes of over 20 different HCMV strains have been sequenced, including those of both laboratory strains and clinical isolates. For example, the following strains of HCMV have been sequenced: Towne (GI:239909366), AD169 (GI:219879600), Toledo (GI:290564358) and Merlin (GI:155573956). HCMV strains AD169, Towne and Merlin can be obtained from the American Type Culture Collection (ATCC VR538, ATCC VR977 and ATCC VR1590, respectively).

HCMV contains an unknown number of membrane protein complexes. Of the approximately 30 known glycoproteins in the viral envelope, gH and gL have emerged as particularly interesting due to their presence in several different complexes: dimeric gH/gL, trimeric gH/gL/gO (also known as the gCIII complex) and the pentameric gH/gL/pUL128/pUL130/pUL131A (the latter protein is

also referred to as pUL131). HCMV is thought to use the pentameric complexes to enter epithelial and endothelial cells by endocytosis and low-pH-dependent fusion but it is thought to enter fibroblasts by direct fusion at the plasma membrane in a process involving gH/gL or possibly gH/gL/gO. The gH/gL and/or gH/gL/gO complex(es) is/are sufficient for fibroblast infection, whereas the pentameric complex is required to infect endothelial and epithelial cells (Ryckman, Rainish, et al. 2008).

Genini *et al.* (2011) discloses a serum antibody response to the pentameric complex of HCMV in primary and reactivated HCMV infections. The response was determined by both indirect immunofluorescence (IFA) and ELISA, using fixed or lysed epithelial (ARPE-19) cells infected with one or more adenoviral vectors, each carrying one HCMV gene and, in parallel, with a control adenovirus vector. The specificity of results was determined by the reactivity of human neutralizing monoclonal antibodies recognizing two, three, or four proteins of the complex. In 14 cases of primary infection, an IgG antibody seroconversion to the UL128–131 gene products was consistently detected within 2–4 weeks after onset of infection, while antibodies persisted for at least 12 months. The IgG antibody response to UL128–131 gene products was generally superior to the response to gH and appeared to follow the neutralizing antibody response (as determined in epithelial cells). In reactivated infections, the antibody response showed a trend reminiscent of a booster response. IgG antibodies were detected in HCMV-seropositive healthy adult controls, but not in HCMV-seronegative individuals.

Kinzler *et al.* (2002) co-expressed gH, gL, and gO in insect cells using a recombinant baculovirus, but were unable to produce the gH/gL/gO tripartite complex. Instead, only gH/gL heterodimers, gH/gL heteromultimers, and gO homomultimers were detected. In contrast, co-expression of gH, gL, and gO in mammalian cells produced high molecular weight complexes that closely resemble gH/gL/gO complexes formed in HCMV infected cells. Cell surface immunofluorescence showed that these complexes are expressed and displayed on the surface of transfected cells.

US7704510 discloses that pUL131A is required for epithelial cell tropism. US7704510 also discloses that pUL128 and pUL130 form a complex with gH/gL, which is incorporated into virions. This complex is required to infect endothelial and epithelial cells but not fibroblasts. Anti-CD46 antibodies were found to inhibit HCMV infection of epithelial cells. However, US7704510 does not disclose isolation of HCMV complexes.

To date, researchers have been unable to purify complexes comprising HCMV gH, gL and at least one more HCMV glycoprotein, such as the trimeric gH/gL/gO complex or the pentameric gH/gL/pUL128/pUL130/pUL131A complex. Such purified complexes would be useful as antigens for diagnostic application and as immunogens for vaccines against HCMV.

DISCLOSURE OF THE INVENTION

The invention is based on the recombinant expression and purification of HCMV membrane protein complexes, wherein said complexes comprise gH, gL and at least one more HCMV glycoprotein.

5 The invention provides a process for producing an HCMV membrane protein complex comprising gH, gL and at least one more HCMV glycoprotein, wherein said process comprises the recombinant co-expression of an HCMV gH protein, an HCMV gL protein, and at least one more HCMV glycoprotein, under conditions in which said gH, gL and at least one more HCMV glycoprotein assemble to form a protein complex. This process may optionally involve isolation of
10 the protein complex, so that it can be prepared in purified form. In some embodiments, the process does not involve co-expression of any non-envelope HCMV proteins, such as the tegument or capsid proteins.

The invention also provides a protein complex produced by this process. For instance, the complex can comprise (i) gH, gL and gO or (ii) gH, gL, pUL128, pUL130 and pUL131A.

15 In some embodiments, the complexes of the invention can be produced at high yields. For example, in processes involving growing cells of the invention in growth medium, the membrane protein complex of the invention may accumulate to a level of more than 0.4 mg per litre of growth medium (*e.g.* 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1.0, 1.2, 1.4, 1.6, 1.8, 2, 2.5, 3, 3.5, 4, 4.5 or 5 mg per litre of growth medium or more).

20 The invention also provides an isolated protein complex comprising an HCMV gH, an HCMV gL and at least one more HCMV glycoprotein. One example of a protein complex that comprises gH and gL is the trimeric complex consisting of gH, gL and gO. Another example of a protein complex that comprises gH and gL is the pentameric complex consisting of gH, gL, pUL128, pUL130 and pUL131A.

25 The invention also provides a composition comprising a protein complex of the invention. In some embodiments, the composition does not contain polyacrylamide. In some embodiments, the composition is a liquid *e.g.* an aqueous liquid, not a gel. In some embodiments, the protein complex is not immobilised within the composition. For example, said HCMV membrane protein complex may not be present in a gel, or on a film, membrane, paper or slide.

30 The invention also provides a composition comprising a protein complex of the invention, wherein said composition does not contain any non-envelope HCMV proteins, such as the HCMV tegument or HCMV capsid proteins.

35 The invention also provides a modified HCMV gH polypeptide, wherein said polypeptide lacks a transmembrane (TM) domain. The absence of a TM domain means that this modified polypeptide cannot reside within a lipid bilayer. In some embodiments, the gH polypeptide lacks the full-length natural TM domain; in other embodiments, it can retain a portion of the natural TM domain, but not enough to let the protein reside in a lipid bilayer. Thus the polypeptide can contain

up to 10 amino acids (*e.g.* 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 amino acids) of the natural gH TM domain. In addition to lacking some or all of the TM domain, the polypeptide may also lack the natural C-terminal domain of HCMV gH or may lack a portion of the C-terminal domain. The invention also provides nucleic acid molecules encoding said modified gH polypeptide, and processes for producing said modified gH polypeptide *e.g.* by recombinant expression, or by chemical synthesis (at least in part).

The invention provides a method of transfection involving the introduction of one or more nucleic acid molecules which encode the protein components of said HCMV membrane protein complex into cells. Said one or more nucleic acid molecules may be stably introduced into the cells (*e.g.* by chromosomal integration) or may be transiently introduced into the cells. The invention also provides a cell which is derived from said transfection. Preferably, a cell which is derived from said transfection is a cell in which said one or more nucleic acid molecules have been stably introduced. The invention also provides progeny of said cell, a clone of said cell or a cell which has been passaged from said cell. These cells can be cultured to express complexes of the invention, which may then be purified.

The invention also provides a cell that produces an HCMV membrane protein complex, wherein the cell does not (i) contain an HCMV genome, and/or (ii) produce HCMV virions, and/or (iii) express any non-envelope HCMV proteins. Ideally the cell lacks one of (i), (ii) or (iii); preferably, it lacks two; more preferably, it lacks all three of (i), (ii) and (iii).

The invention provides antibodies which recognise an isolated HCMV membrane protein complex of the invention, but do not bind to any of isolated gH, gL, gO, pUL128, pUL130 or pUL131A polypeptides and/or do not bind to isolated gH-gL heterodimers. The antibodies of the invention may have been raised using an isolated HCMV membrane protein complex of the invention as an antigen. Preferably, the antibodies of the invention are neutralizing antibodies. The antibodies of the invention may have been identified using *in vitro* selection methods, such as phage display using diverse antibody libraries. As described below, antibodies of the invention may be human or humanised antibodies and/or they may be monoclonal or polyclonal antibodies.

The invention also provides a method for raising antibodies using an isolated HCMV membrane protein complex of the invention. Alternatively, isolated HCMV membrane protein complex of the invention may be used to identify antibodies using *in vitro* selection methods, such as phage display using diverse antibody libraries.

The antibodies of the invention may be used in a diagnostic assay and may be labelled directly or indirectly. In some embodiments, the antibodies of the invention may be used in therapy, for example in the treatment of HCMV infection.

Proteins of the invention

HCMV glycoprotein H (gH), which is encoded by the UL75 gene, is a virion glycoprotein that is essential for infectivity and which is conserved among members of the alpha-, beta- and

gammaherpesviruses. It forms a stable complex with gL, and the formation of this complex facilitates the cell surface expression of gH. Based on the crystal structures of HSV-2 and EBV gH/gL complexes, the gL subunit and N-terminal residues of gH form a globular domain at one end of the structure (the 'head'), which is implicated in interactions with gB and activation of membrane fusion. The C-terminal domain of gH, proximal to the viral membrane (the 'tail'), is also implicated in membrane fusion. gH displays determinants that are recognized by the host factor TLR2, and it directly interacts with a heterodimer formed between the host factors TLR2 and TLR1. TLR2 mediates NF- κ B activation and inflammatory cytokine responses from cells (Boehme, Guerrero and Compton 2006).

The gH from HCMV strain Merlin has been reported (GI:52139248, SEQ ID NO: 1) to consist of 742 amino acids. The gH from HCMV strain Towne (GI:138314, SEQ ID NO: 2) also consists of 742 amino acids, and has been reported to have six N-glycosylation sites (at residues 55, 62, 67, 192, 641 and 700), and consist of a hydrophobic signal sequence at its N-terminus (amino acid residues 1-23), an ectodomain (residues 24-717) that projects out of the cell into the extracellular space, a hydrophobic TM domain (residues 718-736) and a C-terminal cytoplasmic domain (residues 737-742). SEQ ID NO: 2 shares 99% and 96% amino acid similarity to SEQ ID NO: 1 and the gH from HCMV strain AD169 (GI:138313, SEQ ID NO: 3), respectively.

Typically, the N-terminal signal sequence of gH proteins is cleaved by a host cell signal peptidase to produce mature gH proteins. The gH proteins in HCMV membrane complexes of the invention may lack an N-terminal signal sequences. Preferably, mature forms of gH (as found in isolated HCMV membrane complexes of the invention) lack the N-terminal signal sequence, the TM domain and the C-terminal domain.

Expression of the full-length UL75 gene sequence hinders purification of soluble complexes comprising gH. Rather, complexes comprising gH can be purified at high yield and purity by omitting at least a portion of the TM domain of gH. For example, constructs encoding just the N-terminal signal sequence and the ectodomain of gH (but not the TM domain) can be used to express a form of gH which is easily purified. Said constructs may encode the majority (*e.g.* 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) of the ectodomain of gH, but none or only a small portion of the TM domain. gH proteins of the invention may include the whole of the gH ectodomain or a truncated form of the gH ectodomain. Said truncated forms of the ectodomain may lack between 1 and 20 amino acids (*e.g.* 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 amino acid residues) at their N-termini and/or C-termini relative to a full-length HCMV gH protein. An example of a gH protein of the invention is SEQ ID NO: 4, which consists of amino acid residues 1-715 of SEQ ID NO: 1. An example of a preferred gH protein of the invention is SEQ ID NO: 29, which lacks the N-terminal signal sequence, TM domain and C-terminal domain of gH and consists of amino acid residues 24-715 of SEQ ID NO: 1.

gH proteins of the invention may contain additional amino acid residues, such as N-terminal or C-terminal extensions. Such extensions may include one or more tags, which can facilitate detection

(e.g. an epitope tag for detection by monoclonal antibodies) and/or purification (e.g. a polyhistidine-tag to allow purification on a nickel-chelating resin) of the gH protein. An example of a C-terminal extension which includes a myc-tag and a polyhistidine-tag is given as SEQ ID NO: 5. Thus, gH proteins of the invention (e.g. SEQ ID NOs: 6 and 30) may include at their C-termini an amino acid sequence which is at least 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 5. gH proteins of the invention may comprise a truncated gH ectodomain fused to a C-terminal extension.

The ectodomain of gH corresponds to the portion of gH which lacks the hydrophobic TM. The location and length of the ectodomain can be predicted based on pairwise alignment of a given sequence to SEQ ID NO: 1, for example by aligning the amino acid sequence of a gH polypeptide of interest to SEQ ID NO: 1 and identifying the sequence that aligns to residues 24-717 of SEQ ID NO: 1. Similarly, the locations of the TM and C-terminal domains can be predicted by aligning the amino acid sequence of a gH polypeptide of interest to SEQ ID NO: 1 and identifying the sequences that align to residues 718-736 and 737-742 of SEQ ID NO: 1, respectively. Alternatively, the location and length of the ectodomain, the signal sequence and the TM domain can be predicted based on computational analysis of the hydrophobicity along the length of a given gH protein sequence. The signal sequence and the TM domain have the highest levels of hydrophobicity and these two regions flank the ectodomain, which is less hydrophobic.

gH proteins of the invention can have various degrees of identity to SEQ ID NO: 4 such as at least 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 4. gH proteins of the invention can have various degrees of identity to SEQ ID NO: 6 such as at least 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 6. gH proteins of the invention can have various degrees of identity to SEQ ID NO: 29 such as at least 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 29. gH proteins of the invention can have various degrees of identity to SEQ ID NO: 30 such as at least 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 30. Preferred gH proteins: (i) can dimerise with HCMV gL; (ii) form part of the trimeric gH/gL/gO complex; (iii) form part of the pentameric gH/gL/pUL128/pUL130/pUL131A complex; (iv) comprise at least one epitope from SEQ ID NO: 4 or SEQ ID NO: 29; and/or (v) can elicit antibodies *in vivo* which immunologically cross-react with an HCMV virion.

HCMV glycoprotein L (gL) is encoded by the UL115 gene. gL is thought to be essential for viral replication and all known functional properties of gL are directly associated with its dimerization with gH. The gL/gH complex is required for the fusion of viral and plasma membranes leading to virus entry into the host cell. gL from HCMV strain Merlin (GI:39842115, SEQ ID NO: 7) and HCMV strain Towne (GI:239909463, SEQ ID NO: 8) have been reported to be 278 amino acids in length. gL from HCMV strain AD169 (GI:2506510, SEQ ID NO: 9) has been reported to be

278 amino acids in length, include a signal sequence at its N-terminus (amino acid residues 1-35), have two N-glycosylation sites (at residues 74 and 114) and lack a TM domain (Rigoutsos, et al. 2003). The N-terminal signal sequence in SEQ ID NO: 7 is predicted to comprise amino acid residues 1-30. SEQ ID NO: 8 shares 98% amino acid identity with SEQ ID NO: 7. Sequencing of the full-length gL gene from 22 to 39 clinical isolates, as well as laboratory strains AD169, Towne and Toledo revealed less than 2% variation in the amino acid sequences among the isolates (Rasmussen, et al. 2002).

Typically, the N-terminal signal sequence of gL proteins is cleaved by a host cell signal peptidase to produce mature gL proteins. The gL proteins in HCMV membrane complexes of the invention may lack an N-terminal signal sequences. An example of a preferred gL protein of the invention is SEQ ID NO: 31, which lacks an N-terminal signal sequence and consists of amino acid residues 31-278 of SEQ ID NO: 7.

gL proteins of the invention can have various degrees of identity to SEQ ID NO: 7 such as at least 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 7. gL proteins of the invention can have various degrees of identity to SEQ ID NO: 31 such as at least 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 31. Preferred gL proteins: (i) can dimerise with HCMV gH; (ii) form part of the trimeric gH/gL/gO complex; (iii) form part of the pentameric gH/gL/pUL128/pUL130/pUL131A complex; (iv) comprise at least one epitope from SEQ ID NO: 7 or SEQ ID NO: 31; and/or (v) can elicit antibodies *in vivo* which immunologically cross-react with an HCMV virion.

HCMV glycoprotein O (gO), which is encoded by the UL74 gene, has been reported to act as a molecular chaperone, increasing gH/gL ER export and incorporation into virions. It has been proposed that gO competes with pUL128-131A for binding onto gH/gL but is released from gH/gL, so that gH/gL (lacking pUL128-131A) is incorporated into virions (Ryckman, Chase and Johnson 2010). Compared with other viral genes, HCMV gO is unusually variable among different HCMV strains: the variability of the gO amino acid sequence among 22 to 39 clinical isolates, as well as laboratory strains AD169, Towne and Toledo approached 45% (*i.e.* there was only 55% identity between the gO amino acid sequences between different isolates) (Rasmussen, et al. 2002). The gO from HCMV strains Merlin (GI:39842082, SEQ ID NO: 10), AD169 (GI:136968, SEQ ID NO: 11) and Towne have been reported (GI:239909431, SEQ ID NO: 12) to consist of 472, 466 and 457 amino acids, respectively. The gO of HCMV strain AD169, which shares a 73% amino acid similarity to SEQ ID NO: 10, has 18 N-glycosylation sites (at residues 75, 83, 87, 103, 130, 157, 162, 171, 219, 242, 288, 292, 350, 385, 392, 399, 433 and 454), and may include a cleavable signal sequence at its N-terminus (predicted to consist of amino acid residues 1-30), which is absent from the mature polypeptide. Rigoutsos (2003) predicted the presence of TM domains (in regions 10-28 and 190-212) and a coiled coil region (residues 240-272).

Typically, the N-terminal signal sequence of gO proteins is cleaved by a host cell signal peptidase to produce mature gO proteins. The gO proteins in HCMV membrane complexes of the invention may lack an N-terminal signal sequences. An example of a preferred gO protein of the invention is SEQ ID NO: 32, which lacks an N-terminal signal sequence and consists of amino acid residues 31-472 of SEQ ID NO: 10.

gO proteins of the invention can have various degrees of identity to SEQ ID NO: 10 such as at least 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 10. gO proteins of the invention can have various degrees of identity to SEQ ID NO: 32 such as at least 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 32. Preferred gO proteins: (i) can form part of the trimeric gH/gL/gO complex; (ii) cannot form part of the pentameric gH/gL/pUL128/pUL130/pUL131A complex, (iii) comprise at least one epitope of SEQ ID NO: 10 or SEQ ID NO: 32; and/or (iv) can elicit antibodies *in vivo* which immunologically cross-react with an HCMV virion.

The pUL128 from HCMV strain Merlin has been reported (GI:39842124, SEQ ID NO: 13) to consist of 130 amino acids and to contain a 1 nucleotide substitution causing premature termination. The pUL128 from HCMV strains Towne (GI:39841882, SEQ ID NO: 14) and AD169 (GI:59803078, SEQ ID NO: 15) have been reported to consist of 171 amino acids. Due to the premature termination of SEQ ID NO: 13, SEQ ID NOs: 13 and 15 only share 75% identity over the full-length of SEQ ID NO: 15. pUL128 is predicted to have an N-terminal signal sequence, which is located at residues 1-27 of SEQ ID NO: 13, but it is predicted to lack a TM domain. An example of a preferred pUL128 protein of the invention is SEQ ID NO: 33, which lacks an N-terminal signal sequence and consists of amino acid residues 28-171 of SEQ ID NO: 14. SEQ ID NO: 33 also consists of amino acid residues 28-171 of SEQ ID NO: 15.

pUL128 proteins of the invention can have various degrees of identity to SEQ ID NO: 15 such as at least 60%, 70%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 15. pUL128 proteins of the invention can have various degrees of identity to SEQ ID NO: 33 such as at least 60%, 70%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 33. Preferred pUL128 proteins: (i) can form part of the pentameric gH/gL/pUL128/pUL130/pUL131A complex, (ii) comprise at least one epitope of SEQ ID NO: 15 or SEQ ID NO: 33, and/or (iii) can elicit antibodies *in vivo* which immunologically cross-react with an HCMV virion.

UL130 is the central and the largest (214 codons) gene of the UL131A-128 locus. Conceptual translation of the gene predicts a long (25 amino acids) N-terminal signal sequence that precedes a hydrophilic protein containing two potential N-linked glycosylation sites (Asn85 and Asn118) within a putative chemokine domain (amino acids 46 to 120) and an additional N-glycosylation site

(Asn201) close to the end of a unique C-terminal region. pUL130 is predicted to lack a TM domain. It has been reported to be a luminal glycoprotein that is inefficiently secreted from infected cells but is incorporated into the virion envelope as a Golgi-matured form (Patrone, et al. 2005). The sequences of pUL130 from HCMV strain Merlin and Towne are publicly available (GI:39842125, SEQ ID NO: 16 and GI:239909473, SEQ ID NO: 17, respectively) and they consist of 214 and 229 amino acids, respectively. SEQ ID NO: 17 has been reported to contain a frameshift mutation in the C-terminal region of pUL130, and it shares 94% identity to the HCMV SEQ ID NO: 16 over the full-length of SEQ ID NO: 16.

Typically, the N-terminal signal sequence of pUL130 proteins is cleaved by a host cell signal peptidase to produce mature pUL130 proteins. The pUL130 proteins in HCMV membrane complexes of the invention may lack an N-terminal signal sequences. An example of a preferred pUL130 protein of the invention is SEQ ID NO: 34, which lacks an N-terminal signal sequence and consists of amino acid residues 26-214 of SEQ ID NO: 16.

pUL130 proteins of the invention can have various degrees of identity to SEQ ID NO: 16 such as at least 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 16. pUL130 proteins of the invention can have various degrees of identity to SEQ ID NO: 34 such as at least 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 34. As an alternative, pUL130 proteins of the invention can have various degrees of identity to SEQ ID NO: 17 such as at least 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 17. Preferred pUL130 proteins: (i) can form a pentameric gH/gL/pUL128/pUL130/pUL131A complex; (ii) comprise at least one epitope of SEQ ID NO: 16, SEQ ID NO: 34 or SEQ ID NO: 17, respectively; and/or (iii) can elicit antibodies *in vivo* which immunologically cross-react with an HCMV virion.

pUL131A function is required for HCMV replication not only in endothelial cells but also in epithelial cells. The pUL131A from HCMV strains Merlin (GI:39842126, SEQ ID NO: 18) and Towne (GI:239909474, SEQ ID NO: 19) and AD169 (GI:219879712, SEQ ID NO: 20) have been reported to consist of 129, 129 and 76 amino acids, respectively. pUL131A is predicted to contain an N-terminal signal sequence, which is located at residues 1-18 of SEQ ID NO: 18, and to lack a TM domain. The UL131A from strain AD169 has been reported to contain a 1-base-pair insertion, which causes a frame-shift (Wang and Shenk 2005). SEQ ID NO: 18 is 96% identical to SEQ ID NO: 20 over the N-terminal 28 amino acids, but it is only 36% identical to SEQ ID NO: 20 over the full-length of SEQ ID NO: 18 due to the frame-shift in the AD169 UL131A gene.

Typically, the N-terminal signal sequence of pUL131A proteins is cleaved by a host cell signal peptidase to produce mature pUL131A proteins. The pUL131A proteins in HCMV membrane complexes of the invention may lack an N-terminal signal sequences. An example of a preferred pUL131A protein of the invention is SEQ ID NO: 35, which lacks an N-terminal signal sequence and

consists of amino acid residues 19-129 of SEQ ID NO: 18. SEQ ID NO: 35 also consists of amino acid residues 19-129 of SEQ ID NO: 19.

pUL131A proteins of the invention can have various degrees of identity to SEQ ID NO: 17 such as at least 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 18. pUL131A proteins of the invention can have various degrees of identity to SEQ ID NO: 35 such as at least 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 35. Preferred pUL131A proteins: (i) can form pentameric gH/gL/pUL128/pUL130/pUL131A complexes, (ii) comprise at least one epitope of SEQ ID NO: 18 or SEQ ID NO: 35; and/or (iii) can elicit antibodies *in vivo* which immunologically cross-react with an HCMV virion.

If a clinical isolate is passaged in fibroblasts, it can very rapidly accumulate a mutation in the UL131A-128 locus, precluding its ability to infect other cell types. Indeed, as few as three passages in fibroblasts can be sufficient for such mutations to appear in virus stocks. For example, compared to clinical isolates which have not been passaged in fibroblasts, Merlin and Toledo carry mutations in UL128, Towne carries a mutation in the UL130 ORF and AD169 contains a mutated UL131A ORF. Since clinical isolates, but not laboratory strains, efficiently infect and generate infectious progeny in epithelial cells, this cell type, like endothelial cells, holds promise as a laboratory host for the production of clinical HCMV stocks that are not selected for mutations in the UL131A-UL128 locus (Wang and Shenk 2005).

The invention provides an immunogenic composition comprising HCMV complexes, wherein said complexes comprise gH, gL and at least one more HCMV glycoprotein, *e.g.* gH/gL/gO or gH/gL/pUL128/pUL130/pUL131A. Such immunogenic compositions may additionally comprise other HCMV proteins (but preferably not non-envelope HCMV proteins), such as glycoprotein B (gB).

gB is encoded by UL55 and mediates fusion between the virus and the cell membrane. It therefore has a key role to play in entry and infection of the virus. Like many other viral fusion proteins, gB contains hydrophobic loops that insert into the cell membrane and it undergoes a large structural change (pre- and post- fusion conformation) during entry. Like gH, gB displays determinants that are recognized by the host factor TLR2, and it directly interacts with a heterodimer formed between the host factors TLR2 and TLR1. TLR2 mediates NF- κ B activation and inflammatory cytokine responses from cells (Boehme, Guerrero and Compton 2006).

Glycoprotein B (gB) is the most highly conserved of the envelope glycoproteins of human herpesviruses. Although the structure of HCMV gB is currently unknown, it is assumed that the structure of HCMV gB is similar to that of the gBs of HSV and EBV based on sequence homology. The postfusion conformations of HSV-1 and EBV gBs also show a surprising degree of structural homology to the postfusion conformation of fusion protein (G) of vesicular stomatitis virus (VSV) protein, despite the lack of sequence similarity between the gBs of HSV-1 and EBV and VSV-G.

The gB from HCMV strains Merlin (GI:39842076, SEQ ID NO: 21) and Towne (GI:138193, SEQ ID NO: 22) have been reported to consist of 907 amino acids. The gB from HCMV strain AD169 (GI:138192, SEQ ID NO: 23) has been reported to consist of 906 amino acids, have 19 N-glycosylation sites (at residues 37, 68, 73, 85, 208, 281, 286, 302, 341, 383, 405, 409, 417, 447, 452, 464, 465, 554, and 585) and consists of a signal sequence at its N-terminus (at amino acid residues 1-25), an extracellular region (residues 26-751), a TM domain (residues 752-772) and a cytoplasmic domain (residues 773-907) (Rigoutsos, et al. 2003). In a study of 53 women, five subtypes of gB were found among with nucleotide polymorphisms between them ranging from 28 to 124 bp (Murthy, et al. 2011). The gBs of HCMV strain Merlin and AD169 share 95% amino acid similarity. The N-terminal signal sequence is predicted to consist of amino acid residues 1-22 of SEQ ID NO: 21 in the HCMV strain Merlin gB.

Typically, the N-terminal signal sequence of gB proteins is cleaved by a host cell signal peptidase to produce mature gB proteins. The gB proteins in HCMV membrane complexes of the invention may lack an N-terminal signal sequences. An example of a preferred gB protein of the invention is SEQ ID NO: 36, which lacks an N-terminal signal sequence and consists of amino acid residues 23-907 of SEQ ID NO: 21.

gB proteins of the invention can have various degrees of identity to SEQ ID NO: 21 such as at least 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 21. gB proteins of the invention can have various degrees of identity to SEQ ID NO: 36 such as at least 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 36. Preferred gB proteins: (i) comprise at least one epitope of SEQ ID NO: 18 or SEQ ID NO: 36; and/or (ii) can elicit antibodies *in vivo* which immunologically cross-react with an HCMV virion.

Glycosylation

Although gH, gL, gO, gB and pUL130 may be referred to as glycoproteins, this nomenclature should not be taken to mean that these proteins must be glycosylated when used with the invention. On the contrary, in some embodiments of the invention, one or more of polypeptides are not glycosylated. Usually, however, one or more (or all) polypeptides in a complex of the invention are glycosylated. In some embodiments, one or more (or all) polypeptides in a complex of the invention are glycosylated by glycosylation mutants of cultured cells, such as mutated mammalian cells. Such glycosylation mutants produce a pattern of polypeptide glycosylation which differs from a wild-type pattern of glycosylation, i.e. the resulting polypeptide glycoforms differs from wild-type glycoforms.

The level and type of glycosylation is dependent on the species of the host cell. In general, the species most distant to humans in evolutionary terms, such as bacteria, yeasts, fungi, insects and plants, have glycosylation repertoires that are least similar to that of humans. Proteins are usually not glycosylated in bacterial cells, although the transfer of N-linked glycosylation systems into *Escherichia coli* has been reported (Langdon, Cuccui and Wren 2009). Proteins can be glycosylated

in insect cells. However, unlike vertebrate cells, insect cells are unable to produce complex N-linked side chains with penultimate galactose and terminal sialic acid. Hence, the type of glycosylation in insect cells can be sub-optimal for therapeutic proteins. Yeast cells can perform N-linked (to asparagine) and O-linked (to serine/threonine) glycosylation using mannose. Hyperglycosylation (outer chain extension) in the Golgi is a characteristic feature of yeast cells which is not typical of mammalian cells, and this can lead to problems with antibody reactivity. Also, unlike mammalian cells, yeast cells are unable to incorporate sugars other than mannose. In contrast to yeast and insect cells, mammalian glycoproteins expressed in mammalian cells are authentically glycosylated resulting in a recombinant product most similar to that formed *in vivo*.

Hence, preferably glycosylated polypeptides in complexes of the invention: (i) have a mammalian glycosylation pattern; and/or (ii) do not contain an insect cell pattern of glycosylation. In some embodiments, one or more of the proteins of the invention contain complex N-linked side chains with penultimate galactose and terminal sialic acid.

Membrane protein complexes of the invention

HCMV membrane protein complexes of the invention are hetero-oligomeric associations between gH, gL and at least one additional HCMV protein. The proteins in these complexes may be associated by non-covalent and/or covalent interactions. In the gH/gL/gO trimeric complex, disulfide bonds link gH to gO and gL. In the pentameric complex of the invention, gH, gL and pUL128 are typically linked through disulfide bonds, but pUL130 and pUL131A are typically incorporated into the pentameric complex by non-covalent interactions (as shown in Example 7). In some embodiments, the pUL130 protein of the invention and/or pUL131A protein of the invention is incorporated into the pentameric complex by non-covalent interactions. Furthermore, the pUL130 protein of the invention and/or pUL131A may be inter-linked by non-covalent interactions.

The stoichiometries of the trimeric and pentameric complexes are assumed to be 1:1:1 (Huber and Compton 1999) and 1:1:1:1:1 (Ryckman, Chase and Johnson 2010), respectively, but this has yet to be definitively confirmed.

The inventors have discovered that the membrane protein complexes of the invention are able to induce an immunogenic response. Membrane protein complexes of the invention may thus be able to induce immunity against HCMV infection. These two functions are dependent on the retention of epitopes on the membrane protein complexes of the invention that can elicit the production of antibodies, including neutralizing antibodies. A range of conformational epitopes for the pentameric complex are known. For example, Macagno (2010) isolated a panel of human monoclonal antibodies that neutralized HCMV infection of endothelial, epithelial, and myeloid cells. With the single exception of an antibody that bound to a conserved epitope in the UL128 gene product, all other antibodies bound to conformational epitopes that required expression of two or more proteins of the gH/gL/UL128-131A complex. Preferably, the pentameric complexes of the invention possess one or more of the conformational epitopes identified by Macagno (2010).

Each protein of the invention may contain mutations, such as insertions, deletions and substitutions relative to the Merlin strain and/or the AD169 strain of HCMV, so long as these mutations are not detrimental to the use of the proteins as antigens, in particular so long as they retain one or more epitopes that can elicit the production of antibodies that can bind to at least a membrane protein complex of the Merlin and/or AD169 strain of HCMV and/or antibodies that can neutralize the biological effects of said HCMV membrane protein complex. In addition, such mutations should not prevent the capacity of the proteins to form a membrane protein complex of the invention. The ability to form a membrane protein complex of the invention can be tested by performing protein purification, and analyzing the proteins by non-reducing PAGE, Western blot and/or size exclusion chromatography. If the proteins form part of a complex, they may all be present in a single band on a native PAGE gel and/or be present in a single peak in a size exclusion chromatogram.

The HCMV membrane protein complex of the invention can be prepared at various levels of purity *e.g.* at least 80%, 85%, 90%, 95%, or 99% of total protein by mass, *i.e.* the complex makes up at least 80% of the total proteinaceous mass in a composition. The composition may be free from polyacrylamide.

Expression systems

In one aspect, the invention provides a process for expressing the membrane protein complex of the invention. Suitable expression systems for use in the present invention are well known to those of skill in the art and many are described in detail in Doyle (2008). Generally, any system or vector that is suitable to maintain, propagate and express nucleic acid molecules to produce a polypeptide in the required host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those described in Sambrook (2000). Generally, the encoding gene can be placed under the control of a control element such as a promoter, and, optionally, an operator, so that the DNA sequence encoding the desired peptide is transcribed into RNA in the transformed host cell.

Examples of suitable expression systems include, for example, chromosomal, episomal and virus-derived systems, including, for example, vectors derived from: bacterial plasmids, bacteriophage, transposons, yeast episomes, insertion elements, yeast chromosomal elements, viruses such as baculoviruses, papova viruses such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, or combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, including cosmids and phagemids. Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid.

Suitable expression systems include microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected or transfected with virus expression vectors (for example, baculovirus); plant cell systems transformed with virus expression vectors (for example, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors

(for example, Ti or pBR322 plasmids); or animal cell systems. Cell-free translation systems can also be employed to produce the proteins of the invention. Preferably, the proteins of the invention are produced in eukaryotic cells, such as mammalian cells.

Recombinant polypeptides may be expressed transiently or stably. Preferably, the recombinant proteins are expressed stably. For example, cell lines that stably express the peptide of interest may be transfected using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells that successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalised cell lines available from the American Type Culture Collection (ATCC) including, but not limited to, Chinese hamster ovary (CHO), HeLa, baby hamster kidney (BHK), monkey kidney (COS), C127, 3T3, BHK, human embryonic kidney (HEK) 293, Bowes melanoma and human hepatocellular carcinoma (for example Hep G2) cells and a number of other cell lines. Expression in mammalian cells is preferable because the proteins that are produced will have authentic mammalian glycosylation patterns, and thus possess epitopes that are present on infectious HCMV particles. Accordingly, production of membrane protein complexes of the invention in mammalian cells will lead to the production of antibodies that are able to bind to naturally occurring HCMV particles during infection.

In the baculovirus system, the materials for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA (the "MaxBac" kit). These techniques are generally known to those skilled in the art and are described fully in Summers *et al.* (Summers and Smith 1987). Particularly suitable host cells for use in this system include insect cells such as *Drosophila* S2 (*i.e.* by recombinant baculovirus infection of stably transfected *Drosophila* S2 cells) and *Spodoptera* Sf9 cells. In some embodiments, the proteins of the invention are not produced in insect cells.

There are many plant cell culture and whole plant genetic expression systems known in the art. Examples of suitable plant cellular genetic expression systems include those described in US Patent 5,693,506; US Patent 5,659,122; US Patent 5,608,143 and Zenk (1991)²³. In particular, all plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be utilised, so that whole plants are recovered which contain the transferred gene. Practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugar cane, sugar beet, cotton, fruit and other trees, legumes and vegetables.

Examples of prokaryotic expression systems include those that use *streptococci*, *staphylococci*, *E. coli*, *Streptomyces* and *Bacillus subtilis* as host cells.

Examples of fungal expression systems include those that use yeast (for example, *S. cerevisiae*) and *Aspergillus* as host cells.

5 HEK293 cells are suitable for transient expression of the HCMV proteins of the invention due to their high transfectability by various techniques, including the calcium phosphate and polyethylenimine (PEI) methods. A useful cell line of HEK293 is one that expresses the EBNA1 protein of EBV, such as 293-6E (Loignon, et al. 2008). Transformed HEK293 cells have been shown to secrete high levels of the protein complexes of the invention into the growth medium, thus
10 allowing the purification of such protein complexes directly from the growth medium.

CHO cells are particularly suitable mammalian hosts for industrial production of the HCMV proteins of the invention for use as immunogens or antigens because they allow long-term, stable gene expression and high yields of proteins.

15 In some embodiments, the membrane protein complexes of the invention are secreted from the cells in which they are expressed. In other embodiments of the invention, the proteins of the invention are not secreted. In *E. coli*, for example, non-secreted proteins may accumulate in inclusion bodies. Methods for purifying recombinant proteins from inclusion bodies are well known in the art.

20 Transfection can be carried out by a range of methods including using calcium phosphate, electroporation, or by mixing a cationic lipid with the material to produce liposomes which fuse with the cell membrane and deposit their cargo inside.

Nucleic acid constructs

The invention provides a recombinant nucleic acid which encodes gL, gH that lacks a TM domain, and at least one additional HCMV glycoprotein. Preferably, said recombinant nucleic acid:
25 (a) is not a self-replicating RNA molecule; (b) is not an alphavirus replicon; (c) does not encode any alphavirus nonstructural proteins, such as NSP1, NSP2, NSP3 and NSP4; (d) does not contain an Internal Ribosomal Entry Site (IRES), such as EMCV or EV71; and/or (e) does not contain a viral 2A site, such as FMDV. An example of said recombinant nucleic acid may be a single construct which encodes a gL protein of the invention, a gH protein of the invention, a pUL128 protein of the
30 invention, a pUL130 protein of the invention and a pUL131 protein of the invention.

The invention also provides a plurality of recombinant nucleic acids which encode one or more proteins of the invention. For example, in one embodiment the invention provides two nucleic acid constructs: the first construct encoding a gH protein of the invention and a gL protein of the invention, and the second construct encoding a pUL128 protein of the invention, a pUL130 protein
35 of the invention and a pUL131A protein of the invention.

The invention also provides a plurality of recombinant nucleic acids comprising: a first recombinant nucleic acid molecule which encodes a gL protein of the invention; a second recombinant nucleic acid molecule which encodes a gH protein of the invention; and one or more third recombinant nucleic acid molecules which encode one or more additional HCMV proteins.

5 Preferably, said first, second and/or third recombinant nucleic acid molecule(s): (a) is/are not a self-replicating RNA molecule; (b) is/are not (an) alphavirus replicon(s); (c) do(es) not encode any alphavirus nonstructural proteins, such as NSP1, NSP2, NSP3 and NSP4; (d) do(es) not contain: an Internal Ribosomal Entry Site (IRES), such as EMCV or EV71; and/or (e) do(es) not contain a viral 2A site, such as FMDV.

10 In one embodiment, the third recombinant nucleic acid molecule encodes a gO protein of the invention and in another embodiment the third recombinant nucleic acid molecule(s) encode(s) pUL128, pUL130 and pUL131A proteins of the invention. Thus the sequences encoding each individual polypeptide in a complex can be present in a single nucleic acid molecule, or distributed among two or more nucleic acid molecules.

15 In one embodiment, the invention provides a plurality of recombinant nucleic acids comprising: (i) a first recombinant nucleic acid molecule which encodes a gL protein of the invention, (ii) a second recombinant nucleic acid molecule which encodes a gH protein of the invention, (iii) a third recombinant nucleic acid molecule which encodes a pUL128 protein of the invention, (iv) a fourth recombinant nucleic acid molecule which encodes a pUL130 protein of the invention, and (v) a fifth recombinant nucleic acid molecule which encodes a pUL131A protein of the invention. Preferably, said first, second, third, fourth and/or fifth recombinant nucleic acid molecule(s): (a) is/are not a self-replicating RNA molecule; (b) is/are not (an) alphavirus replicon(s); (b) do(es) not encode any alphavirus nonstructural proteins, such as NSP1, NSP2, NSP3 and NSP4; (c) do(es) not contain: an Internal Ribosomal Entry Site (IRES), such as EMCV or EV71; and/or (d) 25 do(es) not contain a viral 2A site, such as FMDV.

Nucleic acid molecules which encode a gH protein of the invention can have various degrees of identity to SEQ ID NO: 24 such as at least 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 24. Nucleic acid molecules which encode a gL protein of the invention can have various degrees of identity to SEQ ID NO: 25 such as 30 at least 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 25. Nucleic acid molecules which encode a pUL128 protein of the invention can have various degrees of identity to SEQ ID NO: 26 such as at least 74%, 76%, 78%, 80%, 82%, 84%, 86%, 88%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 26. Nucleic acid molecules which encode a pUL130 protein of the invention can have various degrees of identity to SEQ ID NO: 27 such as at least 74%, 76%, 78%, 80%, 82%, 84%, 86%, 88%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 27. Nucleic acid molecules which encode a pUL131A protein of the invention can have various degrees of identity to SEQ ID NO: 28 such as at least 74%, 76%,

78%, 80%, 82%, 84%, 86%, 88%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 28.

The nucleic acids of the invention may comprise genomic DNA and/or cDNAs. Unlike cDNA, genomic DNA may contain introns. Some genes are expressed more efficiently when introns are present. Genomic UL128 and UL131A genes each consist of two exons, whereas UL130 does not contain any introns. If genomic sequences are used, the proteins that are produced will depend on splicing, which can vary according to which expression system is used.

The invention provides vectors that comprise said nucleic acids, wherein said vectors include suitable promoters and terminators. Said recombinant nucleic acids may be plasmids, or may be incorporated into the genome of a cell. The promoters in these vectors can be HCMV promoters or non-HCMV promoters.

The invention also provides a process for expressing an HCMV membrane protein complex comprising gH, gL and at least one more HCMV glycoprotein by introducing one or more recombinant nucleic acid molecules which encode gH, gL and at least one more HCMV glycoprotein into an expression system; expressing said one or more nucleic acids in said expression system; and purifying said HCMV membrane protein complex. In some embodiments, this process comprises transfecting cells with a first nucleic acid construct which encodes: either gH, gL, pUL128, pUL130 and pUL131A proteins of the invention or gH, gL and gO proteins of the invention. In some embodiments, this process may comprise transfecting cells with a first nucleic acid construct which encodes a gH protein of the invention, a second nucleic acid construct which encodes a gL protein of the invention; and one or more third nucleic acid construct(s) which encode(s) one or more additional HCMV glycoprotein(s) of the invention. In some embodiments, this process may comprise transfecting cells with a first nucleic acid construct which encodes a gH protein of the invention and a gL protein of the invention; and a second nucleic acid construct(s) which encode(s) one or more additional HCMV glycoprotein(s) of the invention, such as gO or pUL128, pUL130 and pUL131A.

Said HCMV membrane protein complex may be expressed in a mammalian cell. Said isolated HCMV membrane protein complex may optionally be purified.

Cells of the invention

The invention also provides a cell that expresses a nucleic acid molecule or plurality of nucleic acid molecules of the invention, wherein said cell does not comprise the full HCMV genome. Said cell may be stably transformed with said nucleic acid molecule or plurality of nucleic acid molecules of the invention. Preferably, said cell is a mammalian cell, for example a CHO cell.

The invention provides a cell comprising gH, gL and at least one additional HCMV glycoprotein, wherein said cell does not contain the HCMV genome and/or does not produce HCMV virions and/or does not express any non-envelope HCMV proteins.

Isolation and purification of membrane protein complexes

Complexes of the invention are preferably prepared and used in isolated form. The term “isolated” as used herein means removed from its natural environment. Hence, an “isolated HCMV membrane protein complex” does not encompass the HCMV membrane protein complex on the surface of HCMV infected cells or within an infectious HCMV virion.

Using the expression methods described in the examples, the complexes of the invention can be produced at high yields. [See above].

The invention provides processes for purifying HCMV membrane complexes of the invention. Such processes of the invention allow for production of the HCMV membrane protein complex at a purity of >85%, >86%, >87%, >88%, >89%, >90%, >91%, >92%, >93%, >94% or >95% of total protein by mass, as determined by gel electrophoresis. These high levels of purity make the complexes suitable for use as an immunogen in diagnostic applications or as an antigen in vaccine formulations.

The invention provides a process for purifying an HCMV membrane protein complex of the invention, wherein said purification comprises one or more chromatographic steps. Said chromatographic steps may comprise affinity chromatography, such as Ni^{2+} affinity chromatography and/or size exclusion chromatography.

Compositions

The invention also provides compositions comprising the isolated HCMV membrane protein complexes of the invention. The invention also provides compositions comprising the purified HCMV membrane protein complexes of the invention.

The HCMV membrane protein complex can be incorporated into an immunogenic composition, or a vaccine composition. Such compositions can be used to raise antibodies in a mammal (*e.g.* a human).

The invention provides pharmaceutical compositions comprising an HCMV membrane protein complex of the invention. Similarly, the invention provides processes for making a pharmaceutical composition involving combining an HCMV membrane protein complex of the invention with a pharmaceutically acceptable carrier.

In addition to their antigens, immunogenic and pharmaceutical compositions of the invention typically include a pharmaceutically acceptable carrier, and a thorough discussion of such carriers is available in *Remington: The Science and Practice of Pharmacy*.

The pH of the composition is usually between 6 and 8, and more preferably between 6.5 and 7.5 (*e.g.* about 7). Stable pH may be maintained by the use of a buffer *e.g.* a Tris buffer, a citrate buffer, phosphate buffer, or a histidine buffer. Thus a composition will generally include a buffer.

A composition may be sterile and/or pyrogen-free. Compositions may be isotonic with respect to humans.

5 A composition comprises an immunologically effective amount of its antigen(s). An 'immunologically effective amount' is an amount which, when administered to a subject, is effective for eliciting an antibody response against the antigen. This amount can vary depending upon the health and physical condition of the individual to be treated, their age, the capacity of the individual's immune system to synthesise 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 antigen content of compositions of the invention will generally be expressed in terms of the mass of protein per dose. A dose of 10-500µg (*e.g.* 50µg) per antigen can be useful.

15 Immunogenic compositions may include an immunological adjuvant. Thus, for example, they may include an aluminium salt adjuvant or an oil-in-water emulsion (*e.g.* an oil-in-water emulsion comprising squalene, such as MF59 or AS03). Suitable aluminium salts include hydroxides (*e.g.* oxyhydroxides), phosphates (*e.g.* hydroxyphosphates, orthophosphates), (*e.g.* see chapters 8 & 9 of *Vaccine Design...* (1995) eds. Powell & Newman. ISBN: 030644867X. Plenum), or mixtures thereof. The salts can take any suitable form (*e.g.* gel, crystalline, amorphous, *etc.*), with adsorption of antigen to the salt being preferred. The concentration of Al⁺⁺⁺ in a composition for administration to a patient is preferably less than 5mg/ml *e.g.* <4 mg/ml, <3 mg/ml, <2 mg/ml, <1 mg/ml, *etc.* A preferred range is between 0.3 and 1mg/ml. A maximum of 0.85mg/dose is preferred. Aluminium hydroxide and aluminium phosphate adjuvants are particularly suitable for use with the invention.

20 One suitable immunological adjuvant comprises a compound of Formula (I) as defined in WO2011/027222, or a pharmaceutically acceptable salt thereof, adsorbed to an aluminum salt. Many further adjuvants can be used, including any of those disclosed in Powell & Newman (1995).

25 Compositions may include an antimicrobial, particularly when packaged in multiple dose format. Antimicrobials such as thiomersal and 2-phenoxyethanol are commonly found in vaccines, but it is preferred to use either a mercury-free preservative or no preservative at all.

Compositions may comprise detergent *e.g.* a polysorbate, such as polysorbate 80. Detergents are generally present at low levels *e.g.* <0.01%.

30 Compositions may include sodium salts (*e.g.* sodium chloride) to give tonicity. A concentration of 10+2 mg/ml NaCl is typical *e.g.* about 9 mg/ml.

35 Compositions of the invention will generally be administered directly to a subject. Direct delivery may be accomplished by parenteral injection (*e.g.* subcutaneously, intraperitoneally, intravenously, intramuscularly, or to the interstitial space of a tissue), or by any other suitable route. Intramuscular administration is preferred *e.g.* to the thigh or the upper arm. Injection may be via a needle (*e.g.* a hypodermic needle), but needle-free injection may alternatively be used. A typical intramuscular dosage volume is 0.5 ml.

Administration can involve a single dose schedule or a multiple dose schedule.

The subject who is immunised is a human being, who may be any age *e.g.* 0-12 months old, 1-5 years old, 5-18 years old, 18-55 years old, or more than 55 years old.

5 Vaccines of the invention may be prophylactic (*i.e.* to prevent disease) or therapeutic (*i.e.* to reduce or eliminate the symptoms of a disease).

10 Isolated and/or purified HCMV membrane protein complexes of the invention can be administered alone or as either prime or boost in mixed-modality regimes, such as a RNA prime followed by a protein boost. Benefits of the RNA prime protein boost strategy, as compared to a protein prime protein boost strategy, include, for example, increased antibody titers, a more balanced IgG1:IgG2a subtype profile, induction of TH1-type CD4+ T cell-mediated immune response that was similar to that of viral particles, and reduced production of non-neutralizing antibodies. The RNA prime can increase the immunogenicity of compositions regardless of whether they contain or do not contain an adjuvant.

15 In the RNA prime-protein boost strategy, the RNA and the protein are directed to the same target antigen. Examples of suitable modes of delivering RNAs include virus-like replicon particles (VRPs), alphavirus RNA, replicons encapsulated in lipid nanoparticles (LNPs) or formulated RNAs, such as replicons formulated with cationic nanoemulsions (CNEs). Suitable cationic oil-in-water nanoemulsions are disclosed in WO2012/006380 *e.g.* comprising an oil core (*e.g.* comprising squalene) and a cationic lipid (*e.g.* DOTAP, DMTAP, DSTAP, DC-cholesterol, *etc.*).

20 WO2012/051211 discloses that antibodies to the pentameric complex are produced in mice that have been immunized with VRPs and formulated RNAs (CNEs and LNPs) that encode the protein constituents of the pentameric complex. These antibodies have been found to be capable of neutralizing HCMV infection in epithelial cells. The RNA prime-protein boost regimen may involve first (*e.g.* at weeks 0-8) performing one or more priming immunization(s) with RNA (which could be delivered as VRPs, LNPs, CNEs, *etc.*) that encodes one or more of the protein components of an HCMV membrane protein complex of the invention and then perform one or more boosting immunization(s) later (*e.g.* at weeks 24-58) with: an isolated HCMV membrane protein complex of the invention, optionally formulated with an adjuvant or a purified HCMV membrane protein complex of the invention, optionally formulated with an adjuvant.

30 The invention thus provides an immunogenic composition comprising: a self-replicating RNA molecule that encodes a first polypeptide HCMV antigen comprising a first epitope; and a second polypeptide HCMV antigen comprising a second epitope. The invention also relates to kits comprising: (i) a priming composition comprising a self-replicating RNA molecule that encodes a first polypeptide antigen that comprises a first epitope from a pathogen; and (ii) a boosting composition comprising a second polypeptide antigen that comprises a second epitope from the pathogen.

Antigens may be independently selected from the group consisting of gB, gH, gL, gO, pUL128, pUL130 and pUL131A. Said first polypeptide HCMV antigen is preferably an HCMV membrane protein complex of the invention, such as the trimeric gH/gL/gO complex or the pentameric complex. Said second polypeptide HCMV antigen is preferably: an isolated HCMV membrane protein complex of the invention, such as the trimeric gH/gL/gO complex or the pentameric complex or a purified HCMV membrane protein complex of the invention, such as the trimeric gH/gL/gO complex or the pentameric complex. The first and second polypeptide antigens can be substantially the same. The first polypeptide antigen can be a soluble or membrane anchored polypeptide, and the second polypeptide antigen can be a soluble polypeptide. The first polypeptide antigen can be a fusion polypeptide. The second polypeptide antigen can be a fusion polypeptide. The self-replicating RNA can be an alphavirus-derived RNA replicon.

The self-replicating RNA molecule can comprise one or more modified nucleotides. In some embodiments, the self-replicating RNA molecule encodes an HCMV membrane protein complex of the invention, such as the trimeric gH/gL/gO complex or the pentameric complex of the invention. In some embodiments, the second polypeptide antigen is a purified HCMV membrane protein complex of the invention, such as a purified trimeric gH/gL/gO complex or purified pentameric complex of the invention.

In some embodiments, the RNA molecule is encapsulated in, bound to or adsorbed on a cationic lipid, a liposome, a cochleate, a virosome, an immune-stimulating complex, a microparticle, a microsphere, a nanosphere, a unilamellar vesicle, a multilamellar vesicle, an oil-in-water emulsion, a water-in-oil emulsion, an emulsome, a polycationic peptide, a cationic nanoemulsion or combinations thereof.

In some embodiments, the priming composition of the kit or the immunogenic composition further comprises a cationic lipid, a liposome, a cochleate, a virosome, an immune-stimulating complex, a microparticle, a microsphere, a nanosphere, a unilamellar vesicle, a multilamellar vesicle, an oil-in-water emulsion, a water-in-oil emulsion, an emulsome, a polycationic peptide, or a cationic nanoemulsion.

Antibodies of the invention

The invention provides antibodies which recognise an isolated and/or purified HCMV membrane protein complex of the invention, but which do not bind to any of isolated gH, gL, gO, pUL128, pUL130 or pUL131A polypeptides and/or do not bind to isolated gH-gL heterodimers.

As described below, antibodies of the invention may be human or humanised antibodies and/or they may be monoclonal or polyclonal antibodies.

Antibodies of the invention may be polyclonal or monoclonal. Monoclonal antibodies (mAbs) are preferred for many situations. The term "monoclonal" as originally used in relation to antibodies referred to antibodies produced by a single clonal line of immune cells, as opposed to "polyclonal" antibodies that, while all recognizing the same target protein, were produced by different B cells and

would be directed to different epitopes on that protein. As used herein, the word “monoclonal” does not imply any particular cellular origin, but refers to any population of antibodies that display a single binding specificity and affinity for a particular epitope in the same target protein. This usage is normal in the art.

Thus a mAb may be produced using any suitable protein synthesis system, including immune cells, non-immune cells, acellular systems, *etc.* A mAb can thus be produced by a variety of techniques, including conventional monoclonal antibody methodology (*e.g.* the standard somatic cell hybridization technique of Kohler & Milstein), by viral or oncogenic transformation of B lymphocytes, by combinatorial synthesis, by phage display, *etc.* Thus antibodies of the invention may be raised *in vivo* using: an isolated HCMV membrane protein complex of the invention as an antigen or a purified HCMV membrane protein complex of the invention as an antigen. The animal which raises the antibodies can be a mouse, a rat, a rabbit, a goat, *etc.* As an alternative approach, antibodies may be identified using *in vitro* selection methods, such as phage display of antibodies.

Antibodies of the invention can take various forms. For instance, they may be native antibodies, as naturally found in mammals. Native antibodies are made up of heavy chains and light chains. The heavy and light chains are both divided into variable domains and constant domains. The ability of different antibodies to recognize different antigens arises from differences in their variable domains, in both the light and heavy chains. Light chains of native antibodies in vertebrate species are either kappa (κ) or lambda (λ), based on the amino acid sequences of their constant domains. The constant domain of a native antibody's heavy chains will be α , δ , ϵ , γ or μ , giving rise respectively to antibodies of IgA, IgD, IgE, IgG, or IgM class. Classes may be further divided into subclasses or isotypes *e.g.* IgG1, IgG2, IgG3, IgG4, IgA, IgA2, *etc.* Antibodies may also be classified by allotype *e.g.* a γ heavy chain may have G1m allotype a, f, x or z, G2m allotype n, or G3m allotype b0, b1, b3, b4, b5, c3, c5, g1, g5, s, t, u, or v; a κ light chain may have a Km(1), Km(2) or Km(3) allotype. A native IgG antibody has two identical light chains (one constant domain CL and one variable domain VL) and two identical heavy chains (three constant domains CH1 CH2 & CH3 and one variable domain VH), held together by disulfide bridges. The domain and three-dimensional structures of the different classes of native antibodies are well known.

Where an antibody of the invention has a light chain with a constant domain, it may be a κ or λ light chain. Where an antibody of the invention has a heavy chain with a constant domain, it may be an α , δ , ϵ , γ or μ heavy chain. Heavy chains in the γ class (*i.e.* IgG antibodies) are preferred.

Antibodies of the invention may be fragments of native antibodies that retain antigen binding activity. For instance, papain digestion of native antibodies produces two identical antigen-binding fragments, called “Fab” fragments, each with a single antigen-binding site, and a residual “Fc” fragment without antigen-binding activity. Pepsin treatment yields a “F(ab')₂” fragment that has two antigen-binding sites. “Fv” is the minimum fragment of a native antibody that contains a complete antigen-binding site, consisting of a dimer of one heavy chain and one light chain variable domain.

Thus an antibody of the invention may be Fab, Fab', F(ab')₂, Fv, or any other type, or fragment of a native antibody.

An antibody of the invention may be a "single-chain Fv" ("scFv" or "sFv"), comprising a VH and VL domain as a single polypeptide chain. Typically the VH and VL domains are joined by a short polypeptide linker (e.g. >12 amino acids) between the VH and VL domains that enables the scFv to form the desired structure for antigen binding. A typical way of expressing scFv proteins, at least for initial selection, is in the context of a phage display library or other combinatorial library. Multiple scFvs can be linked in a single polypeptide chain.

An antibody of the invention may be a "diabody" or "triabody" *etc.*, comprising multiple linked Fv (scFv) fragments. By using a linker between the VH and VL domains that is too short to allow them to pair with each other (e.g. <12 amino acids), they are forced instead to pair with the complementary domains of another Fv fragment and thus create two antigen-binding sites. These antibodies may include CH and/or CL domains.

An antibody of the invention may be a single variable domain or VHH antibody. Antibodies naturally found in camelids (e.g. camels and llamas) and in sharks contain a heavy chain but no light chain. Thus antigen recognition is determined by a single variable domain, unlike a mammalian native antibody. The constant domain of such antibodies can be omitted while retaining antigen binding activity. One way of expressing single variable domain antibodies, at least for initial selection, is in the context of a phage display library or other combinatorial library.

An antibody of the invention may be a "domain antibody" (dAb). Such dAbs are based on the variable domains of either a heavy or light chain of a human antibody and have a molecular weight of approximately 13 kDa (less than one-tenth the size of a full antibody). By pairing heavy and light chain dAbs that recognize different targets, antibodies with dual specificity can be made. dAbs are cleared from the body quickly and so benefit from a sustained release system, but can additionally be sustained in circulation by fusion to a second dAb that binds to a blood protein (e.g. to serum albumin), by conjugation to polymers (e.g. to a polyethylene glycol), or by other techniques.

The antibody may have a scaffold which is based on the fibronectin type III domain *e.g.* an adnectin or trinectin. The fibronectin-based scaffold is not an immunoglobulin, although the overall fold is closely related to that of the smallest functional antibody fragment. Because of this structure the non-immunoglobulin antibody mimics antigen binding properties that are similar in nature and affinity to those of natural antibodies. The FnIII domain has 7 or 8 beta strands which are distributed between two beta sheets, which themselves pack against each other to form the core of the protein, and further containing loops (analogous to antibody CDRs) which connect the beta strands to each other and are solvent exposed. There are at least three such loops at each edge of the beta sheet sandwich, where the edge is the boundary of the protein perpendicular to the direction of the beta strands. The FnIII loops can be replaced with immunoglobulin CDRs using standard cloning techniques, and can be used in a loop randomization and shuffling strategy *in vitro* that is similar to

the process of affinity maturation of antibodies *in vivo*. The FnIII scaffold may be based on the tenth module of fibronectin type III (*i.e.* ¹⁰Fn3).

Thus the term “antibody” as used herein encompasses a range of proteins having diverse structural features, but usually including at least one immunoglobulin domain, having an all-β protein fold with a 2 layer sandwich of anti-parallel β-strands arranged in two β-sheets.

Antibodies used with the invention may include a single antigen binding site (*e.g.* as in a Fab fragment or a scFv) or multiple antigen binding sites (*e.g.* as in a F(ab')₂ fragment or a diabody or a native antibody). Where an antibody has more than one antigen binding site then advantageously it can result in cross-linking of antigens.

Where an antibody has more than one antigen-binding site, the antibody may be mono specific (*i.e.* all antigen-binding sites recognize the same antigen) or it may be multi specific (*i.e.* the antigen-binding sites recognise more than one antigen).

An antibody of the invention may include a non-protein substance *e.g.* via covalent conjugation. For example, an antibody may include a radio isotope *e.g.* the ZEVALIN™ and BEXXAR™ products include ⁹⁰Y and ¹³¹I isotopes, respectively. As a further example, an antibody may include a cytotoxic molecule *e.g.* MYLOTARG™ is linked to N-acetyl-γ calicheamicin, a bacterial toxin. As a further example, an antibody may include a covalently-attached polymer *e.g.* attachment of polyoxyethylated polyols or polyethylene glycol (PEG) has been reported to increase the circulating half-life of antibodies.

In some embodiments, an antibody can include one or more constant domains (*e.g.* including CH or CL domains). As mentioned above, the constant domains may form a κ or λ light chain or an α, δ, ε, γ or μ heavy chain. Where an antibody includes a constant domain, it may be a native constant domain or a modified constant domain. A heavy chain may include either three (as in α, γ, δ classes) or four (as in μ, ε classes) constant domains. Constant domains are not involved directly in the binding interaction between an antibody and an antigen, but they can provide various effector functions, including but not limited to: participation of the antibody in antibody dependent cellular cytotoxicity (ADCC); C1q binding; complement dependent cytotoxicity; Fc receptor binding; phagocytosis; and down regulation of cell surface receptors.

The constant domains can form a “Fc region”, which is the C-terminal region of a native antibody's heavy chain. Where an antibody of the invention includes a Fc region, it may be a native Fc region or a modified Fc region. A Fc region is important for some antibodies' functions *e.g.* the activity of HERCEPTIN™ is Fc dependent. Although the boundaries of the Fc region of a native antibody may vary, the human IgG heavy chain Fc region is usually defined to stretch from an amino acid residue at position Cys226 or Pro230 to the heavy chain's C-terminus. The Fc region will typically be able to bind one or more Fc receptors, such as a FcγRI (CD64), a FcγRII (*e.g.* FcγRIIA, FcγRIIB1, FcγRIIB2, FcγRIIC), a FcγRIII (*e.g.* FcγRIIIA, FcγRIIIB), a FcRn, FcαR (CD89), FcδR, FcμR, a FcεRI (*e.g.* FcεRIαβγ2 or FcεRIαγ2), FcεRII (*e.g.* FcεRIIA or FcεRIIB), *etc.* The Fc region

may also or alternatively be able to bind to a complement protein, such as C1q. Modifications to an antibody's Fc region can be used to change its effector function(s) *e.g.* to increase or decrease receptor binding affinity.

Antibodies will typically be glycosylated. N-linked glycans attached to the CH2 domain of a heavy chain, for instance, can influence C1q and FcR binding, with aglycosylated antibodies having lower affinity for these receptors. The glycan structure can also affect activity *e.g.* differences in complement-mediated cell death may be seen depending on the number of galactose sugars (0, 1 or 2) at the terminus of a glycan's biantennary chain. An antibody's glycans preferably do not lead to a human immunogenic response after administration.

Antibodies can be prepared in a form free from products with which they would naturally be associated. Contaminant components of an antibody's natural environment include materials such as enzymes, hormones, or other host cell proteins.

Useful antibodies have nanomolar or picomolar affinity constants for their target antigens *e.g.* 10^{-9} M, 10^{-10} M, 10^{-11} M, 10^{-12} M, 10^{-13} M or tighter). Such affinities can be determined using conventional analytical techniques *e.g.* using surface plasmon resonance techniques as embodied in BIAcore™ instrumentation and operated according to the manufacturer's instructions.

The monoclonal antibody used with the invention may be a human antibody, a humanized antibody, a chimeric antibody or (*e.g.* for veterinary purposes) a non-human antibody.

In some embodiments the antibodies are human mAbs. These can be prepared by various means. For example, human B cells producing an antigen of interest can be immortalized *e.g.* by infection with Epstein Barr Virus (EBV), optionally in the presence of a polyclonal B cell activator. Human monoclonal antibodies can also be produced in non-human hosts by replacing the host's own immune system with a functioning human immune system *e.g.* into Scid mice or Trimera mice. Transgenic and transchromosomal mice have been successfully used for generating human monoclonal antibodies, including the "humab mouse" from Medarex and the "xeno-mouse" from Abgenix, collectively referred to herein as "human Ig mice". Phage display has also been successful for this purpose. Unlike non-human antibodies, human antibodies will not elicit an immune response directed against their constant domains when administered to humans. Furthermore, the variable domains of these human antibodies are fully human (in particular the framework regions of the variable domains are fully human, in addition to the complementarity determining regions [CDRs]) and so will not elicit an immune response directed against the variable domain framework regions when administered to humans (except, potentially, for any anti-idiotypic response). Human antibodies do not include any sequences that do not have a human origin.

In some embodiments the antibodies are humanised mAbs, CDR-grafted mAbs or chimeric mAbs. These can be prepared by various means. For example, they may be prepared based on the sequence of a non-human (*e.g.* murine) monoclonal antibody. DNA encoding the non-human heavy and light chain immunoglobulins can be obtained and engineered to contain human immunoglobulin

sequences using standard molecular biology techniques. For example, to create a chimeric antibody, the murine variable regions can be linked to human constant regions using methods known in the art. To create a CDR-grafted antibody, the murine CDR regions can be inserted into a human framework. To create a humanized antibody, one or more non-CDR variable framework residue(s) is also altered.

5 The H1, H2 and H3 CDRs may be transferred together into an acceptor VH domain, but it may also be adequate to transfer only one or two of them. Similarly, one two or all three of the L1, L2 and L3 CDRs may be transferred into an acceptor VL domain. Preferred antibodies will have 1, 2, 3, 4, 5 or all 6 of the donor CDRs. Where only one CDR is transferred, it will typically not be the L2 CDR, which is usually the shortest of the six. Typically the donor CDRs will all be from the same human antibody, but it is also possible to mix them *e.g.* to transfer the light chain CDRs from a first antibody and the heavy chain CDRs from a second antibody.

In some embodiments the antibodies are non-human mAbs. These can be prepared by various means *e.g.* the original Kohler & Milstein technique for preparing murine mAbs.

Methods for raising antibodies of the invention

15 The invention also provides a method for raising antibodies using: isolated HCMV membrane protein complexes of the invention or purified HCMV membrane protein complexes of the invention. These antibodies may be human or humanised. Preferably, these antibodies are specific to the isolated HCMV membrane protein complexes of the invention, and do not bind to isolated gH, gL, gO, pUL128, pUL130 or pUL131A. Such antibodies may be used for diagnostic assays, and may be labelled directly or indirectly. A wide range of antibody labels are known in the art. In other
20 embodiments of the invention, antibodies of the invention can be used for therapy, *e.g.* in the treatment of HCMV infection, and may be in the form of neutralizing antibodies, which can inhibit or neutralize the antigen's biological activity.

Definitions

25 "Recombinant" as used herein to describe a polynucleotide means a polynucleotide of genomic, cDNA, semisynthetic, or synthetic origin which, by virtue of its origin or manipulation: (1) is not associated with all or a portion of the polynucleotide with which it is associated in nature; and/or (2) is linked to a polynucleotide other than that to which it is linked in nature. The term "recombinant" as used with respect to a protein or polypeptide means a polypeptide produced by
30 expression of a recombinant polynucleotide.

The term "comprising" encompasses "including" as well as "consisting" *e.g.* a composition "comprising" X may consist exclusively of X or may include something additional *e.g.* X + Y.

The word "substantially" does not exclude "completely" *e.g.* a composition which is "substantially free" from Y may be completely free from Y. Where necessary, the word
35 "substantially" may be omitted from the definition of the invention.

The term "about" in relation to a numerical value *x* is optional and means, for example, $x \pm 10\%$.

Unless specifically stated, a process comprising a step of mixing two or more components does not require any specific order of mixing. Thus components can be mixed in any order. Where there are three components then two components can be combined with each other, and then the combination may be combined with the third component, *etc.*

5 Where animal (and particularly bovine) materials are used in the culture of cells, they should be obtained from sources that are free from transmissible spongiform encephalopathies (TSEs), and in particular free from bovine spongiform encephalopathy (BSE). Overall, it is preferred to culture cells in the total absence of animal-derived materials.

10 Where a compound is administered to the body as part of a composition then that compound may alternatively be replaced by a suitable prodrug.

Sequence identity between polypeptide sequences is preferably determined by pairwise alignment algorithm using the Needleman-Wunsch global alignment algorithm (Needleman and Wunsch 1970), using default parameters (*e.g.* with Gap opening penalty = 10.0, and with Gap extension penalty = 0.5, using the EBLOSUM62 scoring matrix). This algorithm is conveniently
15 implemented in the *needle* tool in the EMBOSS package (Rice, Longden and Bleasby 2000). Sequence identity should be calculated over the entire length of the polypeptide sequence of the invention.

Particular embodiments of the invention

Particular embodiments of the invention include:

- 20 1. A process for producing an isolated HCMV membrane protein complex comprising gH, gL and at least one additional HCMV glycoprotein, wherein said process comprises recombinant expression of said gH, gL and at least one more HCMV glycoprotein.
2. The process of claim 1, wherein said process comprises the purification of the HCMV membrane protein complex.
- 25 3. A process for expressing an HCMV membrane protein complex comprising gH, gL and at least one more HCMV glycoprotein by:
 - introducing one or more recombinant nucleic acid molecules which encode gH, gL and at least one more HCMV glycoprotein into an expression system;
 - expressing said one or more nucleic acids in said expression system; and
 - 30 - purifying said membrane protein complex.
4. The process of claim 3, which comprises the step of transfecting cells with a first nucleic acid construct which encodes a fragment of gH that lacks the transmembrane domain, a second nucleic acid construct which encodes the gL protein; and a third nucleic acid construct which encodes at least one more HCMV glycoprotein.

5. The process of claim 3 or claim 4, wherein said HCMV membrane protein complex is expressed in a mammalian cell.
6. A process for producing a purified HCMV membrane protein complex comprising gH, gL and at least one more HCMV glycoprotein, wherein said process comprises expressing said HCMV membrane protein complex according to the process of claim 3 and purifying the isolated HCMV membrane protein complex.
7. The process of any preceding claim, wherein the HCMV membrane protein complex consists of gH, gL and gO.
8. The process of any preceding claim, wherein the HCMV membrane protein complex consists of gH, gL, pUL128, pUL130 and pUL131A.

The process of any preceding claim, wherein said gH comprises or consists of any one of the sequences recited in SEQ ID NOs: 1, 2, 3, 4, 6, 29 or 30; and/or said gL comprises or consists of any one of the sequences recited in SEQ ID NOs: 7, 8, 9 or 31.

9. The process of claim 7, wherein said gO comprises or consists of any one of the sequences recited in SEQ ID NOs: 10, 11, 12 or 32.

10. The process of claim 8 or claim 9, wherein said:

- pUL128 comprises or consists of any one of the sequences recited in SEQ ID NOs: 13, 14, 15 or 33;
- pUL130 comprises or consists of any one of the sequences recited in SEQ ID NOs: 16, 17 or 34; and/or
- pUL131A comprises or consists of any one of the sequences recited in SEQ ID NOs: 18, 19, 20 or 35.

11. The process of any preceding claim, wherein said:

- gH comprises or consists of sequences which is at least 70% identical to any one of the sequences recited in SEQ ID NOs: 1, 2, 3, 4, 6, 29 or 30; and/or
- gL comprises or consists of sequences which is at least 70% identical to any one of the sequences recited in SEQ ID NOs: SEQ ID NOs: 7, 8, 9 or 31.

12. The process of claim 11, wherein said gO comprises or consists of sequences which is at least 70% identical to any one of the sequences recited in SEQ ID NOs: SEQ ID NOs: 10, 11, 12 or 32.

13. The process of claim 12, wherein said:

- pUL128 comprises or consists of sequences which is at least 70% identical to any one of the sequences recited in SEQ ID NOs: 13, 14, 15 or 33;

- pUL130 comprises or consists of sequences which is at least 70% identical to any one of the sequences recited in SEQ ID NOs: 16, 17 or 34; and/or
 - pUL131A comprises or consists of sequences which is at least 70% identical to any one of the sequences recited in SEQ ID NOs: 18, 19, 20 or 35.
- 5 14. A process for purifying an HCMV membrane protein complex as defined in any one of the preceding claims, wherein said purification comprises one or more chromatographic steps.
15. The process of claim 14, wherein said chromatographic steps comprise affinity chromatography, preferably Ni^{2+} affinity chromatography and/or size exclusion chromatography.
16. The process of any preceding claim, wherein the HCMV membrane protein complex has a purity
10 of >85%, >86%, >87%, >88%, >89%, >90%, >91%, >92%, >93%, >94% or >95% by mass.
17. The process of any preceding claim, wherein one or more of gH, gL, gO, pUL128, pUL130 and pUL131A in said HCMV membrane protein complex:
- have a mammalian glycosylation pattern; and/or
 - do not contain an insect cell pattern of glycosylation.
- 15 18. A purified HCMV membrane protein complex comprising gH, gL and at least one more HCMV glycoprotein.
19. An HCMV membrane protein complex comprising gH, gL and at least one more HCMV glycoprotein, wherein said complex is produced by the process of any preceding claim.
20. A composition comprising the isolated HCMV membrane protein complex of claim 18 or claim
20 19.
21. The composition of claim 20, wherein said composition does not contain polyacrylamide.
22. The composition of claim 20 or claim 21, wherein said composition does not contain HCMV tegument or capsid proteins.
23. The composition of any one of claims 20-22, wherein said composition is a liquid.
- 25 24. The composition of any one of claims 20-23, wherein said composition is an immunogenic composition.
25. The immunogenic composition of claim 24, which comprises gB.
26. The immunogenic composition of claim 25, wherein said gB comprises or consists of any one of the sequences recited in SEQ ID NOs: 21, 22, 23 or 36.
- 30 27. The immunogenic composition of claim 26, wherein said gB comprises or consists of a sequence which is at least 70% identical to any one of the sequences recited in SEQ ID NOs: 21, 22, 23 or 36.

28. The immunogenic composition of any one of claims 25-27, wherein said gB:

- has a mammalian glycosylation pattern; and/or
- does not contain an insect cell pattern of glycosylation.

29. The immunogenic composition of any one of claim 24-28, wherein said composition is a vaccine.

5 30. The immunogenic composition of any one of claims 24-29, which comprises an adjuvant.

31. The immunogenic composition of claim 30, wherein said adjuvant is an oil-in-water emulsion or an aluminium salt.

32. An immunogenic composition comprising:

- a self-replicating RNA molecule that encodes an HCMV membrane protein complex; and
- 10 - the HCMV membrane protein complex of claim 18 or claim 19.

33. A kit comprising:

- a priming composition comprising a self-replicating RNA molecule that encodes an HCMV membrane protein complex; and
- 15 - a boosting composition comprising the HCMV membrane protein complex of claim 18 or claim 19.

34. A recombinant nucleic acid molecule which encodes gL, gH that lacks a transmembrane domain, and at least one additional HCMV glycoprotein, wherein said recombinant nucleic acid:

- (a) is not a self-replicating RNA molecule;
- (b) is not an alphavirus replicon;
- 20 (c) does not encode any alphavirus nonstructural proteins, such as NSP1, NSP2, NSP3 and NSP4;
- (d) does not contain: an Internal Ribosomal Entry Site (IRES), such as EMCV or EV71; and/or
- (e) does not contain a viral 2A site, such as FMDV.

25 35. The recombinant nucleic acid molecule of claim 34, wherein said recombinant nucleic acid molecule encodes:

- gL, gH that lacks a transmembrane domain, pUL128, pUL130 and pUL131A; or
- gL, gH that lacks a transmembrane domain and gO.

30 36. A plurality of recombinant nucleic acids, wherein said plurality of recombinant nucleic acids encode gL, gH that lacks a transmembrane domain, and at least one additional HCMV glycoprotein, wherein one or more or all of said plurality of recombinant nucleic acids:

(a) is not a self-replicating RNA molecule;

(b) is not an alphavirus replicon;

(c) does not encode any alphavirus nonstructural proteins, such as NSP1, NSP2, NSP3 and NSP4;

5 (d) does not contain: an Internal Ribosomal Entry Site (IRES), such as EMCV or EV71; and/or

(e) does not contain a viral 2A site, such as FMDV.

37. The plurality of recombinant nucleic acids of claim 36 comprising:

- a first construct encoding gH that lacks a transmembrane domain and gL; and
- 10 - a second construct encoding one additional HCMV glycoprotein.

38. The plurality of recombinant nucleic acids of claim 37, wherein said second construct encodes:

- pUL128, pUL130 and pUL131A; or
- gO.

39. The plurality of recombinant nucleic acids of claim 36 comprising:

- 15 - a first recombinant nucleic acid molecule which encodes gL;
- a second recombinant nucleic acid molecule which encodes a fragment of gH that lacks a transmembrane domain; and
- one or more third recombinant nucleic acid molecules which encode one or more additional HCMV proteins.

20 40. The recombinant nucleic acid molecule of claim 34 or claim 35, or the plurality of recombinant nucleic acids molecules of any one of claims 36-39, wherein:

(a) said nucleic acid molecule which encodes gH is 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 24;

25 (b) said nucleic acid molecule which encodes gL is 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 25;

(c) said nucleic acid molecule which encodes pUL128 is 74%, 76%, 78%, 80%, 82%, 84%, 86%, 88%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 26;

30 (d) said nucleic acid molecule which encodes pUL130 is 74%, 76%, 78%, 80%, 82%, 84%, 86%, 88%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 27; and/or

(e) said nucleic acid molecule which encodes pUL131A is 74%, 76%, 78%, 80%, 82%, 84%, 86%, 88%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the sequence recited in SEQ ID NO: 28.

41. A cell that expresses the recombinant nucleic acid or plurality of recombinant nucleic acids of
5 any one of claims 34-40, wherein said cell does not comprise the full HCMV genome.
42. The cell of claim 41, wherein said cell is stably transformed with said recombinant nucleic acid or plurality of recombinant nucleic acids.
43. The cell of claim 41 or claim 42, wherein said cell is a mammalian cell.
44. A cell comprising gH, gL and at least one additional HCMV glycoprotein, wherein said cell does
10 not:
- (a) contain the HCMV genome;
 - (b) produce HCMV virions;
 - (c) contain self-replicating RNA molecules encoding said gH, gL and at least one additional HCMV glycoprotein; and/or
 - 15 (d) contain alphavirus replicons.
45. A process for producing an isolated or a purified HCMV membrane protein complex comprising gH, gL and at least one additional HCMV glycoprotein, wherein said process involves growing the cell of any one of claims 41-44 in growth medium.
46. The process of claim 45, wherein said HCMV membrane protein complex is secreted into said
20 growth medium.
47. The process of claim 46, wherein said HCMV membrane protein complex accumulates to a concentration of >0.8 mg, >0.85 mg, >0.88mg, >0.9 mg, >0.95 mg, >1mg, >1.5mg, >2mg, >2.5mg, >3mg, >3.5mg, >4mg, >4.5mg, >5mg of complex per litre of growth medium.
48. The process of any one of claims 45-47, wherein said process comprises purifying said HCMV
25 membrane protein complex from said growth medium.
49. A method for raising antibodies using the HCMV membrane protein complex of claim 18 or claim 19.
50. The method of claim 49, wherein said antibodies are human or humanised.
51. The method of claim 49 or 50, wherein said antibodies are neutralizing antibodies.
- 30 52. An antibody produced by the method of any one of claims 49-51.

53. An antibody produced by the method of any one of claims 49-52, wherein said antibody binds to the isolated HCMV membrane protein complexes of any preceding claim, but not isolated gH, gL, gO, pUL128, pUL130 or pUL131A.

54. The antibody of claim 52 or claim 53, wherein said antibody is for use in a diagnostic assay.

55. The antibody of any one of claims 52-54, wherein said antibody is labelled directly or indirectly.

56. The antibody of any one of claims 52-53, wherein said antibody is for use in therapy.

58. An RNA prime-protein boost regimen comprising:

- performing one or more priming immunization(s) with RNA that encodes one or more of the protein components of an HCMV membrane protein complex, wherein said HCMV membrane protein complex comprises gH, gL and at least one additional HCMV glycoprotein;
- performing one or more boosting immunization(s) later with a purified HCMV membrane protein complex, wherein said purified HCMV membrane protein complex comprises gH, gL and at least one additional HCMV glycoprotein.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 shows the plasmid map of the 7591 bp construct used for the expression of His-tagged, soluble gH in mammalian cells. The nucleotide sequence of this construct is given as SEQ ID NO: 23. The construct comprises a CMV promoter, a gene which encodes SEQ ID NO: 4 (the gH soluble protein, which consists of amino acid residues 1-715 of the full-length gH protein) fused to a myc-polyhistidine tag, bovine growth hormone (BGH) polyadenylation (bgh-PolyA) signal termination sequence, F1 phage origin of replication, SV40 origin of replication, Neomycin resistance gene, pUC subgenomic promoter and an ampicillin resistance gene.

FIG. 2 shows the plasmid map of the 5156 bp construct used for the expression of gL in mammalian cells. The nucleotide sequence of this construct is given as SEQ ID NO: 24. The construct comprises a CMV promoter/enhancer which contains intronA, a gene which encodes SEQ ID NO: 7 (gL), bovine growth hormone (BGH) polyadenylation (bgh-PolyA) signal termination sequence, a kanamycin resistance gene and the ColE1 origin of replication.

FIG. 3 shows the plasmid map of the 4835 bp construct used for the expression of UL128 in mammalian cells. The nucleotide sequence of this construct is given as SEQ ID NO: 25. The construct comprises a CMV promoter/enhancer which contains intronA, the UL128 gene, bovine growth hormone (BGH) polyadenylation (bgh-PolyA) signal termination sequence, a kanamycin resistance gene and the ColE1 origin of replication.

FIG. 4 shows the plasmid map of the 4964 bp construct used for the expression of UL130 in mammalian cells. The nucleotide sequence of this construct is given as SEQ ID NO: 26. The

construct comprises a CMV promoter/enhancer which contains intronA, the UL130 gene, bovine growth hormone (BGH) polyadenylation (bgh-PolyA) signal termination sequence, a kanamycin resistance gene and the ColE1 origin of replication.

FIG. 5 shows the plasmid map of the 4709 bp construct used for the expression of UL131A in mammalian cells. The nucleotide sequence of this construct is given as SEQ ID NO: 27. The construct comprises a CMV promoter/enhancer which contains intronA, the UL131A gene, bovine growth hormone (BGH) polyadenylation (bgh-PolyA) signal termination sequence, a kanamycin resistance gene and the ColE1 origin of replication.

FIG. 6 lanes A-C correspond to a silver-stained SDS-PAGE, whereas lanes E-K correspond to Western blot analysis. In lanes E-G, green = Anti-APtag (gH) and red = Anti-gL, whereas in lanes H-K, red = anti-6His (gO). The ladder is shown in lanes A, E and H. Lanes B, F and I correspond to samples that have been briefly heated to near boiling in the presence of dithiothreitol (DTT), whereas lanes C, G and K correspond to samples that have not been subjected to heat or DTT.

FIG. 7 shows a size exclusion chromatogram. All five proteins were eluted as a single peak, thus demonstrating the presence of an intact pentameric complex.

FIG. 8 shows SDS-PAGE and Western blot analysis.

FIG. 9 shows a comparison between the neutralization titers elicited by gH/gL either alone or formulated with MF59 versus those elicited by the pentameric complex either alone or formulated with MF59. Purified pentameric complex formulated with MF59 elicited higher neutralizing titers than the unformulated protein.

FIG. 10 shows the neutralization titers elicited by pentameric complex alone (A) and pentameric complex formulated with MF59 (B), aluminium oxyhydroxide (C) and aluminium hydroxide to which a TLR7 agonist is adsorbed (D).

FIG. 11 is a graph showing that self amplifying RNA and subunit, alone or in combination, elicit high neutralizing antibody titers. In particular, the neutralization titers elicited by: the self amplifying RNA (self replicating RNA, encapsulated in LNPs) encoding the HCMV pentameric complex; purified pentameric subunit adjuvanted with MF59; different sequences of self amplifying RNA followed by subunit in MF59; or a combination of self amplifying RNA and subunit are shown. Self amplifying RNA and subunit dose were 1 μ g, mixed dose was 1+1 μ g. Neutralizing assay: VR1814 infection of ARPE-19 cells in presence of complement.

FIG. 12 are graphs showing the CD4⁺ T cell responses (in terms of the net % of CD4⁺ T cells, and the % of Th0, Th1 and Th2 CD4) to the vaccinations using purified gH/gL and pentameric subunits at (A) 3wp3 (day 64) and (B) 4wp3 (day 71).

FIG. 13 are graphs showing the CD4⁺ T cell responses (in terms of the net % of CD4⁺ T cells) to the vaccinations using purified pentameric complex at (A) 3wp3 (day 64) and (B) 4wp3 (day 71).

FIG. 14 are graphs showing the CD8+ T cell responses (in terms of the net % of CD8+ T cells) to the vaccinations using purified pentameric complex at (A) 3wp3 (day 64) and (B) 4wp3 (day 71).

FIG. 15 are graphs showing the CD8+ T cell responses (in terms of the net % of CD8+ T cells) to the vaccinations using gH peptide pool 2 at (A) 3wp3 (day 64) and (B) 4wp3 (day 71).

5 EXAMPLES

Example 1 – Immunogenicity of replicons expressing pentameric complex

In WO 2012/051211, an alphavirus replicon vector expressing all five proteins of the pentameric complex (gH, gL, pUL128, pUL130 and pUL131A) from a single construct was produced. The RNAs expressed by this vector were either packaged into VRPs or formulated for RNA vaccination either by complexing replicons with CNEs or by encapsulating replicons in LNPs. The VRPs and formulated RNAs were used to immunize BALB/c mice at three-week intervals. Sera from immunized mice were used in microneutralization assays to block infection of epithelial cells with HCMV TB40 (in the absence of complement). The HCMV TB40 strain is similar to clinical strains and infects endothelial and epithelial cells, natural target cells of HCMV *in vivo* (17). Microneutralization assay data demonstrated that replicons expressing the pentameric complex elicited more potently neutralizing antibodies than replicons expressing gH/gL. Microneutralisation data also showed that antibodies elicited by RNA expressing pentameric complex are able to neutralize HCMV infection in epithelial cells (because they target the pentameric complex), but not in fibroblasts (in which infection does not require the pentameric complex), thus demonstrating that RNA expressing the pentameric complex elicits antibodies that specifically target the intact pentameric complex rather than a gH/gL dimer. This work demonstrates that antibodies can be raised against the pentameric complex, and these antibodies are capable of neutralizing HCMV infection.

Example 2 – Constructs for stable expression of the pentameric complex in mammalian cells

Five nucleic acid constructs were produced to enable to expression and purification of the pentameric complex in mammalian cells. Previous attempts to purify the pentameric complex when constructs including the gene encoding the full-length gH protein were unsuccessful. In an attempt to overcome this problem, the inventors produced constructs that encode only the ectodomain of gH (gHecto) with a C-terminal myc-(His)₆ tag (SEQ ID NO: 6) rather than the full-length sequence. The following five constructs were used to produce the pentameric complex: a construct encoding SEQ ID NO: 6 (Fig. 1 and SEQ ID NO: 23), a construct encoding full-length gL (Fig. 2 and SEQ ID NO: 24), a construct encoding full-length pUL128 (Fig. 3 and SEQ ID NO: 25), a construct encoding full-length pUL130 (Fig. 4 and SEQ ID NO: 26) and a construct encoding full-length pUL131A (Fig. 5 and SEQ ID NO: 27).

Example 3 – Protocol for transfection and expression of protein complexes in 293-6E cells*Materials:*

Mammalian 293-6E cells (Gibco FreeStyle 293 Expression Medium; Opti-MEM and Polyethylenimine Linear (PEI), MW 25,000.

5 *Preparation of Cells*

293-6E cells were maintained in serum-free 293 Expression Medium. Once the cells are doubling every 24 hours and at more than 90% viability, the cells were diluted to a density of 1×10^6 /1mL media.

Transfection

10 DNA corresponding to each construct was diluted in Opti-MEM using a volume of Opti-MEM that is 2.5% of the volume of cell culture to be transfected. DNA constructs were combined in an appropriate ratio such that total DNA was equal to 1 μ g /1 mL culture volume. For example, for the stable expression of the pentameric complex using 1 L cell culture, 200 μ g of each of SEQ ID NOs: 23-27 were added to 25 mL Opti-MEM.

15 PEI was diluted in Opti-MEM using a volume of Opti-MEM that is 2.5% the volume of cell culture to be transfected. The diluted PEI was incubated for 5 min at room temperature with occasional swirling to mix. 3 μ g PEI was used per 1mL culture (*e.g.*, for 1L cell culture 3mg PEI was diluted in 25mL Opti-MEM).

20 The DNA mix was added to the PEI mix (so that 1 μ g total DNA + 3 μ g PEI were used per 1 mL culture), swirled and incubated at room temperature for 30min. The DNA-PEI mix was added to cells by gradually adding mixture and swirling cells occasionally such that mixture was added evenly to culture. The cells were then immediately returned to the original growth conditions.

Expression and Harvest

25 Three days post-transfection, the media was harvested by spinning the cells down at 2,000 rpm. The media was then concentrated approximately 10x and diafiltered into buffer containing 300mM NaCl, 25mM Tris pH 7.5. Finally, the dialyzed media was frozen at -80° C. Fresh media were added to the culture and three more days later the media was harvested and concentrated/diafiltered as above.

Example 4 – Protocol for Purification of the HCMV complexes30 *Materials:*

35 GE AKTExpress; Qiagen Ni-NTA Superflow Cartridges, 5ml; 96 Well Clear V-Bottom 2 mL Polypropylene Block; Buffer A (= binding buffer) 50 mM Tris-HCl pH 7.5, 300 mM NaCl, 5 mM Imidazole; Buffer B (= elution buffer) 50mM Tris-HCl pH 7.5, 300 mM NaCl, 1 M Imidazole; SEC buffer (= for buffer exchange and size exclusion chromatography); 2 x 500 ml 0.5 M NaOH solution for system cleaning; Invitrogen NuPAGE® Novex 4-12% Bis-Tris Gel 1.0 mm, 12 well; NuPAGE® LDS Sample Buffer (4X) and NuPAGE® Sample Reducing Agent (10X).

Procedures

Buffers were prepared with endotoxin-free stock solutions and filtered Milli-Q water.

The medium to be purified was thawed in a warm water bath. Meanwhile AKTAexpress was cleaned with 0.5M NaOH, to get rid of possible endotoxin contamination as well as residual protein/media, and then rinsed with filtered endotoxin-free Milli-Q water.

The Ni-NTA superflow cartridge was connected to the AKTAexpress system, and the water and buffers A and B were set in place. The program “Ni-NTA prep” was then started in order to flush the system with buffers, wash away ethanol in the column, and run buffer A through the column to equilibrate it.

The fraction collecting 96-well was prepared by putting 3.5µl 500mM EDTA solution in each well. The loading sample was also prepared immediately before loading. When the medium had thawed, 1/500 volume of 2.5M Imidazole stock was added, and mixed gently. The loading sample was then set in place.

The purification program was then started. This program performed the following steps: sample loaded onto the column; washed away the unbound compounds with buffer A until the baseline settles; washed away non-specifically binding compounds with 15 column volumes of 2.5% buffer B (= 30mM Imidazole); elute the HCMV protein complex with 10 column volumes of 25% buffer B (= 254 mM Imidazole); and finally washed the column with 5 column volumes of 100 % buffer B (= 1 M Imidazole). The sample loading rate was 2.5 ml/min, whereas the wash and elution rate was 5 mL/ min. The flow through in whole and 1.75 ml each of fractions from wash and elution were collected).

Six or seven fractions from the 250 mM Imidazole elution peak were selected to be analysed by SDS-PAGE. The four or five fractions which had the highest amount of protein according to the SDS-PAGE gel were pooled together, and dialyzed against 2 L SEC buffer for 1h at RT, twice. The dialysate was recovered and the concentration measured using the BCA method. The presence of all components of the complex in the purified protein pool was verified by Western blot.

In order to increase the purity of some samples, size exclusion chromatography (SEC) was performed. On a Superdex 200 10/300 GL (GE Healthcare, 17-5175-01) equilibrated with SEC buffer for more than two column volumes. The dialyzed pool was loaded one column volume of buffer was run through the column, and 1 mL fractions were collected. SDS-PAGE was performed to determine which fractions to pool and keep.

Example 5 – Expression, purification and characterization of the trimeric gH/gL/gO complex

The following three constructs were produced: gH ectodomain with C-terminal APP tag, full-length gL and full-length gO with a C-terminal (His)6 tag. These three constructs were co-expressed in HEK 293 cells according to the method described in Example 3. The purification method of Example 4, involving Ni²⁺ affinity chromatography, was performed.

The purified samples were then subjected to SDS-PAGE, followed by Western blot analysis using anti-APptag (gH), anti-gL and anti-6His (gO) antibodies. These three antibodies bound to different proteins under reducing conditions (+heat, +DTT), but all bound to a single complex in non-reducing conditions (-heat, -DTT). These results demonstrate the successful purification of gH/gL/gO as a trimeric complex. Approximately 0.5mg/L media of the complex was purified from HEK 293 cells. SEC increased the purity of the gH/gL/gO complex.

Example 6 – Expression and purification of the pentameric complex

HEK293 cells were co-transfected with the five constructs described in Example 2 according to the method described in Example 3, the media were harvested 3 and 6 days post transfection, and the expressed protein was purified by Ni-NTA chromatography according to the method of Example 4.

A single peak in the size exclusion chromatogram for the purified pentameric complex (Fig. 7) indicated that an intact, monomeric complex was successfully purified. To assess whether all five members of the pentameric complex: gHecto-His, gL, pUL128, pUL130 and pUL131A were present within the purified complex, SDS-PAGE and Western blot analysis was performed.

Example 7 –Western blotting of HCMV pentameric complex

Materials:

BioRad Trans-Blot SD Semi-Dry Electrophoretic Transfer Cell; Invitrogen Nitrocellulose membranes, 0.2µm pore; NuPAGE® Transfer Buffer (20X); Methanol; Odyssey® Blocking Buffer; DPBS; 10x PBS; Primary antibodies: Mouse anti-His tag, rabbit anti-gL 27-46, mouse anti-pUL128 4B10, mouse anti-UL130 3E3, and rabbit anti-UL131A 90-136; secondary antibodies: IRDye 800CW Goat anti-Rabbit IgG (H + L), IRDye 680LT Goat anti-Mouse IgG (H + L).

Procedures

Three sets of antibodies were used, one for the detection of gHecto-His/gL, one for the detection of pUL128/pUL131A, and one for the detection of pUL130.

9 µl of the protein was mixed with 3 µl of LDS sample buffer/reducing agent mixture (9:1) and boiled at 100°C for 3 minutes. The boiled samples were loaded on the wells and run at 200 V for 35 minutes. The protein was then transferred to a 0.2 µm nitrocellulose membrane using the Trans-Blot SD Semi-Dry Electrophoretic Transfer Cell at 20 V for 35 minutes. The membrane was blocked with blocking buffer at room temperature for 30 minutes. The membranes were incubated with primary antibody solutions at RT for 1 hr. Primary antibody solutions consisted of: (A) 1:10000 dilution of anti-His tag and 1:5000 dilution of anti-gL, (B) 1:500 dilution of anti-pUL128 and 1:1000 dilution of anti-UL131A, and (C) 1:500 dilution of anti-UL130, all diluted in 1:1 mixture of the blocking buffer and DPBS. The membranes were rinsed three times with PBS + 0.1% Tween at RT, and then incubated with secondary antibody solutions at RT for 1hr. For membranes (A) and (B), 1:25000 dilution each of anti-mouse and anti-rabbit antibodies in a 1:1 mixture with the blocking buffer and DPBS was used. For membrane (C), 1:25000 dilution of anti-mouse antibodies were used

only. The membranes were rinsed three times with PBS+ 0.1% Tween at RT, and then rinsed with DPBS once at RT. The membrane was scanned with Odyssey Infrared Imaging System (Li-Cor 9201) and analyzed using Odyssey software version 2.1.12.

All five members of the pentameric complex: gHecto-His, gL, pUL128, pUL130 and pUL131A were present within the purified complex, were identified by SDS-PAGE and Western blot analysis (Fig. 8), thus confirming the successful purification of the pentameric complex.

gHecto-His, gL and pUL128 co-migrate in an SDS Bis-tris gel without heat nor DTT, but the association of pUL128 and gH/gL completely disappears in the presence of DTT (data not shown), thus demonstrating that pUL128 associates with gH/gL through a disulfide bond. pUL130/pUL131A does not co-migrate with gH/gL/pUL128, indicating that it is incorporated into the pentameric complex by non-covalent bonds.

The most dominant band on the gels was near the gH/gL complex position with non-reducing, non-boiling conditions, and was thus thought to correspond to the HCMV pentameric complex. The purity was estimated to be 90%, by mass. With SEC purification the purity increased to nearly 100%. Approximately 0.6 mg of pentameric complex per litre of media could be purified by Ni-NTA purification.

Example 8 - Recombinant pentameric complex binds to conformation dependent neutralizing antibodies

A panel of human neutralizing antibodies (HumAb) were isolated from memory B cells of seropositive individuals. Direct ELISA was performed in which pentameric complex protein was immobilized on a plate and neutralizing antibodies were added in a 10-fold series dilution. The results of this ELISA are shown in Table 1. The HumAb bound to several UL proteins and gH/gL/pUL128/pUL130, which confirmed the correct conformation of the pentameric complex forms. Binding of HumAb against gH suggests that these epitopes are exposed on the recombinant complex.

Table 1

HuMab	10P3	5A2	4I22	8J16	7I13	15D8	8I21	3G16	11B12	cytotect	Chick. Lyso.
epitope	pUL130/ pUL131A site 3	pUL130/ pUL131 A site 2	pUL130/ pUL131 A site 1	pUL128/ pUL130/ pUL131A site 1	pUL128/ pUL130/ pUL131 A site 2	pUL128	gH/gL/ pUL128/ pUL130	gH/gL site A	gH/gL site B		
KD (nM)	0.17	0.1	0.1	0.13	0.11	0.33	0.57	0.06	0.21	9.8	-

Example 9 – The pentameric complex elicits higher neutralizing titers than gH/gL

0.1/1 µg purified gH/gL and 0.16 µg/1.6 µg pentameric complex protein were formulated with or without MF59 and used for immunization of mice. The neutralizing titer of 3wp3 showed approximately a 4-fold increase for pentameric complex compared with gH/gL (which in turn elicits

higher neutralizing titers compared with gB) (Fig. 9). Purified pentameric complex formulated with MF59 elicited higher neutralizing titers than the unformulated protein (Fig. 9). Formulation of the pentameric complex with aluminium hydroxide to which a TLR7 agonist is adsorbed elicits even higher neutralizing titers than formulations with aluminium oxyhydroxide and/or MF59 (Fig. 10).

5 **Example 10 – Production of monoclonal antibodies using purified gH/gL/gO as an antigen**

The purified gH/gL/gO complex is diluted to 0.4 mg/mL in 150 mM NaCl, 25 mM Tris (pH 7.5), 1 mM EDTA and frozen. The gH/gL/gO complex is thawed on the day of vaccination and 438 µL of adjuvant is added to the thawed gH/gL/gO complex and mixed well by inverting the tubes at least 10 times. The resulting composition is used within 1 hour of mixing.

10 Two groups of three 6-8 week-old female BALB/c mice are immunized with a composition comprising 50 µg purified gH/gL/gO, and MF59 adjuvant. Each mouse is immunized with 125 µL in each quadriceps muscle (*i.e.* each mouse received 250 µL) and bled from their orbital sinuses.

For each group of three mice, the immunization schedule is summarised in Table 2 below:

Table 2

<i>Weeks</i>	0	0	3	5	6	8
<i>Procedure</i>	Bleed 0	Immunization 1	Immunization 2	Bleed 2	Immunization 3	Bleed 3

15 In order to purify monoclonal antibodies against gH/gL/gO, the gH/gL/gO antigen is used for primary ELISA screening and then the positive clones from the primary screening are then further screened using the gH/gL antigen.

A similar method can be employed to produce monoclonal antibodies using the purified pentameric complex as an antigen.

Example 11: A SAM™ Vaccine Prime, Protein Boost and Coadministration of RNA and subunit Using the HCMV Pentameric Antigen

25 Mice were immunized three times, three weeks apart with a SAM vaccine, which is a self-replicating RNA as described herein, encoding the HCMV pentameric complex, purified pentameric subunit adjuvanted with MF59, different sequences of SAM vaccine followed by subunit in MF59, or a combination of the two (Table 3). The SAM vaccine was encapsulated in synthetic LNPs for non-viral delivery. A group of control mice did not receive any vaccine.

Table 3

Group	No. Mice	No. Doses	Antigen	Formulation	Dose
1	4	-	3	-	1
2	8	3	SAM vaccine encoding pentameric complex (Penta SAM vaccine)	Lipid nanoparticle	1 microgram

				(LNP)	
3	8	3	Purified pentameric complex (Penta subunit)	MF59	1 microgram
4	8	3	1st Penta SAM vaccine 2nd and 3rd Penta subunit	LNP MF59	1 microgram
5	8	3	1st and 2nd Penta SAM vaccine 3rd Penta subunit	LNP MF59	1 microgram
6	8	3	Penta SAM + Penta subunit (mixed)	LNP	1 microgram + 1 microgram

Sera were harvested three weeks after each immunization and used for ELISA to determine binding antibody titers, using the same purified antigen in the assay as in the subunit vaccine. The sera were also used for HCMV microneutralization assay using TB40 or VR1814 infection of ARPE-19 epithelial cells. Three or four weeks after the third immunization, spleens were extracted from sacrificed mice. Spleen cells were stimulated in vitro with purified protein or a pool of 15-mer peptides (overlapping by 11 amino acids) corresponding to the c-terminal half of the gH protein, stained for cytokine expression, and analyzed using flow cytometry.

SAM vaccine and subunit/MF59 alone elicited potently neutralizing antibody responses after three doses (Figure 11). The pentameric subunit in MF59 did not respond as well as pentameric SAM vaccine to the first and second dose of vaccine, but titers elicited by subunit/MF59 surpassed titers elicited by SAM vaccine after the third dose. One SAM vaccine prime followed by a single dose of subunit/MF59 elicited stronger neutralizing responses than two doses of subunit/MF59, but was equal to SAM alone. A second subunit/MF59 boost administered to these animals raised neutralizing responses to a level that exceeded those seen after three doses of subunit/MF59 or SAM vaccine. Two doses of SAM vaccine followed by a single dose of subunit/MF59 did not appear to benefit neutralizing responses compared to either subunit/MF or SAM vaccine alone. Mixing SAM vaccine with subunit, without MF59, elicited a strong response after the first dose, similar to RNA alone, and elicited strong neutralizing titers after two and three doses.

CD4⁺ T cell responses to the vaccinations using purified gH/gL and pentameric subunits were analyzed. SAM vaccine prime protein boost and mixed SAM vaccine + subunit elicited more CD4⁺ T cells responding to gH/gL re-stimulation than SAM vaccine or subunit/MF59 alone (Figure 12). CD4⁺ responses to RNA alone were insignificant, whereas responses to subunit/MF59 alone were Th2/Th0 phenotype. The phenotype of the responding cells from mice immunized combinations of SAM vaccine and subunit was primarily Th1/Th0. Similar trends were seen when re-stimulating cells with purified pentameric complex, although the responses were generally stronger (Figure 13).

CD8⁺ T cell responses to the vaccinations using purified pentameric subunit or a pool of peptides to gH were also analyzed. The only significant CD8 responses seen when re-stimulating with pentameric subunit was in the mice immunized with two doses of SAM vaccine followed by one dose of subunit/MF59 (Figure 14). Cells from these animals also showed the strongest responses

when re-stimulated with gH peptides (Figure 15). Mice immunized with the SAM vaccine + subunit, with one dose of SAM vaccine followed by two doses of subunit/MF59, or with SAM vaccine alone, also showed significant responses to re-stimulation (Figure 15).

Conclusions: One dose of SAM vaccine followed by two doses of subunit/MF59, as well as SAM vaccine + subunit, elicited higher neutralizing titers than subunit/MF59 alone. The response to SAM vaccine + subunit did not require addition of MF59 adjuvant. The largest impact of SAM vaccine prime subunit/MF59 boost was on cellular immune responses. Any combination including the SAM vaccine produced primarily a Th1/Th0 CD4+ response. Moreover, two immunizations with SAM vaccine followed by one immunization with subunit/MF59 produced the strongest CD8+ responses. This study shows that a SAM vaccine prime protein boost regimen can be optimized to produce a desired immune response, *i.e.* cellular or humoral.

REFERENCES:

- Akter, P, et al. "Two novel spliced genes in human cytomegalovirus." *Journal of General Virology* 84 (2003): 1117-1122.
- Boehme, KW, M Guerrero, and T Compton. "Human cytomegalovirus envelope glycoproteins B and H are necessary for TLR2 activation in permissive cells." *Journal of Immunology* 177 (2006): 7094-7102.
- Doyle, Sharon A, ed. *High Throughput Protein Expression and Purification: Methods and Protocols (Methods in Molecular Biology)*. Humana Press, 2008.
- Genini, E, E Percivalle, A Sarasini, MG Revello, F Baldanti, and G Gerna. "Serum antibody response to the gH/gL/pUL128-131 five-protein complex of human cytomegalovirus (HCMV) in primary and reactivated HCMV infections." *Journal of Clinical Virology* 52 (2011): 113-118.
- Huber, MT, and T Compton. "Intracellular Formation and Processing of the Heterotrimeric gH-gL-gO (gCIII) Glycoprotein Envelope Complex of Human Cytomegalovirus." *Journal of Virology* 73 (1999): 3886-3892.
- Kinzler, Eric R, Regan N Theiler, and Teresa Compton. "Expression and reconstitution of the gH/gL/gO complex of human cytomegalovirus." *Journal of Clinical Virology*, 2002: 87-95.
- Langdon, RH, J Cuccui, and BW. Wren. "N-linked glycosylation in bacteria: an unexpected application." *Future Microbiology* 4 (2009): 401-412.
- Loignon, M, et al. "Stable high volumetric production of glycosylated human recombinant IFN α 2b in HEK293 cells." *BMC Biotechnology* 8 (2008): 65.

- Macagno, A, et al. "Isolation of human monoclonal antibodies that potently neutralize human cytomegalovirus infection by targeting different epitopes on the gH/gL/UL128-131A complex." *Journal of Virology* 84 (2010): 1005-13.
- 5 Melnicka, M, PP Sedghizadehb, CM Allenc, and T Jaskolla. "Human cytomegalovirus and mucoepidermoid carcinoma of salivary glands: Cell-specific localization of active viral and oncogenic signaling proteins is confirmatory of a causal relationship." *Experimental and Molecular Pathology* 92 (2012): 118-125.
- Mocarski, ES, T Shenk, and RF Pass. "Cytomegalovirus." In *Fields Virology*, edited by David M Knipe and Peter M Howley. Philadelphia, Pa., USA: Lippincott Williams and Wilkins, 2006.
- 10 Murthy, S, et al. "Detection of a Single Identical Cytomegalovirus (CMV) Strain in Recently Seroconverted Young Women." *PLoS One* 6 (2011): e15949.
- Needleman, SB, and CD Wunsch. "A general method applicable to the search for similarities in the amino acid sequences of two proteins." *Journal of Molecular Biology* 48 (1970): 443-453.
- Patrone, M, M Secchi, L Fiorina, M Ierardi, G Milanesi, and A Gallina. "Human Cytomegalovirus
15 UL130 Protein Promotes Endothelial Cell Infection through a Producer Cell Modification of the Virion." *Journal of Virology* 79 (2005): 8361-8373.
- Rasmussen, L, A Geissler, C Cowan, A Chase, and M Winters. "The Genes Encoding the gCIII Complex of Human Cytomegalovirus Exist in Highly Diverse Combinations in Clinical Isolates." *Journal of Virology* 76 (2002): 10841-10888.
- 20 Rice, P, I Longden, and A Bleasby. "EMBOSS: The European Molecular Biology Open Software Suite." *Trends Genetics* 16 (2000): 276-277.
- Rigoutsos, I, et al. "In silico pattern-based analysis of the human cytomegalovirus genome." *Journal of Virology* 77 (2003): 4326-44.
- Ryckman, BJ, et al. "Characterization of the Human Cytomegalovirus gH/gL/UL128-131 Complex
25 That Mediates Entry into Epithelial and Endothelial Cells." *Journal of Virology* 82 (2008): 60-70.
- Ryckman, BJ, MC Chase, and DC Johnson. "Human Cytomegalovirus TR Strain Glycoprotein O Acts as a Chaperone Promoting gH/gL Incorporation into Virions but Is Not Present in Virions." *Journal of Virology* 84 (2010): 2597-2609.
- 30 Sambrook, J. *Molecular Cloning: A Laboratory Manual*. 3rd. Cold Spring Harbor Laboratory Press, 2000.
- Simanek, Amanda M., Jennifer Beam Dowd, Graham Pawelec, David Melzer, Ambarish Dutta, and Allison E. Aiello. "Seropositivity to Cytomegalovirus, Inflammation, All-Cause and Cardiovascular Disease-Related Mortality in the United States." *PLoS ONE* 6 ((2011)):
35 e16103.

Summers, MD, and GE Smith. "A manual of methods for baculovirus vectors and insect cell culture procedures." *Texas Agricultural Experiment Station Bulletin No. 1555*, 1987.

Wang, D, and T Shenk. "Human Cytomegalovirus UL131 Open Reading Frame Is Required for Epithelial Cell Tropism. Human Cytomegalovirus UL131 Open Reading Frame Is Required for Epithelial Cell Tropism." *Journal of Virology* 79 (2005): 10330–10338.

5

SEQUENCE LISTING

SEQ ID NO: 1 (gH from HCMV strain Merlin = GI:52139248)

MRPGLPSYLIILAVCLFSLHLLSSRYGAEAVSEPLDKAFHLLLNTYGRPIRFLRENTTQCTYNSSLRNSTVVRENAIS
 5 FNFFQSYNQYYVFHMPRCLFAGPLAEQFLNQVDLTETLERYQQRNLNTYALVSKDLASYRSFSQQLKAQDSLGEQPTT
 VPPPIDLSIPHVMPPQTTPHGWTESHTTSGLHRPHFNQTCILFDGHDLLFSTVTPCLHQGFYLIDELRYVKITLTE
 DFFVVTVSIDDDTPMLLIFGHLPRVLFKAPYQRDNFILRQTEKHELLVLVKKDQLNRHSYKDPDFLDAALDFNYLD
 LSALLRNSFHRYAVDVLKSGRCQMLDRRTVEMAFAYALALFAAARQEEAGAQVSVPRALDRQAALLQIQEFMITCLS
 10 QTPPRTTLLLYPTAVDLAKRALWTPNQITDITSLVRLVYILSKQNQQHLIPQWALRQIADFALKLHKTHLASFLSAF
 ARQELYLMGSLVHSMVLHTTERREIFIVETGLCSLAELSHFTQLLAHPHHEYLSLDYTPCSSSGRRDHSLERLTRLF
 PDATVPATVPAALSILSTMQPSTLETFPDLFCLPLGESFSALTVSEHVSIVTNQYLIKGISYPVSTTVVGQSLIIT
 QTDSQTKCELTRNMHTTHSITVALNISLENCAFCQSALLEYDDTQGVINIMYMHDSDDVLFALDPYNEVVVSSPRTH
 YLMLLKNGTVLEVTDVVVDATDSRLLMMSVYALSIIIGIYLLYRMLKTC

SEQ ID NO: 2 (gH from HCMV strain Towne = GI:138314)

MRPGLPSYLIIVLAVCLLHLLSSRYGAEAISEPLDKAFHLLLNTYGRPIRFLRENTTQCTYNSSLRNSTVVRENAIS
 FNFFQSYNQYYVFHMPRCLFAGPLAEQFLNQVDLTETLERYQQRNLNTYALVSKDLASYRSFSQQLKAQDSLGEQPTT
 VPPPIDLSIPHVMPPQTTPHGWTESHTTSGLHRPHFNQTCILFDGHDLLFSTVTPCLHQGFYLIDELRYVKITLTE
 20 DFFVVTVSIDDDTPMLLIFGHLPRVLFKAPYQRDNFILRQTEKHELLVLVKKDQLNRHSYKDPDFLDAALDFNYLD
 LSALLRNSFHRYAVDVLKSGRCQMLDRRTVEMAFAYALALFAAARQEEAGAQVSVPRALDRQAALLQIQEFMITCLS
 QTPPRTTLLLYPTAVDLAKRALWTPNQITDITSLVRLVYILSKQNQQHLIPQWALRQIADFALKLHKTHLASFLSAF
 ARQELYLMGSLVHSMVLHTTERREIFIVETGLCSLAELSHFTQLLAHPHHEYLSLDYTPCSSSGRRDHSLERLTRLF
 PDATVPPTVPAALSILSTMQPSTLETFPDLFCLPLGESFSALTVSEHVSIVVTNQYLIKGISYPVSTTVVGQSLIIT
 25 QTDSQTKCELTRNMHTTHSITAALNISLENCAFCQSALLEYDDTQGVINIMYMHDSDDVLFALDPYNEVVVSSPRTH
 YLMLLKNGTVLEVTDVVVDATDSRLLMMSVYALSIIIGIYLLYRMLKTC

SEQ ID NO: 3 (gH from HCMV strain AD169 = GI:138313)

MRPGLPPYLTVTFTVYLLSHLPSQRYGADAASEALDPHAFHLLLNTYGRPIRFLRENTTQCTYNSSLRNSTVVRENAI
 SFNFFQSYNQYYVFHMPRCLFAGPLAEQFLNQVDLTETLERYQQRNLNTYALVSKDLASYRSFSQQLKAQDSLGGQPT
 30 TVPPPIDLSIPHVMPPQTTPHWDKGSHTTSGLHRPHFNQTCILFDGHDLLFSTVTPCLHQGFYLMDELRYVKITLT
 EDFFVVTVSIDDDTPMLLIFGHLPRVLFKAPYQRDNFILRQTEKHELLVLVKKQAQLNRHSYKDSDFLDAALDFNYL
 DLSALLRNSFHRYAVDVLKSGRCQMLDRRTVEMAFAYALALFAAARQEEAGTEISIPRALDRQAALLQIQEFMITCL
 SQTPPRTTLLLYPTAVDLAKRALWTPDQITDITSLVRLVYILSKQNQQHLIPQWALRQIADFALQLHKTHLASFLSA
 FARQELYLMGSLVHSMVLHTTERREIFIVETGLCSLAELSHFTQLLAHPHHEYLSLDYTPCSSSGRRDHSLERLTRL
 35 FPDATVPATVPAALSILSTMQPSTLETFPDLFCLPLGESFSALTVSEHVSIVVTNQYLIKGISYPVSTTVVGQSLII
 TQTDSQTKCELTRNMHTTHSITAALNISLENCAFCQSALLEYDDTQGVINIMYMHDSDDVLFALDPYNEVVVSSPRTH
 HYLMLLKNGTVLEVTDVVVDATDSRLLMMSVYALSIIIGIYLLYRMLKTC

SEQ ID NO: 4 (gH protein consisting of amino acid residues 1-715 of SEQ ID NO: 1)

MRPGLPSYLIILAVCLFSLHLLSSRYGAEAVSEPLDKAFHLLLNTYGRPIRFLRENTTQCTYNSSLRNSTVVRENAIS
 FNFFQSYNQYYVFHMPRCLFAGPLAEQFLNQVDLTETLERYQQRNLNTYALVSKDLASYRSFSQQLKAQDSLGEQPTT
 VPPPIDLSIPHVMPPQTTPHGWTESHTTSGLHRPHFNQTCILFDGHDLLFSTVTPCLHQGFYLIDELRYVKITLTE
 45 DFFVVTVSIDDDTPMLLIFGHLPRVLFKAPYQRDNFILRQTEKHELLVLVKKDQLNRHSYKDPDFLDAALDFNYLD
 LSALLRNSFHRYAVDVLKSGRCQMLDRRTVEMAFAYALALFAAARQEEAGAQVSVPRALDRQAALLQIQEFMITCLS
 QTPPRTTLLLYPTAVDLAKRALWTPNQITDITSLVRLVYILSKQNQQHLIPQWALRQIADFALKLHKTHLASFLSAF
 ARQELYLMGSLVHSMVLHTTERREIFIVETGLCSLAELSHFTQLLAHPHHEYLSLDYTPCSSSGRRDHSLERLTRLF
 PDATVPATVPAALSILSTMQPSTLETFPDLFCLPLGESFSALTVSEHVSIVTNQYLIKGISYPVSTTVVGQSLIIT
 50 QTDSQTKCELTRNMHTTHSITVALNISLENCAFCQSALLEYDDTQGVINIMYMHDSDDVLFALDPYNEVVVSSPRTH
 YLMLLKNGTVLEVTDVVVDATD

SEQ ID NO: 5 (C-terminal extension which includes a myc-tag and a polyhistidine-tag)

GTKLGPEQKLISEEDLNSAVDHHHHHH

SEQ ID NO: 6 (gH protein comprising SEQ ID NOs: 4 and 5)

MRPGLPSYLIILAVCLFSLHLLSSRYGAEAVSEPLDKAFHLLLNTYGRPIRFLRENTTQCTYNSSLRNSTVVRENAIS
 FNFFQSYNQYYVFHMPRCLFAGPLAEQFLNQVDLTETLERYQQRNLNTYALVSKDLASYRSFSQQLKAQDSLGEQPTT
 VPPPIDLSIPHVMPPQTTPHGWTESHTTSGLHRPHFNQTCILFDGHDLLFSTVTPCLHQGFYLIDELRYVKITLTE
 60 DFFVVTVSIDDDTPMLLIFGHLPRVLFKAPYQRDNFILRQTEKHELLVLVKKDQLNRHSYKDPDFLDAALDFNYLD

LSALLRNSFHRYAVDVLKSGRCQMLDRRTVEMAFAYALALFAAARQEEAGAQVSVPRALDRQAALLQIQEFMITCLS
 QTPPRTTLLLYPTAVDLAKRALWTPNQITDITSLVRLVYILSKQNQQHLIPQWALRQIADFAKLKHLKTHLASFLSAF
 ARQELYLMGSLVHSMVLHTTERREIFIVETGLCSLAELSHFTQLLAHPHHEYLSDLYTPCSSSGRRDHSLERLTRLF
 PDATVPATVPAALSILSTMQPSTLETFFDLFCLPLGESFSALTVEHVSYIVTNQYLIKGISYPVSTTVVGQSLIIT
 5 QTDSQTKCELTRNMHTTHSITVALNISLENCAFCQSALLEYDDTQGVINIMYMHDSDDVLFALDPYNEVVVSSPRTH
 YLMLLNKGTVEVTDVVVDATDGTKLGPEQKLISEEDLNSAVDHHHHHH

SEQ ID NO: 7 (gL from HCMV strain Merlin = GI:39842115)

MCRRPDCGFSFSPGPVILLWCCLLLPIVSSAAVSVAPTAAEKVPAECPPELTRRCLLGEVFEGDKYESWLRPLVNVTG
 10 RDGPLSQLIRYRPVTPEAANSVLLDEAFDLTLALLYNNPDQLRALLTLLSSDTAPRWMTVMRGYSEC GDGSPAVYTC
 VDDLCRGYDLTRLSYGRSIFTEHVLGFELVPPSLFNVVVAIRNEATRTNRAVRLPVSTAAAPEGITLFYGLYNAVKE
 FCLRHLQDLPPLLRHLDKYYAGLPPELKQTRVNLPASRYGPQAVDAR

SEQ ID NO: 8 (gL from HCMV strain Towne = GI:239909463)

MCRRPDCGFSFSPGPVALLWCCLLLPIVSSATVSVAPTVAEKVPAECPPELTRRCLLGEVFQGD KYESWLRPLVNVTR
 15 RDGPLSQLIRYRPVTPEAANSVLLDDAFDLTLALLYNNPDQLRALLTLLSSDTAPRWMTVMRGYSEC GDGSPAVYTC
 VDDLCRGYDLTRLSYGRSIFTEHVLGFELVPPSLFNVVVAIRNEATRTNRAVRLPVSTAAAPEGITLFYGLYNAVKE
 FCLRHLQDLPPLLRHLDKYYAGLPPELKQTRVNLPASRYGPQAVDAR

SEQ ID NO: 9 (gL from HCMV strain AD169 = GI:2506510)

MCRRPDCGFSFSPGPVLLWCCLLLPIVSSVAVSVAPTAAEKVPAECPPELTRRCLLGEVFQGD KYESWLRPLVNVTR
 20 RDGPLSQLIRYRPVTPEAANSVLLDDAFDLTLALLYNNPDQLRALLTLLSSDTAPRWMTVMRGYSEC GDGSPAVYTC
 VDDLCRGYDLTRLSYGRSIFTEHVLGFELVPPSLFNVVVAIRNEATRTNRAVRLPVSTAAAPEGITLFYGLYNAVKE
 FCLRHLQDLPPLLRHLDKYYAGLPPELKQTRVNLPASRYGPQAVDAR

SEQ ID NO: 10 (gO from HCMV strain Merlin = GI:39842082)

MGKKEMIMVKGIPKIMLLISITFLLLSLINCNVLVNSRGTTRSWPYTVLSYRGKEILKKQKEDILKRLMSTSSDGYR
 25 FLMPYPSQQKFHAIVISMDFKFPQDYILAGPIRNDSTHMFDFYSTQLRKPAKYVYSEYNHTAHKITLRPPPCGTVP
 MNCLSEMLNVSKRNDTGEKGCNFTTFNPMFFNVPRWNTKLYIGSNKVNVDSTIYFLGLTALLRYAQRNCTRSFY
 30 LVNAMSRLNFRVPKYINGTKLKNTMRKLKRKQALVKEQPQKKNKKSQSTTTPYLSYTTSTAFNVTTNVYSATAAVT
 RVATSTTGYPDSNFMKSIMATQLRDLATWVYTTLRYNPFCKPDRNRTAVSEFMKNTHVLIRNETPYTIYGTLD
 SSLYNETMSVENETASDNNETTPTSPSTRFQRTFIDPLWDYLDLSLLFLDKIRNFSQLPAYGNLTPPEHRAANLS
 TLNSLWWSQ

SEQ ID NO: 11 (gO from HCMV strain AD169 = GI:136968)

MGRKEMMVRDVPKMVFLISISFLLVSFINCKVMSKALYNRPWRGLVLSKIGKYKLDQLKLEILRQLETTISTKYNVS
 35 KQPVKNLTMMNTEFPQYYILAGPIQNYISITYLWDFDYSTQLRKPAKYVYSQYNHTAKTITFRPPPCGTVP
 MLNVSKRNDTGEQGCNFTTFNPMFFNVPRWNTKLYVGPTKVNVDSTIYFLGLTALLRYAQRNCTHSFYLVNAMS
 RNLNFRVPKYINGTKLKNTMRKLKRKQAPVKEQFEKKAKKTQSTTTPYFSYTTSAALNVTTNVYSITTAARRVSTST
 40 IAYRPDSSFMKSIMATQLRDLATWVYTTLRYNPFCEPSRNRTAVSEFMKNTHVLIRNETPYTIYGTLD
 ETMFVENKTASDSNKTTPSPSMGFQRTFIDPLWDYLDLSLLFLDEIRNFSLSRSPTYVNLTPPEHRAVNLSLNSLW
 WWLQ

SEQ ID NO: 12 (gO from HCMV strain Towne = GI:239909431)

MGRKGEMRGVFNLFLLMSLTFLFSFINCKIAVARFRVKSQAKEEERQLKLRILQELASKTGDIYKFFTFPSQQKL
 45 YNITVEMKQFPNSILAGPIRNHSITHLWDFHTTQLRKPAKYVYSEYNHTGQKITFRPPSCGTIPSMTCLEMLNV
 SRRNNTGEENCNFTTFNPMFFNVPRWNTKLYVGPSKVNVDSTIYFLGLAALLRYAQRNCTRSFYLVNAMSRLN
 RVPKYINSTKLKNTMRKLKRKQAPVKSISKKS RVSTTTPYSSYTSTIFNVSTNVYSPIVPTRIPTSTIGYRPDENF
 50 MKSILTTQLKDLATWVYTTLRYNPFCKPNRNRTAVSEFMKNTHVLIRNETPYTIYGTLD
 ASDNNKTTPSPSTRFQRTFIDPMWDYLDLSLLFLSEIRNFSLSQSSTYGNLTPPEHRAVNLSLNSLW
 WWLQ

SEQ ID NO: 13 (pUL128 from HCMV strain Merlin = GI:39842124)

MSPKDLTPFLTALWLLGHRSRVPRVRAEECCFEFINVNHPPERCYDFKMCNRFTVALRCPDGEVCYSPEKTAEIRGIV
 55 TTMTSHLTRQVVHNKLTSCNYNPLYLEADGRIRCGKVNDKAQYLLGAAGSVPY

SEQ ID NO: 14 (pUL128 from HCMV strain Towne = GI:39841882)

MSPKNLTPFLTALWLLGHRSRVPRVRAEECCFEFINVNHPPERCYDFKMCNRFTVALRCPDGEVCYSPEKTAEIRGIV
 60 TTMTSHLTRQVVHNKLTSCNYNPLYLEADGRIRCGKVNDKAQYLLGAAGSVPYRWINLEYDKITRIVGLDQYLESVK
 KHKRLDVCRAKMGYMLQ

SEQ ID NO: 15 (pUL128 from HCMV strain AD169 = GI:59803078)

MSPKDLTPFLTTLLWLLLGHSRVPRVRAEECCFEINVNHPPERCYDFKMCNRFTVALRCPDGEVCYSPEKTAEIRGIV
 TTMTHSLTRQVVHNKLTSCNYNPLYLEADGRIRCCKVNDKAQYLLGAAGSVPRWINLEYDKITRIVGLDQYLESVK
 KHKRLDVCRAKMGYMLQ

SEQ ID NO: 16 (pUL130 from HCMV strain Merlin = GI:39842125)

MLRLLLLRHFFHCLLLCAVWATPCLASPWSTLTANQNPSPPWSKLTYSKPHDAATFYCPFLYSPPPRSPLQFSGFQRV
 STGPECRNETLYLLYNREGQTLVERSSTWVKKVIWYLSGRNQITILQRMPTASKPSDGNVQISVEDAKIFGAHMVPK
 QTKLLRFVVNDGTRYQMCVMKLESWAHVFRDYSVSFQVRLTFTEANNQTYTFCTHPNLIV

SEQ ID NO: 17 (pUL130 from HCMV strain Towne = GI:239909473)

MLRLLLLRHFFHCLLLCAVWATPCLASPWSTLTANQNPSPPWSKLTYSKPHDAATFYCPFLYSPPPRSPLQFSGFQRV
 LTGPECRNETLYLLYNREGQTLVERSSTWVKKVIWYLSGRNQITILQRMPTASKPSDGNVQISVEDAKIFGAHMVPK
 QTKLLRFVVNDGTRYQMCVMKLESWAHVFRDYSVSFQVRLTFTEANNQTFPSAPIPISSFEFVARAGNFENRAS

SEQ ID NO: 18 (pUL131A from HCMV strain Merlin = GI:39842126)

MRLCRVWLSVCLCAVVLGQCQRETAEKNDYYRVPHYWDACSRALPDQTRYKYVEQLVDLTNLNYHDASHGLDNFDVL
 KRINTEVSLLISDFRRQNRRGGTNKRTTFNAAGSLAPHARSLEFSVRLFAN

SEQ ID NO: 19 (pUL131A from HCMV strain Towne = GI:239909474)

MRLCRVWLSVCLCAVVLGQCQRETAEKNDYYRVPHYWDACSRALPDQTRYKYVEQLVDLTNLNYHDASHGLDNFDVL
 KRINTEVSLLISDFRRQNRRGGTNKRTTFNAAGSLAPHARSLEFSVRLFAN

SEQ ID NO: 20 (pUL131A from HCMV strain AD169 = GI:219879712)

MRLCRVWLSVCLCAVVLGQCQRETAEEKRLLPSTALLGRVLSRAARNPLQVCGTARGPHVELPLRCEPRLGQL

SEQ ID NO: 21 (gB from HCMV strain Merlin = GI:39842076)

MESRIWCLVVCVNLICIVCLGAAVSSSSSTRGTSATHSHSSHTTSAAHRSRSGSVSQRTVSSHGVNETIYNTTLK
 YGDVVGVTNTKYPYRVCSMAQGTDLIRFERNIVCTSMKPINEDLDEGIMVYKRNIVAHTFKVRVYQKVLTFRRSYA
 YIHTTYLLGSNTEYVAPPMWEIHHINSHSQCYSSYSRVIACTVGFVAYHRDSYENKTMQLMPDDYSNTHSTRYVTVKD
 QWHSRGSTWLYRETCLNLCMVTITARSKYPIHFFATSTGDVVDISPFFYNGTNRNASYFGENADKFFIFPNYITIVSD
 FGRPN SALETHRLVAFLERADSVISWDIQDEKNVTCQLTFWEASERTIRSEAEDSYHFSSAKMTATFLSKKQEVNMS
 DSALDCVRDEAINKLQQIFNTSYNQTYEKYGNVSVFETTGGVLVFWQGIKQKSLVELERLANRSSNLTHNRTKRST
 DGNNATHLSNMESVHNLVYAQLQFTYDTRLGYINRALAQIAEAWCVDQRRTELVFKELSKINPSAILSIAIYNKPIAA
 RFMGDVLGLASCVTINQTSVKVLRDMNVKESPGRCYSRPVVIFFNFANSSYVQYQGLGEDNEILLGNHRTEECQLPSL
 KIFIAGNSAYEYVDYLFKRMIDLSSISTVDSMIALDIDPLENTDFRVLELYSQKELRSSNVFDLEEIMREFNSYKQR
 VKYVEDKVVDPLPPYLKGLDDLMSGLGAAGKAVGVAIGAVGGAVASVVEGVATFLKNPFGAFTIILVAIAVVIITYL
 IYTRQRRCTQPLQNLFPYLVADGTTVTSGSTKDTSLQAPPSYEEVSYNRGRKGPSPSSDASTAAPPYTNEQAYQ
 MLLALARLDAEQRAQQNGTDSL DGRGTGTQDKGQKPNLLDRLRHRKNGYRHLKDSDEEENV

SEQ ID NO: 22 (gB from HCMV strain Towne = GI:138193)

MESRIWCLVVCVNLICIVCLGAAVSSSSSTRGTSATHSHSSHTTSAAHRSRSGSVSQRTVSSHGVNETIYNTTLK
 YGDVVGVTNTKYPYRVCSMAQGTDLIRFERNIVCTSMKPINEDLDEGIMVYKRNIVAHTFKVRVYQKVLTFRRSYA
 YIHTTYLLGSNTEYVAPPMWEIHHINSHSQCYSSYSRVIACTVGFVAYHRDSYENKTMQLMPDDYSNTHSTRYVTVKD
 QWHSRGSTWLYRETCLNLCMVTITARSKYPIHFFATSTGDVVDISPFFYNGTNRNASYFGENADKFFIFPNYITIVSD
 FGRPN SALETHRLVAFLERADSVISWDIQDEKNVTCQLTFWEASERTIRSEAEDSYHFSSAKMTATFLSKKQEVNMS
 DSALDCVRDEAINKLQQIFNTSYNQTYEKYGNVSVFETTGGVLVFWQGIKQKSLVELERLANRSSNLTHNRTKRST
 DGNNATHLSNMESVHNLVYAQLQFTYDTRLGYINRALAQIAEAWCVDQRRTELVFKELSKINPSAILSIAIYNKPIAA
 RFMGDVLGLASCVTINQTSVKVLRDMNVKESPGRCYSRPVVIFFNFANSSYVQYQGLGEDNEILLGNHRTEECQLPSL
 KIFIAGNSAYEYVDYLFKRMIDLSSISTVDSMIALDIDPLENTDFRVLELYSQKELRSSNVFDLEEIMREFNSYKQR
 VKYVEDKVVDPLPPYLKGLDDLMSGLGAAGKAVGVAIGAVGGAVASVVEGVATFLKNPFGAFTIILVAIAVVIITYL
 IYTRQRRCTQPLQNLFPYLVADGTTVTSGSTKDTSLQAPPSYEEVSYNRGRKGPSPSSDASTAAPPYTNEQAYQ
 MLLALVRLDAEQRAQQNGTDSL DGTGTGTQDKGQKPNLLDRLRHRKNGYRHLKDSDEEENV

SEQ ID NO: 23 (gB from HCMV strain AD169 = GI:138192)

MESRIWCLVVCVNLICIVCLGAAVSSSSSTSHATSSHTNGSHTSRTTSAQTRSVYSQHVTSSSEAVSHRANETIYNTTLK
 YGDVVGVTNTKYPYRVCSMAQGTDLIRFERNI ICTSMKPINEDLDEGIMVYKRNIVAHTFKVRVYQKVLTFRRSYA
 YIYTTYLLGSNTEYVAPPMWEIHHINKFAQCYSSYSRVIGTVGFVAYHRDSYENKTMQLIPDDYSNTHSTRYVTVKD
 QWHSRGSTWLYRETCLNLCMLTITARSKYPIHFFATSTGDVVYISPFFYNGTNRNASYFGENADKFFIFPNYITIVSD

FGRPNAAPETHRLVAFLEADSVISWDIQDEKNVTCQLTFWEASERTIRSEAEDSYHFSSAKMTATFLSKKQEVNMS
 DSALDCVRDEAINKLQQIFNTSYNQTYEKEYGNVSFETSGGLVFWQGIKQKSLVELERLANRSSLNITHRRRSTS
 DNNTTHLSSMESVHNLVYAQLQFTYDTRLGYINRALAQIAEAWCVDQRRTELVFKELSKINPSAILSIAINYKPIAAR
 FMGDVLGLASCVTINQTSVKVLRDMNVKESPGRCYSRPVVI FNFANSSYVQYQQLGEDNEILLGNHRTEECQLPSLK
 5 IFIAGNSAYEYVDYLFKRMIDLSSI STVDSMIALDIDPLENTDFRVLELYSQKELRSSNVFDLEEIMREFNSYKQRV
 KYVEDKVVDPLPPYLLKGLDDLM SGLGAAGKAVGVAIGAVGGAVASVVEGVATFLKNPFGAFTIILVAIAVVIITYLI
 YTRQRRRLCTQPLQNLFPYLV SADGTTVTSGSTKDTSLQAPPSYEE SVYNSGRKGPGPPSSDASTAAPPYTNEQAYQM
 LLALARLDAEQRAQQNGTDSL DGQTGTQDKGQKPNLLDRLRHRKNGYRHLKDSDEEENV

SEQ ID NO: 24 (a construct encoding gH(ecto) fused to a C-terminal myc-(His)6 tag)

gacggatcgggagatctcccgatcccctatggtgcactctcagtacaatctgctctgatgccgcatagttaagccag
 tatctgctccctgcttgtgtgtgttgaggctcgtgagtagtgccgcgagcaaaatttaagctacaacaaggcaaggctt
 gaccgacaattgcatgaagaatctgcttagggtaggcgttttgcgctgcttcgcatgtacgggccagatatacgc
 15 gttgacattgattattgactagttattaatagtaaatcaattacgggggtcattagttcatagcccatatatggagttc
 cgcggttacataacttacggtaaatggccgcctggtgacggcccaacgacccccgccattgacgtcaataatgac
 gtatgttcccatagtaacgccaatagggaactttccattgacgtcaatgggtggagattttacggtaaaactgcccact
 tggcagtacatcaagtgtatcatatgccaaagtacgccccctattgacgtcaatgacggtaaatggccgcctggcat
 tatgccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcacgtattaccatgggt
 20 gatgcgggttttggcagtacatcaatggcggtggatagcgggtttgactcacggggattttccaaagtctccacccccattg
 acgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacg
 caaatgggcggttaggcgtgtacgggtgggaggtctatataagcagagctctctggctaactagagaacccactgctta
 ctggcttatcgaaattaatacagactcactatagggagacccaagctggctagcgccaccatgaggcctggcctgccc
 25 tctacctgatcatcctggcgtgtgctgttcagccacctgctgtccagcagatacggcgccgaggccgtgagcga
 gccctggacaaggctttccacctgctgctgaacacctacggcagacccatccggtttctgcgggagaaacaccaccc
 agtgcacctacaacagcagcctgcggaacagcacgctcgtgagagagaacgccatcagcttcaactttttccagagc
 tacaaccagtactacgtgttccacatgccagatgcctgtttgcccggccctctggccgagcagttcctgaaccaggt
 ggacctgaccgagacactggaaagataaccagcagcggctgaatacctacgccctgggtgtccaaggacctggccagct
 accggtccttttagccagcagctcaaggctcaggatagcctcggcgagcagcctaccaccggtgccccctcccatcgac
 30 ctgagcatccccacgtgtggatgcctccccagaccacccctcacggctggaccgagagccacaccacctccggcct
 gcacagacccccacttcaaccagacctgcacctgttcgacggccacgacctgctgttttagcaccgtgacccccctgcc
 tgcaccagggtcttctacctgatcgacgagctgagatacgtgaagatcacctgaccgaggtatttcttctgtgtcacc
 gtgtccatcgacgacgacacccccatgctgctgatcttcggccacctgcccagagtgctgttcaaggccccctacca
 35 gcgggacaacttcatcctgcggcagacccagagaagcagcagctgctggtgctggtcaagaaggaccagctgaaccggc
 actcctacctgaaggaccccgacttccctggacgcgcctggacttcaactacctggacctgagccgctgctgtgaga
 aacagcttccacagatacgcctggacgtgctgaagtccggacgggtgccagatgctcgatcggcgggacctgtgagat
 ggcccttcgcctatgccctgcctgttgcgcgtgccagacaggaagaggctggcgcccagggtgtcagtgcccagag
 ccctggatagacaggccgcctgctgcagatccaggaattcatgatcacctgcctgagccagacccccctagaacc
 accctgctgctgtacccccacagccgtggatctggccaagaggccctgtggaccccccaaccagatcacccgacatcac
 40 aagcctcgtgcggctcgtgtacatcctgagcaagcagaaccagcagcactgatccccagtgggccctgagacaga
 tcgcccacttcgcccctgaagctgcacaagacccatctggccagctttctgagcgcttcgcccaggcaggaactgtac
 ctgatgggcagcctggtccacagcatgctggtgcataaccacggagcggcgaggagatcttcatcgtggagacaggcct
 gtgtagcctggccgagctgtccactttaccagctgctggccaccctcaccacgagtagctgagcgacctgtaca
 45 cccccctgcagcagcagcggcagacgggaccacagcctggaacggctgaccagactgttccccgatgccaccgtgcct
 gctacagtgctgcccgcctgtccatcctgtccaccatgcagcccagcaccctggaaaccttccccgacctgttctg
 cctgccccctggcgagagcttttagcgccctgacctgtccgagcagctgtcctacatcgtgaccaatcagtacctga
 tcaagggcacatcagctaccccgctgtccaccacagctcgtgggcccagagcctgatcatccccagaccgacagccagacc
 aagtgcgagctgaccgggaacatgcacaccacacacagcatcacctggccctgaacatcagcctggaaaactgcgc
 50 tttctgtcagctctgcctgctggaatacgaagatacccaggcggtgatcaacatcatgtacatgcacgacagcgacg
 acgtgctgttcgcccctggacccctacaacgaggtggtggtgtccagcccccgacccactacctgatgctgctgaag
 aacggcaccgctgctggaagtgaccgacgtggtggtggacgccaccgacgggtaccaagcttgggcccgaacaaaaact
 catctcagaagaggatctgaatagcgccgctgacccatcatcatcatcattgagtttaaaggctctccagcttaa
 gtttaaaccgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctcccccgctgccttcc
 55 ttgacctggaagggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcatgtctgagtaggtg
 tcatttctattctggggggtgggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctgggg
 atgcggtgggctctatggcttctgaggcggaagaaccagctggggctctaggggtatccccacgcgcccctgtagc
 ggcgcattaaagcgggcggggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgtcc
 60 tttcgcttttctcccttcccttctctcgccacgttcgcccgttttccccgtcaagctctaaatcgggggctccctttag
 ggttccgatttagtgctttacggcacctcgacccccaaaaaacttgattaggggtgatggttcacgtagtgggccatcg
 ccctgatagacgggttttctgccttttgacgttggagtcacgttctttaatagtgagctcttgttccaaactggaac
 aacactcaaccctatctcggctctattcttttgatttataagggattttgcccgatctcggcctattgggttaaaaaatg

agctgattttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttaggggtgtggaaaagtccccaggct
 cccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccagggtgtggaaaagtccccaggctcccca
 gcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcgccgccctaaactccgccccatcccgcc
 cctaactccgcccagttccgcccattctccgccccatggctgactaattttttttatattatgcagaggccgaggccg
 5 cctctgcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaaagctccccggga
 gcttgatatatccattttcggatctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgca
 cgcagggttctccggccgcttgggtggagaggctattcggctatgactgggcacacagacaatcggtgctctgatg
 ccgccgtgttccggctgtcagcgcagggggcgcccggttcttttgtcaagaccgacctgtccggtgccctgaatgaa
 ctgcaggacgaggcagcgcggctatcggtggtggccacgacgggctccttgcgcagctgtgctcgacgttgtcac
 10 tgaagcgggaagggactggctgctattgggcgaagtgcggggcaggatctcctgtcatctcaccttgctcctgccc
 agaaagtatccatcatggctgatgcaatgcggcggtgcatacgttgatccggctacctgcccattcgaccaccaa
 gcgaaacatcgcatcgagcgcagcagctactcggatggaagccggtcttgcgatcaggatgatctggacgaagagca
 tcaggggctcgcgccagccgaactgttcgccagggtcaaggcgcgcatgcccgcaggcgaggatctcgctcgtagccc
 atggcgatgcttgccttgcgaatatcatgggtggaaaatggccgcttttctggattcatcgactgtggccggctgggt
 15 gtggcggaccgctatcaggacatagcgttgggtaccgctgatattgctgaagagcttggcgcggaatgggctgaccg
 ctctcctcgctgtttacgggtatcgccgctcccgattcgcagcgcacgccttctatcgcttcttgacgagttcttct
 gagcgggactctgggggttcgaaatgaccgaccaagcgacgcccacctgccatcacgagatttcgattccaccgccc
 ccttctatgaaagggttgggcttcggaatcggttttccgggacgcccggctggatgatcctccagcggtggatctcatg
 ctggagttcttgcgccaccccaacttgtttattgcagcttataatgggttacaataaagcaatagcatcacaaattt
 20 cacaaataaagcatttttttactgcattctagttgtgggttgtccaaactcatcaatgtatcttatcatgtctgta
 taccgtcgacctctagctagagcttggcgtaatacatgggtcatagctgttctcgtgtgaaattgttatccgctcaca
 attccacacacatacgcagccggaagcataaagtgtaaagcctggggtgcctaataagtgagctaaactcacattaat
 tgcgttgcgctcactgcccgtttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaaacgcgcgg
 ggagaggcggttggcgatttggcgctcttccgcttctcgtcactgactcgctgcgctcggtcggtcggctgcgg
 25 cgagcggatcatcagctcactcaaaggcggttaatacgggttatccacagaatcaggggataacgcaggaaagaacatgtg
 agcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcttgcgtggcggttttccataggctccgccccctg
 acgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaaccgcagagactataaagataccaggcggttccc
 cctggaagctccctcgctgcgctctcctgttccgacctgcccgttaccgggataacctgtccgcctttctcccttcggg
 aagcgtggcgctttctcatagctcacgctgtaggtatctcagttcgggtgtaggtcggtcgcctccaagctgggctgtg
 30 tgcacgaaccccccggttcagcccgaccgctgcgccttatccggtaactatcgctttaggtccaacccggtaagacac
 gacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttctt
 gaagtgggtggcctaactacggctacactagaagaacagttatttgggtatctgcgctctgctgaagccagttaccttcg
 gaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtgggttttttggtttgcaagcagcag
 attacgcgcagaaaaaaaggatctcaagaagatccttggatcttttctacggggtctgacgctcagtggaacgaaaa
 35 ctcacgttaagggttttgggtcatgagattatcaaaaaggatcttcacctagatccttttaattaaaaaatgaagtt
 ttaaatcaatctaaagtatatatgagtaaacttgggtctgacagttaccaatgcttaatcagtgaggcacctatctca
 gcgatctgtctattttcggttcattccatagttgcctgactccccgctgctgtagataactacgatacgggagggttacc
 atctggccccagtgctgcaatgataccgcgagacccacgctcacccggtccagatttatcagcaataaaccagccag
 ccggaaggggccgagcgcagaagtgggtcctgcaactttatccgcctccatccagttctattaattgttgccgggaagct
 40 agagtaagtagttcggcagtttaatagtttgcgcaacggttgggtgacattgctacagggcatcggtgtgtcacgctcgtc
 gtttgggtatggcttcattcagctccgggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaag
 cggttagctccttcggctcctccgatcggttgtcagaagtaagttggccgcagtggttatcactcatggttatggcagca
 ctgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactgggtgagtactcaaccaagtcatctctg
 agaatagtgatgctggcgacccaggttgccttgcggcgctcaatacgggataataccgcgccacatagcagaactt
 45 taaaagtgtcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcg
 atgtaaccctcgtgcacccaactgatcttcagcatcttttactttcaccagcggttctgggtgagcaaaaaacagg
 aaggcaaaatgcccgaaaaaagggaataaggcgacacggaaatgttgaaactcatactcttctttttcaatatt
 attgaagcatttatcagggttattgtctcatgagcgggatacatatttgaaatgtatttagaaaaataacaaataggg
 gttccgcgcacattttccccgaaaagtgcacactgacgtc

SEQ ID NO: 25 (a construct encoding full-length gL)

gccgcggaatttcgactctaggccattgcatacgttggtatctatatcataatatgtacatttatattgggtcatgtc
 caatatgaccgcatgttgacattgattattgactagttattaatagtaatacaattacggggtcattagttcatagc
 ccatatatggagttccgcttacataacttacggtaaatggcccgctgggtgaccgcccacgacccccgcccatt
 55 gacgtcaataatgacgtatgttcccatagtaacgcgaatagggactttccattgacgtcaatgggtggagttattac
 ggtaaaactgccacttggcagttacatcaagtgtatcatatgccaagtccgccccctattgacgtcaatgacggtaaa
 tggcccgcttggcattatgccagttacatgaccttacgggactttcctacttggcagttacatctacgtattagtcac
 cgctattaccatgggtgatgcgggttttggcagttacaccaatggcggtggatagcggttgactcacggggatttccaa
 gtctccacccattgacgtcaatgggagtttgggttggcaccaaaatcaacgggactttccaaaatgtcgtaataac
 60 cccgccccgttgacgcaaatggcggttaggcgtgtacgggtgggaggtctatataagcagagctcggttagtgaaaccg
 tcagatcgctggagacgccatccacgctgttttgacctccatagaagacaccgggaccgatccagcctccgcgcc

gggaaacggtgcattggaacgcggattccccgtgccaagagtgcgtaagtaccgcctatagactctataggcacacc
 cctttggctcttatgcatgctatactgtttttggcttggggcctatacacccccgcttcccttatgctataggatgatg
 gtatagcttagcctatagggtgtgggttattgaccattattgaccactcccctattgggtgacgatactttccattact
 aatccataacatggctcttttgccacaactatctctattggctatatgccaatactctgtccttcagagactgacacg
 5 gactctgtatttttacaggatgggggtcccatttattattttacaaattcacatatacaacaacgcggtcccccggtgcc
 cgcagttttttatataacatagcgtggggtatctccacgcgaatctcgggtacgtgttcgggacatggggtcttctccgg
 tagcggcgagacttccacatccgagccctgggtcccatgcctccagcgggtcatgggtcgctcggcagctccttgctcc
 taacagtggaggccagacttaggcacagcacaatgccaccaccaccagtggtgcccgcacaaggccgtggcggtaggg
 tatgtgtctgaaaatgagctcggagattgggtcgcaccgctgacgcagatggaagacttaaggcagcggcagaaga
 10 agatgcaggcagctgagttgttgtattctgataagagtcagaggttaactcccgttgcgggtgctgttaacgggtggagg
 gcagtgtagtctgagcagtaactcgttgcgtgccgcgcgcgccaccagacataatagctgacagactaacagactgttc
 ctttccatgggtcttttctgcagtcacccgtcgtcgcagcgcaccatgtgcagaaggcccgactgcggcttcagcttca
 gccctggaccgcgtgatcctgctgtggtgctgcctgctgctgcctatcgtgtcctctgccgcggtgtctgtggccct
 acagccgcgcgagaaggtgccagccgagtgccccgagctgaccagaagatgcctgctgggcgaggtgttcgaggggcga
 15 caagtacgagagctgggtgccccctgggtcaacgtgaccggcagagatggccccctgagccagctgatccggtaca
 gaccgctgacccccgaggccgccaatagcgtgctgctggacgagggccttccctggataccctggccctgctgtacaac
 aaccccgaccagctgagagccctgctgacccctgctgtccagcgcacccgccccagatggatgaccgtgatgcgggg
 ctacagtcgagtggtggagatggcagccctgccgtgtacacctgctggagcgcctgtgcagagggctacgacctgacca
 gactgagctacggccggtccatcttcacagagcagctgctgggtctcgagctgggtgccccccagcctgttcaacgtg
 20 gtgggtggccatccggaacgaggccaccagaaccaacagagccgtgcgggtgcctgtgtctacagccgctgcacctga
 gggcatcacactgttctacggcctgtacaacgcgctgaaagagttctgcctccggcaccagctggatccccccctgc
 tgagacacctggacaagtaactacgcggcctgccccagagctgaagcagaccagagtgaaacctgcccgcacagc
 agatatggccctcaggccgtggagccagatgataatctagaaagccatggatccgactacgcgttagagc
 tcgctgatcagcctcgactgtgccttctagtgtccagccatctgttgtttgcccctcccccgctgccttccctgaccc
 25 tggaaggtgccactcccactgtcctttcctaataaaaatgaggaaattgcacgcattgtctgagtaggtgtcattct
 attctgggggggtgggggtggggcaggacagcaagggggaggattgggaagacaatagcaggggggtgggcgaagaact
 ccagcatgagatccccgcgctggaggatcatccagccggcgtcccggaaaacgattccgaagcccaacctttcatag
 aaggcggcggtggaatcgaaatctcgtgatggcaggttgggcgctcgcttgggtcgggtcatttcgaaccccagagtccc
 gctcagaagaactcgtcaagaaggcgatagaaggcgatgcgctgcgaatcgggagcggcgataaccgtaaagcacgag
 30 gaagcggtcagcccattcgcgcgcaagctcttcagcaatatcacgggtagccaacgctatgtcctgatagcgggtccg
 ccacaccagcgggcccacagtcgatgaatccagaaaagcggccattttccaccatgatattcggcaagcaggcatcg
 ccatgggtcacgacgagatcctcgcgctcgggcatgcgcgccttgagcctggcgaacagttcgggtggcgcgagccc
 ctgatgtctcttcgtccagatcatcctgatcgacaagaccggcttccatccgagtagctgctcgtcgtcgtatggt
 tcgcttgggtggatcgaaatgggcaggttagccggatcaagcgtatgcagccgcgcgattgcacagccatgatggatact
 35 ttctcggcaggagcaaggtgagatgacaggagatcctgccccggcacttcgcccataagcagccagtccttcccg
 ttcatgtgacaacgtcgagcacagctgcgcaaggaaacgcgcgctcgtggccagccacgatagccgcgctgcctcgtcct
 gcagttcattcagggcaccggacaggtcgggtcttgacaaaaagaacccggcgccccctgcgctgacagccggaaacacg
 gcggcatcagagcagccgattgtctgttgtgcccagtcatagccgaatagcctctccacccaagcggccgggagaacc
 tgctgcaatccatcttgttcaatcatgcgaaacgatcctcatcctgtctcttgatcagatcttgatccccctgcgcc
 40 atcagatccttggcggcaagaaagccatccagtttactttgcagggttcccaaccttaccagagggcgccccagct
 ggcaattccgggttcgcttgcgtgtccataaaaacgcgccagcttagctatcgccatgtaagccactgcaagctacctg
 ctttctctttgcgcttgcgttttcccttgcagatagcccagtagctgacattcatccgggggtcagcaccgtttct
 gcggactggctttctacgtgttccgcttccctttagcagcccttgccgctgagtgcttgcggcagcgtgaagctaatt
 tcatggttaaatttttgttaaatacagctcattttttaaccaataggccgaaatcggcaaaatcccttataaatcaaa
 45 agaatagcccagataggggttagtggttgcagtttggaaacaagagtcactattaaagaacgtggactccaacg
 tcaaagggcgaaaaaacgctctatcaggggcgatggccggatcagcttatgcgggtgtgaaataaccgcacagatgcgtaa
 ggagaaaaataccgcatcaggcgtcttccgcttccctcgtcactgactcgtgcgctcgggtcgttcgggtgcggcga
 gcgggtatcagctcactcaaaggcggtaatacgggttatccacagaatcaggggataacgcaggaaagaacatgtgagc
 aaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgcgtggcggtttttccataggctccgccccctgacg
 50 agcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataaccaggcgtttccccct
 ggaagctccctcgtgcgctctcctgttccgacctgcccgttaccggataacctgtccgcttctccttccgttgcggaag
 cgtggcgctttctcatagctcacgctgtaggtatctcagttcgggtgtaggtcgcttcgctccaagctgggctgtgtgc
 acgaacccccgcttcagcccagccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgac
 ttatcgccactggcagcagccactggtaacaggattagcagagcagaggtatgtaggcgggtgctacagagttcttgaa
 55 gtgggtggcctaactacggctacactagaaggacagtatgttgggtatctgcgctctgctgaagccagttaccttcggaa
 aaagagttggtagctcttgatccggcaaaacaaaccaccgctggtagcgggtggtttttttgtttgcaagcagcagatt
 acgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctactgaacggtgatccccaccggaattgcg

SEQ ID NO: 26 (a construct encoding full-length pUL128)

60 gccgcggaatttcgactctaggccattgcatacgttgtatctatatcataatatgtacatttatattggctcatgtc
 caatatgaccgcatgttgacattgattattgactagttattaatagtaatacaattacgggggtcatttagttcatagc

ccatatatggagttccgcgttacataacttacggtaaatggcccgccctggctgaccgcccacgacccccgcccatt
gacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagattttac
ggtaaaactgccacttggcagtagacatcaagtgtatcatatgccaaagtccgccccctattgacgtcaatgacggtaaa
5 tggcccgccctggcattatgccagtagacatgaccttacgggactttcctacttggcagtagacatctacgtattagtcat
cgctattaccatgggtgatgcggttttggcagtagaccaatgggcgtggatagcggtttgactcacgggggatttccaa
gtctccaccccatgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtataaac
cccgccccgttgacgcaaatgggcgttaggcgtgtacgggtgggaggtctatataagcagagctcgtttagtgaaccg
tcagatcgccctggagacgccatccacgctgttttgacctccatagaagacaccgggaccgatccagcctccgcggcc
10 gggaaacggtgcatgggaacgcggattccccgtgccaaagagtacgtaagtaccgcctatagactctataggcacacc
cctttggctcttatgcatgtatactgttttggcttggggcctatacacccccgcttcccttatgctataggatgatg
gtatagcttagcctatagggtgtgggttattgaccattattgaccactccccctattgggtgacgatactttccattact
aatccataacatggctcttttgccacaactatctctattggctatatgccaaactctgtccttcagagactgacacg
gactctgtatttttacaggatgggggtcccatttattatttacaattcacatatacaacaacgccgtcccccggtgcc
15 cgcagtttttattaaacatagcgtggggtctccacgcgaatctcgggtacgtgttccggacatgggctcttctccgg
tagcggcgaggacttccacatccgagccctgggtcccatgcctccagcggctcatggctcgctcggcagctccttgcctcc
taacagtggaggccagacttaggcacagcacaaatgccaccaccaccagtggtgccgcacaaggccgtggcggtaggg
tatgtgtctgaaaatgagctcgggagattgggctcgcaccgctgacgcagatggaagacttaaggcagcggcagaaga
agatgcaggcagctgagttgttattctgataagagtcagaggttaactccggttgccggtgtgttaacggtggagg
20 gcagtgtagtctgagcagtagctgttgcgtgccgcgcgcgccaccagacataatagctgacagactaacagactgttc
ctttccatgggtcttttctgcagtcaccgctcgtcgcagcgcaccatgagccccaaggacctgaccccccttccctgacaa
ccctgtgggtgtccttgggcatagcagagtgccctagagtgccgggcccagggaatgctgcgagttcatcaacgtgaac
cacccccccgagcgggtgtacgacttcaagatgtgcaaccgggttcaccgctggccctgagatgccccgcagcgcgaagt
gtgctacagccccgagaaaaccgcgcgagatccggggcatcgtgaccaccatgaccacagcctgacccggcaggtgg
25 tgcacaacaagctgaccagctgcaactacaacccccctgtacctggaagccgacggccggatcagatgcggcaaatg
aacgacaaggccagtagctgtgggagccgcgcggaagcgtgccctaccggtggatcaacctggaatacgcacaagat
cacccggtatcgtgggcctggaccagtagctggaagcgtgaagaagcacaagcggctggacgtgtgcagagccaaga
tgggctacatgctgcagtgataatctagaaagccatggatatacggatccactacgcgttagagctcgctgatcagcc
tcgactgtgccttctagttgccagccatctgttgtttgccccctcccccgctgccttccctgaccctggaaggtgccac
30 tcccactgtcctttcctaataaaaatgaggaaaatgcatcgcatgtctgagtaggtgtcattctattctgggggggtg
gggtggggcaggacagcaagggggaggattgggaagacaatagcaggggggtgggcgaagaactccagcatgagatc
cccgcgctggaggatcatccagccggcgtcccggaacacgattccgaagcccaacctttcatagaaggcggcggtgg
aatcgaaatctcgtgatggcaggttgggcgtcgttgggtcgggtcatttgcgaaccccagagtagccgctcagaagaact
cgtcaagaaggcgatagaaggcgatgcgtctgcaatcgggagcggcgataccgtaaaagcagaggaagcggctcagcc
35 cattcgccgccaagctctctcagcaatatcaccgggtagccaacgctatgtcctgatagcggtccgccaacaccagccg
gccacagtcgatgaatccagaaaagcggccattttccaccatgatattcggcaagcagggcatcgccatgggtcacga
cgagatcctcgccgtcgggcatgcgcgccttgagcctggcgaacagttcgggtggcgcgagccctgatgtctctcg
tccagatcatcctgatcgacaagaccggcttccatccgagtagctgctcgctcgatgcgatgtttcgcttgggtggtc
gaatgggcaggtagccggatcaagcgtatgcagccgcgcgcatgcatcagccatgatggatactttctcggcaggag
40 caaggtgagatgacaggagatcctgccccggcacttcgccaatagcagccagtccttccccgcttcagtgacaacg
tcgagcacagctgcgcgaaggaaacgcccgtcgtggccagccacgatagccgcgctgcctcgtcctgcagttcattcag
ggcaccggacaggtcgggtcttgacaaaaagaacggggcgccccctgcgtgacagccggaacacggcggtcatcagagc
agccgattgtctgttgtgcccagtcatagcgaatagcctctccacccaagcggccggagaaacctgcgtgcaatcca
tcttgttcaatcatgcgaacgatcctcatcctgtctcttgatcagatcttgatccccctgcgccaatcagatccttgg
45 cgggaagaaagccatccagtttactttgcagggcttcccaaccttaccagagggcgccccagctggcaattccgggtt
cgcttgcgtgtccataaaaccgcccagctctagctatcgccatgtaagccactgcaagctacctgctttctcttgcg
cttgcgttttcccttgcagatagcccagtagctgacattcatccggggtcagcacccgtttctgcggactggcttt
ctacgtgttccgcttccctttagcagcccttgcgccttgagtgcttgcggcagcgtgaagctaatcatggttaaatt
tttgttaaatacagctcattttttaaccaataggccgaaatcggaacaaatcccttataaatcaaaaagaatagcccag
50 ataggggttagtggttgcagtttgggaacaagagtcactattaaagaacgtggactccaacgtcaaaagggcgaaa
aaccgtctatcagggcgatggccggatcagcttatgcgggtgtgaaataccgcacagatgcgtaaggagaaaaatccg
catcaggcgctcttccgcttccctcgtcactgactcgtcgcgtcgggtcgttcgggtgcggcgagcggatcagctc
actcaaaggcggttaatacgggttatccacagaatcaggggataacgcaggaagaacatgtgagcaaaaaggccagcaa
aaggccaggaaccgtaaaaaggccgcgttgcgtggcggttttccataggctccgccccctgacgagcatcacaaaaa
55 tcgacgctcaagtacagaggtggcgaaaccgcagaggactataaagataaccaggcggtttccccctggaagctccctcg
tgcgctctcctgttccgacctgcgcgttaaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttct
catagctcacgctgtaggtatctcagttcgggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccg
tcagcccgaccgctgcgccttatccggtaactatcgtcttgagttccaacccggtaagacacgacttatcgccactgg
cagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaac
tacggctacactagaaggacagtatttgggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtag
60 ctcttgatccggcaacaaaccaccgctggttagcgggtgggttttttgggttgcaagcagcagattacgcgcagaaaaa
aaggatctcaagaagatcctttgatcttttctactgaacgggtgatccccaccggaattgcg

SEQ ID NO: 27 (a construct encoding full-length pUL130)

gccgcggaatttcgactctaggccattgcatacgttgatatctatatcataatatgtacatttatattgggtcatgtc
caatatgaccgccatgttgacattgattattgactagttattaatagtaatacaattacggggcatttagttcatagc
5 ccatatatggagttccgcgttacataacttacggtaaatggccgcctggctgaccgcccacgacccccgccatt
gacgtcaataatgacgtatgttcccatagtaacgccaatagggaactttccattgacgtcaatgggtggagtatttac
ggtaaaactgccacttggcagtagcatcaagtgtatcatatgccaaagtcgccccctattgacgtcaatgacggtaaa
tggccgcctggcattatgccagtagcatgaccttacgggactttcctacttggcagtagcatctacgtattagtcac
10 cgctattaccatgggtgatgcggttttggcagtagaccaatgggcgtggatagcggtttgactcacggggatttccaa
gtctccacccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaataac
cccgccccgttgacgcaaatgggcgttaggcgtgtacgggtgggaggtctatataagcagagctcggttagtgaaccg
tcagatcgctggagacgccatccacgctgttttgacctccatagaagacaccgggacccgatccagcctccgcggcc
gggaacgggtgcattggaacgcggattccccgtgccaagagtgcgtaagtaccgcctatagactctataggcacacc
15 cctttggctcttatgcatgctatactgttttggccttggggcctatacacccccgcttcccttatgctataggtgatg
gtatagcttagcctataggtgtgggttattgaccattattgaccactcccctattgggtgacgatactttccattact
aatccataacatggctctttggcacaactatctctatttggctatatgccaaactctgtccttcagagactgacacg
gactctgtatttttacaggatgggttccatttatttatttacaattcacatatacaacaacgcgctcccccgtagcc
cgagtttttattaaacatagcgtgggatctccacgcgaatctcggtacgtgttcgggacatgggtcttctccgg
20 tagcggcgagcttccacatccgagccctgggtcccatgcctccagcggtcatggctcgctcggcagctccttgctcc
taacagtggaggccagacttaggcacagcacaaatgccaccaccaccagtggtgcccgcacaaggccgtggcggtagg
tatgtgtctgaaaatgagctcgagattgggtcgcaccgctgacgcagatggaagacttaaggcagcggcagaaga
agatgcaggcagctgagttgttctgataagagtcagaggttaactcccgttgcgggtgctgttaacgggtggagg
gcagtgtagtctgagcagtagctgttgcgtgcgcgcgcgcaccagacataatagctgacagactaacagactgttc
25 ctttccatgggtcttttctgcagtcaccgctcgtagcgcaccatgctgcggctgctgctgagacaccacttccact
gcctgctgctgtgtgcgtgtgggcccacccttgtctggccagcccttggagcaccctgaccgccaaccagaaccct
agcccccttgggtccaagctgacctacagcaagccccacgacgcgcgcacccttctactgccccttctgtaccccag
ccctcccagaagccccctgcagttcagcggcttccagagagtggtccaccggccctgagtgccggaacgagacactgt
acctgctgtacaaccgggagggccagacactggtggagcgggagcagcacctgggtgaaaaaagtgtatctgt
30 agcggccggaaccagaccatcctgcagcggatgcccagaaccgccagcaagcccagcgacggcaacgtgcagatcag
cgtggaggacgcaaaaatcttcggcgcccatggtgcccagcagaccaagctgctgagattcgtggtcaacgacg
gcaccagatatcagatgtgctgatgaagctggaaagctgggcccacgtgttccgggactactccgtgagcttccag
gtccggctgaccttcaccgagggccaacaaccagacctacaccttctgcaccaccccaacctgatcgtgtgataatc
tagaaagccatggatatcggtaccactacgcgttagagctcgctgatcagcctcgactgtgccttctagttgccagc
35 catctgttgtttgccccctccccgtgccttccctgacctggaaggtgccactcccactgtcctttcctaataaaaat
gagaaattgcatcgcatgtctgtagtgggtgcattctattcttgggggtgggtgggagcagcaaggggga
ggattgggaagacaatagcaggggggtggggaagaactccagcatgagatccccgcgctggaggatcatccagccg
gcgtcccgaaaaacgattccgaagcccaacctttcatagaaggcggcggtggaatcgaaatctcgtgatggcaggtt
gggcgtcgcttgggtcggtcatttccgaaccccagagctcccgctcagaagaactcgtcaagaaggcgatagaaggcgat
40 gcgctgcgaatcgggagcggcgataccgtaaagcacgaggaagcggtcagcccatcgcgcgcaagctcttcagcaa
tatcacgggtagccaacgctatgtcctgatagcgggtccgccacaccagccggccacagtcgatgaatccagaaaag
cggccattttccaccatgatattcggcaagcagggcatcgccatgggtcacgacgagatcctcgccgtcgggcagtcg
cgcttgagcctggcgaacagttcggctggcgcgagccctgatgctcttcgtccagatcatcctgatcgacaagac
cggcttccatccgagtagctgctcgctcgatgcgatgtttcgcttgggtggatgggcaggtagccggatcaagc
45 gtatgcagccgcgcattgcatcagccatgatggatactttctcggcaggagcaaggtgagatgacaggagatcctg
ccccggcacttgcgccaatagcagccagtccttcccgttccagtgcacaacgtcgagcacagctgcgcaaggaacgc
cgtcgtggccagccacgatagccgcgtgcctcgtcctgcagttcattcagggcacccggacaggtcggtcttgaca
aaaagaaccgggcgccccctgcgtgacagccggaacacggcggcacatcagagcagccgattgtctgttggccagtc
atagccgaatagcctctccaccaagcggcggaacacgtgcgtgcaatccatcttgttcaatcatgcgaaacgatc
50 ctcatcctgtctcttgatcagatcttgatccccctgcgccatcagatccttggcggaagaaagccatccagtttact
ttgcagggttcccaaccttaccagagggcgccccagctggcaattccgggttcgcttgcgtgtccataaaaccgcca
gtctagctatcgccatgtaagcccactgcaagctacctgcttctcttgcgcttgcgttttcccttgcagatag
cccagtagctgacattcatccgggtcagcaccgcttctcgggactggcttctacgtgttcgcgttccctttagcag
cccttgcgcccgtgagtgcttgcggcagcgtgaagctaattcatgggttaaattttgttaaatcagctcatttttttaa
ccaataggccgaaatcggaacaaatcccttataaatcaaaagaatagcccgagataggggtgagtggttccagttt
55 ggaacaagagtccactattaaagaacgtggactccaacgtcaaaggcgcaaaaaccgtctatcaggggcgatggccgg
atcagcttatgcggtgtgaaataccgcacagatgcgtaaggagaaaaataccgcacagggcgtcttccgcttccctcg
ctcactgactcgctgcgtcggctcgttccggctgcggcgagcggatcagctcactcaaaggcggttaatacggttatc
cacagaatcaggggataacgcaggaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccg
cgttgcgtggcggttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcga
60 aaccgcagaggactataaagataccaggcggttccccctggaagctccctcgtagcgtctcctgttccgaccctgcc
gcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctca

gttcggtgtaggtcggttcgctccaagctgggctgtgtgcacgaaccccccggttcagcccgaccgctgcgcccttatcc
 ggtaactatcgctcttgagtgccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattag
 cagagcgaggtatgtaggcggtgctacagagttcttgaaagtgggtggcctaactacggctacactagaaggacagtat
 ttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcgaacaaaccacc
 5 gctggtagcggtgggtttttttgtttgcaagcagcagattacgcgcgagaaaaaaggatctcaagaagatcctttgat
 cttttctactgaacggtgatccccaccggaattgcg

SEQ ID NO: 28 (a construct encoding full-length pUL131A)

gccgcggaatttcgactctaggccattgcatacgttgtatctatatcataatatgtacatttatattggctcatgtc
 10 caatatgaccgccatgttgacattgattattgactagttattaatagtaatacaattacggggtcattagttcatagc
 ccatatatggagttccgcgttacataacttacggtaaatggccccgcctggctgaccgccccacgacccccgcccatt
 gacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatattac
 ggtaactgcccacttggcagtagcatcaagtgtatcatatgccaaagtccgccccctattgacgtcaatgacggtaaa
 tggccccgcctggcattatgcccagtagcatgaccttacgggactttcctacttggcagtagcatctacgtattagtcac
 15 cgctattaccatgggtgatgcggttttggcagtagaccaatgggctggatagcggtttgactcacggggattttcaa
 gtctccaccccatgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaataac
 cccgccccgttgacgcaaatgggctggtaggcgtgtacgggtgggaggtctatataagcagagctcgttttagtgaaccg
 tcagatcgccctggagacgccatccacgctgttttgacctccatagaagacaccgggaccgatccagcctccgcggcc
 gggaacggtgcattggaacgcggattccccgtgccaagagtgcgtaagtaccgcctatagactctataggcacacc
 20 cctttggctcttatgcatgctatactgttttggcttggggcctatacacccccgccttcccttatgctatagggtgatg
 gtatagcttagcctatagggtgtgggttattgaccattattgaccactcccctattgggtgacgatactttccattact
 aatccataacatggctctttgcccacaactatctctattggctatatgccaaactctgtccttcagagactgacacg
 gactctgtatttttacaggatggggtccattttattatttacaattcacatatacaacaacgcggtcccccggtgcc
 cgcagttttttatataacatagcgtgggatctccacgcgaatctcgggtacgtgttcgggacatgggctcttctccgg
 25 tagcggcgagcttccacatccgagccctgggtcccatgcctccagcggctcatggctcgtcggcagctccttgctcc
 taacagtggaggccagacttaggcacagcacaatgccaccaccaccagtggtgccgcacaaggccgtggcggtaggg
 tatgtgtctgaaaatgagctcggagattgggctcgcaccgctgacgcagatggaagacttaaggcagcggcagaaga
 agatgcaggcagctgagttgttgtattctgataagagtcagaggttaactcccgttgcgggtgctgttaacggtggagg
 gcagtgtagtctgagcagtagctcgttgcgtgcgcgcgcgcgccaccagacataatagctgacagactaacagactgttc
 30 ctttccatgggtcttttctgcagtcaccgctcgtcgcagccaccatgcggctgtgcagagtggtgctgtccgtgtgcc
 tgtgtgcggtggtgctgggcccagtgccagagagagacagccgagaagaacgactactaccgggtgccccactactgg
 gatgctgcagcagagccctgcccagaccagaccgggtacaaatacgtggagcagctcgtggacctgacctgaacta
 ccactacgacgcccagccacggcctggacaacttcgacgtgctgaagcggatcaacgtgaccgaggtgtccctgtcga
 tcagcgacttccggcggcagaacagaagaggcggcaccacaacgcggaccacctcaacgcgcgtggctctctggcc
 35 cctcacgccagatccctggaattcagcgtgcggctgttcgccaactgataatctagaaagccatggatatcggatcc
 actacgcgttagagctcgtgtagcagcctcgactgtgccttctagttgccagccatctgttgtttgccccctcccccg
 tgccttcccttgaccctggaaggtgccactcccactgtcctttccataaaaatgaggaaaattgcatcgcatgtctg
 agtaggtgtcattctattctggggggtgggggtggggcaggacagcaagggggaggattgggaagacaatagcagggg
 ggtggggaagaactccagcatgagatccccgcgctggaggatcatccagccggcgtcccggaacagattccgaag
 40 cccaacctttcatagaaggcggcggtggaatcgaaatctcgtgatggcaggttgggctcgttgggtcggtcatttc
 gaacccagagtcctcgtcagaagaactcgtcaagaaggcgatagaaggcgatgcgctgcgaatcgggagcggcgat
 accgtaagcacgaggaagcggtcagccatttcgcccgaagctcttcagcaatatcacgggtagccaacgctatgt
 cctgatagcgggtccgccacaccagcggccacagtcgatgaatccagaaaagcggccattttccaccatgatattc
 ggcaagcaggcatcgccatgggtcacgacgagatcctcgcgcgtcgggcatgcgcgccttgagcctggcgaacagttc
 45 ggctggcgcgagcccctgatgctcttcgtccagatcatcctgatcgacaagaccggcttccatccgagtagctgctc
 gctcgatgcgatgtttcgttgggtggtcgaatgggcaggtagccggatcaagcgtatgcagccgcccgcattgcatca
 gccatgatggatactttctcggcaggagcaaggtgagatgacaggagatcctgccccggcacttcgcccataagcag
 ccagtcccttcccgttccagtgcacaacgtcagacacagctgcgcaaggaacgcccgtcgtggccagccacgatagcc
 gcgctgcctcgtcctgcagttcattcagggcaccggacaggtcggctcttgacaaaagaaccgggcccctgcgct
 50 gacagccggaacacggcggcatcagagcagccgattgtctgttgtgcccagtcataagccgaatagcctctccacca
 agcggccggagaacctgcgtgcaatccatcttgttcaatcatgcgaaacgatcctcatcctgtctcttgatcagatc
 ttgatccccctgcgcatcagatccttggcggaagaaagccatccagtttactttgcaagggttcccaaccttacc
 gagggcgccccagctggcaattccggttcgcttgcgtgtccataaaaaccgcccagctctagctatcgccatgtaagccc
 actgcaagctacctgttttcttttgcgcttgcgttttcccttgcagatagcccagtagctgacattcatccggg
 55 gtcagcacggtttctgcggaactggctttctacgtgttccgcttcccttagcagcccttgcgcccgtgagtgcttgcg
 cagcgtgaagctaattcatgggttaatttttgttaaatcagctcattttttaaccaataggccgaaatcggaacaaat
 cccttataaatcaaaagaatagcccagatagggttgagtggttccagtttgggaacaagagtcactattaaaga
 acgtggactccaacgtcaaaggcgcaaaaaccgtctatcagggcgatggccggatcagcttatgcggtgtgaaatac
 cgcacagatgcgtaaggagaaaataccgcacagggcgtcttccgcttccctcgtcactgactcgtcgcgtcggtc
 60 gttcggctgcggcgagcggatcagctcactcaaaggcggttaatacgggttatccacagaatcaggggataacgcagg
 aaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcttgctggcggtttttccataggc

tccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagatac
 caggcggtttccccctggaagctccctcgctgcgtctcctgttccgaccctgccgcttacccggatacctgtccgcctt
 tctcccttcggaagcggtggcgctttctcatagctcacgctgtaggtatctcagttcgggtgtaggtcggtccgctcca
 agctgggctgtgtgcacgaaccccccggttcagcccgaccgctgcgccttatccggttaactatcgctcttgagtccaac
 5 cccgtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgc
 tacagagttcttgaagtgggtggcctaactacggctacactagaaggacagtatgttggtatctgcgctctgctgaagc
 cagttaccttcggaaaaagagttggtagctcttgatccggcaaacaccaccgctggtagcggtggtttttttgtt
 tgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctactgaacggtgatcccc
 accggaattgcg

SEQ ID NO: 29 (gH mature protein consisting of amino acid residues 24-715 of SEQ ID NO: 1)

RYGAEAVSEPLDKAFHLLNTYGRPIRFLRENTTQCTYNSSLRNSTVVRENAISFNFFQSYNQYYVFHMPRCLFAGP
 LAEQFLNQVDLTETLERYQQRNLNTYALVSKDLASYRSFSQQLKAQDSLGEQPTTVPPPIDLSIPHVWMPPTTPHGW
 15 TESHTTSGLHRPHFNQTCILFDGHDLLFSTVTPCLHQGFYLI DELRYVKITLTEDFFVVTVSIDDDTPMLLI FGHLP
 RVLFKAPYQRDNFILRQTEKHELLVLVKKDQLNRHSYLKDPDFLDAALDFNYLDLSALLRNSFHRYAVDVLKSGRCQ
 MLDRRTVEMAFAYALALFAAARQEEAGAQVSVPRALDRQAALLQIQEFMITCLSQTTPRPTLLLYPTAVDLAKRALW
 TPNQITDITSLVRLVYILSKQNQQHLIPQWALRQIADFAKLKHLKTHLASFLSAFARQELYLMGSLVHSMVLHTTERR
 EIFIVETGLCSLAELSHFTQLLAHPHHEYLSDLTPCSSSGRRDHSRLRLTRLPDATVPATVPAALSILSTMQPST
 20 LETFPDLFCLPLGESFSALTVEHSVSIYVTNQYLIKISYPVSTTVVGQSLIITQTDSTQKCELRNMHTTHSITVA
 LNISLENCAFCQSALLEYDDTQGVINIMYMHDSDDVLFALDPYNEVVVSSPRTHYLMMLKNGTVLEVTDVVVDATD

SEQ ID NO: 30 (gH mature protein comprising SEQ ID NOs: 29 and 5)

RYGAEAVSEPLDKAFHLLNTYGRPIRFLRENTTQCTYNSSLRNSTVVRENAISFNFFQSYNQYYVFHMPRCLFAGP
 25 LAEQFLNQVDLTETLERYQQRNLNTYALVSKDLASYRSFSQQLKAQDSLGEQPTTVPPPIDLSIPHVWMPPTTPHGW
 TESHTTSGLHRPHFNQTCILFDGHDLLFSTVTPCLHQGFYLI DELRYVKITLTEDFFVVTVSIDDDTPMLLI FGHLP
 RVLFKAPYQRDNFILRQTEKHELLVLVKKDQLNRHSYLKDPDFLDAALDFNYLDLSALLRNSFHRYAVDVLKSGRCQ
 MLDRRTVEMAFAYALALFAAARQEEAGAQVSVPRALDRQAALLQIQEFMITCLSQTTPRPTLLLYPTAVDLAKRALW
 30 TPNQITDITSLVRLVYILSKQNQQHLIPQWALRQIADFAKLKHLKTHLASFLSAFARQELYLMGSLVHSMVLHTTERR
 EIFIVETGLCSLAELSHFTQLLAHPHHEYLSDLTPCSSSGRRDHSRLRLTRLPDATVPATVPAALSILSTMQPST
 LETFPDLFCLPLGESFSALTVEHSVSIYVTNQYLIKISYPVSTTVVGQSLIITQTDSTQKCELRNMHTTHSITVA
 LNISLENCAFCQSALLEYDDTQGVINIMYMHDSDDVLFALDPYNEVVVSSPRTHYLMMLKNGTVLEVTDVVVDATD
 TKLGPEQKLI SEEDLNSAVDHHHHHH

SEQ ID NO: 31 (gL mature protein consisting of amino acid residues 31-278 of SEQ ID NO: 7)

AAVSVAPTAAEKVPAECPELTRRCLLGEVFEGDKYESWLRPLVNVTVGRDGPLSQLIRYRPVTPPEAANSVLLDEAFLD
 TLALLYNNPDQLRALLTLLSSDTAPRWMTVMRGYSEC GDGSPAVYTCVDDLCRGYDLTRLSYGRSIFTEHVLGFELV
 40 PPSLFNVVVAIRNEATRTNRAVRLPVSTAAAPGITLFGYGLYNAVKEFCLRHQLDPPLLRHLDKYYAGLPPELKQTR
 VNLPAHSRYGPQAVDAR

SEQ ID NO: 32 (gO mature protein consisting of amino acid residues 31-472 of SEQ ID NO: 10)

CNVLVNSRGTRRSWPYTVLSYRGKEILKKQKEDILKRLMSTSSDGYRFLMYPSSQKFHAIVISMDKFPQDYILAGPI
 45 RND SITHMWFDFYSTQLRKPAKYVYSEYNHTAHKITLRPPPCGTVPSMNCLEMLNVSKRNDTGEKGCNFTTFNPM
 FFNVPRWNTKLYIGSNKVNVDSTIYFLGLTALLRYAQRNCTRSFYLVNAMSRLFRVPKYINGTKLKNMTRKLR
 KQALVKEQPQKNKKSQSTTTPYLSYTTSTAFNVTTNVTYSATAAVTRVATSTTGYRPDSNFMKSIMATQLRDLATW
 VYTTLRYRNEPFCKPDRNRTAVSEFMKNTHVLIRNETPYTIYGLTDMSSLYNETMSVENETASDNNETTPTSPSTR
 50 FQRTFIDPLWDYLDLFLDKIRNFSQLPAYGNLTPEHRRANLSTLNSLWWSQ

SEQ ID NO: 33 (pUL128 mature protein consisting of amino acid residues 28-171 of SEQ ID NO: 14 and 15)

EECCFEINVNHPPERCYDFKMCNRFTVALRCPDGEVCYSPEKTAEIRGIVTTMTHSLTRQVVHNKLTSCNYNPLYLE
 55 ADGRIRCGKVNDKAQYLLGAAGSVPYRWINLEYDKITRIVGLDQYLESVKKHKRLDVCRAKMGYMLQ

SEQ ID NO: 34 (pUL130 mature protein consisting of amino acid residues 26-214 of SEQ ID NO: 16)

SPWSTLTANQNPPWSKLTYSKPHDAATFYCPFLYSPPPRSPLQFSGFQRVSTGPECRNETLYLLYNREGQTLVER
 60 SSTWVKKVIWYLSGRNQITILQRMPTASKPSDGNVQISVEDAKIFGAHMPKQTKLLRFVNDGTRYQMCVMKLESW
 AHVFRDYSVSFQVRLTFTEANNQTYTFCTHPNLIV

SEQ ID NO: 35 (pUL131A mature protein consisting of amino acid residues 19-129 of SEQ ID NO: 18 and SEQ ID NO: 19)

QCQRETAEKNDYYRVPHYWDACSRALPDQTRYKYVEQLVDLTNLNYHDASHGLDNFDVLKRINVTEVSLISDFRRQ
NRRGGTNKRRTTFNAAGSLAPHARSLEFSVRLFAN

5

SEQ ID NO: 36 (gB mature protein consisting of amino acid residues 23-907 of SEQ ID NO: 21)

VSSSSTRGTSATHSHHSSHTTSAAHSRSGSVSQRVTSSTQTVSHGVNETIYNTTLKYGDVVGVNNTTKYPYRVCSMAQG
TDLIRFERNIVCTSMKPINEDLDEGIMVVYKRNIVAHTFKVRVYQKVLTFRRSYAYIHTTYLLGSNTEYVAPPMWEI
10 HHINSHSQCYSSYSRVIAGTVFVAYHRDSYENKTMQLMPDDYSNTHSTRYVTVKDQWHSRGSTWLYRETCNLNCMVT
ITTARSKYPYHFFATSTGDIVDISPFYNGTNRNASYFGENADKFFIFPNYTIVSDFGRPNSALETHRLVAFLEADER
VISWDIQDEKNVTCQLTFWEASERTIRSEAEDSYHFSSAKMTATFLSKKQEVNMSDSALDCVRDEAINKLQQIFNTS
YNQTYEKYGNVSVFETTGGVLVFWQGIKQKSLVELERLANRSSNLNTHNRTRKSTDGNNATHLSNMESVHNLVYAQL
QFTYDTLRGYINRALAQIAEAWCVDQRRTLEVFKELSKINPSAILSAIYNKPIAARFMGDVGLASCVTINQTSVKV
15 LRDMNVKESPGRCYSRPVVIFNFANSSYVQYQGLGEDNEILLGNHRTEECQLPSLKIFIAGNSAYEYVDYLFKRMID
LSSIISTVDSMIALDIDPLENTDFRVLELYSQKELRSSNVFDLEEIMREFNSYKQRVKYVEDKVVDPLPPYLKGLDDL
MSGLGAAGKAVGVAIGAVGGAVASVVEGVATFLKNPFGAFTIILVAIAVVIITYLIYTRQRRRLCTQPLQNLFPYLV
ADGTTVTSGSTKDTSLQAPPSYEEVSYNSGRKGPSPSSDASTAAPPYTNEQAYQMLLALARLDAEQRAQQNGTDSL
20 DGRTGTQDKGQKPNLLDRLRHRKNGYRHLKDSDEENV

20

Claims

1. A composition comprising purified human Cytomegalovirus (HCMV) pentameric complex comprising: HCMV gH, HCMV gL, HCMV pUL128, HCMV pUL130, and HCMV pUL131A, wherein the purity of said pentamer is at least 80% of the total proteinaceous mass in said composition.
2. The composition of claim 1, wherein said gH comprises a sequence that is at least 80% identical to any one of SEQ ID NOs: 4, 6, 29 and 30.
3. The composition of claim 1 or 2, wherein said gH comprises the whole of the gH ectodomain.
4. The composition of claim 1 or 2, wherein said gH comprises a truncated form of the gH ectodomain.
5. The composition of any one of claims 1-4, wherein said gH lacks a transmembrane domain.
6. The composition of any one of claims 1-5, wherein said gL comprises a sequence that is at least 80% identical to any one of SEQ ID NOs: 7 and 31.
7. The composition of any one of claims 1-6, wherein said gL lacks N-terminal signal sequence.
8. The composition of any one of claims 1-7, wherein said pUL128 comprises a sequence that is at least 80% identical to any one of SEQ ID NOs: 15 and 33.
9. The composition of any one of claims 1-8, wherein said pUL130 comprises a sequence that is at least 80% identical to any one of SEQ ID NOs: 16, 17 and 34.
10. The composition of any one of claims 1-9, wherein said pUL130 lacks N-terminal signal sequence.
11. The composition of any one of claims 1-10, wherein said pUL131A comprises a sequence that is at least 80% identical to any one of SEQ ID NOs: 18 and 35.
12. The composition of any one of claims 1-11, wherein said pUL131A lacks N-terminal signal sequence.
13. The composition of any one of claims 1-12, wherein the pH is between 6 and 8.
14. The composition of claim 13, wherein the pH is between 6.5 and 7.5.

15. The composition of claim 13 or 14, wherein the pH is maintained by a buffer selected from the group consisting of Tris buffer, citrate buffer, phosphate buffer and histidine buffer.
16. The composition of any one of claims 1-15, further comprising sodium chloride.
17. A combination of:
 - (a) (i) a recombinant nucleic acid construct encoding HCMV gH and HCMV gL, wherein said gH lacks a transmembrane domain; and (ii) a recombinant nucleic acid construct encoding HCMV pUL128, HCMV pUL130, and HCMV pUL131A; or
 - (b) (i) a first recombinant nucleic acid construct encoding HCMV gL; (ii) a second recombinant nucleic acid construct encoding HCMV gH, wherein said gH lacks a transmembrane domain; and (iii) a third recombinant nucleic acid construct encoding HCMV pUL128, HCMV pUL130, and HCMV pUL131A.
18. A recombinant nucleic acid construct encoding HCMV gH, HCMV gL, HCMV pUL128, HCMV pUL130, and HCMV pUL131A, wherein said gH lacks a transmembrane domain.
19. The combination of the recombinant nucleic acid constructs of claim 17, or the recombinant nucleic acid construct of claim 18, wherein said gH comprises a sequence that is at least 80% identical to any one of SEQ ID NOs: 4, 6, 29 and 30.
20. The combination of the recombinant nucleic acid constructs or the recombinant nucleic acid construct of any of claims 17-19, wherein said gH comprises the ectodomain of the gH.
21. The combination of the recombinant nucleic acid constructs or the recombinant nucleic acid construct of any one of claims 17-20, wherein said nucleic acid construct encoding HCMV gH comprises a sequence that is at least 89% identical to SEQ ID NO: 24.
22. The combination of the recombinant nucleic acid constructs or the recombinant nucleic acid construct of any one of claims 17-21, wherein said gL, comprises a sequence that is at least 80% identical to any one of SEQ ID NOs: 7 and 31.
23. The combination of the recombinant nucleic acid constructs or the recombinant nucleic acid construct of any one of claims 17-22, wherein said gL lacks N-terminal signal sequence.

24. The combination of the recombinant nucleic acid constructs or the recombinant nucleic acid construct of any one of claims 17-23, wherein said nucleic acid construct encoding HCMV gL comprise a sequence that is at least 89% identical to SEQ ID NO: 25.
25. The combination of the recombinant nucleic acid constructs or the recombinant nucleic acid construct of any one of claims 17-24, wherein said pUL128, comprises a sequence that is at least 80% identical to any one of SEQ ID NOs: 15 and 33.
26. The combination of the recombinant nucleic acid constructs or the recombinant nucleic acid construct of any one of claims 17-25, wherein said nucleic acid construct encoding HCMV UL128 comprise a sequence that is at least 74% identical to SEQ ID NO: 26.
27. The combination of the recombinant nucleic acid constructs or the recombinant nucleic acid construct of any one of claims 17-26, wherein said pUL130, comprises a sequence that is at least 80% identical to any one of SEQ ID NOs: 16, 17, and 34.
28. The combination of the recombinant nucleic acid constructs or the recombinant nucleic acid construct of any one of claims 17-27, wherein said pUL130 lacks N-terminal signal sequence.
29. The combination of the recombinant nucleic acid constructs or the recombinant nucleic acid construct of any one of claims 17-28, wherein said nucleic acid construct encoding HCMV UL130 comprise a sequence that is at least 74% identical to SEQ ID NO: 27.
30. The combination of the recombinant nucleic acid constructs or the recombinant nucleic acid construct of any one of claims 17-29, wherein said pUL131A, comprises a sequence that is at least 80% identical to any one of SEQ ID NOs: 18, and 35.
31. The combination of the recombinant nucleic acid constructs or the recombinant nucleic acid construct of any one of claims 17-30, wherein said pUL131A lacks N-terminal signal sequence.
32. The combination of the recombinant nucleic acid constructs or the recombinant nucleic acid construct of any one of claims 17-31, wherein said nucleic acid construct encoding HCMV UL131A comprise a sequence that is at least 74% identical to SEQ ID NO: 28.
33. A host cell comprising the combination of the recombinant nucleic acid constructs or the recombinant nucleic acid construct of any one of claims 17-32.
34. The host cell of claim 33, wherein said host cell is a mammalian host cell.
35. The host cell of claim 33 or 34, wherein said host cell is a HEK293 cell.

36. A cell culture for producing HCMV pentameric complex, comprising any one of host cell of claims 33-35 and a growth medium, wherein the amount of said pentameric complex is at a level of more than 0.4 mg per litre of growth medium.
37. The composition of any one of claims 1-16, which is immunogenic and further comprises an adjuvant.
38. The composition of any one of claims 1-16 and 37, for use in raising an antibody.
39. The use of claim 38, wherein said antibody is a neutralizing antibody.
40. A process for producing an HCMV pentameric complex as defined in any one of claims 1-16, wherein said process comprises recombinant expression of HCMV gH, HCMV gL, HCMV pUL128, HCMV pUL130, and HCMV pUL131A.
41. A process for expressing an HCMV pentameric complex comprising:
 - introducing the combination of the recombinant nucleic acid constructs or the recombinant nucleic acid construct according to any one of claims 17-32 into an expression system;
 - expressing said nucleic acid constructs in said expression system; and
 - purifying said HCMV pentameric complex.
42. The process of claim 41, which comprises the step of transfecting cells with the combination of nucleic acid constructs or the recombinant nucleic acid construct of any one of claims 17-32.
43. The process of claim 42, wherein the cells are mammalian cells.
44. The process of claim 43, wherein the cells are HEK293 cells.
45. The process of any one of claims 40-44, wherein the HCMV membrane protein complex has a purity of at least 80% of the total proteinaceous mass.
46. A process for producing an isolated or purified HCMV pentameric complex according to any one of claims 1-16, wherein said process involves growing the host cell of any one of claims 33-35 in growth medium.
47. The process of claim 46 wherein said HCMV pentameric complex is secreted into said growth medium.
48. The process of claim 47 wherein said HCMV complex accumulates to a concentration of more than 0.4mg of complex per litre of growth medium.

49. The process of any one of claims 40-48, wherein the HCMV membrane protein complex has a purity of >95% by mass.
50. The process of claim 49, wherein purity is determined by gel electrophoresis.