INVESTIGATION OF NON-PROTEIN-CODING REGIONS IN THE HUMAN CYTOMEGALOVIRUS GENOME

Ralph David Hector

September 2005

A thesis presented to the Faculty of Biomedical and Life Sciences at the University of Glasgow for the degree of Doctor of Philosophy

Institute of Virology
Church Street
Glasgow G11 5JR

ABSTRACT

Human cytomegalovirus (HCMV) has the largest genome of the human herpesviruses, and its gene content is imperfectly understood. The gene content of HCMV strain AD169 was recently re-evaluated, discounting 51 previously proposed protein-coding open reading frames (ORFs) because they have no counterparts in chimpanzee cytomegalovirus (CCMV) and lack any other convincing evidence for expression. Some of the discounted ORFs were located in blocks, and their omission left three large 'empty' regions in the AD169 genome. One of these regions, located between genes RL1 and RL10, has been investigated previously leading to the discovery of a new gene. The other two regions are investigated in this study. The first, termed region X, is located between genes UL105 and UL112, which are firmly accepted as encoding proteins. The second, termed region O, is located between genes UL57 and UL69. Initial analysis of the AD169 genome predicted six (UL106-UL111) and eleven (UL58-UL68) small ORFs in regions X and O, respectively, and subsequent analyses have predicted three further ORFs in each region (C-ORF16-C-ORF18 and ORF3-ORF5, respectively).

Sequence data were generated from seven HCMV strains by PCR amplification and cloning of region X (approximately 6 kbp) and region O (approximately 8 kbp). Sequence comparisons were used to identify which ORFs in these regions are conserved and which are disrupted by insertions, deletions or substitutions leading to in-frame termination codons. All of the ORFs are frameshifted in certain strains, with the exception of UL108, C-ORF16 and C-ORF18 in region X, and UL66 in region O. The disrupted ORFs are unlikely to represent protein-coding genes. Moreover, the few ORFs unaffected by frameshifts remain unlikely to encode proteins as they are predicted to encode very small proteins and lack counterparts in CCMV. Furthermore, in region X no transcripts corresponding to the cognate ORFs were detected by northern blotting, and in region O UL66 is completely overlapped by the previously identified pp67 transcript. This transcript, which is routinely detected by PCR assays, is disrupted in all the strains analysed and is therefore also unlikely to encode a protein.

Transcript mapping in region X detected a spliced 1.1 kb polyadenylated RNA and a 4.6 kb intron, which covers most of the region. 5'- and 3'-ends of the 1.1 kb RNA were identified, the former located 25 bp from a TATA box and the latter to two sites located 20 and 34 bp from a polyadenylation signal. These results consolidate the findings of two previous studies that had partially characterised the 5 kb and 1.1 kb RNAs, respectively. The splice sites, 3'-polyadenylation signal, and 5'-TATA box of the 1.1 kb RNA are conserved in the corresponding region of CCMV, suggesting that a similar RNA with a large intron should be expressed in CCMV. However, this RNA is unlikely to encode a protein as no amino acid sequences are conserved between the two genomes.

A third region of the HCMV genome where coding potential was not clear was also investigated. Region G, at the end of the unique short (U_S) sequence, is located between genes US32 and TRS1, which are firmly accepted as encoding proteins. On the basis of conservation in CCMV, three originally defined ORFs (US33, US35 and US36) had been discounted and a novel ORF (US34A) identified. Again, sequence data were generated from seven HCMV strains by PCR amplification and cloning. Comparisons showed that the sequence between US34A and TRS1 is highly variable, and that US35 and US36 are frameshifted in multiple strains. US33 is conserved between strains, but no relevant transcript was detected. However, a transcript from a small, novel ORF (US33A) was detected on the opposing strand. 5'- and 3'-ends were identified for this RNA, the 5'-end located 31 bp downstream from a TATA box. The US33A ORF and the TATA box are also conserved in CCMV, suggesting that US33A may constitute a novel HCMV gene. The US33A transcript was shown to be 3'-co-terminal with that of US34. Transcripts from US31 and US32, which are conserved ORFs immediately to the left of region G, were also mapped. No convincing evidence was found for transcription of US34A.

This work extends the understanding of the genetic content of HCMV and has identified novel transcripts in HCMV, providing a basis on which to develop future experiments aimed at determining their function.

ACKNOWLEDGEMENTS

I would like to thank the following people for their help in completing this thesis.

My supervisor, Dr. Andrew Davison, for his support, advice, good humour, for his patience during the critical evaluation of this thesis, and also for supplying me with a number of figures. Professor Duncan McGeoch for giving me the chance to do my PhD at the Institute of Virology. Everyone in lab 200: Charles Cunningham, Aidan Dolan, Clare Addison, Parvis Akter and Edgar Sevilla-Reyes. Their support on a daily basis was greatly appreciated and encouraged me enormously. Dr. Derrick Dargan, Fiona Jamieson and Elaine Douglas in lab 204 for helping me get my cell culture work off the ground. Thanks also to the Institute of Virology support staff.

I am especially indebted to my Mum and Dad, and even more so to Sylvia Brown, whose support, encouragement and love have been everything to me.

Funding for this study was provided by the Medical Research Council.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

mM	millimolar
moi	multiplicity of infection
mRNA	messenger ribonucleic acid
μCi	microcuries
μg	micrograms
μl	microlitres
μM	micromolar
nm	nanometres
nt	nucleotides
ORF	open reading frame
PCR	polymerase chain reaction
$polyA^+$	polyadenylated
$polyA^-$	non-polyadenylated
p.i.	post infection
RACE	rapid amplification of cDNA ends
RCMV	rat cytomegalovirus
RhCMV	rhesus cytomegalovirus
RNA	ribonucleic acid
RT	reverse transcriptase
s	seconds
SCMV	simian (African green monkey) cytomegalovirus
SDS	sodium dodecyl sulphate
SSC	standard saline citrate
ssRNA	single-stranded ribonucleic acid
T	thymine
TBE	tris borate ethylenediaminetetra-acetic acid
TR_L	long terminal repeat
TR_S	short terminal repeat
TuHV-1	tupaiid herpesvirus 1
U	uracil
U_L	unique-long
U_S	unique-short
UV	ultraviolet
VZV	varicella-zoster virus
v/v	volume/volume
w/v	weight/volume

CHAPTER 1: INTRODUCTION

1.1 The family *Herpesviridae*

The International Committee on Taxonomy of Viruses (ICTV) has divided the family *Herpesviridae* into three subfamilies (Davison *et al.*, 2005a). The family *Herpesviridae* comprises over 120 large DNA viruses that infect vertebrates and at least one invertebrate, the oyster (Davison *et al.*, 2005a). Members of the subfamily *Alphaherpesvirinae* (known colloquially as α-herpesviruses) infect epithelial cells *in vivo* and establish latency in the nervous system. Many cause evident epithelial lesions in their natural hosts. Members of the subfamily *Betaherpesvirinae* (β-herpesviruses) are often clinically silent in immunocompetent hosts. Many species show a restricted host range *in vitro* with a long productive cycle and are usually associated with enlarged (cytomegalic) cells. Members of the subfamily *Gammaherpesvirinae* (γ-herpesviruses) establish latent infection in lymphocytes and infect lymphocytes *in vitro*, and are often associated with lymphoproliferative disease and carcinomas. Genera within the three subfamilies are shown in Table 1.1. A single fish herpesvirus has been classified into a genus (*Ictalurivirus*) separate from the three subfamilies.

Subfamily	Genus	Example species
Alphaherpesvirinae	*Simplexvirus*	Herpes simplex virus type 1
	Varicellovirus	Varicella-zoster virus
	Mardivirus	Marek's disease virus
	Iltovirus	Infectious laryngotracheitis-like virus
Betaherpesvirinae	*Cytomegalovirus*	Human cytomegalovirus
	Muromegalovirus	Murine cytomegalovirus
	Roseolovirus	Human herpesvirus 6
Gammaherpesvirinae	*Lymphocryptovirus*	Epstein-Barr virus
	Rhadinovirus	Kaposi's sarcoma-associated herpesvirus
(Unassigned genus)	*Ictalurivirus*	Channel catfish virus

Table 1.1. Subfamilies and genera of the family *Herpesviridae*.

1.2 Herpesvirus evolution, phylogeny and classification

The species that currently constitute the family *Herpesviridae* probably represent only a small fraction of the number in existence, as each herpesvirus is usually associated with a single host species in nature, and several distinct herpesviruses may infect a host (Davison, 2002). The host specificity displayed by herpesviruses implies that they have evolved with their hosts over long periods of time and are well adapted to them. This also means that many herpeviruses are likely to have become extinct as their hosts became extinct, and that modern viruses represent only a small proportion of all the herpesviruses that have ever existed. In structuring the taxa within *Herpesviridae*, relationships between viral DNA and amino acid sequences have superseded comparisons of viral antigenic and biological properties as the main criterion (McGeoch and Davison, 1999a). The phylogenetic trees shown in Figure 1.1 illustrate the division of each of the three subfamilies and further into genera and species.

Genetic differences are readily identified between members of different genera within a subfamily, but the gene complements of these viruses are, for the most part, similar. For example, the α-herpesviruses herpes simplex virus type 1 (HSV-1, genus *Simplexvirus*) and varicella-zoster virus (VZV, genus *Varicellovirus*) share 64 genes (Davison, 2002). However, sublineages of α- and β-herpesviruses show extensive speciation with host lineages, where the branching patterns and extent of divergence are similar. However, γ-herpesviruses show a more complex picture, with higher and lower substitution rates in different sublineages (McGeoch *et al.*, 2000).

Modern α-, β- and γ-herpesviruses are thought to contain 43 genes contributed by an ancestral virus (Table 1.2) (Davison, 2002). These conserved, 'core' genes tend to be located in the central region of the genomes and can be arranged into seven blocks, which have different orders and orientations in each of the subfamilies. The functions of the core genes, which are involved in many vital aspects of the viral life cycle, such as cell entry, DNA replication, capsid formation and genome packaging, indicate that the ancestral virus was recognisably a herpesvirus (McGeoch and Davison, 1999a). Genes that are lineage-specific tend to be located towards the ends of the genome.

Some of these genes may have arisen de novo, but many appear to have been captured from the cell. Gene duplication is also apparent, particularly in cytomegaloviruses (β-herpesviruses), which contain large arrays of related genes (Davison, 2002). Original assessments of the timescale of herpesvirus evolution estimated that the three subfamilies arose approximately 180 to 220 million years ago (McGeoch *et al.*, 1995). The root of the subfamilies is now thought to be around 400 million years ago (McGeoch and Gatherer, 2005). The major sublineages within the subfamilies probably arose before mammalian radiation of 60-80 million years ago, and speciations within the sublineages took place after this (McGeoch *et al.*, 1995). A recent study of reptilian herpesvirus evolution has thrown up some new ideas about the evolution of the herpesviruses, particularly among the α-herpesviruses (McGeoch and Gatherer, 2005). Reptilian and mammalian viruses could have coevolved with their host lineages, but the two avian genera (*Mardivirus* and *Iltovirus*) would have arisen by a transfer mechanism. Alternatively, the reptilian and avian lineages could have coevolved with their host lineages, but the mammalian clade would have arisen by a transfer mechanism. Both scenarios are possible but mutually exclusive.

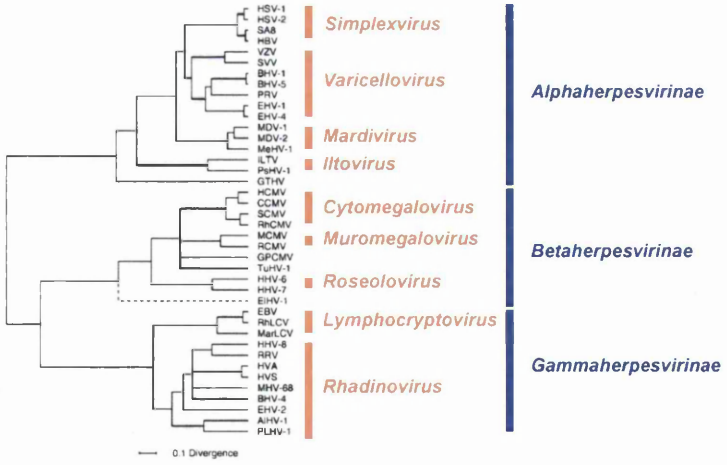


Figure 1.1. Phylogenetic tree of the α-, β- and γ-herpesviruses.

Phylogenetic tree for the *Herpesviridae* based on alignments of six sets of homologous proteins, obtained by a maximum likelihood evaluation, with molecular clock imposed. The dashed line, the branch for EIHV-1, was interpolated from a tree based on two available proteins. Unpublished data of A. Davison and D. McGeoch, MRC Virology Unit, Glasgow.

Group	Genome and Gene HSV-1	HCMV	EBV	Core gene function
Control and modulation	UL13	UL97	BGLF4	Serine-threonine protein-kinase
	UL54	UL69	BMLF1	Post-transcriptional regulator
DNA replication machinery	UL5	UL105	BBLF4	Helicase (DNA helicase-primase complex)
	UL8	UL102	BBLF2/3	Component of DNA helicase-primase complex
	UL9			*Ori*-binding helicase
	UL29	UL57	BALF2	Single-stranded DNA-binding protein
	UL30	UL54	BALF5	Catalytic subunit of DNA polymerase
	UL42	UL44	BMRF1	Processivity subunit of DNA polymerase
	UL52	UL70	BSLF1	Primase (DNA helicase-primase complex)
Peripheral enzymes	UL2	UL114	BKRF3	Uracil-DNA glycosylase
	UL23		BXLF1	Thymidine kinase
	UL39	UL45	BORF2	Ribonucleotide reductase; large subunit
	UL40		BaRF1	Ribonucleotide reductase; small subunit
	UL50	UL72	BLLF3	Deoxyuridine triphosphate
Processing and packaging of DNA	UL6	UL104	BBRF1	Portal protein
	UL12	UL98	BGLF5	Deoxyribonuclease
	UL15	UL89	BGRF1	ATPase subunit of terminase
	UL17	UL93	BGLF1	Nuclear capsid localisation
	UL25	UL77	BVRF1	Putative portal-capping protein
	UL28	UL56	BALF3	Terminase subunit
	UL32	UL52	BFLF1	Nuclear capsid localisation
	UL33	UL51	BFRF1A	Interacts with terminase
Capsid assembly and structure	UL18	UL85	BDLF1	Capsid triplex protein
	UL19	UL86	BcLF1	Major capsid protein
	UL26	UL80	BVRF2	Maturational protease
	UL26.5	UL80.5	BdRF1	Scaffolding protein
	UL35	UL48A	BFRF3	Capsid protein on hexon tips
	UL38	UL46	BORF1	Capsid triplex protein
Egress of capsid from nucleus	UL31	UL53	BFLF2	Nuclear matrix protein
	UL34	UL50	BFRF1	Inner nuclear membrane protein
Tegument	UL7	UL103	BBRF2	Tegument protein
	UL11	UL99	BBLF1	Myristylated tegument protein
	UL14	UL96	BGLF3.5	Tegument protein
	UL16	UL94	BGLF2	Tegument protein
	UL36	UL48	BPLF1	Tegument protein
	UL37	UL47	BOLF1	Tegument protein
	UL51	UL71	BSRF1	Tegument protein
Surface and Membrane	UL1	UL115	BKRF2	Virion glycoprotein L
	UL10	UL100	BBRF3	Virion glycoprotein M
	UL22	UL75	BXLF2	Virion glycoprotein H
	UL27	UL55	BALF4	Virion glycoprotein B
	UL49A	UL73	BLRF1	Virion glycoprotein N
Unknown	UL24	UL76	BXRF1	Nuclear protein

Table 1.2. Genes conserved in α-, β- and γ-herpesviruses.

Adapted from Davison (2002). Core genes are grouped according to function. HSV-1 is an α-herpesvirus, HCMV is a β-herpesvirus and EBV is a γ-herpesvirus. Three genes (UL9, UL23 and UL40 in HSV-1) are considered ancestral, but have been lost from some lineages.

Two other major herpesvirus lineages have been identified, one leading to the bony fish and amphibian viruses and the other to the only known herpesvirus that infects an invertebrate, oyster herpesvirus (OsHV-1) (Davison, 1992, 1998; Bernard and Mercier, 1993; Davison *et al.*, 1999). However, these two lineages are extremely diverged and recent sequence comparisons have shown that OsHV-1 is, at best, tenuously related to the two classes of vertebrate herpesviruses (Davison *et al.*, 2005b). It is thought that the best method of taxonomic classification would be to establish three families, incorporating the three lineages described above, within an order, *Herpesvirales* (Davison *et al.*, 2005b).

1.3 Herpesvirus genetic content

Herpesviruses have large, linear, double-stranded DNA genomes. Virus genomes range in size from 124 kbp to over 295 kbp; the smallest is simian varicella virus (SVV) (Gray *et al.*, 2001), and the largest is koi herpesvirus (Waltzek *et al.*, 2005). The largest mammalian herpesvirus genome, of 240 kbp, is chimpanzee cytomegalovirus (Davison *et al.*, 2003a). The herpesvirus genomes are thought to contain between about 70 and 200 genes which are tightly packed and arranged in about equal numbers on each DNA strand. Most are expressed as single exons, and spliced genes, of which relatively few have been identified, are more common among the β- and γ-herpesviruses than the α-herpesviruses (Davison and Clements, 1996). Overlaps between genes in different reading frames on the same strand or on the opposing strand are infrequent and usually not extensive. Families of genes tend to be arranged tandemly on the same strand and often share the same polyadenylation site.

A wide range of base compositions is also exhibited, ranging from 32-75% in G+C content (Honess, 1984). A characteristic feature of herpesvirus genome structure is the presence of terminal or internal repeated sequences, in direct or inverse orientations. Figure 1.2 shows the herpesvirus genome types, which differ in the layouts of the repeat elements in relation to unique sequences.

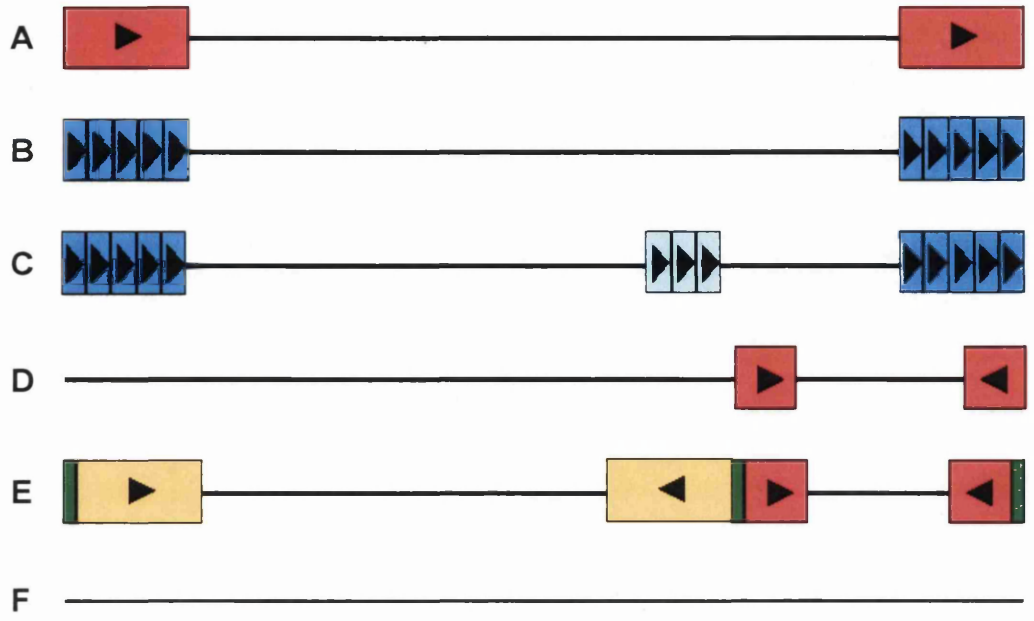


Figure 1.2. Herpesvirus genome structures.

Black lines show unique sequences and coloured blocks show repeat elements in the genomes, which are not to scale. Arrows show orientations of the repeated regions. Adapted from Roizman and Pellett (2001).

(A) A long unique sequence flanked by large direct terminal repeats (Example genome: Human herpesvirus 6).

(B) A long unique sequence flanked by smaller tandem direct terminal repeats that are variable in number (Kaposi's sarcoma-associated herpesvirus).

(C) A long unique sequence flanked by smaller tandem direct terminal repeats, plus a different, internal sequence. The repeats are present in variable numbers (Epstein-Barr virus).

(D) A short unique sequence flanked by large inverted terminal repeats and a long unique sequence flanked by very small or no terminal repeats (Varicella-zoster virus).

(E) Long and short unique sequences, each flanked by large inverted terminal repeats, and a terminal direct repeat, the *a* sequence (green), which is found at the ends of the genome and in inverted orientation between the internal repeats (Human cytomegalovirus).

(F) A single unique sequence with very small or no terminal repeats (Tupaia herpesvirus).

1.4 Herpesvirion structure

Herpesvirions are characteristic, but complex, spherical structures that comprise a core, capsid, tegument and envelope (Figure 1.3) (Wildy *et al.*, 1960; Roizman and Pellett, 2001). In HSV-1, they can range in size from 160 to 230 nm, averaging 180 nm (Szilágyi and Berriman, 1994). The core comprises the linear double-stranded DNA herpesvirus genome, packed in a liquid crystalline array that fills the entire volume of the preformed icosahedral capsid. The mature capsid is composed of 162 capsomers: 12 pentavalent capsomers (pentons) at the vertices, 60 hexavalent capsomers (hexons) at the 20 faces and 90 hexavalent capsomers along the 30 edges (Wildy *et al.*, 1960; Gibson, 1996). Between the nucleocapsid and envelope is a largely amorphous proteinaceous layer, the tegument, which consists of at least 18 proteins (Zhou *et al.*, 1999). Some of these proteins have functions related to morphogenesis, uncoating and regulation of gene expression, but the functions of many of the tegument proteins are not known. The envelope, a lipid bilayer enclosing the virion, is coated with glycoprotein spikes (Wildy et al., 1960).

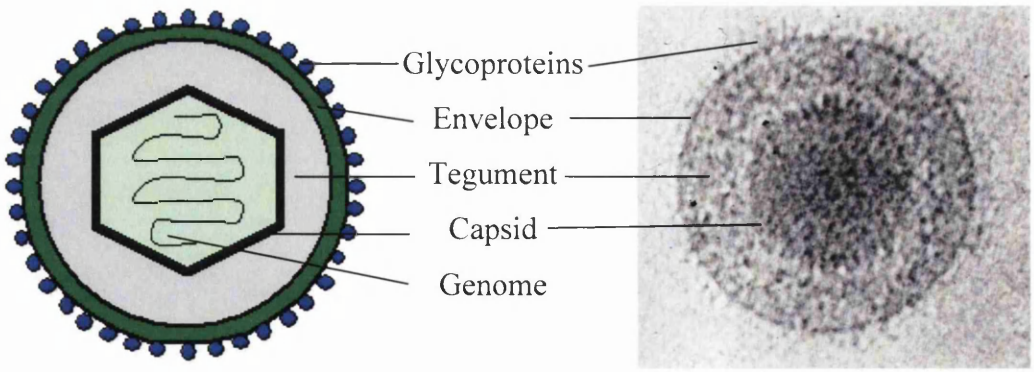


Figure 1.3 Herpesvirion structure

(Left) A representation of a herpesvirion. (Right) A cryo-electron microscope image of an HSV-1 virion (courtesy of F. Rixon, MRC Virology Unit, Glasgow).

1.5 Herpesvirus life cycle

The herpesvirus replicative life cycle, which consists of lytic and latent phases, is conveniently described in the case of HSV-1. The virus initially binds to glycosaminoglycans on cell surface proteoglycans via glycoproteins gB and/or gC (Shukla and Spear, 2001). This interaction, usually with heparan sulphate, is not thought to be essential for viral entry, but does greatly increase the efficiency of viral entry, thought to be due to concentrating virus on the cell surface so the appropriate entry receptor viral ligands find these receptors (Spear and Longnecker, 2003). Viral envelope glycoproteins are essential to virus entry, and mediate virus attachment and entry to the host cell. Three essential glycoproteins, gB, gH and gL, are conserved in other herpesviruses (Spear and Longnecker, 2003). HSV requires a fourth essential glycoprotein, gD, which acts as a ligand for several cell surface receptors. Consequently, all four essential glycoproteins are required to stimulate fusion of the viral envelope with a cell membrane. The capsid is then released into the cytoplasm.

On entry, the capsid migrates to the nucleus by binding to microtubules (Batterson *et al.*, 1983; Sodeik *et al.*, 1997). A microtubule-dependent motor, dynein, also attaches to the incoming capsids to drive them from the edge of the cell to the nucleus (Sodeik *et al.*, 1997). At the nucleus, the nuclear pore complex spans the double membrane of the nuclear envelope. The capsid, essentially tegument-free at this stage, docks via importin β to cytoplasmic nuclear-pore-complex proteins (Greber and Fassati, 2003). It is thought that the cytosolic factors and energy are also recruited, and the capsid opens up at one side to eject the viral DNA into the nucleus (Dasgupta and Wilson, 1999; Greber and Fassati, 2003). Within the nucleus, transcription, viral DNA replication, capsid formation and packaging of viral DNA take place.

Viral gene expression takes the form of a co-ordinated cascade that is regulated and sequentially ordered, with three main phases; immediate-early (IE), early (E) and late (L) (Honess and Roizman, 1974, 1975, Jones and Roizman, 1979; Sears and Roizman, 1990). Transcription takes place in the nucleus, but viral proteins are synthesised in the cytoplasm. IE genes are the first viral genes expressed after infection, and their transcription is enhanced by the virion structural protein VP16.

Most of the IE proteins expressed are *trans*-acting regulators which initiate the expression of E genes. E genes encode several enzymes involved in DNA replication and nucleotide metabolism, as well as a subset of glycoproteins. L gene expression follows production of E proteins, and is enhanced by viral DNA synthesis (Honess and Roizman, 1975).

Viral replication starts in the nucleus, near ND10 structures, with the circularisation of genomic DNA by direct ligation of the termini (Garber *et al.*, 1993; Strang and Stow, 2005). DNA synthesis is then initiated from the viral origins of lytic replication (*ori*S and *ori*L in HSV-1) to produce DNA in an endless conformation, probably as head-to-tail concatemers produced by a rolling circle mechanism (Jacob *et al.*, 1979; Poffenberger and Roizman, 1985). Cleavage of these concatamers is site-specific, at unique sites (*pac* sites) within *a* sequences. The *a* sequences, which are present at the ends of the viral genome and at the internal repeated sequences, are necessary and sufficient for HSV-1 DNA cleavage and packaging. Unit-length molecules of viral DNA are now ready to be packaged in capsids. Seven HSV-1 proteins which have been identified which are involved in the processing and packaging of viral DNA: UL6, UL15, UL17, UL25, UL28, UL32 and UL33 (Brown *et al.*, 2002). The functions of some of these genes have now been elucidated, and they are thought to have roles in introducing specific cuts in the DNA concatemer, forming the portal through which the virus DNA enters the capsid, assisting DNA translocation into the capsid, and sealing the capsid once it is filled.

Final capsid assembly occurs in the nucleus, although some capsid proteins can begin the initial stages of assembly in the cytoplasm (Rixon *et al.*, 1996). Partial procapsids and procapsids are intermediates formed as the mature capsid is constructed. In understanding the process of capsid assembly, much of the work was carried out from studies of herpesvirus proteins expressed by recombinant baculoviruses (Tatman *et al.*, 1994; Thomsen *et al.*, 1994). Partial procapsids are angular pieces of capsid shell which partially surround a portion of core. The shell contains the major capsid protein (VP5), the two triplex proteins (VP19C and VP23) and the core contains the scaffolding proteins (pre-VP22a and VP21) (Newcomb *et al.*, 1996; Homa and Brown, 1997). The partial procapsid grows by progressive addition of shell and core proteins, in small increments. Once this structure is closed, the procapsid is formed.

The procapsid is composed of capsomers, has the same diameter and the same symmetry (T=16) as the mature capsid. However, at this stage the procapsid is spherical, unlike the mature capsid, and the morphology of the hexons is also different. In addition, the triplexes are thought to constitute the main links by which the capsomers are held together as the floor layer found in mature capsid is largely absent in the procapsid (Newcomb *et al.*, 1996; Homa and Brown, 1997). The procapsid then undergoes extensive conformational changes, angularising to create the mature, icosahedral capsid. The scaffolding protein is cleaved and exits the capsid during this transformation. Although these changes generally do not involve the addition of further protein molecules, if the small capsid protein is present, it is added only as a result of angularisation (Chi and Wilson, 2000). The final capsid also has a much greater structural stability upon maturation (Newcomb *et al.*, 1996).

Viral DNA enters the capsid through the portal, a ring-shaped structure composed of 12 UL6 protein molecules. Each mature capsid contains a single portal, as does each procapsid, indicating that it is incorporated during capsid assembly, into one of the 12 capsid vertices (Sheaffer *et al.*, 2001). It is thought, from electron microscopy of HSV-1 capsids, that DNA packaging begins with the procapsid, and that procapsid angularisation may occur as the DNA is encapsidated (Brown *et al.*, 2002).

Capsid egress from the nucleus begins at the nuclear membrane. The first budding process occurs here, in which the capsid acquires an envelope derived from the inner nuclear membrane (Mettenleiter, 2004). This process involves a complex of the UL31 and UL34 gene products (Chang *et al.*, 1997; Roller *et al.*, 2000). Consequent to this budding process is the formation of primary enveloped virions in the perinuclear space. At this stage, the primary virion is different both morphologically and biochemically from the secondary virion (Mettenleiter, 2004). This primary envelope then fuses with the outer nuclear membrane, releasing the nucleocapsid into the cytoplasm. It is thought that the UL34 protein, a component of primary virions, and the US3 protein of HSV-1 plays a role in this fusion process. However, these processes occur in the absence of other viral envelope and tegument proteins which are present in the mature virion (Mettenleiter, 2004).

Once in the cytoplasm, the nucleocapsid must acquire more than 15 tegument proteins during the process of tegumentation, a complex network of protein-protein interactions. It is thought that the UL36 gene product interacts with the pentons of the capsid to form the first layer of tegument (McNabb & Courtney, 1992). This inner shell of tegument shows icosahedral symmetry, with part of it anchored at the vertices of the capsid (Zhou *et al.*, 1999). A second layer of tegument is thought to be composed of the UL37 gene product (Mettenleiter, 2004). Subsequent tegumentation steps involving these other tegument proteins are not fully defined, however, a major structural role is predicted of UL48, which is critical for tegument formation. Secondary envelopment occurs at the Golgi apparatus, from where the vesicles into which budding occurs are derived. Although it is unclear how tegumented capsids are directed to this site and how viral glycoproteins are assembled, two viral proteins have been implicated. UL11 is thought to direct tegument components to the budding site, and gM is thought to aid accumulation of envelope glycoproteins at this site (Mettenleiter, 2004). This secondary envelopment process results in an enveloped virion within a secretory vesicle, which is then transported to the plasma membrane. These processes are not fully understood, however, the HSV-1 proteins UL20 and gK have been implicated (Baines *et al.*, 1991; Foster and Kousoulas, 1999). The vesicle and plasma membranes fuse, and mature virions are released from the infected cell.

All the viruses are also able to establish life long infections in their host via various mechanisms. In the case of the human herpesviruses, HSV-1, HSV-2 and VZV are neurotropic, HHV-6, HHV-7 and EBV are lymphotropic, and HCMV establishes latency in the monocytic lineage (Roizman and Pellett, 2001). In its latent form, HSV-1 DNA is non-replicating, does not integrate into the host DNA and is present either as a circular molecule or as a concatemer (Rock and Fraser, 1983, 1985; Efstathiou *et al.*, 1986; Mellerick and Fraser, 1987). During latency, expression from the HSV-1 genome is repressed, except for the latency-associated transcripts (LATs), which accumulate in the neurons and help to maintain latency in a process that is not fully understood (Stevens *et al.*, 1988; Preston, 2000). Reactivation, which may occur in response to stress, fever or sunlight, causes lytic infection to proceed. It appears that ICP0, an initiator of lytic infection, plays an important role in reactivation in the neuronal nucleus, as ICP0-deficient viruses reactivate with reduced efficiency (Leib *et al.*, 1989; Cai *et al.*, 1993; Everett, 2000).

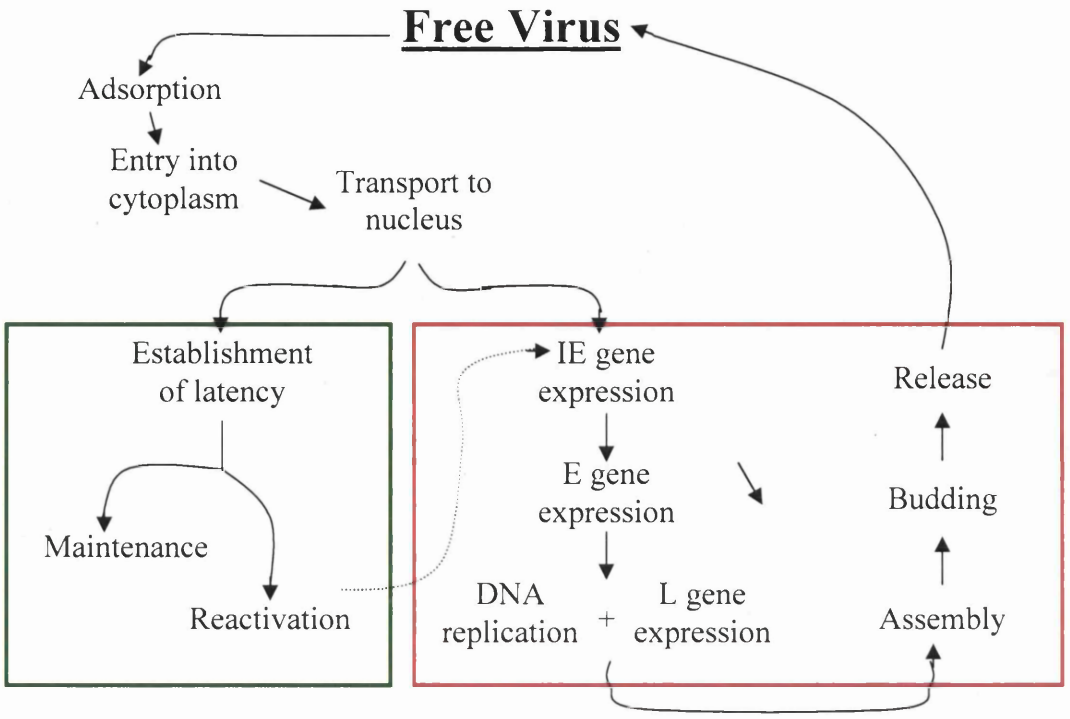


Figure 1.4 Herpesvirus life cycle.

Adapted from Davison and Clements (1996). After transport to the nucleus, the virus can either establish latency (the green box) or start productive infection (the red box).

1.6 Human herpesviruses

Eight human herpesviruses have been identified, representing all three subfamilies (Table 1.3). The effects of herpesvirus infection can range from asymptomatic to severe, depending on the virus and the host, but they are often disabling or fatal in the foetus, the very young and the immunosuppressed. A wide variety of diseases may result, ranging from minor lesions to severe or life-threatening encephalopathies.

Virus	Abbreviation of formal name	Subfamily	Genus	Genome Size	Genome Type
HSV-1	HHV-1	*Alphaherpesvirinae*	*Simplexvirus*	152	E
HSV-2	HHV-2	*Alphaherpesvirinae*	*Simplexvirus*	155	E
VZV	HHV-3	*Alphaherpesvirinae*	*Varicellovirus*	125	D
EBV	HHV-4	*Gammaherpesvirinae*	*Lymphocryptovirus*	172	C
HCMV	HHV-5	*Betaherpesvirinae*	*Cytomegalovirus*	236	E
HHV-6	HHV-6	*Betaherpesvirinae*	*Roseolovirus*	159	A
HHV-7	HHV-7	*Betaherpesvirinae*	*Roseolovirus*	153	A
KSHV	HHV-8	*Gammaherpesvirinae*	*Rhadinovirus*	165	B

Table 1.3. The eight human herpesviruses.

The formal names are those specified by the International Committee of Taxonomy of Viruses. The approximate genome size is in kbp and the genome type refers to the arrangements shown in Figure 1.2.

The clinical manifestations of HSV-1 and HSV-2 overlap greatly, although each has different levels of prevalence. Oral lesions, or "cold sores", are more commonly associated with HSV-1 (Whitley and Roizman, 1997), and genital lesions are more commonly associated with HSV-2 (Reeves *et al.*, 1981; Corey *et al.*, 1983). In rare cases, conjunctivitis, herpetic whitlow, keratitis or encephalitis may result. Primary infection and reactivation result in similar symptoms and neonatal infection is often life threatening (Whitley and Roizman, 1997). Primary infection with VZV causes chickenpox, with a rash appearing 14-15 days post-infection often accompanied by a fever (Gershon and Silverstein, 1997). The reactivated infection, "shingles", later appears at the relevant dermatome, often causing severe pain and sometimes, and, if the ophthalmic branch of the trigeminal nerve is involved, corneal ulceration (Donahue *et al.*, 1995). Primary HCMV infection often causes enlargement and fusion of macrophages, but is usually asymptomatic. However, infection can be problematic to the offspring of a mother infected during pregnancy, and in immunocompromised individuals, where symptoms commonly include retinitis and gastro-enteritis (Griffiths and Emery, 1997). HHV-6 and HHV-7 cause febrile illness and may act as opportunistic agents in immunosuppresed patients (Carrigan *et al.*, 1991; Cone *et al.*, 1993). HHV-6 has two variants, HHV-6A and HHV-6B, which are

known to cause exanthem subitum, pneumonia and encephalitis (Yamanishi *et al.*, 1988, 1997). EBV and KSHV are both associated with carcinomas. Primary EBV infection is often asymptomatic in young children but causes infectious mononucleosis in older children and adults. Infection is associated with Burkitt's lymphoma, nasopharyngeal carcinoma and other lymphoproliferative disorders and carcinomas (Neri *et al.*, 1991; Beaulieu and Sullivan, 1997). KSHV is associated with Kaposi's sarcoma, a vasculated skin lesion, and Castleman's disease (Renne *et al.*, 1996; Moore and Chang, 1997).

Most of the human herpesviruses are highly prevalent in the developed and developing world. However, HHV-8 is common only in some populations in sub-Saharan Africa and southern Europe (Sarmati, 2004). HHV-8 is closely associated with HIV-1 populations, and can be identified in most patients with HIV-associated Kaposi's sarcoma (KS), and a high incidence of KS is reported among HIV-infected homosexual men (Schulz, 1998).

1.7 General characteristics of HCMV

HCMV, which is formally designated HHV-5 (Table 1.3), was originally isolated by three groups in the 1950s (Smith, 1956; Rowe *et al.*, 1956; Craig *et al.*, 1957). The observations of cytomegalic inclusions in the nuclei of infected salivary gland cells led Weller (1971) to coin the term "cytomegalovirus", although "generalised cytomegalic inclusion disease" was used in a study by Wyatt *et al.* (1950) to describe specific infantile multi-organ disease. Although a variety of cell culture systems are permissive, the virus is propagated most efficiently in human fibroblasts (Pass, 2001). Virus growth in these cells is characteristically slow and causes cell enlargement and rounding (Figure 1.5). Cytopathic effect (CPE) can be well developed by 48 to 72 hours post-infection, and it can take around 7 to 14 days to produce a generalised CPE involving the entire monolayer.

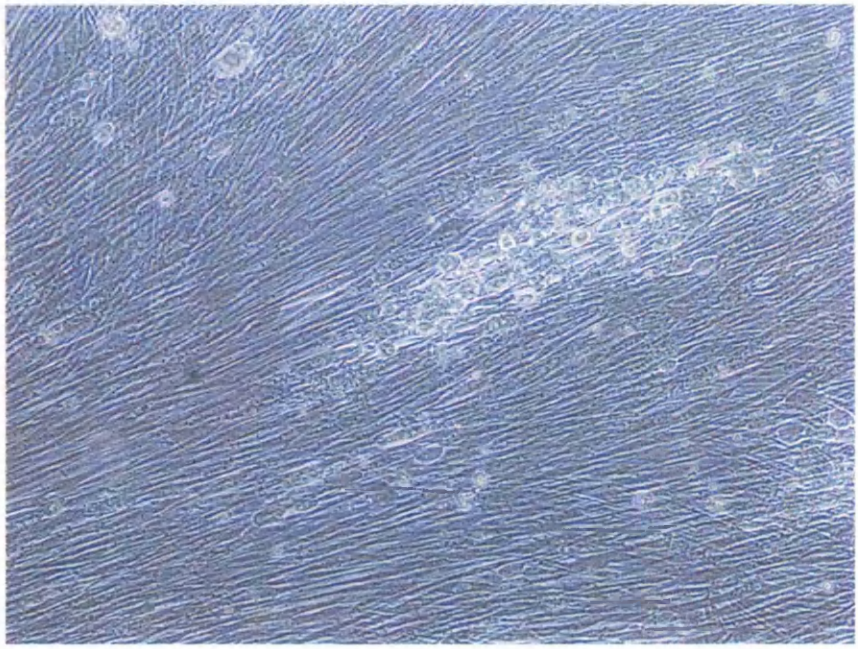


Figure 1.5. HCMV CPE in a monolayer of human fibroblasts.

Courtesy of D. Dargan, MRC Virology Unit, Glasgow. The fibroblast monolayer, infected with HCMV strain AD169, is shown at 10× magnification. Virus CPE and plaque formation can be seen as cells become enlarged and round.

HCMV has the ability to cause a lifelong subclinical infection in healthy adults, but uniquely among human herpesviruses establishes latency in monocytes (Kondo *et al.*, 1994; Hahn *et al.*, 1998). Although the majority of infections are asymptomatic, HCMV can lead to significant morbidity and mortality in neonates and immunocompromised hosts. The significance of HCMV as a pathogen has increased in the past 20 years with the increased prominence of transplants and immunosuppressive post-transplant therapies, and AIDS. These conditions predispose individuals to primary HCMV infection or to reactivation of a latent infection, both of which can result in acute, life-threatening disease.

1.8 HCMV disease and epidemiology

In some cases, primary HCMV infection may manifest as a mononucleosis-like syndrome. Approximately 8% of all cases of mononucleosis are caused by HCMV,

and show similar clinical symptoms to EBV-induced disease, including malaise, headaches, fatigue and fever (Klemola *et al.*, 1970; Jordan *et al.*, 1973). Congenital HCMV infection is clinically important because of the potential damage to the organs and the CNS in neonates. However, only 5-10% of neonates with congenital CMV infection exhibit clinical manifestations, the most common of these being petechiae, jaundice, hepatosplenomegaly, abnormal hearing and microcephaly (Bale *et al.*, 1990; Ramsay *et al.*, 1991; Boppana *et al.*, 1992). Most symptomatic neonates exhibit a combination of these conditions. The involvement of multiple organ systems *in utero* corresponds to the range of symptoms exhibited (Boppana *et al.*, 1992). In infants and young children, virus transmission via breast milk may cause symptoms, but only in rare cases. However, acquired HCMV infection is more likely to cause disease in infants with a low birth weight (Yeager, 1983).

In the immunocompromised host, primary infection or reactivation of latent HCMV can cause major complications (Hebart *et al.*, 1998; Ljungman, 1996; Prentice and Kho, 1997). In transplant patients, the transplanted organ and transfusion can be a source of infection at a time when the patient is maximally immunosuppressed. Disease severity tends to correspond to the degree of host immunosuppression, with symptoms ranging from febrile illness to life-threatening multisystem disease. The most common manifestations in transplant patients are fever, leukopenia, malaise, arthralgias and macular rash (Griffiths and Emery, 1997).

Life-threatening diseases such as pneumonitis (particularly in bone marrow transplantation), gastrointestinal ulceration and severe hepatic dysfunction may also result (Meyers *et al.*, 1982). The majority of morbidity for HCMV disease occurs when the recipient is seronegative but the donor is seropositive (Ho *et al.*, 1975; Smiley *et al.*, 1985; Meyers *et al.*, 1986; Apperley and Goldman, 1988; Winston *et al.*, 1990). HCMV is also thought to be the leading opportunistic infection among adults with AIDS. Almost all adults and approximately 50% of children with HIV infection show serological evidence of HCMV infection, and the risk of developing HCMV disease in people with advanced AIDS is directly related to quantity of HCMV DNA in the plasma (Spector *et al.*, 1998). In a study of AIDS patients with HCMV disease, approximately 85% developed retinitis, and just under 10% developed oesophagitis or colitis (Cheung and Teich, 1999; Jacobson and Mills, 1988;

Jabs *et al.*, 1989; Gallant *et al.*, 1992; Spector *et al.*, 1993; Boppana *et al.*, 1992). HCMV may also be a cofactor in the development of AIDS by enhancing HIV infection (Pass, 2001).

HCMV is universally distributed among human populations and antibody prevalence in normal adults ranges from 40-100% (Krech, 1973). In developing countries and in areas of developed countries with lower socio-economic factors, HCMV is generally acquired earlier in life and is of greater prevalence (Griffiths and Emery, 1997). Infection is not seasonal, but is endemic and present throughout the year (Gold *et al.*, 1990).

Transmission and infection via person to person contact can occur at any age and commonly takes place in settings where close contact with body fluids may occur. Infants and young children excrete HCMV in their respiratory tract and urine, and in settings such as day care centres the virus can be easily transmitted between children and adults (Stagno and Cloud, 1994). Transmission from mother to foetus occurs by three main routes: transplacental, intrapartum and breast milk. If the woman was infected long before pregnancy or acquires a primary infection during pregnancy, the virus may be transmitted across the placenta (Pass, 2001). Alternatively, local shedding of the virus, which appears to increase closer to the time of delivery, may result in intrapartum transmission (Stagno *et al.*, 1982).

Breast milk is the most common route of transmission from mother to neonate. With seropositive mothers, infants breast-fed for less than one month were extremely unlikely to be infected compared to a 39% rate of infection for infants breast-fed for more than one month (Dworsky *et al.*, 1983). Furthermore, almost 70% of infants whose seropositive mothers have detectable HCMV in their breast milk become infected. These routes of transmission from the mother do not generally result in neonatal disease, but they are common, and are therefore important modes of vertical transmission for maintaining HCMV infection in the population.

Post puberty, the most important mode of transmission is sexual, particularly among homosexuals (Handsfield *et al.*, 1985; Drew *et al.*, 1981; Collier *et al.*, 1987). Reinfection is also common in sexually active populations (Drew *et al.*, 1984; Spector

et al., 1984). The prevalence of HCMV infection is also higher among populations with other markers of sexual activity such as greater numbers of partners and sexually transmitted diseases (Chandler *et al.*, 1985; Pereira *et al.*, 1990; Sohn *et al.*, 1991). Transmission of HCMV to immunocompromised hosts may occur via blood products or transplanted organs (Van der Meer *et al.*, 1996). However, since this has been established as a source of infection, it is now routine to screen blood products and organs before transfusion and transplantation.

The quantity of viral load is directly related to the development of HCMV disease. An initial study of congenital HCMV showed that children with congenital infection had approximately one log higher viral load in their urine if they were asymptomatic when compared to children with asymptomatic perinatal infection (Stagno *et al.*, 1975). Significantly, symptomatic children had approximately two logs higher viral load in their urine. Subsequent studies have shown that high viral load, detected in urine and blood, is also the major determinant of HCMV disease in solid organ and bone marrow transplant populations (Cope *et al.*, 1997a, 1997b; Gor *et al.*, 1998). In the solid organ population, when peak viral load in urine for hosts was plotted against the presence of disease for the same individuals, the relationship between viral load and HCMV disease is clear (Griffiths *et al.*, 1999). However, as high viral load coincides with HCMV disease, qualitative PCR and monitoring CMV viremia are recommended as a marker to initiate pre-emptive therapy (Griffiths *et al.*, 1999).

1.9 The HCMV virion

The HCMV virion is generally typical of herpesviruses, but the capsid, tegument and envelope each have unique features. The mature HCMV capsid is 130 nm in diameter, 5% larger than in HSV-1, to create an internal capsid volume 17% larger (Butcher *et al.*, 1998; Chen *et al.*, 1999). This accommodates the HCMV genome, which is approximately 37% longer than HSV-1 but is more densely packed (Bhella *et al.*, 2000). Other differences include hexamer spacing, relative tilt, morphology of the tips of the hexons and the average diameter of the scaffold (Butcher *et al.*, 1998). However, like HSV-1, the HCMV capsid is composed of 162 capsomers: 12 pentavalent capsomers (pentons) at the vertices, and 150 hexavalent capsomers

(hexons) at the faces and edges (Butcher *et al.*, 1998). The 3D structure of the HCMV AD169 capsid, determined by cryo-electron microscopy (cryo-EM) and image reconstruction, is shown in Figure 1.6.

The capsid contains at least six proteins: the minor capsid protein (mCP), the mCP-binding protein (mC-BP), both of which are triplex proteins, the major capsid protein (MCP), the small capsid protein (SCP), and two assembly proteins (Britt and Boppana, 2004; Mocarski and Tan Courcelle, 2001). The MCP (pUL86) constitutes approximately 90% of the capsid protein mass and comprises the pentamers and hexamers (Chen *et al.*, 1999). The mCP (pUL85) and mC-BP (pUL46) are present in a 2:1 ratio and constitute the inter-capsomeric triplex that links adjacant capsomers (Chen *et al.*, 1999). The SCP (pUL48A) is essential for the generation of infectious progeny in tissue culture cells and coats the tips of the hexons (Borst *et al.*, 2001).

The assembly proteins are encoded by UL80 and UL80.5 (Oien *et al.*, 1997; Butcher *et al.*, 1998). The UL80.5 protein, 'assemblin', forms a scaffold by self-interaction via N-terminal sequences, and interacts with the MCP via C-terminal sequences (Gibson, 1996; Irmiere and Gibson, 1983; Oien *et al.*, 1997). Once nucleocapsids are assembled, the scaffolding proteins vacate the capsid following protease cleavage of a short C-terminal region, leaving space within the capsid to package the genome.

The tegument is an amorphous layer that lies between the nucleocapsid and envelope (Gibson, 1996; Zhou *et al.*, 1999). The HCMV tegument consists of many phosphorylated proteins, of which pp65 (UL83) and pp150 (UL32) are the major ones. Other tegument proteins which have been identified have roles in regulating viral gene expression, modifying host cell responses, and capsid egress (Britt and Boppana, 2004). The functions for many have still to be determined.

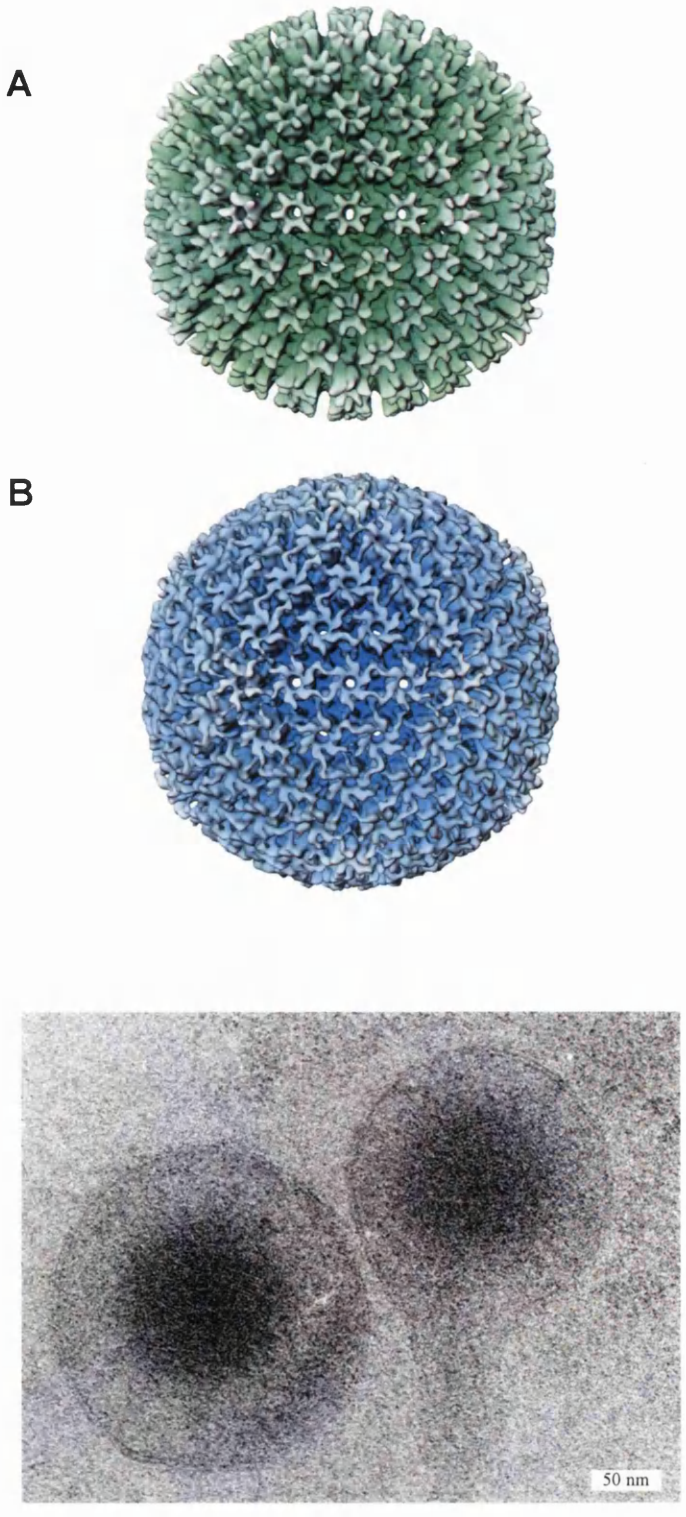


Figure 1.6 The HCMV virion.

Courtesy of D. Bhella, MRC Virology Unit, Glasgow. **(A, B)** Reconstructions of the capsid and virion, respectively, from cryo-EM images. **(C)** A cryo-EM image of HCMV virions. A protrusion from the envelope of one of the virions is evident.

The envelope, which encloses the virion, is a lipid bilayer derived from the host (Wildy *et al.*, 1960). Glycoprotein spikes protrude from the envelope, a number of which have so far been identified. Three glycoprotein complexes are known to form: one comprising covalently linked dimers of gB, one comprising gM and gN, and the third comprising gH, gL and gO (Britt and Mach, 1996; Spaete *et al.*, 1994). These glycoproteins (with the exception of gO) have homologues in other herpesviruses that form complexes in the virion envelope (Stinksi, 1976; Gibson, 1983; Pachl *et al.*, 1989; Britt and Mach, 1996).

The mature infectious virion is one of three types of particle produced in HCMV-infected cells and released into the extracellular medium in culture (Mocarski and Stinski, 1979; Irmiere and Gibson, 1983). The other two, non-infectious enveloped particles (NIEPs) and dense bodies, can be distinguished by physical differences (Irmiere and Gibson, 1983; Sarov and Abady, 1975; Gibson, 1996). NIEPs, which resemble virions, lack the viral genome but retain one scaffolding protein; pUL80.5, the capsid assembly protein (Gibson, 1996; Irmiere and Gibson, 1983; Oien *et al.*, 1997). Dense bodies are present only in the cytoplasm of infected cells, and lack both nucleocapsids and viral DNA (Irmiere and Gibson, 1983; Sarov and Abady, 1975). They are composed of tegument proteins, but principally of a single protein (pp65, encoded by UL83), and enclosed in a cytoplasmic membrane-derived envelope (Craighead *et al.*, 1972; Sarov and Abady, 1975). In AD169 virus preparations, they are found in approximately equal numbers to virions. It is thought that dense bodies may consist of surplus cellular and viral products, which accumulate in a storage vacuole that is subsequently voided into extracellular medium (Severi *et al.*, 1992).

1.10 HCMV attachment and entry into host cells

When a herpesvirus initially associates with the host cell, viral glycoproteins on the surface of the envelope have important roles in adsorption and penetration of virus into the cell, as has been shown for HSV-1 (Campadelli-Fiume, 1994). The most highly conserved herpesvirus glycoprotein, gB, is a multifunctional protein that has been shown to promote virion penetration into cells and transmission of infection from cell to cell (Compton *et al.*, 1992; Navarro *et al.*, 1993; Bold *et al.*, 1996).

In HCMV, the virus envelope fuses with the host cell plasma membrane in a pH-independent process that probably involves the gH/gL complex (Compton *et al.*, 1992; Keay *et al.*, 1991). Heparan sulphate interaction appears to play a role in this initial fusion event, as shown from studies of heparinase-treated cells, in which viral attachment was blocked (Neyts *et al.*, 1992; Compton *et al.*, 1993). The nucleocapsid is then released into the cytoplasm and migrates to the nucleus where DNA enters via nuclear pores. Tegument proteins, such as pp65 (UL83) also enter the nucleus, and can be detected within 30 minutes after virus inoculum exposure (Dal Monte *et al.*, 1996).

A different mechanism, that of endocytosis, has been proposed for HCMV entry into retinal pigmented epithelial (RPE) cells and endothelial (HUVEC) cells (Bodaghi *et al.*, 1999). In contrast to fibroblasts, where nucleocapsids are released directly into the cytosol, the virus appears to enter endocytotic vesicles in RPEs and HUVECs. In addition, pp65 could only be detected in the nucleus 1 to 2 hours after virus-cell contact, much later than in fibroblasts.

1.11 Transcription and translation

As in other herpesviruses, HCMV transcription is temporally regulated. Genes are expressed in a cascade, in three kinetic classes. Genes expressed following viral entry, independent of any other viral gene expression, are termed immediate-early (IE). Early (E) gene expression is dependent on IE products, and late (L) gene expression occurs last. UL123 (IE1), UL122 (IE2), UL36-37, TRS1/IRS1 and US3 are all IE genes, of which IE1, IE2 and US3 appear to be involved in subsequent gene regulation steps (Stenberg *et al.*, 1985; Weston, 1988). Translation of IE mRNAs is a prerequisite for transcription of E genes, which are found throughout the genome and have roles in DNA replication, nucleotide metabolism and immune evasion (Demarchi, 1981). L genes are transcribed from about 24 h p.i., effectively after the inception of viral DNA replication (Stamminger and Fleckenstein, 1990). L gene expression is divided into two subclasses; E-L (leaky late) or L (true late). These can be classified quantitatively by the signal intensity from HCMV microarrays, as shown

by Chambers *et al.* (1999). True L transcription is strictly dependent on DNA replication (Stasiak and Mocarski, 1992). Expression of L genes comes from throughout the genome and contributes proteins involved in virion assembly and morphogenesis. In cell culture, the different kinetic classes of transcripts can be separated using chemical inhibitors. Cycloheximide inhibits protein synthesis, stopping E transcription so that only IE transcripts are generated. Phosphonoacetic acid (PAA) or ganciclovir inhibit viral DNA replication, inhibiting L transcription so that only IE and E transcripts are generated.

Viral transcription trans-activators help to create a complex regulatory network that controls gene expression. The major HCMV trans-activators are the nuclear phosphoproteins $IE1_{491aa}$ 72 kDa (UL123) and $IE2_{579aa}$ 86 kDa (UL122). These IE proteins are expressed by differential splicing from the major immediate-early (MIE) gene locus (Figure 1.7). Expression from this locus generates differentially spliced polyadenylated transcripts. $IE1_{491aa}$ shares an 85 residue N-terminal domain with $IE2_{579aa}$, both of which have key roles in activating early and late gene expression. (Spector, 1996; Stenberg, 1996). $IE2_{579aa}$ plays a key role in switching from IE to E gene expression. In addition to $IE2_{579aa}$, $IE2_{425aa}$ is made from a transcript in which an additional intron is removed from UL122, and is less abundant. The most abundant IE2 gene product, $\gamma IE2_{338aa}$, is expressed exclusively from UL122 late in infection (Figure 1.7) (Stenberg, 1996; Jenkins *et al.*, 1994). $IE2_{338aa}$ both represses IE gene expression and trans-activates E and L genes.

The MIE locus is under the control of a strong promoter-enhancer (MIEP), which is subject to positive and negative regulation. The MIEP contains a large number of binding sites for cellular transcription factors and specific viral proteins (Mocarski, 1996). Similar enhancer elements control US3, another IE gene (Biegalke, 1999). $IE1_{491aa}$ and $IE2_{579aa}$ autoregulate IE1/IE2 and US3 expression by activating via NF-κB binding sites and repressing via IE2 cis-repression signals in these promoter-enhancers. The MIEP can be divided into three main regions; the enhancer region, the specific NF-1-binding region, and the modulator region. The complex array of binding sites in the enhancer and variety of cellular factors binding to these allows the MIEP to respond to the intracellular environment in a range of cell types. Binding of

transcription factors to the modulator region has been shown to repress transcription, but the role of this region is not fully understood.

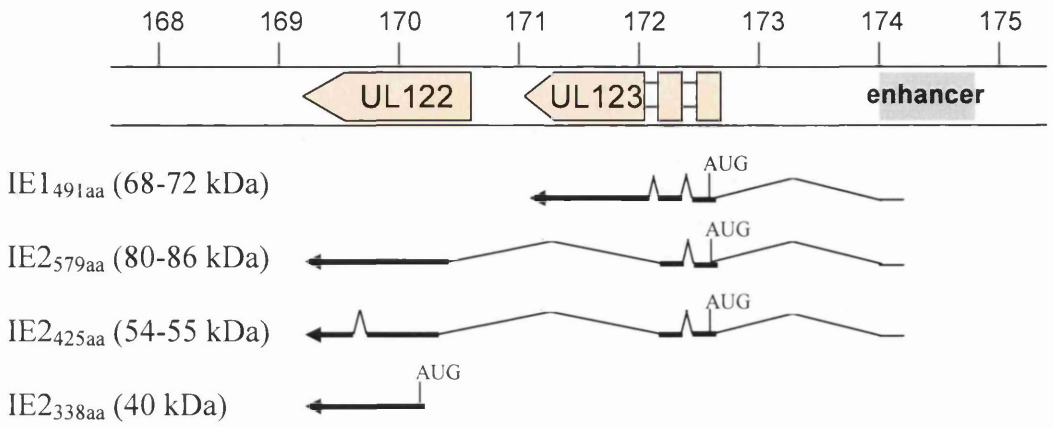


Figure 1.7 Differential splicing from the MIE gene locus.

Adapted from Mocarski (1996). The HCMV strain AD169 genome is shown on the top line (the scale is in kbp). The principal spliced IE transcripts are shown with the coding regions as thick black lines. In the spliced transcripts, the two exons immediately after the AUG codon make up the 85 residue N-terminal domain shared by the IE1 and IE2 transcripts.

1.12 Evasion of host defences

Natural killer (NK) cells and specific cellular and humoral immunity all play important roles in controlling HCMV disease. Many studies of host defence utilised MCMV, which provides a tractable animal model for CMV infection. Beige mice, which are genetically deficient in functional NK cells, showed increased susceptibility to MCMV, and depletion of NK cells from non-beige mice leads to a more serious infection after challenge (Shellam *et al.*, 1981; Bukowski *et al.*, 1984). Furthermore, spleen cells with an NK cell phenotype were shown to confer protection from a lethal MCMV challenge (Bukowski *et al.*, 1984, 1985). An *in vitro* study of NK responses to HCMV showed they were mediated by large granular lymphocytes, independent of interferon, and were manifest at the time early viral proteins were expressed (Borysiewicz *et al.*, 1985).

Severe HCMV disease (with the exception of congenital infection) is seen almost exclusively in patients with profoundly impaired cell-mediated immunity, such as bone-marrow transplant recipients or AIDS patients. Cytotoxic T-lymphocyte (CTL) responses to CMV in humans also appear to be noticeably suppressed after mononucleosis in infants and immunocompromised patients with HCMV disease (Levin *et al.*, 1979; Pass *et al.*, 1983; Pollard *et al.*, 1978; Zanghellini *et al.*, 1999). Consequently, important roles for CTL responses, and CD4+ helper T-lymphocytes in γ-interferon production were suggested (Quinnan *et al.*, 1978; Hengel *et al.*, 1994). CD4+ helper T-lymphocyte responses to multiple structural and non-structural HCMV proteins have been demonstrated in humans with naturally acquired infection, and CTL response to HCMV was associated with clinical recovery from disease (Quinnan *et al.*, 1982). Disease severity corresponds to the degree of impairment of T-lymphocyte responses. The humoral antibody response is important in settings such as neonatal disease and solid organ transplant (Yeager *et al.*, 1981; Snydman and Falagas, 1996).

To avoid this variety of host defence mechanisms, HCMV has developed strategies to circumvent elimination of active infection, via gene products that can interfere with immune response (Huang *et al.*, 1980; Wiertz *et al.*, 1997). A number of related mechanisms are thought to assist HCMV escape from CTL response. Phosphorylation of an important CTL target, the 72 kD IE protein, by pp65 (UL83) prevents antigen processing (Gilbert *et al.*, 1996). The US3 protein interferes with MHC class I complexes, either by destroying them or sequestering them in the endoplasmic reticulum (Jones *et al.*, 1996). The US6 protein inhibits antigen presentation, and the US2 and US11 proteins interact to re-import class I complexes from the endoplasmic reticulum to the cytosol where they are degraded in the proteosome (Wiertz *et al.*, 1996; Ahn *et al.*, 1996).

HCMV escape from NK cell attack is becoming better understood. One mechanism is the up-regulation of surface expression of HLA-E by the UL40-encoded glycoprotein, which is homologous to a class I MHC sequence. HLA-E expressed on the cell surface protects potential target from NK cell-mediated lysis by binding to inhibitory

sensors on surface of NK cells (Tomasec *et al.*, 2000). A protein encoded by UL18 also has a similar function (Hassan-Walker *et al.*, 1998). More recently, UL141 has been shown to block the surface expression of CD155, a ligand for two NK cell-activating receptors (Tomasec *et al.*, 2005). In addition, UL16 has been shown to sequester another ligand, NK62D, to assist virus escape from NK cell attack (Odeberg et al., 2003; Vales-Gomez et al., 2003).

Other mechanisms of HCMV immune evasion have been proposed. These include the binding and sequestering of β-chemokines by US27, UL28 and UL33 (Bodaghi *et al.*, 1998), down-regulating MHC class II expression on the surface of macrophages (Fish *et al.*, 1996) and competing with the TNF receptor superfamily by UL144 (Benedict *et al.*, 1999).

HCMV can also prevent cell destruction, either by interfering with apoptosis or protecting infected cells from complement-mediated lysis (Zhu *et al.*, 1995; Spear *et al.*, 1995; Spiller *et al.*, 1996). The UL36-UL38 IE gene locus encodes two anti-apoptotic genes: gpUL37 and pUL37x1 (Colberg-Poley *et al.*, 2000). These co-localise with Bax, which is involved in amplifying the apoptotic signal, in the mitochondria of the cell (Poncet *et al.*, 2004). UL146 encodes a interleukin-8-like chemokine (vCXC-1) which influences the behaviour of leukocytes during infection (Penfold *et al.*, 1999), and UL111A encodes an interleukin-10 homologue which possesses immunosuppressive properties, by inhibiting cytokine synthesis (Spencer *et al.*, 2002).

1.13 Treating HCMV infection

There are several antiviral drugs used to treat the clinical manifestations of HCMV, but no vaccine is in use. In the USA there are currently only four approved antiviral therapies; ganciclovir, foscarnet and cidofovir, all of which target the viral DNA polymerase, and fomivirsen, a novel antisense compound (De Clercq, 2004). Ganciclovir is a nucleoside analogue, a derivative of acyclovir, which is used to treat retinitis in immunocompromised patients (De Clercq, 2004). Ganciclovir is also

available as a pro-drug (valganciclovir) with improved gastrointestinal absorption (Cvetkovic and Wellington, 2005). Foscarnet, a pyrophosphate analogue, is effectively a second-line treatment for HCMV infection due to associated nephrotoxicity (Palestine *et al.*, 1991). Cidofovir is also a nucleoside analogue, and an alternative to ganciclovir, that is used mainly to treat retinitis in AIDS patients and cutaneous disease, but is linked to neutropenia and nephrotoxicity (Lea and Bryson, 1996). Fomivirsen is the latest HCMV antiviral drug to be approved, and the first novel antisense compound of any kind to be approved for therapeutic use (Galderisi *et al.*, 1999). It is a 21-nucleotide phosphorothioate oligonucleotide that operates differently from other antiviral drugs by targeting mRNA from the MIE transcriptional unit (UL123) and inhibiting translation (de Smet *et al.*, 1999; Jabs and Griffiths, 2002).

Use of anti-HCMV treatments has increased in the last 20 years with the increased prominence of transplants and immunosuppressive post-transplant therapies, and AIDS. The most common use of these drugs is the treatment of retinitis in AIDS patients. Intravenous treatment with ganciclovir, foscarnet or cidofovir are the most common therapies for slowing the progression of retinitis or bringing acute retinitis into remission (Lalezari *et al.*, 1998; Palestine *et al.*, 1991; Spector *et al.*, 1993). Antiviral treatment lengthens the progression of retinitis from two to three weeks (in untreated cases) to two to three months. Local treatment of retinitis via intravitreal injection with ganciclovir is an additional therapeutic approach (Masur *et al.*, 1996). Fomivirsen is used exclusively for intravitreal injection. As fomivirsen cannot suppress HCMV infections elsewhere in the body, it is recommended that therapy is combined with oral ganciclovir treatment (Jabs and Griffiths, 2002). Combination therapy, generally of foscarnet and ganciclovir, is widely used, and is specifically recommended for patients with CNS or neurological disease (Arribas *et al.*, 1996). In addition to treating retinitis, intravenous ganciclovir and foscarnet therapy has also been shown to be beneficial for other clinical manifestations; oesophagitis, enterocolitis, encephalitis, peripheral neuropathy and pneumonitis.

Antiviral treatment of bone marrow and solid organ transplant patients is largely similar. In the pre-antiviral era, mortality due to HCMV was much higher in bone marrow transplant patients than solid organ transplant patients. Ganciclovir treatment

has been shown to reduce the incidence of HCMV disease in the first 100 days post-transplant, but overall mortality and graft survival were not improved (Boeckh *et al.*, 1996). Detecting pp65 antigenemia in white blood cells and PCR detection of HCMV in blood are two methods used to suggest pre-emptive antiviral therapy in transplant patients to help combat disease (Hebart *et al.*, 1998). Antiviral therapy for congenital HCMV infection is not a common approach, but ganciclovir therapy has been shown to ameliorate some disease syndromes, and may help to limit neurodevelopment injury, particularly sensorineural hearing loss (Schleiss, 2004; Kimberlin *et al.*, 2003).

The toxic side effects of ganciclovir (myelotoxicity), foscarnet and cidofovir (nephrotoxicity) can create major obstacles when treating HCMV disease. Another major concern of long-term therapy is the emergence of resistant viral isolates (Drew *et al*, 1991; Baldanti *et al.*, 2004). Resistance to ganciclovir is associated with a variety of mutations in UL97 (protein kinase) and, in approximately 5% of cases, UL54 (DNA polymerase) (Perez, 1997; Erice, 1999). Resistance to foscarnet is associated with mutations in UL54 (Chou *et al.*, 2003). Therefore, the emergence of fomivirsen as a novel antiviral therapy is important when resistance to other antiviral drugs emerges, particularly as it works by an alternative mechanism. The development of anti-HCMV therapies is necessary to combat the emergence of antiviral drug resistance, with a priority of limiting the type of toxic side effects produced by the current therapies.

Currently, there is no vaccine available for HCMV. One major problem in developing a vaccine is the ability of HCMV to establish a persistent, latent infection. HCMV strain Towne, originally isolated from a child with congenital infection, was passaged more than 125 times *in vitro* to achieve attenuation, and did not produce a febrile illness in challenge studies (Plotkin *et al.*, 1975). Immunisation with this strain as a vaccine does induce antibody response and lymphocyte proliferative responses (Starr *et al.*, 1981). However, the Towne vaccine appears to induce only limited protection and does not confer the protection from reinfection that comes from naturally acquired infection (Adler *et al.*, 1996). MCMV and GPCMV models have been used to show that immunisation with viral glycoproteins and virus mutants can provide protection from virus challenge (Rapp *et al.*, 1992; Harrison *et al.*, 1995; Gill *et al.*,

2000; Morley *et al.*, 2002). Some studies have focused on subunit vaccines, specifically directed against HCMV glycoprotein B (Speckner *et al.*, 1999).

1.14 Cell tropism

HCMV is able to infect many organ systems, causing a variety of diseases including mental retardation, retinitis and vascular disorders. Consequently, HCMV is able to replicate in a variety of cell types: fibroblasts (Albrecht *et al.*, 1980), macrophages (Ibanez *et al.*, 1991), smooth muscle cells (Plachter *et al.*, 1996), endothelial cells (Sinzger *et al.*, 1995), leukocytes (Gerna *et al.*, 2003), epithelial cells (Heieren *et al.*, 1988; Waldman *et al.*, 1991), astrocytes (Poland *et al.*, 1990) and trophoblasts (Hemmings *et al.*, 1998). However, the virus can normally be propagated efficiently *in vitro* only in human fibroblast cells such as HFL1 (human foetal lung) and HFFF2 (human foetal foreskin fibroblasts), where CPE develops slowly and focally. Studies of HCMV cell tropism have subsequently shown that highly passaged laboratory strains, such as AD169, Towne and Davis, have lost certain host range capabilities of low passage strains and clinical isolates, and infection in other cell culture systems is often abortive (Sinzger and Jahn, 1996).

In contrast, low passage HCMV strains and clinical isolates have been shown to infect endothelial cells and leukocytes (Hahn *et al.*, 2002). Accordingly, growth restrictions in cell culture appear to correspond with extensive passaging of laboratory stocks in fibroblasts (Brown *et al.*, 1995; Sinzger *et al.*, 2000). Furthermore, extensive passaging of the laboratory strains is associated with changes in the virus genome, including large-scale deletions (Cha *et al.*, 1996).

Virus cell tropism is commonly determined during the initial stage of the replicative cycle or at the initiation of viral gene transcription (Tyler and Fields, 1996). However, the factor determining HCMV endothelial cell tropism, as suggested by differences between strains, appears to be transport of penetrated capsids towards the nucleus of endothelial cells (Sinzger *et al.*, 2000).

It seems likely that genes susceptible to ready mutation upon passage in cell culture have roles in tropism. UL128, UL130 and UL131A are genes that share a polyadenylation site, and are predicted to encode secreted proteins. UL128 and UL130 have sequence characteristics of a CC-chemokine (Akter *et al.*, 2003; Novotny *et al.*, 2001). Multiple strain analysis showed that many fibroblast-passaged strains have visibly disabling mutations in UL128, UL130 or UL131A (Akter *et al.*, 2003; Hahn *et al.*, 2004; Dolan *et al.*, 2004) Adaptation of HCMV to fibroblasts is associated with disruption of any one of these genes, and each is required for growth in endothelial cells and transmission to leukocytes (Hahn *et al.*, 2004). Furthermore, when these genes were supplied in *trans*-complementation experiments, growth rescue was observed (Hahn *et al.*, 2004). Other herpesviruses passaged in cell culture (CCMV and SCMV strain Colburn) also exhibit disruptions of UL128 (Davison et al., 2003a). Most passaged HCMV strains also have mutations in at least one member of the RL11 gene family (usually RL13), which mostly encode glycoproteins and have also been suggested to have roles in tropism (Dunn *et al.*, 2003; Davison *et al.*, 2003b). These studies suggest that fundamental biological properties may be lost *in vitro* according to cell culture system used. AD169 may reacquire tropism for endothelial cells and leukocytes *in vitro* (Gerna *et al.*, 2003). As AD169 lacks a 15 kbp DNA segment in U_L, this suggests that factors specifying tropism in these cells are encoded by viral genes elsewhere in the genome.

1.15 Latency

Like all herpesviruses, HCMV persists indefinitely within its host. Cells in bone marrow and peripheral blood are a key reservoir (Kondo *et al.*, 1994). Myelomonocytic stem cells resident in bone marrow maintain the latent infection, and latently infected CD14+ monocytes circulate in peripheral blood (Hahn *et al.*, 1998). Differentiation of these latently infected monocytes into macrophages then leads to reactivation and productive infection. Monocytes are the primary cell type infected in the blood during acute HCMV infection and are the predominant infiltrating cell type found in infected organs (Smith *et al.*, 2004). Epithelial cells are also thought to play a part in this model of HCMV latency. Epithelial cells are infected by HCMV-containing bodily fluids, then the virus replicates and spreads to monocytes in the

peripheral blood (Jarvis and Nelson, 2002; Smith *et al.*, 2004). Latency could then be established by monocyte migration into bone marrow (Kondo et al., 1994; Smith *et al.*, 2004). It had been thought that vascular endothelial and smooth muscle cells may harbour latent HCMV virus. However, sensitive PCR assays by Reeves *et al.* (2004) could not detect viral sequences (specific to the MIE locus) in any of the individuals tested. Another site of HCMV latency is myeloid dendritic cell progenitors (Hahn *et al.*, 1998; Reeves *et al.*, 2005). Differentiation of these cells to mature dendritic cells leads to reactivation, leading to viral lytic gene expression and an increase in viral copy number (Reeves *et al.*, 2005).

Induction of IE lytic gene expression appears to be preceded by chromatin remodelling of the MIE region of the viral genome. Latent HCMV in dendritic cell progenitors is thought to associate with a closed, transcriptionally silent chromatin conformation, as indicated by association with HP1, a silencing protein (Murphy *et al.*, 2002; Reeves *et al.*, 2005). On differentiation to mature dendritic cells, an association with acetylated histones was made which is consistent with an open, transcriptionally active chromatin conformation. Thus, viral lytic IE gene expression is reactivated. This mechanism is also linked to the down-regulation of HDAC1 protein, a corepressor of other transcriptional repressors (Reeves *et al.*, 2005).

In contrast to this model of transcriptional silencing, some HCMV transcripts appear to be expressed during latency. Although viral replication is not detectable during latency, viral transcripts from the major regulatory gene locus (IE1/IE2) have been detected in granulocyte-macrophage progenitors (Kondo *et al.*, 1994, 1996). These viral transcripts have different structures from those transcribed during productive infection, and antibodies to the proteins encoded by these transcripts could be detected in the serum of seropositive individuals (Kondo *et al.*, 1996). In addition, a vIL-10 transcript that differed from that reported during productive infection has been detected using an experimental model of latency in which human hematopoietic cells were latently infected with HCMV (Jenkins *et al.*, 2004). An antisense transcript spanning the UL81-82 locus has also been identified, but in this study no IE genes were expressed (Bego *et al.*, 2005). This creates a confusing picture of events during latency and much further work is required to elucidate a functional model.

1.16 Genome analysis of HCMV strain AD169

The complete DNA sequence of HCMV strain AD169 was determined by chain termination sequencing of M13 clones generated from *Hin*dIII fragments cloned in plasmids (Chee *et al.*, 1990; Oram *et al.*, 1982). Sequence analysis identified ORFs on the basis of three criteria: (1) an ORF size of 300 bp or greater, (2) overlap with other ORFs of not more than 60%, and (3) codon usage patterns. Consequently, 778 ORFs were identified, of which 581 were extensively overlapped by larger ORFs, leaving 197 candidate putative protein-coding ORFs. The gaps in the sequence were then examined for ORFs of less than 300 bp, for which additional criteria were employed: the presence of potential transcription signals and initiation codons, homology to other ORFs or known genes, the presence of protein structural or functional motifs in the amino acid sequence, and codon usage patterns. Consequently, a genome map of 189 ORFs was proposed (208 if diploid genes were counted twice). This map is shown in Figure 1.8.

The ORFs were numbered sequentially and prefixed to identify which segment of the genome they relate to: TR_L (11,247 bp; TRL1-14), U_L (166,972 bp; UL1-UL132), IR_L (11,247 bp; IRL1-14), IR_S (2,524 bp; IRS1), U_S (35,418 bp; US1-US36) and TR_S (2,524 bp; TRS1). The J ORFs (J1L, J1I and J1S), present in direct copies at the ends of the genome and an inverted copy between IR_L and IR_S, contain the *a* sequence which carries signals for cleavage and packaging of the viral genome (*pac*-1 and *pac*-2; Mocarski *et al.*, 1987). The 229,354 bp genome has an average G+C content of 57.2% and a class E structure (Figure 1.2). The ORFs, of which four were predicted to be spliced, were arranged about equally on both strands of the genome with little overlap. Further analysis revealed sets of related genes that were grouped into nine families (Chee *et al.*, 1990). Of the nine families identified, three were pairs of genes (the UL25, UL82 and US2 families) and the others were groups of three to fourteen genes (the RL11, US6, US22, GCR, US1 and US12 families). Relatively few polyadenylation signals were postulated in the genome, and many genes, especially in gene families, were predicted to have coterminal 3'-ends.

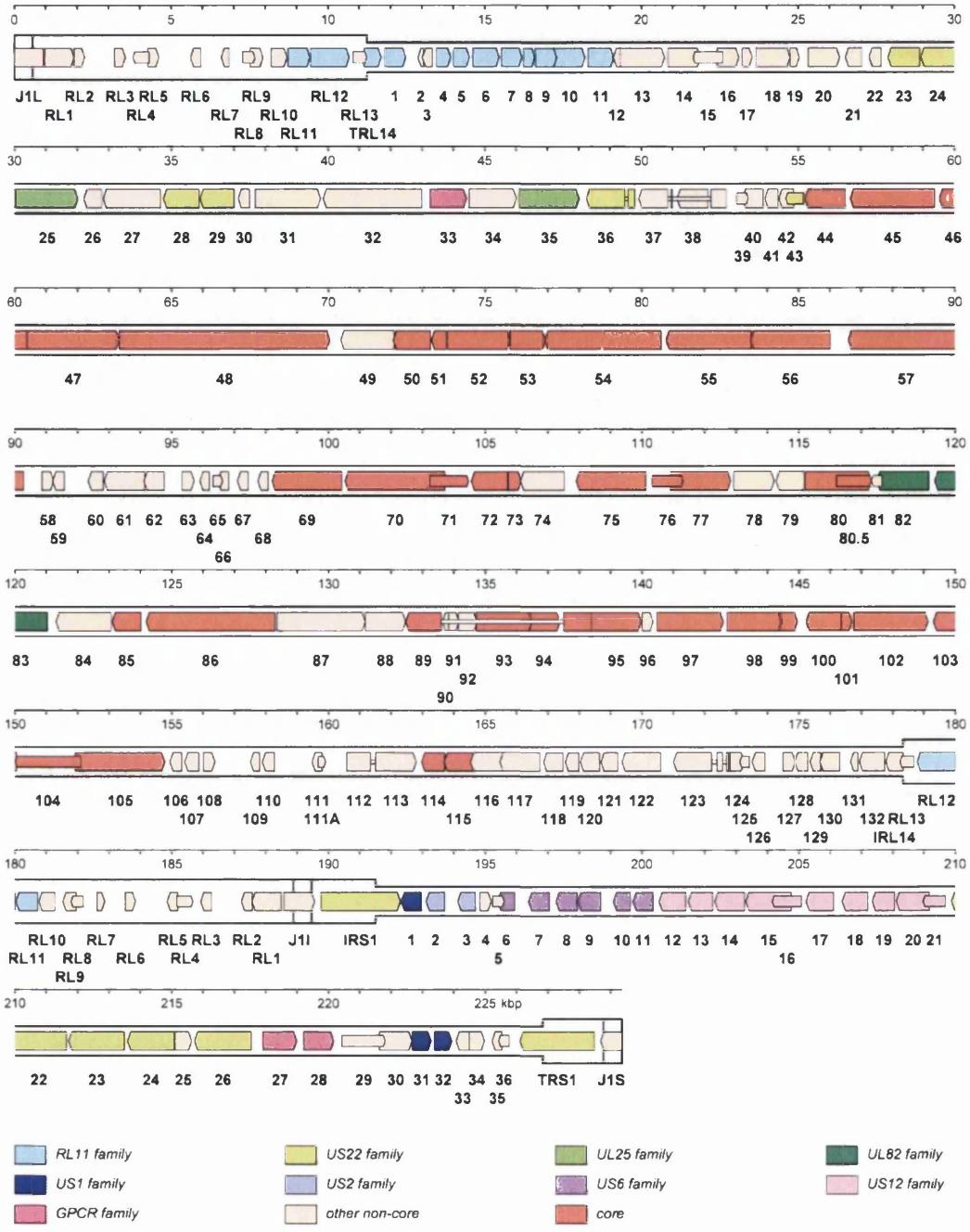


Figure 1.8. Original genome map of AD169.

The scale above the ORFs is in kbp. The thicker regions denote the inverted repeats and the thinner regions denote the unique long (U_L) and unique short (U_S) regions. Protein-coding regions are shown as open arrows above the gene nomenclature with prefixes omitted. The exons of spliced genes are connected by narrow horizontal bars. HCMV core genes and gene families are shown in different colours as indicated. Provided by A. Davison, MRC Virology Unit, Glasgow. Adapted from Chee *et al.* (1990).

1.17 Errors and developments

The analysis of the AD169 genome by Chee *et al.* (1990) was a landmark. Subsequent analyses of this laboratory strain have identified a few errors in the sequence, leading to the reinterpretation of the affected ORFs. Two differences were identified in UL102, one extending the 5'-end and the other affecting the reading frame of the ORF (Smith and Pari, 1995). These changes effectively replaced UL101 and UL102 with a single, larger ORF, also called UL102 (Figure 1.9). The 3'-end of US28 was also extended when an error in the original sequence was identified (Neote *et al.*, 1993). Further differences were identified in AD169 in the form of single residue insertions (Davison *et al.*, 2003a). Two of these disrupt UL15 and contribute to the emergence of an ORF on the other strand (UL15A, Figure 1.9), and one difference in US22 causes extension of the 5'-end. All of these differences represent errors in the original AD169 sequence.

Substantial differences between strains were identified in a comparative study of strains AD169, Toledo and Towne (Cha *et al.*, 1996). Southern blots indicated that Toledo and Towne contained additional sequence at the right end of U_L. The low-passage, virulent Toledo strain was found to contain a DNA segment absent from the high passage laboratory strains. Approximately 13 kbp of this additional U_L sequence was not present in Towne and approximately 15 kbp was not present in AD169. Using the same criteria employed by Chee *et al.* (1990), 19 additional ORFs in Toledo and four in Towne were identified, compared to AD169. Most of the novel ORFs in Toledo were predicted to encode membrane proteins, and none was found to be homologous to ORFs in other β-herpesviruses (Cha *et al.*, 1996). From these additional genes, three further HCMV gene families were identified (Davison *et al.*, 2003a).

Although Toledo contains an additional 15 kbp DNA segment, the genome is only 5 kbp larger than AD169 (approximately 235 kbp) since IR_L is much shorter in Toledo than in AD169 or Towne (Cha *et al.*, 1996). This is because the diploid genes in AD169 TR_L/IR_L are haploid in Toledo. A subsequent study by Lurain *et al.* (1999) identified these 19 ORFs in a range of clinical isolates of HCMV, indicating that they

are highly likely to be part of wild-type HCMV. It was speculated that part of the additional DNA segment in Toledo is in an inverted orientation compared to the corresponding region in clinical isolates (Cha *et al.*, 1996; Lurain *et al.*, 1999). Analysis of the corresponding region in chimpanzee CMV (CCMV) supported the view that Toledo contains an inverted region (Davison *et al.*, 2003a). An error in this extra segment in Toledo was also subsequently identified, as a single residue insertion in UL145 caused extension of the 5'-end of this ORF (Davison *et al.*, 2003a). Some stocks of AD169, including that analysed by Chee *et al.* (1990), were also found to contain a 929 bp deletion affecting two genes: UL42 and UL43 (Dargan *et al.*, 1997; Mocarski *et al.*, 1997). This changes the reading frame at the 5'-end of UL42, expands UL43 from 187 to 423 codons and replaces UL41 with a new ORF, UL41A.

ORF	Donor	Acceptor	Reference
UL22A	27,193	27,277	Rawlinson and Barrell, 1993
UL33	43,087	43,209	Margulies *et al.*, 1996
UL36	49,575	49,471	Kouzarides *et al.*, 1988
UL37	52,219 50,947	50,989 50,842	Kouzarides *et al.*, 1988
UL89	137,502	133,599	Chee *et al.*, 1990
UL111A	159,857 160,134	159,934 160,218	Kotenko *et al.*, 2000
UL112	161,345	161,503	Wright *et al.*, 1988
UL119	167,563	167,474	Leatham *et al.*, 1991
UL122	172,396	172,225	Stenberg et al., 1989
UL123	173,610 172,695 172,396	172,782 172,580 172,225	Stenberg et al., 1989
UL128	175,459 175,201	175,335 175,081	Akter *et al.*, 2003
UL131A	176,589	176,480	Akter *et al.*, 2003

Table 1.4. Spliced HCMV genes.

The splice donor and acceptor sites refer to strain AD169.

Although the original analysis of AD169 identified only four spliced genes (Chee *et al.*, 1990), a total of 12 have now been characterised (Figure 1.9). The splice donor and acceptor sites within these genes are shown in Table 1.4. Other reports of spliced ORFs remain to be fully characterised (Rawlinson and Barrell, 1993). A number of additional ORFs were also characterised, such as UL48A (Gibson *et al.*, 1996).

1.18 Reinterpreting the genome

Construction of an infectious bacterial artificial chromosome (BAC) clone of AD169 was reported by Yu *et al.* (2002). The authors used this clone in a subsequent study to introduce defined mutations by random or site-directed transposon mutagenesis, producing viral mutants for all ORFs deemed to have a high probability of encoding proteins (Yu *et al.*, 2003). Mutants were selected and their growth assayed in cultured human fibroblasts to identify 41 essential, 88 non-essential and 27 augmenting ORFs in AD169. A similar approach was used at the same time by Dunn *et al.* (2003), who generated a collection of viral mutants by deleting 162 ORFs from a BAC clone of Towne. In assessing the requirement for each ORF for replication in fibroblasts *in vitro*, 45 ORFs were identified as essential. Of these, 39 were also identified by Yu *et al.* (2003).

The differences are in six ORFs which were identified as essential for Towne replication but as augmenting in AD169 (UL48, UL71, UL76, UL90, UL94 and UL96), and two ORFs which were identified as essential for AD169 replication but not Towne (UL61, which is probably non-protein coding and is located within *ori*Lyt, and UL80.5, a surprising result as it is completely overlapped by UL80, an essential ORF). Almost all the essential ORFs, for which no growth of the mutant was observed in fibroblasts, are core genes (conserved in herpesviruses) and are located in the central region of HCMV. Furthermore, genes UL64 and UL24 were required for replication in retinal pigment epithelial (RPE) cells and microvascular endothelial cells (HMVEC), respectively, but not in fibroblasts, indicating roles in viral tropism in different cell types (Dunn *et al.*, 2003).

A study by Murphy *et al.* (2003a) utilised a gene finder algorithm, the Bio-Dictionary-based Gene Finder, to identify potential new protein-coding ORFs based on matches to amino acid patterns derived from the large collection of proteins in the Swiss-Prot/TrEMBL database (Murphy *et al.*, 2003a; Shibuya and Rigoutsos, 2002). Protein-coding ORFs were identified and then screened for orthologues in CCMV, RhCMV and MCMV. In total, 37 previously annotated ORFs were discarded, and 12 ORFs with coding potential were added. This view was amended in a subsequent study by the same group, in which further strains of HCMV were BAC-cloned (Murphy *et al.*, 2003b). In addition to AD169 and Towne, four low passage isolates (Toledo, PH, TR and FIX) were BAC-cloned and the genomes sequenced. The sequences were compared and a total of 252 ORFs were identified that were conserved in all four clinical isolates. ORFs were subsequently discounted if the ATG initiation codon was less than 80 codons from a termination signal, or if the ORF was overlapped by another ORF by more than 396 bp (the size of the largest overlap between recognised HCMV ORFs, involving UL76 and UL77). As a result, a list of 29 newly recognised ORFs with the potential for coding proteins were identified (C-ORF1 to C-ORF29). Of the 12 new ORFs recognised in the previous study (Murphy *et al.*, 2003a), two were discarded in the later work (ORF4 and ORF12) to make a total of 39 newly recognised ORFs in the genome. However, most of these ORFs introduced by these analyses are small, overlap recognised HCMV genes and, most significantly, are not conserved in CCMV. In addition, assignment of these ORFs to HCMV is based on the predictions of a single algorithm.

At the same time as these studies, the sequence of CCMV, the closest known relative of HCMV, was compared with AD169 and Toledo on the basis of conservation of potential protein-coding regions and other sequence features, and the content of these genomes interpreted (Davison *et al.*, 2003a). Of the previously proposed HCMV ORFs by Chee *et al.* (1990), 51 were discounted, modified interpretations were made for 24 (including the assignment of multiple exons), and 10 novel ORFs were proposed (Davison *et al.*, 2003a). Within the additional region in Toledo, the genes recognised by Cha *et* al. (1996) were reinterpreted by redefining four, discounting five and introducing five novel ORFs. This region of Toledo is not collinear with CCMV due to an apparent inversion in the region, originally identified by Prichard *et al.* (2001).

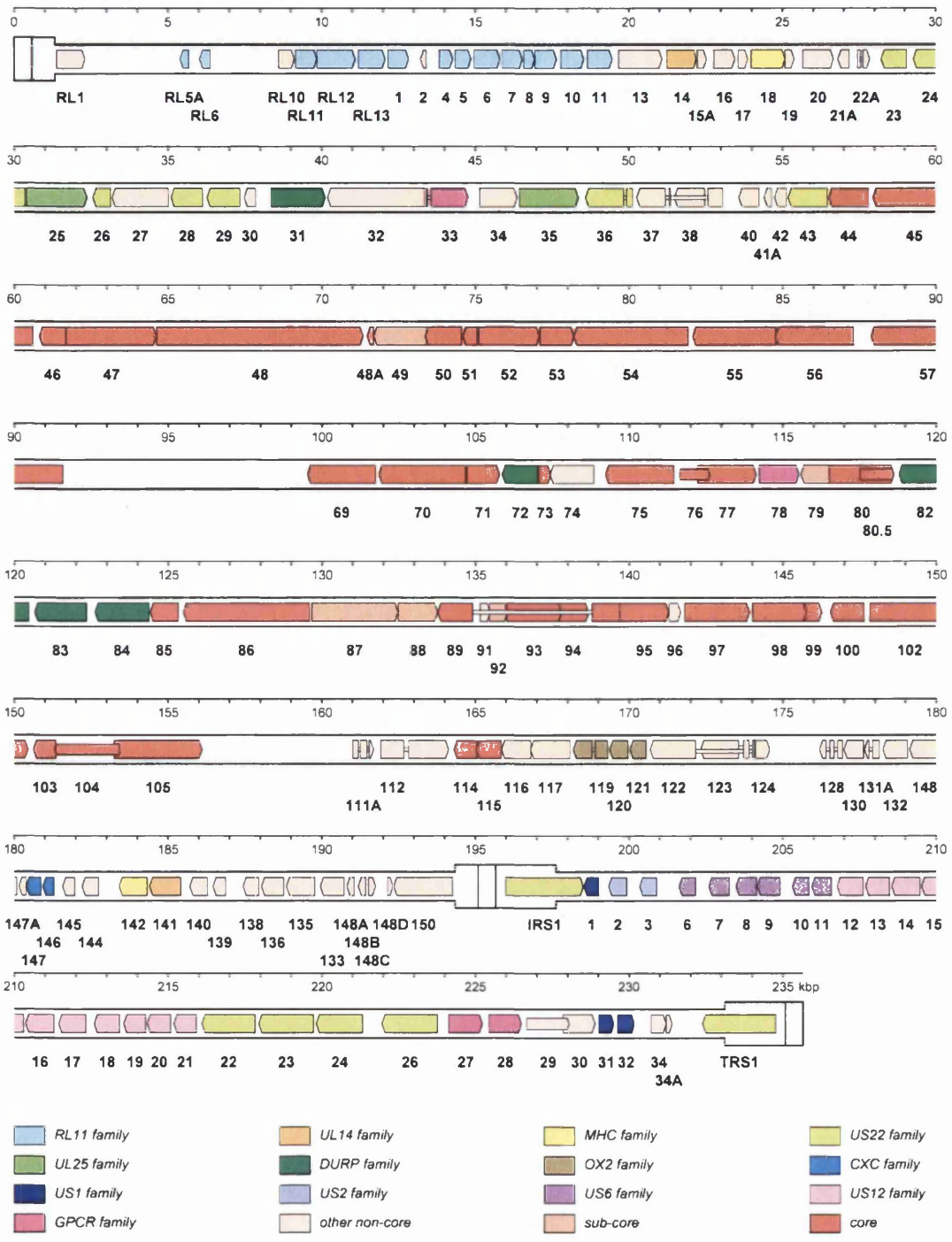


Figure 1.9. Genome map of Merlin.

The scale above the ORFs is in kbp. The thicker regions denote the inverted repeats and the thinner regions denote the unique long (U_L) and unique short (U_S) regions. Protein-coding regions are shown as open arrows above the gene nomenclature with prefixes omitted. The exons of spliced genes are connected by narrow horizontal bars. HCMV core genes and gene families are shown in different colours as indicated. Provided by A. Davison, MRC Virology Unit, Glasgow.

Consequently, the genetic content of HCMV was significantly altered from the original analysis of Chee *et al.* (1990), and a total of 164-167 genes predicted. Although the basis of this analysis was conservation of ORFs, virus-specific genes were recognised in each genome. CCMV lacks counterparts of HCMV genes UL1, a member of the RL11 glycoprotein family (Davison *et al.*, 2003b) and UL111A, which encodes an interleukin-10 homologue (Kotenko *et al.*, 2000).

Sequencing of a low passage HCMV strain, Merlin, by the same group has further clarified the genetic content of wild-type HCMV (Dolan *et al.*, 2004). The genome of Merlin is 235,645 bp, approximately 5 kbp larger than AD169, and contains the extra DNA segment identified in Toledo without the inversion (Cha *et al.*, 1996), and no internal repeat of ORFs RL1-RL13. Comparative analysis with AD169 and Toledo identified a total of 165 genes, with Merlin containing no obvious mutations except a single nucleotide substitution that truncates UL128. The map of Merlin, shown in Figure 1.9, differs a great deal from the AD169 genome described by Chee *et al.* (1990) (Figure 1.8), and presents a more conservative view of wild-type HCMV than the contemporaneous genome re-analyses (Murphy *et al.*, 2003a, 2003b). The comparative analysis with CCMV contributes strongly to the Merlin genome map, containing the extra ORFs identified in the 19 kbp absent from AD169, retaining the 10 novel ORFs with counterparts in CCMV, and discounting a large number of ORFs (Figure 1.10). Other new ORFs that have been well characterised since the original HCMV analysis (those discussed in Section 1.17), are also included. This map is thought to represent the most reliable account of the gene complement of wild-type HCMV, with a best estimate of 165 genes (Dolan *et al.*, 2004).

1.19 HCMV strain variation

The most extensively used strains are also the most highly passaged strains. Davis was isolated via liver biopsy from a 3 month old girl with microencephaly (Weller *et al.*, 1957). Towne, a vaccine candidate, was isolated from the urine of a congenital infection (Plotkin *et al.*, 1975). AD169, isolated via tonsillectomy-adenoidectomy from a 7-year-old girl (Rowe *et al.*, 1956), has become the most extensively used strain because it replicates so efficiently in cultured cells, and one form

(AD169*var*UK) was used by Chee et al. (1990) in the original HCMV genome analysis. Low passage strains such as Toledo were subsequently used in HCMV analyses (Quinnan *et al.*, 1984). Clinical isolates are not inclined to yield much extracellular virus unless they have been adapted to cell culture growth by serial passage, and can take much longer to achieve CPE than highly passaged laboratory strains. As discussed in sections 1.17 and 1.14, highly passaged strains have large scale deletions and gene disruptions or rearrangements associated with extensive passaging in cell culture.

In addition to large scale mutations that occur *in vitro*, many natural variations have been observed between strains. In a study of the genetic content of wild-type HCMV, a subset of genes showed unusually high variation between strains Merlin and AD169 or Toledo (Dolan *et al.*, 2004). Ka values, a measure of nonsynonymous nucleotide substitution which results directly in amino acid sequence divergence, were computed for pairwise alignments of coding sequences. This approach identified a subset of genes displaying unusually high variation, comprising genes which are either known or predicted to encode secreted or membrane-associated proteins. All 25 genes with Ka values of 0.03 or more are in this category (Figure 1.10). For example, analysis of the RL12 proteins encoded by the AD169 and Merlin genomes show only 49% identity in their amino acid sequences.

As previously discussed, envelope glycoproteins form three complexes: gB, gM/gN and gH/gL/gO (Section 1.9). Each of UL55 (gB), UL73 (gN), UL75 (gH), UL115 (gL) and UL74 (gO) is highly variable between HCMV strains. Sequence analysis has shown that each of these glycoproteins has four genotypes, with the exception of gH, which has two and gO, which has seven (Pignatelli *et al.*, 2004). Genetic polymorphism has also been observed in other viral glycoproteins. US9 and UL4 have four and three distinct groups, respectively (Rasmussen *et al.*, 2003; Bar *et al.*, 2001). UL11, a transmembrane glycoprotein, and UL37, which encodes two anti-apoptotic proteins, both have highly variable extracellular domains (Hitomi *et al.*, 1997; Hayajneh *et al.*, 2001).

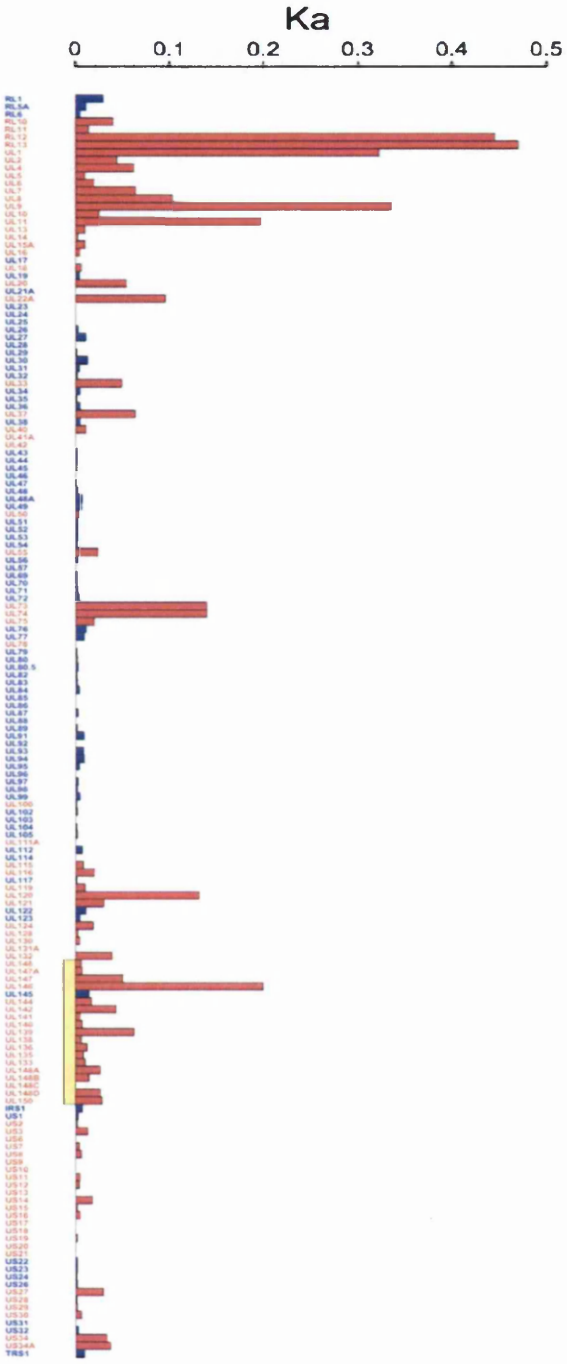


Figure 1.10 Nonsynonymous divergence.

Nonsynonymous divergence (Ka) between Merlin and AD169 genes (Dolan *et al.* , 2004). The gene order proceeds from the left end of the genome (top of the figure). Genes predicted to encode secreted or membrane-associated proteins are shown in red, and all other genes are shown in blue. The yellow bar indicates where Merlin was compared with Toledo as AD169 lacks these genes. Comparisons of all other genes were carried out between Merlin and AD169 genes.

UL146, which encodes a CXC-chemokine, is particularly variable. It has been sequenced in 40 strains, identifying 14 genotypes (Penfold *et al.*, 1999; Prichard *et al.*, 2001; Dolan *et al.*, 2004). US28, which encodes a β-chemokine receptor, has been sequenced in 25 strains to identify four genotypes (Rasmussen *et al.*, 2003). UL144 encodes a homologue of the tumour necrosis receptor superfamily (Arav-Boger *et al.*,

2002; Bale *et al.*, 2001; Lurain *et al.*, 1999). In a study of 45 low-passage clinical isolates, an amino acid substitution rate of approximately 21% was found, most of the changes occurring in the 5' half of the gene (Lurain *et al.*, 1999). The differences identified three genotpyes groups, with one group containing three subgroups.

The genetic polymorphism seen in numerous HCMV genes could have many different implications. For example, the polymorphism evident in the envelope glycoproteins could affect viral cell tropism. In particular, gB has important roles in virus penetration into the cell and in cell-to-cell fusion, and these functions could be affected. Other highly variable genes are known to have roles in immune evasion. UL146, US28 and UL144 could have developed genetic polymorphisms to help evade host immune surveillance (Pignatelli *et al.*, 2004). Polymorphism also potentially increases the variety of ways in which HCMV is presented as an antigen, thus potentially helping to evade the host immune defence. There is also potential for polymorphism to affect viral replication efficiency and regulation, as some variation has also been observed in UL54, UL122 and UL123 (Zweygberg Wirgart *et al.*, 1998; Sorg and Stamminger, 1998).

1.20 HCMV relationship to other herpesviruses

Comparing sequences to a close relative is one of the most significant ways of improving the interpretation of genetic content, on the basis that authentic protein-coding regions will have been conserved during evolution whereas spurious ORFs will not. The genetic content of CCMV, the closest known relative of HCMV, was reported by Davison *et al.* (2003a). The 241,087 bp genome was shown to be moderately diverged and essentially colinear with HCMV, with a similar genomic structure. CCMV is thought to contain only four genes that are not present in HCMV; UL146A and UL157, which are related to UL146 in HCMV (Penfold *et al.*, 1999), UL155 and UL156. The comparative analysis used to modify the interpretation of the HCMV genome is discussed in Section 1.18.

Host Species	Name	Reference	Accession no.	Genome size	Estimated gene no.
Human	HCMV	Chee *et al.* (1990)	X17403	229	208
		Dolan *et al.* (2004)	AY446894	236	165
Chimpanzee	CCMV	Davison *et al.* (2003a)	AF480884	241	165
Rhesus	RhCMV	Hansen *et al.* (2003)	AY186194	221	230
Murine	MCMV	Rawlinson *et al.* (1996)	U68299	230	170
Rat	RCMV	Vink *et al.* (2000)	AF232689	230	166
Tupaia	TuHV-1	Bahr and Darai (2001)	AF281817	196	158
HHV-6	HHV-6	Gompels *et al.* (1995)	X683413	159	119
		Dominguez *et al.* (1999)	AF157706	162	119
		Isegawa *et al.* (1999)	AB021506	162	115
HHV-7	HHV-7	Nicholas (1996)	U43400	144	109
		Megaw *et al.* (1998)	AF037218	153	84

Table 1.5. Sequenced *Betaherpesvirus* genomes.
Approximate genome sizes are in kbp.

The 221,469 bp genome of RhCMV strain 68-1 has been reported (Hansen *et al.*, 2003). This sequence provides another important tool for comparative analysis of HCMV as it is evolutionarily closer than most other β-herpesviruses. RhCMV appears to contain a single unique region with no large internal or terminal repeats like those in HCMV, but a virion packaging site (homologous to the *a* sequence in HCMV) separates the counterparts of the U_L and U_S regions of HCMV. Of the 230 ORFs identified, 138 are homologous to known HCMV proteins, including almost all the HCMV gene families and all the *trans*-activating factors required for HCMV lytic replication.

Several other β-herpesviruses more distantly related to HCMV have been completely sequenced (Table 1.5). The genomes of MCMV and RCMV are similar in size (230,278 bp and 229,896 bp, respectively) and both contain a core set of genes that have counterparts to those in HCMV. Analysis of the Smith strain of MCMV shows significant similarities to AD169, particularly in 78 centrally located ORFs, including homologues of the UL25, UL82, US22 and GCR gene families (Rawlinson *et al.*,

1996). The Maastricht strain of RCMV contains counterparts to all but one of the ORFs conserved between MCMV and HCMV (Vink *et al.*, 2000).

MCMV is the most studied CMV with the exception of HCMV. Analyses of the MCMV genome subsequent to that of Rawlinson *et al.* (1996) mirrors that of HCMV, as errors have been identified in the original sequence. Reannotation of the MCMV and RCMV genomes by Brocchieri *et al.* (2005) used a set of criteria that was significantly looser than that used in other genome analyses, but similar to that employed by the analyses of Murphy *et al.* (2003a, 2003b) for HCMV. Consequently, over 100 new ORFs were assigned to each genome by Brocchieri *et al.* (2005) but only some (14 in MCMV and 20 in RCMV) were considered convincing as protein-coding genes.

The genome of TuHV-1 (195,857 bp) was sequenced by Bahr and Darai (2001). Reannotation of the genome by the same authors (Bahr and Darai, 2004) discounted 23 of the original 158 protein-coding ORFs, leaving 76 centrally located ORFs showing significant homology to HCMV, and species-specific ORFs at the ends of the genome.

The U1102 strain of HHV-6 (159,321 bp) is predicted to encode 86 protein-coding ORFs, and lacks the RL11, US1, US2, US6 and US12 gene families (Megaw *et al.*, 1998; Zou *et al.*, 1999). HHV-7 (153,080 bp) is the closest relative of HHV-6, and has a total of 84 genes (Nicholas *et al.* 1996; Megaw *et al.*, 1998).

1.21 Justification of study

Several analyses have led to reinterpretation of the HCMV genome. Of particular significance, comparative analysis with CCMV has discounted a substantial number of genes in the original analysis of Chee *et al.* (1990). This led to the appearance of three large regions in HCMV which apparently do not encode proteins; UL106-UL111, UL58-UL68, and RL2-RL9. The RL2-RL9 region has been investigated by Davison *et al.* (2003b), but the protein-coding potential of the other two regions require further evaluation. The region flanked by UL105 and UL112 has been termed

Region X, and the region flanked by UL57 and UL69 has been termed **Region O**. A 5 kb RNA in region X has been previously reported and requires further characterisation. Region O, which encompasses the HCMV origin of lytic DNA replication, also harbours an RNA (pp67) previously reported to encode a protein. A third region investigated in the current study, which also contains ORFs not thought to encode proteins, is flanked by US32 and TRS1 and has been termed **Region G**. The protein-coding potential of ORFs predicted from the original work of Chee *et al.* (1990) and subsequent analyses (Murphy *et al.*, 2003a, 2003b; Davison *et al.*, 2003a; Dolan *et al.*, 2004) are investigated in the current study by sequence comparisons of multiple HCMV strains (laboratory strains and clinical isolates) and transcript mapping.

CHAPTER 2.1: MATERIALS

2.1.1 Viruses

HCMV strain AD169 (Chee *et al.*, 1990) was provided by Dr. Derrick Dargan (Institute of Virology, Glasgow) and used for all low-multiplicity and high-multiplicity infections in cell culture. Infected cell DNA from strains AD169 and Toledo were also obtained from Dr. Derrick Dargan. Low-passage strains 3157, 6397 and Merlin were isolated from the urine of congenitally infected children, and DNA samples were obtained from Dr. Gavin Wilkinson (University of Cardiff). Strain 3301, isolated directly from urine without passage, was also provided by Dr. Gavin Wilkinson. DNA from strain W was isolated, without passage, from the adrenal gland of an HCMV-infected AIDS patient, and was provided by Dr. Aycan Hassan-Walker and Prof. Paul Griffiths (University College Medical School).

2.1.2 PCR

Advantage 2 Polymerase Mix (50×): Advan 2 PCR buffer dNTPs (10 mM each)	BD Clontech
Primers	Sigma Genosys
PCR Optimiser kit: 16 PCR buffers 2.5 mM dNTP mix DMSO Control primers & template PCR-grade H_2O	Invitrogen
Advantage-GC 2 PCR kit: 50 × Advantage-GC 2 Polymerase Mix 5 × Advantage-GC 2 PCR buffer GC-Melt (5 M) dNTPs (10 mM each) Control primers & template PCR-grade H_2O	BD Clontech

2.1.3 Agarose gel electrophoresis

Agarose	Sigma
10 × TBE	109 g/l Trizma 55 g/l boric acid 9.3 g/l EDTA
DF dyes	37.2 g/l EDTA 100 g/l Ficoll 400 5 × TBE 1% (w/v) bromophenol blue
DNA marker (2 log ladder)	New England BioLabs
Ethidium bromide (10 mg/ml aqueous solution)	Sigma
Low melting point (SeaPlaque) agarose	FMC Bioproducts
Geneclean spin kit: Glassmilk matrix Wash concentrate Elution solution Catch tubes and spin filters	Qbiogene
Turbo spin kit: Turbo salt solution Wash concentrate Elution solution Catch tubes and spin filters	Qbiogene

2.1.4 Plasmid preparation

One Shot TOP10 *Escherichia coli* strain K12 competent cells (Invitrogen). Genotype: F^- *mcr*A Δ(*mrr-hd*RMS-*mcr*BC) ϕ80*lac*ZMΔ15 Δ*lac*X74 *deo*R *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G.

pGEM-T Vector System I: pGEM-T vector T4 DNA ligase 2 × ligation buffer Control insert DNA	Promega

L-broth	177 mM NaCl 1% (w/v) bactopeptone autoclaved
L-broth agar	1.5% (w/v) agar in L-broth autoclaved
SOC medium	Invitrogen
X-gal	40 mg/ml 5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside in N, N'-dimethyl formamide
IPTG	30 mg/ml isopropylthio-β-D-galactoside
2YT broth	85 mM NaCl 1% (w/v) bactopeptone 1% (w/v) yeast extract autoclaved
GTE	50 mM D-glucose 25 mM Tris-HCl, pH 8 10 mM EDTA
NaOH/SDS	200 mM NaOH 1% (w/v) SDS
Potassium acetate solution	3 M potassium acetate 2 M acetic acid
PEG/NaCl	20% (w/v) PEG 6000 2.5 M NaCl
Phenol	phenol equilibrated with TE
Restriction endonucleases and buffers	New England BioLabs

2.1.5 Sequencing

Ethanol mix	75% ethanol 40 µl/ml 3M sodium acetate pH 4.6
Page-plus 40% concentrate	Amresco
Acrylamide	0.48 g/ml urea 12% (v/v) Page-plus 10% (v/v) 10 × TBE

Blue dextran	4 mg/ml blue dextran EDTA
Loading buffer	70% (v/v) formamide 30% (v/v) blue dextran

2.1.6 Cells and media

Human foetal foreskin fibroblast (HFFF-2) cells were used in all cell culture experiments. Cells were initially supplied by Dr. Derrick Dargan (Institute of Virology). HFFF-2 cells were grown in Dulbecco's Modified Eagle's Medium with 10% (v/v) foetal calf serum (FCS), 1% non-essential amino acids and 1% L-glutamine supplements. On occasion, 1% penicillin-streptomycin was also used as a supplement.

Versene (0.6 mM EDTA in PBS, 0.002% (w/v) phenol red) and trypsin (0.25% (w/v) in Tris-saline) were used for washing and removing cell monolayers.

Storage medium	90% (v/v) foetal calf serum (FCS) 10% (v/v) dimethylsulfoxide (DMSO)
Methylcellulose overlay	39% (v/v) carboxymethylcellulose 1.56% (v/v) tryptose phosphate 3.9% (v/v) FCS 24.3% (v/v) $NaHCO_3$ (7.5%) 1.56% (v/v) penicillin/stremptomycin 0.08% (v/v) amphotericin B 7% (v/v) 10×Glasgow modified medium
Giemsa stain	Sigma
Phosphate-buffered saline (PBS)	170 mM NaCl 3.4 mM KCl 10 mM Na_2HPO_4 1.8 mM KH_2PO_4 6.8 mM $CaCl_2$ 4.9 mM $MgCl_2$

2.1.7 RNA preparation and extraction

Cycloheximide	Sigma
Phosphonoacetic acid (PAA)	Sigma
TRIZOL reagent	Gibco Life Technologies
Oligo (dT)-cellulose	Sigma
Spin-X columns	Costar
FastTrack 2.0 mRNA isolation kit: Stock Buffer Protein/RNase degrader Binding buffer Low salt wash buffer Elution buffer 2 M Sodium Acetate 5 M NaCl Glycogen carrier Oligo(dT) cellulose Spin columns	Invitrogen

2.1.8 Northern blotting

10 × MOPS	Sigma
Formaldehyde	Sigma
Formamide	Sigma
Electrophoresis buffer	1 × MOPS
Loading buffer	4 mg/ml bromophenol blue 4 mg/ml xylene cyanol 0.25 g/ml Ficoll 400
20 × SSC	88.2 g/l tri-sodium citrate 174 g/l NaCl
Nytran SuperCharge membrane	Schleicher & Schuell
3MM Paper	Whatman

NonaPrimer kit: NonaPrimer mix dATP dCTP dGTP dTTP Klenow DNA polymerase Adsorb solution DNAprep resin Wash buffer Elute solution	Qbiogene
Lig'nScribe kit: SP6 promoter adaptor T7 promoter adaptor 10 × T4 DNA ligase 10 × ligation buffer PCR adaptor primers 1 and 2 Control template and primer Nuclease-free H_2O	Ambion
MaxiScript SP6/T7 kit: SP6 enzyme mix T7 enzyme mix 10 × transcription buffer ATP solution (10 mM) CTP solution (10 mM) GTP solution (10 mM) UTP solution (10 mM) DNase I Nuclease-free H_2O	Ambion
Rapid hyb buffer	Amersham
Gene-specific primers	Sigma Genosys
$[\alpha^{32}P]$CTP, $[\alpha^{32}P]$GTP, $[\alpha^{32}P]$UTP (3000 Ci/mmol, 10 μCi/ml)	NEN Life Science

2.1.9 RT-PCR and RACE-PCR

RT-PCR Superscript III One-Step kit: Superscript III RT/Taq enzyme mix 2 × reaction mix 5 mM $MgSO_4$	Invitrogen

GeneRacer amplification kit: Invitrogen
- RNaseOut
- Calf intestinal phosphatase (CIP)
- 10 x CIP buffer
- Tobacco acid pyrophosphatase (TAP)
- 10 × TAP buffer
- GeneRacer RNA Oligo
- T4 RNA ligase
- 10 × T4 RNA ligase buffer
- 10 mM ATP
- Phenol/chloroform
- Mussel glycogen
- 3 M Sodium acetate
- GeneRacer 5' primer and nested primer
- GeneRacer 3' primer and nested primer
- Control total RNA and primers

SMARTRACE cDNA amplification kit: BD Clontech
- SMART II oligonucleotide
- 3' RACE CDS primer
- 5' RACE CDS primer
- Powerscript reverse transcriptase
- 5 × first-strand buffer
- DTT
- Deionised H_2O
- 10 × universal primer mix
- Nested universal primer mix
- dNTP mix (10 mM each)
- Tricine-EDTA buffer
- Control total RNA and primers

5'-RACE amplification kit: Invitrogen
- 10 x PCR buffer
- 25 mM $MgCl_2$
- 10 mM dNTP mix
- 0.1 M DTT
- Superscript II Reverse Transcriptase
- RNase mix
- 5 x tailing buffer
- 2 mM dCTP
- Terminal deoxynucleotidyl transferase
- 5'-RACE abridged anchor primer (10 μM)

2.1.10 Chemicals

All general chemicals were obtained from BDH or Sigma.

2.1.11 Autoclaving and glassware sterilisation

Equipment and solutions were sterilised at 15 psi for 20 min by staff in the wash room of the Institute of Virology. Glassware was sterilised by baking in an oven at 180°C for at least 2 h.

CHAPTER 2.2: METHODS

2.2.1 PCR

PCR was performed in 50 µl reactions:

PCR buffer (10 ×)	5 µl
dNTPs (10 mM each)	1 µl
Forward primer (10 µM)	1 µl
Reverse primer (10 µM)	1 µl
Template (genomic DNA)	1 µl
H_2O	40 µl
Advantage II Polymerase	1 µl

Reactions were placed in a thermocycler and denatured at 94°C for 30 s. A standard PCR program was 35 cycles of 94°C/5 s (i.e. 94°C for 5 s), 68°C/30 s and 72°C/1 min per kb, followed by a final extension step of 72°C/7 min. Conditions were as stringent as possible to promote specificity.

2.2.2 PCR Optimiser

Optimising PCR was performed in 45 µl reactions using an Invitrogen kit:

PCR buffer (5×)	10 µl
dNTPs (10 mM each)	1 µl
Forward primer (10 µM)	1 µl
Reverse primer (10 µM)	1 µl
Template (genomic DNA)	1 µl
H_2O	30 µl
Advantage II Polymerase	1 µl

Reactions were placed at 80°C and 5 µl of 2.5 mM dNTP mix was added (to make a total reaction volume of 50 µl) before proceeding directly to the thermocycler. For each set of primers, seven different PCR reactions were prepared using different buffers, each with a specific pH (ranging from 8.5 to 10) and Mg^{2+} concentration (ranging from 1.5 to 3.5 mM). A standard PCR program was 94°C/2 min, 35 cycles of 94°C/1 min, 65°C/1 min and 72°C/1 min per kb, followed by a final extension step of 72°C/7 min. A series of optimiser PCRs is shown in Figure 2.1(a).

2.2.3 High-GC PCR

High-GC PCR was performed in series of 50 µl reactions using a BD Clontech kit:

	final GC-Melt concentration (M)			
	0	0.5	1	1.5
PCR buffer (5×)	10	10	10	10 µl
dNTPs (10 mM each)	1	1	1	1 µl
Forward primer (10 µM)	1	1	1	1 µl
Reverse primer (10 µM)	1	1	1	1 µl
Template (genomic DNA)	1	1	1	1 µl
GC-Melt	0	5	10	15 µl
H_2O	35	30	25	20 µl
Advantage-GC2 Polymerase	1	1	1	1 µl

GC-Melt is a reagent, supplied in the kit, which further weakens base pairing in GC-rich sequences. Reactions were placed in a thermocycler and denatured at 94°C for 3 min. A standard PCR program was 30 cycles of 94°C/30 s, 65°C/30 s and 68°C/1 min per kb, followed by a final extension step of 68°C/3 min. A series of high-GC PCRs is shown in Figure 2.1(B).

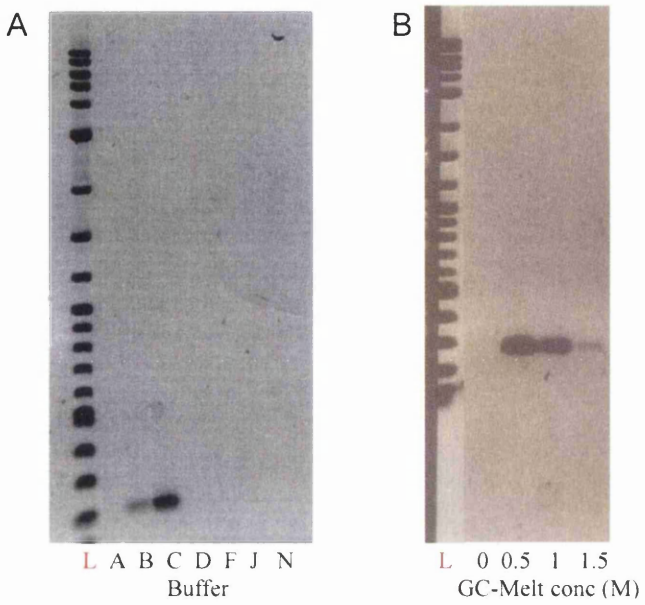


Figure 2.1: Optimiser PCR and High-GC PCR

(A) A series of optimiser PCR reactions using seven different PCR buffers. In this series, primers worked optimally with buffer C. **(B)** A series of high-GC PCR reactions using different concentrations of GC-Melt reagent. In this series, primers work optimally with a GC-Melt concentration of 0.5 M. Lanes L are DNA markers.

2.2.4 Agarose gel electrophoresis

Unless otherwise stated, all agarose gels were at 1% (w/v) concentration. 1 g of agarose was melted in 100 ml of 1 x TBE and allowed to set at 4°C in a gel former. Gels were electrophoresed for approximately 2 h at 120 V, and then stained with 0.5 µg/ml ethidium bromide for 30 min. Gels were visualised under UV light using the GelDoc system (BioRad).

2.2.5 Gel purification: low melting temperature agarose

1.5% (w/v) low-melting-temperature agarose gels electrophoresed overnight (approximately 16 h) at 30 V were stained with 0.5 μg/ml ethidium bromide for 1 h. Appropriate bands were cut out on a long-wave UV lightbox using a sterile scalpel, and placed in 1.5 ml microfuge tubes, each containing 20 μl of 10 × β-agarase buffer. Samples were heated at 65°C for 1 h, and then at 40°C for 5 min, 3 μl of β-agarase was added, and the tubes were left at 40°C for another 1 h. The samples were supplemented with 10 μl of 3M NaAc, incubated on ice for 15 min, and centrifuged at 12,000 × g for 10 min to pellet any remaining agarose. The supernatants were transferred to fresh microfuge tubes, 200 μl of ethanol was added, and the tubes were incubated at -20°C for 1 h. The samples were centrifuged at 12,000 × g for 10 min and the supernatants removed. The pellets were centrifuged in 100 μl of 70% ethanol, and the supernatants removed. The samples were dried at 37°C for 1 h, and the pellets were resuspended in 10 μl of sterile H_2O and stored at -20°C. Sample recovery was checked by electrophoresis of 2 μl on a 1% (w/v) agarose gel.

2.2.6 Gel purification: GENECLEAN spin kit

1% (w/v) agarose gels electrophoresed for approximately 2 h at 120 V were stained with 0.5 μg/ml ethidium bromide for 30 min. Appropriate bands were cut out on a long-wave UV lightbox using a sterile scalpel, and transferred to 1.5 ml microfuge tubes. Gel samples were transferred to Spin Filter Tubes containing 400 μl of Glassmilk matrix solution, which is able to bind DNA. Samples were heated at 55°C for 10 min, and inverted every minute to prevent the matrix settling.

Samples were centrifuged at 12,000 × g for 1 min, and the flow-through discarded. Spin new wash was added (500 μl), samples were centrifuged at 12,000 × g for 1 min, and the flow-through was discarded. A further centrifugation step of 12,000 × g for 2 min was carried out to dry the membrane. The filters were transferred to Elution Catch Tubes, and the Glassmilk resuspended in 20 μl of Spin Elution Solution

(DNase-free H_2O). The samples were centrifuged at 12,000 × g for 1 min, the filter was discarded, and the eluted DNA was stored at -20°C. Sample recovery was checked by the electrophoresis of 2 µl on a 1% (w/v) agarose gel.

2.2.7 Gel purification: GENECLEAN Turbo kit

1% (w/v) agarose gels electrophoresed for approximately 2 h at 120 V were stained with 0.5 µg/ml ethidium bromide for 30 min. Appropriate bands were cut out on a long-wave UV lightbox using a sterile scalpel, and transferred to 1.5 ml microfuge tubes. Turbo salt solution (150 µl) was added to each sample (more for bigger gel pieces) and incubated at 55°C for 10 min, inverting the samples every minute. Up to 600 µl of each sample was then transferred to a Turbo cartridge in a catch tube.

The samples were centrifuged at 12,000 × g for 10 s and the flow-through discarded. Turbo Wash was added (500 µl), the samples were centrifuged at 12,000 × g for 10 s and the flow-through was discarded. A further centrifugation step of 12,000 × g for 4 min was carried out to dry the membrane. The Turbo cartridges were transferred to new catch tubes and the Glassmilk was resuspended in 20 µl of elution solution (DNase-free H_2O). The samples were centrifuged at 12,000 × g for 1 min, the Turbo cartridge was discarded and the eluted DNA was stored at -20°C. Sample recovery was checked by the electrophoresis of 2 µl on a 1% (w/v) agarose gel.

2.2.8 Ligations

Ligations were performed in 5 µl reactions using a Promega kit:

T4 DNA ligase buffer (2×)	2.5 µl
pGEM-T vector	0.25 µl
DNA insert	1.5 µl
H_2O	0.25 µl
T4 DNA ligase	0.5 µl

Each set of ligation reactions included a positive control (a DNA sample supplied in the kit) and a negative control (no DNA insert). H_2O was added to the controls to make a total volume of 5 µl. Reactions were incubated overnight at 4°C or at room temperature for 1 h.

2.2.9 Transformations

Competent cells were moved directly from -70°C to ice and allowed to thaw for 15 min. Cells were divided into aliquots of 25 µl, and 1 µl ligation reaction was added, taking care not to damage the cells by pipetting. The samples were kept on ice for 30 min, heat-shocked at 42°C for 30 s, then moved back to ice for 5 min. SOC medium (250 µl) was added to each sample, which was then incubated in a 37°C incubation shaking cabinet for 1 h. IPTG (10 µl) and X-gal (20 µl) were added to each tube, the sample mixed and the solution spread using a sterilised Pasteur pipette on an ampicillin plate pre-incubated at 37°C. The plates were allowed to stand at room temperature for 1 h, then incubated upside-down overnight at 37°C.

2.2.10 Plasmid purification: alkaline lysis

Bacterial colonies grown on ampicillin plates were picked using sterile cocktail sticks and grown overnight at 37°C in 2YT broth containing ampicillin. Each culture (1 ml) was transferred to a separate microfuge tube and centrifuged at 12,000 × g for 5 min. The supernatant was removed using a Pasteur pipette, and the pellet was resuspended in 100 µl GTE/lysosyme and left at room temperature for 5 min. NaOH/SDS (200 µl) was added to the vortexed mixture, and the tubes were inverted and left at 4°C for 5 min. 3 M KAc pH5 (150 µl) was added, and the tubes were inverted immediately and left at 4°C for 5 min. The tubes were centrifuged at 12,000 × g for 10 min, and the supernatants transferred to fresh microfuge tubes. One volume of phenol/chloroform was added to each tube, which was mixed by inversion for 1 min then centrifuged at 12,000 × g for 5 min. The upper phase was transferred to a fresh tube and 750 µl

100% ethanol added. The samples were allowed to stand at room temperature for 30 min, and then the tubes were centrifuged at 12,000 × g for 10 min. The supernatant was removed and the tubes were allowed to dry at 37°C for 20 min. The samples were resuspended in 50 µl TE and 100 µg/ml RNaseA, and incubated at 37°C for 30 min. The tubes were inverted gently after addition of 60 µl PEG/NaCl, and allowed to stand at 4°C for 30 min. The samples were centrifuged at 12,000 × g for 7 min, and the supernatants removed. The DNA pellets were washed in 100 µl 70% ethanol. The samples were allowed to stand at 37°C for 30 min, and resuspended in 50 µl H_2O and stored at -20°C.

2.2.11 Plasmid purification: QIAspin miniprep

Plasmid purification was also carried out using a Qiagen miniprep kit. Bacterial colonies grown on ampicillin plates were picked using sterile cocktail sticks and grown overnight at 37°C in 2YT broth containing ampicillin. Each culture (1 ml) was transferred to a separate microfuge tube and centrifuged at 12,000 × g for 5 min. The supernatant was removed using a Pasteur pipette, and the pellet was resuspended in 250 µl buffer P1 (containing RNaseA) and left at room temperature for 5 min. Lysis buffer P2 (250 µl) was added to the vortexed samples, which were mixed by inversion and allowed to stand no longer than 5 min. Neutralisation buffer N3 (350 µl) was added to the samples, which were immediately mixed by inversion. The samples were centrifuged at 12,000 × g for 10 min, and the supernatants were transferred to spin columns. The samples were centrifuged at 12,000 × g for 1 min and the flow-through discarded. Wash buffer PE (750 µl) was added to each sample. The tubes were centrifuged at 12,000 × g for 1 min and the flow-through discarded. The empty columns were then centrifuged at 12,000 × g for 1 min, and the columns transferred to fresh microfuge tubes. H_2O (60 µl) was added to each sample in the centre of the column and allowed to stand at room temperature for 2 min. The tubes were centrifuged at 12,000 × g for 1 min, the spin column discarded, and the recovered plasmid solution stored at -20°C.

2.2.12 Restriction endonuclease digests

Each restriction endonuclease digest contained 0.4 µl of restriction enzyme, 2 µl of the appropriate buffer, made up to 20 µl with H_2O. The reactions were incubated at 37°C for 2 h, and the digest was checked by electrophoresis of 2 µl on a 1% (w/v) agarose gel.

2.2.13 Sequencing: Beckman 2000XL

Reactions were prepared in 96-well plastic plates. Each reaction contained 2 µl DTCS mix, 2 µl DNA and 1 µl primer (1 µM). Reactions were placed in a thermocycler. The PCR program used consisted of 30 cycles of 96°C/20 s, 50°C/20 s and 60°C/4 min. After thermocycling, the reactions were ethanol-precipitated by adding 62 µl ethanol mix to each well, then centrifuged immediately at 4,000 × g for 30 min at 4°C. The plates were centrifuged briefly upside down, and the reactions were washed with 70% (v/v) ethanol by centrifuging at 4,000 × g for 15 min at 4°C. The plates were again centrifuged briefly upside down to dispel all ethanol and the pellets allowed to dry at room temperature. Sample (formamide-based) loading solution (20 µl) was added to each well, followed by a mineral oil overlay (500 µl), before the plate was loaded into the Beckman sequencer.

2.2.14 Sequencing: ABI 377

Reactions were performed in 96-well plastic plates. Each reaction contained:

10 × buffer	1 µl
BigDyes	2 µl
DNA	4 µl
Primer (1 µM)	2 µl
H_2O	1 µl

The reactions were placed in a thermocycler. The PCR program consisted of a denaturation step of 94°C/2 min, then 30 cycles of 94°C/10 s, 50°C/5 s and 60°C/4 min. After thermocycling, the reactions were ethanol-precipitated by adding 62 µl ethanol mix to each well, then immediately centrifuged at 4,000 × g for 30 min at 4°C. The plates were centrifuged briefly upside down, and the reactions were washed with 70% (v/v) ethanol by centrifugation at 4,000 × g for 15 min at 4°C. The plates were again centrifuged briefly upside down to dispel all ethanol, and the pellets allowed to dry at room temperature.

Polyacrylamide gels were made using 50 ml acrylamide, 300 µl APS and 30 µl Temed, and allowed to set with a 96-well comb. The gels were pre-electrophoresed for 1 h before sample loading, using 1 × TBE as a buffer. A 3:1 mixture of formamide and blue dextran was made, and 3 µl added to each reaction. The plates were heated on a thermocycler at 96°C for 5 min, and then the samples were loaded into the wells and the gels electrophoresed for 7 h at 4800 V.

2.2.15 Cell culture

Monolayers of human foetal foreskin fibroblast (HFFF-2) cells were grown in plastic flasks in a solution of Dulbecco's modified eagle's medium (DMEM) and 10% (v/v) foetal calf serum (FCS). Cells were grown at 37°C in a humidified atmosphere of 95% (v/v) air and 5% (v/v) CO_2. Cells were harvested from roller bottles by pouring off the medium, washing the monolayer twice with 20 ml versene, then washing with 20 ml trypsin/versene to detach the cells. The cells were then resuspended in fresh medium and aliquots of approximately 10^7 cells were seeded into roller bottles containing 100 ml of medium. Aliquots of cells were also prepared for storage by harvesting, pelleting at 1000 x g and resuspending in storage medium. These aliquots were frozen to -70°C overnight then stored in liquid nitrogen. On recovery, the aliquots were thawed quickly at 37°C then resuspended in growth medium in a medium flask.

2.2.16 Virus Stocks

Virus stocks were prepared from HFFF-2 cells grown in roller bottles containing 100 ml medium until 90% confluent. The medium was replaced with 30 ml fresh medium containing approximately 0.01 p.f.u./cell HCMV strain AD169. The roller bottles were returned to 37°C, and the infected cells incubated until maximal CPE was observed (usually 10-15 days p.i.). The medium containing infected cells was decanted and cells pelleted by centrifugation for 15 min at 1000 x g at 4°C. The supernatent was collected and stored at -70°C, and virus stocks were then titrated.

2.2.17 Titrations

Plaque assay on HFFF-2 cells was used to determine viral titres. HCMV stocks were serially diluted ten-fold using 900 µl of 1% PBS(A) in foetal calf serum. Tissue culture dishes containing HFFF-2 monolayers were infected with 100 µl of each virus dilution. The monolayers were allowed to adsorb virus for 1 h at 37°C, with rocking every 15 min. The inocula were removed and the cells overlaid with 2 ml per dish of cell overlay. The monolayers were incubated at 37°C for a few days until the plaques could be seen easily. The medium was removed, 2 ml fixing solution was added, and then approximately 1 ml Giemsa stain was applied to each dish. After 2 h, the stain was poured off, and the tissue culture dishes were washed and allowed to dry. The plaques were then counted under a light microscope.

2.2.18 Preparing RNA

All infections were carried out in 80cm^2 tissue culture flasks with HFFF-2 cells at approximately 95% confluency. HCMV strain AD169 was used for all infections.

Preparing MI RNA

The medium was replaced with fresh medium, and the flask was incubated at 37°C for 1 h. The medium was then replaced with 10 ml of fresh medium and incubated at

37°C. After 96 h, the medium was removed and the cells washed three times, all the medium was removed, and the flask was stored at -70°C.

Preparing IE RNA

The medium was replaced with fresh medium containing 200 μg/ml cycloheximide, and the flask was incubated at 37°C for 1 h. The medium was then removed and high-titre virus (approximately 5 p.f.u./cell) containing 200 μg/ml cycloheximide was added. The flask was incubated at 37°C for 1 h with gentle shaking every 15 min. The cells were then washed three times with medium containing 200 μg/ml cycloheximide, left and incubated in 10 ml of this medium at 37°C. After 24 h, the medium was removed and the cells were washed three times with medium containing 200 μg/ml cycloheximide. All the medium was removed, and the flask was stored at -70°C.

Preparing E RNA

The medium was replaced with fresh medium containing 300 μg/ml phosphonoacetic acid (PAA), and the flask was incubated at 37°C for 1 h. The medium was then removed and high-titre virus (approximately 5 p.f.u./cell) was added. The flask was incubated at 37°C for 1 h with gentle shaking every 15 min. The cells were then washed three times with medium containing 300 μg/ml PAA, and incubated in 10 ml of this medium at 37°C. After 24 h, the medium was removed and the cells were washed three times with medium containing 300 μg/ml PAA. All the medium was removed, and the flask was stored at -70°C.

Preparing L RNA

The medium was replaced with fresh medium, and the flask was incubated at 37°C for 1 h. The medium was then removed and high-titre virus (approximately 5 p.f.u./cell) was added. The flask was incubated at 37°C for 1 h with gentle shaking every 15 min. The cells were then washed three times with medium, and incubated in 10 ml of this medium at 37°C. After 24 h, the medium was removed and the cells were washed three times with medium. All the medium was removed, and the flask was stored at -70°C.

2.2.19 Extracting total cellular RNA

Infected cells were lysed in the flasks by adding 1.2 ml of TRIZOL and mixing with a pipette. The solutions were transferred to 1.5 ml microfuge tubes, and 200 µl of phenol/chloroform was added. The tubes were shaken, left at room temperature for 3 min, and centrifuged at 12,000 × g for 15 min at 4°C. The upper phase containing the RNA was transferred to a fresh tube, a half-volume of propan-2-ol added, and the sample was left at room temperature for 10 min.

The samples were then centrifuged at 12,000 × g for 10 min at 4°C. The supernatant was removed, and 100 µl of 75% ethanol was added to the RNA pellets. The samples were centrifuged again at 12,000 × g for 10 min at 4°C, the supernatant removed, and the pellets dried at room temperature. When dry, the samples were resuspended in 20 µl TE, incubated at 55°C for 10 min, and stored at -70°C.

2.2.20 Extracting polyA$^+$ mRNA

For each RNA sample (IE, E, L and MI), 1 ml stock buffer and 20 µl Protein/RNase degrader was prepared and incubated at 45°C. The infected 80 cm^2 flasks were defrosted in the presence of this solution and the cells were collected using a cell scraper and passed through a 19 gauge needle using a 2 ml syringe. The samples were incubated at 45°C for 20 min, and centrifuged at 4,000 × g for 5 min. The supernatant was transferred to a fresh microfuge tube, and 63 µl 5M NaCl was added to each. The DNA was sheared by passing the sample through a 19 gauge needle three or four times. The cell lysate was then transferred to a tube of oligo(dT) cellulose, and after 2 min was rotated for 20 min at room temperature to allow RNA to bind to the oligo(dT). The samples were centrifuged at 4,000 × g for 5 min, and the supernatants were removed carefully. At this point, the unbound RNA in the supernatant could be recovered by ethanol precipitation.

The oligo(dT) cellulose was resuspended in 1.3 ml binding buffer, centrifuged at 4,000 × g for 5 min, and the supernatant removed carefully. This washing step was

repeated twice more. The oligo(dT) cellulose was resuspended in 300 µl washing buffer, and transferred to a spin column. The tubes were centrifuged at 4,000 × g for 5 min and the flow-through discarded. This washing step was repeated thrice more. The samples were gently resuspended in 200 µl low salt wash buffer, centrifuged at 4,000 × g for 5 min and the flow-through discarded. This washing step was repeated once more.

The spin columns were transferred to fresh microfuge tubes. Elution buffer (100 µl) was used to resuspend the oligo(dT) cellulose, and the samples were centrifuged at 4,000 × g for 5 min and the flow-through retained. This elution step was repeated once more. 2M NaAc (30 µl), 2 mg/ml glycogen carrier (10 µl) and 100% (v/v) ethanol (600 µl) were added to each sample, which were frozen in dry ice until solid. The samples were then thawed and centrifuged at 13,000 × g for 15 min at 4°C. The supernatant was removed, the tubes centrifuged again briefly, and the last traces of supernatant removed. The samples were allowed to dry for 10 min, resuspended in 15 µl of elution buffer and stored at -70°C.

2.2.21 Northern blot

A formaldehyde-agarose gel was made by mixing 1 g agarose and 85 ml H_2O, boiling, then cooling to 55°C. A preheated mixture of 5.8 ml formaldehyde and 10ml 10 × MOPS (at 55°C) was added to the agarose, mixed well and set in a gel former with comb at room temperature. RNA samples were prepared as 30 µl reactions:

formaldehyde	5.5 µl
formamide	15 µl
H_2O	6 µl
10 × MOPS	1.5 µl
RNA	2 µl

Gels also included two lanes containing a marker RNA, containing 5 µl of 0.24-9.5 kb RNA ladder (Invitrogen) and H_2O adjusted to the required volume. Samples were

incubated at 55°C for 15 min (the RNA marker at 65°C) and cooled on ice for 5 min. Sterile 10 × loading buffer (3 µl) was added. Samples were mixed and loaded onto the formaldehyde-agarose gel which had been electrophoresed for 15 min at 75 V. The gel was electrophoresed at 75 V for 4 h and stained with 0.05 µg/ml ethidium bromide for 45 min. The RNA ladder was visualised under short wave UV light with a ruler placed along one side of the gel. A picture was taken using the GelDoc system (BioRad) and used later for calculating transcript sizes.

RNA was transferred from the gel to a nylon membrane by capillary blotting. The gel was shaken gently at room temperature in H_2O (twice for 5 min) then 10 × SSC (twice for 15 min). Nylon membrane was cut to size and pre-soaked in H_2O (5 min) and then 20 × SSC (20 min). 3MM paper was also cut to size and pre-soaked in 2 × SSC for 20 min. The blot was set up by sitting the gel on top of a wick in a reservoir, then adding the nylon membrane, three sheets of pre-soaked 3MM paper, two sheets of dry 3MM paper and a stack of absorbant towels. A weight was placed on top of the towels, and the reservoir filled with 20 × SSC. After overnight transfer of the RNA, the membrane was rinsed gently in 10 × SSC, dried for 15 min, and crosslinked using a Stratagene UV Crosslinker (12,000 Jcm^{-2}).

2.2.22 DNA probes

DNA was radiolabelled using a Qbiogene NonaPrimer kit. A mixture of plasmid DNA (4 µl) and H_2O (3 µl) was denatured at 95°C for 10 min, chilled on ice for 5 min, and added to a 20 µl annealing reaction:

DNA/H_2O mix	7 µl
L2 NonaPrimer mix	4 µl
L3 dATP	1 µl
L6 dTTP	1 µl
L5 dGTP	1 µl
α^{32}P -dCTP	5 µl
L7 Klenow enzyme	1 µl

The reaction was incubated at 37°C for 30 min. Adsorb solution P1 (60 µl) and DNAprep resin P2 (2 µl) were added to the sample, mixed, centrifuged briefly, and the supernatant discarded. The sample was resuspended in 100 µl wash buffer P3, centrifuged briefly, the supernatant discarded, and the washing step repeated. Eluting solution P4 (100 µl) was added, and the sample vortexed then incubated at 65°C for five min. The supernatant containing the purified DNA probe was recovered and added directly to the hybridisation mix at 68°C for overnight hybridisation.

2.2.23 RNA Probes

The Lig'n Scribe reaction (Ambion) is a two-step process, as illustrated in Figure 2.2. The first 5 µl reaction step ligates a phage T7 or SP6 promoter to a PCR fragment:

10 × ligation buffer	1 µl
T7/SP6 adaptor	1 µl
PCR product	1 µl
T4 DNA ligase	1 µl
H_2O	1 µl

The reaction was incubated at room temperature for 15 min. The second step is a PCR amplification that uses the ligation reaction, a gene-specific primer (one of the original PCR primers) and a T7/SP6 primer (supplied in the kit). The 50 µl reactions were prepared:

10 × PCR buffer	5 µl
dNTPs (10 mM)	1 µl
GSP (10 µM)	1 µl
Adapter primer	1 µl
T7/SP6 ligation reaction	1 µl
Taq polymerase	1 µl
H_2O	40 µl

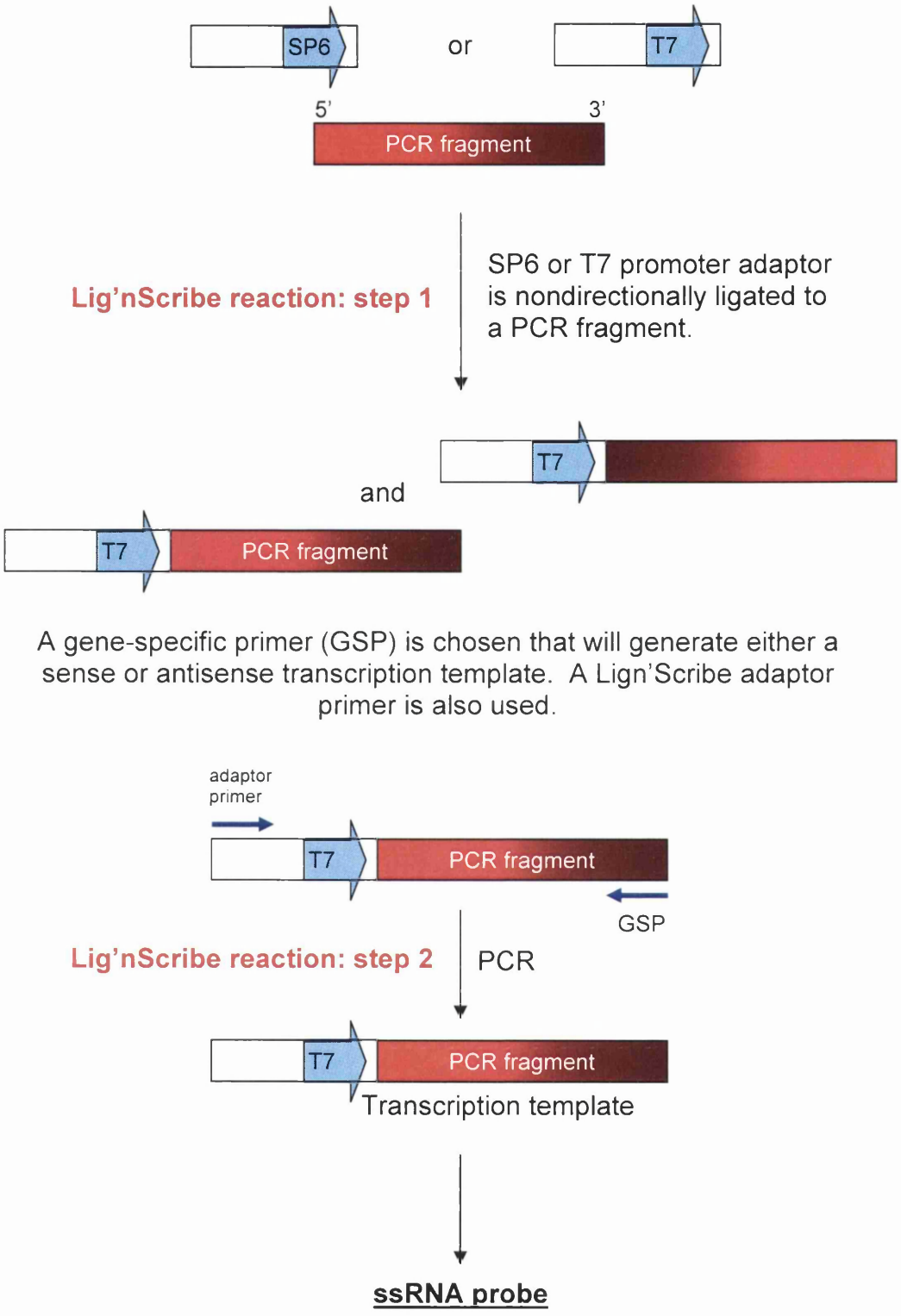


Figure 2.2. Lign'Scribe reaction.

The Lign'Scribe reaction involves a ligation step and a PCR step to prepare a transcription template that can be radiolabelled.

Purified Lig'n Scribe products were subsequently labelled radioactively in 20 µl reactions using the MaxiScript kit (Ambion):

10 × transcription buffer	2 µl
ATP (10mM)	1 µl
CTP (10mM)	1 µl
GTP (10mM)	1 µl
Lig'n Scribe DNA product	1 µl
α^{32}P-dUTP	5 µl
T7/SP6 RNA polymerase	2 µl
H_2O	7 µl

The samples were incubated at 37°C for 20 min, 1 µl of DNase I was added, and incubation was continued at 37°C for a further 15 min. To separate the radiolabelled RNA probe from unincorporated nucleotides, a mixture of 3 M sodium acetate (5 µl), H_2O (30 µl) and 100% (v/v) ethanol (150 µl) was added, and the sample was incubated at -20°C for 30 min. The sample was centrifuged at 12,000 × g for 15 min at 4°C, the supernatent discarded and the pellet washed with (v/v) 70% ethanol. The pellets were allowed to dry, then were resuspended in H_2O and added directly to the hybridisation bottles.

2.2.24 Hybridisation

Membranes containing RNA were prehybridised in 15 ml of hybridisation buffer at 68°C in a Hybaid minioven. After 1 h, radioactive probe was added and the hybridisation continued overnight. The temperature was increased to 71°C when greater stringency was necessary. Hybridised membranes were washed once at room temperature with 2 × SSC/0.1% SDS for 20 min, then twice at 68°C with 0.1 × SSC/0.1% SDS for 20 min. The membranes were placed in a hybridisation bag, and exposed to a phosphoimager screen. The images were visualised using the GelDoc system (BioRad).

2.2.25 RT-PCR

RT-PCR was performed in 50 µl reactions using an Invitrogen kit:

2 × reaction mix	25 µl
Template RNA	2 µl
Sense primer	1 µl
Antisense primer	1 µl
RT/*Taq* Mix	1 µl
H_2O	20 µl

Samples were incubated at 55°C for 30 min (cDNA synthesis), then denatured at 94°C for 2 min. PCR amplification was 40 cycles of 94°C for 15s, 62°C for 30s, 68°C for 1 min/kbp, and a final extension step of 68°C for 5 min.

2.2.26 SMARTRACE

5'-and 3'-RACE reactions were carried out using the BD Clontech SMARTRACE kit. Two separate reactions for 5'-and 3'-RACE were prepared using polyA$^+$ mRNA template and primers supplied in the kit:

for 5'-RACE		*for 3'-RACE*	
RNA sample	1 µl	RNA sample	1 µl
5'-CDS primer	1 µl	3'-CDS primer	1 µl
H_2O	2 µl	H_2O	2 µl
SMART II A oligo	1 µl		

The reactions were incubated at 70°C for 2 min, placed on ice for 2 min and centrifuged briefly. The following reagents were then added to each reaction tube:

5 × First-Strand buffer	2 µl
DTT (20 mM)	1 µl
dNTPs (10 mM each)	1 µl
Reverse transcriptase	1 µl

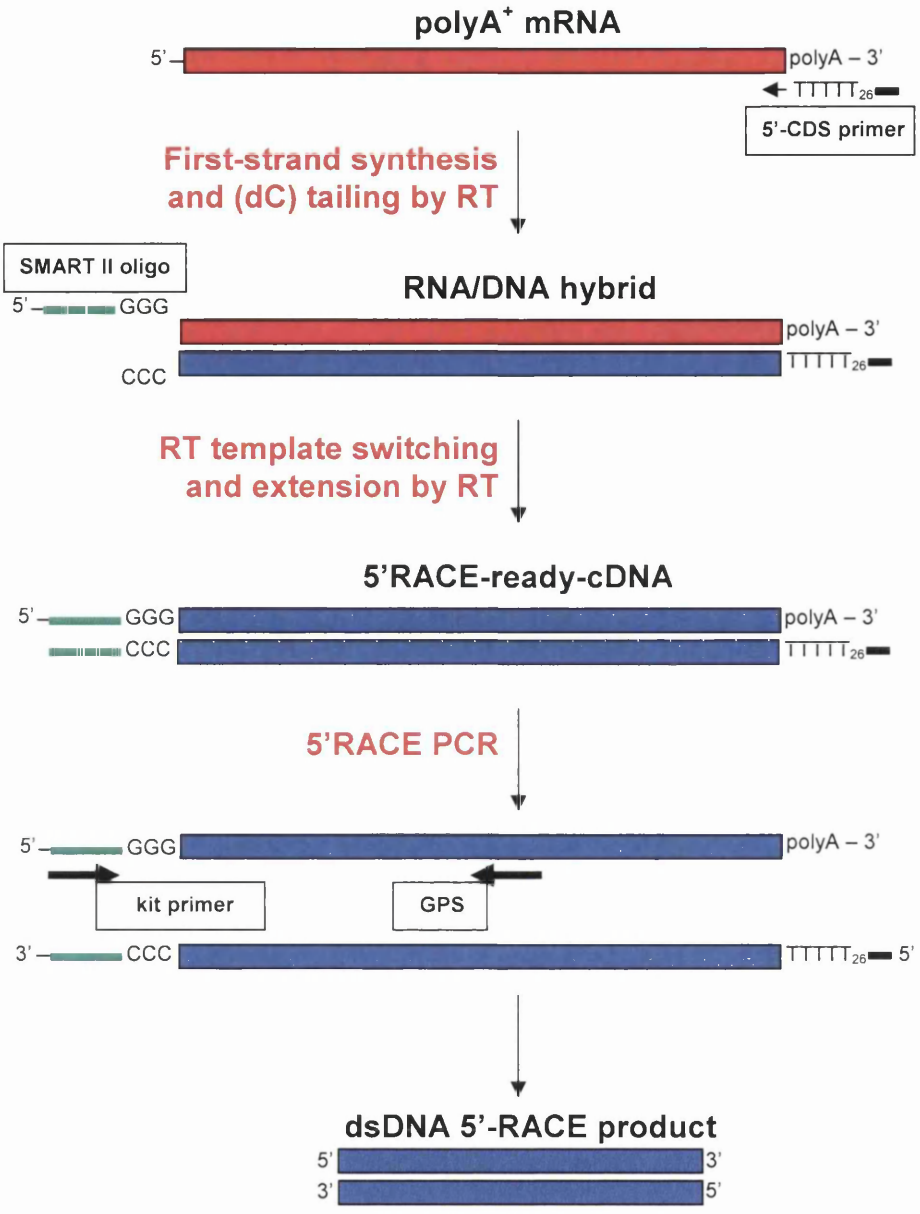


Figure 2.3. Mechanism of 5'-SMARTRACE reaction.

The 5'-CDS primer and SMART II oligonucleotide are used to synthesise the 5'-RACE-ready cDNA. When reverse transcriptase (RT) reaches the end of the mRNA template, several dC residues are added. The SMART II oligonucleotide anneals to the tail of the cDNA and acts as an extended template for RT. The GeneRacer and 5'RACE methods used slightly different methods of polyA$^+$ mRNA modification and purification, as detailed in sections 2.2.27 and 2.2.28. For each PCR reaction, a kit-supplied primer and gene-specific primer (GSP) are used.

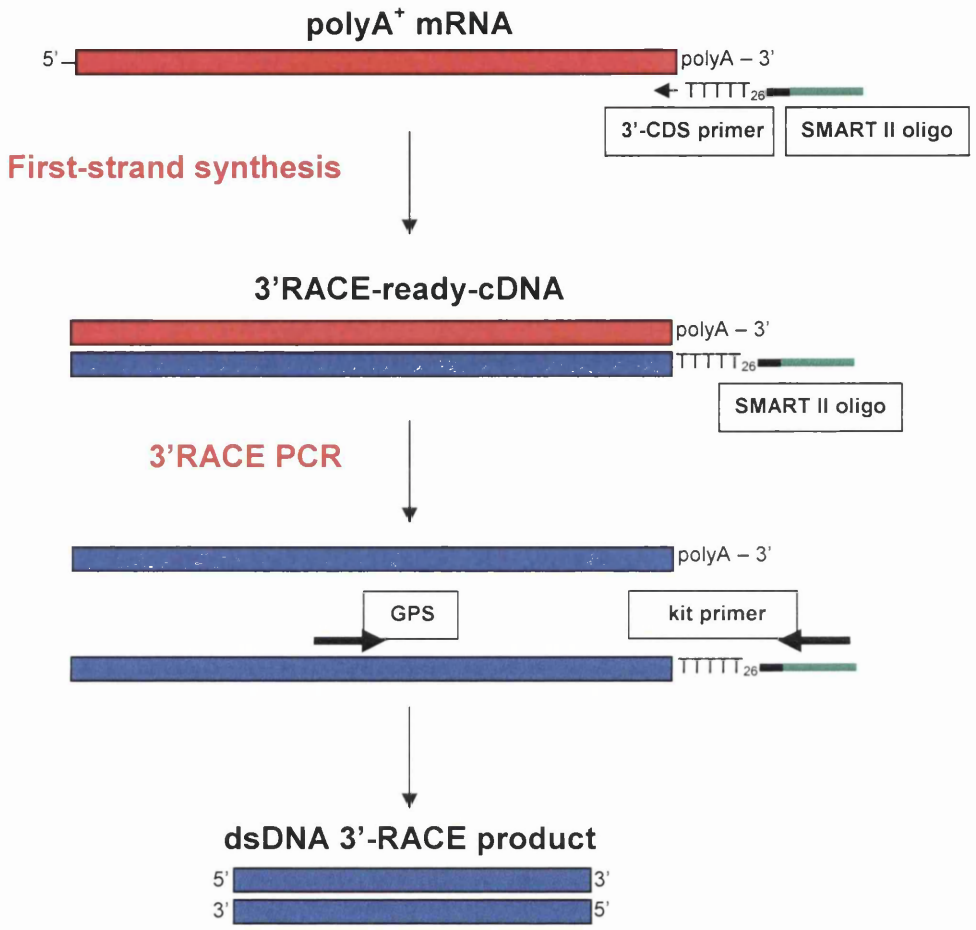


Figure 2.4. Mechanism of 3'-SMARTRACE reaction.

The 3'-CDS primer and SMART II oligonucleotide are used to synthesise the 3'-RACE-ready cDNA. The GeneRacer kit used a slightly different method to generate 3'-RACE-ready cDNA, as detailed in section 2.2.27. For each PCR reaction, a kit-supplied primer and gene-specific primer (GSP) were used.

The reactions were incubated at 42°C for 90 min in a thermocycler. First-strand reaction products were diluted in 200 µl Tricine-EDTA buffer and heated at 72°C for 7 min. RACE-ready cDNA was used immediately for PCR or stored at -20°C.

The control template supplied in the kit was human placental total RNA. A standard SMARTRACE PCR program consisted of a 94°C/2 min denaturation step, 5 cycles of 94°C/30 s, 72°C/3 min, 5 cycles of 94°C/30 s, 70°C/30s, 72°C/3 min, then 20 cycles of 94°C/30 s, 68°C/30 s, 72°C/3 min. 5'- and 3'-RACE-PCR reactions were prepared in 50 µl reactions:

UPM (10×)	5 µl
GSP	1 µl
5'- or 3'-cDNA template	2.5 µl
10 × buffer	5 µl
dNTPs (10 mM each)	1 µl
Advantage II polymerase	1 µl
H_2O	34.5 µl

2.2.27 GeneRacer

The Invitrogen GeneRacer kit was used to make 5'- and 3'-RACE-ready cDNA from polyA$^+$ mRNA. To dephosphorylate mRNA, 10 µl reactions were prepared:

RNA	1 µl
10 × CIP buffer	1 µl
RNaseOut	1 µl
CIP	1 µl
H_2O	6 µl

The samples were mixed, vortexed and centrifuged, and incubated at 50°C for 1 h. The GeneRacer ethanol precipitation step was then carried out. The samples were centrifuged, placed on ice, and 90 µl H_2O and 100 µl phenol:chloroform were added. The samples were vortexed for 30 s, centrifuged for 13,000 × g for 5 min, and the

upper layer was transferred to a fresh tube. The following 232 µl mixture was added to each sample;

10 mg/ml mussel glycogen	2 µl
3 M sodium acetate pH 5.2	10 µl
95% (v/v) ethanol	220 µl

The samples were vortexed and freeze-dried for 10 min. The RNA was pelleted by centrifugation for 13,000 × g for 20 min at 4°C. The supernatant was carefully removed, 500 µl of (v/v) 70% ethanol was added, and the pellet centrifuged at 13,000 × g for 2 min. The supernatant was removed carefully, the pellet was allowed to dry for 5 min and resuspended in 7 µl H_2O. To remove the cap structure, the following 10 µl reaction was performed;

Dephosphorylated RNA	7 µl
10× TAP buffer	1 µl
RNase Out	1 µl
TAP	1 µl

The samples were mixed, vortexed, centrifuged, and incubated at 37°C for 1 h. They were then subjected to the GeneRacer ethanol precipitation step described previously. The 7 µl of decapped mRNA was added to a vial of GeneRacer RNA Oligo, mixed and centrifuged. To ligate the GeneRacer RNA Oligo to the decapped mRNA, the mixture was incubated at 65°C for 5 min to relax the secondary RNA structure, then placed on ice for 2 min. The following reagents were added, in order:

10 × Ligase buffer	1 µl
10 mM ATP	1 µl
RNaseOut	1 µl
T4 RNA ligase	1 µl

The samples were mixed, vortexed, centrifuged, and incubated at 37°C for 1 h. They were then subjected to the same ethanol precipitation step as described previously. The final resuspension in this step, however, was in 10 µl H_2O. To reverse transcribe the mRNA, GSP (1 µl) and dNTAP mix (1 µl) were added to the reaction mix,

incubated at 65°C for 5 min, on ice for 2 min, and centrifuged. The following reagents were added, in order:

5 × RT buffer	4 µl
Cloned AMV RT	1 µl
H_2O	2 µl
RNaseOut	1 µl

The 20 µl reaction was mixed thoroughly and incubated at 45°C for 1 h. A 15 min incubation at 85°C followed in order to inactivate the reverse transcriptase. The RACE-ready cDNA was then used immediately for PCR or stored at -20°C. The control template supplied in the kit was HeLa cell total RNA. A standard GeneRacer PCR program consisted of a 94°C/2 min denaturation step, 5 cycles of 94°C/30 s, 72°C/2 min, 5 cycles of 94°C/30 s, 70°C/2 min, 20 cycles of 94°C/30 s, 65°C/30 s, 68°C/2 min, then a final extension step of 68°C/10 min. 5'- and 3'-RACE-PCR reactions were prepared in 50 µl reactions:

5'-or 3'-RACE primer	3 µl
GSP	1 µl
cDNA template	1 µl
10 × buffer	5 µl
dNTPs (10 mM each)	1 µl
Taq polymerase	1 µl
$MgSO_4$	2 µl
H_2O	36 µl

2.2.28 5'-RACE

The Invitrogen 5'-RACE kit was used to make 5'-RACE-ready cDNA from polyA$^+$ mRNA. Wash buffer (1 ml wash buffer concentrate, 18 ml H_2O and 21 ml 100% (v/v) ethanol) and 70% (v/v) ethanol were prepared and stored at 4°C until immediately before use.

A mixture of template mRNA (1 µl), gene-specific primer (1 µl) and H_2O (13.5 µl) was incubated at 70°C for 10 min and then on ice for 1 minute. The following reagents were added, in order:

10 × PCR buffer	2.5 µl
25 mM MgCl2	2.5 µl
10 mM dNTPs	1 µl
0.1 M DTT	2.5 µl

The reaction was incubated at 42°C for 1 min, and 1 µl SuperScript II RT was added and incubation was continued at 42°C for a further 50 min. The reaction was terminated by incubation at 70°C for 15 min, and then 1 µl of RNase was added. The reaction was incubated at 37°C for a further 30 min and then placed on ice.

Binding solution (6 M NaI) was brought to room temperature, and 120 µl added to each reaction. The cDNA/NaI solution was transferred to a SNAP column and centrifuged at 13,000 × g for 20 s. The flow-through was retained for use as a control reaction. The SNAP column was washed with 0.4 ml cold Wash Buffer, centrifuged at 13,000 × g for 20 s, and the flow-through discarded. This step was repeated three more times. The SNAP column was then washed with 0.4 ml cold 70% (v/v) ethanol, centrifuged at 13,000 × g for 20 s, and the flow-through discarded. This step was repeated, then the empty tube centrifuged for 1 min. The cartridge was transferred to a recovery tube, 50 µl of H_2O (preheated to 65°C) was added, and the tube centrifuged at 13,000 × g for 20 s to elute the cDNA. The following reagents were added:

H_2O	6.5 µl
5× tailing buffer	5 µl
2 mM dCTP	2.5 µl

The cDNA sample was heated for 3 min at 94°C, placed on ice and then centrifuged. TdT (1 µl) was added and the cDNA incubated at 37°C for 10 min. The TdT was heat inactivated by incubation at 65°C for 10 min. The RACE-ready cDNA was used immediately for PCR or stored at -20°C.

Template RNA supplied in the kit was used as a control. A standard 5'-RACE PCR program consisted of a 94°C/2 min denaturation step, 35 cycles of 94°C/1 min, 63°C/30 s, 72°C/2 min, then a final extension step of 72°C/10 min. 5'-RACE-PCR was prepared on ice in 50 µl reactions:

Abridged anchor primer	2 µl
GSP	2 µl
dC-tailed cDNA	5 µl
10× PCR buffer	5 µl
dNTPs (10 mM each)	1 µl
Taq polymerase	1 µl
25 mM MgCl2	3 µl
H_2O	31 µl

2.2.28 Computer analysis

The sequence database was compiled from electropherograms using PreGap4 and Gap4 (Staden *et al.*, 2000) and Phred (Ewing and Green, 1998; Ewing *et al.*, 1998). Sequences were analysed using the GCG suite of programs by Accelrys.

CHAPTER 3: REGION X

3.1 Protein-coding genes

A comparative analysis of the HCMV and CCMV genome (Davison *et al.*, 2003) significantly altered the proposed gene complement of HCMV from that described in the original analysis of strain AD169 (Chee *et al.*,1990). The study proposed ten novel genes, modified the interpretations of 24 genes and discounted 51 previously proposed ORFs because they had no counterparts in CCMV and lacked any other convincing evidence for expression. Some of the discounted ORFs were located together in the genome, and their omission left 'empty' regions in the genome. One of these regions is located between ORFs UL105 and UL112, which are accepted as protein-coding genes (Figure 3.1). This is here termed region X.

Prior to and since the original analysis of the AD169 genome (Chee *et al.*, 1990), functions have been assigned to many HCMV genes. However, few have been mapped within region X. The original analysis of AD169 predicted six ORFs in region X: UL106 to UL111 (Figure 3.1). There is no evidence for expression of proteins from these ORFs. However, a new protein-coding gene was identified in region X. A viral interleukin-10 homologue, vIL-10, is encoded by UL111A (Kotenko *et al.*, 2000). UL111A retains the name and start site from the original analysis, but was subsequently shown to be spliced. Three further novel, protein-coding ORFs have been predicted in region X. A recent study of the coding content of the HCMV genome analysed six BAC-cloned strains for sequence comparisons, four of which were low-passage isolates (Murphy *et al.*, 2003b). From a total of 252 ORFs that were conserved in all four clinical isolates, ORFs were subsequently discounted if they failed to meet specific criteria. This process is discussed in Section 1.18. From the list of 29 newly recognised ORFs with the potential for coding proteins, three are in region X: C-ORF16, C-ORF17 and C-ORF18 (Figure 3.1).

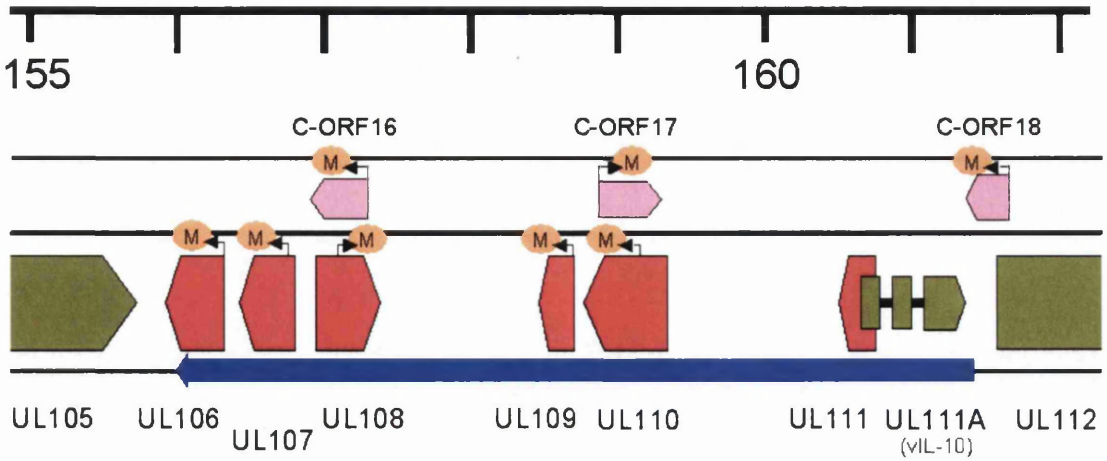


Figure 3.1. Map of the characterised genes and predicted genes in region X.

The scale at the top shows the genomic location (kbp) in the AD169 genome. Dark green shows parts of ORFs of known functions. Introns in UL111A are shown as narrow black bars. Pink shows novel ORFs predicted by Murphy *et al.* (2003b). Red shows ORFs predicted by Chee *et al.* (1990) that have no known function and have been discounted because they have no counterparts in CCMV (Davison *et al.*, 2003). This leaves a region of 4.9 kbp between UL105 and UL111A that appears to contain no protein-coding genes. The 3'-end of each ORF has a point to show its orientation, and the M shows the first in-frame ATG initiation codon in each red or pink ORF. Dark green ORFs start with ATG initiation codons. The blue line underneath the ORFs shows the 5 kb transcript and its orientation as described by Plachter *et al.* (1988).

The only confirmed gene in region X, the vIL-10 gene (UL111A), is composed of three exons and two introns (Kotenko *et al.*, 2000). The splice junctions correspond to those between exons 1 and 2 and exons 3 and 4 in the human IL-10 gene, which is composed of five exons and four introns. In contrast, the first herpesvirus IL-10 homologue that was identified, in EBV, contains no introns (Moore *et al.*, 1990; Hsu *et al.*, 1990). IL-10 gene homologues have also been identified in the corresponding region of the genomes of other primate CMVs (RhCMV, BCMV and SCMV), but not in CCMV (Lockridge *et al.*, 2000; Davison *et al.*, 2003a). vIL-10 can bind to the human IL-10 receptor and compete with human IL-10 for binding sites, despite the fact that there is only 27% amino-acid sequence identity between the two proteins (Kotenko *et al.*, 2000). IL-10 is a potent suppressor of aspects of the cellular immune

response, and the presence of a virus-encoded homologue helps HCMV to circumvent detection and destruction by the host immune system. It has recently been shown that vIL-10 suppresses production of CXCL10 (a T-cell chemoattractant) during viral replication (Cheeran *et al.*, 2003; Hu *et al.*, 2002). Other studies have shown that vIL-10 is able to suppress lymphocyte proliferation, proinflammatory cytokine production and cell-surface expression of major histocompatability complex molecules (Spencer *et al.*, 2002; Lockridge *et al.*, 2000). Using an experimental model of latency in which human hematopoietic cells were latently infected with HCMV, a vIL-10 transcript was detected that differed from that reported during productive infection. In this transcript the second intron was not excised from the mRNA (Jenkins *et al.*, 2004). The authors suggested that latent CMV might encode a functional, truncated vIL-10 that differs from that expressed during lytic infection.

At the left end of region X, UL105 specifies a 3.4 kb RNA, which was detected at 24 h p.i. but not in the presence of cycloheximide (Smith *et al.*, 1996). UL105 encodes the helicase activity of the HCMV helicase-primase complex, which associates with two other proteins encoded by UL70 (primase) and UL102 (primase-associated factor) (Smith *et al.*, 1996). At the right end of region X, UL112 is a spliced gene which incorporates UL112 and UL113, originally predicted to be two separate overlapping ORFs (Chee *et al.*, 1990; Rodems *et al.*, 1988). Alternative splicing of UL112 specifies different 2.2 kb RNAs which encode four phosphorylated proteins expressed at early times p.i.: pp34, pp43, pp50 and pp84 (Staprans *et al.*, 1988; Wright *et al.*, 1988). The UL112 proteins appear to be required for efficient viral DNA replication, although they may not be essential, and they act as transactivators of the UL54 (DNA polymerase) promoter.

Of the other ORFs originally defined in region X, only UL106 and UL109 commence with ATG initiation codons. The other ORFs are preceded by termination codons, with UL107, UL108 and UL110 containing in-frame ATG codons internally and UL111 lacking such codons entirely (Figure 3.1). The three novel C-ORFs all commence with ATG initiation codons.

3.2 The 5 kb RNA

A number of studies in the early 1980s identified and mapped the major HCMV transcripts. One of these was a 5 kb transcript mapping within region X (Figure 3.1). This transcript was first described in strains Towne (Wathen & Stinski, 1982) and Davis (DeMarchi *et al.*, 1983), and subsequently in strain AD169 (Nelson *et al.*, 1984; Jahn *et al.*, 1984). Plachter *et al.* (1988) mapped the 5 kb RNA, confirming its presence at high levels at IE and L times p.i., and at low levels at E times p.i. (in total RNA). The transcript was also found in equal abundance in cytoplasmic and nuclear RNA fractions, and evenly distributed between polyadenylated and nonpolyadenylated fractions in total infected cell RNA. To determine the size of the RNA, gels of a reduced agarose concentration were electrophoresed under a low voltage gradient. This led to the identification of two bands of approximately 5.0 and 4.6 kb. Northern blot, nuclease protection and primer extension experiments were used to map the 5'- and 3'-ends of the 5 kb transcript. The 5'-end was mapped to a G residue at nucleotide position 160,570 in AD169, and the 3'-end to two sites downstream from a consensus polyadenylation signal at 155,759: A residues at 155,750 and 155,737. S1 analysis of a plasmid clone in the middle of the 5 kb RNA had previously indicated that splicing did not occur (Nelson *et al.*, 1984), and no further indication of splicing at the ends of the RNA was found by Plachter *et al.* (1988).

The 5 kb RNA was also identified as one of a number of viral RNAs specifically incorporated into virions (Bresnahan *et al.*, 2000; Greijer *et al.*, 2000). The RNAs were potentially delivered to the host cell on infection, suggesting that they could function immediately after virus entry, independent of transcription from the viral genome. However, a subsequent study found that RNAs were packaged into HCMV virions in proportion to their cellular concentration, independent of a *cis*-acting packaging element, and that viral and cellular RNAs were non-specifically incorporated into HCMV through interactions with several virion proteins (Terhune *et al.*, 2004). The functional significance of virion-associated RNAs is therefore questionable. Northern blot analysis of the 5 kb RNA detected the transcript in virions, but at a decreased proportion from that in total infected cell RNA (Terhune *et al.*, 2004).

3.3 Sequence comparisons

UL105, UL111A and UL112 are well characterised protein-coding genes that would be expected to be conserved between HCMV strains. Other ORFs in region X, UL106-UL111 (Figure 3.1), are not considered to be protein-coding because they lack counterparts in CCMV. However, if these, or any other previously unrecognised ORFs in the region, are protein-coding, they would also be expected to be conserved between HCMV strains. Sequence comparisons of multiple HCMV strains could be used to identify which of ORFs UL106-UL111 are conserved and which are disrupted by insertions or deletions. Sequence comparisons could also be used to investigate the 5 kb RNA (Figure 3.1).

Region X was sequenced in seven HCMV strains: high passage strain AD169, low passage strains Toledo, Merlin, 3157 and 6397, and strains W and 3301 obtained directly from human tissue (Section 2.1.1). Initially, a single PCR product, originating from within UL105 and terminating in UL112, was amplified for each strain. However, very few plasmid clones were generated. Consequently, three overlapping PCR products were amplified (Figure 3.2) and cloned into plasmids, and at least four plasmids were sequenced for each product. Sequencing of multiple clones was performed to rule out PCR errors. Consensus sequences were compiled from the sequence data and PCR errors were identified where a single clone was different from the others at a specific position. When PCR fragments were not obtained from certain strains, alternative primers were used. The third fragment of strain 3301 was amplified using primers X30 and X18, and the three fragments of strain W were amplified using primers X2 and X7, X6 and X13, and X30 and X18 respectively (Appendix). Within the first PCR fragment (Figure 3.2), a 350 bp region failed to generate readable sequence. Therefore, PCR products were amplified for each strain using primers adjacent to this region (X16 and X32, Appendix). These products were cloned and sequenced using dGTP Big Dye. The sequence database was compiled from electropherograms using PreGap4 and Gap4 (Staden *et al.*, 2000) and Phred (Ewing and Green, 1998; Ewing *et al.*, 1998).

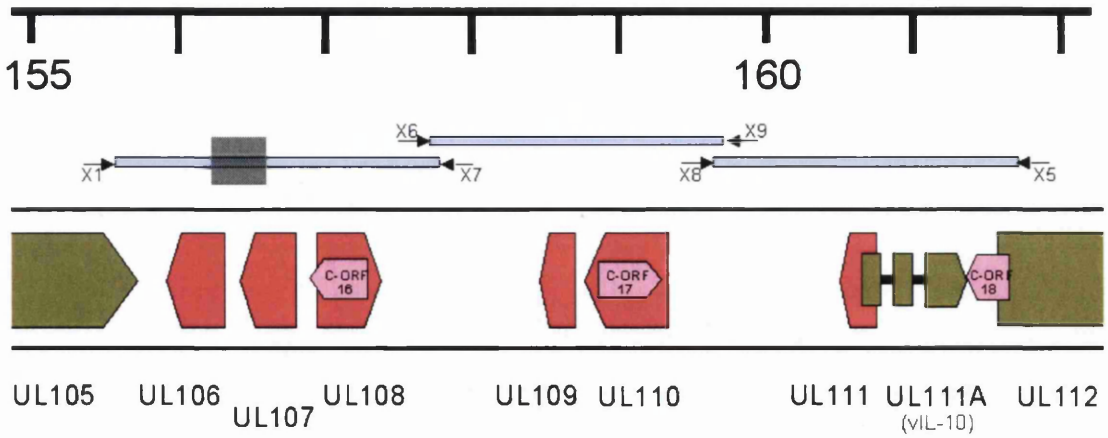


Figure 3.2. PCR amplification of region X for sequence comparisons.

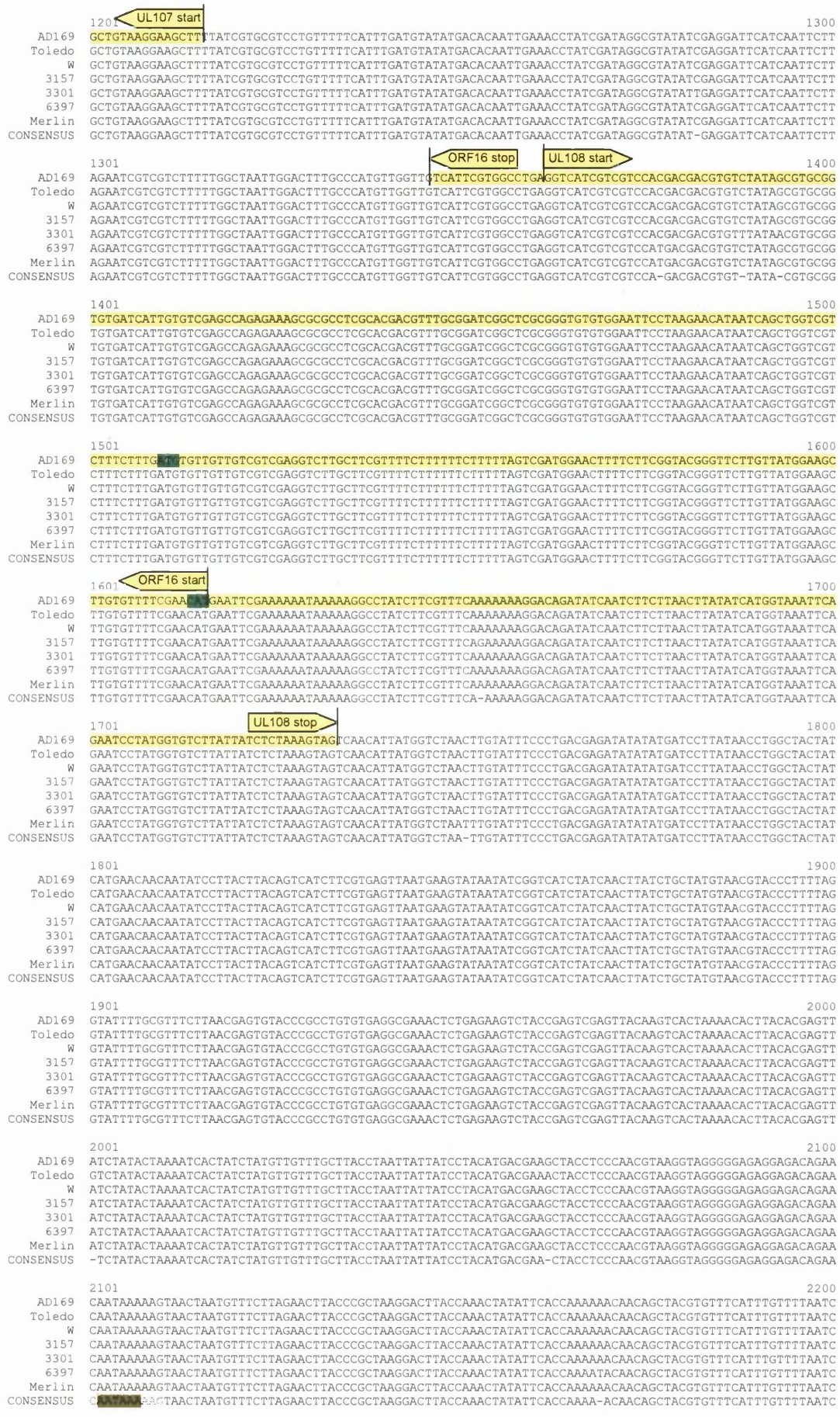
Blue bars show the three overlapping PCR fragments that were amplified using primers X1 and X7, X6 and X9, and X8 and X5 respectively (Appendix). The grey box in the first PCR fragment is a 350 bp region that initially failed to generate readable sequence. This region was amplified using primers X16 and X32 (Appendix).

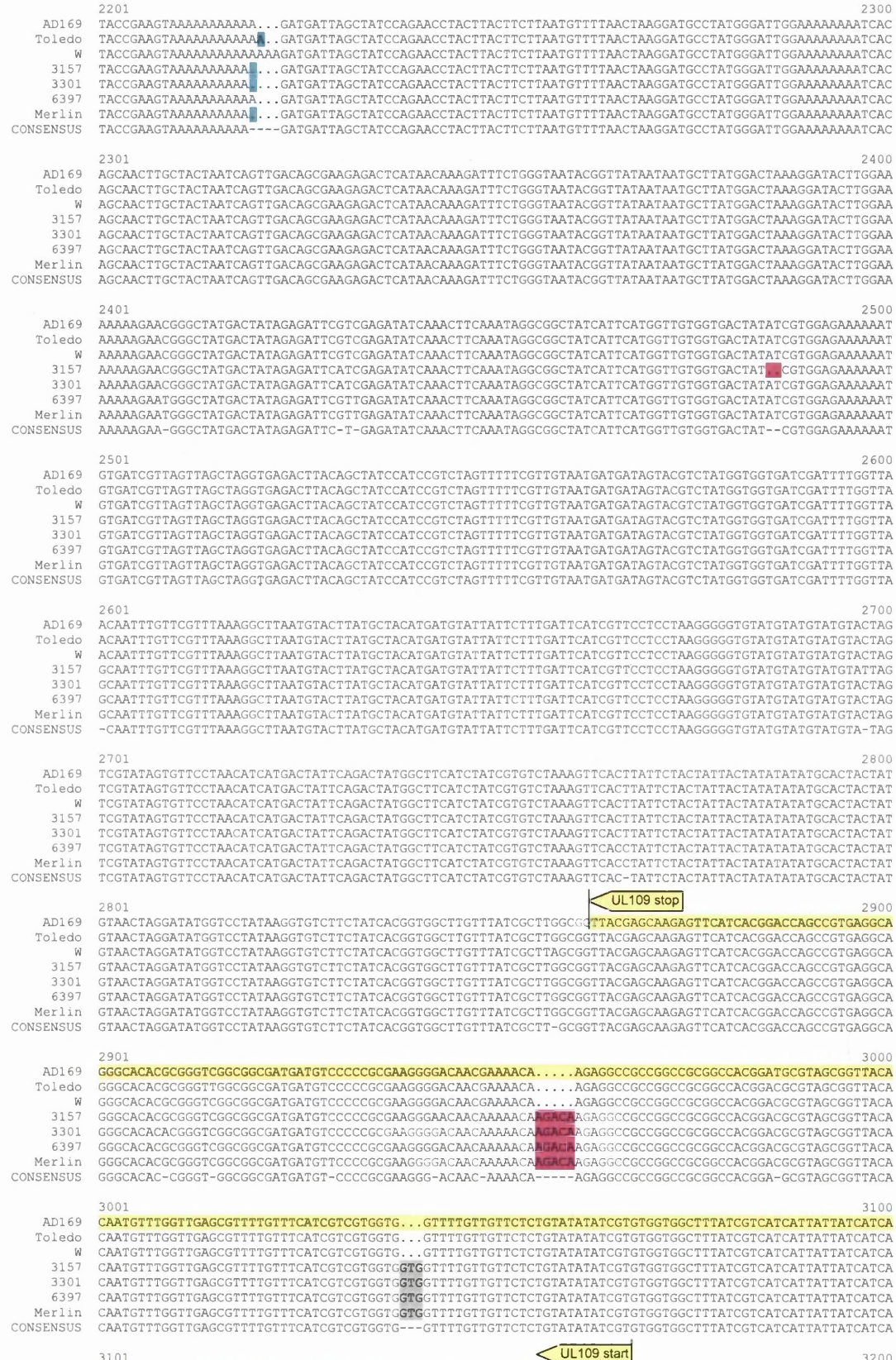
Figure 3.3. Sequence comparisons of seven HCMV strains in region X.

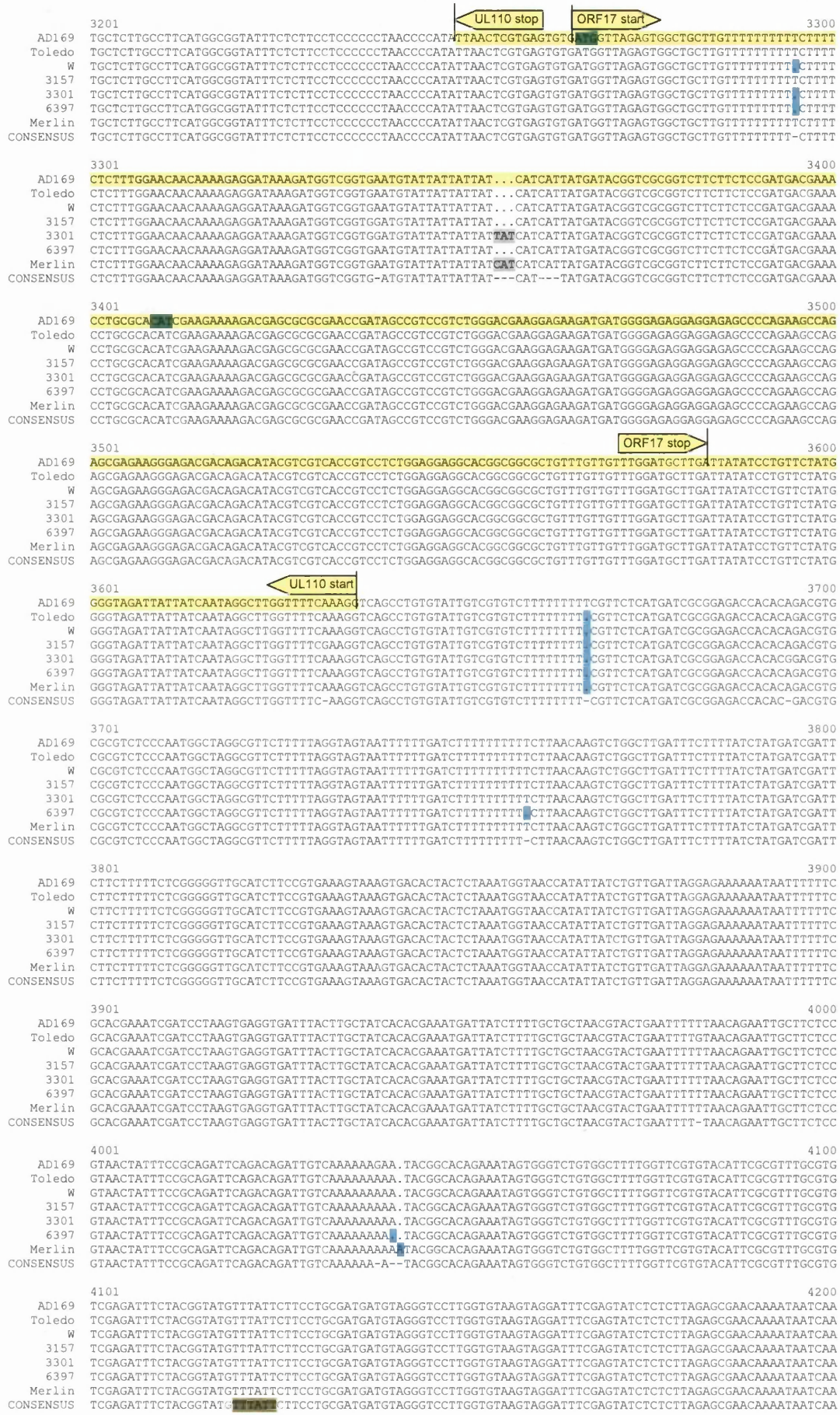
Alignments were made using ClustalW and displayed using Pretty. Yellow AD169 sequence shows all characterised and predicted ORFs in the region. The start and stop codons for each ORF are shown, the arrows indicating their orientation. Green shows the first ATG initiation codon for each ORF. Dark blue AD169 sequence shows introns for UL111A. Purple shows insertions/deletions that cause frameshifts in comparison to AD169. Sky blue shows variation in polynucleotide A:T tracts of eight or more residues that cause frameshifts in coding frames in comparison to AD169. Grey shows in-frame insertions/deletions. Brown consensus residues show potential polyadenylation signals (AATAAA). Sequence for strain W is not complete at the ends because different primers were used for PCR amplification. The Figure is continued on the following pages.

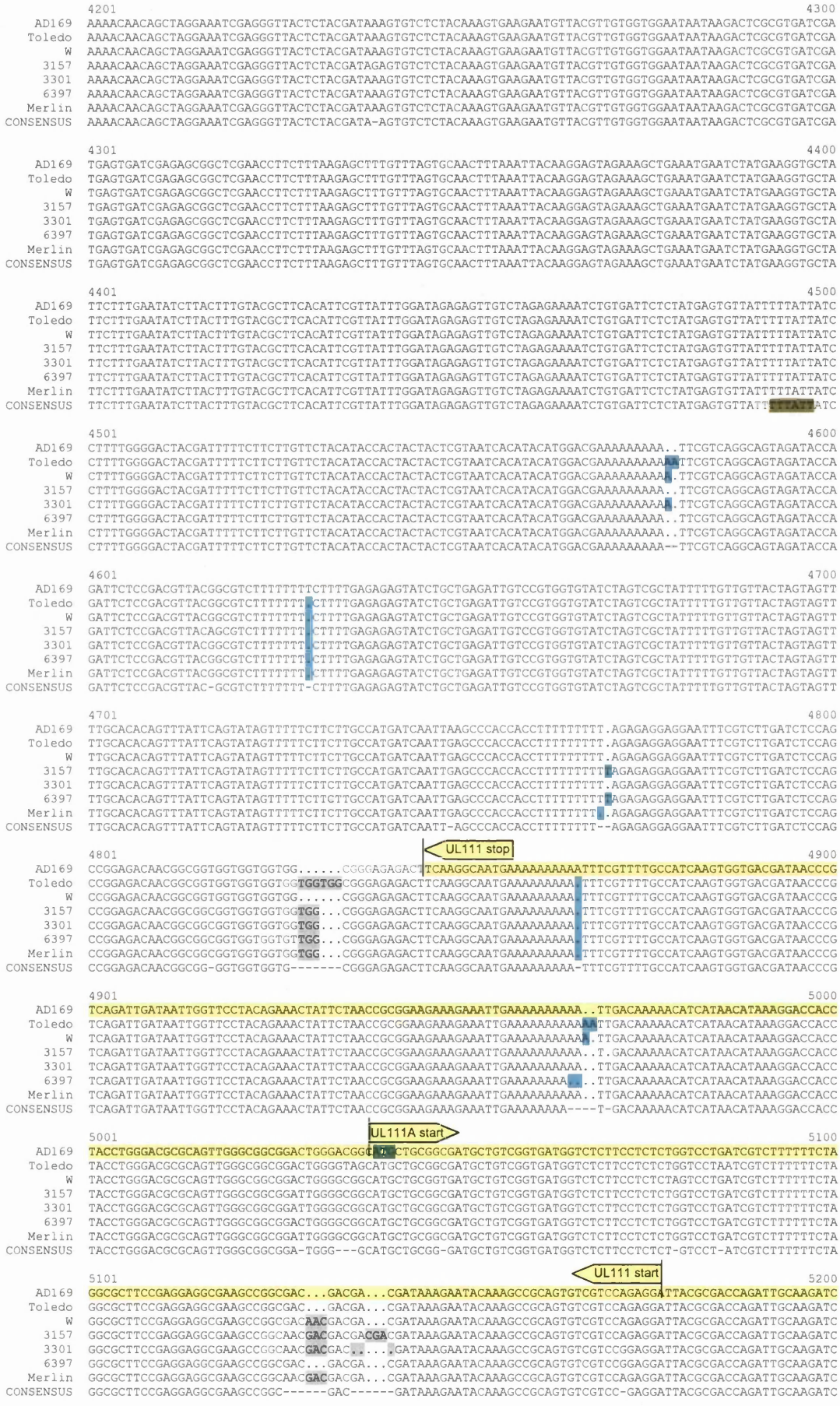
```
          1                                                                                                  100
AD169     GTCGCGAGTCACGGACCCCGAACACCTCATGATGAACGTTAACCCGTTGCGACTGCCCTATGAGAAGAACACCGCTATCACCCCCTATATCTGTCGCGCG
Toledo    GTCGCGAGTCACGGACCCCGAGCACCTCATGATGAACGTTAACCCGTTGCGACTGCCCTATGAGAAGAACACCGCCATCACCCCCTATATCTGTCGCGCG
W         ..................................................GCCCTATGAGAAGAACACCGCTATCACCCCCTATATCTGTCGCGCG
3157      GTCGCGAGTCACGGACCCCGAGCACCTCATGATGAACGTCAACCCGTTGCGACTTCCCTATGAGAAGAACACCGCTATCACCCCCTATATCTGTCGCGCG
3301      GTCGCGAGTCACGGACCCCGAGCACCTCATGATGAACGTCAACCCGTTGCGACTTCCCTATGAGAAGAACACCGCTATCACCCCCTATATCTGTCGCGCG
6397      GTCGCGAGTCACGGACCCCGAGCACCTCATGATGAACGTCAACCCGTTGCGACTTCCCTATGAGAAGAACACCGCTATCACCCCCTATATCTGTCGCGCG
Merlin    GTCGCGAGTCACGGACCCCGAGCACCTCATGATGAACGTCAACCCGTTGCGACTTCCCTATGAGAAGAACACCGCTATCACCCCCTATATCTGTCGCGCG
CONSENSUS ----------------------------------------------------CCCTATGAGAAGAACACCGC-ATCACCCCCTATATCTGTCGCGCG

                                UL105 stop
          101                                                                                                200
AD169     CTCAAAGACAAACGCACCACGCTTATTTTTTGACACAACACCGTGTAAGGAAAACGTGACTTTATTGAGCAGGGTAAAAACCACGTACAAGAACCACGTT
Toledo    CTCAAAGACAAACGCACCACGCTTATTTTTTGACACAACACCGTGTAATGAAAACGTGACTTTATTGAGCAGGGTAAAAACCACGTACAAGAACCACGTT
W         CTCAAAGACAAACGCACCACGCTTATTTTTTGACACAACACCGTGTAAGGAAAACGTGACTTTATTGAGCAGGGTAAAAACCACGTACAAGAACCACGTT
3157      CTCAAAGACAAACGCACCACGCTTATTTTTTGACACAACACCGTGTAAGGAAAACGTGACTTTATTGAGCAGGGTAAAAACCACGTACAAGAACCACGTT
3301      CTCAAAGACAAACGCACCACGCTTATTTTTTGACACAACACCGTGTAAGGAAAACGTGACTTTATTGAGCAGGGTAAAAACCACGTACAAGAACCACGTT
6397      CTCAAAGACAAACGCACCACGCTTATTTTTTGACACAACACCGTGTAAGGAAAACGTGACTTTATTGAGCAGGGTAAAAACCACGTACAAGAACCACGTT
Merlin    CTCAAAGACAAACGCACCACGCTTATTTTTTGACACAACACCGTGTAAGGAAAACGTGACTTTATTGAGCAGGGTAAAAACCACGTACAAGAACCACGTT
CONSENSUS CTCAAAGACAAACGCACCACGCTTATTTTTTGACACAACACCGTGTAA-GAAAACGTGACTTTATTGAGCAGGGTAAAAACCACGTACAAGAACCACGTT
```

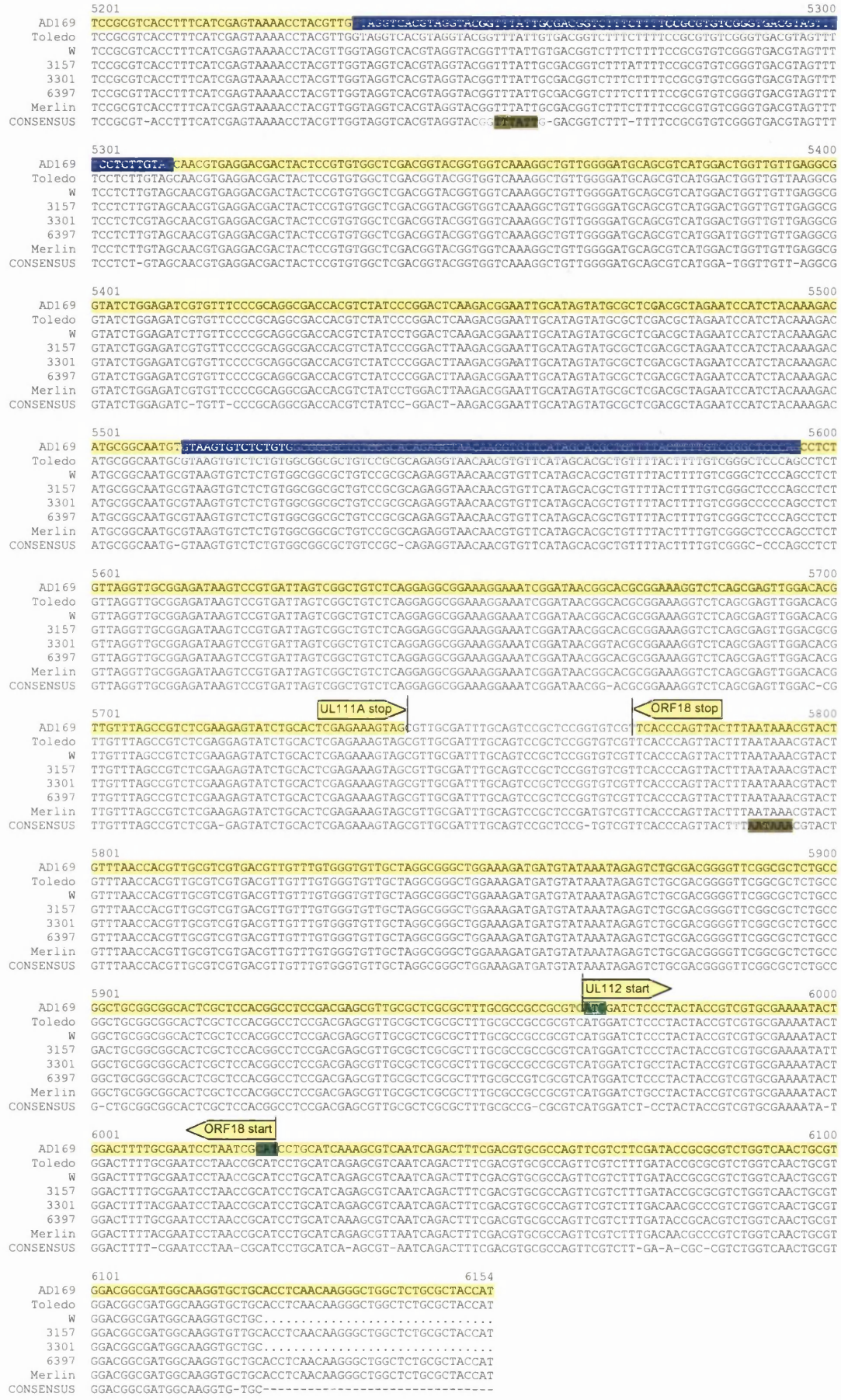

UL106 stop
UL106 start
UL107 stop
AD169
Toledo
W
3157
3301
6397
Merlin
CONSENSUS


UL107 start
ORF16 stop
UL108 start
ORF16 start
UL108 stop
AD169
Toledo
W
3157
3301
6397
Merlin
CONSENSUS


UL109 stop
UL109 start


UL110 stop
ORF17 start
ORF17 stop
UL110 start
AD169
Toledo
W
3157
3301
6397
Merlin
CONSENSUS


UL111 stop
UL111A start
UL111 start


UL111A stop
ORF18 stop
UL112 start
ORF18 start

Region X is well conserved, with over 99% identity between strains. The data for AD169 confirmed the previously published sequence in this region (Chee *et al.*, 1990). Data for Merlin confirmed the recently published sequence for this strain (Dolan *et al.*, 2004) with the exception of three differences, all found in the length of polynucleotide A:T tracts. Variation in the length of A:T tracts can be an artefact of PCR amplification, as the fidelity of DNA synthesis is affected by factors including the polymerase and 3'-5' exonuclease activity (Hite *et al.*, 1996). The variation occurs due to slipped strand extension by *Taq* DNA polymerase (Shinde *et al.*, 2001). During amplification, insertion/deletion events can produce 'stutter' products that differ from the original template by multiples of the repeat unit length (in this case, one nucleotide). The mutation rate of $(A)_n$ repeats was estimated at 1.5×10^{-2} per repeat per PCR cycle with contractions outnumbering expansions by more than five-fold (Shinde *et al.*, 2001). The type of polymerase used for PCR amplification in the current study was a mixture of *Taq* polymerases (Section 2.1.2) and could be expected to have a higher fidelity than other *Taq* polymerases, but PCR errors and variation in the length of A:T tracts between clones were nonetheless observed. The threshold for detecting stutter products was found to be eight repeat units (Shinde *et al.*, 2001), and sequence analysis of the HCMV strains in region X showed that only A:T tracts of eight residues or more showed repeat length variation between clones. These differences could not be considered a reliable indicator of sequence variation between strains. Polynucleotide A:T tracts of eight or more residues are very rare in protein-coding regions in HCMV (discussed in Chapter 6), and are found at a much higher frequency in region X than in the rest of the genome (Table 3.1). Similar patterns are observed in the corresponding regions of other CMV genomes (Table 3.1).

During the course of this work, an analysis of the coding content of the HCMV genome was carried out sequencing six strains that had been cloned as infectious bacterial artificial chromosomes (BACs) (Murphy *et al.*, 2003b). Two of these strains were extensively passaged laboratory strains (AD169 and Towne) and four were low-passage isolates (Toledo, FIX, PH and TR). Each BAC DNA was sequenced using a shotgun cloning method, thus avoiding PCR-generated length variation in A:T tracts. The corresponding sequences of the four strains not utilised for comparison in the current study (Towne, FIX, PH and TR) were subsequently compared to the seven

strains analysed in the present study in order to investigate variation in the A:T tracts and other differences between strains that would cause frameshifts.

	Number of A:T tracts of ≥ 8 residues/kbp	
Genome	**Region X**	**Rest of genome***
HCMV	1.85	0.20
CCMV	1.33	0.04
RhCMV	1.28	0.06
MCMV	1.11	0.03
RCMV	1.40	0.10
TCMV	0.42	0.03

Table 3.1. The distribution of A:T tracts in different CMV genomes.

* Region O, the area between UL57 and UL69, which is discussed in Chapter 4, also contains several A:T tracts and is excluded.

Despite the high level of sequence identity between the strains in the present study, each of UL106-UL111, with the exception of UL108 (and C-ORFs 16 and 18), were disrupted in certain strains by nucleotide insertions or deletions that shift the reading frame in comparison with AD169. UL108, which was not disrupted by a frameshift, contains a single nucleotide substitution that introduces an in-frame termination codon. As expected of recognised genes, no frameshifts were registered in UL105, UL111A or UL112. The data in Table 3.2 shows all the frameshifts in each ORF caused by insertions or deletions. The Table also specifies which of these are caused by differences in A:T tract length, where strains differ from each other in the number of A or T residues. As explained above, variation in A:T tract length within strains could be due to PCR amplification artefacts, and these differences were an unreliable indicator of real variation between strains. However, data from the four strains of HCMV cloned as BACs were used to identify variation in the A:T tracts (Table 3.2, Figure 3.4).

In addition to the frameshifts identified in UL106-UL111, a total of nine differences that would cause frameshifts were also found in the regions between the ORFs (Figure 3.4). Further analysis of region X not affected by frameshift mutations revealed no novel ORFs potentially encoding proteins. All ORFs in strains AD169 and Merlin are shown in Figure 3.5. The potential of region X to contain protein-coding ORFs is limited by the high frequency of stop codons in all six coding frames and by frameshifts in the few ORFs conserved in the HCMV strains. Consequently, it is unlikely that ORFs UL106-UL111, or any other ORFs in region X, code for proteins.

ORF	Number of frameshift mutations affecting ORF	Number of these frameshifts caused by variation in A:T tract length (≥ 8 bp)
UL105	**0**	**0**
UL106	4	0
UL107	2	1
C-ORF 16	0	0
UL108	0	0
UL109	1	0
UL110	1	1
C-ORF 17	1	1
UL111	2	2
UL111A	**0**	**0**
C-ORF 18	0	0
UL112	**0**	**0**

Table 3.2. Frameshift mutations in the ORFs of region X, identified by sequence comparisons.

The central column concerns all frameshift mutations which affect the ORF. The right column concerns the number of these frameshift mutations caused by differences in A:T tracts of 8 or more residues.

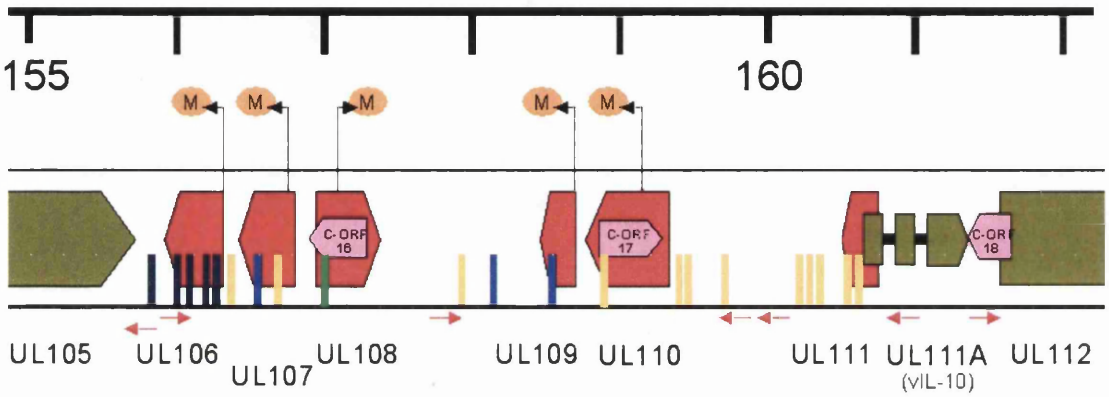


Figure 3.4. Frameshifts between HCMV strains in region X.

Navy blue lines show insertions/deletions that cause frameshifts in comparison to AD169. Gold lines show A:T tracts of variable length that cause frameshifts in comparison to AD169. Green lines show single base substitutions that introduce an in-frame stop codon. Red arrows show potential polyadenylation signals (AATAAA) and the M shows the first in-frame ATG codon of each ORF. No frameshifts were detected in UL105, UL111A or UL112.

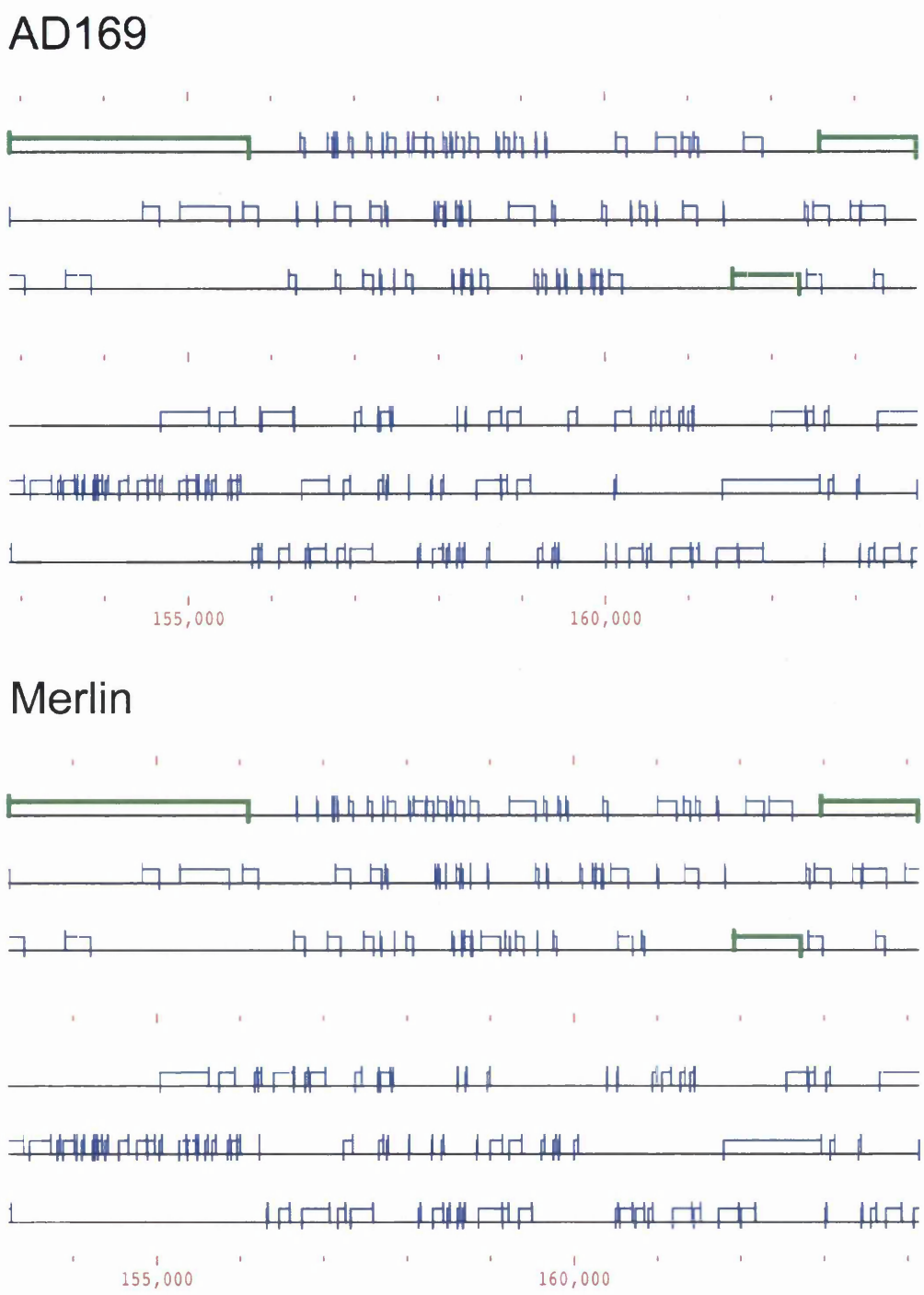


Figure 3.5. ORFs in region X.

ORFs for the six translation frames of the DNA sequences of AD169 and Merlin have been computed using Frames. Start codons are shown as short lines extending above the box, and stop codons are shown as short lines extending below the box. The region shown is larger than that shown in Figures 3.1 to 3.4, to include ORFs UL105 and UL112/UL113, which are highlighted in green.

3.4 Transcripts in region X: northern blot and RT-PCR analysis

Northern blots were carried out to determine the expression kinetics and sizes of transcripts from region X. Equal amounts of polyA$^+$ AD169 mRNA were electrophoresed in agarose-formaldehyde gels, blotted to membranes overnight, and fixed by UV irradiation. A polyA$^+$ RNA marker ladder was electrophoresed alongside the mRNA samples to aid estimation of transcript sizes (Section 2.1.8).

The three overlapping PCR products which were used to sequence region X (Figure 3.2) were used to make dsDNA probes (section 2.2.22). The 5 kb RNA characterised by Plachter *et al.* (1988) was detected by each, and sized at approximately 4.6 kb. Detection of this in L RNA at high levels, in much lower levels in IE RNA, and not at all in E RNA (Figure 3.6), is consistent with previously published data (Plachter *et al.*, 88). A heterogeneous transcript of approximately 1.1 kb was also detected in L RNA by probe A, possibly by probe C and not by probe B (Figure 3.6).

Probe	ORF	Strand	F primer	R primer	Size (bp)
D	UL106	-	X16	**X40**	321
E	UL108	+	**X25**	X7	566
F	UL109/110	-	X20	**X21**	720
G	UL111A	+	**X57**	X59	457
H	UL111	-	X57	**X59**	457
I	C-ORF18	-	X53	**X54**	149

Table 3.3. ssRNA probes used for northern blots in region X.

Each probe was amplified by Lign'Scribe PCR reaction (Section 2.2.23) using the primers listed in Appendix. The positions of the probes within the region are shown in Figure 3.7. For each of probes D-I, the primer in bold was used in the MaxiScript radiolabelling reaction to make the ssRNA probe (Section 2.2.23).

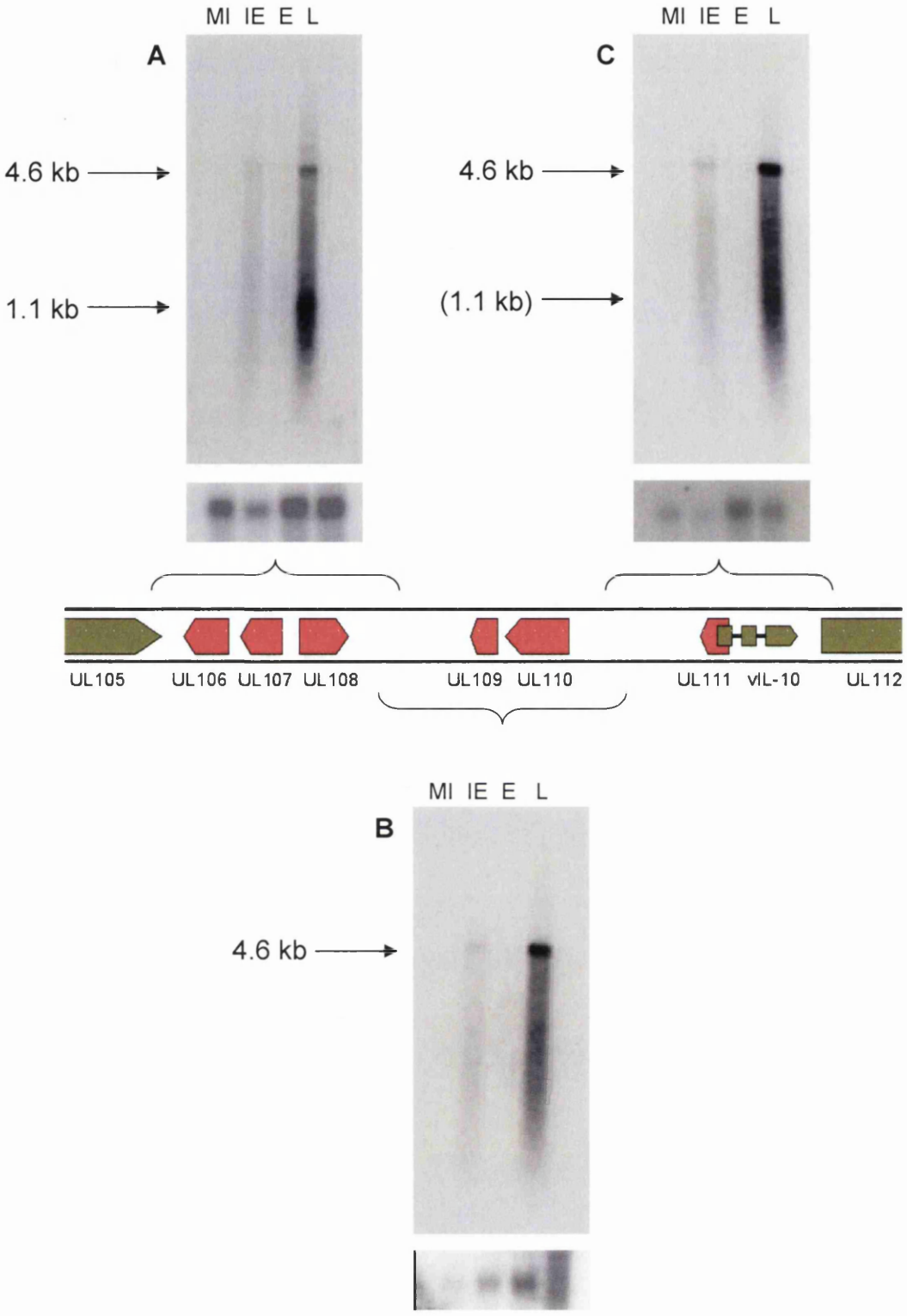


Figure 3.6. Northern blot analysis of region X using dsDNA probes.

The locations of the DNA probes are shown. Each blot included IE, E and L AD169 polyA$^+$ mRNA, and a negative control of MI polyA$^+$ mRNA. The results obtained using a GAPDH probe are shown at the bottom of each blot.

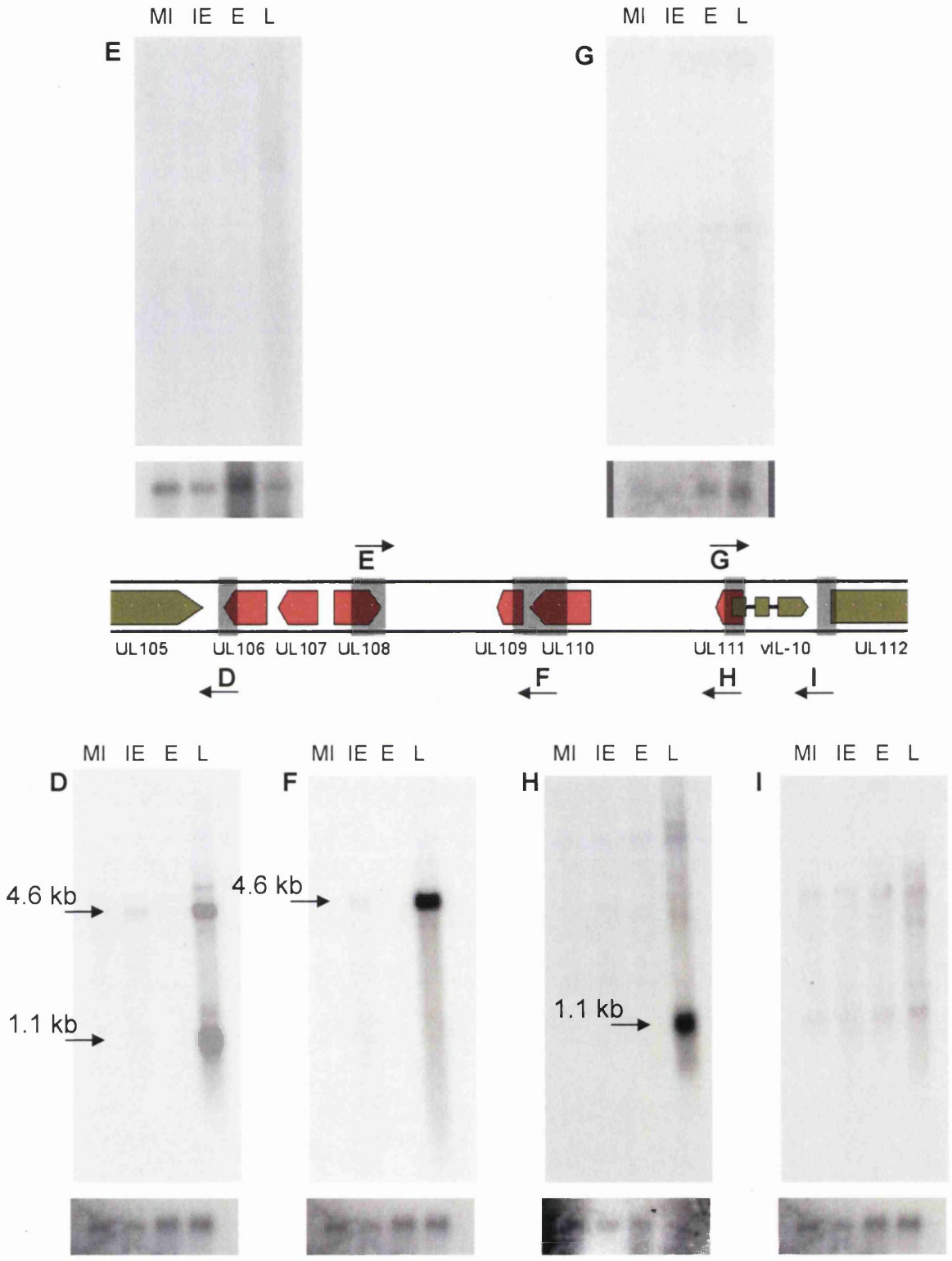


Figure 3.7. Northern blot analysis of region X using ssRNA probes.

The locations of the ssRNA probes are shown (not to scale, see Table 3.3 for accurate sizes), the arrows indicating the direction of any RNAs they would detect. Each blot included IE, E and L AD169 polyA$^+$ mRNA, and a negative control of MI polyA$^+$ mRNA. The results obtained using a GAPDH probe are shown at the bottom of each blot.

Strand-specific ssRNA probes were also used in northern blots to investigate the transcripts in region X. The primers used for generating the ssRNA probes are listed in Table 3.3. PCR products generated by the Lign'Scribe reaction (Section 2.2.23) were radiolabelled using the MaxiScript reaction. When hybridised to northern blots, ssRNA probes create RNA-RNA hybrids that are more stable than the DNA-RNA hybrids. Their strand-specificity also avoids the depletion of available probe by hybridisation to the complementary strand that occurs with dsDNA probes. All blots were performed twice to ensure their reproducibility and accuracy. Probe D (corresponding to UL106) detected the 4.6 kb RNA in IE and L RNA and the 1.1 kb RNA in L RNA, which was also detected by dsDNA probe A (Figures 3.7 and 3.6, respectively). Probe E (UL108) did not detect any transcripts, but the 4.6 kb RNA was detected by probe F (UL109/110). Probe H (UL111) also detected a 1.1 kb RNA in the same orientation as the other transcripts, which was not detected by probe I. Probe G (UL111A) did not appear to detect the UL111A RNA (of approximately 0.7 kb). Previous studies of this gene also did not detect the transcript by northern blot, possibly because it is present in low abundance during lytic infection (Jenkins *et al.*, 2004; Cheeran *et al.*, 2003; Kotenko *et al.*, 2000).

The 4.6 kb RNA was detected in the middle of region X, between UL106 and UL111A, and ssRNA probes at the ends of the region detected a 1.1 kb RNA in the same orientation as the larger transcript. Although two different 1.1 kb RNAs may have been detected, it is possible that this transcript is spliced, with the probes at the ends of the region (probes D and H, Figure 3.7) detecting the exons flanking the 4.6 kb intron.

Plachter *et al.* (1988) described the 5 kb RNA as being evenly distributed between polyA$^+$ and polyA$^-$ fractions of total infected cellular RNA. However, transcript mapping in the present study suggested that the 4.6 kb RNA detected was the intron derived during maturation of a 1.1 kb transcript. To check the polyadenylation status of both transcripts, total cellular RNA was prepared at late times from a high m.o.i. HCMV infection, and equal amounts were separated into polyA$^+$ and polyA$^-$ fractions (Section 2.2.20). Probe A detected the 4.6 kb RNA as a strong signal in the total RNA and polyA$^-$ RNA samples, but as a very faint signal in the polyA$^+$ sample (Figure 3.8). The small amount of 4.6 kb RNA in the polyA$^+$ fraction may be a result

of binding of the RNA to the oligo(dT) matrix via the many poly(A) tracts found throughout the sequence. The 1.1 kb message was detected as a strong signal in the total RNA and polyA$^+$ mRNA samples, but as a very weak signal in the polyA$^-$ samples. The small amount of 1.1 kb RNA in the polyA$^-$ sequence is likely to be contamination from when the polyA$^+$ and polyA$^-$ fractions were separated. Therefore, the 4.6 kb RNA is not polyadenylated, consistent with it being an intron spliced during maturation of the 1.1 kb RNA, which has exons at the ends of region X.

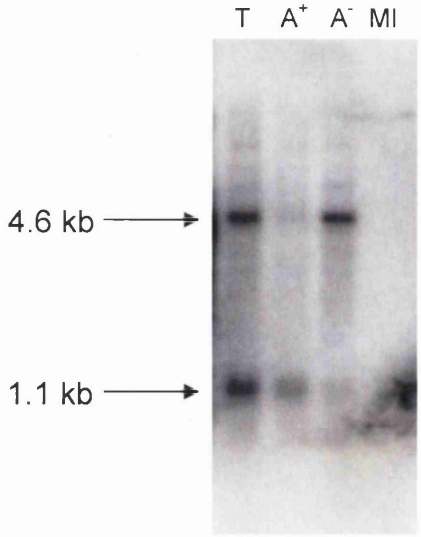


Figure 3.8. Polyadenylation of RNAs in region X.

Lane T contains late total cellular RNA, lane A$^+$ contains the polyA$^+$ fraction of the total RNA sample and lane A$^-$ contains the polyA$^-$ fraction of the total RNA sample. Lane MI contains mock-infected RNA. The blot was probed using dsDNA probe A (Figure 3.6).

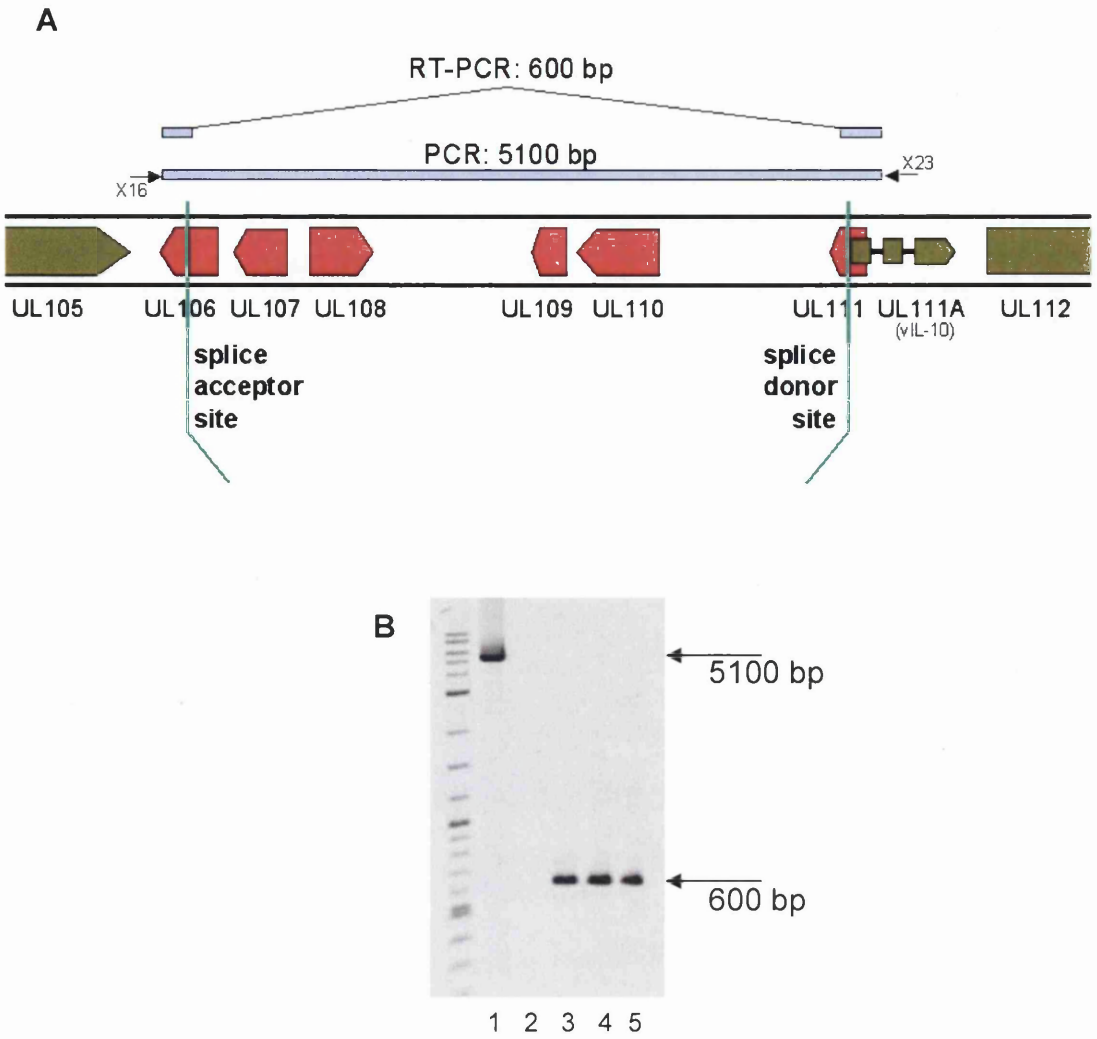


Figure 3.9. Splicing in region X.

(**A**) Positions of the splice donor and acceptor sites in region X. The blue bars show the approximate sizes of the PCR and RT-PCR products amplified using primers X16 and X46 (Appendix). **(B)** PCR and RT-PCR across region X. Lane 1 shows the PCR product generated from a genomic DNA template. Lanes 2 to 5 show the RT-PCR products generated from a polyA$^+$ RNA template (MI, IE, E and L respectively). (**C**) The conserved splice sites in HCMV and CCMV, donor and acceptor respectively, where y is a pyrimidine residue, r is a purine residue and n is any residue. The strand corresponding to the transcript is shown.

Comparative studies have shown that UL106-UL111 do not have counterparts in CCMV (Davison *et al.*, 2003), and that there is limited sequence identity between the two genomes in region X. However, splice donor and splice acceptor sites in UL111 and UL106, respectively, are conserved between HCMV and CCMV (Figure 3.9), and in all of the sequenced HCMV strains. These splice sites are potentially involved in production of the 4.6 kb and 1.1 kb RNAs identified by northern blotting. Primers X16 and X23 (Appendix) were designed in the predicted exons of the transcript for PCR amplification across the region, and across the predicted intron. Using these primers and AD169 DNA as a template, a product of approximately 5.1 kbp was amplified (Figure 3.9). RT-PCR reactions also used the same set of primers, with a template of AD169 IE, E or L polyA$^+$ mRNA, and a negative control of MI polyA$^+$ mRNA. RT-PCR products of approximately 600 bp were amplified using the IE, E and L templates (Figure 3.9). The amplification of a smaller product from polyA$^+$ mRNA than from genomic DNA suggested that a processed transcript had been detected, minus an intron removed by splicing. Sequencing of the RT-PCR products confirmed that the transcript detected was spliced, with a large intron of 4528 bp excised from the splice sites postulated from the sequence comparisons.

On identification of this novel spliced RNA, a detailed literature search of previously identified HCMV spliced RNAs was carried out. Consequently, it was found that a study by Rawlinson *et al.* (1993) had sequenced a single clone derived from L RNA, identifying a 1.1 kb RNA with a large intron of 4528 nt. This RNA was termed R160461. The splice junction identified by Rawlinson *et al.* (1993) is congruent with that in the present study. The ends of the 1.1 kb RNA were subsequently determined in the present study by RACE analysis to verify if these were also congruent with the ends of R160461 (Section 3.5).

RT-PCR was also used to confirm splicing of the UL111A (vIL-10) transcript. Amplification reactions employed a primer at the start of the UL111A transcript (X57, Appendix) and a primer from the SMARTRACE kit (Section 2.1.9) that anneals to the poly(A) tail at the 3'-end of the transcript. Sequencing of the RT-PCR products identified fully spliced UL111A transcripts, in agreement with the results previously reported by Kotenko *et al.* (2000), partially spliced transcripts (in which only the first intron was excised) and unprocessed transcripts (in which no splicing had occurred).

3.5 Transcripts in region X: RACE analysis of 3'- and 5'-ends

To characterise the transcripts in region X identified by northern blotting and RT-PCR, RACE was performed using a combination of two different kits to identify the 3'- and 5'-ends: SMARTRACE and GeneRacer (Sections 2.2.26 and 2.2.27, respectively). RACE reactions used a template of polyA$^+$ L mRNA harvested from total infected cell RNA (Section 2.2.20). The primers used to map the ends of the 1.1 kb transcript are shown in Figure 3.10. 3'-RACE products of approximately 450 bp were amplified using primer X48 (Figure 3.10). To map the 5'-end of the transcript, RACE products were amplified using five different primers. Using primers X57, X66 and X28, 5'-RACE products of approximately 850, 710 and 300 bp were amplified (Figure 3.10). No products were amplified using primers X53 and X50, suggesting that these primers were either directly adjacent to or upstream from the 5'-end. Each RACE-PCR product was gel purified and cloned, and approximately 10 clones were sequenced for each to ensure that an accurate 3'- or 5'-end was obtained.

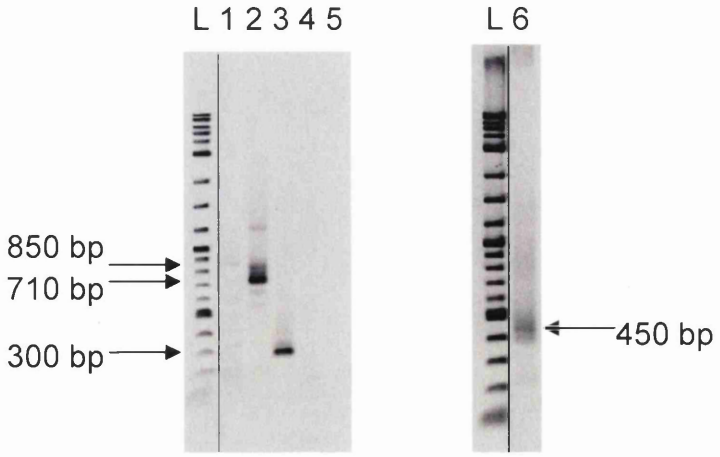


Figure 3.10. 3'- and 5'-RACE amplification of the ends of the 1.1 kb RNA.

Lanes 1 to 5 are 5'-RACE products amplified using primers X57, X66, X28, X53 and X50 respectively. Lane 6 is a 3'-RACE product for the 1.1 kb RNA, using primer X48. Lanes L are DNA markers.

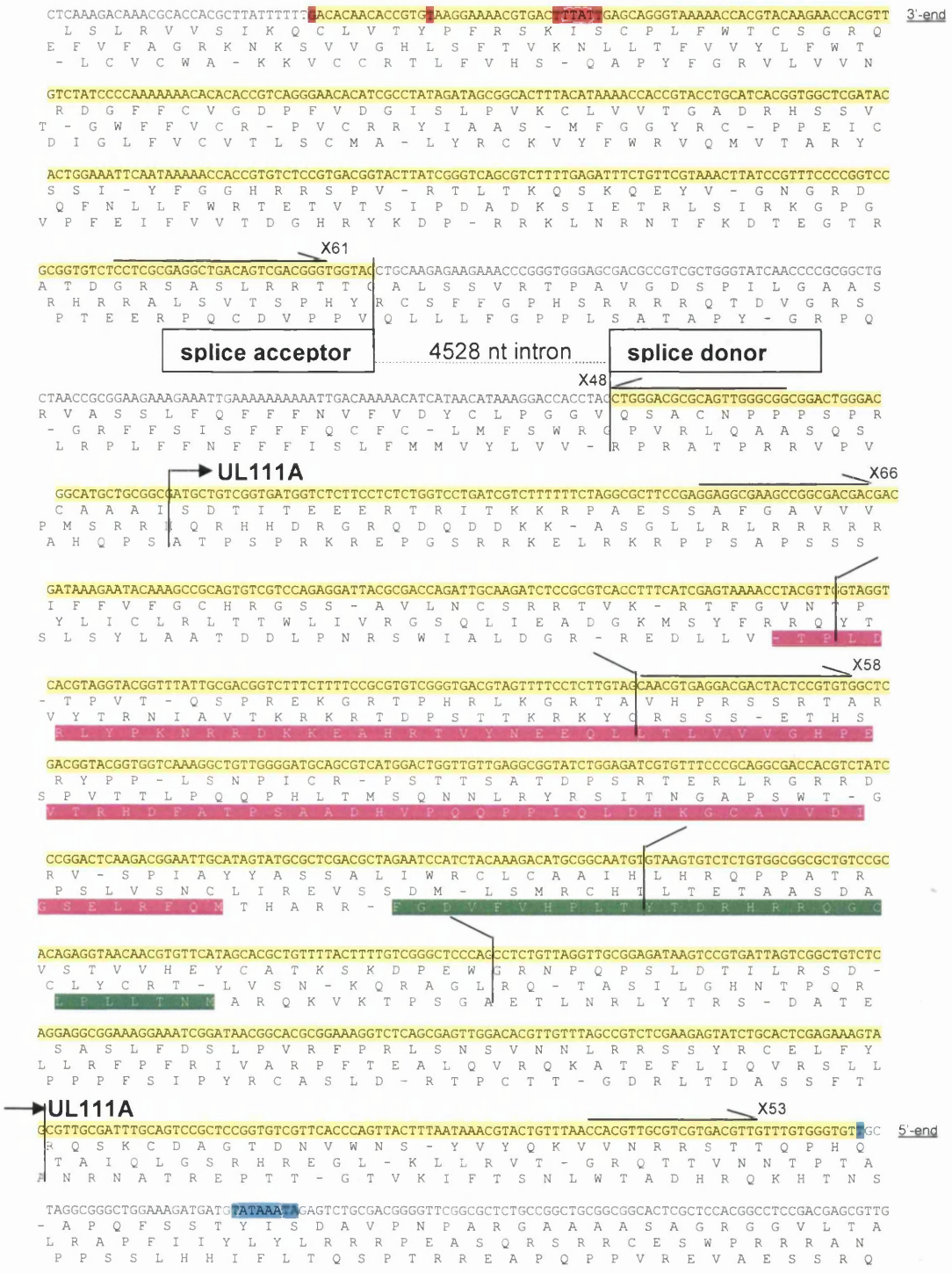


Figure 3.11. Three-frame translation of the spliced 1.1 kb RNA in AD169.

Yellow shows exon sequence of the 1.1 kb RNA. The splice donor and acceptor sites are separated by a 4528 nt intron. Green shows a 27 aa protein starting on the first ATG codon and purple shows the 78 aa protein (the largest coding region starting on an ATG). At the 3'-end, red shows the poly(A) signal (AATAAA) and the two downstream RNA cleavage sites. At the 5'-end, blue shows the T residue mapped at the 5'-end and the upstream TATA box (TATTTATA). Lines indicate the UL111A ORF and the internal splice sites. Also shown are primers X61, X48, X58 and X53, used for RACE-PCR.

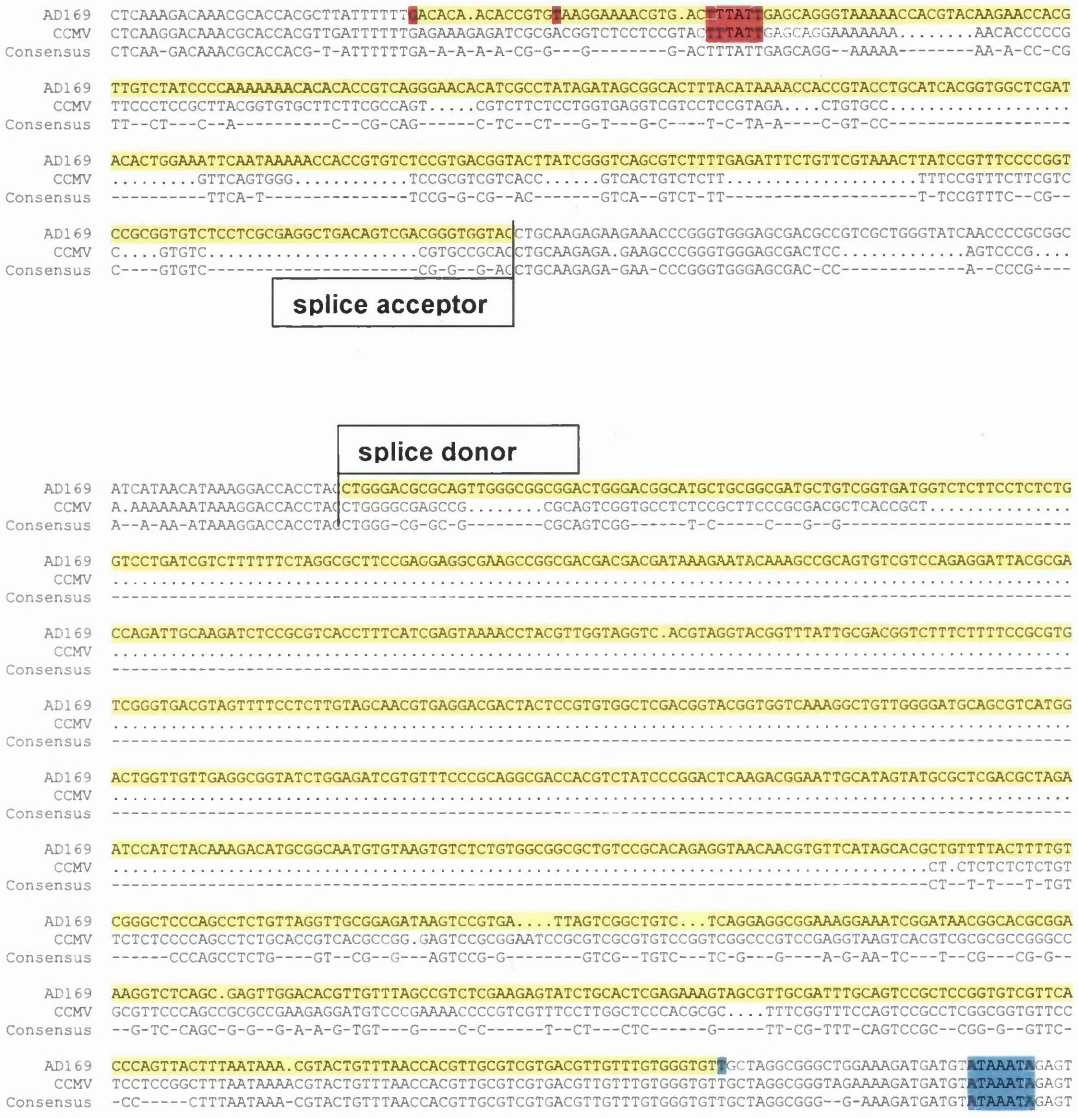


Figure 3.12. Signals for the 1.1 kb HCMV RNA are conserved in CCMV.

An alignment of AD169 and CCMV sequences. Yellow shows the exons of the AD169 1.1 kb RNA. Red shows the mapped 3'-ends of the RNA and the consensus poly(A) signal (AATAAA) conserved in CCMV. Blue shows the mapped 5'-end of the RNA and the TATA box (TATTTAT) conserved in CCMV.

The 3'-end of the 1.1 kb transcript mapped to two sites: a T residue at nucleotide position 155,740 in AD169, and a G residue at 155,726. These cleavage sites were 20 bp and 34 bp, respectively, downstream from a consensus poly(A) signal AATAAA (Figure 3.11). The 3'-end identified by Rawlinson *et al.* (1993) corresponds to the second of these two RNA cleavage sites. The 5'-end mapped to a T residue at 161,392, 830 bp upstream of the splice donor site, the same as the end mapped by Rawlinson *et al.* (1993) (Figure 3.11). A TATA box was identified 25 bp upstream from the 5'-end of the transcript. The two exons combine to make a transcript of approximately 1.14 kb. Potential translation products of the transcript in strain AD169 in the three reading frames are shown in Figure 3.11.

On characterising the 1.1 kb RNA, Rawlinson *et al.* (1993) identified a 31 aa protein, crossing the splice site, as the likely product encoded by the mRNA. However, this does not use an in-frame ATG initiation codon. The first ATG initiation codon in the mRNA specifies a 27 aa protein (highlighted in green in Figure 3.11) and the largest product would be a 78 aa protein (highlighted in purple in Figure 3.11). Two frameshifts, potentially affecting any protein coding region, were found towards the 3'-end of the mRNA. Neither of these frameshifts affect the 27 aa or the 78 aa coding regions, but the 31 aa region is disrupted. None of these three amino acid sequences is significantly similar to any proteins in the NCBI database.

Sequence comparisons show that the poly(A) signal at the 3'-end, the splice site, and the TATA box at the 5'-end of the 1.1 kb RNA are also conserved in CCMV (Figure 3.12). This suggests that a similar RNA, of approximately 0.5 kb with a 5 kb intron, would be transcribed from the corresponding region in CCMV. However, within this putative CCMV RNA, there are no ATG codons (Figure 3.12).

3.6 Discussion

It is unlikely that ORFs UL106-UL111, assigned to region X in the original analysis of HCMV strain AD169 (Chee *et al.*, 1990), actually encode proteins. Only UL106 and UL109 commence with ATG codons. The other ORFs are preceded by

termination codons, with UL107, UL108 and UL110 containing initiation codons internally and UL111 lacking such codons (Figure 3.1). Sequence comparisons of seven HCMV strains in region X showed that these ORFs, with the exception of UL108, are not intact in all HCMV strains, being disrupted by frameshifts in comparison with AD169 (Figure 3.4). At many of the frameshifts, differences between the two clinical (unpassaged) samples, W and 3301, were observed. Moreover, no RNAs specified by these ORFs were detected in previous transcript mapping studies of the region, and none were detected in the present study. Of the three ORFs identified by Murphy *et al.* (2003b), only C-ORF17 is frameshifted, in strains 3301, W, 6397 and BAC-cloned strain TR. C-ORF16 and C-ORF18 remain intact, although neither has a counterpart in CCMV. Furthermore, ssRNA probes D and I, used for northern blot analysis in the present analysis, did not detect transcripts corresponding to these ORFs (Figure 3.7). Probes D and I should detect transcripts C-ORF16 and C-ORF18 respectively, based on the positions of poly(A) signals in region X. The three ORFs in region X unaffected by frameshifts (UL108, C-ORF16 and C-ORF18) are small compared to other HCMV protein-coding ORFs, potentially encoding proteins of only 46 aa, 90 aa and 84 aa, respectively, and none of these are similar to other proteins in the NCBI database. In addition to the frameshifts identified in UL106-UL111, other frameshifts occurred throughout the region and no novel ORFs emerged as potentially encoding proteins from the sequence comparisons. The known functional ORFs (UL105, UL111A and UL112) were conserved.

Region X has a distinct nucleotide composition, consisting of local regions of high G+C content embedded in sequences of low G+C content (Figure 3.13). The regions of high G+C content (and therefore where termination codons are rarer) tend to coincide with UL106-UL111. Overall, the G+C content of the region falls well below the average level for the rest of the genome. Similar patterns of unusual nucleotide composition are exhibited in the corresponding regions of other CMV genomes, and to a lesser extent in HHV-6 and HHV-7 (Figure 3.13). In addition, polynucleotide A:T tracts of eight or more residues are more common in region X than in the rest of the genome (Table 3.1).

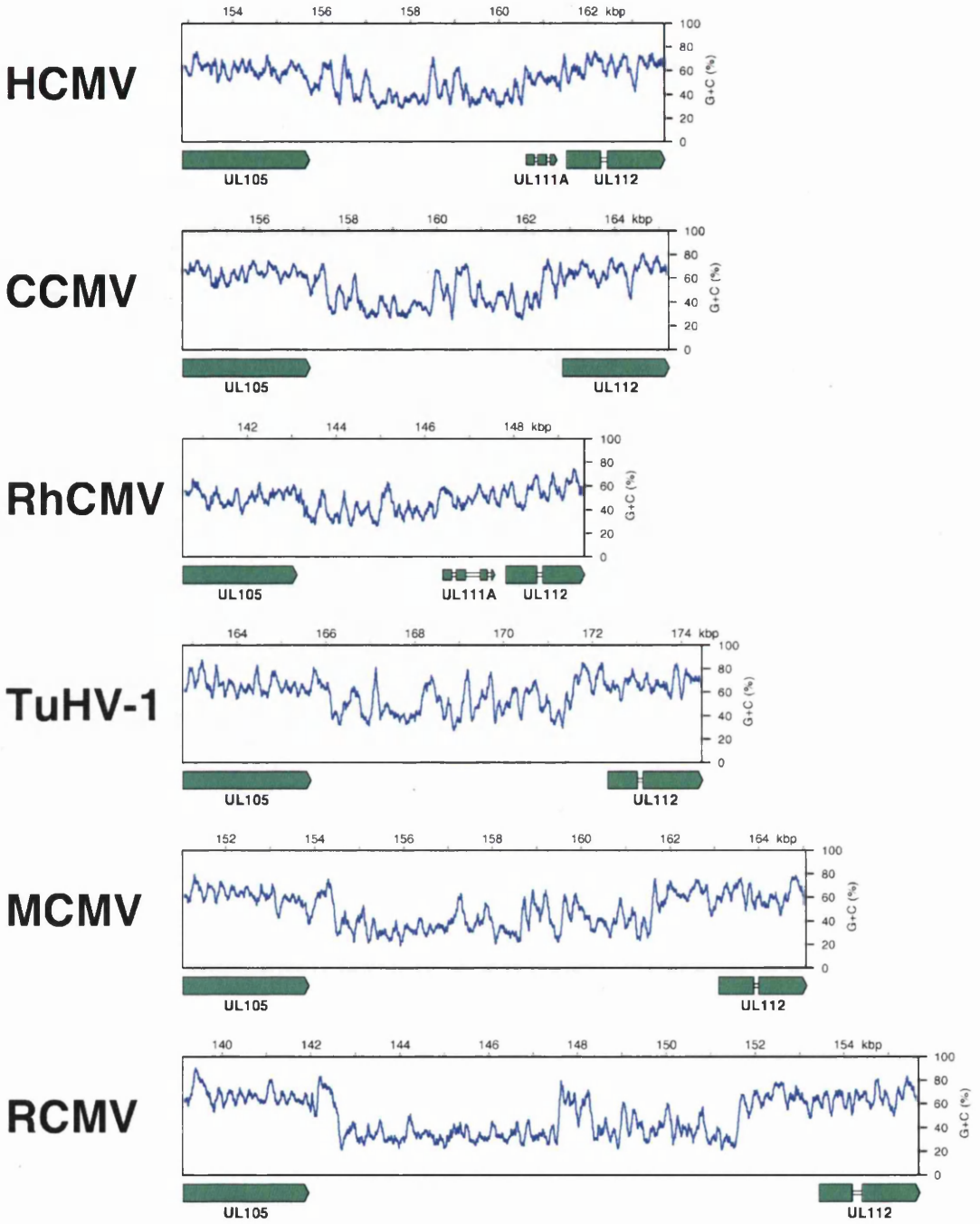


Figure 3.13. The nucleotide composition of region X and corresponding regions in other CMVs.

The plots were computed using Window (100 nucleotides shifted by 3 nucleotide increments) and displayed using Statplot. Known genes UL105, UL111A (present only in HCMV and RhCMV) and UL112 are marked on each plot.

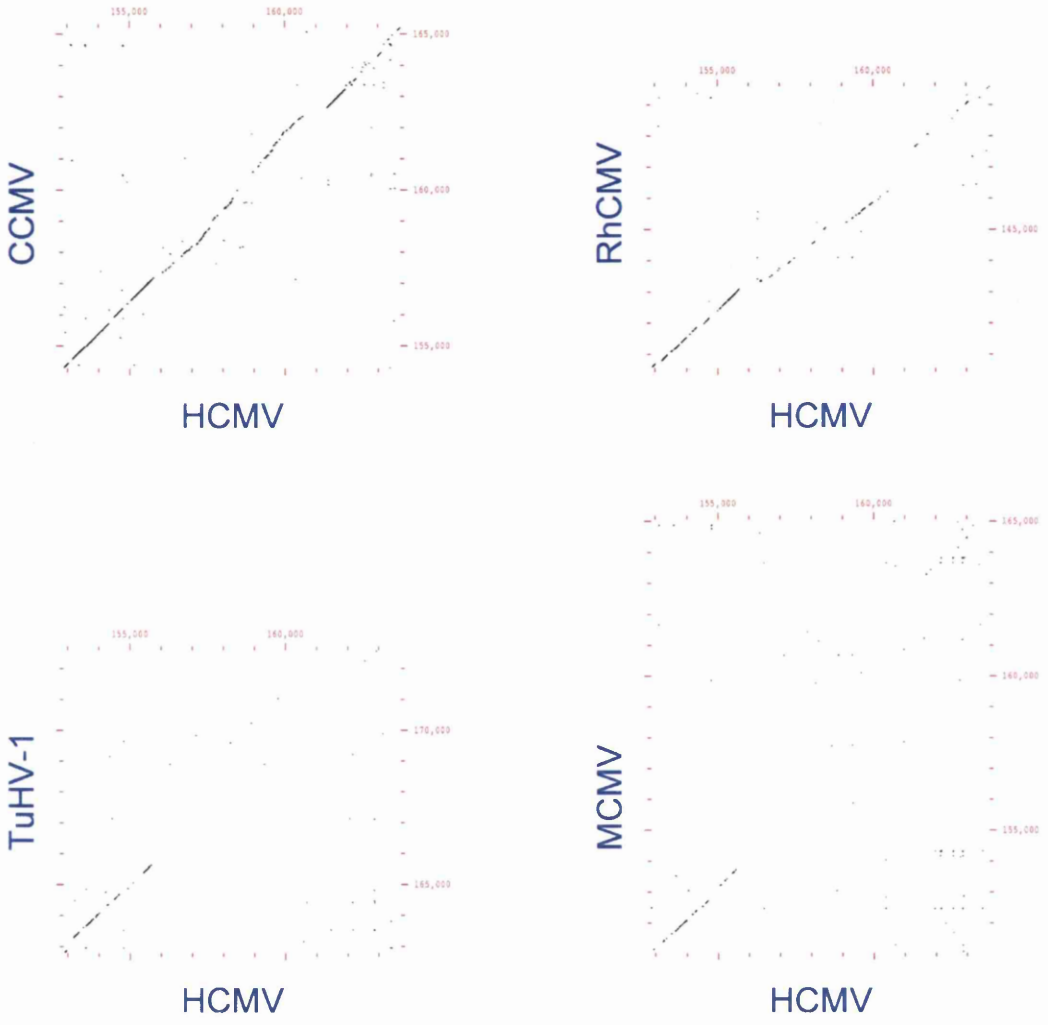


Figure 3.14. Sequence comparisons of region X between HCMV and other CMVs.

Comparisons were computed using Compare (window of 21 nucleotides with a minimum match of 17 nucleotides) and displayed using Dotplot.

Sequences corresponding to region X are present in other CMVs, where they are similar in size and position in the genome. DNA sequence comparisons between HCMV and other CMV sequences are shown as matrix plots in Figure 3.14. The genomes of HCMV and CCMV (the closest known relative of HCMV) are only moderately diverged and are essentially collinear (Davison *et al.*, 2003; discussed in Section 1.20). None of the ORFs predicted to be in region X (UL106-UL111 and C-ORFs 16-18) are conserved in CCMV, and the regions share limited sequence similarity. In addition, a distinct discontinuity is apparent near the right end of the region, where CCMV lacks a homologue of HCMV UL111A (vIL-10). Sequences corresponding to region X can also be identified in the more distantly related genomes of RhCMV (approximately 3.4 kbp), MCMV (9.2 kbp), RCMV (11.5 kbp) and TuHV1 (6.7 kbp) (Figure 3.13). When compared with HCMV, these genomes are extremely divergent in region X, much more so than in the flanking genes (UL105 and UL112).

Transcript analysis identified a spliced 1.1 kb RNA and its 4.6 kb intron as the only RNAs detected from region X (Figure 3.15). The splice sites, 3'-poly(A) signal, and 5'-TATA box of the 1.1 kb RNA are conserved in the corresponding region in CCMV, suggesting that a similar RNA with a large intron should be expressed in CCMV. However, it is unlikely that the RNA encodes a functional protein, either in HCMV or in CCMV. In HCMV, the RNA has the potential to encode proteins of 27 aa (utilising the first in-frame ATG initiation codon) or 78 aa (utilising the third in-frame ATG codon). However, neither of these are conserved in the potential CCMV reading frame, which contains no ATG initiation codons.

The 4.6 kb transcript identified by northern blot analysis in the present study corresponds to the 5 kb RNA originally characterised by Plachter *et al.* (1988). The 5'-end of the 5 kb RNA was mapped to a residue four bases from the splice donor site identified in the current study (Figure 3.11). The 3'-end was mapped to two sites downstream from a poly(A) signal at nucleotide position 155,759 in AD169. This is the same poly(A) signal subsequently identified at the 3'-end of the 1.1 kb RNA (Figure 3.11). The nucleotide sequence published by Plachter *et al.* (1988) was compared to that of AD169 in region X from the present study. The comparison identified a 211 bp deletion in the strain originally used to map the 5 kb RNA, from

156,008 to 156,219. This deletion included the splice acceptor site, and thus would compromise processing of the 1.1 kb RNA. This explains why some of the observations and conclusions made by Plachter *et al.* (1988) conflict with the current study.

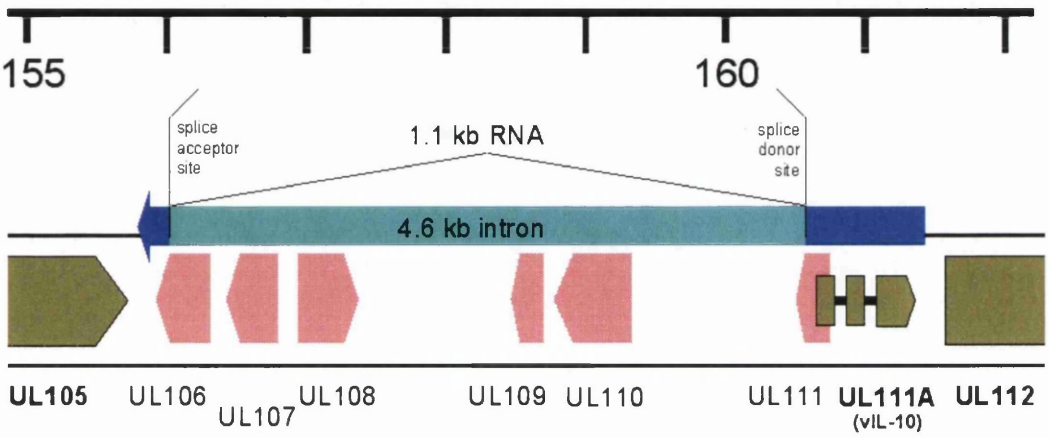


Figure 3.15. A model of the spliced 1.1 kb RNA and its 4.6 kb intron in region X. The exons, at either end of the region, are joined via the splice sites shown in the figure to create a 1.1 kb RNA, and the 4.6 kb intron is consequently excised.

A study of the 5 kb RNA was recently published, investigating its role as an intron (Kulesza and Shenk, 2004). The authors showed that the transcript was an intron, derived from the processing of a precursor RNA. In transfection experiments, the 5 kb RNA was shown to be spliced from a heterologous reporter gene in transfection. Northern blot analysis confirmed that the majority of the 5 kb RNA is found in the polyA$^-$ fraction of IE and L RNA. A viable mutant was constructed with a mutation in the splice donor site of the RNA, compromising processing of the 1.1 kb RNA. Total RNA prepared from cells infected with this viral mutant did not express the 4.6 kb or the 1.1 kb RNA (Kulesza and Shenk, 2004).

It is unlikely that there are protein-coding ORFs in region X, and the only RNAs identified in the region, a 1.1 kb spliced RNA with a 4.6 kb intron, are also unlikely to encode functional proteins. However, corresponding regions are found in other CMVs, and one study has noted that the expression of a large transcript can be detected in RhCMV and MCMV (Kulesza and Shenk, 2004). Also, sequence comparisons in the present study suggest that a similar RNA with a large intron is expressed from the corresponding region of CCMV. It seems likely that the RNAs produced from region X of HCMV have a function in the virus life cycle that does not require translation.

CHAPTER 4: REGION O

4.1 Protein-coding genes

The second large region in HCMV not thought to contain protein-coding genes is located between ORFs UL57 and UL69. This region is here termed region O. The analysis of AD169 predicted 11 ORFs in region O, UL58-UL68 (Chee *et al.*, 1990). The ORFs flanking region O, UL57 and UL69, have been characterised, and a transcript related to UL65 has been identified, but there is no evidence for expression of proteins by UL58-UL64 and UL66-UL68. Three further novel, protein-coding ORFs have been predicted in region O by the analyses of Murphy *et al.* (2003a; 2003b), which are discussed in Section 1.18. From the list of 39 newly recognised ORFs with the potential for coding proteins, three are in region O: ORF3, ORF4 and ORF5. Part of region O has been shown to have an essential viral *cis*-acting function as the origin of lytic replication (*ori*Lyt). Features of the origin region and its boundaries are discussed in Section 4.2.

At the left end of region O, UL57 specifies transcripts of 10 and 14 kb in infected cell RNA, and a 17 kb RNA of low abundance (Pari and Anders, 1993; Smuda *et al.*, 1997; Kiehl *et al.*, 2003). The 10 kb transcript was detected in cycloheximide-treated infected cells, but not in the presence of anisomycin, a more effective inhibitor (Kiehl *et al.*, 2003). Accordingly, the RNA transcribed is consistent with early kinetics of expression. Three distinct 5'-ends were identified, spread over 800 bp (Kiehl *et al.*, 2003). Two of these sites contain upstream sequences which function as promoters, responsive to HCMV infection in transient assays. The third promoter was predicted, by sequence homology, to be regulated by specific promoter elements often present in cellular, usually TATA-less promoters, but not previously found in HCMV (Burke and Kadonaga, 1997; Kiehl *et al.*, 2003). These sites constitute a large, complex promoter region with novel features which control UL57 expression. Sequences have also been identified which regulate UL57 transcription and *ori*Lyt activation (Kiehl *et al.*, 2003). UL57 encodes an early, nuclear, single-stranded DNA-binding protein of about 140 kDa, which is one of 11 trans-acting factors required for transient complementation of *ori*Lyt-dependent DNA replication (Pari and Anders, 1993).

At the right end of region O, UL69 specifies a complex pattern of transcripts that changes during the HCMV replicative cycle. Two transcripts, of 2.7 and 3.5 kb, classified as E-L genes, could be detected as early as 7 h p.i. but not in the presence of CHX (Winkler *et al.*, 1994). UL69 encodes a regulatory protein which exerts a broad stimulatory effect on gene expression and is required for efficient HCMV replication (Hayashi *et al.*, 2000). The UL69 protein, which is present in dense bodies, was not detected within the membrane fraction of purified virions after treatment with detergent, suggesting that the protein is a tegument constituent (Winkler and Stamminger, 1996). A phosphorylated form of the protein also contained in the virion has been identified by Western blot (Winkler and Stamminger, 1996). In cell culture, UL69 is able to arrest cell cycle progression by causing different cell types, including HFFF cells, to accumulate within the G1 compartment of the cell cycle (Lu and Shenk, 1999; Hayashi *et al.*, 2000). In the presence of UL69, even the introduction of serum fails to induce progression of quiescent HFFF cells into S phase.

A transcript outwith the core *ori*Lyt region is also known to be expressed from region O. The late RNA pp67 is routinely detected in nucleic acid sequence-based amplification (NASBA) assays used to monitor the progression of HCMV infection and the effect of antiviral therapy on viral activity (Blok *et al.*, 1998; Gerna *et al.*, 1999). The initial study by Davis *et al.* (1984) also detected an associated 67 kDa protein in strain Towne. Although the pp67 RNA is routinely detected in NASBA assays, the study by Davis *et al.* (1984) remains the only one to characterise this protein. The potential for this RNA to encoded a protein will also be investigated in the present study by sequence comparisons of seven HCMV strains.

Of the other ORFs originally defined in region O, only UL59, UL67 and UL68 commence with ATG initiation codons. The other ORFs are preceded by termination codons, with UL60-UL62 and UL64-UL66 containing in-frame ATG codons internally and UL58 and UL63 lacking such codons (Figure 4.1).

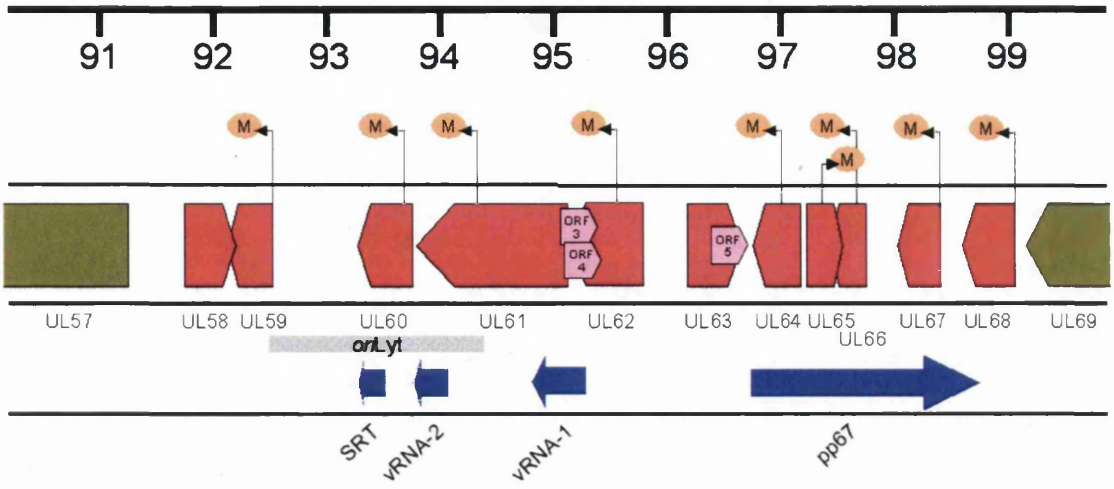


Figure 4.1. Map of the characterised genes and predicted genes in region O.

The scale at the top shows the genomic location (kbp) in the AD169 genome. Dark green shows parts of ORFs of known functions. Red shows ORFs predicted by Chee *et al.*, 1990 that have no known function and have been discounted because they have no counterparts in CCMV (Davison *et al.*, 2003a). Pink shows novel ORFs predicted by Murphy *et al.* (2003a, 2003b). The 3'-end of each ORF has a point to show its orientation, and the M shows the first in-frame ATG initiation codon in each red ORF. ORFs 3,4 and 5 all commence on ATG initiation codons. Dark green and pink ORFs all start with ATG initiation codons. The core region of the HCMV origin of lytic replication (*ori*Lyt) is shown as a grey bar and the three characterised RNAs in the origin are shown as blue arrows (Zhu *et al.*, 1998; Huang *et al.*, 1996; Prichard *et al.*, 1998). A fourth transcript in the region, pp67, has also been previously mapped in AD169 (Scott *et al.*, 2002).

4.2 Origin of lytic DNA replication (*ori*Lyt)

After the original HCMV analysis of Chee *et al.* (1990), a candidate origin of lytic replication was identified in SCMV and HCMV by assessing the ability of cloned restriction fragments to mediate DNA replication after transfection into HFFFs when required trans-acting factors were supplied by infection (Anders and Punturieri, 1991; Hamzeh *et al.*, 1990). In SCMV strain Colburn, sequence analysis identified four distinct domains containing different repeat elements. Clustered sets of repeat elements, similar to those in SCMV, were also identified in HCMV *ori*Lyt (Anders *et al.*, 1992). Masse *et al.* (1992) identified a region between nucleotide positions 93,000 and 94,400 in AD169, residing within *ori*Lyt, which contains the greatest

content of inverted and direct repeats in HCMV. Various studies, using transient replication assays, identified a core region of *ori*Lyt that contains elements required for *ori*Lyt function, including multiple copies of elements similar to known transcription factor binding sites, but requires flanking sequences for full activity (Masse *et al.*, 1992; Anders *et al.*, 1992; Zhu *et al.*, 1998). The core region of *ori*Lyt shown in Figure 4.1 was determined to be between 92,682 and 94,230 from the most recent analysis (Zhu *et al.*, 1998). Cotransfection assays of *ori*Lyt replication in HCMV cosmids identified eleven ORFs which encode trans-acting factors sufficient for transient complementation of *ori*Lyt-dependent DNA replication (Pari and Anders, 1993). UL57, which is to the left of region O, is one of these essential ORFs in which engineered frameshift mutations prevent complementation of DNA replication.

Transcript analysis studies identified at least four transcripts that cross or originate within *ori*Lyt, but only one of these has been characterised (Huang *et al.*, 1996). The smallest replicator transcript (SRT) is an RNA of approximately 0.25 kb, localised to approximately the centre of the core region of *ori*Lyt by northern blot analysis. The RNA was detected as early as 2 h p.i. but not in the presence of CHX. A single 5'-end was identified, and sequence immediately upstream of SRT was shown to function as a promoter, responsive to HCMV infection in transfection assays (Huang *et al.*, 1996). SRT is not polyadenylated and was shown to have heterogeneous 3'-ends which overlap an oligopyrimidine sequence shown by mutation analysis to be essential to *ori*Lyt replicator function (Huang *et al.*, 1996). SRT is likely to play a role in initiating or regulating HCMV lytic-phase DNA synthesis. Two more RNAs, of approximately 300 and 500 bp, were identified in *ori*Lyt by Prichard *et al.* (1998). These were termed vRNAs (virus-associated RNAs) from their association with packaged virion DNA. They were mapped to positions within the origin region, as shown on Figure 4.1. One of these, vRNA-2, was identified as a RNA-DNA hybrid and mapped within the core *ori*Lyt, on the same strand as SRT (Prichard *et al.*, 1998).

4.3 Sequence comparisons

As discussed in Section 3.3, sequence comparisons of multiple HCMV strains can be used to identify ORFs that are conserved or disrupted by insertions or deletions.

Region O was sequenced in seven HCMV strains: high passage strain AD169, low passage strains Toledo, Merlin, 3157 and 6397, and strains obtained directly from human tissue: W and 3301 (Section 2.1.1). Initially, six different PCR product were amplified for each strain (PCR products 1,2 and 4-7, Table 4.1 and Figure 4.2). Larger PCR products proved difficult to amplify and clone, most likely because of the unusual nucleotide composition of the region and the presence of various repeat elements associated with *ori*Lyt.

PCR product	F primer	R primer	Size (kbp)	Strains amplified
1	O7	O13	1.4	A, T, M, 31, 63, 33, W
2	O54	O60	1	A, T, M, 31, 63, 33, W
3a	O78	O84	0.3	A, T, M, 31, 63, 33, W
3b	O3	O72	0.3	A, T, M, 31, 63, 33, W
3c	O86	O73	0.2	A, T, M, 63
3d	O76	O96	0.2	A
3e	O88	O93	0.25	A
3f	O96R	O101R	0.3	A
3g	O90	O91R	0.2	A
3h	O83	O95	0.55	A, T, M, 31, 63, 33
4	O21	O20	1	A, T, M, 31, 63, 33
5	O19	O18	1.4	A, T, M, 31, 63, 33, W
6	O17	O16	1.4	A, T, M, 31, 63, 33, W
7	O28	O6	1.3	A, T, M, 31, 63, 33, W

Table 4.1. PCR amplification of region O for sequence comparisons.

Each PCR amplification product was amplified by the primers shown (Appendix). The approximate size of each PCR fragment and the strains amplified for each set of primers is also shown: A is AD169, T is Toledo, M is Merlin, 31 is 3157, 63 is 6397, 33 is 3301, W is W. PCR products 3a-3h required a high-GC PCR kit for amplification (Section 2.2.3). The positions of these overlapping PCR fragments are shown in Figure 4.2.

PCR products 3a-3h (Table 4.1), which largely correspond to the core *ori*Lyt region, were especially difficult to obtain and could not be amplified for all the strains. Where products were obtained, they were amplified using a high G+C kit (Section 2.2.3) and sequenced using dGTP Big Dye, a sequencing reagent for high G+C templates. Sequence data across the entire region could only be generated for strain AD169, the rest of the strains containing gaps of approximately 500 bp (strains Merlin, Toledo and 6397), 600 bp (strains 3301 and 3157) or 1 kbp (strain W) in UL61 (Figures 4.2 and 4.3). Each PCR product was cloned into plasmids, and at least four plasmids were sequenced for each product. The sequence database was compiled from electropherograms using PreGap4 and Gap4 (Staden *et al.*, 2000) and Phred (Ewing and Green, 1998; Ewing *et al.*, 1998).

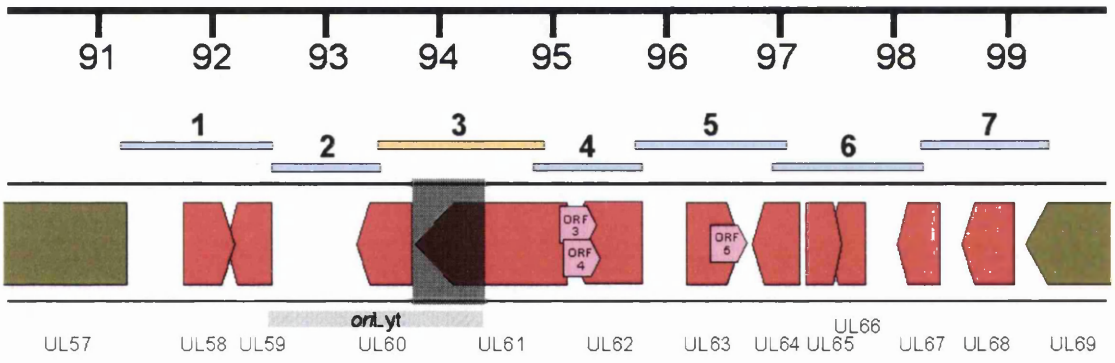


Figure 4.2. PCR amplification of region O for sequence comparisons.

Blue bars show the overlapping PCR fragments, amplified using the primers shown in Table 4.1. Fragment 3 corresponds to the region of high G+C content that was amplified in even smaller fragments (Table 4.1), and not in all strains. The dark grey box over UL61, which corresponds to the core region of *ori*Lyt, shows where complete sequence was only generated for AD169.

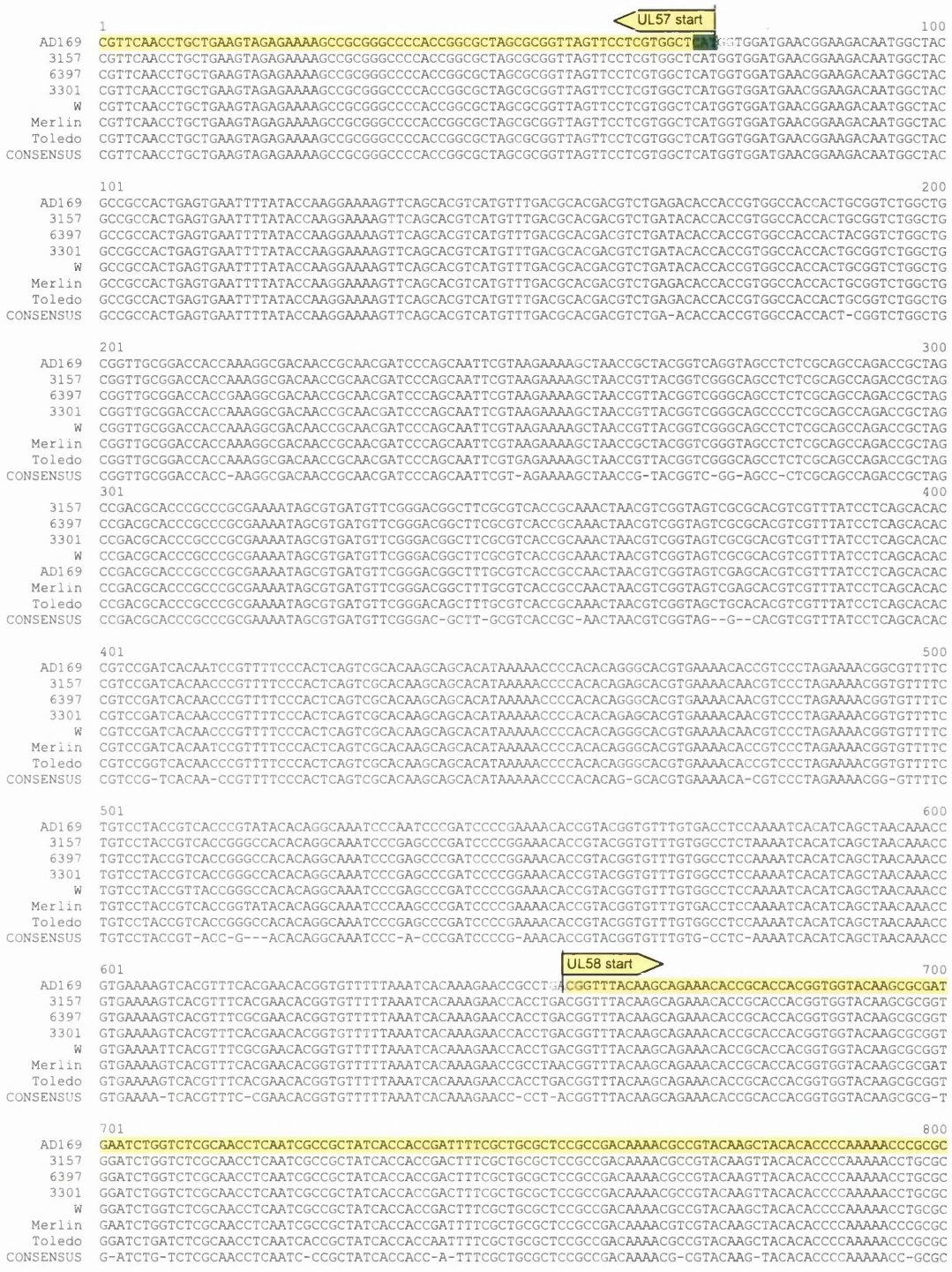
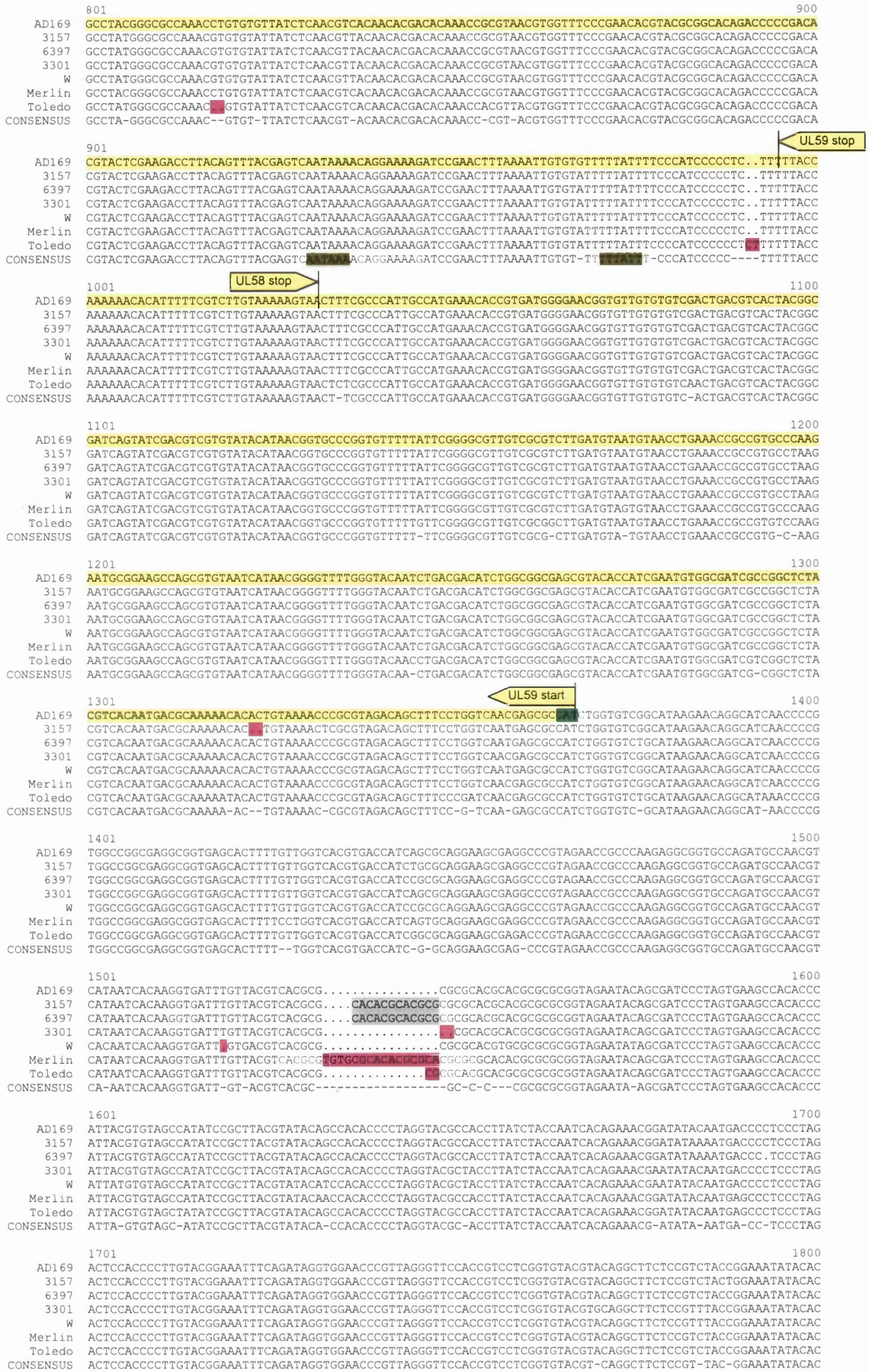
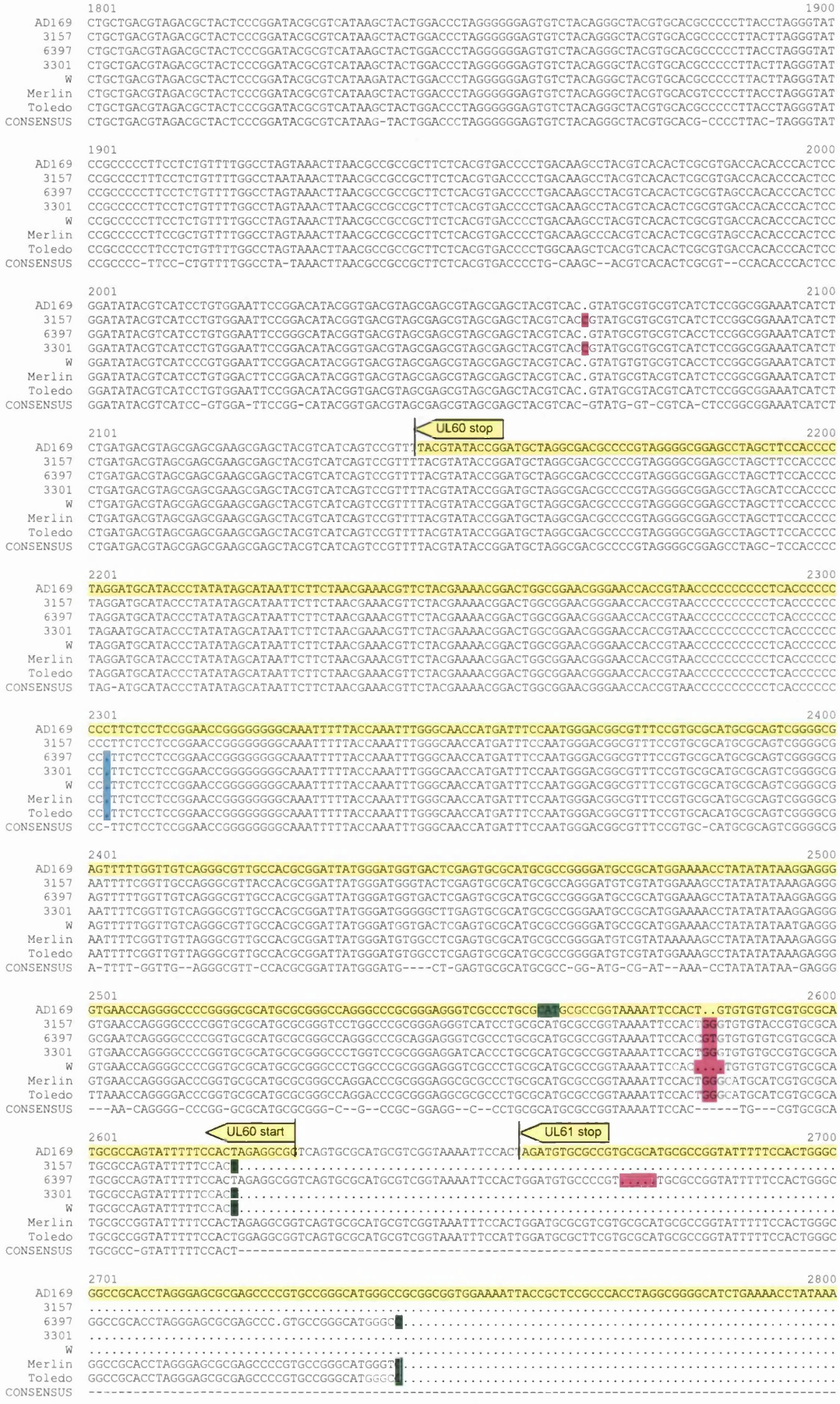
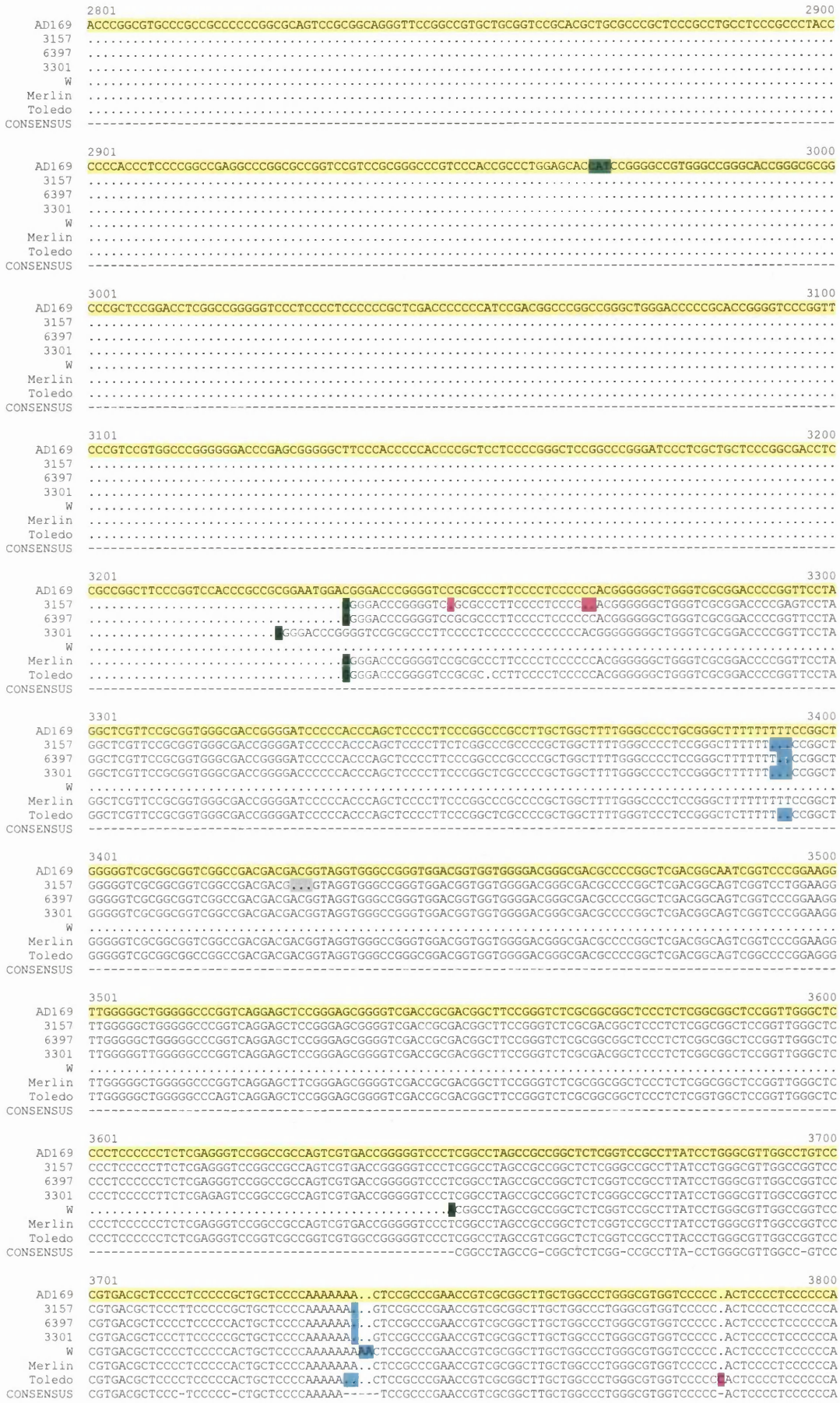


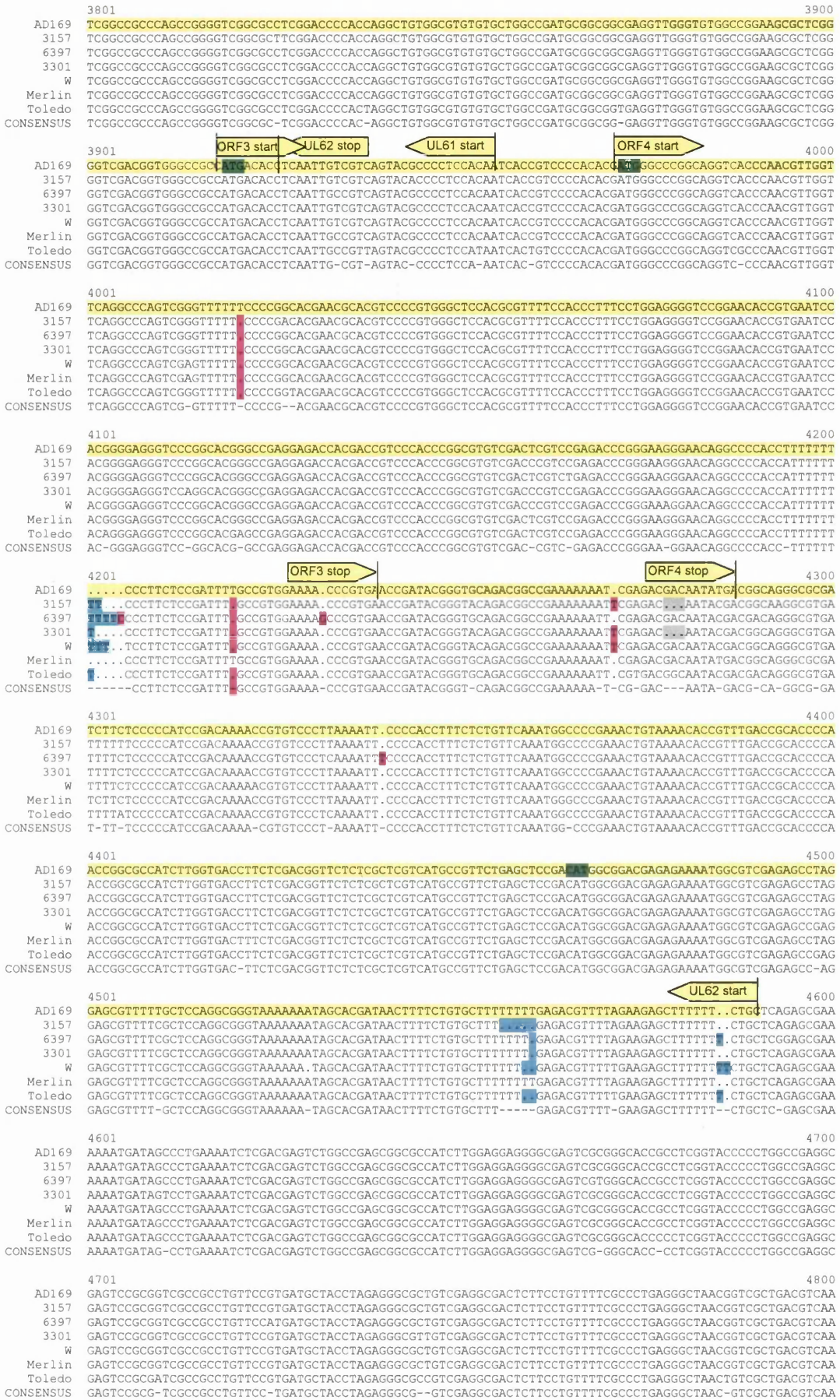
Figure 4.3. Sequence comparisons of seven HCMV strains in region O.

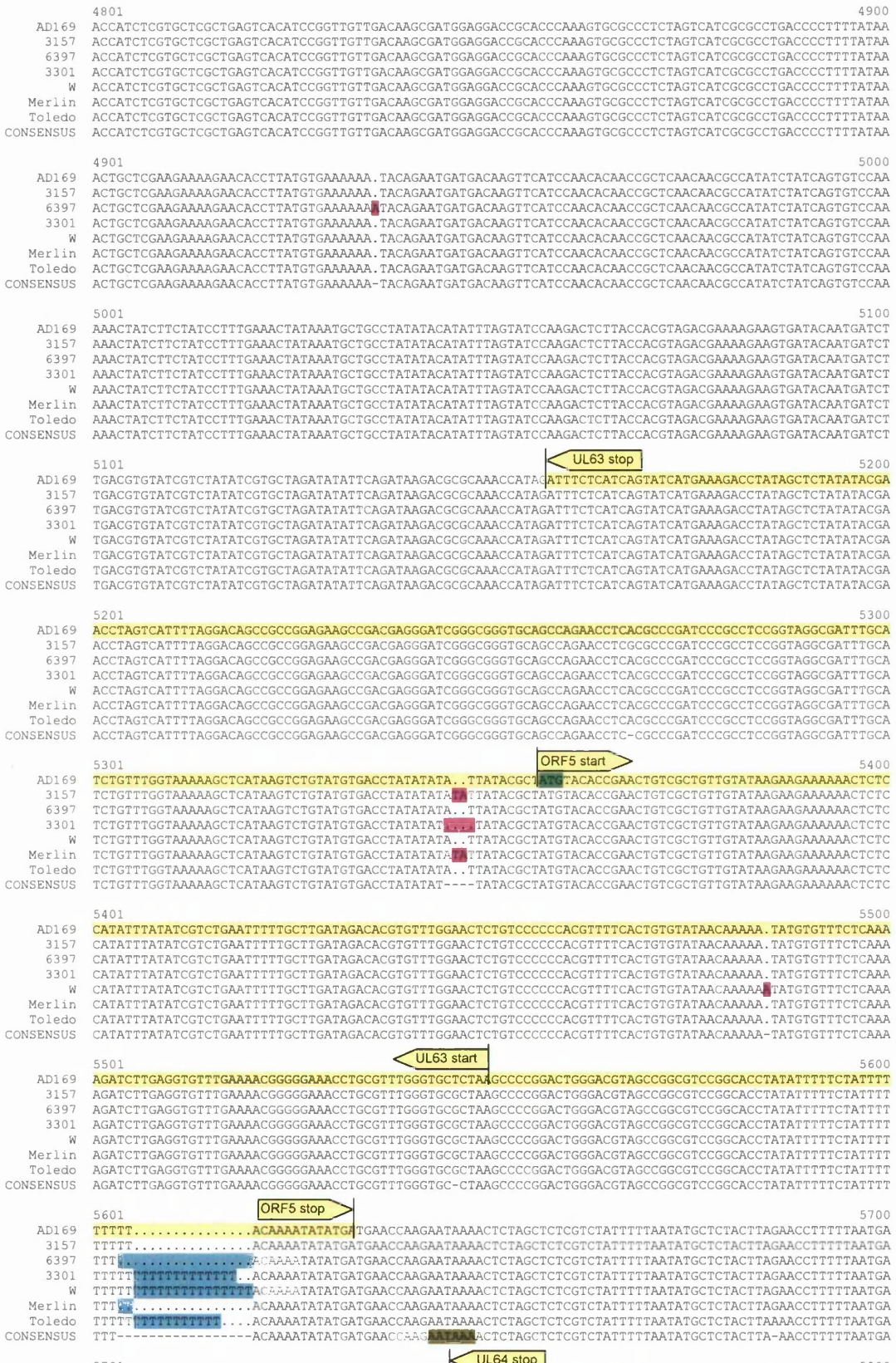
Alignments were made using ClustalW and displayed using Pretty. Yellow AD169 sequence shows all characterised and predicted ORFs in the region. The start and stop codons for each ORF are shown, the arrows indicating their orientation. Purple shows insertions/deletions that cause frameshifts in comparison to AD169. Sky blue shows variation in polynucleotide A:T or G:C tracts of eight or more residues that cause frameshifts in comparison to AD169. Grey shows in-frame insertions/deletions. Brown consensus residues show conserved potential polyadenylation signals (AATAAA). The figure is continued on the following pages.

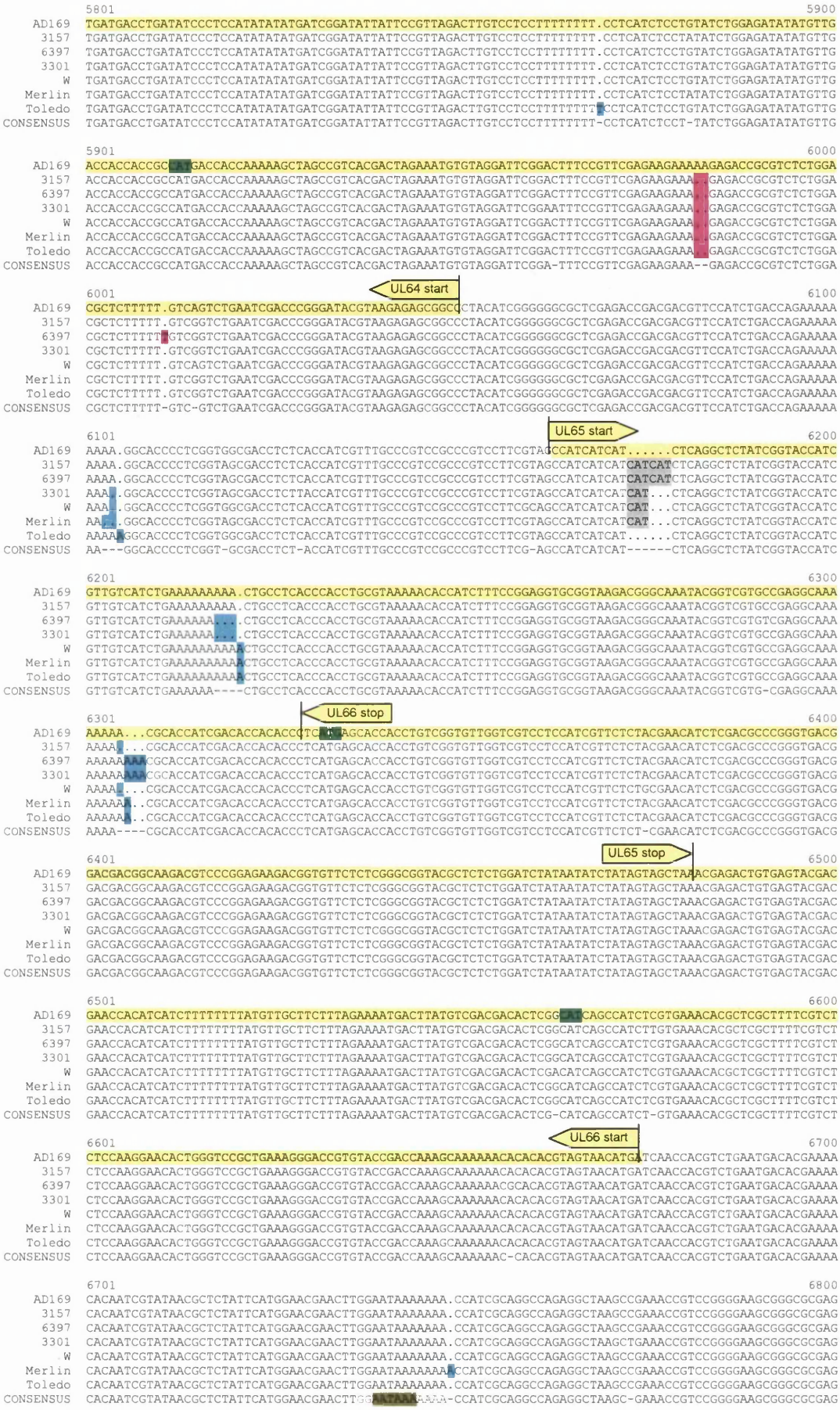

UL59 stop
UL58 stop
UL59 start

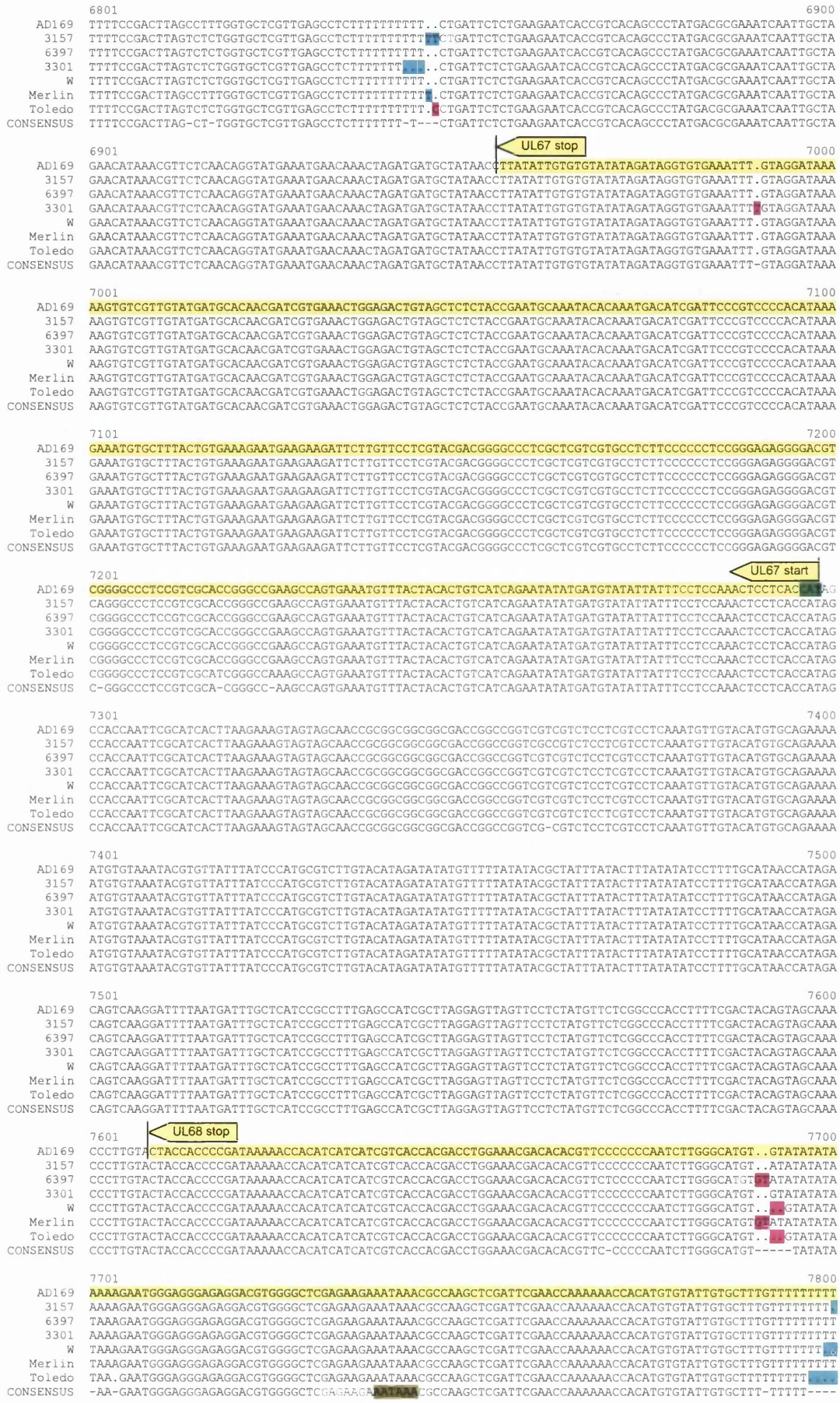

UL60 stop
UL60 start
UL61 stop

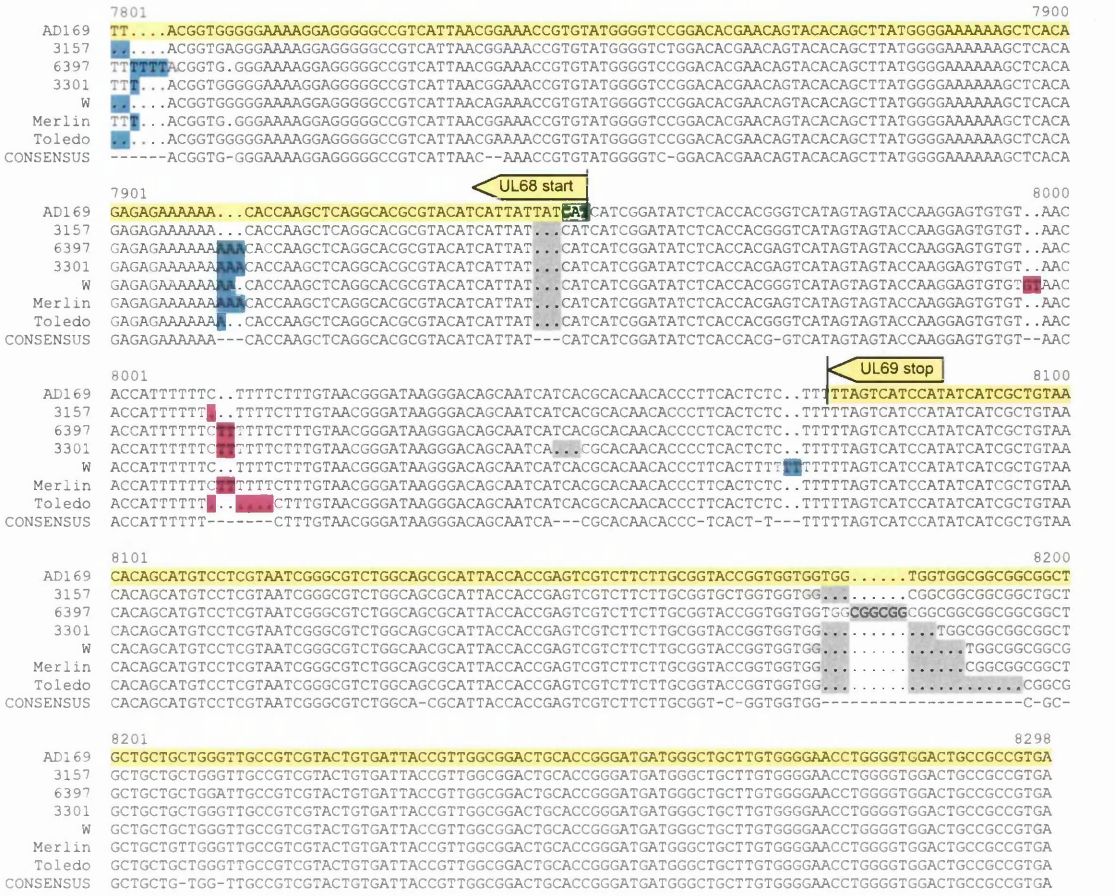

2801 2900
AD169 ACCCGGCGTGCCCGCCGCCCCCCGGCGCAGTCCGCGGCAGGGTTCCGGCCGTGCTGCGGTCCGCACGCTGCGCCCGCTCCCGCCTGCCTCCCGCCCTACC
3157 ..... 6397 ..... 3301 ..... W ..... Merlin ..... Toledo ..... CONSENSUS -----
2901 3000
AD169 CCCCACCCTCCCCGGCCGAGGCCCGGCGCCGGTCCGTCCGCGGGCCCGTCCCACCGCCCTGGAGCACCATCCGGGGCCGTGGGCCGGGCACCGGGCGCGG
3157 ..... 6397 ..... 3301 ..... W ..... Merlin ..... Toledo ..... CONSENSUS -----
3001 3100
AD169 CCCGCTCCGGACCTCGGCCGGGGGGTCCCTCCCCTCCCCCCGCTCGACCCCCCCATCCGACGGCCCGGCCGGGCTGGGACCCCCGCACCGGGGTCCCGGTT
3157 ..... 6397 ..... 3301 ..... W ..... Merlin ..... Toledo ..... CONSENSUS -----
3101 3200
AD169 CCCGTCCGTGGCCCGGGGGGACCCGAGCGGGGGCTTCCCACCCCCACCCCGCTCCTCCCCGGGCTCCGGCCCGGGATCCCTCGCTGCTCCCGGCGACCTC
3157 ..... 6397 ..... 3301 ..... W ..... Merlin ..... Toledo ..... CONSENSUS -----
3201 3300
AD169 CGCCGGCTTCCCGGTCCACCCGCCGCGGAATGGACGGGACCCGGGGGTCCGCGCCCTTCCCCTCCCCCCACGGGGGGGCTGGGTCGCGGACCCCGGTTCCTA
3157 .....GGGACCCGGGGTCGCGCCCTTCCCCTCCCCCACGGGGGGGCTGGGTCGCGGACCCCGAGTCCTA
6397 .....GGGACCCGGGGTCCGCGCCCTTCCCCTCCCCCACGGGGGGGCTGGGTCGCGGACCCCGGTTCCTA
3301 .....GGGACCCGGGGTCCGCGCCCTTCCCCTCCCCCCCCCCCACGGGGGGGCTGGGTCGCGGACCCCGGTTCCTA
W .....
Merlin .....GGGACCCGGGGTCCGCGCCCTTCCCCTCCCCCACGGGGGGGCTGGGTCGCGGACCCCGGTTCCTA
Toledo .....GGGACCCGGGGTCCGCGC.CCTTCCCCTCCCCCACGGGGGGGCTGGGTCGCGGACCCCGGTTCCTA
CONSENSUS -----
3301 3400
AD169 GGCTCGTTCCGCGGTGGGCGACCGGGGATCCCCCACCCAGCTCCCCTTCCCGGCCCGCCTTGCTGGCTTTTGGGCCCCTGCGGGCTTTTTTTTTCCGGCT
3157 GGCTCGTTCCGCGGTGGGCGACCGGGGATCCCCCACCCAGCTCCCCTTCTCGGCCCGCCCCGCTGGCTTTTGGGCCCCTCCGGGCTTTTTTTTCCGGCT
6397 GGCTCGTTCCGCGGTGGGCGACCGGGGATCCCCCACCCAGCTCCCCTTCCCGGCCCGCCCCGCTGGCTTTTGGGCCCCTCCGGGCTTTTTTTTCCGGCT
3301 GGCTCGTTCCGCGGTGGGCGACCGGGGACCCCCCACCCAGCTCCCCTTCCCGGCTCGCCCCGCTGGCTTTTGGGCCCCTCCGGGCTTTTTTTTCCGGCT
W .....
Merlin GGCTCGTTCCGCGGTGGGCGACCGGGGATCCCCCACCCAGCTCCCCTTCCCGGCCCGCCCCGCTGGCTTTTGGGCCCCTCCGGGCTTTTTTTTTCCGGCT
Toledo GGCTCGTTCCGCGGTGGGCGACCGGGGATCCCCCACCCAGCTCCCCTTCCCGGCTCGCCCCGCTGGCTTTTGGGTCCCTCCGGGCTCTTTTTTCCGGCT
CONSENSUS -----
3401 3500
AD169 GGGGGTCGCGGCGGTCGGCCGACGACGACGGTAGGTGGGCCGGGTGGACGGTGGTGGGGACGGGCGACGCCCCGGCTCGACGGCAATCGGTCCCGGAAGG
3157 GGGGGTCGCGGCGGTCGGCCGACGACG...GTAGGTGGGCCGGGTGGACGGTGGTGGGGACGGGCGACGCCCCGGCTCGACGGCAGTCGGTCCTGGAAGG
6397 GGGGGTCGCGGCGGTCGGCCGACGACGACGGTAGGTGGGCCGGGTGGACGGTGGTGGGGACGGGCGACGCCCCGGCTCGACGGCAGTCGGTCCCGGAAGG
3301 GGGGGTCGCGGCGGTCGGCCGACGACGACGGTAGGTGGGCCGGGTGGACGGTGGTGGGGACGGGCGACGCCCCGGCTCGACGGCAGTCGGTCCCGGAAGG
W .....
Merlin GGGGGTCGCGGCGGTCGGCCGACGACGACGGTAGGTGGGCCGGGTGGACGGTGGTGGGGACGGGCGACGCCCCGGCTCGACGGCAGTCGGTCCCGGAAGG
Toledo GGGGGTCGCGGCGGCCGGCCGACGACGACGGTAGGTGGGCCGGGCGGACGGTGGTGGGGACGGGCGACGCCCCGGCTCGACGGCAGTCGGCCCCGGAGGG
CONSENSUS -----
3501 3600
AD169 TTGGGGGCTGGGGGCCCGGTCAGGAGCTCCGGGAGCGGGGTCGACCGCGACGGCTTCCGGGTCTCGCGGCGGCTCCCTCTCGGCGGCTCCGGTTGGGCTC
3157 TTGGGGGCTGGGGGCCCGGTCAGGAGCTCCGGGAGCGGGGTCGACCGCGACGGCTTCCGGGTCTCGCGACGGCTCCCTCTCGGCGGCTCCGGTTGGGCTC
6397 TTGGGGGCTGGGGGCCCGGTCAGGAGCTCCGGGAGCGGGGTCGACCGCGACGGCTTCCGGGTCTCGCGGCGGCTCCCTCTCGGCGGCTCCGGTTGGGCTC
3301 TTGGGGGTTGGGGGCCCGGTCAGGAGCTCCGGGAGCGGGGTCGACCGCGACGGCTTCCGGGTCTCGCGACGGCTCCCTCTCGGCGGCTCCGGTTGGGCTC
W .....
Merlin TTGGGGGCTGGGGGCCCGGTCAGGAGCTTCGGGAGCGGGGTCGACCGCGACGGCTTCCGGGTCTCGCGGCGGCTCCCTCTCGGCGGCTCCGGTTGGGCTC
Toledo TTGGGGGCTGGGGGCCCAGTCAGGAGCTCCGGGAGCGGGGTCGACCGCGACGGCTTCCGGGTCTCGCGGCGGCTCCCTCTCGGTGGCTCCGGTTGGGCTC
CONSENSUS -----
3601 3700
AD169 CCCTCCCCCCTCTCGAGGGTCCGGCCGCCAGTCGTGACCGGGGGTCCCTCGGCCTAGCCGCCGGCTCTCGGTCCGCCTTATCCTGGGCGTTGGCCTGTCC
3157 CCCTCCCCCTTCTCGAGGGTCCGGCCGCCAGTCGTGACCGGGGGTCCCTCGGCCTAGCCGCCGGCTCTCGGGCCGCCTTATCCTGGGCGTTGGCCGGTCC
6397 CCCTCCCCCCTCTCGAGGGTCCGGCCGCCAGTCGTGACCGGGGGTCCCTCGGCCTAGCCGCCGGCTCTCGGTCCGCCTTATCCTGGGCGTTGGCCGGTCC
3301 CCCTCCCCCTTCTCGAGAGTCCGGCCGCCAGTCGTGACCGGGGGTCCCTCGGCCTAGCCGCCGGCTCTCGGGCCGCCTTATCCTGGGCGTTGGCCGGTCC
W .....CGGCCTAGCCGCCGGCTCTCGGTCCGCCTTATCCTGGGCGTTGGCCGGTCC
Merlin CCCTCCCCCCTCTCGAGGGTCCGGCCGCCAGTCGTGACCGGGGGTCCCTCGGCCTAGCCGCCGGCTCTCGGTCCGCCTTATCCTGGGCGTTGGCCGGTCC
Toledo CCCTCCCCCCTCTCGAGGGTCCGGTCGCCGGTCGTGGCCGGGGGTCCCTCGGCCTAGCCGTCGGCTCTCGGTCCGCCTTACCCTGGGCGTTGGCCGGTCC
CONSENSUS -----CGGCCTAGCCG-CGGCTCTCGG-CCGCCTTA-CCTGGGCGTTGGCC-GTCC
3701 3800
AD169 CGTGACGCTCCCCTCCCCCGCTGCTCCCCAAAAAAA..CTCCGCCCGAACCGTCGCGGCTTGCTGGCCCTGGGCGTGGTCCCCC.ACTCCCCTCCCCCCA
3157 CGTGACGCTCCCTTCCCCCGCTGCTCCCCAAAAAAT..GTCCGCCCGAACCGTCGCGGCTTGCTGGCCCTGGGCGTGGTCCCCC.ACTCCCCTCCCCCCA
6397 CGTGACGCTCCCCTCCCCCACTGCTCCCCAAAAAAT..CTCCGCCCGAACCGTCGCGGCTTGCTGGCCCTGGGCGTGGTCCCCC.ACTCCCCTCCCCCCA
3301 CGTGACGCTCCCTTCCCCCGCTGCTCCCCAAAAAAT..GTCCGCCCGAACCGTCGCGGCTTGCTGGCCCTGGGCGTGGTCCCCC.ACTCCCCTCCCCCCA
W CGTGACGCTCCCCTCCCCCACTGCTCCCCAAAAAAAT..CTCCGCCCGAACCGTCGCGGCTTGCTGGCCCTGGGCGTGGTCCCCC.ACTCCCCTCCCCCCA
Merlin CGTGACGCTCCCCTCCCCCACTGCTCCCCAAAAAAA..CTCCGCCCGAACCGTCGCGGCTTGCTGGCCCTGGGCGTGGTCCCCC.ACTCCCCTCCCCCCA
Toledo CGTGACGCTCCCCTCCCCCACTGCTCCCCAAAAAT..CTCCGCCCGAACCGTCGCGGCTTGCTGGCCCTGGGCGTGGTCCCCCCACTCCCCTCCCCCCA
CONSENSUS CGTGACGCTCCC-TCCCCC-CTGCTCCCCAAAAA-----TCCGCCCGAACCGTCGCGGCTTGCTGGCCCTGGGCGTGGTCCCCC-ACTCCCCTCCCCCCA


ORF3 start
UL62 stop
UL61 start
ORF4 start
ORF3 stop
ORF4 stop
UL62 start


UL63 stop
ORF5 start
UL63 start
ORF5 stop
UL64 stop


UL64 start
UL65 start
UL66 stop
UL65 stop
UL66 start


UL67 stop
UL67 start
UL68 stop



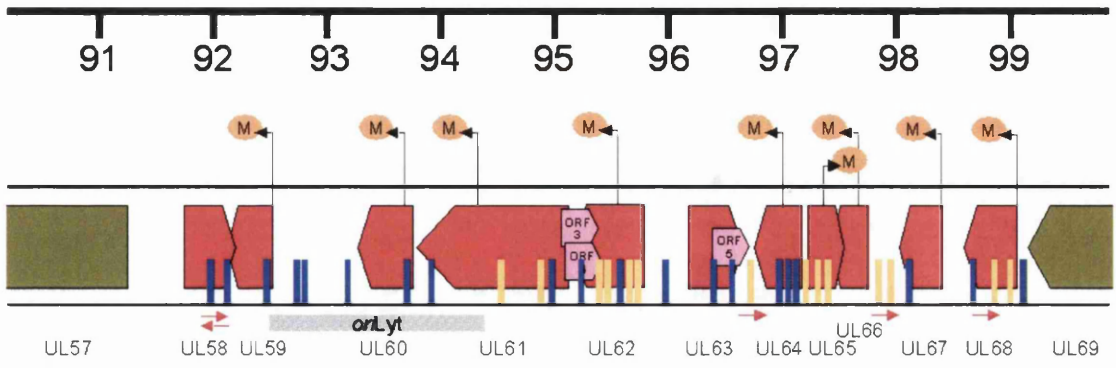


Figure 4.4. Frameshifts between HCMV strains in region O.

Navy blue lines show insertions/deletions that cause frameshifts in comparison to AD169. Gold lines show A:T tracts of variable length that cause frameshifts in comparison to AD169. Red arrows show polyadenylation signals and the first in-frame ATG codon of each ORF is shown by the yellow M.

ORF	Number of frameshift mutations affecting ORF	Number of these frameshifts caused by variation in A:T tract length (≥ 8 bp)
UL57	**0**	**0**
UL58	2	0
UL59	1	0
UL60	1	0
UL61	4	2
UL62	8	3
ORF3	4	1
ORF4	5	1
UL63	2	0
ORF5	2	1
UL64	3	0
UL65	2	2
UL66	0	0
UL67	1	0
UL58	3	2
UL69	**0**	**0**

Table 4.2. Frameshift mutations in the ORFs of region O, identified by sequence comparisons.

The central column concerns all frameshift mutations which affect the ORF. The right column concerns the number of these frameshift mutations caused by differences in A:T tracts of 8 or more residues.

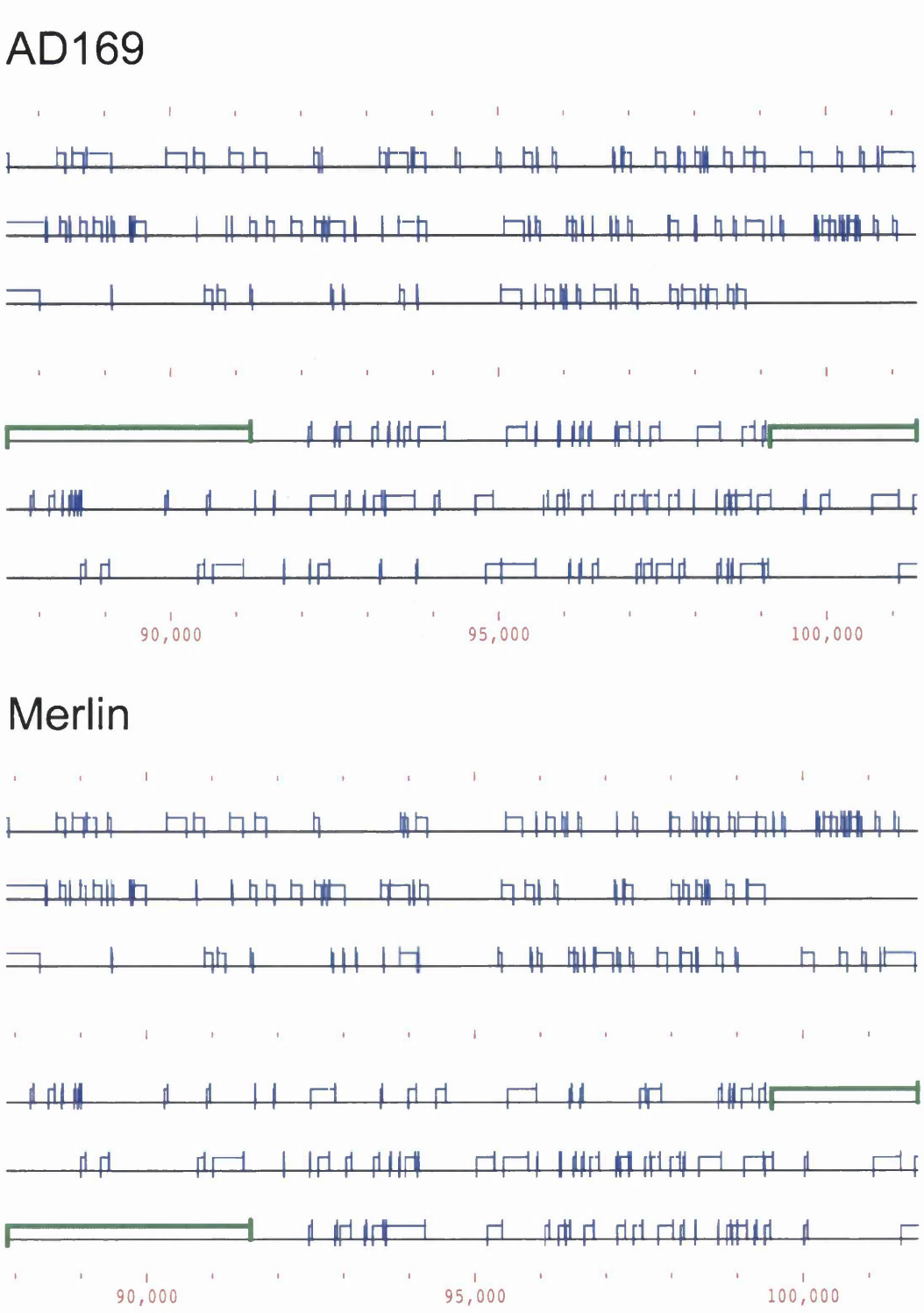


Figure 4.5. ORFs in region O.

ORFs for the six translation frames of the DNA sequences of AD169 and Merlin have been computed using Frames. Start codons are shown as short lines extending above the box, and stop codons are shown as short lines extending below the box. The region shown is larger than that shown in Figures 4.1 to 4.4, to include ORFs UL57 and UL69, which are highlighted in green.

Region O is well conserved, with a very high level of identity between strains. The data for AD169 and Merlin confirmed the previously published sequence for these strains in this region (Chee *et al.*, 1990; Dolan *et al.*, 2004). Like region X, region O contains a number of polynucleotide A:T tracts, in which length variation was observed between clones. As previously discussed in Section 3.3, variation in these tracts (of eight repeat units or more) appears to be an artefact of PCR amplification. Consequently, the corresponding sequences of four strains (Towne, FIX, PH and TR) sequenced by Murphy *et al.* (2003b) from cloned BACs were compared to the seven strains in the sequence comparisons in region O to more reliably investigate variation in these tracts.

Each of UL58-UL68, with the exception of UL66, was disrupted in certain strains by nucleotide insertions or deletions that shift the reading frame, in comparison with AD169. Although approximately 600 bp of sequence for six of the seven strains sequenced in the present study is missing from UL61, four frameshifts were registered in other regions of this ORF UL61. This part of region O is of particularly high G+C content and contains numerous inverted and direct repeats within the core region of *ori*Lyt (Section 4.2).

The data in Table 4.2 shows all the frameshifts in each ORF caused by insertions or deletions, also specifying which frameshifts are caused by differences in A:T tract length. In addition to the frameshifts identified in UL58-UL68, a total of nine differences that would cause frameshifts were also found in the areas between the predicted ORFs (Figure 4.4). Analysis of the areas of region O not affected by frameshift mutations concluded that no novel ORFs emerged as potentially encoding proteins from the sequence comparisons. All ORFs in strains AD169 and Merlin are shown in Figure 4.5. The potential of region O to contain protein-coding ORFs is limited by the high frequency of stop codons in all six coding frames and by frameshifts in the few ORFs conserved in the HCMV strains. Consequently, it is unlikely that ORFs UL58-UL68, or any other ORFs in region O, code for proteins. As expected of recognised genes, no frameshifts were registered in UL57 or UL69.

4.4 Discussion

It seems unlikely that UL58-UL68, which were assigned to region O in the original analysis of HCMV strain AD169 (Chee *et al.*, 1990), actually encode proteins. Only UL59, UL67 and UL68 commence with ATG codons. The other ORFs are preceded by termination codons, with UL60-UL62 and UL64-UL66 containing in-frame ATG codons internally and UL58 and UL63 lacking such codons (Figure 4.1). Sequence comparisons of seven HCMV strains in region O showed that each of these ORFs, with the exception of UL66, is not intact in all HCMV strains and is disrupted by frameshifts, in comparison with AD169 (Figure 3.4). In addition, the three ORFs defined by Murphy *et al.* (2003a) are all disrupted by frameshifts. Using an internal ATG initiation codon, UL66 would be predicted to encode a protein of 78 aa, which is not similar to other proteins in the NCBI database, and UL66 is completely overlapped by pp67, a transcript originating from region O (Figure 4.1).

Sequence comparisons show that frameshifts also affect pp67, a late RNA associated with UL65. This RNA was originally sequenced and mapped in strain Towne by Davis *et al.* (1984, 1985), then later in strain AD169 by Scott *et al.* (2002). The pp67 RNA is routinely detected in nucleic acid sequence-based amplification (NASBA) assays to monitor the progression of HCMV infections and the effect of antiviral therapy on viral activity (Blok *et al.*, 1998; Gerna *et al.*, 1999). The initial study by Davis *et al.* (1984) reported that this RNA encodes a 67 kDa protein, which could be detected by monoclonal antibodies in total protein preparations from Towne-infected cells. Although the pp67 RNA is routinely detected in NASBA assays, the work by Davis *et al.* (1984) remains the only study suggesting that a protein is encoded, and the authors did not describe the process by which monoclonal antibodies were prepared. A more recent study by Scott *et al.* (2002) characterised pp67 transcripts in AD169, identifying multiple 5'-ends. The authors noted that multiple strain analysis would be required to confirm initial differences observed between the Towne and AD169 ORFs. Sequence comparisons of seven strains in the present study suggest that pp67 does not encode a protein.

Figure 4.6. Sequence comparisons of region O in HCMV and other CMVs.

Comparisons are displayed in a matrix plot computed using Compare (window of 21 nucleotides with a minimum match of 17 nucleotides) and displayed using Dotplot.

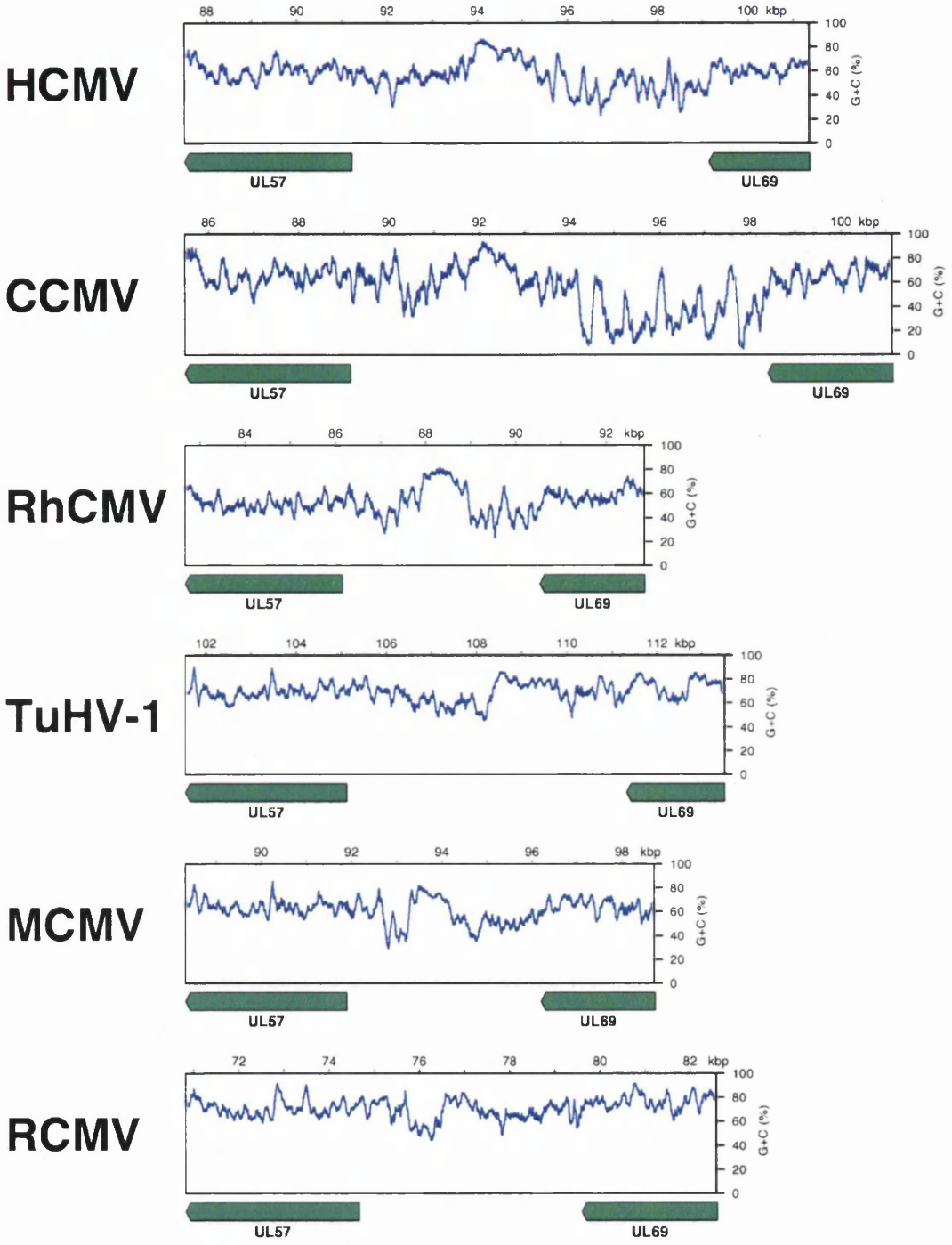


Figure 4.7. The nucleotide composition of region O and corresponding regions in other CMVs.

The plots were computed using Window (100 nucleotides shifted by 3 nucleotide increments) and displayed using Statplot. Known genes UL57 and UL69 are marked on each plot.

Sequence comparisons did not identify any novel ORFs emerging as potentially encoding proteins because of other frameshifts which were identified throughout the region, and the number of stop codons. The lack of protein-coding ORFs in region O leaves a large region of 7.9 kbp that contains *ori*Lyt, a region of unusual nucleotide composition and possibly the highest content of inverted and direct repeats in HCMV (Masse *et al.*, 1992). Corresponding regions are present in other CMVs, where they are similar in size and position in the genome (Table 4.3). DNA sequence comparisons between HCMV and other CMV sequences are shown as two-dimensional comparisons on matrix plots in Figure 4.6. The corresponding region in CCMV, the closest relative of HCMV, is similar in size and position but the genomes show little collinearity in UL58-UL68, in contrast to the sequence similarity evident in the flanking genes UL57 and UL69 (Figure 4.6). Regions corresponding to region O can also be identified in the more distantly related CMV genomes of RhCMV (4.4 kb), MCMV (3.3 kb), RCMV (4.9 kbp) and TuHV-1 (6.1 kbp) (Table 4.3, Figure 4.6). When compared with HCMV, these genomes are extremely divergent in region O, much more so than in the flanking genes (UL57 and UL69).

Region O in HCMV has an asymmetrical nucleotide composition, with high G+C content in the core *ori*Lyt region and lower G+C content elsewhere (Figure 4.7). Similar patterns of unusual nucleotide composition are exhibited in the corresponding regions of other CMV genomes. Even in distant relatives such as TuHV-1, which has little sequence similarity to HCMV, a similar pattern can be seen (Figure 4.7). Polynucleotide A:T tracts of eight or more residues are a common feature in the low G+C part of region O, as in region X (Chapter 3). In contrast, such tracts are not common in recognised HCMV genes.

Located in the left half of region O, *ori*Lyt functions as an essential *cis*-acting factor required for lytic-phase DNA replication, as previously described (Section 4.2). The origin regions in the other CMVs, where they have been identified, map to the corresponding region. The core region of *ori*Lyt is flanked by sequences that are required for full activity but are not essential. Each of UL58-UL62, sequences essential to or augmenting *ori*Lyt function are disrupted by frameshift mutations. In the right half of region O, UL63-UL68 and ORF3-ORF5 do not encode proteins as

they are disrupted by frameshift mutations. The only ORF in this half of region O reported to encode a transcript, pp67, is also disrupted by frameshift mutations and is therefore unlikely to be protein-coding. Transcript mapping would be required to further investigate the possibility of other RNAs originating from this region, although none would be expected to be protein-coding.

CHAPTER 5: REGION G

5.1 Protein-coding genes

In addition to regions X and O, which are investigated in chapters 3 and 4, the proposed gene content of a region at the end of the U_S sequence of the viral genome was affected by the comparative analysis with CCMV (Davison *et al.*, 2003a). This region, US33-US36, is here termed region G. ORFs US33, US35 and US36 have been discounted as they have no counterparts in CCMV, and a novel ORF, US34A, has been identified as having a counterpart in CCMV (Davison *et al.*, 2003a). The region is flanked by ORFs US32 and TRS1, and contains US34 internally (Figure 5.1).

TRS1 extends from TR_S into U_S, and has a counterpart gene, IRS1, which extends from IR_S into the U_S (Chee *et al.*, 1990). Transcription is controlled by identical promoters, and both genes are expressed at IE times p.i., although their mRNAs and encoded proteins could be detected during all phases of the viral replication cycle (Romanowski and Shenk, 1997).

TRS1 and IRS1 encode proteins with identical N-terminal domains (549 aa) and different C-terminal regions (240 aa and 298 aa, respectively). TRS1 encodes a 789 aa, 84 kDa protein, which is present in the nucleus and cytoplasm IE and E times p.i., but mostly in the cytoplasm at L times p.i. (Romanowski and Shenk, 1997). Both the TRS1 and IRS1 proteins are associated with HCMV virions, and were not degraded when intact virions were treated with trypsin (Romanowski *et al.*, 1997). As IE proteins, pIRS1 and pTRS1 are transcriptional trans-activators that co-operate with the major IE1 and IE2 proteins to create a complex regulatory network that controls gene expression (Pari and Anders, 1993; Pari *et al.*, 1993). However, TRS1-expressing plasmids were shown to produce several-fold-higher levels of activity than IRS1-expressing plasmids, as shown using reporter genes (Kerry *et al.*, 1996).

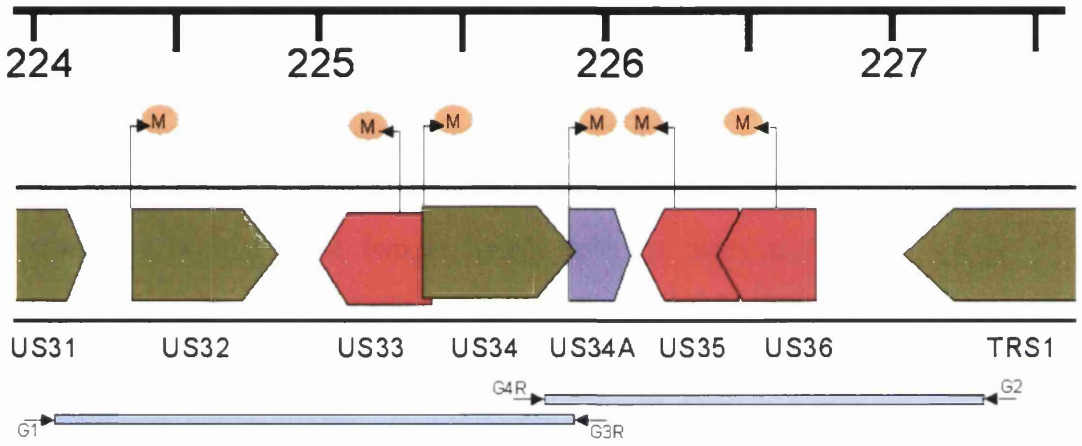


Figure 5.1. Map of the characterised genes and predicted genes in region G and flanking sequences.

The scale at the top shows the location (kbp) in the AD169 genome. Dark green shows ORFs thought to encode proteins. Red shows ORFs predicted by Chee *et al.* (1990) that have no known function and have been discounted because they lack counterparts in CCMV (Davison *et al.*, 2003). Purple shows a novel ORF that has a counterpart in CCMV. Blue bars show the two overlapping PCR products amplified using the primers shown in Appendix. The 3'-end of each ORF has a point to show its orientation, and the M shows the first in-frame ATG initiation codon in each ORF.

An IE gene product encoded within IRS1 gene, termed pIRS1^{263}, was identified by Romanowski and Shenk (1997). This new protein was encoded within the 3'-end of the IRS1 gene in same reading frame as the large pIRS1 protein. Expression of the smaller protein was controlled by a promoter residing within IRS1, in the U_S region. The protein pIRS1^{263}, maximally expressed at early times p.i. and residing in the nucleus, can antagonise the transcriptional activating capabilities of pTRS1 (Romanowski and Shenk, 1997).

Deletion mutants which failed to produce pIRS1 and pIRS1^{263} proteins were shown to replicate in HFFF cells with normal growth kinetics (Jones and Muzithras, 1992; Blankenship and Shenk, 2002). A TRS1 C-terminal deletion mutant, however, produced ~200-fold less virus than wild-type HCMV, apparently due to a packaging defect (Blankenship and Shenk, 2002). On infection of this mutant, viral DNA accumulated normally but two tegument proteins were abnormally localised and infectious viral particles did not accumulate to normal levels. It was concluded that the structure of the mutant particles was aberrant, indicating that the TRS1 protein was required, either directly or indirectly, for efficient assembly of virus particles (Blankenship and Shenk, 2002). More recent studies have shown that TRS1 and IRS1 are also able to counteract critical host cell antiviral responses, indicating that TRS1 is a complex gene with multiple distinct functions (Child *et al.*, 2002; Child *et al.*, 2004).

US32, which flanks the left end of region G, is a member of the HCMV US1 gene family, with US1 and US31 (Chee *et al.*, 1990). US31 was one of 24 HCMV ORFs revised by the analysis of Davison *et al.* (2003a) in which a simple correction was made at the start of the reading frame to reduce the size to 486 bp, allowing the putative protein (revised to 162 aa) to commence on an ATG initiation codon. US34 has been partially characterised in a study by Scott *et al.* (2002). The transcript, sequenced from a late cDNA library, was found to be mostly congruent with the US34 ORF predicted by Chee *et al.* (1990), although a definitive 5'-end was not determined. US34A was introduced to region G as one of four new AD169 genes predicted from the comparative study with CCMV (Davison *et* al., 2003a).

All of these characterised or partially characterised genes within or flanking region G commence with ATG codons (Figure 5.1). Only US33, US35 and US36, which were discounted by the comparative analysis with CCMV, do not commence with ATG initiation codons. These ORFs are preceded by termination codons, and contain in-frame ATG codons internally (Figure 5.1).

5.2 Sequence comparisons

Sequence comparisons of this region were used to investigate the protein-coding potential of the three ORFs discounted by comparative analysis with CCMV (US33, US35 and US36), and the novel ORF introduced (US34A) (Davison *et al.*, 2003a). Sequence data were also obtained for flanking ORFs. Two overlapping PCR products were amplified for each strain (Figure 5.1), one originating from US31 and terminating in US34A, and the other originating from US34A and terminating in TRS1. The seven HCMV strains sequenced were high passage strain AD169, low passage strains Toledo, Merlin, 3157 and 6397, and strains obtained directly from human tissue, W and 3301 (Section 2.1.1). The PCR products for each strain were cloned into plasmids and at least four plasmids were sequenced for each PCR product, as described previously (Section 3.3). Electrophoresis of the PCR products identified one distinct difference between strains, where the amplification product of strain 6397 (using primers G4R and G2; Figure 5.1) was approximately 350 bp larger than the corresponding product for all the other strains.

The sequence is very well conserved between strains in US31-US34A and TRS1 (Figure 5.2). Data for AD169 confirmed the previously published sequence for this strain (Chee *et al.*, 1990). Data for Merlin also confirmed the recently published sequence for this strain with the exception of two differences, both found in polynucleotide G:C tracts (Dolan *et al.*, 2004). Three G:C tracts were identified in the region, all of variable length between strains, and all located in US35 or US36 (Figure 5.2). Sequence comparisons in regions X and O (Chapters 3 and 4, respectively) showed higher frequencies of A:T tracts in apparently non-protein-coding regions. Region G contains one A:T tract of variable length in US35, in addition to the three G:C tracts.

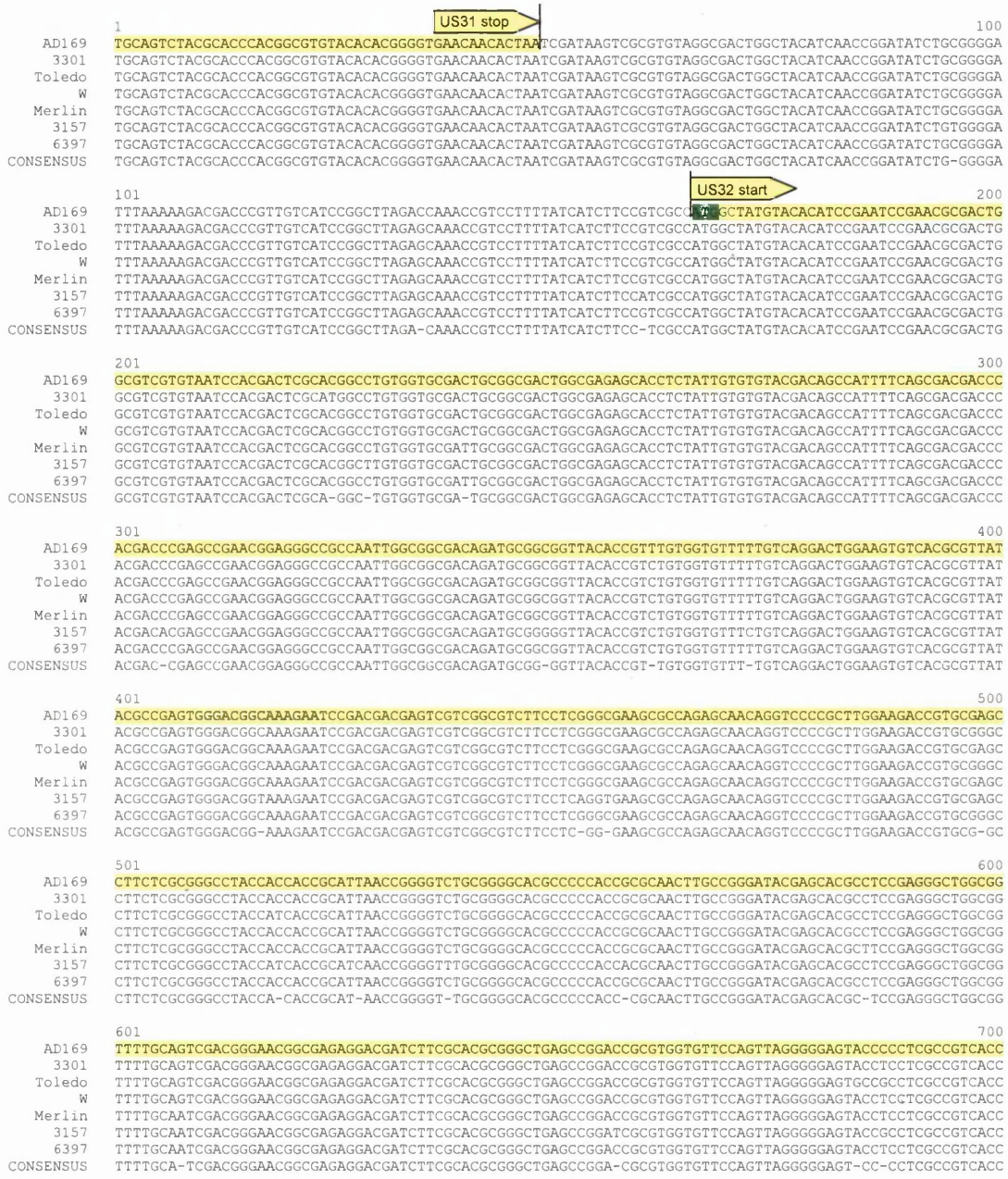
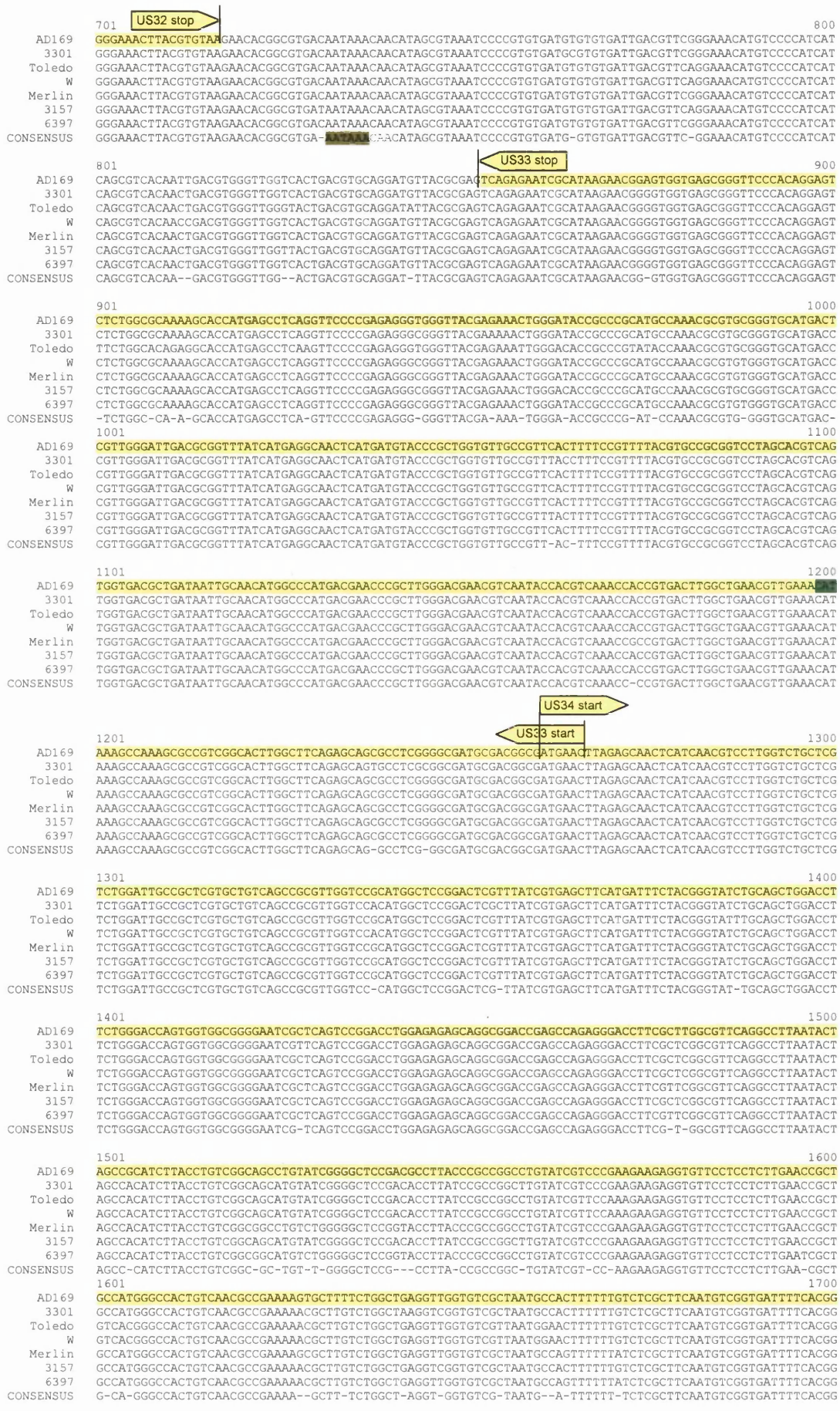
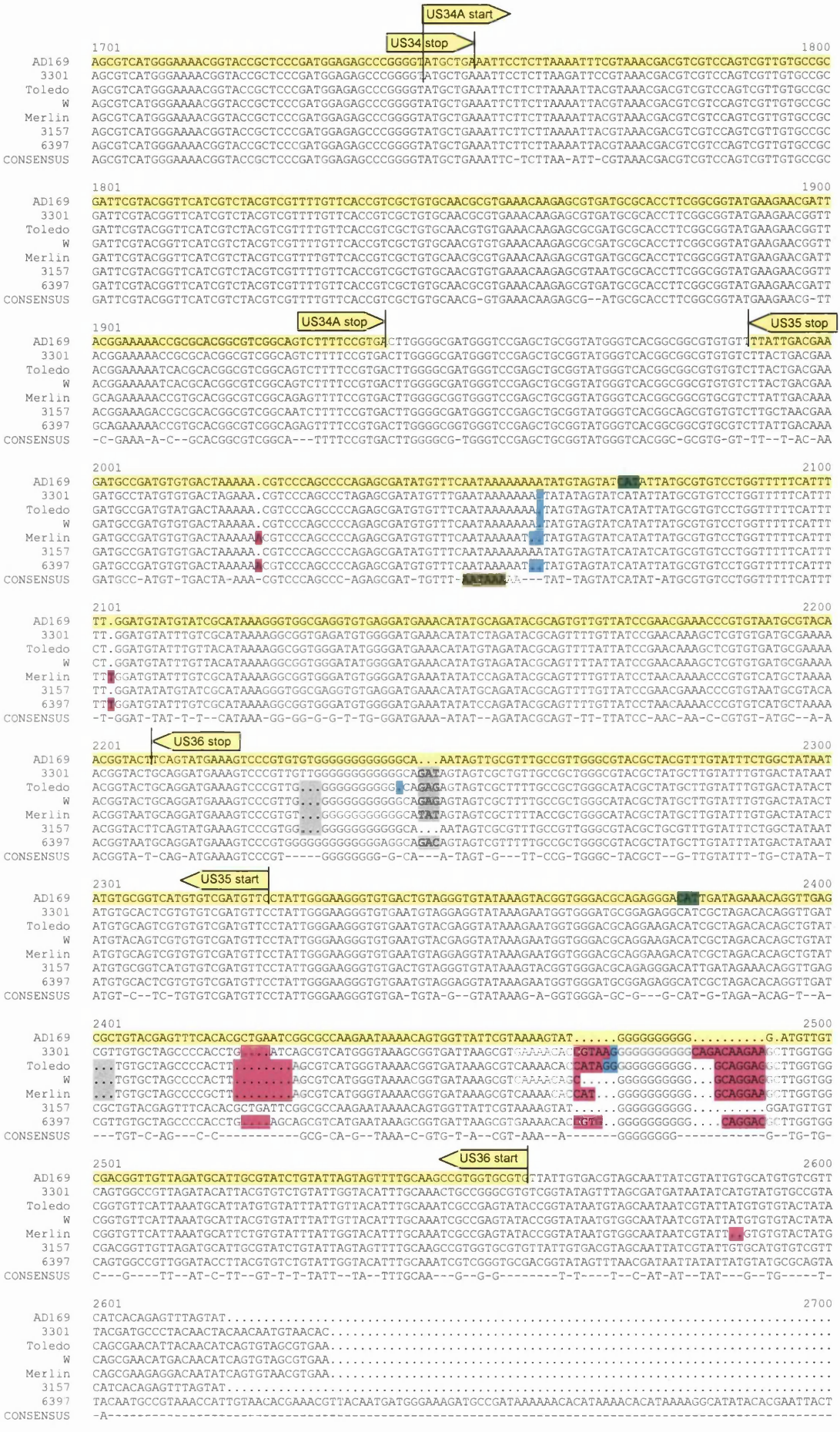
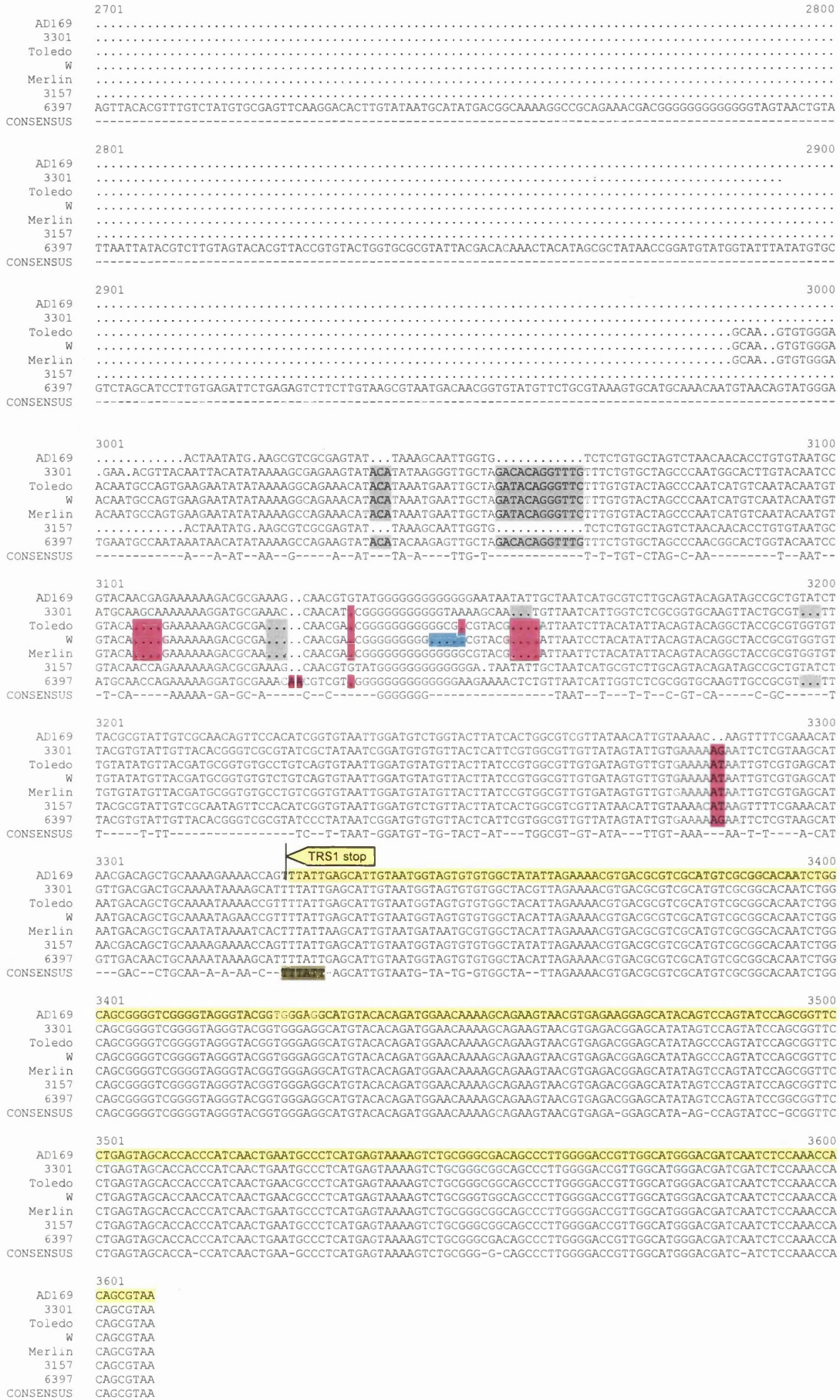


Figure 5.2. Sequence comparisons of seven HCMV strains in region G.

Alignments were made using ClustalW and displayed using Pretty. Yellow AD169 sequence shows all characterised and predicted ORFs in the region. The start and stop codons for each ORF are shown, the arrows indicating their orientation. Purple shows insertions/deletions that cause frameshifts relative to AD169. Sky blue shows length variation in polynucleotide A:T or G:C tracts of eight or more residues that cause frameshifts in comparison to AD169. Grey shows in-frame insertions/deletions. Brown consensus residues show conserved potential poly(A) signals (AATAAA). Green shows the internal ATG initiation codon in ORFs US33, US35 and US36. The Figure is continued on the following pages.


US32 stop
US33 stop
US34 start
US33 start


US34A start
US34 stop
US34A stop
US35 stop
US36 stop
US35 start
US36 start
AD169
Toledo
Merlin
CONSENSUS


TRS1 stop

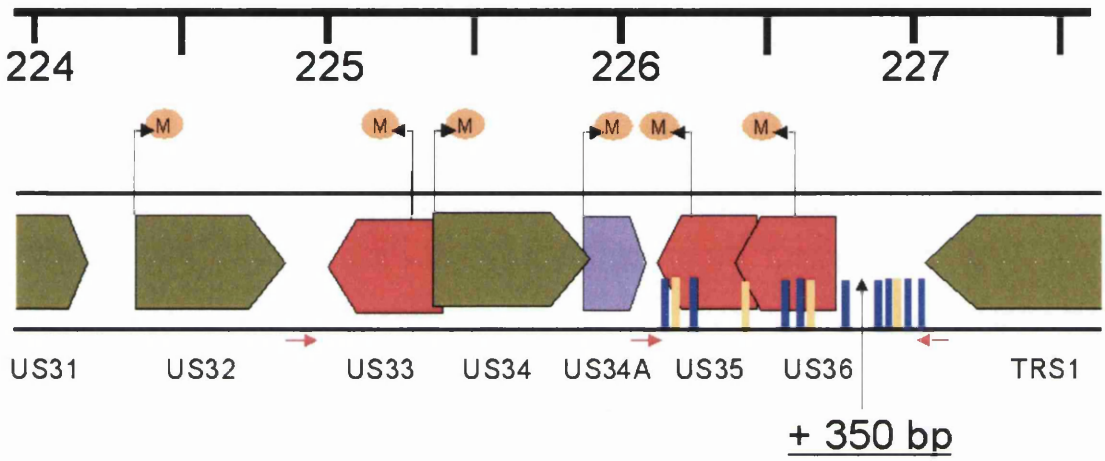


Figure 5.3. Frameshifts between HCMV strains in region G.

Navy blue lines show insertions/deletions that cause frameshifts in comparison to AD169. Gold lines show A:T tracts of variable length that cause frameshifts in comparison to AD169. Red arrows show poly(A) signals and the first in-frame ATG codon of each ORF is shown by the yellow M. No frameshifts were detected in US31-US34A, or in TRS1. In strain 6397, the location of the additional sequence (of approximately 350 bp) is shown.

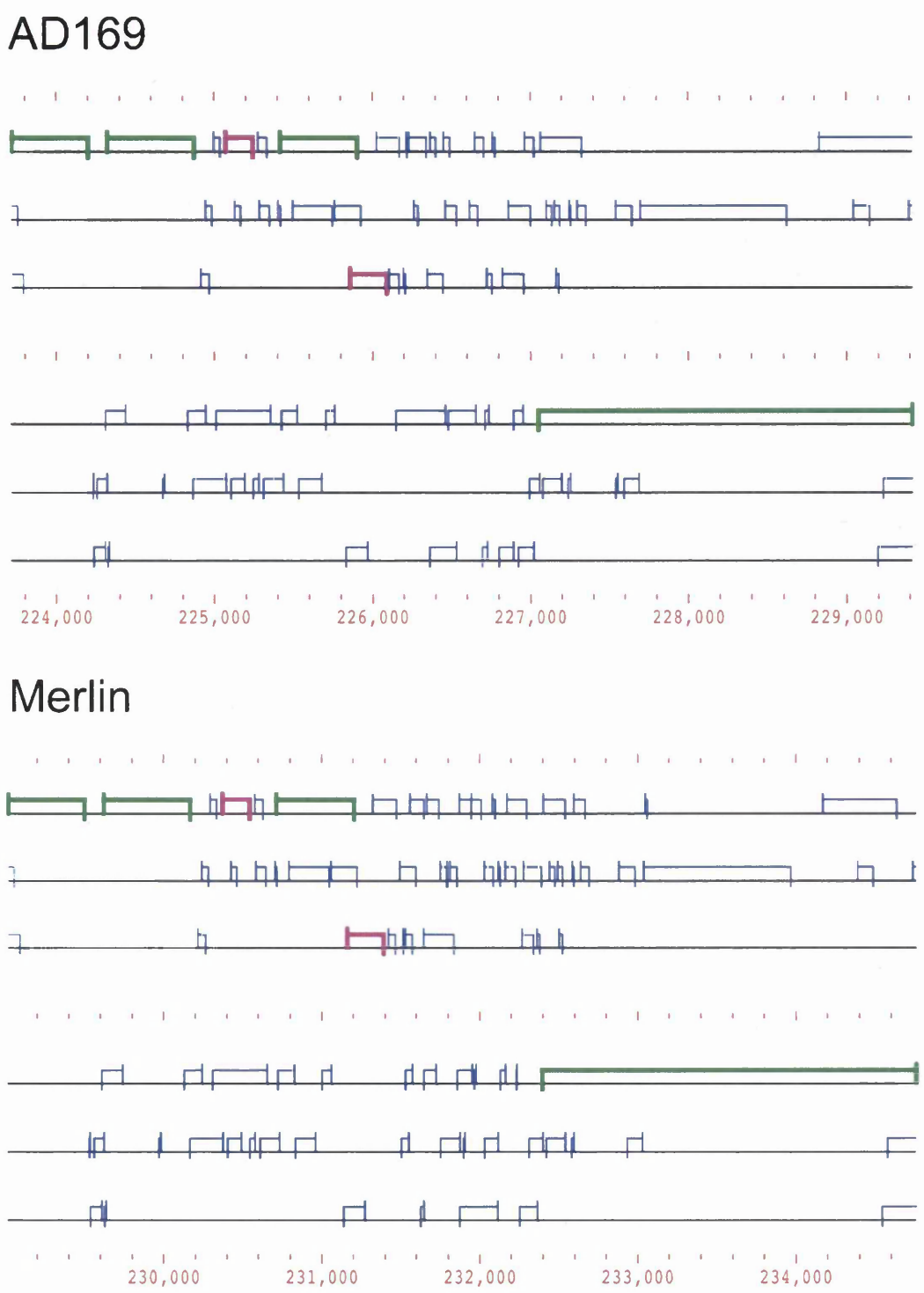


Figure 5.4. ORFs in region G.

ORFs for the six translation frames of the DNA sequences of AD169 and Merlin have been computed using Frames. Start codons are shown as short lines extending above the box, and stop codons are shown as short lines extending below the box. The region shown is larger than that shown in Figures 3.1 to 3.4, to include ORFs UL31 and TRS1, which are highlighted in green. In addition, US32 and US34 are highlighted in green and US33A (described in Section 5.4) and US34A are highlighted in purple.

Sequence data confirmed previous observations from gel electrophoresis of the PCR products across region G that strain 6397 harboured extra sequence. An insertion of approximately 350 bp was identified in 6397, between the end of TRS1 and the start of ORF US36 (Figure 5.3). A recent analysis of the coding content of the HCMV genome sequenced six strains of HCMV which had been cloned as infectious bacterial artificial chromosomes (Murphy *et al.*, 2003b). BAC DNA was subcloned and sequenced using a shotgun cloning method so the length of polynucleotide tracts was not affected by the fidelity of a PCR amplification reaction, as described in Section 3.3. The corresponding sequences of four of the strains (Towne, FIX, PH and TR) were subsequently compared to the seven strains sequenced in the current study to investigate variation in the polynucleotide tracts and other differences between strains that would cause frameshifts.

In comparison with AD169, one or more strains exhibit frameshifts in US35 and US36. These ORFs are disrupted in certain strains by nucleotide insertions or deletions that shift the reading frame in comparison with AD169 (Figures 5.2 and 5.3). As expected, a number of in-frame insertions/deletions which do not cause frameshifts were identified. Strain 6397, which contains additional sequence between US36 and TRS1, is distinctly different from the other strains in this area of region G. However, this strain alone did not account for the frameshifts observed, since variation was found in all strains. Data for BACs Towne and TR were mostly congruent with that of strain 6397, containing the additional ~350 bp in region G compared to the other strains. Although similar in size, the additional sequence in these three strains is also quite variable. As each of US35-US36 exhibits frameshifts in comparison with AD169, it is very unlikely that these ORFs actually encode proteins since they are not intact in all HCMV strains. Frameshifts were also identified in the sequence between US36 and TRS1. Furthermore, the presence of three G:C tracts of eight or more residues would be unusual for HCMV protein-coding regions. The only other G:C tracts of eight or more residues in the genome are grouped within the non-coding, highly repetitive *ori*Lyt (Section 4.3). Consequently, analysis of the region found no novel ORFs emerging as potentially encoding proteins from the sequence comparisons. All ORFs in strains AD169 and Merlin are shown in Figure 5.4. The sequence between US34A and TRS1 was highly variable between strains, and no novel ORFs were found to be conserved in all the HCMV strains. Nor

were any novel ORFs identified in the additional sequence of strain 6397. Significantly, no frameshifts were registered in US31-US34A or in TRS1, as expected of protein-coding genes.

In addition to US35 and US36, US33 also has no counterpart in CCMV. In contrast to US35 and US36, this ORF was well conserved between strains and not frameshifted (Figure 5.2). The ORF as defined by Chee *et* al. (1990) is preceded by a stop codon and contains an in-frame ATG initiation codon internally, 66 bp downstream from the predicted start site. Transcript analysis was undertaken to evaluate US33 further.

5.3 Transcript analysis in region G and flanking ORFs: northern blot analysis

Northern blots were carried out to determine the expression kinetics and sizes of transcripts from the sequence containing region G. Equal amounts of polyA$^+$ AD169 mRNA were electrophoresed in agarose-formaldehyde gels, blotted to membranes overnight, and fixed by UV irradiation. A polyA$^+$ RNA ladder was electrophoresed alongside the mRNA samples to aid estimations of transcript sizes (Section 2.1.8).

The two overlapping PCR products which were used to sequence region G, shown in Figure 5.1, were used to make dsDNA probes (section 2.2.22). The TRS1 RNA, characterised by Romanowski *et al.* (1997), was detected by probe K (Figure 5.5). The kinetics of expression were consistent with the published data, detecting an IE RNA of approximately 2.8 kb. A smear, also detected by probe K, appears to correspond to transcripts in E and L RNA of between approximately 0.7 kb and ~1.2 kb. Although not well defined by the dsDNA probe, the RNAs detected might correspond to US34 and US34A. Probe J also detected transcripts similar in size and kinetics of expression to the smear detected by probe K (Figure 5.5). An RNA of approximately 1.2 kb was detected in E and L RNA. In L RNA, this RNA was more abundant, and a smaller, L transcript of approximately 0.8 kb was also detected. The two different sized transcripts might correspond to US31 and US32 and/or US34 and US34A.

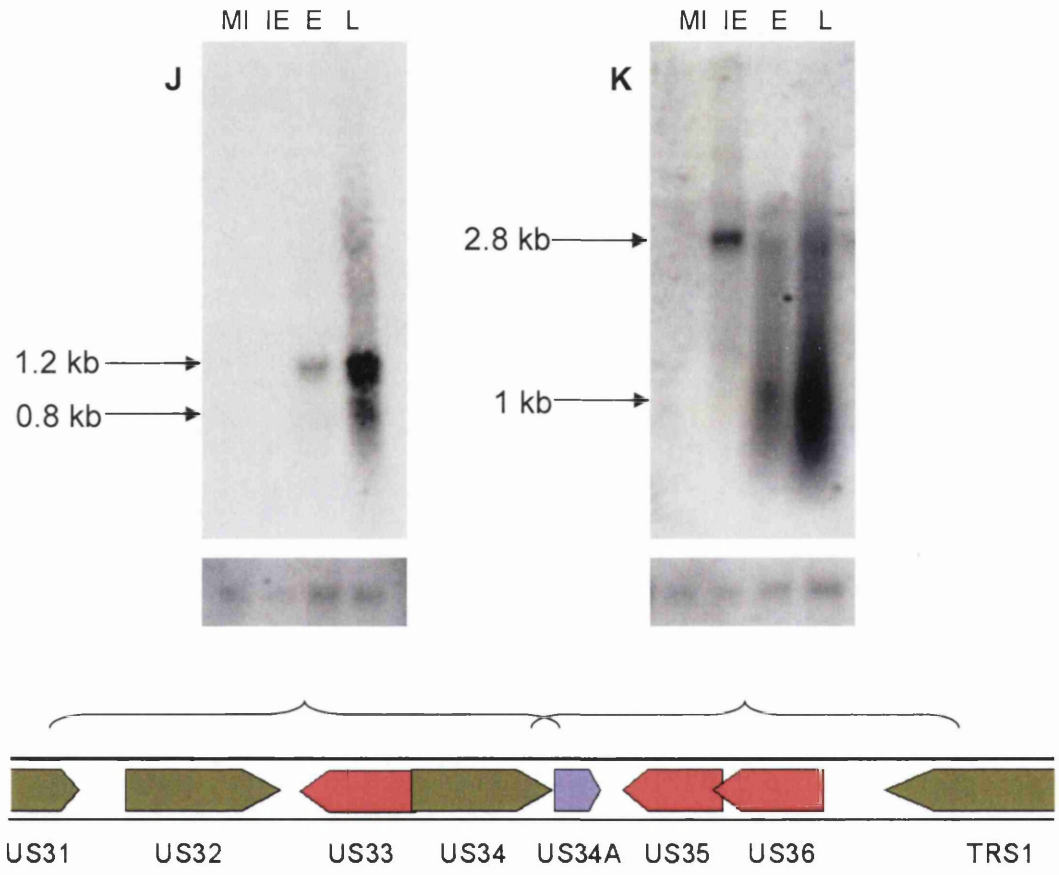


Figure 5.5. Northern blot analysis of region G using dsDNA probes.

The locations of the DNA probes are shown. Each blot included IE, E and L AD169 polyA$^+$ mRNA, and a negative control of MI polyA$^+$ mRNA. The results obtained using a GAPDH probe are shown at the bottom of each blot.

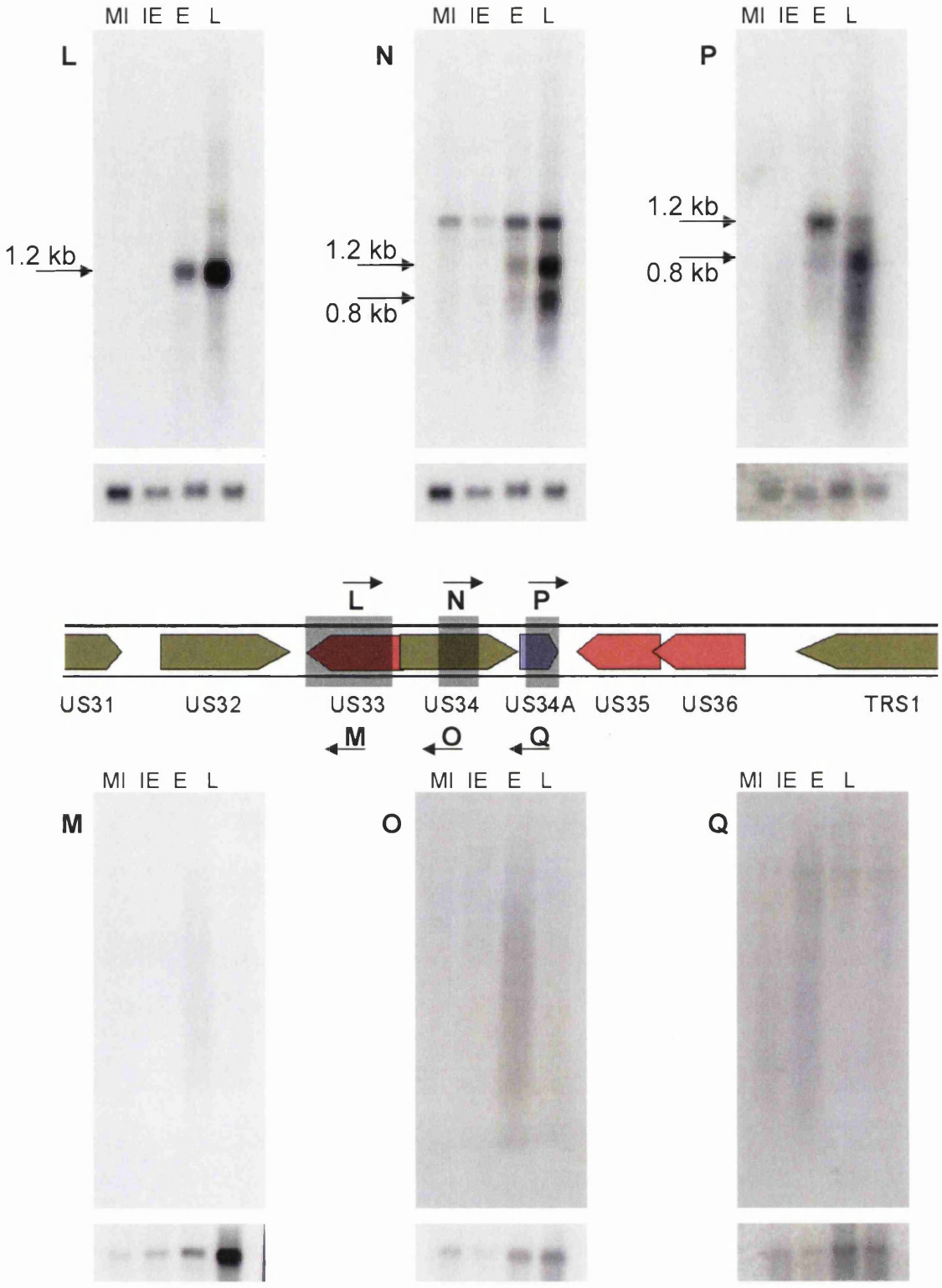


Figure 5.6. Northern blot analysis of region G using ssRNA probes.

The locations of the ssRNA probes are shown (not to scale, see Table 3.3 for accurate sizes), the arrows indicating the direction of any RNAs they would detect. Each blot included IE, E and L AD169 polyA$^+$ mRNA, and a negative control of MI polyA$^+$ mRNA. The results obtained using a GAPDH probe are shown at the bottom of each blot.

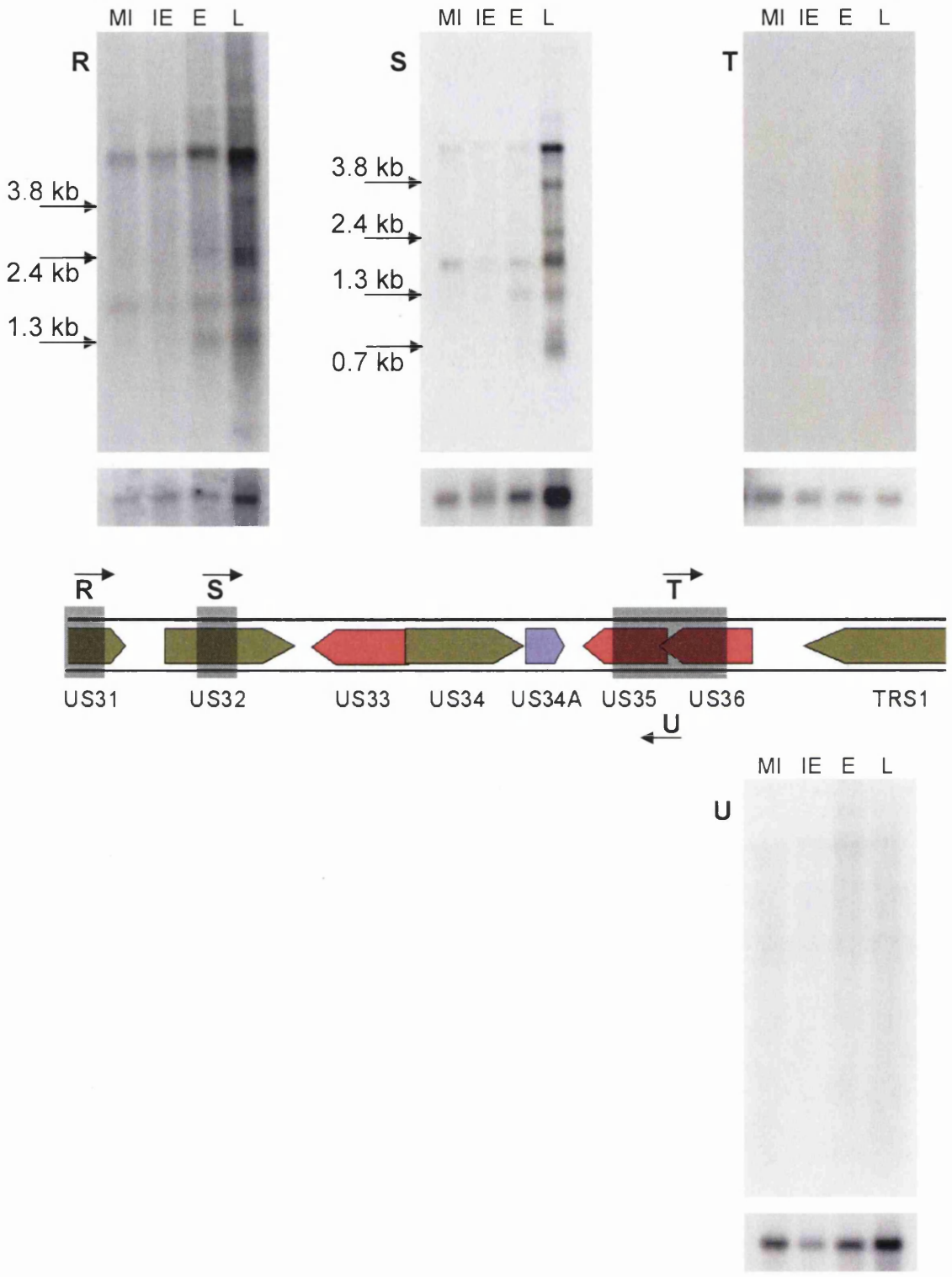


Figure 5.7. Further northern blot analysis of region G using ssRNA probes.

The locations of the ssRNA probes are shown (not to scale, see Table 3.3 for accurate sizes), the arrows indicating the direction of any RNAs they would detect. Each blot included IE, E and L AD169 polyA$^+$ mRNA, and a negative control of MI polyA$^+$ mRNA. The results obtained using a GAPDH probe are shown at the bottom of each blot.

Probe	ORF	strand	F primer	R primer	size (bp)
L	US33	+	**G10**	G24	337
M	US33	-	G10	**G24**	337
N	US34	+	**G14**	G7	194
O	US34	-	G14	**G7**	194
P	US34A	+	**G4R**	G21	213
Q	US34A	-	G4R	**G21**	213
R	US31	+	**G23**	G1R	234
S	US32	+	**G6**	G15	176
T	US35/US36	+	**G12**	G22	270
U	US35/US36	-	G12	**G22**	270

Table 5.1. ssRNA probes used for northern blot in region G and flanking ORFs.

Each probe was amplified by Lign'Scribe PCR reaction (Section 2.2.23) using the primers listed in Appendix. The positions of the probes within the region are shown in Figure 3.8. For each of probes L-U, the primer in bold was used in the MaxiScript radiolabelling reaction to make the probe single-stranded and therefore strand-specific (Section 2.2.23).

Sequence comparisons identified only three consensus poly(A) signals (AATAAA) in region G that were conserved in all the strains, as shown in Figure 5.3. From the positions of the poly(A) signals, it was predicted that US31 and US32 would express 3'-co-terminal transcripts utilising the poly(A) signal at nucleotide position 224,884 in AD169, and producing RNAs of approximately 1.3 and 0.7 kb, respectively. US34 and US34A were also predicted to express co-terminal transcripts, utilising a poly(A) signal at position 226,202, and producing RNAs of approximately 0.9 and 0.4 kb, respectively.

Strand-specific ssRNA probes were then used in northern blots to investigate further the transcripts in region G. All blots were performed twice to ensure their reproducibility and accuracy. Probe M, which corresponds to the US33 ORF predicted by Chee *et al.*, (1990), did not detect any transcripts. However, probe L (corresponding to the opposite strand of US33) detected an RNA of approximately 1.2

kb in E RNA, accumulating to a higher level in L RNA (Figure 5.6). Probe N (US34) detected a band of approximately 2 kb in all the polyA$^+$ mRNA samples (Figure 5.6), which probably is the 18S ribosomal RNA (1.9 kb). Probe N also detected two other transcripts, of approximately 1.2 and 0.8 kb. The size and kinetics of expression of the 1.2 kb RNA were the same as the RNA detected by probe L. The 0.8 kb RNA was detected in L RNA, and at low levels in E RNA. Probe P (US34A) also detected the 1.2 and 0.8 kb RNAs. Probes O and Q, corresponding to the opposite strands of US34 and US34A, respectively, did not detect any transcripts.

Probes R and S were specific to US31 and US32, which were predicted to be 3'-co-terminal genes (Figure 5.7). Probe R (US31) detected a transcript of approximately 1.3 kb in E and L RNA. Probe S (US32) detected the same RNA, with the same kinetics of expression, and a smaller transcript of approximately 0.7 kb in L RNA. Both probes also detected a number of additional RNAs. Two of these were present in all RNA samples, and probably correspond to cellular ribosomal RNAs 18S (1.9 kb) and 28S (5 kb) RNAs. RNAs of approximately 2.4 and 3.8 kb were also detected in L RNA by both probes. These could correspond to US29 and US30, upstream of US31 and in the same orientation, if these ORFs utilise the poly(A) site downstream of US32 and not a poly(A) site 2 bp downstream of the end of US30. Probes T and U, covering ORFs US35 and US36 and corresponding to the two strands, did not detect any transcripts (Figure 5.7).

Northern blot analysis identified strong candidates for the RNAs of US31, US32 and US34, each corresponding to the size predicted from the sequence. RNAs from US33, US35 and US36 were not detected. A transcript from US34A was not clearly detected. A novel RNA was mapped to the region occupied by US33, but on the opposite strand. This 1.2 kb E RNA, identified by probe L (Figure 5.6), is in the same orientation as US34, and consequently was detected by probes N and P, located downstream. The detection of this RNA by all three probes, and the sequence comparison data, suggested the transcript shared the same poly(A) signal as US34, making them co-terminal. RACE analysis was required to map accurately the ends of this novel transcript and those of the other RNAs identified in the region.

5.4 Transcript analysis in region G and flanking ORFs: 3'- and 5'-RACE

RACE was performed using a combination of three different kits to identify 3'- and 5'-ends: SMARTRACE, GeneRacer and 5'RACE (Sections 2.2.26, 2.2.27 and 2.2.28, respectively). RACE reactions utilised a template of L polyA$^+$ mRNA harvested from total infected cell RNA (Section 2.2.20). The primers used to map the ends of region G RNAs are shown in Table 5.2.

The 1.2 kb transcript, mapping to the opposite strand from US33, was termed US33A. This has not been previously recognised as an ORF. Sequence and northern blot analyses suggested that the US33A, US34 and possibly the US34A RNAs are expressed as co-terminal transcripts. Primers G29, G14 and G4R were used to amplify 3'-RACE products, and primers G28, G7 and G27 were used to amplify 5'-RACE products for these three ORFs, respectively (Table 5.2 and Figure 5.8). Each RACE-PCR product was gel purified and cloned, and approximately 10 clones were sequenced. All three 3'-RACE amplification products mapped to an A residue at nucleotide position 226,222 in AD169, 20 bp downstream from a consensus poly(A) signal (AATAAA). However, as US33A and US34 utilise the same poly(A) signal downstream from US34A, the presence of a US34A RNA could not be confirmed.

The 5'-end of US34 mapped to an A residue at position 225,383 and the 5'-end of novel ORF US33A mapped to an A residue at position 225,048. Two potential non-consensus TATA elements were identified upstream of the 5'-ends of the genes: CATAA motifs 31 bp and 33 bp upstream of the transcription start sites of US33A and US34, respectively (Figure 5.10). Using primer G27, 5'-RACE products of approximately 700 bp and 1050 bp were amplified that corresponded to US34 and US33A, respectively. From the same reaction, a 5'-RACE product of approximately 350 bp was amplified that corresponded to the predicted start of novel ORF US34A. This 5'-RACE product, which could only be identified using SMARTRACE reactions (Section 2.2.26), were gel purified and cloned. Each sequenced clone identified a different 5'-end. None of the potential 5'-ends was associated closely with an upstream consensus or non-consensus TATA box (Figure 5.10).

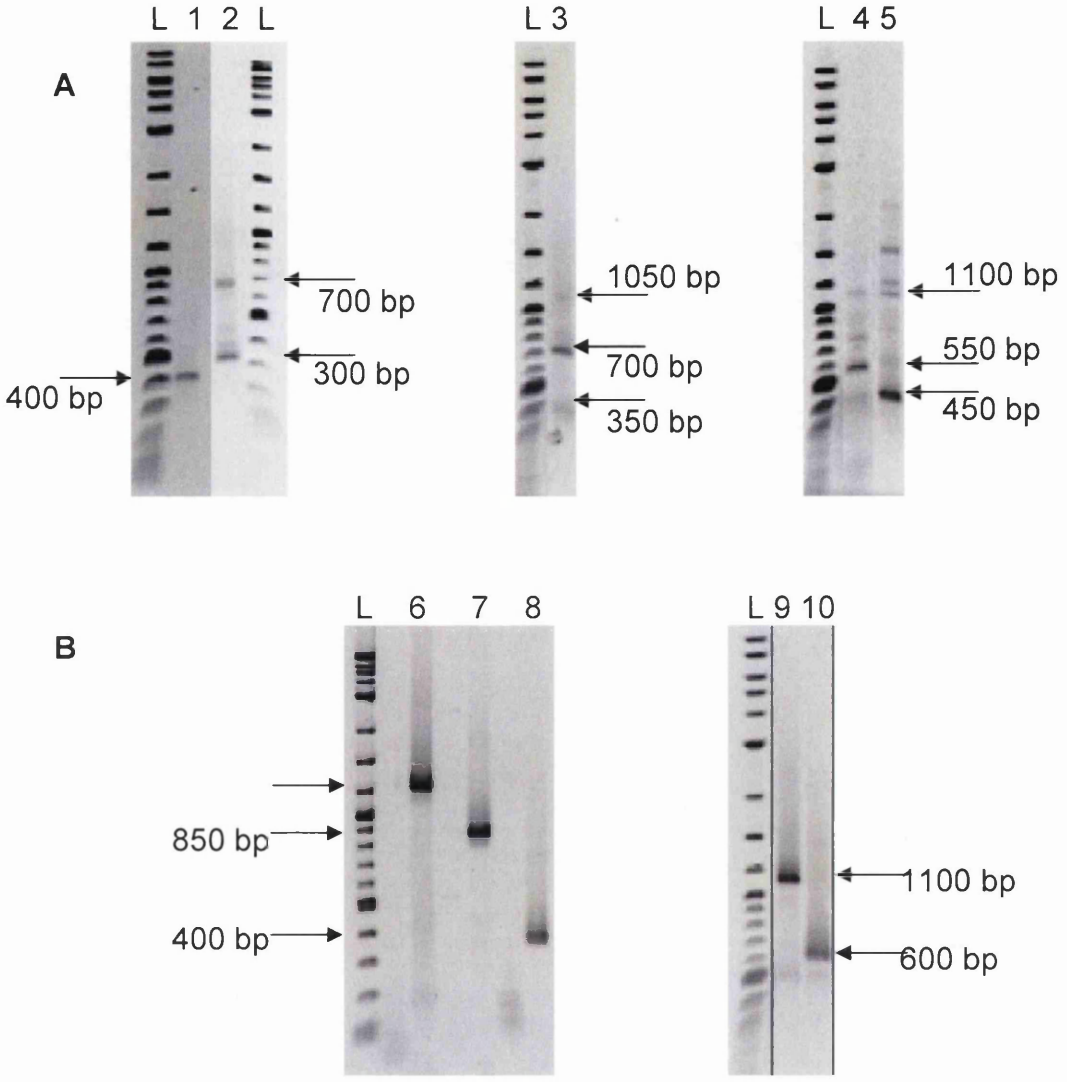


Figure 5.8. 3'- and 5'-RACE amplification of the ends of RNAs in region G and flanking ORFs.

(A) 5'-RACE amplification products. Lanes 1-5 show RACE products amplified using primers specific to US33A, US34, US34A, US31 and US32 respectively (Table 5.2). **(B)** 3'-RACE amplification products. Lanes 6-10 show RACE products amplified using primers specific to US33A, US34, US34A, US31 and US32, respectively (Table 5.2). Lanes L are DNA markers. Sequencing of additional bands in these gels showed that these were non-specific amplifications.

	5'-RACE		3'-RACE	
ORF	**primer**	**product sizes (bp)**	**primer**	**product sizes (bp)**
US31	G1R	550	G23	1100
US32	G15	450, 1100	G6	600
US33A	G28	400	G29	1200
US34	G7	300, 700	G14	850
US34A	G27	350, 700, 1050	G4R	400

Table 5.2. 3'- and 5'-RACE amplification of the ends of RNAs in region G and flanking ORFs.

The primers used for RACE amplification are detailed in Appendix. The approximate product sizes refer to the bands detected in the agarose gels that correspond to the predicted 5'- and 3'-ends of the ORFs. Multiple product sizes indicate the primer detecting multiple 5'-ends, from overlapping transcripts.

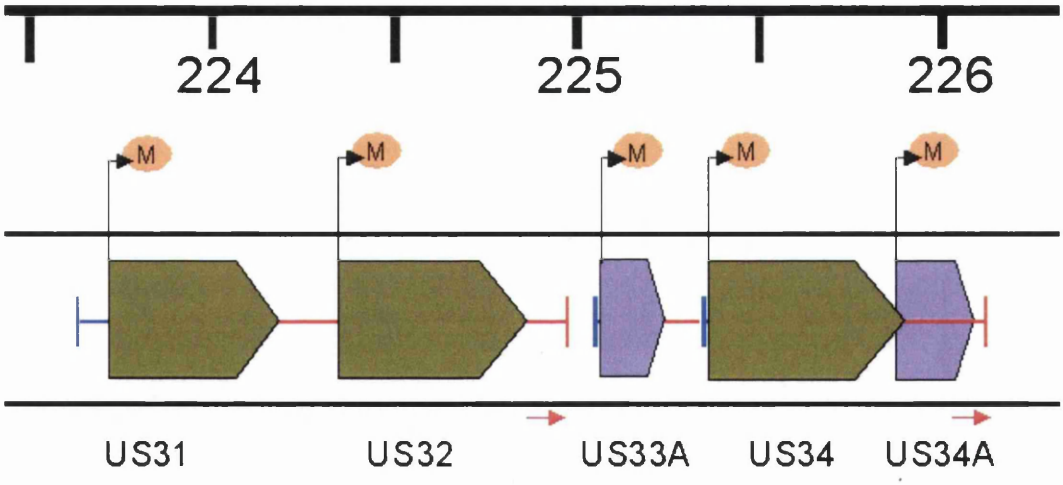


Figure 5.9. Summary of RACE transcript mapping in region G and flanking ORFs.

RACE transcript mapping identified two sets of 3'-co-terminal ORFs: US33A, US34 and US34A in region G, and US31 and US32 flanking region G. Blue bars show the 5'-ends of the ORFs and red bars show the 3'-ends of the ORFs. We were unable to clone and sequence the 5'-ends of US32 and US34A, so these ends could not be determined.

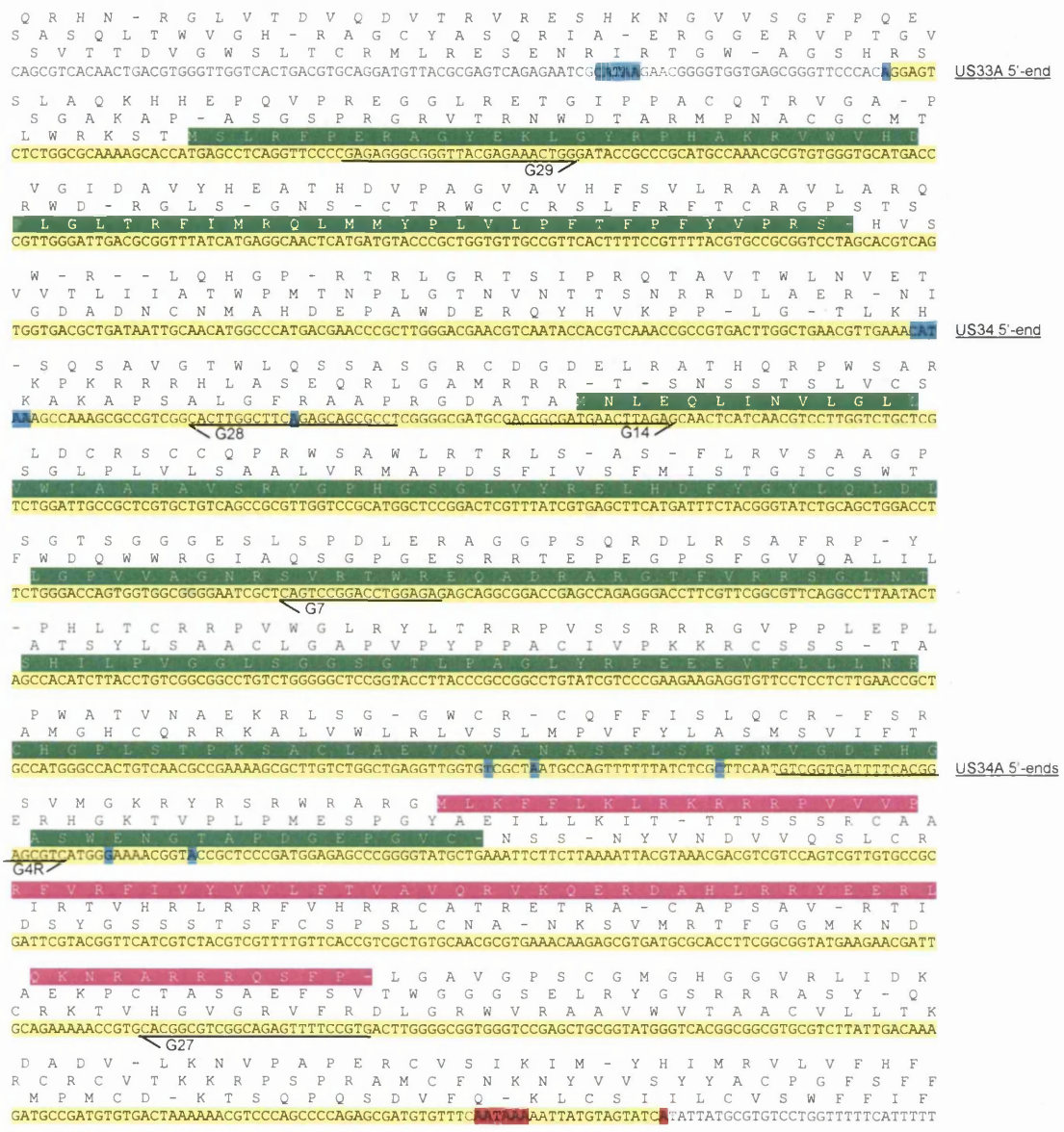


Figure 5.10. Three-frame translation of co-terminal RNAs US33A and US34.

Yellow shows transcribed sequence. At the 3'-end of the co-terminal RNAs, red shows the poly(A) signal (AATAAA) and the downstream RNA cleavage site. At the 5'-ends, sky blue shows the residue mapped at the 5'-end of each ORF and the upstream TATA boxes (CATAA) of each. The five different 5'-ends determined for US34A are also shown in blue. Green shows the amino acid sequences of the proteins encoded by US33A and US34. Purple shows the protein predicted to be encoded by US34A and blue shows the different residues mapped for the 5'-end of this ORF. Also shown are primers used for RACE-PCR.

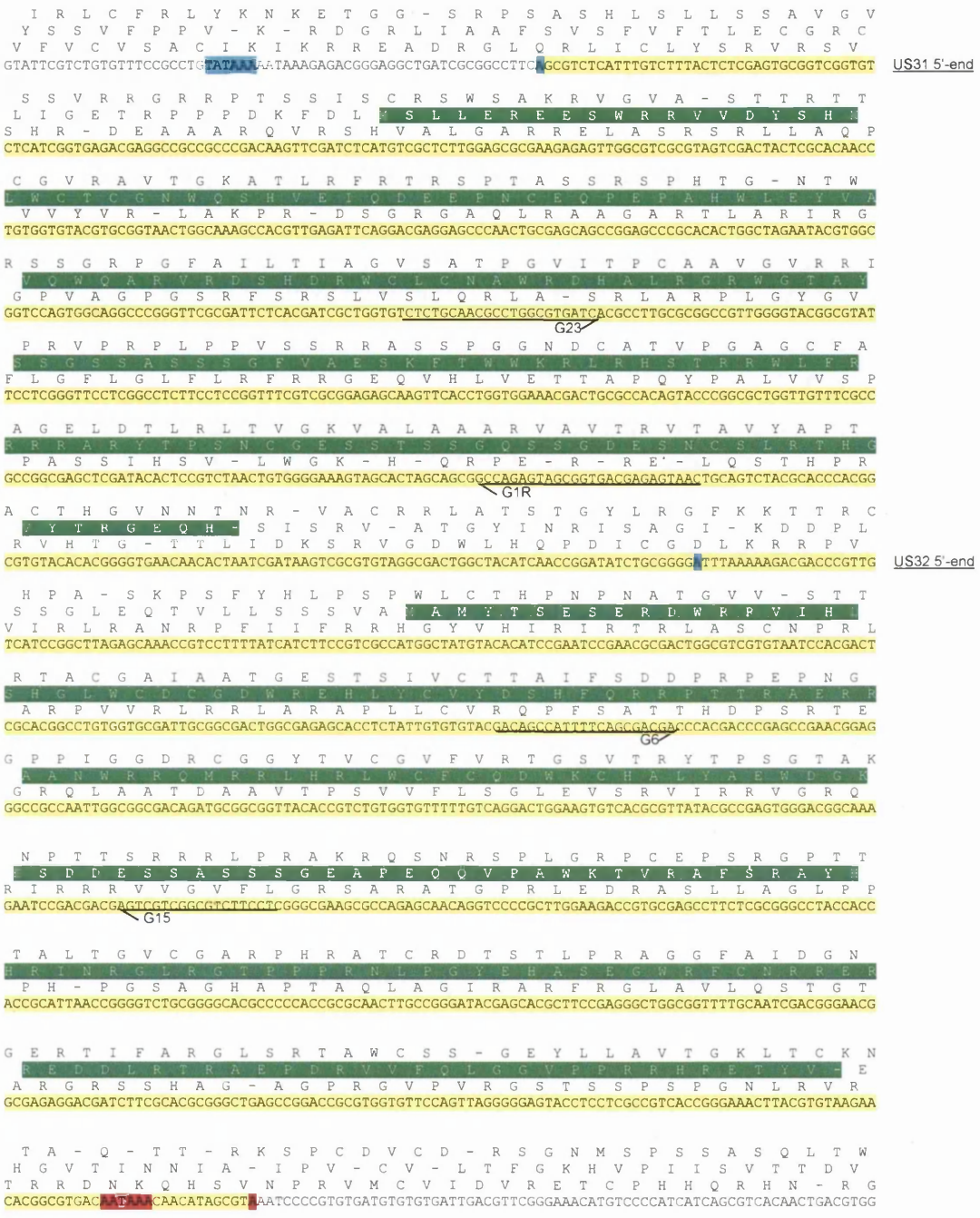


Figure 5.11. Three-frame translation of co-terminal RNAs US31 and US32.

Yellow shows transcribed sequence. At the 3'-end of the co-terminal RNAs, red shows the poly(A) signal (AATAAA) and the downstream RNA cleavage site. At the 5'-ends, sky blue shows the residue mapped at the 5'-end of each ORF and the upstream TATA box for US31. Green shows the amino acid sequences of the proteins encoded by US31 and US32. Also shown are primers used for RACE-PCR.

The amino acid sequences encoded by these co-terminal transcripts are shown in Figure 5.10. From the RACE analysis, US34 was predicted to encode a 163 aa protein, as predicted by Chee *et al.* (1990). This protein is conserved in CCMV (Davison *et al.*, 2003a). RACE analysis also mapped the ends of US33A, which was predicted to encode a small protein (57 aa). BLAST analysis of the US33A protein did not show significant similarities to other proteins in the NCBI database. However, a counterpart of US33A was identified in CCMV. The CCMV ORF is of similar size and position and contains a CATAA motif upstream of the potential 5'-end (Figure 5.12). These represent novel ORFs in HCMV and CCMV.

RACE analysis determined multiple 5'-ends for US34A, a novel ORF in region G proposed by Davison et al. (2003a). However, each 5'-end was determined from a single clone, and consequently a definitive 5'-end could not be identified. Furthermore, none of these candidate 5'-ends was associated with TATA elements. Therefore, although HCMV US34A has a counterpart in CCMV has been identified (Davison et al., 2003a), transcript mapping did not yield evidence for its expression.

US31 and US32 were also predicted to encode co-terminal messages and these genes encode related proteins in the US1 family. Primers G23 and G6 were used to amplify 3'-RACE products, and primers G1R and G15 were used to amplify 5'-RACE products (Table 5.2 and Figure 5.8). Again, each RACE-PCR product was gel purified and cloned to ensure that an accurate 3'- or 5'-end was obtained. The two 3'-RACE amplification products for these ORFs mapped to a T residue at nucleotide position 224,901 in AD169, 17 bp downstream of a poly(A) signal at position 235,675. The 5'-end of US31 mapped to an A residue at position 223,634, 35 bp downstream from a consensus TATA box. However, only one plasmid clone was generated from the 5'-RACE products for US32, mapping to an A residue at position 224,252. This site is not associated with a TATA box and is therefore not a convincing 5'-end. The reading frames of these co-terminal transcripts are shown in Figure 5.11. From the RACE analysis, US31 was predicted to encode a 162 aa protein, as predicted for the revised form of the ORF (Davison *et al.*, 2003).

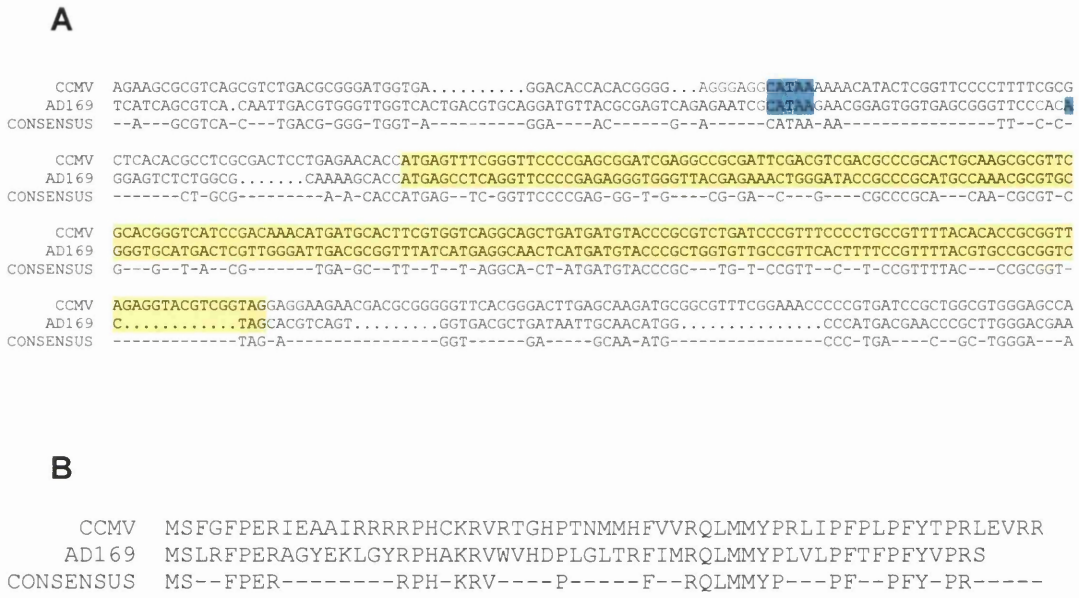


Figure 5.12. DNA and amino acid sequence alignments of HCMV and CCMV US33A.

(A) DNA sequence alignment. Yellow shows the protein-coding regions in each genome. Blue shows the 5'-end of US33A in HCMV and the conserved TATA box element in both genomes. **(B)** Amino acid alignment indicating 47% identity.

5.5 Discussion

The original analysis of HCMV strain AD169 assigned four ORFs to region G: US33-US36 (Chee *et al.*, 1990). Comparisons with the CCMV genome led to the discounting of US33, US35 and US36 (Davison *et al.*, 2003a). This study also introduced a new ORF to the region: US34A. Of the ORFs flanking region G, only TRS1 has been functionally characterised (Romanowski and Shenk, 1997).

Sequence analysis of region G and the flanking ORFs showed that US31-US34A and TRS1 are well conserved and not disrupted by frameshifts. However, US35 and US36 are not intact in all HCMV strains, being disrupted by frameshifts in certain strains in comparison with AD169 (Figure 5.3). Neither of these ORFs commence with ATG codons, and are instead preceded by termination codons, containing initiation codons internally. A high level of variation between strains was observed in US35-US36, and frameshifts were identified throughout. Moreover, no novel ORFs emerged in the region between US34A and TRS1 as potentially encoding proteins. Also, RNAs specified by these ORFs were not detected in previous transcript mapping studies of the region, or in the current study. Therefore, it is very unlikely that ORFs US35 and US36 actually encode proteins.

A region corresponding to region G can be identified in CCMV, the closest known relative of HCMV. Corresponding regions cannot be identified in more distantly related CMV genomes. These genomes are more diverged and contain a single unique region in contrast to the U_L and U_S regions of HCMV and CCMV, which are divided by repeat regions. CCMV is only moderately diverged and is essentially collinear to HCMV, and the corresponding region in CCMV is similar in size and position (Davison *et al.*, 2003a). DNA sequence comparisons between HCMV and CCMV sequences in region G are shown as a two-dimensional comparison on a matrix plot in Figure 5.11A. CCMV shares significant sequence similarity with HCMV in the corresponding genes, but not in US35-36.

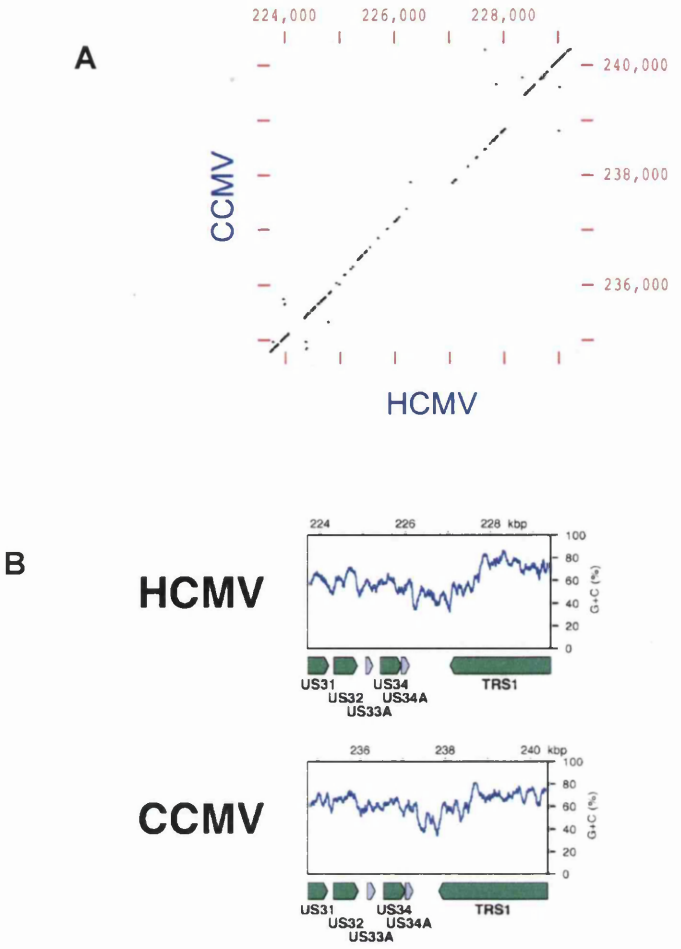


Figure 5.11. Sequence comparisons and nucleotide composition of region G in HCMV and CCMV.

(A) The sequence comparison is displayed in a matrix plot computed using Compare (a window of 21 nucleotides with a minimum match of 17 nucleotides) and displayed using Dotplot. **(B)** G+C plots computed using Window (100 nucleotides shifted by 3 nucleotide increments) and displayed using Statplot.

The G+C content of US31-US34A and TRS1 is similar to the rest of the HCMV genome. However, in the region between US34A and TRS1, the G+C content is very low (Figure 5.11B). The G+C content is also very low in the additional sequence in strain 6397. Within this area of low G+C content lie three G:C polynucleotide tracts of eight or more residues (four in strain 6397). The only other G:C tracts of eight or more residues in the HCMV genome are found in a group within *ori*Lyt (Chapter 4). A similar G+C distribution is exhibited in the corresponding region of CCMV (Figure 5.12).

US33 was discounted by the analysis of Davison *et al.* (2003a). However, in contrast to US35 and US36, US33 is well conserved (Figure 5.3), and has a G+C content similar to other HCMV genes. Although US33 does not have a counterpart in CCMV, much of the sequence in this region is conserved between the two genomes (Figure 5.11A). Transcript analysis identified a 1.2 kb RNA on the opposite strand from UL33. The 3'-end of this transcript (US33A) was mapped, showing the RNA to be 3'-co-terminal with US34. The 5'-end was mapped approximately 340 bp upstream from the start site of US34, and 31 bp downstream from a TATA box (CATAA). The predicted US33A protein did not show strong similarities to other proteins in the NCBI database. Significantly, a counterpart was identified in CCMV, where the ORF is of similar size. US33A is therefore a candidate for a novel HCMV protein-coding ORF.

US34A was incorporated previously because a counterpart was identified in CCMV (Davison *et al.*, 2003a). Sequence comparisons show that this ORF is not frameshifted in seven HCMV strains. However, a definitive transcript corresponding to U34A was not determined by northern blot. Also, although a 5'-RACE product corresponding to the ORF was amplified, five 5'-ends were identified from five clones, none associated with TATA elements. Despite the presence of a US34A homologue in CCMV, the lack of convincing evidence from transcript analysis for transcription of HCMV US34A casts doubt on this ORF as protein-coding. This leaves US34 as 3'-co-terminal with US33A.

US31 and US32 were mapped in the current study as 3'-co-terminal genes that belong to the US1 gene family. The 5'-end of the US31 mRNA agreed with the revised form of the ORF (Davison *et al.*, 2003a). Although a definitive 5'-end was not identified for US32, its relationship to US1 and US31 and the northern blot analysis indicate that this ORF encodes a functional protein.

CHAPTER 6: DISCUSSION

The original genome analysis of HCMV strain AD169 was carried out by Chee *et al.* (1990). Subsequent studies have characterised many HCMV genes, modifying the interpretation of some and identifying novel genes. This has led to a more detailed and accurate picture of the genetic content of HCMV. Comparative genomics is one of the most useful ways of annotating genes, on the basis that authentic protein-coding regions will have been conserved during evolution whereas spurious features, such as non-functional ORFs, will not. The aim of the present study was to investigate the protein-coding potential of three regions in the HCMV genome, each of which contains ORFs identified originally by Chee *et al.* (1990) that were subsequently discounted by a comparative analysis with CCMV (Davison *et al.*, 2003a). Region X (Chapter 3) is located between UL105 and UL112, region O (Chapter 4) between UL57 and UL69, and region G (Chapter 5), between US32 and TRS1. The genes flanking these regions are accepted as encoding functional proteins. In addition to containing ORFs discounted by the CCMV analysis, all three regions have been predicted to contain novel, small ORFs potentially encoding proteins (Murphy *et al.*, 2003a, 2003b; Davison *et al.*, 2003a). Only US34A in region G has a counterpart in CCMV (Davison *et al.*, 2003a).

Table 6.1 summarises the results of sequence comparisons and transcript mapping of the ORFs in regions X, O and G, and gives an evaluation of whether the ORFs are likely to encode functional proteins. In regions X and O, frameshifts affect most of the small ORFs identified by Chee *et al.* (1990) and Murphy *et al.* (2003a, 2003b). Only four ORFs (UL66, UL108, C-ORF16 and C-ORF18) are unaffected by frameshifts. None of these are conserved in CCMV. In regions X and O, sequence analysis of regions not affected by frameshift mutations revealed no novel ORFs potentially encoding proteins. Although the presence of highly spliced mRNAs with small exons cannot be completely ruled out, careful inspection of the sequence did not reveal candidate splice donor and acceptor sites. Furthermore, no transcripts were identified in region X other than the 1.1 and 4.6 kb RNAs, representing the two spliced exons and intron generated from a single primary transcript covering the entire region.

Region	ORF	Conserved in: HCMV	CCMV	Transcript identified	Likelihood of protein-coding
X	UL106	×	×	×	×
	UL107	×	×	×	×
	C-ORF16	✓	×	×	×
	UL108	✓	×	×	×
	UL109	×	×	×	×
	UL110	×	×	×	×
	C-ORF17	×	×	×	×
	UL111	×	×	×	×
	C-ORF18	✓	×	×	×
O	UL58	×	×	×	×
	UL59	×	×	×	×
	UL60	×	×	×	×
	UL61	×	×	×	×
	UL62	×	×	×	×
	ORF3	×	×	×	×
	ORF4	×	×	×	×
	UL63	×	×	×	×
	ORF5	×	×	×	×
	UL64	×	×	×	×
	UL65	×	×	×	×
	UL66	✓	×	×	×
	UL67	×	×	×	×
	UL68	×	×	×	×
G	US31	✓	✓	✓	✓
	US32	✓	✓	✓	✓
	US33	✓	×	×	×
	US33A	✓	✓	✓	✓
	US34	✓	✓	✓	✓
	US34A	✓	✓	×	?
	US35	×	×	×	×
	US36	×	×	×	×

Table 6.1. Evaluation of the protein-coding potential of ORFs in regions X, O and G.

✓ is yes, × is no, ? is marginal.

If regions X and O, which are approximately 4.9 and 8 kbp in size respectively, do not contain protein-coding genes, what are their functions? In region X, the splice sites for generating the equivalent of the 1.1 and 4.6 kb RNAs are conserved in CCMV, and large transcripts in the corresponding region were reported in RhCMV and MCMV by Kulesza and Shenk (2004) during the course of the present study. As it is unlikely that a functional protein is encoded by the 1.1 kb RNA (Section 3.6), it seems probable that one of these RNAs has a non-translated role. This role is unlikely to be essential to viral replication *in vitro*, as a mutant lacking the splice donor site was shown to replicate efficiently in cultured fibroblasts (Kulesza and Shenk, 2004). The conclusion of Kulesza and Shenk (2004), that the 4.6 kb RNA represents a stable intron has parallels with the latency-associated transcripts (LATs) of HSV-1. The most abundant LAT is a stable 2 kb intron, spliced from an 8.5 kb transcript present in the long repeats of HSV-1, although alternative splicing produces further transcripts of 1.4 kb and 1.5 kb (Farrell *et al.*, 1991; Krummenacher *et al.*, 1997; Mador *et al.*, 1995). The 2 kb LAT does not encode a protein (Lock *et al.*, 2001). LAT is the major viral transcript detected by *in situ* hybridisation of latently infected neurons (Croen *et al.* 1987; Stevens *et al.* 1988; Krause *et al.* 1988, 1991). LAT is reported to have a number of functions, one of which is promoting latency by protecting neuronal cells from apoptosis (Perng *et al.*, 2000; Thompson and Sawtell, 2001; Ahmed *et al.*, 2002). Roles in regulating translation and splicing have also been suggested (Ahmed *et al.*, 2001). It is possible that the 4.6 kb intron in HCMV region X has a function similar to one of those of the LAT transcript. A study by Goodrum *et al.* (2002) suggested that the 4.6 kb RNA (detected using a UL110-specific probe) may be expressed during latency (in CD34+ cells). However, this study detected gene expression from many ORFs throughout the HCMV genome, in contrast to other studies of HCMV latency. Therefore, investigating the expression of the 1.1 kb or the 4.6 kb RNA during HCMV latency would be a useful direction in determining their function.

A non-coding viral function is also likely to be employed by pp67, a transcript originating from region O. Sequence comparisons in the present study suggest that pp67 does not encode a protein, in contrast to the conclusions of Davis *et al.* (1984, 1985). Sequence analysis indicates that it is unlikely that any protein-coding ORFs

are present in the region, but this does not rule out the possibly of other transcripts in this part of region O, which could be investigated by transcript mapping.

The last few years have seen a large increase in the number of RNA transcripts detected that have no apparent protein-coding potential. These noncoding RNAs (ncRNAs) have roles in a variety of biological processes, including transcriptional regulation, chromosome replication, RNA processing and modification, mRNA stability (Eddy, 1999, 2002; Huttenhofer *et al.*, 2002, 2005; Storz, 2002). Consequently, RNAs that do not function via translation may be far more abundant than previously predicted. At least four mammalian genes have been identified as carrying a number of intron-encoded ncRNA genes: *gas5*, *UHG*, *U17HG* and *U19H* (Eddy, 1999). It is possible the 5 kb RNA could have a similar role, although ncRNAs associated with region X were not identified in studies of HCMV microRNAs. Pfeffer *et al.* (2005) identified nine HCMV miRNAs, spread across the viral genome. Grey *et al.* (2005) characterised three of these by northern blot analysis, and identified a further ten HCMV-encoded miRNAs. None of the miRNAs identified map to regions X, O or G. Clearly, much work is still to be done to elucidate the functions, if any, of HCMV miRNAs.

Unusual nucleotide composition and a high prevalence of homopolymeric tracts are characteristic features of regions X, O and G. Table 6.2 shows the distribution of the polynucleotide tracts in the Merlin genome. Although there are relatively few G:C tracts in the genome, all are found in apparently non-protein coding regions. A:T tracts, especially those of nine or more residues, are also found predominantly in non-protein coding regions of the genome. Therefore, it would appear that there is a form of evolutionary selection against long homopolymeric tracts in protein-coding genes, since slippage during DNA replication could promote length instability. Region X consists of local regions of high G+C content embedded in sequences of low G+C content (Figure 3.12), and region O consists of high G+C content in the *ori*Lyt region and lower G+C content elsewhere (Figure 4.7). In addition, the two ORFs discounted in region G, US35 and US36, are found in a highly variable region of approximately 1 kb which also exhibits low G+C content (Figure 5.11).

Polynucleotide tract		Number of polynucleotide tracts in:				
					Rest of genome:	
Minimum length	Type	Region X	Region O	Region G	Protein-coding sequence	Non-protein-coding sequence
8	A:T	11	10	1	5	36
	G:C	0	3	3	0	1
9	A:T	8	6	0	1	26
	G:C	0	2	3	0	0
10	A:T	4	2	0	0	10
	G:C	0	1	3	0	0

Table 6.2. Distribution of polynucleotide tracts in the Merlin genome.

Sequence comparisons of seven HCMV strains in three regions of the genome have confirmed and extended the view of the genetic content of HCMV from comparisons with CCMV, by indicating that many ORFs are unlikely to encode proteins. As discussed in Chapter 4, sequencing across region O is incomplete due to difficulties caused by high G+C content and highly repetitive and palindromic sequence. However, to extend this part of the current study, transcript analysis of region O should be carried out, using northern blot and RACE analysis to identify and map any further transcripts in the region. Outwith the minimum *ori*Lyt region, one other transcript (pp67, discussed in Chapter 4) has been detected in region O. Transcript mapping in this region may identify further transcripts in this region, identifying it a source of important non-coding, functional RNAs in HCMV.

A novel, spliced transcript from region X was characterised and most likely has a role as a non-coding functional RNA. As previously discussed, the spliced transcript has parallels with the LATs of HSV-1 and further work would be required to ascertain if this spliced RNA has a role in HCMV latency. Alternatively, this RNA, and the pp67 RNA detected in region O, may have other roles. Functions such as transcriptional activation, regulation of gene expression, RNA processing, gene silencing, and even

genomic imprinting have been described for non-coding RNAs in eukaryotes and one of these functions could be performed by the RNAs transcribed in regions X and O.

Alternatively, it is possible that regions X and O may have a *cis*-acting function in HCMV, as both regions are well conserved between strains. To date, one origin of replication has been identified in HCMV (*ori*Lyt, discussed in Section 4). In contrast, three origins have been identified in HSV-1. It is possible that within region X, or within the right half of region O (outwith the HCMV *ori*Lyt region), there could be important *cis*-acting sequences which have a function in an alternative origin of replication. This role could be investigated using replication assays.

The non-coding area of region G is different to the non-coding areas of regions X and O in both size and sequence conservation. The non-coding area of region G is smaller, at only 1 kbp, and is not nearly as well conserved between different strains of HCMV as regions X and O are. In addition, no transcripts have been detected across this area of region G in the current analyses, whereas transcripts have been detected in regions X (1.1 kb and 5 kb RNAs) and O (pp67). Although the lack of transcripts across this region may suggest a *cis*-acting function, the lack of sequence conservation casts doubt on this assumption. It is possible that this 1 kbp region may harbour micro-RNAs previously unidentified in HCMV, or perhaps it is simply non-functional, and an example of 'junk' DNA in herpesvirus genomes.

In the case of US33A in region G, further experiments, such as western blotting, are required to detect and characterise any encoded protein. This would complement the transcript mapping data of the current study, which has detected this novel transcript by northern blot analysis, and mapped the 5'- and 3'-ends. Like most other genes in the U_S region of the HCMV genome, US33A is unlikely to be an essential or core gene, and BLAST analysis of the US33A protein did not show significant similarities to other proteins in the NCBI database. Although no function is therefore suggested from the protein sequence, US33A is detected as an E transcript and may have a similar role to other E proteins, such as a glycoprotein or an enzyme involved in DNA replication or nucleotide metabolism.

Adler, S. P., Finney, J. W., Manganello, A. M. & Best, A. M. (1996). Prevention of child-to-mother transmission of cytomegalovirus by changing behaviors: a randomized controlled trial. *Pediatr Infect Dis J* **15**, 240-6.

Ahmed, M. & Fraser, N. W. (2001). Herpes simplex virus type 1 2-kilobase latency-associated transcript intron associates with ribosomal proteins and splicing factors. *J Virol* **75**, 12070-80.

Ahmed, M., Lock, M., Miller, C. G. & Fraser, N. W. (2002). Regions of the herpes simplex virus type 1 latency-associated transcript that protect cells from apoptosis in vitro and protect neuronal cells in vivo. *J Virol* **76**, 717-29.

Ahn, K., Angulo, A., Ghazal, P., Peterson, P. A., Yang, Y. & Fruh, K. (1996). Human cytomegalovirus inhibits antigen presentation by a sequential multistep process. *Proc Natl Acad Sci U S A* **93**, 10990-5.

Akter, P., Cunningham, C., McSharry, B. P., Dolan, A., Addison, C., Dargan, D. J., Hassan-Walker, A. F., Emery, V. C., Griffiths, P. D., Wilkinson, G. W. & Davison, A. J. (2003). Two novel spliced genes in human cytomegalovirus. *J Gen Virol* **84**, 1117-22.

Albrecht, T., Cavallo, T., Cole, N. L. & Graves, K. (1980). Cytomegalovirus: development and progression of cytopathic effects in human cell culture. *Lab Invest* **42**, 1-7.

Anders, D. G., Kacica, M. A., Pari, G. & Punturieri, S. M. (1992). Boundaries and structure of human cytomegalovirus oriLyt, a complex origin for lytic-phase DNA replication. *J Virol* **66**, 3373-84.

Anders, D. G. & Punturieri, S. M. (1991). Multicomponent origin of cytomegalovirus lytic-phase DNA replication. *J Virol* **65**, 931-7.

Apperley, J. F. & Goldman, J. M. (1988). Cytomegalovirus: biology, clinical features and methods for diagnosis. *Bone Marrow Transplant* **3**, 253-64.

Arav-Boger, R., Willoughby, R. E., Pass, R. F., Zong, J. C., Jang, W. J., Alcendor, D. & Hayward, G. S. (2002). Polymorphisms of the cytomegalovirus (CMV)-encoded tumor necrosis factor-alpha and beta-chemokine receptors in congenital CMV disease. *J Infect Dis* **186**, 1057-64.

Arribas, J. R., Storch, G. A., Clifford, D. B. & Tselis, A. C. (1996). Cytomegalovirus encephalitis. *Ann Intern Med* **125**, 577-87.

Bahr, U. & Darai, G. (2001). Analysis and characterization of the complete genome of tupaia (tree shrew) herpesvirus. *J Virol* **75**, 4854-70.

Bahr, U. & Darai, G. (2004). Re-evaluation and in silico annotation of the Tupaia herpesvirus proteins. *Virus Genes* **28**, 99-120.

Baines, J. D., Ward, P. L., Campadelli-Fiume, G. & Roizman, B. (1991). The UL20 gene of herpes simplex virus 1 encodes a function necessary for viral egress. *J Virol* **65**, 6414-24.

Baldanti, F., Lurain, N. & Gerna, G. (2004). Clinical and biologic aspects of human cytomegalovirus resistance to antiviral drugs. *Hum Immunol* **65**, 403-9.

Bale, J. F., Jr., Blackman, J. A. & Sato, Y. (1990). Outcome in children with symptomatic congenital cytomegalovirus infection. *J Child Neurol* **5**, 131-6.

Bale, J. F., Jr., Petheram, S. J., Robertson, M., Murph, J. R. & Demmler, G. (2001). Human cytomegalovirus a sequence and UL144 variability in strains from infected children. *J Med Virol* **65**, 90-6.

Bar, M., Shannon-Lowe, C. & Geballe, A. P. (2001). Differentiation of human cytomegalovirus genotypes in immunocompromised patients on the basis of UL4 gene polymorphisms. *J Infect Dis* **183**, 218-225.

Batterson, W., Furlong, D. & Roizman, B. (1983). Molecular genetics of herpes simplex virus. VIII. further characterization of a temperature-sensitive mutant defective in release of viral DNA and in other stages of the viral reproductive cycle. *J Virol* **45**, 397-407.

Beaulieu, B. L. & Sullivan, J. L. (1997). Epstein-Barr Virus. In *Clinical Virology*, pp. 485-507: Churchill Livingstone.

Bego, M., Maciejewski, J., Khaiboullina, S., Pari, G. & St Jeor, S. (2005). Characterization of an antisense transcript spanning the UL81-82 locus of human cytomegalovirus. *J Virol* **79**, 11022-34.

Benedict, C. A., Butrovich, K. D., Lurain, N. S., Corbeil, J., Rooney, I., Schneider, P., Tschopp, J. & Ware, C. F. (1999). Cutting edge: a novel viral TNF receptor superfamily member in virulent strains of human cytomegalovirus. *J Immunol* **162**, 6967-70.

Bernard, J. & Mercier, A. (1993). Sequence of two Eco RI fragments from Salmonis herpesvirus 2 and comparison with Ictalurid herpesvirus 1. *Arch Virol* **132**, 437-42.

Bhella, D., Rixon, F. J. & Dargan, D. J. (2000). Cryomicroscopy of human cytomegalovirus virions reveals more densely packed genomic DNA than in herpes simplex virus type 1. *J Mol Biol* **295**, 155-61.

Biegalke, B. J. (1999). Human cytomegalovirus US3 gene expression is regulated by a complex network of positive and negative regulators. *Virology* **261**, 155-64.

Blankenship, C. A. & Shenk, T. (2002). Mutant human cytomegalovirus lacking the immediate-early TRS1 coding region exhibits a late defect. *J Virol* **76**, 12290-9.

Blok, M. J., Goossens, V. J., Vanherle, S. J., Top, B., Tacken, N., Middeldorp, J. M., Christiaans, M. H., van Hooff, J. P. & Bruggeman, C. A. (1998). Diagnostic value of monitoring human cytomegalovirus late pp67 mRNA expression in renal-allograft recipients by nucleic acid sequence-based amplification. *J Clin Microbiol* **36**, 1341-6.

Bodaghi, B., Jones, T. R., Zipeto, D., Vita, C., Sun, L., Laurent, L., Arenzana-Seisdedos, F., Virelizier, J. L. & Michelson, S. (1998). Chemokine sequestration by viral chemoreceptors as a novel viral escape strategy: withdrawal of chemokines from the environment of cytomegalovirus-infected cells. *J Exp Med* **188**, 855-66.

Bodaghi, B., Slobbe-van Drunen, M. E., Topilko, A., Perret, E., Vossen, R. C., van Dam-Mieras, M. C., Zipeto, D., Virelizier, J. L., LeHoang, P., Bruggeman, C. A. & Michelson, S. (1999). Entry of human cytomegalovirus into retinal pigment epithelial and endothelial cells by endocytosis. *Invest Ophthalmol Vis Sci* **40**, 2598-607.

Boeckh, M., Gooley, T. A., Myerson, D., Cunningham, T., Schoch, G. & Bowden, R. A. (1996). Cytomegalovirus pp65 antigenemia-guided early treatment with ganciclovir versus ganciclovir at engraftment after allogeneic marrow transplantation: a randomized double-blind study. *Blood* **88**, 4063-71.

Bold, S., Ohlin, M., Garten, W. & Radsak, K. (1996). Structural domains involved in human cytomegalovirus glycoprotein B-mediated cell-cell fusion. *J Gen Virol* **77 (Pt 9)**, 2297-302.

Boppana, S. B., Pass, R. F., Britt, W. J., Stagno, S. & Alford, C. A. (1992). Symptomatic congenital cytomegalovirus infection: neonatal morbidity and mortality. *Pediatr Infect Dis J* **11**, 93-9.

Borst, E. M., Mathys, S., Wagner, M., Muranyi, W. & Messerle, M. (2001). Genetic evidence of an essential role for cytomegalovirus small capsid protein in viral growth. *J Virol* **75**, 1450-8.

Borysiewicz, L. K., Rodgers, B., Morris, S., Graham, S. & Sissons, J. G. (1985). Lysis of human cytomegalovirus infected fibroblasts by natural killer cells: demonstration of an interferon-independent component requiring expression of early viral proteins and characterization of effector cells. *J Immunol* **134**, 2695-701.

Bresnahan, W. A. & Shenk, T. (2000). A subset of viral transcripts packaged within human cytomegalovirus particles. *Science* **288**, 2373-6.

Britt, W. J. & Boppana, S. (2004). Human cytomegalovirus virion proteins. *Hum Immunol* **65**, 395-402.

Britt, W. J. & Mach, M. (1996). Human cytomegalovirus glycoproteins. *Intervirology* **39**, 401-12.

Brocchieri, L., Kledal, T. N., Karlin, S. & Mocarski, E. S. (2005). Predicting coding potential from genome sequence: application to betaherpesviruses infecting rats and mice. *J Virol* **79**, 7570-96.

Brown, J. M., Kaneshima, H. & Mocarski, E. S. (1995). Dramatic interstrain differences in the replication of human cytomegalovirus in SCID-hu mice. *J Infect Dis* **171**, 1599-603.

Brown, J. C., McVoy, M. A. & Homa, F. L. (2002). Packaging DNA into Herpesvirus Capsids. In *Structure-Function Relationships of Human Pathogenic Viruses*, pp. 111-53: Kluwer Academic/Plenum Publishers.

Bukowski, J. F., Warner, J. F., Dennert, G. & Welsh, R. M. (1985). Adoptive transfer studies demonstrating the antiviral effect of natural killer cells in vivo. *J Exp Med* **161**, 40-52.

Bukowski, J. F., Woda, B. A. & Welsh, R. M. (1984). Pathogenesis of murine cytomegalovirus infection in natural killer cell-depleted mice. *J Virol* **52**, 119-28.

Burke, T. W. & Kadonaga, J. T. (1997). The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. *Genes Dev* **11**, 3020-31.

Butcher, S. J., Aitken, J., Mitchell, J., Gowen, B. & Dargan, D. J. (1998). Structure of the human cytomegalovirus B capsid by electron cryomicroscopy and image reconstruction. *J Struct Biol* **124**, 70-6.

Cai, W., Astor, T. L., Liptak, L. M., Cho, C., Coen, D. M. & Schaffer, P. A. (1993). The herpes simplex virus type 1 regulatory protein ICP0 enhances virus replication during acute infection and reactivation from latency. *J Virol* **67**, 7501-12.

Campadelli-Fiume, G., Cocchi, F., Menotti, L. & Lopez, M. (2000). The novel receptors that mediate the entry of herpes simplex viruses and animal alphaherpesviruses into cells. *Rev Med Virol* **10**, 305-19.

Carrigan, D. R., Drobyski, W. R., Russler, S. K., Tapper, M. A., Knox, K. K. & Ash, R. C. (1991). Interstitial pneumonitis associated with human herpesvirus-6 infection after marrow transplantation. *Lancet* **338**, 147-9.

Cha, T. A., Tom, E., Kemble, G. W., Duke, G. M., Mocarski, E. S. & Spaete, R. R. (1996). Human cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains. *J Virol* **70**, 78-83.

Chambers, J., Angulo, A., Amaratunga, D., Guo, H., Jiang, Y., Wan, J. S., Bittner, A., Frueh, K., Jackson, M. R., Peterson, P. A., Erlander, M. G. & Ghazal, P. (1999). DNA microarrays of the complex human cytomegalovirus genome: profiling kinetic class with drug sensitivity of viral gene expression. *J Virol* **73**, 5757-66.

Chandler, S. H., Holmes, K. K., Wentworth, B. B., Gutman, L. T., Wiesner, P. J., Alexander, E. R. & Handsfield, H. H. (1985). The epidemiology of cytomegaloviral infection in women attending a sexually transmitted disease clinic. *J Infect Dis* **152**, 597-605.

Chang, Y. E., Van Sant, C., Krug, P. W., Sears, A. E. & Roizman, B. (1997). The null mutant of the UL31 gene of herpes simplex I: construction and phenotype in infected cells. *J Virol* **71**, 8307-15.

Chee, M. S., Bankier, A. T., Beck, S., Bohni, R., Brown, C. M., Cerny, R., Horsnell, T., Hutchison, C. A., 3rd, Kouzarides, T., Martignetti, J. A. & et al. (1990). Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. *Curr Top Microbiol Immunol* **154**, 125-69.

Cheeran, M. C., Hu, S., Sheng, W. S., Peterson, P. K. & Lokensgard, J. R. (2003). CXCL10 production from cytomegalovirus-stimulated microglia is regulated by both human and viral interleukin-10. *J Virol* **77**, 4502-15.

Chen, D. H., Jiang, H., Lee, M., Liu, F. & Zhou, Z. H. (1999). Three-dimensional visualization of tegument/capsid interactions in the intact human cytomegalovirus. *Virology* **260**, 10-6.

Cheung, T. W. & Teich, S. A. (1999). Cytomegalovirus infection in patients with HIV infection. *Mt Sinai J Med* **66**, 113-24.

Chi, J. H., & Wilson, D. W. (2000). ATP-dependent localization of the herpes simplex virus capsid protein VP26 to sites of procapsid maturation. *J Virol* **74**, 1468-76.

Child, S. J., Hakki, M., De Niro, K. L. & Geballe, A. P. (2004). Evasion of cellular antiviral responses by human cytomegalovirus TRS1 and IRS1. *J Virol* **78**, 197-205.

Chou, S., Lurain, N. S., Thompson, K. D., Miner, R. C. & Drew, W. L. (2003). Viral DNA polymerase mutations associated with drug resistance in human cytomegalovirus. *J Infect Dis* **188**, 32-9.

Colberg-Poley, A. M., Patel, M. B., Erezo, D. P. & Slater, J. E. (2000). Human cytomegalovirus UL37 immediate-early regulatory proteins traffic through the secretory apparatus and to mitochondria. *J Gen Virol* **81**, 1779-89.

Collier, A. C., Meyers, J. D., Corey, L., Murphy, V. L., Roberts, P. L. & Handsfield, H. H. (1987). Cytomegalovirus infection in homosexual men. Relationship to sexual practices, antibody to human immunodeficiency virus, and cell-mediated immunity. *Am J Med* **82**, 593-601.

Compton, T., Nepomuceno, R. R. & Nowlin, D. M. (1992). Human cytomegalovirus penetrates host cells by pH-independent fusion at the cell surface. *Virology* **191**, 387-95.

Compton, T., Nowlin, D. M. & Cooper, N. R. (1993). Initiation of human cytomegalovirus infection requires initial interaction with cell surface heparan sulfate. *Virology* **193**, 834-41.

Cone, R. W., Hackman, R. C., Huang, M. L., Bowden, R. A., Meyers, J. D., Metcalf, M., Zeh, J., Ashley, R. & Corey, L. (1993). Human herpesvirus 6 in lung tissue from patients with pneumonitis after bone marrow transplantation. *N Engl J Med* **329**, 156-61.

Cope, A. V., Sabin, C., Burroughs, A., Rolles, K., Griffiths, P. D. & Emery, V. C. (1997). Interrelationships among quantity of human cytomegalovirus (HCMV) DNA in blood, donor-recipient serostatus, and administration of methylprednisolone as risk factors for HCMV disease following liver transplantation. *J Infect Dis* **176**, 1484-90.

Cope, A. V., Sweny, P., Sabin, C., Rees, L., Griffiths, P. D. & Emery, V. C. (1997). Quantity of cytomegalovirus viruria is a major risk factor for cytomegalovirus disease after renal transplantation. *J Med Virol* **52**, 200-5.

Corey, L., Adams, H. G., Brown, Z. A. & Holmes, K. K. (1983). Genital herpes simplex virus infections: clinical manifestations, course, and complications. *Ann Intern Med* **98**, 958-72.

Craig, J. M., Macauley, J. C., Weller, T. H. & Wirth, P. (1957). Isolation of intranuclear inclusion producing agents from infants with illnesses resembling cytomegalic inclusion disease. *Proc Soc Exp Biol Med* **94**, 4-12.

Craighead, J. E., Kanich, R. E. & Almeida, J. D. (1972). Nonviral microbodies with viral antigenicity produced in cytomegalovirus-infected cells. *J Virol* **10**, 766-75.

Croen, K. D., Ostrove, J. M., Dragovic, L. J., Smialek, J. E. & Straus, S. E. (1987). Latent herpes simplex virus in human trigeminal ganglia. Detection of an immediate early gene "anti-sense" transcript by in situ hybridization. *N Engl J Med* **317**, 1427-32.

Cvetkovic, R. S. & Wellington, K. (2005). Valganciclovir: a review of its use in the management of CMV infection and disease in immunocompromised patients. *Drugs* **65**, 859-78.

Dal Monte, P., Bessia, C., Landini, M. P. & Michelson, S. (1996). Expression of human cytomegalovirus ppUL83 (pp65) in a stable cell line and its association with metaphase chromosomes. *J Gen Virol* **77 (Pt 10)**, 2591-6.

Dargan, D. J., Jamieson, F. E., MacLean, J., Dolan, A., Addison, C. & McGeoch, D. J. (1997). The published DNA sequence of human cytomegalovirus strain AD169 lacks 929 base pairs affecting genes UL42 and UL43. *J Virol* **71**, 9833-6.

Dasgupta, A., Wilson, D. W. (1999). ATP depletion blocks herpes simplex virus DNA packaging and capsid maturation. *J Virol* **73**, 2006-15.

Davis, M. G., Mar, E. C., Wu, Y. M. & Huang, E. S. (1984). Mapping and expression of a human cytomegalovirus major viral protein. *J Virol* **52**, 129-35.

Davis, M. G. & Huang, E. S. (1985). Nucleotide sequence of a human cytomegalovirus DNA fragment encoding a 67-kilodalton phosphorylated viral protein. *J Virol* **56**, 7-11.

Davison, A. J. (1992). Channel catfish virus: a new type of herpesvirus. *Virology* **186**, 9-14.

Davison, A. J. & Clements, J. B. (1996). Herpesviruses: general properties. In *Topley and Wilsons Microbiology and Microbiol Infections*, pp. 309-323: Oxford University Press.

Davison, A. J. (1998). The genome of salmonid herpesvirus 1. *J Virol* **72**, 1974-82.

Davison, A. J., Sauerbier, W., Dolan, A., Addison, C. & McKinnell, R. G. (1999). Genomic studies of the Lucke tumor herpesvirus (RaHV-1). *J Cancer Res Clin Oncol* **125**, 232-8.

Davison, A. J. (2002). Evolution of the herpesviruses. *Vet Microbiol* **86**, 69-88.

Davison, A. J., Dolan, A., Akter, P., Addison, C., Dargan, D. J., Alcendor, D. J., McGeoch, D. J. & Hayward, G. S. (2003a). The human cytomegalovirus genome revisited: comparison with the chimpanzee cytomegalovirus genome. *J Gen Virol* **84**, 17-28.

Davison, A. J., Akter, P., Cunningham, C., Dolan, A., Addison, C., Dargan, D. J., Hassan-Walker, A. F., Emery, V. C., Griffiths, P. D. & Wilkinson, G. W. (2003b). Homology between the human cytomegalovirus RL11 gene family and human adenovirus E3 genes. *J Gen Virol* **84**, 657-63.

Davison, A. J., Eberle, R., Hayward, G. S., McGeoch, D. J., Minson, A. C., Pellett, P. E., Roizman, B., Studdert, M. J. & Thiry, E. (2005a). Herpesviridae. In *Virus Taxonomy, VIIIth Report of the ICTV*, pp. 193-212: Elsevier/Academic Press, London.

Davison, A. J., Trus, B. L., Cheng, N., Steven, A. C., Watson, M. S., Cunningham, C., Le Deuff, R. M. & Renault, T. (2005b). A novel class of herpesvirus with bivalve hosts. *J Gen Virol* **86**, 41-53.

De Clercq, E. (2004). Antiviral drugs in current clinical use. *J Clin Virol* **30**, 115-33.

de Smet, M. D., Meenken, C. J. & van den Horn, G. J. (1999). Fomivirsen - a phosphorothioate oligonucleotide for the treatment of CMV retinitis. *Ocul Immunol Inflamm* **7**, 189-98.

DeMarchi, J. M. (1981). Human cytomegalovirus DNA: restriction enzyme cleavage maps and map locations for immediate-early, early, and late RNAs. *Virology* **114**, 23-38.

DeMarchi, J. M. (1983). Post-transcriptional control of human cytomegalovirus gene expression. *Virology* **124**, 390-402.

Dolan, A., Cunningham, C., Hector, R. D., Hassan-Walker, A. F., Lee, L., Addison, C., Dargan, D. J., McGeoch, D. J., Gatherer, D., Emery, V. C., Griffiths, P. D., Sinzger, C., McSharry, B. P., Wilkinson, G. W. & Davison, A. J. (2004). Genetic content of wild-type human cytomegalovirus. *J Gen Virol* **85**, 1301-12.

Dominguez, G., Dambaugh, T. R., Stamey, F. R., Dewhurst, S., Inoue, N. & Pellett, P. E. (1999). Human herpesvirus 6B genome sequence: coding content and comparison with human herpesvirus 6A. *J Virol* **73**, 8040-52.

Donahue, J. G., Choo, P. W., Manson, J. E. & Platt, R. (1995). The incidence of herpes zoster. *Arch Intern Med* **155**, 1605-9.

Drew, W. L., Mintz, L., Miner, R. C., Sands, M. & Ketterer, B. (1981). Prevalence of cytomegalovirus infection in homosexual men. *J Infect Dis* **143**, 188-92.

Drew, W. L. (1984). Sexual transmission of CMV and its relationship to Kaposi sarcoma in homosexual men. *Birth Defects Orig Artic Ser* **20**, 121-9.

Drew, W. L. (1991). Clinical use of ganciclovir for cytomegalovirus infection and the development of drug resistance. *J Acquir Immune Defic Syndr* **4 Suppl 1**, S42-6.

Dunn, W., Chou, C., Li, H., Hai, R., Patterson, D., Stolc, V., Zhu, H. & Liu, F. (2003). Functional profiling of a human cytomegalovirus genome. *Proc Natl Acad Sci U S A* **100**, 14223-8.

Dworsky, M., Yow, M., Stagno, S., Pass, R. F. & Alford, C. (1983). Cytomegalovirus infection of breast milk and transmission in infancy. *Pediatrics* **72**, 295-9.

Eddy, S. R. (1999). Noncoding RNA genes. *Curr Opin Genet Dev* **9**, 695-9.

Eddy, S. R. (2002). Computational genomics of noncoding RNA genes. *Cell* **109**, 137-40.

Efstathiou, S., Minson, A. C., Field, H. J., Anderson, J. R. & Wildy, P. (1986). Detection of herpes simplex virus-specific DNA sequences in latently infected mice and in humans. *J Virol* **57**, 446-55.

Elliott, G., Mouzakitis, G. & O'Hare, P. (1995). VP16 interacts via its activation domain with VP22, a tegument protein of herpes simplex virus, and is relocated to a novel macromolecular assembly in coexpressing cells. *J Virol* **69**, 7932-41.

Enquist, L. W., Husak, P. J., Banfield, B. W. & Smith, G. A. (1999). Infection and spread of alphaviruses in the nervous system. *Adv Virus Res* **51**, 237-347.

Erice, A. (1999). Resistance of human cytomegalovirus to antiviral drugs. *Clin Microbiol Rev* **12**, 286-97.

Everett, R. D. (2000). ICP0, a regulator of herpes simplex virus during lytic and latent infection. *Bioessays* **22**, 761-70.

Ewing, B. & Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8**, 186-94.

Ewing, B., Hillier, L., Wendl, M. C. & Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8**, 175-85.

Farrell, M. J., Dobson, A. T. & Feldman, L. T. (1991). Herpes simplex virus latency-associated transcript is a stable intron. *Proc Natl Acad Sci U S A* **88**, 790-4.

Fish, K. N., Britt, W. & Nelson, J. A. (1996). A novel mechanism for persistence of human cytomegalovirus in macrophages. *J Virol* **70**, 1855-62.

Foster, T. P. & Kousoulas, K. G. (1999). Genetic analysis of the role of herpes virus type 1 glycoprotein K in infectious virus production and egress. *J Virol* **73**, 8457-68.

Galderisi, U., Cascino, A. & Giordano, A. (1999). Antisense oligonucleotides as therapeutic agents. *J Cell Physiol* **181**, 251-7.

Gallant, J. E., Moore, R. D., Richman, D. D., Keruly, J. & Chaisson, R. E. (1992). Incidence and natural history of cytomegalovirus disease in patients with advanced human immunodeficiency virus disease treated with zidovudine. The Zidovudine Epidemiology Study Group. *J Infect Dis* **166**, 1223-7.

Garber, D. A., Beverley, S. M. & Coen, D. M. (1993). Demonstration of circularization of herpes simplex virus DNA following infection using pulsed field gel electrophoresis. *Virology* **197**, 459-62.

Gardner, M. B., Officer, J. E., Parker, J., Estes, J. D. & Rongey, R. W. (1974). Induction of disseminated virulent cytomegalovirus infection by immunosuppression of naturally chronically infected wild mice. *Infect Immun* **10**, 966-9.

Gerna, G., Baldanti, F., Middeldorp, J. M., Furione, M., Zavattoni, M., Lilleri, D. & Revello, M. G. (1999). Clinical significance of expression of human cytomegalovirus pp67 late transcript in heart, lung, and bone marrow transplant recipients as determined by nucleic acid sequence-based amplification. *J Clin Microbiol* **37**, 902-11.

Gerna, G., Percivalle, E., Sarasini, A., Baldanti, F., Campanini, G. & Revello, M. G. (2003). Rescue of human cytomegalovirus strain AD169 tropism for both leukocytes and human endothelial cells. *J Gen Virol* **84**, 1431-6.

Gershon, A. A. & Silverstein, S. J. (1997). Varicella-Zoster Virus. In *Clinical Virology*, pp. 421-44: Churchill Livingstone.

Gibson, M. G. & Spear, P. G. (1983). Insertion mutants of herpes simplex virus have a duplication of the glycoprotein D gene and express two different forms of glycoprotein D. *J Virol* **48**, 396-404.

Gibson, W. (1996). Structure and assembly of the virion. *Intervirology* **39**, 389-400.

Gibson, W., Clopper, K. S., Britt, W. J. & Baxter, M. K. (1996). Human cytomegalovirus (HCMV) smallest capsid protein identified as product of short open reading frame located between HCMV UL48 and UL49. *J Virol* **70**, 5680-3.

Gibson, W. & Roizman, B. (1972). Proteins specified by herpes simplex virus. 8. Characterization and composition of multiple capsid forms of subtypes 1 and 2. *J Virol* **10**, 1044-52.

Gilbert, M. J., Riddell, S. R., Plachter, B. & Greenberg, P. D. (1996). **Cytomegalovirus** selectively blocks antigen processing and presentation of its immediate-early gene product. *Nature* **383**, 720-2.

Gold, D., Bowden, R., Sixbey, J., Riggs, R., Katon, W. J., Ashley, R., Obrigewitch, R. M. & Corey, L. (1990). Chronic fatigue. A prospective clinical and virologic study. *Jama* **264**, 48-53.

Gompels, U. A., Nicholas, J., Lawrence, G., Jones, M., Thomson, B. J., Martin, M. E., Efstathiou, S., Craxton, M. & Macaulay, H. A. (1995). The DNA sequence of human herpesvirus-6: structure, coding content, and genome evolution. *Virology* **209**, 29-51.

Goodrum, F. D., Jordan, C. T., High, K. & Shenk, T. (2002). Human cytomegalovirus gene expression during infection of primary hematopoietic progenitor cells: a model for latency. *Proc Natl Acad Sci U S A* **99**, 16255-60.

Gor, D., Sabin, C., Prentice, H. G., Vyas, N., Man, S., Griffiths, P. D. & Emery, V. C. (1998). Longitudinal fluctuations in cytomegalovirus load in bone marrow transplant patients: relationship between peak virus load, donor/recipient serostatus, acute GVHD and CMV disease. *Bone Marrow Transplant* **21**, 597-605.

Gray, W. L., Starnes, B., White, M. W. & Mahalingam, R. (2001). The DNA sequence of the simian varicella virus genome. *Virology* **284**, 123-30.

Greber, U. F., Fassati, A. (2003). Nuclear import of viral DNA genomes. *Traffic* **4**, 136-43.

Greijer, A. E., Dekkers, C. A. & Middeldorp, J. M. (2000). Human cytomegalovirus virions differentially incorporate viral and host cell RNA during the assembly process. *J Virol* **74**, 9078-82.

Grey, F., Antoniewicz, A., Allen, E., Saugstad, J., McShea, A., Carrington, J. C. & Nelson, J. (2005). Identification and characterization of human cytomegalovirus-encoded microRNAs. *J Virol* **79**, 12095-9.

Griffiths, P. D., Cope, A. V., Hassan-Walker, A. F. & Emery, V. C. (1999). Diagnostic approaches to cytomegalovirus infection in bone marrow and organ transplantation. *Transpl Infect Dis* **1**, 179-86.

Griffiths, P. D. & Emery, V. C. (1997). Cytomegalovirus. In *Clinical Virology*, pp. 445-70: Churchill Livingstone.

Haarr, L. & Skulstad, S. (1994). The herpes simplex virus type 1 particle: structure and molecular functions. Review article. *Apmis* **102**, 321-46.

Hahn, G., Jores, R. & Mocarski, E. S. (1998). Cytomegalovirus remains latent in a common precursor of dendritic and myeloid cells. *Proc Natl Acad Sci U S A* **95**, 3937-42.

Hahn, G., Khan, H., Baldanti, F., Koszinowski, U. H., Revello, M. G. & Gerna, G. (2002). The human cytomegalovirus ribonucleotide reductase homolog UL45 is dispensable for growth in endothelial cells, as determined by a BAC-cloned clinical isolate of human cytomegalovirus with preserved wild-type characteristics. *J Virol* **76**, 9551-5.

Hahn, G., Revello, M. G., Patrone, M., Percivalle, E., Campanini, G., Sarasini, A., Wagner, M., Gallina, A., Milanesi, G., Koszinowski, U., Baldanti, F. & Gerna, G. (2004). Human cytomegalovirus UL131-128 genes are indispensable for virus growth in endothelial cells and virus transfer to leukocytes. *J Virol* **78**, 10023-33.

Hamzeh, F. M., Lietman, P. S., Gibson, W. & Hayward, G. S. (1990). Identification of the lytic origin of DNA replication in human cytomegalovirus by a novel approach utilizing ganciclovir-induced chain termination. *J Virol* **64**, 6184-95.

Handsfield, H. H., Chandler, S. H., Caine, V. A., Meyers, J. D., Corey, L., Medeiros, E. & McDougall, J. K. (1985). Cytomegalovirus infection in sex partners: evidence for sexual transmission. *J Infect Dis* **151**, 344-8.

Hansen, S. G., Strelow, L. I., Franchi, D. C., Anders, D. G. & Wong, S. W. (2003). Complete sequence and genomic analysis of rhesus cytomegalovirus. *J Virol* **77**, 6620-36.

Harrison, C. J., Britt, W. J., Chapman, N. M., Mullican, J. & Tracy, S. (1995). Reduced congenital cytomegalovirus (CMV) infection after maternal immunization with a guinea pig CMV glycoprotein before gestational primary CMV infection in the guinea pig model. *J Infect Dis* **172**, 1212-20.

Hassan-Walker, A. F., Cope, A. V., Griffiths, P. D. & Emery, V. C. (1998). Transcription of the human cytomegalovirus natural killer decoy gene, UL18, in vitro and in vivo. *J Gen Virol* **79 (Pt 9)**, 2113-6.

Hayajneh, W. A., Colberg-Poley, A. M., Skaletskaya, A., Bartle, L. M., Lesperance, M. M., Contopoulos-Ioannidis, D. G., Kedersha, N. L. & Goldmacher, V. S. (2001). The sequence and antiapoptotic functional domains of the human cytomegalovirus UL37 exon 1 immediate early protein are conserved in multiple primary strains. *Virology* **279**, 233-40.

Hayashi, M. L., Blankenship, C. & Shenk, T. (2000). Human cytomegalovirus UL69 protein is required for efficient accumulation of infected cells in the G1 phase of the cell cycle. *Proc Natl Acad Sci U S A* **97**, 2692-6.

Hebart, H. & Einsele, H. (1998). Diagnosis and treatment of cytomegalovirus infection. *Curr Opin Hematol* **5**, 483-7.

Hebart, H., Kanz, L., Jahn, G. & Einsele, H. (1998). Management of cytomegalovirus infection after solid-organ or stem-cell transplantation. Current guidelines and future prospects. *Drugs* **55**, 59-72.

Heieren, M. H., Kim, Y. K. & Balfour, H. H., Jr. (1988). Human cytomegalovirus infection of kidney glomerular visceral epithelial and tubular epithelial cells in culture. *Transplantation* **46**, 426-32.

Hemmings, D. G., Kilani, R., Nykiforuk, C., Preiksaitis, J. & Guilbert, L. J. (1998). Permissive cytomegalovirus infection of primary villous term and first trimester trophoblasts. *J Virol* **72**, 4970-9.

Hengel, H., Lucin, P., Jonjic, S., Ruppert, T. & Koszinowski, U. H. (1994). Restoration of cytomegalovirus antigen presentation by gamma interferon combats viral escape. *J Virol* **68**, 289-97.

Hite, J. M., Eckert, K. A. & Cheng, K. C. (1996). Factors affecting fidelity of DNA synthesis during PCR amplification of d(C-A)n.d(G-T)n microsatellite repeats. *Nucleic Acids Res* **24**, 2429-34.

Hitomi, S., Kozuka-Hata, H., Chen, Z., Sugano, S., Yamaguchi, N. & Watanabe, S. (1997). Human cytomegalovirus open reading frame UL11 encodes a highly polymorphic protein expressed on the infected cell surface. *Arch Virol* **142**, 1407-27.

Ho, M., Suwansirikul, S., Dowling, J. N., Youngblood, L. A. & Armstrong, J. A. (1975). The transplanted kidney as a source of cytomegalovirus infection. *N Engl J Med* **293**, 1109-12.

Homa, F. L. & Brown, J. C. (1997). Capsid assembly and DNA packaging in herpes simplex virus. *Rev Med Virol* **7**, 107-122.

Honess, R. W. & Roizman, B. (1974). Regulation of herpesvirus macromolecular synthesis. I. Cascade regulation of the synthesis of three groups of viral proteins. *J Virol* **14**, 8-19.

Honess, R. W. & Roizman, B. (1975). Proteins specified by herpes simplex virus. XIII. Glycosylation of viral polypeptides. *J Virol* **16**, 1308-26.

Honess, R. W. (1984). Herpes simplex and 'the herpes complex': diverse observations and a unifying hypothesis. The eighth Fleming lecture. *J Gen Virol* **65 (Pt 12)**, 2077-107.

Hsu, D. H., de Waal Malefyt, R., Fiorentino, D. F., Dang, M. N., Vieira, P., de Vries, J., Spits, H., Mosmann, T. R. & Moore, K. W. (1990). Expression of interleukin-10 activity by Epstein-Barr virus protein BCRF1. *Science* **250**, 830-2.

Huang, E. S., Huong, S. M., Tegtmeier, G. E. & Alford, C. (1980). Cytomegalovirus: genetic variation of viral genomes. *Ann N Y Acad Sci* **354**, 332-46.

Huang, L., Zhu, Y. & Anders, D. G. (1996). The variable 3' ends of a human cytomegalovirus oriLyt transcript (SRT) overlap an essential, conserved replicator element. *J Virol* **70**, 5272-81.

Huttenhofer, A., Brosius, J. & Bachellerie, J. P. (2002). RNomics: identification and function of small, non-messenger RNAs. *Curr Opin Chem Biol* **6**, 835-43.

Huttenhofer, A., Schattner, P. & Polacek, N. (2005). Non-coding RNAs: hope or hype? *Trends Genet* **21**, 289-97.

Ibanez, C. E., Schrier, R., Ghazal, P., Wiley, C. & Nelson, J. A. (1991). Human cytomegalovirus productively infects primary differentiated macrophages. *J Virol* **65**, 6581-8.

Irmiere, A. & Gibson, W. (1983). Isolation and characterization of a noninfectious virion-like particle released from cells infected with human strains of cytomegalovirus. *Virology* **130**, 118-33.

Isegawa, Y., Mukai, T., Nakano, K., Kagawa, M., Chen, J., Mori, Y., Sunagawa, T., Kawanishi, K., Sashihara, J., Hata, A., Zou, P., Kosuge, H. & Yamanishi, K. (1999). Comparison of the complete DNA sequences of human herpesvirus 6 variants A and B. *J Virol* **73**, 8053-63.

Ishov, A. M. & Maul, G. G. (1996). The periphery of nuclear domain 10 (ND10) as site of DNA virus deposition. *J Cell Biol* **134**, 815-26.

Jabs, D. A., Enger, C. & Bartlett, J. G. (1989). Cytomegalovirus retinitis and acquired immunodeficiency syndrome. *Arch Ophthalmol* **107**, 75-80.

Jabs, D. A. & Griffiths, P. D. (2002). Fomivirsen for the treatment of cytomegalovirus retinitis. *Am J Ophthalmol* **133**, 552-6.

Jacob, R. J., Morse, L. S. & Roizman, B. (1979). Anatomy of herpes simplex virus DNA. XII. Accumulation of head-to-tail concatemers in nuclei of infected cells and their role in the generation of the four isomeric arrangements of viral DNA. *J Virol* **29**, 448-57.

Jacobson, M. A. & Mills, J. (1988). Serious cytomegalovirus disease in the acquired immunodeficiency syndrome (AIDS). Clinical findings, diagnosis, and treatment. *Ann Intern Med* **108**, 585-94.

Jahn, G., Knust, E., Schmolla, H., Sarre, T., Nelson, J. A., McDougall, J. K. & Fleckenstein, B. (1984). Predominant immediate-early transcripts of human cytomegalovirus AD 169. *J Virol* **49**, 363-70.

Jarvis, M. A. & Nelson, J. A. (2002). Human cytomegalovirus persistence and latency in endothelial cells and macrophages. *Curr Opin Microbiol* **5**, 403-7.

Jenkins, C., Abendroth, A. & Slobedman, B. (2004). A novel viral transcript with homology to human interleukin-10 is expressed during latent human cytomegalovirus infection. *J Virol* **78**, 1440-7.

Jenkins, D. E., Martens, C. L. & Mocarski, E. S. (1994). Human cytomegalovirus late protein encoded by ie2: a trans-activator as well as a repressor of gene expression. *J Gen Virol* **75 (9)**, 2337-48.

Jones, P. C. & Roizman, B. (1979). Regulation of herpesvirus macromolecular synthesis. VIII. The transcription program consists of three phases during which both extent of transcription and accumulation of RNA in the cytoplasm are regulated. *J Virol* **31**, 299-314.

Jones, T. R. & Muzithras, V. P. (1992). A cluster of dispensable genes within the human cytomegalovirus genome short component: IRS1, US1 through US5, and the US6 family. *J Virol* **66**, 2541-6.

Jones, T. R., Wiertz, E. J., Sun, L., Fish, K. N., Nelson, J. A. & Ploegh, H. L. (1996). Human cytomegalovirus US3 impairs transport and maturation of major histocompatibility complex class I heavy chains. *Proc Natl Acad Sci U S A* **93**, 11327-33.

Jordan, M. C., Rousseau, W., Stewart, J. A., Noble, G. R. & Chin, T. D. (1973). Spontaneous cytomegalovirus mononucleosis. Clinical and laboratory observations in nine cases. *Ann Intern Med* **79**, 153-60.

Keay, S. & Baldwin, B. (1991). Anti-idiotype antibodies that mimic gp86 of human cytomegalovirus inhibit viral fusion but not attachment. *J Virol* **65**, 5124-8.

Kerry, J. A., Priddy, M. A., Jervey, T. Y., Kohler, C. P., Staley, T. L., Vanson, C. D., Jones, T. R., Iskenderian, A. C., Anders, D. G. & Stenberg, R. M. (1996). Multiple regulatory events influence human cytomegalovirus DNA polymerase (UL54) expression during viral infection. *J Virol* **70**, 373-82.

Kiehl, A., Huang, L., Franchi, D. & Anders, D. G. (2003). Multiple 5' ends of human cytomegalovirus UL57 transcripts identify a complex, cycloheximide-resistant promoter region that activates oriLyt. *Virology* **314**, 410-22.

Kimberlin, D. W., Lin, C. Y., Sanchez, P. J., Demmler, G. J., Dankner, W., Shelton, M., Jacobs, R. F., Vaudry, W., Pass, R. F., Kiell, J. M., Soong, S. J. & Whitley, R. J. (2003). Effect of ganciclovir therapy on hearing in symptomatic congenital cytomegalovirus disease involving the central nervous system: a randomized, controlled trial. *J Pediatr* **143**, 16-25.

Klemola, E., Von Essen, R., Henle, G. & Henle, W. (1970). Infectious-mononucleosis-like disease with negative heterophil agglutination test. Clinical features in relation to Epstein-Barr virus and cytomegalovirus antibodies. *J Infect Dis* **121**, 608-14.

Kollert-Jons, A., Bogner, E. & Radsak, K. (1991). A 15-kilobase-pair region of the human cytomegalovirus genome which includes US1 through US13 is dispensable for growth in cell culture. *J Virol* **65**, 5184-9.

Kondo, K., Kaneshima, H. & Mocarski, E. S. (1994). Human cytomegalovirus latent infection of granulocyte-macrophage progenitors. *Proc Natl Acad Sci U S A* **91**, 11879-83.

Kondo, K., Xu, J. & Mocarski, E. S. (1996). Human cytomegalovirus latent gene expression in granulocyte-macrophage progenitors in culture and in seropositive individuals. *Proc Natl Acad Sci U S A* **93**, 11137-42.

Kotenko, S. V., Saccani, S., Izotova, L. S., Mirochnitchenko, O. V. & Pestka, S. (2000). Human cytomegalovirus harbors its own unique IL-10 homolog (cmvIL-10). *Proc Natl Acad Sci U S A* **97**, 1695-700.

Kouzarides, T., Bankier, A. T., Satchwell, S. C., Preddy, E. & Barrell, B. G. (1988). An immediate early gene of human cytomegalovirus encodes a potential membrane glycoprotein. *Virology* **165**, 151-64.

Krause, P. R., Croen, K. D., Straus, S. E. & Ostrove, J. M. (1988). Detection and preliminary characterization of herpes simplex virus type 1 transcripts in latently infected human trigeminal ganglia. *J Virol* **62**, 4819-23.

Krause, P. R., Ostrove, J. M. & Straus, S. E. (1991). The nucleotide sequence, 5' end, promoter domain, and kinetics of expression of the gene encoding the herpes simplex virus type 2 latency-associated transcript. *J Virol* **65**, 5619-23.

Krech, U. (1973). Complement-fixing antibodies against cytomegalovirus in different parts of the world. *Bull World Health Organ* **49**, 103-6.

Kulesza, C. A. & Shenk, T. (2004). Human cytomegalovirus 5-kilobase immediate-early RNA is a stable intron. *J Virol* **78**, 13182-9.

Lalezari, J. P., Holland, G. N., Kramer, F., McKinley, G. F., Kemper, C. A., Ives, D. V., Nelson, R., Hardy, W. D., Kuppermann, B. D., Northfelt, D. W., Youle, M., Johnson, M., Lewis, R. A., Weinberg, D. V., Simon, G. L., Wolitz, R. A., Ruby, A. E., Stagg, R. J. & Jaffe, H. S. (1998). Randomized, controlled study of the safety and efficacy of intravenous cidofovir for the treatment of relapsing cytomegalovirus retinitis in patients with AIDS. *J Acquir Immune Defic Syndr Hum Retrovirol* **17**, 339-44.

Lea, A. P. & Bryson, H. M. (1996). Cidofovir. *Drugs* **52**, 225-230; discussion 231.

Leatham, M. P., Witte, P. R. & Stinski, M. F. (1991). Alternate promoter selection within a human cytomegalovirus immediate-early and early transcription unit (UL119-115) defines true late transcripts containing open reading frames for putative viral glycoproteins. *J Virol* **65**, 6144-53.

Leib, D. A., Coen, D. M., Bogard, C. L., Hicks, K. A., Yager, D. R., Knipe, D. M., Tyler, K. L. & Schaffer, P. A. (1989). Immediate-early regulatory gene mutants define different stages in the establishment and reactivation of herpes simplex virus latency. *J Virol* **63**, 759-68.

Levin, M. J., Rinaldo, C. R., Jr., Leary, P. L., Zaia, J. A. & Hirsch, M. S. (1979). Immune response to herpesvirus antigens in adults with acute cytomegaloviral mononucleosis. *J Infect Dis* **140**, 851-7.

Ljungman, P. (1996). Cytomegalovirus infections in transplant patients. *Scand J Infect Dis Suppl* **100**, 59-63.

Lock, M., Miller, C. & Fraser, N. W. (2001). Analysis of protein expression from within the region encoding the 2.0-kilobase latency-associated transcript of herpes simplex virus type 1. *J Virol* **75**, 3413-26.

Lockridge, K. M., Zhou, S. S., Kravitz, R. H., Johnson, J. L., Sawai, E. T., Blewett, E. L. & Barry, P. A. (2000). Primate cytomegaloviruses encode and express an IL-10-like protein. *Virology* **268**, 272-80.

Lu, M. & Shenk, T. (1999). Human cytomegalovirus UL69 protein induces cells to accumulate in G1 phase of the cell cycle. *J Virol* **73**, 676-83.

Lurain, N. S., Kapell, K. S., Huang, D. D., Short, J. A., Paintsil, J., Winkfield, E., Benedict, C. A., Ware, C. F. & Bremer, J. W. (1999). Human cytomegalovirus UL144 open reading frame: sequence hypervariability in low-passage clinical isolates. *J Virol* **73**, 10040-50.

Mador, N., Panet, A., Latchman, D. & Steiner, I. (1995). Expression and splicing of the latency-associated transcripts of herpes simplex virus type 1 in neuronal and non-neuronal cell lines. *J Biochem (Tokyo)* **117**, 1288-97.

Margulies, B. J., Browne, H. & Gibson, W. (1996). Identification of the human cytomegalovirus G protein-coupled receptor homologue encoded by UL33 in infected cells and enveloped virus particles. *Virology* **225**, 111-25.

Masse, M. J., Karlin, S., Schachtel, G. A. & Mocarski, E. S. (1992). Human cytomegalovirus origin of DNA replication (oriLyt) resides within a highly complex repetitive region. *Proc Natl Acad Sci U S A* **89**, 5246-50.

Masur, H., Whitcup, S. M., Cartwright, C., Polis, M. & Nussenblatt, R. (1996). Advances in the management of AIDS-related cytomegalovirus retinitis. *Ann Intern Med* **125**, 126-36.

McGeoch, D. J. (1987). The genome of herpes simplex virus: structure, replication and evolution. *J Cell Sci Suppl* **7**, 67-94.

McGeoch, D. J., Cook, S., Dolan, A., Jamieson, F. E. & Telford, E. A. (1995). Molecular phylogeny and evolutionary timescale for the family of mammalian herpesviruses. *J Mol Biol* **247**, 443-58.

McGeoch, D. J. & Davison, A. J. (1999). The descent of human herpesvirus 8. *Semin Cancer Biol* **9**, 201-9.

McGeoch, D. J. & Davison, A. J. (1999). The molecular evolutionary history of the herpesviruses. In *Origin and evolution of viruses*, pp. 441-465: Academic press.

McGeoch, D. J., Dolan, A. & Ralph, A. C. (2000). Toward a comprehensive phylogeny for mammalian and avian herpesviruses. *J Virol* **74**, 10401-6.

McGeoch, D. J. & Gatherer, D. (2005). Integrating reptilian herpesviruses into the family herpesviridae. *J Virol* **79**, 725-31.

McNabb, D. & Courtney, R. (1992). Characterisation of the large tegument protein (ICP1/2) of herpes simplex virus type 1. *Virology* **190**, 221-32.

Megaw, A. G., Rapaport, D., Avidor, B., Frenkel, N. & Davison, A. J. (1998). The DNA sequence of the RK strain of human herpesvirus 7. *Virology* **244**, 119-32.

Mellerick, D. M. & Fraser, N. W. (1987). Physical state of the latent herpes simplex virus genome in a mouse model system: evidence suggesting an episomal state. *Virology* **158**, 265-75.

Meyers, J. D., Flournoy, N. & Thomas, E. D. (1986). Risk factors for cytomegalovirus infection after human marrow transplantation. *J Infect Dis* **153**, 478-88.

Meyers, J. D., McGuffin, R. W., Bryson, Y. J., Cantell, K. & Thomas, E. D. (1982). Treatment of cytomegalovirus pneumonia after marrow transplant with combined vidarabine and human leukocyte interferon. *J Infect Dis* **146**, 80-4.

Mocarski, E. S., Liu, A. C. & Spaete, R. R. (1987). Structure and variability of the a sequence in the genome of human cytomegalovirus (Towne strain). *J Gen Virol* **68 (Pt 8)**, 2223-30.

Mocarski, E. S. (1996). Cytomegaloviruses and their replication. In *Fields Virology*, pp. 2447-91: Lippincott Williams and Wilkins.

Mocarski, E. S., Prichard, M. N., Tan, C. S. & Brown, J. M. (1997). Reassessing the organization of the UL42-UL43 region of the human cytomegalovirus strain AD169 genome. *Virology* **239**, 169-75.

Mocarski, E. S. & Tan Courcelle, C. (2001). Cytomegaloviruses and their replication. In *Fields Virology*, pp. 2629-73: Lippincott Williams and Wilkins.

Moore, K. W., Vieira, P., Fiorentino, D. F., Trounstine, M. L., Khan, T. A. & Mosmann, T. R. (1990). Homology of cytokine synthesis inhibitory factor (IL-10) to the Epstein-Barr virus gene BCRFI. *Science* **248**, 1230-4.

Moore, P. S. & Chang, Y. (1997). Kaposi's sarcoma-associated herpesvirus. In *Clinical Virology*, pp. 509-24: Churchill Livingstone.

Murphy, E., Rigoutsos, I., Shibuya, T. & Shenk, T. E. (2003a). Reevaluation of human cytomegalovirus coding potential. *Proc Natl Acad Sci U S A* **100**, 13585-90.

Murphy, E., Yu, D., Grimwood, J., Schmutz, J., Dickson, M., Jarvis, M. A., Hahn, G., Nelson, J. A., Myers, R. M. & Shenk, T. E. (2003b). Coding potential of laboratory and clinical strains of human cytomegalovirus. *Proc Natl Acad Sci U S A* **100**, 14976-81.

Murphy, J. C., Fischle, W., Verdin, E. & Sinclair, J. H. (2002). Control of cytomegalovirus lytic gene expression by histone acetylation. *Embo J* **21**, 1112-20.

Navarro, D., Paz, P., Tugizov, S., Topp, K., La Vail, J. & Pereira, L. (1993). Glycoprotein B of human cytomegalovirus promotes virion penetration into cells, transmission of infection from cell to cell, and fusion of infected cells. *Virology* **197**, 143-58.

Nelson, J. A., Fleckenstein, B., Jahn, G., Galloway, D. A. & McDougall, J. K. (1984). Structure of the transforming region of human cytomegalovirus AD169. *J Virol* **49**, 109-15.

Neri, A., Barriga, F., Inghirami, G., Knowles, D. M., Neequaye, J., Magrath, I. T. & Dalla-Favera, R. (1991). Epstein-Barr virus infection precedes clonal expansion in Burkitt's and acquired immunodeficiency syndrome-associated lymphoma. *Blood* **77**, 1092-5.

Newcomb, W. W., Homa, F. L., Thomsen, D. R., Booy, F. P., Trus, B. L., Steven, A. C., Spencer, J. V. & Brown, J. C. (1996). Assembly of the herpes simplex virus capsid: characterization of intermediates observed during cell-free capsid formation. *J Mol Biol* **263**, 432-46.

Newcomb, W. W., Homa, F. L., Thomsen, D. R., Trus, B. L., Cheng, N., Steven, A. C., Booy, F. P., & Brown, J. C. (1996). Assembly of the herpes simplex virus procapsid from purified components and identification of small complexes containing the major capsid and scaffolding proteins. *J Virol* **73**, 4239-50.

Neyts, J., Snoeck, R., Schols, D., Balzarini, J., Esko, J. D., Van Schepdael, A. & De Clercq, E. (1992). Sulfated polymers inhibit the interaction of human cytomegalovirus with cell surface heparan sulfate. *Virology* **189**, 48-58.

Nicholas, J. (1996). Determination and analysis of the complete nucleotide sequence of human herpesvirus. *J Virol* **70**, 5975-89.

Novotny, J., Rigoutsos, I., Coleman, D. & Shenk, T. (2001). In silico structural and functional analysis of the human cytomegalovirus (HHV5) genome. *J Mol Biol* **310**, 1151-66.

Odeberg, J., Browne, H., Metkar, S., Froelich, C. J., Branden, L., Cosman, D. & Soderberg-Naucler, C. (2003). The human cytomegalovirus protein UL16 mediates increased resistance to natural killer cell cytotoxicity through resistance to cytolytic proteins. *J Virol* **77**, 4539-45.

Oien, N. L., Thomsen, D. R., Wathen, M. W., Newcomb, W. W., Brown, J. C. & Homa, F. L. (1997). Assembly of herpes simplex virus capsids using the human cytomegalovirus scaffold protein: critical role of the C terminus. *J Virol* **71**, 1281-91.

Oram, J. D., Downing, R. G., Akrigg, A., Dollery, A. A., Duggleby, C. J., Wilkinson, G. W. & Greenaway, P. J. (1982). Use of recombinant plasmids to investigate the structure of the human cytomegalovirus genome. *J Gen Virol* **59**, 111-29.

Pachl, C., Probert, W. S., Hermsen, K. M., Masiarz, F. R., Rasmussen, L., Merigan, T. C. & Spaete, R. R. (1989). The human cytomegalovirus strain Towne glycoprotein H gene encodes glycoprotein p86. *Virology* **169**, 418-26.

Palestine, A. G., Polis, M. A., De Smet, M. D., Baird, B. F., Falloon, J., Kovacs, J. A., Davey, R. T., Zurlo, J. J., Zunich, K. M., Davis, M. & et al. (1991). A randomized, controlled trial of foscarnet in the treatment of cytomegalovirus retinitis in patients with AIDS. *Ann Intern Med* **115**, 665-73.

Pari, G. S. & Anders, D. G. (1993). Eleven loci encoding trans-acting factors are required for transient complementation of human cytomegalovirus oriLyt-dependent DNA replication. *J Virol* **67**, 6979-88.

Pari, G. S., Kacica, M. A. & Anders, D. G. (1993). Open reading frames UL44, IRS1/TRS1, and UL36-38 are required for transient complementation of human cytomegalovirus oriLyt-dependent DNA synthesis. *J Virol* **67**, 2575-82.

Pass, R. F., Griffiths, P. D. & August, A. M. (1983). Antibody response to cytomegalovirus after renal transplantation: comparison of patients with primary and recurrent infections. *J Infect Dis* **147**, 40-6.

Pass, R. F. (2001). Cytomegalovirus. In *Fields Virology*, pp. 2675-705: Lippincott Williams and Wilkins.

Penfold, M. E., Dairaghi, D. J., Duke, G. M., Saederup, N., Mocarski, E. S., Kemble, G. W. & Schall, T. J. (1999). Cytomegalovirus encodes a potent alpha chemokine. *Proc Natl Acad Sci U S A* **96**, 9839-44.

Pereira, L. H., Embil, J. A., Haase, D. A. & Manley, K. M. (1990). Cytomegalovirus infection among women attending a sexually transmitted disease clinic: association with clinical symptoms and other sexually transmitted diseases. *Am J Epidemiol* **131**, 683-92.

Perez, J. L. (1997). Resistance to antivirals in human cytomegalovirus: mechanisms and clinical significance. *Microbiologia* **13**, 343-52.

Perng, G. C., Slanina, S. M., Yukht, A., Ghiasi, H., Nesburn, A. B. & Wechsler, S. L. (2000). The latency-associated transcript gene enhances establishment of herpes simplex virus type 1 latency in rabbits. *J Virol* **74**, 1885-91.

Pfeffer, S., Sewer, A., Lagos-Quintana, M., Sheridan, R., Sander, C., Grasser, F. A., van Dyk, L. F., Ho, C. K., Shuman, S., Chien, M., Russo, J. J., Ju, J., Randall, G., Lindenbach, B. D., Rice, C. M., Simon, V., Ho, D. D., Zavolan, M. & Tuschl, T. (2005). Identification of microRNAs of the herpesvirus family. *Nat Methods* **2**, 269-76.

Pignatelli, S., Dal Monte, P., Rossini, G. & Landini, M. P. (2004). Genetic polymorphisms among human cytomegalovirus (HCMV) wild-type strains. *Rev Med Virol* **14**, 383-410.

Plachter, B., Traupe, B., Albrecht, J. & Jahn, G. (1988). Abundant 5 kb RNA of human cytomegalovirus without a major translational reading frame. *J Gen Virol* **69 (Pt 9)**, 2251-66.

Plachter, B., Sinzger, C. & Jahn, G. (1996). Cell types involved in replication and distribution of human cytomegalovirus. *Adv Virus Res* **46**, 195-261.

Plotkin, S. A., Furukawa, T., Zygraich, N. & Huygelen, C. (1975). Candidate cytomegalovirus strain for human vaccination. *Infect Immun* **12**, 521-7.

Poffenberger, K. L. & Roizman, B. (1985). A noninverting genome of a viable herpes simplex virus 1: presence of head-to-tail linkages in packaged genomes and requirements for circularization after infection. *J Virol* **53**, 587-95.

Poland, S. D., Costello, P., Dekaban, G. A. & Rice, G. P. (1990). Cytomegalovirus in the brain: in vitro infection of human brain-derived cells. *J Infect Dis* **162**, 1252-62.

Pollard, R. B., Rand, K. H., Arvin, A. M. & Merigan, T. C. (1978). Cell-mediated immunity of cytomegalovirus infection in normal subjects and cardiac transplant patients. *J Infect Dis* **137**, 541-9.

Poncet, D., Larochette, N., Pauleau, A. L., Boya, P., Jalil, A. A., Cartron, P. F., Vallette, F., Schnebelen, C., Bartle, L. M., Skaletskaya, A., Boutolleau, D., Martinou, J. C., Goldmacher, V. S., Kroemer, G. & Zamzami, N. (2004). An anti-apoptotic viral protein that recruits Bax to mitochondria. *J Biol Chem* **279**, 22605-14.

Prentice, H. G. & Kho, P. (1997). Clinical strategies for the management of cytomegalovirus infection and disease in allogeneic bone marrow transplant. *Bone Marrow Transplant* **19**, 135-42.

Preston, C. M. (2000). Repression of viral transcription during herpes simplex virus latency. *J Gen Virol* **81**, 1-19.

Prichard, M. N., Jairath, S., Penfold, M. E., St Jeor, S., Bohlman, M. C. & Pari, G. S. (1998). Identification of persistent RNA-DNA hybrid structures within the origin of replication of human cytomegalovirus. *J Virol* **72**, 6997-7004.

Prichard, M. N., Penfold, M. E., Duke, G. M., Spaete, R. R. & Kemble, G. W. (2001). A review of genetic differences between limited and extensively passaged human cytomegalovirus strains. *Rev Med Virol* **11**, 191-200.

Quinnan, G. V., Manischewitz, J. E. & Ennis, F. A. (1978). Cytotoxic T lymphocyte response to murine cytomegalovirus infection. *Nature* **273**, 541-3.

Quinnan, G. V., Jr., Kirmani, N., Rook, A. H., Manischewitz, J. F., Jackson, L., Moreschi, G., Santos, G. W., Saral, R. & Burns, W. H. (1982). Cytotoxic t cells in cytomegalovirus infection: HLA-restricted T-lymphocyte and non-T-lymphocyte cytotoxic responses correlate with recovery from cytomegalovirus infection in bone-marrow-transplant recipients. *N Engl J Med* **307**, 7-13.

Quinnan, G. V., Jr., Delery, M., Rook, A. H., Frederick, W. R., Epstein, J. S., Manischewitz, J. F., Jackson, L., Ramsey, K. M., Mittal, K., Plotkin, S. A. & et al. (1984). Comparative virulence and immunogenicity of the Towne strain and a nonattenuated strain of cytomegalovirus. *Ann Intern Med* **101**, 478-83.

Ramsay, M. E., Miller, E. & Peckham, C. S. (1991). Outcome of confirmed symptomatic congenital cytomegalovirus infection. *Arch Dis Child* **66**, 1068-9.

Rapp, M., Messerle, M., Buhler, B., Tannheimer, M., Keil, G. M. & Koszinowski, U. H. (1992). Identification of the murine cytomegalovirus glycoprotein B gene and its expression by recombinant vaccinia virus. *J Virol* **66**, 4399-406.

Rasmussen, L., Geissler, A. & Winters, M. (2003). Inter- and intragenic variations complicate the molecular epidemiology of human cytomegalovirus. *J Infect Dis* **187**, 809-19.

Rawlinson, W. D. & Barrell, B. G. (1993). Spliced transcripts of human cytomegalovirus. *J Virol* **67**, 5502-13.

Rawlinson, W. D., Farrell, H. E. & Barrell, B. G. (1996). Analysis of the complete DNA sequence of murine cytomegalovirus. *J Virol* **70**, 8833-49.

Reeves, M. B., Coleman, H., Chadderton, J., Goddard, M., Sissons, J. G. & Sinclair, J. H. (2004). Vascular endothelial and smooth muscle cells are unlikely to be major sites of latency of human cytomegalovirus in vivo. *J Gen Virol* **85**, 3337-41.

Reeves, M. B., MacAry, P. A., Lehner, P. J., Sissons, J. G. & Sinclair, J. H. (2005). Latency, chromatin remodeling, and reactivation of human cytomegalovirus in the dendritic cells of healthy carriers. *Proc Natl Acad Sci U S A* **102**, 4140-5.

Reeves, W. C., Corey, L., Adams, H. G., Vontver, L. A. & Holmes, K. K. (1981). Risk of recurrence after first episodes of genital herpes. Relation to HSV type and antibody response. *N Engl J Med* **305**, 315-9.

Rixon, F. J., Addison, C., McGregor, A., Macnab, S. J., Nicholson, P., Preston, V. G. & Tatman, J. D. (1996). Multiple interactions control the intracellular localization of the herpes simplex virus type 1 capsid proteins. *J Gen Virol* **77 (Pt 9)**, 2251-60.

Rock, D. L. & Fraser, N. W. (1983). Detection of HSV-1 genome in central nervous system of latently infected mice. *Nature* **302**, 523-5.

Rock, D. L. & Fraser, N. W. (1985). Latent herpes simplex virus type 1 DNA contains two copies of the virion DNA joint region. *J Virol* **55**, 849-52.

Rodems, S. M., Clark, C. L. & Spector, D. H. (1998). Separate DNA elements containing ATF/CREB and IE86 binding sites differentially regulate the human cytomegalovirus UL112-113 promoter at early and late times in the infection. *J Virol* **72**, 2697-707.

Roizman, B. (1979). The structure and isomerization of herpes simplex virus genomes. *Cell* **16**, 481-94.

Roizman, B. & Pellett, P. E. (2001). The Family Herpesviridae: A Brief Introduction. In *Fields Virology*, 4th edn, pp. 2381-2397: Lippincott Williams and Wilkins.

Roller, R., Zhou, Y., Schnetzer, R., Ferguson, J. & DeSalvo, J. R. (2003). CXCL10 production from cytomegalovirus-stimulated microglia is regulated by both human and viral interleukin-10. *J Virol* **77**, 4502-15.

Romanowski, M. J., Garrido-Guerrero, E. & Shenk, T. (1997). pIRS1 and pTRS1 are present in human cytomegalovirus virions. *J Virol* **71**, 5703-5.

Romanowski, M. J. & Shenk, T. (1997). Characterization of the human cytomegalovirus irs1 and trs1 genes: a second immediate-early transcription unit within irs1 whose product antagonizes transcriptional activation. *J Virol* **71**, 1485-96.

Rowe, W. P., Hartley, J. W., Waterman, S., Turner, H. C. & Huebner, R. J. (1956). Cytopathogenic agent resembling human salivary gland virus recovered from tissue cultures of human adenoids. *Proc Soc Exp Biol Med* **92**, 418-24.

Sarmati, L. (2004). HHV-8 infection in African children. *Herpes* **11**, 50-3.

Sarov, I. & Abady, I. (1975). The morphogenesis of human cytomegalovirus. Isolation and polypeptide characterization of cytomegalovirions and dense bodies. *Virology* **66**, 464-73.

Schleiss, M. R. & McVoy, M. A. (2004). Overview of congenitally and perinatally acquired cytomegalovirus infections: recent advances in antiviral therapy. *Expert Rev Anti Infect Ther* **2**, 389-403.

Schulz, T. F. (1998). Kaposi's sarcoma-associated herpesvirus (human herpesvirus-8). *J Gen Virol* **79 (Pt 7)**, 1573-91.

Scott, G. M., Barrell, B. G., Oram, J. & Rawlinson, W. D. (2002). Characterisation of transcripts from the human cytomegalovirus genes TRL7, UL20a, UL36, UL65, UL94, US3 and US34. *Virus Genes* **24**, 39-48.

Sears, A. E. & Roizman, B. (1990). Amplification by host cell factors of a sequence contained within the herpes simplex virus 1 genome. *Proc Natl Acad Sci U S A* **87**, 9441-4.

Severi, B., Landini, M. P., Cenacchi, G., Zini, N. & Maraldi, N. M. (1992). Human cytomegalovirus nuclear and cytoplasmic dense bodies. *Arch Virol* **123**, 193-207.

Sheaffer, A. K., Newcomb, W. W., Gao, M., Yu, D., Weller, S. K., Brown, J. C. & Tenney, D. J. (2001). Herpes simplex virus DNA cleavage and packaging proteins associate with the procapsid prior to its maturation. *J Virol* **75**, 687-98.

Shellam, G. R., Allan, J. E., Papadimitriou, J. M. & Bancroft, G. J. (1981). Increased susceptibility to cytomegalovirus infection in beige mutant mice. *Proc Natl Acad Sci U S A* **78**, 5104-8.

Shibuya, T. & Rigoutsos, I. (2002). Dictionary-driven prokaryotic gene finding. *Nucleic Acids Res* **30**, 2710-25.

Shinde, D., Lai, Y., Sun, F. & Arnheim, N. (2003). Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)n and (A/T)n microsatellites. *Nucleic Acids Res* **31**, 974-80.

Shukla, D., Spear, P. G. (2001). Herpesviruses and heparan sulfate: an intimate relationship in aid of viral entry. *J Clin Ivestig* **108**, 503-10.

Siminoff, P. & Menefee, M. G. (1966). Normal and 5-bromodeoxyuridine-inhibited development of herpes simplex virus. An electron microscope study. *Exp Cell Res* **44**, 241-55.

Sinzger, C., Grefte, A., Plachter, B., Gouw, A. S., The, T. H. & Jahn, G. (1995). Fibroblasts, epithelial cells, endothelial cells and smooth muscle cells are major targets of human cytomegalovirus infection in lung and gastrointestinal tissues. *J Gen Virol* **76 (Pt 4)**, 741-50.

Sinzger, C. & Jahn, G. (1996). Human cytomegalovirus cell tropism and pathogenesis. *Intervirology* **39**, 302-19.

Sinzger, C., Kahl, M., Laib, K., Klingel, K., Rieger, P., Plachter, B. & Jahn, G. (2000). Tropism of human cytomegalovirus for endothelial cells is determined by a post-entry step dependent on efficient translocation to the nucleus. *J Gen Virol* **81**, 3021-35.

Smiley, M. L., Wlodaver, C. G., Grossman, R. A., Barker, C. F., Perloff, L. J., Tustin, N. B., Starr, S. E., Plotkin, S. A. & Friedman, H. M. (1985). The role of pretransplant immunity in protection from cytomegalovirus disease following renal transplantation. *Transplantation* **40**, 157-61.

Smith, J. A., Jairath, S., Crute, J. J. & Pari, G. S. (1996). Characterization of the human cytomegalovirus UL105 gene and identification of the putative helicase protein. *Virology* **220**, 251-5.

Smith, J. A. & Pari, G. S. (1995). Human cytomegalovirus UL102 gene. *J Virol* **69**, 1734-40.

Smith, M. G. (1956). Propagation in tissue cultures of a cytopathogenic virus from human salivary gland virus (SGV) disease. *Proc Soc Exp Biol Med* **92**, 424-30.

Smith, M. S., Bentz, G. L., Alexander, J. S. & Yurochko, A. D. (2004). Human cytomegalovirus induces monocyte differentiation and migration as a strategy for dissemination and persistence. *J Virol* **78**, 4444-53.

Smuda, C., Bogner, E. & Radsak, K. (1997). The human cytomegalovirus glycoprotein B gene (ORF UL55) is expressed early in the infectious cycle. *J Gen Virol* **78 (Pt 8)**, 1981-92.

Snydman, D. R. & Falagas, M. E. (1996). Prevention of cytomegalovirus disease in transplant recipients. *Lancet* **347**, 268-9.

Sodeik, B., Ebersold, M. W. & Helenius, A. (1997). Microtubule-mediated transport of incoming herpes simplex virus 1 capsids to the nucleus. *J Cell Biol* **136**, 1007-21.

Sohn, Y. M., Oh, M. K., Balcarek, K. B., Cloud, G. A. & Pass, R. F. (1991). Cytomegalovirus infection in sexually active adolescents. *J Infect Dis* **163**, 460-3.

Sorg, G. & Stamminger, T. (1998). Strong conservation of the constitutive activity of the IE1/2 transcriptional control region in wild-type strains of human cytomegalovirus. *J Gen Virol* **79 (Pt 12)**, 3039-47.

Spaete, R. R., Gehrz, R. C. & Landini, M. P. (1994). Human cytomegalovirus structural proteins. *J Gen Virol* **75 (Pt 12)**, 3287-308.

Spear, G. T., Lurain, N. S., Parker, C. J., Ghassemi, M., Payne, G. H. & Saifuddin, M. (1995). Host cell-derived complement control proteins CD55 and CD59 are incorporated into the virions of two unrelated enveloped viruses. Human T cell leukemia/lymphoma virus type I (HTLV-I) and human cytomegalovirus (HCMV). *J Immunol* **155**, 4376-81.

Spear, P. G. & Longnecker, R. (2003). Herpesvirus Entry: an update. *J Virol* **77**, 10179-85.

Speckner, A., Glykofrydes, D., Ohlin, M. & Mach, M. (1999). Antigenic domain 1 of human cytomegalovirus glycoprotein B induces a multitude of different antibodies which, when combined, results in incomplete virus neutralization. *J Gen Virol* **80 (Pt 8)**, 2183-91.

Spector, D. H. (1996). Activation and regulation of human cytomegalovirus early genes. *Intervirology* **39**, 361-77.

Spector, S. A., Hirata, K. K. & Newman, T. R. (1984). Identification of multiple cytomegalovirus strains in homosexual men with acquired immunodeficiency syndrome. *J Infect Dis* **150**, 953-6.

Spector, S. A., Weingeist, T., Pollard, R. B., Dieterich, D. T., Samo, T., Benson, C. A., Busch, D. F., Freeman, W. R., Montague, P., Kaplan, H. J. & et al. (1993). A randomized, controlled study of intravenous ganciclovir therapy for cytomegalovirus peripheral retinitis in patients with AIDS. AIDS Clinical Trials Group and Cytomegalovirus Cooperative Study Group. *J Infect Dis* **168**, 557-63.

Spector, S. A., Wong, R., Hsia, K., Pilcher, M. & Stempien, M. J. (1998). Plasma cytomegalovirus (CMV) DNA load predicts CMV disease and survival in AIDS patients. *J Clin Invest* **101**, 497-502.

Spencer, J. V., Lockridge, K. M., Barry, P. A., Lin, G., Tsang, M., Penfold, M. E. & Schall, T. J. (2002). Potent immunosuppressive activities of cytomegalovirus-encoded interleukin-10. *J Virol* **76**, 1285-92.

Spencer, J. V., Lockridge, K. M., Barry, P. A., Lin, G., Tsang, M., Penfold, M. E. & Schall, T. J. (2002). Potent immunosuppressive activities of cytomegalovirus-encoded interleukin-10. *J Virol* **76**, 1285-92.

Spiller, O. B., Morgan, B. P., Tufaro, F. & Devine, D. V. (1996). Altered expression of host-encoded complement regulators on human cytomegalovirus-infected cells. *Eur J Immunol* **26**, 1532-8.

Staden, R., Beal, K. F. & Bonfield, J. K. (2000). The Staden package, 1998. *Methods Mol Biol* **132**, 115-30.

Stagno, S., Reynolds, D. W., Tsiantos, A., Fuccillo, D. A., Long, W. & Alford, C. A. (1975). Comparative serial virologic and serologic studies of symptomatic and subclinical congenitally and natally acquired cytomegalovirus infections. *J Infect Dis* **132**, 568-77.

Stagno, S., Pass, R. F., Dworsky, M. E. & Alford, C. A., Jr. (1982). Maternal cytomegalovirus infection and perinatal transmission. *Clin Obstet Gynecol* **25**, 563-76.

Stagno, S. & Cloud, G. A. (1994). Working parents: the impact of day care and breast-feeding on cytomegalovirus infections in offspring. *Proc Natl Acad Sci U S A* **91**, 2384-9.

Stamminger, T. & Fleckenstein, B. (1990). Immediate-early transcription regulation of human cytomegalovirus. *Curr Top Microbiol Immunol* **154**, 3-19.

Stannard, L. M., Fuller, A. O. & Spear, P. G. (1987). Herpes simplex virus glycoproteins associated with different morphological entities projecting from the virion envelope. *J Gen Virol* **68 (Pt 3)**, 715-25.

Staprans, S. I., Rabert, D. K. & Spector, D. H. (1988). Identification of sequence requirements and trans-acting functions necessary for regulated expression of a human cytomegalovirus early gene. *J Virol* **62**, 3463-73.

Starr, S. E., Glazer, J. P., Friedman, H. M., Farquhar, J. D. & Plotkin, S. A. (1981). Specific cellular and humoral immunity after immunization with live Towne strain cytomegalovirus vaccine. *J Infect Dis* **143**, 585-9.

Stasiak, P. C. & Mocarski, E. S. (1992). Transactivation of the cytomegalovirus ICP36 gene promoter requires the alpha gene product TRS1 in addition to IE1 and IE2. *J Virol* **66**, 1050-8.

Stenberg, R. M. (1996). The human cytomegalovirus major immediate-early gene. *Intervirology* **39**, 343-9.

Stenberg, R. M., Depto, A. S., Fortney, J. & Nelson, J. A. (1989). Regulated expression of early and late RNAs and proteins from the human cytomegalovirus immediate-early gene region. *J Virol* **63**, 2699-708.

Stenberg, R. M., Witte, P. R. & Stinski, M. F. (1985). Multiple spliced and unspliced transcripts from human cytomegalovirus immediate-early region 2 and evidence for a common initiation site within immediate-early region 1. *J Virol* **56**, 665-75.

Stevens, J. G., Haarr, L., Porter, D. D., Cook, M. L. & Wagner, E. K. (1988). Prominence of the herpes simplex virus latency-associated transcript in trigeminal ganglia from seropositive humans. *J Infect Dis* **158**, 117-23.

Stinski, M. F. (1976). Human cytomegalovirus: glycoproteins associated with virions and dense bodies. *J Virol* **19**, 594-609.

Storz, G. (2002). An expanding universe of noncoding RNAs. *Science* **296**, 1260-3.

Strang, B. L. & Stow, N. D. (2005). Circularization of the Herpes Simplex Virus Type 1 Genome upon Lytic Infection. *J Virol* **79**, 12487-94.

Szilagyi, J. F. & Berriman, J. (1994). Herpes simplex virus L particles contain spherical membrane-enclosed inclusion vesicles. *J Gen Virol* **75 (Pt 7)**, 1749-53.

Tatman, J. D., Preston, V. G., Nicholson, P., Elliott, R. M. & Rixon, F. J. (1994). Assembly of herpes simplex virus type 1 capsids using a panel of recombinant baculoviruses. *J Gen Virol* **75 (Pt 5)**, 1101-13.

Terhune, S. S., Schroer, J. & Shenk, T. (2004). RNAs are packaged into human cytomegalovirus virions in proportion to their intracellular concentration. *J Virol* **78**, 10390-8.

Thompson, R. L. & Sawtell, N. M. (2001). Herpes simplex virus type 1 latency-associated transcript gene promotes neuronal survival. *J Virol* **75**, 6660-75.

Thomsen, D. R., Roof, L. L. & Homa, F. L. (1994). Assembly of herpes simplex virus (HSV) intermediate capsids in insect cells infected with recombinant baculoviruses expressing HSV capsid proteins. *J Virol* **68**, 2442-57.

Tomasec, P., Braud, V. M., Rickards, C., Powell, M. B., McSharry, B. P., Gadola, S., Cerundolo, V., Borysiewicz, L. K., McMichael, A. J. & Wilkinson, G. W. (2000). Surface expression of HLA-E, an inhibitor of natural killer cells, enhanced by human cytomegalovirus gpUL40. *Science* **287**, 1031.

Tomasec, P., Wang, E. C., Davison, A. J., Vojtesek, B., Armstrong, M., Griffin, C., McSharry, B. P., Morris, R. J., Llewellyn-Lacey, S., Rickards, C., Nomoto, A., Sinzger, C. & Wilkinson, G. W. (2005). Downregulation of natural killer cell-activating ligand CD155 by human cytomegalovirus UL141. *Nat Immunol* **6**, 181-8.

Trus, B. L., Booy, F. P., Newcomb, W. W., Brown, J. C., Homa, F. L., Thomsen, D. R. & Steven, A. C. (1996). The herpes simplex virus procapsid: structure, conformational changes upon maturation, and roles of the triplex proteins VP19c and VP23 in assembly. *J Mol Biol* **263**, 447-62.

Tyler, K. L. & Fields, B. N. (1996). Pathogenesis of viral infections. In *Fields Virology*, pp. 173-217: Lippincott Williams and Wilkins.

Vales-Gomez, M., Browne, H. & Reyburn, H. T. (2003). Expression of the UL16 glycoprotein of Human Cytomegalovirus protects the virus-infected cell from attack by natural killer cells. *BMC Immunol* **4**, 4.

van der Meer, J. T., Drew, W. L., Bowden, R. A., Galasso, G. J., Griffiths, P. D., Jabs, D. A., Katlama, C., Spector, S. A. & Whitley, R. J. (1996). Summary of the International Consensus Symposium on Advances in the Diagnosis, Treatment and Prophylaxis and Cytomegalovirus Infection. *Antiviral Res* **32**, 119-40.

Vink, C., Beuken, E. & Bruggeman, C. A. (2000). Complete DNA sequence of the rat cytomegalovirus genome. *J Virol* **74**, 7656-65.

Vlazny, D. A., Kwong, A. & Frenkel, N. (1982). Site-specific cleavage/packaging of herpes simplex virus DNA and the selective maturation of nucleocapsids containing full-length viral DNA. *Proc Natl Acad Sci U S A* **79**, 1423-7.

Waldman, W. J., Roberts, W. H., Davis, D. H., Williams, M. V., Sedmak, D. D. & Stephens, R. E. (1991). Preservation of natural endothelial cytopathogenicity of cytomegalovirus by propagation in endothelial cells. *Arch Virol* **117**, 143-64.

Waltzek, T. B., Kelley, G. O., Stone, D. M., Way, K., Hanson, L., Fukuda, H., Hirono, I., Aoki, T., Davison, A. J. & Hedrick, R. P. (2005). Koi herpesvirus represents a third cyprinid herpesvirus (CyHV-3) in the family Herpesviridae. *J Gen Virol* **86**, 1659-67.

Wathen, M. W. & Stinski, M. F. (1982). Temporal patterns of human cytomegalovirus transcription: mapping the viral RNAs synthesized at immediate early, early, and late times after infection. *J Virol* **41**, 462-77.

Weller, T. H. (1971). The cytomegaloviruses: ubiquitous agents with protean clinical manifestations. II. *N Engl J Med* **285**, 267-74.

Weston, K. (1988). An enhancer element in the short unique region of human cytomegalovirus regulates the production of a group of abundant immediate early transcripts. *Virology* **162**, 406-16.

Whitley, R. J. & Roizman, B. (1997). Herpes simplex viruses. In *Clinical Virology*, pp. 375-410: Churchill Livinstone.

Wiertz, E., Hill, A., Tortorella, D. & Ploegh, H. (1997). Cytomegaloviruses use multiple mechanisms to elude the host immune response. *Immunol Lett* **57**, 213-6.

Wiertz, E. J., Jones, T. R., Sun, L., Bogyo, M., Geuze, H. J. & Ploegh, H. L. (1996). The human cytomegalovirus US11 gene product dislocates MHC class I heavy chains from the endoplasmic reticulum to the cytosol. *Cell* **84**, 769-79.

Wildy, P., Russell, W. C. & Horne, R. W. (1960). The morphology of herpes virus. *Virology* **12**, 204-22.

Winkler, M., Rice, S. A. & Stamminger, T. (1994). UL69 of human cytomegalovirus, an open reading frame with homology to ICP27 of herpes simplex virus, encodes a transactivator of gene expression. *J Virol* **68**, 3943-54.

Winkler, M. & Stamminger, T. (1996). A specific subform of the human cytomegalovirus transactivator protein pUL69 is contained within the tegument of virus particles. *J Virol* **70**, 8984-7.

Winston, D. J., Ho, W. G. & Champlin, R. E. (1990). Cytomegalovirus infections after allogeneic bone marrow transplantation. *Rev Infect Dis* **12 Suppl 7**, S776-92.

Wright, D. A., Staprans, S. I. & Spector, D. H. (1988). Four phosphoproteins with common amino termini are encoded by human cytomegalovirus AD169. *J Virol* **62**, 331-40.

Wyatt, J. P., Saxton, J. & et al. (1950). Generalized cytomegalic inclusion disease. *J Pediatr* **36**, 271-94, illust.

Yamanishi, K., Okuno, T., Shiraki, K., Takahashi, M., Kondo, T., Asano, Y. & Kurata, T. (1988). Identification of human herpesvirus-6 as a causal agent for exanthem subitum. *Lancet* **1**, 1065-7.

Yamanishi, K. (1997). Human herpesvirus 6 and human herpesvirus 7. In *Cliniacl Virology*, pp. 471-83: Churchill Livingstone.

Yeager, A. S., Grumet, F. C., Hafleigh, E. B., Arvin, A. M., Bradley, J. S. & Prober, C. G. (1981). Prevention of transfusion-acquired cytomegalovirus infections in newborn infants. *J Pediatr* **98**, 281-7.

Yeager, A. S., Palumbo, P. E., Malachowski, N., Ariagno, R. L. & Stevenson, D. K. (1983). Sequelae of maternally derived cytomegalovirus infections in premature infants. *J Pediatr* **102**, 918-22.

Yu, D., Smith, G. A., Enquist, L. W. & Shenk, T. (2002). Construction of a self-excisable bacterial artificial chromosome containing the human cytomegalovirus genome and mutagenesis of the diploid TRL/IRL13 gene. *J Virol* **76**, 2316-28.

Yu, D., Silva, M. C. & Shenk, T. (2003). Functional map of human cytomegalovirus AD169 defined by global mutational analysis. *Proc Natl Acad Sci U S A* **100**, 12396-401.

Zabolotny, J. M., Krummenacher, C. & Fraser, N. W. (1997). The herpes simplex virus type 1 2.0-kilobase latency-associated transcript is a stable intron which branches at a guanosine. *J Virol* **71**, 4199-208.

Zanghellini, F., Boppana, S. B., Emery, V. C., Griffiths, P. D. & Pass, R. F. (1999). Asymptomatic primary cytomegalovirus infection: virologic and immunologic features. *J Infect Dis* **180**, 702-7.

Zhou, Z. H., Chen, D. H., Jakana, J., Rixon, F. J. & Chiu, W. (1999). Visualization of tegument-capsid interactions and DNA in intact herpes simplex virus type 1 virions. *J Virol* **73**, 3210-8.

Zhu, H., Shen, Y. & Shenk, T. (1995). Human cytomegalovirus IE1 and IE2 proteins block apoptosis. *J Virol* **69**, 7960-70.

Zhu, Y., Huang, L. & Anders, D. G. (1998). Human cytomegalovirus oriLyt sequence requirements. *J Virol* **72**, 4989-96.

Zou, P., Isegawa, Y., Nakano, K., Haque, M., Horiguchi, Y. & Yamanishi, K. (1999). Human herpesvirus 6 open reading frame U83 encodes a functional chemokine. *J Virol* **73**, 5926-33.

Zweygberg Wirgart, B., Brytting, M., Linde, A., Wahren, B. & Grillner, L. (1998). Sequence variation within three important cytomegalovirus gene regions in isolates from four different patient populations. *J Clin Microbiol* **36**, 3662-9.

APPENDIX

The position refers to the AD169 genome (accession number NC_001347).

Region X.

Region X was amplified in three major fragments (Section 3.3). Also included in this list are internal sequencing primers (*italics*), primers used to generate ssRNA probes (Table 3.3) and primers used for 5'- and 3'-RACE (Section 3.5).

Primer	Sequences (5'-3')	Position	Strand
X1	GATGAGCCACATCTACGTGGCCAT	155,571	+
X2	GATGAACGTTAACCCGTTGCG	155,625	+
X5	GCCGAAGCCTCGCCGTGCTGCATA	161,731	-
X6	AGAAGTCTACCGAGTCGAGTTACA	157,537	+
X7	CGTTGGGAGGTAGCTTCGTCATGT	157,660	-
X8	CGCACGAAATCGATCCTAAGTGAG	159,471	+
X9	ACACGCAAACGCGAATGTACACGA	159,671	-
X10	*TCTTCGTAAACTTATCCGTT*	*155,967*	+
X11	*AGACCATAATGTTGACTACT*	*157,333*	-
X12	*TCAGTTGACAGCGAAGAGAC*	*157,899*	+
X13	*TTACTTTCACGGAAGATGCA*	*159,409*	-
X14	*ATCGAGAGCGGCTCGAACCT*	*159,877*	+
X15	*TTTCCAGCCCGCCTAGCAAC*	*161,409*	-
X16	*CCTGCATCACGGTGGCTCGA*	*155,872*	+
X18	AGAGCCAGCCCTTGTTGAGGT	161,697	-
X19	*GACGACGATTCTAAGAATTG*	*156,897*	-
X20	*CTATGGCTTCATCTATCGTG*	*158,317*	+
X21	*CTTCTCCTTCGTCCCAGACG*	*159,037*	-
X22	*TTCGTCTTGATCTCCAGCCG*	*160,351*	+
X23	*TTGACCACCGTACCGTCGAG*	*160,911*	-
X25	*ATGTGTTGTTGTCGTCGAGG*	*156,192*	+
X28	*GATTAGTCGGCTGTCTCAGG*	*161,178*	+
X29	*TCTGGTATCTACTGCCTGAC*	*160,170*	-
X30	*TCGTTCTCATGATCGCGGAG*	*159,238*	+
X32	*CCTCTCTCTAATTCCCTGGA*	*156,465*	-
X34	*GTGTGGCAGAGAGGATGCCTT*	*156,329*	+
X35	*GTGATGTGGTGGTACATTGA*	*156,667*	+
X36	*TCATGGCGGTATTTCTCTTCC*	*158,786*	+

X37	*CTGACGGGTTATCGTCACCA*	*160,465*	-
X38	*CCTACGTTGGTAGGTCACGT*	*160,780*	+
X40	AACAACCAAAGGACGACC	156,192	-
X41	*AACAACCAAAGGACGACC*	*156,192*	-
X43	*CTTAGGAGGAACGATGAATC*	*158,258*	-
X48	CCACCCGGGTTTCTTCTCTTGCAG	156,058	-
X50	CAAGACGCAGCAGGGCTTCATGTC	161,738	+
X53	CCACGTTGCGTCGTGACGTTG	161,360	+
X54	GCAAAGCGCGAGCGCAAC	161,508	-
X57	ATGCTGTCGGTGATGGTCTCTTCC	160,609	+
X59	ACATTGCCGCATGTCTTTGTAG	161,065	-
X61	CCTCGCGAGGCTGACAGTCGACGG	156,004	+
X66	GGAGGCGAAGCCGGCGACGACGAC	160,673	+

Region O.

Region O was amplified in 12 fragments (Table 4.1). Also included in this list are internal sequencing primers (*italics*).

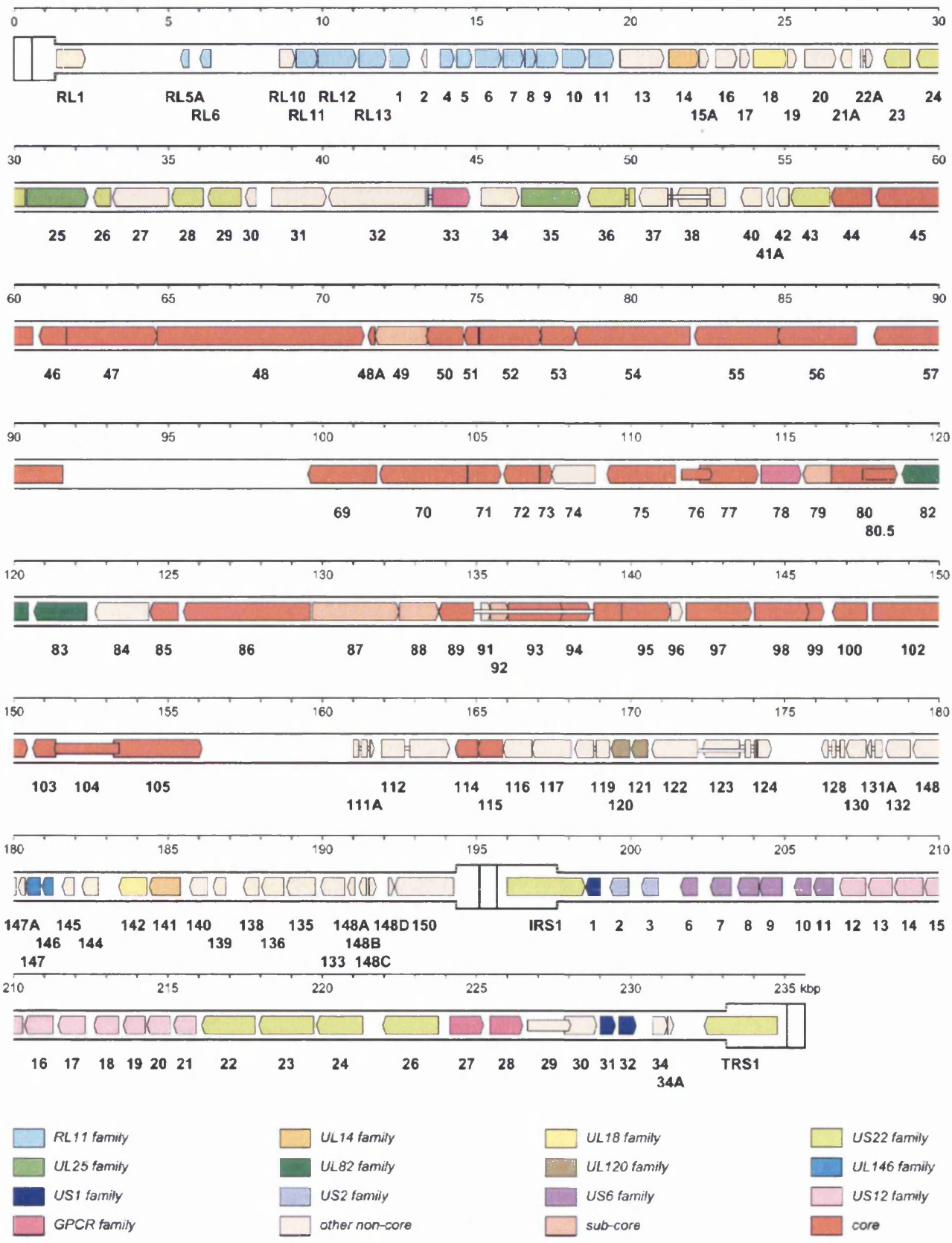
Primer	Sequences (5'-3')	Position	Strand
O3	TCCAATGGGACGGCGTTTCC	93,479	+
O6	TTGATGACGCCGTCGCCTTC	99,366	-
O7	CCAGAATCTCCTGCGTTTCTG	91,119	+
O13	CTCGTTGACCAGGAAAGCTGTC	92,498	-
O16	CCCGTCGTACGAGGAACAAG	98,224	-
O17	CCGTATCCTTCCCTGTTGTAC	96,845	+
O18	TGGTGGTCATGGCGGTGGTGGTCA	97,007	-
O19	GAGAAAATGGCGTCGAGAGCCTAG	95,584	+
O20	CTCGTCGAGATTTTCAGGGCTATC	95,734	-
O21	AGTCGTGACCGGGGGTCCCT	94,748	+
O24	*TCTCGCAGCCAGACCGCTA*	*91,420*	+
O25	*GTGTTTCATGGCAATGGG*	*92,193*	-
O26	*CTGACGTCAAACCATCTC*	*95,896*	+
O27	*GTCCGCCCGTCCTTCGTA*	*97,227*	+
O28	CAACGATCGTGAAACTGGAGACTG	98,090	+
O29	*GACGGTGATTCTTCAGAGAATC*	*97,941*	-
O30	*ATAGGTGCCGGACGCCGGCTAC*	*96,688*	-
O31	*GTAACATGATCAACCACGTCTG*	*97,739*	+
O33	*CCTACACATTTCTAGTCGTGAC*	*97,040*	-

O36	CACCGTGATGGGGAACGGTGTT	92,191	+
O41	GGTAGGCGATTTGCATCTG	96,390	+
O42	CCAAGAATAAAACTCTAGCTC	96,724	+
O45	GGTTTCCCGAACACGTAC	92,003	+
O46	CCTCGGCACGACCGTATTTG	97,372	-
O47	CACGGTTTGTTAGCTGATGTG	91,742	-
O51	CCATAAGCTGTGTACTGTTC	98,948	-
O52	CCTCAAATGTTGTACATGTG	98,443	+
O54	TGGGACGGTCGTGGTCTCCTC	95,260	-
O55	CATCATCGGATATCTCACCAC	99,008	+
O56	GATGCGAATTGGTGGCTATGGTG	98,384	-
O60	GTTCCGCCAGTCCGTTTTCGTAG	93,387	-
O61	CAGCACACACGCCACAGCCT	94,973	-
O63	CTCCACGCGTTTTCCACCCT	95,165	+
O64	GTTCTCTCGCTCGTCATG	95,538	+
O65	TGAACAGAGAAAGGTGGG	95,466	-
O66	CTACCAATCACAGAAACGG	92,781	+
O67	CAGGGGTCACGTGAGAAG	93,083	-
O68	TGGGTGTGGCTTCACTAGG	92,722	-
O69	GGTGCCAGATGCCAACGT	92,620	+
O72	GACGCATGCGCACTGACCGCCTC	93,762	-
O73	GAGCGGTAATTTTCCACCGCC	93,883	-
O76	GGTATTTTTCCACTGGGCGG	93,801	+
O78	GAAGCGAGCTACGTCATCAGTC	93,238	+
O83	CTTCCCGGTCCACCCGCC	94,325	+
O84	TCGCCCCGACTGCGCATG	93,521	-
O86	CTGCGCATGCGCCGGTAAAATTCC	93,676	+
O88	CACCTAGGGAGCGCGAG	93,824	+
O90	GACCCGAGCGGGGGCTTC	94,238	+
O91R	GGAAGGGGAGCTGGGTGG	94,468	-
O93	CTGGAGCACCATCCGGGG	94,077	+
O95	GACCACGCCCAGGGCCAGCA	94,896	-
O96	GGGGAGGGTGGGGGGTAG	94,031	-
O96R	CTACCCCCCACCCTCCCC	94,014	+
O101R	GGGAGCAGCGAGGGATCC	94,309	-

Region G.

Region G was amplified in two major fragments (Section 5.2). Also included in this list are internal sequencing primers (*italics*), primers used to generate ssRNA probes (Table 5.1) and primers used for 5'- and 3'-RACE (Table 5.2). Primers labelled .1, .2, etc. are strain-specific.

Primer	Sequences (5'-3')	Position	Strand
G1	CCAGAGTAGCGGTGACGAGAGTAAC	155,571	+
G1R	GTTACTCTCGTCACCGCTACTCTGG	155,594	-
G2	CAACGACGTTGGAAGAAAACGGTG	227,350	-
G3R	GAATCGCGGCACAACGACTGGAC	225,957	-
G4R	GTCGGTGATTTTCACGGAGCGTC	225,836	+
G5	*CCAATTGGCGGCGACAGA*	*224,478*	+
G6	GACAGCCATTTTCAGCGAC	224,429	+
G7	*CTCTCCAGGTCCGGACTG*	*225,599*	-
G8	*GCGTGTCCTGGTTTTTCA*	*226,231*	+
G9	*TGCGACGCGTCACGTTTTC*	*227,100*	-
G10	*GTCCCCATCATCAGCGTC*	*224,942*	+
G11	*CTCCTGTGGGAACCCGCTC*	*225,051*	-
G12.1	*CAAGCCGTGGTGCGTGTT*	*226,676*	+
G12.2	*CAACATCAGTGTAGCGTG*	*226,676*	+
G12.3	*CAAATCGTCGGGTGCGAC*	*226,676*	+
G12.4	*CAAACTGCGGGGCGTGTC*	*226,676*	+
G13.1	*CCACTGTTTTATTCTTGGCG*	*226,596*	-
G13.2	*CACCGTTTTACCCATGAC*	*226,596*	-
G13.3	*CACCGCTTTACCCATGAC*	*226,596*	-
G14	*GACGGCGATGAACTTAGAG*	*225,406*	+
G15	*GAGGAAGACGCCGACGAC*	*224,604*	-
G16.1	*CTTATCACTGGCGTCGTT*	*226,971*	+
G16.2	*CTTATCCGTGGCGTTGTG*	*226,971*	+
G16.3	*CTCATTCGTGGCGTTGTT*	*226,971*	+
G17	*GAAAAACCAGGACACGCA*	*226,247*	+
G21	GTTCTTCATACCGCCGAAGG	226,048	-
G22	GTGGAACTGTTGCGACAATAC	226,946	-
G23	CTCTGCAACGCCTGGCGTGATC	223,919	+
G24	GTCCCATGCCAACGGTCCCCAAG	227,301	-
G27	CACGGAAAAGACTGCCGACGCCGTG	226,091	-
G28	GAGGCGCTGCTCTGAAGCCAAGTG	225,395	-
G29	GAGAGGGTGGGTTACGAGAAACTG	225,089	+


0 5 10 15 20 25 30
RL1 RL5A RL6 RL10 RL11 RL12 RL13 1 2 4 5 6 7 8 9 10 11 13 14 15A 16 17 18 19 20 21A 22A 23 24
30 35 40 45 50 55 60
25 26 27 28 29 30 31 32 33 34 35 36 37 38 40 41A 42 43 44 45
60 65 70 75 80 85 90
46 47 48 48A 49 50 51 52 53 54 55 56 57
90 95 100 105 110 115 120
69 70 71 72 73 74 75 76 77 78 79 80 80.5 82
120 125 130 135 140 145 150
83 84 85 86 87 88 89 91 92 93 94 95 96 97 98 99 100 102
150 155 160 165 170 175 180
103 104 105 111A 112 114 115 116 117 119 120 121 122 123 124 128 130 131A 132 148
180 185 190 195 200 205 210
147A 146 147 145 144 142 141 140 139 138 136 135 148A 133 148B 148C 148D 150 IRS1 1 2 3 6 7 8 9 10 11 12 13 14 15
210 215 220 225 230 235 kbp
16 17 18 19 20 21 22 23 24 26 27 28 29 30 31 32 34 34A TRS1
RL11 family
UL14 family
UL18 family
US22 family
UL25 family
UL82 family
UL120 family
UL146 family
US1 family
US2 family
US6 family
US12 family
GPCR family
other non-core
sub-core
core