

(12) STANDARD PATENT
(19) AUSTRALIAN PATENT OFFICE

(11) Application No. **AU 2013286093 B2**

(54) Title
Complexes of cytomegalovirus proteins

(51) International Patent Classification(s)
C07K 14/045 (2006.01) **C12N 7/02** (2006.01)
A61K 39/245 (2006.01)

(21) Application No: **2013286093** (22) Date of Filing: **2013.06.29**

(87) WIPO No: **WO14/005959**

(30) Priority Data

(31)	Number	(32)	Date	(33)	Country
	61/770,257		2013.02.27		US
	61/668,975		2012.07.06		US

(43) Publication Date: **2014.01.09**

(44) Accepted Journal Date: **2018.01.04**

(71) Applicant(s)
Novartis AG

(72) Inventor(s)
Carfi, Andrea;Wen, Yingxia

(74) Agent / Attorney
FB Rice Pty Ltd, L 23 44 Market St, Sydney, NSW, 2000, AU

(56) Related Art
WO 2012/051211 A2
RYCKMAN, B. J. et al. "Characterization of the human cytomegalovirus gH/gL/UL128-131 complex that mediates entry into epithelial and endothelial cells"
Journal of Virology, vol. 82, no. 1.
ERIC R. K. et al. "Expression and reconstitution of the gH/gL/gO complex of human cytomegalovirus", Journal of clinical virology, (2002), vol. 25, no. Supp.2, pages S87 - S95.
WO 2007/146024 A2
WO 2013/006842 A2
WO 2013/033563 A1
WO 2013/006837 A1



(51) International Patent Classification:

A61K 39/245 (2006.01) C12N 7/02 (2006.01)
C07K 14/045 (2006.01)

(21) International Application Number:

PCT/EP2013/063750

(22) International Filing Date:

29 June 2013 (29.06.2013)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/668,975 6 July 2012 (06.07.2012) US
61/770,257 27 February 2013 (27.02.2013) US

(71) Applicant: **NOVARTIS AG** [CH/CH]; Lichtstrasse 35,
CH-4056 Basel (CH).

(72) Inventors: **CARFI, Andrea**; c/o Novartis Vaccines and
Diagnostics, Inc., 350 Massachusetts Avenue, Cambridge,
Massachusetts 02139 (US). **WEN, Yingxia**; c/o Novartis
Vaccines and Diagnostics, Inc., 350 Massachusetts Avenue,
Cambridge, Massachusetts 02139 (US).

(74) Agents: **EVANS, Stephen John Eves** et al.; Novartis Vac-
cines and Diagnostics S.r.L., Via Fiorentina 1, I-53100 Si-
ena (IT).

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,

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

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

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments (Rule 48.2(h))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: COMPLEXES OF CYTOMEGALOVIRUS PROTEINS

(57) Abstract: This disclosure provides an isolated human cytomegalovirus (HCMV) membrane protein complex, wherein said complex comprises gH, gL and at least one more HCMV glycoprotein. In some embodiments the complex consists of gH, gL and gO. In other embodiments the complex consists of gH, gL, p UL128, p UL130 and p UL131A. The disclosure also provides processes for expressing and purifying said complexes, and subsequent uses of said complexes in immunogenic compositions and vaccines.



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. In one example, the invention provides a purified isolated HCMV pentameric complex
25 comprising HCMV gH, HCMV gL, HCMV pUL128, HCMV pUL130, and HCMV pUL131A, wherein said gH lacks a transmembrane domain.

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
30 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.

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). In one example, the invention provides a recombinant nucleic acid molecule which encodes HCMV gL, HCMV gH that lacks a transmembrane domain, HCMV pUL128, HCMV pUL130 and HCMV pUL131A, wherein said recombinant nucleic acid molecule:

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

In another example, the invention provides a plurality of recombinant nucleic acids comprising a first construct encoding HCMV gH that lacks a transmembrane domain and HCMV gL, and a second construct encoding HCMV pUL128, HCMV pUL130 and HCMV pUL131A, 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 alphavirus non-structural proteins selected from NSP1, NSP2, NSP3 and NSP4;
- (d) does not contain an Internal Ribosomal Entry Site (IRES) selected from EMCV and EV71; and/or
- (e) does not contain viral 2A site FMDV.

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). In one example, the invention provides a cell expressing the recombinant nucleic acid of the invention or the plurality of recombinant nucleic acids of the invention, wherein said cell does not comprise the full HCMV genome.

The invention also provides a process for producing the isolated HCMV pentameric complex of the invention, said process comprising: (i) growing the cell of the invention in a growth medium; and (ii) purifying said HCMV pentameric complex from the growth medium.

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. In one example, the invention provides a method of raising antibodies in a mammal, comprising administering to the mammal an immunologically effective amount of an isolated pentameric complex of the invention or an immunogenic composition of the invention.

The invention also provides a method of treating HCMV infection in a subject, comprising administering to the subject an immunologically effective amount of an isolated pentameric complex of the invention or an immunogenic composition of the invention.

The invention also provides for use of an isolated pentameric complex of the invention or an immunogenic composition of the invention or a recombinant nucleic acid molecule of the invention or a plurality of recombinant nucleic acids of the invention or a cell of the invention in the manufacture of a medicament for raising antibodies against HCMV in a mammal.

The invention also provides for use of an isolated pentameric complex of the invention or an immunogenic composition of the invention or a recombinant nucleic acid molecule of the invention or a plurality of recombinant nucleic acids of the invention or a cell of the invention in the manufacture of a medicament for treatment of HCMV infection.

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.

Any discussion of documents, acts, materials, devices, articles or the like which has been included in the present specification is not to be taken as an admission that any or all of these matters form part of the prior art base or were common general knowledge in the field relevant to the present disclosure as it existed before the priority date of each of the appended claims.

Throughout this specification the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element, integer or step, or group of elements, integers or steps, but not the exclusion of any other element, integer or step, or group of elements, integers or steps

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%, 35

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 Ni²⁺ 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 10-500 μ g (*e.g.* 50 μ g) 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).

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
10 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.

In the RNA prime-protein boost strategy, the RNA and the protein are directed to the same
15 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
25 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
35 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.

5 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
10 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
15 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
20 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
25 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
30 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”
35 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
10 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
20 labelled directly or indirectly. A wide range of antibody labels are known in the art. In other 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.*

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.

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 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:

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 embodiment 1, wherein said process comprises the purification of the HCMV membrane protein complex.
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
 - purifying said membrane protein complex.
4. The process of embodiment 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 embodiment 3 or embodiment 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 embodiment 3 and purifying the isolated HCMV membrane protein complex.
7. The process of any preceding embodiment, wherein the HCMV membrane protein complex consists of gH, gL and gO.
8. The process of any preceding embodiment, wherein the HCMV membrane protein complex consists of gH, gL, pUL128, pUL130 and pUL131A.

The process of any preceding embodiment, 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 embodiment 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 embodiment 8 or embodiment 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 embodiment, 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 embodiment 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 embodiment 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.
14. A process for purifying an HCMV membrane protein complex as defined in any one of the preceding embodiments, wherein said purification comprises one or more chromatographic steps.
 15. The process of embodiment 14, wherein said chromatographic steps comprise affinity chromatography, preferably Ni²⁺ affinity chromatography and/or size exclusion chromatography.
 16. The process of any preceding embodiment, wherein the HCMV membrane protein complex has a purity of >85%, >86%, >87%, >88%, >89%, >90%, >91%, >92%, >93%, >94% or >95% by mass.
 17. The process of any preceding embodiment, 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.
 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 embodiment.
 20. A composition comprising the isolated HCMV membrane protein complex of embodiment 18 or embodiment 19.
 21. The composition of embodiment 20, wherein said composition does not contain polyacrylamide.
 22. The composition of embodiment 20 or embodiment 21, wherein said composition does not contain HCMV tegument or capsid proteins.
 23. The composition of any one of embodiments 20-22, wherein said composition is a liquid.
 24. The composition of any one of embodiments 20-23, wherein said composition is an immunogenic composition.
 25. The immunogenic composition of embodiment 24, which comprises gB.
 26. The immunogenic composition of embodiment 25, wherein said gB comprises or consists of any one of the sequences recited in SEQ ID NOs: 21, 22, 23 or 36.
 27. The immunogenic composition of embodiment 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 embodiments 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 embodiments 24-28, wherein said composition is a vaccine.
30. The immunogenic composition of any one of embodiments 24-29, which comprises an adjuvant.
31. The immunogenic composition of embodiment 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
 - the HCMV membrane protein complex of embodiment 18 or embodiment 19.
33. A kit comprising:
 - a priming composition comprising a self-replicating RNA molecule that encodes an HCMV membrane protein complex; and
 - a boosting composition comprising the HCMV membrane protein complex of embodiment 18 or embodiment 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;
 - (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.
35. The recombinant nucleic acid molecule of embodiment 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.

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;
 - (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 embodiment 36 comprising:
 - a first construct encoding gH that lacks a transmembrane domain and gL; and
 - a second construct encoding one additional HCMV glycoprotein.
38. The plurality of recombinant nucleic acids of embodiment 37, wherein said second construct encodes:
 - pUL128, pUL130 and pUL131A; or
 - gO.
39. The plurality of recombinant nucleic acids of embodiment 36 comprising:
 - 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.
40. The recombinant nucleic acid molecule of embodiment 34 or embodiment 35, or the plurality of recombinant nucleic acids molecules of any one of embodiment 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;
 - (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;

(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 any one of embodiments 34-40, wherein said cell does not comprise the full HCMV genome.
42. The cell of embodiment 41, wherein said cell is stably transformed with said recombinant nucleic acid or plurality of recombinant nucleic acids.
43. The cell of embodiment 41 or embodiment 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 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
 - (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 embodiments 41-44 in growth medium.
46. The process of embodiment 45, wherein said HCMV membrane protein complex is secreted into said growth medium.
47. The process of embodiment 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 embodiments 45-47, wherein said process comprises purifying said HCMV membrane protein complex from said growth medium.
49. A method for raising antibodies using the HCMV membrane protein complex of embodiment 18 or embodiment 19.
50. The method of embodiment 49, wherein said antibodies are human or humanised.
51. The method of embodiment 49 or 50, wherein said antibodies are neutralizing antibodies.
52. An antibody produced by the method of any one of embodiments 49-51.

53. An antibody produced by the method of any one of embodiments 49-52, wherein said antibody binds to the isolated HCMV membrane protein complexes of any preceding embodiment, but not isolated gH, gL, gO, pUL128, pUL130 or pUL131A.
54. The antibody of embodiment 52 or embodiment 53, wherein said antibody is for use in a diagnostic assay.
55. The antibody of any one of embodiments 52-54, wherein said antibody is labelled directly or indirectly.
56. The antibody of any one of embodiments 52-53, wherein said antibody is for use in therapy.
57. 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 1µg, mixed dose was 1+1 µg. 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:*

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
35 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
20 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:

- 15 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.
- 20 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.
- 25 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.
- 30 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 IFNalpha2b 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.
- 15 Patrone, M, M Secchi, L Fiorina, M Ierardi, G Milanesi, and A Gallina. "Human Cytomegalovirus 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.
- 25 Ryckman, BJ, et al. "Characterization of the Human Cytomegalovirus gH/gL/UL128-131 Complex 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.
- 35 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)): 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)

MRPGLPSYLIILAVCLFSHLLSSRYGAEAVSEPLDKAFHLLLNTYGRPIRFLRENTTQCTYNSSLRNSTVVRENAIS
 5 FNFFQSYNQYYVFHMPRCLFAGPLAEQFLNQVDLTETLERYQQRNLNTYALVSKDLASYRSFSQQLKAQDSLGEQPTT
 VPPPIDLSIPHVWMPPTTPHGWTESHTTSGLHRPHFNQTCILFDGHDLLFSTVTPCLHQGFYLIDELRYVKITLTE
 DFFVVTVSIDDDTPMLLIFGHLPRVLFKAPYQRDNFILRQTEKHELLVLVKKDQLNRHSYKDPDFLDAALDFNYLD
 LSALLRNSFHRYAVDVLKSGRCQMLDRRTVEMAFAYALALFAAARQEEAGAQVSVPRALDRQAALLQIQEFMITCLS
 10 QTPPRTTLLLYPTAVDLAKRALWTPNQITDITSLVRLVYILSKQNQQHLIPQWALRQIADFALKLHKTHLASFLSAF
 ARQELYLMGSLVHSMVLHTTERREIFIVETGLCSLAELSHFTQLLAHPHHEYLSLDLYTPCSSSGRRDHSLERLTRLF
 PDATVPATVPAALSILSTMQPSTLETFPDLFCLPLGESFSALTVSEHSVSYIVTNQYLIKGISYPVSTTVVGQSLIIT
 QTDSQTKCELTRNMHTTHSITVALNISLENCAFCQSALLEYDDTQGVINIMYMHDSDDVLFALDPYNEVVVSSPRTH
 YLMLLKNGTVLEVTDVVVDATDSRLLMMSVYALSIIIGIYLLYRMLKTC

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

MRPGLPSYLIIVLAVCLLSHLLSSRYGAEAISEPLDKAFHLLLNTYGRPIRFLRENTTQCTYNSSLRNSTVVRENAIS
 FNFFQSYNQYYVFHMPRCLFAGPLAEQFLNQVDLTETLERYQQRNLNTYALVSKDLASYRSFSQQLKAQDSLGEQPTT
 VPPPIDLSIPHVWMPPTTPHGWTESHTTSGLHRPHFNQTCILFDGHDLLFSTVTPCLHQGFYLIDELRYVKITLTE
 DFFVVTVSIDDDTPMLLIFGHLPRVLFKAPYQRDNFILRQTEKHELLVLVKKDQLNRHSYKDPDFLDAALDFNYLD
 20 LSALLRNSFHRYAVDVLKSGRCQMLDRRTVEMAFAYALALFAAARQEEAGAQVSVPRALDRQAALLQIQEFMITCLS
 QTPPRTTLLLYPTAVDLAKRALWTPNQITDITSLVRLVYILSKQNQQHLIPQWALRQIADFALKLHKTHLASFLSAF
 ARQELYLMGSLVHSMVLHTTERREIFIVETGLCSLAELSHFTQLLAHPHHEYLSLDLYTPCSSSGRRDHSLERLTRLF
 PDATVPPTVPAALSILSTMQPSTLETFPDLFCLPLGESFSALTVSEHSVSYVVTNQYLIKGISYPVSTTVVGQSLIIT
 25 QTDSQTKCELTRNMHTTHSITAALNISLENCAFCQSALLEYDDTQGVINIMYMHDSDDVLFALDPYNEVVVSSPRTH
 YLMLLKNGTVLEVTDVVVDATDSRLLMMSVYALSIIIGIYLLYRMLKTC

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

MRPGLPPYLTVTFTVYLLSHLPSQRYGADAASEALDPHAFHLLLNTYGRPIRFLRENTTQCTYNSSLRNSTVVRENAI
 SFNFFQSYNQYYVFHMPRCLFAGPLAEQFLNQVDLTETLERYQQRNLNTYALVSKDLASYRSFSQQLKAQDSLGGQPT
 30 TVPPPIDLSIPHVWMPPTTPHWDWKSHTTSGLHRPHFNQTCILFDGHDLLFSTVTPCLHQGFYLMDELRYVKITLT
 EDFVVTVSIDDDTPMLLIFGHLPRVLFKAPYQRDNFILRQTEKHELLVLVKKQAQLNRHSYKDSDFLDAALDFNYL
 DLSALLRNSFHRYAVDVLKSGRCQMLDRRTVEMAFAYALALFAAARQEEAGTEISIPRALDRQAALLQIQEFMITCL
 SQTPPRTTLLLYPTAVDLAKRALWTPDQITDITSLVRLVYILSKQNQQHLIPQWALRQIADFALQLHKTHLASFLSA
 FARQELYLMGSLVHSMVLHTTERREIFIVETGLCSLAELSHFTQLLAHPHHEYLSLDLYTPCSSSGRRDHSLERLTRL
 35 FPDATVPATVPAALSILSTMQPSTLETFPDLFCLPLGESFSALTVSEHSVSYVVTNQYLIKGISYPVSTTVVGQSLII
 TQTDSQTKCELTRNMHTTHSITAALNISLENCAFCQSALLEYDDTQGVINIMYMHDSDDVLFALDPYNEVVVSSPR
 HYLMLLKNGTVLEVTDVVVDATDSRLLMMSVYALSIIIGIYLLYRMLKTC

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

MRPGLPSYLIILAVCLFSHLLSSRYGAEAVSEPLDKAFHLLLNTYGRPIRFLRENTTQCTYNSSLRNSTVVRENAIS
 FNFFQSYNQYYVFHMPRCLFAGPLAEQFLNQVDLTETLERYQQRNLNTYALVSKDLASYRSFSQQLKAQDSLGEQPTT
 VPPPIDLSIPHVWMPPTTPHGWTESHTTSGLHRPHFNQTCILFDGHDLLFSTVTPCLHQGFYLIDELRYVKITLTE
 DFFVVTVSIDDDTPMLLIFGHLPRVLFKAPYQRDNFILRQTEKHELLVLVKKDQLNRHSYKDPDFLDAALDFNYLD
 45 LSALLRNSFHRYAVDVLKSGRCQMLDRRTVEMAFAYALALFAAARQEEAGAQVSVPRALDRQAALLQIQEFMITCLS
 QTPPRTTLLLYPTAVDLAKRALWTPNQITDITSLVRLVYILSKQNQQHLIPQWALRQIADFALKLHKTHLASFLSAF
 ARQELYLMGSLVHSMVLHTTERREIFIVETGLCSLAELSHFTQLLAHPHHEYLSLDLYTPCSSSGRRDHSLERLTRLF
 PDATVPATVPAALSILSTMQPSTLETFPDLFCLPLGESFSALTVSEHSVSYIVTNQYLIKGISYPVSTTVVGQSLIIT
 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)

MRPGLPSYLIILAVCLFSHLLSSRYGAEAVSEPLDKAFHLLLNTYGRPIRFLRENTTQCTYNSSLRNSTVVRENAIS
 FNFFQSYNQYYVFHMPRCLFAGPLAEQFLNQVDLTETLERYQQRNLNTYALVSKDLASYRSFSQQLKAQDSLGEQPTT
 VPPPIDLSIPHVWMPPTTPHGWTESHTTSGLHRPHFNQTCILFDGHDLLFSTVTPCLHQGFYLIDELRYVKITLTE
 60 DFFVVTVSIDDDTPMLLIFGHLPRVLFKAPYQRDNFILRQTEKHELLVLVKKDQLNRHSYKDPDFLDAALDFNYLD

LSALLRNSFHRYAVDVLKSGRCQMLDRRTVEMAFAYALALFAAARQEEAGAQVSVPRALDRQAALLQIQEFMITCLS
QTPPRTTLLLYPTAVDLAKRALWTPNQITDITSLVRLVYIILSKQNQQHLIPQWALRQIADFAKLKHLKTHLASFLSAF
ARQELYLMGSLVHSMVLHTTERREIFIVETGLCSLAELSHFTQLLAHPHHEYLSDLYTPCSSSGRRDHSLERLTRLF
PDATVPATVPAALSILSTMQPSTLETFFDLFCLPLGESFSALTVEHVSYIVTNQYLIKISYPVSTTVVGQSLIIT
5 QTDSQTKCELTRNMHTTHSITVALNISLENCAFCQSALLEYDDTQGVINIMYMHDSDDVLFALDPYNEVVVSSPRTH
YLMLLNKGTVEVTDVVVDATDGTGLGPEQKLISEEDLNSAVDHHHHHH

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

MCRRPDCGFSFSPGPVILLWCCLLLPIVSSAAVSVAPTAAEKVPAECPPELTRRCLLGEVFEGDKYESWLRPLVNVTG
10 RDGPLSQLIRYRPVTPAANSVLLDEAFDLTLALLYNPDQLRALLTLLSSDTAPRWMTVMRGYSEC GDGSPAVYTC
VDDLRCGYDLTRLSYGRSIFTEHVLGFELVPPSLFNVVVAIRNEATRTNRAVRLPVSTAAAPEGITLFYGLYNAVKE
FCLRHLQDLPPLLRHLDKYYAGLPPELKQTRVNLPASRYGPQAVDAR

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

MCRRPDCGFSFSPGPVALLWCCLLLPIVSSATVSVAPTVAEKVPAECPPELTRRCLLGEVFQGD KYESWLRPLVNVTR
15 RDGPLSQLIRYRPVTPAANSVLLDDAFDLTLALLYNPDQLRALLTLLSSDTAPRWMTVMRGYSEC GDGSPAVYTC
VDDLRCGYDLTRLSYGRSIFTEHVLGFELVPPSLFNVVVAIRNEATRTNRAVRLPVSTAAAPEGITLFYGLYNAVKE
FCLRHLQDLPPLLRHLDKYYAGLPPELKQTRVNLPASRYGPQAVDAR

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

MCRRPDCGFSFSPGPVLLWCCLLLPIVSSVAVSVAPTAAEKVPAECPPELTRRCLLGEVFQGD KYESWLRPLVNVTR
20 RDGPLSQLIRYRPVTPAANSVLLDDAFDLTLALLYNPDQLRALLTLLSSDTAPRWMTVMRGYSEC GDGSPAVYTC
VDDLRCGYDLTRLSYGRSIFTEHVLGFELVPPSLFNVVVAIRNEATRTNRAVRLPVSTAAAPEGITLFYGLYNAVKE
FCLRHLQDLPPLLRHLDKYYAGLPPELKQTRVNLPASRYGPQAVDAR

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

MGKKEMIMVKGIPKIMLLISITFLLLSLINCNVLVNSRGTTRSWPYTVLSYRGKEILKKQKEDILKRLMSTSSDGYR
FLMYPQQKFHAIIVISMDKFPQDYILAGPIRNDSTHMFDFYSTQLRKPAKYVYSEYNHTAHKITLRPPPCGTVPS
MNCLSEMLNVSKRNDTGEKGCNFTTFNPMFFNVPRWNTKLYIGSNKVNVDSTIYFLGLTALLRYAQRNCTRSFY
30 LVNAMSRLNFRVPKYINGTKLKNTMRKLKRKQALVKEQPQKKNKKSQSTTTPYLSYTTSTAFNVTNTVYSATAAVT
RVATSTTGYPDSNFMKSIMATQLRDLATWVYTTLRYNPFCKPDRNRTAVSEFMKNTHVLIRNETPYTIYGTLDL
SSLYNETMSVENETASDNNETTPTSPSTRFQRTFIDPLWDYLDLSLLFLDKIRNFSLQLPAYGNLTPPEHRAANLS
TLNSLWWSQ

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

MGRKEMMVRDVPKMVFLISISFLLVSFINCKVMSKALYNRPWRGLVLSKIGKYKLDQLKLEILRQLETTISTKYNV
KQPVKNLTMMNTEFPQYYILAGPIQNYISITYLWDFDYSTQLRKPAKYVYSQYNHTAKTITFRPPPCGTVPSMTCLSE
MLNVSKRNDTGEQGCNFTTFNPMFFNVPRWNTKLYVGPTKVNVDSTIYFLGLTALLRYAQRNCTHSFYLVNAMS
RNLFRVPKYINGTKLKNTMRKLKRKQAPVKEQFEKKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTAARRVSTST
40 IAYRPDSSFMKSIMATQLRDLATWVYTTLRYPQNPFCPSRNRTAVSEFMKNTHVLIRNETPYTIYGTLDL
ETMFVENKTASDSNKTTPSPSMGFQRTFIDPLWDYLDLSLLFLDEIRNFSLSRPTYVNLTPPEHRAVNLSLNSLW
WWLQ

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

MGRKGEMRGVFNLFLLMSLTFLLSFINCKIAVARFRVKSQKAKEEERQLKLRLILQELASKTGDYKFFFTFPSQK
YNITVEMKQFPNSILAGPIRNHSITHLWDFHTTQLRKPAKYVYSEYNHTGQKITFRPPSCGTIPSMTCLEMLNV
SRRNNTGEENCNFTTFNPMFFNVPRWNTKLYVGPSKVNVDSTIYFLGLAALLRYAQRNCTRSFYLVNAMSRLN
RVPKYINSTKLKNTMRKLKRKQAPVKSISKKSRVSTTTPYSSYSTIFNVSTNTVYSPIVPTRIPTSTIGYRPDENF
MKSILTTQLKDLATWVYTTLRYPQNPFCPSRNRTAVSEFMKNTHVLIRNETPYTIYGTLDLSSLYYNDTMPVENET
50 ASDNNKTTPSPSTRFQRTFIDPMWDYLDLSLLFLSEIRNFSLSQSSTYGNLTPPEHRAVNLSLNSLW
WWWLQ

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

MSPKDLTPFLTALWLLLGHSRVPRVRAEECCFEINVNHPERCYDFKMCNRFTVALRCPDGEVCYSPEKTAEIRGIV
TTMTHSLTRQVVHNKLTSCNYNPLYLEADGRIRCGKVNDKAQYLLGAAGSVPY

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

MSPKNLTPFLTALWLLLGHSRVPRVRAEECCFEINVNHPERCYDFKMCNRFTVALRCPDGEVCYSPEKTAEIRGIV
TTMTHSLTRQVVHNKLTSCNYNPLYLEADGRIRCGKVNDKAQYLLGAAGSVPYRWINLEYDKITRIVGLDQYLESVK
60 KHKRLDVCRAKMGYMLQ

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

MSPKDLTPFLTTLLWLLLGHSRVPRVRAEECCFEINVNHPPERCYDFKMCNRFTVALRCPDGEVCYSPEKTAEIRGIV
 TTMTHSLTRQVVHNKLTSCNYPNPLYLEADGRIRCCKVNDKAQYLLGAAGSVPRWINLEYDKITRIVGLDQYLESVK
 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)

MRLCRVWLSVCLCAVVLGQCQRETAEKNDYYRVPHYWDACSRALPDQTRYKYVEQLVDLTLNYHYDASHGLDNFVDVL
 KRINTEVSLLISDFRRQNRGGTNKRTTFNAAGSLAPHARSLEFSVRLFAN

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

MRLCRVWLSVCLCAVVLGQCQRETAEKNDYYRVPHYWDACSRALPDQTRYKYVEQLVDLTLNYHYDASHGLDNFVDVL
 KRINTEVSLLISDFRRQNRGGTNKRTTFNAAGSLAPHARSLEFSVRLFAN

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

MRLCRVWLSVCLCAVVLGQCQRETAEEKRLLPSTALLGRVLSRAARNPLQVCGTARGPHVELPLRCEPRLGQL

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

MESRIWCLVVCVNLICIVCLGAAVSSSSSTRGTSATHSHSSHTTSAAHRSRSGSVSQRVTSSTQTVSHGVNETIYNTTLK
 YGDVVGVTNTKYPYRVCSMAQGTDLIRFERNIVCTSMKPINEDLDEGIMVYKRNIVAHTFKVRVYQKVLTFRRSYA
 YIHTTYLLGSNTEYVAPPMWEIHHINSHSQCYSSYSRVIACTVGFVAYHRDSYENKTMQLMPDDYSNTHSTRYVTVKD
 QWHSRGSTWLYRETCNLNLCMVTITARSKYPHYHFFATSTGDVVDISPFYNGTNRNASYFGENADKFFIFPNYTIIVSD
 FGRPN SALETHRLVAFLE RADSVISWDIQDEKNVTCQLTFWEASERTIRSEAEDSYHFSSAKMTATFLSKKQEVNMS
 DSALDCVRDEAINKLQQIFNTSYNQTYEKGYNVSVFETTGGVLVFWQGIKQKSLVELERLANRSSNLTHNRTRKST
 DGNNATHLSNMESVHNLVYAQLQFTYDTRLGYINRALAQIAEAWCVDQRRTELVFKELSKINPSAILSAINKPIAA
 RFMGDVLGLASCVTINQTSVKVLRDMNVKESPGRCYSRPVVI FNFANSSYVQYQGLGEDNEILLGNHRTEECQLPSL
 KIFIAGNSAYEYVDYLFKRMIDLSSI STVDSMIALDIDPLENTDFRVLELYSQKELRSSNVFDLEEIMREFNSYKQR
 VKYVEDKVVDPLPPYLKGLDDLMSGLGAAGKAVGVAIGAVGGAVASVVEGVATFLKNPFGAFTIILVAIAVVIITYL
 IYTRQRRCTQPLQNLFPYLVADGTTVTSGSTKDTSLQAPPSYEE SVYNSGRKGPGPPSSDASTAAPPYTNEQAYQ
 MLLALARLDAEQRAQQNGTDSL DGRGTGTQDKGQKPNLLDRLRHRKNGYRHLKDSDEEENV

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

MESRIWCLVVCVNLICIVCLGAAVSSSSSTRGTSATHSHSSHTTSAAHRSRSGSVSQRVTSSTQTVSHGVNETIYNTTLK
 YGDVVGVTNTKYPYRVCSMAQGTDLIRFERNIVCTSMKPINEDLDEGIMVYKRNIVAHTFKVRVYQKVLTFRRSYA
 YIHTTYLLGSNTEYVAPPMWEIHHINSHSQCYSSYSRVIACTVGFVAYHRDSYENKTMQLMPDDYSNTHSTRYVTVKD
 QWHSRGSTWLYRETCNLNLCMVTITARSKYPHYHFFATSTGDVVDISPFYNGTNRNASYFGENADKFFIFPNYTIIVSD
 FGRPN SALETHRLVAFLE RADSVISWDIQDEKNVTCQLTFWEASERTIRSEAEDSYHFSSAKMTATFLSKKQEVNMS
 DSALDCVRDEAINKLQQIFNTSYNQTYEKGYNVSVFETTGGVLVFWQGIKQKSLVELERLANRSSNLTHNRTRKST
 DGNNATHLSNMESVHNLVYAQLQFTYDTRLGYINRALAQIAEAWCVDQRRTELVFKELSKINPSAILSAINKPIAA
 RFMGDVLGLASCVTINQTSVKVLRDMNVKESPGRCYSRPVVI FNFANSSYVQYQGLGEDNEILLGNHRTEECQLPSL
 KIFIAGNSAYEYVDYLFKRMIDLSSI STVDSMIALDIDPLENTDFRVLELYSQKELRSSNVFDLEEIMREFNSYKQR
 VKYVEDKVVDPLPPYLKGLDDLMSGLGAAGKAVGVAIGAVGGAVASVVEGVATFLKNPFGAFTIILVAIAVVIITYL
 IYTRQRRCTQPLQNLFPYLVADGTTVTSGSTKDTSLQAPPSYEE SVYNSGRKGPGPPSSDASTAAPPYTNEQAYQ
 MLLALVRLDAEQRAQQNGTDSL DGTGTGTQDKGQKPNLLDRLRHRKNGYRHLKDSDEEENV

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

MESRIWCLVVCVNLICIVCLGAAVSSSSSTSHATSSHTNGSHTSRTTSAQTRSVYSQHVTSSSEAVSHRANETIYNTTLK
 YGDVVGVTNTKYPYRVCSMAQGTDLIRFERNI ICTSMKPINEDLDEGIMVYKRNIVAHTFKVRVYQKVLTFRRSYA
 YIYTTYLLGSNTEYVAPPMWEIHHINKFAQCYSSYSRVIGGVTFVAYHRDSYENKTMQLIPDDYSNTHSTRYVTVKD
 QWHSRGSTWLYRETCNLNLCMLTITARSKYPHYHFFATSTGDVVYISPFYNGTNRNASYFGENADKFFIFPNYTIIVSD

FGRPNAAPETHRLVAFLEADSVISWDIQDEKNVTCQLTFWEASERTIRSEAEDSYHFSSAKMTATFLSKKQEVNMS
 DSALDCVRDEAINKLQQIFNTSYNQTYEKGNSVVFETSGGLVFWQGIKQKSLVELERLANRSSLNITHRTRRSTS
 DNNTTHLSSMESVHNLVYAQLQFTYDTRLGYINRALAQIAEAWCVDQRRTELVFKELSKINPSAILSAIYNKPIAAR
 FMGDVLGLASCVTINQTSVKVLRDMNVKESPGRCYSRPVVI FNFANSSYVQYQGLGEDNEILLGNHRTTEECQLPSLK
 5 IFIAGNSAYEYVDYLFKRIMDLSSI STVDSMIALDIDPLENTDFRVLELYSQKELRSSNVFDLEEIMREFNSYKQRV
 KYVEDKVVDPLPPYLLKGLDDLM SGLGAAGKAVGVAIGAVGGAVASVVEGVATFLKNPFGAFTIILVAIAVVIITYLI
 YTRQRRLLCTQPLQNLFPYLV SADGTTVTSGSTKDTSLQAPPSYEE SVYNSGRKGPGPPSSDASTAAPPYTNEQAYQM
 LLALARLDAEQRAQQNGTDSL DGTGTQDKGQKPNLLDRLRHRKNGYRHLKDSDEEENV

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

gacggatcgggagatctcccgatcccctatggtgcactctcagtacaatctgctctgatgccgcatagttaagccag
 tatctgctccctgcttgtgtgtgttgaggctcgtgagtagtgccgcgagcaaaatttaagctacaacaaggcaaggctt
 gaccgacaattgcatgaagaatctgcttagggtaggcgttttgcgctgcttcgcgatgtacgggccagatatacgc
 15 gttgacattgattattgactagttattaatagtaaatcaattacgggggtcattagttcatagcccatatatggagttc
 cgcgttacataacttacggtaaatggccgcctggtgacggcccaacgacccccgcccattgacgtcaataatgac
 gtatgttcccataagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacggtaaactgccact
 tggcagtagcatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggccgcctggcat
 tatgcccagtagcatgaccttatgggactttcctacttggcagtagcatctacgtattagtcacgtattaccatgggt
 20 gatgcgggttttggcagtagcatcaatggcggtggatagcgggtttgactcacggggattttccaaagtctccacccccattg
 acgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacg
 caaatgggcggttaggcgtgtacgggtgggaggtctatataagcagagctctctggctaactagagaacccactgctta
 ctggcttatcgaaattaatacagactcactatagggagacccaagctggctagcgccaccatgaggcctggcctgcc
 tcctacctgatcatcctggcgtgtgctgttcagccacctgctgtccagcagatacggcgccgagggcgtgagcga
 25 gcccctggacaaggctttccacctgctgctgaacacctacggcagacccatccggtttctgcgggagaaacaccaccc
 agtgcacctacaacagcagcctgcggaacagcaccgtcgtgagagagaacgccatcagcttcaacttttccagagc
 tacaaccagtactacgtgttccacatgccagatgcctgtttgcccggccctctggccgagcagttcctgaaccaggt
 ggacctgaccgagacactggaaagataaccagcagcggtgaatacctacgccctgggtgtccaaggacctggccagct
 accggtccttttagccagcagctcaaggctcaggatagcctcggcgagcagcctaccaccggtgccccctcccacgcac
 30 ctgagcatccccacgtgtggatgcctccccagaccacccctcacggctggaccgagagccacaccacctccggcct
 gcacagacccccacttcaaccagacctgcacctgttcgacggccacgacctgctgttttagcaccgtgacccccctgcc
 tgcaccagggtcttctacctgatcgacgagctgagatacgtgaagatcacctgaccgaggtatttcttctgtgtcacc
 gtgtccatcgacgacgacacccccatgctgctgatcttcggccacctgcccagagtgctgttcaaggccccctacca
 gcgggacaacttcatcctgcggcagaccgagaagcagcagctgctggtgctggtcaagaaggaccagctgaaccggc
 35 actcctacctgaaggaccccgacttccctggacgcgcgcctggacttcaactacctggacctgagcgccctgctgaga
 aacagcttccacagatacgcctggacgtgctgaagtccggacgggtgccagatgctcgatcggcgacctgtggagat
 ggcccttcgcctatgccctgcgcctgttcgcgcgtgccagacaggaagaggctggcgcccagggtgtcagtgcccagag
 ccctggatagacaggccgcctgctgcagatccaggaattcatgatcacctgcctgagccagacccccctagaacc
 accctgctgctgtacccccacagccgtggatctggccaagaggccctgtggaccccccaaccagatcacccgacatcac
 40 aagcctcgtgcggctcgtgtacatcctgagcaagcagaaccagcagcactgatccccagtgggccctgagacaga
 tcgcccacttcgcccctgaagctgcacaagacccatctggccagctttctgagcgcttcgcccaggcaggaactgtac
 ctgatgggcagcctggtccacagcatgctggtgcataaccacggagcggcgaggagatcttcatcgtggagacaggcct
 gtgtagcctggccgagctgtccactttaccagctgctggcccaccctcaccacgagtagctgagcgacctgtaca
 cccccctgcagcagcagcggcagacgggaccacagcctggaacggctgaccagactgttccccgatgccacctgtcct
 45 gctacagtgctgcccgcctgtccatcctgtccaccatgcagcccagcaccctggaaaccttccccgacctgttctg
 cctgccccctgggagagacttttagcgccctgacctgtccgagcagctgtcctacatcgtgaccaatcagtagctga
 tcaaggcatcagctaccccgctgtccaccacagctcgtgggcccagagcctgatcatccccagaccgacagccagacc
 aagtgcgagctgaccgggaacatgcacaccacacacagcatcacctggccctgaacatcagcctggaaaactgcgc
 tttctgtcagctcgtccctgctggaatacagcagatacccaggcgctgatcaacatcatgtacatgcacgacagcgacg
 50 acgtgctgttcgcccctggacccctacaacgaggtggtggtgtccagcccccgaccactacctgatgctgtgaag
 aacggcacctgctggaagtacccgacgtggtggtggacgcccaccgacggtaccaagcttgggcccgaacaaaaact
 catctcagaagaggatctgaatagcgccgctgaccatcatcatcatcatcattgagtttaaacgggtctccagcttaa
 gtttaaacgctgatcagcctcgactgtgccttctagttgccagccatctgtgtttgcccctcccccgctgccttcc
 ttgacctggaagggtgccactcccactgtcctttcctaataaaatgaggaaaattgcatcgcatgtgtcagtaggtg
 55 tcattctattctggggggtgggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctgggg
 atgcggtgggctctatggcttctgaggcggaagaaccagctggggctctagggggtatccccacgcgcccctgtagc
 ggcgcattaaagcgcgggcggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgcccctagcgcccgtcc
 tttcgcttttctcccttcccttctcgccacgttcgcccgttttccccgtcaagctctaaatcgggggctccctttag
 ggttccgatttagtgctttacggcacctcgacccccaaaaaacttgattaggggtgatggttcacgtagtgggccatcg
 60 ccctgatagacgggttttctgccccttgacgttggagtcacgttctttaaagtgaggactctgttccaaactggaac
 aacactcaaccctatctcgggtctattcttttgatttataagggattttgcccgatctcggcctattggttaaaaaatg

agctgattttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttaggggtgtggaagtccccaggct
 cccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccagggtgtggaagtccccaggctcccca
 gcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcgcccccctaactccgcccacccgcc
 cctaactccgcccagttccgcccattctccgcccattggtgactaattttttttatattatgcagaggccgaggccg
 5 cctctgcctctgagctattccagaagtagtgaggaggcttttttgaggcctaggttttgcaaaaagctccccggga
 gcttgatatatccattttcggatctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgca
 cgcagggttctccggccgcttggttgagaggctattcggctatgactgggcacacagacaatcggtgctctgatg
 ccgccgtgttccggctgtcagcgcagggcgcccggttcttttgtcaagaccgacctgtccggtgacctgaatgaa
 ctgcaggacgaggcagcgcggtatcggtggtggccacgacggcggttccttgccgagctgtgctcgacgttgtcac
 10 tgaagcgggaagggactggctgctattgggcgaagtgcggggcaggatctcctgtcatctcaccttgctcctgccc
 agaaagtatccatcatggctgatgcaatgcggcggtgcatacgttgatccggctacctgcccattcgaccacca
 gcgaacatcgcatcgagcagcagctactcggatggaagccggtcttgctcgatcaggatgatctggacgaagagca
 tcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatgcccgcaggcgaggatctcgctcgtagccc
 atggcgatgctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcgactgtggccggctgggt
 15 gtggcggaccgctatcaggacatagcgttggttaccgctgatattgctgaagagcttgccggcgaatgggtgaccg
 ctctcctcgctgtttacgggtatcgccgctcccgattcgcagcgcacgccttctatcgcttcttgacgagttcttct
 gagcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccgcg
 ccttctatgaaaggttggttctcggaatcgtttccgggacgcggctggatgatcctccagcgcggggatctcatg
 ctggagtcttctgcccccacccaacttggtttattgcagcttataatgggttacaataaagcaatagcatcacaaattt
 20 cacaaataaagcatttttttactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgta
 taccgtcgacctctagctagagcttggtgaatcatggtcatagctgttctcgtgtgaaattgttatccgctcaca
 attccacacacatacagagccggaagcataaagtgtaaagcctggggtgcctaataagtgagtaactcacattaat
 tgcgttgcgctcactgcccgtttccagtcgggaaacctgtcgtgccagctgcattaatgaatcgcccaacgcgcg
 ggagaggcggtttgctgattgggcgctcttccgcttctcgtcactgactcgctgcgctcggtcggtcggtgcg
 25 cgagcggatcagctcactcaaaggcggttaatacgggtatccacagaatcaggggataacgcaggaaagaacatgtg
 agcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcttgctggcggttttccataggctccgccccctg
 acgagcatcacaaaaatcgacgctcaagtcagagggtggcgaaaccgcagagactataaagataaccaggcggttccc
 cctggaagctccctcgctgcgctctcctgttccgacctgcccgttaccgggataacctgtccgcctttctcccttcggg
 aagcgtggcgctttctcatagctcacgctgtaggtatctcagttcgggtgtaggtcggtcgctccaagctgggctgtg
 30 tgcacgaaccccccggttcagcccgaccgctgcgccttatccggtaactatcgctcttgagtccaacccggtaagacac
 gacttatcgccactggcagcagccactggtaacaggattagcagagcgaggatgtaggcggtgctacagagttctt
 gaagtgggtggcctaactacggctacactagaagaacagttatttggtatctgcgctctgctgaagccagttacctcg
 gaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtgggttttttgtttgcaagcagcag
 attacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaa
 35 ttcacggttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaattaaaaaatgaagtt
 ttaaatcaatctaaagtatatatgagtaaaccttggtctgacagttaccaatgcttaatcagtgaggcacctatctca
 gcgatctgtctattttcggtcatccatagttgcctgactccccgctcggtgtagataactacgatacgggagggttacc
 atctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccag
 ccggaaggggccgagcgcagaagtgggtcctgcaactttatccgcctccatccagttatattaattgttgccgggaagct
 40 agagtaagtagttcgccagtttaatagtttgcgcaacggttggtgacattgctacaggcatcggtggtgcacgctcgctc
 gtttggtatggcttcattcagctccgggttcccaacgatcaaggcgagttacatgatccccatgttggtgcaaaaaag
 cggttagctccttcggtcctccgatcggttgctcagaagtaagttggccgcagtggttatcactcatggttatggcagca
 ctgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactgggtgagtactcaaccaagtcattctg
 agaatagtgatgctggcgaccgagttgctcttgccggcgctcaatacgggataataccgcgccacatagcagaactt
 45 taaaagtgtcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctggttgagatccagttcg
 atgtaaccctcctgaccccaactgatcttcagcatcttttactttcaccagcggttctggttgagcaaaaaacagg
 aaggcaaatgcccgaaaaaagggaataaggcgacacggaaatgttgaaactcatactcttcttttcaatatt
 attgaagcatttatcagggttattgtctcatgagcggatataatttgatgtatttagaaaaataaacaatataggg
 gttccgcgcacatttccccgaaaagtgcacactgacgtc

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

gccgcggaatttcgactctaggccattgcatacgttggtatctatatcataatatgtacattttatattggctcatgtc
 caatatgaccgcatgttgacattgattattgactagttattaatagtaatacaattacggggtcattagttcatagc
 ccatatatggagttccgcgttacataacttacggtaaatggcccgctgggtgaccgcccacgacccccgcccatt
 55 gacgtcaataatgacgtatgttcccatagtaacgcccaatagggactttccattgacgtcaatgggtggagtatttac
 ggtaaactgccacttgccagtagcatcaagtgtatcatatgccaaagtcgccccctattgacgtcaatgacggtaaa
 tggcccgctggcattatgccagtagcatgaccttacgggactttcctacttgccagtagcatctacgtattagtcac
 cgctattaccatggtgatgctggttttgccagtagaccaatggcggtgagtagcggttgactcacggggatttccaa
 gtctccacccattgacgtcaatgggagtttggtttggcaccaaaatcaacgggactttccaaaatgtcgtaataac
 60 cccgccccgttgacgcaaatggcggttaggcgtgtacgggtgggaggtctatataagcagagctcggttagtgaaaccg
 tcagatcgctggagacgccatccacgctgttttgacctccatagaagacaccgggaccgatccagcctccgcgcc

gggaacggtgcattggaacgcggattccccgtgccaagagtgcgtaagtaccgcctatagactctataggcacacc
 cctttggctcttatgcatgctatactgtttttggcttggggcctatacacccttcccttatgctataggatgatg
 gtatagcttagcctatagggtgtgggttattgaccattattgaccactcccctattgggtgacgatactttccattact
 aatccataacatggctcttttgcacaactatctctattggctatatgccaatactctgtccttcagagactgacacg
 5 gactctgtatttttacaggatgggggtccattttattttacaaattcacatatacaacaacgcggtcccccggtgcc
 cgcagttttttatataacatagcgtgggatctccacgcgaatctcggtacgtgttccggacatggggtcttctccgg
 tagcggcgagcttccacatccgagccctgggtcccatgcctccagcgggtcatgggtcgctcggcagctccttgcctc
 taacagtggaggccagacttaggcacagcacaatgccaccaccaccagtggtgcccgcacaaggccgtggcggtaggg
 tatgtgtctgaaaatgagctcggagattgggtcgcaccgctgacgcagatggaagacttaaggcagcggcagaaga
 10 agatgcaggcagctgagttgttgtattctgataagagtcagaggtaactcccgttgcgggtgctgttaacggtggagg
 gcagtgtagtctgagcagctactcgttgcctgcccgcgcgcgccaccagacataatagctgacagactaacagactgttc
 ctttccatgggtcttttctgcagtcacccgtcgtcgcagcgcaccatgtgcagaaggcccgactgcggcttcagcttca
 gccctggaccgcgtgatcctgctgtggtgctgcctgctgctgcctatcgtgtcctctgccgcccgtgtctgtggccct
 acagccgcgcgagaagggtgccagccgagtgccccgagctgaccagaagatgcctgctgggcgaggtgttcgagggcga
 15 caagtacgagagctgggtgcccgccttgggtcaacgtgaccggcagagatggccccctgagccagctgatccgggtaca
 gaccgtgacccccgagggccccaatagcgtgctgctggacgagggccttccctggataccctggccctgctgtacaac
 aaccccgaccagctgagagccctgctgacctgctgctccagcgacaccgccccagatggatgacctgagtcgaggg
 ctacagcgagtggtggagatggcagccctgcccgtgtacacctgctggacgacctgtgcagaggtctacgacctgacca
 gactgagctacggccggtccatcttcacagagcagctgctgggtctcgagctgggtgccccccagcctgttcaacgtg
 20 gtggtggccatccggaacgaggccaccagaaccaacagagccgtgcccgtgctgtgtctacagccgctgcacctga
 gggcatcacactgttctacggcctgtacaacgcgctgaaagagtcttgcctccggcaccagctggatccccccctgc
 tgagacacctggacaagtaactacgcggcctgccccagagctgaagcagaccagagtgaaacctgcccgcacagc
 agatatggccctcaggccgtggacgcccagatgataatctagaaagccatggatccgactacgcgttagagc
 tcgctgatcagcctcgactgtgccttctagtgtccagccatctgttgtttgcccctccccgtgccttccctgaccc
 25 tgggaagggtgccactcccactgtcctttcctaataaaaatgaggaaattgcacgcattgtctgagtaggtgtcattct
 attctgggggggtgggggtggggcaggacagcaagggggaggattgggaagacaatagcaggggggtggcggaagaact
 ccagcatgagatccccgcgctggaggatcatccagccggcgtcccggaaaacgattccgaagcccaacctttcatag
 aaggcggcggtggaatcgaaatctcgtgatggcaggttggggcgtcgttgggtcgggtcatttcgaaccccagagtccc
 gctcagaagaactcgtcaagaaggcgatagaaggcgatgcgctgcgaatcgggagcggcgataaccgtaaagcacgag
 30 gaagcggtcagcccatttcgcgcgaagctcttcagcaatatcacgggtagccaacgctatgtcctgatagcgggtccg
 ccacaccagccggccacagtcgatgaatccagaaaagcggccattttccaccatgatattcggaagcagggcatcg
 ccattgggtcacgacgagatcctcgcgctcgggcatgcgcgccttgagcctggcgaacagttcgggtggcgagccc
 ctgatgtctcttcgtccagatcatcctgatcgacaagaccggcttccatccgagtaogtgcctcgtcgtatgcgatgtt
 tcgcttgggtgggtcgaaatgggcaggttagccggatcaagcgtatgcagccgcgcgattgcacagccatgatggatact
 35 ttctcggcaggagcaaggtgagatgacaggagatcctgccccggcacttcgcccataagcagccagtccttccccgc
 ttcagtgacaacgtcgagcacagctgcgcaaggaaacgcgcgctggtggccagccacgatagccgcgctgcctcgtcct
 gcagttcattcagggcaccggacaggtcgggtcttgacaaaaagaaccggggcgccctgcgctgacagccggaacacg
 gcggcatcagagcagccgattgtctgttgtgcccagtcatagccgaatagcctctccacccaagcggccgggagaacc
 tgctgcaatccatcttgttcaatcatgcgaaacgatcctcatcctgtctcttgatcagatcttgatccccctgcgcc
 40 atcagatccttggcggcaagaaagccatccagtttactttgcagggttcccaaccttaccagaggcgccccagct
 ggcaattccgggttcgcttgcgtgtccataaaaaccgcccagcttagctatcgccatgtaagccactgcaagctacctg
 ctttctctttgcgcttgcgttttcccttgcagatagcccagtagctgacattcatccgggggtcagcacccgtttct
 gcggactggctttctacgtgttccgcttccctttagcagcccttgccgctgagtgcttgcggcagcgtgaagctaatt
 tcatggttaaaattttgttaaatcagctcattttttaaccaataggccgaaatcggcaaaatcccttataaatcaaa
 45 agaatagcccagataggggttagtggttgcagtttggaaacaagagtcactattaaagaacgtggactccaacg
 tcaaagggcgaaaaaacgctctatcaggggcgatggccggatcagcttatgcggtgtgaaataccgcacagatgcgtaa
 ggagaaaaataccgcatcaggcgctcttccgcttccctcgtcactgactcgtgcgctcgggtcgttcgggtgcggcga
 gcgggtatcagctcactcaaaggcggtaatacgggttatccacagaatcaggggataacgcaggaaagaacatgtgagc
 aaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgcgtggcgtttttccataggctccgccccctgacg
 50 agcatcacaaaaatcgacgctcaagtcagagggtggcgaaacccgacaggactataaagataaccaggcgtttccccct
 ggaagctccctcgtgcgctctcctgttccgacctgcccgttaccggatacctgtccgcttctccttccgttccgggaag
 cgtggcgctttctcatagctcacgctgtaggtatctcagttcgggtgtaggtcgttcgctccaagctgggctgtgtgc
 acgaacccccgttcagcccagccgctgcgccttatccggtaactatcgtcttgagtccaacccggttaagacacgac
 ttatcgccactggcagcagccactggtaacaggattagcagagcagggatgtaggcgggtgtacagagttcttgaa
 55 gtggtggcctaactacggctacactagaaggacagtatgttgggtatctgcgctctgctgaagccagttaccttccgaa
 aaagagttggtagctcttgatccggcaaaacaaaccaccgctggtagcgggtgtttttttgtttgcaagcagcagatt
 acgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctactgaacggtgatccccaccggaattgcg

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

60 gccgcggaatttcgactctaggccattgcatacgttgtatctatatcataatatgtacatttatattggctcatgtc
 caatatgaccgcatgttgacattgattattgactagttattaatagtaatacaattacgggggtcattagttcatagc

ccatatatggagttccgcgttacataacttacggtaaattggcccgctggctgaccgcccacgacccccgcccatt
 gacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttac
 ggtaaaactgccacttggcagtagcatcaagtgtatcatatgccaaagtcgccccctattgacgtcaatgacggtaaa
 5 tggcccgctggcattatgccagtagacacgttacgggactttcctacttggcagtagcatctacgtattagtcat
 cgctattaccatgggtgatgcggttttggcagtagaccaatgggcgtggatagcggtttgactcacggggatttccaa
 gtctccaccccatgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtataaac
 cccgccccgttgacgcaaatgggcgttaggcgtgtacgggtgggaggtctatataagcagagctcggttagtgaaaccg
 tcagatcgctggagacgccatccacgctgttttgacctccatagaagacaccgggaccgatccagcctccgcgcc
 10 gggaacggtgcatgggaacgcggattccccgtgccaaagagtacgtaagtaccgcctatagactctataggcacacc
 cctttggctcttatgcatgtatactgttttggcttggggcctatacacccccgcttcccttatgctatagggtgatg
 gtatagcttagcctatagggtgtgggttattgaccattattgaccactccccctattgggtgacgatactttccattact
 aatccataacatggctctttggccaaactatctctattggctatatgccaaactctgtccttcagagactgacacg
 gactctgtatttttacaggatgggggtcccatttattatttacaaattcacatatacaacaacgccgtcccccggtgcc
 15 cgcagtttttattaaacatagcgtggggtctccacgcgaatctcgggtacgtgttccggacatggggtcttctccgg
 tagcggcgagcttccacatccgagccctgggtcccatgcctccagcgggtcatgggtcgctcggcagctccttgctcc
 taacagtggaggccagacttaggcacagcacaaatgccaccaccaccagtggtgcgcgacaaggccgtggcggtaggg
 tatgtgtctgaaaatgagctcggagattgggtcgcacgcgtacgcagatggaagacttaaggcagcggcagaaga
 agatgcaggcagctgagttgttattctgataagagtcagaggttaactccggttgcggtgtgttaacggtggagg
 gcagtgtagtctgagcagtagctgttgcgtgcgcgcgcgccaccagacataatagctgacagactaacagactgttc
 20 ctttccatgggtcttttctgcagtcacgcgtcgctgcagcgcaccatgagcccaaggacctgaccccccttctgacaa
 ccctgtgggtgctcctgggcatagcagagtgcttagagtgctggggccgaggaatgctgcgagttcatcaacgtgaac
 ccccccccgagcgggtgctacgacttcaagatgtgcaaccgggttcaccgtggccctgagatgccccgacggcgaagt
 gtgctacagccccgagaaaaccgcgcgagatccggggcatcgtgaccaccatgaccacagcctgacccggcaggtgg
 tgcacaacaagctgaccagctgcaactacaacccccctgtacctggaagccgacggccggatcagatgcggcaaagt
 25 aacgacaaggccagtagctgtgggagcgcgcggaagcgtgcctaccggtggatcaacctggaatacagacaagat
 caccggatcggtgggcctggaccagtagctggaagcgtgaagaagcacaagcggctggacgtgtgcagagccaaga
 tgggctacatgctgcagtgataatctagaaagccatggatatacggatccactacgcgttagagctcgctgatcagcc
 tcgactgtgccttctagttgccagccatctgttgtttgccccctcccccggtgccttccctgaccctggaaggtgccac
 tcccactgtcctttcctaataaaaatgaggaaaattgcatcgcatgtctgagtaggtgtcattctattctgggggggtg
 30 ggggtggggcaggacagcaagggggaggattgggaagacaatagcaggggggtgggccaagaactccagcatgagatc
 cccgcgtggaggatcatccagccggcgtcccggaacacgattccgaagcccaacctttcatagaaggcggcggtgg
 aatcgaaatctcgtgatggcaggttgggcgtcgcttgggtcggtcatttccgaaccccagagtcggcgtcagaagaact
 cgtcaagaaggcgatagaaggcgatgcgtcgcgaatcgggagcggcgataccgtaaaagcacgaggaagcggtcagcc
 cattcgccgccaagctctctcagcaatatcacgggtagccaacgctatgtcctgatagcggtccgcccacaccagccg
 35 gccacagtcgatgaatccagaaaagcggccattttccaccatgatattcggaagcagggcatcgccatgggtcacga
 cgagatcctcgccgtcgggcatgcgcgccttgagcctggcgaacagttcggtggcgcgagccctgatgtcttctg
 tccagatcatcctgatcgacaagaccggttccatccagtagcgtgctcgctcgatgcgatgtttcgcttgggtggtc
 gaattgggcaggtagccggatcaagcgtatgcagccgcgcattgcatcagccatgatggatactttctcggcaggag
 40 caaggtgagatgacaggagatcctgccccggcacttcgcccataagcagccagtccttccccgcttcagtgacaacg
 tcgagcacagctgcgcaaggaaacgcccgtcggtggccagccacgatagccgcgtgcctcgctcctgcagttcattcag
 ggcacccggacaggtcggtcttgacaaaaagaacggggcgccccctgcgtgacagccggaacacggcggtcatcagagc
 agccgattgtctgttgtgcccagtcatagcgaatagcctctccacccaagcggccggagaaacctgcgtgcaatcca
 tcttgttcaatcatgcgaacgatcctcatcctgtctcttgatcagatcttgatccccctgcgccatcagatccttgg
 45 cggcaagaaagccatccagtttactttgcaggggttcccaaccttaccagagggcgccccagctggcaattccgggtt
 cgcttgcgtgtccataaaaccgcccagtcctagctatcgccatgtaagcccactgcaagctacctgctttctcttgcg
 cttgcgttttcccttgcagatagcccagtagctgacattcatccggggtcagcacccgtttctgcggactggcctt
 ctacgtgttccgcttccctttagcagcccttgcgcctgagtgcttgcggcagcgtgaagctaatcatgggttaaatt
 tttgttaaatacagctcattttttaaccaataggccgaatcggaacaaatcccttataaatcaaaaagaatagcccag
 50 ataggggttagtggttgcagtttgggaacaagagtcactattaaagaacgtggactccaacgtcaaagggcgaaa
 aaccgtctatcagggcgatggccggatcagcttatgcgggtgtgaaataccgcacagatgcgtaaggagaaaaatccg
 catcaggcgctcttccgcttccctcgctcactgactcgctgcgtcggtcggttcgggtgcggcgagcggatcagctc
 actcaaaaggcggtaatacgggttatccacagaatcaggggataacgcaggaagaacatgtgagcaaaaaggccagcaa
 aaggccaggaaccgtaaaaaggcgcgttgcgtggcggttttccataggctccgccccctgacgagcatcacaaaaa
 55 tcgacgctcaagtacagaggtggcgaaacccgacaggaactataaagataaccaggcggttccccctggaagctccctcg
 tgcgtctcctgttccgacctgcgcgttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttct
 catagctcacgctgtaggtatctcagttcggtgtaggtcgcttccgctccagctggggtgtgtgcacgaaccccccg
 tcagcccgaccgctgcgccttatccggtaactatcgctttaggtccaacccggtaagacacgacttatcgccactgg
 cagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaac
 60 tacggctacactagaaggacagtatgttggtatctgcgtctgctgaagccagttaccttcggaaaaagagttggtag
 ctcttgatccggcaaacaaaccacgcgtggttagcgggtgggttttttggtttgcaagcagcagattacgcgcagaaaa
 aaggatctcaagaagatcctttgatcttttctactgaacgggtgatccccaccggaattgcg

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

gccgcggaatttcgactctaggccattgcatacgttgatatctatatcataatatgtacatttatattggctcatgtc
caatatgaccgccatgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagc
5 ccatatatggagttccgcgttacataacttacggtaaatggccgcctggctgaccgcccaacgacccccgccatt
gacgtcaataatgacgtatgttcccatagtaacgccaatagggaactttccattgacgtcaatgggtggagtatttac
ggtaaaactgccacttggcagtagatcaagtgtatcatatgccaaagtcgccccctattgacgtcaatgacggtaaa
tggccgcctggcattatgccagtagatgaccttacgggactttcctacttggcagtagatctacgtattagtcac
10 cgctattaccatgggtgatgcggttttggcagtagaccaatgggcgtggatagcggtttgactcacggggatttccaa
gtctccacccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtataaac
cccgccccgttgacgcaaatgggcgttaggcgtgtacgggtgggaggtctatataagcagagctcgtttagtgaaccg
tcagatcgctggagacgccatccacgctgttttgacctccatagaagacaccgggaccgatccagcctccgcggcc
gggaacgggtgcattggaacgcggattccccgtgccaagagtgtacgtaagtaccgcctatagactctataggcacacc
cctttggctcttatgcatgctatactgttttggcgttggggcctatacacccccgcttccttatgctatagggtgatg
15 gtatagcttagcctatagggtgtgggttattgaccattattgaccactcccctattgggtgacgatactttccattact
aatccataacatggctctttggcacaactatctctatttggctatatgccaaactctgtccttcagagactgacacg
gactctgtatttttacaggatgggtcccatttattatttacaattcacatatatacaacaacgcgcgtcccccgtagcc
cgagtttttattaaacatagcgtgggtatctccacgcgaatctcggtacgtgttccggacatgggtccttctccgg
tagcggcggagcttccacatccgagccctgggtcccatgcctccagcggtcatgggtcgctcggaactccttgcctcc
20 taacagtggaggccagacttaggcacagcaccaatgccaccaccaccagtggtgccgcacaaggccgtggcggtaggg
tatgtgtctgaaaatgagctcgagattgggtcgcaccgctgacgcagatggaagacttaaggcagcggcagaaga
agatgcaggcagctgagttgttattctgataagagtcagaggtaactcccgttgcgggtgctgttaacgggtggagg
gcagtgtagtctgagcagtagctgttgcgtgcgcgcgcgcaccagacataatagctgacagactaacagactgttc
ctttccatgggtcttttctgcagtcacgcgtcgtagcgcaccatgctgcggtgctgctgtagacaccacttccact
25 gcctgctgctgtgtgcgtgtgggcccacccttgcgtgcccagcccttggagcaccctgaccgccaaccagaaccct
agcccccttgggtccaagctgacctacagcaagccccacgacgcgcgcacccttctactgcccccttctgtaccccag
ccctcccagaagccccctgcagttcagcggttccagagagtggtccaccggccctgagtgccggaacgagacactgt
acctgctgtacaaccgggagggccagacactggtggagcgggagcagcacctgggtgaaaaaagtgtatctgtatctg
agcgcccggaaccagaccatcctgcagcggtatgccagaaccgccagcaagcccagcgacggcaacgtgcagatcag
30 cgtggaggacgcaaaaatcttcggcgcccatggtgcccaagcagaccaagctgctgagattcgtgggtcaacgacg
gcaccagatatcagatgtgctgatgaagctggaaagctgggcccacgtgttccgggactactccgtgagcttccag
gtccggctgaccttcaccgagggccaacaaccagacctacaccttctgcaccaccccaacctgatcgtgtgataatc
tagaaagccatggatccggtccactacgcgttagagctcgctgatcagcctcgactgtgccttctagttgcccagc
catctgttgtttgccccctccccgtgccttccctgacctggaaggtgccactcccactgtcctttcctaataaaaat
35 gagaaattgcatcgcatgtgtctgagtaggtgtcattctattctgggggtgagatccccgcgctggaggtcatccagccg
gcgtcccggaaaacgattccgaagcccaacctttcatagaaggcggtggaatcgaaatctcgtgatggcaggtt
gggcgtcgcttgggtcggtcatttccgaacccagagtcctcgctcagaagaactcgtcaagaaggcgatagaaggcgat
gcgtgcgaatcgggagcggtcgataccgtaaagcacgaggaagcggtcagccattcgcgcgaagctcttcagcaa
40 tatcacgggtagccaacgctatgtcctgatagcggtccgccacaccagccggccacagtcgatgaatccagaaaag
cggccattttccaccatgatattcggcaagcaggtatcgccatgggtcacgacgagatcctcgccgtcgggcatgcg
cgctttagcctggcgaacagttcgggtggcgagccctgatgctcttctcggtccagatcatcctgatcgacaagac
cggcttccatccgagtagctgctcgctcgatgcgatgttctcgcttgggtggatgggcaggttagccggatcaagc
gtatgcagccgcccattgcatcagccatgatggatactttctcggcagagcaaggtgagatgacaggagatcctg
45 ccccggcacttccgccaatagcagccagtccttcccgttccagtgcacaacgtcgagcacagctgcgcaaggaaacgc
ccgtcggtggccagccacgatagccgcgtgcctcgctcctgcagttcattcagggcacccggacaggtcggtcttgaca
aaaagaaccgggcgccccctgcgtgacagccggaacacggcggtcatcagagcagccgattgtctgttgtgcccagtc
atagccgaatagcctctccaccaagcggcggaacacgtgcgtgcaatccatcttgttcaatcatgcgaacgatc
ctcatcctgtctcttgatcagatcttgatccccctgcgccatcagatccttggcggaagaaagccatccagtttact
50 ttgcagggttcccaaccttaccagagggcgccccagctggcaattccgggttcgcttgcgtgtccataaaaccgcca
gtctagctatcgccatgtaagcccactgcaagctacctgcttctcttgcgcttgcgttttcccttgtccagatag
cccagtagctgacattcatccgggtcagcacccgttctcggcgactggcttctacgtgttccgcttccctttagcag
cccttgcgcccgtgagtgcttgcggcagcgtgaagctaattcatgggttaaattttgttaaatacagctcattttttaa
ccaataggccgaaatcggaacaaatcccttataaatcaaaagaatagcccgagatagggttgagtggttccagttt
55 ggaacaagagtcactattaaagaacgtggactccaacgtcaaaggcgcaaaaaccgtctatcaggggcgatggccgg
atcagcttatgcggtgtgaaataccgcacagatgcgtgaaggagaaaaataccgcacagggcgtcttccgcttccctcg
ctcactgactcgctgcgtcggtcggttccggtgcggcgagcggtatcagctcactcaaaggcggttaatacgggttatc
cacagaatcaggggataacgcaggaagaacatgtgagcaaaaggccagcaaaaggccaggaacgtaaaaaggccg
cggttgcgtggcggttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcga
60 aaccgcagaggactataaagataccaggcggttccccctggaagctccctcgtagcgtctcctgttccgaccctgcc
gcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctca

gttcggtgtaggtcggttcgctccaagctgggctgtgtgcacgaaccccccggttcagcccgaccgctgcgcccttatcc
 ggtaactatcgctcttgagtgccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattag
 cagagcgaggtatgtaggcgggtgctacagagttcttgaaagtgggtggcctaactacggctacactagaaggacagtat
 ttggatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcacaacaaaccacc
 5 gctggtagcgggtgggtttttttgtttgcaagcagcagattacgcgcagaaaaaaggatctcaagaagatcctttgat
 cttttctactgaacggtgatccccaccggaattgcg

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

gccgcggaatttcgactctaggccattgcatacgttgtatctatatcataatatgtacatttatattggctcatgtc
 10 caatatgaccgccatgttgacattgattattgactagttattaatagtaatacaattacggggtcattagttcatagc
 ccatatatggagttccgcgttacataacttacggtaaatggccccgctggctgaccgccccacgacccccgcccatt
 gacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttac
 ggtaaactgcccacttggcagtagcatcaagtgtatcatatgccaaagtccgccccctattgacgtcaatgacggtaaa
 15 tggccccgctggcattatgcccagtagacgttacgggactttcctacttggcagtagcatctacgtatttagtcat
 cgctattaccatgggtgatgcggttttggcagtagaccaatgggcgtggatagcgggttgactcacggggatttccaa
 gtctccaccccatgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgttaataac
 cccgccccgttgacgcaaatgggcgttaggcgtgtacgggtgggaggtctatataagcagagctcgttttagtgaaccg
 tcagatcgccctggagacgccatccacgctgttttgacctccatagaaagacaccgggaccgcatccagcctccgcggcc
 20 gggaacggtgcattggaacgcggattccccgtgccaaagagtgcgtaagtaccgcctatagactctataggcacacc
 cctttggctcttatgcatgctatactgttttggcttggggcctatacacccccgcttcccttatgctatagggtgatg
 gtatagcttagcctatagggtgtgggttattgaccattattgaccactcccctattgggtgacgatactttccattact
 aatccataacatggctctttgcccacaactatctctattggctatatgccaaactctgtccttcagagactgacacg
 gactctgtatttttacaggatgggggtccattttattatttacaattcacatatacaacaacgcggtcccccggtgcc
 25 cgcagtttttattaaacatagcgtgggatctccacgcgaatctcgggtacgtgttccggacatgggctcttctccgg
 tagcggcgagcttccacatccgagccctgggtcccatgcctccagcggctcatggctcgtcggcagctccttgctcc
 taacagtggaggccagacttaggcacagcacaatgccaccaccaccagtggtgccgcacaaggccgtggcggtaggg
 tatgtgtctgaaaatgagctcggagattgggctcgcaccgctgacgcagatggaagacttaaggcagcggcagaaga
 agatgcaggcagctgagttgttgtattctgataagagtcagaggttaactcccgttgcgggtgctgttaacgggtggagg
 gcagtgtagtctgagcagtagctcgttgcgtgccgcgcgcgccaccagacataatagctgacagactaacagactgttc
 30 ctttccatgggtcttttctgcagtcaccgctcgtcgcagcgcaccatgcggctgtgcagagtggtggctgtccgtgtgcc
 tgtgtgccgtgggtgctgggcccagtgccagagagagacagccgagaagaacgactactaccgggtgccccactactgg
 gatgctgcagcagagccctgcccagaccagaccgggtacaaatacgtggagcagctcgtggacctgacctgaacta
 ccactacgacgcccagccacggcctggacaacttcgacgtgctgaagcggatcaacgtgaccgaggtgtccctgtctga
 tcagcgacttccggcggcagaacagaagaggcggcaccacaacgcggaccacctcaacgcgcgtggctctctgtggcc
 35 cctcacgcccagatccctggaattcagcgtgcggctgttcgccaaactgataatctagaaagccatggatctcggatcc
 actacgcgttagagctcgtgtagcagcctcgactgtgccttctagttgccagccatctgttgtttgccccctcccccg
 tgccttcccttgaccctggaaggtgccactcccactgtcctttccataaaaatgaggaaaattgcatcgcatgtgtctg
 agtaggtgtcattctattctggggggtgggggtggggcaggacagcaagggggaggattgggaagacaatagcagggg
 ggtggggaagaactccagcatgagatccccgcgctggaggatcatccagccggcgtcccggaacacgattccgaag
 40 cccaacctttcatagaaggcggcggtggaatcgaaatctcgtgatggcaggttgggcgtcgttgggtcggtcatttc
 gaacccagagtcctcgtcagaagaactcgtcaagaaggcgatagaaggcgatgcgctgcgaatcgggagcggcgat
 accgtaagcacgaggaagcggtcagccatttcgccgccaagctcttcagcaatatcacgggtagccaacgctatgt
 cctgatagcgggtccgccacaccagcggccacagtcgatgaatccagaaaagcggccattttccaccatgatattc
 ggcaagcaggcatcgccatgggtcacgacgagatcctcgcgcgtcgggcagtcgcgccttgagcctggcgaacagttc
 45 ggctggcgcgagcccctgatgctcttcgtccagatcatcctgatcgacaagaccggcttccatccgagtacgtgctc
 gctcgatgcgatgttttcgcttgggtgggtcgaatgggcaggttagccggatcaagcgtatgcagccgcccgcattgcatca
 gccatgatggatactttctcggcaggagcaaggtgagatgacaggagatcctgccccggcacttcgcccataagcag
 ccagtcccttcccgttccagtgcacaacgtcgcagcagcagctgcgcaaggaacgcccgtcgtggccagccacgatagcc
 gcgctgcctcgtcctgcagttcattcagggcaccggacaggtcggctcttgacaaaaagaacggggcgcccctgcgct
 50 gacagccggaacacggcggtcatcagagcagccgattgtctgttgtgcccagtcataagccgaatagcctctccaccca
 agcggcgccgagaacctgcgtgcaatccatcttgttcaatcatgcgaaacgatcctcatcctgtctcttgatcagatc
 ttgatccccctgcgcatcagatccttggcggaagaaagccatccagtttacttttgcaagggttcccaaccttacc
 gagggcgccccagctggcaattccgggttcgcttgcgtgtccataaaaacgcccagtcagctatcgccatgtaagccc
 actgcaagctacctgctttctctttgcgcttgcgttttcccttgcagatagcccagtagctgacattcatccggg
 55 gtcagcacggtttctgcggactggctttctacgtgttccgcttcccttagcagcccttgcgcccgtgagtgcttgcgg
 cagcgtgaagctaattcatgggttaattttgttaaatcagctcattttttaaccaataggccgaaatcggcacaat
 cccttataaatcaaaagaatagcccagatagggttgagtggttccagtttgggaacaagagtcactattaaaga
 acgtggactccaacgtcaaaggcgcaaaaacgctctatcagggcgatggccggatcagcttatgcgggtgtgaaatac
 cgcacagatgcgtaaggagaaaataccgcacagggcgtcttccgcttccctcgtcactgactcgtcgcgtcgggtc
 60 gttcggctgcggcgagcgggtatcagctcactcaaaggcggttaatacgggtatccacagaatcaggggataacgcagg
 aaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcttgctggcggttttccataggc

tccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagatac
 caggcggtttccccctggaagctccctcgctgcgctctcctgttccgaccctgccgcttacccggatacctgtccgcctt
 tctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcgggtgtaggtcggtcgcctca
 agctgggctgtgtgcacgaaccccccggttcagcccgaccgctgcgccttatccggtaactatcgctcttgagtccaac
 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
 LAEQFLNQVDLTETLERYQQLNNTYALVSKDLASYRSFSQQLKAQDSLGEQPTTVPPPIDLSIPHVWMPPTTPHGW
 15 TESHTTSGLRPHFNQTCILFDGHDLLFSTVTPCLHQGFYLI DELRYVKITLTEDFFVVTVSIDDDTPMLLI FGHLP
 RVLFKAPYQRDNFILRQTEKHELLVLVKKDQLNRHSYLKDPDFLDAALDFNYLDLSALLRNSFHRYAVDVLKSGRCQ
 MLDRRTVEMAFAYALALFAAARQEEAGAQVSVPRALDRQAALLQIQEFMITCLSQTTPPRTLLLYPTAVDLAKRALW
 TPNQITDITSLVRLVYILSKQNQQHLIPQWALRQIADFAKLKHLKTHLASFLSAFARQELYLMGSLVHSMVHTTERR
 EIFIVETGLCSLAELSHFTQLLAHPHHEYLSDLTPCSSSGRRDHSLERLTRLFPDATVPATVPAALSILSTMQPST
 20 LETFPDLFCLPLGESFSALTVEHSVSIYVTNQYLIKGISYPVSTTVVGQSLIITQTDSTQKCELRNMHTTHSITVA
 LNISLENCAFCQSALLEYDDTQGVINIMYMHDSDDVLFALDPYNEVVVSSPRTHYLMLLKNGTVLEVTDVVVDATD

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

RYGAEAVSEPLDKAFHLLNTYGRPIRFLRENTTQCTYNSSLRNSTVVRENAISFNFFQSYNQYYVFHMPRCLFAGP
 25 LAEQFLNQVDLTETLERYQQLNNTYALVSKDLASYRSFSQQLKAQDSLGEQPTTVPPPIDLSIPHVWMPPTTPHGW
 TESHTTSGLRPHFNQTCILFDGHDLLFSTVTPCLHQGFYLI DELRYVKITLTEDFFVVTVSIDDDTPMLLI FGHLP
 RVLFKAPYQRDNFILRQTEKHELLVLVKKDQLNRHSYLKDPDFLDAALDFNYLDLSALLRNSFHRYAVDVLKSGRCQ
 MLDRRTVEMAFAYALALFAAARQEEAGAQVSVPRALDRQAALLQIQEFMITCLSQTTPPRTLLLYPTAVDLAKRALW
 TPNQITDITSLVRLVYILSKQNQQHLIPQWALRQIADFAKLKHLKTHLASFLSAFARQELYLMGSLVHSMVHTTERR
 30 EIFIVETGLCSLAELSHFTQLLAHPHHEYLSDLTPCSSSGRRDHSLERLTRLFPDATVPATVPAALSILSTMQPST
 LETFPDLFCLPLGESFSALTVEHSVSIYVTNQYLIKGISYPVSTTVVGQSLIITQTDSTQKCELRNMHTTHSITVA
 LNISLENCAFCQSALLEYDDTQGVINIMYMHDSDDVLFALDPYNEVVVSSPRTHYLMLLKNGTVLEVTDVVVDATD
 TKLGPEQKLI SEEDLNSAVDHHHHHH

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

AAVSVAPTAAEKVPAECPELTRRCLLGEVFEGDKYESWLRPLVNVTVGRDGPLSQLIRYRPVTPEAANSVLLDEAFLD
 TLALLYNPDQLRALLTLLSSDTAPRWMTVMRGYSEC GDGSPAVYTCVDDLCRGYDLTRLSYGRSIFTEHVLGFELV
 40 PPSLFNVVVAIRNEATRTNRAVRLPVSTAAAPGITLFGYGLYNAVKEFCLRHQLDPPLLRHLDKYYAGLPPELKQTR
 VNLPAHSRYGPQAVDAR

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

CNVLVNSRGTRRSWPYTVLSYRGKEILKKQKEDILKRLMSTSSDGYRFLMYPSSQKFAHIVISMDKFPQDYILAGPI
 45 RND SITHMWFDFYSTQLRKPAKYVYSEYNHTAHKITLRPPPCGTVPSMNCLSEMLNVSKRNDTGEKGCGNFTTFNPM
 FFNVPRWNTKLYIGSNKVNVDSTIYFLGLTALLRYAQRNCTRSFYLVNAMSRLFRVPKYINGTKLKNTRMKLKR
 KQALVKEQPQKKNKKSQSTTTPYLSYTTSTAFNVTTNTVYSATAAVTRVATSTTGYRDPDSNFMKSIMATQLRDLATW
 VYTTLRYRNEPFCKPDRNRTAVSEFMKNTHVLIRNETPYTIYGTLDMSLYNETMSVENETASDNNETTPTSPSTR
 50 FQRTFIDPLWDYLDLFLDKIRNFSQLPAYGNLTPEHRRANLSTLNSLWWSQ

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

EECCFEINVNHPPERCYDFKMCNRFTVALRCPDGEVCYSPEKTAEIRGIVTTMTHSLTRQVVHNKLTSCNYPPLYE
 55 ADGRIRCGKVNDKAQYLLGAAGSVPYRWINLEYDKITRIVGLDQYLESVKKKHRLDVCRAKMGYMLQ

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)

QCQRETAEKNDYYRVPHYWDACSRALPDQTRYKYVEQLVDLTLNYHYDASHGLDNFDVLKRINVTEVSLLISDFRRQ
NRRGGTNKRRTTFNAAGSLAPHARSLEFSVRLFAN

5

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

VSSSSTRGTSATHSHSSHTTSAAHSRSGSVSQRVTSSTQTVSHGVNETIYNTTLKYGDVVGVNNTTKYPYRVCSMAQG
TDLIRFERNIVCTSMKPINEDLDEGIMVVYKRNIVAHTFKVRVYQKVLTFRRSYAYIHTTYLLGSNTEYVAPPMWEI
10 HHINSHSQCYSSYSRVIAGTVFVAYHRDSYENKTMQLMPDDYSNTHSTRYVTVKDQWHSRGSTWLYRETCNLNCMVT
ITTARSKYPYHFFATSTGDVVDISPFYNGTNRNASYFGENADKFFIFPNYTIVSDFGRPNSALETHRLVAFLERADS
VISWDIQDEKNVTCQLTFWEASERTIRSEAEDSYHFSSAKMTATFLSKKQEVNMSDSALDCVRDEAINKLQQIFNTS
YNQTYEKYGNVSVFETTGGVLVFWQGIKQKSLVELERLANRSSLNLTHNRTRKSTDGNNATHLSNMESVHNLVYAQL
QFTYDTRLGYINRALAQIAEAWCVDQRRTLEVFKELSKINPSAILSAIYNKPIAARFMGDVLGLASCVTINQTSVKV
15 LRDMNVKESPGRCYSRPVVIFNFANSSYVQYQLGEDNEILLGNHRTEECQLPSLKIFIAGNSAYEYVDYLFKRMID
LSSIISTVDSMIALDIDPLENTDFRVLELYSQKELRSSNVFDLEEIMREFNSYKQRVKYVEDKVVDPLPPYLKGLDDL
MSGLGAAGKAVGVAIGAVGGAVASVVEGVATFLKNPFGAFTIILVAIAVVIITYLIYTRQRRRLCTQPLQNLFPYLV
ADGTTVTSGSTKDTSLQAPPSYEEVSYNSGRKGPSPSSDASTAAPPYTNEQAYQMLLALARLDAEQRAQQNGTDSL
20 DGRTGTQDKGQKPNLLDRLRHRKNGYRHLKDSDEENV

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. A purified isolated human Cytomegalovirus (HCMV) pentameric complex comprising HCMV glycoprotein H (gH), HCMV glycoprotein L (gL), HCMV pUL128, HCMV pUL130, and HCMV pUL131A, wherein said gH lacks a transmembrane domain.
2. The isolated pentameric complex of claim 1, wherein said gH comprises the ectodomain of the full length gH encoded by UL 75 gene.
3. The isolated pentameric complex of claim 1 or 2, wherein:
 - (i) said gH comprises a sequence that is at least 80% identical to a sequence set forth in any one of SEQ ID NOs: 4, 6, 29 and 30;
 - (ii) said gL comprises a sequence that is at least 80% identical to a sequence set forth in any one of SEQ ID NOs: 7 and 31;
 - (iii) said pUL128 comprises a sequence that is at least 80% identical to any a sequence set forth in one of SEQ ID NOs: 15 and 33;
 - (iv) said pUL130 comprises a sequence that is at least 80% identical to a sequence set forth in any one of SEQ ID NOs: 16, 17, and 34; and/or
 - (v) said pUL131A comprises a sequence that is at least 80% identical to a sequence set forth in any one of SEQ ID NOs: 17, 18 and 35.
4. The isolated pentameric complex of any one of claims 1 to 3, wherein:
 - (i) said gH comprises a sequence set forth in any one of SEQ ID NOs: 4, 6, 29 and 30;
 - (ii) said gL comprises a sequence set forth in any one of SEQ ID NOs: 7 and 31;
 - (iii) said pUL128 comprises a sequence set forth in any one of SEQ ID NOs: 15 and 33;
 - (iv) said pUL130 comprises a sequence set forth in any one of SEQ ID NOs: 16, 17, and 34; and/or
 - (v) said pUL131A comprises a sequence set forth in any one of SEQ ID NOs: 17, 18 and 35.
5. An immunogenic composition comprising the isolated pentameric complex of any one of claims 1 to 4.
6. The immunogenic composition of claim 5, further comprising an adjuvant.
7. The immunogenic composition of claim 6, wherein said adjuvant is an oil-in-water emulsion or an aluminium salt.

8. A recombinant nucleic acid molecule which encodes human Cytomegalovirus (HCMV) glycoprotein L (gL), HCMV glycoprotein H (gH) that lacks a transmembrane domain, HCMV pUL128, HCMV pUL130 and HCMV pUL131A, wherein said recombinant nucleic acid molecule:
 - (a) is not a self-replicating RNA molecule;
 - (b) is not an alphavirus replicon;
 - (c) does not encode alphavirus non-structural proteins selected from NSP1, NSP2, NSP3 and NSP4;
 - (d) does not contain an Internal Ribosomal Entry Site (IRES) selected from EMCV and EV71; and/or
 - (e) does not contain viral 2A site FMDV.
9. A plurality of recombinant nucleic acids comprising a first construct encoding human Cytomegalovirus (HCMV) glycoprotein H (gH) that lacks a transmembrane domain and HCMV glycoprotein L (gL), and a second construct encoding HCMV pUL128, HCMV pUL130 and HCMV pUL131A, 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 alphavirus non-structural proteins selected from NSP1, NSP2, NSP3 and NSP4;
 - (d) does not contain an Internal Ribosomal Entry Site (IRES) selected from EMCV and EV71; and/or
 - (e) does not contain viral 2A site FMDV.
10. The recombinant nucleic acid molecule of claim 8 or the plurality of recombinant nucleic acids of claim 9, wherein:
 - (i) the recombinant nucleic acid molecule or plurality of recombinant nucleic acids do/es not encode any alphavirus non-structural proteins;
 - (ii) the recombinant nucleic acid molecule or plurality of recombinant nucleic acids do/es not contain any IRES;
 - (iii) the recombinant nucleic acid molecule or plurality of recombinant nucleic acids do/es not contain any viral 2A site.
11. A cell expressing the recombinant nucleic acid of claim 8 or the plurality of recombinant nucleic acids of claim 9, wherein said cell does not comprise the full human Cytomegalovirus (HCMV) genome.
12. The cell of claim 11, wherein said cell is a mammalian cell.

13. The cell of claim 12, wherein the mammalian cell is an HEK-293 cell.
14. A process for producing the isolated human Cytomegalovirus (HCMV) pentameric complex of any one of claims 1 to 4, said process comprising:
 - (i) growing the cell of any one of claims 11 to 13 in a growth medium; and
 - (ii) purifying said HCMV pentameric complex from the growth medium.
15. The process of claim 14, wherein the isolated HCMV pentameric complex is purified to a purity of >90% by mass.
16. A method of treating human Cytomegalovirus (HCMV) infection in a subject, comprising administering to the subject an immunologically effective amount of an isolated pentameric complex of any one of claims 1 to 4 or an immunogenic composition of any one of claims 5 to 7.
17. A method of raising antibodies in a mammal, comprising administering to the mammal an immunologically effective amount of an isolated pentameric complex of any one of claims 1 to 4 or an immunogenic composition of any one of claims 5 to 7.
18. Use of an isolated pentameric complex of any one of claims 1 to 4 or an immunogenic composition of any one of claims 5 to 7 or a recombinant nucleic acid molecule of claim 8 or 10 or a plurality of recombinant nucleic acids of claim 9 or 10 or a cell of any one of claims 11 to 13 in the manufacture of a medicament for the treatment of human Cytomegalovirus (HCMV) infection.
19. Use of an isolated pentameric complex of any one of claims 1 to 4 or an immunogenic composition of any one of claims 5 to 7 or a recombinant nucleic acid molecule of claim 8 or 10 or a plurality of recombinant nucleic acids of claim 9 or 10 or a cell of any one of claims 11 to 13 in the manufacture of a medicament for raising antibodies against human Cytomegalovirus (HCMV) in a mammal.

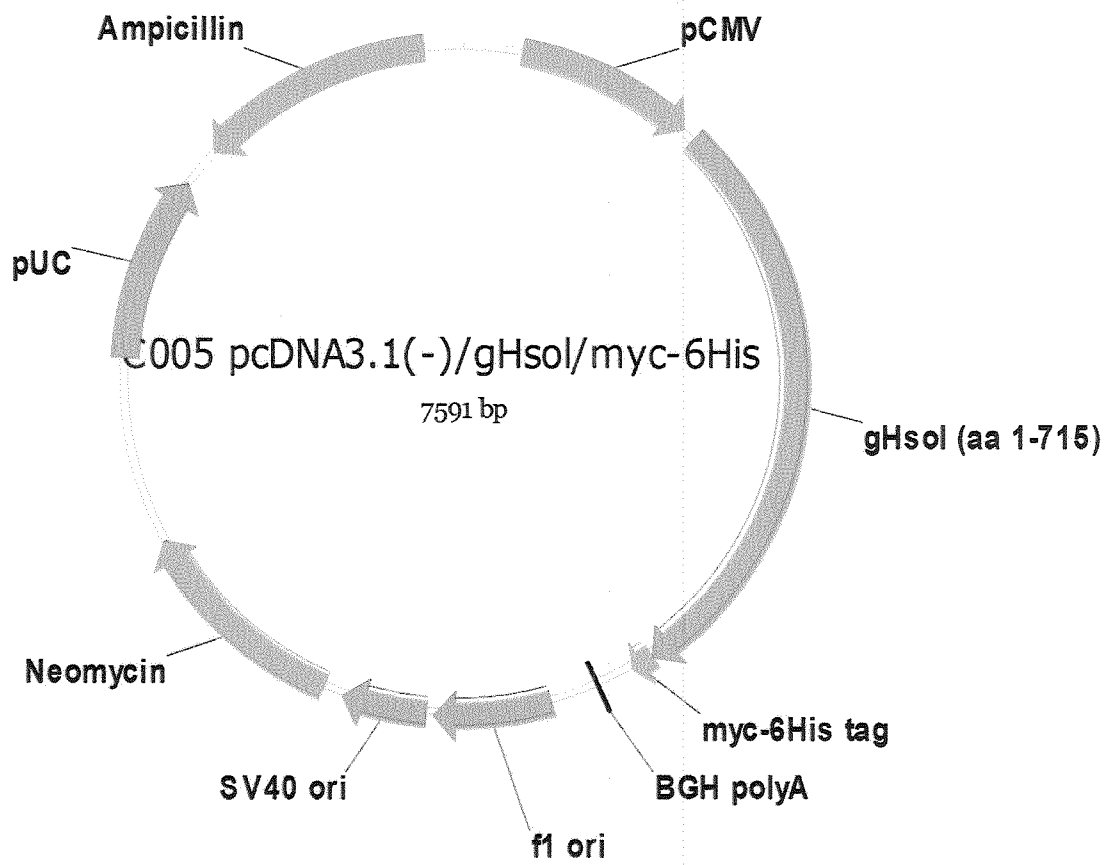


Figure 1

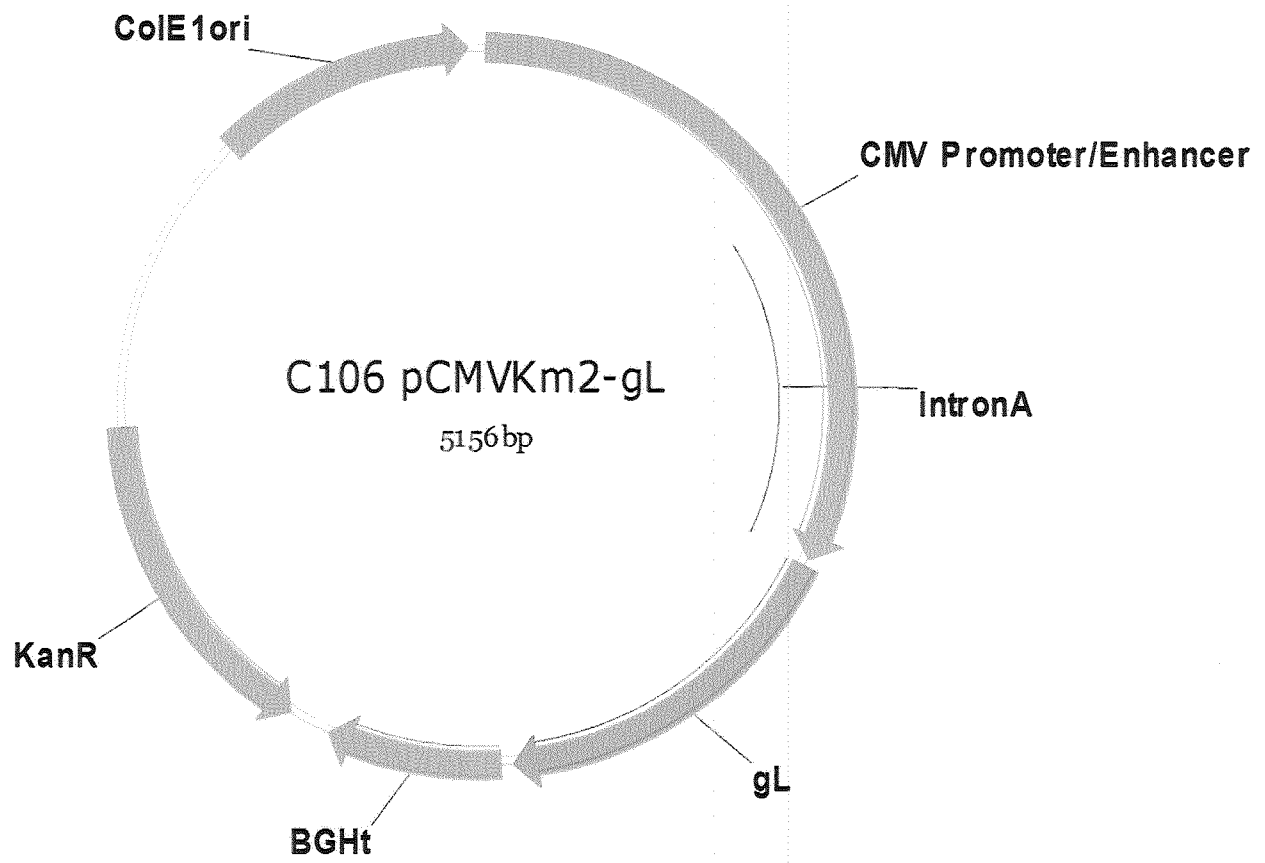


Figure 2

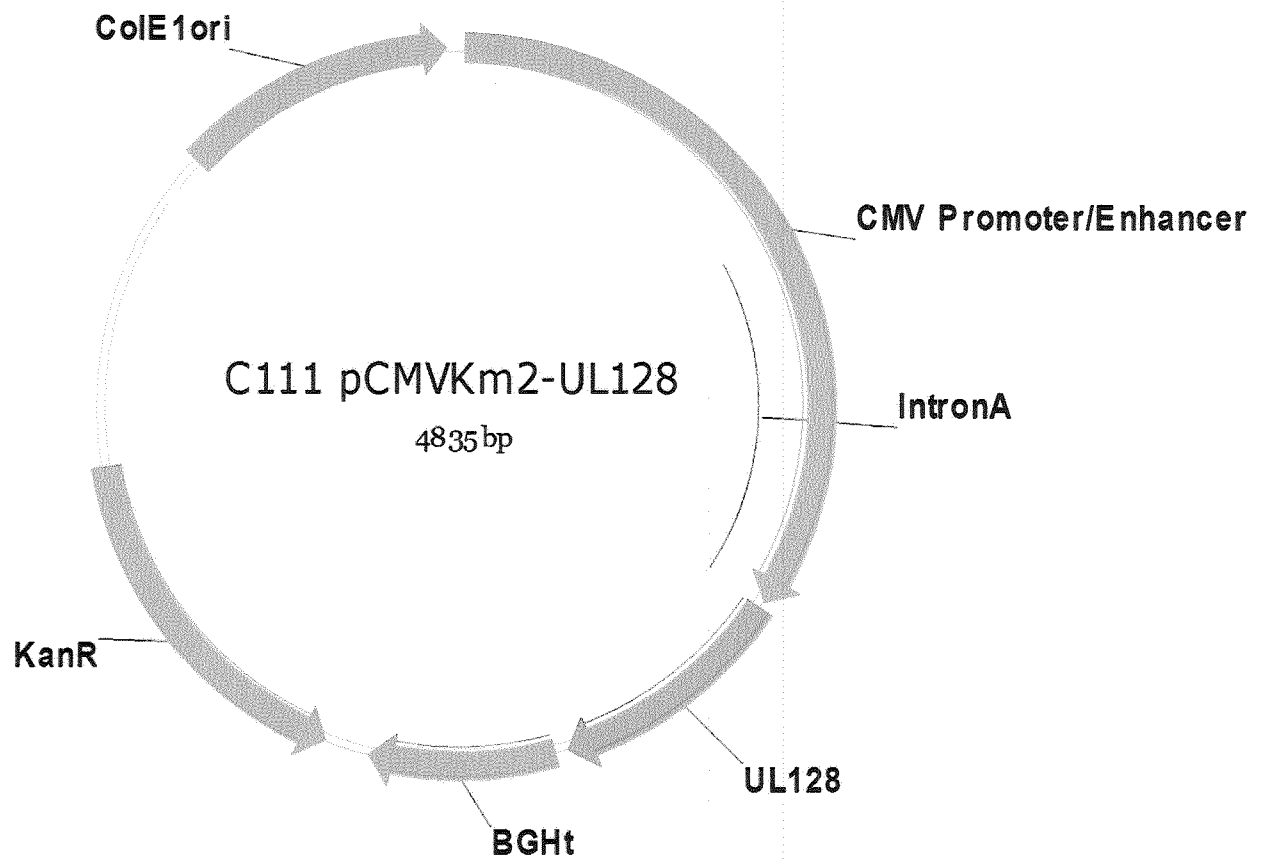


Figure 3

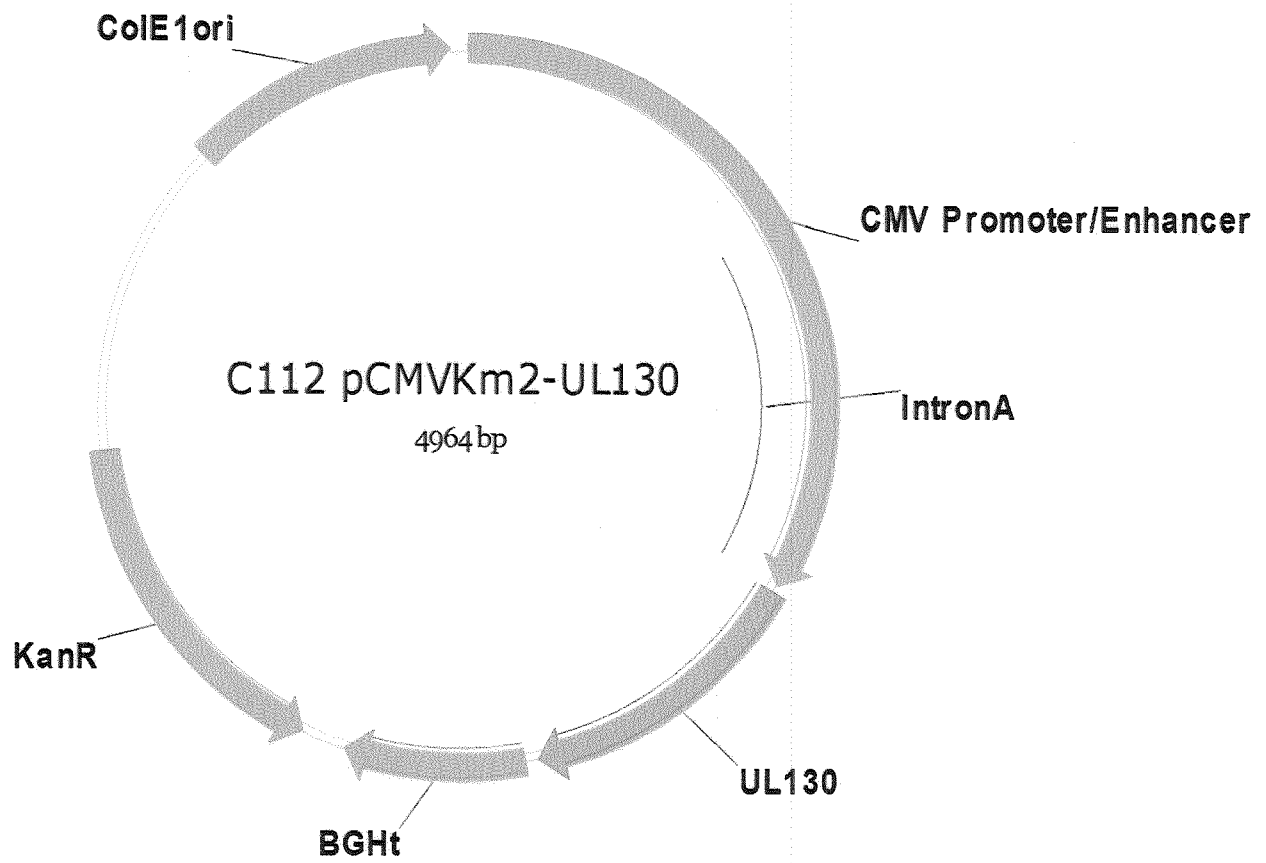


Figure 4

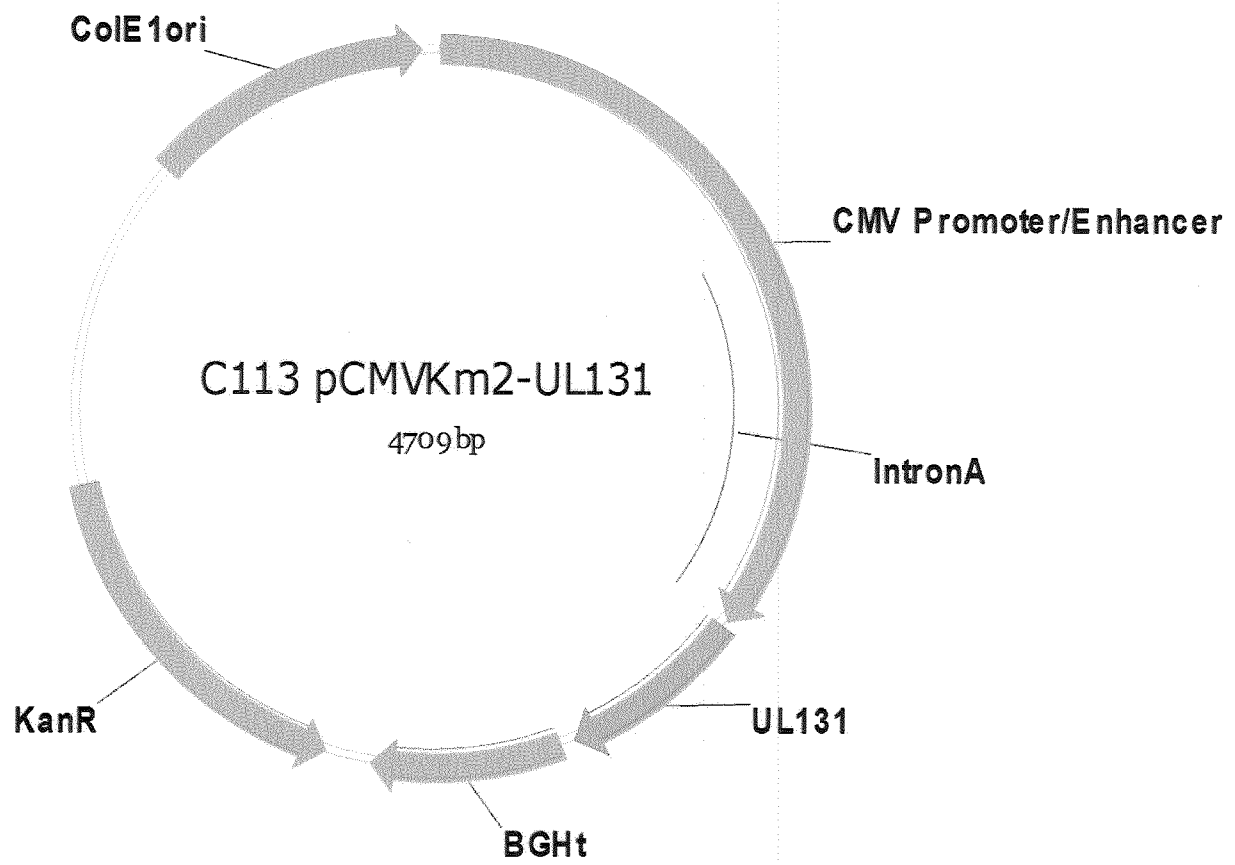


Figure 5

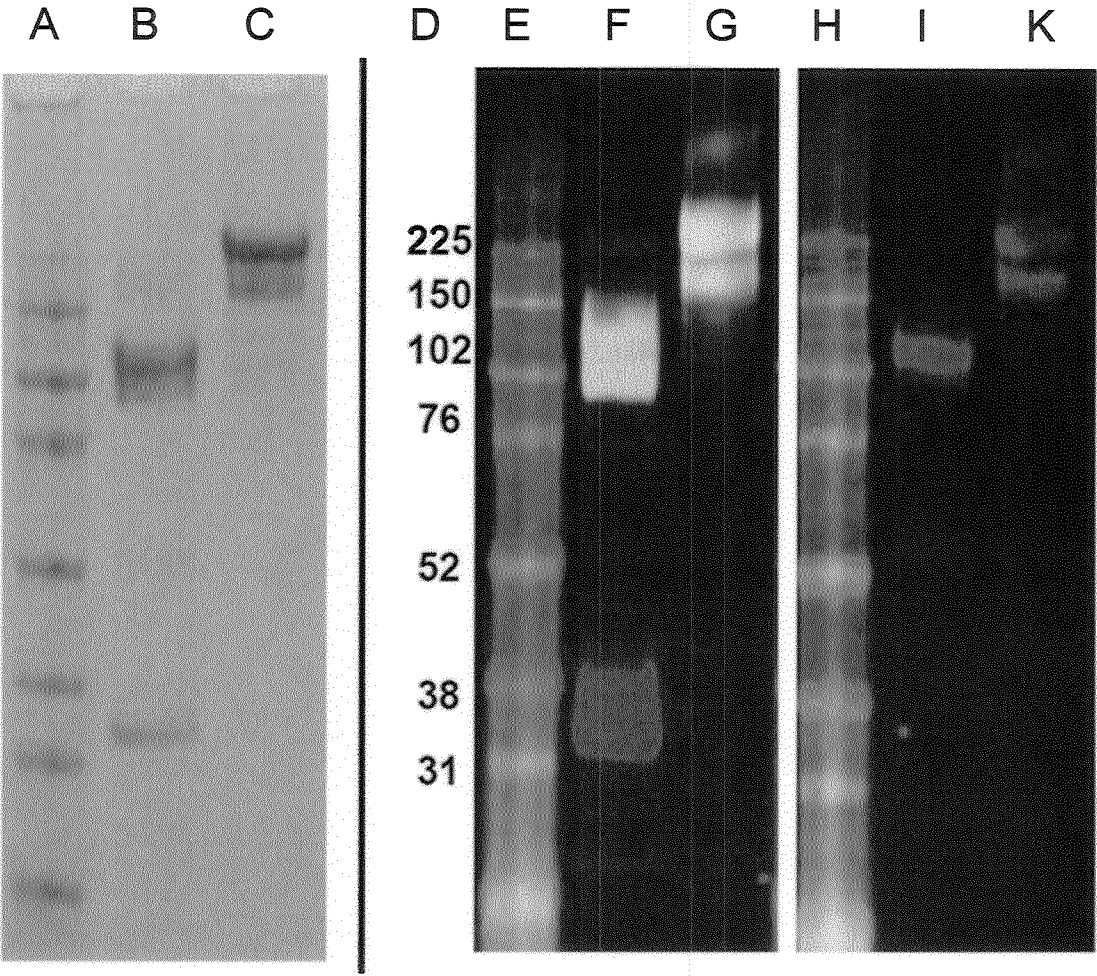


Figure 6

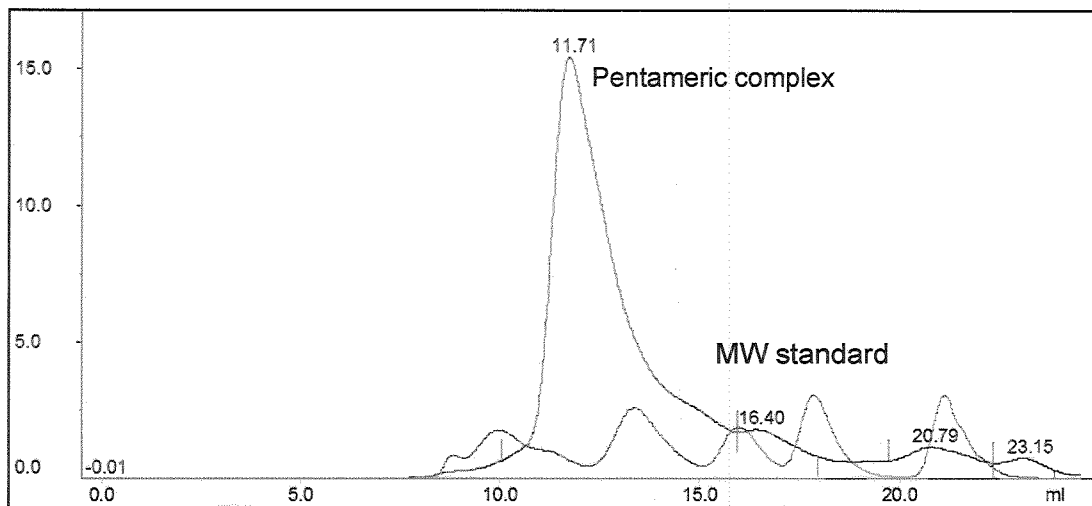


Figure 7

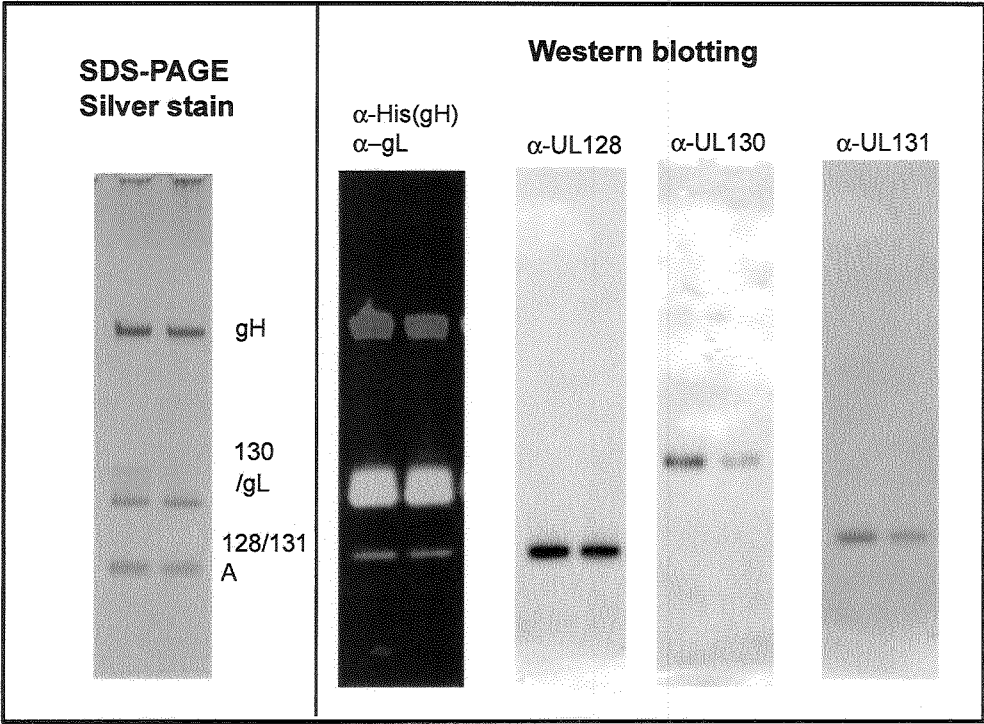


Figure 8

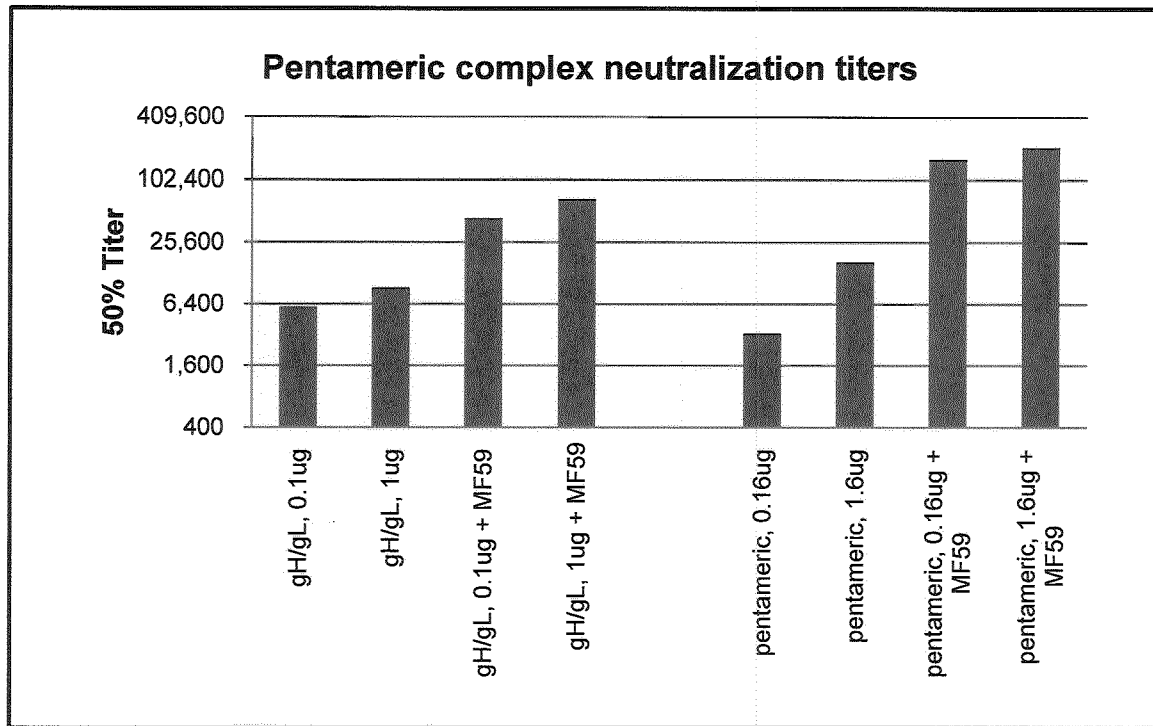


Figure 9

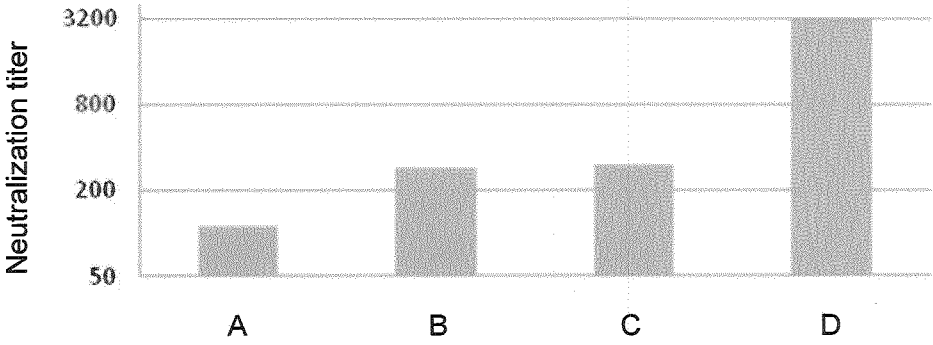


Figure 10

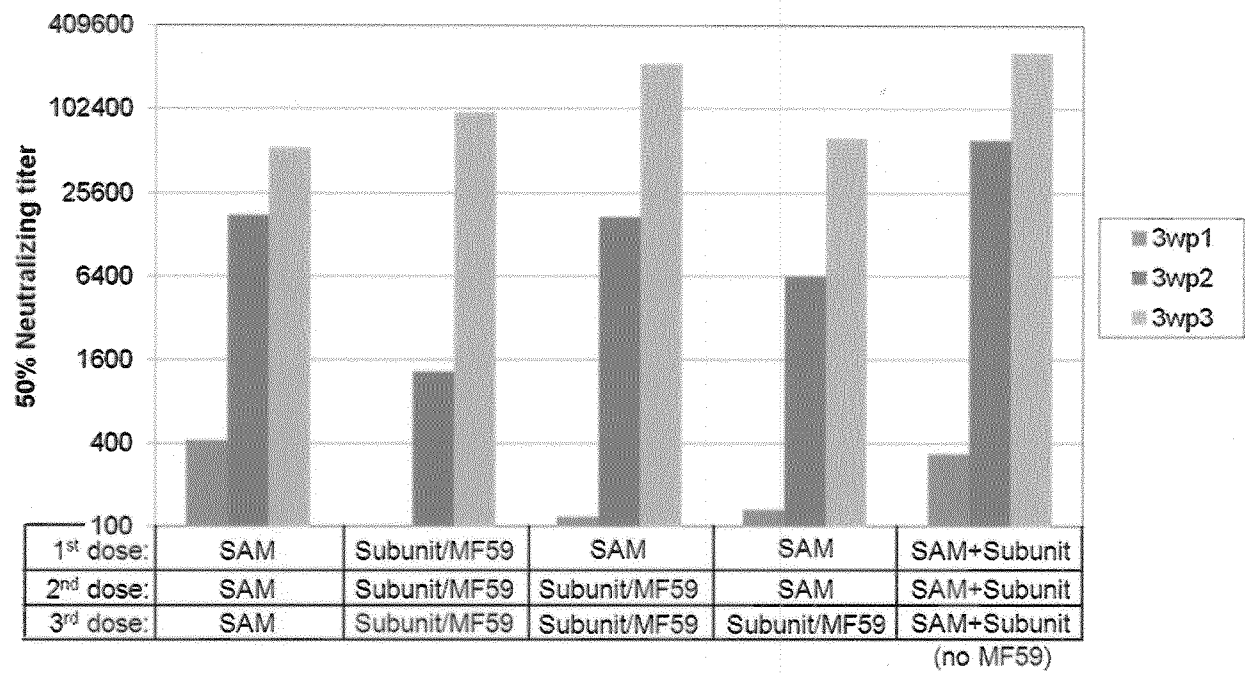


Figure 11

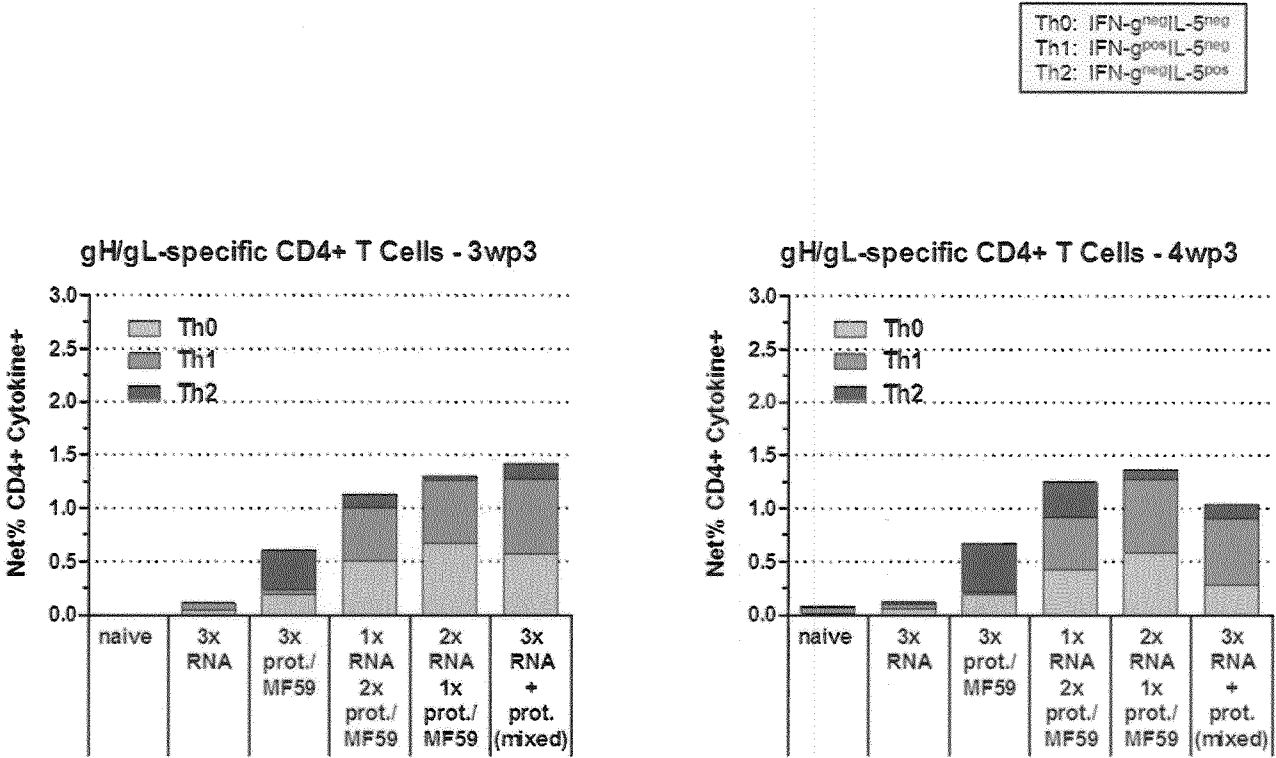


Figure 12

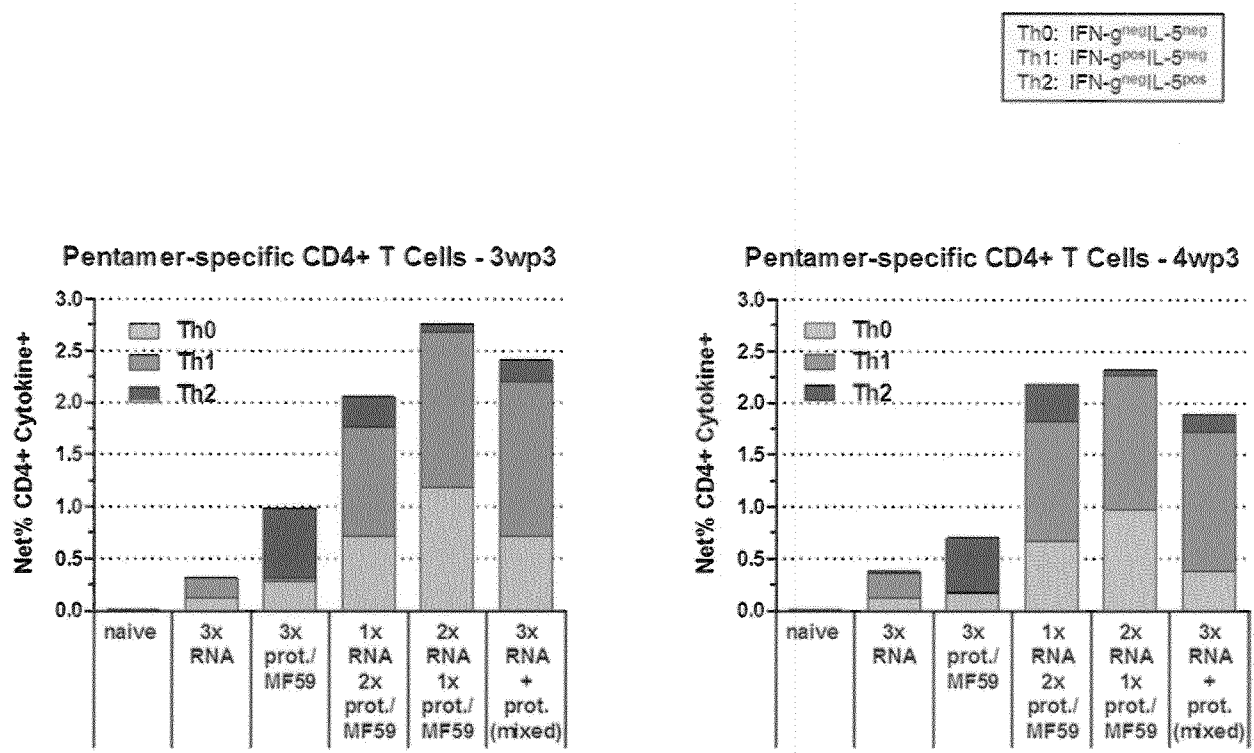


Figure 13

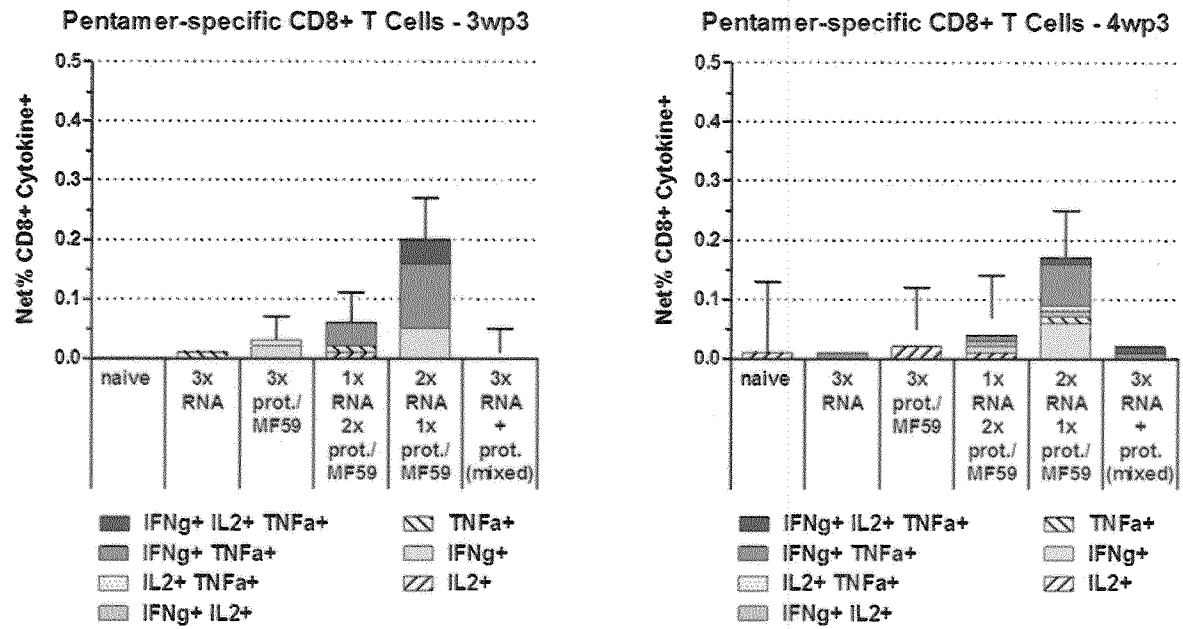
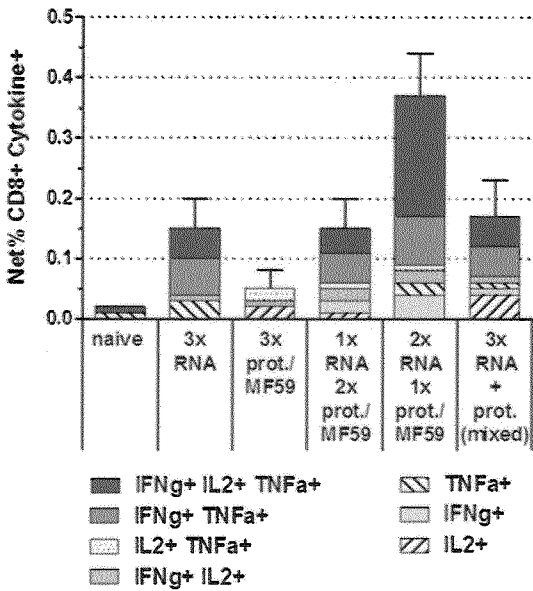


Figure 14

gH Peptide Pool 2-specific CD8+ T Cells - 3wp3



gH Peptide Pool 2-specific CD8+ T Cells - 4wp3

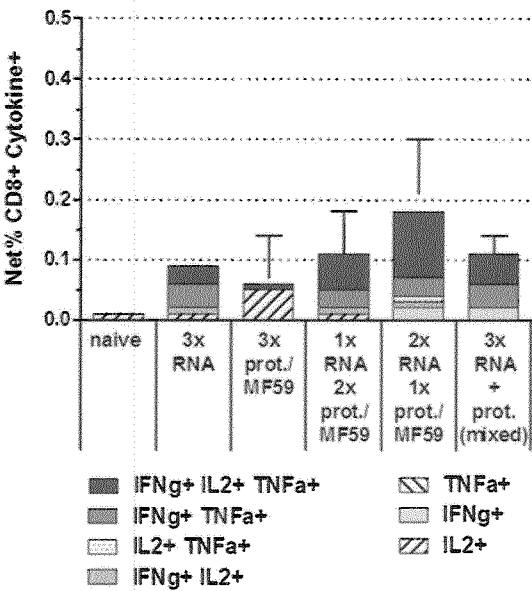


Figure 15

eolf-seql.txt
SEQUENCE LISTING

```

<110>  NOVARTIS AG
<120>  COMPLEXES OF CYTOMEGALOVIRUS PROTEINS
<130>  PAT054805-WO-PCT
<140>
<141>  2013-06-28
<150>  US 61/668,975
<151>  2012-07-06
<150>  US 61/770,257
<151>  2013-02-27
<160>  36
<170>  PatentIn version 3.5
<210>  1
<211>  742
<212>  PRT
<213>  Human cytomegalovirus, strain Merlin
<400>  1
Met Arg Pro Gly Leu Pro Ser Tyr Leu Ile Ile Leu Ala Val Cys Leu
1          5          10          15

Phe Ser His Leu Leu Ser Ser Arg Tyr Gly Ala Glu Ala Val Ser Glu
          20          25          30

Pro Leu Asp Lys Ala Phe His Leu Leu Leu Asn Thr Tyr Gly Arg Pro
          35          40          45

Ile Arg Phe Leu Arg Glu Asn Thr Thr Gln Cys Thr Tyr Asn Ser Ser
50          55          60

Leu Arg Asn Ser Thr Val Val Arg Glu Asn Ala Ile Ser Phe Asn Phe
65          70          75          80

Phe Gln Ser Tyr Asn Gln Tyr Tyr Val Phe His Met Pro Arg Cys Leu
          85          90          95

Phe Ala Gly Pro Leu Ala Glu Gln Phe Leu Asn Gln Val Asp Leu Thr
          100          105          110

Glu Thr Leu Glu Arg Tyr Gln Gln Arg Leu Asn Thr Tyr Ala Leu Val
          115          120          125

Ser Lys Asp Leu Ala Ser Tyr Arg Ser Phe Ser Gln Gln Leu Lys Ala
          130          135          140

Gln Asp Ser Leu Gly Glu Gln Pro Thr Thr Val Pro Pro Pro Ile Asp
145          150          155          160

```

eolf-seql.txt

Leu Ser Ile Pro His Val Trp Met Pro Pro Gln Thr Thr Pro His Gly
165 170 175

Trp Thr Glu Ser His Thr Thr Ser Gly Leu His Arg Pro His Phe Asn
180 185 190

Gln Thr Cys Ile Leu Phe Asp Gly His Asp Leu Leu Phe Ser Thr Val
195 200 205

Thr Pro Cys Leu His Gln Gly Phe Tyr Leu Ile Asp Glu Leu Arg Tyr
210 215 220

Val Lys Ile Thr Leu Thr Glu Asp Phe Phe Val Val Thr Val Ser Ile
225 230 235 240

Asp Asp Asp Thr Pro Met Leu Leu Ile Phe Gly His Leu Pro Arg Val
245 250 255

Leu Phe Lys Ala Pro Tyr Gln Arg Asp Asn Phe Ile Leu Arg Gln Thr
260 265 270

Glu Lys His Glu Leu Leu Val Leu Val Lys Lys Asp Gln Leu Asn Arg
275 280 285

His Ser Tyr Leu Lys Asp Pro Asp Phe Leu Asp Ala Ala Leu Asp Phe
290 295 300

Asn Tyr Leu Asp Leu Ser Ala Leu Leu Arg Asn Ser Phe His Arg Tyr
305 310 315 320

Ala Val Asp Val Leu Lys Ser Gly Arg Cys Gln Met Leu Asp Arg Arg
325 330 335

Thr Val Glu Met Ala Phe Ala Tyr Ala Leu Ala Leu Phe Ala Ala Ala
340 345 350

Arg Gln Glu Glu Ala Gly Ala Gln Val Ser Val Pro Arg Ala Leu Asp
355 360 365

Arg Gln Ala Ala Leu Leu Gln Ile Gln Glu Phe Met Ile Thr Cys Leu
370 375 380

Ser Gln Thr Pro Pro Arg Thr Thr Leu Leu Leu Tyr Pro Thr Ala Val
385 390 395 400

Asp Leu Ala Lys Arg Ala Leu Trp Thr Pro Asn Gln Ile Thr Asp Ile
405 410 415

Thr Ser Leu Val Arg Leu Val Tyr Ile Leu Ser Lys Gln Asn Gln Gln
420 425 430

eolf-seql.txt

His Leu Ile Pro Gln Trp Ala Leu Arg Gln Ile Ala Asp Phe Ala Leu
435 440 445

Lys Leu His Lys Thr His Leu Ala Ser Phe Leu Ser Ala Phe Ala Arg
450 455 460

Gln Glu Leu Tyr Leu Met Gly Ser Leu Val His Ser Met Leu Val His
465 470 475 480

Thr Thr Glu Arg Arg Glu Ile Phe Ile Val Glu Thr Gly Leu Cys Ser
485 490 495

Leu Ala Glu Leu Ser His Phe Thr Gln Leu Leu Ala His Pro His His
500 505 510

Glu Tyr Leu Ser Asp Leu Tyr Thr Pro Cys Ser Ser Ser Gly Arg Arg
515 520 525

Asp His Ser Leu Glu Arg Leu Thr Arg Leu Phe Pro Asp Ala Thr Val
530 535 540

Pro Ala Thr Val Pro Ala Ala Leu Ser Ile Leu Ser Thr Met Gln Pro
545 550 555 560

Ser Thr Leu Glu Thr Phe Pro Asp Leu Phe Cys Leu Pro Leu Gly Glu
565 570 575

Ser Phe Ser Ala Leu Thr Val Ser Glu His Val Ser Tyr Ile Val Thr
580 585 590

Asn Gln Tyr Leu Ile Lys Gly Ile Ser Tyr Pro Val Ser Thr Thr Val
595 600 605

Val Gly Gln Ser Leu Ile Ile Thr Gln Thr Asp Ser Gln Thr Lys Cys
610 615 620

Glu Leu Thr Arg Asn Met His Thr Thr His Ser Ile Thr Val Ala Leu
625 630 635 640

Asn Ile Ser Leu Glu Asn Cys Ala Phe Cys Gln Ser Ala Leu Leu Glu
645 650 655

Tyr Asp Asp Thr Gln Gly Val Ile Asn Ile Met Tyr Met His Asp Ser
660 665 670

Asp Asp Val Leu Phe Ala Leu Asp Pro Tyr Asn Glu Val Val Val Ser
675 680 685

Ser Pro Arg Thr His Tyr Leu Met Leu Leu Lys Asn Gly Thr Val Leu
690 695 700

eolf-seql.txt

Glu Val Thr Asp Val Val Val Asp Ala Thr Asp Ser Arg Leu Leu Met
705 710 715 720

Met Ser Val Tyr Ala Leu Ser Ala Ile Ile Gly Ile Tyr Leu Leu Tyr
725 730 735

Arg Met Leu Lys Thr Cys
740

<210> 2
<211> 742
<212> PRT
<213> Human cytomegalovirus, strain Towne

<400> 2

Met Arg Pro Gly Leu Pro Ser Tyr Leu Ile Val Leu Ala Val Cys Leu
1 5 10 15

Leu Ser His Leu Leu Ser Ser Arg Tyr Gly Ala Glu Ala Ile Ser Glu
20 25 30

Pro Leu Asp Lys Ala Phe His Leu Leu Leu Asn Thr Tyr Gly Arg Pro
35 40 45

Ile Arg Phe Leu Arg Glu Asn Thr Thr Gln Cys Thr Tyr Asn Ser Ser
50 55 60

Leu Arg Asn Ser Thr Val Val Arg Glu Asn Ala Ile Ser Phe Asn Phe
65 70 75 80

Phe Gln Ser Tyr Asn Gln Tyr Tyr Val Phe His Met Pro Arg Cys Leu
85 90 95

Phe Ala Gly Pro Leu Ala Glu Gln Phe Leu Asn Gln Val Asp Leu Thr
100 105 110

Glu Thr Leu Glu Arg Tyr Gln Gln Arg Leu Asn Thr Tyr Ala Leu Val
115 120 125

Ser Lys Asp Leu Ala Ser Tyr Arg Ser Phe Ser Gln Gln Leu Lys Ala
130 135 140

Gln Asp Ser Leu Gly Glu Gln Pro Thr Thr Val Pro Pro Pro Ile Asp
145 150 155 160

Leu Ser Ile Pro His Val Trp Met Pro Pro Gln Thr Thr Pro His Gly
165 170 175

Trp Thr Glu Ser His Thr Thr Ser Gly Leu His Arg Pro His Phe Asn
180 185 190

Gln Thr Cys Ile Leu Phe Asp Gly His Asp Leu Leu Phe Ser Thr Val

195

200

205

Thr Pro Cys Leu His Gln Gly Phe Tyr Leu Ile Asp Glu Leu Arg Tyr
 210 215 220

Val Lys Ile Thr Leu Thr Glu Asp Phe Phe Val Val Thr Val Ser Ile
 225 230 235 240

Asp Asp Asp Thr Pro Met Leu Leu Ile Phe Gly His Leu Pro Arg Val
 245 250 255

Leu Phe Lys Ala Pro Tyr Gln Arg Asp Asn Phe Ile Leu Arg Gln Thr
 260 265 270

Glu Lys His Glu Leu Leu Val Leu Val Lys Lys Asp Gln Leu Asn Arg
 275 280 285

His Ser Tyr Leu Lys Asp Pro Asp Phe Leu Asp Ala Ala Leu Asp Phe
 290 295 300

Asn Tyr Leu Asp Leu Ser Ala Leu Leu Arg Asn Ser Phe His Arg Tyr
 305 310 315 320

Ala Val Asp Val Leu Lys Ser Gly Arg Cys Gln Met Leu Asp Arg Arg
 325 330 335

Thr Val Glu Met Ala Phe Ala Tyr Ala Leu Ala Leu Phe Ala Ala Ala
 340 345 350

Arg Gln Glu Glu Ala Gly Ala Gln Val Ser Val Pro Arg Ala Leu Asp
 355 360 365

Arg Gln Ala Ala Leu Leu Gln Ile Gln Glu Phe Met Ile Thr Cys Leu
 370 375 380

Ser Gln Thr Pro Pro Arg Thr Thr Leu Leu Leu Tyr Pro Thr Ala Val
 385 390 395 400

Asp Leu Ala Lys Arg Ala Leu Trp Thr Pro Asn Gln Ile Thr Asp Ile
 405 410 415

Thr Ser Leu Val Arg Leu Val Tyr Ile Leu Ser Lys Gln Asn Gln Gln
 420 425 430

His Leu Ile Pro Gln Trp Ala Leu Arg Gln Ile Ala Asp Phe Ala Leu
 435 440 445

Lys Leu His Lys Thr His Leu Ala Ser Phe Leu Ser Ala Phe Ala Arg
 450 455 460

Gln Glu Leu Tyr Leu Met Gly Ser Leu Val His Ser Met Leu Val His
 Page 5

eolf-seql.txt

465 470 475 480
 Thr Thr Glu Arg Arg Glu Ile Phe Ile Val Glu Thr Gly Leu Cys Ser
 485 490 495
 Leu Ala Glu Leu Ser His Phe Thr Gln Leu Leu Ala His Pro His His
 500 505 510
 Glu Tyr Leu Ser Asp Leu Tyr Thr Pro Cys Ser Ser Ser Gly Arg Arg
 515 520 525
 Asp His Ser Leu Glu Arg Leu Thr Arg Leu Phe Pro Asp Ala Thr Val
 530 535 540
 Pro Thr Thr Val Pro Ala Ala Leu Ser Ile Leu Ser Thr Met Gln Pro
 545 550 555 560
 Ser Thr Leu Glu Thr Phe Pro Asp Leu Phe Cys Leu Pro Leu Gly Glu
 565 570 575
 Ser Phe Ser Ala Leu Thr Val Ser Glu His Val Ser Tyr Val Val Thr
 580 585 590
 Asn Gln Tyr Leu Ile Lys Gly Ile Ser Tyr Pro Val Ser Thr Thr Val
 595 600 605
 Val Gly Gln Ser Leu Ile Ile Thr Gln Thr Asp Ser Gln Thr Lys Cys
 610 615 620
 Glu Leu Thr Arg Asn Met His Thr Thr His Ser Ile Thr Ala Ala Leu
 625 630 635 640
 Asn Ile Ser Leu Glu Asn Cys Ala Phe Cys Gln Ser Ala Leu Leu Glu
 645 650 655
 Tyr Asp Asp Thr Gln Gly Val Ile Asn Ile Met Tyr Met His Asp Ser
 660 665 670
 Asp Asp Val Leu Phe Ala Leu Asp Pro Tyr Asn Glu Val Val Val Ser
 675 680 685
 Ser Pro Arg Thr His Tyr Leu Met Leu Leu Lys Asn Gly Thr Val Leu
 690 695 700
 Glu Val Thr Asp Val Val Val Asp Ala Thr Asp Ser Arg Leu Leu Met
 705 710 715 720
 Met Ser Val Tyr Ala Leu Ser Ala Ile Ile Gly Ile Tyr Leu Leu Tyr
 725 730 735

Arg Met Leu Lys Thr Cys

740

<210> 3
 <211> 743
 <212> PRT
 <213> Human cytomegalovirus, strain AD169

<400> 3

Met Arg Pro Gly Leu Pro Pro Tyr Leu Thr Val Phe Thr Val Tyr Leu
 1 5 10 15

Leu Ser His Leu Pro Ser Gln Arg Tyr Gly Ala Asp Ala Ala Ser Glu
 20 25 30

Ala Leu Asp Pro His Ala Phe His Leu Leu Leu Asn Thr Tyr Gly Arg
 35 40 45

Pro Ile Arg Phe Leu Arg Glu Asn Thr Thr Gln Cys Thr Tyr Asn Ser
 50 55 60

Ser Leu Arg Asn Ser Thr Val Val Arg Glu Asn Ala Ile Ser Phe Asn
 65 70 75 80

Phe Phe Gln Ser Tyr Asn Gln Tyr Tyr Val Phe His Met Pro Arg Cys
 85 90 95

Leu Phe Ala Gly Pro Leu Ala Glu Gln Phe Leu Asn Gln Val Asp Leu
 100 105 110

Thr Glu Thr Leu Glu Arg Tyr Gln Gln Arg Leu Asn Thr Tyr Ala Leu
 115 120 125

Val Ser Lys Asp Leu Ala Ser Tyr Arg Ser Phe Ser Gln Gln Leu Lys
 130 135 140

Ala Gln Asp Ser Leu Gly Gln Gln Pro Thr Thr Val Pro Pro Pro Ile
 145 150 155 160

Asp Leu Ser Ile Pro His Val Trp Met Pro Pro Gln Thr Thr Pro His
 165 170 175

Asp Trp Lys Gly Ser His Thr Thr Ser Gly Leu His Arg Pro His Phe
 180 185 190

Asn Gln Thr Cys Ile Leu Phe Asp Gly His Asp Leu Leu Phe Ser Thr
 195 200 205

Val Thr Pro Cys Leu His Gln Gly Phe Tyr Leu Met Asp Glu Leu Arg
 210 215 220

Tyr Val Lys Ile Thr Leu Thr Glu Asp Phe Phe Val Val Thr Val Ser
 225 230 235 240

eolf-seql.txt

Ile Asp Asp Asp Thr Pro Met Leu Leu Ile Phe Gly His Leu Pro Arg
245 250 255

Val Leu Phe Lys Ala Pro Tyr Gln Arg Asp Asn Phe Ile Leu Arg Gln
260 265 270

Thr Glu Lys His Glu Leu Leu Val Leu Val Lys Lys Ala Gln Leu Asn
275 280 285

Arg His Ser Tyr Leu Lys Asp Ser Asp Phe Leu Asp Ala Ala Leu Asp
290 295 300

Phe Asn Tyr Leu Asp Leu Ser Ala Leu Leu Arg Asn Ser Phe His Arg
305 310 315 320

Tyr Ala Val Asp Val Leu Lys Ser Gly Arg Cys Gln Met Leu Asp Arg
325 330 335

Arg Thr Val Glu Met Ala Phe Ala Tyr Ala Leu Ala Leu Phe Ala Ala
340 345 350

Ala Arg Gln Glu Glu Ala Gly Thr Glu Ile Ser Ile Pro Arg Ala Leu
355 360 365

Asp Arg Gln Ala Ala Leu Leu Gln Ile Gln Glu Phe Met Ile Thr Cys
370 375 380

Leu Ser Gln Thr Pro Pro Arg Thr Thr Leu Leu Leu Tyr Pro Thr Ala
385 390 395 400

Val Asp Leu Ala Lys Arg Ala Leu Trp Thr Pro Asp Gln Ile Thr Asp
405 410 415

Ile Thr Ser Leu Val Arg Leu Val Tyr Ile Leu Ser Lys Gln Asn Gln
420 425 430

Gln His Leu Ile Pro Gln Trp Ala Leu Arg Gln Ile Ala Asp Phe Ala
435 440 445

Leu Gln Leu His Lys Thr His Leu Ala Ser Phe Leu Ser Ala Phe Ala
450 455 460

Arg Gln Glu Leu Tyr Leu Met Gly Ser Leu Val His Ser Met Leu Val
465 470 475 480

His Thr Thr Glu Arg Arg Glu Ile Phe Ile Val Glu Thr Gly Leu Cys
485 490 495

Ser Leu Ala Glu Leu Ser His Phe Thr Gln Leu Leu Ala His Pro His
500 505 510

eolf-seql.txt

His Glu Tyr Leu Ser Asp Leu Tyr Thr Pro Cys Ser Ser Ser Gly Arg
515 520 525

Arg Asp His Ser Leu Glu Arg Leu Thr Arg Leu Phe Pro Asp Ala Thr
530 535 540

Val Pro Ala Thr Val Pro Ala Ala Leu Ser Ile Leu Ser Thr Met Gln
545 550 555 560

Pro Ser Thr Leu Glu Thr Phe Pro Asp Leu Phe Cys Leu Pro Leu Gly
565 570 575

Glu Ser Phe Ser Ala Leu Thr Val Ser Glu His Val Ser Tyr Val Val
580 585 590

Thr Asn Gln Tyr Leu Ile Lys Gly Ile Ser Tyr Pro Val Ser Thr Thr
595 600 605

Val Val Gly Gln Ser Leu Ile Ile Thr Gln Thr Asp Ser Gln Thr Lys
610 615 620

Cys Glu Leu Thr Arg Asn Met His Thr Thr His Ser Ile Thr Ala Ala
625 630 635 640

Leu Asn Ile Ser Leu Glu Asn Cys Ala Phe Cys Gln Ser Ala Leu Leu
645 650 655

Glu Tyr Asp Asp Thr Gln Gly Val Ile Asn Ile Met Tyr Met His Asp
660 665 670

Ser Asp Asp Val Leu Phe Ala Leu Asp Pro Tyr Asn Glu Val Val Val
675 680 685

Ser Ser Pro Arg Thr His Tyr Leu Met Leu Leu Lys Asn Gly Thr Val
690 695 700

Leu Glu Val Thr Asp Val Val Val Asp Ala Thr Asp Ser Arg Leu Leu
705 710 715 720

Met Met Ser Val Tyr Ala Leu Ser Ala Ile Ile Gly Ile Tyr Leu Leu
725 730 735

Tyr Arg Met Leu Lys Thr Cys
740

<210> 4
<211> 715
<212> PRT
<213> Human cytomegalovirus, strain Merlin
<400> 4

eolf-seql.txt

Met Arg Pro Gly Leu Pro Ser Tyr Leu Ile Ile Leu Ala Val Cys Leu
1 5 10 15

Phe Ser His Leu Leu Ser Ser Arg Tyr Gly Ala Glu Ala Val Ser Glu
20 25 30

Pro Leu Asp Lys Ala Phe His Leu Leu Leu Asn Thr Tyr Gly Arg Pro
35 40 45

Ile Arg Phe Leu Arg Glu Asn Thr Thr Gln Cys Thr Tyr Asn Ser Ser
50 55 60

Leu Arg Asn Ser Thr Val Val Arg Glu Asn Ala Ile Ser Phe Asn Phe
65 70 75 80

Phe Gln Ser Tyr Asn Gln Tyr Tyr Val Phe His Met Pro Arg Cys Leu
85 90 95

Phe Ala Gly Pro Leu Ala Glu Gln Phe Leu Asn Gln Val Asp Leu Thr
100 105 110

Glu Thr Leu Glu Arg Tyr Gln Gln Arg Leu Asn Thr Tyr Ala Leu Val
115 120 125

Ser Lys Asp Leu Ala Ser Tyr Arg Ser Phe Ser Gln Gln Leu Lys Ala
130 135 140

Gln Asp Ser Leu Gly Glu Gln Pro Thr Thr Val Pro Pro Pro Ile Asp
145 150 155 160

Leu Ser Ile Pro His Val Trp Met Pro Pro Gln Thr Thr Pro His Gly
165 170 175

Trp Thr Glu Ser His Thr Thr Ser Gly Leu His Arg Pro His Phe Asn
180 185 190

Gln Thr Cys Ile Leu Phe Asp Gly His Asp Leu Leu Phe Ser Thr Val
195 200 205

Thr Pro Cys Leu His Gln Gly Phe Tyr Leu Ile Asp Glu Leu Arg Tyr
210 215 220

Val Lys Ile Thr Leu Thr Glu Asp Phe Phe Val Val Thr Val Ser Ile
225 230 235 240

Asp Asp Asp Thr Pro Met Leu Leu Ile Phe Gly His Leu Pro Arg Val
245 250 255

Leu Phe Lys Ala Pro Tyr Gln Arg Asp Asn Phe Ile Leu Arg Gln Thr
260 265 270

eolf-seql.txt

Glu Lys His Glu Leu Leu Val Leu Val Lys Lys Asp Gln Leu Asn Arg
275 280 285

His Ser Tyr Leu Lys Asp Pro Asp Phe Leu Asp Ala Ala Leu Asp Phe
290 295 300

Asn Tyr Leu Asp Leu Ser Ala Leu Leu Arg Asn Ser Phe His Arg Tyr
305 310 315 320

Ala Val Asp Val Leu Lys Ser Gly Arg Cys Gln Met Leu Asp Arg Arg
325 330 335

Thr Val Glu Met Ala Phe Ala Tyr Ala Leu Ala Leu Phe Ala Ala Ala
340 345 350

Arg Gln Glu Glu Ala Gly Ala Gln Val Ser Val Pro Arg Ala Leu Asp
355 360 365

Arg Gln Ala Ala Leu Leu Gln Ile Gln Glu Phe Met Ile Thr Cys Leu
370 375 380

Ser Gln Thr Pro Pro Arg Thr Thr Leu Leu Leu Tyr Pro Thr Ala Val
385 390 395 400

Asp Leu Ala Lys Arg Ala Leu Trp Thr Pro Asn Gln Ile Thr Asp Ile
405 410 415

Thr Ser Leu Val Arg Leu Val Tyr Ile Leu Ser Lys Gln Asn Gln Gln
420 425 430

His Leu Ile Pro Gln Trp Ala Leu Arg Gln Ile Ala Asp Phe Ala Leu
435 440 445

Lys Leu His Lys Thr His Leu Ala Ser Phe Leu Ser Ala Phe Ala Arg
450 455 460

Gln Glu Leu Tyr Leu Met Gly Ser Leu Val His Ser Met Leu Val His
465 470 475 480

Thr Thr Glu Arg Arg Glu Ile Phe Ile Val Glu Thr Gly Leu Cys Ser
485 490 495

Leu Ala Glu Leu Ser His Phe Thr Gln Leu Leu Ala His Pro His His
500 505 510

Glu Tyr Leu Ser Asp Leu Tyr Thr Pro Cys Ser Ser Ser Gly Arg Arg
515 520 525

Asp His Ser Leu Glu Arg Leu Thr Arg Leu Phe Pro Asp Ala Thr Val
530 535 540

eolf-seql.txt

Pro Ala Thr Val Pro Ala Ala Leu Ser Ile Leu Ser Thr Met Gln Pro
545 550 555 560

Ser Thr Leu Glu Thr Phe Pro Asp Leu Phe Cys Leu Pro Leu Gly Glu
565 570 575

Ser Phe Ser Ala Leu Thr Val Ser Glu His Val Ser Tyr Ile Val Thr
580 585 590

Asn Gln Tyr Leu Ile Lys Gly Ile Ser Tyr Pro Val Ser Thr Thr Val
595 600 605

Val Gly Gln Ser Leu Ile Ile Thr Gln Thr Asp Ser Gln Thr Lys Cys
610 615 620

Glu Leu Thr Arg Asn Met His Thr Thr His Ser Ile Thr Val Ala Leu
625 630 635 640

Asn Ile Ser Leu Glu Asn Cys Ala Phe Cys Gln Ser Ala Leu Leu Glu
645 650 655

Tyr Asp Asp Thr Gln Gly Val Ile Asn Ile Met Tyr Met His Asp Ser
660 665 670

Asp Asp Val Leu Phe Ala Leu Asp Pro Tyr Asn Glu Val Val Val Ser
675 680 685

Ser Pro Arg Thr His Tyr Leu Met Leu Leu Lys Asn Gly Thr Val Leu
690 695 700

Glu Val Thr Asp Val Val Val Asp Ala Thr Asp
705 710 715

<210> 5
<211> 27
<212> PRT
<213> Artificial sequence

<220>
<223> C-terminal extension including myc-tag and his-tag

<400> 5

Gly Thr Lys Leu Gly Pro Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu
1 5 10 15

Asn Ser Ala Val Asp His His His His His His
20 25

<210> 6
<211> 742
<212> PRT
<213> Human cytomegalovirus

eolf-seql.txt

<400> 6

Met Arg Pro Gly Leu Pro Ser Tyr Leu Ile Ile Leu Ala Val Cys Leu
 1 5 10 15

Phe Ser His Leu Leu Ser Ser Arg Tyr Gly Ala Glu Ala Val Ser Glu
 20 25 30

Pro Leu Asp Lys Ala Phe His Leu Leu Leu Asn Thr Tyr Gly Arg Pro
 35 40 45

Ile Arg Phe Leu Arg Glu Asn Thr Thr Gln Cys Thr Tyr Asn Ser Ser
 50 55 60

Leu Arg Asn Ser Thr Val Val Arg Glu Asn Ala Ile Ser Phe Asn Phe
 65 70 75 80

Phe Gln Ser Tyr Asn Gln Tyr Tyr Val Phe His Met Pro Arg Cys Leu
 85 90 95

Phe Ala Gly Pro Leu Ala Glu Gln Phe Leu Asn Gln Val Asp Leu Thr
 100 105 110

Glu Thr Leu Glu Arg Tyr Gln Gln Arg Leu Asn Thr Tyr Ala Leu Val
 115 120 125

Ser Lys Asp Leu Ala Ser Tyr Arg Ser Phe Ser Gln Gln Leu Lys Ala
 130 135 140

Gln Asp Ser Leu Gly Glu Gln Pro Thr Thr Val Pro Pro Pro Ile Asp
 145 150 155 160

Leu Ser Ile Pro His Val Trp Met Pro Pro Gln Thr Thr Pro His Gly
 165 170 175

Trp Thr Glu Ser His Thr Thr Ser Gly Leu His Arg Pro His Phe Asn
 180 185 190

Gln Thr Cys Ile Leu Phe Asp Gly His Asp Leu Leu Phe Ser Thr Val
 195 200 205

Thr Pro Cys Leu His Gln Gly Phe Tyr Leu Ile Asp Glu Leu Arg Tyr
 210 215 220

Val Lys Ile Thr Leu Thr Glu Asp Phe Phe Val Val Thr Val Ser Ile
 225 230 235 240

Asp Asp Asp Thr Pro Met Leu Leu Ile Phe Gly His Leu Pro Arg Val
 245 250 255

Leu Phe Lys Ala Pro Tyr Gln Arg Asp Asn Phe Ile Leu Arg Gln Thr
 260 265 270

eolf-seql.txt

Glu Lys His Glu Leu Leu Val Leu Val Lys Lys Asp Gln Leu Asn Arg
275 280 285

His Ser Tyr Leu Lys Asp Pro Asp Phe Leu Asp Ala Ala Leu Asp Phe
290 295 300

Asn Tyr Leu Asp Leu Ser Ala Leu Leu Arg Asn Ser Phe His Arg Tyr
305 310 315 320

Ala Val Asp Val Leu Lys Ser Gly Arg Cys Gln Met Leu Asp Arg Arg
325 330 335

Thr Val Glu Met Ala Phe Ala Tyr Ala Leu Ala Leu Phe Ala Ala Ala
340 345 350

Arg Gln Glu Glu Ala Gly Ala Gln Val Ser Val Pro Arg Ala Leu Asp
355 360 365

Arg Gln Ala Ala Leu Leu Gln Ile Gln Glu Phe Met Ile Thr Cys Leu
370 375 380

Ser Gln Thr Pro Pro Arg Thr Thr Leu Leu Leu Tyr Pro Thr Ala Val
385 390 395 400

Asp Leu Ala Lys Arg Ala Leu Trp Thr Pro Asn Gln Ile Thr Asp Ile
405 410 415

Thr Ser Leu Val Arg Leu Val Tyr Ile Leu Ser Lys Gln Asn Gln Gln
420 425 430

His Leu Ile Pro Gln Trp Ala Leu Arg Gln Ile Ala Asp Phe Ala Leu
435 440 445

Lys Leu His Lys Thr His Leu Ala Ser Phe Leu Ser Ala Phe Ala Arg
450 455 460

Gln Glu Leu Tyr Leu Met Gly Ser Leu Val His Ser Met Leu Val His
465 470 475 480

Thr Thr Glu Arg Arg Glu Ile Phe Ile Val Glu Thr Gly Leu Cys Ser
485 490 495

Leu Ala Glu Leu Ser His Phe Thr Gln Leu Leu Ala His Pro His His
500 505 510

Glu Tyr Leu Ser Asp Leu Tyr Thr Pro Cys Ser Ser Ser Gly Arg Arg
515 520 525

Asp His Ser Leu Glu Arg Leu Thr Arg Leu Phe Pro Asp Ala Thr Val
530 535 540

eolf-seql.txt

Pro Ala Thr Val Pro Ala Ala Leu Ser Ile Leu Ser Thr Met Gln Pro
545 550 555 560

Ser Thr Leu Glu Thr Phe Pro Asp Leu Phe Cys Leu Pro Leu Gly Glu
565 570 575

Ser Phe Ser Ala Leu Thr Val Ser Glu His Val Ser Tyr Ile Val Thr
580 585 590

Asn Gln Tyr Leu Ile Lys Gly Ile Ser Tyr Pro Val Ser Thr Thr Val
595 600 605

Val Gly Gln Ser Leu Ile Ile Thr Gln Thr Asp Ser Gln Thr Lys Cys
610 615 620

Glu Leu Thr Arg Asn Met His Thr Thr His Ser Ile Thr Val Ala Leu
625 630 635 640

Asn Ile Ser Leu Glu Asn Cys Ala Phe Cys Gln Ser Ala Leu Leu Glu
645 650 655

Tyr Asp Asp Thr Gln Gly Val Ile Asn Ile Met Tyr Met His Asp Ser
660 665 670

Asp Asp Val Leu Phe Ala Leu Asp Pro Tyr Asn Glu Val Val Val Ser
675 680 685

Ser Pro Arg Thr His Tyr Leu Met Leu Leu Lys Asn Gly Thr Val Leu
690 695 700

Glu Val Thr Asp Val Val Val Asp Ala Thr Asp Gly Thr Lys Leu Gly
705 710 715 720

Pro Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu Asn Ser Ala Val Asp
725 730 735

His His His His His His
740

<210> 7
<211> 278
<212> PRT
<213> Human cytomegalovirus, strain Merlin

<400> 7

Met Cys Arg Arg Pro Asp Cys Gly Phe Ser Phe Ser Pro Gly Pro Val
1 5 10 15

Ile Leu Leu Trp Cys Cys Leu Leu Leu Pro Ile Val Ser Ser Ala Ala
20 25 30

eolf-seql.txt

Val Ser Val Ala Pro Thr Ala Ala Glu Lys Val Pro Ala Glu Cys Pro
35 40 45

Glu Leu Thr Arg Arg Cys Leu Leu Gly Glu Val Phe Glu Gly Asp Lys
50 55 60

Tyr Glu Ser Trp Leu Arg Pro Leu Val Asn Val Thr Gly Arg Asp Gly
65 70 75 80

Pro Leu Ser Gln Leu Ile Arg Tyr Arg Pro Val Thr Pro Glu Ala Ala
85 90 95

Asn Ser Val Leu Leu Asp Glu Ala Phe Leu Asp Thr Leu Ala Leu Leu
100 105 110

Tyr Asn Asn Pro Asp Gln Leu Arg Ala Leu Leu Thr Leu Leu Ser Ser
115 120 125

Asp Thr Ala Pro Arg Trp Met Thr Val Met Arg Gly Tyr Ser Glu Cys
130 135 140

Gly Asp Gly Ser Pro Ala Val Tyr Thr Cys Val Asp Asp Leu Cys Arg
145 150 155 160

Gly Tyr Asp Leu Thr Arg Leu Ser Tyr Gly Arg Ser Ile Phe Thr Glu
165 170 175

His Val Leu Gly Phe Glu Leu Val Pro Pro Ser Leu Phe Asn Val Val
180 185 190

Val Ala Ile Arg Asn Glu Ala Thr Arg Thr Asn Arg Ala Val Arg Leu
195 200 205

Pro Val Ser Thr Ala Ala Ala Pro Glu Gly Ile Thr Leu Phe Tyr Gly
210 215 220

Leu Tyr Asn Ala Val Lys Glu Phe Cys Leu Arg His Gln Leu Asp Pro
225 230 235 240

Pro Leu Leu Arg His Leu Asp Lys Tyr Tyr Ala Gly Leu Pro Pro Glu
245 250 255

Leu Lys Gln Thr Arg Val Asn Leu Pro Ala His Ser Arg Tyr Gly Pro
260 265 270

Gln Ala Val Asp Ala Arg
275

<210> 8
<211> 278
<212> PRT

<213> Human cytomegalovirus, strain Towne

<400> 8

Met Cys Arg Arg Pro Asp Cys Gly Phe Ser Phe Ser Pro Gly Pro Val
 1 5 10 15

Ala Leu Leu Trp Cys Cys Leu Leu Leu Pro Ile Val Ser Ser Ala Thr
 20 25 30

Val Ser Val Ala Pro Thr Val Ala Glu Lys Val Pro Ala Glu Cys Pro
 35 40 45

Glu Leu Thr Arg Arg Cys Leu Leu Gly Glu Val Phe Gln Gly Asp Lys
 50 55 60

Tyr Glu Ser Trp Leu Arg Pro Leu Val Asn Val Thr Arg Arg Asp Gly
 65 70 75 80

Pro Leu Ser Gln Leu Ile Arg Tyr Arg Pro Val Thr Pro Glu Ala Ala
 85 90 95

Asn Ser Val Leu Leu Asp Asp Ala Phe Leu Asp Thr Leu Ala Leu Leu
 100 105 110

Tyr Asn Asn Pro Asp Gln Leu Arg Ala Leu Leu Thr Leu Leu Ser Ser
 115 120 125

Asp Thr Ala Pro Arg Trp Met Thr Val Met Arg Gly Tyr Ser Glu Cys
 130 135 140

Gly Asp Gly Ser Pro Ala Val Tyr Thr Cys Val Asp Asp Leu Cys Arg
 145 150 155 160

Gly Tyr Asp Leu Thr Arg Leu Ser Tyr Gly Arg Ser Ile Phe Thr Glu
 165 170 175

His Val Leu Gly Phe Glu Leu Val Pro Pro Ser Leu Phe Asn Val Val
 180 185 190

Val Ala Ile Arg Asn Glu Ala Thr Arg Thr Asn Arg Ala Val Arg Leu
 195 200 205

Pro Val Ser Thr Ala Ala Ala Pro Glu Gly Ile Thr Leu Phe Tyr Gly
 210 215 220

Leu Tyr Asn Ala Val Lys Glu Phe Cys Leu Arg His Gln Leu Asp Pro
 225 230 235 240

Pro Leu Leu Arg His Leu Asp Lys Tyr Tyr Ala Gly Leu Pro Pro Glu
 245 250 255

eolf-seql.txt

Leu Lys Gln Thr Arg Val Asn Leu Pro Ala His Ser Arg Tyr Gly Pro
260 265 270

Gln Ala Val Asp Ala Arg
275

<210> 9
<211> 278
<212> PRT
<213> Human cytomegalovirus, strain AD169

<400> 9

Met Cys Arg Arg Pro Asp Cys Gly Phe Ser Phe Ser Pro Gly Pro Val
1 5 10 15

Val Leu Leu Trp Cys Cys Leu Leu Leu Pro Ile Val Ser Ser Val Ala
20 25 30

Val Ser Val Ala Pro Thr Ala Ala Glu Lys Val Pro Ala Glu Cys Pro
35 40 45

Glu Leu Thr Arg Arg Cys Leu Leu Gly Glu Val Phe Gln Gly Asp Lys
50 55 60

Tyr Glu Ser Trp Leu Arg Pro Leu Val Asn Val Thr Arg Arg Asp Gly
65 70 75 80

Pro Leu Ser Gln Leu Ile Arg Tyr Arg Pro Val Thr Pro Glu Ala Ala
85 90 95

Asn Ser Val Leu Leu Asp Asp Ala Phe Leu Asp Thr Leu Ala Leu Leu
100 105 110

Tyr Asn Asn Pro Asp Gln Leu Arg Ala Leu Leu Thr Leu Leu Ser Ser
115 120 125

Asp Thr Ala Pro Arg Trp Met Thr Val Met Arg Gly Tyr Ser Glu Cys
130 135 140

Gly Asp Gly Ser Pro Ala Val Tyr Thr Cys Val Asp Asp Leu Cys Arg
145 150 155 160

Gly Tyr Asp Leu Thr Arg Leu Ser Tyr Gly Arg Ser Ile Phe Thr Glu
165 170 175

His Val Leu Gly Phe Glu Leu Val Pro Pro Ser Leu Phe Asn Val Val
180 185 190

Val Ala Ile Arg Asn Glu Ala Thr Arg Thr Asn Arg Ala Val Arg Leu
195 200 205

Pro Val Ser Thr Ala Ala Ala Pro Glu Gly Ile Thr Leu Phe Tyr Gly

210

215

Leu Tyr Asn Ala Val Lys Glu Phe Cys Leu Arg His Gln Leu Asp Pro
225 230 235 240

Pro Leu Leu Arg His Leu Asp Lys Tyr Tyr Ala Gly Leu Pro Pro Glu
245 250 255

Leu Lys Gln Thr Arg Val Asn Leu Pro Ala His Ser Arg Tyr Gly Pro
260 265 270

Gln Ala Val Asp Ala Arg
275

<210> 10
<211> 472
<212> PRT
<213> Human cytomegalovirus, strain Merlin
<400> 10

Met Gly Lys Lys Glu Met Ile Met Val Lys Gly Ile Pro Lys Ile Met
1 5 10 15

Leu Leu Ile Ser Ile Thr Phe Leu Leu Ser Leu Ile Asn Cys Asn
20 25 30

Val Leu Val Asn Ser Arg Gly Thr Arg Arg Ser Trp Pro Tyr Thr Val
35 40 45

Leu Ser Tyr Arg Gly Lys Glu Ile Leu Lys Lys Gln Lys Glu Asp Ile
50 55 60

Leu Lys Arg Leu Met Ser Thr Ser Ser Asp Gly Tyr Arg Phe Leu Met
65 70 75 80

Tyr Pro Ser Gln Gln Lys Phe His Ala Ile Val Ile Ser Met Asp Lys
85 90 95

Phe Pro Gln Asp Tyr Ile Leu Ala Gly Pro Ile Arg Asn Asp Ser Ile
100 105 110

Thr His Met Trp Phe Asp Phe Tyr Ser Thr Gln Leu Arg Lys Pro Ala
115 120 125

Lys Tyr Val Tyr Ser Glu Tyr Asn His Thr Ala His Lys Ile Thr Leu
130 135 140

Arg Pro Pro Pro Cys Gly Thr Val Pro Ser Met Asn Cys Leu Ser Glu
145 150 155 160

Met Leu Asn Val Ser Lys Arg Asn Asp Thr Gly Glu Lys Gly Cys Gly
165 170 175

eolf-seql.txt

Asn Phe Thr Thr Phe Asn Pro Met Phe Phe Asn Val Pro Arg Trp Asn
180 185 190

Thr Lys Leu Tyr Ile Gly Ser Asn Lys Val Asn Val Asp Ser Gln Thr
195 200 205

Ile Tyr Phe Leu Gly Leu Thr Ala Leu Leu Leu Arg Tyr Ala Gln Arg
210 215 220

Asn Cys Thr Arg Ser Phe Tyr Leu Val Asn Ala Met Ser Arg Asn Leu
225 230 235 240

Phe Arg Val Pro Lys Tyr Ile Asn Gly Thr Lys Leu Lys Asn Thr Met
245 250 255

Arg Lys Leu Lys Arg Lys Gln Ala Leu Val Lys Glu Gln Pro Gln Lys
260 265 270

Lys Asn Lys Lys Ser Gln Ser Thr Thr Thr Pro Tyr Leu Ser Tyr Thr
275 280 285

Thr Ser Thr Ala Phe Asn Val Thr Thr Asn Val Thr Tyr Ser Ala Thr
290 295 300

Ala Ala Val Thr Arg Val Ala Thr Ser Thr Thr Gly Tyr Arg Pro Asp
305 310 315 320

Ser Asn Phe Met Lys Ser Ile Met Ala Thr Gln Leu Arg Asp Leu Ala
325 330 335

Thr Trp Val Tyr Thr Thr Leu Arg Tyr Arg Asn Glu Pro Phe Cys Lys
340 345 350

Pro Asp Arg Asn Arg Thr Ala Val Ser Glu Phe Met Lys Asn Thr His
355 360 365

Val Leu Ile Arg Asn Glu Thr Pro Tyr Thr Ile Tyr Gly Thr Leu Asp
370 375 380

Met Ser Ser Leu Tyr Tyr Asn Glu Thr Met Ser Val Glu Asn Glu Thr
385 390 395 400

Ala Ser Asp Asn Asn Glu Thr Thr Pro Thr Ser Pro Ser Thr Arg Phe
405 410 415

Gln Arg Thr Phe Ile Asp Pro Leu Trp Asp Tyr Leu Asp Ser Leu Leu
420 425 430

Phe Leu Asp Lys Ile Arg Asn Phe Ser Leu Gln Leu Pro Ala Tyr Gly
435 440 445

eolf-seql.txt

Asn Leu Thr Pro Pro Glu His Arg Arg Ala Ala Asn Leu Ser Thr Leu
450 455 460

Asn Ser Leu Trp Trp Trp Ser Gln
465 470

<210> 11
<211> 466
<212> PRT
<213> Human cytomegalovirus, strain AD169

<400> 11

Met Gly Arg Lys Glu Met Met Val Arg Asp Val Pro Lys Met Val Phe
1 5 10 15

Leu Ile Ser Ile Ser Phe Leu Leu Val Ser Phe Ile Asn Cys Lys Val
20 25 30

Met Ser Lys Ala Leu Tyr Asn Arg Pro Trp Arg Gly Leu Val Leu Ser
35 40 45

Lys Ile Gly Lys Tyr Lys Leu Asp Gln Leu Lys Leu Glu Ile Leu Arg
50 55 60

Gln Leu Glu Thr Thr Ile Ser Thr Lys Tyr Asn Val Ser Lys Gln Pro
65 70 75 80

Val Lys Asn Leu Thr Met Asn Met Thr Glu Phe Pro Gln Tyr Tyr Ile
85 90 95

Leu Ala Gly Pro Ile Gln Asn Tyr Ser Ile Thr Tyr Leu Trp Phe Asp
100 105 110

Phe Tyr Ser Thr Gln Leu Arg Lys Pro Ala Lys Tyr Val Tyr Ser Gln
115 120 125

Tyr Asn His Thr Ala Lys Thr Ile Thr Phe Arg Pro Pro Pro Cys Gly
130 135 140

Thr Val Pro Ser Met Thr Cys Leu Ser Glu Met Leu Asn Val Ser Lys
145 150 155 160

Arg Asn Asp Thr Gly Glu Gln Gly Cys Gly Asn Phe Thr Thr Phe Asn
165 170 175

Pro Met Phe Phe Asn Val Pro Arg Trp Asn Thr Lys Leu Tyr Val Gly
180 185 190

Pro Thr Lys Val Asn Val Asp Ser Gln Thr Ile Tyr Phe Leu Gly Leu
195 200 205

eolf-seql.txt

Thr Ala Leu Leu Leu Arg Tyr Ala Gln Arg Asn Cys Thr His Ser Phe
210 215 220

Tyr Leu Val Asn Ala Met Ser Arg Asn Leu Phe Arg Val Pro Lys Tyr
225 230 235 240

Ile Asn Gly Thr Lys Leu Lys Asn Thr Met Arg Lys Leu Lys Arg Lys
245 250 255

Gln Ala Pro Val Lys Glu Gln Phe Glu Lys Lys Ala Lys Lys Thr Gln
260 265 270

Ser Thr Thr Thr Pro Tyr Phe Ser Tyr Thr Thr Ser Ala Ala Leu Asn
275 280 285

Val Thr Thr Asn Val Thr Tyr Ser Ile Thr Thr Ala Ala Arg Arg Val
290 295 300

Ser Thr Ser Thr Ile Ala Tyr Arg Pro Asp Ser Ser Phe Met Lys Ser
305 310 315 320

Ile Met Ala Thr Gln Leu Arg Asp Leu Ala Thr Trp Val Tyr Thr Thr
325 330 335

Leu Arg Tyr Arg Gln Asn Pro Phe Cys Glu Pro Ser Arg Asn Arg Thr
340 345 350

Ala Val Ser Glu Phe Met Lys Asn Thr His Val Leu Ile Arg Asn Glu
355 360 365

Thr Pro Tyr Thr Ile Tyr Gly Thr Leu Asp Met Ser Ser Leu Tyr Tyr
370 375 380

Asn Glu Thr Met Phe Val Glu Asn Lys Thr Ala Ser Asp Ser Asn Lys
385 390 395 400

Thr Thr Pro Thr Ser Pro Ser Met Gly Phe Gln Arg Thr Phe Ile Asp
405 410 415

Pro Leu Trp Asp Tyr Leu Asp Ser Leu Leu Phe Leu Asp Glu Ile Arg
420 425 430

Asn Phe Ser Leu Arg Ser Pro Thr Tyr Val Asn Leu Thr Pro Pro Glu
435 440 445

His Arg Arg Ala Val Asn Leu Ser Thr Leu Asn Ser Leu Trp Trp Trp
450 455 460

Leu Gln
465

eolf-seql.txt

```

<210> 12
<211> 457
<212> PRT
<213> Human cytomegalovirus, strain Towne

<400> 12

Met Gly Arg Lys Gly Glu Met Arg Gly Val Phe Asn Leu Phe Phe Leu
1      5      10      15

Met Ser Leu Thr Phe Leu Leu Phe Ser Phe Ile Asn Cys Lys Ile Ala
20      25      30

Val Ala Arg Phe Arg Val Lys Ser Gln Lys Ala Lys Glu Glu Glu Arg
35      40      45

Gln Leu Lys Leu Arg Ile Leu Gln Glu Leu Ala Ser Lys Thr Gly Asp
50      55      60

Tyr Tyr Lys Phe Phe Thr Phe Pro Ser Gln Gln Lys Leu Tyr Asn Ile
65      70      75      80

Thr Val Glu Met Lys Gln Phe Pro Pro Asn Ser Ile Leu Ala Gly Pro
85      90      95

Ile Arg Asn His Ser Ile Thr His Leu Trp Phe Asp Phe His Thr Thr
100     105     110

Gln Leu Arg Lys Pro Ala Lys Tyr Val Tyr Ser Glu Tyr Asn His Thr
115     120     125

Gly Gln Lys Ile Thr Phe Arg Pro Pro Ser Cys Gly Thr Ile Pro Ser
130     135     140

Met Thr Cys Leu Ser Glu Met Leu Asn Val Ser Arg Arg Asn Asn Thr
145     150     155     160

Gly Glu Glu Asn Cys Gly Asn Phe Thr Thr Phe Asn Pro Met Phe Phe
165     170     175

Asn Val Pro Arg Trp Asn Thr Lys Leu Tyr Val Gly Pro Ser Lys Val
180     185     190

Asn Val Asp Ser Gln Thr Ile Tyr Phe Leu Gly Leu Ala Ala Leu Leu
195     200     205

Leu Arg Tyr Ala Gln Arg Asn Cys Thr Arg Ser Phe Tyr Leu Val Asn
210     215     220

Ala Met Ser Arg Asn Ile Phe Arg Val Pro Lys Tyr Ile Asn Ser Thr
225     230     235     240

```

eolf-seql.txt

Lys Leu Lys Asn Thr Met Arg Lys Leu Lys Arg Lys Gln Ala Pro Val
245 250 255

Lys Ser Ile Ser Lys Lys Ser Arg Val Ser Thr Thr Thr Pro Tyr Ser
260 265 270

Ser Tyr Thr Ser Thr Ile Phe Asn Val Ser Thr Asn Val Thr Tyr Ser
275 280 285

Pro Ile Val Pro Thr Arg Ile Pro Thr Ser Thr Ile Gly Tyr Arg Pro
290 295 300

Asp Glu Asn Phe Met Lys Ser Ile Leu Thr Thr Gln Leu Lys Asp Leu
305 310 315 320

Ala Thr Trp Val Tyr Thr Thr Leu Arg Tyr Arg Asp Glu Pro Phe Cys
325 330 335

Lys Pro Asn Arg Asn Arg Thr Ala Val Ser Glu Phe Met Lys Asn Thr
340 345 350

His Val Leu Ile Arg Asn Glu Thr Pro Tyr Thr Ile Tyr Gly Thr Leu
355 360 365

Asp Met Ser Ser Leu Tyr Tyr Asn Asp Thr Met Pro Val Glu Asn Glu
370 375 380

Thr Ala Ser Asp Asn Asn Lys Thr Thr Pro Thr Ser Pro Ser Thr Arg
385 390 395 400

Phe Gln Arg Thr Phe Ile Asp Pro Met Trp Asp Tyr Leu Asp Ser Leu
405 410 415

Leu Phe Leu Ser Glu Ile Arg Asn Phe Ser Leu Gln Ser Ser Thr Tyr
420 425 430

Gly Asn Leu Thr Pro Pro Glu His Arg Arg Ala Val Asn Leu Ser Thr
435 440 445

Leu Asn Ser Leu Trp Trp Trp Leu Gln
450 455

<210> 13
<211> 130
<212> PRT
<213> Human cytomegalovirus, strain Merlin

<400> 13

Met Ser Pro Lys Asp Leu Thr Pro Phe Leu Thr Ala Leu Trp Leu Leu
1 5 10 15

Leu Gly His Ser Arg Val Pro Arg Val Arg Ala Glu Glu Cys Cys Glu
Page 24

20

25

30

Phe Ile Asn Val Asn His Pro Pro Glu Arg Cys Tyr Asp Phe Lys Met
 35 40 45

Cys Asn Arg Phe Thr Val Ala Leu Arg Cys Pro Asp Gly Glu Val Cys
 50 55 60

Tyr Ser Pro Glu Lys Thr Ala Glu Ile Arg Gly Ile Val Thr Thr Met
 65 70 75 80

Thr His Ser Leu Thr Arg Gln Val Val His Asn Lys Leu Thr Ser Cys
 85 90 95

Asn Tyr Asn Pro Leu Tyr Leu Glu Ala Asp Gly Arg Ile Arg Cys Gly
 100 105 110

Lys Val Asn Asp Lys Ala Gln Tyr Leu Leu Gly Ala Ala Gly Ser Val
 115 120 125

Pro Tyr
 130

<210> 14

<211> 171

<212> PRT

<213> Human cytomegalovirus, strain Towne

<400> 14

Met Ser Pro Lys Asn Leu Thr Pro Phe Leu Thr Ala Leu Trp Leu Leu
 1 5 10 15

Leu Gly His Ser Arg Val Pro Arg Val Arg Ala Glu Glu Cys Cys Glu
 20 25 30

Phe Ile Asn Val Asn His Pro Pro Glu Arg Cys Tyr Asp Phe Lys Met
 35 40 45

Cys Asn Arg Phe Thr Val Ala Leu Arg Cys Pro Asp Gly Glu Val Cys
 50 55 60

Tyr Ser Pro Glu Lys Thr Ala Glu Ile Arg Gly Ile Val Thr Thr Met
 65 70 75 80

Thr His Ser Leu Thr Arg Gln Val Val His Asn Lys Leu Thr Ser Cys
 85 90 95

Asn Tyr Asn Pro Leu Tyr Leu Glu Ala Asp Gly Arg Ile Arg Cys Gly
 100 105 110

Lys Val Asn Asp Lys Ala Gln Tyr Leu Leu Gly Ala Ala Gly Ser Val
 115 120 125

eolf-seql.txt

Pro Tyr Arg Trp Ile Asn Leu Glu Tyr Asp Lys Ile Thr Arg Ile Val
130 135 140

Gly Leu Asp Gln Tyr Leu Glu Ser Val Lys Lys His Lys Arg Leu Asp
145 150 155 160

Val Cys Arg Ala Lys Met Gly Tyr Met Leu Gln
165 170

<210> 15
<211> 171
<212> PRT
<213> Human cytomegalovirus, strain AD196
<400> 15

Met Ser Pro Lys Asp Leu Thr Pro Phe Leu Thr Thr Leu Trp Leu Leu
1 5 10 15

Leu Gly His Ser Arg Val Pro Arg Val Arg Ala Glu Glu Cys Cys Glu
20 25 30

Phe Ile Asn Val Asn His Pro Pro Glu Arg Cys Tyr Asp Phe Lys Met
35 40 45

Cys Asn Arg Phe Thr Val Ala Leu Arg Cys Pro Asp Gly Glu Val Cys
50 55 60

Tyr Ser Pro Glu Lys Thr Ala Glu Ile Arg Gly Ile Val Thr Thr Met
65 70 75 80

Thr His Ser Leu Thr Arg Gln Val Val His Asn Lys Leu Thr Ser Cys
85 90 95

Asn Tyr Asn Pro Leu Tyr Leu Glu Ala Asp Gly Arg Ile Arg Cys Gly
100 105 110

Lys Val Asn Asp Lys Ala Gln Tyr Leu Leu Gly Ala Ala Gly Ser Val
115 120 125

Pro Tyr Arg Trp Ile Asn Leu Glu Tyr Asp Lys Ile Thr Arg Ile Val
130 135 140

Gly Leu Asp Gln Tyr Leu Glu Ser Val Lys Lys His Lys Arg Leu Asp
145 150 155 160

Val Cys Arg Ala Lys Met Gly Tyr Met Leu Gln
165 170

<210> 16
<211> 214
<212> PRT

<213> Human cytomegalovirus, strain Merlin

<400> 16

Met Leu Arg Leu Leu Leu Arg His His Phe His Cys Leu Leu Leu Cys
1 5 10 15

Ala Val Trp Ala Thr Pro Cys Leu Ala Ser Pro Trp Ser Thr Leu Thr
20 25 30

Ala Asn Gln Asn Pro Ser Pro Pro Trp Ser Lys Leu Thr Tyr Ser Lys
35 40 45

Pro His Asp Ala Ala Thr Phe Tyr Cys Pro Phe Leu Tyr Pro Ser Pro
50 55 60

Pro Arg Ser Pro Leu Gln Phe Ser Gly Phe Gln Arg Val Ser Thr Gly
65 70 75 80

Pro Glu Cys Arg Asn Glu Thr Leu Tyr Leu Leu Tyr Asn Arg Glu Gly
85 90 95

Gln Thr Leu Val Glu Arg Ser Ser Thr Trp Val Lys Lys Val Ile Trp
100 105 110

Tyr Leu Ser Gly Arg Asn Gln Thr Ile Leu Gln Arg Met Pro Arg Thr
115 120 125

Ala Ser Lys Pro Ser Asp Gly Asn Val Gln Ile Ser Val Glu Asp Ala
130 135 140

Lys Ile Phe Gly Ala His Met Val Pro Lys Gln Thr Lys Leu Leu Arg
145 150 155 160

Phe Val Val Asn Asp Gly Thr Arg Tyr Gln Met Cys Val Met Lys Leu
165 170 175

Glu Ser Trp Ala His Val Phe Arg Asp Tyr Ser Val Ser Phe Gln Val
180 185 190

Arg Leu Thr Phe Thr Glu Ala Asn Asn Gln Thr Tyr Thr Phe Cys Thr
195 200 205

His Pro Asn Leu Ile Val
210

<210> 17

<211> 229

<212> PRT

<213> Human cytomegalovirus, strain Towne

<400> 17

Met Leu Arg Leu Leu Leu Arg His His Phe His Cys Leu Leu Leu Cys

eo1f-seql.txt

```

1              5              10              15
Ala Val Trp Ala Thr Pro Cys Leu Ala Ser Pro Trp Ser Thr Leu Thr
  20              25              30
Ala Asn Gln Asn Pro Ser Pro Pro Trp Ser Lys Leu Thr Tyr Ser Lys
  35              40              45
Pro His Asp Ala Ala Thr Phe Tyr Cys Pro Phe Leu Tyr Pro Ser Pro
  50              55              60
Pro Arg Ser Pro Leu Gln Phe Ser Gly Phe Gln Arg Val Leu Thr Gly
  65              70              75              80
Pro Glu Cys Arg Asn Glu Thr Leu Tyr Leu Leu Tyr Asn Arg Glu Gly
  85              90              95
Gln Thr Leu Val Glu Arg Ser Ser Thr Trp Val Lys Lys Val Ile Trp
 100              105              110
Tyr Leu Ser Gly Arg Asn Gln Thr Ile Leu Gln Arg Met Pro Arg Thr
 115              120              125
Ala Ser Lys Pro Ser Asp Gly Asn Val Gln Ile Ser Val Glu Asp Ala
 130              135              140
Lys Ile Phe Gly Ala His Met Val Pro Lys Gln Thr Lys Leu Leu Arg
 145              150              155              160
Phe Val Val Asn Asp Gly Thr Arg Tyr Gln Met Cys Val Met Lys Leu
 165              170              175
Glu Ser Trp Ala His Val Phe Arg Asp Tyr Ser Val Ser Phe Gln Val
 180              185              190
Arg Leu Thr Phe Thr Glu Ala Asn Asn Gln Thr Phe Thr Pro Ser Ala
 195              200              205
Pro Ile Pro Ile Ser Ser Phe Glu Pro Val Ala Arg Ala Gly Asn Phe
 210              215              220
Glu Asn Arg Ala Ser
 225

```

```

<210> 18
<211> 129
<212> PRT
<213> Human cytomegalovirus; strain Merlin
<400> 18

```

```

Met Arg Leu Cys Arg Val Trp Leu Ser Val Cys Leu Cys Ala Val Val
1              5              10              15

```

eolf-seql.txt

Leu Gly Gln Cys Gln Arg Glu Thr Ala Glu Lys Asn Asp Tyr Tyr Arg
20 25 30

Val Pro His Tyr Trp Asp Ala Cys Ser Arg Ala Leu Pro Asp Gln Thr
35 40 45

Arg Tyr Lys Tyr Val Glu Gln Leu Val Asp Leu Thr Leu Asn Tyr His
50 55 60

Tyr Asp Ala Ser His Gly Leu Asp Asn Phe Asp Val Leu Lys Arg Ile
65 70 75 80

Asn Val Thr Glu Val Ser Leu Leu Ile Ser Asp Phe Arg Arg Gln Asn
85 90 95

Arg Arg Gly Gly Thr Asn Lys Arg Thr Thr Phe Asn Ala Ala Gly Ser
100 105 110

Leu Ala Pro His Ala Arg Ser Leu Glu Phe Ser Val Arg Leu Phe Ala
115 120 125

Asn

<210> 19
<211> 129
<212> PRT
<213> Human cytomegalovirus, strain Towne

<400> 19

Met Arg Leu Cys Arg Val Trp Leu Ser Val Cys Leu Cys Ala Val Val
1 5 10 15

Leu Gly Gln Cys Gln Arg Glu Thr Ala Glu Lys Asn Asp Tyr Tyr Arg
20 25 30

Val Pro His Tyr Trp Asp Ala Cys Ser Arg Ala Leu Pro Asp Gln Thr
35 40 45

Arg Tyr Lys Tyr Val Glu Gln Leu Val Asp Leu Thr Leu Asn Tyr His
50 55 60

Tyr Asp Ala Ser His Gly Leu Asp Asn Phe Asp Val Leu Lys Arg Ile
65 70 75 80

Asn Val Thr Glu Val Ser Leu Leu Ile Ser Asp Phe Arg Arg Gln Asn
85 90 95

Arg Arg Gly Gly Thr Asn Lys Arg Thr Thr Phe Asn Ala Ala Gly Ser
100 105 110

eolf-seql.txt

Leu Ala Pro His Ala Arg Ser Leu Glu Phe Ser Val Arg Leu Phe Ala
 115 120 125

Asn

<210> 20
 <211> 74
 <212> PRT
 <213> Human cytomegalovirus, strain AD169

<400> 20

Met Arg Leu Cys Arg Val Trp Leu Ser Val Cys Leu Cys Ala Val Val
 1 5 10 15

Leu Gly Gln Cys Gln Arg Glu Thr Ala Glu Lys Lys Arg Leu Leu Pro
 20 25 30

Ser Thr Ala Leu Leu Gly Arg Val Leu Ser Arg Ala Ala Arg Pro Asn
 35 40 45

Pro Leu Gln Val Cys Gly Thr Ala Arg Gly Pro His Val Glu Leu Pro
 50 55 60

Leu Arg Cys Glu Pro Arg Leu Gly Gln Leu
 65 70

<210> 21
 <211> 907
 <212> PRT
 <213> Human cytomegalovirus; strain Merlin

<400> 21

Met Glu Ser Arg Ile Trp Cys Leu Val Val Cys Val Asn Leu Cys Ile
 1 5 10 15

Val Cys Leu Gly Ala Ala Val Ser Ser Ser Thr Arg Gly Thr Ser
 20 25 30

Ala Thr His Ser His His Ser Ser His Thr Thr Ser Ala Ala His Ser
 35 40 45

Arg Ser Gly Ser Val Ser Gln Arg Val Thr Ser Ser Gln Thr Val Ser
 50 55 60

His Gly Val Asn Glu Thr Ile Tyr Asn Thr Thr Leu Lys Tyr Gly Asp
 65 70 75 80

Val Val Gly Val Asn Thr Thr Lys Tyr Pro Tyr Arg Val Cys Ser Met
 85 90 95

Ala Gln Gly Thr Asp Leu Ile Arg Phe Glu Arg Asn Ile Val Cys Thr
 Page 30

eofl-seql.txt

100

105

110

Ser Met Lys Pro Ile Asn Glu Asp Leu Asp Glu Gly Ile Met Val Val
115 120 125

Tyr Lys Arg Asn Ile Val Ala His Thr Phe Lys Val Arg Val Tyr Gln
130 135 140

Lys Val Leu Thr Phe Arg Arg Ser Tyr Ala Tyr Ile His Thr Thr Tyr
145 150 155 160

Leu Leu Gly Ser Asn Thr Glu Tyr Val Ala Pro Pro Met Trp Glu Ile
165 170 175

His His Ile Asn Ser His Ser Gln Cys Tyr Ser Ser Tyr Ser Arg Val
180 185 190

Ile Ala Gly Thr Val Phe Val Ala Tyr His Arg Asp Ser Tyr Glu Asn
195 200 205

Lys Thr Met Gln Leu Met Pro Asp Asp Tyr Ser Asn Thr His Ser Thr
210 215 220

Arg Tyr Val Thr Val Lys Asp Gln Trp His Ser Arg Gly Ser Thr Trp
225 230 235 240

Leu Tyr Arg Glu Thr Cys Asn Leu Asn Cys Met Val Thr Ile Thr Thr
245 250 255

Ala Arg Ser Lys Tyr Pro Tyr His Phe Phe Ala Thr Ser Thr Gly Asp
260 265 270

Val Val Asp Ile Ser Pro Phe Tyr Asn Gly Thr Asn Arg Asn Ala Ser
275 280 285

Tyr Phe Gly Glu Asn Ala Asp Lys Phe Phe Ile Phe Pro Asn Tyr Thr
290 295 300

Ile Val Ser Asp Phe Gly Arg Pro Asn Ser Ala Leu Glu Thr His Arg
305 310 315 320

Leu Val Ala Phe Leu Glu Arg Ala Asp Ser Val Ile Ser Trp Asp Ile
325 330 335

Gln Asp Glu Lys Asn Val Thr Cys Gln Leu Thr Phe Trp Glu Ala Ser
340 345 350

Glu Arg Thr Ile Arg Ser Glu Ala Glu Asp Ser Tyr His Phe Ser Ser
355 360 365

Ala Lys Met Thr Ala Thr Phe Leu Ser Lys Lys Gln Glu Val Asn Met

370

375

Ser Asp Ser Ala Leu Asp Cys Val Arg Asp Glu Ala Ile Asn Lys Leu
385 390 395 400

Gln Gln Ile Phe Asn Thr Ser Tyr Asn Gln Thr Tyr Glu Lys Tyr Gly
405 410 415

Asn Val Ser Val Phe Glu Thr Thr Gly Gly Leu Val Val Phe Trp Gln
420 425 430

Gly Ile Lys Gln Lys Ser Leu Val Glu Leu Glu Arg Leu Ala Asn Arg
435 440 445

Ser Ser Leu Asn Leu Thr His Asn Arg Thr Lys Arg Ser Thr Asp Gly
450 455 460

Asn Asn Ala Thr His Leu Ser Asn Met Glu Ser Val His Asn Leu Val
465 470 475 480

Tyr Ala Gln Leu Gln Phe Thr Tyr Asp Thr Leu Arg Gly Tyr Ile Asn
485 490 495

Arg Ala Leu Ala Gln Ile Ala Glu Ala Trp Cys Val Asp Gln Arg Arg
500 505 510

Thr Leu Glu Val Phe Lys Glu Leu Ser Lys Ile Asn Pro Ser Ala Ile
515 520 525

Leu Ser Ala Ile Tyr Asn Lys Pro Ile Ala Ala Arg Phe Met Gly Asp
530 535 540

Val Leu Gly Leu Ala Ser Cys Val Thr Ile Asn Gln Thr Ser Val Lys
545 550 555 560

Val Leu Arg Asp Met Asn Val Lys Glu Ser Pro Gly Arg Cys Tyr Ser
565 570 575

Arg Pro Val Val Ile Phe Asn Phe Ala Asn Ser Ser Tyr Val Gln Tyr
580 585 590

Gly Gln Leu Gly Glu Asp Asn Glu Ile Leu Leu Gly Asn His Arg Thr
595 600 605

Glu Glu Cys Gln Leu Pro Ser Leu Lys Ile Phe Ile Ala Gly Asn Ser
610 615 620

Ala Tyr Glu Tyr Val Asp Tyr Leu Phe Lys Arg Met Ile Asp Leu Ser
625 630 635 640

Ser Ile Ser Thr Val Asp Ser Met Ile Ala Leu Asp Ile Asp Pro Leu

eolf-seql.txt

645

650

655

Glu Asn Thr Asp Phe Arg Val Leu Glu Leu Tyr Ser Gln Lys Glu Leu
 660 665 670
 Arg Ser Ser Asn Val Phe Asp Leu Glu Glu Ile Met Arg Glu Phe Asn
 675 680 685
 Ser Tyr Lys Gln Arg Val Lys Tyr Val Glu Asp Lys Val Val Asp Pro
 690 695 700
 Leu Pro Pro Tyr Leu Lys Gly Leu Asp Asp Leu Met Ser Gly Leu Gly
 705 710 715 720
 Ala Ala Gly Lys Ala Val Gly Val Ala Ile Gly Ala Val Gly Gly Ala
 725 730 735
 Val Ala Ser Val Val Glu Gly Val Ala Thr Phe Leu Lys Asn Pro Phe
 740 745 750
 Gly Ala Phe Thr Ile Ile Leu Val Ala Ile Ala Val Val Ile Ile Thr
 755 760 765
 Tyr Leu Ile Tyr Thr Arg Gln Arg Arg Leu Cys Thr Gln Pro Leu Gln
 770 775 780
 Asn Leu Phe Pro Tyr Leu Val Ser Ala Asp Gly Thr Thr Val Thr Ser
 785 790 795 800
 Gly Ser Thr Lys Asp Thr Ser Leu Gln Ala Pro Pro Ser Tyr Glu Glu
 805 810 815
 Ser Val Tyr Asn Ser Gly Arg Lys Gly Pro Gly Pro Pro Ser Ser Asp
 820 825 830
 Ala Ser Thr Ala Ala Pro Pro Tyr Thr Asn Glu Gln Ala Tyr Gln Met
 835 840 845
 Leu Leu Ala Leu Ala Arg Leu Asp Ala Glu Gln Arg Ala Gln Gln Asn
 850 855 860
 Gly Thr Asp Ser Leu Asp Gly Arg Thr Gly Thr Gln Asp Lys Gly Gln
 865 870 875 880
 Lys Pro Asn Leu Leu Asp Arg Leu Arg His Arg Lys Asn Gly Tyr Arg
 885 890 895
 His Leu Lys Asp Ser Asp Glu Glu Glu Asn Val
 900 905

<210> 22

eolf-seql.txt

<211> 907

<212> PRT

<213> Human cytomegalovirus, strain Towne

<400> 22

Met Glu Ser Arg Ile Trp Cys Leu Val Val Cys Val Asn Leu Cys Ile
1 5 10 15

Val Cys Leu Gly Ala Ala Val Ser Ser Ser Ser Thr Arg Gly Thr Ser
20 25 30

Ala Thr His Ser His His Ser Ser His Thr Thr Ser Ala Ala His Ser
35 40 45

Arg Ser Gly Ser Val Ser Gln Arg Val Thr Ser Ser Gln Thr Val Ser
50 55 60

His Gly Val Asn Glu Thr Ile Tyr Asn Thr Thr Leu Lys Tyr Gly Asp
65 70 75 80

Val Val Gly Val Asn Thr Thr Lys Tyr Pro Tyr Arg Val Cys Ser Met
85 90 95

Ala Gln Gly Thr Asp Leu Ile Arg Phe Glu Arg Asn Ile Val Cys Thr
100 105 110

Ser Met Lys Pro Ile Asn Glu Asp Leu Asp Glu Gly Ile Met Val Val
115 120 125

Tyr Lys Arg Asn Ile Val Ala His Thr Phe Lys Val Arg Val Tyr Gln
130 135 140

Lys Val Leu Thr Phe Arg Arg Ser Tyr Ala Tyr Ile His Thr Thr Tyr
145 150 155 160

Leu Leu Gly Ser Asn Thr Glu Tyr Val Ala Pro Pro Met Trp Glu Ile
165 170 175

His His Ile Asn Ser His Ser Gln Cys Tyr Ser Ser Tyr Ser Arg Val
180 185 190

Ile Ala Gly Thr Val Phe Val Ala Tyr His Arg Asp Ser Tyr Glu Asn
195 200 205

Lys Thr Met Gln Leu Met Pro Asp Asp Tyr Ser Asn Thr His Ser Thr
210 215 220

Arg Tyr Val Thr Val Lys Asp Gln Trp His Ser Arg Gly Ser Thr Trp
225 230 235 240

Leu Tyr Arg Glu Thr Cys Asn Leu Asn Cys Met Val Thr Ile Thr Thr
245 250 255

eolf-seql.txt

Ala Arg Ser Lys Tyr Pro Tyr His Phe Phe Ala Thr Ser Thr Gly Asp
260 265 270

Val Val Asp Ile Ser Pro Phe Tyr Asn Gly Thr Asn Arg Asn Ala Ser
275 280 285

Tyr Phe Gly Glu Asn Ala Asp Lys Phe Phe Ile Phe Pro Asn Tyr Thr
290 295 300

Ile Val Ser Asp Phe Gly Arg Pro Asn Ser Ala Leu Glu Thr His Arg
305 310 315 320

Leu Val Ala Phe Leu Glu Arg Ala Asp Ser Val Ile Ser Trp Asp Ile
325 330 335

Gln Asp Glu Lys Asn Val Thr Cys Gln Leu Thr Phe Trp Glu Ala Ser
340 345 350

Glu Arg Thr Ile Arg Ser Glu Ala Glu Asp Ser Tyr His Phe Ser Ser
355 360 365

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

Ser Asp Ser Ala Leu Asp Cys Val Arg Asp Glu Ala Ile Asn Lys Leu
385 390 395 400

Gln Gln Ile Phe Asn Thr Ser Tyr Asn Gln Thr Tyr Glu Lys Tyr Gly
405 410 415

Asn Val Ser Val Phe Glu Thr Thr Gly Gly Leu Val Val Phe Trp Gln
420 425 430

Gly Ile Lys Gln Lys Ser Leu Val Glu Leu Glu Arg Leu Ala Asn Arg
435 440 445

Ser Ser Leu Asn Leu Thr His Asn Arg Thr Lys Arg Ser Thr Asp Gly
450 455 460

Asn Asn Ala Thr His Leu Ser Asn Met Glu Ser Val His Asn Leu Val
465 470 475 480

Tyr Ala Gln Leu Gln Phe Thr Tyr Asp Thr Leu Arg Gly Tyr Ile Asn
485 490 495

Arg Ala Leu Ala Gln Ile Ala Glu Ala Trp Cys Val Asp Gln Arg Arg
500 505 510

Thr Leu Glu Val Phe Lys Glu Leu Ser Lys Ile Asn Pro Ser Ala Ile
515 520 525

eolf-seql.txt

Leu Ser Ala Ile Tyr Asn Lys Pro Ile Ala Ala Arg Phe Met Gly Asp
 530 535 540
 Val Leu Gly Leu Ala Ser Cys Val Thr Ile Asn Gln Thr Ser Val Lys
 545 550 555 560
 Val Leu Arg Asp Met Asn Val Lys Glu Ser Pro Gly Arg Cys Tyr Ser
 565 570 575
 Arg Pro Val Val Ile Phe Asn Phe Ala Asn Ser Ser Tyr Val Gln Tyr
 580 585 590
 Gly Gln Leu Gly Glu Asp Asn Glu Ile Leu Leu Gly Asn His Arg Thr
 595 600 605
 Glu Glu Cys Gln Leu Pro Ser Leu Lys Ile Phe Ile Ala Gly Asn Ser
 610 615 620
 Ala Tyr Glu Tyr Val Asp Tyr Leu Phe Lys Arg Met Ile Asp Leu Ser
 625 630 635 640
 Ser Ile Ser Thr Val Asp Ser Met Ile Ala Leu Asp Ile Asp Pro Leu
 645 650 655
 Glu Asn Thr Asp Phe Arg Val Leu Glu Leu Tyr Ser Gln Lys Glu Leu
 660 665 670
 Arg Ser Ser Asn Val Phe Asp Leu Glu Glu Ile Met Arg Glu Phe Asn
 675 680 685
 Ser Tyr Lys Gln Arg Val Lys Tyr Val Glu Asp Lys Val Val Asp Pro
 690 695 700
 Leu Pro Pro Tyr Leu Lys Gly Leu Asp Asp Leu Met Ser Gly Leu Gly
 705 710 715 720
 Ala Ala Gly Lys Ala Val Gly Val Ala Ile Gly Ala Val Gly Gly Ala
 725 730 735
 Val Ala Ser Val Val Glu Gly Val Ala Thr Phe Leu Lys Asn Pro Phe
 740 745 750
 Gly Ala Phe Thr Ile Ile Leu Val Ala Ile Ala Val Val Ile Ile Ile
 755 760 765
 Tyr Leu Ile Tyr Thr Arg Gln Arg Arg Leu Cys Met Gln Pro Leu Gln
 770 775 780
 Asn Leu Phe Pro Tyr Leu Val Ser Ala Asp Gly Thr Thr Val Thr Ser
 785 790 795 800

eolf-seql.txt

Gly Asn Thr Lys Asp Thr Ser Leu Gln Ala Pro Pro Ser Tyr Glu Glu
805 810 815

Ser Val Tyr Asn Ser Gly Arg Lys Gly Pro Gly Pro Pro Ser Ser Asp
820 825 830

Ala Ser Thr Ala Ala Pro Pro Tyr Thr Asn Glu Gln Ala Tyr Gln Met
835 840 845

Leu Leu Ala Leu Val Arg Leu Asp Ala Glu Gln Arg Ala Gln Gln Asn
850 855 860

Gly Thr Asp Ser Leu Asp Gly Gln Thr Gly Thr Gln Asp Lys Gly Gln
865 870 875 880

Lys Pro Asn Leu Leu Asp Arg Leu Arg His Arg Lys Asn Gly Tyr Arg
885 890 895

His Leu Lys Asp Ser Asp Glu Glu Glu Asn Val
900 905

<210> 23
<211> 906
<212> PRT
<213> Human cytomegalovirus, strain AD169
<400> 23

Met Glu Ser Arg Ile Trp Cys Leu Val Val Cys Val Asn Leu Cys Ile
1 5 10 15

Val Cys Leu Gly Ala Ala Val Ser Ser Ser Thr Ser His Ala Thr
20 25 30

Ser Ser Thr His Asn Gly Ser His Thr Ser Arg Thr Thr Ser Ala Gln
35 40 45

Thr Arg Ser Val Tyr Ser Gln His Val Thr Ser Ser Glu Ala Val Ser
50 55 60

His Arg Ala Asn Glu Thr Ile Tyr Asn Thr Thr Leu Lys Tyr Gly Asp
65 70 75 80

Val Val Gly Val Asn Thr Thr Lys Tyr Pro Tyr Arg Val Cys Ser Met
85 90 95

Ala Gln Gly Thr Asp Leu Ile Arg Phe Glu Arg Asn Ile Ile Cys Thr
100 105 110

Ser Met Lys Pro Ile Asn Glu Asp Leu Asp Glu Gly Ile Met Val Val
115 120 125

eolf-seql.txt

Tyr Lys Arg Asn Ile Val Ala His Thr Phe Lys Val Arg Val Tyr Gln
 130 135 140
 Lys Val Leu Thr Phe Arg Arg Ser Tyr Ala Tyr Ile Tyr Thr Thr Tyr
 145 150 155 160
 Leu Leu Gly Ser Asn Thr Glu Tyr Val Ala Pro Pro Met Trp Glu Ile
 165 170 175
 His His Ile Asn Lys Phe Ala Gln Cys Tyr Ser Ser Tyr Ser Arg Val
 180 185 190
 Ile Gly Gly Thr Val Phe Val Ala Tyr His Arg Asp Ser Tyr Glu Asn
 195 200 205
 Lys Thr Met Gln Leu Ile Pro Asp Asp Tyr Ser Asn Thr His Ser Thr
 210 215 220
 Arg Tyr Val Thr Val Lys Asp Gln Trp His Ser Arg Gly Ser Thr Trp
 225 230 235 240
 Leu Tyr Arg Glu Thr Cys Asn Leu Asn Cys Met Leu Thr Ile Thr Thr
 245 250 255
 Ala Arg Ser Lys Tyr Pro Tyr His Phe Phe Ala Thr Ser Thr Gly Asp
 260 265 270
 Val Val Tyr Ile Ser Pro Phe Tyr Asn Gly Thr Asn Arg Asn Ala Ser
 275 280 285
 Tyr Phe Gly Glu Asn Ala Asp Lys Phe Phe Ile Phe Pro Asn Tyr Thr
 290 295 300
 Ile Val Ser Asp Phe Gly Arg Pro Asn Ala Ala Pro Glu Thr His Arg
 305 310 315 320
 Leu Val Ala Phe Leu Glu Arg Ala Asp Ser Val Ile Ser Trp Asp Ile
 325 330 335
 Gln Asp Glu Lys Asn Val Thr Cys Gln Leu Thr Phe Trp Glu Ala Ser
 340 345 350
 Glu Arg Thr Ile Arg Ser Glu Ala Glu Asp Ser Tyr His Phe Ser Ser
 355 360 365
 Ala Lys Met Thr Ala Thr Phe Leu Ser Lys Lys Gln Glu Val Asn Met
 370 375 380
 Ser Asp Ser Ala Leu Asp Cys Val Arg Asp Glu Ala Ile Asn Lys Leu
 385 390 395 400

eolf-seql.txt

Gln Gln Ile Phe Asn Thr Ser Tyr Asn Gln Thr Tyr Glu Lys Tyr Gly
405 410 415

Asn Val Ser Val Phe Glu Thr Ser Gly Gly Leu Val Val Phe Trp Gln
420 425 430

Gly Ile Lys Gln Lys Ser Leu Val Glu Leu Glu Arg Leu Ala Asn Arg
435 440 445

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

Asn Thr Thr His Leu Ser Ser Met Glu Ser Val His Asn Leu Val Tyr
465 470 475 480

Ala Gln Leu Gln Phe Thr Tyr Asp Thr Leu Arg Gly Tyr Ile Asn Arg
485 490 495

Ala Leu Ala Gln Ile Ala Glu Ala Trp Cys Val Asp Gln Arg Arg Thr
500 505 510

Leu Glu Val Phe Lys Glu Leu Ser Lys Ile Asn Pro Ser Ala Ile Leu
515 520 525

Ser Ala Ile Tyr Asn Lys Pro Ile Ala Ala Arg Phe Met Gly Asp Val
530 535 540

Leu Gly Leu Ala Ser Cys Val Thr Ile Asn Gln Thr Ser Val Lys Val
545 550 555 560

Leu Arg Asp Met Asn Val Lys Glu Ser Pro Gly Arg Cys Tyr Ser Arg
565 570 575

Pro Val Val Ile Phe Asn Phe Ala Asn Ser Ser Tyr Val Gln Tyr Gly
580 585 590

Gln Leu Gly Glu Asp Asn Glu Ile Leu Leu Gly Asn His Arg Thr Glu
595 600 605

Glu Cys Gln Leu Pro Ser Leu Lys Ile Phe Ile Ala Gly Asn Ser Ala
610 615 620

Tyr Glu Tyr Val Asp Tyr Leu Phe Lys Arg Met Ile Asp Leu Ser Ser
625 630 635 640

Ile Ser Thr Val Asp Ser Met Ile Ala Leu Asp Ile Asp Pro Leu Glu
645 650 655

Asn Thr Asp Phe Arg Val Leu Glu Leu Tyr Ser Gln Lys Glu Leu Arg
660 665 670

eolf-seql.txt

Ser Ser Asn Val Phe Asp Leu Glu Glu Ile Met Arg Glu Phe Asn Ser
675 680 685

Tyr Lys Gln Arg Val Lys Tyr Val Glu Asp Lys Val Val Asp Pro Leu
690 695 700

Pro Pro Tyr Leu Lys Gly Leu Asp Asp Leu Met Ser Gly Leu Gly Ala
705 710 715 720

Ala Gly Lys Ala Val Gly Val Ala Ile Gly Ala Val Gly Gly Ala Val
725 730 735

Ala Ser Val Val Glu Gly Val Ala Thr Phe Leu Lys Asn Pro Phe Gly
740 745 750

Ala Phe Thr Ile Ile Leu Val Ala Ile Ala Val Val Ile Ile Thr Tyr
755 760 765

Leu Ile Tyr Thr Arg Gln Arg Arg Leu Cys Thr Gln Pro Leu Gln Asn
770 775 780

Leu Phe Pro Tyr Leu Val Ser Ala Asp Gly Thr Thr Val Thr Ser Gly
785 790 795 800

Ser Thr Lys Asp Thr Ser Leu Gln Ala Pro Pro Ser Tyr Glu Glu Ser
805 810 815

Val Tyr Asn Ser Gly Arg Lys Gly Pro Gly Pro Pro Ser Ser Asp Ala
820 825 830

Ser Thr Ala Ala Pro Pro Tyr Thr Asn Glu Gln Ala Tyr Gln Met Leu
835 840 845

Leu Ala Leu Ala Arg Leu Asp Ala Glu Gln Arg Ala Gln Gln Asn Gly
850 855 860

Thr Asp Ser Leu Asp Gly Gln Thr Gly Thr Gln Asp Lys Gly Gln Lys
865 870 875 880

Pro Asn Leu Leu Asp Arg Leu Arg His Arg Lys Asn Gly Tyr Arg His
885 890 895

Leu Lys Asp Ser Asp Glu Glu Glu Asn Val
900 905

<210> 24
<211> 7585
<212> DNA
<213> Artificial sequence

<220>
<223> HCMV gH ecto domain fused to C-terminal myc-(his)6 tag
Page 40

eolf-seql.txt

```

<400> 24
gacggatcgg gagatctccc gatcccctat ggtgcactct cagtacaatc tgctctgatg      60
ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gagtagtgcg      120
cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc      180
ttagggttag gcgttttgcg ctgcttcgcg atgtacgggc cagatatacg cgttgacatt      240
gattattgac tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata      300
tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc      360
cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc      420
attgacgtca atgggtggag tatttacggg aaactgccc cttggcagta catcaagtgt      480
atcatatgcc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt      540
atgcccagta catgacctta tgggactttc ctacttgga gtacatctac gtattagtca      600
tcgctattac catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg      660
actcacgggg atttccaagt ctccaccca ttgacgtcaa tgggagtttg ttttggcacc      720
aaaatcaacg ggactttcca aaatgtcgta acaactccgc ccattgacg caaatgggcg      780
gtaggcgtgt acgggtggag gtctatataa gcagagctct ctggctaact agagaacca      840
ctgcttactg gcttatcgaa attaatacga ctactatag ggagaccaa gctggctagc      900
gccaccatga ggctggcct gccctcctac ctgatcatcc tggccgtgtg cctgttcagc      960
cacctgctgt ccagcagata cggcgccgag gccgtgagcg agcccctgga caaggctttc     1020
cacctgctgc tgaacaccta cggcagacct atccggtttc tgcgggagaa caccaccag      1080
tgcacctaca acagcagcct gcggaacagc accgtcgtga gagagaacgc catcagcttc      1140
aactttttcc agagctacaa ccagtactac gtgttccaca tgcccagatg cctgtttgcc      1200
ggccctctgg ccgagcagtt cctgaaccag gtggacctga ccgagacact ggaaagatac      1260
cagcagcggc tgaataccta cgccctggtg tccaaggacc tggccagcta ccggtccttt      1320
agccagcagc tcaaggctca ggatagcctc ggcgagcagc ctaccaccgt gcccctccc      1380
atcgacctga gcatcccca cgtgtggatg cctcccaga ccacctca cggctggacc      1440
gagagccaca ccacctccg cctgcacaga cccacttca accagacctg catcctgttc      1500
gacggccacg acctgctgtt tagcaccgtg acccctgcc tgcaccaggg cttctacctg      1560
atcgacgagc tgagatacgt gaagatcacc ctgaccgagg atttcttcgt ggtcaccgtg      1620
tccatcgacg acgacacccc catgctgctg atcttcggcc acctgcccag agtgctgttc      1680
aaggccccct accagcggga caacttcac ctgcggcaga ccgagaagca cgagctgctg      1740
gtgctggtca agaaggacca gctgaaccgg cactcctacc tgaaggaccc cgacttcctg      1800
gacgccgccc tggacttcaa ctacctggac ctgagcgccc tgctgagaaa cagcttcac      1860
agatacgccg tggacgtgct gaagtccgga cggtgccaga tgctcgatcg gcggaccgtg      1920
gagatggcct tcgcctatgc cctcgccctg ttcgccgctg ccagacagga agaggctggc      1980

```

eolf-seql.txt						
gcccaggtgt	cagtgccag	agccctggat	agacaggccg	ccctgctgca	gatccaggaa	2040
ttcatgatca	cctgcctgag	ccagaccccc	cctagaacca	ccctgctgct	gtaccccaca	2100
gccgtggatc	tggccaagag	ggccctgtgg	acccccaacc	agatcaccca	catcacaagc	2160
ctcgtgcggc	tcgtgtacat	cctgagcaag	cagaaccagc	agcacctgat	ccccagtg	2220
gccctgagac	agatcgccga	cttcgccctg	aagctgcaca	agacccatct	ggccagcttt	2280
ctgagcgct	tcgccaggca	ggaactgtac	ctgatgggca	gcctgggtcca	cagcatgctg	2340
gtgcatacca	ccgagcggcg	ggagatcttc	atcgtggaga	caggcctgtg	tagcctggcc	2400
gagctgtccc	actttaccca	gctgctggcc	caccctcacc	acgagtacct	gagcgacctg	2460
tacaccccct	gcagcagcag	cggcagacgg	gaccacagcc	tggaacggct	gaccagactg	2520
ttccccgatg	ccaccgtgcc	tgctacagt	cctgccgccc	tgtccatcct	gtccaccatg	2580
cagcccagca	ccctggaaac	cttccccgac	ctgttctgcc	tggccctggg	cgagagcttt	2640
agcgccctga	ccgtgtccga	gcacgtgtcc	tacatcgtga	ccaatcagta	cctgatcaag	2700
ggcatcagct	accccgtgtc	caccacagtc	gtgggcccaga	gcctgatcat	caccagacc	2760
gacagccaga	ccaagtgcga	gctgacccgg	aacatgcaca	ccacacacag	catcacctg	2820
gccctgaaca	tcagcctgga	aaactgcgct	ttctgtcagt	ctgccctgct	ggaatacgac	2880
gatacccagg	gcgtgatcaa	catcatgtac	atgcacgaca	gcgacgacgt	gctgttcgcc	2940
ctggaccccct	acaacgaggt	ggtggtgtcc	agcccccgga	cccactacct	gatgctgctg	3000
aagaacggca	ccgtgctgga	agtgaccgac	gtggtgggtg	acgccaccga	cggtaccaag	3060
cttggggccg	aacaaaaact	catctcagaa	gaggatctga	atagcgccgt	cgaccatcat	3120
catcatcatc	attgagttta	aacggtctcc	agcttaagtt	taaaccgctg	atcagcctcg	3180
actgtgcctt	ctagttgcca	gccatctgtt	gtttgcccct	cccccgctgc	ttccttgacc	3240
ctggaagggtg	ccactcccac	tgtcctttcc	taataaaatg	aggaaattgc	atcgcatgtg	3300
ctgagtaggt	gtcattctat	tctggggggg	gggggtgggg	aggacagcaa	gggggaggat	3360
tggaagaca	atagcaggca	tgctggggat	gcggtgggct	ctatggcttc	tgaggcgga	3420
agaaccagct	ggggctctag	ggggtatccc	cacgcgccct	gtagcggcgc	attaagcgcg	3480
gcgggtgtgg	tggttacgcg	cagcgtgacc	gctacacttg	ccagcgccct	agcgcccgt	3540
cctttcgctt	tcttcccttc	ctttctcgcc	acgttcgccg	gctttccccg	tcaagctcta	3600
aatcgggggc	tccctttagg	gttccgattt	agtgttttac	ggcacctcga	ccccaaaaaa	3660
cttgattagg	gtgatggttc	acgtagtggg	ccatcgccct	gatagacggg	ttttcgccct	3720
ttgacgttgg	agtccacgtt	ctttaatagt	ggactcttgt	tccaaactgg	aacaacactc	3780
aaccctatct	cggtctattc	ttttgattta	taagggattt	tgccgatttc	ggcctattgg	3840
ttaaaaaatg	agctgattta	acaaaaattt	aacgcgaatt	aattctgtgg	aatgtgtgtc	3900
agttaggggtg	tggaaagtcc	ccaggctccc	cagcaggcag	aagtatgcaa	agcatgcac	3960
tcaattagtc	agcaaccagg	tgtggaaagt	ccccaggctc	cccagcaggc	agaagtatgc	4020

eolf-seql.txt						
aaagcatgca	tctcaattag	tcagcaacca	tagtcccgcc	cctaactccg	cccatcccg	4080
ccctaactcc	gccagttcc	gccattctc	cgcccatgg	ctgactaatt	ttttttat	4140
atgcagaggc	cgaggccgc	tctgcctctg	agctattcca	gaagtagtga	ggaggcttt	4200
ttggaggcct	aggcttttgc	aaaaagctcc	cgggagcttg	tatatccatt	ttcggatctg	4260
atcaagagac	aggatgagga	tcgtttcgca	tgattgaaca	agatggattg	cacgcaggtt	4320
ctccggccgc	ttgggtggag	aggctattcg	gctatgactg	ggcacaacag	acaatcggct	4380
gctctgatgc	cgccgtgttc	cggctgtcag	cgcagggg	cccggttctt	tttgtcaaga	4440
ccgacctgtc	cggtgccctg	aatgaactgc	aggacgaggc	agcgcggtta	tcgtggctgg	4500
ccacgacggg	cgttccttgc	gcagctgtgc	tcgacgttgt	cactgaagcg	ggaagggact	4560
ggctgctatt	gggcgaagtg	ccggggcagg	atctcctgtc	atctcacctt	gctcctgccg	4620
agaaagtatc	catcatggct	gatgcaatgc	ggcggctgca	tacgcttgat	ccggctacct	4680
gccattcga	ccaccaagcg	aaacatcgca	tcgagcgagc	acgtactcgg	atggaagccg	4740
gtcttgtcga	tcaggatgat	ctggacgaag	agcatcaggg	gctcgcgcca	gccgaactgt	4800
tcgccaggct	caaggcgcgc	atgcccgacg	gcgaggatct	cgtcgtgacc	catggcgatg	4860
cctgcttgcc	gaatatcatg	gtggaaaatg	gccgcttttc	tggattcatc	gactgtggcc	4920
ggctgggtgt	ggcggaccgc	tatcaggaca	tagcgttggc	taccctgat	attgctgaag	4980
agcttggcgg	cgaatgggct	gaccgcttcc	tcgtgcttta	cggtatcgcc	gctcccgatt	5040
cgcagcgcat	cgccttctat	cgccttcttg	acgagttctt	ctgagcggga	ctctgggggt	5100
cgaaatgacc	gaccaagcga	cgcccaacct	gccatcacga	gatttcgatt	ccaccgccgc	5160
cttctatgaa	aggttgggct	tcggaatcgt	tttccgggac	gccggctgga	tgatcctcca	5220
gcgcggggat	ctcatgctgg	agttcttcgc	ccaccccaac	ttgtttattg	cagcttataa	5280
tggttacaaa	taaagcaata	gcatcacaaa	tttcacaaat	aaagcatttt	tttactgca	5340
ttctagtgtg	ggtttgtcca	aactcatcaa	tgtatcttat	catgtctgta	taccgtcgac	5400
ctctagctag	agcttggcgt	aatcatggtc	atagctgttt	cctgtgtgaa	attgttatcc	5460
gctcacaatt	ccacacaaca	tacgagccgg	aagcataaag	tgtaaagcct	ggggtgccta	5520
atgagtgagc	taactcacat	taattgcgtt	gcgctcactg	cccgccttcc	agtcgggaaa	5580
cctgtcgtgc	cagctgcatt	aatgaatcgg	ccaacgcgcg	gggagaggcg	gtttgcgtat	5640
tgggcgctct	tccgcttcct	cgctcactga	ctcgtgcgc	tcggtcgttc	ggctgcggcg	5700
agcggtatca	gctcactcaa	aggcggtaat	acggttatcc	acagaatcag	gggataacgc	5760
aggaaagaac	atgtgagcaa	aaggccagca	aaaggccagg	aaccgtaaaa	aggccgcgtt	5820
gctggcgttt	ttccataggc	tccgcccccc	tgacgagcat	cacaaaaatc	gacgctcaag	5880
tcagaggtgg	cgaaacccga	caggactata	aagataccag	gcgtttcccc	ctggaagctc	5940
cctcgtgcgc	tctcctgttc	cgaccctgcc	gcttaccgga	tacctgtccg	cctttctccc	6000
ttcgggaagc	gtggcgcttt	ctcatagctc	acgctgtagg	tatctcagtt	cgggtgtaggt	6060

eolf-seql.txt

cgttcgctcc aagctgggct gtgtgcacga acccccgcgtt cagcccgcacc gctgcgcctt	6120
atccggtaac tatcgtcttg agtccaaccc ggtaagacac gacttatcgc cactggcagc	6180
agccactggg aacaggatta gcagagcgag gtatgtaggc ggtgctacag agttcttgaa	6240
gtggtggcct aactacggct aactagaag aacagtattt ggtatctgcg ctctgctgaa	6300
gccagttacc ttcggaaaaa gagttggtag ctcttgatcc ggcaaaaaa ccaccgctgg	6360
tagcggtggt ttttttgttt gcaagcagca gattacgcgc agaaaaaaag gatctcaaga	6420
agatcctttg atcttttcta cggggctctga cgctcagtgg aacgaaaact cacgttaagg	6480
gattttggtc atgagattat caaaaaggat cttcacctag atccttttaa attaaaaatg	6540
aagttttaa tcaatctaaa gtatatatga gtaaacttgg tctgacagtt accaatgctt	6600
aatcagtgag gcacctatct cagcgatctg tctatttcgt tcatccatag ttgcctgact	6660
ccccgtcgtg tagataacta cgatacggga gggcttacca tctggcccca gtgctgcaat	6720
gataccgcga gaccacgct caccggctcc agatttatca gcaataaacc agccagccgg	6780
aagggccgag cgcagaagtg gtcctgcaac tttatccgcc tccatccagt ctattaattg	6840
ttgccgggaa gctagagtaa gtagttcgcc agttaatagt ttgcgcaacg ttgttgccat	6900
tgctacaggc atcgtggtgt cacgctcgtc gtttggtatg gcttcattca gctccggttc	6960
ccaacgatca aggcgagtta catgatcccc catgttgtgc aaaaaagcgg ttagctcctt	7020
cggtcctccg atcgttgtca gaagtaagtt ggccgcagtg ttatcactca tggttatggc	7080
agcactgcat aattctctta ctgtcatgcc atccgtaaga tgcttttctg tgactggtga	7140
gtactcaacc aagtcattct gagaatagtg tatgcggcga ccgagttgct cttgcccggc	7200
gtcaatacgg gataatacgg cgccacatag cagaacttta aaagtgtca tcattggaaa	7260
acgttcttcg gggcgaaaac tctcaaggat cttaccgctg ttgagatcca gttcgatgta	7320
accactcgt gcaccaact gatcttcagc atcttttact ttcaccagcg tttctgggtg	7380
agcaaaaaca ggaaggcaaa atgccgcaaa aaaggggaata agggcgacac ggaaatgttg	7440
aatactcata ctcttccttt ttcaatatta ttgaagcatt tatcagggtt attgtctcat	7500
gagcggatac atatttgaat gtatttagaa aaataaaca ataggggttc cgcgcacatt	7560
tccccgaaaa gtgccacctg acgtc	7585

<210> 25
 <211> 5156
 <212> DNA
 <213> Human cytomegalovirus

<400> 25	
gccgcggaat ttcgactcta ggccattgca tacgttgtat ctatatcata atatgtacat	60
ttatattggc tcatgtccaa tatgaccgcc atgttgacat tgattattga ctagttatta	120
atagtaatca attacggggt ctagtttca tagcccatat atggagttcc gcgttacata	180
acttacggta aatggccgc ctggctgacc gcccaacgac ccccgcccat tgacgtcaat	240
aatgacgtat gttcccatag taacgccaat agggactttc cattgacgtc aatgggtgga	300

eolf-seql.txt

gtatttacgg	taaactgccc	acttggcagt	acatcaagtg	tatcatatgc	caagtccgcc	360
ccctattgac	gtcaatgacg	gtaaatggcc	cgcttggcat	tatgcccagt	acatgacctt	420
acgggacttt	cctacttggc	agtacatcta	cgtattagtc	atcgctatta	ccatggtgat	480
gcggttttgg	cagtacacca	atgggcgtgg	atagcggttt	gactcacggg	gatttccaag	540
tctccacccc	attgacgtca	atgggagttt	gttttggcac	caaaatcaac	gggactttcc	600
aaaatgtcgt	aataaccccc	ccccgttgac	gcaaatgggc	ggtaggcgtg	tacggtggga	660
ggtctatata	agcagagctc	gtttagttaa	ccgtcagatc	gcctggagac	gccatccacg	720
ctgttttgac	ctccatagaa	gacaccggga	ccgatccagc	ctccgcggcc	gggaacggtg	780
cattggaacg	cggattcccc	gtgccaagag	tgacgtaagt	accgcctata	gactctatag	840
gcacaccctt	ttggctctta	tgcatgctat	actgtttttg	gcttggggcc	tatacacccc	900
cgcttcctta	tgctataggt	gatggtatag	cttagcctat	aggtgtgggt	tattgaccat	960
tattgaccac	tcccctattg	gtgacgatac	tttccattac	taatccataa	catggctctt	1020
tgccacaact	atctctattg	gctatatgcc	aatactctgt	ccttcagaga	ctgacacgga	1080
ctctgtatth	ttacaggatg	gggtcccat	tattatthac	aaattcacat	atacaacaac	1140
gccgtccccc	gtgcccgcag	tttttattaa	acatagcgtg	ggatctccac	gcgaatctcg	1200
ggtacgtgtt	ccggacatgg	gctcttctcc	ggtagcggcg	gagcttccac	atccgagccc	1260
tggtcccatg	cctccagcgg	ctcatggctg	ctcggcagct	ccttgctcct	aacagtggag	1320
gccagactta	ggcacagcac	aatgcccacc	accaccagtg	tgccgcacaa	ggccgtggcg	1380
gtagggatatg	tgtctgaaaa	tgagctcgga	gattgggctc	gcaccgctga	cgcagatgga	1440
agacttaagg	cagcggcaga	agaagatgca	ggcagctgag	ttgttgattt	ctgataagag	1500
tcagaggtaa	ctcccgttgc	ggtgctgtta	acggtggagg	gcagtgtagt	ctgagcagta	1560
ctcgttgctg	ccgcgcgcgc	caccagacat	aatagctgac	agactaacag	actgttcctt	1620
tccatgggtc	ttttctgcag	tcaccgtcgt	cgacgccacc	atgtgcagaa	ggcccgaactg	1680
cggcttcagc	ttcagccctg	gaccctgat	cctgctgtgg	tgctgcctgc	tgctgcctat	1740
cgtgtcctct	gccgccgtgt	ctgtggcccc	tacagccgcc	gagaagggtg	cagccgagtg	1800
ccccgagctg	accagaagat	gcctgctggg	cgaggtgttc	gagggcgaca	agtacgagag	1860
ctggctgcgg	cccctggtca	acgtgaccgg	cagagatggc	cccctgagcc	agctgatccg	1920
gtacagaccc	gtgacccccg	aggccgccaa	tagcgtgctg	ctggacgagg	ccttcctgga	1980
taccctggcc	ctgctgtaca	acaaccccga	ccagctgaga	gccctgctga	ccctgctgtc	2040
cagcgacacc	gccccagat	ggatgaccgt	gatgcggggc	tacagcgagt	gtggagatgg	2100
cagccctgcc	gtgtacacct	gcgtggacga	cctgtgcaga	ggctacgacc	tgaccagact	2160
gagctacggc	cgggtccatct	tcacagagca	cgtgctgggc	ttcgagctgg	tgccccccag	2220
cctgttcaac	gtggtggtgg	ccatccggaa	cgaggccacc	agaaccaaca	gagccgtgcg	2280
gctgcctgtg	tctacagccg	ctgcacctga	gggcatcaca	ctgttctacg	gcctgtacaa	2340

eolf-seql.txt

cgccgtgaaa	gagttctgcc	tccggcacca	gctggatccc	cccctgctga	gacacctgga	2400
caagtactac	gccggcctgc	ccccagagct	gaagcagacc	agagtgaacc	tgcccgccca	2460
cagcagatat	ggccctcagg	ccgtggacgc	cagatgataa	tctagaaagc	catggatata	2520
ggatccacta	cgcgttagag	ctcgtgatc	agcctcgact	gtgccttcta	gttgccagcc	2580
atctgttgtt	tgccctcccc	ccgtgccttc	cttgaccctg	gaaggtgcca	ctcccactgt	2640
cctttcctaa	taaaatgagg	aaattgcatc	gcattgtctg	agtaggtgtc	attctattct	2700
gggggggtggg	gtggggcagg	acagcaaggg	ggaggattgg	gaagacaata	gcaggggggt	2760
gggcgaagaa	ctccagcatg	agatccccgc	gctggaggat	catccagccg	gcgtcccgga	2820
aaacgattcc	gaagcccaac	ctttcataga	aggcggcggt	ggaatcgaag	tctcgtgatg	2880
gcaggttggg	cgtcgcttgg	tcggtcattt	cgaacccag	agtcccgtc	agaagaactc	2940
gtcaagaagg	cgatagaagg	cgatgcgctg	cgaatcgga	gcggcgatac	cgtaaagcac	3000
gaggaagcgg	tcagcccatt	cgccgccaag	ctcttcagca	atatcacggg	tagccaacgc	3060
tatgtcctga	tagcgggccg	ccacaccag	ccggccacag	tcgatgaatc	cagaaaagcg	3120
gccattttcc	accatgatata	tcggcaagca	ggcatcgcca	tgggtcacga	cgagatcctc	3180
gccgtcgggc	atgcgcgcct	tgagcctggc	gaacagttcg	gctggcgcca	gcccctgatg	3240
ctcttcgtcc	agatcatcct	gatcgacaag	accggcttcc	atccgagtac	gtgctcgctc	3300
gatgcgatgt	ttcgcttggg	ggtcgaatgg	gcaggtagcc	ggatcaagcg	tatgcagccg	3360
ccgcattgca	tcagccatga	tggatacttt	ctcggcagga	gcaaggtgag	atgacaggag	3420
atcctgcccc	ggcacttcgc	ccaatagcag	ccagtccctt	cccgtttcag	tgacaacgtc	3480
gagcacagct	gcgcaaggaa	cgcccgctcg	ggccagccac	gatagccgcg	ctgcctcgctc	3540
ctgcagttca	ttcagggcac	cggacaggtc	ggtcttgaca	aaaagaaccg	ggcgcccctg	3600
cgctgacagc	cggaaacacg	cggcatcaga	gcagccgatt	gtctgtttgt	cccagtcata	3660
gccgaatagc	ctctccaccc	aagcggcccg	agaacctg	tgcaatccat	cttgttcaat	3720
catgcgaaac	gatcctcatc	ctgtctcttg	atcagatctt	gatcccctgc	gccatcagat	3780
ccttggcggc	aagaaagcca	tccagtttac	tttgcagggc	ttcccaacct	taccagaggg	3840
cgccccagct	ggcaattccg	gttcgcttgc	tgtccataaa	accgcccagt	ctagctatcg	3900
ccatgtaagc	ccactgcaag	ctacctgctt	tctctttg	cttgcgtttt	cccttgtcca	3960
gatagcccag	tagctgacat	tcacccggg	tcagcaccgt	ttctgcggac	tggctttcta	4020
cgtgttccgc	ttcctttagc	agcccttg	ccctgagtg	ttgcggcagc	gtgaagctaa	4080
ttcatgggta	aatttttgtt	aatcagctc	atcttttaac	caataggccg	aaatcggcaa	4140
aatcccttat	aatcaaaaag	aatagcccga	gataggggtg	agtgttggtc	cagtttgtaa	4200
caagagtcca	ctattaaaga	acgtggactc	caacgtcaaa	gggcgaaaaa	ccgtctatca	4260
gggcgatggc	cggatcagct	tatgcggtgt	gaaataccgc	acagatgcgt	aaggagaaaa	4320
taccgcatca	ggcgctcttc	cgcttcctcg	ctcactgact	cgctgcgctc	ggcgttcggt	4380

eolf-seql.txt

ctgcggcgag	cggtatcagc	tcactcaaag	gcggtaat	ggttatccac	agaatcaggg	4440
gataacgcag	gaaagaacat	gtgagcaaaa	ggccagcaaa	aggccaggaa	ccgtaaaaag	4500
gccgcgttgc	tggcgttttt	ccataggctc	cgccccctg	acgagcatca	caaaaatcga	4560
cgctcaagtc	agaggtggcg	aaacccgaca	ggactataaa	gataccaggc	gtttccccct	4620
ggaagctccc	tcgtgcgctc	tcctgtttccg	accctgccgc	ttaccggata	cctgtccgcc	4680
tttctccctt	cgggaagcgt	ggcgctttct	catagctcac	gctgtaggta	tctcagttcg	4740
gtgtaggtcg	ttcgctccaa	gctgggctgt	gtgcacgaac	ccccgttca	gcccgaccgc	4800
tgcgccttat	ccggtaaact	tcgtcttgag	tccaacccgg	taagacacga	cttatcgcca	4860
ctggcagcag	ccactggtaa	caggattagc	agagcgaggt	atgtaggcgg	tgctacagag	4920
ttcttgaagt	ggtggcctaa	ctacggctac	actagaagga	cagtatttgg	tatctgcgct	4980
ctgctgaagc	cagttacctt	cggaaaaaga	gttggtagct	cttgatccgg	caaacaaacc	5040
accgctggta	gcggtggttt	ttttgtttgc	aagcagcaga	ttacgcgcag	aaaaaaagga	5100
tctcaagaag	atcctttgat	cttttctact	gaacggtgat	ccccaccgga	attgcg	5156

<210> 26
 <211> 4835
 <212> DNA
 <213> Human cytomegalovirus

<400> 26						
gccgcggaat	ttcgactcta	ggccattgca	tacgttgtat	ctatatcata	atatgtacat	60
ttatattggc	tcatgtccaa	tatgaccgcc	atgttgacat	tgattattga	ctagttatta	120
atagtaatca	attacgggggt	cattagttca	tagcccatat	atggagttcc	gcgttacata	180
acttacggta	aatggcccg	ctggctgacc	gccaacgac	ccccgccc	tgacgtcaat	240
aatgacgtat	gttcccatag	taacgccaat	agggactttc	cattgacgtc	aatgggtgga	300
gtatttacgg	taaactgccc	acttggcagt	acatcaagt	tatcatatgc	caagtccgcc	360
ccctattgac	gtcaatgacg	gtaaatggcc	cgcctggcat	tatgccag	acatgacctt	420
acgggacttt	cctacttggc	agtacatcta	cgtattagtc	atcgctatta	ccatggtgat	480
gcggttttgg	cagtacacca	atgggcgtgg	atagcggttt	gactcacggg	gatttccaag	540
tctccacccc	attgacgtca	atgggagttt	gttttggcac	caaaatcaac	gggactttcc	600
aaaatgtcgt	aataaccccg	ccccgttgac	gcaaattgggc	ggtaggcgtg	tacggtggga	660
ggtctatata	agcagagctc	gtttagtga	ccgtcagatc	gcctggagac	gccatccacg	720
ctgttttgac	ctccatagaa	gacaccggga	ccgatccagc	ctccgcggcc	gggaacggtg	780
cattggaacg	cggattcccc	gtgccaagag	tgacgtaagt	accgcctata	gactctatag	840
gcacaccctt	ttggctctta	tgcattgctat	actgtttttg	gcttggggcc	tatacacc	900
cgcttcctta	tgctataggt	gatggtatag	cttagcctat	aggtgtgggt	tattgaccat	960
tattgaccac	tcccctattg	gtgacgatac	tttccattac	taatccataa	catggctctt	1020

eolf-seql.txt						
tgccacaact	atctctattg	gctatatgcc	aatactctgt	ccttcagaga	ctgacacgga	1080
ctctgtat	ttacaggatg	gggtcccat	tattatttac	aaattcacat	atacaacaac	1140
gccgtcccc	gtgcccgcag	tttttattaa	acatagcg	ggatctccac	gcgaatctcg	1200
ggtacgtgt	ccggacatgg	gctcttctcc	ggtagcggcg	gagcttccac	atccgagccc	1260
tgggtcccat	cctccagcgg	ctcatgggtc	ctcggcagct	ccttgctcct	aacagtggag	1320
gccagactta	ggcacagcac	aatgcccacc	accaccagtg	tgccgcacaa	ggccgtggcg	1380
gtaggggatg	tgtctgaaaa	tgagctcgga	gattgggctc	gcaccgctga	cgcagatgga	1440
agacttaagg	cagcggcaga	agaagatgca	ggcagctgag	ttgttgtatt	ctgataagag	1500
tcagaggtaa	ctcccgttgc	ggtgctgtta	acggtggagg	gcagtgtagt	ctgagcagta	1560
ctcgttgctg	ccgcgcgcgc	caccagacat	aatagctgac	agactaacag	actgttcctt	1620
tccatgggtc	ttttctgcag	tcaccgtcgt	cgacgccacc	atgagcccca	aggacctgac	1680
ccccttcctg	acaaccctgt	ggctgctcct	gggccatagc	agagtgccta	gagtgcgggc	1740
cgaggaatgc	tgcgagttca	tcaacgtgaa	ccaccccccc	gagcggtgct	acgacttcaa	1800
gatgtgcaac	cgggttcaccg	tggccctgag	atgccccgac	ggcgaagtgt	gctacagccc	1860
cgagaaaacc	gccgagatcc	ggggcatcgt	gaccaccatg	accacagcc	tgacccgga	1920
ggtggtgcac	aacaagctga	ccagctgcaa	ctacaacccc	ctgtacctgg	aagccgacgg	1980
ccggatcaga	tgcgggcaag	tgaacgacaa	ggcccagtag	ctgctgggag	ccgccggaag	2040
cgtgccctac	cgggtggatca	acctggaata	cgacaagatc	accgggatcg	tgggcctgga	2100
ccagtacctg	gaaagcgtga	agaagcacaa	gcggctggac	gtgtgcagag	ccaagatggg	2160
ctacatgctg	cagtataat	ctagaaagcc	atggatatcg	gatccactac	gcgttagagc	2220
tcgctgatca	gcctcgactg	tgccttctag	ttgccagcca	tctgttgttt	gcccctcccc	2280
cgtgccttcc	ttgaccctgg	aaggtgccac	tcccactgtc	ctttccta	aaaatgagga	2340
aattgcatcg	cattgtctga	gtaggtgtca	ttctattctg	gggggtgggg	tggggcagga	2400
cagcaagggg	gaggattggg	aagacaatag	caggggggtg	ggcgaagaac	tccagcatga	2460
gatccccgcg	ctggaggatc	atccagccgg	cgtcccggaa	aacgattccg	aagcccaacc	2520
tttcatagaa	ggcggcggtg	gaatcgaaat	ctcgtgatgg	caggttgggc	gtcgcttggt	2580
cggtcatttc	gaaccccaga	gtcccgtc	gaagaactcg	tcaagaaggc	gatagaaggc	2640
gatgcgctgc	gaatcgggag	cggcgatacc	gtaaagcacg	aggaagcgg	cagcccattc	2700
gccgccaagc	tcttcagcaa	tatcacgggt	agccaacgct	atgtcctgat	agcggtcgc	2760
cacacccagc	cggccacagt	cgatgaatcc	agaaaagcgg	ccattttcca	ccatgatatt	2820
cggcaagcag	gcatcgccat	gggtcacgac	gagatcctcg	ccgtcgggca	tgcgcgcctt	2880
gagcctggcg	aacagttcgg	ctggcgcgag	cccctgatgc	tcttcgtcca	gatcatcctg	2940
atcgacaaga	ccggcttcca	tccgagtacg	tgctcgctcg	atgcgatgtt	tcgcttggtg	3000
gtcgaatggg	caggtagccg	gatcaagcgt	atgcagccgc	cgcattgcat	cagccatgat	3060

eolf-seql.txt

ggatactttc	tcggcaggag	caaggtgaga	tgacaggaga	tcctgccccg	gcacttcgcc	3120
caatagcagc	cagtcccttc	ccgcttcagt	gacaacgtcg	agcacagctg	cgcaaggaac	3180
gcccgtcgtg	gccagccacg	atagccgcgc	tgctctgtcc	tgcagttcat	tcagggcacc	3240
ggacaggtcg	gtcttgacaa	aaagaaccgg	gcgcccctgc	gctgacagcc	ggaacacggc	3300
ggcatcagag	cagccgattg	tctgtttgtg	ccagtcatag	ccgaatagcc	tctccaccca	3360
agcggccgga	gaacctgcgt	gcaatccatc	ttgttcaatc	atgcgaaacg	atcctcatcc	3420
tgtctcttga	tcagatcttg	atccccctgc	ccatcagatc	cttggcggca	agaaagccat	3480
ccagtttact	ttgcaggggt	tcccaacctt	accagagggc	gccccagctg	gcaattccgg	3540
ttcgcttgct	gtccataaaa	ccgcccagtc	tagctatcgc	catgtaagcc	cactgcaagc	3600
tacctgcttt	ctctttgcgc	ttgcgttttc	ccttgtccag	atagcccagt	agctgacatt	3660
catccggggg	cagcaccgtt	tctgcggact	ggctttctac	gtgttccgct	tccttttagca	3720
gcccttgctc	cctgagtgct	tgccggcagc	tgaagcta	tcattggtta	atctttgtta	3780
aatcagctca	ttttttaacc	aataggccga	aatcggcaaa	atcccttata	aatcaaaaga	3840
atagcccag	ataggggtga	gtgtttgttc	agtttggaa	aagagtccac	tattaaagaa	3900
cgtggactcc	aacgtcaaag	ggcgaaaaac	cgtctatcag	ggcgatggcc	ggatcagctt	3960
atgcggtgtg	aaataccgca	cagatgcgta	aggagaaaat	accgcatcag	gcgctcttcc	4020
gcttcctcgc	tactgactc	gctgcgctcg	gtcgttcggc	tgccggcagc	ggatcagctt	4080
cactcaaagg	cggtaatacg	gttatccaca	gaatcagggg	ataacgcagg	aaagaacatg	4140
tgagcaaaa	gccagcaaaa	ggccaggaac	cgtaaaaagg	ccgcgttgct	ggcgtttttc	4200
cataggctcc	gccccctga	cgagcatcac	aaaaatcgac	gctcaagtca	gaggtggcga	4260
aacccgacag	gactataaag	ataccaggcg	tttccccctg	gaagctccct	cgtgcgctct	4320
cctgttccga	ccctgccgct	taccggatac	ctgtccgcct	ttctcccttc	gggaagcgtg	4380
gcgctttctc	atagctcacg	ctgtaggtat	ctcagttcgg	tgtaggtcgt	tcgctccaag	4440
ctgggctgtg	tgcacgaacc	ccccgttcag	cccgaccgct	gcgccttata	cggtaactat	4500
cgtcttgagt	ccaacccggt	aagacacgac	ttatcgccac	tggcagcagc	cactggtaac	4560
aggattagca	gagcgaggta	tgtaggcggt	gctacagagt	tcttgaagtg	gtggcctaac	4620
tacggctaca	ctagaaggac	agtatttggg	atctgcgctc	tgctgaagcc	agttaccttc	4680
ggaaaaagag	ttggtagctc	ttgatccggc	aaacaaacca	ccgctggtag	cggtggtttt	4740
tttgtttgca	agcagcagat	tacgcgcaga	aaaaaaggat	ctcaagaaga	tcctttgatc	4800
ttttctactg	aacggtgatc	cccaccggaa	ttgcg			4835

<210> 27
 <211> 4964
 <212> DNA
 <213> Human cytomegalovirus

<400> 27
 gccgcggaat ttcgactcta ggccattgca tacgttgtat ctatatcata atatgtacat 60

eolf-seql.txt

ttatatattggc	tcatgtccaa	tatgaccgcc	atgttgacat	tgattattga	ctagttatta	120
atagtaatca	attacgggggt	cattagttca	tagcccatat	atggagttcc	gcgttacata	180
acttacggta	aatggcccg	ctggctgacc	gccaacgac	ccccgccc	tgacgtcaat	240
aatgacgtat	gttcccatag	taacgccaat	agggactttc	cattgacgtc	aatgggtgga	300
gtatttacgg	taaactgccc	acttggcagt	acatcaagt	tatcatatgc	caagtccgcc	360
ccctattgac	gtcaatgacg	gtaaatggcc	cgcctggcat	tatgcccagt	acatgacctt	420
acgggacttt	cctacttggc	agtacatcta	cgtattagtc	atcgctatta	ccatgggtgat	480
gcggttttgg	cagtacacca	atgggcgtgg	atagcggttt	gactcacggg	gattttccaag	540
tctccacccc	attgacgtca	atgggagttt	gttttggcac	caaaatcaac	gggactttcc	600
aaaatgtcgt	aataacccc	ccccgttgac	gcaaattggc	ggtaggcgtg	tacggtggga	660
ggctctatata	agcagagctc	gtttagtga	ccgtcagatc	gcctggagac	gccatccacg	720
ctgttttgac	ctccatagaa	gacaccggga	ccgatccagc	ctccgcgccc	gggaacgggtg	780
cattggaacg	cggattcccc	gtgccaagag	tgacgtaagt	accgcctata	gactctatag	840
gcacaccctt	ttggctctta	tgcattgctat	actgtttttg	gcttggggcc	tatacacccc	900
cgcttcctta	tgctataggt	gatgggtatag	cttagcctat	agggtgtgggt	tattgaccat	960
tattgaccac	tcccctattg	gtgacgatac	tttccattac	taatccataa	catggctctt	1020
tgccacaact	atctctattg	gctatatgcc	aatactctgt	ccttcagaga	ctgacacgga	1080
ctctgtattt	ttacaggatg	gggtcccatt	tattattttac	aaattcacat	atacaacaac	1140
gccgtcccc	gtgcccgcag	tttttattaa	acatagcgtg	ggatctccac	gcgaatctcg	1200
ggtagctgtt	ccggacatgg	gctcttctcc	ggtagcggcg	gagcttccac	atccgagccc	1260
tggtcccatg	cctccagcgg	ctcatggctg	ctcggcagct	ccttgctcct	aacagtggag	1320
gccagactta	ggcacagcac	aatgcccacc	accaccagt	tgccgcacaa	ggccgtggcg	1380
gtagggtatg	tgtctgaaaa	tgagctcgga	gattgggctc	gcaccgctga	cgcagatgga	1440
agacttaagg	cagcggcaga	agaagatgca	ggcagctgag	ttgttgtatt	ctgataagag	1500
tcagaggtaa	ctcccgttgc	ggtgctgtta	acgggtggagg	gcagtgtagt	ctgagcagta	1560
ctcgttgctg	ccgcgcgcgc	caccagacat	aatagctgac	agactaacag	actgttcctt	1620
tccatgggtc	ttttctgcag	tcaccgtcgt	cgacgccacc	atgctgcggc	tgctgctgag	1680
acaccacttc	cactgcctgc	tgctgtgtgc	cgtgtggggc	acccttgctc	tgccagccc	1740
ttggagcacc	ctgaccgcca	accagaacc	tagccccctt	tggtccaagc	tgacctacag	1800
caagccccac	gacgccgcca	ccttctactg	cccctttctg	tacccagcc	ctcccagaag	1860
ccccctgcag	ttcagcggct	tccagagagt	gtccaccggc	cctgagtgcc	ggaacgagac	1920
actgtacctg	ctgtacaacc	gggagggcca	gacactgggt	gagcggagca	gcacctgggt	1980
gaaaaaagt	atctggatc	tgagcggccg	gaaccagacc	atcctgcagc	ggatgcccag	2040
aaccgccagc	aagcccagcg	acggcaacgt	gcagatcagc	gtggaggacg	ccaaaatctt	2100

eolf-seql.txt

cggcgcccac	atggtgccc	agcagaccaa	gctgctgaga	ttcgtggtca	acgacggcac	2160
cagatatcag	atgtgcgtga	tgaagctgga	aagctgggcc	cacgtgttcc	gggactactc	2220
cgtgagcttc	caggtccggc	tgaccttcac	cgaggccaac	aaccagacct	acaccttctg	2280
caccacccc	aacctgatcg	tgtgataatc	tagaaagcca	tggatatcgg	atccactacg	2340
cgtagagct	cgctgatcag	cctcgactgt	gccttctagt	tgccagccat	ctgttgtttg	2400
cccctcccc	gtgccttcct	tgaccctgga	aggtgccact	cccactgtcc	tttcctaata	2460
aaatgaggaa	attgcatcgc	attgtctgag	taggtgtcat	tctattctgg	ggggtgggg	2520
ggggcaggac	agcaaggggg	aggattggga	agacaatagc	aggggggtgg	gcgaagaact	2580
ccagcatgag	atccccgcgc	tggaggatca	tccagccggc	gtcccggaaa	acgattccga	2640
agcccaacct	ttcatagaag	gcggcggtgg	aatcgaaatc	tcgtgatggc	aggttggg	2700
tcgcttggtc	ggtcatttcg	aaccccagag	tcccgctcag	aagaactcgt	caagaaggcg	2760
atagaaggcg	atgcgctgcg	aatcgggagc	ggcgataccg	taaagcacga	ggaagcggtc	2820
agcccatcgc	ccgccaagct	cttcagcaat	atcacgggta	gccaacgcta	tgtcctgata	2880
gcgggtccgc	acaccagcc	ggccacagtc	gatgaatcca	gaaaagcggc	cattttccac	2940
catgatattc	ggcaagcagg	catcgccatg	ggtcacgacg	agatcctcgc	cgtcgggcat	3000
gcgcgccttg	agcctggcga	acagttcggc	tggcgcgagc	ccctgatgct	cttcgtccag	3060
atcatcctga	tcgacaagac	cggcttccat	ccgagtacgt	gctcgctcga	tgcgatgttt	3120
cgcttggtgg	tcgaatgggc	aggtagccgg	atcaagcgta	tgcagccgcc	gcattgcac	3180
agccatgatg	gatactttct	cggcaggagc	aaggtagagat	gacaggagat	cctgccccgg	3240
cacttcgccc	aatagcagcc	agtcccttcc	cgcttcagtg	acaacgtcga	gcacagctgc	3300
gcaaggaacg	cccgtcgtgg	ccagccacga	tagccgcgct	gcctcgtcct	gcagttcatt	3360
cagggcaccg	gacaggtcgg	tcttgacaaa	aagaaccggg	cgcccctgcg	ctgacagccg	3420
gaacacggcg	gcatcagagc	agccgattgt	ctgttggtgc	cagtcatagc	cgaatagcct	3480
ctccacccaa	gcggccggag	aacctgcgtg	caatccatct	tgttcaatca	tgcgaaacga	3540
tcctcatcct	gtctcttgat	cagatcttga	tcccctgcgc	catcagatcc	ttggcggcaa	3600
gaaagccatc	cagtttactt	tgcagggtct	cccaacctta	ccagagggcg	ccccagctgg	3660
caattccggt	tcgcttgctg	tccataaaac	cgcccagctc	agctatcgcc	atgtaagccc	3720
actgcaagct	acctgctttc	tctttgcgct	tgcgttttcc	cttgtccaga	tagcccagta	3780
gctgacattc	atccgggggtc	agcaccgttt	ctgcggactg	gctttctacg	tgttccgctt	3840
cctttagcag	cccttgccgc	ctgagtgtct	gcggcagcgt	gaagctaatt	catggttaaa	3900
tttttgttaa	atcagctcat	tttttaacca	ataggccgaa	atcggaacaa	ttccttataa	3960
atcaaaagaa	tagcccgaga	taggggtgag	tgttggtcca	gtttggaaca	agagtccact	4020
attaaagaac	gtggactcca	acgtcaaagg	gcgaaaaacc	gtctatcagg	gcgatggccg	4080
gatcagctta	tgcgggtgtga	aataccgcac	agatgcgtaa	ggagaaaata	ccgcatcagg	4140

eolf-seql.txt

cgctcttccg	cttcctcget	cactgactcg	ctgcgctcgg	tcgttcggct	gcggcgagcg	4200
gtatcagctc	actcaaaggc	ggtaatacgg	ttatccacag	aatcagggga	taacgcagga	4260
aagaacatgt	gagcaaaaagg	ccagcaaaaag	gccaggaacc	gtaaaaaggc	cgcgttgctg	4320
gcgtttttcc	ataggctccg	ccccctgac	gagcatcaca	aaaatcgacg	ctcaagtcag	4380
aggtggcgaa	acccgacagg	actataaaga	taccaggcgt	ttccccctgg	aagctccctc	4440
gtgcgctctc	ctgttccgac	cctgccgctt	accggatacc	tgtccgcctt	tctcccttcg	4500
ggaagcgtgg	cgcttttctca	tagctcacgc	tgtaggtatc	tcagttcggg	gtaggtcggt	4560
cgctccaagc	tgggctgtgt	gcacgaaccc	cccgttcagc	ccgaccgctg	cgctttatcc	4620
ggtaactatc	gtcttgagtc	caacccggta	agacacgact	tatcgccact	ggcagcagcc	4680
actggtaaca	ggattagcag	agcgaggtat	gtaggcggtg	ctacagagtt	cttgaagtgg	4740
tggcctaact	acggctacac	tagaaggaca	gtattttggt	tctgcgctct	gctgaagcca	4800
gttaccttcg	gaaaaagagt	tggtagctct	tgatccggca	aacaaaccac	cgctggtagc	4860
ggtggttttt	ttgtttgcaa	gcagcagatt	acgcgcagaa	aaaaaggatc	tcaagaagat	4920
cctttgatct	tttctactga	acggtgatcc	ccaccggaat	tgcg		4964

<210> 28
 <211> 4709
 <212> DNA
 <213> Human cytomegalovirus

<400> 28	
gccgcggaat	ttcgactcta ggccattgca tacgtttgat ctatatcata atatgtacat 60
ttatattggc	tcatgtccaa tatgaccgcc atgttgacat tgattattga ctagttatta 120
atagtaatca	attacgggggt cattagttca tagcccatat atggagttcc gcgttacata 180
acttacggta	aatggcccg cctggctgacc gcccaacgac ccccgcccat tgacgtcaat 240
aatgacgtat	gttcccatag taacgccaat agggactttc cattgacgtc aatgggtgga 300
gtatttacgg	taaactgccc acttggcagt acatcaagtg tatcatatgc caagtccgcc 360
ccctattgac	gtcaatgacg gtaaattggc cgcttggcat tatgcccagt acatgacctt 420
acgggacttt	cctacttggc agtacatcta cgtattagtc atcgctatta ccatggtgat 480
gcggtttttg	cagtacacca atgggcgtgg atagcggttt gactcacggg gattttccaag 540
tctccacccc	attgacgtca atgggagttt gttttggcac caaatcaac gggactttcc 600
aaaatgtcgt	aataaccccg ccccgttgac gcaaatgggc ggtaggcgtg tacggtggga 660
ggtctatata	agcagagctc gtttagtgaa ccgtcagatc gcctggagac gccatccacg 720
ctgttttgac	ctccatagaa gacaccggga ccgatccagc ctccgcggcc gggaacggtg 780
cattggaacg	cggattcccc gtgccaagag tgacgtaagt accgcctata gactctatag 840
gcacaccctt	ttggctctta tgcattgctat actgtttttg gcttggggcc tatacacccc 900
cgcttcctta	tgctataggt gatggtatag cttagcctat aggtgtgggt tattgaccat 960

eolf-seql.txt						
tattgaccac	tcccctattg	gtgacgatac	tttccattac	taatccataa	catggctctt	1020
tgccacaact	atctctattg	gctatatgcc	aatactctgt	ccttcagaga	ctgacacgga	1080
ctctgtatth	ttacaggatg	gggtcccat	tattatthac	aaattcacat	atacaacaac	1140
gccgtcccc	gtgcccgcag	tttttattaa	acatagcgtg	ggatctccac	gcgaatctcg	1200
ggtacgtgtt	ccggacatgg	gctcttctcc	ggtagcggcg	gagcttccac	atccgagccc	1260
tggtcccatg	cctccagcgg	ctcatggctg	ctcggcagct	ccttgctcct	aacagtggag	1320
gccagactta	ggcacagcac	aatgcccacc	accaccagtg	tgccgcacaa	ggccgtggcg	1380
gtaggggatg	tgtctgaaaa	tgagctcgga	gattgggctc	gcaccgctga	cgcagatgga	1440
agacttaagg	cagcggcaga	agaagatgca	ggcagctgag	ttgttgatt	ctgataagag	1500
tcagaggtaa	ctcccgttgc	ggtgctgtta	acgggtggagg	gcagtgtagt	ctgagcagta	1560
ctcgttgctg	ccgcgcgcgc	caccagacat	aatagctgac	agactaacag	actgttcctt	1620
tccatgggtc	ttttctgcag	tcaccgtcgt	cgacgccacc	atgcggctgt	gcagagtgtg	1680
gctgtccgtg	tgctgtgtg	ccgtggtgct	gggccagtgc	cagagagaga	cagccgagaa	1740
gaacgactac	taccgggtgc	cccactactg	ggatgcctgc	agcagagccc	tgcccgacca	1800
gacccggtac	aaatacgtgg	agcagctcgt	ggacctgacc	ctgaactacc	actacgacgc	1860
cagccacggc	ctggacaact	tcgacgtgct	gaagcggatc	aacgtgaccg	aggtgtccct	1920
gctgatcagc	gacttccggc	ggcagaacag	aagaggcggc	accaacaagc	ggaccacctt	1980
caacgccgct	ggctctctgg	cccctcacgc	cagatccctg	gaattcagcg	tgccgctgtt	2040
cgccaactga	taatctagaa	agccatggat	atcggatcca	ctacgcgtta	gagctcgtg	2100
atcagcctcg	actgtgcctt	ctagttgcca	gccatctgtt	gtttgcccct	ccccgctgcc	2160
ttccttgacc	ctggaagggt	ccactcccac	tgtcctttcc	taataaaatg	aggaaattgc	2220
atcgcattht	ctgagtaggt	gtcattctat	tctggggggg	ggggtggggc	aggacagcaa	2280
gggggaggat	tggaagaca	atagcagggg	ggtgggcgaa	gaactccagc	atgagatccc	2340
cgcgctggag	gatcatccag	ccggcgtccc	ggaaaacgat	tccgaagccc	aacctttcat	2400
agaaggcggc	ggtggaatcg	aaatctcgtg	atggcagggt	gggcgtcgct	tggtcggcca	2460
tttcgaaccc	cagagtcccg	ctcagaagaa	ctcgtcaaga	aggcgataga	aggcgatgcg	2520
ctgcgaatcg	ggagcggcga	taccgtaaag	cacgaggaag	cggtcagccc	attcgccgcc	2580
aagctcttca	gcaatatcac	gggtagccaa	cgctatgtcc	tgatagcggg	ccgccacacc	2640
cagccggcca	cagtcgatga	atccagaaaa	gcggccatth	tccaccatga	tattcggcaa	2700
gcaggcatcg	ccatgggtca	cgacgagatc	ctcgccgtcg	ggcatgcgcg	ccttgagcct	2760
ggcgaacagt	tcggctggcg	cgagcccctg	atgctcttcg	tccagatcat	cctgatcgac	2820
aagaccgggt	tccatccgag	tacgtgctcg	ctcgatgcga	tgtttcgctt	ggtggtcgaa	2880
tgggcaggta	gccggatcaa	gcgtatgcag	ccgccgcatt	gcatcagcca	tgatggatac	2940
tttctcggca	ggagcaagggt	gagatgacag	gagatcctgc	cccggcactt	cgcccaatag	3000

eolf-seql.txt

```

cagccagtcc cttcccgtt cagtgacaac gtcgagcaca gctgcgcaag gaacgcccgt 3060
cgtggccagc cacgatagcc gcgctgcctc gtcctgcagt tcattcaggg caccggacag 3120
gtcgggtcttg acaaaaagaa ccggggcgccc ctgctgtgac agccggaaca cggcggcatc 3180
agagcagccg attgtctgtt gtgcccagtc atagccgaat agcctctcca cccaagcggc 3240
cggagaacct gcgtgcaatc catcttgttc aatcatgcga aacgatcctc atcctgtctc 3300
ttgatcagat cttgatcccc tgcgccatca gatccttggc ggcaagaaaag ccatccagtt 3360
tactttgcag ggcttcccaa ccttaccaga gggcgcccca gctggcaatt ccggttcgct 3420
tgctgtccat aaaaccgccc agtctagcta tcgccatgta agcccactgc aagctacctg 3480
ctttctcttt gcgcttgctt tttcccttgt ccgatagcc cagtagctga cattcatccg 3540
gggtcagcac cgtttctgct gactggcttt ctacgtgttc cgcttccttt agcagccctt 3600
gcgccctgag tgcttgctggc agcgtgaagc taattcatgg ttaaattttt gttaaatacag 3660
ctcatttttt aaccaatagg ccgaaatcgg caaaatccct tataaatcaa aagaatagcc 3720
cgagataggg ttgagtgttg ttccagtttg gaacaagagt ccactattaa agaacgtgga 3780
ctccaacgtc aaagggcgaa aaaccgtcta tcagggcgat ggccggatca gcttatgcgg 3840
tgtgaaatac cgcacagatg cgtaaggaga aaataccgca tcaggcgctc ttccgcttcc 3900
tcgctcactg actcgtgctg ctcggtcgtt cggctgcggc gagcggtatc agctcactca 3960
aaggcggtaa tacggttatc cacagaatca ggggataacg caggaaagaa catgtgagca 4020
aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt tgctggcggt tttccatagg 4080
ctccgcccc ctgacgagca tcacaaaaat cgacgtcaa gtcagaggtg gcgaaacccg 4140
acaggactat aaagatacca ggcgtttccc cctggaagct ccctcgtgct ctctcctgtt 4200
ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc cttcgggaag cgtggcgctt 4260
tctcatagct cacgctgtag gtatctcagt tcggtgtagg tcgttcgctc caagctgggc 4320
tgtgtgcacg aacccccgt tcagcccgac cgctgcgcct tatccggtaa ctatcgtctt 4380
gagtccaacc cggtagaca cgacttatcg ccaactggcag cagccactgg taacaggatt 4440
agcagagcga ggtatgtagg cgggtctaca gagttcttga agtggtggcc taactacggc 4500
tacactagaa ggacagtatt tggatatctg gctctgtga agccagttac cttcggaaaa 4560
agagttggta gctcttgatc cggcaaacaa accaccgctg gtagcgggtg tttttttgtt 4620
tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag aagatccttt gatcttttct 4680
actgaacggt gatccccacc ggaattgctg 4709

```

<210> 29
 <211> 692
 <212> PRT
 <213> Human cytomegalovirus

<400> 29

Arg Tyr Gly Ala Glu Ala Val Ser Glu Pro Leu Asp Lys Ala Phe His

1

5

10

15

eolf-seql.txt

Leu Leu Leu Asn Thr Tyr Gly Arg Pro Ile Arg Phe Leu Arg Glu Asn
 20 25 30
 Thr Thr Gln Cys Thr Tyr Asn Ser Ser Leu Arg Asn Ser Thr Val Val
 35 40 45
 Arg Glu Asn Ala Ile Ser Phe Asn Phe Phe Gln Ser Tyr Asn Gln Tyr
 50 55 60
 Tyr Val Phe His Met Pro Arg Cys Leu Phe Ala Gly Pro Leu Ala Glu
 65 70 75 80
 Gln Phe Leu Asn Gln Val Asp Leu Thr Glu Thr Leu Glu Arg Tyr Gln
 85 90 95
 Gln Arg Leu Asn Thr Tyr Ala Leu Val Ser Lys Asp Leu Ala Ser Tyr
 100 105 110
 Arg Ser Phe Ser Gln Gln Leu Lys Ala Gln Asp Ser Leu Gly Glu Gln
 115 120 125
 Pro Thr Thr Val Pro Pro Pro Ile Asp Leu Ser Ile Pro His Val Trp
 130 135 140
 Met Pro Pro Gln Thr Thr Pro His Gly Trp Thr Glu Ser His Thr Thr
 145 150 155 160
 Ser Gly Leu His Arg Pro His Phe Asn Gln Thr Cys Ile Leu Phe Asp
 165 170 175
 Gly His Asp Leu Leu Phe Ser Thr Val Thr Pro Cys Leu His Gln Gly
 180 185 190
 Phe Tyr Leu Ile Asp Glu Leu Arg Tyr Val Lys Ile Thr Leu Thr Glu
 195 200 205
 Asp Phe Phe Val Val Thr Val Ser Ile Asp Asp Asp Thr Pro Met Leu
 210 215 220
 Leu Ile Phe Gly His Leu Pro Arg Val Leu Phe Lys Ala Pro Tyr Gln
 225 230 235 240
 Arg Asp Asn Phe Ile Leu Arg Gln Thr Glu Lys His Glu Leu Leu Val
 245 250 255
 Leu Val Lys Lys Asp Gln Leu Asn Arg His Ser Tyr Leu Lys Asp Pro
 260 265 270
 Asp Phe Leu Asp Ala Ala Leu Asp Phe Asn Tyr Leu Asp Leu Ser Ala
 275 280 285

eolf-seql.txt

Leu Leu Arg Asn Ser Phe His Arg Tyr Ala Val Asp Val Leu Lys Ser
 290 295 300
 Gly Arg Cys Gln Met Leu Asp Arg Arg Thr Val Glu Met Ala Phe Ala
 305 310 315 320
 Tyr Ala Leu Ala Leu Phe Ala Ala Ala Arg Gln Glu Glu Ala Gly Ala
 325 330 335
 Gln Val Ser Val Pro Arg Ala Leu Asp Arg Gln Ala Ala Leu Leu Gln
 340 345 350
 Ile Gln Glu Phe Met Ile Thr Cys Leu Ser Gln Thr Pro Pro Arg Thr
 355 360 365
 Thr Leu Leu Leu Tyr Pro Thr Ala Val Asp Leu Ala Lys Arg Ala Leu
 370 375 380
 Trp Thr Pro Asn Gln Ile Thr Asp Ile Thr Ser Leu Val Arg Leu Val
 385 390 395 400
 Tyr Ile Leu Ser Lys Gln Asn Gln Gln His Leu Ile Pro Gln Trp Ala
 405 410 415
 Leu Arg Gln Ile Ala Asp Phe Ala Leu Lys Leu His Lys Thr His Leu
 420 425 430
 Ala Ser Phe Leu Ser Ala Phe Ala Arg Gln Glu Leu Tyr Leu Met Gly
 435 440 445
 Ser Leu Val His Ser Met Leu Val His Thr Thr Glu Arg Arg Glu Ile
 450 455 460
 Phe Ile Val Glu Thr Gly Leu Cys Ser Leu Ala Glu Leu Ser His Phe
 465 470 475 480
 Thr Gln Leu Leu Ala His Pro His His Glu Tyr Leu Ser Asp Leu Tyr
 485 490 495
 Thr Pro Cys Ser Ser Ser Gly Arg Arg Asp His Ser Leu Glu Arg Leu
 500 505 510
 Thr Arg Leu Phe Pro Asp Ala Thr Val Pro Ala Thr Val Pro Ala Ala
 515 520 525
 Leu Ser Ile Leu Ser Thr Met Gln Pro Ser Thr Leu Glu Thr Phe Pro
 530 535 540
 Asp Leu Phe Cys Leu Pro Leu Gly Glu Ser Phe Ser Ala Leu Thr Val
 545 550 555 560

eolf-seql.txt

Ser Glu His Val Ser Tyr Ile Val Thr Asn Gln Tyr Leu Ile Lys Gly
565 570 575

Ile Ser Tyr Pro Val Ser Thr Thr Val Val Gly Gln Ser Leu Ile Ile
580 585 590

Thr Gln Thr Asp Ser Gln Thr Lys Cys Glu Leu Thr Arg Asn Met His
595 600 605

Thr Thr His Ser Ile Thr Val Ala Leu Asn Ile Ser Leu Glu Asn Cys
610 615 620

Ala Phe Cys Gln Ser Ala Leu Leu Glu Tyr Asp Asp Thr Gln Gly Val
625 630 635 640

Ile Asn Ile Met Tyr Met His Asp Ser Asp Asp Val Leu Phe Ala Leu
645 650 655

Asp Pro Tyr Asn Glu Val Val Val Ser Ser Pro Arg Thr His Tyr Leu
660 665 670

Met Leu Leu Lys Asn Gly Thr Val Leu Glu Val Thr Asp Val Val Val
675 680 685

Asp Ala Thr Asp
690

<210> 30
<211> 719
<212> PRT
<213> Human cytomegalovirus

<400> 30

Arg Tyr Gly Ala Glu Ala Val Ser Glu Pro Leu Asp Lys Ala Phe His
1 5 10 15

Leu Leu Leu Asn Thr Tyr Gly Arg Pro Ile Arg Phe Leu Arg Glu Asn
20 25 30

Thr Thr Gln Cys Thr Tyr Asn Ser Ser Leu Arg Asn Ser Thr Val Val
35 40 45

Arg Glu Asn Ala Ile Ser Phe Asn Phe Phe Gln Ser Tyr Asn Gln Tyr
50 55 60

Tyr Val Phe His Met Pro Arg Cys Leu Phe Ala Gly Pro Leu Ala Glu
65 70 75 80

Gln Phe Leu Asn Gln Val Asp Leu Thr Glu Thr Leu Glu Arg Tyr Gln
85 90 95

eolf-seql.txt

Gln Arg Leu Asn Thr Tyr Ala Leu Val Ser Lys Asp Leu Ala Ser Tyr
100 105 110

Arg Ser Phe Ser Gln Gln Leu Lys Ala Gln Asp Ser Leu Gly Glu Gln
115 120 125

Pro Thr Thr Val Pro Pro Pro Ile Asp Leu Ser Ile Pro His Val Trp
130 135 140

Met Pro Pro Gln Thr Thr Pro His Gly Trp Thr Glu Ser His Thr Thr
145 150 155 160

Ser Gly Leu His Arg Pro His Phe Asn Gln Thr Cys Ile Leu Phe Asp
165 170 175

Gly His Asp Leu Leu Phe Ser Thr Val Thr Pro Cys Leu His Gln Gly
180 185 190

Phe Tyr Leu Ile Asp Glu Leu Arg Tyr Val Lys Ile Thr Leu Thr Glu
195 200 205

Asp Phe Phe Val Val Thr Val Ser Ile Asp Asp Asp Thr Pro Met Leu
210 215 220

Leu Ile Phe Gly His Leu Pro Arg Val Leu Phe Lys Ala Pro Tyr Gln
225 230 235 240

Arg Asp Asn Phe Ile Leu Arg Gln Thr Glu Lys His Glu Leu Leu Val
245 250 255

Leu Val Lys Lys Asp Gln Leu Asn Arg His Ser Tyr Leu Lys Asp Pro
260 265 270

Asp Phe Leu Asp Ala Ala Leu Asp Phe Asn Tyr Leu Asp Leu Ser Ala
275 280 285

Leu Leu Arg Asn Ser Phe His Arg Tyr Ala Val Asp Val Leu Lys Ser
290 295 300

Gly Arg Cys Gln Met Leu Asp Arg Arg Thr Val Glu Met Ala Phe Ala
305 310 315 320

Tyr Ala Leu Ala Leu Phe Ala Ala Ala Arg Gln Glu Glu Ala Gly Ala
325 330 335

Gln Val Ser Val Pro Arg Ala Leu Asp Arg Gln Ala Ala Leu Leu Gln
340 345 350

Ile Gln Glu Phe Met Ile Thr Cys Leu Ser Gln Thr Pro Pro Arg Thr
355 360 365

eolf-seql.txt

Thr Leu Leu Leu Tyr Pro Thr Ala Val Asp Leu Ala Lys Arg Ala Leu
370 375 380

Trp Thr Pro Asn Gln Ile Thr Asp Ile Thr Ser Leu Val Arg Leu Val
385 390 395 400

Tyr Ile Leu Ser Lys Gln Asn Gln Gln His Leu Ile Pro Gln Trp Ala
405 410 415

Leu Arg Gln Ile Ala Asp Phe Ala Leu Lys Leu His Lys Thr His Leu
420 425 430

Ala Ser Phe Leu Ser Ala Phe Ala Arg Gln Glu Leu Tyr Leu Met Gly
435 440 445

Ser Leu Val His Ser Met Leu Val His Thr Thr Glu Arg Arg Glu Ile
450 455 460

Phe Ile Val Glu Thr Gly Leu Cys Ser Leu Ala Glu Leu Ser His Phe
465 470 475 480

Thr Gln Leu Leu Ala His Pro His His Glu Tyr Leu Ser Asp Leu Tyr
485 490 495

Thr Pro Cys Ser Ser Ser Gly Arg Arg Asp His Ser Leu Glu Arg Leu
500 505 510

Thr Arg Leu Phe Pro Asp Ala Thr Val Pro Ala Thr Val Pro Ala Ala
515 520 525

Leu Ser Ile Leu Ser Thr Met Gln Pro Ser Thr Leu Glu Thr Phe Pro
530 535 540

Asp Leu Phe Cys Leu Pro Leu Gly Glu Ser Phe Ser Ala Leu Thr Val
545 550 555 560

Ser Glu His Val Ser Tyr Ile Val Thr Asn Gln Tyr Leu Ile Lys Gly
565 570 575

Ile Ser Tyr Pro Val Ser Thr Thr Val Val Gly Gln Ser Leu Ile Ile
580 585 590

Thr Gln Thr Asp Ser Gln Thr Lys Cys Glu Leu Thr Arg Asn Met His
595 600 605

Thr Thr His Ser Ile Thr Val Ala Leu Asn Ile Ser Leu Glu Asn Cys
610 615 620

Ala Phe Cys Gln Ser Ala Leu Leu Glu Tyr Asp Asp Thr Gln Gly Val
625 630 635 640

eolf-seql.txt

Ile Asn Ile Met Tyr Met His Asp Ser Asp Asp Val Leu Phe Ala Leu
645 650 655

Asp Pro Tyr Asn Glu Val Val Val Ser Ser Pro Arg Thr His Tyr Leu
660 665 670

Met Leu Leu Lys Asn Gly Thr Val Leu Glu Val Thr Asp Val Val Val
675 680 685

Asp Ala Thr Asp Gly Thr Lys Leu Gly Pro Glu Gln Lys Leu Ile Ser
690 695 700

Glu Glu Asp Leu Asn Ser Ala Val Asp His His His His His His
705 710 715

<210> 31
<211> 248
<212> PRT
<213> Human cytomegalovirus
<400> 31

Ala Ala Val Ser Val Ala Pro Thr Ala Ala Glu Lys Val Pro Ala Glu
1 5 10 15

Cys Pro Glu Leu Thr Arg Arg Cys Leu Leu Gly Glu Val Phe Glu Gly
20 25 30

Asp Lys Tyr Glu Ser Trp Leu Arg Pro Leu Val Asn Val Thr Gly Arg
35 40 45

Asp Gly Pro Leu Ser Gln Leu Ile Arg Tyr Arg Pro Val Thr Pro Glu
50 55 60

Ala Ala Asn Ser Val Leu Leu Asp Glu Ala Phe Leu Asp Thr Leu Ala
65 70 75 80

Leu Leu Tyr Asn Asn Pro Asp Gln Leu Arg Ala Leu Leu Thr Leu Leu
85 90 95

Ser Ser Asp Thr Ala Pro Arg Trp Met Thr Val Met Arg Gly Tyr Ser
100 105 110

Glu Cys Gly Asp Gly Ser Pro Ala Val Tyr Thr Cys Val Asp Asp Leu
115 120 125

Cys Arg Gly Tyr Asp Leu Thr Arg Leu Ser Tyr Gly Arg Ser Ile Phe
130 135 140

Thr Glu His Val Leu Gly Phe Glu Leu Val Pro Pro Ser Leu Phe Asn
145 150 155 160

eolf-seql.txt

Val Val Val Ala Ile Arg Asn Glu Ala Thr Arg Thr Asn Arg Ala Val
165 170 175

Arg Leu Pro Val Ser Thr Ala Ala Ala Pro Glu Gly Ile Thr Leu Phe
180 185 190

Tyr Gly Leu Tyr Asn Ala Val Lys Glu Phe Cys Leu Arg His Gln Leu
195 200 205

Asp Pro Pro Leu Leu Arg His Leu Asp Lys Tyr Tyr Ala Gly Leu Pro
210 215 220

Pro Glu Leu Lys Gln Thr Arg Val Asn Leu Pro Ala His Ser Arg Tyr
225 230 235 240

Gly Pro Gln Ala Val Asp Ala Arg
245

<210> 32
<211> 442
<212> PRT
<213> Human cytomegalovirus

<400> 32

Cys Asn Val Leu Val Asn Ser Arg Gly Thr Arg Arg Ser Trp Pro Tyr
1 5 10 15

Thr Val Leu Ser Tyr Arg Gly Lys Glu Ile Leu Lys Lys Gln Lys Glu
20 25 30

Asp Ile Leu Lys Arg Leu Met Ser Thr Ser Ser Asp Gly Tyr Arg Phe
35 40 45

Leu Met Tyr Pro Ser Gln Gln Lys Phe His Ala Ile Val Ile Ser Met
50 55 60

Asp Lys Phe Pro Gln Asp Tyr Ile Leu Ala Gly Pro Ile Arg Asn Asp
65 70 75 80

Ser Ile Thr His Met Trp Phe Asp Phe Tyr Ser Thr Gln Leu Arg Lys
85 90 95

Pro Ala Lys Tyr Val Tyr Ser Glu Tyr Asn His Thr Ala His Lys Ile
100 105 110

Thr Leu Arg Pro Pro Pro Cys Gly Thr Val Pro Ser Met Asn Cys Leu
115 120 125

Ser Glu Met Leu Asn Val Ser Lys Arg Asn Asp Thr Gly Glu Lys Gly
130 135 140

Cys Gly Asn Phe Thr Thr Phe Asn Pro Met Phe Phe Asn Val Pro Arg

eolf-seql.txt

145		150		155		160									
Trp	Asn	Thr	Lys	Leu	Tyr	Ile	Gly	Ser	Asn	Lys	Val	Asn	Val	Asp	Ser
				165					170					175	
Gln	Thr	Ile	Tyr	Phe	Leu	Gly	Leu	Thr	Ala	Leu	Leu	Leu	Arg	Tyr	Ala
			180					185					190		
Gln	Arg	Asn	Cys	Thr	Arg	Ser	Phe	Tyr	Leu	Val	Asn	Ala	Met	Ser	Arg
		195					200					205			
Asn	Leu	Phe	Arg	Val	Pro	Lys	Tyr	Ile	Asn	Gly	Thr	Lys	Leu	Lys	Asn
	210					215					220				
Thr	Met	Arg	Lys	Leu	Lys	Arg	Lys	Gln	Ala	Leu	Val	Lys	Glu	Gln	Pro
225					230					235					240
Gln	Lys	Lys	Asn	Lys	Lys	Ser	Gln	Ser	Thr	Thr	Thr	Pro	Tyr	Leu	Ser
				245					250					255	
Tyr	Thr	Thr	Ser	Thr	Ala	Phe	Asn	Val	Thr	Thr	Asn	Val	Thr	Tyr	Ser
			260					265					270		
Ala	Thr	Ala	Ala	Val	Thr	Arg	Val	Ala	Thr	Ser	Thr	Thr	Gly	Tyr	Arg
		275					280						285		
Pro	Asp	Ser	Asn	Phe	Met	Lys	Ser	Ile	Met	Ala	Thr	Gln	Leu	Arg	Asp
	290					295					300				
Leu	Ala	Thr	Trp	Val	Tyr	Thr	Thr	Leu	Arg	Tyr	Arg	Asn	Glu	Pro	Phe
305					310					315					320
Cys	Lys	Pro	Asp	Arg	Asn	Arg	Thr	Ala	Val	Ser	Glu	Phe	Met	Lys	Asn
				325					330					335	
Thr	His	Val	Leu	Ile	Arg	Asn	Glu	Thr	Pro	Tyr	Thr	Ile	Tyr	Gly	Thr
			340					345					350		
Leu	Asp	Met	Ser	Ser	Leu	Tyr	Tyr	Asn	Glu	Thr	Met	Ser	Val	Glu	Asn
		355					360					365			
Glu	Thr	Ala	Ser	Asp	Asn	Asn	Glu	Thr	Thr	Pro	Thr	Ser	Pro	Ser	Thr
	370					375					380				
Arg	Phe	Gln	Arg	Thr	Phe	Ile	Asp	Pro	Leu	Trp	Asp	Tyr	Leu	Asp	Ser
385					390					395					400
Leu	Leu	Phe	Leu	Asp	Lys	Ile	Arg	Asn	Phe	Ser	Leu	Gln	Leu	Pro	Ala
			405						410					415	
Tyr	Gly	Asn	Leu	Thr	Pro	Pro	Glu	His	Arg	Arg	Ala	Ala	Asn	Leu	Ser

420

425

430

Thr Leu Asn Ser Leu Trp Trp Trp Ser Gln
 435 440

<210> 33
 <211> 144
 <212> PRT
 <213> Human cytomegalovirus

<400> 33

Glu Glu Cys Cys Glu Phe Ile Asn Val Asn His Pro Pro Glu Arg Cys
 1 5 10 15

Tyr Asp Phe Lys Met Cys Asn Arg Phe Thr Val Ala Leu Arg Cys Pro
 20 25 30

Asp Gly Glu Val Cys Tyr Ser Pro Glu Lys Thr Ala Glu Ile Arg Gly
 35 40 45

Ile Val Thr Thr Met Thr His Ser Leu Thr Arg Gln Val Val His Asn
 50 55 60

Lys Leu Thr Ser Cys Asn Tyr Asn Pro Leu Tyr Leu Glu Ala Asp Gly
 65 70 75 80

Arg Ile Arg Cys Gly Lys Val Asn Asp Lys Ala Gln Tyr Leu Leu Gly
 85 90 95

Ala Ala Gly Ser Val Pro Tyr Arg Trp Ile Asn Leu Glu Tyr Asp Lys
 100 105 110

Ile Thr Arg Ile Val Gly Leu Asp Gln Tyr Leu Glu Ser Val Lys Lys
 115 120 125

His Lys Arg Leu Asp Val Cys Arg Ala Lys Met Gly Tyr Met Leu Gln
 130 135 140

<210> 34
 <211> 189
 <212> PRT
 <213> Human cytomegalovirus

<400> 34

Ser Pro Trp Ser Thr Leu Thr Ala Asn Gln Asn Pro Ser Pro Pro Trp
 1 5 10 15

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

Pro Phe Leu Tyr Pro Ser Pro Pro Arg Ser Pro Leu Gln Phe Ser Gly
 35 40 45

eolf-seql.txt

Phe Gln Arg Val Ser Thr Gly Pro Glu Cys Arg Asn Glu Thr Leu Tyr
50 55 60

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

Trp Val Lys Lys Val Ile Trp Tyr Leu Ser Gly Arg Asn Gln Thr Ile
85 90 95

Leu Gln Arg Met Pro Arg Thr Ala Ser Lys Pro Ser Asp Gly Asn Val
100 105 110

Gln Ile Ser Val Glu Asp Ala Lys Ile Phe Gly Ala His Met Val Pro
115 120 125

Lys Gln Thr Lys Leu Leu Arg Phe Val Val Asn Asp Gly Thr Arg Tyr
130 135 140

Gln Met Cys Val Met Lys Leu Glu Ser Trp Ala His Val Phe Arg Asp
145 150 155 160

Tyr Ser Val Ser Phe Gln Val Arg Leu Thr Phe Thr Glu Ala Asn Asn
165 170 175

Gln Thr Tyr Thr Phe Cys Thr His Pro Asn Leu Ile Val
180 185

<210> 35
<211> 111
<212> PRT
<213> Human cytomegalovirus

<400> 35

Gln Cys Gln Arg Glu Thr Ala Glu Lys Asn Asp Tyr Tyr Arg Val Pro
1 5 10 15

His Tyr Trp Asp Ala Cys Ser Arg Ala Leu Pro Asp Gln Thr Arg Tyr
20 25 30

Lys Tyr Val Glu Gln Leu Val Asp Leu Thr Leu Asn Tyr His Tyr Asp
35 40 45

Ala Ser His Gly Leu Asp Asn Phe Asp Val Leu Lys Arg Ile Asn Val
50 55 60

Thr Glu Val Ser Leu Leu Ile Ser Asp Phe Arg Arg Gln Asn Arg Arg
65 70 75 80

Gly Gly Thr Asn Lys Arg Thr Thr Phe Asn Ala Ala Gly Ser Leu Ala
85 90 95

eolf-seql.txt

Pro His Ala Arg Ser Leu Glu Phe Ser Val Arg Leu Phe Ala Asn
 100 105 110

<210> 36
 <211> 884
 <212> PRT
 <213> Human cytomegalovirus

<400> 36

Val Ser Ser Ser Ser Thr Arg Gly Thr Ser Ala Thr His Ser His His
 1 5 10 15

Ser Ser His Thr Thr Ser Ala Ala His Ser Arg Ser Gly Ser Val Ser
 20 25 30

Gln Arg Val Thr Ser Ser Gln Thr Val Ser His Gly Val Asn Glu Thr
 35 40 45

Ile Tyr Asn Thr Thr Leu Lys Tyr Gly Asp Val Val Gly Val Asn Thr
 50 55 60

Thr Lys Tyr Pro Tyr Arg Val Cys Ser Met Ala Gln Gly Thr Asp Leu
 65 70 75 80

Ile Arg Phe Glu Arg Asn Ile Val Cys Thr Ser Met Lys Pro Ile Asn
 85 90 95

Glu Asp Leu Asp Glu Gly Ile Met Val Val Tyr Lys Arg Asn Ile Val
 100 105 110

Ala His Thr Phe Lys Val Arg Val Tyr Gln Lys Val Leu Thr Phe Arg
 115 120 125

Arg Ser Tyr Ala Tyr Ile His Thr Thr Tyr Leu Leu Gly Ser Asn Thr
 130 135 140

Glu Tyr Val Ala Pro Pro Met Trp Glu Ile His His Ile Asn Ser His
 145 150 155 160

Ser Gln Cys Tyr Ser Ser Tyr Ser Arg Val Ile Ala Gly Thr Val Phe
 165 170 175

Val Ala Tyr His Arg Asp Ser Tyr Glu Asn Lys Thr Met Gln Leu Met
 180 185 190

Pro Asp Asp Tyr Ser Asn Thr His Ser Thr Arg Tyr Val Thr Val Lys
 195 200 205

Asp Gln Trp His Ser Arg Gly Ser Thr Trp Leu Tyr Arg Glu Thr Cys
 210 215 220

Asn Leu Asn Cys Met Val Thr Ile Thr Thr Ala Arg Ser Lys Tyr Pro

eolf-seql.txt

```

225                230                235                240

Tyr His Phe Phe Ala Thr Ser Thr Gly Asp Val Val Asp Ile Ser Pro
      245      250      255

Phe Tyr Asn Gly Thr Asn Arg Asn Ala Ser Tyr Phe Gly Glu Asn Ala
      260      265      270

Asp Lys Phe Phe Ile Phe Pro Asn Tyr Thr Ile Val Ser Asp Phe Gly
      275      280      285

Arg Pro Asn Ser Ala Leu Glu Thr His Arg Leu Val Ala Phe Leu Glu
      290      295      300

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

Thr Cys Gln Leu Thr Phe Trp Glu Ala Ser Glu Arg Thr Ile Arg Ser
      325      330      335

Glu Ala Glu Asp Ser Tyr His Phe Ser Ser Ala Lys Met Thr Ala Thr
      340      345      350

Phe Leu Ser Lys Lys Gln Glu Val Asn Met Ser Asp Ser Ala Leu Asp
      355      360      365

Cys Val Arg Asp Glu Ala Ile Asn Lys Leu Gln Gln Ile Phe Asn Thr
      370      375      380

Ser Tyr Asn Gln Thr Tyr Glu Lys Tyr Gly Asn Val Ser Val Phe Glu
385      390      395      400

Thr Thr Gly Gly Leu Val Val Phe Trp Gln Gly Ile Lys Gln Lys Ser
      405      410      415

Leu Val Glu Leu Glu Arg Leu Ala Asn Arg Ser Ser Leu Asn Leu Thr
      420      425      430

His Asn Arg Thr Lys Arg Ser Thr Asp Gly Asn Asn Ala Thr His Leu
      435      440      445

Ser Asn Met Glu Ser Val His Asn Leu Val Tyr Ala Gln Leu Gln Phe
      450      455      460

Thr Tyr Asp Thr Leu Arg Gly Tyr Ile Asn Arg Ala Leu Ala Gln Ile
465      470      475      480

Ala Glu Ala Trp Cys Val Asp Gln Arg Arg Thr Leu Glu Val Phe Lys
      485      490      495

Glu Leu Ser Lys Ile Asn Pro Ser Ala Ile Leu Ser Ala Ile Tyr Asn

```


500

505

510

Lys Pro Ile Ala Ala Arg Phe Met Gly Asp Val Leu Gly Leu Ala Ser
 515 520 525

Cys Val Thr Ile Asn Gln Thr Ser Val Lys Val Leu Arg Asp Met Asn
 530 535 540

Val Lys Glu Ser Pro Gly Arg Cys Tyr Ser Arg Pro Val Val Ile Phe
 545 550 555 560

Asn Phe Ala Asn Ser Ser Tyr Val Gln Tyr Gly Gln Leu Gly Glu Asp
 565 570 575

Asn Glu Ile Leu Leu Gly Asn His Arg Thr Glu Glu Cys Gln Leu Pro
 580 585 590

Ser Leu Lys Ile Phe Ile Ala Gly Asn Ser Ala Tyr Glu Tyr Val Asp
 595 600 605

Tyr Leu Phe Lys Arg Met Ile Asp Leu Ser Ser Ile Ser Thr Val Asp
 610 615 620

Ser Met Ile Ala Leu Asp Ile Asp Pro Leu Glu Asn Thr Asp Phe Arg
 625 630 635 640

Val Leu Glu Leu Tyr Ser Gln Lys Glu Leu Arg Ser Ser Asn Val Phe
 645 650 655

Asp Leu Glu Glu Ile Met Arg Glu Phe Asn Ser Tyr Lys Gln Arg Val
 660 665 670

Lys Tyr Val Glu Asp Lys Val Val Asp Pro Leu Pro Pro Tyr Leu Lys
 675 680 685

Gly Leu Asp Asp Leu Met Ser Gly Leu Gly Ala Ala Gly Lys Ala Val
 690 695 700

Gly Val Ala Ile Gly Ala Val Gly Gly Ala Val Ala Ser Val Val Glu
 705 710 715 720

Gly Val Ala Thr Phe Leu Lys Asn Pro Phe Gly Ala Phe Thr Ile Ile
 725 730 735

Leu Val Ala Ile Ala Val Val Ile Ile Thr Tyr Leu Ile Tyr Thr Arg
 740 745 750

Gln Arg Arg Leu Cys Thr Gln Pro Leu Gln Asn Leu Phe Pro Tyr Leu
 755 760 765

Val Ser Ala Asp Gly Thr Thr Val Thr Ser Gly Ser Thr Lys Asp Thr

770

775

Ser Leu Gln Ala Pro Pro Ser Tyr Glu Glu Ser Val Tyr Asn Ser Gly
785 790 795 800

Arg Lys Gly Pro Gly Pro Pro Ser Ser Asp Ala Ser Thr Ala Ala Pro
805 810 815

Pro Tyr Thr Asn Glu Gln Ala Tyr Gln Met Leu Leu Ala Leu Ala Arg
820 825 830

Leu Asp Ala Glu Gln Arg Ala Gln Gln Asn Gly Thr Asp Ser Leu Asp
835 840 845

Gly Arg Thr Gly Thr Gln Asp Lys Gly Gln Lys Pro Asn Leu Leu Asp
850 855 860

Arg Leu Arg His Arg Lys Asn Gly Tyr Arg His Leu Lys Asp Ser Asp
865 870 875 880

Glu Glu Glu Asn