

**TRANSCRIPT MAPPING IN
HUMAN CYTOMEGALOVIRUS STRAIN AD169**

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

**by
Parvis Akter
February 2002**

**Division of Virology
Church Street
Glasgow G11 5JR**

ProQuest Number: 13818485

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 13818485

Published by ProQuest LLC (2018). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

GLASGOW
UNIVERSITY
LIBRARY:

12651

copy 1

ACKNOWLEDGEMENTS

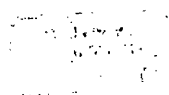
First and foremost, I thank God for His mercy and grace to me.

I would like to thank the following people who have contributed to the completion of this thesis.

My supervisor, Dr. A. J. Davison, for his support, wisdom and advice throughout the course of this work and also for his thorough criticism of this manuscript; your patience is a virtue. My assessor, Dr. J. Parkinson, for her support and advice throughout the project. Prof. D. J. McGeoch for his words of encouragement and support during the difficult stage of writing up. Dr. D. Dargan for his continuous support and interest in my work throughout the course of these studies. Dr. G. Wilkinson for provision of DNA from HCMV strain GW. Lesley Taylor and Mairi Hope for doing all the sequencing in these studies. Charles Cunningham for being a great lab father. The washroom and media staff for provision of glassware and general solutions used in this project. Everyone else who helped a little bit every day (Fiona, Richard, Dorothy, Wei-Li, Hugo, Ralph and all others) - thank you.

My Mum, Dad, sisters and brothers for their constant support and endurance. My son Ruhan, who faced the unpleasant experience of missing his Mum at an early age in his life. I shall always be indebted to him. Most of all, I would like to thank my husband, Sulaman. Without his unfaltering help and support I certainly would not be here.

During the course of this study, the author was in receipt of a Commonwealth Scholarship, and unless otherwise stated, all results were obtained by the author's own efforts.



SUMMARY

Human cytomegalovirus (HCMV) is of considerable medical importance, with infection *in utero* being a major health risk for the developing foetus, causing a variety of neonatal abnormalities including deafness, physical abnormality and mental retardation. HCMV also poses a life-threatening risk to immunosuppressed individuals such as allograft recipients and HIV-infected people. HCMV is responsible for the blindness due to retinitis that can affect some AIDS patients.

The gene content of HCMV is less well understood than that of any other human herpesvirus. This reflects the large size and complexity of the genome, and also the lack of a laboratory strain with the full genetic complement of wild type virus. The complete DNA sequence (229,354 bp) of HCMV strain AD169 was published in 1990, and the genome was predicted to contain 208 protein-coding open reading frames (ORFs). This is not likely to be an accurate estimate of the actual number of genes, as the criteria employed to identify coding regions were necessarily arbitrary and applied without the benefit of comparisons with other betaherpesviruses. Moreover, HCMV strain Toledo and other low passage isolates were later found to possess a 15 kbp genome segment absent from AD169. Recently, the gene content of HCMV has been revised by comparison to the chimpanzee cytomegalovirus (CCMV) sequence (241,087 bp), and the number of protein-coding genes in AD169 is now estimated at 145, several of which are novel. It is anticipated that this picture of the gene content of HCMV will be improved further.

The HCMV genome contains a set of 41 conserved herpesvirus-common 'core' genes, which are arranged in blocks that maintain relative position and orientation in different herpesviruses and reflect evolution from a common ancestor. The majority of genes are not spliced and overall the genome has relatively few polyadenylation signals. At the outset of this project, 12 HCMV genes had been shown experimentally to be spliced, and more spliced genes probably remained to be identified. Ten different families of related genes (RL11, US6, US22, GCR, UL25, UL82, UL146, US1, US12 and US22) have been recognised in HCMV that appear to have been generated by gene duplication events.

The US22 gene family contains 13 distantly related members (UL23, UL24, UL26, UL36, UL43, US22, US23, US24, US26, TRS1 and IRS1) sharing one or more of four conserved amino acid sequence motifs. Three members of this family (UL36, TRS1 and IRS1) have been reported as exhibiting transcriptional trans-activating properties in transient transfection assays, indicating that US22 genes are likely to be regulatory proteins. Moreover, since each of the sequenced betaherpesviruses contains a similar number of US22 genes, it is anticipated that these genes provide important functions during virus replication.

Although the AD169 genome was sequenced over ten years ago, the products of a large number of HCMV genes have not been identified, and the assignment of gene function is largely based on sequence similarity to homologous genes in herpes simplex virus type 1 (HSV-1). Transcript mapping data are also fragmentary. The purpose of this study was to evaluate transcription of a selection of AD169 genes, including several that are conserved in CCMV and some that appear unlikely to encode functional proteins because the HCMV ORFs are not conserved in CCMV. Primary use was made of northern blotting, RT-PCR and RACE techniques, employing RNA isolated from infected human fibroblasts. Three groups of genes were analysed: the 13 members of the US22 gene family; the 14 ORFs in TR_L and 30 adjacent ORFs at the left end of U_L; and the novel spliced genes UL128 and UL131A.

Transcripts were detected by northern blotting for nine members of the US22 family, and 5'- and 3'-ends were identified for eight. Failure to obtain data for the other members analysed was probably due to transcription at low levels. RNAs were identified for most ORFs in TR_L and the adjacent part of U_L. Most of the 5'-ends are located 20-30 bp downstream from TATA elements, and all the 3'-ends are located 20-24 bp downstream from polyadenylation signals. The 5'-ends of two genes (UL18 and US24) appeared to be located downstream from the first ATG codon in the relevant ORF. Transcripts were detected for five ORFs in TR_L and one in U_L that appear unlikely to encode proteins. Certain ORFs in TR_L and U_L have more than one 5'-end, suggesting that they are transcribed in a complex manner.

PCR and the RACE analysis confirmed expression of three novel genes predicted from comparisons of HCMV with CCMV (UL21A, UL128 and UL131A). RT-PCR analysis confirmed that UL128 and UL131A are spliced genes, the former consisting of three exons and the latter of two. The 5'- and 3'-ends of UL21A and UL128 was mapped, but attempts to detect the 5'-end of UL131A were unsuccessful.

PCR and sequence analysis led to the identification of errors in the original sequences of ORFs UL15, UL102, UL145 and US22, as predicted from comparisons between HCMV and CCMV. In addition, two other ORFs in AD169 (UL131A and TRL13/14) were shown to be naturally frameshifted.

This work consolidates our understanding of the genetic content of HCMV and transcription of HCMV genes. It also provides a foundation for further experimental studies of gene expression in HCMV.

CONTENTS

	Page
Contents	i
List of Figures	vi
List of Tables	ix
Abbreviations	xi
 CHAPTER 1: INTRODUCTION	
1.1 The <i>Herpesviridae</i>: overview	1
1.1.1 General properties.....	1
1.1.2 Virion structure.....	1
1.1.3 Herpesvirus classification.....	3
1.1.4 Genetic content.....	5
1.1.5 Life cycle.....	8
1.1.5.1 <i>The lytic cycle</i>	8
1.1.5.2 <i>The latent cycle</i>	10
1.2 Human herpesviruses	12
1.3 Human cytomegalovirus	15
1.3.1 General characteristics.....	15
1.3.2 Clinical relevance.....	16
1.3.2.1 <i>Epidemiology</i>	16
1.3.2.2 <i>Disease and therapy</i>	17
1.3.3 Cell tropism.....	19
1.3.4 HCMV particles.....	21
1.3.4.1 <i>The capsid</i>	21
1.3.4.2 <i>The tegument</i>	24
1.3.4.3 <i>The envelope</i>	26
1.3.5 HCMV genome.....	27
1.3.5.1 <i>Smith and Pari (1995)</i>	31
1.3.5.2 <i>Cha et al. (1996)</i>	32
1.3.5.3 <i>Dargan et al. (1997) and Mocarski et al. (1997)</i>	32
1.3.5.4 <i>Davison et al. (unpublished data)</i>	33

1.3.6	Regions in HCMV genome not encoding protein.....	33
1.3.6.1	<i>TRL2-TRL9</i>	36
1.3.6.2	<i>UL58-UL68</i>	36
1.3.6.3	<i>UL106-UL111</i>	37
1.3.7	Relationship to other herpesviruses.....	38
1.3.8	Spliced transcripts in HCMV.....	45
1.3.9	The HCMV US22 gene family.....	47
1.3.10	Gene expression.....	51
1.3.11	Transcriptional trans-activators in HCMV.....	53
1.3.11.1	<i>Major immediate early (MIE) gene locus</i>	53
1.3.11.2	<i>Other transactivators</i>	58
1.4	Justification of the study	60
1.5	Specific objectives of the study	61

CHAPTER 2: MATERIALS AND METHODS

2.1	Materials	62
2.1.1	Viruses.....	62
2.1.2	Cells.....	62
2.1.3	Cell culture media.....	62
2.1.4	Cosmids.....	62
2.1.5	Chemicals.....	62
2.1.6	Solutions and buffers.....	63
2.1.7	RNA extraction.....	67
2.1.8	Northern blotting.....	68
2.1.9	PCR techniques: PCR, RT-PCR and SMART RACE PCR.....	69
2.1.10	Agarose gel electrophoresis.....	69
2.1.11	Recovery, cloning and sequencing of DNA fragments.....	70
2.1.12	Primer extension.....	70
2.1.13	Bacterial strains.....	71
2.1.14	Bacterial growth media.....	71
2.1.15	Autoclaving and glassware sterilisation.....	71
2.2	Methods	72
2.2.1	Cell culture.....	72

2.2.2	Preparation of virus stocks.....	72
2.2.3	Titration of virus.....	72
2.2.4	Preparation of immediate early (IE), early (E), late (L) and mock infected (MI) RNA.....	73
2.2.4.1	<i>Preparation of IE RNA</i>	73
2.2.4.2	<i>Preparation of E RNA</i>	73
2.2.4.3	<i>Preparation of L RNA</i>	74
2.2.4.4	<i>Preparation of MI RNA</i>	74
2.2.5	Preparation of total cellular RNA.....	74
2.2.6	Isolation of polyA ⁺ mRNA.....	75
2.2.6.1	<i>Isolation of polyA⁺ mRNA from total cellular RNA</i>	75
2.2.6.2	<i>Isolation of polyA⁺ mRNA directly from cells</i>	75
2.2.6.3	<i>Determination of RNA yield</i>	77
2.2.7	Northern blotting.....	77
2.2.7.1	<i>Agarose gel electrophoresis of polyA⁺ RNA</i>	77
2.2.7.2	<i>Transfer of RNA to membrane</i>	78
2.2.7.3	<i>Preparation of RNA probes</i>	79
2.2.7.4	<i>Preparation of DNA probe</i>	83
2.2.7.5	<i>Nucleic acid hybridisation</i>	83
2.2.7.6	<i>Phosphorimager analysis</i>	84
2.2.8	5'- and 3'-RACE.....	84
2.2.9	RT-PCR.....	88
2.2.10	PCR.....	89
2.2.11	Agarose gel electrophoresis.....	90
2.2.12	Recovery of DNA fragments and cloning.....	90
2.2.13	Miniprep plasmid DNA preparation.....	90
2.2.14	Restriction endonuclease digests.....	91
2.2.15	DNA sequencing.....	91
2.2.16	DNA sequence analysis.....	91
2.2.17	Primer extension.....	92
2.2.17.1	<i>5'-labelling of primer</i>	92
2.2.17.2	<i>Labelling of DNA markers</i>	93

2.2.17.3	Primer extension reaction.....	93
2.2.17.4	Polyacrylamide gel electrophoresis.....	94

CHAPTER 3: POTENTIAL ERRORS IN THE HCMV DNA SEQUENCE

3.1	Sequences of PCR products.....	96
3.1.1	UL15.....	96
3.1.2	UL102.....	99
3.1.3	UL145.....	108
3.1.4	US22.....	108
3.1.5	UL131A.....	117
3.1.6	TRL13/TRL14.....	121
3.2	Discussion.....	124

CHAPTER 4: VALIDATION OF THE METHODS

4.1	SMART RACE.....	126
4.2	Northern blot.....	130
4.3	Primer extension.....	136
4.4	RT-PCR.....	137
4.5	Discussion.....	140

CHAPTER 5: TRANSCRIPTION OF GENES IN THE HCMV US22 FAMILY

5.1	Mapping the 5'- and 3'-ends of US22 genes by RACE.....	141
5.1.1	Sequences of 5'-ends.....	141
5.1.2	Sequences of 3'-ends.....	151
5.2	Mapping the 5'-ends of US22 genes by primer extension analysis.....	157
5.3	Mapping US22 genes by northern blot.....	160
5.4	Spliced genes in the US22 gene family.....	166
5.5	Discussion.....	170

CHAPTER 6: TRANSCRIPTION OF HCMV ORFs TRL1-TRL14

6.1	Sequences of 5'-ends in TR_L.....	175
6.1.1	5'-ends of ORFs that are conserved in CCMV.....	175

6.1.2	5'-ends of ORFs that are not conserved in CCMV.....	182
6.2	Mapping the 5'-ends of genes in the TR_L region by primer extension.....	185
6.3	Discussion.....	185
 CHAPTER 7: TRANSCRIPTION OF HCMV ORFs UL1-UL30		
7.1	Mapping 5'- ends of genes in the U_L region.....	189
7.2	Discussion.....	208
 CHAPTER 8: TRANSCRIPT MAPPING OF TWO NOVEL HCMV GENES		
8.1	UL128.....	211
8.1.1	Characterisation of spliced transcripts from UL128.....	211
8.1.2	Mapping the 5'- and 3'-ends of spliced UL128.....	218
8.1.3	Northern blot analysis of UL128 transcript.....	218
8.2	UL131A.....	221
8.2.1	Characterisation of spliced transcripts from UL131A.....	221
8.2.2	Mapping the 5'- and 3'-ends of spliced UL131A.....	221
8.2.3	Northern blot analysis of UL131A transcript.....	227
8.3	Discussion.....	227
 CHAPTER 9: DISCUSSION		
9.1	Overall discussion.....	231
9.1.1	Purpose of the study.....	231
9.1.2	Sequencing errors.....	234
9.1.3	Novel genes in HCMV.....	235
9.1.4	US22 family genes.....	236
9.1.5	ORFs in TR _L and at left end of U _L	237
9.1.6	HCMV ORFs that are transcribed but may not encode functional proteins.....	238
9.2	Conclusion.....	241
 REFERENCES.....		242

LIST OF FIGURES

		Page
Figure 1.1	Herpesvirus morphology.....	2
Figure 1.2	Phylogeny of herpesviruses.....	6
Figure 1.3	Types of herpesvirus genome structure.....	7
Figure 1.4	Organisation of conserved gene blocks in members of different herpesvirus subfamilies.....	9
Figure 1.5	The replication cycle of HSV-1.....	11
Figure 1.6	HCMV particles.....	20
Figure 1.7	HCMV morphology.....	23
Figure 1.8	Putative protein-coding ORFs in the AD169 genome as predicted by Chee <i>et al.</i> (1990).....	29
Figure 1.9	Modern interpretation of the AD169 genome according to Davison <i>et al.</i> (unpublished data).....	34
Figure 1.10	Three conserved motifs present in the HCMV US22 gene family members.....	50
Figure 1.11	Structural organisation of the AD169 major IE gene locus.....	54
Figure 1.12	Regulation of HCMV gene expression.....	55
Figure 2.1	Summary of Lig'nScribe and MAXIscript reactions.....	81
Figure 2.2	Mechanism of 5'-RACE reactions.....	86
Figure 2.3	Mechanism of 3'-RACE reactions.....	87
Figure 3.1	Gel electrophoresis of PCR products from regions of the HCMV genome containing potential errors.....	98
Figure 3.2	Comparison of the region containing AD169 UL15 with the cognate CCMV sequence.....	100
Figure 3.3	Comparison of the corrected sequence of the region containing AD169 UL15A with the cognate CCMV sequence.....	102
Figure 3.4	Comparison of the region containing AD169 UL102 with the cognate CCMV sequence.....	104
Figure 3.5	Comparison of the corrected sequence of the region containing AD169 UL102 with the cognate CCMV sequence.....	106

Figure 3.6	Comparison of the region containing Toledo UL145 with the cognate CCMV sequence.....	109
Figure 3.7	Comparison of the corrected sequence of the region containing Toledo UL145 with the cognate CCMV sequence.....	111
Figure 3.8	Comparison of the region containing AD169 US22 with the cognate CCMV sequence.....	113
Figure 3.9	Comparison of the corrected sequence of the region containing AD169 US22 with the cognate CCMV sequence.....	115
Figure 3.10	Comparison of the region containing AD169 UL131A with the cognate CCMV sequence.....	118
Figure 3.11	Corrected sequence of AD169 UL131A.....	120
Figure 3.12	Comparison of the region containing AD169 TRL13/14 with the cognate CCMV sequence.....	122
Figure 4.1	5'-RACE of control HCMV genes.....	131
Figure 4.2	3'-RACE of control HCMV genes.....	132
Figure 4.3	Sequences at the 5'- and 3'-ends of the UL22A mRNA.....	133
Figure 4.4	Northern blot analysis of control HCMV genes.....	135
Figure 4.5	Gel electrophoresis of RT-PCR products from control HCMV genes.....	138
Figure 4.6	Interpretation of splicing in HCMV UL36 by RT-PCR.....	139
Figure 5.1	5'-RACE of HCMV US22 genes.....	144
Figure 5.2	Sequences at the 5'-end of the US22 mRNA.....	149
Figure 5.3	3'-RACE of HCMV US22 genes.....	153
Figure 5.4	Sequences at the 3'-end of the US22 mRNA.....	155
Figure 5.5	Primer extension analysis of transcripts from the US22 gene family.....	158
Figure 5.6	Gel electrophoresis of Lig'nScribe PCR products used to generate RNA probes.....	161
Figure 5.7	Northern blot analysis of transcripts from members of the US22 gene family.....	163
Figure 5.8	Comparison of the region containing AD169 UL28/UL29 with the cognate CCMV sequence.....	167

Figure 5.9	Gel electrophoresis of RT-PCR products potentially from a spliced UL28 transcript.....	169
Figure 5.10	Positions of the 5'- and 3'-ends of US22 gene family members...	171
Figure 6.1	5'-RACE of ORFs in the TR _L region.....	179
Figure 6.2	Position of 5'-ends as represented in 5' cDNA clones of TRL11..	183
Figure 6.3	Positions of the 5'-ends of HCMV ORFs TRL1-TRL14.....	186
Figure 7.1	Positions of the 5'-ends of HCMV ORFs UL1-UL30.....	193
Figure 7.2	5'-RACE of HCMV genes UL1-UL30.....	196
Figure 7.3	5'-RACE of the HCMV UL18 mRNA.....	203
Figure 7.4	Sequence at a 5'-end of the UL18 mRNA.....	204
Figure 7.5	Comparison of the region containing AD169 UL21 with the cognate CCMV sequence.....	206
Figure 8.1	Comparison of the region containing AD169 UL128 with the cognate CCMV sequence.....	213
Figure 8.2	Gel electrophoresis of RT-PCR products from HCMV UL128....	215
Figure 8.3	Interpretation of splicing in HCMV UL128 by RT-PCR.....	216
Figure 8.4	5'- and 3'-RACE of HCMV UL128 mRNA.....	219
Figure 8.5	Northern blot analysis of HCMV UL128.....	220
Figure 8.6	Gel electrophoresis of RT-PCR products from HCMV UL131A..	222
Figure 8.7	Interpretation of splicing in HCMV UL131A by RT-PCR.....	223
Figure 8.8	5'- and 3'-RACE of HCMV UL131A mRNA.....	225
Figure 8.9	Northern blot analysis of HCMV UL131A.....	226
Figure 8.10	Transcription from HCMV novel genes, UL128 and UL131A....	228
Figure 9.1	Transcription from HCMV genes.....	232

LIST OF TABLES

		Page
Table 1.1	Herpesvirus classification and accession numbers for sequenced genomes.....	4
Table 1.2	The eight herpesviruses known to infect humans and their corresponding designations by the ICTV.....	13
Table 1.3	Summary of the clinical manifestations of the eight human herpesviruses.....	14
Table 1.4	HCMV strains in research use.....	16
Table 1.5	Functions of HCMV genes.....	39
Table 1.6	Published splice sites in AD169.....	46
Table 2.1	Gene-specific primers for making RNA probes.....	80
Table 2.2	Primers used in RT-PCR.....	89
Table 2.3	Primers used in primer extension experiments.....	93
Table 3.1	PCR primers designed to investigate potential errors in the HCMV sequence.....	97
Table 3.2	Corrections to HCMV DNA sequences.....	124
Table 4.1	Primers used in RACE to map 5'- and 3'-ends of transcript from UL22A and UL36.....	127
Table 5.1	Primers used in 5'-SMART RACE PCR to map the 5'-ends of US22 genes.....	142
Table 5.2	5'-ends of HCMV US22 genes.....	150
Table 5.3	Primers used in 3'-RACE to map the 3'-ends of US22 genes.....	152
Table 5.4	3'-ends mapped for HCMV US22 genes.....	156
Table 5.5	Primer extension analysis of transcripts from the US22 gene family.....	159
Table 5.6	Kinetics of expression of HCMV US22 genes.....	165
Table 6.1	Primers used in 5'-RACE to map 5'-ends of transcripts from ORFs in TR _L	176
Table 6.2	5'-ends of HCMV ORFs TRL1-TRL14.....	177

Table 7.1	Primers used in 5'-RACE to map 5'-ends of transcripts from UL1-UL30.....	190
Table 7.2	5'-ends of HCMV ORFs UL1-UL30.....	191
Table 8.1	Primers used in RACE to map 5'- and 3'-ends of transcript from UL128 and UL131A.....	212

Abbreviations

aa	amino acid
A	adenine
AIDS	acquired immune deficiency syndrome
ATP	adenosine triphosphate
BPB	bromophenol blue
bp	base pair
C	cytosine
°C	degrees Celsius
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dNTPs	deoxyribonucleoside triphosphates
DR	direct repeat
DTT	dithiothreitol
E	early
EDTA	ethylenediaminetetra-acetic acid
EM	electron micrograph
EtBr	ethidium bromide
EtOH	ethanol
FCS	foetal calf serum
g	gram
G	guanine
g/l	gram/litre
gp	glycoprotein
GTP	guanosine triphosphate
h	hour
IE	immediate-early
IPTG	isopropyl- β -D-thiogalactoside
IR _L	internal long repeat
IR _S	internal short repeat
kb	kilobase
kbp	kilobase pair
L	late
l	litre
M	molar
MI	mock infected
MOPS	[3-(N-morpholino) propanesulfonic acid]
mg	milligram
min	minute
ml	millilitre
mM	millimolar
moi	multiplicity of infection
mRNA	messenger ribonucleic acid
Mw	molecular weight
μ Ci	microcurie

μg	microgram
μl	microlitre
μM	micromolar
nm	nanomolar
nt	nucleotide
ORF	open reading frame
^{32}P	phosphorus-32 radioisotope
PAA	phosphonoacetic acid
PCR	polymerase chain reaction
PEG	polyethylene glycol
pfu	plaque forming unit
RACE	rapid amplification of cDNA ends
RNA	ribonucleic acid
RT	reverse transcriptase
SMART	switching mechanism at 5' end of RNA transcript
s	second
TR _L	long terminal repeat
TR _S	short terminal repeat
U _L	long unique
UPM	universal primer mix
U _S	short unique
UV	ultraviolet
VP	virion protein
v/v	volume/volume
w/v	weight/volume

CHAPTER 1:

Introduction

Introduction

This chapter is divided into three sections. Sections 1.1 and 1.2 give an overview of the *Herpesviridae* and human herpesviruses, respectively. Section 1.3 provides a brief description of human cytomegalovirus (HCMV), including discovery and classification, general characteristics, clinical relevance, virion structure, genome content, gene expression and gene functions. The chapter concludes with short description of the justification and objectives of the study.

1.1 The *Herpesviridae*: overview

1.1.1 General properties

The general characteristics of the *Herpesviridae* have been reviewed in the recent report of the International Committee on the Taxonomy of Viruses (ICTV) (Minson *et al.*, 2000). The family *Herpesviridae* comprises over 120 large DNA viruses that infect vertebrates, including humans (Minson *et al.*, 2000), and at least one invertebrate, the oyster (Comps and Cochenec, 1993). Herpesviruses are well adapted to their hosts. Most display a high degree of host specificity, and a single host may be infected by several different herpesviruses; for example, eight human herpesviruses have been identified.

1.1.2 Virion structure

Virions of herpesviruses have complex and characteristic structures consisting of both symmetric and nonsymmetric components (Homa and Brown, 1997; Steven and Spear, 1997). The spherical virion of herpes simplex virus type 1 (HSV-1) comprises core, capsid, tegument and envelope (Fig 1.1). The core consists of the viral genome packed in a liquid crystalline array that fills the entire volume of the preformed icosahedral capsid. The mature capsid is 125 nm in diameter and 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).

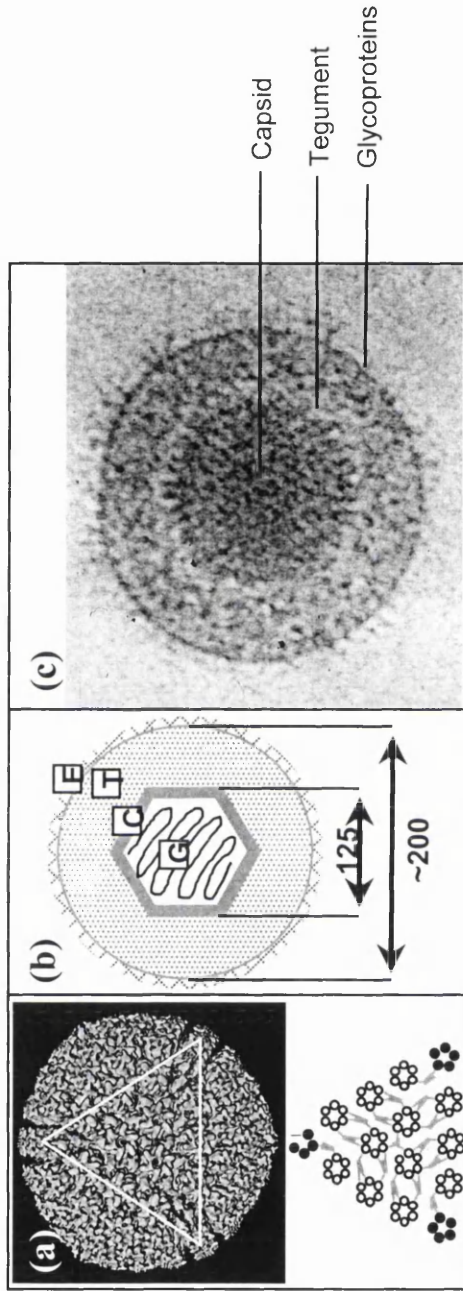


Figure. 1.1 Herpesvirus morphology.

(a) The top panel shows a reconstruction of an HSV-1 capsid generated from cryo-electron micrograph (EM) images. The lower panel shows a cartoon of one of the 20 icosahedral faces. Penton subunits are shown in black, hexon subunits in white and the pattern of triplexes and their connections with adjacent capsomers in grey. (b) A schematic representation of a virion with diameters in nm. G, genome; C, capsid; T, tegument; E, envelope. (c) A cryo-EM image of an HSV-1 virion. (Courtesy of F. Rixon, MRC Virology Unit, Glasgow)

The tegument is an amorphous proteinaceous layer which lies between the nucleocapsid and envelope (Roizman and Furlong, 1974; Zhou *et al.*, 1999). It consists of at least 18 proteins, the functions of many of which have not been fully elucidated. Certain functions related to morphogenesis, uncoating and regulation of gene expression have been assigned to its component proteins.

The virion is enclosed by a host-derived lipid bilayer, the envelope (Roizman and Furlong, 1974; Wildy *et al.*, 1960). The envelope contains protruding glycoprotein spikes (Stannard *et al.*, 1987) which vary in size from 8-24 nm. Thus far, 11 glycoproteins have been identified (reviewed by Spear, 1993; Haarr and Skulstad, 1994): gB, gC, gD, gE, gG, gH, gI, gJ, gK, gL and gM. Virions range in diameter between 160 nm and 230 nm, averaging 180 nm (Szilagyi and Berriman, 1994).

1.1.3 Herpesvirus classification

A formal system of classification for the herpesviruses has been developed by the ICTV (Roizman *et al.*, 1981, 1992, 1995; Minson *et al.*, 2000), in which the family Herpesviridae is divided into three subfamilies, the *Alphaherpesvirinae* (informally known as the α -herpesviruses), the *Betaherpesvirinae* (β -herpesviruses) and *Gammaherpesvirinae* (γ -herpesviruses). The *Alphaherpesvirinae* exhibit rapid cytolytic growth *in vitro* and establish latency in the nervous system. The *Betaherpesvirinae* have a restricted host range *in vitro* and have a long productive cycle, and are usually associated with the formation of enlarged (cytomegalic) cells. The *Gammaherpesvirinae* establish latent infection in lymphocytes and are often associated with lymphoproliferative disease. The subfamilies are further divided into genera. For example, the *Alphaherpesvirinae* subfamily consists of four genera: Simplexvirus (informally α_1), Varicellovirus (α_2), 'Marek's disease-like virus' (α_3) and 'Infectious laryngotracheitis-like virus' (α_4). Table 1.1 gives an outline of the current herpesvirus classification and accession numbers for the sequenced genomes.

Limited sequence data from herpesviruses isolated from turtles (Quackenbush *et al.*, 1998) suggest that these viruses belong to the α -herpesviruses. However, herpesviruses of fish (Davison, 1992, 1998), amphibians (Davison *et al.*, 1999) and one invertebrate (the oyster)

Common name/abbreviation ^a	ICTV designation/abbreviation	GenBank accession no.
ALPHAHERPESVIRINAE		
Simplexvirus (α_1)		
Herpes simplex virus type 1/ (HSV-1)	<i>Human herpesvirus 1/HHV-1</i>	X14112
Herpes simplex virus type 2/ (HSV-2)	<i>Human herpesvirus 2/HHV-2</i>	Z86099
Bovine mammillitis virus/ BHV-2	<i>Bovine herpesvirus 2/ BoHV-2</i>	NA ^b
Varicellovirus (α_2)		
Varicella-zoster virus/ (VZV)	<i>Human herpesvirus 3/HHV-3</i>	X04370
Simian varicella virus/(SVV)	<i>Cercopithecine herpesvirus 9/CeHV9</i>	AF275348
Infectious bovine rhinotracheitis virus/ BHV-1	<i>Bovine herpesvirus 1/ BoHV-1</i>	AJ004801
Pseudorabies virus/PRV	<i>Suid herpesvirus 1/SuHV-1</i>	NA ^b
Feline rhinotracheitis virus/FHV-1	<i>Felid herpesvirus 1/FeHV-1</i>	NA ^b
Equine abortion virus/EHV-1	<i>Equid herpesvirus 1/EHV-1</i>	M86664
Equine rhinopneumonitis virus/EHV-4	<i>Equid herpesvirus 4/EHV-4</i>	AF030027
“Marek’s disease-like virus”(α_3)		
Marek’s disease virus 1/(MDV-1)	<i>Gallid herpesvirus 2/GaHV-2</i>	AF243438
Marek’s disease virus 2/(MDV-2)	<i>Gallid herpesvirus 3/GaHV-3</i>	AB049735
Herpesvirus of turkeys/HVT	<i>Meleagrid herpesvirus 1/MeHV-1</i>	AF291866
“Infectious laryngotracheitis-like virus”(α_4)		
Infectious laryngotracheitis virus/(ILTV)	<i>Gallid herpesvirus 1/GaHV-1</i>	NA ^b
BETAHERPESVIRINAE		
Cytomegalovirus (β_1)		
Human cytomegalovirus/(HCMV)	<i>Human herpesvirus 5/HHV-5</i>	X17403
Chimpanzee cytomegalovirus/CCMV	<i>None/None</i>	Unpublished
Rhesus cytomegalovirus/RHCM	<i>Cercopithecine herpesvirus 8/CeHV-8</i>	NA ^b
Muromegalovirus (β_1)		
Murine cytomegalovirus/(MCMV)	<i>Murid herpesvirus 1/MuHV-1</i>	U68299
Rat cytomegalovirus/RCMV	<i>Murid herpesvirus 2/MuHV-2</i>	AF232689
Roseolovirus (β_2)		
Human herpesvirus 6/(HHV-6)	<i>Human herpesvirus 6/HHV-6</i>	X83413, AB021506, AF157706
Human herpesvirus 7/(HHV-7)	<i>Human herpesvirus 7/HHV-7</i>	U43400, AF037218
Unassigned member		
Tupaia herpesvirus/TuHV	<i>Tupaiaid herpesvirus 1/TuHV-1</i>	AF281817
GAMMAHERPESVIRINAE		
Lymphocryptovirus (γ_1)		
Epstein-Barr virus/(EBV)	<i>Human herpesvirus 4/HHV-4</i>	M35547, V01555, M80517
Rhadinovirus (γ_2)		
Herpesvirus saimiri/(HVS)	<i>Saimirine herpesvirus 2/SaHV-2</i>	X64346
Herpesvirus ateles/HVA	<i>Ateline herpesvirus 2/AtHV-2</i>	AF083424
Kapsi’s sarcoma-associated herpesvirus/KSHV	<i>Human herpesvirus 8/HHV-8</i>	U75698, U93872
Rhesus rhadinovirus/RRV	<i>Cercopithecine herpesvirus 17/CeHV-17</i>	AF083501, AF210726
Mouse herpesvirus 68/MHV68	<i>Murid herpesvirus 4/MuHV-4</i>	U97553, AF105037
Equine herpesvirus 2/EHV-2	<i>Equid herpesvirus 2/EHV-2</i>	U20824
Malignant catarrhal fever virus/ AHV-1	<i>Alcelaphine herpesvirus 1/AiHV-1</i>	AF005370
Bovine herpesvirus 4/BHV-4	<i>Bovine herpesvirus 4/BoHV-4</i>	AF318573
UNDEFINED SUBFAMILY		
“Ictalurid herpes-like virus”		
Channel catfish virus/(CCV)	<i>Ictalurid herpesvirus 1/IcHV-1</i>	M75136

Table 1.1: Herpesvirus classification and accession numbers for sequenced genomes.

^aGeneric names and their alternative synonyms (lowercase bold) are given for each subfamily (uppercase bold). Abbreviations for the names of virus species are given. ^bComplete sequence not available. Adapted from Davison (2001).

(A.J. Davison, unpublished data) are very remotely related to the mammalian/avian group. Fig. 1.2 shows an illustrative phylogenetic tree of the Herpesviridae, indicating evolution of three major groups from a putative ancestor (McGeoch and Davison, 1999a).

Although herpesviruses of mammals and birds are highly divergent, conservation of a subset of about 40 core genes across the three subfamilies suggests a common evolutionary origin. Herpesvirus evolution is thought to have proceeded through mutational and recombinational processes, including large scale genomic rearrangements, gene capture and gene duplications (McGeoch, 1989; McGeoch and Cook, 1994). An in-depth phylogenetic analysis of mammalian and avian herpesviruses based on a set of eight conserved genes has recently been accomplished by McGeoch *et al.* (2000), confirming the overall high level of congruence between the virus hosts' lineages and the herpesvirus phylogenetic tree. This supports the inference made previously that herpesviruses have evolved with their hosts. Based on this hypothesis and using the divergence times of the host lineages (Benton, 1990; Kumar and Hedges, 1998), it is estimated that the α -, β - and γ -herpesviruses arose approximately 180-220 million years ago and that the genera diverged 60-80 million years ago.

1.1.4 Genetic content

Herpesviruses have large, linear, double-stranded DNA genomes which range in size from 124 kbp (e.g. simian varicella virus (SVV)) to 240 kbp (e.g. HCMV) (McGeoch and Davison, 1999a). The viruses also exhibit an impressively wide range of base compositions ranging from 32-75% in G+C content (Honest, 1984; Minson *et al.*, 2000). A characteristic of herpesvirus genome structure is the presence of terminal or internal repeated sequences, in direct or inverse orientation. The layouts of these repeat elements in relation to unique sequences define at least six genome types (Fig 1.3).

Herpesviruses reportedly contain between about 70 and 200 genes. For example, the HSV-1 genome contains 74 unique protein-coding genes, and the HCMV genome contains approximately 165. The genes are densely packed and arranged in about equal numbers on each DNA strand. Most are expressed as single exons from their own promoters, although

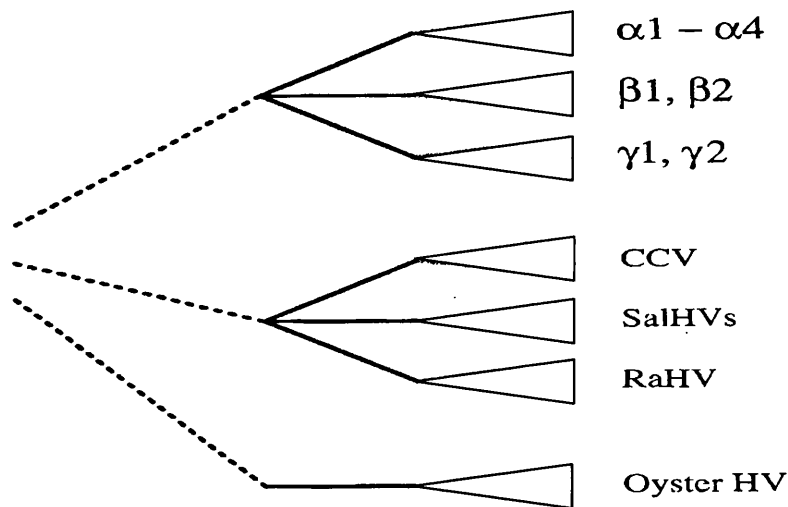


Figure 1.2: Phylogeny of herpesviruses.

A schematic view is given of the organisation of the three major lineages of herpesviruses and their possible relationships. Subfamily groupings of the mammalian/avian viruses, the counterparts of these for fish and amphibian viruses, and the invertebrate (oyster) virus, are shown as grey triangles. Evolution of such groupings from common ancestors is indicated by solid lines and a hypothetical overall common origin by dashed, incompletely joined lines. SalHV; salmonid herpesvirus, RaHV; ranid (frog) herpesvirus. Adapted from McGeoch and Davison (1999).

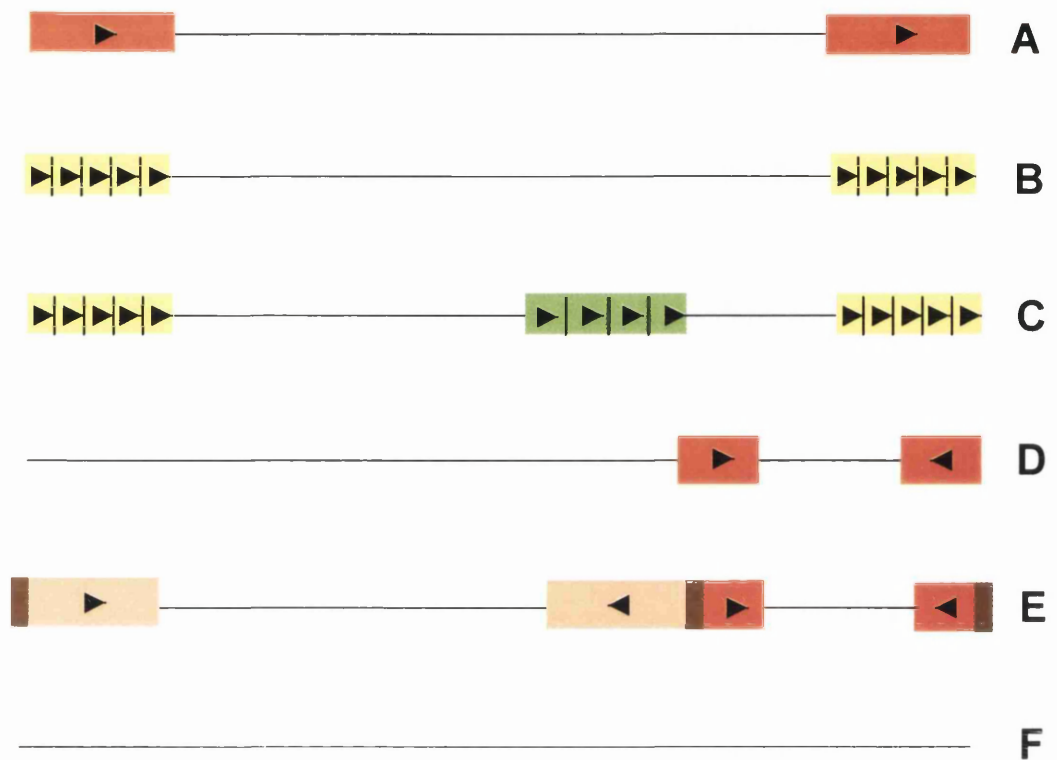


Figure 1.3: Types of herpesvirus genome structure.

Unique sequences and repeat elements are represented by black lines and coloured rectangles, respectively. The number of terminal repeats in types B and C and the number of internal repeats in type C are variable. Arrows indicate locations and orientations of repeated regions. The genomes are not to scale.

A. A long unique sequence flanked by large direct terminal repeats (CCV, EHV-2, HHV-6 and HHV-7) (Chousterman *et al.*, 1979; Davison, 1992; Browning and Studdert, 1989; Telford *et al.*, 1995; Martin *et al.*, 1991; Gompels *et al.*, 1995; Nicholas, 1996).

B. A long unique sequence flanked by multiple smaller tandem direct repeats at the genome termini (HVS and HHV-8) (Bornkamm *et al.*, 1976; Albrecht *et al.*, 1992; Russo *et al.*, 1996).

C. Multiple smaller direct tandem repeats at the genome termini plus a different sequence repeated a variable number of times at an internal location (EBV) (Given and Kieff, 1979; Baer *et al.*, 1984).

D. A short unique sequence flanked by large inverted repeats, and a long unique sequence flanked by very small or no terminal repeats (EHV-1 and VZV) (Whalley *et al.*, 1981; Telford *et al.*, 1991; Dumas *et al.*, 1981; Davison and Scott, 1986).

E. Short and long unique sequences each flanked by a pair of large inverted repeats. A terminal direct repeat, the *a* sequence, is found at the genomic termini and in inverted orientation at the junction between the internal repeats (IR_L and IR_S) (HSV-1 and HCMV) (McGeoch *et al.*, 1988; Weststrate *et al.*, 1983; Chee *et al.*, 1990).

F. A single unique region with very small or no terminal repeats (TuHV-1) (Roizman *et al.*, 1995).

Adapted from McGeoch and Davison (1999).

families of genes arranged tandemly on the same strand commonly share a single polyadenylation site (Wagner, 1985). Overlaps between genes in different reading frames on the same strand or on opposing strands are infrequent and usually not extensive. Few genes are spliced in the α -herpesviruses, but splicing is more common in the β - and γ -herpesviruses (Davison and Clements, 1996).

The genes that have counterparts in α -, β - and γ -herpesviruses (core genes; McGeoch and Davison, 1999b) are arranged in seven gene blocks (Fig. 1.4). The great majority of core genes are involved in vital aspects of the viral life cycle, such as entry into the cell, viral DNA replication, packaging of the genome and virion assembly. Nine genes (of which all but one are involved in DNA metabolism and replication) are clearly related by similarities in encoded amino acid sequences to cellular genes (McGeoch and Davison, 1999b) and are probably originated from the cellular genome. Viral structural genes, on the other hand, do not appear to have any obvious cellular homologues. Non-core genes are involved in control of the life cycle, latency, pathogenicity and modification of the immune response.

1.1.5 Life cycle

The life cycle of herpesviruses replication consists of lytic and latent phases. It can best be illustrated by briefly considering the life cycle of the best-studied herpesvirus, HSV-1 (Fig. 1.5); its general aspects apply to other herpesviruses.

1.1.5.1 The lytic cycle

Initial association of HSV-1 with the host cell is mediated by viral envelope glycoproteins, which have important roles in adsorption and penetration of virus into the cell (Campadelli-Fiume, 1994). After the virus envelope fuses with the plasma membrane, the nucleocapsid is released into the cytoplasm and migrate to the nucleus where the DNA enters via nuclear pores. At least some of the tegument proteins also enter the nucleus, although the process by which this occurs is not clear. Transcription and replication of viral DNA and capsid assembly take place in the nucleus.

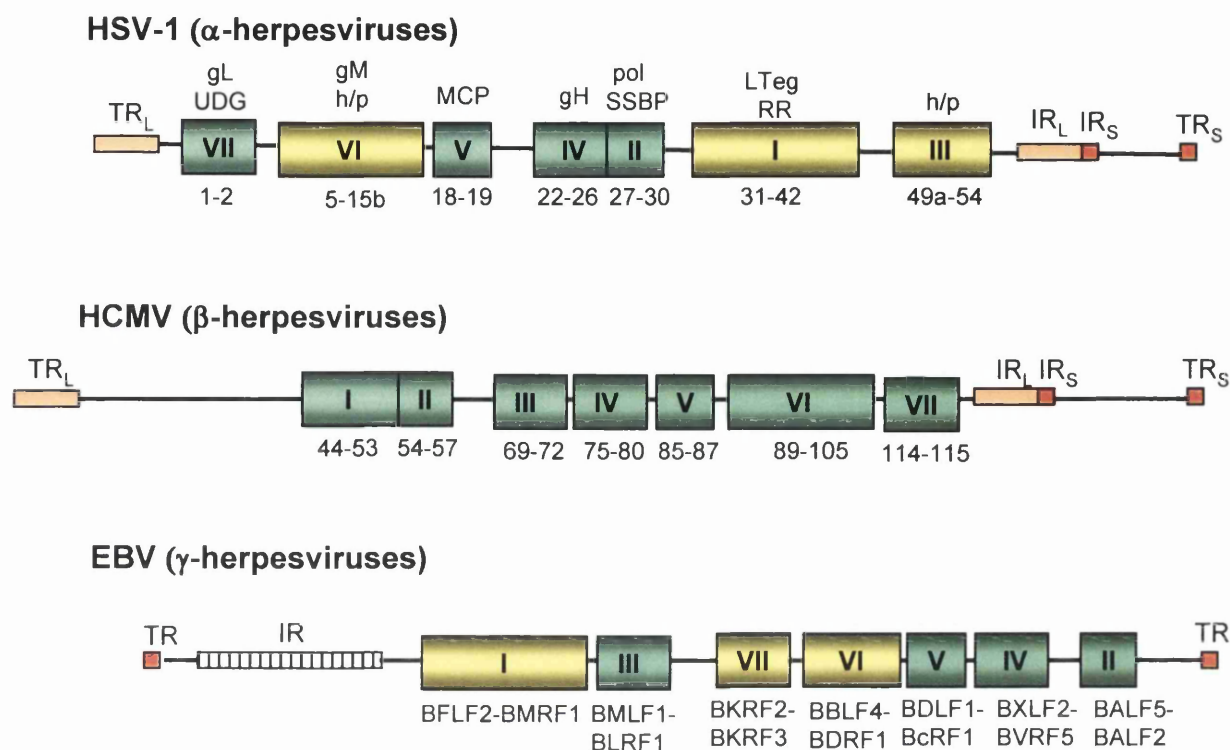


Figure 1.4: Organisation of conserved gene blocks in members of different herpesvirus subfamilies.

Prototype genome arrangements (not to scale) are shown of HSV-1, HCMV and EBV. Seven conserved gene blocks are shown. Yellow blocks in HSV-1 and EBV genomes denote the inverted orientation of the blocks relative to HCMV. The numbers below the sequence blocks for HCMV and HSV-1 denote the UL ORFs that bracket each conserved block in these viruses. The names listed below the blocks for EBV denote the ORFs bracketing each conserved block in that virus. Examples of genes in each block are indicated for HSV-1. UDG; uracil-DNA glycosylase, gL, gM, gH; glycoproteins L, M, H, h/p; subunit of the helicase-primase complex; pol; catalytic subunit of DNA polymerase, SSBP; single-stranded DNA binding protein, RR; ribonucleotide reductase, MCP; major capsid protein, LTeg; large tegument protein. Adapted from Minson *et al.* (2000).

Gene expression occurs from the viral genome in a co-ordinated, regulated and sequentially ordered cascade, with three main phases: immediate early (IE or α), early (E or β) and late (L or γ) (Hones and Roizman, 1975; Roizman and Sears, 1990). Some IE proteins are trans-acting regulators of virus genes, and initiate cascade expression. E genes encode several enzymes involved in DNA replication and nucleotide metabolism, as well as a subset of glycoproteins and some uncharacterised proteins. L genes encode many virion structural proteins (Hones and Roizman, 1975). The tegument protein (VP16) encoded by the UL48 gene transactivates IE promoters upon infection (O'Hare, 1993; Arnosti *et al.*, 1993).

Replication starts in the nucleus, with circularisation of viral genomic DNA by direct ligation of the termini. Viral DNA synthesis is initiated from the viral origins of replication (oriS and oriL) to produce DNA in an endless conformation, probably as head-to-tail concatemers, by a rolling circle mechanism (Roizman, 1979; Jacob *et al.*, 1979). Replicated DNA is cleaved specifically into unit-length molecules and packaged into preformed capsids.

The processes by which the capsid acquires tegument and envelope are not fully understood. The envelope is derived from altered host membranes and contains viral glycoproteins, which are processed into their mature forms in the Golgi apparatus. New viral progeny are released from the cell by exocytosis (Rixon, 1993).

1.1.5.2 The latent cycle

Latent infections are produced by all of the human herpesviruses, but are established at sites that are specific to the virus. For example, HSV-1 and VZV become latent in sensory ganglia, whereas EBV latently infects circulating B lymphocytes (Rock, 1993; Kieff and Liebowitz, 1990; Bastian *et al.*, 1972). The monocyte is the probable site of HCMV latency (Kondo *et al.*, 1994).

In its latent form, HSV-1 DNA is present either as a circular molecule or as a concatemer (Rock and Fraser, 1983, 1985; Efstathiou *et al.*, 1986) and is not thought to integrate into

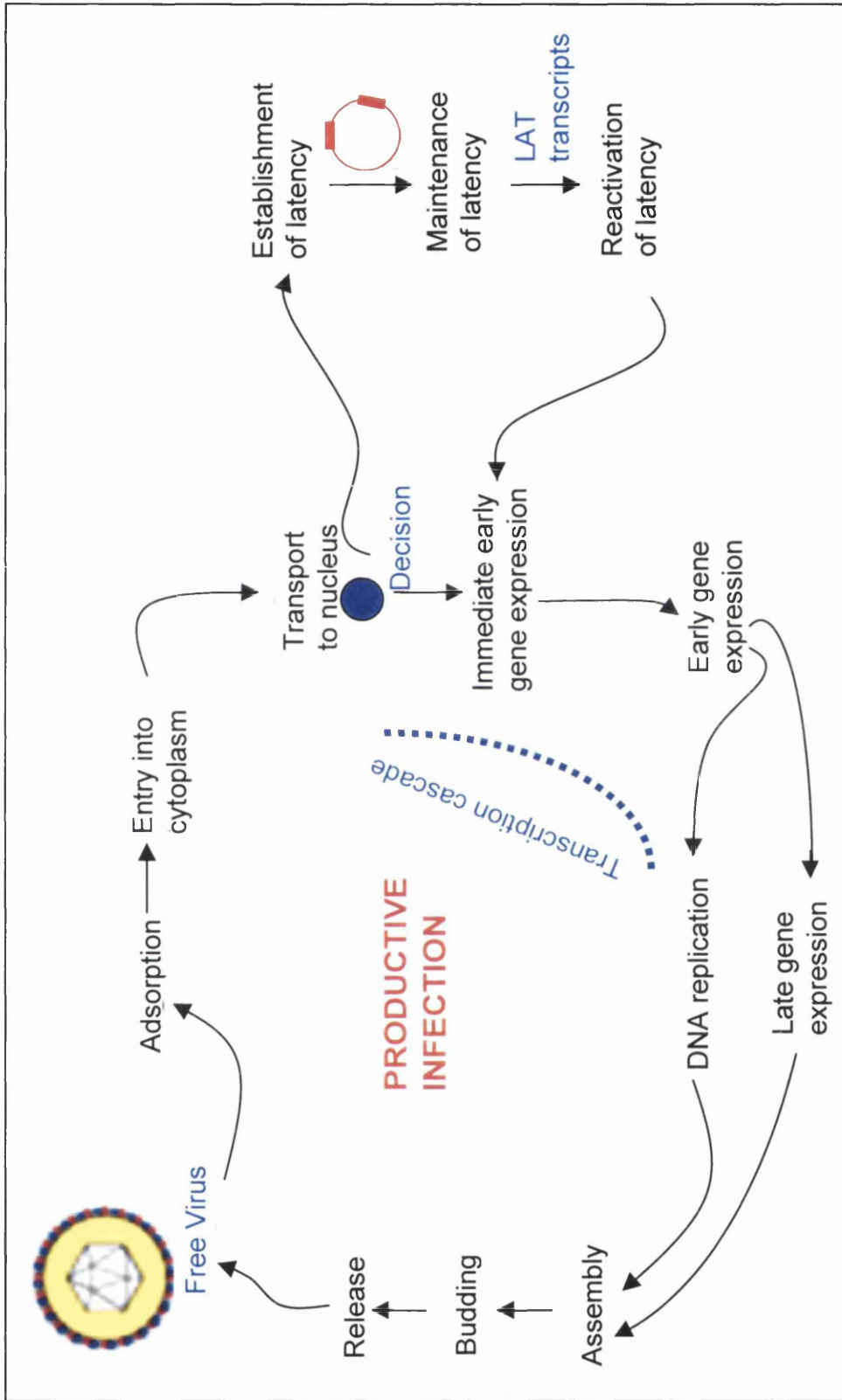


Figure 1.5: The replication cycle of HSV-1.

(Adapted from Davison and Clements, 1996).

the host DNA (Mellerick and Fraser, 1987). EBV latent DNA is maintained as covalently closed circular episomes (Lindhal *et al.*, 1976), and EBV infection can result in cell transformation and proliferation that may lead to carcinomas or lymphomas (Zur Hausen and Schultz-Hausen, 1970).

The products of viral genes expressed during the latent state appear to function in maintenance of and reactivation from latency (Steiner *et al.*, 1989; Dambaugh and Hennessy, 1986; Speck and Strominger, 1989), but differ between HSV-1 (latency-associated transcripts or LATs: Spivack and Fraser, 1987; Deatly *et al.*, 1988) and EBV (Epstein-Barr virus nuclear antigen or EBNA proteins: Lindahl *et al.*, 1974).

1.2 Human herpesviruses

Eight herpesviruses (Table 1.2) of humans have been isolated to date, including members of all three subfamilies (Table 1.1). Most are widespread in both the developed and the developing world (Whitely and Schlitt, 1991), and humans remain the sole known reservoir for transmission, which occurs via person to person contact. These viruses cause a wide variety of diseases and possess varied molecular properties. The infections range from inapparent to severe, with disabling or fatal infections in the foetus, the very young or the immunosuppressed. Symptoms, which are summarised in Table 1.3, cover a broad spectrum of clinical manifestations, from minor lesions to severe and life-threatening encephalopathies. Most of these viruses occur at high prevalence levels in human populations with few exceptions, such as, HHV-8 occurs at high prevalence only in some populations, such as those in central and southern Africa (>50%). The viruses establish a life long infection in their host, utilizing varied mechanisms of latency; HSV-1, HSV-2 and VZV are neurotropic, HHV-6, HHV-7 and EBV are lymphotropic, while HCMV establishes latency in the monocytic lineage. The situation with HCMV will be discussed in detail below. EBV and HHV-8 are implicated in human cancers.

Virus	ICTV designation	Subfamily	Genus	Genome size	Genome type	Review
HSV-1	HHV-1	α -herpesvirus	α_1	152	E	Subak-Sharpe and Dargan (1998)
HSV-2	HHV-2	α -herpesvirus	α_1	155	E	Levine (1992)
VZV	HHV-3	α -herpesvirus	α_2	124	D	Gelb (1990)
EBV	HHV-4	γ -herpesvirus	γ_1	184	C	Miller (1990)
HCMV	HHV-5	β -herpesvirus	β_1	230	E	Britt (1996)
HHV-6	HHV-6	β -herpesvirus	β_2	159	A	Levy (1997)
HHV-7	HHV-7	β -herpesvirus	β_2	153	A	Levy (1997)
KSHV	HHV-8	γ -herpesvirus	γ_2	141	B	Levy (1997)

Table 1.2: The eight herpesviruses known to infect humans and their corresponding designations by the ICTV.

Virus	Associated illness
HSV-1	80-95% of oral lesions, or "cold sores," and 30-50% of genital lesions are caused by HSV-1. Primary infection and reactivation give similar symptoms. In rare cases it causes conjunctivitis, herpetic whitlow, keratitis and encephalitis. Neonatal infections are often life threatening.
HSV-2	5-20% of oral lesions and 50-70% of genital lesions are caused by HSV-2. Primary infection and reactivation give similar symptoms. In rare cases it causes conjunctivitis, herpetic whitlow, keratitis and encephalitis. Neonatal infections are often life threatening.
VZV	Primary infection: chicken pox, a rash which appears 14-15 days post-infection, accompanied by a fever. Reactivated infection, "shingles", appears at the relevant dermatome often accompanied by pain and is sometimes followed by post-herpetic neuralgia.
HCMV	Primary infection: enlargement and fusion of macrophages often occurs. Usually asymptomatic but can be fatal. Reactivation can occur. Infection is problematic in immunocompromised individuals. Symptoms can include gastro-enteritis and retinitis.
HHV-6	Febrile illness and infant rash exanthem subitum. May act as an opportunistic agent in immunosuppressed patients suffering from AIDS or undergoing organ or bone marrow transplants. Symptoms can include pneumonia and encephalitis.
HHV-7	Febrile illness. May act as an opportunistic agent in immunosuppressed patients.
EBV	Primary infection is often asymptomatic in young children, but occurs as infectious mononucleosis in older children and adults. Associated with Burkitt's lymphoma and nasopharyngeal carcinoma.
KSHV	Associated with Kaposi's sarcoma, a vasculated skin lesion. Associated with Castleman's disease.

Table 1.3: Summary of the clinical manifestations of the eight human herpesviruses.

1.3.1 General characteristics

The isolation of HCMV was independently accomplished by Smith (1956), Rowe *et al.* (1956) and Weller *et al.* (1957). The phrase “cytomegalovirus” was coined by Weller (1971), who observed salivary gland tropism and cytomegalic inclusions in the nuclei of infected cells. Human cytomegalovirus (HCMV) is a member of the *Betaherpesvirinae* subfamily of the Herpesviridae (Roizman, 1996), and is formally designated human herpesvirus 5 (HHV-5) (see Table 1.1).

HCMV shares a number of features with other herpesviruses including a similar virion structure, a shared gene subset and the ability to establish a latent infection within the host. HCMV also has some distinguishing features such as salivary gland tropism (Roizman *et al.*, 1981), slow growth in cell culture and a highly species-specific host range.

DNA sequence-based analysis indicates that a large number of different HCMV strains are in circulation worldwide. The AD169 strain was isolated almost half a century ago (Rowe *et al.*, 1956) and became a widely used strain because of its efficient replication in cultured cells. However, little care was taken to ensure that a single, uniform stock was distributed to laboratories. At least two forms of AD169 are currently available, AD169*var*UK and AD169*var*ATCC. The Towne strain (passage 125), a vaccine candidate that has also become a common laboratory strain (Plotkin *et al.*, 1975), induces seroconversion without producing disease. However, other forms of Towne at different passage numbers are also available. In contrast a low passage HCMV strain, Toledo, produced clinically apparent disease when administered to healthy volunteers (Quinnan *et al.*, 1984). HCMV strains that are widely used in laboratory research are listed in Table 1.4. In each case, passage was carried out on human fibroblasts.

HCMV strain	ATCC ^a number	Isolation	Comments
AD169	VR-538	Tonsillectomy-adenoidectomy, 7-year-old girl.	Attenuated high passage strain.
Towne	VR-977	Urine from patient with congenital cytomegalic inclusion.	Attenuated high passage strain.
Toledo		Not known.	Low passage strain.
Davis	VR-807	Liver biopsy from 3 month old girl with microencephaly and other symptoms resembling congenital toxoplasmosis.	High passage strain.

Table 1.4: HCMV strains in research use.

^aAmerican Type Culture Collection.

1.3.2 Clinical relevance

The importance of HCMV as a pathogen has increased over the past two decades as immunosuppressive post-transplant therapies, as well as acquired immunodeficiency syndrome (AIDS) and other immunodeficiency states, have become prominent medical concerns. These conditions predispose individuals to primary HCMV infection or reactivation of latent infection resulting in fulminant, life-threatening, acute disease.

1.3.2.1 Epidemiology

HCMV is ubiquitous in the human population and the prevalence of antibodies in normal adults ranges from 40 to 100% (Krech, 1973). Transmission of HCMV occurs by person to person contact and infection can occur at any age. Transmission during birth and from breast milk is very common in developing countries. Infants and young children excrete HCMV in their urine and respiratory tract and may be the source of infection for adults and other children (Stango and Cloud, 1994). The most important mode of transmission post puberty is probably sexual (Handsfield *et al.*, 1985) especially among sexually active male homosexuals (Drew *et al.*, 1981; Collier *et al.*, 1987), and reinfection is common in

sexually active populations (Drew *et al.*, 1984; Spector *et al.*, 1984). HCMV can also be transmitted via blood products and transplanted organs (Van der Meer *et al.*, 1996).

Primary HCMV infection is endemic and is present throughout the year rather than being seasonal (Gold *et al.*, 1990). Ill-defined socioeconomic factors predispose certain populations to higher infection rates, both by vertical and horizontal transmission.

HCMV is the most common congenital viral infection in humans, owing to the high prevalence of the virus in the general population. 60% of the UK population has experienced infection by the age of 40. In North America, HCMV infects about 50% of the population outside urban centres and up to 90% of the population within cities (Mocarski, 1996). HCMV is the major cause of viral congenital infection and an estimated 40000 infected infants are born each year in the United States, with 4000 to 6000 exhibiting long term neurological sequelae (Britt and Alford, 1996). Acute infection and reactivation of latent infection presents a life-threatening risk to allograft recipients, HIV-infected individuals and immunosuppressed persons. Prior to recent advances in diagnosis and treatment, HCMV was the major infectious cause of death in bone marrow allograft recipients, with mortality rates from untreated HCMV pneumonia of approximately 90% (Meyers *et al.*, 1982)

1.3.2.2 Disease and therapy

Primary infection with HCMV in normal adults is often asymptomatic, but may be manifested by a self-limiting mononucleosis-like syndrome in less than 1:1000 cases (Horwitz *et al.*, 1986). In contrast, infections in immunocompromised hosts range from similar asymptomatic infection to fulminant life-threatening disease. HCMV is associated with different diseases depending on the type and level of immunosuppression.

In AIDS patients, HCMV infection is commonly manifested as retinitis and gastrointestinal dysfunction (Jacobson and Mills, 1988; Jabs *et al.*, 1989; Dietrich and Rehmin, 1991; Gallant *et al.*, 1992; Spector *et al.*, 1993; Boppana *et al.*, 1992). The risk of developing HCMV disease and death in persons with advanced AIDS is directly related to the quantity

of HCMV DNA in plasma (Spector *et al.*, 1998). In allograft transplantations 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; Appereley and Goldman, 1988; Winston *et al.*, 1990). The risk for, and the manifestations of, HCMV disease are associated with the organ of transplant. For example, pneumonia is a major problem in bone marrow or lung transplant recipients, whereas hepatitis is quite common after liver transplantation.

The risk of foetal infection with HCMV correlates with the lack of pre-existing seroimmunity in women infected with the virus during pregnancy. Infants with congenital infection exhibit many of the findings of both AIDS patients and allograft recipients infected with HCMV (Becroft, 1981; Boppana *et al.*, 1992). Post-natally, HCMV infections are self-limiting, but frequent central nervous system (CNS) sequelae result from damage secondary to foetal infection (Stango *et al.*, 1983; Boppana *et al.*, 1992; Williamson *et al.*, 1992).

Resources generated to control HCMV infection include HCMV vaccines and anti-HCMV drugs (Harper and Kinchington, 1998). The live HCMV vaccine developed for the Towne laboratory strain, however, appears to induce only limited protection (reviewed in Britt and Alford, 1996). Attention has now focused on subunit vaccines, particularly those directed against the viral glycoprotein gB, which is the major antigen elucidating viral neutralising antibodies (Speckner *et al.*, 1999). The anti-HCMV drug ganciclovir, a nucleoside analogue and a derivative of acyclovir, is licensed to treat HCMV-induced retinitis in immunocompromised patients but has been associated with myelotoxicity (Flint *et al.*, 2000). Foscarnet, a pyrophosphate analogue, has also been used to treat HCMV infections but is nephrotoxic and is used only in life-threatening cases or when other anti-viral drugs are no longer effective (Chan *et al.*, 2001). Clearly, there is a continued need to develop more effective and less toxic anti-HCMV therapies. The molecular genetic investigation of HCMV, in so much as it identifies new potential antiviral targets, will play an important role in facilitating new intervention strategies for HCMV. One characteristic of herpesviruses that makes anti-herpesvirus intervention strategies difficult to develop is their ability to establish life-long persistent infections. HCMV latency is associated with bone marrow CD34+ myeloid progenitor cells. However it has not been conclusively

established whether HCMV establishes a true non-replicating latent infection in a specific cell type or persists by low-level replication (Mocarski, 1996).

Murine cytomegalovirus (MCMV) has been extensively used as a surrogate animal model to represent HCMV infection with respect to viral transmission, pathogenesis, acute infection and/or reactivation after immunosuppression, transfusion or transplantation (Xiao *et al.*, 2000). However, as MCMV does not cross the placenta, its utility as a model for congenital HCMV infection is questionable. More recently, however, guinea pig cytomegalovirus (GPCMV), which does cross the placenta and shares many biological similarities with HCMV, has been proposed as a better model for HCMV foetal infection (McGregor and Schliess, 2001).

1.3.3 Cell tropism

HCMV can infect almost all organ systems, but can normally be propagated *in vitro* only in human fibroblast cells such as HFL1 (human foetal lung) and HFFF2 (human foetal foreskin fibroblast). HCMV has also been reported to infect a wide range of other cell types *in vitro*, although the infection tends to be abortive (reviewed by Sinzger and Jahn, 1996). Various strains also vary in competencies in tropism for endothelial cells (Brown *et al.*, 1995). Towne or AD169 have lost certain host range capabilities of low passaged strain Toledo.

It has been shown that the growth restrictions in cell culture may be partly due to passaging laboratory stocks in fibroblasts (Waldman *et al.*, 1989; Brown *et al.*, 1995; Sinzger *et al.*, 2000). Passaging HCMV in cell culture may be associated with gross changes, especially large-scale deletions, within the virus genome (Cha *et al.*, 1996), and loss of certain genes may, as seen for MCMV (Cavanaugh *et al.*, 1996), result in altered tropism.

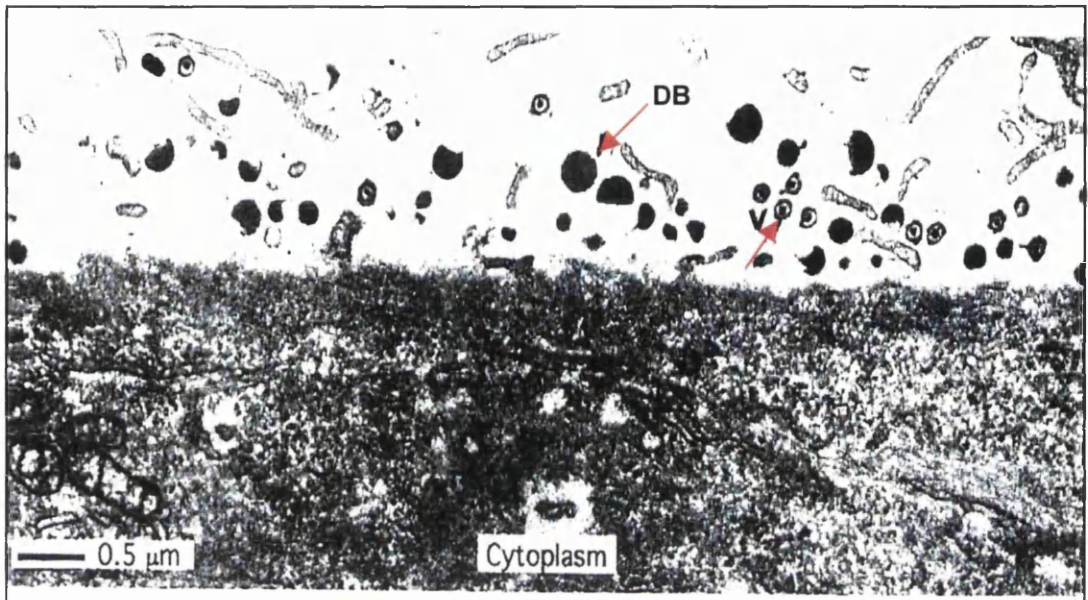


Figure 1.6: HCMV particles.

Surface of a cell onto which virions (V) and dense bodies (DB) of HCMV (Towne strain) have attached. This illustrates the sizes and approximate ratio of these two types of particles in a virus preparation. Adapted from Mocarski and Stinski (1979).

1.3.4 HCMV particles

Three types of virus-related particles are released into the extracellular medium from HCMV-infected cells in culture. These are the mature infectious virion and two non-infectious particles: non-infectious enveloped particles (NIEPs) and dense bodies (Fig.1.6) (Mocarski and Stinski, 1979; Irmiere and Gibson, 1983). Dense bodies and NIEPs can be distinguished from virions and each other by morphology and by biochemical analysis. Dense bodies, which are present only in the cytoplasm of infected cells, lack nucleocapsids and DNA and are predominantly composed of a single tegument protein, pp65 (pUL83). Whether dense bodies are produced in natural HCMV infections or whether they represent an artifact of *in vitro* culture systems is controversial. In support of the latter view it has been noted that dense bodies are not produced in tissue culture cells infected with MCMV. It has been suggested that dense bodies may consist of surplus cellular and viral products accumulated in a storage vacuole that is subsequently voided into the extra-cellular medium (Severi *et al.*, 1992). NIEPs are morphologically indistinguishable from virions but lack the virus genome and retain one of the scaffolding proteins (pUL80.5). As has been reported for HSV-1 (Roizman, 1996), several forms of HCMV capsid structures (A, B and C capsids) can be identified in infected cells. Type A capsids lack the DNA and accumulate because of a failure to stably package the viral genome, whereas, B capsids lack DNA but contain the viral scaffolding/assembly protein, and are found in the nucleus as a precursor of mature capsids. Type C capsids are fully mature nucleocapsids and are present in the cytoplasm of infected cells (Gibson, 1996).

HCMV has a virion structure typical of herpesviruses, consisting of a double-stranded DNA genome packaged inside the icosahedral capsid, which is embedded in the tegument, and the whole then enclosed within a lipid envelope decorated with surface viral glycoprotein spikes (Roizman, 1996). To date, approximately 30 HCMV genes are reported to encode products that are present in the virion particle.

1.3.4.1 The capsid

The three dimensional structure of the HCMV strain AD169 capsid has been determined by electron cryomicroscopy and image reconstruction (Fig. 1.7). Comparisons indicate that

certain features of the HCMV capsid differ from those of HSV-1: the HCMV capsid shell is larger (diameter of 130 nm) and has to accommodate a genome that is some 37% larger. Differences between HCMV and HSV-1 in capsid shell architecture have been reported, including hexamer spacing, relative tilt, morphology of the tips of the hexons, and the average diameter of the scaffold (Butcher *et al.*, 1998). The internal volume of the HCMV capsid is 17% larger than that of HSV-1, but this is insufficient by itself to account for packaging of the larger HCMV genome. It has been shown that the HCMV genome is more densely packed than that of HSV-1 (Bhella *et al.*, 2000).

The HCMV capsid shell contains at least four proteins: the major capsid protein (MCP) (pUL86), the minor capsid protein (mCP) (pUL85), the mCP binding protein (mC-BP) (pUL46) and the small capsid protein (SCP) (pUL48A) (Gibson, 1996). The HCMV MCP is the largest of the four capsid proteins. It constitutes approximately 90% of the capsid protein mass and comprises the pentamers and hexamers. The mC-BP and mCP are present in a 1:2 ratio and constitute the inter-capsomeric triplex that links adjacent capsomeres. The HSV-1 SCP is not essential for the production of a stable capsid shell structure, and its location at the tips of the hexon is thought to direct tegument attachment (Zhou *et al.*, 1995), but the HCMV SCP has been shown to be essential for generation of infectious progeny in tissue culture cells (Borst *et al.*, 2001). The role of capsid shell proteins in interaction with tegument proteins is relatively little studied. However, it has been demonstrated that HCMV tegument proteins are attached to the pentons, hexons and triplexes (Chen *et al.*, 1999; Trus *et al.*, 1999). HSV-1 tegument proteins have also been shown to be attached to the pentons (Thomsen *et al.*, 1994).

The HCMV 'assemblin' or 'assembly protein' (pUL80.5), is a 36 kDa scaffolding protein and, corresponding to the C-terminal half of the 85 kDa UL80 gene product. Like its HSV-1 counterpart pUL26.5, pUL80.5 forms a scaffold by self-interaction, via N-terminal sequences, and also interacts with the MCP via C-terminal sequences (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.

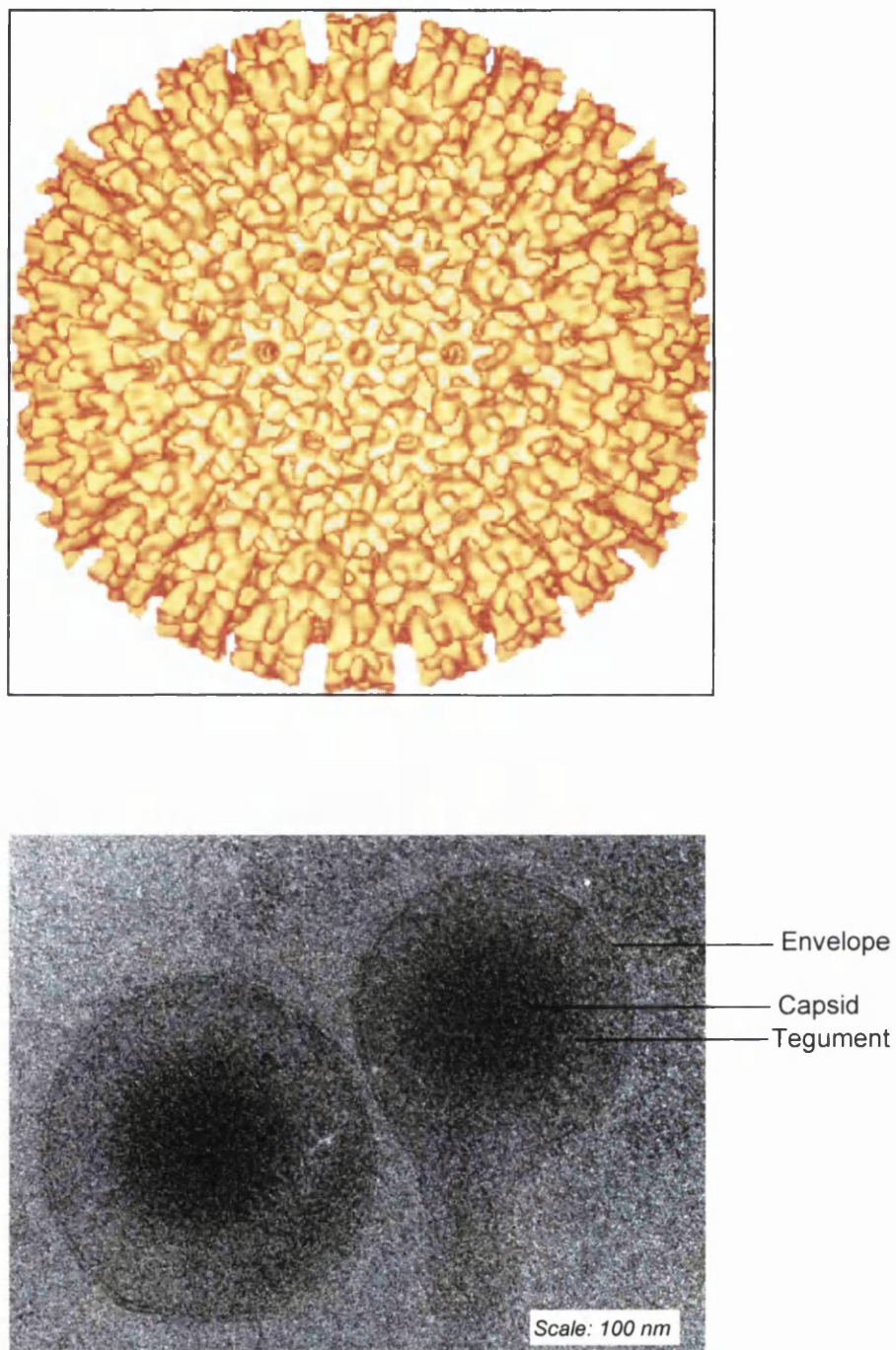


Figure 1.7: HCMV morphology.

A. A reconstruction of an HCMV capsid generated from cryo-electron micrograph (EM) images. B. A cryo-EM image of HCMV virions. (Courtesy of D. Bhella, MRC Virology Unit, Glasgow.)

1.3.4.2 The tegument

The tegument of HCMV particles lies between the capsid and the envelope and contains approximately 40% of the virion protein mass (Gibson, 1996). At least 25 proteins are located in the tegument layer. The products of 11 ORFs (UL25, UL26, UL32, UL47, UL48, UL48A, UL82, UL83, UL85, UL88 and UL99) appear to be both phosphorylated and highly immunogenic (Gibson, 1983; Gibson, 1996). Certain HCMV tegument proteins have counterparts in other betaherpesviruses and in the alphaherpesviruses, indicating important conserved functions. Many genes, for example, UL32, UL36, UL82, UL88, appear to have no counterparts in HSV-1, but have homologues in the other betaherpesviruses, and this indicates an important function for the gene products in this subfamily.

Two tegument proteins, pp65 (pUL83 or 'lower matrix protein') and pp150 (pUL32 or 'basic phosphoprotein'), are the major HCMV tegument components and are highly immunogenic (Gibson, 1996; Mocarski, 1996). pp65 is an abundant virion protein (approximately 18% of total virion protein mass) and is the main protein constituent of dense bodies (approximately 95% of particle mass) (Irmiere and Gibson, 1983). There is evidence, however, that the amount of this protein is reduced in virions of low passage, clinical isolates (Klages *et al.*, 1989). pp65 appears to be dispensable for growth in cell culture (Schmolke *et al.*, 1995). pp65 localises to the nucleus immediately after infection with virus or dense bodies, prior to IE protein synthesis (i.e. in cells treated with cycloheximide (Hensel *et al.*, 1995)), and is therefore believed to have a role in initiating lytic infection. pp65 is strongly associated with the nuclear matrix of infected cells, and accumulating along the nuclear periphery and possibly associating with nuclear lamins (Sanchez *et al.*, 1998).

pp150 is an essential gene product, constituting approximately 20% of total virion mass. It is one of the most immunogenic proteins of virion (Gibson and Irmiere, 1984; Jahn *et al.*, 1987; Pereira *et al.*, 1993). HCMV seropositive individuals have high titres of anti-pp150 antibodies at up to 48 h post-infection. Hensel *et al.* (1995) reported that pp150 is present in the nuclei of infected cells, where it associates with viral assembly compartments or with the nuclear membrane, but at 72 h post-infection it is also cytoplasmic. However, more

recent investigation has failed to detect pp150 in the nucleus but shown it to be present in the cytoplasm throughout infection, accumulating in a stable juxtannuclear structure (Sanchez *et al.*, 2000). The function of pp150 is unknown, but its simian CMV (SCMV) homologue (Trus *et al.*, 1999) is tightly bound to the capsid by an interaction involving a conserved sequence near the N terminus of the pp150 protein and the MCP (Baxter and Gibson, 2001). It is possible that pp150 provides an anchor for attachment of additional tegument proteins that are built subsequently around the particle.

The roles performed by protein constituents of the tegument layer may be multifunctional and complex. Certain tegument proteins are thought to perform key architectural roles or function to facilitate virus egress. Some, delivered to cells at the time of infection, may operate prior to IE protein synthesis to induce a favourable intracellular environment for lytic replication in the infected cell. These tegument proteins may be determinants of tissue tropism or be involved in abrogating host defences. Certain HCMV tegument proteins (e.g. pp71 (pUL82), pUL69 and pTRS1/pIRS1), can trans-activate the major IE promoter and/or other viral or cellular promoters.

The tegument of HCMV particles contains an antibody Fc receptor that can bind all subtypes of human IgG, and is likely to be involved in preventing antibody-mediated clearance of HCMV (Stannard and Hardie, 1991; Antonsson and Johansson, 2001). Recently, HCMV TRL11/IRL11 has been shown to encode a glycoprotein which is thought to encode the Fc receptor (Lilley *et al.*, 2001).

Unexpectedly, the tegument has also been reported to contain a specific subset of viral transcripts, termed 'virion RNAs', originating from ORFs UL21.5 (termed UL22A in this thesis), UL109, IRL4 and IRL7 (Bresnahan and Shenk, 2000; Greijer *et al.*, 2000). These genes are of unknown function, although UL22A is known to encode a protein that localizes to the Golgi network. TRL4/IRL4 is a major early transcript, but no protein product has been identified, and it is therefore possible that it operates as a functional RNA molecule. It is assumed that delivery of virion RNAs to the host cell allows for their expression immediately after virus entry in the absence of new transcription. Several host cell proteins, including β_2 -microglobulin, annexin II, an actin-like protein, and enzymes, such as capsid-associated protein phosphatase PP2A and DNA polymerase, have also been

reported to be associated with particles (Baldick and Shenk, 1996; Gallina *et al.*, 1999; Michelson *et al.*, 1996).

1.3.4.3 The envelope

The HCMV envelope is composed of viral proteins embedded in host-derived cytoplasmic membranes (Mocarski, 1996). HCMV encodes an abundance of putative glycoprotein (gp) genes (approximately 60), many of which appear to lack homologues in other herpesviruses (Chee *et al.*, 1990). Most of these glycoproteins have not been studied, and it is not yet clear which are expressed on the surface of the infected cell or in the virion envelope. Early studies revealed approximately eight major HCMV glycoproteins (Stinski, 1976; Gibson, 1983) with homologues in other herpesviruses which are known to be involved in forming glycoprotein complexes (gC1, gC2 and gC3) in the virion envelope (Gibson, 1983; Pahl *et al.*, 1989; Britt and Mach, 1996).

HCMV gB (gpUL55) is the major envelope glycoprotein. Homodimeric molecules of gB are linked by disulphide bonds to form glycoprotein complex gC1. gB binds to heparin sulphate on the cell surface and is essential for virus entry, cell to cell transmission of infection and fusion of infected cells (Boyle and Compton, 1998; Mocarski, 1996). gB is the principal target of neutralising antibodies against HCMV and is an important candidate of subunit vaccines initiatives. Four major genotypes of gB are known, along with several rare variants (reviewed in Pignatelli *et al.*, 2001). Most of the HCMV-infected individuals develop anti-gB antibodies, targeted primarily to a region called AD-1 (amino acids 552-635), although neutralization is complement-dependent (Speckner *et al.* 1999). Neutralization reaches 50% efficiency at best, indicating that during natural infection the virus may escape the immune system by inducing production of non-neutralizing antibodies, or by producing viruses with variant gBs (Speckner *et al.* 1999).

Glycoprotein complex gC2, which consists of gM (gpUL100) and gN (gpUL73), is another important and abundant constituent of the HCMV envelope (Mach *et al.*, 2000). Hobom *et al.* (2000) reported that the gC2 complex probably contains more than a single gM/gN dimer. The gC2 complex binds heparin and is thought to play a role in virus entry and membrane fusion. Conuterparts of gM and gN in HSV-1 also form a complex. The gM/gN

complex is essential for replication of HCMV (Kari *et al.*, 1994), whereas HSV-1 mutants lacking this complex replicated to lower titres and exhibited reduced virulence (MacLean *et al.*, 1993).

Glycoproteins gH (gpUL75), gL (gpUL115) and gO (gpUL74) form complex gCIII, which is involved in membrane fusion and virus penetration. Homologues of gH and gL are also present in HSV-1, where they are essential genes with a role in virus release and entry. gH is also a target for complement-independent anti-HCMV neutralising antibodies (Mocarski, 1996) and anti-gH antibody can prevent cell to cell transmission of virus (Ramussen *et al.*, 1984). Antibodies that mimic the gH/gL complex also block virus penetration (Keay and Baldwin, 1992). Thus all three components of the HCMV gH/gL/gO are likely to be essential for viral entry.

The HCMV envelope also contains gene products that exhibit amino acid sequence similarity to seven-transmembrane-spanning (7TM) receptors (UL33, UL78, US27 and US28) (Gompels *et al.*, 1995; Margulies *et al.*, 1996). Three of these are related to G protein-coupled receptors (GCRs) (UL33, US27, and the CC/CX₃C chemokine receptor US28) (Margulies *et al.*, 1996; Beisser *et al.*, 2001). GCRs transduce extracellular signals to alter intracellular processes, and the HCMV GCRs may function similarly during the initial virus binding stage of infection. The HCMV envelope-associated chemokine receptor (US28) is expressed on the surface of latently-infected THP-1 monocytes (Beisser *et al.*, 2001), and it has been suggested that circulating infected monocytes expressing US28 *in vivo* might promote dissemination of HCMV by adhering to CX₃C-expressing endothelial cells.

1.3.5 HCMV genome

The double-stranded linear DNA genome of HCMV (around 230 kbp for AD169) is larger than that of any other known herpesvirus. It exhibits a class E structure (see Fig. 1.3) and is capable of genome inversion, forming four genome isomers (Mocarski, 1996).

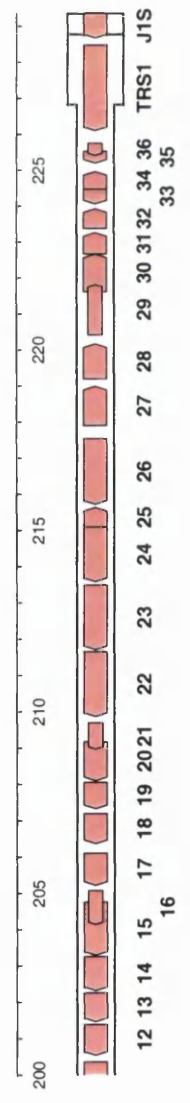
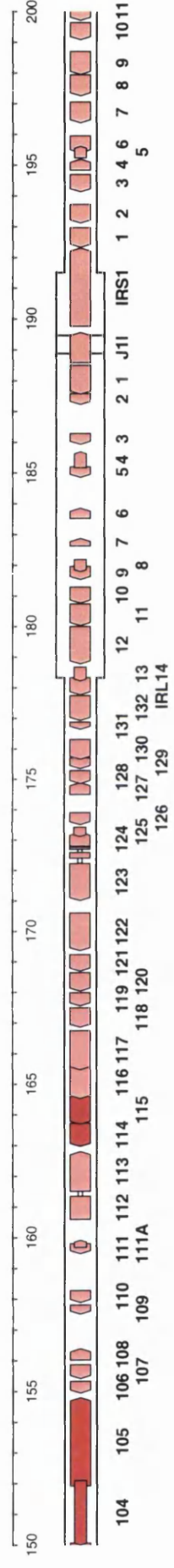
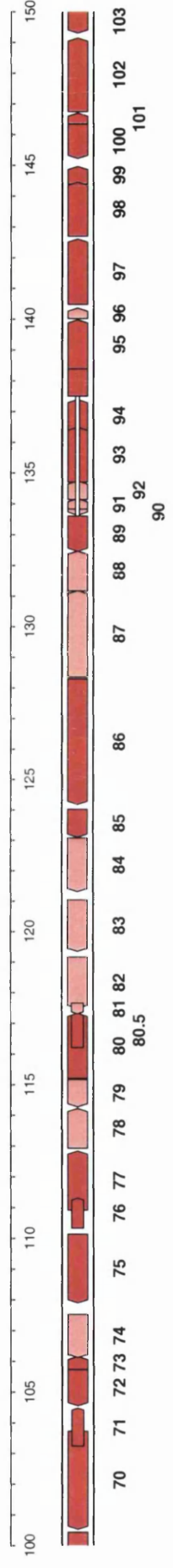
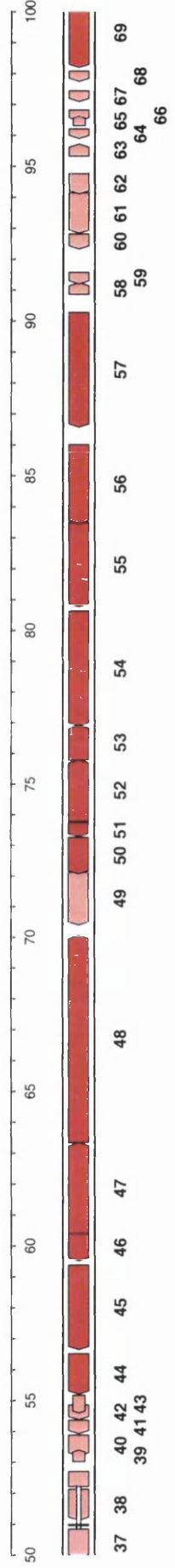
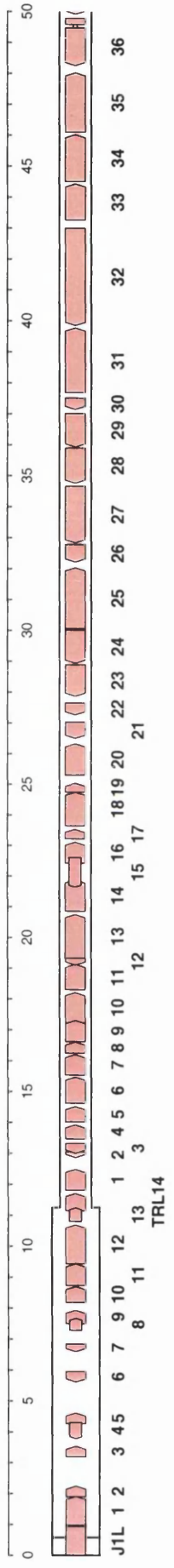
The complete DNA sequence of HCMV strain AD169 has been determined (Chee *et al.*, 1990) by chain termination sequencing of M13 clones generated from *Hind*III fragments cloned in plasmids (Oram *et al.*, 1982). Double-stranded sequencing on appropriate overlapping plasmid clones (Fleckenstein *et al.*, 1982) confirmed that the sequence was contiguous except for an extra 393 bp *Hind*III fragment that was absent from the clone library (Chee *et al.*, 1990). The sequence of AD169 as determined by Chee *et al.* (1990) consists of 229354 bp. U_L is 166972 bp, the flanking repeats (IR_L and TR_L) are 11247 bp each. U_S is 35418 bp and is flanked by 2524 bp repeats (IR_S and TR_S). A directly repeated *a* sequence is present at the genome termini and also in inverted orientation at the IR_L - IR_S junction (Mocarski *et al.*, 1987). The sizes given above, IR_L and IR_S are conserved as overlapping by one copy of the *a* sequence. The *a* sequence contains *cis*-signals, called *pac-1* and *pac-2*, for cleavage and packaging of the viral genome (Mocarski *et al.*, 1987; McVoy *et al.*, 1997). The average nucleotide composition is 57.2% G+C.

Open reading frames (ORFs) was identified in the sequence on the basis of three criteria (Chee *et al.*, 1990): length of the ORF (300 bp or larger), overlap between ORFs (not more than 60%) and codon usage. Chee *et al.* (1990) found 778 ORFs that are over 300 bp, but 581 of these are extensively overlapped by larger ORFs. Consequently 197 candidate protein-coding ORFs were initially proposed. The sequence was then examined for ORFs of less than 300 bp in the gaps, and these were assessed using additional criteria, for example, the presence of potential transcriptional signals, and initiation codons, homology to other ORFs or known genes, presence of protein structural or functional motifs in the amino acid sequence, and patterns of codon usage. From this analysis, Chee *et al.* (1990) proposed a map of 208 ORFs (Fig. 1.8) that are predicted to encode proteins. This amounts to 189 ORFs if diploid genes are counted once only. The ORFs are arranged about equally on both strands of the genome with little overlap. ORFs were numbered sequentially and named with a prefix related to their location in the genome: TRL ($TRL1$ - $TRL14$), UL ($UL1$ - $UL132$), IRL ($IRL14$ - $IRL1$), IRS ($IRS1$), US ($US1$ - $US36$), TRS ($TRS1$) and J (the junction between TR_L and TR_S) (Fig. 1.8).

HCMV is noted for its extensive gene families (Chee *et al.*, 1990). Fifty-two of the ORFs shown in Fig. 1.8 can be grouped into nine families of related genes ($RL11$, $US6$, $US22$,

Figure 1.8: Putative protein-coding ORFs in the AD169 genome as predicted by Chee *et al.* (1990).

The genome is shown in five sections, with a scale in kbp. The thinner and thicker portions in the genome denote the unique regions and inverted repeats, respectively. The approximate locations of genome components are: TR_L; 11 kbp, U_L; 11-178 kbp, IR_L; 178-189 kbp, IR_S; 189-192 kbp, U_S; 192-227 kbp; TR_S; 227-229 kbp. The α sequence (encoding J1L and J1S) is shown at the termini and at the junction of IR_L and IR_S. Protein-coding regions (TRL1-TRL14, UL1-JUL132, IRL14-IRL1, IRS1, US1-US36, TRS1) are shown as open arrows above the gene nomenclature with prefixes omitted for clarity. ORFs predicted to be expressed as spliced mRNAs are connected by narrow horizontal bars. HCMV core genes that are common to α -, β - and γ -herpesviruses are shown in red. Non-core genes are shown in pink. Adapted from Chee *et al.* (1990) and the corresponding entry (X17403). (Provided by A. Davison, MRC Virology Unit, Glasgow).



other non-core
 core

GCR, UL25, UL82, US1, US12 and US2) that appear to have been generated by gene duplication events. There are three pairs (UL25, UL82 and US2) and six larger groups containing 3-14 genes in each family. The gene families (represented by colour coding in Fig. 1.9.) are clustered in U_S and at one end of U_L, regions of the genome that appear to be unique to the β -herpesviruses. The US22 gene family is discussed in section 1.3.9. An additional gene family (UL146) was identified in Toledo and Towne (Penfold *et al.*, 1999; Davison *et al.*, unpublished data), but is absent from AD169.

The analysis of Chee *et al.* (1990) indicates that the majority of HCMV genes appear not to be spliced, and overall the genome has relatively few polyadenylation signals. Consequently, many genes have a polycistronic 3'-coterminal structure (Chee *et al.*, 1990). This pattern can also be seen in other herpesviruses.

A number of studies since 1990 have produced data that have required amendments to be made to the AD169 sequence or to its interpretation. Major changes are described in the following sections.

1.3.5.1 Smith and Pari (1995)

UL102, together with an upstream region of 735 bp, was found to be absolutely necessary for origin-dependent HCMV DNA replication (Pari and Anders, 1993). Subsequently, this upstream region was shown to contain a small ORF (UL101X; overlaps UL101 in a different reading frame) comprising 208 bp immediately upstream from the first ATG codon in UL102. Smith and Pari (1995) mapped a single mRNA species extending through UL102 with its 5'-end located 20 bp upstream from the first ATG codon in UL101X. Sequencing of this region from the Towne and AD169 genomes showed that the UL101X stop codon (TAG) was actually a TAC codon. Moreover, the subsequent codon GGT was a GCT codon. Smith and Pari (1995) found that a cDNA clone started upstream from the first ATG codon in UL101X, continued through the first ATG codon in UL102, and ended downstream from the UL102 stop codon. Therefore, UL101 and UL102 as originally defined were replaced by a larger single ORF, also called UL102.

1.3.5.2 Cha *et al.* (1996)

Cha *et al.* (1996) looked at genome variation in different strains of HCMV and found that a large DNA sequence of Toledo was absent from high passage laboratory strains. AD169 lacks about 15 kbp and Towne about 13 kbp of DNA. Southern blotting indicated that this additional sequence is at the right end of U_L. Sequence analysis of this extra region, using the criteria of Chee *et al.* (1990), identified 19 additional ORFs in Toledo and 4 in Towne compared to AD169. Most of the ORFs contained in the novel 15 kbp Toledo sequence appear to code for membrane proteins, and are obviously not required for replication in cell culture, giving rise to the speculation that they may have roles in cell tropism, pathogenesis or latency (Cha *et al.*, 1996). Lurain *et al.* (1999) reported that the extra 19 ORFs found in Toledo by Cha *et al.* (1996) are also present in clinical isolates. The extra sequence in Toledo and other low passage isolates is therefore likely to be present in wild-type unpassaged HCMV.

Despite having a novel 15 kbp of DNA sequence not present in AD169, the Toledo genome is likely to have a size of about 235 kbp (5 kbp larger than AD169) since IR_L is much shorter than in AD169 or Towne (Cha *et al.*, 1996). Consequently the diploid genes in AD169 TR_L/IR_L are haploid in Towne and Toledo and presumably also in clinical isolates. None of the extra ORFs in Toledo show homology to ORFs in other β -herpesviruses.

Cha *et al.* (1996) speculated that part of the additional sequence present in Toledo but missing from AD169 is in inverted orientation compared with other low passage isolates. Davison *et al.* (unpublished data) confirmed this by sequencing this region from three low passage and two clinical isolates. Moreover, the gene order in CCMV supports the presence of an inverted region in Toledo (Davison *et al.*, unpublished data). It is now clear, therefore, that Toledo is a mutant, even though it is a low passage strain.

1.3.5.3 Dargan *et al.* (1997) and Mocarski *et al.* (1997)

These workers reported a 929 bp segment of AD169 DNA that is absent from the published sequence. This segment was found to be present in most stocks of AD169 at a location within the left part of U_L.

Several AD169 stocks (including *AD169varUK* and *AD169varATCC*) were obtained from different laboratories and digested with *EcoRI*, revealing a novel fragment of 2.8 kbp in place of a 1.9 kbp fragment present in the sequenced strain (Chee *et al.*, 1990). Some stocks consisted of a mixed population of both genomes. The 2.8 kbp fragment was exactly the same as the published sequence of the 1.9 kbp fragment except for an extra 929 bp inserted after nucleotide 54612. The deletion point lies within coding region for UL43 (Chee *et al.*, 1990). Since this extra sequence was present in the majority of AD169 stocks, it is probable that the stock sequenced by Chee *et al.* (1990) had been passaged further than others. Therefore, an amendment to the published sequence was made by Dargan *et al.* (1997) and Mocarski *et al.* (1997). The insertion lengthens UL42 and UL43 and alters the total sequence length to 230283 bp. Dargan *et al.* (1997) also provided a reinterpretation of this region of the genome that replaced UL41 by an alternative ORF (UL41A). UL43 is now a complete member of the US22 gene family.

1.3.5.4 Davison *et al.* (unpublished data)

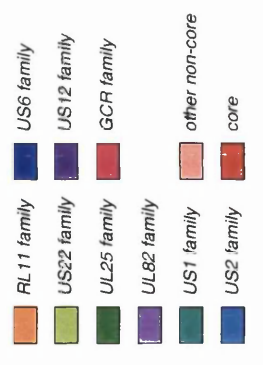
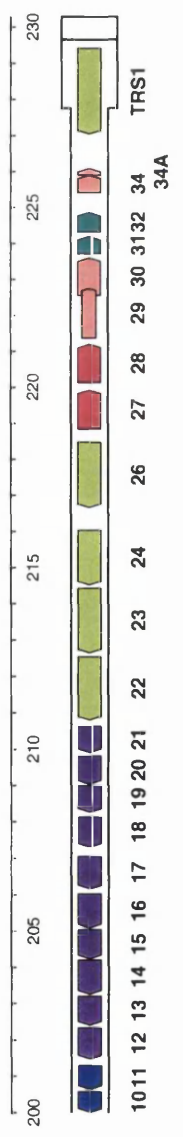
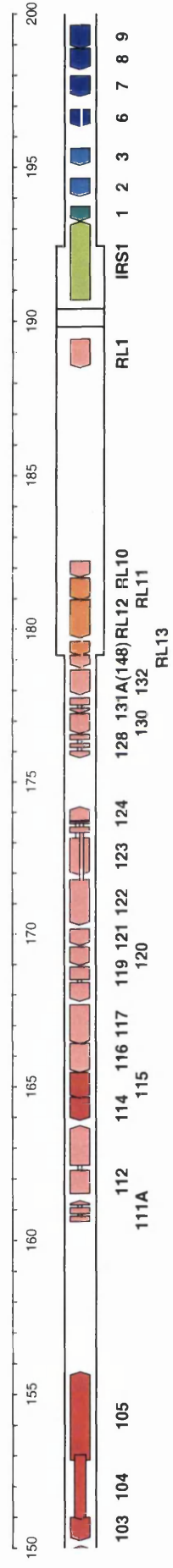
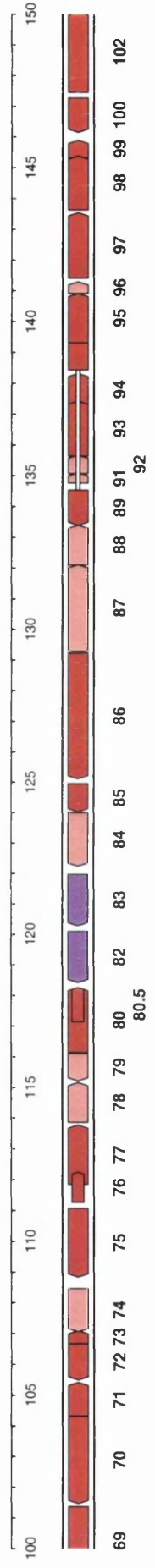
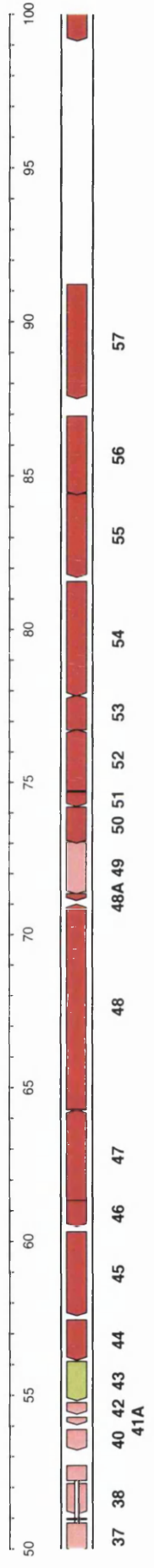
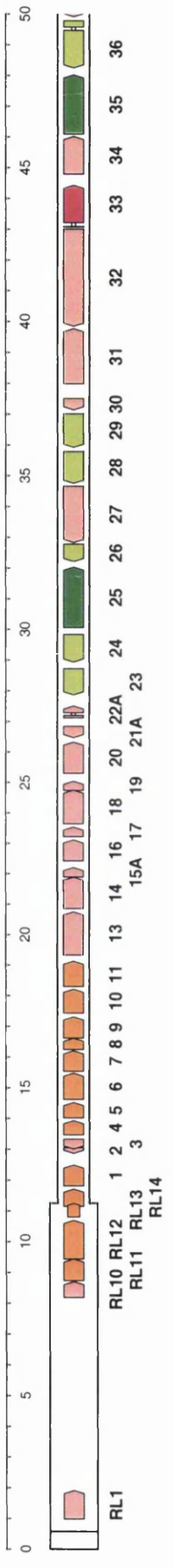
Comparison of published HCMV sequences with those of other betaherpesviruses is a powerful tool for evaluating the likelihood of ORFs encoding protein (Davison, 1993). Recently, the gene content of HCMV has been re-evaluated by comparison to the CCMV sequence (241,087 bp). On the basis of the conservation of ORFs and other sequence features, some previously identified ORFs were excluded, some were modified, and ten novel genes were proposed (five in AD169 and five in Toledo). Davison, A., Dolan, A., Akter, P., Wright, K., Addison, C., Alcendor, D., McGeoch, D. and Hayward, G. (unpublished data) recognised 145 unique genes in AD169, compared with the 189 listed by Chee *et al.* (1990). The number of genes in wild type HCMV is estimated at 164. This analysis is considered a more accurate representation of the genetic content of HCMV than was previously available. The deduced layout of these genes is given in Fig. 1.9.

1.3.6 Regions in the HCMV genome not encoding protein

Amino acid sequence comparisons carried out between AD169 and CCMV revealed three substantial regions in AD169 that are unlikely to encode proteins because the ORFs and

Figure 1.9: Modern interpretation of the AD169 genome according to Davison *et al.* (unpublished data).

The genome is shown in five sections, with a scale in kbp. The thinner and thicker portions in the genome denote the unique regions and inverted repeats, respectively. The approximate locations of genome components are: TR_L; 11 kbp, U_L; 11-179 kbp, IR_L; 179-190 kbp, IR_S; 190-192 kbp, U_S; 192-227 kbp; TR_S; 227-229 kbp. The *a* sequence is shown at the genomic termini and at the junction of IR_L and IR_S. Protein-coding regions (TRL1-TRL14, UL1-UL132 and portion of UL148, IRL14-IRL1, IRS1, US1-US34A, TRS1) are shown as open arrows above the gene nomenclature, with prefixes omitted for clarity. ORFs predicted to be expressed as spliced mRNAs are connected by narrow horizontal bars. HCMV core genes that are common to α -, β - and γ -herpesviruses are shown in red. Gene families are shown in different colors as indicated. (Provided by A. Davison, MRC Virology Unit, Glasgow).



their amino acid sequences are not conserved (Davison *et al.*, unpublished data). These regions contain ORFs TRL2-TRL9, UL58-UL68 and UL106-UL111 identified by Chee *et al.* (1990) (Fig. 1.9).

1.3.6.1 TRL2-TRL9

Chambers *et al.* (1999) determined the kinetics of expression of most of these ORFs using DNA microarray technology, and in some cases, northern blotting. Hutchinson *et al.* (1986) identified TRL5 and TRL6 as expressing an E and E-L RNA by northern blotting. The most abundantly transcribed E gene of HCMV (termed $\beta 2.7$) is located in the TR_L/IR_L and represents approximately 20-40% of the total viral RNA. This RNA appears not to encode any protein, although the RNA is polysome-associated (Greenway and Wilkinson, 1987; Staprans *et al.*, 1988). There is some evidence that the $\beta 2.7$ transcript might be a functional RNA, with a role in specific interactions with cell proteins. It is possible that one role of this RNA is to interact with the host translational machinery to regulate viral and host translation differentially during infection (Greenway and Wilkinson, 1987; Sinclair *et al.*, 2000). A recent study detected a 1.2 kb E RNA from TRL7 that is associated with the virion (Bresnahan and Shenk, 2000).

1.3.6.2 UL58-UL68

RNAs mapped in this region (except UL58 and UL63) fall into the L expression class (Chambers *et al.*, 1999).

The HCMV origin of lytic DNA replication (*oriLyt*) is located in the region UL58-UL68. This element covers 2.4 kbp of a structurally complex region which encodes several transcripts (Huang *et al.*, 1996). The minimal core region of *oriLyt* is located in a region approximately 1.5 kbp between nucleotides 91751 and 93299 in the AD169 genome (Huang *et al.*, 1996), and contains two essential regions (I and II) of 364 and 166 bp separated by non-essential sequence of approximately 400 bp. The central non-essential segment and sequences upstream of essential region I contribute significantly to origin function.

Essential region I (92209-92573) contains the Y-block, a polypyrimidine tract (92471-92501), which is immediately upstream of the region of termination of transcription of the smallest replicator transcript (SRT; see below) (Huang *et al.*, 1996). Region I also contains numerous repeated sequences (Zhu *et al.*, 1998).

Essential region II (92979-93145) has been less intensively studied. This region contains Sp1 binding sites and has been shown to contribute to SRT promoter activity (Huang *et al.*, 1996; Zhu *et al.*, 1998). Region II also contains vRNA2 (Prichard *et al.*, 1998). vRNAs are two small *oriLyt*-derived RNAs reported to be covalently integrated into the replication origins of some packaged HCMV genomes (Prichard *et al.*, 1998). The literature suggests that the 500 bp vRNA1 lies between nucleotides 93700 and 94340 and the 300 bp long vRNA2 between nucleotides 92949 and 93290.

SRT refers to a series of non-polyadenylated early transcripts, with a common 5'-end and loosely defined 3'-ends, which are transcribed from *oriLyt* (Huang *et al.*, 1996). SRT transcription is cycloheximide-sensitive but phosphonoformic acid (PFA)-insensitive, and is probably directed by host-cell RNA polymerase (Zhu *et al.*, 1998). A number of factors (e.g. location in the centre of the *oriLyt* core, E expression, similarity to replicator transcripts in other systems, etc.) suggest that SRT plays a role in initiating or regulating lytic phase HCMV DNA synthesis (Zhu *et al.*, 1998)

1.3.6.3 UL106-UL111

A 5 kb IE RNA was described in this region of AD169 by Plachter *et al.* (1988) before the published sequence of AD169 became available. Nuclease protection assay, primer extension and RNA ligase-mediated 3'-RACE showed that the transcript appeared in the polyA⁺ and polyA⁻ RNA fractions. This is consistent with the previous findings by Jahn *et al.* (1984) that part of this RNA is polyadenylated, but most is not. The 5 kb RNA was also found in other strains, such as Towne (Wathen *et al.*, 1981), Davis (DeMarchi, 1983) and many other low passage isolates (Chandler and McDougall, 1986; Somogyi, *et al.*, 1987). Chambers *et al.* (1999) detected the kinetics of expression of all the ORFs in the UL106-UL111 region.

Wilkinson *et al.* (1984) and Stinski *et al.* (1983) reported that the 5 kb transcript was likely to be originated by splicing from the MIEP. However, no indication of splicing at either the 3'- or 5'-end of this 5 kb transcript was found by Plachter *et al.* (1988). No signals known to be important elements in eukaryotic RNA polymerase II promoters (such as TATA or CCAAT), are present near the 5'-end. Since sequence analysis did not reveal significant evidence for a protein encoded by this transcript, it is likely to represent a large non-coding IE transcript (Plachter *et al.*, 1988). The 5 kb RNA copurified with virions (Bresnahan and Shenk, 2000).

1.3.7 Relationships to other herpesviruses

Chee *et al.* (1990) identified homologs of HCMV ORFs by screening predicted protein sequences against the PIR (George *et al.*, 1986) and SWISSPROT (Barloch, 1988) libraries using the FastA program (Pearson and Lipman (1988). Searches were also performed against a herpesvirus protein library that contained HSV-1, VZV and EBV sequences. More recently, the DNA sequences of five β -herpesviruses have been available and have been compared to HCMV.

Analysis of the DNA sequence of the Smith strain of MCMV (230,278 bp; Rawlinson *et al.*, 1996) shows significant genetic similarities to HCMV strain AD169, particularly in the 78 centrally located ORFs. Homologues of the HCMV UL25, UL82, US22 and GCR gene families exist in MCMV, but there appear to be no homologues of the HCMV RL11, US1, US2, US6 or US12 gene families. The Maastricht strain of rat CMV (229,896 bp; Vink *et al.*, 2000) contains counterparts to all but one of the ORFs that are conserved between MCMV and HCMV.

The DNA sequence of strain U1102 of HHV-6 (159,321 bp; Gompels *et al.*, 1995) was predicted to contain 102 protein-coding ORFs, More recently, Megaw *et al.* (1998) revised the gene content to 85 genes. The HHV-6 genome lacks the RL11, US1, US2, US6 and US12 families. The genome of HHV-7 (144,861 bp; Nicholas, 1996 or 153,080 bp; Megaw *et al.*, 1998) is related to that of HHV-6. Megaw *et al.* (1998) predicted a total of 84 genes

ORF	HSV-1 homologue	Kinetic class ^a	Function or characteristics	References
TRL1		E	NK ^c	
TRL10		E-L	Structural glycoprotein	Spaderna <i>et al.</i> (2002)
TRL11		L	Glycoprotein with IgG Fc-binding capabilities	Lilley <i>et al.</i> (2001)
TRL12		E-L	Glycoprotein	Chee <i>et al.</i> (1990)
TRL13		E-L	Glycoprotein	Chee <i>et al.</i> (1990)
UL1		E-L	Glycoprotein	
UL2		L	NK	
UL3		L	NK	
UL4		E	Glycoprotein	Alderete <i>et al.</i> (1999); Chang <i>et al.</i> (1989)
UL5		E	Glycoprotein	
UL6			Glycoprotein	
UL7		L	Glycoprotein	
UL8			Glycoprotein	
UL9		L	Glycoprotein	
UL10			Glycoprotein	
UL11			Glycoprotein	
UL13		E		
UL14		L	Glycoprotein	
UL15A			Hydrophobic domain	
UL16			Glycoprotein	Kaye <i>et al.</i> (1992a)
UL17		E	NK	
UL18		L	Immune evasion; MHC-1 homologue	Chapman <i>et al.</i> (1999); Browne <i>et al.</i> (1992)
UL19			NK	
UL20			Glycoprotein, Similar to T cell receptor γ chain	Beck and Barrell (1991)
UL21A			NK	
UL22A			NK	
UL23			NK	
UL24			NK	
UL25			Associated with virion, possibly involved in transcriptional activation or interaction with chromatin	Baldick and Shenk (1996)
UL26		E	Virion protein	Baldick and Shenk (1996)
UL27		E	NK	
UL28			NK	
UL29		L	NK	
UL30			NK	
UL31		L	NK	

UL32		L	Virion basic phosphoprotein (BPP), possibly binds to the capsid triplex structure	Greis <i>et al.</i> (1994)
UL33		E	Glycoprotein, homology to GCR	Margulies <i>et al.</i> (1996)
UL34		E-L	Represses transcription of US3.	LaPierre and Biegelke (2001)
UL35		E	NK	
UL36		E	Transcriptional transactivator, anti-apoptotic protein	Patterson and Shenk (1999); Goldmacher <i>et al.</i> (1999); Colberg-Poley <i>et al.</i> (2001)
UL37			Anti-apoptotic glycoprotein	Goldmacher <i>et al.</i> (1999); Colberg-Poley <i>et al.</i> (2001)
UL38			UL36-UL38 region acts as a transcriptional transactivator	
UL39			NK	
UL40		E-L	Glycoprotein. Involved in immune evasion by protecting infected cells from NK lysis	Tomasec <i>et al.</i> (2000)
UL41A			Hydrophobic domain	
UL42			Hydrophobic domain	
UL43		L	NK	
UL44	UL42	E-L	DNA polymerase processivity factor	Ripalti <i>et al.</i> (1995)
UL45	UL39		Homologue of large subunit of ribonucleotide reductase but probably not functional.	Lembo <i>et al.</i> (2000)
UL46	UL38	E-L	Minor capsid binding protein (mC-BP), forms intercapsomeric triplex with pUL85	Gibson <i>et al.</i> (1996)
UL47	UL37	E-L	Virion protein	Baldick and Shenk (1996)
UL48	UL36	L	<i>Tegument protein; required for release of viral DNA from capsid</i>	Batterson <i>et al.</i> (1983)
UL48A	UL35		Small capsid protein (SCP) bound to tips of hexons	Borst <i>et al.</i> (2001)
UL49		E-L	NK	
UL50	UL34		<i>Inner nuclear membrane protein, interacts with UL31 protein</i>	
UL51	UL33		<i>Processing and packaging of DNA</i>	
UL52	UL32	L	<i>Processing and packaging of DNA</i>	
UL53	UL31	E	<i>Nuclear matrix protein, interacts with UL34 protein</i>	
UL54	UL30	E	Catalytic subunit of DNA polymerase	Kouzaride <i>et al.</i> (1987)

UL55	UL27	E	Glycoprotein (gB), binding to ligand up regulates cellular transcription factors Sp-1 and NF-kB	Yurochko <i>et al.</i> (1997)
UL56	UL28	E	Terminase subunit, interacts with pUL89	Buerger <i>et al.</i> (2001)
UL57	UL29	E	Single-stranded DNA-binding protein	Anders and McCue (1996); Penfold and Mocarski (1997)
UL69	UL54	E-L	Transactivator, synergy with pp71 to activate MIEP. Induces cells to accumulate in G ₁ . One form is a tegument protein	Winkler <i>et al.</i> (2000)
UL70	UL52		Component of helicase/primase	Anders and McCue (1996); Pari and Anders (1993)
UL71	UL51		<i>Tegument protein</i>	
UL72	UL50	E-L	Related to dUTPase, but lacks active site motifs.	McGeoch (1990)
UL73	UL49A	E-L	Glycoprotein (gN). Complexed with gM.	Mach <i>et al.</i> (2000)
UL74			Glycoprotein (gO), required for envelope structure.	Huber and Compton (1998)
UL75	UL22	E-L	Glycoprotein (gH). Mediates fusion of the envelope for virus entry.	Yurochko <i>et al.</i> (1997)
UL76	UL24		NK	
UL77	UL25	E	<i>Processing and packaging of DNA</i>	
UL78		E	Glycoprotein, homology to GCR	Milne <i>et al.</i> (2000)
UL79			NK	
UL80	UL26		Assemblin precursor	Gibson <i>et al.</i> (1996)
UL80.5	UL26.5	L	Scaffolding protein	Oien <i>et al.</i> (1997)
UL82		L	pp71, phosphorylated upper matrix protein, transactivator of MIEP	Liu and Stinski (1992); Rawlinson <i>et al.</i> (1996); Bresnahan <i>et al.</i> (2000)
UL83		L	pp65, phosphorylated lower matrix protein	Schmolke <i>et al.</i> (1995)
UL84		E-L	IE2/UL84 complex enhances negative regulation of MIEP	Gebert <i>et al.</i> (1997)
UL85	UL18	E-L	Minor capsid protein (mCP), forms intercapsomeric triplex with pUL46	Baldick and Shenk (1996)
UL86	UL19	E-L	Major capsid protein (MCP), forms hexons and pentons	Rudolph <i>et al.</i> (1990)
UL87			NK	
UL88			Virion component	Baldick and Shenk (1996)

UL89	UL15	E-L	Terminase subunit	Underwood <i>et al.</i> (1998)
UL91		L	NK	
UL92		L	NK	
UL93	UL17	L	<i>Tegument protein; role in DNA packaging</i>	
UL94	UL16	L	<i>Tegument protein</i>	
UL95	UL14		<i>Tegument protein</i>	
UL96		E-L	NK	
UL97	UL13	E-L	Serine/threonine protein kinase, phosphorylates ganciclovir	Prichard <i>et al.</i> (1999)
UL98	UL12	E-L	Alkaline exonuclease.	Sheaffer <i>et al.</i> (1997)
UL99	UL11		<i>Myristylated tegument protein, role in envelopment and transport of virions</i>	
UL100	UL10	E-L	Glycoprotein (gM), integral membrane protein, complexes with gN	Mach <i>et al.</i> (2000)
UL102	UL8	L	Component of helicase-primase	Smith and Pari (1995)
UL103	UL7	L	<i>Tegument protein</i>	
UL104	UL6	E	<i>Capsid portal protein</i>	
UL105	UL5	E	ATPase component of helicase-primase complex	Anders and McCue (1996)
UL111A		E-L	IL-10 homolog (cmvIL-10)	Kotenko <i>et al.</i> (2000)
UL112		E	Encodes 4 phosphoproteins, pp34, pp43, pp50, pp84. pp43 is responsible for IE2 dependent activation of UL54 promoter. Accumulates with IE2 in replication compartments that initiate from PODs	Ahn <i>et al.</i> (1999)
UL114	UL2	E	Uracil-DNA glycosylase	Prichard <i>et al.</i> (1996)
UL115	UL1		Glycoprotein (gL), complexes with gH, role in cell entry	Kaye <i>et al.</i> (1992b); Spaete <i>et al.</i> (1993)
UL116		E-L	NK	
UL117			NK	
UL119		E	Glycoprotein	
UL120		L	Glycoprotein	
UL121		L	Glycoprotein	
UL122		IE, L	Major IE gene, IE2. Transcriptional transactivator	Romanowski <i>et al.</i> (1997); Spector (1996); Stenberg (1996); Liu and Stinski (1992); Stenberg <i>et al.</i> (1989)

UL123	IE	Major IE gene, IE1. Transcriptional transactivator	Romanowski <i>et al.</i> (1997); Spector (1996); Stenberg (1996); Liu and Stinski (1992); Stenberg <i>et al.</i> (1989)
UL124	E	Glycoprotein	
UL128	E	NK	
UL130	E-L	NK	
UL131A	L	NK	
UL132	E-L	Glycoprotein	
UL133 ^b		NK	
UL135 ^b		NK	
UL136 ^b		NK	
UL138 ^b		NK	
UL139 ^b		NK	
UL140 ^b		NK	
UL141 ^b		Hydrophobic domain	
UL142 ^b		Hydrophobic domain	
UL144 ^b		Homolog of TNFR	Benedict <i>et al.</i> (1999)
UL145 ^b		NK	
UL146 ^b		CXC chemokine	Penfold <i>et al.</i> (1999)
UL147 ^b	E-L	Putative CXC chemokine	Penfold <i>et al.</i> (1999)
UL147A ^b		Hydrophobic domain	
UL148 ^b		NK	
UL148A ^b		Hydrophobic domain	
UL148B ^b		Hydrophobic domain	
UL148C ^b		Hydrophobic domain	
UL148D ^b		Hydrophobic domain	
UL150 ^b		NK	
IRS1/		Transcriptional transactivator	Pari <i>et al.</i> (1993);
TRS1			Romanowski <i>et al.</i> (1997)
US1		NK	
US2		Glycoprotein, immune evasion, US2 and US11 induce rapid export of MHC-1 from ER to cytosol where it is degraded by proteasomes	Wiertz <i>et al.</i> (1996)
US3	IE	Glycoprotein, immune evasion, involved in activating expression of cellular <i>hsp70</i> and causing retention of MHC-1 chains in the ER	Jones <i>et al.</i> (1996); Biegelke (1999)
US6	E-L	Glycoprotein, immune evasion, inhibits with TAP-mediated peptide translocation	Ahn <i>et al.</i> (1997)

US7	E-L	Glycoprotein	
US8	E	Glycoprotein	
US9	E	Glycoprotein, implicated in cell-to-cell spread in epithelial cells	Maidji <i>et al.</i> (1998)
US10	E	Glycoprotein	
US11	E	Glycoprotein, US2 and US11 induce rapid export of MHC-1 from ER to cytosol where it is degraded by proteasomes	
US12	E	Multiply hydrophobic	
US13	E	Multiply hydrophobic	
US14	E	Multiply hydrophobic	
US15	E-L	Multiply hydrophobic	
US16	E	Multiply hydrophobic	
US17	E	Multiply hydrophobic	
US18	E	Multiply hydrophobic	
US19	E	Multiply hydrophobic	
US20	E	Multiply hydrophobic	
US21		Multiply hydrophobic	
US22	E	Secreted protein	Mocarski <i>et al.</i> (1988)
US23	E	NK	
US24	E	NK	
US26	E	NK	
US27	E	Multiply hydrophobic, homology to G protein-coupled receptor (GCR)	Kledal <i>et al.</i> (1998); Chee <i>et al.</i> (1990)
US28	E	Multiply hydrophobic, chemokine receptor	Gao and Murphy (1994); Bodaghi <i>et al.</i> (1998)
US29	E-L	NK	
US30	E	NK	
US31		NK	
US32	L	NK	
US34	E	NK	
US34A		NK	

Table 1.5: Functions of HCMV genes.

HCMV gene function is provided with references where applicable. The function of the homologous gene in HSV-1 is provided (pink italic font) where the function of HCMV gene is unknown. The proposed gene content of HCMV is derived from Davison *et al.* (unpublished data). Essential and non-essential genes are shown in red and blue, respectively. ^a From Chambers *et al.* (1999). ^b Gene present only in Toledo strain (Cha *et al.*, 1996; Davison *et al.*, unpublished). ^c Function not known.

arranged co-linearly with those of HCMV. The DNA sequence of tupaïid herpesvirus (195,857 bp; Bahr and Darai, 2001) has been published recently. The sequence is predicted to contain 158 protein-coding ORFs, of which 76 genes have significant homology to genes in HCMV. Homologues of most of the members of the HCMV US22 gene family exist, but there appear to be no counterparts of the TRS1, IRS1, UL24 or UL43.

The HCMV genome contains 41 of the core genes common to α -, β - and γ -herpesviruses (Fig. 1.9). These are arranged in blocks (see section 1.1.4) in different herpesviruses and reflect evolution from a common ancestor (McGeoch and Davison, 1999; Davison, 1993).

Chambers *et al.* (1999) determined the kinetic class of approximately 75% of transcripts specified by the ORFs identified by Chee *et al.* (1990). However, much work is still required to characterise gene functions. The functions ascribed to many HCMV genes depend upon their homology with identified HSV-1 gene products. Many such genes are involved in DNA replication, nucleotide metabolism, capsid structure and virion morphogenesis. Little or nothing is known about the functions of more than half of the predicted HCMV genes. At least 54 genes have characteristics of glycoproteins (Chee *et al.*, 1990). The products of at least 44 AD169 ORFs (TRL4-TRL14, UL1-UL10, UL16, UL18, UL20, UL33, UL128, IRS1, US1-US13, US27; UL24; UL36; UL43; those coded by the additional region in Toledo) are dispensable for virus growth in cell culture (Mocarski, 1996; Patterson and Shenk, 1999; Dargan *et al.*, 1997; Cha *et al.*, 1996). A list of HCMV genes and their functions is given in Table 1.5.

Novotny *et al.* (2001) showed possible matches between predicted HCMV proteins and a library of three-dimensional protein structures, using a threading program. However, the significance of this analysis in being able to discriminate meaningful findings is uncertain.

1.3.8 Spliced transcripts in HCMV

HCMV encodes a number of spliced mRNAs (Elliot *et al.*, 1991). The donor and acceptor sequences of published splice sites are shown in Table 1.6.

ORF	Position ^a		Sequence ^b (5'-3')		References
	Donor	Acceptor	Donor	Acceptor	
UL22A	3323	3253	AC <u>GG</u> TGAAT	ATCGTGT <u>TTTTGCAGC</u>	1, 7
UL33	43087	43209	TC <u>GG</u> TGAGC	TGCTGTTCCGCC <u>CAGA</u>	5
UL36	49575	49471	A <u>AG</u> GTAAGC	TCTATTCTCTACC <u>AGG</u>	4
UL37	52219	50989	CC <u>AG</u> TAAAGC	TCATTTTCTTTCTAG <u>T</u>	3, 4
	50947	50842	C <u>AG</u> GTAATA	GTCGTCTCCACGTAG <u>G</u>	
UL89	137502	133599	AA <u>CG</u> TGAGT	CTCCTCTCTACACAG <u>A</u>	2, 7
UL111A	159857	159934	TT <u>GG</u> TAGGT	GTTTTCCTCTTGTAG <u>C</u>	6
	160134	160218	TG <u>TG</u> TAAAGT	TTGTCGGGCTCCCAG <u>C</u>	
UL112-UL113	161345	161503	AC <u>GG</u> TGAGT	TCGTCCCGTCTGTAG <u>G</u>	9
UL119-UL118	167563	167474	A <u>AG</u> GTAAGT	GAATTTTATCCACAG <u>G</u>	7
UL122	172396	172225	TC <u>GG</u> TAAAGT	TCTTATCACCATCAG <u>G</u>	9
UL123	173610	172782	GAC <u>CG</u> TAAAGT	GGGTCTTTTCTGCAG <u>T</u>	8
	172695	172580	AC <u>GG</u> TACGT	TTCTCATGTGTTTAG <u>G</u>	
	172396	172225	TC <u>GG</u> TAAAGT	TATCCTCCTCTACAG <u>T</u>	

Table 1.6: Published splice sites in AD169.

^a Underlining indicates the last base of the upstream exon and the first base the next exon.

^b Donor and acceptor splice sequences as shown for mapped mRNAs.

References: 1; Bresnahan and Shenk (2000), 2; Costa *et al.* (1985), 3; Goldmacher *et al.* (1999), 4; Kouzarides *et al.* (1988), 5; Margulies *et al.* (1996), 6; Muralidhar *et al.* (1996), 7; Rawlinson and Barrell (1993), 8; Stenberg *et al.* (1989). 9; Wright *et al.* (1988).

Splicing has been reported for additional ORFs, e.g. UL116-UL115, UL118-UL117, US3 and US6 (Leatham *et al.*, 1991; Jones and Muzithras, 1991; Rawlinson and Barrell, 1993), but in these cases the functional significance is not clear. In some cases, locations of intron-exon boundaries are not known precisely (e.g. US6) or the putative acceptor and donor splices in HCMV are not conserved in CCMV.

Splicing patterns may differ between β -herpesviruses. For example, the homolog of UL112 is spliced differently in HHV-6 and HHV-7 (U79). U79 contains three exons (Nicholas, 1996; Megaw *et al.*, 1998), whereas only two have been identified in HCMV. The MCMV counterpart of UL112 (M112) also consists of three exons (Rawlinson *et al.*, 1996).

It is probable that more spliced genes remain to be identified in HCMV. For example, two novel spliced genes (UL128 and UL131A) were identified from comparisons between HCMV and CCMV (Davison *et al.*, unpublished data).

1.3.9 The HCMV US22 gene family

The HCMV US22 gene family contains 13 members: US22, US23, US24, US26, TRS1, IRS1, UL23, UL24, UL26, UL28, UL29, UL36, and UL43. Among these, UL26 is a new member identified by Davison *et al.* (unpublished data). Most of the HCMV US22 genes have counterparts in other β -herpesviruses (HHV-6, Gompels *et al.*, 1995; HHV-7, Nicholas, 1996; Megaw *et al.*, 1998; MCMV, Rawlinson *et al.*, 1996; Rat CMV, Vink *et al.*, 2000; and Tupaiid herpesvirus, Bahr and Darai, 2001), although the precise correspondence is difficult to establish for some, particularly those in TR_S/IR_S and U_S. Homologues have not been identified in other herpesviruses, but the presence of a similar number of US22 genes in the sequenced β -herpesviruses suggests that these genes provide important functions during replication.

Genes belonging to the US22 gene family contain one or more of four conserved amino acid sequence motifs (Chee *et al.*, 1990; Nicholas, 1996). Among these, motifs I and II are well conserved in most US22 genes (Fig. 1.10). Motif I differs between UL and US members and is defined as GXXOXOXWP in UL members and OXOXXPXXW in the latter (where O is a hydrophobic residue and X any residue). Motif II, OOCXXXLXXOG (Kouzarides *et al.*, 1988; Chee *et al.*, 1990) is present in all US22 genes except UL28, UL29, TRS1 and IRS1. Motif III, originally defined as OOCXD/E(X)₁₋₄OXXOG by Nicholas (1996) and recently revised by Davison *et al.* (unpublished data), is present in all US22 genes. Motif IV consists of a short run of charged residues in the C-terminal region (Efstathiou *et al.*, 1992). This motif is poorly defined and therefore not shown in Fig. 1.10.

The functions provided by the conserved motifs are unknown, but their existence argues for some similarity in biochemistry or biology of the US22 family proteins. Interestingly, the spacing of motifs I and II appears to be conserved (20-22 amino acids) and this is presumably reflected in protein folding and function. The direct repeat (DR) regions of

HHV-6 and HHV-7 appear to be related to sequences located in the US region of HCMV and contain US22 gene homologues (Gompels *et al.*, 1995; Nicholas, 1996; Megaw *et al.*, 1998). Homologues of HCMV UL36 (both exons) exist in HHV-6 and HHV-7, and the positions of the splice donor and acceptor sequences are well conserved (Nicholas, 1996). Splicing may also be important for expression of certain HHV-6 and HHV-7 US22 gene family members, since splicing of two pairs of ORFs would result in proteins containing the US22 motifs in their usual order (Megaw *et al.*, 1998).

Approximately twelve MCMV ORFs belong to the US22 gene family (Rawlinson *et al.*, 1996), but as in HCMV, their functions largely unknown. However, MCMV M43 (UL43 homologue) has been shown to be a determinant for viral growth in the salivary gland (Xiao *et al.*, 2000). M139, M140 and M141 (US22, US23 and US24 homologues, respectively) are early proteins that are distributed throughout the infected cell (Hanson *et al.*, 2001). Deletion of M140 or M141 results in impaired replication in macrophages and spleen tissue, whereas M139 does not appear to be required for replication in macrophages. M140 and M141 appear to act individually and together to regulate MCMV replication in a cell-type specific manner and are predicted to influence viral pathogenesis.

To date viral mutants in two HCMV US22 family genes, (IRS1 and UL36), have been generated (Jones and Muzithras, 1992; Patterson and Shenk, 1999). However a naturally occurring mutant with a partial deletion of gene UL43 has been identified (Dargan *et al.*, 1997) (see section 1.3.5.3). Few antibodies have been generated to identify US22 family proteins; those reported are against pTRS1/pIRS1 (Romanowski *et al.*, 1997), pUS22 (Mocarski *et al.*, 1988) and pUL36 (Patterson and Shenk, 1999).

HCMV gene US22, after which the gene family was named, specifies a protein of approximately 76 kDa in infected cells (Mocarski *et al.*, 1988; Dal Monte *et al.*, 1998). US22 is transcribed as an E gene, and pUS22 can be detected from 24 h post-infection (Mocarski *et al.*, 1988). Immune fluorescence showed that pUS22 is present in the nuclei of infected cells at E and L times in infection. However, cell fractionation studies showed that pUS22 is cytoplasmic with a proportion released from cells as a soluble protein at E (24 h) and L (72 to 120 h) times post-infection.

The UL36-38 gene locus is a complex region that codes a number of gene products, only one of which (pUL36) is a member of the US22 family. Both UL36 and UL37 are spliced genes, and UL38 is contained within one of the UL37 introns. Proteins from the UL36-38 locus have previously been shown to be essential for transient complementation of HCMV *oriLyt*-dependent DNA replication (Pari *et al.*, 1993; Pari and Anders, 1993). One or more protein products from the UL36-38 locus have also been reported to co-operate with the major IE genes IE1 (72 kDa) and IE2 (86 kDa) in a synergistic manner to enhance transcriptional trans-activation from some cellular (heat shock protein 70 (*hsp70*) and brain creatine kinase) and HCMV gene (UL112) promoters (Colberg-Poley *et al.*, 1992). The extent to which the transcriptional trans-activating activity of the UL36-38 locus resides in the UL36 gene product has yet to be determined. Recent investigations, however, have indicated that the acidic domain of pUL37x1 and gpUL37 is probably responsible for the trans-activation properties of the UL36-38 locus and the synergism observed in the presence of HCMV IE1/IE2 proteins (Colberg-Poley *et al.*, 1998). Zhang *et al.* (1996), have shown that the acidic domain of pUL37x1 is dispensable for trans-activation of the *hsp70* promoter but is required for IE1/IE2 synergism.

gpUL37 is a type I membrane-glycoprotein, modified by N-linked glycosylation (Al-Barazi and Colberg-Poley, 1996). After being processed through the secretory apparatus gpUL37 is deposited in mitochondria in infected cells (Colberg-Poley *et al.*, 2000). Recently, each of the spliced variants of UL37 (gpUL37, pUL37x1 and gpUL37_M) has been shown to function as a mitochondria-localised inhibitor of apoptosis (vMIA), inhibiting Fas-mediated apoptosis (Goldmacher *et al.*, 1999). The Towne strain of CMV does not express gpUL37 or gpUL37_M, indicating that pUL37x1 alone is sufficient to block apoptosis during infection.

UL36 is a tegument component that is dispensable for viral replication in cell culture (Patterson and Shenk, 1999). pUL36 is expressed under IE conditions and is present throughout infection. On SDS-PAGE gels, pUL36 migrates as a doublet of bands, suggesting possible post-translational modification. The role of pUL36 immediately after delivery to the infected cell is almost certainly to inhibit the initiation of apoptosis. The

Motif I

	UL GXXOXOXWP		US OXOXXPXXW
UL23	GORVALVWP	US22	VALRNPANW
UL24	GQVLPVWVP	US23	LSLGIPHNW
UL28	GRWLPLCWP	US24	LSLGPPKGW
UL29	GSCVSLGWP	US26	LPISAPPGW
UL36	GTRLHVAWP		
UL43	GTVLRLSWP		

(X= any residue, O=hydrophobic residue)

Motif II

	OOCXXXXLXXOG
UL23	YLCCPEPLRFVG
UL24	YLCCQTRLAFVG
UL36	YVCCQEYLHPFG
UL43	FLCCDKFLLPVG
US22	YLCCDDTLEAVG
US23	VICCPERLIVLG
US24	LICCREPLTPLG
US26	LICCEESLES LG

Motif III

	oo00oo g OyOy	oyr0A so eFor G1	y
UL23	SELYLGASGAMYLWTDHIYSDSLTFVAESITEFLNIGL		RRCNF
UL24	AVCLISDEGYVFCY VREDTAVYYLARNLMEFARVGL		RAVETLH
UL26	SVAGVAADGSVLCY EISRENFVVRAADSLPQLLERGL		LHSY
UL28	PEIWVSGHGHAFAY LPGEDKVYVLGLSFGEFFENGL		FAVY
UL29	RILCGDTGTVYAAL VGQDKLVRLARDLRGFVRVGL		ALLIDDFRY
UL36	LIVLIGQRGGIYCY DDLRDCVYELAPTMKDFLRNGF		RHRD FH
UL43	ISILVNECGLVRGV HPETNRAHFLARGLQTFNNGF		LRFNNNY
US22	GLVLLDKFGVVYLH KIEDSDLYRIADNFHMFLKCGL		LKLRGLCRF
US23	VVIVKSHLDRSPPL QRLAGEIYRLADSLEELFRAGL		MKVYVRRRY
US24	VLVLLDWFGAVYAIQMDDPNHYVRRVANTITEFFRMGL		LKMVF
US26	LVLLGRYETVWCL DRDRGVLYYLAHSLDDFARHGL		LHCEAIY
TRS1/IRS1	LVVLLDELGAVFGY CPLDGHVYPLAAELSHFLRAGV		LGALALG

Figure 1.10: Three conserved motifs present in HCMV US22 gene family members.

Adapted from Chee *et al.* (1990), Nicholas (1996) and Davison *et al.* (unpublished data).

UL36 gene encodes an inhibitor of caspase-8 activated (vICA) apoptosis (Skaletskaya *et al.*, 2001). pUL36 binds the pro-domain of caspase-8 and prevents activation of Fas-mediated apoptosis. vICA does not share sequence homology with FLICE-inhibitory proteins (FLIPs) or other known suppressors of apoptosis, and therefore UL36, like UL37x1, appears to encode a new class of cell-death suppressor.

Two additional members of the US22 gene family, pIRS1 and pTRS1, are known to be tegument proteins (Romanowski *et al.*, 1997). The TRS1 gene is transcribed as an unspliced 2.7 kb mRNA. pIRS1 has been shown to be dispensable (Jones and Muzithras, 1992). pIRS1 and pTRS1 are present throughout infected cells at IE and E times post-infection, but become more abundant and predominantly cytoplasmic at L times post-infection (Romanowski and Shenk, 1997). Overall, pIRS1 and pTRS1 are 55% identical and, despite divergence after amino acid 549, the C terminus of each protein is charged, which may be important for protein-protein or protein-DNA interactions.

1.3.10 Gene Expression

Transcription of HCMV genes upon infection is temporally controlled and regulated in a cascade system. In common with other herpesviruses, gene expression falls into three major kinetic classes; immediate early (IE or α), early (E or β) and late (L or γ).

IE genes are those transcribed by host cell RNA polymerase II in the presence of inhibitors of protein synthesis. They include UL123 (IE1), UL122 (IE2), UL36-38, TRS1/IRS1 and US3, whose products appear to be involved largely in subsequent gene regulation steps (Stenberg *et al.*, 1985; Weston, 1988; Kieff and Liebowitz, 1990).

Transcription of E genes requires the presence of IE proteins and is not affected by inhibitors of viral DNA replication (Depto and Stenberg, 1989). E genes are distributed throughout the genome, although TR_L/IR_L is the most transcriptionally active region (DeMarchi, 1981; Staprans *et al.*, 1988). The products of this class of genes have roles in DNA replication, nucleotide metabolism and immune evasion.

L genes are expressed essentially after the onset of viral DNA replication, from around 24 h postinfection (Stamminger and Fleckenstein, 1990). L gene expression may be divided into two sub-classes; γ_1 (leaky late or E-L) and γ_2 (true late or L). E-L transcription and translation start prior to viral DNA replication, but maximal expression occurs subsequently. L gene expression is strictly dependent on viral DNA replication (Stasiak and Mocarski, 1992). Expression at L times occurs from the entire genome and results in the production of transcripts encoding proteins that contribute to the assembly and morphogenesis of the virion. (e.g. virion proteins such as pUL86, phosphoproteins such as pp150 (pUL32) and pp65 (pUL83) and glycoproteins such as gB (pUL55)).

The kinetic classes of HCMV genes can be determined using chemical inhibitors. For example, in the presence of cycloheximide (a protein synthesis inhibitor), only IE RNA will accumulate. In the presence of phosphonoacetic acid (PAA) or ganciclovir (GCV) (inhibitors of viral DNA replication), IE and E RNA will accumulate. Further classification of transcripts as E-L or L is done quantitatively (Chambers *et al.*, 1999).

Although this simplistic scheme provides a convenient system by which individual HCMV genes can be grouped, it is based on highly artificial inhibitor studies. The reality is more complex, with overlap between transcriptional classes. Indeed some genes are expressed in two classes from different promoters (e.g. UL122). Moreover, with some genes there is a delay between transcript production and protein synthesis, indicating that translational control mechanisms contribute to the overall regulation of gene expression. For example, expression of gpUL4 is repressed at the translational level by the presence of a short upstream open reading frame (uORF2) within the UL4 transcript leader, and mutation of uORF2 eliminates ribosomal stalling and results in abundant early expression of gpUL4 (Alderete *et al.*, 2001).

Chambers *et al.* (1999) used DNA microarray technology to detect the expression of 151 HCMV ORFs, 16% of which are IE, 26% of which are E, 26% of which are E-L and 32% of which are L. Some of the differences between microarray data and previously published results are probably due to the inability of the microarray technology to distinguish between overlapping transcripts. This is a serious drawback to the technique, and indicates

the importance of using other techniques to determine the individual kinetic classes, sizes and 5'- and 3'-ends of transcripts.

1.3.11 Transcriptional trans-activators in HCMV

Several proteins encoded by HCMV function as transcriptional trans-activators and thereby create a complex regulatory network that controls gene expression and establishes a progressively changing environment inside the infected cell. Some of trans-activators are expressed at IE times post-infection (IE1 72, IE2 86), whereas others are virion components (pIRS1, pTRS1, pUL69, pp71 (pUL82)) (Romanowski *et al.*, 1997; Winkler *et al.*, 1995; Liu and Stinski, 1992). Certain trans-activators (pIRS1 and pTRS1) are delivered to the infected cell as particle components and are also expressed at IE times.

1.3.11.1 Major immediate early (MIE) gene locus

The major HCMV trans-activators are the nuclear phosphoproteins IE1_{491aa} 72 kDa (IE1 72) and IE2_{579aa} 86 kDa (IE2 86), which are expressed by differential splicing from the major immediate-early (MIE) gene locus. The MIE gene region of HCMV is located between 167 and 175 kbp on the HCMV genome (Fig. 1.9). The first sequence data for this region were reported for HCMV Towne (Stenberg *et al.*, 1984). This region is abundantly expressed from a strong and complex enhancer/promoter, and expression results in the generation of differentially spliced polyadenylated transcripts (exon 4 corresponds to UL123 and exon 5 to UL122 in Fig. 1.11). These transcripts encode two major products, IE1_{491aa} which shares an 85 residue N-terminal domain with IE2_{579aa} (Spector, 1996; Stenberg, 1996). IE1_{491aa} exhibits an apparent size of 68 to 72 kDa and is phosphorylated (Blanton and Tevethia, 1981). IE2_{579aa} also a nuclear phosphoprotein (Hermiston *et al.*, 1987). In addition to IE2_{579aa}, a less abundant protein (IE2_{425aa}) is made from a transcript in which an additional intron is removed from exon 5. The most abundant IE2 gene product (γ IE2_{338aa}) is expressed exclusively late in infection from a non-spliced transcript originating within exon 5 (Stenberg, 1996; Jenkins and Mocarski, 1994). IE2_{338aa} is both a repressor of IE gene expression and a transactivator of E and L genes, although the mechanism of action is different from that of IE2_{579aa} (Barracchini *et al.*, 1992; Mocarski, 1996).

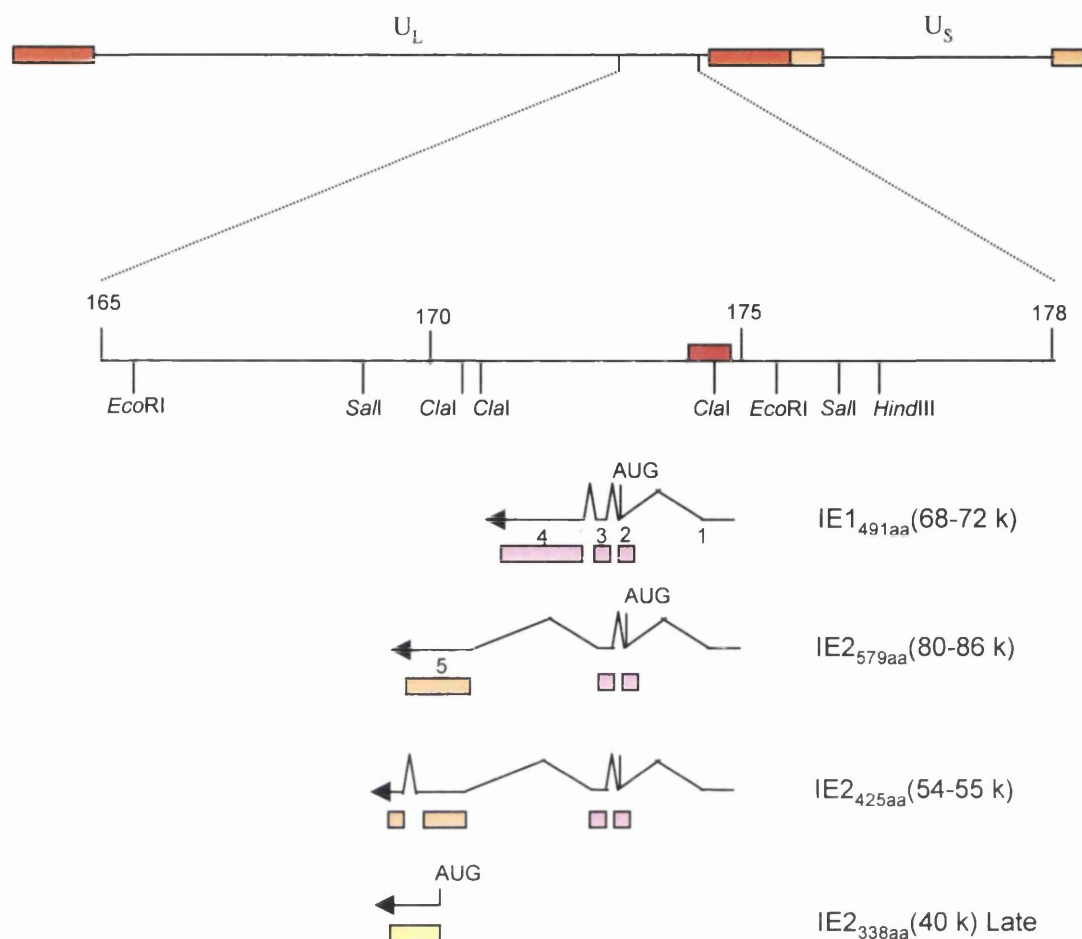
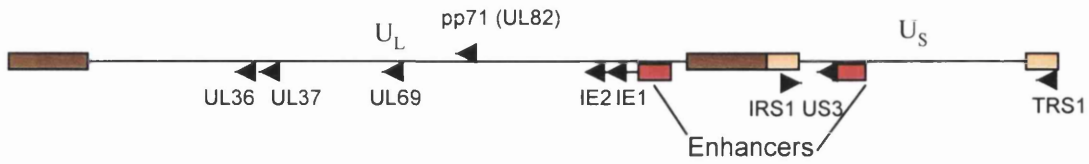
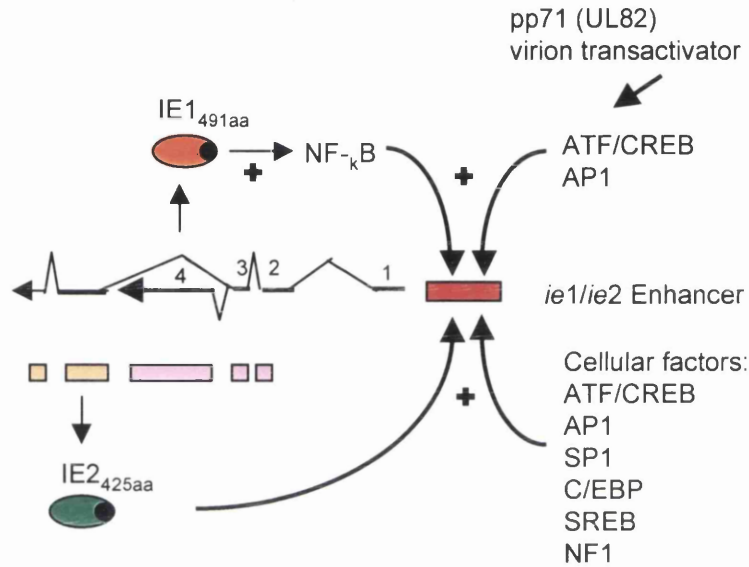


Figure 1.11: Structural organisation of the AD169 MIE gene locus.

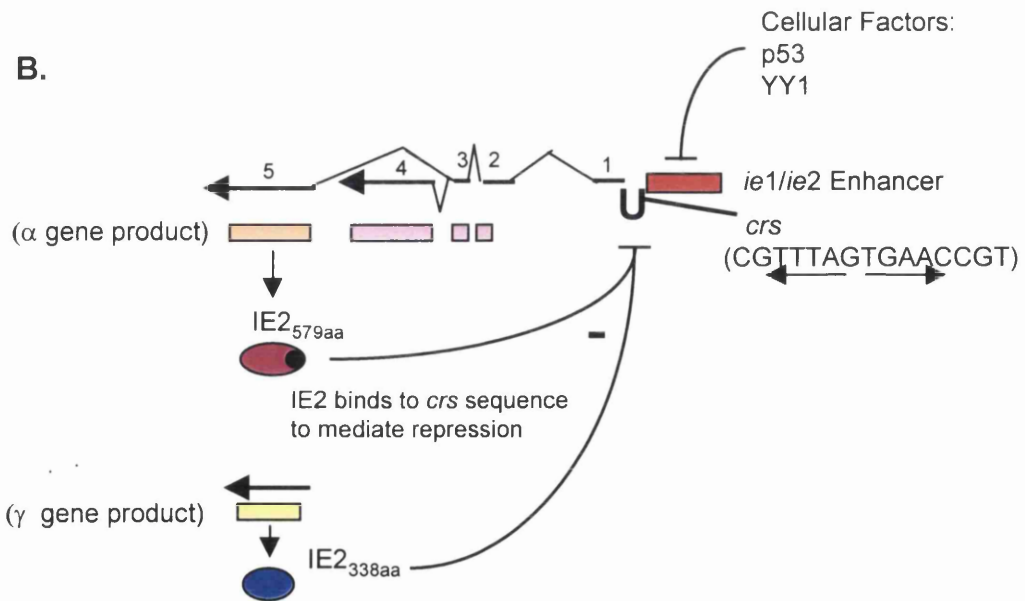
The AD169 genome is depicted on the top line with the region between 165 and 178 kbp expanded below (not to scale). *EcoRI*, *Sall*, *Clal* and *HindIII* restriction sites common to AD169 and Towne are shown. The position of the enhancer is indicated by a red box. Three principal spliced IE transcripts are depicted by arrows with exons numbered and coding regions are indicated by the coloured boxes below. Adapted from Edward and Mocarski (1996).



A.



B.



C.

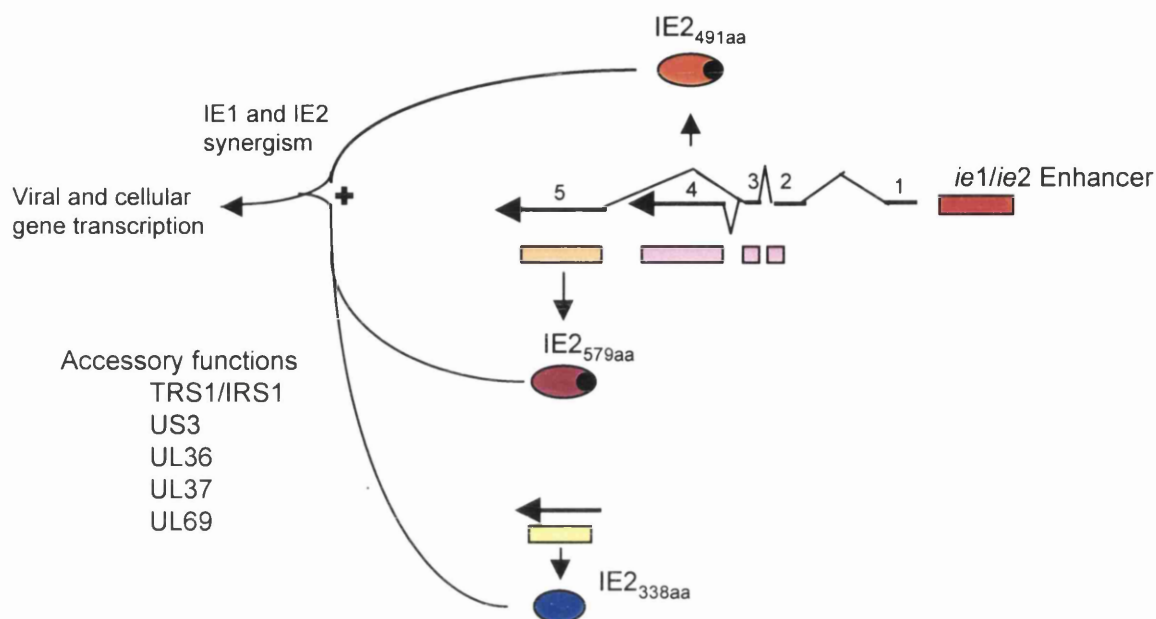


Figure 1.13: Regulation of HCMV gene expression.

The top line depicts the structure of the HCMV genome (not to scale) with the virion transactivator gene UL82, depicted above the line and IE genes (IE1 and IE2, UL36, UL37, UL69, TRS1, IRS1 and US3) along with both IE gene enhancers (red boxes) depicted below the line. Both enhancers contain a high concentration of binding sites for cellular transcription factors (described in text). The lower three parts of the diagram show the IE1/IE2 locus expanded and depict models of the role of regulatory functions in the activation and repression of HCMV gene expression as described in the text. A: Activation of the IE1/IE2 enhancer. B: Repression of IE1/IE2 gene expression. C: Activation of E and L gene expression. Adapted from Edward and Mocarski (1996).

The MIE locus is under the control of the MIE promoter-enhancer (MIEP), which is subject to both positive and negative regulation. The MIEP contains a large number of binding sites for cellular transcription factors and certain viral proteins (Mocarski, 1996). Analogous promoter-enhancers are found upstream of the predominant IE gene locus in MCMV and SCMV, and similar enhancer elements control expression of HCMV US3 (Biegelke, 1999; Colberg-Poley, 1996). The MIEP can be divided into the enhancer region (-500 to + 1, relative to the IE1/IE2 transcription start site), the specific NF-1 binding region (-800 to -500) and the modulator region (-1,100 to -800). Several cellular factors (e.g. p53, YY1, NF- κ B, ATF, CREB, AP1, Sp1, C/EBP, SREB, NF1, TNF- α) have been shown to bind within and upstream of the enhancer region and downstream of the transcriptional start site (Mocarski, 1996; Prosch *et al.*, 1996; Zhang *et al.*, 2001). The enhancer contains repeated sequence elements (16, 18, 19 and 21 bp repeats) interspersed with unique sequences. This complex array of binding sites allows the MIEP to respond to the intracellular environment in a broad range of different cell types. Deletion of the distal enhancer region has been shown to correspond to a deficiency in production of MIE and US3 RNAs, but only during low m.o.i. infections (Meier and Pruessner, 2000). The role of the modulator region is not fully understood, but binding of transcription factors, such as Yin Yang-1 (YY1), has been shown to repress transcription.

IE1 is known to activate its own promoter through the 18 bp repeats and NF- κ B elements and to act via a TATA box-independent mechanism. The IE1 and IE2 gene products have been implicated in both positive and negative regulation of viral gene expression (Fig. 1.12). The critical player in switching from α to β gene expression is IE2_{579aa}. IE2_{579aa} downregulates expression of gene products by binding to the *cis*-repression (*crs*) signal, a 15 bp target sequence. The *crs* sequence is orientation-independent and can confer IE2-dependent repression on heterologous promoters (Cherrington *et al.*, 1991). Repression occurs at the transcriptional level, possibly by IE2 altering RNA polymerase II before the initiation of complex formation. Only the C-terminal half of IE2_{579aa} is required for this repression, whereas it appears that both N and C terminal regions of IE2_{579aa} are important for trans-activation (Mocarski, 1996).

IE2_{338aa} is both a repressor of a gene expression and a trans-activator of β and γ genes. The target sequence for activation is different from the sequence for IE2_{579aa} repression (Fig.

1.12) (Baracchini *et al.*, 1992). The mechanism of synergistic co-operation between IE1_{491aa} and IE2_{579aa} is not fully understood, although protein-protein interactions with host transcription factors may be more important for both than direct binding to DNA (Mocarski, 1996). Experiments on activation of HIV-LTR and *hsp* 70 promoters reveal that trans-activation by IE2 is TATA box-dependent (Caswell *et al.*, 1993, Hagemeyer *et al.*, 1992). IE2 can overcome Dr-1 mediated repression of the *hsp*-70 promoter via a novel derepression mechanism and can bind to DNA and counter histone-mediated repression (Klucher *et al.*, 1993). The ability to bind DNA and engage in protein-protein interactions may allow IE2_{579aa} to serve as the active recruiter for the formation of multiprotein-DNA complexes on HCMV E promoters to ensure that the proteins needed for high level HCMV transcription are present.

1.3.11.2 Other trans-activators

The virion component, pp71 (pUL82) is encoded by a 1.9 kb mRNA that is 3' co-terminal with UL83 transcript (Nowak *et al.*, 1984). pp71 is located at the inner part of the tegument layer attached to the major capsid protein and is translocated into the nucleus immediately after infection (Hensel *et al.*, 1996; Trus *et al.*, 1999). pp71 is critical for initiating lytic infection since it is a potent trans-activator of the MIEP and can also activate a number of viral promoters in conjunction with other viral proteins, including pUL69 (Homer *et al.*, 1999; Winkler *et al.*, 1995). pp71 activates MIE promoters with upstream ATF or AP-1 binding sites and, has been shown to enhance the infectivity of HCMV DNA up to 80-fold (Liu and Stinski, 1992; Baldick *et al.*, 1997).

All mammalian herpesviruses encode a protein that is homologous to HCMV pUL69 (Winkler *et al.*, 2000). Winkler *et al.* (1994) reported that pUL69 is expressed at E-L conditions, but Mocarski (1996) showed that pUL69 is actually expressed with IE kinetics. Three isoforms of pUL69 are expressed: 105 kDa, 110 kDa, 116 kDa. However, only the 110 kDa isoform is incorporated into the tegument (Winkler and Stamminger, 1996). pUL69 stimulates activation of the MIEP in synergy with the tegument trans-activator pp71 (Winkler *et al.*, 1995). This suggests that pp71 and pUL69 may augment the initiation of lytic infection until sufficient IE1 and IE2 have accumulated for their own auto-stimulation. pUL69 exhibits broad trans-activating properties and has a stimulatory effect on several

viral E and heterologous promoters. It can induce cells to arrest in G₁ phase of the cell cycle (Lu and Shenk, 1999; Hayashi *et al.*, 2000). pUL69 localises to nuclear and DNA replication compartments and interacts with a cellular protein that is homologous to yeast factor SPT6, which is thought to regulate chromatin structure. Mutagenesis of the central domain of pUL69, disrupts the pUL69-SPT6 interaction and also results in loss of the trans-activation potential of pUL69, thus implicating the interaction as being essential for trans-activation by pUL69 (Winkler *et al.*, 2000). pUL69 was shown to antagonise the interaction of the C terminus of SPT6 with histone H3, indicating that pUL69 exhibits broad trans-activating properties and affects lytic DNA replication by targeting a chromatin regulatory protein, SPT6, that acts as a repressor of gene expression.

pIRS1 and pTRS1, have been reported to activate the MIEP modestly in transient transfection assays (Romanowski and Shenk, 1997). pIRS1/pTRS1, in combination with pUL69, can activate reporter expression driven by the IRS1/TRS1 promoter. Although pIRS1/pTRS1 can synergise with IE1 and IE2 to activate a range of heterologous promoters, they do not appear to be a broad acting trans-activators. pTRS1 is required for activation of the E UL44 (ICP36) promoter, in combination with IE1 and IE2 (Stasiak and Mocarski, 1992). pIRS1²⁶³ is a truncated version of pIRS1 that is expressed from a promoter within the IRS1 ORF and encoded entirely within the C-terminal domain and same reading frame as pIRS1 (Romanowski and Shenk, 1997). pIRS1²⁶³ is entirely nuclear and is part of a regulatory loop, since it antagonizes transcriptional trans-activation associated with pIRS1 and pTRS1.

The interactions of the HCMV trans-activators are evidently complex and not fully understood. It is known that the tegument proteins pp71, pUL69, pIRS1 and pTRS1 are delivered to the cell immediately after infection and are probably immediately translocated to the nucleus. pp71 can activate the MIEP, in combination with pUL69, inducing expression of IE1 and IE2. pp71 and pUL69 only modestly activate expression of gene US3, although expression is up-regulated by the combination of IE1, IE2, pIRS1 and pTRS1 (Biegelke, 1999). This synergistic activation of US3 expression is inhibited by products from genes UL37x1, UL38 and UL84, the latter also inhibiting expression of IE2 86, and this probably represents a critical step in the switch from IE to E gene expression.

To date, trans-activation has been studied using transient transfection assays (Romanowski and Shenk, 1997). Such an artificial system cannot reveal all the complex interactions that occur *in vivo*, particularly since it is clear that viral gene products intimately interact with cellular products to create the appropriate cellular environment for viral growth. The recent introduction of gene micro-array technology (Bresnahan and Shenk, 1999) will enable the total cellular response to HCMV infection to be monitored and may provide more information on the complex pattern of transcript activation.

1.4 Justification of the study

HCMV is a subject of intense research effort, yet its gene content is less well understood than that of any other human herpesvirus. This reflects the large size and complexity of the genome, and also lack of a laboratory strain with the full genetic content of wild type virus. The genetic content of HCMV strain AD169 as described by Chee *et al.* (1990) was the best available at the time. However, it is not likely to be complete as the criteria employed to identify protein-coding regions were necessarily arbitrary and applied without the benefit of comparisons with other β -herpesviruses.

Moreover, AD169 is a vaccine strain derived by multiple passage in human fibroblasts. Many studies have shown a link between laboratory passaging and genetic loss. Substantial variability in U_L was reported even before the AD169 sequence was available (Weststrate *et al.*, 1983; Takekoshi *et al.*, 1987). Subsequently, Cha *et al.* (1996) found that a large region in the Toledo genome was absent from high passage strains; AD169 lacks about 15 kbp and Towne strain lacks about 13 kbp. This was confirmed by Lurain *et al.* (1999) who reported that the extra 19 ORFs in Toledo were also present in a set of clinical isolates. In 1997, Mocarski *et al.* (1997) and Dargan *et al.* (1997) reported the presence of an extra 929 bp in most AD169 stocks but not in the stock sequenced by Chee *et al.* (1990). Wild type HCMV has therefore been surmised to contain over 220 genes.

Recently, Davison *et al.* (unpublished data) re-evaluated the gene content of HCMV on the basis of amino acid sequence comparisons between HCMV and CCMV, and revised the

gene content of wild type HCMV to 164 genes. It is anticipated that this picture of HCMV gene content will be improved further.

Although the AD169 genome was sequenced over ten years ago, transcriptional and functional data are available for a minority of genes. Recently, Chambers *et al.* (1999) determined the kinetic class of the majority of predicted HCMV transcripts by DNA microarray. However, much work is still required to characterise these transcripts, as this technique does not lead to a full understanding of transcript structure. Therefore, an accurate transcript map for AD169 is still required. The purpose of my study was to evaluate the transcription of a selection of HCMV ORFs, including several that are conserved in CCMV, some that appear unlikely to encode functional proteins because they are not conserved in CCMV, and two novel genes.

1.5 Specific objectives of the study

1. To check potential errors in the HCMV genome sequence indicated from sequence comparisons between HCMV and CCMV.
2. To determine the transcriptional expression patterns of genes in the HCMV US22 family.
3. To assess transcription of other HCMV genes in TR_L and at left end of U_L.
4. To investigate transcription of two novel spliced HCMV genes (UL128 and UL131A) that emerged from sequence comparisons between HCMV and CCMV.

CHAPTER 2:

Materials and methods

2.1 Materials

2.1.1 Viruses

HCMV strain AD169 (Chee *et al.*, 1990) was used in all studies. HCMV strain Toledo (Cha *et al.*, 1996) and DNA from virus isolated directly from the urine of an HCMV-infected child (strain GW; kindly provided by Dr. Gavin Wilkinson, University of Cardiff) were also used for comparative studies.

2.1.2 Cells

The HFFF-2 cell line (initially supplied by J. Mitchell, Institute of Virology) was used for growth of virus and studies of viral gene expression.

2.1.3 Cell culture media

HFFF-2 cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% (v/v) foetal calf serum (FCS), 1% non-essential amino acids, 1% L-glutamine (stock solution of 200 mM) and 1% penicillin-streptomycin (stock solution of 1000 IU/ml and 100 µg/ml, respectively).

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 removal of cell monolayers during passaging or setting up confluent monolayers for infections.

2.1.4 Cosmids

Cosmid 1B (contains the region in the AD169 genome at nucleotides 197042-229354 and 1-7854) was supplied by Dr. D. Dargan.

2.1.5 Chemicals

All general chemicals were obtained from BDH, Boehringer-Mannheim or Sigma.

2.1.6 Solutions and buffers

Cell culture growth media

Eagles A	0.23 g/l CaCl ₂ ·2H ₂ O 0.23 g/l MgSO ₄ ·7H ₂ O 0.1 ml/l HCl
Eagles B	50% (v/v) salts/plus 40% (v/v) amino acids/plus 3.2% (v/v) vitamins
Salts/plus	10.24 g/l NaCl 0.64 g/l KCl 0.24 g/l NaH ₂ PO ₄ ·2H ₂ O 7.2 g/l glucose 0.00016% (v/v) Fe ₂ (NO ₃) ₃ 0.468 g/l L-glutamine 0.016% (v/v) penicillin 0.016 g/l streptomycin 0.00032% (v/v) amphotericin B
Amino acids/plus	0.84 g/l arginine mono-HCl 0.48 g/l cystine 0.384 g/l histidine mono-HCl 1.048 g/l isoleucine 1.048 g/l leucine 1.462 g/l lycine mono-HCl 0.66 g/l phenylalanine 0.952 g/l threonine 0.16 g/l tryptophan 0.724 g/l tyrosine 0.936 g/l valine 0.3 g/l methionine 0.07% (v/v) inositol 0.03% (v/v) phenol red 55 g/l NaHCO ₃
Vitamins	0.05 g/l choline chloride 0.05 g/l folic acid 0.05 g/l nicotinamide 0.05 g/l pantothenic acid, Ca salt 0.05 g/l pyridoxal-HCl 0.05 g/l thiamine-HCl 0.005 g/l riboflavine

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/streptomycin (10000 IU/ml) 0.08% (v/v) amphotericin B 7% (v/v) 10 x Glasgow's modified medium
Giemsa stain	1.5% (w/v) suspension of Giemsa in glycerol heated at 65°C for 2 h, diluted with an equal volume of methanol
Phosphate buffered saline (PBS)	170 mM NaCl 3.4 mM KCl 10 mM Na ₂ HPO ₄ 1.8 mM KH ₂ PO ₄ 6.8 mM CaCl ₂ 4.9 mM MgCl ₂
Storage medium	90% (v/v) foetal calf serum (FCS) 10% (v/v) dimethylsulfoxide (DMSO)
Tris-saline	140 mM NaCl 30 mM KCl 280 mM Na ₂ HPO ₄ 25 mM Tris-HCl 1 mg/ml glucose 0.1 mg/ml streptomycin 100 u/ml penicillin 0.0015% (w/v) phenol red
<u>RNA extraction</u>	
Stock buffer	0.2 M NaCl 0.2 M Tris-HCl, pH 7.5 1.5 mM MgCl ₂ 2% (v/v) SDS
Binding buffer	0.5 M NaCl 10 mM Tris-HCl, pH 7.5

Plasmid preparation

X-gal	40 mg/ml 5-bromo-4-chloro-3-indoyl β -D-galactoside in N, N'-dimethyl formamide
IPTG	30 mg/ml isopropylthio- β -D- galactoside
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 TE	Phenol equilibrated with TE

Primer extension

10 x T4 polynucleotide kinase (PNK) buffer	500 mM Tris-HCl, pH 7.5 100 mM MgCl ₂ 50 mM dithiothreitol 1 mM spermidine
2 x AMV Primer extension buffer	100 mM Tris-HCl, pH 8.3 at 42°C 100 mM KCl 20 mM MgCl ₂ 20 mM dithiothreitol 2 mM each dNTP 1 mM spermidine
Loading dyes	98% (v/v) formamide 10 mM EDTA 0.1% (w/v) xylene cyanol 0.1% (w/v) bromophenol blue

Acrylamide solution	40% (w/v) Acrylogel 5 premix, deionized
Top gel mix (TGM)	6 ml 10 x TBE 18 ml acrylamide solution 55.2 g urea Final volume 120 ml

2.1.7 RNA extraction

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 Microfuge tubes	Invitrogen
TRIZOL Reagent	Gibco Life Technologies
Oligo (dT)-cellulose	Sigma
Spin-X column	Costar
Cycloheximide [3-{2-(3,5-dimethyl-2-oxocyclohexyl)-2-hydroxyethyl} glutarimide]	Sigma
Phosphonoacetic acid (PAA)	Sigma

2.1.8 Northern blotting

Nytran SuperCharge Membrane	Schleicher & Schuell
3MM Paper	Whatman
Lign'Scribe kit:- 10 x Ligation buffer 10 x T4 DNA ligase T7 Promoter Adapter PCR Adapter primer 1 (10 μ M) Control DNA template (mouse cyclophilin; 1 ng/ μ l) 5'-Cyclophilin PCR primer (10 μ M) 3'-Cyclophilin PCR primer (10 μ M)	Ambion
MaxiScript T7 kit:- RNA polymerase (15 U/ μ l) Ribonuclease inhibitor (5 U/ μ l) 10 x transcription buffer ATP (10 mM) CTP (10 mM) GTP (10 mM) UTP (10 mM) DNase I (Rnase-free; 2 U/ μ l) Loading buffer PTRI-actin-mouse (0.5 mg/ml)	Ambion
Rapid hyb buffer	Amersham
Gene-specific primers (GSPs)	MWG Biotech AG
[α ³² P]CTP, [α ³² P]GTP and [α ³² P]UTP (3000 Ci/mmol, 10 μ Ci/ml)	NEN Life Science Products

2.1.9 PCR techniques: PCR, RT-PCR and SMART RACE PCR

Taq DNA polymerase, 10 x PCR buffer with MgCl₂ Boehringer-Mannheim

Titan One Tube RT-PCR System:- Boehringer-Mannheim
 Enzyme mix, AMV-Reverse transcriptase, *Taq* DNA
 polymerase, *Pwo* DNA polymerase
 RT-PCR buffer with 7.5 mM MgCl₂ and DMSO
 Dithiothreitol (100 mM)

SMART RACE cDNA Amplification kit:- Clontech
 SMART II oligonucleotide (10 μM)
 3'-CDS (10 μM)
 5'-CDS (10 μM)
 5 x first strand buffer
 20 mM dithiothreitol
 Deionized water
 10 x Universal primer mix (UPM)
 Human placental total RNA (1 μg/μl)
 5'-RACE transferrin receptor (TFR) primer (10 μM)
 3'-RACE TFR primer (10 μM)
 dNTP (each at 10 mM)
 Tricine-EDTA buffer

50 x Advantage 2 Polymerase Mix Clontech
 and 2 x Advan 2 PCR buffer

dNTPs (10 mM each) Amersham Life
 Sciences

Primers MWG-Biotech AG

2.1.10 Agarose gel electrophoresis

Standard agarose and ethidium bromide (10 mg/ml aqueous solution) Sigma

Low melting point (SeaPlaque) agarose FMC Bioproducts

DNA markers (1 kbp and 100 bp ladders) New England BioLabs

2.1.11 Recovery, cloning and sequencing of DNA fragments

β -agarase I and 10 x β -agarase buffer	New England Biolabs
pGEM-T Vector System I	Promega
One Shot TOP10 competent <i>E. coli</i> cells	Invitrogen
Restriction endonucleases and other enzymes and buffers	New England Biolabs
Internal gene primers for sequencing	MWG-Biotech AG
M13 universal sequencing primers	Pharmacia Biotech

2.1.12 Primer extension

Primer Extension System-AMV Reverse Transcriptase :- 2 x AMV primer extension buffer Sodium pyrophosphate (40 mM) 1.2 kb kanamycin positive control RNA Primer extension control primer (5 pmol/ μ l) AMV reverse transcriptase (AMV-RT) Dephosphorylated ϕ X174 DNA / <i>Hinf</i> I markers T4 polynucleotide kinase (T4 PNK) 10 x T4 PNK buffer Loading dye Nuclease-free water	Promega
Sigmacote	Sigma
N,N,N',N'-tetramethylethylenediamine	Sigma

2.1.13 Bacterial strains

One shot TOP10 *Escherichia coli* strain K12 competent cells (genotype: F⁻ *mcrA* Δ (*mrr-hsdRMS-mcrBC*) ϕ 80*lacZ* Δ M15 Δ *lacX74deoR* *recA1* *araD139* Δ (*ara-leu*)7697 *galU galK rpsL* (Str^R) *end A1 nupG*) (Invitrogen) were used for maintenance and propagation of plasmids.

2.1.14 Bacterial growth media

2YT broth	85 mM NaCl 1% (w/v) bactopectone 1% (w/v) yeast extract autoclaved
L-broth	177 mM NaCl 1% (w/v) bactopectone autoclaved
L-broth agar	1.5% (w/v) agar in L-broth autoclaved

2.1.15 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. Heat-labile solutions were sterilised by filtration through a Whatman syringe filter (0.2 μ m) into a sterile tube.

2.2 Methods

2.2.1 Cell culture

HFFF-2 cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% (v/v) FCS in plastic roller bottles. Cells were grown at 37°C in a humidified atmosphere comprising 90% (v/v) air and 10% (v/v) CO₂. The cells were harvested from the roller bottles by pouring off the medium and washing the monolayer with 20 ml versene. They were then washed with 20 ml trypsin/versene to remove the monolayer and resuspended in fresh medium at a concentration of approximately 10⁷ cells/ml. Aliquots of 10⁷ cells were seeded into roller bottles containing 100 ml of medium. For storage, cells were harvested, pelleted at 1000 x g in a Sorvall RC-5B superspeed centrifuge, and resuspended in storage medium.

Aliquoted cells were frozen slowly to -70°C overnight and then stored in liquid nitrogen. For recovery, cells were thawed quickly and resuspended in growth medium.

2.2.2 Preparation of virus stocks

HFFF-2 cells were grown in roller bottles containing 100 ml medium until 90% confluent. The medium was replaced with 50 ml fresh medium. The appropriate amount of HCMV (usually 0.01 p.f.u./cell) was added to the roller bottles, which were then returned to 37°C. The infected cells were incubated until maximal c.p.e. was observed (usually 10-15 days p.i.). The medium containing infected cells were decanted and the cells pelleted by centrifugation for 15 min at 1000 x g at 4°C in a Sorvall RC-5B superspeed centrifuge. The supernatant was collected and stored at -70°C. Virus stocks were titrated before use.

2.2.3 Titration of virus

Titres of HCMV stocks were determined by plaque assay on HFFF-2 cells. HCMV strain AD169 stocks were serially diluted ten-fold, and 200 µl of each dilution was used to infect an HFFF-2 monolayer in a 35 mm diameter tissue culture dish. Each

dilution was usually titrated in duplicate. After 1 h adsorption with rocking every 20 min. to prevent drying of the monolayers, the inocula were removed and the cells were overlaid with 2 ml¹ per dish of DMEM. Monolayers were incubated at 37°C until plaques were easily visible under a low magnification light microscope. The medium was removed and approximately 1 ml Giemsa stain was added to each dish. After staining for 1 h at room temperature, the cells were washed with water and the plaques counted under a dissection microscope.

2.2.4 Preparation of immediate early (IE), early (E), late (L) and mock infected (MI) RNA

2.2.4.1 Preparation of IE RNA

HFFF-2 cells (10^7 per flask) were inoculated into large flasks and grown overnight. The medium was replaced by fresh medium containing 200 µg/ml cycloheximide, and the flasks incubated at 37°C for 1 h. The monolayers were then infected with HCMV at 5 p.f.u./cell and the flasks incubated at 37°C for 1 h with gentle shaking every 15 min. The monolayers were washed three times with fresh medium containing 200 µg/ml cycloheximide, supplemented with fresh medium containing 200 µg/ml cycloheximide, and incubated at 37°C for 24 h. The flasks were then washed three times with fresh medium containing 200 µg/ml cycloheximide, drained and stored at -70°C.

2.2.4.2 Preparation of E RNA

HFFF-2 cells (10^7 per flask) were inoculated into large flasks and grown overnight. The medium was replaced by fresh medium. The monolayers were then infected with HCMV at 5 p.f.u./cell and the flasks incubated at 37°C for 1 h with gentle shaking every 15 min. The monolayers were washed three times with fresh medium, supplemented with fresh medium containing 300 µg/ml phosphonoacetic acid (PAA), and incubated at 37°C for 48 h. The flasks were then washed three times with fresh medium containing 300 µg/ml PAA, drained and stored at -70°C.

2.2.4.3 Preparation of L RNA

HFFF-2 cells (10^7 per flask) were inoculated into large flasks and grown overnight. The medium was replaced by fresh medium. The monolayers were then infected with HCMV at 5 p.f.u./cell and the flasks were incubated at 37°C for 1 h with gentle shaking every 15 min. The monolayers were washed three times with fresh medium, supplemented with fresh medium, and incubated at 37°C for 72 h. The flasks were washed three times with fresh DMEM, drained and stored at -70°C.

2.2.4.4 Preparation of MI RNA

HFFF-2 cells (10^7 per flask) were inoculated into large flasks and grown overnight. The medium was replaced by fresh medium, and the flasks were incubated at 37°C for 72 h. The flasks were washed three times with fresh medium, drained and stored at -70°C.

2.2.5 Preparation of total cellular RNA

Total cellular RNA was made from infected cells using TRIZOL extraction. Use of this reagent, a mono-phasic solution of phenol and guanidine isothiocyanate, offers an improvement to the single-step RNA isolation method developed by Chomczynski and Sacchi (1987).

Drained infected cells containing MI, IE, E or L RNA were lysed directly in the flasks by adding TRIZOL (2 ml/ 10^7 cells) and passing the cell lysate several times through a 1 ml pipette. The samples were transferred to 1.5 ml microfuge tubes, and incubated at room temperature for 5 min to permit complete dissociation of nucleoprotein complexes. Chloroform (200 μ l/ml of TRIZOL) was then added to each tube. The tubes were capped securely, shaken vigorously by hand for 30 sec and incubated at room temperature for 3 min. Centrifugation was then carried out in a microfuge at 12000 x g for 15 min at 4°C. During centrifugation, the mixture separated into a lower phenol-chloroform phase, an interphase and a colourless upper aqueous phase. The aqueous phase (containing RNA) was transferred to a fresh tube.

RNA was precipitated by mixing with a half volume of propan-2-ol and incubating at room temperature for 10 min. RNA was pelleted in a microfuge at 12000 x g for 10 min at 4°C. The pellet was rinsed with 75% ethanol, air dried, resuspended in 20 µl sterile TE and incubated at 55°C for 10 min to dissolve RNA. Total cellular RNA was stored at -70°C.

2.2.6 Isolation of polyA⁺ mRNA

PolyA⁺ mRNA was purified in two ways.

2.2.6.1 Isolation of polyA⁺ mRNA from total cellular RNA

Total cellular RNA (200 µg) prepared from infected cells was added to 800 µl binding buffer and transferred to a 1.5 ml microfuge tube containing 15 mg oligo (dT) cellulose. The tube was rotated in a rotator at least for 30 min and centrifuged at 4000 x g for 3 min. A microfuge was used throughout for centrifugation steps. The supernatant was discarded and the pellet was resuspended in 800 µl binding buffer. This step was repeated at least twice before transferring the pellet to a Spin-X column in binding buffer. Centrifugation was carried out at 4000 x g for 3 min, and the flow-through was discarded. After repeating this step three times, the pellet was resuspended in wash buffer and centrifuged at 4000 rpm for 3 min. DNA, proteins and cell debris were eluted by treating the pellet in this way with wash buffer several times. The Spin-X column was then placed in a fresh tube and the pellet was twice resuspended in 100 µl 10 mM Tris-HCl (pH 7.5) and centrifuged at 4000 x g for 2 min. 3 M sodium acetate (20 µl) and 500 µl ethanol (100%) were added to the flow-through, and the tube was incubated at -20°C for 3 h. PolyA⁺ mRNA was pelleted at 12000 x g for 15 min at 4°C, washed with 70% ethanol, air-dried for 15 min and resuspended in sterile TE (at about 0.3 µg/µl) and stored at -70°C.

2.2.6.2 Isolation of polyA⁺ mRNA directly from cells

PolyA⁺ selected mRNA was isolated using the FastTrack 2.0 mRNA isolation kit (Invitrogen). This kit allows polyA⁺ mRNA to be isolated directly from cells (10⁷ – 10⁸) or tissue (0.5 – 1.0 g) in 2-3 h. Cells were lysed in the flasks by adding lysis

buffer, and the lysate was passed several times through a 18-21 gauge needle using a 2 ml syringe. Lysis buffer was made by combining stock buffer with 20 μ l/ml RNase/protein degrader. For complete digestion of ribonucleases and other proteins, the samples were incubated at 45°C for 30 min. Centrifugation was carried out at 3000 x g for 5 min and the supernatant was transferred to a fresh tube. A microfuge was used throughout for centrifugation steps.

The NaCl concentration of the lysate was adjusted to 0.5 M by adding 5 M NaCl (63 μ l/ml). The lysate was then passed through a 18-21 gauge needle several times using a 2 ml syringe to shear any remaining DNA. Oligo (dT) cellulose (25 mg) was added and allowed to swell for 2 min. To increase the efficiency of mRNA binding to oligo (dT) cellulose, the tube was rotated gently at room temperature for at least 30 min using a rotator. The oligo (dT) cellulose was then pelleted at 4000 x g for 5 min at room temperature, and the supernatant removed carefully. The oligo (dT) cellulose was resuspended in binding buffer and centrifuged at 4000 x g for 5 min. This step was repeated until the supernatant was no longer cloudy. The oligo (dT) cellulose was gently resuspended in binding buffer, transferred to a spin-column, and centrifuged at 4000 x g for 10 sec.

The spin-column was removed from the tube, the liquid inside the tube was discarded, and the spin-column was placed back into the tube. Binding buffer was added to the spin-column, after centrifugation at 4000 x g for 10 sec, and the flow-through was discarded. This step was repeated several (at least three) times. The oligo (dT) cellulose was resuspended carefully in wash buffer using a sterile pipette tip, centrifuged at 4000 x g for 10 sec and the flow through discarded. This step was repeated at least twice to remove DNA, degraded proteins and non-polyadenylated RNA. The spin-column was placed into a new (sterile and RNase-free) tube and 100 μ l elution buffer was added to the cellulose bed. The tube was centrifuged at 4000 x g for 10 sec, and the flow-through containing polyA⁺ mRNA was retained. A second 100 μ l aliquot of elution buffer was added to the column, mixed well with the oligo (dT) cellulose, and centrifuged at 4000 x g for 10 sec. The column was removed from the tube, and the polyA⁺ RNA was precipitated with 10 μ l of 2 mg/ml glycogen carrier, 30 μ l of 2 M sodium acetate and 600 μ l of 100% ethanol. The sample was

frozen on dry ice until solid, thawed, and centrifuged at 12000 x g for 15 min at 4°C. The supernatant was discarded, and traces of ethanol removed from the RNA pellet after another brief centrifugation. The polyA⁺ RNA was resuspended in 20 µl of elution buffer and stored at -70°C. Typical yields obtained using this procedure were 10–20 µg polyA⁺ RNA per 10⁸ cells.

2.2.6.3 Determination of RNA yield

To determine the concentration of the eluted mRNA in a Beckman DU-62 spectrophotometer, the sample was diluted by adding 1 µl of sample to 99 µl of H₂O. A sample of H₂O was used to establish the baseline of the spectrophotometer at 260 nm, and absorbance of the diluted RNA sample was determined. The RNA concentration in the undiluted RNA sample was determined from the following formula:

$$\text{Concentration of RNA } (\mu\text{g}/\mu\text{l}) = 4 \times A_{260}$$

A₂₆₀ is the absorbance of sample at 260 nm.

2.2.7 Northern blotting

2.2.7.1 Agarose gel electrophoresis of polyA⁺ RNA

PolyA⁺ RNA was separated on a formaldehyde-agarose gel. To avoid contamination with RNase, sterile disposable plastics were used wherever possible, and all glassware was autoclaved before use.

To prepare the gel, agarose (1 g) was dissolved in 85 ml dH₂O by boiling and cooled to 55°C. Formaldehyde (5.8 ml) and 10 x MOPS (10 ml) were preheated at 55°C and added to the agarose, which was mixed well and poured immediately into a gel former with comb and allowed to set at room temperature. A standard 30 µl RNA sample was prepared as follows:

RNA

3 µg polyA⁺ RNA or 3 µl RNA marker

Formaldehyde	5.5 μ l
Formamide	15 μ l
10 x MOPS	1.5 μ l
dH ₂ O	to 30 μ l

Samples were incubated at 55°C (65°C for the marker) for 15 min, and cooled on ice for 5 min. Sterile 10 x loading buffer (3 μ l) was added to the samples and mixed well. The gel was electrophoresed at 80 V for 10 min using 1 x MOPS as electrophoresis buffer prior to loading the samples. The samples were electrophoresed at 80 V for 3 h until the bromophenol blue reached the bottom of the gel. Following electrophoresis, the gel was stained with 0.05 μ g/ml ethidium bromide for 20 min. RNA was visualised with a ruler placed at one side of the gel using short wave UV light and photographed using the Gel Doc system.

2.2.7.2 Transfer of RNA to membrane

RNA was transferred from the gel to a membrane by capillary blotting (Southern, 1975). The gel was incubated in distilled water twice for 5 min and then twice in 10 x SSC for 15 min, with shaking at room temperature. RNA was transferred overnight to a Nytran Supercharge membrane which was pre-soaked in water and then in 20 x SSC. The gel was placed on a wick of Whatman 3MM paper (soaked in 2 x SSC and descending from a platform into a 20 x SSC reservoir). The membrane, three sheets of Whatman 3MM paper presoaked in 20 x SSC and a stack of absorbent paper towels were added sequentially to the top of the gel. Care was taken to eliminate air bubbles between layers, and a plastic tray and a weight were added to the top of the stack. After transfer overnight, the membrane was rinsed gently in 10 x SSC for 1 min and air-dried. The RNA was fixed to the membrane using a Stratagene UV crosslinker set to 'auto-crosslink' (12000 Jcm⁻²). The blots were then either prehybridised immediately or wrapped in clingfilm and stored at 4°C.

2.2.7.3 Preparation of RNA probes

Since the use of RNA probes in northern blot analysis reportedly provides a significant increase in sensitivity compared to using randomly-primed labelled DNA probes, strand-specific RNA probes were used in hybridisation reactions. RNA probes were prepared using the Lig'n Scribe and MAXIscript T7 kit (Ambion). The Lig'nScribe reaction is a two step process (Fig 2.1) which adds a phage T7 RNA polymerase promoter in a chosen orientation to a PCR fragment. The first step was T4 ligase-mediated ligation of the T7 promoter adapter to an appropriate 5'-RACE product (see section 2.2.8), or PCR product if a 5'-RACE product was not available. A standard 10 μ l ligation reaction was as follows:

10 x ligation buffer	1 μ l
T7 adapter	1 μ l
5'-RACE PCR product	2 μ l
T4 DNA ligase	1 μ l
H ₂ O	5 μ l

The reaction was incubated at room temperature for 15 min, and was either used in the second step of the Lig'nScribe protocol or stored at -20°C . The second step involved PCR amplification using a gene specific primer (GSP) (Table 2.1) and adapter primer 1 (AP1). A standard 50 μ l Lig'nScribe PCR reaction was as follows:

Ligation reaction	2 μ l
AP1	1 μ l
GSP (10 μ M)	1 μ l
H ₂ O	39.8 μ l
10 x PCR buffer	5 μ l
dNTPs (10 mM)	1 μ l
Taq polymerase	0.25 μ l

ORF	Primer ^a	Sequences (5' - 3')	Position
UL22A	RUL22	AGCAGCGACGCCGACAAGACCTCA	27082-27105
UL23	RUL23	TTTGTCCACCGACGCGCGACCGCG	28741-28718
UL24	RUL24	CTCGCCCACTGACTCGGCGCGTCA	29860-29837
UL26	RUL26	GTCATCGACCGACAAGGCGCGGCG	32799-32776
UL29	RUL29	TGGAGCGGGTCGCCGAGGCTACTG	37075-37052
UL36	RUL36	GCGAGCGAAGGTAGAGGAGTCCGT	49801-49778
UL43	RUL43	TGGGAACGCGCGCACGGCGCGGTC	56022-55999
US22	US22-II	TCCGACTCGCTGTCGAGACGGCTC	211631-211608
US23	RUS23	GGGGTGTCTAGCTGGCGGCCTCTT	213521-213498
US24	RUS24	AGGAACTGTTGTTTTGCCTGGAGC	214969-214946
UL128	UL128-S3	GCGTCATGAGTCCCAAAGACCTGA	175627-175604
UL131A	UL131-S4	TATTACCGAGTACCGCATTACTGG	176737-176714

Table 2.1: Gene-specific primers for making RNA probes.

^a Position of primer in the AD169 genome (5'-3').

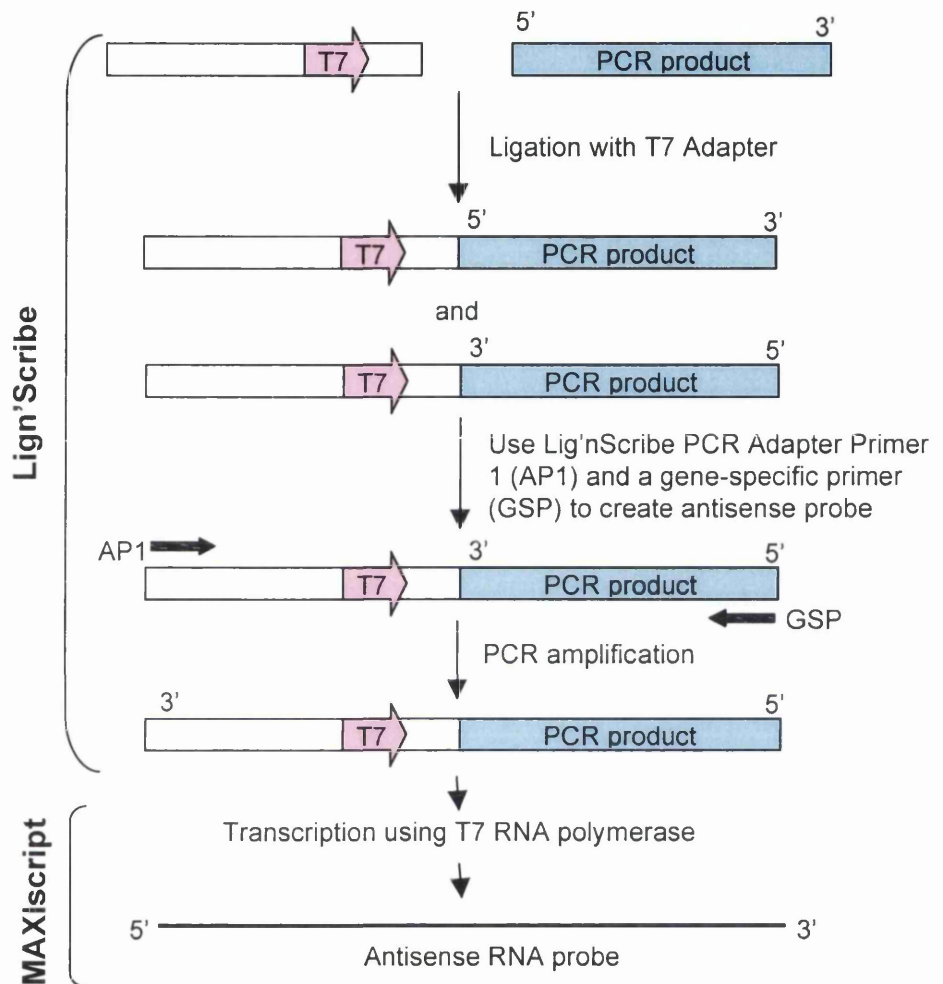


Figure 2.1: summary of Lig'nScribe and MAXiscript reactions.

The standard set of cycling temperature was as follows:

(94°C, 3 min) x 1

(94°C, 30 sec / 57°C, 30 sec / 72°C, 1 min 30 sec) x 30

Hold at 4°C.

The PCR product was purified from a 1.5% (w/v) low melting point agarose gel using β -agarase (section 2.2.12). The size of the adapter-ligated PCR product was that of the 5'-RACE product plus 64 bp. This fragment was then used directly as DNA template in an *in vitro* transcription reaction to produce radiolabelled RNA probe, using a MAXIscript kit (Ambion). A standard 20 μ l MAXIscript reaction was as follows:

10 x Transcription buffer	2 μ l
ATP (10mM)	1 μ l
CTP (10 mM)	1 μ l
GTP (10 mM)	1 μ l
DNA template (from Lig'nScribe PCR)	1 μ g
[α^{32} P]UTP	5 μ l
T7 RNA polymerase	2 μ l

The MAXIscript reaction was incubated at 37°C for 20 min, and then DNaseI (1 μ l) was added and incubated at 37°C for 15 min. T7 RNA polymerase binds to the double-stranded T7 promoter and uses the 3' to 5' strand as a template to synthesize a complementary 5' to 3' RNA transcript (Fig 2.1). Transcription stops at the end of the DNA template. To separate radiolabelled RNA probe from unincorporated nucleotides, water (30 μ l), 5 M ammonium acetate (5 μ l) and 100% ethanol (150 μ l) were added and incubated at -20°C for 30 min. Radiolabelled RNA was pelleted in a microfuge at 12000 x g for 15 min at 4°C, washed with 70% ethanol, air dried and resuspended in 20 μ l RNase-free water. The probe was used immediately or stored at -20°C.

A positive control reaction was carried out to ensure the kit was functioning properly. The control consisted of a plasmid containing sequence from the mouse cyclophilin gene and appropriate 3' and 5' primers, and was supplied with the kit.

2.2.7.4 Preparation of DNA probe

Plasmid pSP64 (Promega) containing the glyceraldehyde 3'-phosphate dehydrogenase (GAPDH₂) coding region was labelled with 20 μ Ci each of [α^{32} P]dGTP and [α^{32} P]dCTP using the Nonaprimer random labelling kit (Appligene). Approximately 500 ng of plasmid DNA was made up to a volume of 10 μ l in a 1.5 ml microfuge tube with H₂O and heated to 95°C for 10 min to denature the dsDNA. The tube was chilled on ice for 5 min and then centrifuged briefly in a microfuge at 12000 x g. The denatured plasmid DNA was mixed with 4 μ l of primer solution (random nonanucleotide mixture), 1 μ l each of unlabelled dATP and dTTP, 2 μ l each of [α^{32} P]dGTP and [α^{32} P]dCTP and 1 μ l of enzyme solution (2.5 U/ μ l 'Klenow' fragment in buffer). The tube was incubated at 37°C for 30 min.

Following incubation, 2 μ l of DNAprep resin (P2) and 60 μ l of adsorption solution (P1) were added, mixed well by vortexing and incubated at room temperature for 2 min to absorb radiolabelled DNA onto the resin. The tube was centrifuged briefly at 12000 x g and the supernatant was discarded. The pellet was resuspended in 100 μ l wash buffer, mixed well by vortexing, and centrifuged briefly at 12000 x g. This step was repeated at least three times in order to separate the radiolabelled DNA from unincorporated dNTPs. The pellet was resuspended in 100 μ l of 'elute' solution (2.5 mM EDTA in 50% deionized formamide), and incubated at 100°C for 10 min. The tube was centrifuged briefly at 12000 x g and the supernatant (containing the purified denatured DNA probe) was recovered. The probe was used immediately or stored at -20°C.

2.2.7.5 Nucleic acid hybridisation

Membranes containing RNA were prehybridised in a hybridisation oven (Hybaid minioven MK II) for 2-4 h at 68°C in 15 ml Rapid-Hyb buffer (Amersham) in glass hybridisation tubes. Denatured probe (DNA or RNA) was added and hybridisation continued overnight. The membranes were washed twice with 2 x SSC, 0.1% (w/v) SDS at room temperature for 30 min, then with 0.5 x SSC, 0.1% (w/v) SDS at 68°C for 1 h. The membrane was then air-dried.

2.2.7.6 Phosphorimager analysis

The dried membrane was wrapped in clingfilm and exposed to a phosphorimager screen for 3-16 hr. The image was visualised using a Bio-Rad Personal FX phosphorimager running the Quantity One software package.

2.2.8 5'- and 3'- RACE

cDNAs were synthesised from polyA⁺ mRNAs using the Marathon SMART RACE cDNA amplification kit (Clontech). This kit integrates the Marathon cDNA Amplification kit (Chenchik *et al.*, 1995, 1996, 1998) with the SMART (Switching Mechanism At 5' end of RNA Transcript) cDNA synthesis technology. Clontech's SMART technology provides a mechanism for generating full length cDNAs (5'-RACE-Ready cDNA or 3'-RACE-Ready cDNA) in reverse transcription reactions (Fig. 2.2). The cDNA for 5'-RACE is synthesized using a 5'-RACE cDNA synthesis primer (5'-CDS) and the SMART II oligonucleotide. The 5'-CDS has two degenerate nucleotide positions at the 3'-end. These nucleotides position the primer at the start of the poly-A tail. When certain MMLV-RT variants reach at the end of an RNA template, they exhibit a terminal transferase activity that adds 3-5 residues (predominantly dC) to the 3'-end of the first strand cDNA. This feature is exploited by the SMART oligonucleotide, whose terminal stretch of dG residues can anneal to the dC-rich cDNA tail and serve as an extended template for RT. After RT switches templates from the mRNA molecule to the SMART oligonucleotides, a complete double-stranded cDNA copy of the original RNA is synthesized with the additional SMART sequence at the end. Following reverse transcription, this cDNA is used directly in 5'-RACE PCR to amplify the 5' ends using a universal primer (UPM in Fig. 2.2) and a gene-specific primer (GSP) in each reaction. This process ensures that the use of high quality RNA will result in products that contains the 5'-end sequences of the RNA.

The 3'-RACE cDNA is synthesized using a traditional reverse transcription procedure (Fig. 2.3), but with a special oligo (dT) primer, 3'-RACE cDNA synthesis (3'-CDS) primer. This 3'-CDS primer includes the lock-docking nucleotide positions as in the

5'-CDS primer and also has a portion of the SMART sequence at its 5'-end. Following reverse transcription, this cDNA is used directly in 3'-RACE PCR to amplify the 3' ends using a universal primer (UPM in Fig. 2.3) and a gene-specific primer (GSP) in each reaction.

IE, E, L or MI RNA (1 μ g) was combined with 3' or 5'-CDS (1 μ l), SMART II oligo (1 μ l) and water (to 5 μ l) in a sterile 1.5 ml microfuge tube and incubated at 70°C for 2 min to make the first strand cDNA. The tube was cooled on ice for 2 min and centrifuged briefly. A mixture of 5 x first strand cDNA synthesis buffer (2 μ l), dNTP (1 μ l), DTT (1 μ l) and Superscript II RT (1 μ l) was added, mixed well and incubated for 90 min at 42°C. Tricine-EDTA buffer (20 μ l) was added and the tube was incubated for 7 min at 72°C. The first strand cDNA (also called 3' or 5'-RACE-ready cDNA) was used immediately for 3' or 5'-RACE PCR or stored at -20°C.

To perform 5'- or 3'-RACE PCR, a 24 bp GSP was designed from the region approximately 500 bp downstream from the putative initiation ATG codon or upstream from the stop codon, respectively, of the target gene. A standard 50 μ l reaction was as follows:

5'- or 3'-RACE-ready cDNA	2 μ l
10 μ M GSP (designed from the appropriate end)	1 μ l
PCR grade H ₂ O	35 μ l
10 x UPM	5 μ l
dNTP mix	1 μ l
10 x Advan 2 PCR buffer	5 μ l
50 x Advantage 2 polymerase mix	1 μ l

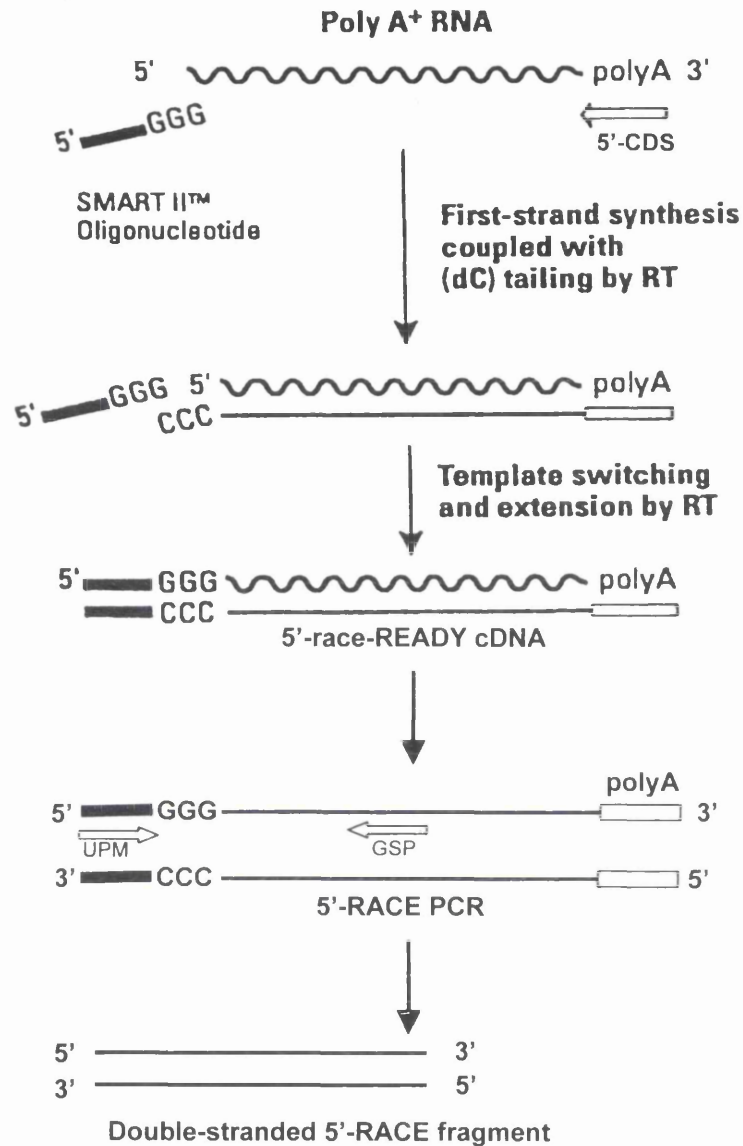


Figure 2.2: Mechanism of 5'-RACE reactions.

5'-RACE-ready cDNA is synthesized using the 5'-CDS primer and the SMART II oligonucleotide. After RT reaches the end of the mRNA template, it adds several dC residues. The SMART II oligonucleotide anneals to the tail of the cDNA and serve as an extended template for RT. Adapted from Chenchick *et al.* (1998).

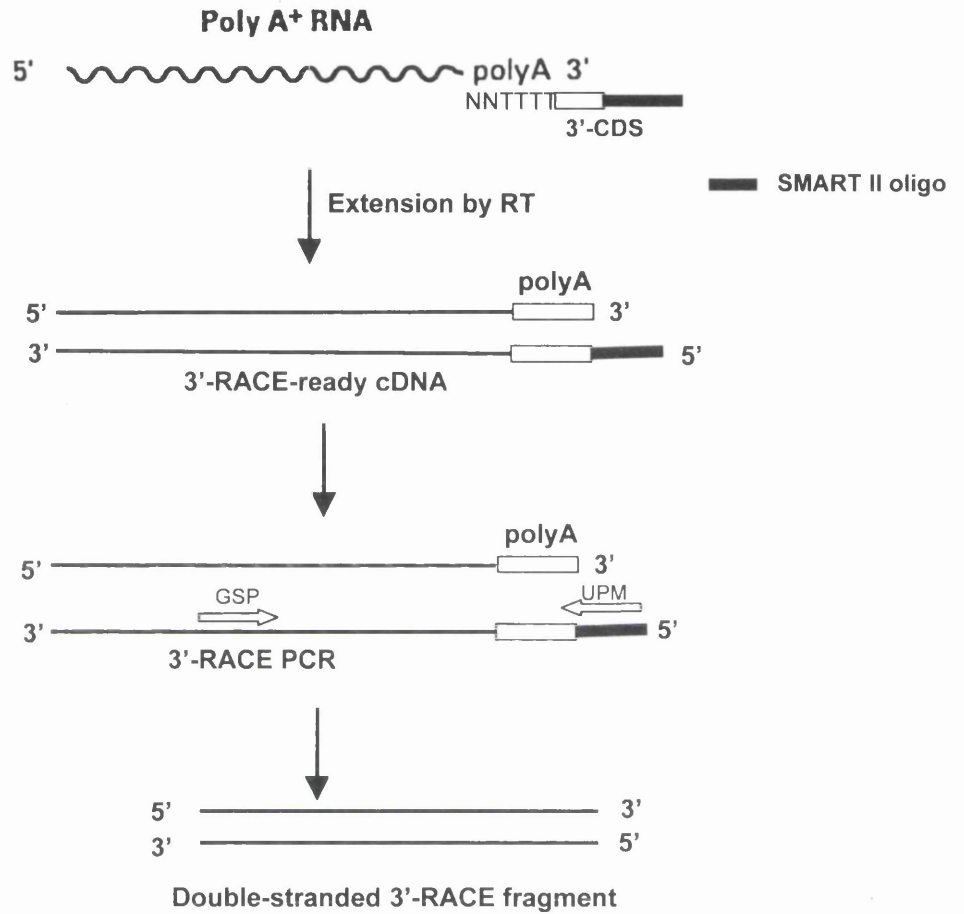


Figure 2.3: Mechanism of 3'-RACE reactions.

3'-RACE-ready cDNA is synthesized using the 3'-CDS primer in the reverse transcription reaction. Adapted from Chenchick *et al.* (1998).

The standard set of cycling temperatures was as follows:

(94°C, 30 sec / 72°C, 3 min) x 5

(94°C, 30 sec / 70°C, 30 sec / 72°C, 3 min) x 5

(94°C, 30 sec / 68°C, 30 sec / 72°C, 3 min) x 20

Hold at 4°C.

A positive control reaction was carried out to ensure the kit was functioning properly. The control consisted of a control human placenta total RNA and appropriate 3' and 5'- RACE primers, and was supplied with the kit.

PCR products (5 µl) were analysed on a 1.2% agarose gel (section 2.2.11). Major bands were then excised and cloned into pGEM-T (Promega) for sequencing (section 2.2.12).

2.2.9 RT-PCR

Splicing patterns in certain genes were confirmed by RT-PCR. IE, E and L RNAs were used as templates with two primers (listed in Table 2.2) mapping in the exons flanking predicted introns. RT-PCR was carried out using the Titan™ One Tube RT-PCR System kit (Boehringer-Mannheim). The standard 50 µl reaction was as follows:

RNase-free H ₂ O	32.5 µl
dNTPs (10 mM each)	1 µl
Primers (10 µM)	1 µl each
Template RNA	1 µl
100 mM DTT	2.5 µl
5 x buffer	10 µl
Enzyme mix (AMV reverse transcriptase in storage buffer)	1 µl

The standard set of cycling temperatures was as follows:

Pre-heat block to 50°C.

(50°C, 30 min) x 1

(94°C, 30 sec / 55°C, 30 sec / 68°C, 1 min) x 10

(94°C, 30 sec / 55°C, 30 sec / 68°C, 1 min) x 25,
 increasing the 68°C step by 5 sec at each cycle
 (94°C, 30 sec / 55°C, 30 sec / 68°C, 7 min) x 1
 Hold at 4°C.

2.2.10 PCR

PCR was performed using *Taq* polymerase in a total volume of 50 µl containing 20-200 ng of template DNA (cosmid or virion DNA). Stringent measures were undertaken to prevent PCR contamination and sterile conditions were used for pre-PCR, PCR and post-PCR steps. The conditions used to amplify the products are as follows:

(96°C, 4 min / 55°C, 2 min / 72°C, 3 min) x 1
 (96°C, 1 min / 55°C, 2 min / 72°C, 3 min) x 25
 Hold at 4°C.

ORF	Primer	Sequence (5'-3')	Position ^a
UL22A	UL22A	TTTGGCTTGATTTTCTTTGTGTTTC	27478-27455
	UL22A-3'	GGCTTTGGCGGCACCTTCTCAGAA	27158-27181
UL36	UL36-S1	ATGGGCCGCTGGTAGTCGCGCATA	49244-49267
	UL36-S2	GCCTACGGCTGCATCGCCATCCGA	49749-49726
UL29	UL29S1	GCGCACCAACCCAAAGTACTGAGT	35537-35560
	UL29S2	TGCGGATTCTCTGCGGCGACACGG	36137-36114
	UL29SP1	GATGCCGTGCAGCGCGGCCAGCA	35573-35596
	UL29SP2	GGCGCTGGTGGGGCAGGATAAGTT	36100-36077
	UL28 (II)	CGGCGTGGAGCTGACATACGCGCA	35176-35199
UL29P1	TGGAGCGGGTCGCCGAGGCTACTG	37075-37052	
UL128	UL128-S1	CGCCCGTCAGCTTCGAGGTATAAC	175058-175081
	UL128-S3	GCGTCATGAGTCCCAAAGACCTGA	175627-175604
UL131	UL131-S3	CGCTGAACCTCGAGGCTCCGGGCGT	176347-176370
	UL131-S2	CATGCGGCTGTGTCCGGTGTGGCT	176862-176803

Table 2.2: Primers used in RT-PCR.

^a Position of primer in the AD169 genome (5'-3').

2.2.11 Agarose gel electrophoresis

DNA fragments generated by PCR (RACE PCR, RT-PCR or PCR) or restriction endonuclease digestion were analysed on 1.0-1.5% agarose gels in 1 x TBE. Samples were prepared in DF dyes and electrophoresis was performed at 100 V for 3 h. DNA bands were visualised using short wave UV light and photographed using the Gel-Doc system (Bio-Rad).

2.2.12 Recovery of DNA fragments and cloning

DNA fragments were separated by gel electrophoresis using 1% (w/v) SeaPlaque agarose. Fragments were visualized under long wave UV light for short periods to minimise DNA damage. Appropriate bands were excised, and the DNA was recovered using β -agarase (New England Biolabs). In a sterile 1.5 ml microfuge tube, 1/10 volume 10 x β -agarase buffer (New England Biolabs) was added to the excised agarose. The tube was incubated for 10 min at 65°C and then at 40°C for 10 min. β -agarase (2 μ l) was added, and incubation was continued for 60 min at 40°C. β -agarase acts by cleaving carbohydrate bonds in the agarose and releasing trapped DNA. 3 M sodium acetate (1/10 volume) was added, and the tube was incubated on ice for 15 min. Centrifugation was then carried out in a microfuge at 12000 x g for 10 min to pellet residual agarose. The supernatant was transferred to a fresh tube and two volumes of 100% ethanol were added. The tube was incubated for 60 min at -20°C and centrifuged at 12000 x g for 15 min. The pellet was washed with 70% ethanol, dried at room temperature and resuspended in 20 μ l TE. Recovered DNA was cloned into pGEM-T using pGEM-T Vector System I (Promega).

2.2.13 Miniprep plasmid DNA preparation

A culture (1.5 ml) of *E. coli* bearing a plasmid was grown overnight at 37°C with shaking, pelleted in a microfuge at 12000 x g for 15 min, resuspended in 100 μ l GTE containing lysozyme (4 mg/ml) and incubated at room temperature for 5 min. NaOH/SDS (200 μ l) was added and the tubes were incubated for 5 min on ice. The solution was neutralised by the addition of potassium acetate solution (200 μ l), and

incubated on ice for 5 min. After centrifugation at 12000 x g for 10 min, the supernatant was transferred to a fresh tube and extracted with equal volume of phenol/chloroform. The DNA was precipitated for 30 min after the addition of two volumes of ethanol, pelleted by centrifugation at 12000 x g for 10 min, washed with 70% ethanol, dried and dissolved in 100 µl TE containing 100 µg/ml RNase. Digestion of the plasmid DNAs using restriction enzymes (see section 2.2.14) were then carried out to ascertain the size of the insert.

For DNA sequencing miniprep DNA from selected clones was incubated at 37°C for 30 min. DNA was precipitated by the addition of 120 µl PEG/NaCl and incubation on ice for 1 h. The DNA was pelleted by centrifugation at 12000 x g for 5 min, washed with 70% ethanol, air-dried, and resuspended in 20 µl TE.

2.2.14 Restriction endonuclease digests

Digestion of plasmid DNAs using restriction enzymes was carried out using commercial restriction enzymes and the buffers supplied, in accordance with the manufacturer's instructions. A typical digest would involve 1 U of enzyme/0.5 µg DNA in a volume of 20 µl, incubated for 1-4 h at the recommended temperature (usually 37°C).

2.2.15 DNA sequencing

Sequencing was performed using the M13 universal primer (Pharmacia Biotech) and gene specific primers. Some PCR products were sequenced directly using the appropriate PCR primers. DNA samples were provided for sequencing at 0.4 µg in 2 µl (usually 1/10 of a miniprep) with 1.6 pmol primer. All the samples were sequenced by Lesley Taylor and Mairi Hope (Institute of Virology) using an ABI PRISM 377 DNA sequencer.

2.2.16 DNA sequence analysis

Sequences were analysed for matches with the HCMV genome sequence using the Basic Local Alignment Search Tool (BLAST) similarity search program. BLAST uses

a heuristic algorithm which seeks local as opposed to global alignments, and is therefore able to detect relationships among sequences which share only isolated regions of similarity with the nucleotide sequence databases (Altschul *et al.*, 1990). BLAST was used online at the National Center for Biotechnology Information (NCBI) at www.ncbi.nlm.nih.gov.

The DNA sequence of HCMV (strains AD169 and Toledo) and CCMV were conceptually translated into amino acid sequences using Ptrans (Taylor, 1986).

2.2.17 Primer extension

Primer extension experiments were carried out using the Primer Extension System-AMV Reverse Transcriptase kit (Promega). A 5'-end-labelled 24 bp GSP (see section 2.2.17.1) mapping in the region approximately 100 nucleotides downstream from the proposed 5'-end of the mRNA was utilised as a primer for RT (section 2.2.17.3). The labelled cDNA was analysed on a denaturing 8% polyacrylamide gel (section 2.2.17.4). Labelled DNA size markers were included. The 1.2 kb kanamycin positive control RNA (10 ng/ μ l) and corresponding primer (5 pmol/ μ l) supplied with the kit were used as a positive control and H₂O was used as a negative control. The length of the cDNA is equivalent to the number of nucleotides between the 5'-end of the primer and the 5' end of the mRNA.

2.2.17.1 5'-labelling of primer

In order to end-label the primer (listed in Table 2.3), the following reaction was assembled in a sterile microfuge tube.

Control primer or GSP	10 pmol
10 x T4 PNK buffer	1 μ l
[γ ³² P]ATP (3,000 Ci/ml)	3 μ l
T4 polynucleotide kinase (10 u/ μ l)	1 μ l
Nuclease-free water	to 10 μ l

The tube was incubated for 10 min at 37°C. It was then incubated at 90°C for 2 min to inactivate the T4 polynucleotide kinase and centrifuged briefly at 12000 x g in a microfuge. The final concentration of end-labeled primer was adjusted to 100 fmol/μl by adding 90 μl nuclease-free water. The end-labeled primer was used immediately or stored at -20°C.

2.2.17.2 Labelling of DNA markers

The protocol described above (section 2.2.17.1) was followed to label 5 μl of the DNA marker. The marker was diluted by adding 190 μl nuclease-free water and stored at -20°C. One μl of the diluted labelled marker was loaded onto the gel.

ORF	Primer	Sequence (5' to 3')	Position ^a
TRL2	PET2	GCCTTCACCCTCTTCACTCCCTGT	2159-2136
TRL5	PET5	TCTTCCGCGTCTCCCGCCGTACC	4326-4303
TRL8	PET8	TGTGGGTGATACCCAACCGGACGC	7426-7403
TRL9	PET9	CTCCGTTTCTCCTCAGCTGCCGTA	7793-7770
UL24	PEL24	CGCGCCAGCGTGGCGAGGACACT	29905-29928
UL36	PEL36	TCGGATGGCGATGCAGCCGTAGGC	49726-49749
US22	PES22	AGCATCTTCGCAGCGCTCGTCGCG	212524-211548
US24	PES24	ACAGCGTCCCTCGTTACGATGAAC	214902-214925

Table 2.3: Primers used in primer extension experiments.

^a Position of primer in the AD169 genome (5'-3').

2.2.17.3 Primer extension reaction

Control RNA (50 ng) or sample RNA (1 μg) was combined in a sterile microfuge tube with 1 μl of control or specific labelled primer, 5 μl of 2 x AMV primer extension buffer and nuclease-free water (to 11 μl). The control RNA sample was incubated for 30 min at 58°C. However, as the melting temperature (T_m) of each primer is the optimal temperature for annealing that primer to RNA (Sambrook *et al.*, 1989), each

tube was incubated at its calculated T_m for 30-60 min. Following incubation, the tubes were cooled for 10 min at room temperature. A 'master' RT extension mixture was prepared in a sterile microfuge tube by combining 5 μ l 2 x AMV primer extension buffer, 1.4 μ l sodium pyrophosphate (40 mM), 1.6 μ l nuclease-free water and 1 μ l AMV-RT per reaction. AMV-RT was added last and the solution was mixed well by pipetting and by gently inverting the tube. Immediately after mixing, 9 μ l of 'master' RT mixture was added to each reaction tube containing the annealed primer and RNA. The tubes were incubated for 30 min at 42°C. After incubation, sodium acetate (1/10 volume) and two volumes of 100% ethanol were added to the tube and incubated at -20°C for at least 3 h. Centrifugation was then carried out at 12000 x g for 15 min at 4°C, and the supernatant was removed carefully. The pellet was washed with 70% ethanol and resuspended in nuclease-free H₂O. Samples were used immediately or stored at -20°C.

2.2.17.4 Polyacrylamide gel electrophoresis

Primer extension products were analyzed by electrophoresis on an 8% polyacrylamide gel. Two siliconised glass plates (33 x 37 cm), polished with 95% ethanol and separated by 1.5 mm spacers, were sealed with 38 mm Scotch electrical tape and secured by two large foldback clips positioned over each spacer. TGM (80 ml) was mixed with 160 μ l 25% (w/v) APS and 160 μ l TEMED. The mixture was poured into the gel sandwich, and bubbles were dislodged by striking the glass plates vigorously. The gel sandwich was placed almost horizontal and a 20 tooth Teflon well-forming sharks tooth comb was inserted, 0.5 cm into the gel (teeth uppermost). Three large foldback clips were positioned to secure the combs, and the gel was allowed to set for 30 min before removing the clips and the tape at the bottom of the gel. The gel sandwich was clamped into a BRL S2 electrophoresis kit and the reservoirs were filled with 0.5 x TBE. The comb was removed from the gel and the well was washed out three times with 0.5 x TBE using a 60 ml syringe fitted with a needle. The gel was pre-electrophoresed at 40 W for 15 min.

Primer extension product (5 μ l) was mixed with 5 μ l of loading dyes, denatured for 10 min at 90°C and cooled on ice for 5 min. The surface of the gel was washed and the

Teflon comb was repositioned with its teeth touching the surface of the gel. Aliquots (10 μ l) of the samples were loaded onto the wells. One μ l of labelled denatured DNA marker was loaded onto well alongside the samples. The samples were electrophoresed at 40 W for approximately 3 h until the bromophenol blue was 2 cm from the bottom of the gel. The gel sandwich was removed from the electrophoresis apparatus and the plates were separated using the point of a pair of scissors. The gel was soaked 5 times with 10% acetic acid for 5 min in a fume hood, transferred onto two sheets of 3MM paper, and covered with clingfilm. The gel was dried on a gel drier for 45 min at 80°C under vacuum and exposed to a phosphorimager screen at room temperature for 16-24 h. The image was analysed using a Personal FX phosphorimager (Bio-Rad).

CHAPTER 3: RESULTS 1

Potential errors in the HCMV DNA sequence

Potential Errors in the HCMV DNA Sequence

Amino acid sequence comparisons carried out between translated forms of the CCMV and HCMV genome sequences indicated the presence of five potential frameshift errors in the AD169 sequence, in TRL14, UL15, UL102, UL131 and US22 (Davison *et al.*, unpublished data). Errors in UL102 were recognised previously by Smith and Pari (1995). A sixth error was identified in UL145, which is absent from AD169 but present in Toledo. To investigate these potential errors, cognate HCMV sequences were analysed by PCR and sequencing. AD169-infected cellular DNA was used as template for PCR in all cases except for UL145, where Toledo-infected cellular DNA was used. In the case of US22, AD169 cosmid 1B was also used. PCR primers were designed to amplify fragments containing the potential errors and are listed in Table 3.1. PCR products (208-712 bp) were separated by gel electrophoresis (Fig. 3.1), excised and purified, and cloned into pGEM-T. The fragments were sequenced in both orientations. Sequences from five clones were analysed for each region analysed.

3.1 Sequences of PCR products

3.1.1 UL15

Fig. 3.2 shows six-frame translations of the relevant regions of the AD169 and CCMV genomes. UL15 as defined in AD169 and shown in blue in panel I is not conserved in CCMV (panel II). However, a CCMV ORF (UL15A) located on the other strand and shown in red in panel II appears to be conserved in AD169 in three different reading frames (bold black font with conserved residues underlined). This analysis indicates that the potential presence of two frameshift errors in the AD169 sequence within the region marked in yellow in panel I. Sequence analysis of a PCR product derived from this region of the AD169 genome using the primers shown by arrows in Fig. 3.2 (panel I) indicated the presence of one extra G residue after nucleotide 21955 and

ORF	Primer	Sequence (5' - 3')	Position in genome ^a	Region analysed ^c	PCR product (bp)
TRL14	ETRL14	GCCCATGAGCGAGAATAATATT	10769-10792	11481-10769	714
	TRL14	TTACAGGGCATGTGCCAAGAGTGG	11481-11458	<u>(11201-11209)</u>	434
UL15	UL15-I	ATCGTACCGGCGTCTGAACACCGA	21757-21780	21800-22200	465
	UL15-II	ACACGCGTTCCTTGACATCGATTTC	22222-22199	<u>(21956, 22005)</u>	
UL102	UL102-1	TAGAGGCCTCGGTGCCTATCGTAC	146607-146630	146915-146607	309
	UL102-2	TCTCCACGCCACTAGCGTCCGTCT	146915-146892	<u>(146744-146749)</u>	
UL131	UL131-S3	GCTGATGAGCAACGACACCTCGGT	176547-176570	176347-176826	379
	UL131-S2	CATGCGGCTGTGTGGGTGTGGCT	176826-176803	<u>(176742-176749)</u>	
UL145	UL145-1	AATAAGTGTGCGCTACATAAGAGC	8715-8738 ^b	8715-9015	301
	UL145-2	ACGGCTGTACGGCACTGTATCGA	9015-8992	<u>(8884 - 8889)</u>	
US22	US22	TCCAGTCGTAGACGTAGACGGTG	211144-211167	211631-211171	488
	US22-II	TCCGACTCGCTGTGAGACGGCTC	211631-211608	<u>(211536)</u>	

Table 3.1: PCR primers designed to investigate potential errors in the HCMV sequence.

^a Position of primer in the AD169 genome (Chee *et al.*, 1990).

^b Position of primer in the Toledo sequence (Cha *et al.*, 1996).

^c Region containing potential errors is underlined.

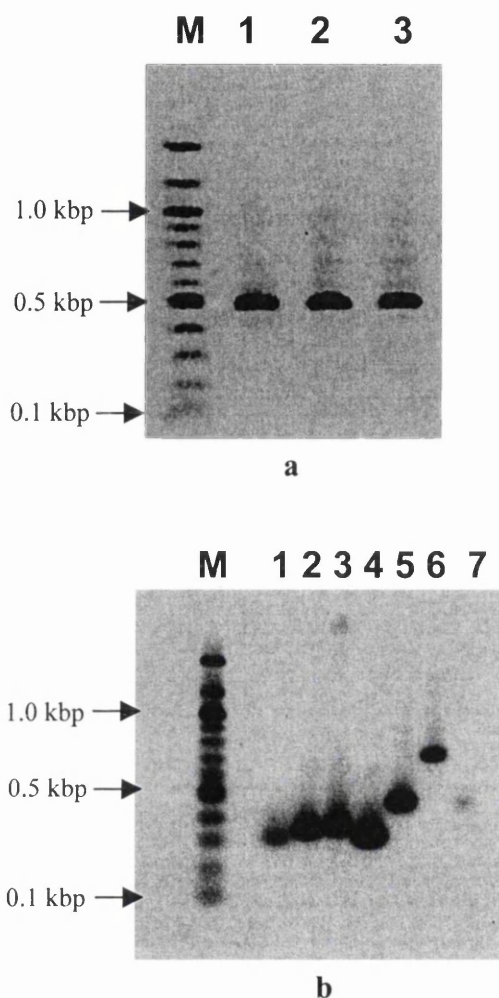


Figure 3.1 : Gel electrophoresis of PCR products from regions of the HCMV genome containing potential errors.

Gel a: Products from US22 (488 bp). M, 100 bp ladder. 1 & 2; AD169 template. 3; cosmid 1B template.

Gel b: Products from other regions. M, 100 bp ladder. 1, UL145 (301 bp) with Toledo template. 2, UL131 (379 bp) with AD169 template. 3, UL131 (~379 bp) with GW template. 4, UL102 (309 bp) with AD169 template. 5, UL15 (465 bp) with AD169 template. 6, TRL14 (714 bp) with AD169 template. 7, TRL14 (approx. 450 bp) with Toledo template.

another extra G residue after 22004. The electrophoregrams are shown in Fig. 3.3 (panel II) and the additional nucleotides are highlighted in yellow in Fig. 3.3 (panel I). These corrections disrupt AD169 UL15 and introduce a novel ORF corresponding to CCMV UL15A. The amino acid sequences of AD169 and CCMV UL15A are shown in red in Fig. 3.3 (panels I and III); each is followed by a putative polyadenylation signal (AATAAA or ATTTAAA) underlined in green. Inability to detect the 5'-end of an mRNA for AD169 UL15 (section 7.1) is consistent with the conclusion that this ORF does not encode a protein. Consequently, UL15 has been removed from the list of AD169 genes, and a novel gene (UL15A) introduced. The proteins encoded by UL15A in AD169 and CCMV possess potential transmembrane domains near their C termini, 18 residues in AD169 and 20 residues in CCMV (shown in boxes in Fig 3.3 panels I and III), and lack homologues in public protein databases.

3.1.2 UL102

HCMV UL102, together with a large upstream region (735 bp), was found to be absolutely necessary for origin-dependent DNA replication (Pari and Anders, 1993). Subsequently, this upstream region was shown to contain a small ORF (UL101X) comprising 208 bp immediately upstream from UL102 (Smith and Pari, 1994). Fig. 3.4 shows the relevant regions of the AD169 and CCMV genomes. AD169 UL102 (Fig. 3.4 panel I) is well conserved in CCMV (Fig. 3.4 panel II). However, the coding capacity of an additional region of approximately 200 bp near the 5'-terminus of CCMV UL102 appears to be conserved in AD169 UL101X (Fig 3.4 panel I). This analysis indicates the potential presence of errors in the AD169 sequence within the region marked in yellow in panel I. Consistent with this, Smith and Pari (1995) mapped a single mRNA species extending through UL102 with its 5'-end located 20 bp upstream of the first ATG codon in the UL101X. Sequence analysis of a PCR product derived from this region of the AD169 genome indicated that the G residue in the UL101X stop codon at nucleotide 146746, is actually a C residue, changing a TAG to a TAC codon. Moreover, the G residue in the subsequent codon at 146748, is actually a C residue, changing this codon from GGT to GCT. These two codons now encode Y and A residues, respectively, and replace UL101X and UL102 by a single ORF (Fig. 3.5 panels I and II). These results confirm the findings of Smith and Pari (1995), who detected these errors by sequencing

UL14 T T T A A A G I E L T A L G G A I I Y G G I T G - R I R R R S G C W P
 N N C S R R R H R V D G A R R R Y H L R R D Y W L T D P K I G L L A
 Q Q L Q P P A S S - R R - A A L S S T E G L L A D G S E D R V A G
 CAACAAC TGCAGCCGCCGCTCAGTTGACGGCGTAGCGCGCTATCATCTACGGAGGATTACTGGCTGACGGATCCGGAAGATCCGGTGTCTGGCC 21700
 L L Q L R R C R T S P A L R R - - R R L S - Q S V S G F I P N S A
 V V A A A P M S N V A S P P A I M - P P I V P Q R I R L D P Q Q G
 C C S C G G A D L Q R R - A A S D D V S P N S A S P D S S R T A P R
 R D R W P - P P S A T C C A T G V P N R T G V - T P R R K A R R R
 A G S V A L T S L C H L L C Y W C S E S Y R R L N T E E E S E A A
 R G I G G P D L P L P P A V L L V F R I V P A S E H R G G K R G G G
 GCGGATCGGTGCCCTGACCTCCCTCTGCCACTGTGTCTACTGGTTCGGA **ATCGTACCGCGCTCTGAACACCGG**AGGAAAGCGAGCGCGG 21800
 A P S D T A R V E R Q W R S H - Q H E S D Y R R R R F V S S L S A A S
 R S R H G Q V G E A V P Q A V P T G F R V P T Q V G L L F A L R R
 P I P P G S R G R G G A T S S T N R I T G A D S C R P P F R P P P
 R K L P R E K P L R - R R R R P S L R K S S G G S K R G E P - S G -
 E E T A A G E A S A V A A A A V S E E E Q R R E - T R R A **M K R M I**
 G N A C R G R S L C V S G G G R L - G R A A A G V N E E S H E A D D
 AGGAACTGCCGCGGAGAAGCCTCTGCGGTAGCGCGCGCGCTCTGAGGAAGAGCAGCGGGAGTAACGAGGAGGCCATGAACCGGATGAT 21900
 S V A A P S A E A T A A A A T E S S S C R R S Y V L L A M F R I I
 L F S G R S F G R R Y R R R G D R L F L L P P L L R P S S G H L P H N
 P F Q P L L R Q P L P P P R R Q L P L A A A P T F S L W S A S S
 F A V T A G K R N V R - R A P A S D A L R R R C A H L R Q R Y P T R
R S H G R K T E C Q M T S A G E R R A P P S V R P S A A A V P D A
 S Q S R Q E N G M S D D E R R R A T **R S A V G A P I C G C S G T R R**
 TCGCAGTCACGGCAGAAACGGAATGTCCAGTACGAGCGCGCGGAGCG **ACCGCTCCGCGCTCGGTGCGCCATCTGCGGCACCGGTACCCGACGCG** 22000
 R L - P L F V S H - I V L A P S R R A G G D T R G D A A A T G S A
 A T V A P F R F T L H R A G A L S A S R R R R H A W R R C R Y G V R
 E C D R C S F P I D S S S R R R A V R E A T P A G M Q P L P V R R P
 H G A N E R R D S D V G R P I A H S S **G T R R G S A N E T S A C T R**
 A R R Q R T P R L R P S H R P Q - R Y Q T R F G E - N V R L Y A
 G T A P T N A A T P T S V P S P T V A V P D A V R R M K R P P V R G
GCACGGCCCAACGACCGCGACTCCGACGTCGGTCCCATCGCCACAGTACGGTACCAGACGGGTTCGGCGAATGAACCTCCGCGCTGACGCGG 22100
 A R R W R V R S R R R D G W R G C Y R P V W V R N P S H F T R R Y A S
 C P A L S R R S E S T P G M A W L R P V L R P E A F S V D A Q V R
 V A G V F A A V G V D T G D G V T A T G S A T R R I F R G G T R P
 T D H Q K A D I G L W F M F L V F G L C S W L A M R Y R A Q - I L
 D R S P E G G H W A V V H V S G F W T V F V V G D A V S R T I N F E
 P I T R R R T L G C G S C F W F L D C V R G W R C G I A H N K F
 ACCGATCACCAGAAGCGGACATTGGGCTGTGGTTCATGTTCTGGTTTTGGACTGTGTCGTTGGTGGCGATCGGTATCGGCACAATAAATTT**CGA** 22200
 R D G S P P C Q A T T - T E P K Q V T N T T P S A T D R V I F K S
 V S - W F A S M P S H N M N R T K P S H E H N A I R Y R A C Y I K F
 G I V L L R V N P Q P E H K Q N K S Q T R P Q R H P I A C L L N Q
 N R C Q G T R V L Y F I G N I G G D K P V S D V Y P - S Y R G P R C
 S M S R N A C F V F Y I W R G - T C F G C L P L L L P G T S L
 I D V K E R V F C I L L W G I L A G I N L F R M F T L N L T G D L V
ATCGATGTCAAGGAACCGGTGTTTGTATTTTATTGGGAATATTGGCGGGATAAACCTGTTTCGGATGTTTACCCCTAATCTTACCGGGACCTCGTTG 22300
 D I D L F A H K T N - Q S Y Q R P Y V Q K P H K G K I K G P V E N
 R H - P V R T K Y I K I P P S I L G T E S T - G - D - R P G R Q
 I S T L S R T N Q I K N P I N A P I F R N R I N V R L R V P S R T T
 P L L L L P R T P G S M L T - V P T G V K S L G T Y E E R A Q R T
 S S P P S S S D T G L H A D V G T D W G Q K P G Y L - G A R T K D
 V L S S F F L G H R A P C - R R Y R L G S K A W V L M T R R S A H K G P
 TCCTCTCCTCTTCTCGGACACCGGCTCCATGCTGACGTAGTACCGACTGGGCTCAAAGCCTGGTACTTATGAGGAGCGCCACAAAGGAC 22400
 D E G G T E E S V P S W A S T P V S Q P - F G P Y K H P A R V F S R
 G R R R R G R V G P E M S V Y T G V P T L L R P V - S S R A C L V
 R E E K K R P C R A G H Q R L Y R S P D F A Q T S I L L A C L P G
 V R R R H G A S P R Y G T A G M G V F C S L L I C L F L V C C - P
 R - A P A W S V A E V R Y R W D G C F L F F A Y L P L P V L L T W
 UL16 L G A G M E R R R G T V P L G W V F F V L C L S A S S S C A V D L 22500
 GTTAGCGCGGCATGGAGCGTCGCCGAGGTACGGTACCGCTGGGATGGGTGTTTTTGTCTTGTCTATCTGCCTCTCCTCGTGTCTGTGACCTG
 - A G A H L T A S T R Y R Q S P H K K N K A - R G R G R T S N V Q G P
 T L R C R P A D G L Y P V A P I P T K Q E K S I Q R K R T H Q Q G P
 N P A P M S R R R P V T G S P H T N K T R Q K D A E E E H A T S R
 G - Q V L Q L D L P L E C D G V G L D P S W G N V D V T R D V Y F Y
 V A S P P T R P A A - M - R S W P R S I L G K R G R Y T G C V F L
 G S K S S N S T C R L N V T E L A S I H P G E T W T L H G M C I S
 GGTAGCAAGTCTCCAAC TCGACCTGCCGCTTGAATGTGACGGAGTGGCCCTCGATCCATCCTGGGAAACGTGGACGTTACACGGGATGTGTATTCTA 22600
 T A L G G V R G A A Q I H R L Q G R D M R P F R P R - V P H T N R
 Y C T R W S S R G S H S P T P R S G D Q P F T S T V R S T Y K -
 P L L D E L E V Q R K F T V S N A E I W G P S V H V N C P I H I E I
 L L L R E C D R G R D H R R G F Y L A A - R V C G - P V V V P E R
 S A T T R M - P R T R S S A W L L L G S I T S L W L T C G C T R T
 I C Y Y E N V T E D E I I G V A F T W Q H N E S V V D L W L Y Q N D
 TCTGCTACTACGAGAATGTACCGAGGACGAGATCATCGCGTGGCTTTTACTTGGCAGCATAACGAGTCTGTGGTTGACCTGTGGTTGACCGAAGCA 22700
 D A V V L I H G L V L D D A H S K S P L M Y L R H N V Q P Q V L V V
 R S S R S H S R P R S - R R P K - K A A Y R T Q P Q G T T T G S R
 UL15 Q - - S F T V S S S I M P T A K V Q C C L S D T T S R H N Y W F S

```

UL15A R P R N L T G Q A G T R A M R R V I R R H T L K T S D Q Q L P H A
UL14 E T E E S D R P G R D K G D E K S D P S S H L K D E - S T T T S C C
      D R G I - Q A R P G Q G R - E E - S V V T P - R R V I N N Y L M L
AGACCGAGGAATCTGACAGGCCAGGCCGGACAAGGGCGATGAGAAGAGTGATCCGTCGTACACCTTAAAGACGAGTGATCAACAACACTACCTCATGCTG 19800
      V S S D S L G P R S L P S S F L S G D D C R L S S H D V V V E H Q
      L G L F R V P W A P V L A I L L T I R R - V K F V L S - C S G - A A
      S R P I Q C A L G P C P R H S S H D T T V G - L R T I L L - R M S

      A Y P G S P Q D G E R R R S S G L L S S P P F S G S G G Y S I L D K
      I P R F T T R W R A P S Q F W P P I V S S L L R Q R R V L D S R -
      H T Q V H H K M A S A V A V L A S Y R L L P S P A A A G T R F - I
CATACCCAGGTTACCACAAGATGGCGAGCGCCGTCGCAGTTCTGGCCTCCTATCGTCTCCTCCCTTCTCCGGCAGCGCGGGTACTCGATTCTAGATAA 19900
      M G L N V V L H R A G D C N Q G G I T E E R R R C R P P T S S E L Y
      Y G P E G C S P S R R R L E P R R D D G G K E P L P P Y E I R S L
      C V W T - W L I A L A T A T R A E - R R R G K E G A A A P V E R N - I L

      K T I I I T T T A C G S G W W S W T P Q Q K A E A Y L W L M L L V
      E N N H H H D N G V R E R L V V L D A A A E G G S L P V A H V T G
      R K Q S S S R Q R R A G A A G G P G R R S R R R K P T C G S C Y W F
GAAAACAATCATCATCAGCAACGGCGTGCGGGAGCGGCTGGTGGTCTGGACCGCGCAGCAGAAGCGGAAGCCTACCTGTGGCTCATGTTACTGGTT 20000
      S F L - - S L P T R S R S T T R S A A A S P P L R G T A - T V P K
      F V I M M V V V A H P L P Q H D Q V G C C F A S A - R H S M N S T
      F C D D D R C R R A P A A P P G P R R L L L R F G V Q P E H - Q N

      L G L C S W L A A W I H A Y - K E I Y L L K P F C V Y G V V F Y F
      F G T V F V A G G V D S R I L K R D L P A E T L L C V W C G F L F R
      W D C V R G W R R G F T H I K K R F T C - N P F V C M V W F F I S
TTGGGACTGTGTTCTGGCTGGCGCGTGGATTACGCATATTAAGAGATTTACCTGCTGAAACCCTTTGTGTATGGTGTGGTTTTTATTTCG 20100
      P V T N T A P P T S E R M N F L S K G A S V R K H T H P K K N R
      K P S H E H S A A H I - A Y - F S I - R S F G K Q T Y P T T - K P
      Q S Q T R P Q R R P N V C I L F L N V Q Q F G K T H I T H N K I E

      G G L D R V D V G R D L K E T L S T - E E T A Q R I A A Q H G E G Q
      G S R P R - R R T - S Q G D A K Y L G R D R T A H R G S A W R R T
UL16 G V - T A L T - D V I S R R R R - V L R K R P H S A S R L S M A K D 20200
      GGGGTCTAGACCGGTTGACGTAGGACGTGATCTCAAGGAGACGCTAAGTACTTAGGAAGAGACCGCACAGCGCATCGGGCTCAGCATGGCGAAGGACA
      P D L G R Q R L V H D - P S A L Y K P L S R V A C R P E A H R L V
      P R S R T S T P R S R L S V S L V - S S V A C R M A A - C P S P C
      P T - V A N V Y S T I E L L R - T S L F L G C L A D R S L M A F S L

      R E R S H D G N R V D P F R R G M F R V D R P G K V L Q L H V Q R
      E G A F P - W E S R G S F S S G D V S G - - T W K S T P T A R T Y T
      R G S V P M M G I A W I L F V G G C F G L I D L E K Y S N C T Y N A
GAGGAGCGTTCATGATGGGAATCGCGTGGATCCTTTTCGTCCGGGGATGTTTCGGGTTGATAGACCTGAAAAGTACTCCAACCTGCACGTACAACGC 20300
      S P A N G H H S D R P D K E D P S T E P Q Y V Q F L V G V A R V V G
      L S R E W S P F R T S G K R R P I N R T S L G P F T S W S C T C R
      P L T G M I P I A H I R K T P P H K P N I S R S F Y E L Q V Y L A
  
```

II

Figure 3.2: Comparison of the region containing AD169 UL15 with the cognate CCMV sequence.

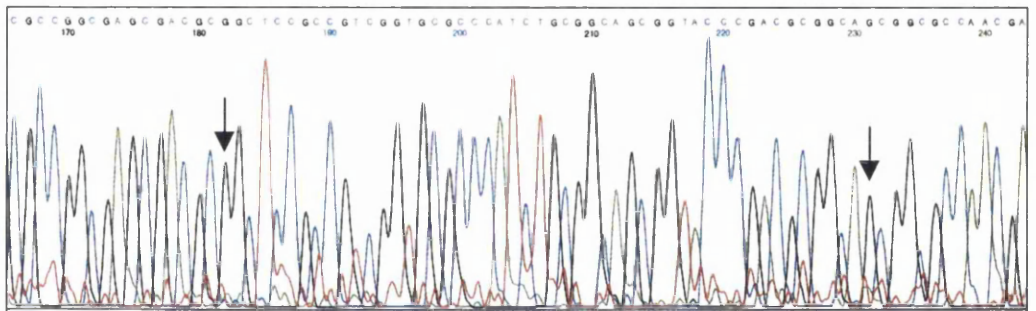
Panels I and II show six-frame translations of the relevant regions of the AD169 and CCMV sequences respectively, with nucleotide coordinates on the right. The locations of PCR primers for AD169 are indicated by arrows in panel I. AD169 UL15 is shown in blue in panel I, and CCMV UL15A is shown in red in panel II. Regions in AD169 that appear to be conserved in CCMV UL15A are shown in bold black font in panel I, with underlined amino acid residues. The region containing possible errors in the AD169 sequence is highlighted in yellow. ORF UL14 and UL16 are represented in green in both panels.

Figure 3.3: Comparison of the corrected sequence of the region containing AD169 UL15A with the cognate CCMV sequence.

Panel I shows the corrected sequence of AD169 and panel III shows the corresponding CCMV sequence, with nucleotide coordinates on the right. In each panel, UL15A is shown in red with conserved residues underlined and putative polyadenylation signals are underlined in green. The hydrophobic domain is boxed in each panel. The two additional G residues are highlighted in yellow in panel I. Panel II shows an electropherogram of the sequence of the relevant region of AD169 (upper strand) with the additional residues indicated by arrows.

UL15A M K R M I
 AGGAAACTGCCGCGGAGAAGCCTCTGCGGTAGCGGCGGCGCGCTCTCTGAGGAAGAGCAGCGGCGGGAGTAAACGAGGAGAGCCATGAAGCGGATGAT 21900
R S H G R K T E C Q M T S A G E R R R S A V G A P I C G S G T R R
 TCGCAGTCACGGCAGGAAAACGGAAATGTCAGATGACGAGCGCGCGGAGCGACCGGCTCCGCGGTCCGTGCGCCCATCTGCGGCAGCGGTACCCGACGC 22000
G S G A N E R R D S D V G P I A H S S G T R R G S A N E T S A C T
 GGCAAGCGGCGCCAACGAACGCCGCGACTCCGACGTCCGTCCCATCGCCACAGTAGCGGTACCAGACGCGGTTCCGGCGAATGAAACGTCGCCCTGTACGC 22100
R T D H Q K A D I G L W F M F L V F G L C S W L A M R Y R A Q -
 GGACCGATCACCAGAAGGCGGACATTGGGCTGTGGTTCATGTTCTGGTTTTGGACTGTGTTTCGTGGTTGGCGATGCGGTATCGCGCACATAAATTTT 22200
 GAATCGATGTCAAGGAACCGGTGTTTTGTATTTTATGGGAATATGGCGGGGATAAACCTGTTTCGGATGTTTACCCTTAATCTTACCGGGGACCTCGT 22300
 TGTCTCTCCTCCTTCTCCTCGGACACCGGGCTCCATGCTGACGTAGGTACCGACTGGGGTCAAAGCCTGGGTACTTATGAGGAGCGCGCAAAAGGA 22400
 UL16 M E R R R G T V P L G W V F F V L C L S A S S S C A V D
 CCGTTAGGCGCCGGCATGGAGCGTCGCCGAGGTACGGTACCCTGGGATGGGTGTTTTTGTCTTTGCTTATCTGCCTCTTCTCGTGTGCTGTTGACC 22500

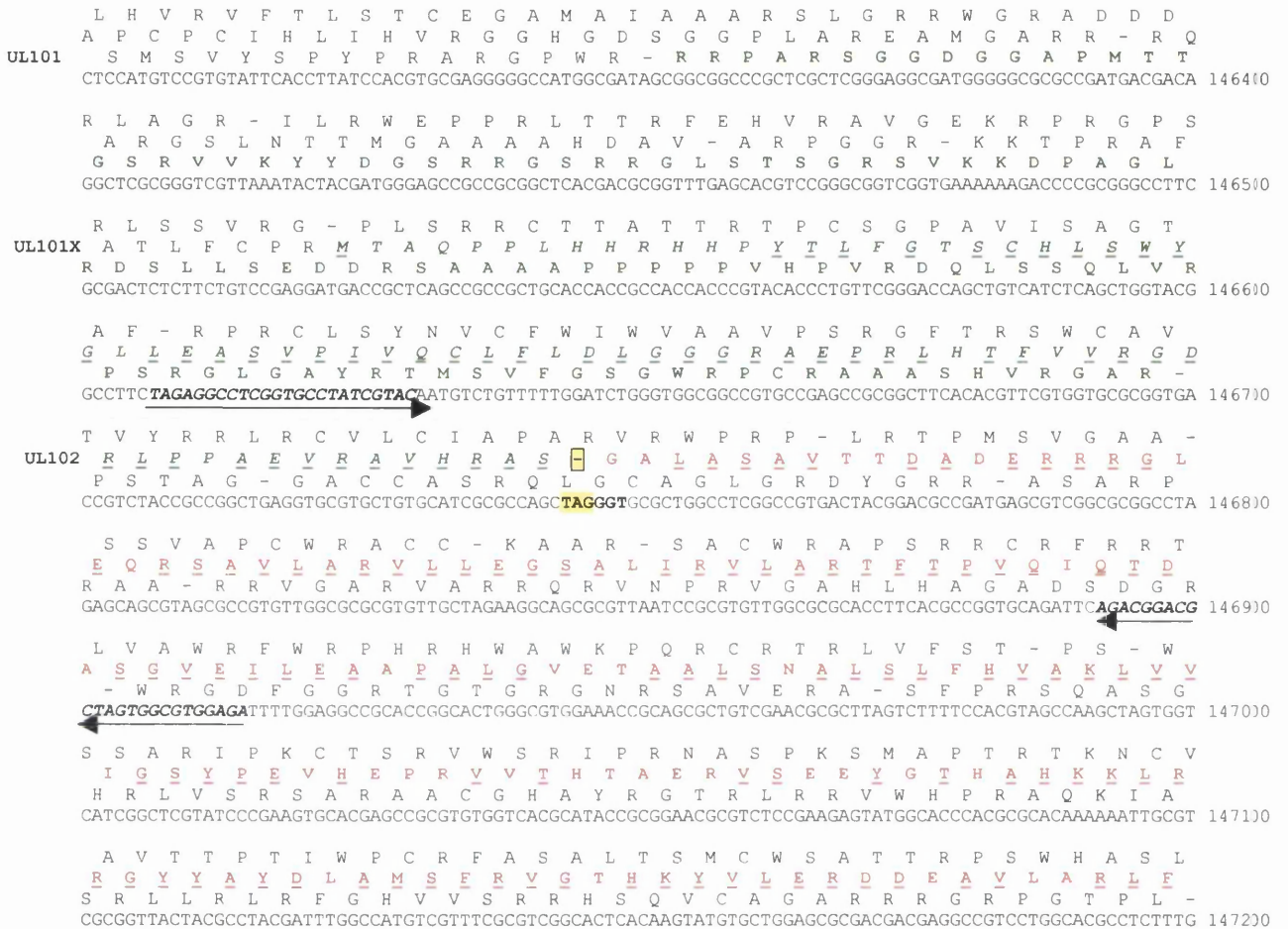
I



II

UL15A M R R V I R R H T L K T S D Q Q L P H A
 AGACCGAGGAATCTGACAGGCCAGGCCGGACAAGGGCGATGAGAAGAGTGATCCGTCGTCACACCTTAAAGACGAGTGATCAACAACCTACCTCATGCTG 19800
A Y P G S P Q D G E R R R S S G L L S S P P F S G S G G Y S I L D K
 CATACCCAGGTTCAACACAAGATGGCGAGCGCCGTCGAGTTCTGGCCTCCTATCGTCTCCTCCCTTCTCCGGCAGCGCGGGTACTCGATTCTAGATAA 19900
K T I I I T T T A C G S G W W S W T P Q Q K A E A Y L W L M L L V
 GAAAAAATCATCATCACGACAACGGCGTGGGGAGCGGCTGGTGGTCTGGACCGCCGAGCAGAAGGCGGAAGCCTACCTGTGGCTCATGTTACTGGTT 20000
L G L C S W L A A W I H A Y -
 TTGGGACTGTGTTTCGTGGCTGGCGGCGTGATTACGCATATTAAGAAAGAGATTACCTGCTGAAACCCCTTTTGTGTGTATGGTGTGGTTTTTTATTTTCG 20100
 UL16 M A K D
 GGGGTCTAGACCGGTTGACGTAGGACGTGATCTCAAGGAGACGCTAAGTACTTAGGAAGAGACCCACAGCGCATCGCGGCTCAGCATGGCGAAGGACA 20200

III



I

Figure 3.4: Comparison of the region containing AD169 UL102 with the cognate CCMV sequence.

Panels I and II show three-frame translations (upper strand) of the relevant regions of the AD169 and CCMV sequences, respectively, with nucleotide coordinates on the right. The locations of PCR primers for AD169 are indicated by arrows in panel I. UL102 is shown in red in panels I and II. Regions in AD169 that appear to be conserved in CCMV UL102 are shown in italic green font in panel I, with conserved amino acid residues underlined. The region containing possible errors in AD169 sequence is highlighted in yellow. UL101 and UL101X are shown in green in panel I.

G G R C V L M T K S P R V V K Y Y Y G G P C R G C A S S R A F P E
 G W P V R A D D E K P A G R - I L L W R P M P R L R L L P R L P R
 G V A G A C - - R K A R G S L N T T M E A H A A A A P P P A P S P R
 GGGTGGCCGGTGCCTGATGACGAAAAGCCCGGGTTCGTTAAATACTACTATGGAGGCCATGCCGCGGCTGCCCTCCTCCCGCCTTCCCCGAG 148700

D A V - T R R R D R P K D R A A P V G P S Q R S S R G R A A K N G
 G R G L N A S T R P A K R P R G A R R A L T A L V A R P G G E K R P
 T R F E R V D A T G Q K T A R R P S G P H S A R R A A G R R K T A
 GACGCGGTTTGAACGCGTCGACGCGACCGGCCAAAAGACCGCGGGCGCCCGTCGGCCCTCACAGCGCTCGTCGCGCGGCGGGGGCGGAAAACGGCC 148800

P E L G G D G R Q P L H A R R G Q L P P Q L V R H P R G L R A H R A
 R A R R G W P P T T T R S S G P A A T S A G T A S S R P P C P S C
 P S S A G M A A N H Y T L V G A S C H L S W Y G I L E A S V P I V
 CCGAGCTCGGCGGGATGGCGCAACCACTACACGCTCGTCGGGGCCAGCTGCCACCTCAGCTGGTACGGCATCCTCGAGGCTCCGTGCCCATCGTGC 148900

V P L H G S G C P R R S R R G A A A A D V H R A R R P P A L G R G
 S A S S W I W V P S A V A P R S R G C G R S S C A A T A C P R P R
 Q C L F M D L G A L G C R A E E P R L R T F I V R G D R L P S A E V
 AGTGCCTCTCATGGATCTGGTGCCCTCGGCGGTCGCGCGGAGGAGCGCGGCTGCGGACGTTTCATCGTCGCGGCGACCGCCTGCCCTCGGCGGAGGT 149000

A C R A P R Q L R Q P G L G R G R R R G A P P R H R A P Q R R R A G
 C V P C T A P A T P A W P R P W P T P R S A A A A S S T A A P C W R
 R A V H R A S Y A S L A S A V A D A E E R R R G I E H R S A V L A
 GCGTCCGTGCACCGGCCAGCTACGCCAGCCTGGCCTCGGCGGTCGGCGAGCGCCGAGGAGCGCCGCCCGGCGATCGAGCACCAGCGCCGCTGCTGGCG 149100

A C A A G G Q R H H P R A G A H L H A R P D P D R P Q R R G D S R G
 V C C W R A A P S S A C W R A P S R P S R S R P T P A A W R F S R
 R V L L E G S A I I R V L A R T F T P V Q I Q T D P S G V E I L E
 CGTGTGCTGCTGGAGGCGAGCCCATCATCGCGTGTGGCGCGCACCTTCAGCCCGTCCAGATCCAGACCACCCAGCGCGTGGAGATTCTCGAGG 149200

R A R A G L R R G R A L A R P Q P L P R R Q A G G H R L L P G D P
 P C P R W A S T R P R S R T P S A S T T S P S W W S S A P T R R S
 A V P A L G F D A A A L S H A L S L Y H V A K L V V I G S Y P E I H
 CCGTCCCGCGCTGGGCTTCGACGCGCGCGCTCTCGCACGCCCTCAGCCTCTACCAGTCGCCAAGCTGGTGGTTCATCGGCTCCTACCCGGAGATCCA 149300

R R R G H L V V R L G A R R H R R P R G L R H P G P Q E A A A R L
 T T P W P P R R P S R R A S P P A A T R T T P P R P T R S C G A A T
 D A V A T S S S V S A R V A T G G H E D Y A T Q A H K K L R R G Y
 CGACGCGTGGCCACCTCGTCTCGGCGCGCTCGCCACCGCGGCCACGAGGACTACGCCACCCAGGCCACAAGAAGCTGGCGCGGCTAC 149400

L R L R R G H V V P R R R P Q V R A G A R R R G R R G L T P L R G A
 T P T T W P C R S A S A P T S T C W S A T T R P P W S H A S S R C
 Y A Y D V A M S F R V G A H K Y V L E R D D A A A V V S R L F E V
 TACGCTACGAGTGGCCATGTCTCCGCGTGGCGGCCACAAGTACGTGCTGGAGCGGACGCGCGGCGCGGCTGCTCAGCCTCTTCGAGGTGC 149500

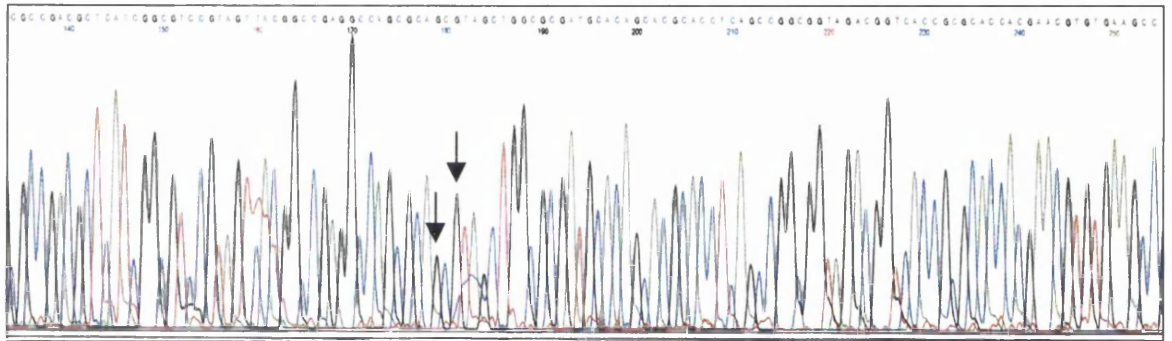
II

Figure 3.5 Comparison of the corrected sequence of the region containing AD169 UL102 with the cognate CCMV sequence.

Panel I shows the corrected sequence of AD169 and panel III shows the corresponding CCMV sequence, with nucleotide coordinates on the right. In each panel, UL102 is shown in red and conserved residues are underlined. The two G residues replaced by C residues are highlighted in yellow in panel I. Panel II shows an electrophoregram of the sequence of the relevant region of AD169 (upper strand) with the altered residues indicated by arrows.

UL102 M T A Q P P L H H R H H P Y T L F G T S C H L S W Y
 GCGACTCTCTTCTGTCCGAGGATGACCGCTCAGCCGCGCTGCACCACCGCCACCACCGGTACACCCTGTTCCGGGACCAGCTGTCATCTCAGCTGGTACG 146600
 G L L E A S V P I V Q C L F L D L G G G R A E P R L H T F V V R G D
 GCCTTCTAGAGGCTCGGTGCCTATCGTACAATGCTGTTTTGGATCTGGGTGGCGGCCGTGCCGAGCCGCGGCTTACACAGTTCGTGGTGC GCGGGA 146700
 R L P P A E V R A V H R A S Y A A L A S A V T T D A D E R R R G L
 CCGTCTACGCCGGCTGAGGTGCGTGTGTCATCGCGCCAGCTA G G T GCGTGGCCTCGGCCGTGACTACGGACGCCGATGAGCGTCGGCGCGGCTA 146800
 E Q R S A V L A R V L L E G S A L I R V L A R T E T P V Q I Q T D A
 GAGCAGCGTAGCGCCGTGTTGGCGCGCGTGTGCTAGAAGCAGCGGTTAATCCGCGTGTGGCGCGCACCTTACGCCGGTGCAGATT CAGACGGACG 146900
 S G V E I L E A A P A L G V E T A A L S N A L S L F H V A K L V V
 CTAGTGGCGTGGAGATTTTGGAGCCGACCCGGCACTGGGCGTGGAAACCGCAGCGTGTGCGAACCGCGCTTAGTCTTTCCAGCTAGCCAAGCTAGTGGT 147000
 I G S Y P E V H E P R V V T H T A E R V S E E Y G T H A H K K L R
 CATCGGCTCGTATCCCGAAGTGACAGGCGCGGTGGTCACGCATACCGCGGAACCGGCTCCGAAGAGTATGGCACCCACGCGCACAAAAAATGCGT 147100
 R G Y Y A Y D L A M S F R V G T H K Y V L E R D D E A V L A R L F
 CGCGGTTACTACGCCTACGATTTGGCCATGTCGTTTCGCGTGGCACTACAAGTATGTGCTGGAGCGCGACGACGAGGCCGCTCCTGGCAGCGCTCTTTG 147200

I



II

UL102 M A A N H Y T L V G A S C H L S W Y G I L E A S V P I V
 CCGAGCTCGGGGGATGGCCCAACCACTACACGCTCGTCGGGGCCAGCTGCCACCTCAGCTGGTACGGCATCCTCGAGGCCTCCGTGCCCATCGTGC 148900
 Q C L F M D L G A L G G R A E E P R L R T F I V R G D R L P S A E V
 AGTGCTCTTCATGGATCTGGGTGCCCTCGGGGTGCGCGCGAGGAGCGCGGCTGCGGACGTTTCATCGTGC GCGGCGACCGCTGCCTCGGCCGAGGT 149000
 R A V H R A S Y A S L A S A V A D A E E R R R G I E H R S A V L A
 GCGTGCCTGCACCGCCAGCTACGCCAGCTGGCCTCGGCCGTGGCCGACGCCGAGGAGCGCCCGCGGCATCGAGCACCGCAGCGCGTGTGGCG 149100
 R V L L E G S A I I R V L A R T E T P V Q I Q T D P S G V E I L E
 CGTGTGCTGCTGGAGGCGAGCCCATCATCGCGTGTGGCGCGCACCTTACGCCCGTCCAGATCCAGACCACCCAGCGCGTGGAGATTCTCGAGG 149200
 A V P A L G F D A A A L S H A L S L Y H V A K L V V I G S Y P E I H
 CCGTGCCCGCGTGGCTTCGACGCGCGCGCTCTCGCAGCCCTCAGCCTTACCAGCTCGCAAGTGGTGGTCATCGGCTCCTACC GCGGATCCA 149300
 D A V A T S S S V S A R V A T G G H E D Y A T Q A H K K L R R G Y
 CGACGCGTGGCCACCTCGTCTCGTCCGCTCTGGCGCGCGTGCACCAGCGCGCCAGGAGTACGCCACCCAGGCCACAAAGAGTGC GCGCGCGGTAC 149400
 Y A Y D V A M S F R V G A H K Y V L E R D D A A A V V S R L F E V
 TACGCCCTACGAGTGGCCATGTCGTTCCGCGTGGCGCCCAAGTACGTGCTGGAGCGCGACGACGCGGCCCGTGGTCTCAGCCTTTCGAGGTGC 149500

III

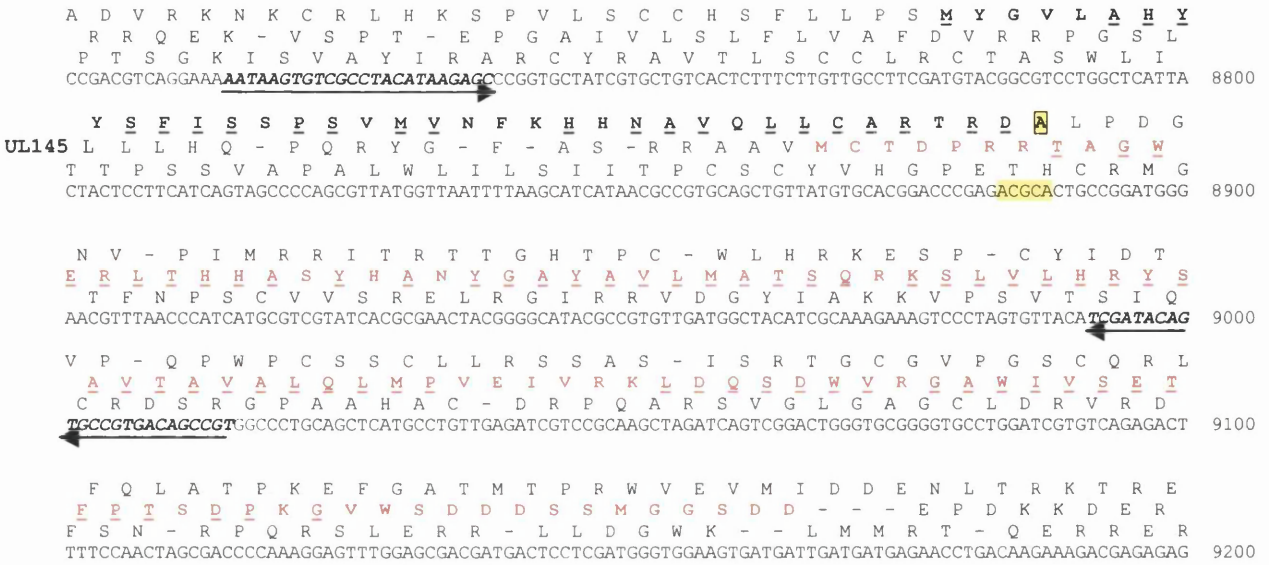
this region of the Towne and AD169 genomes. The predicted protein-coding region for the amended AD169 UL102 contains 870 amino acids. Several other nucleotide substitutions were reported in this region of Towne by Smith and Pari (1995), but none were found in the AD169 DNA sequence in the present study.

3.1.3 UL145

UL145 is present in Toledo but absent from AD169. Fig. 3.6 shows three-frame translations of the relevant regions of the Toledo and CCMV genomes. UL145 as defined in Toledo (Fig. 3.6 panel I) is conserved in CCMV (Fig. 3.6 panel II), but appears to be disrupted in the region of nucleotide 8884-8889. This indicates the potential presence of a frameshift error in the Toledo sequence. Sequence analysis of a PCR product derived from this region of the Toledo genome demonstrated the presence of one extra G residue after nucleotide 8884. The electropherograms are shown in Fig. 3.7 (panel II) and the additional nucleotide is highlighted in yellow in Fig. 3.7 (panel I). This correction extends Toledo UL145 at its 5'-end so that it corresponds to CCMV UL145. The correct ATG is located at position 8777 instead of 8866 as published. These two ORFs are shown in Fig. 3.7 (panels I and III).

3.1.4 US22

AD169 US22 is well conserved in CCMV (Fig. 3.8 panels I and II). However, a region near the 5'-end is conserved in another frame in CCMV. This indicates a possible frameshift error in the sequence of AD169 US22 as marked in yellow in Fig. 3.8 (panel I). Sequence analysis of a PCR product demonstrated the presence of one extra G residue after nucleotide 211535 (Fig. 3.9 panels I and II). This correction extends the protein-coding region of AD169 US22 at its 5'-end and introduces an additional 45 amino acid residues at the N terminus of the protein. The first ATG codon in the US22 is now at 211601 instead of 211462 as originally described by Chee *et al.* (1990). The transcript mapping data for the US22 mRNA (section 5.1) also correlate with the correction made in this region.



I

Figure 3.6 Comparison of the region containing Toledo UL145 with the cognate CCMV sequence.

Panels I and II show three-frame translations of the relevant regions of the Toledo and CCMV sequences, respectively, with nucleotide coordinates on the right. UL145 is oriented left to right in Toledo (panel I) and right to left in CCMV (panel II). In each panel, UL145 is shown in red, and conserved residues are underlined. The locations of PCR primers for Toledo are indicated by arrows in panel I. Regions in Toledo that appear to be conserved in CCMV UL145 are shown in bold black font in panel I, with conserved amino acid residues underlined. The region containing possible errors in the Toledo sequence is highlighted in yellow.

TAAAACGGGAACCACACTTTCACACCAGCTCTACCAACACGGCTCGGACCTTCTCTTTTGCAGACCGCTCTCAAGCGTCACCTCCACCCACCGCGGATT 183900
 F R S G C K - V L E V L V A R V K R K Q L G S E L T V E V W R P N
 - F P F W V K V G A R G V R S P G K E K A S R E - A D G G G V A S E
 L V P V V S E C W S - W C P E S R E R K C V A R L R - R W G G R I

CCCCCTCGCGTCCCTCGTGTCTTATCTCCCTCGTCTTTTCCCCTCGTCCGCGTCTACTGCTGCGACGGTCCCAGAAATCCGTCCGGATCGCAGGT 184000
 G R A D R T D R I E R T T K G R T A A A V A A V T G F D T R I A P
 G E R G E H R K D G E D N E G E D G S S S S R R D R F G D P D C T
 G G R T G R T E - R G R R K G G R R R Q - Q Q S P G S I R G S R L D

CGAAATGTCTCCGACACGATCCACGCGCTTTCACCCAGTCCGACTGGTCCAGTCTCGCCAACACGTTACTCGGCATCAGCTGCAGAGCCACGGCCGTC 184100
 R F H R R C S G R A K - G T R S T W D R W C T V R C - S C L W P R -
 P F T E S V I W A G K V W D S Q D L R A L V N S P M L Q L A V A T
 S I D G V R D V R R E G L G V P G T E G V R - E A D A A S G R G D

ACGGCGCTGTACCGGTGCAACACCAGGACTTGGTCTGCGACGTAGCCATCAGTACGGCGTAGGCCCGTAGTCCGCTGATAGGCCGCTGGTGGGTTA 184200
 P A T G T C C W P S P R R R L W - Y P T P G T T R R I P R T T P -
 V A S Y R H L V L S K T Q S T A M L V A Y A G Y D A Q Y A A H H T L
 R R Q V P A V G P V Q D A V Y G D T R R L G R L G G S L G R P P N

GACGTTCCACCCGAGGTGCGTCCGCGAGCCCTCGCGCACAGCAGCTGGACGGCTTCTGATGTCTGTAGTGGACCATGGCGCTCGGACCGCTGATGAA 184300
 V N G G R P A T A L G R A C C S S P T R I D T T S W P A R V A S S
 R E W G S T G D R A R A C L L Q V A N Q H R Y H V M A S P G S I F
 S T G V G L H R R S G E R V A A P R R E S T Q L P G H R E S R Q H V

CGAGTAATAATGCGCGCCGAGACGGCGTGTGATCGTGCATGGGCGAGCGACGACAACACTACCTGCGGGCTCGTCCATGCTTCGCTCTCACCT 184400
 R T I I R P R S P T N I T C P R A V V V S G A A D D E M S R G R V K
 U1145 S Y Y H A A S V A H Q D H M P S R R C C - R R R R R G G H K T R E G
 L L L A R G L R R T S R A H A L S S L V V Q P T T R W A E D E - R

II

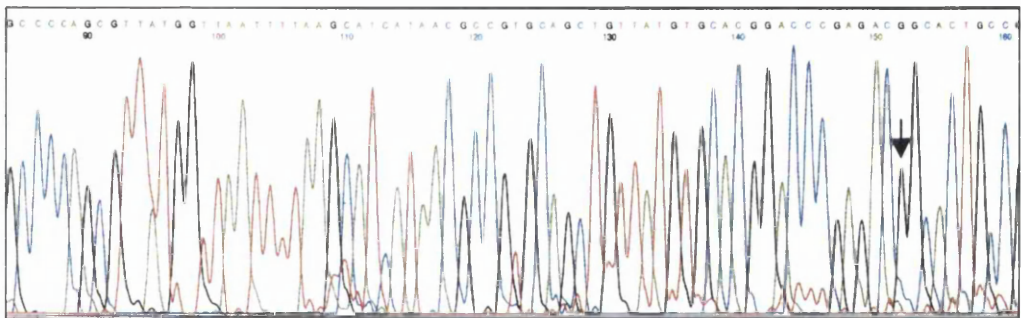
Figure 3.7 Comparison of the corrected sequence of the region containing Toledo UL145 with the cognate CCMV sequence.

Panel I shows the corrected sequence of Toledo and panel III shows the corresponding CCMV sequence, with nucleotide coordinates on the right. UL145 is oriented left to right in Toledo (panel I) and right to left in CCMV (panel III). In each panel, UL145 is shown in red, and conserved residues are underlined. The additional G residue is highlighted in yellow in panel I. Panel II shows an electrophoregram of the sequence of the relevant region of Toledo (upper strand) with the additional residues indicated by arrows.

UL145

CC GACGT CAGG AAAATAAGTGT CGCTACATAAGAGCCCGGTGCTATCGTGTCTACTCTTTCTTGTTCCTTCGATGTACGGCGTCTGGCTCATTAC 8800
 M Y G V L A H Y
 Y S F I S S P S V M V N F K H H N A V Q L L C A R T R D A T A G W 8900
 TACTCCTTCATCAGTAGCCCCAGCGTTATGGTTAATTTTAAGCATCATAACGCCGTGCAGCTGTTATGTGCACGGACCCGAGACG CACTGCCGGATGGG
 E R L T H H A S Y H A N Y G A Y A V L M A T S Q R K S L V L H R Y S 9000
 AACGTTTAACCCATCATGCGTCGTATCAGCGAACTACGGGCATACGCCGTGTTGATGGCTACATCGCAAAGAAAGTCCCTAGTGTACATCGATACAG
 A V T A V A L Q L M P V E I V R K L D Q S D W V R G A W I V S E T 9100
 TGCCGTGACAGCCGTGGCCCTGCAGCTCATGCCTGTTGAGATCGTCCGAAGCTAGATCAGTCGGACTGGGTGCGGGGTGCCTGGATCGTGTGAGAGACT
 F P T S D P K G V W S D D D S S M G G S D D - 9200
 TTCCAAC TAGCGACCCCAAAGGAGTTTGAGCGACGATGACTCCTCGATGGGTGGAAGTGATGATTGATGATGAGAACCTGACAAGAAAGACGAGAGAG

I



II

TAAAACGGGAACCACTTTACACCAGCTCTACCAACACGGCTCGGACCTTTCTCTTTTGCAGACCGCTCTCAAGCGTCACTCCACCCACCGCGGATT 183900
 - A D G G G V A S E
 CCCCTCGCGTCCCTCGTGTCTTTATCTCCCTCGTCGTTTTCCCTCGTCGCCGCTGCTACTGTGCGACGGTCCCGAAATCCGTCCGGATCGCAGGT 184000
 G E R G E H R K D G E D N E G E D G S S S S R R D R F G D P D C T
 CGGAAATGTCTCCGACACGATCCACGCGCCTTTACCCAGTCCGACTGGTCCAGTCTCGCCAACAGTACTCGGCATCAGCTGCAGAGCCACGCGCCGTC 184100
 P F T E S V I W A G K V W D S Q D L R A L V N S P M L Q L A V A T
 ACGGCGCTGTACCGGTGCAACACCAGGGACTTGGTCTGCGACGTAGCCATCAGTACGGCGTAGGCCCCGTAGTCCGCTGATAGGCCGCGTGGTGGGTTA 184200
 V A S Y R H L V L S K T Q S T A M L V A Y A G Y D A Q Y A A H H T L
 GACGTTCCCACCCGAGGTGCCGTGCGGAGCCCTCGCGCACAGCAGCTGGACGGCGTTCTGATGTCTGTAGTGGACCATGGCGCTCGGACCGCTGATGAA 184300
 R E W G S T G D R A R A C L L Q V A N Q H R Y H V M A S P G S I F
 CGAGTAATAATGCGCGCCGAGACGGCGTGTGATCGTGCATGGGCGAGCGACACAACACTACCTGCGGCGTGTCTCCATGCTTCGTCTCTCACCT 184400
 S Y Y H A A S V A H Q D H M

UL145

III



I

Figure 3.8: Comparison of the region containing AD169 US22 with the cognate CCMV sequence.

Panels I and II show three-frame translations of the relevant regions of the AD169 and CCMV sequences, respectively, with nucleotide coordinates on the right. The locations of PCR primers are indicated by arrows in panel I. Part of US22 is shown in red in each panel. In panel I, the first ATG codon in US22 as originally defined is shown in a box, and regions in AD169 that appear to be conserved in CCMV US22 are shown in bold black font, with conserved amino acid residues underlined. The region containing possible errors in the AD169 sequence is highlighted in yellow.

GTTCTTGGCGAAGACCTTGAGGTCGCTGGCTACCTCGTACAGCCCGTCCGCTCTCCAGTCGTAGACGTAGACGGTGCCGTGCTTACCCATCATCAGCACG 223200
 E Q R L G Q P R Q S G R V A R G D G L R L R L R H R A - G D D A R
 T R P S S R S T A P - R T C G T R R G T T S T S P A T S V W - - C A
N K A F V K L D S A V E Y L G D T E W D Y V Y V T G H K G M M L V

CAGGGCACCTCCTGCGCCCACTTGGTATGGCGGTGGGGTTCGTCGCCCGGTGCACGGCCAGCACGCCGACGGCCTCCAGGTACGTACAGCACACA 223300
 L A G G A G V Q Y P P P P R R G G T A R G A R R R G G P - R - L V
 P C R R R G S P I A T P T T T R R D C P W C A S P R W T M T V A C
C P V E Q A W K T H R H P D D D A P R V A L V G V A E L D D D C C L

GATACTCGGAGGCCAGCGAACGTGCGTAGATCTGCGGGATGGCGCCGCTCTCGCGCAGCACCCAGGAGCCAGTTGGCCGGGTACGCAGGGCCACCGTGGT 223400
 S V R L G A F T R L D A P H R G D R A A G P A L Q G P - A P G G H H
 I S P P W R V H T S R R S P P R R A C C W S G T P R T V C P W R P
Y E S A L S R A Y I Q P I A A T E R L V L L W N A P N R L A V T T

GCCCCGTGCTGGCGTACGTAGCTCCGACGCGGTGGGGTTCGTCGCGCAGGAGCCCTGCAGCAGCAGAGCGTGGCGTGGTATAGTGGTGTG 223500
 G P A P T R L E A A R R P R I S A P L A G A A A L A H A P L H H Q
 A G T S A Y T A G C R T P T T N V C T S G R C C R S R A S T S P T A
G R H Q R V Y S R L A D P D Y Q R L Y L G Q L L S L T R Q Y I T N

CGAAACTGAAAGACACCATAGTCGCAGCGGATGGCGTCTCGCAGCGCTCGTCGCGCTGCCGAGATAGGTGCCCGAGGCTCGGCGGCGGCCTTGGTCA 223600
 S V S L C W L R L P H R G R L A R R A A A S L H G L G R R R G Q D
 F S F S V M T A A S P T R A A S T A S G C I P A G P R P P P R P -
R F Q F V G Y D C R I A D E C R E D R Q R L Y T C W A E A A A K T L

GCAGCGACATCGCGCGCCCGACGCGGCGGTGGCGCCCGGCGCGGCGTCAAAGCTTAATATAGGGAGCGGTCCCTCTCGAGTCTGGGCGCGCGG 223700
 A A V D R R G S A P R T A G R R P T L A - Y L S R D R E L R P R R A
 C R C R P A G V R A T H R G P A A D F S L I P L P G E R T Q A A A
L S M A A G R R P G H P A G A R R - L K I Y P A T G R S D P G G R

US22

II

Figure 3.9 Comparison of the corrected sequence of the region containing AD169 US22 with the cognate CCMV sequence.

Panel I shows the corrected sequence of AD169 and panel III shows the corresponding CCMV sequence, with nucleotide coordinates on the right. In each panel, US22 is shown in red, and conserved residues are underlined. The inserted G residue is highlighted in yellow in panel I. Panel II shows an electrophoregram of the sequence of the relevant region of AD169 (lower strand) with the inserted residue indicated by an arrow.

AAAGGCTTTGATATCACTGGCCACCTCGTAGAGCCCGTCGGTCTCCAGTCGTAGACGTAGACGGTGCCGTAATGACTTAGCATGAGCACGCAGGGCAGT 211200
F A K I D S A V E Y L G D T E W D Y V Y V T G Y H S L M L V C P L

TCCTGCGCCTGCTTGGTGTTCGTGTTAGATCGCTGTCGGGTGGACGCACGGCTAGTACACCGACGGCTTCCAGGGTGTTCATCGCAGCAGAGATAGTCGG 211300
E Q A Q K T N R T L D S D P P R V A L V G V A E L T D D C C L Y D A

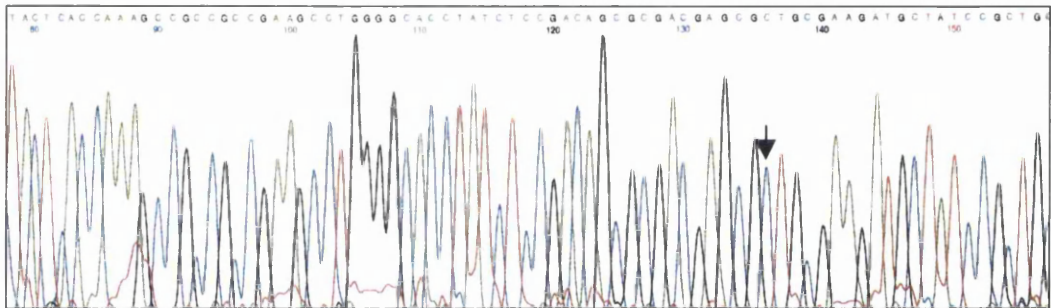
CGGCCAGAGAACGTGCGTAAATCTGCGGGATGGCGGCTGTTGCGGCATCACTAGGAACCAAGTTGGCGGGGTTGCGCAGTGTACGGTGGTTCCTTGGTG 211400
A L S R A Y I Q P I A A Q E R M V L F W N A P N R L A V T T G Q H

CGGTTGCACGTAGGTTCTCAGCGCCGAGGATCGTACTGGCGCAGATAGAGGCCCTTGACGATCGATAACGCTCTTTGAAAGACGGTGTTCCTAAATTGA 211500
R Q V Y T R L A P P D Y Q R L Y L G Q L M S L T K Q F V T N R E Q

US22 AAAACGCCGTAGTCGCAGCGGATAGCATCTTCGCAGCGCTCGTCGCGCTGTCGGAGATAGGTGCCCCAGGCTTCGGCGGGCGGCTTGGTGAGTAGGGACA 211600
F V G Y D C R I A D E C R E D R Q R L Y T G W A E A A A K T L L S M

TGCCGGCGAGCCGCTCTCGACAGCGAGTCGATAAAGCGCGCTGCGCGAAAGCTTAATATAGGAGCAGCGTCAGACGAATCGCGGCTGGTGGCCCGGGG 211700

I



II

GTTCTTGCGAAGACCTTGAGTGCCTGGCTACCTCGTACAGCCCGTCGGTCTCCAGTCGTAGACGTAGACGGTGCCGTGCTTACCCATCATCAGCAG 223200
N K A F V K L D S A V E Y L G D T E W D Y V Y V T G H K G M M L V

CAGGGACCTCCTGCGCCCACTTGGTATGGCGGTGGGGTTCGTGTCGCCCGTGCACGGCCAGCACGCCGACGGCTCCAGGTTCATCGTCACAGCACA 223300
C P V E Q A W K T H R H P D D D A P R V A L V G V A E L D D D C C L

GATACTCGGAGGCCAGCGAACGTGCGTAGATCTGCGGGATGGCGGCGTCTCGCGCAGCACCAGGAGCCAGTTGGCGGGTTACGCAGGGCCACCGTGGT 223400
Y E S A L S R A Y I Q P I A A T E R L V L L W N A P N R L A V T T

GCCCCGTGCTGGCGTACGTAGCTCCGACGCGGTGGGGTTCGTATTGACGACGAGTAGACCCCTGCAGCAGCAGAGCGTGGCTGGTAGATGGTGTG 223500
G R H Q R V Y S R L A D P D Y Q R L Y L G Q L L S L T R Q Y I T N

CGAAACTGAAAGACACCATAGTCGCAGCGGATGGCGTCTCGCAGCGCTCGTCGCGTGCAGCAGATAGGTGCCCCAGGCTTCGGCGGGCCTTGGTCA 223600
R F Q F V G Y D C R I A D E C R E D R Q R L Y T G W A E A A A K T L

US22 GCAGCGACATCGCGGGCCCCGACGCGGGCCGTGTGGCGCCCCGGCGCGCGCTCAAAGCTTAATATAGGAGCGGTCCTCCTCGAGTCTGGGCGCCGCG 223700
L S M

III

3.1.5 UL131A

Fig. 3.10 shows sequences of the relevant regions of the AD169 and CCMV genomes. AD169 UL131 (Fig. 3.10 panel I) is not conserved in CCMV (Fig. 3.10 panel IV). However, a predicted spliced CCMV gene (UL131A) (Fig. 3.10 panel IV) appeared to be conserved in AD169, including the acceptor and donor sites, except that the first exon is in two different reading frames (Fig. 3.10 panel I). This analysis indicated the potential presence of a frameshift error in the AD169 sequence within the region marked in yellow in Fig. 3.10 (panel I). However, sequence analysis of a PCR product derived from this region of the AD169 genome using the primers shown in Fig. 3.10 (panel I) failed to detect errors in this region (Fig. 3.10 panel II). The AD169 sequence is therefore correct. However, the spliced structure of AD169 UL131A was confirmed by transcript mapping (section 8.2.1)

To confirm whether AD169 UL131A is naturally frameshifted, PCR products derived from the corresponding regions of Toledo and a clinical sample (GW) were analysed using the same primers. As can be seen in the electrophoregram shown in Fig. 3.10 (panel III), one T residue is absent from the relevant region in GW in comparison with AD169. Toledo UL131A was also found to have one fewer T residue in this region (data not shown). These corrections disrupt UL131 and introduce a novel ORF corresponding to the first exon of CCMV UL131A. Consequently, it appears that AD169 UL131A is naturally frameshifted, thus failing to encode a functional protein since it lacks an initiation codon. A version of the sequence of AD169 UL131A in which the first exon is in frame, by removal of a T residue, in comparison with CCMV UL131A is shown in Fig. 3.11. In comparison with the AD169 sequence, two other nucleotide substitutions were found in GW which are indicated by green arrows in Fig. 3.10 panel III.

CAGCATATTATTTCCCGTGACGCAGGCTAGTTGGCAAAGAGCCGCA**CGCTGAACTCGAGGCTCCGGGCGT**GTGGCGCCAGCGAACCCGGCGGCGTTGAACG 176400
 A Y - K G H R L S T P L S G C A S S S A G P H R W R V P P T S R
 C C I I E R S A P - N A F L R V S F E L S R A H P A L S G A A N F T
 L M N N G T V C A L Q C L A A R Q V R P E P R T A G A F R R R Q V

TGGTCCTTTTGTGGTGCCGCCGACGGTTCTGACGTCTAAAGTCGCTGATGAGCAACGACACCTCGGTACGTTGATTctgcaagcacaggttccaaa 176500
 P G K T P A A A V T R V D L T A S S C R C R P - T S E A L V P E L
 T R K N T G G R R N Q R R F D S I L L S V E T V N I R C A C T G F
 H D K Q Q H R R S P E S T - L R Q H A V V G R D R Q N Q L C L N W V

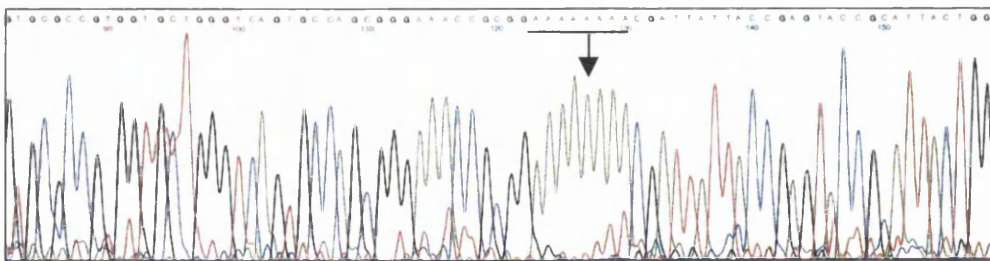
cgtcatttcatacccatgcggttacttagcggttaccggttcgccttaccttccggttgtcatgcaaccttagcgggtaccctcacCTCTTGAGCAGC 176600
 R - K M G W A T V - G N G T R G - R G T T M C R - R T G E G R S C T
 T M E Y G M R N S L R - G N A R V K G N D H V K L A Y G - R K L V
 D N - V G H P - K A T V R E G K G E R Q - A G K A R V R V E Q A R

TCAAAGTTGTCCAAGCCGTGGCTCGCATCGTAGTGGTAGTTCAACGTGAGGTCCACGAGCTGTTCCACATACTTGTAAACGGGTTTGGTGGGGCAGCGCGC 176700
 L T T W A T A R M T T T T - R S T W S S N W M S T V P K T P C R A
 D F N D L G H S A D Y H Y N L T L D V L Q E V Y K Y R T Q D P L A R
 - L Q G L R P E C R L P L E V H P G R A T G C V Q L P N P R A A R

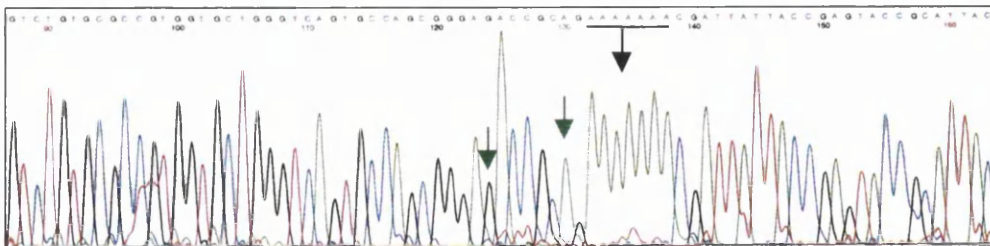
GAGAGCAGCGTCCAGTAATGCGGTACTCGGTAATAATCG**TTTTTTTT**TCGCGGTTTCCCGCTGGCACTGACCCAGCACCAGGGCGCAGACAAACAG 176800
 L A R T G T I R Y E T I I T K K G R N G A P V S G A G R R V S L C
 S C A D W Y H P V R Y Y D N K K R P K G S A S V W C W P A C V F L
 S L V R G L L A T S P L L R K K E A T E R Q C Q G L V V A C L C V S

UL131 **ACAGCCACACCCGACACAGCCGCATG**TGCAGACTGAGAAAGAAGCTTTATTATGAGACATCATACATAGTATAGCGAGGTGATGGGGCGGGGAAA 176900
 V A V G S V A A H Q L S L F F A K N H S M M C M T Y A L H H P P S L
 C G C G V C G C T A S Q S L F S - - S V D Y V Y Y L R P S P A P F
 L W V R C L R M N C V S F S L K I I L C - V C L I P S T I P R P F

I



II



III

```

GGTGCATGACCTCCGCCGGGTTTTACTGGGAAAGAGCCGCACGCTGAATTCAGCATGCGGTAGTGCGGTCCCAGAAAGCCGGAGCGCTGAACGTGG 179400
H M V E A A T K S P F S G C A S N W C A T T R D W F A P P A S R P
P A H G G G R N - Q P F L R V S F E L M R Y H P G L F G S A S F T T
T C S R R R P K V P S L A A R Q I G A H P L A T G S L R L R Q V H

TTCTCTTGGTCCGCCGGGGCCTTCTGGCGCTGAACTCGCCGGCCAGCAGCGTTACCTCCGTTACGTGATCctgcgaaacaaaccccgattggtt 179500
E R T P G A A P R R A G S S A P W C R - R R - T S G A F L G R N T
R K N T G G R A K Q R R F E G A L L T V E T V N I R R V F G S Q N
N E Q Q D R R P G E P A Q V R R G A A N G G N R Q D Q S C V G I P E

cgcgcgcacgtcgttcgggtgccctctttctgctctcccctttcccgaacccgtacgcgtcgcctcggctcaecCTCTTGAGCGTCTTAAAGTTATCCAAAC 179600
R A C T T R T G E K R G G R E R F R V R R E A - R K L T K F N D L G
A R V D N P H G R K T R G K G S V T R T A R S V E Q A D - L - G F
R A R R E P A R K E D E G K G F G Y A D S P E G R S R R L T I W V

CGTGGCTCAGTCGTAGTGGTACTCCACCAACGTGCCACCAGCCGCTTGACGTACTCTGACGGGTACGCTCCGGCAGCGCCGGGAACACGTGTCCA 179700
H S V D Y H Y E V L T D V L R K V Y E Q R T R E P L A P S C T D W
R P E R R L P V G G V H G G A A Q R V G S P Y A G A A G P F V H G L
T A - T T T T S W W R T W W G S S T S R V P V S R C R R P V R T G

GTAGTCCCACGGGCGCGCAAAGTCGTCCCGTTCGGCTTCTCCTCCCTGCGACACTGGCCCCACACCGCGACGCTCAGACACAAAACAGACAGAAACACACGG 179800
Y D W P R A F D D R E A E E R R C Q G W V A V S L C L V S L F V R
L G V P A C L R G A R S G G Q S V P G V G R R E S V F C V S V C P
T T G R A R L T T G S P K G A V S A G C R S A - V C F L C F C V T

TACAACCACATGTTGCACACTGAAGGTCAACTTTATTGTTACACAGGGAAGAATATGGACGTGGGTGGTGGTGGTTTACGGGACTCGGACAGGTGACCG 179900
Y L W M N C V S P - S - Q - V P F F I S T P P P P K V P S P C T V A
V V V H Q V S F T L K I T V C P L I H V H T T T T E R S E S L H G
C G C T A C Q L D V K N N C L S S Y P R P H H H N - P V R V P S R
    
```

UL131A

IV

Figure 3.10: Comparison of the region containing AD169 UL131A with the cognate CCMV sequence.

Panels I and IV show three-frame translations of the relevant regions of the AD169 and CCMV sequences, respectively, with nucleotide coordinates on the right. The locations of PCR primers for AD169 are indicated by arrows in panel I. AD169 UL131 is shown in blue in panel I and CCMV UL131A is shown in red in panel IV. Regions in AD169 that appear to be conserved in CCMV UL131A are shown in bold black font with conserved amino acid residues underlined. Predicted introns are shown in lower case. The region containing possible errors in the AD169 sequence is highlighted in yellow in panel I. Panels II and III show electrophoregrams of the sequences of the PCR-amplified region containing lower strand of AD169 UL131A and GW UL131A, respectively, with the regions containing 7 or 8 T residues, indicated by black arrows. Nucleotide substitutions in GW UL131A are indicated by green arrows in panel III.

CAGCATATTATTTCCCGTGACGCAGGCTAGTTGGCAAAGAGCCGACGCTGAACTCGAGGCTCCGGGCGTGTGGCCGACGAAACCGGCGGCGTTGAACG 176400
 - N A F L R V S F E L S R A H P A L S G A A N F T

TGGTCCTTTTGGTGGCCGCCGCGACGGTTCTGACGTCTAAAAGTCGCTGATGAGCAACGACACCTCGGTACGTTGATTctgcaagcacaggttccaaa 176500
T R K N T G G R R N Q R R F D S I L L S V E T V N I

cgtcatttcataccccatgcggttacttagccgttacccgttcgccttaccttcccgttgtcatgcacctttagcgcgtaccctcacCTCTTGAGCACG 176600
R K L V

TCAAAGTTGTCCAAGCCGTGGCTCGCATCGTAGTGGTAGTTCAACGTGAGGTCCACGAGCTGTCCACATACTTGTAAACGGGTTTGGTCGGGCAGCGCGC 176700
D F N D L G H S A D Y H Y N L T L D V L Q E V Y K Y R T Q D P L A R

GAGAGCACGCGTCCCAGTAATGCGGTACTCGGTAATAATCGTTTTTTCCGCGGTTTCCCGCTGGCACTGACCCAGCACCACGGCGCACAGACAAACAGA 176800
S C A D W Y H P V R Y Y D N K E A T E R Q C Q G L V V A C L C V S

CAGCCACACCCGACACAGCCGATGTTGCAGACTGAGAAAGAAAGCTTTATTATGAGACATCATAACATAGTATAGGCGAGGTGATGGGGCGGGGAAAG 176900
 UL131A L W V R C L R M

I

GGTGCATGACCTCCGCGCGGTTTTACTGGGGAAAAGAGCCGACGCTGAATTCAGCATGCGGTAGTGCGGTCCCAGAAAAGCCGAGGCGCTGAACGTGG 179400
 - Q P F L R V S F E L M R Y H P G L F G S A S F T T

TTCTCTTGTGGTCCCGCCGCGGGCCTTCTGGGCCTGAACTCGCCGCCAGCAGCGTTACCTCCGTTACGTTGATCctgcaacaaaaccccgattggtt 179500
R K N T G G R A K Q R R F E E G A L L T V E T V N I

cgcgcgcacgtcggttcgggtgcctcttttcgtcctccccttcccgaacccgtacgcgtcgctcggtcacCTCTTGAGCGTCTTAAAGTTATCCAAAC 179600
R K L T K F N D L G

CGTGGCTCACGTCGTAGTGGTACTCCACCAACGTGTCCACCAGCCGCTTACGCTACTCCTGACGGGTACGCTCCGGCAGCGCCGGGGAACACGTGTCCA 179700
H S V D Y H Y E V L T D V L R K V Y E Q R T R E P L A P S C T D W

GTAGTCCCAGGGCGCGCAAAGTCGTCCTCGGCTCGGTTCTCTCCCTGCGACTGGCCCCACCCGCGACGCTCAGACACAAAACAGACAGAAACACACGG 179800
Y D W P R A F D D R E A E E R R C Q G W V A V S L C L V S L F V R

TACAACCACATGTTGCACACTGAAGGTCAACTTTATTGTTACACAGGGAAGAATATGGACGTGGGTGGTGGTGGTTTACGGGACTCGGACAGGTGACCG 179900
 UL131A Y L W M

II

Figure 3.11: Corrected sequence of AD169 UL131A.

Panel I shows a version of the sequence of AD169 UL131A in which the first exon is in frame by removal of a T residue, and panel II shows the cognate sequence of CCMV, with nucleotide coordinates on the right. In each panel, UL131A is shown in red and conserved residues are underlined. Hydrophobic domains are shown within boxes. Putative intron sequences are shown in lower case.

3.1.6 TRL13/TRL14

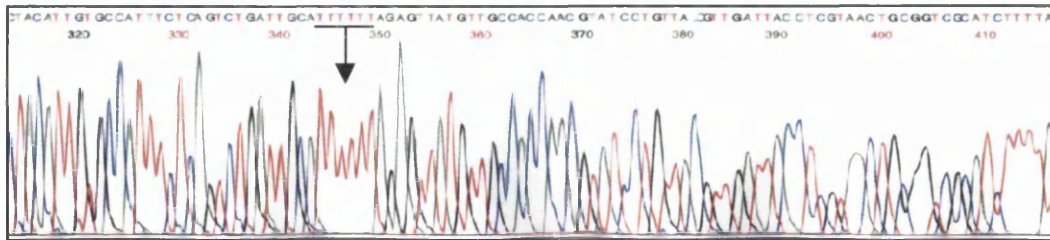
Fig. 3.12 shows three-frame translations of the relevant regions of the AD169 and CCMV genomes. TRL13 and TRL14 are two separate ORFs in AD169 (Fig. 3.12 panel I) and are weakly conserved as a single ORF (TRL13) in CCMV (Fig. 3.12 panel IV). AD169 TRL14 is a member of the RL11 family and contains a hydrophobic domain (shown within a box in Fig. 3.12 panel I). In addition an S+T-rich region is located in TRL13 (at nucleotides 10820-11200). Like AD169 TRL13/14, CCMV TRL13 encodes a hydrophobic domain near the 3'-end and a S+T-rich region near the 5'-end (Fig. 3.12 panel II). Moreover, only one mRNA 5'-end was determined for both AD169 TRL13 and TRL14 by transcript mapping, located upstream from TRL13 (see section 6.1). This suggests the presence of a frameshift error in the AD169 sequence within the region marked in yellow in Fig. 3.12 (panel I). However, sequence analysis of a PCR product derived from this region of the AD169 genome using the primers shown in Fig. 3.12 (panel I) indicated that AD169 sequence is correct (shown in Fig. 3.12 panel II).

To confirm whether AD169 TRL13/14 is naturally frameshifted, a PCR product derived from the corresponding region of Toledo was analysed using the same primers. As can be seen in Fig. 3.12 (panel III), in comparison with AD169 (Fig. 3.12 (panel II)), an additional T residue is present in the Toledo in the region indicated by arrow. This replaces TRL13 and TRL14 by a single ORF. Consequently, it appears that the AD169 TRL13/14 is naturally frameshifted. AD169 TRL13 would encode a truncated protein and TRL14 probably encodes no protein.

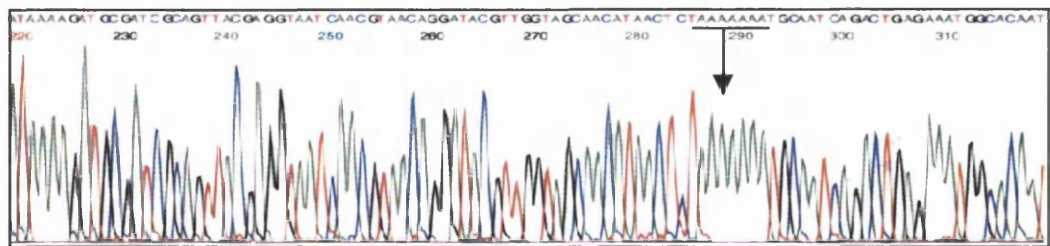
The PCR product from Toledo was smaller than that obtained using AD169 (Fig. 3.1b). Sequence analysis of this region confirms the absence of 276 bp (encoding the region rich in S+T residues) as highlighted in gray in Fig. 3.12 (panel I). Since CCMV TRL13 also encodes an S+T-rich region in a similar position, this is likely to reflect a deletion in Toledo in comparison with AD169.

D I L - S - T I N - N I L - G S A F K G I L M P I E R E L I L Q W
TRL13 R H L M K L N N K L K H S V R L S V Q R N I N A H - A R T N I A M D
 T S Y E A E Q - T K T F C K A Q R S K E Y - C P L S E N - Y C N G
 G A C A T C T T A T G A A G C T G A A C A A T A A A C T T A A A C A T T C T G T A A G G C T C A C G G T T C A A A G G A A T T A A T G C C C A T T G A G C G A G A A C T A A T A T T G C A A T G G A 10800
 T G D L R L C G R Y - Y P R Y Q K A A I K P V P V N V P V V L P L T
 W R F T V M W T I L I S A L S E S C N Q T C S C Q C P C S T T V N
 L A I Y G Y V D D T N I R V I R K L Q S N L F L S M S L - Y Y R -
 C T G G C A T T T A C G G T T A T G T G G A C G A T A C T A A T A T C C G C G T T A T C A G A A G C T G C A A T C A A A C C T G T T C C T G T C A A T G T C C C T G T A G T A C T A C C G T T A A C 10900
 I P L V L R Q P H Q H T V Q Q L S A I K A L Q N L - I A L L Q L H
 Y S T S T E T A T S T Y S T T V I S N K S T S E S I N C S T A T A
 L F H - Y - D S H I N I Q Y N S Y Q Q - K H F R I Y K L L Y C N C T
T A T T C C A C T A G T A C T G A G A C A C C A C A T C A A C A T A C A G T T A T C A G C A A T A A A A G C A C T T C A G A A T C T A T A A A T G C T C T A C T G C A A C T G C A C 11000
 Q Q P P F L Q N R R K Q P H R Y P Q R Q I Q T L R L P H V P T P P
 P A T T V S T K P S K T T T Q I S T T T N T N V E T T T C T N T T T
 S N H R F Y K T V E N N H T D I H N D K Y K R - D Y H M Y Q H H H
G A G C A A C C A C G T T T C T A C A A A C C G T G A A A A C A A C C A C A C A G A T A T C C A A A C G A C A A A T A C A A A C G T T G A G A C T A C C A T T A C C A A C A C C A C C A C 11100
 R P L L V M V S I I Q S I K D A T A V T R - S T - Q D T L V A T - L
 T V T C E G F N Y T V H K R C D R S Y E V I N V T G Y V G G N T
TRL14 D R Y L - W F Q L Y S P - K M R P Q L R G N Q R N R I R W W Q H N
 G A C C G T T A C T T G T G A T G T T T C A A T T A T A C A G T C C A T A A A A G A T G C G A C C G C A G T T A C G A G G T A A T C A A C G T A A C A G G A T A C G T T G G T G G C A A C A T A A C T 11200
 - K M Q S D - E M A Q C R L D S L - V P H A - N V R I R Q L S P N
 L K N A I R L R N G T M - T G F I M S T P R I K C A N - A T I T K
 S K K C N Q T E K W H N V D W I H Y E Y P T H K M C E L G N Y H Q T
C T A A A A A T G C A A T C A G A C T G A G A A A T G G C A C A A T G T A G A C T G G A T T C A T T A T G A G T A C C C C A C G C A T A A A A T G T G C G A A T T A G G C A A C A T A C C C A A A C 11300
 N T T A R H M F - L Q R H L P N Y L Q L N H K K R W K I Y Q A S P
 Q H H G T T Y V L T A T T P P - L S T T - P Q E T L E N I P G I T V
 T P R H D I C F D C N D T S L T I Y N L T T R N A G K Y T R H H R
 A A C A C C A C G G C A G A C A T A T G T T T G A C T G C A A C G A C A C C T C C C T A A C T A T C T A C A A C T T A A C C A A G A A A C G C T G G A A A T A T A C C A G G A T C A C C G T 11400
 - - R S R R K L L R N G V N W R H N V I H S W H M P C K I - R I - E
 I T V K K I T T - R C - L E T Q R Y P L L A H A L - D I K N L G
 D N G Q E E N Y Y V T V L I G D T T L S T L G T C P V R Y K E S R
 G A T A A C G G T C A A G A A G A A A T T A C T A C G T A A C G G T G T A A T T G G A G A C A C A C G T T A T C C A C T C T T G G C A C A T G C C C T G T A A G A T A T A A A G A A T C T A G G A 11500
 H - K H H W K - H H K N H - E S - H S P G N S C C M G R R S G I S
 T L K T P L E V T S - K P L R K L T F P W E F M L Y G Q A - W Y S Q
 N T E N T I G S N I I K T I E K A N I P L G I H A V W A G V V V S V
 A C A C T G A A A A C A C C A T T G G A A G T A A C A T C A T A A A A A C C A T T G A G A A A G C T A A C A T T C C C C T G G G A A T T C A T G C T G T A T G G C A G G C G T A G T G G T A T C A G T 11600
 G A Y S V V H G - P S H S Q K T A L H Q T S Q I - S R - I L D - G
 W R L - R C T W V A I A F P K N R I T P N F P N M I Q M N F G L R L
A L I A L Y M G S H R I P K K P H Y T K L P K Y D P D E F W T K A
 G G C G C T T A T A G C G T T G T A C A T G G G T A G C C A T C G C A T T C C C A A A A A C C G C A T T A C A C C A A A C T T C C C A A A T A T G A T C C A G A T G A A T T T T G G A C T A A G G C T 11700

I



II



III


```

TRL13  A R T R E Y S M V D H F A I T W T L V F L T L A S G T V T S S C H
          C P Y T G V Q H G R S L R D Y L D V G V L N T G I R N S D F F M S
          M P V H G S T A W S I T S R L L G R W C S - H W H Q E Q - L L H V I
          TGCCCGTACACGGGAGTACAGCATGGTCGATCACTTCGCGATTACTTGGACGTTGGTGTCTTAACTGGCATCAGGAACAGTGACTTCTTCATGTCAT 9200

          T S H S V S S T T V A T T S S T P S V N T T N S A D T S N T S T T
          Y E S L R K Q Y Y R R Y Y - F D S L S - Y D E L C R Y E - H V H H R
          R V T P - A V L P S L L L V R L P Q L I R R T L P I R V T R P P Q
          ACGAGTCACTCCGTAAGCAGTACTACCGTCGCTACTACTAGTTCGACTCCCTCAGTAAATACGACGAACCTCTGCCGATACGAGTAACACGTCACCCACAG 9300

          E S S T T S I T S T S N T T S T S L T T G T S A V S S A S T T S K P
          K Q Y Y F Y N I Y I E Y N F Y E S H Y W D I C R F F R L N H I K T
          K A V L L L - H L H R I Q L L R V S L L G H L P F L P P Q P H Q N
          AAAGCAGTACTACTTCTATAACATCTACATCGAATACAACCTTCTACGAGTCTCACTACTGGGACATCTGCCGTTCTTCCGCCTCAACCACATCAAAACC 9400

          T T P S T T I Q S T T T T K T T T V I T S T S T V P A S P T N T T
          N N A I Y H Y S E H Y H Y - D H H R H H - H L Y S T C E S H - H Y
          Q Q R H L P L F R A L P L L R P P S S L A P L Q Y L R V P L T L Q
          AACAACGCCATCTACCACATTCAGAGCACTACCCTACTAAGACCACCACCGTCATCACTAGCACCTCTACAGTACCTGCGAGTCCCCTAACACTACA 9500

          K F E C D T A T T K T W I N I T A K V G D N V T F P A C N T S G K
          K I - M - Y R N N - D V D - Y H S E S W - Q C N V S S L - Y E R Q I
          N L N V I P Q Q L R R G L I S Q R K L V T M - R F Q P V I R A A N
          AAATTTGAATGTGATACCGCAACTAAGAGCTGGATTAATATCACAGCGAAAGTTGGTGACAATGTAACGTTCCAGCCTGTAATACGAGCGGCAAAT 9600

          Y H T A R W T K V V N T K E T D L C L F G P D Y Y S T S P Q A G I C
          P Y R Q M D - G G K Y K R N R P M P L W A G L L L H V T P T S R Y M
          T I P P D G L R W - I Q K K Q T Y A S L G R I I T P R H H K P V Y
          ACCATACGCCAGATGGACTAAGGTGGTAAATACAAAAGAAACAGACCTATGCCTCTTGGGCGGATTATTACTCCACGTACCACAAGCCGGTATATG 9700

          F K C H W Q S M T I Y D V T T E N A G N Y I V R V H G D G N H H Y
          F Q M S L A E Y D Y L - C D D R E C W K L Y R Q S S W - W Q S P L
          V S N V I G R V - L F M M - R P R M L E I I S S E F M V M A I T I M
          TTTCAAATGTCATTGGCAGAGTATGACTATTTATGATGTGACGACCGAGAATGCTGGAATATATATCGTCAGAGTTCATGGTGATGGCAATCACCATTAT 9800

          D K G Y R L Q V T S N H T T G T N R K K K C P N D F T S Y T P D H D
          - - R L P F T S D V Q P H H G N E P K K M S Q - L H F V H S R S - Q
          I K V T P V Y K - R P T P R E R T E K N V P M T S L R T L Q I M T
          GATAAAGTTACCGTTTACAAGTGACGTCCAACCACACCAGCGGAACGAAACAAAATGTCCTCAATGACTTCACTTCGTACACTCCAGATCATGACA 9900

          N K E T K T I E N E F G M N Y D Q P T F P M G M H A I W A A V V V
          K Q R D E N H - K - I W N E L - S T Y F S Y G Y A C H L G G S G S
          K T K R R R K P L K M N L E - T M I N L L F L W V C M P S G R Q W -
          AAAACAAAGAGACGAAAACCATTGAAAATGAATTTGGAATGAACATGATCAACCTACTTTTCCATGGGTATGCATGCCATCTGGCGGCAGTGGTAGT 10000

          I V A L L A L Y M G S R S S V T I V K G G K P R Y K K L S N K D P
          D R S I T G I V H G Q P Q Q R D H C Q G R Q T A L Q E A I - Q G S
          - S - H Y W H C T W A A A A A - P L S R A A N R A T R S Y L T R I R
          GATCGTAGCATTACTGGCATTGTACATGGCAGCGCAGCAGCGTGACCATTGTCAAGGGCGCAAACCGCGCTACAAGAAGTATCTAACAAGGATCCG 10100

          D E Y W A S P - R A S V S M Y N K D I F C P R K S L P M - F Y C P
          G R V L G L S L A R V S V Y V - - R Y I L S Q K K S P D V I L L S C
          T S I G P L P S A R Q C L C I I K I Y F V P E K K V S R C D F I V L
          GACGAGTATTGGGCCTCTCCCTAGCGCGCTCAGTGTCTATGTATAATAAAGATATATTTTGTCCAGAAAAAGTCTCCCGATGTGATTTTATTGTCTG 10200
    
```

IV

Figure 3.12: Comparison of the region containing AD169 TRL13/14 with the cognate CCMV sequence.

Panels I and IV show three-frame translations of the relevant regions of the AD169 and CCMV sequences, respectively, with nucleotide coordinates on the right. The locations of the PCR primers for AD169 are indicated by arrows in panel I. AD169 TRL13 and TRL14 are shown in green and blue respectively in panel I and CCMV TRL13 is shown in red in panel IV. The region containing possible errors in the AD169 sequence is highlighted in yellow and the region that is absent from the Toledo genome is highlighted in grey in panel I. Panels II and III show electrophoregrams of the sequences of the PCR-amplified region containing AD169 TRL13/14 (lower strand) and Toledo TRL13 (upper strand), respectively, with the regions containing one fewer/more T residues indicated by arrows.

3.2 Discussion

On the basis of conservation of putative protein-coding regions between HCMV and CCMV, sequence errors were detected and confirmed in four published ORFs (UL15, UL102, UL145 and US22). Two other ORFs in AD169 (UL131A and TRL13/14) were found to be naturally frameshifted and therefore probably non-functional. The characteristics of the errors and natural frameshifts are listed in Table 3.2. The analysis results in 5'-extension of the protein coding regions of three genes (UL102, UL145 and US22) as well as to the introduction of two novel genes (UL15A and UL131A) to the AD169 genome.

	ORF	Position in sequence ^a	Sequence change to AD169	Outcome
	UL15	21956 22005	Insert G Insert G	Removes UL15 and introduces a novel gene, UL15A
Sequence errors	UL102	146744 146749	C→G C→G	5'-extension of protein-coding region
	UL145 (Toledo)	8886	Insert G	5'-extension of protein-coding region
	US22	211536	Insert G	5'-extension of protein-coding region
Natural frameshifts	TRL13/14	11201-11209	None	No error in AD169, insert T in Toledo
	UL131A	176742-176749	None	No error in AD169, delete T in Toledo and GW

Table 3.2: Corrections to HCMV DNA sequences.

^a AD169 DNA genome (Chee *et al.*, 1990), except for UL145 (Cha *et al.*, 1996).

The proteins encoded by HCMV and CCMV UL15A (Fig. 3.3) possess a potential transmembrane domain near the C terminus which may be a membrane spanning segment (Tomita and Marchesi, 1975). The UL131A protein contains a short hydrophobic domain near its N terminus which may act as a signal sequence for translation on membrane-bound ribosomes (McGeoch, 1985). The UL15A and UL131A proteins lack homologues in public databases. The natural frameshift in the TR_L region of AD169 results in splitting a single ORF (TRL13) into two (TRL13 and TRL14). TRL14 is thus an artifact of the frameshift.

The error in UL102 and an error near the 3'-end of US28 were reported previously (Smith and Pari, 1995; Kuhn *et al.*, 1995). The present work did not re-examine the latter but confirmed the former. It also added errors in three other genes. Detection of these errors depended on analysis of amino acid sequence conservation between HCMV and CCMV. Therefore, errors in non-coding-regions are likely to have escaped detection. It must be considered probable that additional sequencing errors exist in the published AD169 and Toledo sequences.

CHAPTER 4: RESULTS 2

Validation of the methods

Validation of the methods

This chapter describes an assessment of the methods used in this study. The areas covered include reliability of the method, comparisons of the results with published data, drawbacks of the methods and potential sources of artifacts. Greatest emphasis is placed on RACE techniques, as these continue to develop and are less commonly employed than the standard techniques of northern blotting, primer extension and RT-PCR.

4.1 SMART RACE

In order to map the 5'- and 3'-ends of HCMV genes, experiments were carried out using the SMART RACE cDNA amplification kit (Clontech). This kit provides a novel method for performing rapid amplification of cDNA 5'- and 3'-ends (see chapter 2 for a detailed explanation). The integrity and purity of the RNA used as starting material is an important element in high-quality cDNA synthesis. Therefore, RACE was performed using polyA⁺ RNA (IE, E, L and MI) isolated directly from cells using the FastTrack 2.0 mRNA isolation kit (Invitrogen) or from total cellular RNA prepared using Trizol reagent. Prior to performing RACE with the templates, a positive control experiment was successfully performed using RNA provided in the kit (data not shown).

The cDNA for 5'-RACE is synthesised in RT reactions using a modified lock-docking oligo (dT) primer (5'-CDS) and the SMART II oligonucleotide as described in section 2.2.8. The 5'-CDS primer has two degenerate nucleotide positions at the 3' end which position the primer at the start of the polyA tail, thus eliminating the 3' heterogeneity inherent with conventional oligo (dT) priming (Borson *et al.*, 1992). The 3'-RACE cDNA was also synthesised by RT reactions, using a special oligo (dT) primer (3'-CDS) (section 2.2.8). This primer includes the lock-docking nucleotides as in the 5'-CDS primer, and also has a portion of the SMART sequence at its 5'-end. By incorporating the SMART sequence into both the 5'- and 3'-RACE-ready cDNA populations, both ends of the cDNA can be amplified using a universal primer (UPM) that recognises the SMART sequence plus a distinct gene-specific primer (GSP).

ORF	Primer	Sequence (5'-3')	Position ^c	PCR product(s) ^d
UL22A	UL22A ^a	TTTGGCTTGATTTTCTTTGTGTTC	27478-27455	150, 450
UL22A	UL22A-3' ^b	GGCTTTGGCGGCACCTTCTCAGAA	27158-27181	300
UL36	UL36-S1 ^a	ATGGGCCGCTGGTAGTCGCGCATA	49244-49267	550
UL36	UL36-3' ^b	CACGCACCTGTGGCCGCAGGAGCT	48558-48535	500

Table 4.1: Primers used in RACE to map 5'- and 3'-ends of transcripts from UL22A and UL36.

^aPrimer used in 5'-RACE.

^bPrimer used in 3'-RACE.

^cPosition of the primers in the positive strand of the AD169 genome.

^dApproximate sizes (bp) of RACE products excised for cloning (See Fig. 4.1 and 4.2).

The efficiency of the RACE reaction depends on the GSP used, since different primers have different optimal annealing temperatures. GSPs should have a G+C content of 50-70% and a T_m of at least 65°C (Freier *et al.*, 1986). Longer primers closer to the ends of the cDNA with annealing temperatures above 70°C give more robust amplification. Therefore, GSPs were designed to be 24 nucleotides in size, located approximately 500 bp downstream from the putative initiation codon or 500 bp upstream from the putative stop codon of the target gene. Care was also taken to avoid using self-complementary primers which can fold back and form intramolecular hydrogen bonds. Similarly, primers that exhibit complementarity to the UPM were also avoided.

5'- or 3'-RACE PCR products were separated by agarose gel electrophoresis (two examples are shown in Figs. 4.1 and 4.2). Fragments were excised and cloned into pGEM-T, and 5-30 separate clones were sequenced for each PCR product using M13 universal primers and, in some cases, gene-specific primers. The sequence obtained from the highest proportion of clones was considered as the 5'- or 3'-end. Genes can have more than one 5'-end in a local region of a few bp, or even two promoters that give quite separate 5'-ends. The nuances of 5'-end identification are discussed below.

Since the cDNA contains the SMART II oligonucleotide (30 nucleotides) at its 5' end (see section 2.2.8), the size of a 5'-RACE PCR product was the sum of the relevant AD169 sequence plus 30 bp. The 5'-RACE products contained the SMART sequence adjacent to the end of 5'-end of the target gene (an example is underlined in red in Fig. 4.3 panel II). The sequences of 3'-RACE products contained a polyA tract at the 3'-end.

The SMART sequence contains four G residues at its 3'-end designed to anneal to a short tract of C residues introduced at the 3'-end of the cDNA (equivalent to the 5'-end of the mRNA) by RT. This feature introduces ambiguity regarding the precise 5'-ends of certain genes, since one or more of the G residues could have been templated by the RNA rather than having originated by the action of RT in adding non-templated C residues to the 3'-end of the cDNA. In this study, the first nucleotide of HCMV sequence that was found at the end of the G tract was considered as the 5'-end. However, the initiating nucleotide of eukaryotic and prokaryotic mRNAs is usually (>90%) a purine (Lewin, 1994b). Therefore, if the 5'-end identified was a pyrimidine preceded by one or more G residues in the HCMV sequence, the terminal G residue in the SMART sequence was considered as the 5'-end of the target gene. Similarly, since 3'-RACE products contained a polyA tract at the 3'-end, it is possible that one or more A residues are templated. In this study, the last nucleotide of HCMV sequence prior to the polyA tract was considered the 3'-end.

Almost all eukaryotic RNA polymerase II promoters appear to contain a TATA element, but there are examples where TATA boxes are absent (Hall and Brown, 1985; Melton *et al.*, 1984). In this study, indicators predictive of whether an mRNA has a TATA element include the presence of a six residue consensus TATA (TATAAA) element (Breathnach and Chambon, 1981; Carbon *et al.*, 1987; Lewin, 1994a) 20-35 bp upstream from the 5'-end of a target gene. In most instances the consensus TATA was not present, and the sequence most closely related to TATAAA (for example, TATAAG, TATATG) was considered the TATA element. The definition of the polyadenylation signal for a target gene was the presence of a consensus motif, AATAAA (Proudfoot and Brownlee, 1976), ATTAAA (Rawlinson and Barrell, 1993) or CATAAA (Kouzarides *et al.*, 1988) in the region close upstream from the 3'-end of a target gene.

The 5'-end and 3'-ends of the HCMV UL36 and UL22A transcripts were mapped previously by Kouzarides *et al.* (1988) and Rawlinson and Barrell (1993), respectively. Therefore, these transcripts were analysed as controls in this study.

UL36 - 5'-RACE using primer UL36 detected a band of similar size and intensity from IE, E and L RNA but not from MI RNA (Fig. 4.1a). Sequence analysis indicated that the 5'-end of UL36 is located at nucleotide 49863, 21 bp downstream from a potential TATA element (49884-49889). As expected, the 5'-RACE cDNA sequence of UL36 lacked the intron (data not shown). A band corresponding to the 3'-end of the UL36 transcript was obtained from IE, E and L RNA (Fig. 4.2a), again indicating that UL36 is an IE gene. Sequencing indicated that the 3'-end is at 48094, and is preceded 20 nucleotides upstream by a typical polyadenylation signal (AATAAA). The position of the 5'-end of the UL36 mRNA determined by Kouzarides *et al.* (1988) using primer extension is in exactly the same positions as those determined in this study by RACE.

UL22A - ORF UL22 as originally described in AD169 genome was replaced by a spliced gene (R27080s) mapped by Rawlinson and Barrell (1993) and confirmed (and renamed UL21.5) by Bresnahan and Shenk (2000). In this study this gene is named UL22A (see Fig. 1.9). Primer UL22A generated two bands of approximately 150 and 450 bp from L RNA (Fig. 4.1b). The smaller band was cellular in origin, but the larger band mapped the 5'-end of UL22A to 27080, 25 bp downstream from a potential TATA element (Fig. 4.3). The 5'-RACE cDNA sequence lacked the intron (Fig. 4.3). Primer UL22A-3' mapped the 3'-end to 27574, 20 bp downstream from a polyA signal. Rawlinson and Barrell (1993) analysed the UL22A transcript by sequencing a cDNA derived from a library prepared from L RNA, and mapped the 5'-end to 27082 and the 3'-end to 27574.

The results with UL36 and UL22A show that SMART RACE detects the authentic 5'- and 3'-ends of HCMV mRNAs.

In many of the experiments described in later chapters, RACE yielded a single band corresponding to the 5'- or 3'-end. However, in some experiment multiple RACE products were generated, making the interpretation difficult. Multiple bands can have several sources, including alternative splicing, multiple promoters or use of different

polyadenylation signals, amplification of cellular RNA, premature termination of first-strand cDNA and degradation of RNA used as starting material. Alternatively, the gene may be a member of a multigene family, in which case the GSP may simultaneously amplify several cDNAs.

In some experiments, the 5'-RACE product was smaller than expected in that it mapped the 5'-end downstream from the first ATG codon in the ORF. This may indicate that translation is not initiated at the first ATG, or that the RACE product does not correspond to the 5'-end. There are several possible sources of incomplete RACE products generated from correctly primed sites. These include degradation of input RNA, and premature termination of first-strand cDNA due to specific sequences or structures in the RNA. Moreover, the presence of G tract at the 3'-end of the SMART sequence may cause non-specific priming from C tracts within the first strand cDNA, giving a truncated RACE product. This may not be common, however, since no RACE product originating from the C tract (marked as a G tract underlined in yellow on the strand shown in Fig. 4.3) was detected for UL22A. Moreover, during the course of this study no 5'-end (except that for UL18; see chapter 7) mapping to a tract of four or more G residues was identified. However, the possibility that there might be rare situations in which this occurs cannot be ruled out.

The sensitivity of RACE depends in part on the abundance of the target mRNA. Consequently, the studies described here revealed some examples of ORFs that clearly encode proteins but whose 5'-ends could not be mapped.

4.2 Northern blot

RACE experiments give an indication of the kinetic class of a target gene but, since PCR is not necessarily quantitative, the results must be considered as preliminary. Northern blot experiments were carried out to determine the expression kinetics of HCMV genes and to determine mRNA sizes. Equal quantities (usually 1-3 μg) of polyA⁺ mRNA (determined spectrophotometrically) were electrophoresed in agarose-formaldehyde gels. RNA was

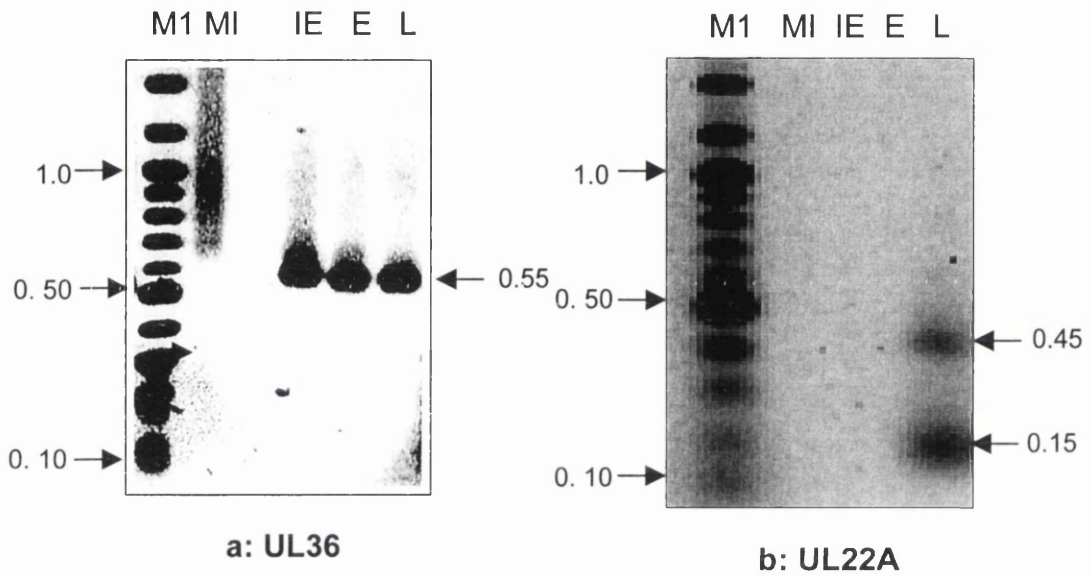


Figure 4.1: 5'-RACE of control HCMV genes.

EtBr-stained 1 or 1.5% (w/v) agarose gels showing representative 5'-RACE products containing regions of the HCMV (a) UL36 and (b) UL22A genes that were amplified from 5'-cDNA with GSPs (listed in Table 4.1) and UPM. Sizes are in kbp. 5'-cDNA was prepared from different kinetic classes (IE, E and L) of AD169 mRNAs and from MI RNA. M1 is a 100 bp ladder.

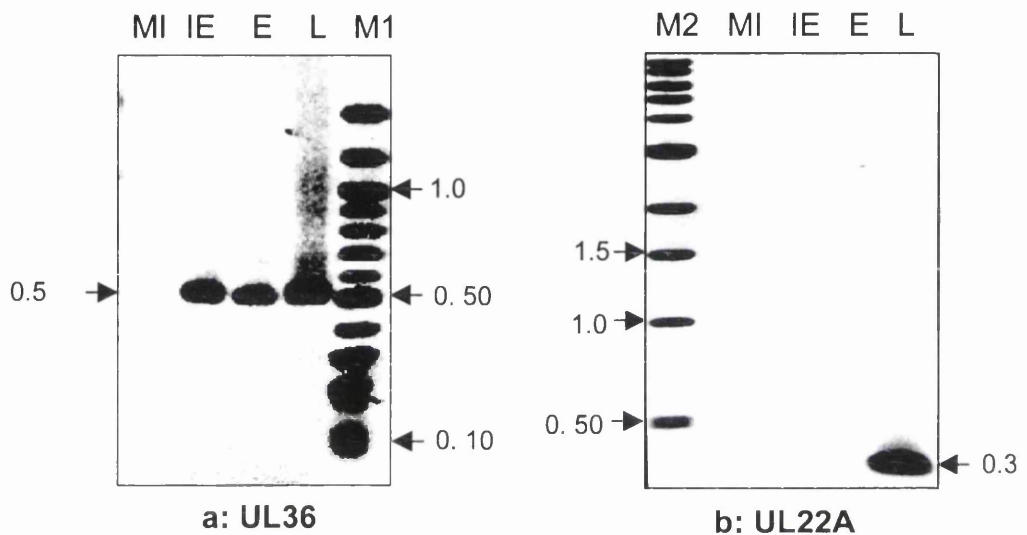
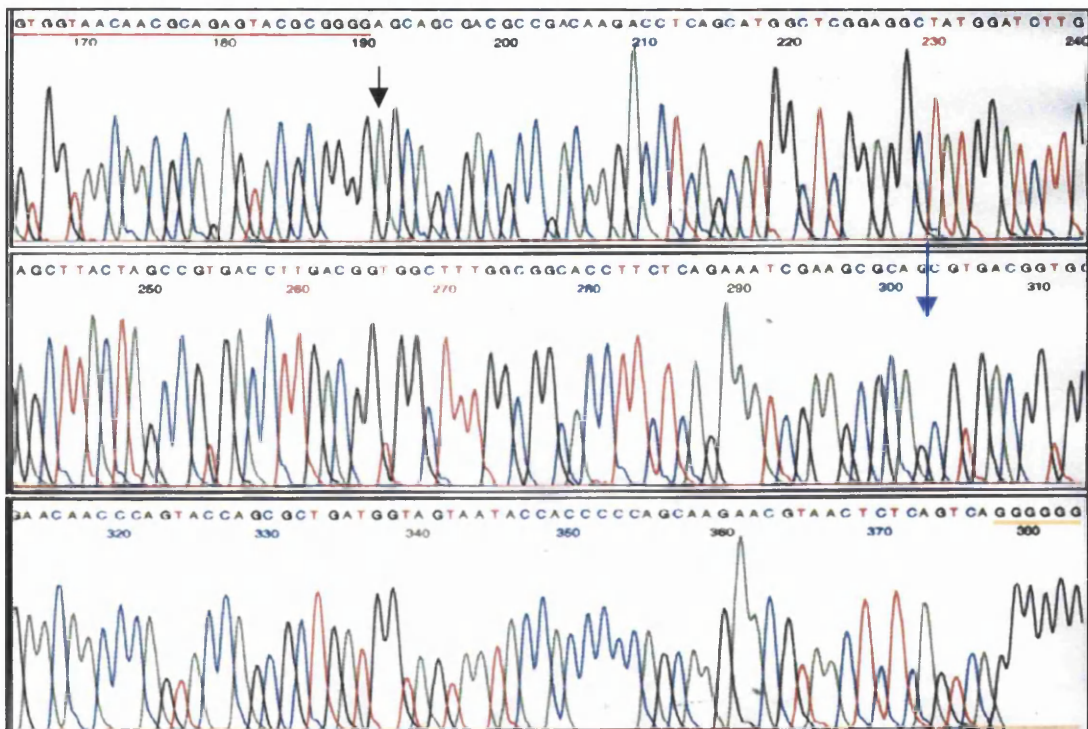


Figure 4.2: 3'-RACE of control HCMV genes.

EtBr-stained 1% (w/v) agarose gels showing representative 3'-RACE products containing regions of the HCMV (a) UL36 and (b) UL22A genes that were amplified from 3'-cDNA with GSPs (listed in Table 4.1) and UPM. Sizes are in kbp. 3'-cDNA was prepared from different kinetic classes (IE, E and L) of AD169 mRNAs and from MI RNA. M1 and M2 are 100 bp and 1 kbp ladders, respectively.

GTTGTTTTTCAGAAGCGGCCACATGACCTCGAGATGTCGTACCCCAAGGTATTTAACGGCACACAGCCAGACGCCTTCGTACGACGACGCCGACAAGA 27100
 M A R R L W I L S L L A V T L T V A L A A P S Q K S K R
 CCTCAGCATGGCTCGGAGGCTATGGATCTTGAGCTACTAGCCGTGACCTTGACGGTGGCTTGGCGGCACCTTCTCAGAAATCGAAGCGCAGgtaaacg 27200
 V T V E Q P S T
 gaatctggggaattcaacacaggtagaatacaaaaaataacgtgattgtgaacgcggttatcgtgtttttgcagCGTGACGGTGAACAACCCAGTAC 27300
 S A D G S N T T P S K N V T L S Q G G G S T T D G D E D Y S G E Y D
 CAGCGCTGATGGTAGTAATACCACCCCGCAAGAACGTAACCTCTCAGTCAGGGGGTCCACCACCGACGGAGACGAAGATTACTCCGGGGAGTATGAC 27400
 V L I T D G D G S E H Q Q P Q K T D E H K E N Q A K E N E K K I Q
 GTTTTGATTACAGACGGAGATGGCAGCGAACATCAGCAACCACAAAAGACTGATGACACAAAGAAAATCAAGCCAAAGAAAATGAAAAGAAGATTTCAGT 27500
 AACAGCAGACCCCAAGGGTTAACGATTATGTTGACTACCTTGTTTTTTATTTAAAAGCTGTAAGGTTTGTCTTAAAAACCCCGCCTCCGGTCTTTTT 27600

I



II

Figure 4.3: Sequences at the 5'- and 3'-ends of the UL22A mRNA.

Panel I shows AD169 UL22A and its encoded amino acid sequence. Coordinates are shown on the right. RACE primers are shown in italic bold font with orientations indicated by arrows. The intron is shown in lower case. The black and red vertical arrows indicate the positions of the 5'- and 3'-ends of the UL22A mRNA, respectively, as represented in the majority of 5'-cDNA clones. The putative TATA element (TATTTA) and polyA signal (ATTTAAA) are shown in italic font and underlined in red and green, respectively. Panel II shows an electropherogram of the sequence of the 5'-RACE product (upper strand of AD169 genome) with the 5'-end and the splice sites indicated by black and blue arrows, respectively. The SMART sequence is underlined in red. The internal G-tract in the RNA, discussed in the text, is underlined in yellow.

blotted to membranes and fixed by UV irradiation. A synthetic polyA⁺ RNA ladder was electrophoresed alongside the mRNA samples to facilitate estimations of transcript sizes. Blots were initially hybridised with transcribed antisense RNA probes corresponding to regions of the target gene. Single-strand RNA probes were preferred to double-stranded DNA probes because RNA-RNA hybrids are more stable than RNA-DNA hybrids and because they are strand specific and are not depleted by hybridisation to the complementary strand of the probe. ³²P-RNA probes were produced using the Lig'nScribe and MAXIscript kit (Ambion). The Lig'nScribe reaction involves the addition of a phage T7 RNA polymerase promoter in a chosen orientation to a PCR fragment (explained in detail in section 2.2.7.3). In general, 5'-RACE products were used for making the probes. The PCR products generated by the Lig'nScribe reaction were examined on agarose gels in order to estimate their concentration and to verify that the products were unique and of the expected size (64 bp larger than the 5'-RACE product). The PCR product was then used as a template in an *in vitro* transcription reaction using the MAXIscript kit (see section 2.2.7.3 for details). A successful positive control reaction was carried out initially to ensure that the kit was functioning properly (data not shown). Most of the blots were rehybridised with a control cellular DNA probe (GAPDH₂) to normalise the amount of RNA loaded in each lane.

UL36- The UL36 probe hybridised to a single band of 1.6 kb (Fig. 4.4a) which was of similar intensity in IE, E and L RNA. Therefore, UL36 is considered to be an IE gene as indicated in the RACE experiment (Fig. 4.1a and 4.2a) and reported previously by Mocarski *et al.* (1988) using northern blotting with a single-strand probe. Chambers *et al.* (1999) detected UL36 mRNA at E and L times post infection using gene microarray technology. The size of the mRNA (1.6 kb) is consistent with the spliced UL36 transcript mapped by RACE (1666 nucleotide transcript; see section 4.1).

UL22A- Probing a northern blot with UL22A sequences yielded a single band of 0.4 kb in L RNA (Fig. 4.4b). Since 5'- and 3'- ends of the UL22A mRNA are located at 27080 and 27574, respectively (495 nucleotide transcript including a 83 nucleotide intron) (see section

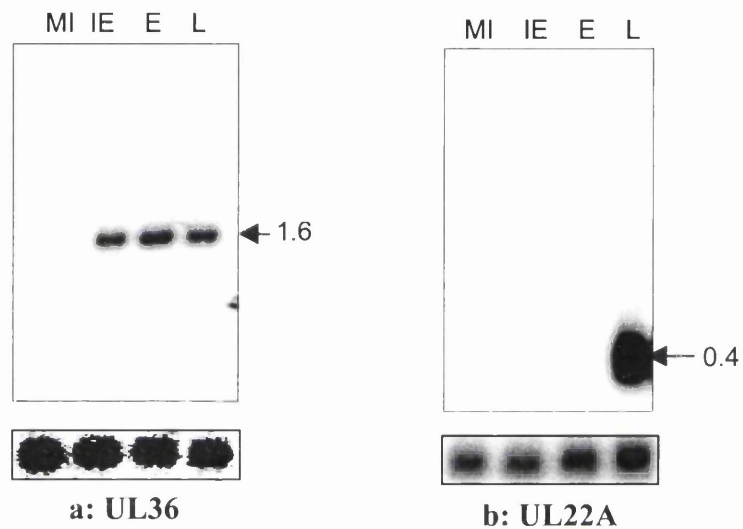


Figure 4.4: Northern blot analysis of control HCMV genes.

Northern blots showing the expression profile of transcripts from HCMV genes (a: UL36 and b: UL22A). IE, E and L RNA prepared from HCMV-infected cells were hybridised with a single-strand specific RNA probe (top) or GAPDH₂ probe (bottom). Sizes are given in kb.

4.1), the 0.4 kb band (Fig. 4.4b) represents the spliced UL22A mRNA. Consistent with this, Bresnahan and Shenk (2000) and Rawlinson and Barrell (1993) detected a 0.4 kb UL22A mRNA at L times post-infection by northern blotting with a DNA probe. Therefore, UL22A is considered to be an L transcript as indicated by the RACE experiment (Fig. 4.1b), previous northern blotting experiments and gene array hybridisation (Chambers *et al.*, 1999).

In many experiments, northern blots probed with single-strand probes yielded a single band corresponding to a specific gene. However, in other experiments more than one band was generated. Multiple bands can have several sources, including alternative splicing, alternative 5'-ends and presence of 3'-coterminal genes.

4.3 Primer extension

To confirm the 5'-ends of transcripts from certain HCMV genes, primer extension experiments were carried out using the Primer Extension System-AMV Reverse Transcriptase kit (Promega). A 24 bp GSP was designed from the region approximately 100 bp downstream from the 5'-end of a target gene determined by RACE (Table 2.3). The primer was 5'-end-labeled with ^{32}P (see section 2.2.17.1) and extended by RT in the presence of the four dNTPs and polyA⁺ RNA (see section 2.2.17.3). The RNA was thus reverse transcribed into end-labelled cDNA. The cDNA was analysed on a denaturing 8% polyacrylamide gel. The size of the cDNA corresponds to the number of nucleotides between the 5'-end of the primer and the 5'-end of the mRNA. A 5'-end for a target gene mapped ± 5 nucleotides from that mapped by RACE was considered an accurate result. A positive control RNA and primer supplied with the kit were successfully tested (data not shown).

The 5'-end of primer PEL36 (Table 2.3) maps 137 nucleotides downstream from the 5'-end of UL36 determined using RACE (49863) and previously reported by Kouzarides *et al.* (1988) using a primer extension assay. This primer yielded a single band of approximately 140 ± 5 bp (measured from gel; see Fig. 5.5), thus mapping the 5'-end for UL36 gene at position 49861-49866, essentially the same as that mapped by 5'-RACE (section 4.1). This

result shows that the primer extension kit used in this study detects authentic 5'-ends. However, other features in successful primer extension include the quality and quantity of the target RNA and the annealing conditions. Moreover, primer extension, unlike RACE, does not involve amplification, and therefore is less sensitive.

4.4 RT-PCR

The predicted splice sites in certain HCMV genes were confirmed by RT-PCR. This strategy has proven useful in identifying the exact locations of splice signals and is widely used. RT-PCR was carried out using the TitanTM One Tube RT-PCR System kit (Boehringer-Mannheim). As a test of this kit, two known spliced HCMV transcripts (UL22A and UL36) were analysed. PolyA⁺ RNAs (IE, E and L) were used as templates with two primers (listed in Table 2.2) mapping in the exons flanking predicted introns. RT-PCR products were cloned into pGEM-T for sequencing.

UL36 – Primers UL36-S1 and UL36-S2 (Table 2.2) were used to amplify the sequence between 49749 and 49244 (Fig. 4.5a). Two products, were generated from IE, E and L RNA (Fig. 4.5a). The smaller (400 bp) represents the spliced RNA and the larger (500 bp) the unspliced RNA. The locations of the splice sites in the DNA sequences and the corresponding electropherogram are shown in Fig. 4.6 (panels I and II). The sites correspond to those detected by Kouzarides *et al.* (1988) using S₁ nuclease and in this study by 5'-RACE (see section 4.1). Since 5'-RACE failed to detect the unspliced sequence, it is likely that the larger RT-PCR product originated from DNA contamination rather than from RNA.

UL22A - Primers UL22A and UL22A-3' (Table 2.2) were used to amplify the sequence between 27157 and 27478 (Fig. 4.5b). Two products were generated from L RNA (Fig. 4.5b). The smaller (250 bp) representing the spliced RNA and the larger (350 bp) the unspliced RNA or contaminating DNA. The locations of the splice sites in the DNA sequences and the corresponding electropherogram are shown in Fig. 4.3 (panels I and II).

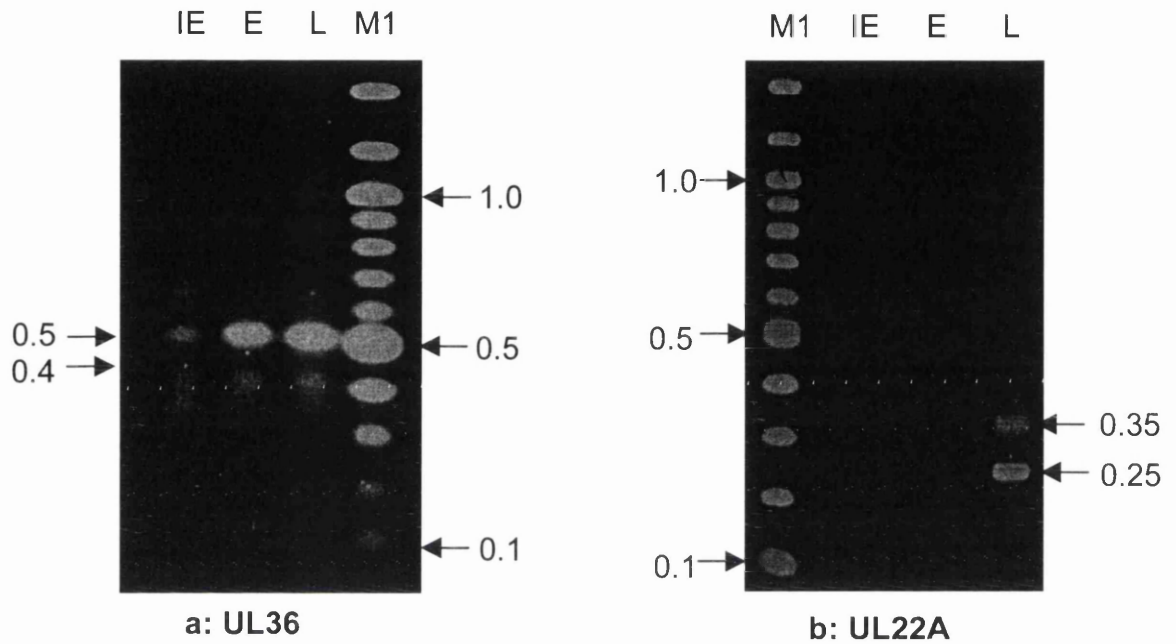
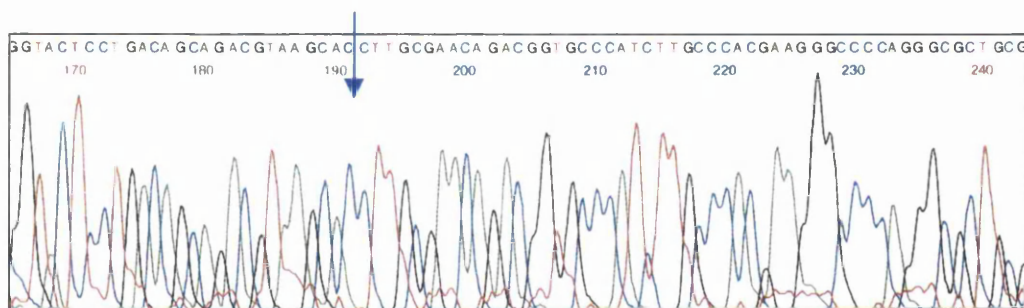


Figure 4.5: Gel electrophoresis of RT-PCR products from control HCMV genes.

EtBr-stained 1% (w/v) agarose gel showing the RT-PCR products (sizes in kbp) amplified from HCMV IE, E and L RNA using primers (listed in Table 2.2) mapping in the exons flanking an intron. (a) UL36 with primers UL36-S1 and UL36-S2. (b) UL22A with primers UL22A and UL22A-3'. M1 is 100 bp ladder.

CCGTAGAGCATGACGGCGTTCCAGTAGTCGTCTACTGCACCA**TGGGCCGCTGGTAGTCGCGCATAGT**GTGGAAGTGGTCGCGGTGACGAAAGCCGTTCC 49300
 R Y L M V A N W Y D D Y Q V M F R Q I D R H I H F H D R H R F G N R
 GCAGAAAGTCCTTCATGGTGGGTGCCAGCTCGTAGACGAGTCGCGCAGGTCATCGTAGCAGTAGATGCCGCCGCGCTGCCCGATGAGCACGATGAGTTG 49400
 L F D K H T P A L E Y V C D R L D D Y C Y I G G R Q G I L V I L Q
 GTAGCGCATAAAGCCCGGACCCTCGACGAAGCCAAAGGGGTGCAGGTA**CTCTGACAGCAGACGTAAGCAC***ctggtagagaatagaaaaatccacgcac* 49500
 Y R H F G P G E V F G F P H L Y E Q C C V Y A G
gttgaaaacacctggaagaacgtgcccagcggaacgtcctctttccaggtgtcttcaacgacgtggggcttacCTTGCGAACAGACGGTGCCCATCTTG 49600
 Q S C V T G M K
 CCCACGAAGGGCCCCAGGGCGCTGCGCGAACGGAGCTGGATGAAGCAGCGTTCGGCCAGGCCACGTGCAGCCGGTGCCGCATTCTGCTCCAGAAAGT 49700
 G V F P G L A S R S R L O I F C R E P W A V H L R T G C E Q E L F D
 CGTTGAGACCGTTAAAGTCCCGGCC***TCGGATGGCGATGCAGCCGTAGGC***CATCAGCGTGTCCCGTAGGTCGTCCATGACGGACTCCTCTACCTTCGCTCG 49800
 N L G N E D G A S I A T C G Y A M L T D R L L D B M

I



II

Figure 4.6: Interpretation of splicing in HCMV UL36 by RT-PCR.

Panel I shows the sequences of the first exon of AD169 UL36 and part of the second exon, with coordinates on the right. The amino acid sequence of UL36 is shown in red. RT-PCR primers are shown in italic bold font and their orientations by arrows. The intron is shown in lower case. Panel II shows an electropherogram of the sequence of the RT-PCR product (upper strand of AD169 genome), with the splice site indicated by a blue arrow.

sequences and the corresponding electropherogram are shown in Fig. 4.3 (panels I and II). They correspond to the splice sites previously detected by Rawlinson and Barrell (1993) using RT-PCR and in this study by 5'-RACE (see section 4.1).

The results presented here confirm that the usefulness of RT-PCR as a strategy for identifying the exact locations of splice signals. The experimental design, however, depends on sound prediction of splicing patterns.

4.5 Discussion

Transcriptional expression patterns of certain previously mapped HCMV genes (UL22A and UL36) were analysed in several ways, including RACE to identify the 5'- and 3'-ends of mRNAs, primer extension, northern blot assay and RT-PCR. The results were in agreement with previously published data, and point to the general reliability of these methods in characterising HCMV mRNAs.

CHAPTER 5: RESULTS 3

Transcription of genes in the HCMV US22 family

Transcription of genes in the HCMV US22 family

The HCMV US22 family is defined on the basis of amino acid sequence homology and consists of 13 members (UL23, UL24, UL26, UL28, UL29, UL36, UL43, IRS1, US22, US23, US24, US26 and TRS1). US22 family members are distributed in U_S, U_L and TR_S, and contain up to four conserved motifs (Kouzarides *et al.*, 1988; Nicholas and Martin, 1994) (see section 1.3.10). Although certain HCMV US22 genes have been investigated with respect to RNA or protein products (e.g. UL36, US22, TRS1 and IRS1), others have not. In this study, transcriptional expression patterns of US22 gene family members were analysed in several ways, including RACE techniques to identify the 5'- and 3'-ends of mRNAs, primer extension, northern blot assay and RT-PCR.

5.1 Mapping the 5'- and 3'-ends of US22 genes by RACE

In order to map the 5'- and 3'-ends of HCMV US22 genes, experiments were carried out using the SMART RACE kit. The 24 bp GSPs used in 5'- and 3'-RACE were designed from the region approximately 500 bp downstream from the putative initiation codon or upstream from the putative stop codon of the target genes and are listed in Tables 5.1 and 5.3, respectively.

5'- or 3'-RACE PCR products were separated by agarose gel electrophoresis (Fig. 5.1 and 5.3). Fragments were excised and cloned into pGEM-T, and 5-30 separate clones were sequenced for each PCR product using M13 universal primers and, in some cases, gene-specific primers. Sequences were compared with the AD169 DNA sequence using BLAST, and 5'- or 3'-ends were defined as those obtained from the highest proportion of clones.

5.1.1 Sequences of 5'-ends

5'-ends were identified for seven of the US22 family genes: UL23, UL24, UL26, UL36, US22, US23 and US24. TRS1 and IRS1 were not examined. The ends are listed in Table 5.2, and a summary of the overall results is listed in Appendix 1 at the back of this thesis.

ORF	Primer	Sequence (5' - 3')	Position ^a	PCR product(s) ^b
UL23	UL23	AAGGTCAACGAGTCAGAGTAGATG	28374-28397	550
UL24	UL24	GATCTGCGGTTGCCGACGCGACGA	29701-29724	350, 450
	UL24 (II)	TCGGCCGTGAGCGCCAGGCTAGCC	29801-29824	350
	PEL24	CGCGGCCAGCGTGCGGAGGACACT	29905-29928	150, 250
UL26	UL26	AAGTTCTCACGGCTGATCTCGTAG	32477-32500	400
UL28	UL28	ACAGTCTCCTCCTTCACGACGCC	34741-34764	No product
	UL28 (II)	CGGCGTGGAGCTGACATACGCGCA	35176-35199	500
	UL28S1	GCGCACCACCCAAAGTACTGAGT	35537-35560	No product
UL29	UL29	CCGGCAGGTGGCGTGCGAGACCGT	36628-36651	300, 450
	UL29 (II)	TGCTCCAGATCGATCTCGGTCAGC	36701-36724	400, 500
	UL29 (III)	AGCGCCGCAAGTAACGCCGCGAGGC	36801-36824	350, 500
UL43	UL43	TAACCAGCTCGACAGCCAGCGCGT	55642-55665	No product
	UL43 (II)	AGCGTGCGTTCCTTCATCTCGTCCG	55671-55694	No product
	UL43 (III)	ATGCTCTGACGGCGCCGCGGATT	55698-55721	450, 650
US22	US22	TCCCAGTCGTAGACGTAGACGGTG	211144-211167	550
US23	US23	AGCACCAGCTCCGTCTCTTCGTAG	213140-213163	400, 550
	US23 (II)	CACCATGATCACGCACTTGCCTAG	213172-213195	350, 500
US24	US24	TCGTAGGCGTAGACGCGAGAAAAG	214672-214695	350
	US24 (II)	AGCGTATCCCAACGGCGTCAACGG	214773-214796	250
	PES24	ACAGCGTCCCTCGTTACGATGAAC	214902-214925	150
US26	US26	AGGACCGCGGCCCGTGTACTCTGG	217209-217232	No product
	US26 (II)	AGCGATTCTTCGCAGCAGATGAGC	217273-217296	200, 1100
	US26 (III)	CGTCGGTGACAGCCGCGCTGCCAA	217319-217342	650, 1000

Table 5.1: Primers used in 5'-SMART RACE PCR to map the 5'-ends of US22 genes.

^a Position of primers in the AD169 genome (upper strand).

^b Approximate sizes (bp) of PCR products excised for cloning.

The 5'-end of the UL36 transcript mapped by Kouzarides *et al.* (1988) was confirmed by RACE (see section 4.1).

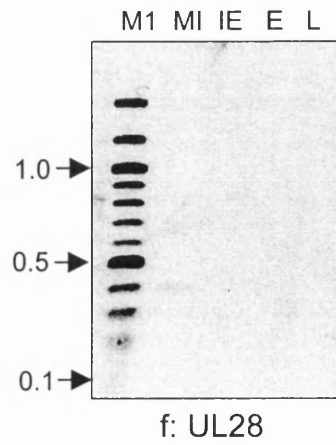
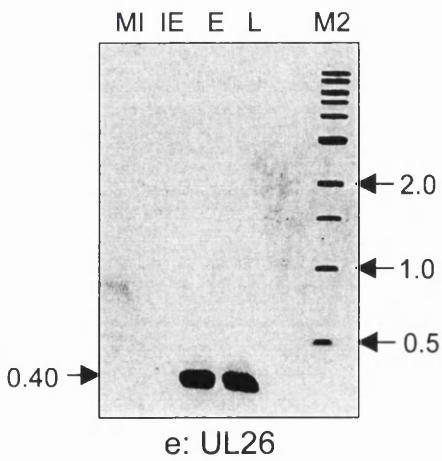
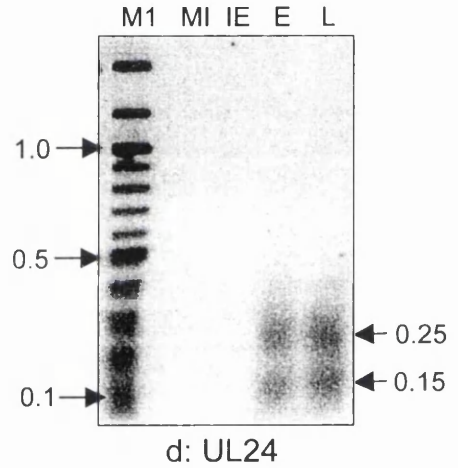
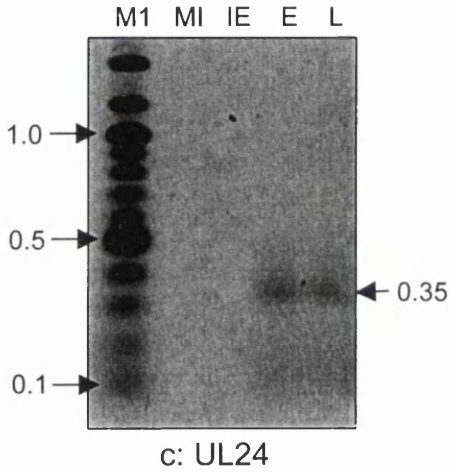
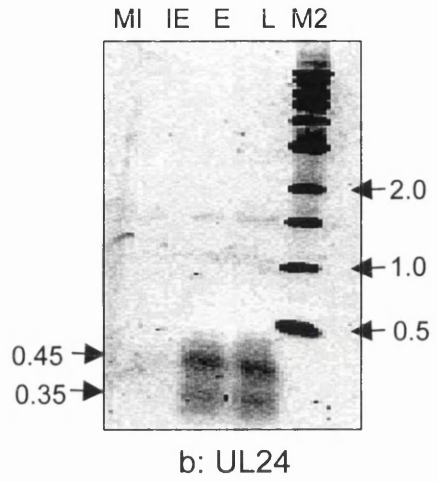
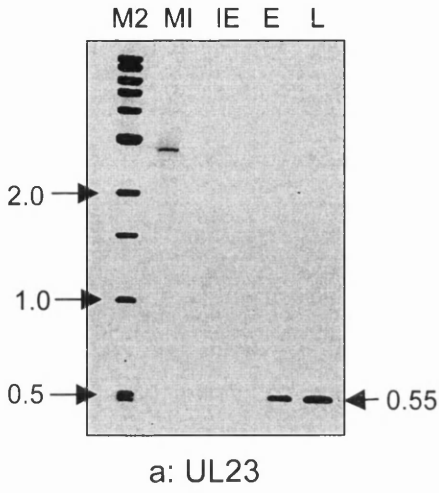
A RACE product representing the 5'-end (at 28855) of the UL23 transcript was detected in approximately equal yields from E and L RNA (Fig. 5.1a). This 5'-end is preceded by a potential TATA element. A higher molecular weight product was obtained from MI RNA and sequencing indicated, as expected, that it did not originate from HCMV. No product was obtained from IE RNA. This suggests that UL23 is an E gene. Chambers *et al.* (1999) were unable to detect UL23 transcripts cells using an HCMV gene microarray, presumably owing to low levels of expression. Inability to detect the UL23 protein in infected fibroblasts (R. Adair, personal communication) is also consistent with the conclusion that UL23 is expressed at low levels. Sequence comparisons between HCMV and CCMV indicate that HCMV UL23 is well conserved in CCMV, but the TATA element is not conserved in the same location in CCMV as it is in HCMV.

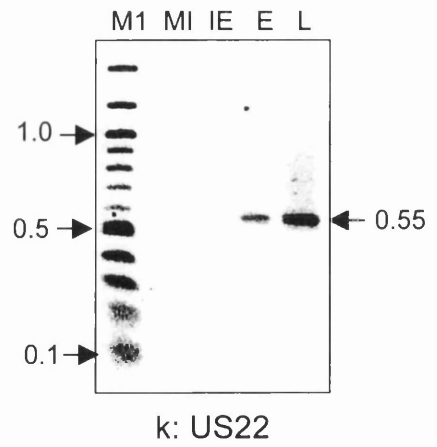
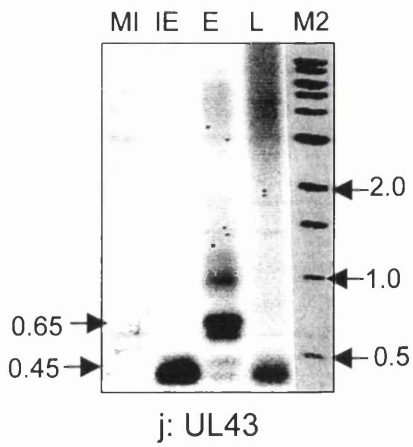
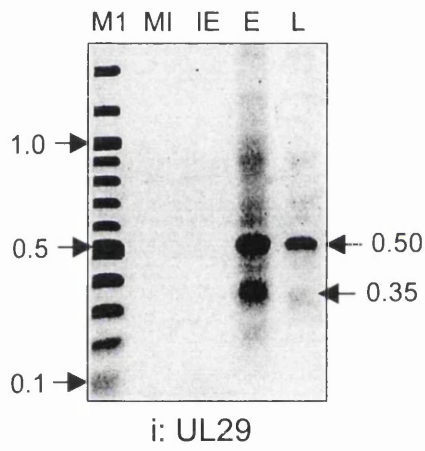
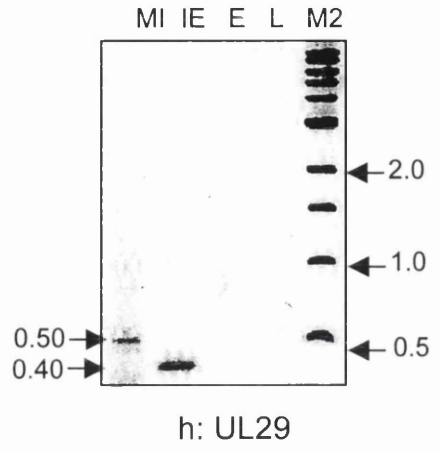
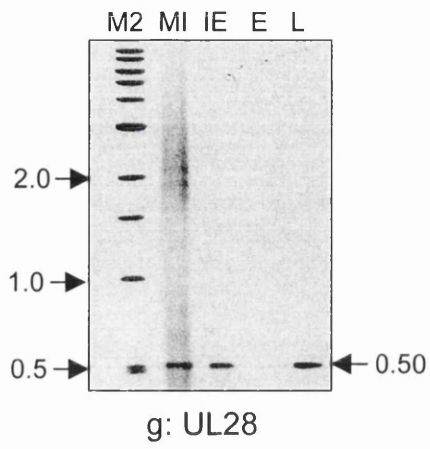
More than one product was obtained from E and L RNA when primer UL24 was used for RACE (Fig. 5.1b). Two bands, the major of 450 bp and the minor of 350 bp, were found to originate from HCMV, indicating the presence of two 5'-ends for UL24 (Table 5.2). The larger product represents a 5'-end located at 30086, which is 25 bp downstream from a potential TATA element, suggesting that the first ATG in UL24 is used as an initiation codon. The smaller product represents a 5'-end at 29994, which is also preceded by a TATA element; in this case the second ATG would be used as an initiation codon. This investigation was extended using two additional GSPs (Fig. 5.1c and 5.1d). Primer UL24 (II) failed to amplify the smaller transcript to detectable levels (Fig. 5.1c), but primer PEL24 amplified both (Fig. 5.1d). HCMV UL24 is conserved in CCMV, including the first ATG codon and the TATA element (TATAAA) of the larger transcript (see Table 5.2). As with UL23, expression of UL24 mRNA in infected cells appeared to be low. Indeed, Chambers *et al.* (1999) failed to detect a transcript from UL24 by gene microarray experiments, and the UL24 protein has proved difficult to detect in infected cells using anti-UL24 monoclonal antibodies (R. Adair, personal communication).

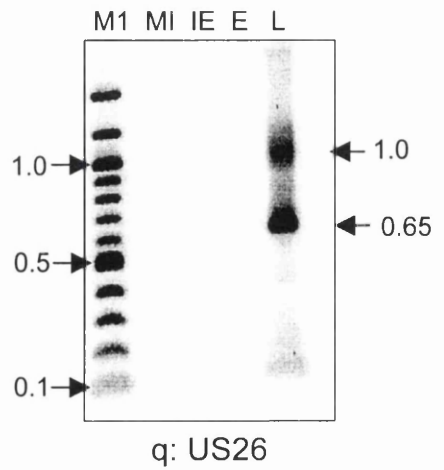
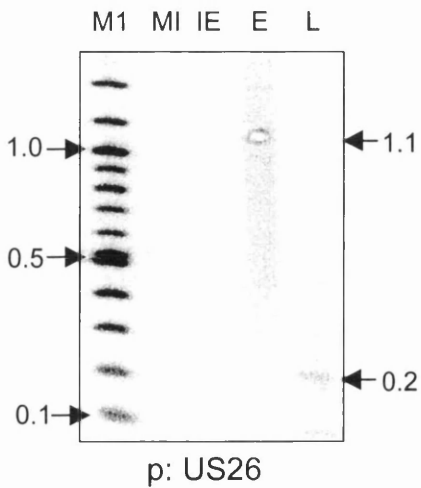
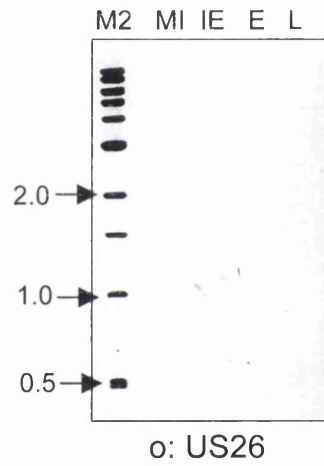
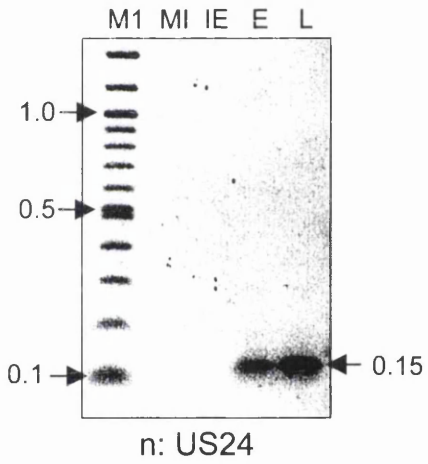
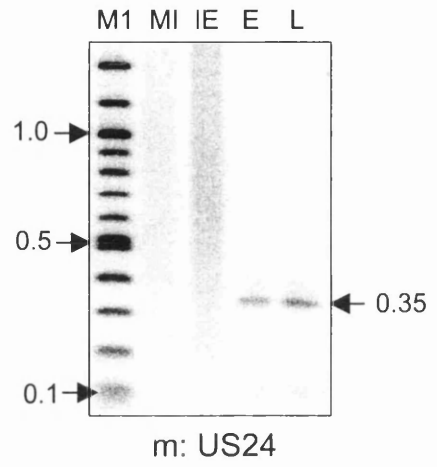
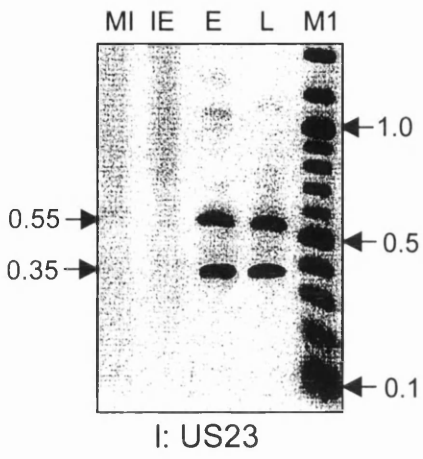
The UL26 protein is a component of the virion (Baldick and Shenk, 1996), and contains a strong transcriptional activation domain. Primer UL26 yielded a single clear band in

Figure 5.1: 5'-RACE of HCMV US22 genes.

EtBr-stained 1-1.5% (w/v) agarose gels, showing representative 5'-RACE PCR products containing regions of the HCMV US22 genes that were amplified from 5'-cDNA with gene-specific primers, (listed in Table 5.1) and UPM. 5'-cDNA was prepared from different kinetic classes (IE, E and L) of AD169 mRNAs and MI RNA. M1 and M2 are 100 bp and 1 kbp ladders, respectively. (a) UL23 with primer UL23. (b) UL24 with primer UL24. (c) UL24 with primer UL24 (II). (d) UL24 with primer PEL24. (e) UL26 with primer UL26. (f) UL28 with primer UL28 (II). (g) UL28 with primer UL28. (h) UL29 with primer UL29 (II). (i) UL29 with primer UL29 (III). (j) UL43 with primer UL43 (III). (k) US22 with primer US22. (l) US23 with primer US23. (m) US24 with primer US24. (n) US24 with primer PES24. (o) US26 with primer US26. (p) US26 with primer US26 (II). (q) US26 with primer US26 (III).







approximately equal amounts from E and L RNA (Fig. 5.1e), indicating that UL26 is an E gene. Sequence analysis of 5'-cDNA clones showed that the 5'-end of UL26 is located at 32785 and is preceded by a potential TATA element. In CCMV, UL26 is highly conserved, including the first ATG and the TATA element.

The HCMV US22 product is an early protein that locates in the cytoplasm and nucleus and is secreted from cells (Mocarski *et al.*, 1988). A sequence error in AD169 US22 (see section 3.1) resulted in 5' extension of the protein-coding region. A 5'-RACE product of 550 bp was obtained from E and L RNA (Fig. 5.1k), suggesting that US22 is an E transcript. Analysis of 5'-sequences indicated a single 5'-end for US22 (Table 5.2), with the first ATG used as an initiation codon. The sequence of 5'-cDNA clones showing the position of the 5'-end of US22 and an electrophoregram of a representative sequence are shown in Fig. 5.2. HCMV US22 is well conserved in CCMV, including the first ATG and the TATA element.

Analysis of the 5'-end of the US23 transcript revealed two major products from E and L RNA (Fig 5.1l). These two bands correspond to initiation at 213512 and 213654 (142 bp apart from each other). The latter 5'-end is preceded by a TATA element, but the former is not. Furthermore, the first ATG would be used as the initiation codon in both mRNAs. The experiment was repeated using a different GSP and led to the same conclusions (data not shown).

Three different primers (US24, US24 (II) and PES24) were used to determine the 5'-end of US24. Each time, a single RACE product was obtained from E and L RNA that corresponds to a 5'-end at 214970, 21 bp downstream from a TATA element (Fig 5.1m and 5.1n). The mapped 5'-end is located downstream from the first ATG codon in the ORF, and the mRNA is likely to use an initiation codon at 214943. Sequence comparisons indicate that HCMV US24 is well conserved in CCMV, including the first ATG (at 215093). It is possible that US24 encodes two transcripts, of which the larger was not detected in this study.

AAAGGCTTTGATATCACTGGCCACCTCGTAGAGCCCGTCGGTCT***TC***CCAGTCGTAGACGTAGACGGT***C***CGTAATGACTTAGCATGAGCACGCAGGGCAGT 211200
 F A K I D S A V E Y L G D T E W D Y V Y V T **C** Y H S L M L V C P L

TCCTGCGCCTGCTTGGTGTTCGTGTTAGATCGCTGTCGGGTGGACGCACGGCTAGTACACCGACGGCTTCCAGGGTGTATCGCAGCAGAGATAGTCGG 211300
 E Q A Q K T N R T L D S D P P R V A L V G V A E L T D D C C L Y D A

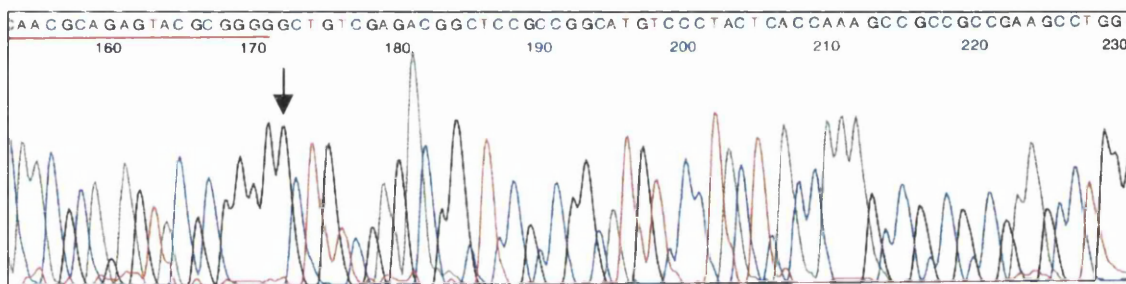
CGGCCAGAGAACGTGCGTAAATCTGCGGGATGGCGCCTGTTTCGCGCATCACTAGGAACCAAGTTGGCGGGGTTCGCGAGTGCTACGGTGGTTCCTTGGTG 211400
 A L S R A Y I Q P I A A Q E R M V L F W N A P N R L A V T T G Q H

GCGTTGCACGTAGGTTCTCAGCGCCGGAGGATCGTACTGGCGCAGATAGAGGCCTTGCAGCATCGATAACGTCTTTTGAAGACGGTGTTCCTAAATGA 211500
 R Q V Y T R L A P P D Y Q R L Y L G Q L M S L T K Q F V T N R F Q

AAAACGCCGTAGTCGCAGCGGATAGCATCTTCGCAGCGCTCGTCGGCTGTCGGAGATAGGTGCCCCAGGCTTCGGCGGGGGCTTGGTGAGTAGGGACA 211600
 F V G Y D C R I A D E C R E D R Q R L Y T G W A E A A A K T L L S M

TGCCGGGGAGCCGTCTCGACAGCGAGTCGGATAAAGCGCGCTGCGCGAAAGCTTAAATATAAGGAGCAGCGTCAGACGAATCGCGGCTGGTGGCCCGGGGG 211700

I



II

Figure 5.2: Sequences at the 5'-end of the US22 mRNA.

Panel I shows the region containing the 5'-end of US22 in the AD169 genome, with coordinates on the right. The gene is oriented right to left. The position of the RACE-PCR primer is shown in bold italic and its orientation by the arrow. The vertical arrow indicates the position of the 5'-end of the US22 mRNA as represented in the majority of 5'-cDNA clones. The putative TATA element is underlined in red. Panel II shows an electropherogram of the sequence of the 5'-RACE product (lower strand) with the 5'-end indicated by an arrow and the SMART sequence underlined in red.

ORF	Condition of expression ^a	Position of 5'-end ^b	Putative TATA element ^c	Position of TATA box ^d	Position of ATG codon	Distance ^e	
						In HCMV	In CCMV
UL23	E	28855	TACTTC	28880-28875 (20)	28717	156	
UL24	E	29994	TGCTTA	30025-30020 (25)	29835	182	
		30086	TATAAA	30117-30112 (25)	30009	102	164
UL26	E	32785	TATAAG	32814-32809 (24)	32775	33	33
UL28		No RACE product					
UL29		Cellular ^g					
UL36 ^f	IE	49863	TATAAA	49890-49885 (21)	49776	108	142
UL43	Not detected	Cellular ^g					
US22	E	211626	TATTAA	211658-211653 (28)	211601	52	49
US23	E	213512			213492		
		213654	TGATTA	213685-213680 (25)	213492	161	
US24	E	214970	TTATAC	214996-214991 (21)	214943	48	
US26		Cellular ^g					

Table 5.2: 5'-ends of HCMV US22 genes.

^a Conditions of expression as determined by RACE. ^b Position of 5'-ends as represented in the majority of 5'-cDNA clones. 5'-end located downstream from the first ATG in the ORF as described by Chee *et al.* (1990) is shown in blue. ^c Position of the putative TATA element in the AD169 genome. The number of nucleotides between the 5'-end and the closest TATA element is shown in brackets. TATA boxes conserved in CCMV are shown in red. ^d Position of the first ATG codon in the ORF. ^e Number of nucleotides between the putative TATA box and the first ATG codon in the ORF. ^f RACE result shown in section 4.1. ^g PCR products not originating from HCMV.

The 5'-ends of the UL28, UL29, UL43 and US26 mRNAs were not detected in RACE experiments (Table 5.1). RACE products produced using primers UL28 (II) and UL28 are shown in Fig. 5.1f and 5.1g, respectively. The single band detected (Fig. 5.1g) was also generated from MI RNA, and sequence analysis showed that it did not originate from HCMV. Primers UL29 and UL29 (II) yielded a single band from L RNA (Fig. 5.1h) and a larger band from MI RNA. Primer UL29 (III) yielded bands only from E and L RNA (Fig. 5.1i). None of these bands originated from HCMV. As can be seen in Fig. 5.1j, primer UL43 (III) yielded multiple bands from IE, E and L RNA, and faint bands from MI RNA. None of these bands originated from HCMV. Primer US26 (II) generated products from E and L RNA (Fig. 5.1p), and primer US26 (III) generated products from L RNA (Fig. 5.1q), but none originated from HCMV. It is likely that expression of UL28, UL29, UL43 and US26 is below detectable levels.

5.1.2 Sequences of 3'-ends

3'-ends were identified for eight of the 13 US22 genes: UL23, UL24, UL26, UL36, UL43, US22, US23 and US24. UL28, UL29, TRS1, IRS1 and US26 were not analysed. The 3'-RACE products are shown in Fig. 5.3 and positions of the 3'-ends of US22 genes are listed in Table 5.4.

The 3'-end of the UL36 transcript mapped by Kouzarides *et al.* (1988) was confirmed by RACE (see section 4.1).

Primers UL23-3' and UL24-3' each yielded a single band from E and L RNA (Fig. 5.3a and 5.3b). However, the UL24 product was approximately 1000 bp larger than that of UL23. Analyses of these products indicated that UL23 and UL24 share a common 3'-end at 27788, preceded by a polyA signal. An identical polyA signal is also present in CCMV UL23 in similar position (Table 5.4).

ORF	Primer	Sequence (5'- 3')	Position ^a	PCR product(s) ^b
UL23	UL23-3'	GACTGGAGGGTCATGGTCGGCAGT	28141-28118	500
UL24	UL24-3'	CGGCCTGGGAACCGTCAGCCTCAA	20200-29177	1500
UL26	UL26-3'	CTACGAGATCAGCCGTGAGAACTT	32500-32477	750
UL43	UL43-3'	AAAGAAGGGGCTCTTTGCTCGAAG	54851-54874	600, 1700
US22	US22-3'	TCTTTCCGCCGCCCATGTGCCGCG	210241-210218	550
US23	US23-3'	GGAAGACGCGGTTGCCGCGGTAT	212014-212037	2300
US24	US24-3'	ACACCGTCACCTGTGCCGCACCA	214053-214076	500

Table 5.3: Primers used in 3'-RACE to map the 3'-ends of US22 genes.

^a Position of primers in the AD169 genome (upper strand).

^b Approximate sizes (bp) of PCR products excised for cloning.

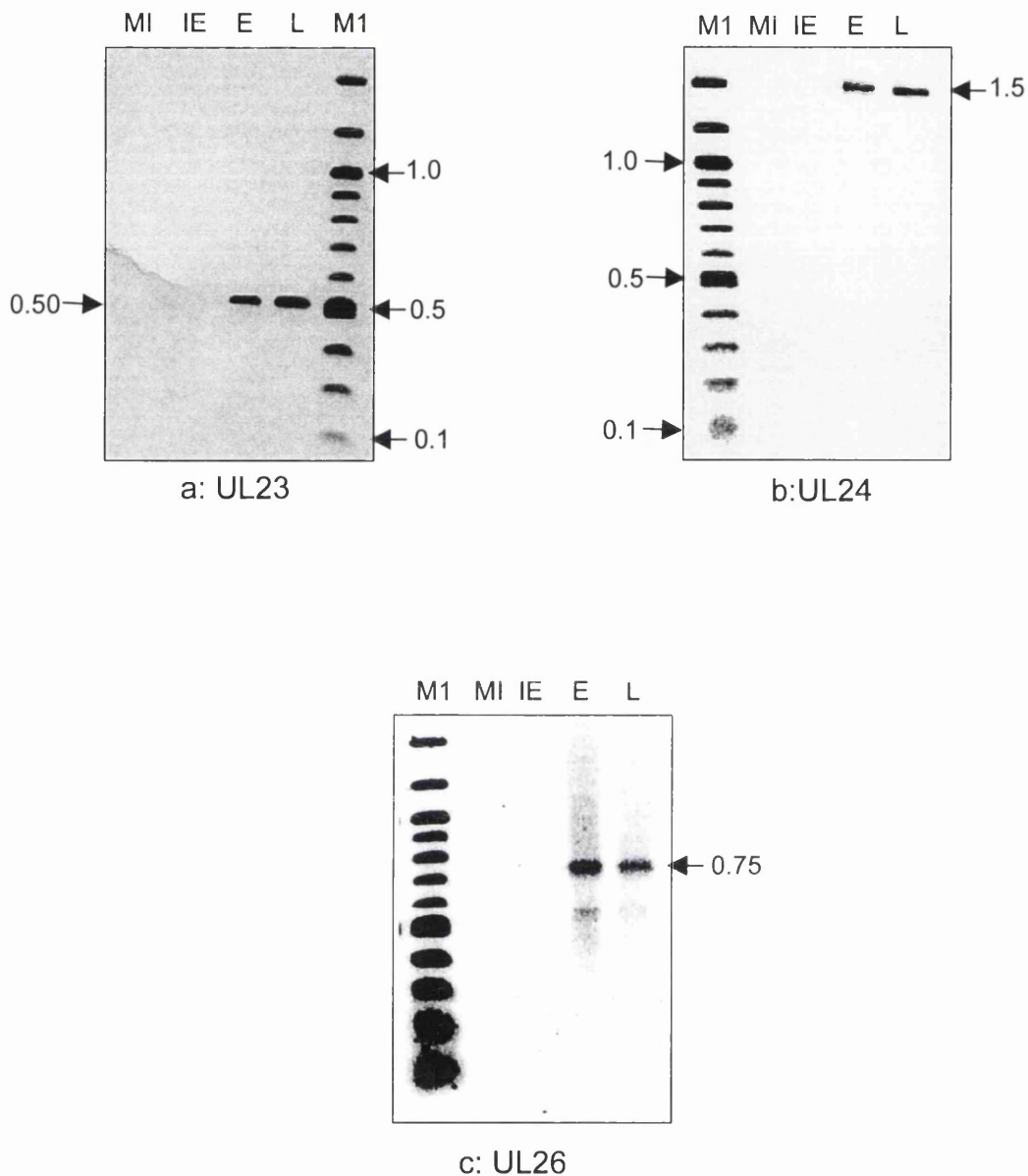
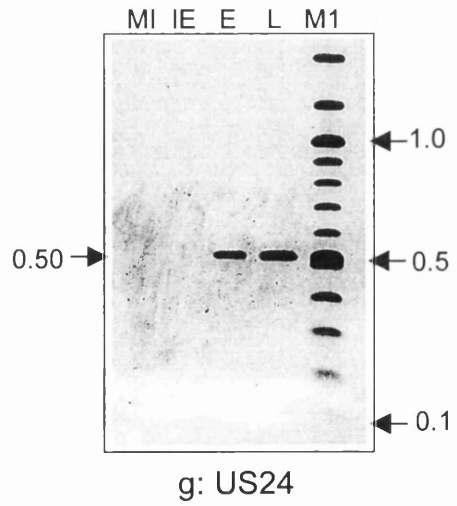
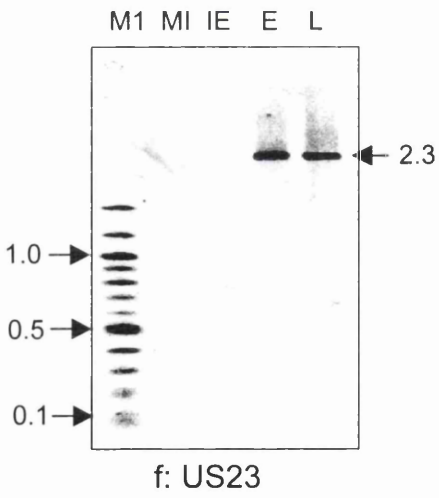
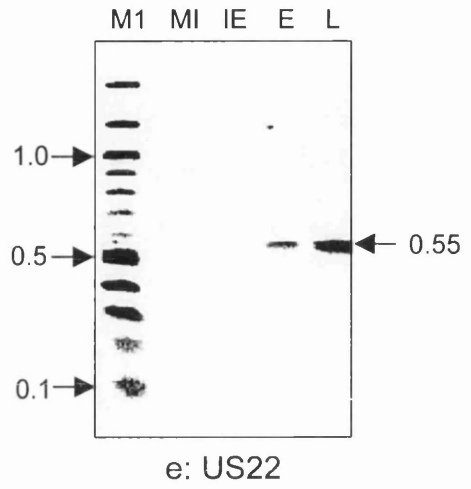
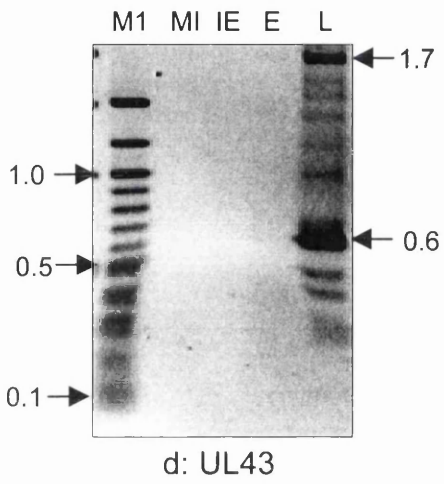


Figure 5.3: 3'-RACE of HCMV US22 genes.

EtBr-stained 1-1.5% (w/v) agarose gels showing representative 3'-RACE PCR products containing regions of HCMV US22 genes that were amplified from 3'-cDNA with GSPs (listed in Table 5.3) and UPM. 3'-cDNA was prepared from different kinetic classes (IE, E and L) of AD169 mRNAs and MI RNA. Sizes are in kbp. M1 is a 100 bp ladder. (a) UL23 with primer UL23-3'. (b) UL24 with primer UL24-3'. (c) UL26 with primer UL26-3'. (d) UL43 with primer UL43-3'. (e) US22 with primer US22-3'. (f) US23 with primer US23-3'. (g) US24 with primer US24-3'.



↓

AACGCCGGCGGTTATCGCCGAGATTCTGTCTAAATACACGAAGCGAACTAGAAAACGCACACACGTGATTTGCAAAAAGAAAGCAGCTGCCGGCTTATTAT 209800

TTTATTAAAAATTTATCTGTGCAGAATCATAAGTTTATGATGAATAAAAACGGGGAAAGGAATCTGCTTTTAGGGACCCGGGTCTGGTCCGTCGTCTCC 209900
 - P G P D P G D D G

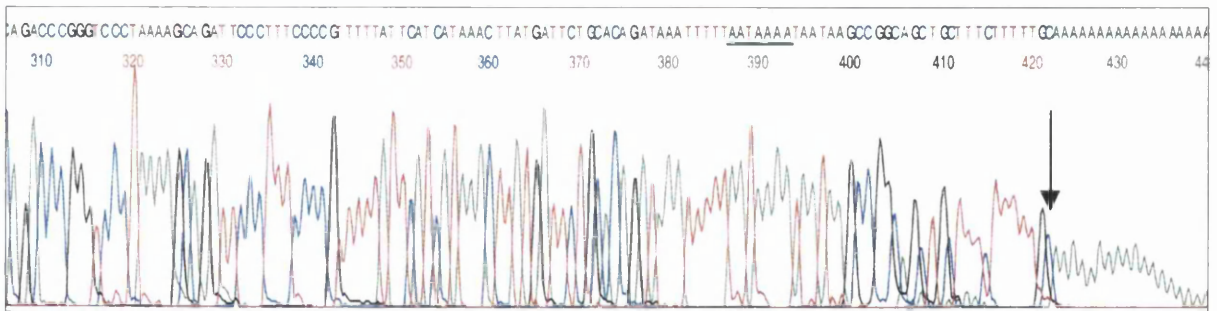
CATCTGGTCGGGTTTCGGGGATGGGGACCTGTTTCAGCGTGTGTCCGCGGGCGTGCATGGCTTTTGTCTCGCCGGCCGCGCTGTAACCAGGCCTCTTTCTCT 210000
 M Q D P E P I P V Q K L T H G R A H M A K A R R G R Q L W A E K E

GTGGTCGGCGAGTCTTCCGACGGTAGGGAGCCTGGGAGTCCATCGCTTCAGGCCACCCTGCTCCCTCGACCGTGTGTCGCTCCTCGTTTTCGCTAT 210100
 T T P S D E S P Y P A Q S D M A E P G G S T G E V T T D D E N E S N

TACACGGGTTTCTGGAGTATCGCCTATACGGTTGGCGATTCTCCGGGGCGGGCGCTCTCGTCCTCGTCGCTGCTATCGCCGCCGGTAA***TTCGACGCC*** 210200
 C P T E P T D G I R N A I R R P R G S E D E D S S D G G P L E V G

GCATTTCGTGTACGGAACGCGGCACATGGGCGCGGAAAGAACTTGGGCATGCGAAAGCAGCGTTGTCCATCCACGGTCTGCGTGGTTTCATCGTTATCC 210300
 C E N Y P V R C M P P P F F K P M R F C R Q G D V T Q T T E D N D

I



II

Figure 5.4: Sequences at the 3'-end of the US22 mRNA.

Panel I shows the region containing the 3'-end of US22 in the AD169 genome, with coordinates at the right. The gene is oriented right to left. The position of the RACE-PCR primer is shown in bold italic and its orientation by an arrow. The vertical arrow indicates the position of the 3'-end of the US22 mRNA as represented in the major proportion of 3'-cDNA clones. The polyA signal is underlined in green. Panel II shows an electropherogram of the sequence of the 3'-RACE product (lower strand of the AD169 genome), with the 3'-end indicated by an arrow and the polyA signal underlined in green.

ORF	Conditions of expression ^a	Position of 3'-end ^b	Polyadenylation signal	Position of polyA signal ^c	Distance ^d	
					In HCMV	In CCMV
UL23	E, L	27788	AATAAA	27814-27809 (20)	48	71
UL24	E,L	27788	AATAAA	27814-27809 (20)	1118	1123
UL26	E, L	31878	CATAAA	31892-31887 (22)	317	410
UL36 ^e	IE, E, L	48094	AATAAA	48119-48114 (20)	123	114
UL43	L	53135	AATAAA	53161-53156 (20)	1660	1566
US22	E, L	209771	AATAAA	209806-209801 (29)	65	41
US23	E, L	209771	AATAAA	209806-209801 (29)	1895	2001
US24	E, L	209771 213539	AATAAA AATAAA	209806-209801 (29) 213523-213518 (20)	3879 20	3904 20

Table 5.4: 3'-ends mapped for HCMV US22 genes.

^a Conditions of expression as determined by 3'-RACE. ^b Position of 3'-end in the AD169 genome as represented by majority of the 3'-cDNA clones. ^c Position of putative polyadenylation signal in the AD169 genome. The number of nucleotides between the 3'-end and polyadenylation signal is shown in brackets. Polyadenylation signals conserved in CCMV are shown in red. ^d Number of nucleotides between the polyadenylation signal and the stop codon in the ORF. ^e RACE result shown in section 4.1.

Primer UL26-3' yielded products from E and L RNA (Fig. 5.3c) that mapped the 3'-end at 31878, 22 bp downstream from a non-canonical polyA signal (CATAAA) that is also conserved in CCMV. Analysis of the AD169 DNA sequence indicated that this polyA signal is likely to be shared by UL27, UL28, UL29 and UL30.

UL43 3'-RACE products were obtained from L RNA (Fig. 5.3d). Only the largest product (1.7 kbp) originated from HCMV, corresponding to a 3'-end at 53135. Consistent with this, Chambers *et al.* (1999) reported that the UL43 transcript was detected only in L mRNA using gene microarray technology. Sequence analysis indicated that the 3'-end of UL43 contains a polyadenylation site 20 bp upstream that is probably shared by UL40-UL46. A similar polyadenylation signal is also present in CCMV UL43, but further upstream (Table 5.4).

3'-RACE products from US22, US23 and US24 were generated from E and L RNA (Fig. 5.3e, 5.3f, 5.3g, respectively), suggesting that they are E transcripts. The products obtained from US22 and US23 corresponded to the same 3'-end at 209771, and the product obtained from US24 corresponded to a different 3'-end at nucleotide 213539. Both 3'-ends have a polyA signal (AATAAA) 20-29 bp upstream.

5.2 Mapping the 5'-ends of US22 genes by primer extension analysis

To confirm the 5'-ends of transcripts from certain US22 genes, a primer extension experiment was carried out using the Primer Extension System-AMV Reverse Transcriptase kit. A 24 bp GSP was designed from the region approximately 100 bp downstream from the 5'-end of each target gene determined by RACE (Table 2.3).

Four US22 genes were analysed using primer extension assay: UL24, UL36, US22 and US24 (Fig 5.5). All except US24 yielded products. Those from UL36 and UL24 were very faint compared to that from US22. This may reflect abundance of cognate mRNAs, efficiency of primer labelling or efficiency of the primer in the RT reaction. The 5'-end of

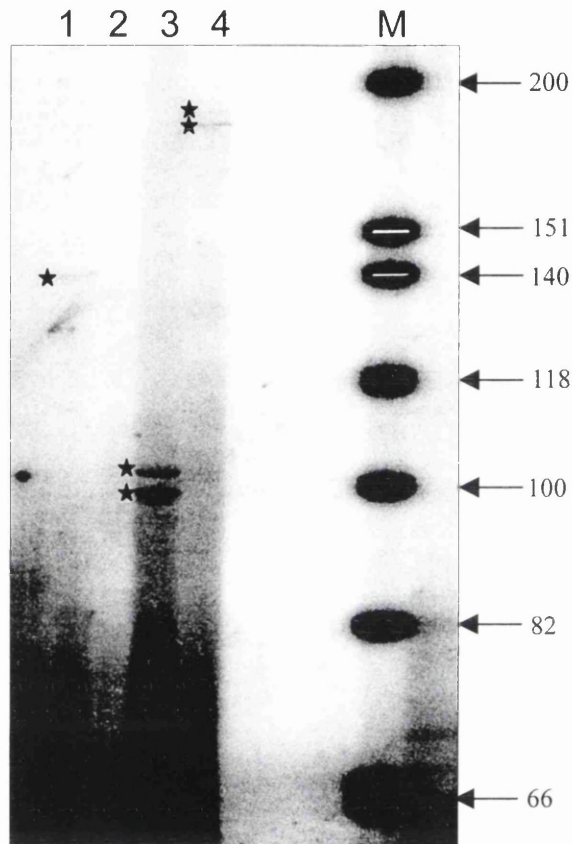


Figure 5.5: Primer extension analysis of transcripts from the US22 gene family.

Phosphorimage showing products (marked with black stars) obtained from primer extension assay. Sizes are shown in nucleotides. (1) UL36 with primer PEL 36. (2) US24 with primer PES 24. (3) US22 with primer PES 22. (4) UL24 with primer PEL 24. M is a ϕ X174 *Hinf* I DNA marker.

ORF	Primer	Sequence (5' to 3')	Position ^a	5'-ends ^b	Distance ^c (bp)	Product ^d size
UL24	PEL24	CGCGGCCAGCGTGGCGAGGACACT	29905-29928	29994 30086	90 182	185,187
UL36	PEL36	TCGGATGGCGATGCAGCCGTAGGC	49726-49749	49863	137	139
US22	PES22	AGCATCTTCGCAGCGCTCGTCGCG	211524-211548	211626	103	98, 102
US24	PES24	ACAGCGTCCCTCGTTACGATGAAC	214902-214925	214970	69	-

Table 5.5: Primers extension analysis of transcripts from the US22 gene family.

^a Position of primer in the AD169 genome (upperstrand).

^b Position of 5'-end in the AD169 genome determined by RACE.

^c Distance between the 5'-end of the primer and the 5'-end of the mRNA determined by RACE.

^d Approximate size of the product (measured from the gel) determined by primer extension assay.

The accuracy of primer extension data is estimated at ± 5 nucleotides.

the UL36 transcript mapped by Kouzarides *et al.* (1988) was confirmed by primer extension (see section 4.3). The overall results are shown in Table 5.5.

Two 5'-ends were determined for UL24 by primer extension assay, differing by approximately 2 nucleotides. These represent 5'-ends at 30089 and 30091 (each ± 5). Consistent with this, four closely spaced 5'-ends within the region 30082-30098 were determined for the larger transcript from UL24 gene using RACE (see Appendix 1), but the 5'-end at 30086 was represented by the majority of 5'-cDNA clones (see section 5.1). The 5'-end (at 29994) of the smaller transcript from UL24 gene identified by 5'-RACE was not detected in this experiment, due to high background in this region of the gel. Two 5'-ends were detected for US22 within approximately 4 nucleotides (Table 5.5) at 211622 and 21625 nucleotides (each ± 5). A single 5'-end (211626) was detected for US22 using RACE (section 5.1).

In conclusion, 5'-ends for UL24, UL36 and US22 determined by primer extension were consistent with those mapped using RACE.

5.3 Mapping US22 genes by northern blot

Northern blot experiments were carried out to determine the expression kinetics and sizes of transcripts from genes in the US22 family (described in section 2 and 4.2). Primers for making single-stranded RNA probes using Lig'nScribe kit (2.2.7.2) are listed in Table 2.1. The PCR products (5'-RACE product of the cognate gene + 64 bp) generated by the Lig'nScribe reaction (for details see section 2.7 and 4.2) were examined on an agarose gel and are shown in Fig. 5.6. The PCR product was then used as a template in an *in vitro* transcription reaction using the MAXIscript kit to produce labeled single-stranded RNA probes (section 2.2.7.3). All blots, except those assaying the UL23 and UL24 mRNAs, were rehybridised with a control DNA probe (GAPDH₂) to normalise the amount of RNA loaded in each lane.

In the present study, the kinetic classes of nine US22 genes (UL23, UL24, UL26, UL28, UL29, UL36, US22, US23 and US24) were determined by northern blotting (Fig. 5.7a to

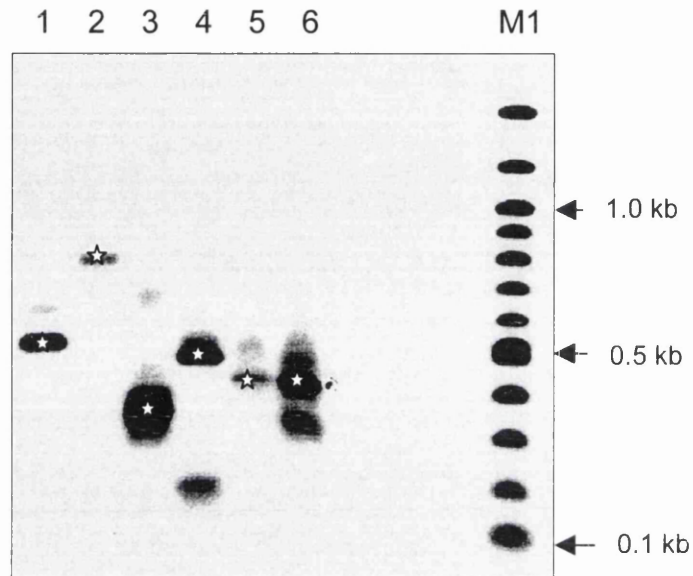


Figure 5.6: Gel electrophoresis of Lig'nScribe PCR products used to generate RNA probes.

EtBr-stained 1% (w/v) agarose gel showing representative Lig'nScribe PCR products amplified from adapter-ligated 5'-RACE products using GSPs (listed in Table 2.1) and adapter primer 1. Fragments of expected sizes (64 bp bigger than the PCR product) are indicated by white stars. (1) UL24. (2) US22. (3) UL29. (4) UL26. (5) US23. (6) UL43. M1 is a 100 bp ladder.

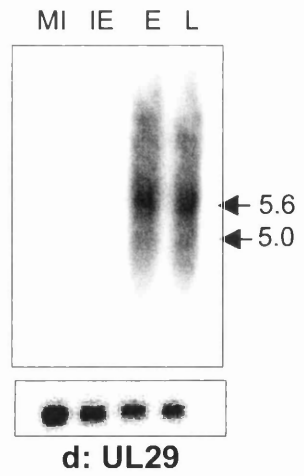
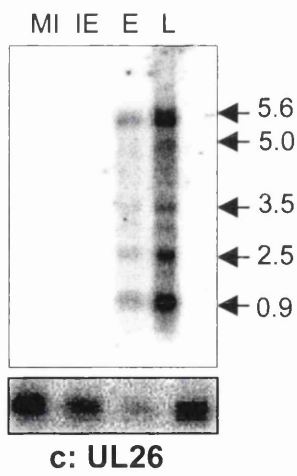
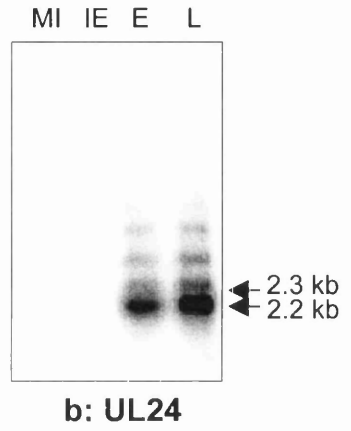
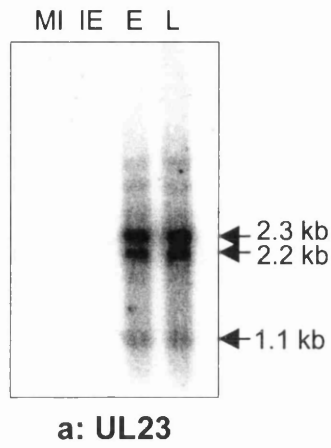
5.7h). TRS1, IRS1 and US26 were not analysed. The results are summarised in Table 5.6 in comparison with published data.

The kinetics of expression and size of the UL36 transcript were consistent with previously published data (Mocarski, 1988) (see section 4.2).

The UL23 probe hybridised to three major bands of 2.3, 2.2 and 1.1 kb (Fig. 5.7a). These bands were of similar intensity in E and L RNA and therefore represent E transcripts. Consistent with this, UL23 RACE products (both 5'- and 3'-) were generated from E and L RNA (see section 5.1.1 and 5.1.2). Since the RACE experiments mapped the 5'-end of UL23 to 28855 and the 3'-end to 27788 (1068 nucleotide transcript), the 1.1 kb band is the UL23 transcript. The 3'-end of UL24 is also at 27788, with two 5'-ends at 30086 and 29994 (2299 and 2207 nucleotide transcripts), indicating that the 2.3 and 2.2 kb bands are from UL24. A blot hybridised with the UL24 probe also identified the 2.3 and 2.2 kb bands (Fig. 5.7b). Only the larger transcript of UL24 was detected in the primer extension experiment (see section 5.2). Chambers *et al.* (1999) failed to detect UL23 and UL24 mRNAs using gene microarray technology.

The UL26 probe hybridised to several bands putatively corresponding to a set of 3'-coterminal transcripts from UL26 to UL30 (Fig. 5.7c). Since the RACE experiments indicated that the 5'-end of UL26 is at 32785 (section 5.1.1) and the 3'-end at 31878 (section 5.1.2; 908 nucleotide transcript), the 0.9 kb band is the UL26 transcript. The 5'-end of UL30 (see section 7.1) is at 37525 (5648 nucleotide transcript), indicating that the 5.6 kb band is from UL30. The 5.0, 3.5, and 2.5 bands may represent the UL29, UL28 and UL27 mRNAs, respectively. Normalisation with a control probe (GAPDH₂) indicated that each transcript was approximately equally abundant in E and L RNA. The UL29 probe hybridised to a barely discernible 5.0 kb band (corresponding to the UL29 transcript) and a 5.6 kb band (corresponding to the UL30 transcript) from E and L RNA (Fig. 5.7d). Thus, UL26, UL29 and UL30, and probably UL27 and UL28, are E genes.

The US22 probe hybridised to five bands of 1.9, 2.5, 3.8, 3.9 and 5.4 kb (Fig. 5.7e). The 3.8, 3.9 and 5.4 kb transcripts were also detected with the US23 probe (Fig. 5.7f). These bands were of similar intensity in E and L RNA and therefore represents E transcripts,



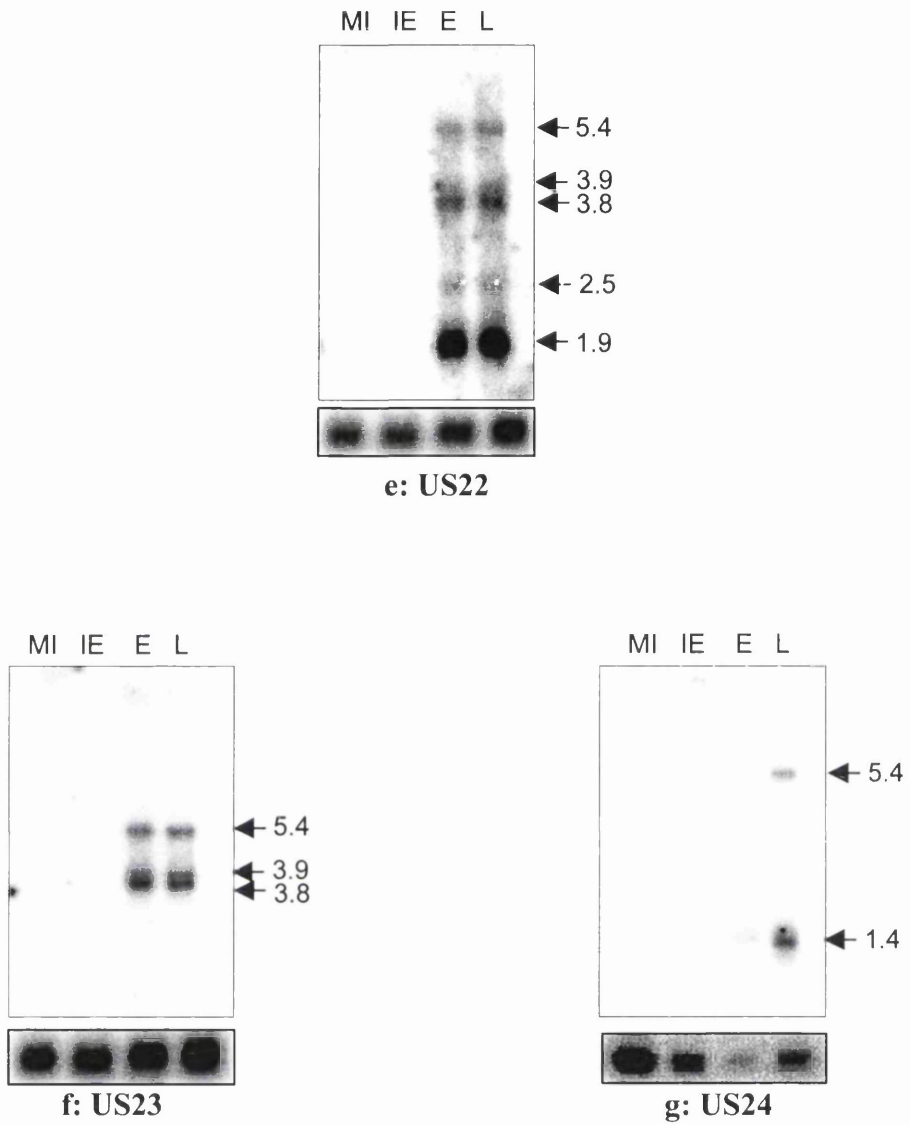


Figure 5.7: Northern blot analysis of transcripts from members of the US22 gene family.

IE, E and L RNA prepared from HCMV-infected cells, and MI RNA, were hybridised with gene-specific probes (top) and rehybridised (except (a) and (b)), with GAPDH₂ probe (bottom). Single-stranded RNA probes were designed from specific genes as indicated below each panel. Sizes are given in kb.

Gene	Transcript size (kb)	Kinetic class		
		Northern blot ^a	DNA microarray ^b	Northern blot ^c
UL23	1.1	E		
UL24	2.2, 2.3	E		
UL26	0.9	E	E	
UL28	3.5	E	E	
UL29	5.0	E	L	
UL36 ^d	1.6	IE	E	IE
UL43	ND			
TRS1	NA			
IRS1	NA			
US22	1.9	E	E	E
US23	3.8, 3.9	E	E	
US24	1.4, 5.4	E	E	

Table 5.6: Kinetics of expression of HCMV US22 genes.

^a Northern blot analysis conducted in this study.

^b DNA microarray analysis by Chambers *et al.* (1999).

^c Northern blot analysis by Mocarski (1996).

^d Northern blot result shown in section 4.2.

NA: Not analysed.

ND: Not detected.

which is consistent with the previous reports of Chambers *et al.* (1999) and Mocarski (1996). Since the RACE experiments indicated that the 5'-end of US22 is at 211626 and the 3'-end at 209771 (1856 nucleotide transcript), the 1.9 kb band is the US22 transcript. The 3'-end of US23 is also at 209771, with two 5'-ends at 213654 and 213544 (3884 and 3774 nucleotide transcripts), indicating that the 3.9 and 3.8 kb bands are US23 transcripts. The 5.4 kb band probably represents the US24 transcript. The US22, US23 and US24 genes are thus 3'-coterminal, sharing the polyA signal at 209771. The 2.5 kb RNA does not correspond to any known gene.

The US24 probe hybridised to two bands of 1.4 and 5.4 kb (Fig. 5.7g), each, after normalisation, of approximately equal abundance in E and L RNA. US24 is thus an E gene. Consistent with this, Chambers *et al.* (1999) reported that US24 falls into the E expression class. The larger transcript (5.4 kb) was also detected using the US22 and US23 probes (Fig. 5.7e and 5.7f). However, the RACE experiments (see section 5.1.1 and 5.1.2) showed a 3'-end for US24 at 213539 and a 5'-end at 214970 (1441 nucleotide transcript), indicating that the 1.4 kb band is a US24 transcript. The larger US24 transcript is thus 3'-coterminal with US23 and US22, and the smaller US24 transcript terminates between US23 and US24.

The UL43 transcript was not detected by northern blotting using two different RNA probes. This result, and the inability to determine the 5'-end of UL43 by RACE (see section 5.1.1), is consistent with the conclusion that UL43 is expressed at low levels in infected fibroblasts.

5.4 Spliced genes in the US22 gene family

Two members of the US22 gene family were predicted to be spliced (UL36 and UL28). Splice sites in UL36 mapped by Kouzarides *et al.* (1988) were confirmed by RT-PCR (see section 4.4).

UL28 – Splicing in UL28 was predicted by Davison *et al.* (unpublished data) on the basis of amino acid sequence comparisons between HCMV and CCMV. Comparisons between the HCMV UL28/UL29 and the cognate sequence of CCMV UL28/UL29 indicated that

```

CAGATTGGCGCGGTTCGTGCAGTATCTGGGAGAGTTCGTACATGCCCGCAAAGGTGTGCTTAAACCACGCGCCCTCTACGATCTCATCCACGTAGTCGCGC 35300
C I P A T T T C Y R P S N T C A R L P T S L G R A R - S R M W T T A S
UL28 L N A R D H L I Q S L E Y M G A F T H K F W A G E V I E D V Y D R
S Q R P R A T D P L T R V H G C L H A - V V R G R R D - G R L R A

TCAAAGAAGCTGTACACGGCAAAGAGCCGTTCTCAAAAACCTCGCCGAACGAGAGCCCCAGCACGTACACCTTGTCTCGCCGGGAGGTACGCAAAGG 35400
L S A T C P L S A T R L F S A S R S G W C T C R T R A P C T R L P
E F S Y V A F L G N E F F E C E F S L G L V Y V K D E G P L I A F A
- L L Q V R C L P R E - F V R R V L A G A R V G Q G R R A P V C L

CGTGCCCGTGCCTGGAGACCCAGATCTCGGGCGCGTGTTCGCGTCCGGCACGCATTTCGTACACACTGACGAGGCCGATAAAGTACAAGCGCCAGCCTG 35500
T G T G P S G S R P R R T Q T R C A N T C V S S A S L T C A A L R
H G H G S V W I E P A T N A D F V C E R Y V S L G I F Y L R D P S W A Q
R A R A R L G L D R A G H K R G A R M R V V C Q R P R Y L V L P W G P

GCGCAGGCACGAGAAGCGCCGTAGTCTTGTGATCGCGCACCCCAAGTACTGAGTGTGCGCCAGCATGATGCCGTGCAGCGCGGCCAGCACAGC 35600
A C A R S A G T P R T I A C W G L T S L T A W C S A T C R R G A C R
R L C S E R R Y T K H D R V V G F Y Q T D G L I G H L E P P W C L
A P V L L A P L D Q S R A G G W L V S H R G A H H R A A A A L V A

GGGAGCCAACGACCCGCGTGGCGCGCAGTAGCGTGCAGGTGAACCCGCTCGCACGCTCGCGCGGCTTCGGGGCGCTTGTGGGTCCAGGCATCACGCA 35700
S G V V R R P A C T A S C T F G A R V S A R S R A S T P G P M V C
F L W R G A T A R V Y R Q L H V G S A K E R P K P R K H T W A D E L
P A L S G G H R A R L A A P S G R E C A R A A E P A Q P D L C - A

GACCGCGCCAGATGCTGCTGAACTTGGGCTGCCCGCGCAGATAGAGCGACGAGAGCGAGTCAAAGTAGCCACGACGAGCctgtcgggagacacaagagc 35800
V A G S A A S S P S G A C I S R R S R T L T A W S S G T P L C L L
G R W I S S E K F Q Q R L Y L S S L S D F Y G V V L R D P S V L A
S R A L H Q Q V Q A R A S L A V L A L - L L G R R A Q R D S V C S R

gcgaaaatcaaacctagagcgcgacggtgaaaaaacccgaccagagcgcggtgtctcaaacacgctacttctcggtataaaaaaccgctcgccctatttc 35900
A F I L G L A V V T F F G V L L A H R L C A V K P - L F V T A R N R
R F D F R S R R R H F F R G S A R T E F V S S E T I F V G D G - K
S F - V - L S S P S F V S W F R T D - V R - K R N Y F C R R G I E

tgggcgcgtgtacactgatgactcacCTACGCTTTTTGAACGGCAGTCTCAGCTCGGGATTGGCCTCGTACAGCGAGCTCGCGTCCACGGGGCCGATGCT 36000
UL29 P A H V S I V - R R K K F P L R L E P N A E Y L S S R D V P G I S
Q A R T C Q H S V - A K Q V A T E A R S Q G R V A L Q P G R P R H E
P R T Y V S S E G V S K S R C D - S P I P R T C R A A T W P A S A

```

I

Figure 5.8: Comparison of the region containing AD169 UL28/UL29 with the cognate CCMV sequence.

Panels I and II show three-frame translations of the relevant regions of the AD169 and CCMV sequences, respectively, with nucleotide coordinates on the right. Genes are oriented right to left. ORF UL28 and UL29 are shown in red, and conserved amino acid residues are underlined. The first ATG codon in the UL28 is shown within a box for each genome. Possible introns are shown in lower case.

AGGCCGGCGCGATCGTGCAGCAGCTGCGAGAGCTCGTACATGCCGGCGAAGGTGTGCTTGAACCAGGTGCTCTCCACCACCTCGTCCACGTAGTCGCGCT 33400
 P R R S R A A A A L A R V H R R L H A Q V L H E G G G R G R L R A
 A P A I T C C S R S S T C A P S P T S S G P A R W W R T W T T A S
 UL28 L G A R D H L L Q S L E Y X G A F T H K F W T S E V V E D Y Y D R E

CGAAGAAGCTGTAGACGGCGAAGAGGCCGTTCTCGTAAACTCGCTGAACGAGAGGCCAGCACGTAGACCTTGTCTCGCCCGCAGGTAGGCGAAGGC 33500
 R L L Q L R R L P R E R L V R Q V L P G A R L G Q G R G A P L R L R
 S S A T S P S S A T R T F S A S R S A W C T S R T R A R C T P S P
 E F S Y V A F L G N E Y F E S E S L G L Y Y V K D E G P L Y A F A

GTGGCCCTGACTGGAGACCCAGATCTCGGGCCGACGCGAGCGTCCGGCAGCACTCATAGACACTGATGAGGCCGATGAAGTAAAGGGCGCGGCTGA 33600
 P G S Q L G L D R A G R S R G A R V - L C Q H P R H L L P P R G S
 T A R V P S G S R P R W A L T R C A S R S S V S S A S S T F A A P R V
 H G O S S V W I E P A V R A D F V C E Y V S I L G I F Y L R G A Q

CGCAGGCACGAGAAGCGCGGTAGACTCTTCTGGTCGCGCACCACCCGAAATAGTTGCTGTCGCCAGCAACAGGCCGTGGAGGGCGGCCAGCAGAGCG 33700
 A P V L L A P L D E P R A G G R F L Q Q R G A V P R P P A A L L A
 C A R S A G T S R R T A C W G S I T A T A W C C A T S P R G A S R
 R L C S E R R Y I K Q D R V V G F Y N S D G L L L G H L P P W C L P

GGATCCAGCAGCCGCGTGGCCCGCACGTAACGCTGCAGCTGGACTCCGCCAAGCGCTCCCGCAACTTCTGACGCTTGTGGATCCAGGCCGTGCGCTAG 33800
 P D L S R G H G A R L A A A P S R G L A G A V E S A Q P D L R R T T
 S G A V P R P G C T V S C S S E A W A S G C S R V S T S G P T A Y
 I W R G A T A R V Y R Q L O V G G L R E R L K Q R K H I W A D R L

TCCCCGCGAGAGGCTGCTGAACTTGGGCTGCCCGCGCAGGTAGAGTGACGAGAGCGGCTCGAAGTAACCGACGATCAGCctgcggtcgacgatagagggga 33900
 G A L P Q Q V Q A A R A P L T V L A A R L L R R D A Q P R R Y L S
 D G G S A A S S P S G A C T S H R S R S S T V S S - G A T S S L P L
 G R W L S S E K P Q G R L Y L S S L P E F Y G V I L R R D V I S P

gatagagaggggaagggcgaagtccaaccgcgacagaggacggtgaaatcacaaaaggaagcagcgtgacttggcacgcaatcttcggttgtaaaact 34000
 I S L S P A L N L G R C L V T F D C L F A A H S P V C N E T T F V
 Y L P F A G L E F R S L P R H F - L P F C R S K A R L K R N H F S
 S L S P L R W T - V A V S S P S I V F S L L T V Q C A I K P Q S F E

ccgtagacccccaggggatgccgggaagtaaaaaataaatcgcgcgcgatttttcaaaaataataaagacgggtactacCTGCGTTCCTTGAAGGGCAGCTT 34100
 G Y V G P S A P F Y F Y F R A R N K L I I F V P V V Q T E Q L A A Q
 R L G W P I G P L L F L I A R S K E F Y Y L R T S G A N R S P C S
 UL29 T S G L P H R S T F I F D R A I K - F L L S P Y - R R K K E P L K

GAGCTCGGGGTTGACCTCGTGCAGGAGGTGTGGTCGATGGGCCGATGTTCTCGTAAAGTTCGTGATGAGCAGGGCCAGGCCACGCGCACGAAG 34200
 A R P Q G R A P L H P R H A R H E R L S L R R H A P G P G R A R L
 S S P T S R T C P P T T S P G S T R T V F T T S S C P W A W A C S P
 L E P N V E H L S T H D I F G I N E Y R F D D I L L A L G V R V E

II

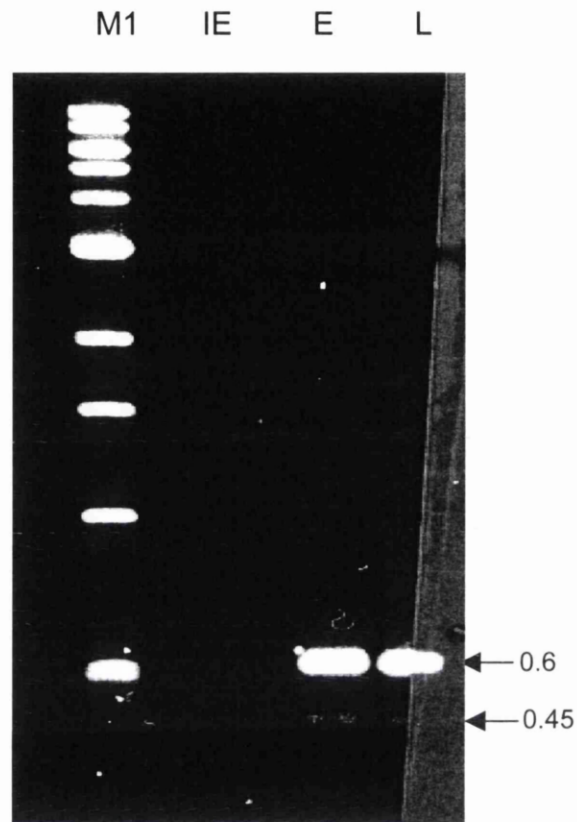


Figure 5.9: Gel electrophoresis of RT-PCR products potentially from a spliced UL28 transcript.

EtBr-stained 1% (w/v) agarose gel showing representative RT-PCR products (sizes in kbp) amplified from AD169 IE, E and L mRNAs using two primers UL29S1 and UL29S2 (Table 2.2). M1 is 1 kbp ladder.

these ORFs are conserved (Fig. 5.8), but that the first ATG codons in UL28 are not located equivalently. Moreover, amino acid sequence homology extends far upstream from both ATG codons, with homology falling off at a splice acceptor site that is conserved in both genomes (Fig. 5.8). Therefore, UL28 is probably spliced, but the location of the upstream exon is not clear. UL29 is a possibility, since there is a potential splice donor site at the 3'-end of CCMV UL29 that is weakly conserved in HCMV.

To investigate whether HCMV UL28 spliced to UL29, RT-PCR was carried out using primers UL29S1 and UL29S2. Two bands were generated in apparently identical sizes and intensities from E and L RNA (Fig. 5.9). Sequencing indicated that the larger product represents an unspliced transcript or contaminating DNA. The smaller product was cellular in origin. The experiment was repeated using two additional sets of primers (listed in Table 2.2), but evidence of splicing of UL29 to UL28 was not detected. Since RACE failed to determine the 5'-ends of UL28 and UL29 (see section 5.1.1), the inability to detect splice sites in the cognate mRNAs may reflect their low abundance. However, the possibility that the exon is spliced to UL28 ~~is~~ further upstream remains a strong possibility.

5.5 Discussion

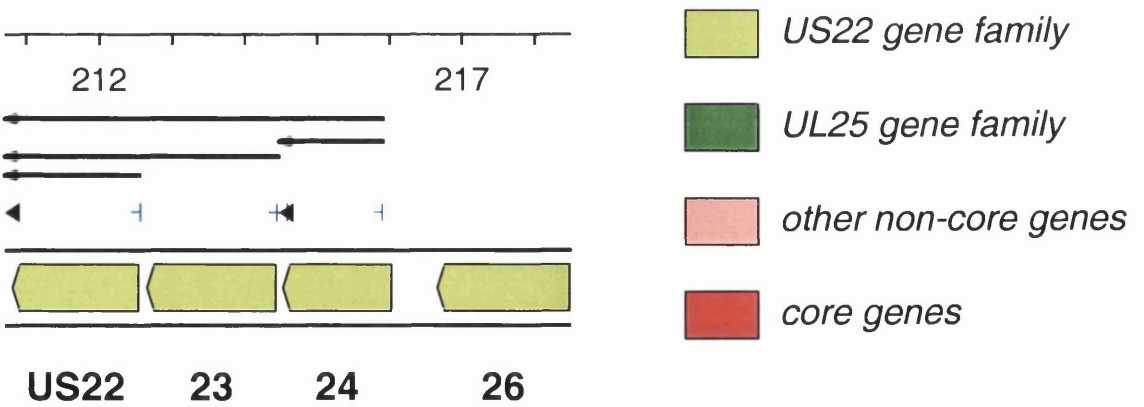
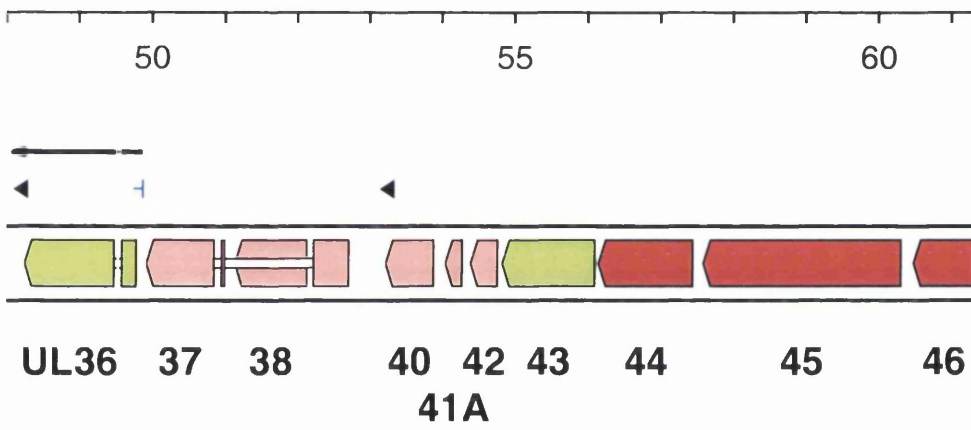
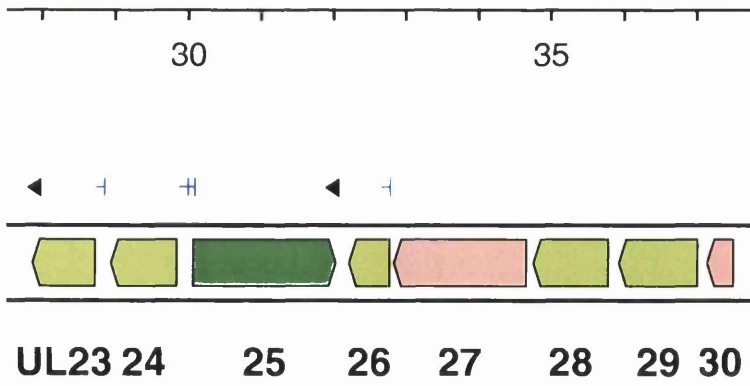
All sequenced β -herpesviruses possess members of the US22 gene family (Weston and Barrell, 1986; Chee *et al.*, 1990; Messerle *et al.*, 1991; Efstathiou *et al.*, 1992; Thomson and Honess, 1992; Gompels *et al.*, 1995; Nicholas, 1996; Rawlinson *et al.*, 1996; Vink *et al.*, 2000). The US22 gene family was initially characterized in HCMV as consisting of 12 distantly related genes of unknown function (UL23, UL24, UL28, UL29, UL36, UL43, IRS1, US22, US23, US24, US26 and TRS1). A thirteenth member (UL26) was recently been added (see Fig. 1.10) (A. Davison, unpublished data). In the present study, transcripts specified by 11 of these genes were investigated: UL23, UL24, UL26, UL28, UL29, UL36, UL43, US22, US23, US24 and US26. Only the UL36 transcript was previously mapped. The results are summarised in Fig. 5.10.

Transcripts of nine US22 genes were detected by northern blotting (Table 5.7); US26, TRS1 and IRS1 were not examined, and the kinetics of expression of UL43 could not be

Figure 5.10: Positions of the 5'- and 3'-ends of US22 gene family members.

Expanded version of parts of the AD169 genome (Fig. 1.9) showing the positions of 5' and 3'-ends of US22 family genes. US22 family genes are shown in three sections, with the genome coordinates in kbp and transcriptional features above the relevant ORFs.

- ⊥ 5'-end
- ◀ 3'-end
- ◀ RNA (and orientation)
- exon
- intron



determined. Chambers *et al.* (1999) failed to detect UL23 and UL24 mRNAs using gene microarray technology. All of these genes are expressed with either IE (UL36) or E kinetics (UL23, UL24, UL26, UL28, UL29, US22, US23 and US24). Most of the US22 family genes in other β -herpesviruses (MCMV, HHV-6 and HHV-7) are also transcribed with IE or E kinetics (Iskederian *et al.*, 1996; Cardin *et al.*, 1995; Nicholas and Martin, 1994; Stasiak and Mocarski, 1992).

Using RACE, 5'-ends were identified for the mRNAs of seven genes in the US22 family: UL23, UL24, UL26, UL36, US22, US23 and US24. In most cases RACE revealed 5'-ends mapping in a short region (less than 10 bp), rather than at a specific nucleotide residue (see Appendix 1). The 5'-ends obtained from the highest proportion of clones are listed in Table 5.2. 5'-ends for certain US22 genes (UL24, UL36 and US22) were confirmed by primer extension. Most of the 5'-ends mapped for US22 genes are located 25-35 bp downstream from potential TATA elements (Table 5.2) and upstream from the first ATG in the ORF, with one exception (US24). Two 5'-ends were determined for each of UL24 and US23, which were further confirmed by northern blotting. Failure to map 5'-ends for four members (UL28, UL29, UL43 and US26) may indicate that these genes are transcribed at low levels.

3'-ends were mapped for eight US22 genes, all of which are located 20-29 bp downstream from a polyA signal that is conserved in the same or a similar position in CCMV (Table 5.4). 3'-RACE revealed three sets of 3'-coterminal genes (UL40-UL47, UL23-UL24 and US22-US23). The latter two were confirmed by northern blotting. Northern blotting also confirmed the presence of an additional set of 3'-coterminal genes (UL26-UL30), and the existence of two 5'-coterminal transcripts from US24, the larger 3'-coterminal with US23 and US22.

Previously published splicing patterns for UL36 was confirmed by RT-PCR. There is also good evidence from sequence analysis that UL28 is spliced at its 5'-end (see Fig. 5.8). However, RACE failed to determine the 5'-ends of UL28 and UL29 (see section 5.1.1), and RT-PCR failed to detect splicing between the two ORFs. However, such splicing would be unusual, since it would result in a protein with two contiguous US22 domains. A similar splicing pattern was predicted in HHV-7 U7 (homologue of UL28) by Megaw *et al.* (1998),

who concluded that U7 might be spliced at its 5'-end, possibly to U8 (resulting in a protein with two contiguous US22 domains), or to an unidentified upstream exon.

Functions have been ascribed to three members of the US22 family (TRS1, IRS1 and UL36) (Stasiak and Mocarski, 1992; Zhang *et al.*, 1996; Colberg-Poley *et al.*, 2000; Goldmacher *et al.*, 2001), and the gene products of three members (US22, UL36 and UL26) have been identified in infected cells (Mocarski *et al.*, 1988; Baldick and Shenk, 1996; Patterson and Shenk, 1999). The expression profiles and transcript mapping data presented in this section will facilitate studies aimed at defining the functions of proteins specified by US22 family genes.

CHAPTER 6: RESULTS 4

Transcription of HCMV ORFs TRL1-TRL14

Transcription of HCMV ORFs TRL1-TRL14

The HCMV genome consists of an arrangement of unique regions and inverted repeats (Greenaway *et al.*, 1982; Spector *et al.*, 1982; Hayward *et al.*, 1984). The inverted repeats (TR_L and IR_L ; IR_S and TR_S) flank the unique sequences (U_L and U_S) in the L and S components (see section 1). TR_L as originally described is 11247 bp in size and contains 14 ORFs: TRL1-TRL14 (Chee *et al.*, 1990). Six ORFs are predicted to encode glycoproteins owing to the presence of a potential signal sequence for membrane insertion at the 5'-end, a transmembrane sequence near the 3'-end and possible N-linked glycosylation sites. Three of these (TRL11, TRL12 and TRL13) are the members of the RL11 family (Chee *et al.*, 1990). Most ORFs in TR_L (TRL2-TRL9) are not conserved in CCMV and therefore appear unlikely to encode proteins (Davison *et al.*, unpublished data). The literature indicates that the most abundantly transcribed E gene of HCMV (the $\beta 2.7$ RNA) is located in TR_L/IR_L and does not encode any protein (Greenway and Wilkinson, 1987; Staprans *et al.*, 1988). The purpose of this study was to evaluate transcription of the ORFs in AD169 TR_L as described by Chee *et al.* (1990), using RACE and primer extension techniques.

6.1 Sequences of 5'-ends in TR_L

5'-RACE was carried out using the SMART RACE kit (section 5.1). GSPs were designed from the region approximately 500 bp downstream from the putative initiation codons of target genes and are listed in Table 6.1. 5'-RACE products were cloned into pGEM-T and at least 5 clones were sequenced for each product. Comparisons of the results with data from Chambers *et al.* (1999) are shown in Table 6.2, and the detailed results are given in Appendix 1 at the back of this thesis.

6.1.1 5'-ends of ORFs that are conserved in CCMV

Five HCMV ORFs are conserved in CCMV (TRL1, TRL10-TRL14). Except for TRL1, a single primer was used to determine the 5'-ends of transcripts. Two primers (TRL1 and TRL1 (II)) were used to determine the 5'-end of TRL1.

	Primer	Sequence (5'- 3')	Position^a	PCR product(s)^b
TRL1	TRL1	AAGGTGACGACAGGTTGGCGCCCC	1427-1404	No product
	TRL1 (II)	TGCGCGAAAGCCTGTGCCGCGGCA	1500-1477	No product
TRL2	TRL2	GCACAAGCTCGGTGCGACGGATTA	2226-2203	150
TRL3	TRL3	TGTATATGCTCCTAGATAAGATTT	3440-3417	No product
	TRL3 (II)	GACTACTAGTACTCCAATCTTAGA	3464-3441	No product
TRL4	TRL4 (II)	GGCAGGAACACCCTTGCGGATTGA	3835-3858	300, 400, 700
	TRL4	CTCTTGTTGGGAATCGTCGACTTT	3943-3966	200, 300, 600
TRL5	TRL5	TCTTCCGCGTCTCCCGGCCGTACC	4324-4303	No product
	TRL5 (II)	ACGCAGGGGTTTAGCAGCTTCCCC	4597-4574	400
TRL6	TRL6	GCCACACATCACCCTGTATTCAT	5629-5652	200, 400, 500
TRL7	TRL7	GGTAACGATGCTACTTTTTTTAAT	6601-6624	No product
	TRL7 (II)	ACTTGCTGGTTCCCTAAAAGTTCGC	6659-6682	No product
TRL8	TRL8	GTCGAGGCCGATCGACTGCCGCAT	7624-7601	350, 450, 750
TRL9	TRL9	GGCCTGCATGATCATCGTCATCAT	7900-7877	100, 200, 350, 600
TRL10	TRL10	CGTGTCAGGCTCCGATAGGCCAGA	8555-8532	500, 600, 800
TRL11	TRL11	CTCGGCAGAAGCTACGTGGGTGGA	9155-9132	550, 600, 650, 750, 800, 1050, 1150
TRL12	TRL12	CATCTGACGTGATAGTTACGGTGA	9875-9852	650, 1350, 1800
TRL13	TRL13	GTGGTGGTGTGGTACATGTGGTA	11100-11077	400, 450, 700,
	TRL13 (II)	GGTTTTGTAGAAACGGTGGTTGCT	11025-11002	750, 1400
TRL14	TRL14	TTACAGGGCATGTGCCAAGAGTGG	11481-11458	400, 750, 800, 1050, 1100

Table 6.1: Primers used in 5'-RACE to map 5'-ends of transcripts from ORFs in TR_L.

^a Position of primers in the positive strand of the AD169 genome.

^b Approximate sizes (bp) of PCR products excised for cloning (see Fig. 6.1).

ORF	Conditions of expression		Position of 5'-end ^c	Putative TATA element (5'-3')	Position of TATA box ^d	Position of ATG codon ^e
	RACE technique ^a	Gene microarray ^b				
TRL1		E	No RACE product			
TRL2	L	E	2083			2094
TRL3		L	No RACE product			
TRL4		E	Cellular			
TRL5	L	Not detected	4252			4266
TRL6	E	L	5800 5990 6080	TATAAC TGATTG	6015-6010 (20) 6106-6101 (20)	5656 5947 5947 (1)
TRL7		E	No RACE product			
TRL8	E	L	7350 7259 6911	TAACTG TACTTT	7224-7229 (29) 6886-6891 (20)	7365 (1) 7284 (2) 7284 (4)
TRL9	E-L	L	7704 7350			7717 7365 (1)
TRL10	E	E-L	7704 8140 8071	TATAAT	8036-8041 (21)	7717 8182 8182 (1)
TRL11	L	L	8140 8071 8452 8509 8560 8605 8662	TATAAT TAATTT TCAATG TAACTC	8036-8041 (21) 8469-8474 (26) 8526-8531 (28) 8634-8639 (22)	8182 8182 (1) 8726 (1) 8726 (1) 8726 8726
TRL12	E	E-L	8140 8509 9342	TAATTT TGTTGG	8469-8474 (26) 9307-9312 (29)	8182 8726 (1) 9434
TRL13	E	E-L	10406 10443 10677 10738	TAGTAA TATTAA TCATTG TATGAA	10377-10382 (23) 10405-10410 (32) 10650-10655 (21) 10708-10713 (24)	10796 (8) 10796 (7) 10796 (2) 10796 (1)
TRL14	E	L	10406 10443 10677 10738	TAGTAA TATTAA TCATTG TATGAA	10377-10382 (23) 10405-10410 (32) 10650-10655 (21) 10708-10713 (24)	10796 (8) 10796 (7) 10796 (2) 10796 (1)

Table 6.2: 5'-ends of HCMV ORFs TRL1-TRL14.

^a Conditions of expression as determined by RACE. ^b Kinetics of expression as previously determined by Chambers *et al.* (1999). ^c Position of 5'-ends as represented in the majority of 5'-cDNA clones. 5'-ends located far upstream and downstream from the first ATG (as described by Chee *et al.*, 1990) are shown in pink and blue, respectively. ^d Position of the putative TATA element in the AD169 genome. The number of nucleotides between the 5'-end and closest TATA element is shown in brackets. ^e Position of the first ATG codon in AD169 (upper strand). Coordinates in brackets indicate the number of ATG codons present in the region between the first ATG codon and the 5'-end. ORFs that are not conserved in CCMV are shown in red.

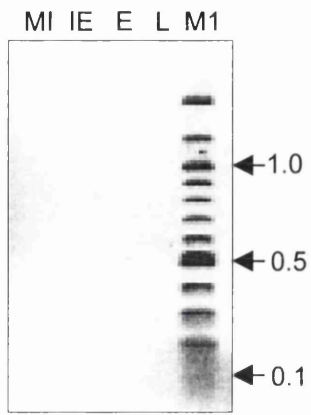
Neither yielded a product (Fig. 6.1a). This suggests that TRL1 is expressed at low levels in infected fibroblasts.

TRL10 codes for a putative structural glycoprotein (Spaderna *et al.*, 2002). Primer TRL10 (Table 6.1) yielded three major bands of approximately 500, 600 and 800 bp in a RACE experiment (Fig. 6.1j). They were produced in similar amounts from E and L RNA, indicating that TRL10 specifies an E transcript. Consistent with this, previous northern blot experiments detected TRL10 transcripts from E and L RNA (Spaderna *et al.*, 2002). Sequencing indicated that all three bands originated from HCMV. The two smaller products represent 5'-ends at 8071 and 8140 (Fig. 6.2) and the larger product represents a 5'-end at 7704, as detected using primer TRL9 (see section 6.1.2). Both 5'-ends of TRL10 are located upstream from the first ATG, but only the former is preceded by a potential TATA element.

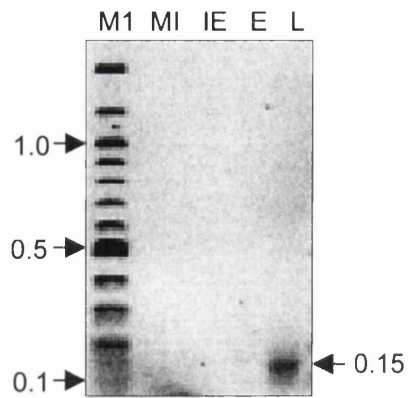
TRL11 is a member of the RL11 family and encodes a glycoprotein that possesses IgG Fc-binding capabilities (Lilley *et al.*, 2001). As can be seen in Fig. 6.1k, primer TRL11 yielded seven major bands from L RNA (Fig. 6.1k). Consistent with this, Chambers *et al.* (1999) reported from the gene microarray experiments that TRL11 specifies an L transcript. Five bands (550, 600, 650, 750 and 800 bp) represent 5'-ends of TRL11 at 8662, 8605, 8560, 8509 and 8452, respectively (indicated by black arrows in Fig. 6.2). Two other bands of approximately 1.05 and 1.15 kb were also generated by primer TRL11, and represent the 5'-ends of TRL10 (indicated with red arrows in Fig. 6.2) as determined using primer TRL10. This indicates that TRL11 is 3'-coterminal with TRL10.

Primer TRL12 yielded several bands from E and L RNA (Fig. 6.1l). The three major bands (650 bp, 1.35 and 1.8 kb) were analysed and mapped the 5'-end of TRL12 to 9342, and those of TRL11 and TRL10 at 8509 and 8140, respectively. This suggests these three genes are 3'-coterminal.

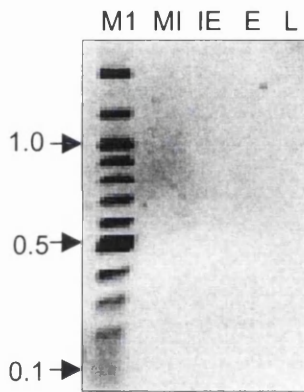
Primer TRL13 yielded several bands from E and L RNA (Fig. 6.1m), indicating that TRL13 is an E transcript. Five major bands (400, 450, 700, 750 and 1400 bp) were analysed. The largest band, originating from L RNA, was cellular in origin. The other four represent the 5'-ends at 10738, 10677, 10443 and 10406,



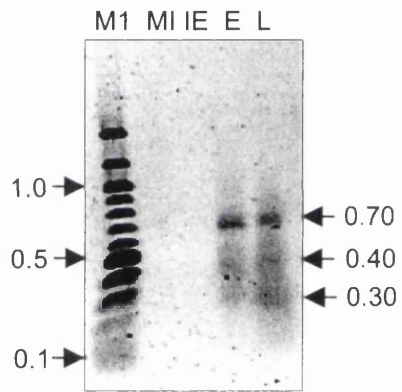
a: TRL1



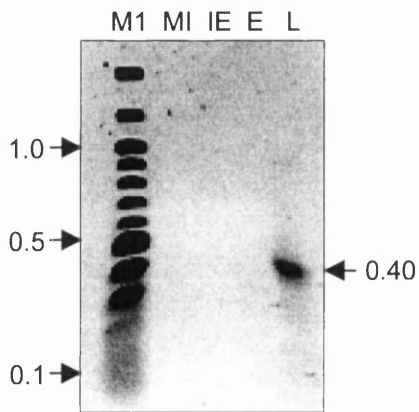
b: TRL2



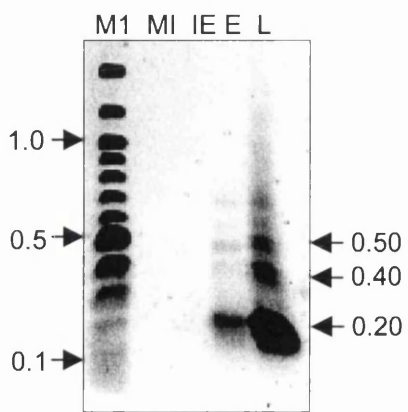
c: TRL3



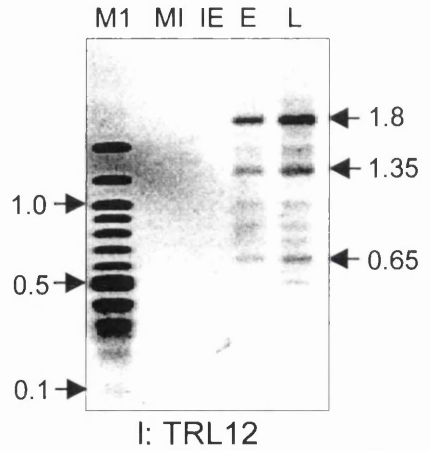
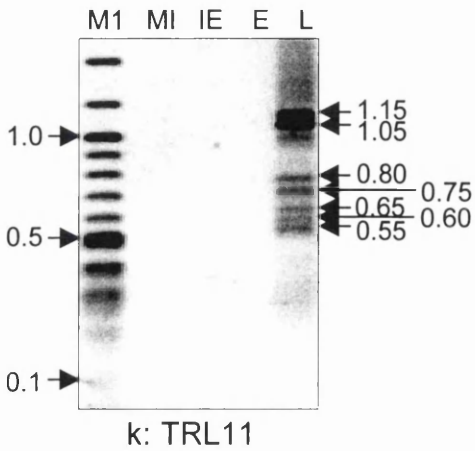
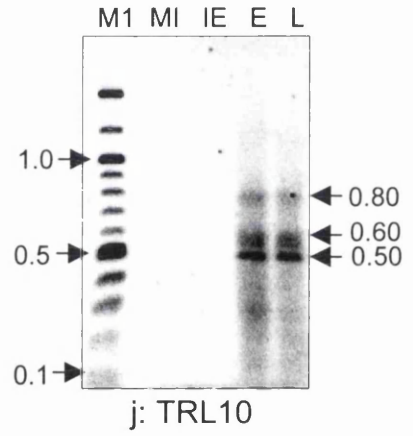
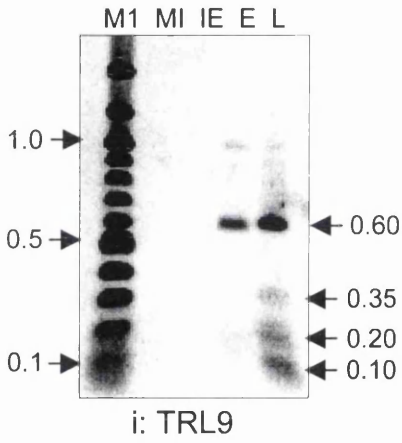
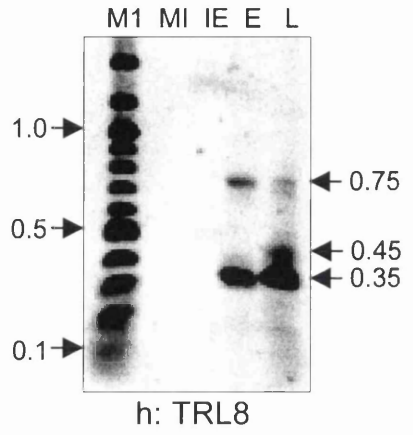
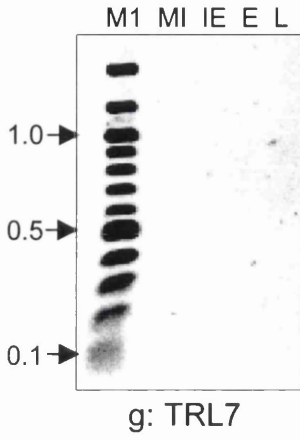
d: TRL4



e: TRL5



f: TRL6



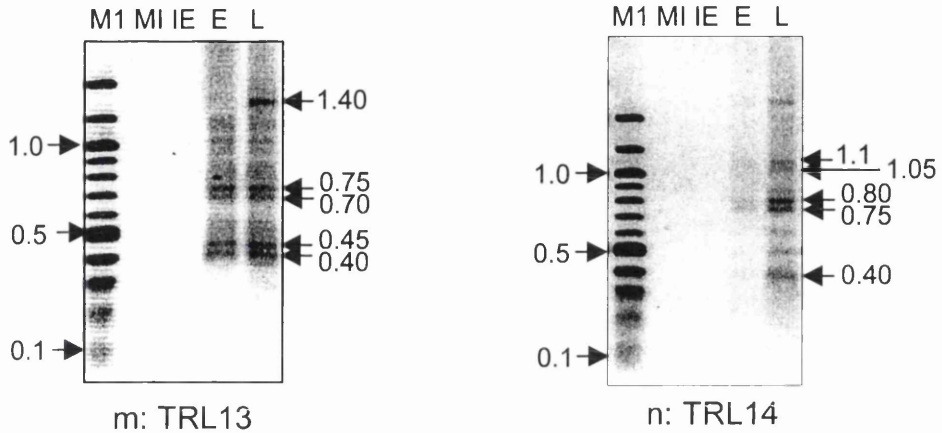


Figure 6.1: 5'-RACE of ORFs in the TR_L region.

EtBr-stained 1% (w/v) agarose gels showing representative 5'-RACE PCR products containing regions of AD169 ORFs TRL1-TRL14 that were amplified from 5'-cDNA with GSPs (listed in Table 6.1) and UPM. 5'-cDNA was prepared from AD169 IE, E or L mRNA or MI RNA. M1 is a 100 bp ladder. (a) TRL1 with primer TRL1. (b) TRL2 with primer TRL2. (c) TRL3 with primer TRL3 (II). (d) TRL4) with primer TRL4. (e) TRL5 with primer TRL5 (II). (f) TRL6 with primer TRL6. (g) TRL7 with primer TRL7 (II). (h) TRL8 with primer TRL8. (i) TRL9 with primer TRL9. (j) TRL10 with primer TRL10. (k) TRL11 with primer TRL11. (l) TRL12 with primer TRL12. (m) TRL13 with primer TRL13. (n) TRL14 with primer TRL14.

respectively. 5'-ends represented by the larger two bands (700 and 750 bp) are located upstream from the first ATG codon in TRL13 and the intervening regions contain an additional eight ATG codons in other reading frames. All the 5'-ends are preceded by a potential TATA element. The experiment was repeated using a different primer TRL13 (II) and yielded a similar result.

Primer TRL14 yielded major five bands (400, 750, 800, 1050 and 1100 bp) from E and L RNA (Fig. 6.1n). The smallest band, originating from L RNA, was cellular in origin. The latter four bands correspond to 5'-ends at 10738, 10677, 10443 and 10406, respectively, the same as those determined for TRL13 using primers TRL13 and TRL13 (II). This is consistent with the view that TRL13 and TRL14 represent the products of a frameshift mutation within an initially larger ORF in AD169 (see section 3.6).

6.1.2 5'-ends of ORFs that are not conserved in CCMV

Eight ORFs (TRL2-TRL9) in AD169 TR_L are not conserved in CCMV in location or sequence. Nevertheless, 5'-ends were mapped for five of these ORFs. Primer TRL2 yielded a single band of 150 bp from L RNA (Fig. 6.1b). Sequence analysis indicated that the 5'-end is located at 2083, but a potential TATA element is not present upstream.

Two primers were used to determine the 5'-end of TRL5. Primer TRL5 yielded no product but primer TRL5 (II) yielded a single band of 400 bp from L RNA (Fig. 6.1e), indicating that this is a L transcript. However, a previous northern blot analysis revealed an E transcript from TRL5 (Hutchinson *et al.*, 1986). Sequence analysis indicated that the 400 bp band represents a 5'-end located at 4252. There is no TATA element upstream.

Primer TRL6 yielded major three bands of 200, 400 and 500 bp (Fig. 6.1f). The abundance of these products was greater in L RNA than in E RNA, suggesting that they are expressed with E-L kinetics. Consistent with this, the temporal class of TRL6 expression was previously classified as E-L by northern blot analysis (Hutchinson *et al.*, 1986). All three of the bands mentioned above originated from HCMV and correspond to 5'-ends at 5800,

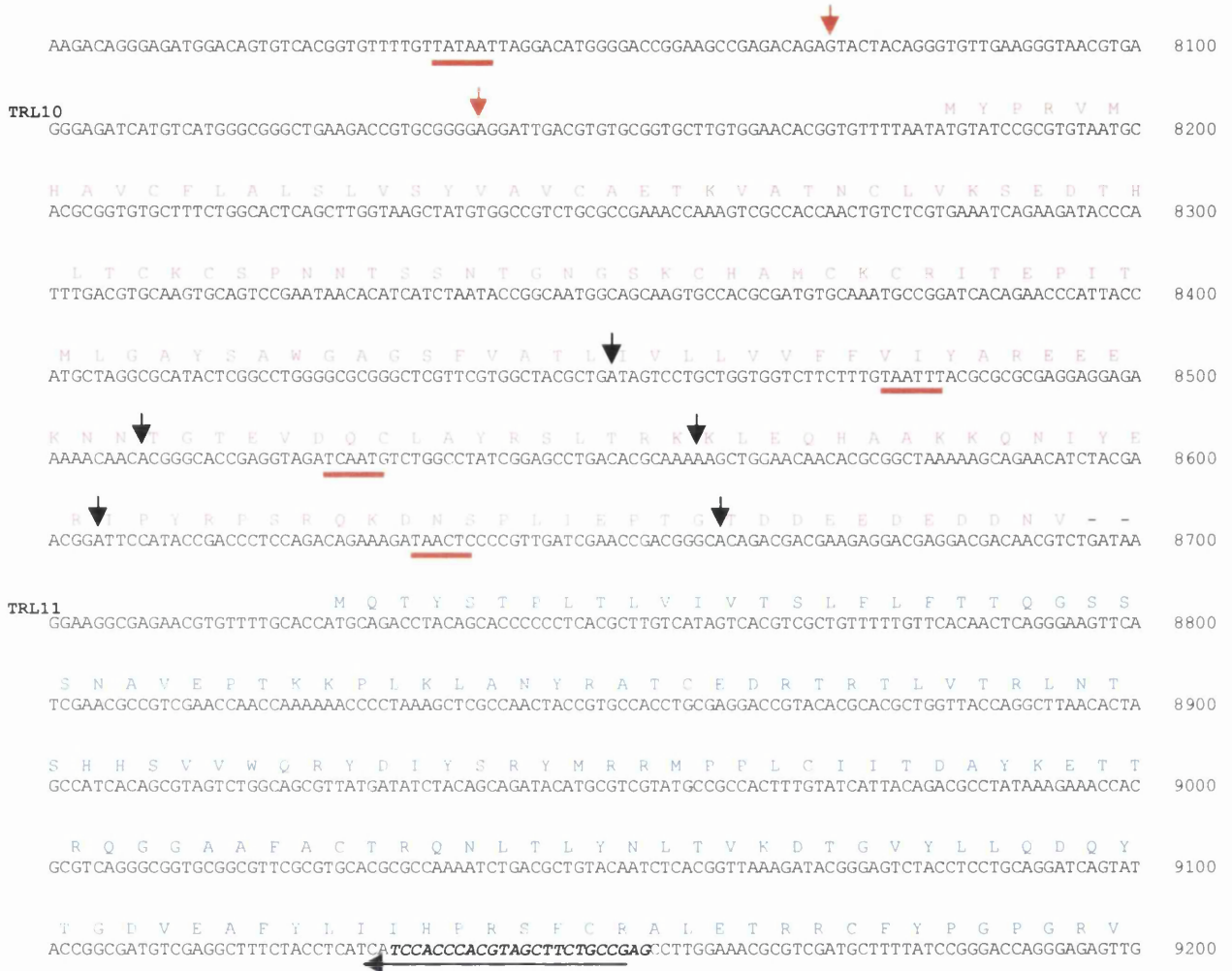


Figure 6.2: Position of 5'-ends as represented in 5'-cDNA clones of TRL11.

Fig. 6.2 shows the region containing the TRL10 and the 5'-ends of TRL11, with coordinates on the right. The predicted amino acid sequences of TRL10 and TRL11 are shown in red and blue, respectively. The position of RACE primer TRL11 is shown in bold italics and its orientation is indicated by an arrow. The red vertical arrows indicate the positions of 5'-ends of TRL10 mRNAs and the black vertical arrows indicate the 5'-ends of TRL11 mRNAs. Potential TATA elements are underlined in red.

5990 and 6080. The 5'-end at 5800 is approximately 150 bp downstream from the first ATG and is not preceded by a TATA box. The only ATG that could be used as initiation codon is located 40 bp upstream from the TRL6 stop codon. The 5'-ends located at 5990 and 6080 are 150 bp and 50 bp upstream, respectively, from the first ATG, and each is preceded by a potential TATA element.

Primer TRL8 yielded two bands of 350 and 750 bp from E and L RNA (Fig. 6.1h). The abundance of smaller product was similar in E and L RNA but that of the larger product was greater in E RNA. An additional minor band of 450 bp was also obtained from L RNA (Fig. 6.1h). Sequence analysis indicated that all three bands (350, 450 and 750 bp) originated from HCMV and represent 5'-ends at 7350, 7259 and 6911, respectively. The former is located downstream from the first ATG. The 5'-end at 7259 is located upstream from the first ATG and has a TATA element 29 bp upstream. The 5'-end at 6911 is located approximately 350 bp upstream from the first ATG and the intervening region contains four ATG codons in other reading frames.

A PCR product of approximately 600 bp was obtained from E and L RNA when primer TRL9 was used (Fig. 6.1i). Sequence analysis indicated that this represents the same 5'-end detected for TRL8 at 7350. This suggests that the TRL8 and TRL9 transcripts are 3'-coterminal. In addition, primer TRL9 yielded two bands (100 and 350 bp) from L RNA and a band of 200 bp from E and L RNA (Fig. 6.1 i). Only the 200 bp band originated from HCMV and indicated a 5'-end at 7704 that is 12 bp upstream from the first ATG codon in the ORF and is not preceded by a TATA element. The greater abundance of the 200 bp band from L RNA indicates that this transcript is in the E-L kinetic class.

Primer TRL4 was used to determine the 5'-end of TRL4. Three major bands were generated in similar proportions from E and L RNA (Fig. 6.1d). Sequencing indicated that none originated from HCMV. Use of a second primer (TRL4 (II)) failed to detect a 5'-end.

In order to determine the 5'-ends of TRL3 and TRL7, two different primers were used for each (Table 6.1). None yielded a product (Fig. 6.1c and 6.1g).

6.2 Mapping the 5'-ends of ORFs in TR_L by primer extension

To confirm the 5'-ends determined by RACE, four ORFs from TR_L were analysed by primer extension: TRL2, TRL5, TRL8 and TRL9. GSPs (listed in Table 2.3) were designed from the region approximately 100 bp downstream from the 5'-end of the RNAs determined by RACE. In two experiments, none yielded a detectable primer extension product. It is therefore possible that these mRNAs are present in low abundance in infected fibroblasts.

6.3 Discussion

Transcripts of the 14 ORFs from AD169 TR_L (TRL1-TRL14) were analysed by 5'-RACE. The data are summarised in Fig. 6.3.

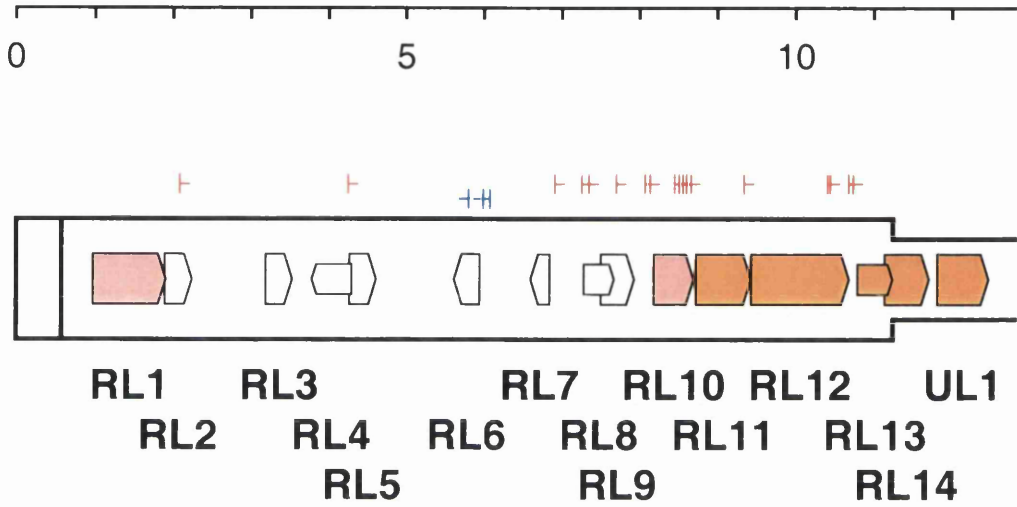
RNAs were identified for five (TRL10, TRL11, TRL12, TRL13 and TRL14) out of six ORFs that are conserved in CCMV. The inability to detect the TRL1 transcript may indicate its poor expression in infected fibroblasts. Expression of all the transcripts detected except that specified by TRL11, fall into the E kinetic class. TRL11 was considered an L transcript. Chambers *et al.* (1999) reported similar findings using gene microarray technology. Transcript mapping data revealed the same 5'-end for TRL13 and TRL14, consistent with TRL13 and TRL14 representing the products of a frameshift mutation within a larger ORF in AD169 as discussed in section 3.6. Transcripts for certain genes (TRL10, TRL11 and TRL13/14) possess multiple 5'-ends. The majority of these 5'-ends are preceded by a potential TATA element 25-35 bp upstream.

Five ORFs (TRL2, TRL5, TRL6, TRL8 and TRL9) were analysed that are not conserved in CCMV and are unlikely to encode proteins. However, Chambers *et al.* (1999) detected transcripts for each ORF using gene microarray technology. Transcript mapping data revealed single 5'-ends for TRL2, TRL5 and TRL9, each located upstream from the first ATG codon in the ORF, but none preceded by a potential TATA element. Some ORFs possess multiple 5'-ends that are located downstream or far upstream (TRL6, TRL8) from

Figure 6.3: Positions of the 5'-ends of HCMV ORFs TRL1-TRL14.

Expanded version of parts of the AD169 genome (Fig. 1.9) showing the positions of 5'-ends of TRL1-TRL14, with the genome coordinates in kbp.

- └ 5'-end of ORFs that are oriented right to left
- ┌ 5'-end of ORFs that are oriented left to right



 *ORF probably not encoding protein*

 *RL11 gene family*

 *other non-core gene*

the first ATG codon in the ORF. The latter arrangement results in several ATG codons in other reading frames preceding the first ATG codon in the ORF.

RACE failed to detect transcripts for three ORFs (TRL3, TRL4 and TRL7) that are not conserved in CCMV. Greenway and Wilkinson (1987) mapped 5'-end of the major E transcript (TRL4 or β 2.7) at 4435 using S₁ nuclease analysis (as described by Akrigg *et al.* (1985) and Wilkinson *et al.* (1984)) and primer extension (as described by Farrell *et al.* (1983)). Rawlinson and Barrell (1993) also detected the TRL4 transcript from E and L RNA using northern blotting with seven probes located along TRL4. Two different primers were used in the present study to determine the 5'-end of TRL4. Each yielded products, but these were cellular in origin. It is not known why the attempts to map 5'-end of the TRL4 transcript was unsuccessful.

The 5'-RACE revealed the existence of a set of 3'-coterminal transcripts in TR_L: TRL8-TRL12. However, it will be necessary to map the 3'-end of these transcripts to confirm this.

Transcript mapping data shown in this study indicates that AD169 TR_L is transcriptionally active, but in a complex manner. TRL10, TRL11, TRL12 and TRL13 appear likely to encode proteins since these are conserved in CCMV. TRL14 is naturally frameshifted in AD169 and encodes a truncated protein. 5'-ends were mapped for the majority of ORFs in the TRL2-TRL9 region, but their expression as functional proteins appear unlikely, since these ORFs are not conserved in CCMV. One transcript in this region (the β 2.7 RNA, mapping to TRL4) does not encode protein and is thought to be a non-translated (Greenway and Wilkinson, 1987). Thus it is possible that other RNAs in the TRL2-TRL9 region may serve similar functions. One possibility, in light of the function of TRL2-TRL9 mRNAs, is that they serve a structural role in assembling and organising proteins in the tegument.

CHAPTER 7: RESULTS 5

Transcription of HCMV ORFs UL1-UL30

Transcription of HCMV ORFs UL1-UL30

The purpose of this study was to map the 5'-ends of transcripts associated with UL1-UL30. Primers used in RACE experiments are listed in Table 7.1.

7.1 Mapping 5'- ends of genes in the U_L region

A detailed summary of the results is given in Appendix 1, and a comparison of the results with data from Chambers *et al.* (1999) is shown in Table 7.2. The data are summarised diagrammatically in Fig. 7.1.

5'-RACE products were not produced from three ORFs (UL1, UL5 and UL9 in Fig. 7.2a, 7.2e and 7.2i, respectively). This experiment was repeated using several GSPs (listed in Table 7.1) for each gene. The inability to map the 5'-ends of these genes suggests that they are transcribed at low levels in infected fibroblasts.

Primers UL2 and UL3 yielded similar patterns of bands from L RNA. (Fig. 7.2b and Fig. 7.2c, respectively). Sequence analysis indicated that the major bands (650 bp in Fig. 7.2b and 500 bp in Fig. 7.2c) correspond to a 5'-end at 13471. The minor bands (500 bp in Fig. 7.2b and 350 bp in Fig. 7.2c) correspond to a 5'-end at 13296. The 5'-end at 13471 is located 20 nucleotides downstream from a potential TATA element and 170 bp upstream from the UL3 ATG codon. The 5'-end at 13296 is located downstream from the first ATG. The 350 bp band (Fig. 7.2b) indicated that the 5'-end of UL2 is at 13186. This 5'-end is 28 bp downstream from a potential TATA box. The abundance of 5'-RACE products in Fig. 7.2b and Fig. 7.2c indicates that the UL2 transcript is rare in comparison with that of UL3, and that these are 3'-coterminal genes. A candidate polyadenylation signal is located near the 3'-end of UL2 (12618-12613).

Three 5'-RACE bands of 550, 650 and 800 bp (Fig. 7.2d) from UL4 were analysed, and indicated a major 5'-ends at 13290 and minor 5'-ends at 13182 and 13046. Each is located upstream (250 to 400 bp) from the first ATG codon in UL4, and several other ATG codons are located between the 5'-ends and first ATG codon. Only the 5'-end at 13046 has TATA

ORF	Primer	Sequence (5'-3')	Position ^a	PCR product(s) ^b
UL1	UL1 UL1 (II)	TGGTGCTATTTATGGTGTAGTTAC G TTCAGTATAAGCACCTGAATCGT	12299-12276 12200-12177	No product
UL2	UL2	AGAGCGTCTCGAAGCAGCTTGAGC	12872-12895	350, 500, 650
UL3	UL3	CGCGACCGAGTCTTCGGCCATGGT	13027-13050	350, 500,
UL4	UL4	AGCAACGTCAAGCGGTTATAGGTG	13825-13802	550, 650, 800
UL5	UL5 UL5 (II)	AGTGATGATGAGACGACTCCACGG ACGAGACGTTGATAGCGGTATTTTC	14324-14301 14500-14477	No product No product
UL6	UL6	GCTACACGGTGTCTTAATGAGACT	15015-14992	450, 750
UL7	UL7	CACCGTTGGTGTCTGTCGTGACGAT	16000-15977	1000, 1700
UL8	UL8	TTTTGAGGTGCTCGCAGACAGATG	16535-16512	400, 500, 1.5
UL9	UL9 UL9 (II)	ACAATGTATGATTCAAAAAACATC TGACATTGTAACATAGAGGTGTCC	16900-16877 16999-16976	No product No product
UL10	UL10	GCCGTTGCTAATGCCGACGCATGA	17609-17632	750
UL11	UL11	GGTGGTAGTGGAAGTGTTCGGTT	18738-18715	500, 1350, 1800
UL12	UL12	CCTCAAACCGATCTGTGAGCAAGT	19105-19128	300, 600
UL13	UL13	ACTGCAGCTCTTGGTACTGCCAAC	19728-19705	450, 600
UL14	UL14	AGGTGTCGCACGTTCGGCCACCAGC	21377-21354	450, 750
UL15	UL15	ATCTTACCGGGGACCTCGTTGTCC	22280-22303	No product
UL16	UL16	AGCGAGCCCAATCGGACGTAGAGG	22895-22872	600, 700, 800
UL17	UL17	TCTTCGGGAGACCGTGCCGGATCC	23518-23495	400, 650, 1200
UL18	PEL18 (II) PEL18 UL18	ATCCCCGTGTACCCGTAACGCAAC ATAGGTA AAAAAGTGTTCGCCGTC TAGTGTCAGGTTATCCTTGAGCTC	23725-23702 23795-23772 24095-24072	150 100, 200 450
UL19	UL19	CGCGACTCACATCTAGCTCGTCTT	25002-25025	800, 1350
UL20	UL20	GTATTGTTGTTCAGCGCAGTTCAGA	25624-25601	400, 500, 600
UL21	UL21 UL21R	GTCGGTTCGATGACGATGTCGATCT AAAGCCAAGCGGCGCAAGGGCCGG	26602-26625 26736-26759	300 200
UL22A	PEL22A	TTCTGAGAAGGTGCCGCCAAAGCC	27181-27158	150, 400
UL25	UL25	CGTCGCCACCGCCCGGCGCCGT	30474-30451	450, 400
UL27	UL27	TCAGACCGCGGATGTCGTAATTGC	34276-34299	550, 1100
UL30	UL30	GTCAACACGCGACTCGCGCGCAA	37319-37342	250

Table 7.1: Primers used in 5'-RACE to map 5'-ends of transcripts from UL1-UL30.

^aPosition of primers in the positive strand of the AD169 genome.

^bApproximate sizes (bp) of PCR products excised for cloning (See Fig. 7.2).

ORF	Conditions of expression		Position of 5'-end ^e	Putative TATA element (5'-3')	Position of TATA box ^d	Position of ATG codon ^e
	RACE technique ^a	Gene microarray ^b				
UL1	No RACE product	E-L	Not detected			
UL2	L	L	13471	TAATAA	13496-13491 (20)	13324
			13296	TATTTG	13321-13316 (20)	None
			13186	TATACA	13220-13215 (28)	13048
UL3	L	L	13471	TAATAA	13496-13491 (20)	13324
			13296	TATTTG	13321-13316 (20)	None
UL4	E	E	13290			13464 (1)
			13182			13464 (7)
			13046	TATAAG	13018-13023 (22)	13464 (10)
UL5	No RACE product	E	Not detected			
UL6	L	Not detected	14271			14612 (1)
			14585	TATTTG	14558-14563 (21)	14612
UL7	L	L	14271			14612 (1)
			15068	TATTTA	15043-15048 (20)	15526 (6)
UL8	L	Not detected	15068	TATTTA	15043-15048 (20)	15526 (6)
			16027			16234 (4)
			16180	TATTAT	16149-16154 (26)	16234 (1)
UL9	No RACE product	L	Not detected			
UL10	E-L	Not detected	16933	TATTAA	16905-16910 (22)	17426 (9)
UL11	E	Not detected	16933	TATTAA	16905-16910 (22)	17426 (9)
			18270	TGAATA	18237-18242 (27)	18295 (2)
UL12	L	Not detected	19699			19321 (8)
			19353			19321 (1)
UL13	E	E	19308	TCATAA	19281-19286 (20)	19320
UL14	L	L	20713	TATGCT	20686-20691 (22)	20843 (1)
			20952			None
UL15	No RACE product	E	Not detected			
UL16	E	Not detected	22325			22378 (1)

UL17	E	E	22325 22896 23171	TAATCG TATAAG	22862-22868 (27) 23145-23150 (20)	22378 23214 (1) 23214
UL18	L	L	23724			23645 (1)
UL19	L	Not detected	23724 ^h 24269	TTATAG	24236-24241 (28)	23645 (1) 24740 (5)
UL20	L	Not detected	25253 25161 24979	TCAATA TATGAT TTAATG	25214-25219 (33) 25136-25141 (20) 24950-24955 (23)	25299 (1) 25299 (3) 25299 (6)
UL21A ¹	L	Not detected	26858	TATTTA	26894-26889 (25)	26839 (1)
UL25	L	Not detected	30016 30065	TATAAC	29982-29987 (28)	30057
UL27	L	E	35295			34657 (9)
UL30	L	Not detected	37525			37500

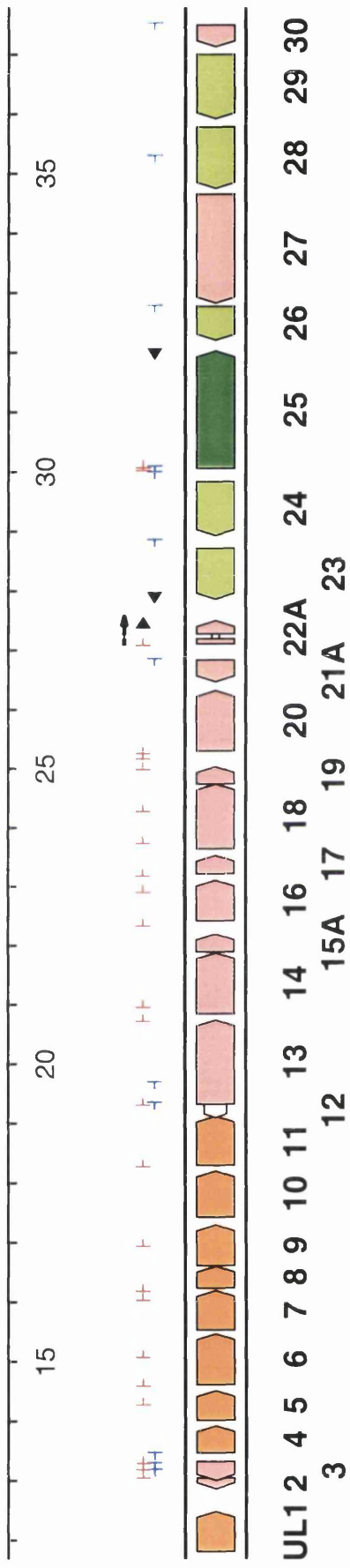
Table 7.2: 5'-ends of HCMV ORFs UL1-UL30.

^a Conditions of expression as determined by RACE. ^b Kinetics of expression as determined by Chambers *et al.* (1999). ^c Position of 5'-ends as represented in the majority of 5'-cDNA clones. 5'-ends located far upstream or downstream from the first ATG in the ORF (as described by Chee *et al.*, 1990) are shown in pink and blue, respectively. ^d Position of the putative TATA element in the AD169 genome. The number of nucleotides between the 5'-end and the closest TATA element is shown in brackets. ^e Position of the first ATG codon in the ORF. Coordinates in brackets indicate the number of ATG codons present in the region between the first ATG codon and the 5'-end. ORFs that are not conserved in CCMV are shown in red.

Figure 7.1: Positions of the 5'- and 3'-ends of HCMV ORFs UL1-UL30.

Expanded version of parts of the AD169 genome (Fig. 1.9) showing the positions of 5' and 3'-ends of UL1-UL30, with the genome coordinates in kbp and transcriptional features above the relevant ORFs. RACE results of US22 family genes and UL22A are shown in chapter 5 and chapter 4, respectively.

- ⊥ 5'-ends of ORFs that are oriented right to left
- ⊥ 5'-ends of ORFs that are oriented left to right
- ◀ 3'-end
- ◀ RNA (and orientation)
- exon
- intron



probably not encoding protein

RL11 gene family

US22 gene family

UL25 gene family

other non-core genes

element upstream. Since UL4 is 3'-coterminal with UL5 and shares a polyA signal at 14747 to 14752 with the polyadenylation signal tail added at 14766 (Chang *et al.*, 1989), the major UL4 transcript should be 1.7 kb in size. Consistent with this, a northern blot probed with a labeled oligonucleotide from the 5'-end of UL4 hybridised to an mRNA of this size (Rawlinson and Barrell, 1993; Alderete *et al.*, 1999).

Primer UL6 yielded several bands at late times (Fig. 7.2f). The major two bands of 450 and 750 bp indicated 5'-ends for UL6 at 14585 and 14271. A potential TATA element is present 21 bp upstream from the former. The latter 5'-end is located within UL5, 350 bp upstream from the first ATG codon in UL6.

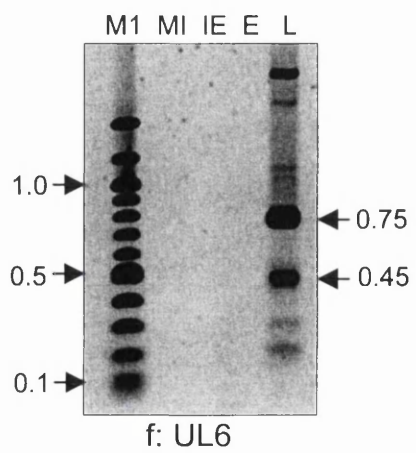
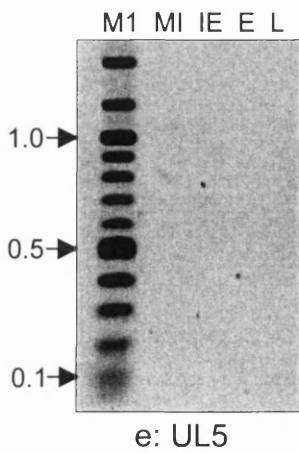
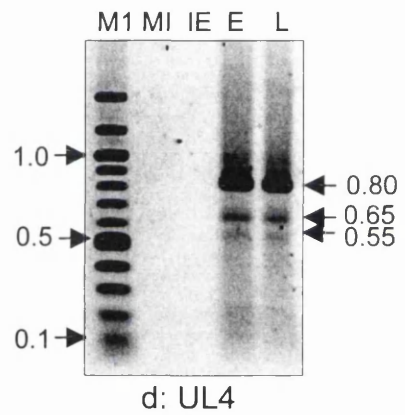
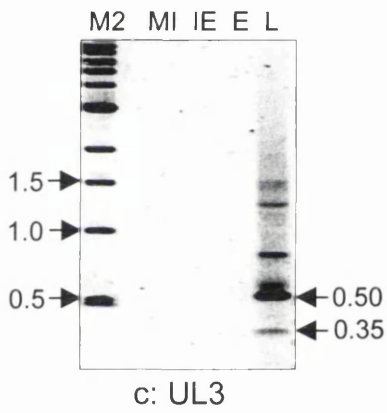
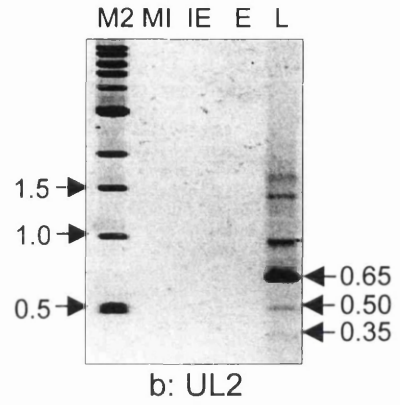
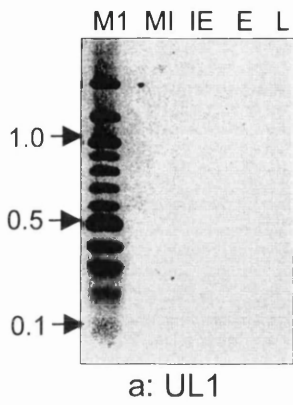
As shown in Fig. 7.2g, primer UL7 yielded several bands from L RNA. Two major bands (1.0 and 1.7 kbp) were analysed, indicating 5'-ends at 15068 and 14271, respectively. The former has a potential TATA element within 20 bp and represents the 5'-end of UL7, and the latter represents the 5'-end of UL6 as determined using primer UL6 (Fig. 7.2f). The 5'-end of UL7 is located far upstream from the first ATG codon in UL7, and the intervening region contains six other ATG codons in different frames. Detection of the 5'-end of UL6 by primer UL7 indicates that UL6 and UL7 are 3'-coterminal.

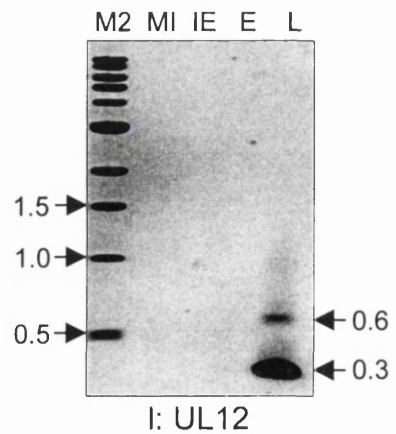
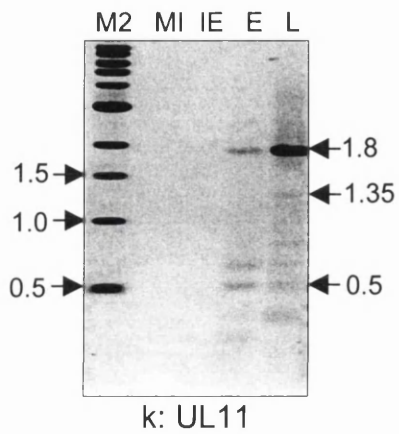
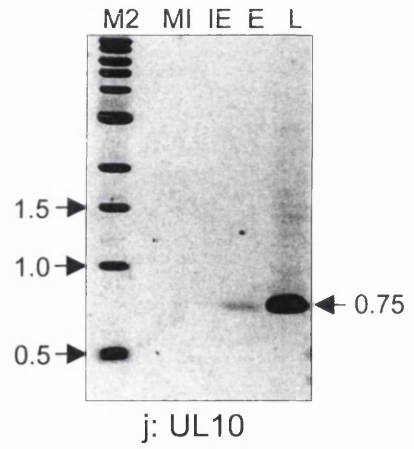
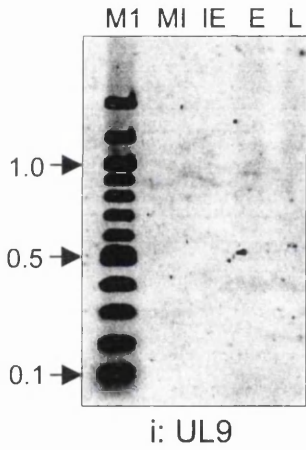
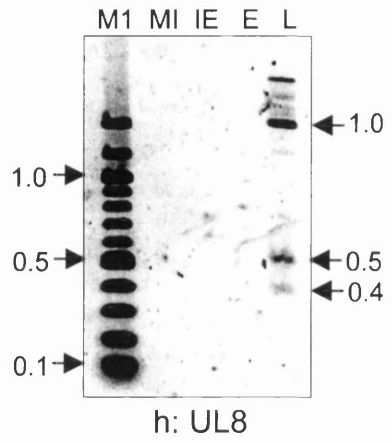
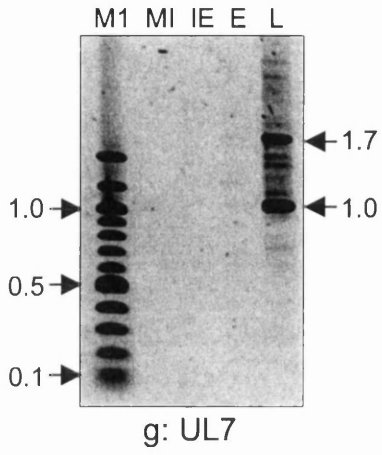
Primer UL8 detected transcripts from L RNA (Fig. 7.2h). The larger products originated from UL6 and UL7, and the lower two bands of 400 and 500 bp correspond to 5'-ends of UL8 at 16180 and 16027. A potential TATA element is present within 26 nucleotides of the former. The data indicate that UL6, UL7 and UL8 are 3'-coterminal, sharing a potential polyadenylation signal at 16558-16563. UL8 contains a splice acceptor sequence (TGATGTGCTTTTATCAGG) at 16212 to 16195, but splicing was not detected by northern blotting and RT-PCR (Chee *et al.*, 1990; Rawlinson and Barrell, 1993). The RACE results also produced no evidence for splicing.

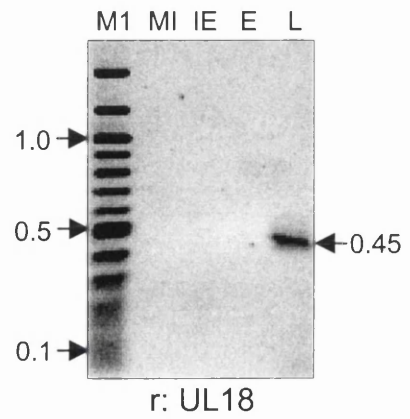
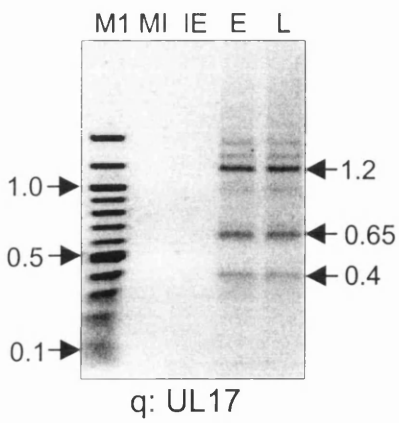
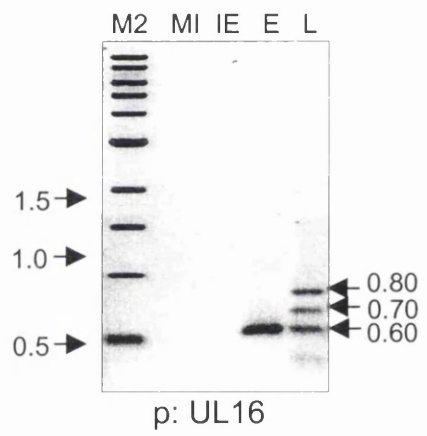
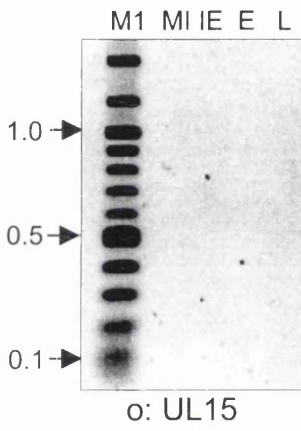
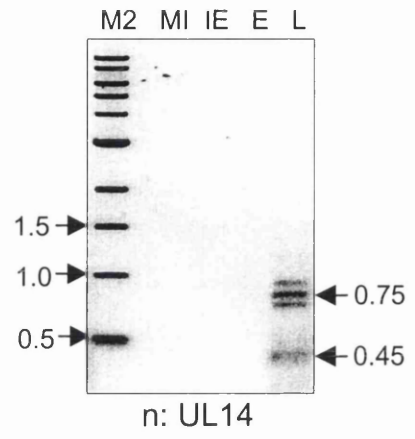
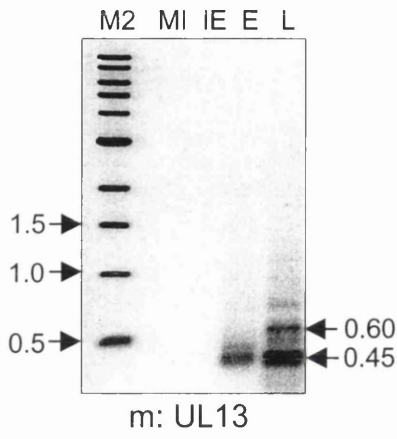
Primer UL10 yielded one major band of approximately 750 bp from E and L RNA (Fig. 7.2j). The intensity of the product was greater in L RNA, indicating that the UL10 transcript may fall into the E-L kinetic class. Sequence analysis indicated the 5'-end of

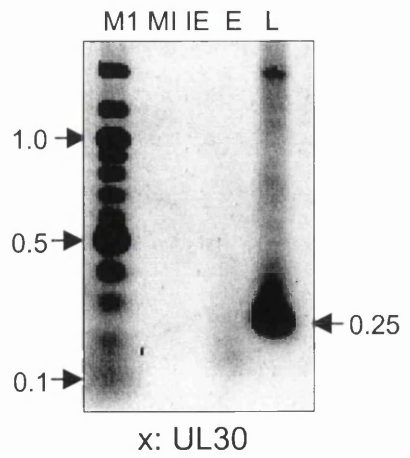
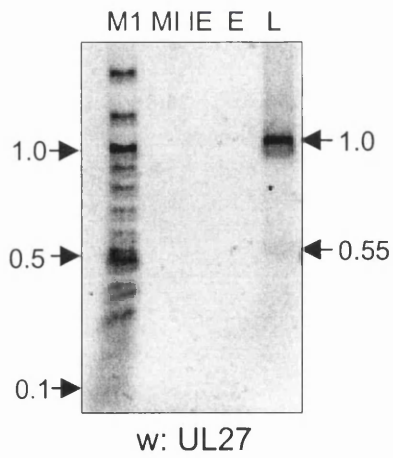
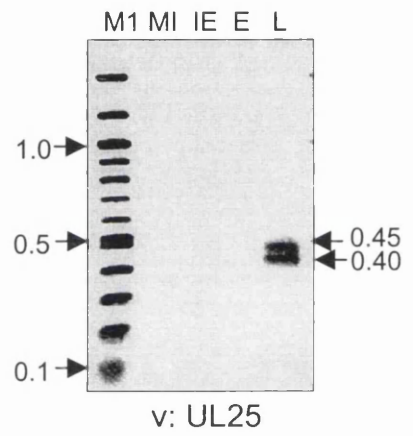
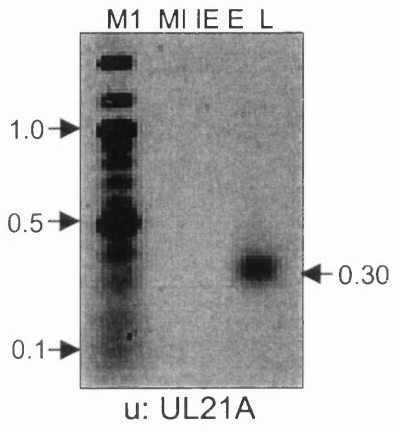
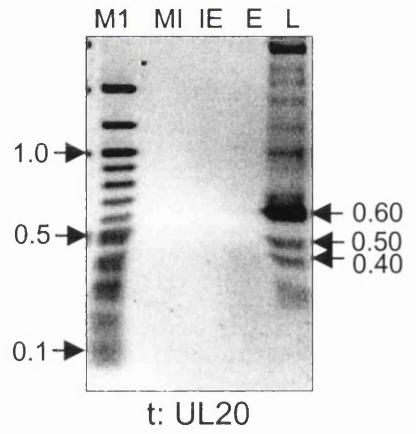
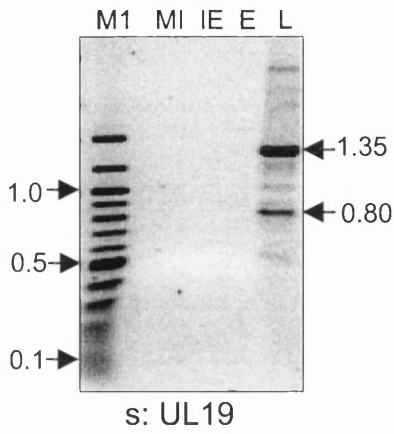
Figure 7.2: 5'-RACE of HCMV genes UL1-UL30.

EtBr-stained 1-1.5% (w/v) agarose gels showing representative 5'-RACE products from UL1-UL30 (except UL22A, UL23, UL24, UL26, UL28 and UL29) produced using GSPs (listed in Table 7.1) and UPM. 5'-cDNA was prepared from IE, E or L AD169 mRNA and from MI RNA. M1 and M2 are 100 bp and 1 kbp ladders, respectively. (a) Primer UL1. (b) Primer UL2. (c) Primer UL3. (d) Primer UL4. (e) Primer UL5. (f) Primer UL6. (g) Primer UL7. (h) Primer UL8. (i) Primer UL9. (j) Primer UL10. (k) Primer UL11. (l) Primer UL12. (m) Primer UL13. (n) Primer UL14. (o) Primer UL15. (p) Primer UL16. (q) Primer UL17. (r) Primer UL18. (s) Primer UL19. (t) Primer UL20. (u) Primer UL21. (v) Primer UL25. (w) Primer UL27. (x) Primer UL30.









UL10 is located at 16933, far upstream from the first ATG codon in this ORF. The region between the 5'-end and the ATG codon contains nine other ATG codons in different frames. A potential TATA element is present within 22 bp from this 5'-end.

Multiple RACE products were detected using primer UL11 (Fig. 7.2k). The largest band corresponded to the 5'-end of UL10 as determined using primer UL10 (Fig. 7.2j). The smaller product of 500 bp represented the 5'-end of UL11 at 18270 and is located 27 bp downstream from a potential TATA box. An additional band of 1.35 kbp from L RNA was found to be cellular in origin.

Primer UL12 yielded two bands of approximately 300 bp and 600 bp from L RNA (Fig. 7.2l) that correspond to 5'-ends at 19353 and 19699. The latter is located far upstream, with the intervening region containing eight other ATG codons in different frames. Both 5'-ends lack TATA elements 25-35 bp upstream. UL12 is not conserved in CCMV and is at present considered not to encode a protein.

Primer UL13 produced two major bands of 450 and 600 bp (Fig. 7.2m). The former was detected early and late times and mapped the 5'-end of UL13 to 19308. The other band was cellular in origin. The 5'-end of UL13 has a potential TATA box at 19281 to 19286.

Primer UL14 yielded four principal bands from L RNA (Fig. 7.2n). The 750 bp band indicated a 5'-end at 20713. A potential TATA element is present 22 bp upstream. The 450 bp band corresponds to a 5'-end at 20952, 110 bp downstream from the first ATG codon in UL14. The two other bands generated by primer UL14 were not analysed.

Two primers were used to determine the 5'-end of UL15. Neither yielded products (one example is shown in Fig. 7.2o). Subsequent to this experiment, two errors were identified in this region that resulted in replacement of UL15 by a new gene, UL15A, on the opposite strand (see section 3.1.1). Nevertheless, Chambers *et al.* (1999) detected a UL15 transcript using gene microarray technology and northern blotting (using a cDNA probe designed from the 3'-end of UL15). The 5'-end of UL15A was not mapped.

The UL16 protein exhibits eight potential N-linked glycosylation sites and is predicted to be a membrane glycoprotein (Chee *et al.*, 1990; Kaye *et al.*, 1992a). It is not detectable in purified enveloped virions and is non-essential for virus growth in cell culture (Kaye *et al.*, 1992a). Primer UL16 yielded a major band of approximately 600 bp from E and L RNA (Fig. 7.2p). Published northern blot data previously showed that UL16 is an E transcript (Kaye *et al.*, 1992a). The 600 bp band corresponds to a 5'-end for UL16 at 22325. Translational initiation would occur at 22378-22380 rather than at 22414-22416 concluded by Chee *et al.* (1990). Two other bands (700 and 800 bp) that were generated from L RNA were cellular in origin.

Primer UL17 yielded three principal bands of 1.2 kbp, 650 bp and 400 bp (Fig. 7.2q). The 1.2 kb band corresponds to the 5'-end of UL16 as mapped using primer UL16 (Fig. 7.2p). This suggests that UL16 and UL17 are 3'-coterminal. Analysis of the 650 and 400 bp bands indicated two 5'-ends for UL17 at 22896 and 23171, respectively. The former 5'-end is located far upstream from the first ATG in UL17, with the intervening sequence containing another ATG codon in a different frame. The 5'-end at 23171 is 20 bp downstream from a potential TATA box.

ORF UL18 exhibits significant sequence homology with the class 1 α -chain genes of HLA. Primer UL18 yielded a major band of 450 bp from L RNA (Fig. 7.2r) that mapped the 5'-end of UL18 to 23724. This places the initiation codon for UL18 at 23745-23747, the sixth ATG codon in the ORF (Fig. 7.4). This experiment was repeated several times using four different primers (listed in Table 7.1) designed from the region 150 bp to 1.4 kbp downstream from the first ATG codon in UL18. As can be seen in the Fig. 7.3, all primers yielded a major band that mapped the 5'-end at same position (23724). Primer PEL18 (Fig. 7.3 lane 1) generated an additional band of 200 bp in apparently similar intensity to the smaller product but was cellular in origin. Additional low abundance bands were also generated (Fig. 7.3) but were not analysed. Therefore, it is formally possible that a major downstream 5'-end for UL18 mRNA has been mapped but a minor upstream 5'-end has been missed. The likely initiation codon for the mapped UL18 is at 23745 to 23747, 21 bp downstream from the mRNA start site. There is no potential TATA box immediately upstream of the initiation site.

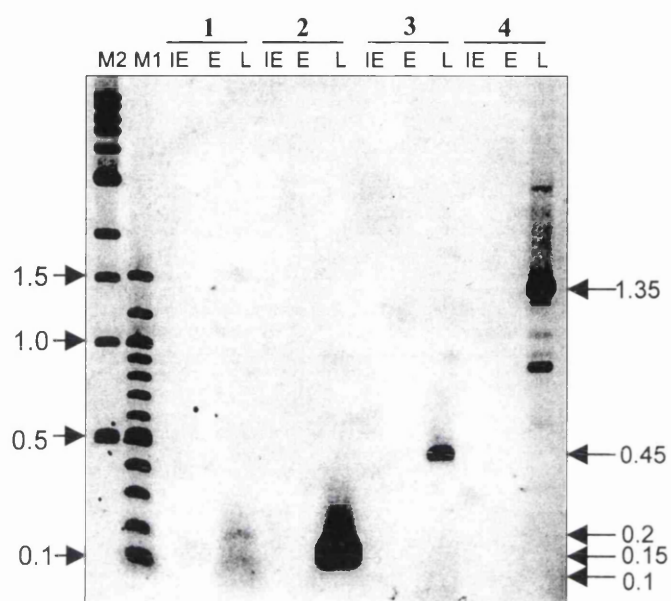
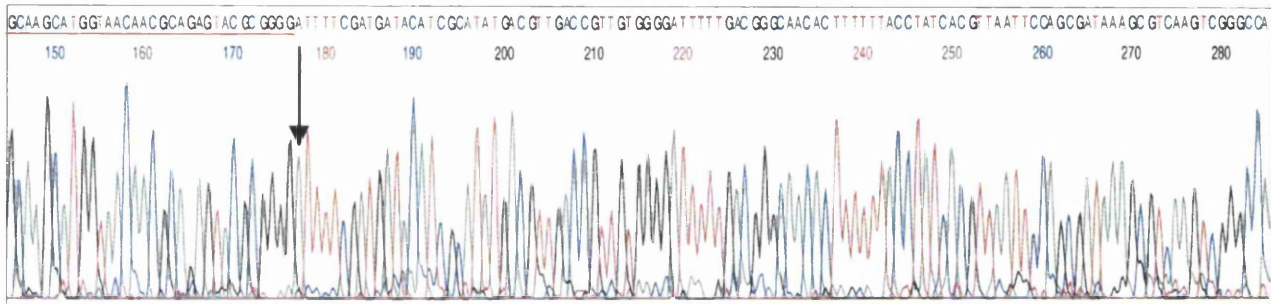


Figure 7.3: 5'-RACE of the HCMV UL18 mRNA.

EtBr-stained 1% (w/v) agarose gels showing representative 5'-RACE products amplified from 5'-cDNA with four different GSPs (listed in Table 7.1). 5'-cDNA was prepared from IE, E or L AD169 mRNA or MI RNA. M1 and M2 are 100 bp and 1 kbp ladders respectively. 1; primer PEL18. 2; primer PEL18 (II). 3; primer UL18. 4; primer UL19.

GGCACGGTCTCCCGAAGAGGAGGAGTAAACAACACACGGCTAAGAGGATACATCATCAAAGAAGATAGGAGGGGTCAAACGCGGACTGAAAGTATATAA 23600
 M T M W C L T L F V L W M L R V V G M H
 CGCCGATCATGTCCGAGGAACGTTAATAAAACGCCATGATGACAAATGGTGTCTGACGTTGTTTGTGCTGGATGTTGAGAGTGGTGGGAATGCACG 23700
 V L R Y G Y T G I F D D T S H T L T V V G I F D G Q H F F T Y H V
 TGTTGCGTTACGGGTACACGGGGATTTCGATGATACATCGCATATGACGTTGACCGTTGTGGGGATTTTTCACGGGCAACACTTTTACCTATCACGT 23800
 N S S D K A S S R A N G T I S W M A N V S A A Y P T Y L D G E R A
 TAATTCACGCGATAAAGCGTCAAGTCGGGCCAACGGTACCATTCTTGATGGCTAACGTCTCGGCGGCCTACCCACCTACCTGGACGGGGAAAGAGCC 23900
 K G D L I F N Q T E Q N L L E L E I A L G Y R S Q S V L T W T H E
 AAAGGTGACCTTATTTTAAACCAAACCGAGCAAACCTGTAGAGCTGAAAATTGCGTTGGGTTACCGGTCACAGAGCGTGTGACGTGGACGCACGAGT 24000

I



II

Figure 7.4: Sequence at a 5'-end of the UL18 mRNA.

Panel I shows a portion of AD169 UL18 and its encoded amino acids, with the coordinates on the right. Amino acid residues conserved in CCMV are underlined. RACE primers are shown in italic bold font with orientations indicated by arrows. The black vertical arrow indicates the position of 5'-end of the UL18 as represented in the majority of 5'-cDNA clones. The first ATG codon in UL18 (as originally described by Chee *et al.*, 1990) and the ATG likely to be used as initiation codon by the mapped 5'-end are shown in red and blue boxes, respectively. A potential TATA element is underlined in red. Panel II shows an electropherogram of the 5'-RACE product (upper strand of AD169 genome) with the 5'-end indicated by an arrow and the SMART sequence underlined in red.

Primer UL19 yielded two major bands of 800 and 1.35 kbp (Fig. 7.2s). The former mapped the 5'-end of UL19 to 24269, which is 28 bp downstream from a potential TATA element. The other band corresponds to the 5'-end of UL18 as described above (Fig. 7.3). This indicates that ORF UL18 and UL19 are 3'-coterminal. A candidate polyA signal (AATAAA) is located at 25216-25221, downstream from UL19.

The product of ORF UL20 is homologous to the T-cell γ -chain receptor (Beck and Barrell, 1991) and is dispensable for replication in cell culture (Mocarski, 1996). Primer UL20 yielded multiple bands from L RNA (Fig. 7.2t). Three bands (approximately 400, 500 and 600 bp) were analysed, placing 5'-ends at 25253, 25161 and 24979, respectively. Each is located upstream from the first ATG codon in UL20 and has a potential TATA box within 24-33 nucleotides. The 5'-end at 24979 represents the major UL20 transcript. It is located 300 bp upstream from the first ATG codon in UL20, with the intervening sequence containing six other ATG codons in different frames.

Primer UL21 generated a single band of 300 bp from L RNA (Fig. 7.2u) that mapped a 5'-end to 26858, downstream from the first ATG codon in UL21. The experiment was repeated using a second primer UL21R, and yielded the same end (data not shown). The 5'-end is preceded by a potential TATA element and the first ATG codon in the mRNA is in an alternative reading frame from UL21. This is consistent with the implication from sequence comparisons that UL21 does not encode a protein (Davison *et al.*, unpublished data). Comparison of three-frame translations of the relevant regions of the AD169 and CCMV genomes indicates that AD169 UL21 (shown in blue in Fig. 7.5 (panel I)) is not well conserved in CCMV (Fig. 7.5 (panel II)). However, a CCMV ORF (UL21A) (shown in red in Fig. 7.5 (panel II)) is conserved in AD169 (bold black font with conserved residues underlined). As a consequence, UL21 is replaced by a novel gene, UL21A, in AD169. The proteins encoded by UL21A in AD169 and CCMV lack homologues in protein databases.

UL25 encodes a virion structural protein (Baldick and Shenk, 1996). Primer UL25 yielded two major bands of approximately 400 and 450 bp from L RNA (Fig. 7.2v). Consistent with this, the UL25 transcript was detected by northern blotting only at L times post infection (Battista *et al.*, 1999). The larger product places the 5'-end of UL25 at 30016, 29

L21A CTGACAATGGACTGAATATACAGACTTTTATATGATCCTTGACAGATGTAATAAATGTTTTATTTAAACTGGTCCCAATGTTCTTCGGGAATCAT 26500
 Q C H V S Y V S K - I I R T C I Y I F H K - K F S T G I N K P F -
 S L P S F I C V K I H D K Y L H L Y F T K I - F Q D W H E E P I M
 S V I S Q I Y L S K Y S G Q V S T F L I N K N L V P G L T R R S D H

L21 GGGGTGGGGACGGGGACCGGTAAAGGAGCAAACCGGGTACATGGGGGGGAACATCGTCCAGCAGTAGCACCAGCGGATTGGGTAGGGGTTGCTGCGGA 26600
 P T P V P S A T L L L V P Y M P P F M T W C Y C W R I P Y P N S R L
 P H P R P V R Y P A F G P V H P P V D D L L L V L P N P L P Q Q P
 P P S P P R P L S C F R T C P P S C R G A T A G A S Q T P T A A S

GTCGGTCGATGACGATGTCGATCTCCATCGGCAGATCCGGCAACATCTCTTCGTCTCCCTCACCAGCAGCACTCGGCGCTGTTCTGGATGTATATGAT 26700
 D T S R S T S W R C I R C C R K T E R V S W C E A S N Q I Y I I
 P R D I V I D I E M P L D P L M E E D G E G V L V R R Q E P H I H N
 T P R H R H R D G D A S G A V D R R R G - R G A S P A T R S T Y S

TTTGAAAAGCCTCCGACGAGCTCGCGGCGCGTAGAAAGCCAAGCGGCGCAAGGGCCGGCGAGCCCAGAAAGTCCATGCGCACAGATGGCATGAGTCTTG 26800
 K S F G G V L E R K T S E W A A P P P L G F T W A C L H C S D K
 Q F L R R R A R P A Y F A L R R L P R R A R F D M R V S P M L G Q
 K P F A E S S S A A R L F G L P A L A P S G S L G H A C I A H T R S

AGTGACGGTGGTGGCTGGGGAACAGGGCTACCTCCCATCGCGACGGTGACAGTGGATCCATGAGAGAGGGCCGACGCTGCATGGCTAAATACCGTGA 26900
 L S P P S S P F L A V E W R S F S L P D M L S A G C A A H S F V T F
 T V T T L Q P V P S G G M A V T V T S G H S L R R V S C P - I G H
 H R H H A P S C P - R G D R R H C H I W S L P A A R Q M A L Y R S

ATCCCCTGACGTCGTCTTTCGTCCCAGACGGTGCATGTTGGGGCGAGGCGTAAACCGTTCGAGGTTGAAAAACCGGTATCTGCGACCCGTCGGGACTAC 27000
 G R V D D K T G F A D H Q P R P T F R R P Q F V A Y R R G T R V V
 I G Q R R R E D R V R - T P P S A Y V T S T S F G R I Q S G D P S R
 D G S T T K R G S R T M N P A L R L G D L N F F R T D A V R G S -

GTTGTTTTTCAGAAGCGCCACATGACCTCGAGATGTCGTACCCAAAGGTATTTAACGGCACACAGCCAGACGCGTTCGTCAGCAGCGACGCCGACAAGA 27100
 N N K L L P W H V E L H R - G L Y K V A C L W V R E D A A V G V L
 Q K E S A A V H G R S T T V W P I - R C V A L R T R - C R R R C S
 T T K - F R G C S R S I D D G L T N L P V C G S A N T L L S A S L V

I


```

TGGTGTGCGTCACCATCTGATGGACATAGACTCTCATTGAGAGACATAAACTACGGCTTCCATGAAAGCACAGAGACTGTACAAAATCCACTTTT 24100
P T R - W R I S M S E - E S V Y C - P K G H F C L S Q V F N W K K
T H T V M Q H V Y V R M - L C L V V A K W S L V S V T C F E V K K
H H A D G D S P C L S E N L S M V S R S E M F A C L S Y L I G S K K

TATTGATGTATACATCATGTAATAAAAATACAAGCTTGTATTTTCTATTAAAAATGTCCAGTTCTCTTCGCGTATCACGGGGTACTCGCGGGGCCT 24200
- Q H I C - T I F Y L S T N E - K F I N G T R K A Y - P T S A P C E
UL21A I S T Y M M Y Y F V L K Y K R I - F H E W N E E R I V P Y E R P V
N I Y V D H L L I C A Q I K K N L F T G L E R R T D R P V R P A S

CGGAAAGGAGCGAAGCCGGGCGGCATGGGCGAACGTCGTCCAACAGCAGCACCAGCGGGTTGGGCAGCGGGCGGTGCGGGCGGTATATCGGCGGCACCT 24300
S L L S A P R C P R F T T W C C C W R T P C R A T P R Y I P P S R
R F E A F G F P M P P V D D L L L V L P N P L P R D F P I D A A V Q
P F S R L R A A H A S R R G V A A G A P Q A A P P R A T Y R R R G

GGATCTCCCTCGGCAGGTCCATCGGTATCGCTTCTCGTCCCCTCCGGGCACAGCGACGATCCGATGAATCGGCGGGCGGACCTGAAGGTGCTGGTGATG 24400
S R G R C T W R Y R K R T G E P C L S S G I F R R R S R F T S T I
I E R F L D M P I A E E D G G F V A V I R H I P P P V Q L H Q H H
P D G E A P G D T D S G R G R R A C R R D S S D A A P G S P A P S A

CTGATGGTAGTTCTCAAAAAGCCGGCGCCGGGCCCGGAGCATAGAAAAGCCAGACGGCGCAGGGCCGGCGGCTCGGAACAGGTCCGCGCGGACCAG 24500
S I T T R L F G A G P G R L M S L W V A C P G A P E S C T R A S W P
Q H Y N E P L R R R A B P A Y F A L R B L P R R A P F L D A R Y V
S P L E - F A P A P G A S C L F G S P A P A P P S P V P G R P G R

GGCTCACACCCTCGGTGGCCACCGAGAGATCGGGTCCGGACTGCCTCCCATATGCGCACTCATGATGATTCCAGGCACGGTACATTACTAAATACCG 24600
G - V R P P W R S I P T P V A E W I A C E H H N W A R Y M V L Y R
P R V G E T A V S L D P D E S G G M N R V - S S E L C P V N S F V T
A E C G R H G G L S R P R S Q R G Y Q A S M I I G P V T C - - I G

```

II

Figure 7.5: Comparison of the region containing AD169 UL21 with the cognate CCMV sequence.

Panels I and II shows three-frame translations of the relevant regions of the AD169 and CCMV sequences, respectively, with nucleotide coordinates on the right. UL21A is oriented right to left. The locations of RACE primers are indicated by arrows in panel I. AD169 UL21 is shown in red in panel I and CCMV UL21A is shown in red in panel II. An ORF in AD169 that appear to be conserved in CCMV UL21A are shown in bold blue font in panel I, with conserved amino acid residues underlined. The vertical arrow indicates the position of the 5'-end of the UL21A mRNA as represented in the majority of 5'-cDNA clones. The first ATG codon in UL21 (as described by Chee *et al.*, 1990) and the first ATG codon in UL21A are shown in red and blue boxes, respectively. The putative TATA element and polyadenylation signals are underlined in red and green, respectively.

bp downstream from a potential TATA element. The smaller product represents a 5'-end at 30065, 7 bp downstream from the first ATG codon in the UL25. This 5'-end has no potential TATA element upstream.

Primer UL27 yielded a major band of 1 kbp (Fig. 7.2w) from L RNA. The corresponding 5'-end is located 650 bp upstream from the first ATG in the ORF, with the intervening sequence containing nine ATG codons in different frames. A potential TATA box is not present immediately upstream from this 5'-end. The less intense band of 550 bp detected in this experiment (Fig. 7.2w) was cellular in origin.

A major band of 250 bp was produced in L RNA using primer UL30 (Fig. 7.2x), mapping the 5'-end of UL30 to 37525. This end is located upstream from the first ATG codon in UL30 but is not preceded by a potential TATA element.

7.2 Discussion

The aim of this chapter was to map the 5'-ends of UL1-UL30, excluding six genes (UL22A, UL23, UL24, UL26, UL28 and UL29) whose transcription is described in chapters 4 and 5. The ORFs within the region of U_L immediately adjacent to TR_L belong to the RL11 family (Chee *et al.*, 1990). This family consists of 12 ORFs, three in TR_L and nine (UL1, UL4, UL5, UL6, UL7, UL8, UL9, UL10 and UL11) in U_L that are predicted to encode glycoproteins.

Of the 25 ORFs analysed, potential 5'-ends were obtained for 22, including six members of the RL11 family (UL4, UL6, UL7, UL8, UL10 and UL11). Most of these 5'-ends have a potential TATA element within 20-35 bp and are positioned such that the first ATG in the ORF is also the first in the mRNA. Kinetic classes indicated by RACE correspond well to those obtained by Chambers *et al.* (1999) using gene microarray technology. However, RACE-PCR is not a quantitative method, and expression kinetics should be confirmed by northern blotting. Transcription from certain genes (e.g. UL4, UL16, UL25) showed perfect concordance with results obtained earlier by Rawlinson and Barrell (1993), Alderete *et al.* (2001), Kaye *et al.* (1992a) and Battista *et al.* (1999) by northern blotting.

5'-ends mapped for certain genes (UL4, UL7, UL10, UL20 and UL27) are located far upstream from the first ATG codon in each ORF, with intervening sequences containing several ATG codons in other reading frames. This is generally considered to be inconsistent with efficient translation. However, all these ORFs are conserved in CCMV, and are presumably expressed. One example is UL20, which has three 5'-ends located 45-319 bp upstream from the first ATG codon with the furthest upstream being the most abundant. Consequently, presence of low level expression of shorter transcripts with more appropriately located 5'-ends cannot be ruled out. UL4 has been shown previously to use a distal promoter to encode a structural glycoprotein. The manner in which translation of this mRNA is controlled is unusual. A 22-codon ORF (uORF2) in the 5'-leader sequence of the UL4 mRNA is conserved in HCMV strains, and the peptide product represses translation of UL4. The nascent uORF2 peptide remains covalently attached to the tRNA that decodes the final uORF2 codon (Cao and Geballe, 1996), and prevents the peptidyl-tRNA hydrolysis reaction during translational termination, resulting in ribosomal stalling and inhibition of downstream translation (Cao and Geballe, 1995; Alderete *et al.*, 1999). It is notable however, that uORF2 is not conserved in CCMV. Nevertheless, a potential exists for similarly complex regulation of the expression of other HCMV genes.

Transcripts of certain genes among those analysed (UL4, UL6, UL8, UL11, UL17 and UL20) appear to be initiated from multiple sites, the majority located downstream from a potential TATA boxes. One small ORF (73 codon) is not conserved in CCMV. As a consequence, it may represent mRNA that does not encode protein. Chambers *et al.* (1999) failed to detect UL12 transcript using gene microarray technology. Further studies are required to elucidate transcription of this ORF and to determine whether it encodes a functional protein unique to HCMV.

A single 5'-end was mapped for one gene (UL18) in this region that is located downstream from the first ATG codon in the ORF and does not have a potential TATA box. The first ATG codon in UL18 as described by Chee *et al.* (1990) is, however, preceded by a TATA box. Since abundance RACE products were not analysed in this study, it is possible that a major downstream 5'-end for UL18 mRNA has been mapped and a minor upstream 5'-end has been missed.

The 5'-RACE technique revealed the existence of five sets of 3'-coterminal genes in the region analysed: UL2-UL3, UL6-UL8, UL10-UL11, UL16-UL17 and UL18-UL19. However, it is necessary to map the 3'-ends of these co-terminal families to confirm these findings. Another set of 3'-coterminal genes: UL26-UL30, was detected by northern blotting and is described in chapter 5. Sets of 3'-coterminal transcripts, with a different promoter for each gene, are well documented in HCMV (Stenberg *et al.*, 1989; Welch *et al.*, 1991; Adam *et al.*, 1995; Wing and Huang, 1995). Rawlinson and Barrell (1993) reported another set of 3'-coterminal genes (UL4-UL5) but this was not confirmed in this study owing to poor expression of UL5.

Except for UL22A and UL36 (see section 4), no evidence of splicing was found in the region analysed. UL28 is predicted to be spliced, but this was not confirmed using RACE and RT-PCR (see chapter 5). In this study, RACE also failed to confirm predicted splicing in UL8 or in the UL4 and UL5 region as predicted by Rawlinson and Barrell (1993).

This study confirmed the existence of novel gene (UL21A) predicted by Davison *et al.* (unpublished data). This replaced an existing ORF (UL21) as originally described in the AD169 sequence. The transcriptional expression pattern of another novel gene (UL15A; Davison *et al.*, unpublished data) remains to be determined.

This study mapped the 5'-ends of 20 genes (Table 7.1) in the left region of U_L. The majority possess potential TATA box 25-35 bp upstream. The 5'-ends of three genes (UL1, UL5 and UL9) were not determined, suggesting that they are probably expressed at low levels in fibroblasts. Further studies using different techniques (e.g. northern blotting) are therefore needed to elucidate transcription of these ORFs.

CHAPTER 8: RESULTS 6

Transcript mapping of two novel HCMV genes

Transcript mapping of two novel HCMV genes

On the basis of the conservation of protein coding regions and other sequence features, Davison *et al.* (unpublished data) proposed five novel genes (UL15A, UL21A, UL128, UL131A and US34A) in AD169, two of which (UL128, UL131A) are predicted to be spliced. In this section, these two novel spliced genes were characterised by northern blotting, RACE and RT-PCR. The GSPs used in Lig'nScribe PCR are listed in Table 2.1. Primers used in RT-PCR and in RACE are listed in Tables 2.2 and 8.1, respectively.

8.1 UL128

8.1.1 Characterisation of spliced transcripts from UL128

ORF UL128, as originally described by Chee *et al.* (1990) (shown in red in Fig. 8.1 panel I) is partially conserved in CCMV at its 3'-end (Fig. 8.1 (panel II)). A HCMV ORF overlapping the 5'-end of UL128, and another overlapping UL129 (bold black font with conserved residues underlined in Fig. 8.1 (panel I)) are conserved in CCMV. Splice donor and acceptor sites are conserved in the appropriate register in CCMV and HCMV. UL128 is thus predicted to be spliced, with exon 3 containing part of ORF UL128 as originally defined. Exon 1 appears to be naturally frameshifted in CCMV.

RT-PCR from this region of the AD169 genome using the primers shown by black arrows in Fig. 8.1 (panel I) generated multiple bands (Fig. 8.2) from E and L RNA. Three bands (0.55, 0.45 and 0.3 kbp) were analysed. The 0.55 kbp band corresponds to the unspliced DNA sequence and could have originated either from unspliced RNA or from contaminating DNA. The 0.45 kbp band represents partially spliced RNA that contains the second intron. The 0.3 kbp band represents the fully spliced form. This result confirms the prediction made by Davison *et al.* (unpublished data). The sequences of the RT-PCR products with the relevant electropherograms are shown in Fig. 8.3 (panels I, II and III). The novel UL128 gene is thus a 5'-extended, spliced form of the original UL128 ORF described by Chee *et al.* (1990), and hence is also named UL128. The coding region is shown in red in Fig. 8.3 (panel I).

ORF	Primer	Sequence (5'-3')	Position ^c	PCR product(s) ^d
UL128	UL128R ^a	TTGTAGTTGCAGCTCGTCAGTTTG	175204-175227	300, 450
UL128	UL128-S2 ^b	GCTGCTATTGGGTCACAGCCGCGT	175581-175558	500, 650
UL131A	UL131R ^a	AACGGGTTTGGTCGGGCAGCGCGC	176677-176700	No product
UL131A	UL131-1 ^a	CGTTGTCATGCACCTTTAGCGCGT	176557-176580	400, 500
UL131A	UL131-S2 ^b	CATGCGGCTGTGTCGGGTGTGGCT	176826-176803	2000 bp

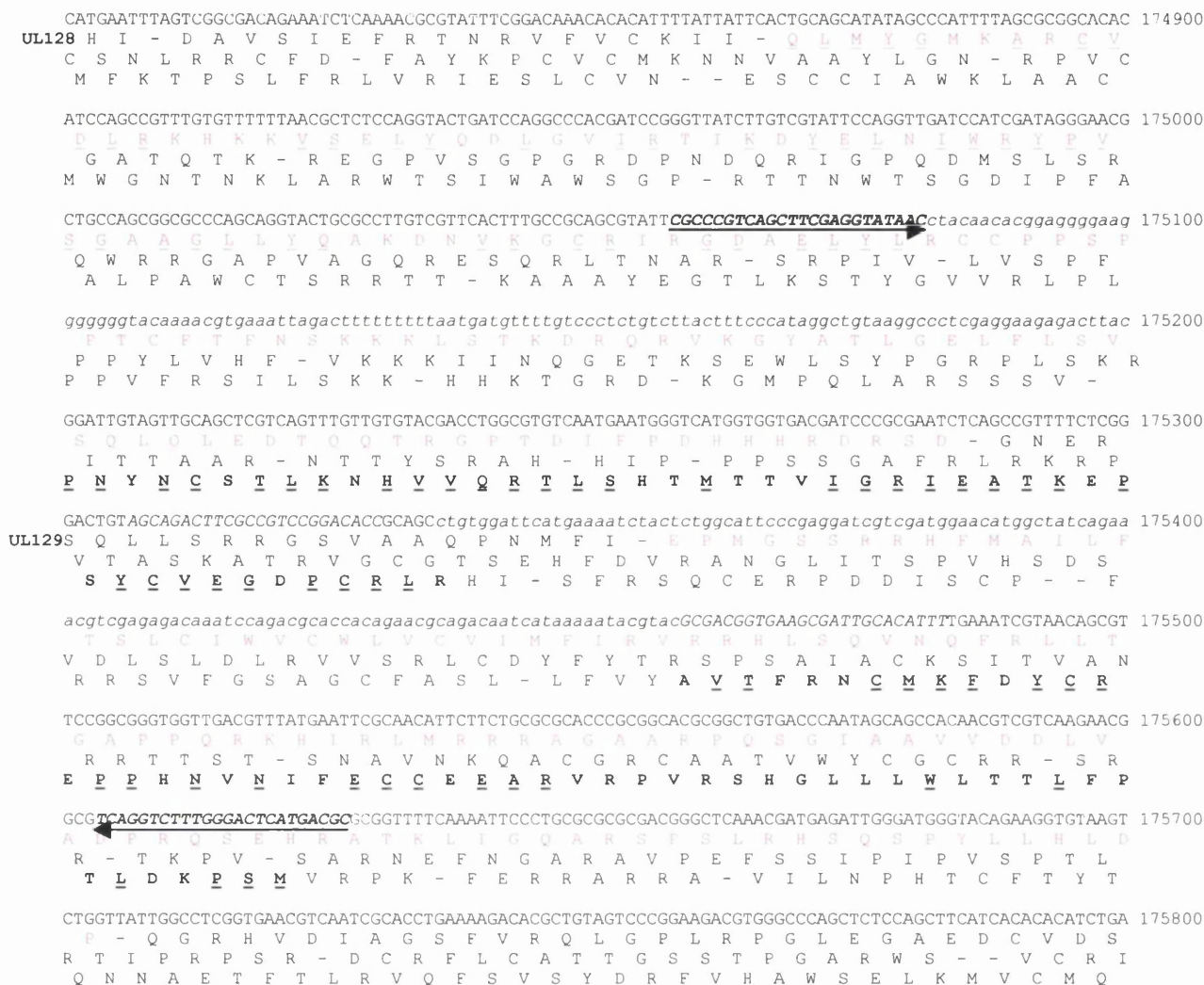
Table 8.1: Primers used in RACE to map 5'- and 3'-ends of transcript from UL128 and UL131A.

^aPrimer used in 5'-RACE.

^bPrimer used in 3'-RACE.

^cPosition of the primers in the positive strand of the AD169 genome.

^dApproximate sizes (bp) of RACE products excised for cloning (See Fig. 8.4 and 8.10).



I

Figure 8.1: Comparison of the region containing AD169 UL128 with the cognate CCMV sequence.

Panel I shows the three-frame translation of the relevant region of the AD169 sequence, with nucleotide coordinates on the right. AD169 UL128 and UL129 as defined by Chee *et al.* (1990) are shown in red, oriented right to left. The predicted introns are shown in lower case. Regions in AD169 that appear to be conserved in CCMV UL128 are shown in bold black font, with conserved amino acid residues underlined. The primers used in RT-PCR and their orientations are indicated by arrows. Panel II shows the three-frame translation of the cognate CCMV sequences with nucleotide coordinates on the right. The CCMV UL128 protein-coding region is shown in red with conserved amino acid residues underlined. The region containing a putative natural frameshift in CCMV UL128 is highlighted in yellow.

TTGGTTTTCTATTGCAGCATGTAGCCAGCTCTGCGTTGCACACGTCCAGTCTCCGGCGGGCCATCACGCTTTCAGGTAGCTGTTCAGGTCGGTGATCC 177900
 P K R N C C T A W S Q T A C T W D G A P W - A R - T A T - T P S G
 N T K - Q L M Y G L E A N C V D L R R R A M V S K L Y S N L D T T I R
 Q N E I A A H L G A R R Q V R G T E P P G D R E E P L Q E P R H D

GCGCGGTTTCTCGTATTCCAGGCTGATCCAGCGGTACGGTACGCTGCCCGTGGCGCCCAAGATGTACTGGGGCTGTTCTTCTTGACCTTAGCGCAGCG 178000
 G P N R T N W A S G A T R Y A A R P A W S T S P S N K R S R L A A
 G P K E Y E L S I W R Y F V S G T A G L I Y Q P Q E E K V K A C R
 A R T E R I G P Q D L P V T R Q G H R G L H V P A T R G Q G - R L A

CAGCTTACCGTCTGATTCCAAGTACAGCctgcggtacaaaaaaggacatgcatatccataactttctctgtcccgcgccccgtaaaactgtaagtat 178100
 C S V T Q N W T C G A P V F F P C A Y G Y S E R D R A G Y V T L N D
 L K G D S E L Y L R R T C F F S M C I W L K R Q G A G R L S N L -
 A - R R I G L V A Q P Y L F L V H M D M V K E T G R G T F Q - T I

cctccttcccgaagcgaggcacttacGGGTTGTGGTTGCACGCGTCCAGCTTATCGTGTACCACCTGCCGGGTCCAGAGCGGGCCATGCCCATGATGA 178200
 E K G F A L C K R T T T A R R - S I T Y W R G P - L A P W A W S S
 G G E R F R P V - P N H N C A T L K D H V V Q R T L S R A M G M I I
 R R G S L S A S V P Q P Q V G D A - R T G G A P D S L P G H G H H

TACCGCGATCTCCGCGTCTTTGGGGGTGTAGCACACTCGCCCTCAGGGCATCGTAGCctggagatatcatcaagcgatttgtgtatgtgttcca 178300
 V A S R R R R K P P T A C R A R L A D Y G P S I M V A I Q T H T G
 G R I E A T K Q P T Y C V E G E P C R L K S I D D C R N T Y T N W
 Y R P D G G D K P P H L V G R G - P M T A Q L Y - - L S K H I H E M

tcaggatagagaagcgcagacgaggatccgcacatacaagcaccgaggacgaacacgcagatacctaCGAGACGGTGTAAAGGGTGCACATCTTGAAG 178400
 D P Y L L R L R P D A C V L V A P R V R L Y R R S P T L P A C R S T
 - S L S S A S S S G C M C A G C S S C A S V - S V T Y P T C M K F
 L I S F V C V L I R V Y L C R L V F V C I G V L R H L P H V D Q L

TTGTAGCACTTCCCGGGGGGTAGGCGATGTTACCATCTCGCAGCATGATTCCGCGCGCGGAACGAGCGGGACGGGTTCCGGCCACATCATATACAGC 178500
 T A T G P F F F L R H E G D R L M I G R A R V L P V P E A V D Y V A
 N Y C N G P P T P S T - W R A A H N R A R S R A P R T R G C - I C C
 Q L L E R P P Y A I N V M E C C S E A R A F S R S P N P W M M Y L

AACGTCAAGGACAGCAGCGGGCCGGAGACATGACGACGCGGGTTCACAAAATCCCGTGCAGCGCGGGGCTCAGACGAGGAGATTGGGGTGC 178600
 V D L V A A P G S V H R R R R N G F I G T R A R R A - V L L N P H
 R - P C C R A R L C S S S P P K W F D G H A R P P S L R P S Q P A
 L T L S L L P G P S M V V V A T E L F G R A R A A P E S S S I P T R

II

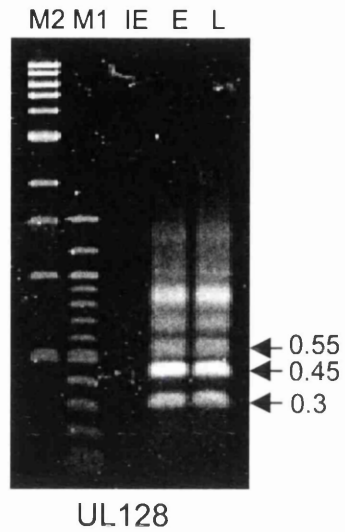


Figure 8.2: Gel electrophoresis of RT-PCR products from HCMV UL128.

EtBr-stained 1% (w/v) agarose gel showing RT-PCR products (sizes in kbp) amplified from HCMV IE, E and L RNA using primers UL128-S1 and UL128-S3. M1 and M2 are 100 bp and 1 kbp ladders, respectively.

128

↓

CATGAATTTAGTCGGCGACAGAAATCTCAAACGCGTATTTCCGGACAAACACACATTTTATTATTCACTGCAGCATATAGCCATTTTAGCGCGGCACAC 174900
 - Q L M Y G M K A R C V

ATCCAGCCGTTTGTGTTTTTAAACGCTCTCCAGGTACTGATCCAGGCCACGATCCGGGTATCTTGTGCGTATTCCAGGTTGATCCATCGATAGGGAAAC 175000
 D L R K H K K V S E L Y Q D L G V I R T I K D Y E L N I W R Y P V

CTGCCAGCGGGCCAGCAGGTACTGCGCCTTGTCTTCACTTTGCCGAGCGTATTCCGCCGTCAGCTTCGAGGTATAACctacaacacggagggaag 175100
 S G A A G L L Y Q A K D N V K G C R I R G D A E L Y L

gggggtacaaaactgaaattagacttttttttaatgatgttttgcctctgtcttactttcccataggctgtaaggcctcgaggaagagacttac 175200

GGATTGTAGTTGCAGCTCGTCAGTTTGTGTGTACGACCTGGCGTGTCAATGAATGGGTCATGGTGGTGACGATCCCAGCAATCTCAGCCGTTTTCTCGG 175300
 P N Y N C S T L K N H V V Q R T L S H T M T T V I G R I E A T K E P

GACTGTAGCAGACTTCGCCGTCGGACACCGCAGCctgtggattcatgaaatctactctggcattcccaggatcgtcgatggaacatggctatcagaa 175400
 S Y C V E G D P C R L

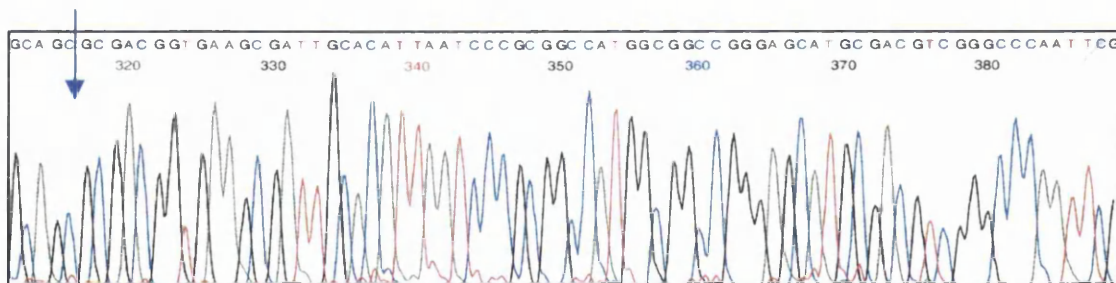
acgtcgagagacaaaatccagacgcaccaagacagacaatcataaaaatacgtacGCGACGGTGAAGCGATTGCACATTTGAAATCGTAACAGCGT 175500
 A V T F R N C M K F D Y C R

TCCGCGGGTGGTTGACGTTTTATGAATTCGCAACATTCTTCTGCGCGCACCCGCGGCACGCGGTGTGACCAATAGCAGCCACAACGTCGTCAAGAACG 175600
 E P P H N V N I F E C C E E A R V R P V R S H G L L L W L T T L F P

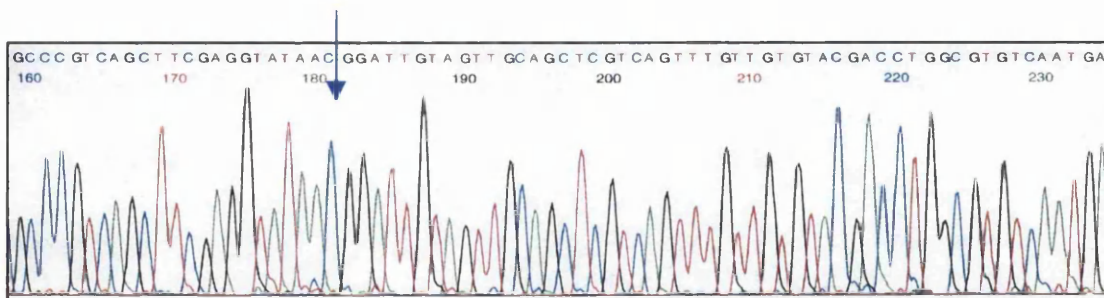
GCGTCAGGTCTTTGGACTCATGACGCGCGGTTTTCAAATCCCTGCGCGCGCAGGGCTCAAACGATGAGATTGGGATGGGTACAGAAGGTGTAAGT 175700
 T L D K P S M

CTGGTTATTGCCTCGGTGAAACGTAATCGCACCTGAAAAGACACGCTGTAGTCCCAGGAAAGACGTGGGCCAGCTCTCCAGCTTCATCACACATCTGA 175800

I



II



III

```

GATCTCTCATACACAGAAGCGCCGCGACGAGTTTTTTATTATTATATCTTTATTACACAGCGACAAGACACGGCAACAGTGTAAATGGTAGCGGTGCGTTA 177800
TTGGTTTTCTATTGCAGCATGTAGCCAGCTCTGCGTTGCACACGTCCAGTCTCCGGCGGGCCATCAGCTCTTCAGGTAGCTGTTTCAGGTGCGGTGATCC 177900
- Q L M Y S L E A N C V D L R R R A M V S K L Y S N L D T I R
GCCCGGGTTTCTCGTATTCCAGGCTGATCCAGCGGTACGGTACGGTACGGTACGGTACGGTACGGTACGGTACGGTACGGTACGGTACGGTACGGTACGGT 178000
G P K E Y E L S I W R Y P V S G T A C L I Y Q P Q E E K V K A C R
CAGCTTACCGTCTGATTCCAAGTACAGCctgcggttacaataaaaggacatgcatacctaactttctctgtcccgcgccccgtaaaactgttaagttat 178100
L K G D S E L Y L
cctcctcccgaagcgaggcacttacGGGTTGTGGTTGCACGCCGTGAGCTTATCGTGTACCACCTGCCGGGTGAGAGAGCGGCCATGCCCATGATGA 178200
P N H N C A T L K D H V V Q R T L S R A M G M I I
TACCGCGGATCTCCGCCGTCTTTTGGGGGTGTAGCACACCTCGCCCTCAGGGCATCGTAGCctggagatatcatcacagcagatttgtgtatgtgttcca 178300
G R I E A T K Q P T Y C Y E G E F C R L
tcaggatagagaagacgcagacgagatccgcacatacaagcaccgaggacgaacacgcagatacctacGAGACGGTGTAAAGGGTGCACATCTTGAAG 178400
S V T Y F T C M K F
TTGTAGCAGTTCCCCGGGGGGGTAGCGGATGTTACCATCTCGCAGCATGATTCGCGCGCGCAACGAGCGGGGCGGGTTCGGCCACATCATATACAGC 178500
N Y C N P P P Y A I N V M E C C S E A R A P S R S P N P W M M Y L
AACGTCAAGGACAGCAGCGGGCCCGGAGACATGACGACGACGGCGGTTTCCAAAAATCCCCGTGCGCGCGCGGGGCTCAGACGAGGAGATTGGGGTGC 178600
L T L S L L P G P S M

```

IV

Figure 8.3: Interpretation of splicing in HCMV UL128 by RT-PCR.

Panel I shows the sequence of AD169 UL128, with nucleotide coordinates on the right. ORF UL128 is shown in red with conserved residues underlined. The gene is oriented right to left. Introns are shown in lower case. The red and green vertical arrows accordingly indicate the position of the 5'- and 3'-ends of UL128 mRNA as represented by the majority of 5'- or 3'-cDNA clones. A potential TATA box and polyadenylation signal are underlined in red and green respectively. Panel IV shows the corresponding CCMV sequence, with coordinates on the right and introns in lower case. The region containing a possible natural frameshift in CCMV UL128 is highlighted in yellow. Panels II and III show electropherograms of the RT-PCR product (upper strand of AD169 genome), with the splice sites indicated by blue arrows.

8.1.2 Mapping the 5'- and 3'-ends of the UL128 mRNA

Primer UL128R, which maps in exon 2, yielded two major bands (450 and 300 bp) from E and L RNA in 5'-RACE (Fig. 8.4a). The 450 bp band corresponds to the sequence lacking the first intron, and mapped a 5'-end of UL128 to 175681. A potential TATA element is present 24 bp upstream from this 5'-end. The 300 bp band that originated from L RNA was cellular in origin.

Primer UL128-S2, which maps in exon 1, yielded two major products (500 and 650 bp) from E and L RNA in 3'-RACE (Fig. 8.4b). Each band corresponds to a polyadenylation site at 174842, 15 bp downstream from a putative polyadenylation signal (AATAAA). The 3'-RACE cDNA sequence obtained from the smaller product (500 bp) lacked both introns, but the 650 bp band corresponded to the sequence containing second intron and lacking the first intron.

8.1.3 Northern blot analysis of the UL128 mRNA

Northern blotting with a UL128 probe identified multiple bands from E and L RNA (Fig. 8.5). The RACE experiments indicated that the 5'-end of the UL128 transcript is at 175681 and the 3'-end is at 174842 (880 nucleotide unspliced transcript), and RT-PCR indicated that size of the first and second introns are 123 and 119 bp, respectively (638 nucleotide fully spliced transcript). Therefore the 0.8 and 0.6 kb bands in L RNA probably correspond to the unspliced or partially spliced and fully spliced UL128 transcripts, respectively. The bands of 1.2, 2.0 and 3.0 kb in Fig. 8.5 probably correspond to 3'-coterminal UL130, UL131A and UL132 transcripts, respectively, spliced in the UL128 region (see section 8.2). The origins of the 4.0, 6.5 and 7.5 kb bands are unknown.

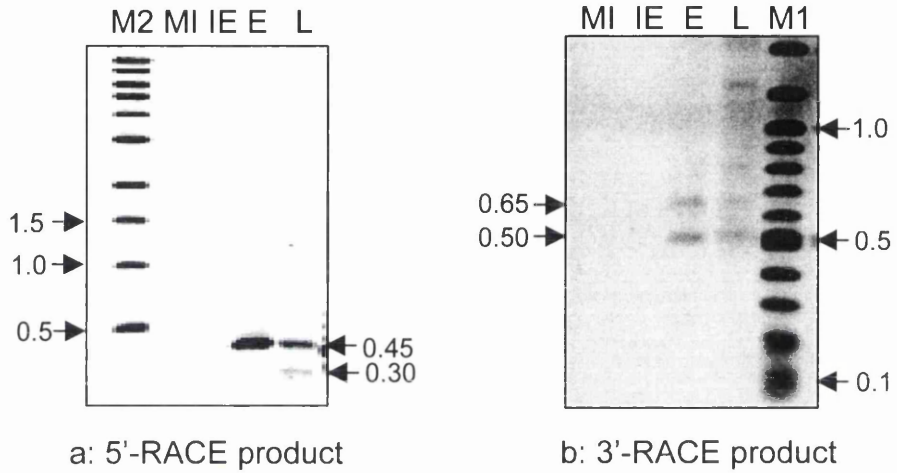


Figure 8.4: 5'- and 3'-RACE of HCMV UL128 mRNA.

EtBr-stained 1% (w/v) agarose gels showing RACE products of HCMV UL128 that were amplified from 5' or 3'-cDNA with GSPs (listed in Table 8.1) and UPM. Sizes are in kbp. (a) 5'-RACE products with primer UL128R. (b) 3'-RACE products with primer UL128-S2. 5' or 3'-cDNA was prepared from different kinetic classes (IE, E and L) of AD169 mRNAs and from MI RNA. M1 and M2 are 100 bp and 1 kbp ladders, respectively.

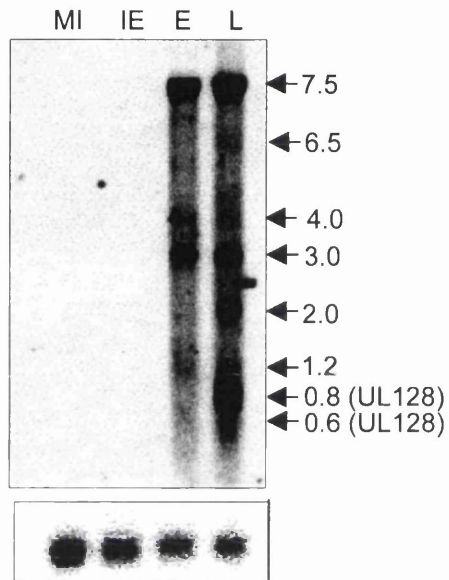


Figure 8.5: Northern blot analysis of HCMV UL128 mRNA.

Northern blot showing the expression profile of transcripts from HCMV UL128. MI, IE, E and L RNA were hybridised with a single-stranded RNA probe (top) or a GAPDH₂ probe (bottom). Transcript sizes are given in kb.

8.2 UL131A

8.2.1 Characterisation of spliced transcripts from UL131A

The novel gene UL131A is naturally frameshifted in AD169 (see section 3.1.5). Amino acid sequence comparisons indicated the presence of a spliced gene in CCMV corresponding to HCMV UL131A (see Fig. 3.11; section 3.1.5), with the predicted splice donor and acceptor sites in the same register in both genomes. Therefore, HCMV UL131A is predicted to be spliced.

RT-PCR amplification of the relevant region in AD169 revealed two bands (Fig. 8.6) from L RNA. The 0.4 kbp band corresponded to the unspliced sequence and could have originated either from unspliced mRNA or from contaminating DNA. The 0.3 kbp band corresponded to the spliced mRNA. The splicing patterns in AD169 UL131A in comparison with CCMV UL131A are shown in Fig. 8.7 (panels I and III).

8.2.2 Mapping the 5'- and 3'-ends of the UL131A mRNA

Two primers (Table 8.1) were used in 5'-RACE to determine the 5'-end of AD169 UL131A. Primer UL131R yielded no product (data not shown). Primer UL131-2 yielded a band of 500 bp (Fig. 8.8a) from L RNA and a band of 400 bp from MI, IE and E RNA. Both were cellular in origin. The 5'-end of UL131A was therefore not detected.

Primer UL131-S2, which maps in the first exon, yielded a single band of 2 kbp (Fig. 8.8b) from L RNA in 3'-RACE which corresponds to a 3'-end at 174842, identical to that determined for UL128 using primer UL128-S2 (see section 8.1.2). Therefore, the polyadenylation signal (AATAAA) at 174857-174862 is shared by UL131A and UL128. Northern blotting with a UL128 probe also suggested that UL128, UL130, UL131A and UL132 are 3'-coterminal genes (see section 8.1.3).

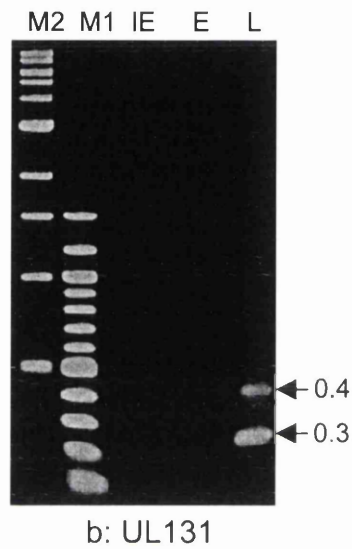


Figure 8.6: Gel electrophoresis of RT-PCR products from HCMV UL131A.

EtBr-stained 1% (w/v) agarose gel showing RT-PCR products (sizes in kbp) amplified from HCMV IE, E and L RNA using primers UL131-S3 and UL131-S2. M1 and M2 are 100 bp and 1 kbp ladders, respectively.

CAGCATATTATTTCCCGTGACGCAGGCTAGTTGGCAAAGAGCCGACGCTGAACTCGAGGCTCCGGGCGTGTGGCGCCAGCGAACCGGGCGTGAACG 176400
 - N A F L R V S F E L S R A H P A L S G A A N E T

TGGTCCTTTTGTGGTGCCGCCGCGACGGTCTGACGTCTAAAGTCGCTGATGAGCAACGACACCTCGGTACGTTGATTctgcaagcacaggttccaaa 176500
 T R K N T G G R R N Q R R F D S I L L S V E T V N I

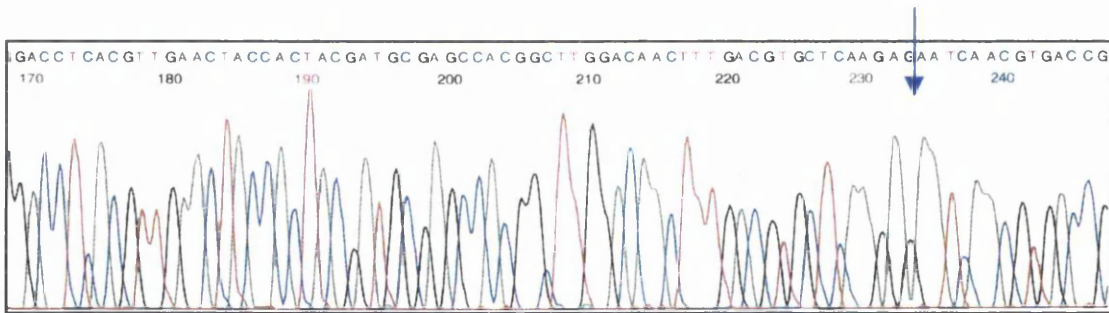
cgtcatttcatacccacatgcggttacttagccgttaccogttccoccttacctcccgttgcacaccttagcgcgtaacctcacCTCTTGAGCAGC 176600
 R K L V

TCAAAGTTGTCCAAGCCGTGGCTCGCATCGTAGTGGTAGTTC AACGTGAGGTCCACGAGCTGTTCCACATACTTGTAACTGGGTTTGGTCGGGCAGCGCGC 176700
 D F N D L G H S A D Y H Y N L T L D V L Q E V Y K Y R T Q D P L A R

GAGAGCACGCGTCCCAGTAATGCGGTACTCGGTAATAATCGTTTTTTTCCGCGGTTTCCCGTGGCACTGACCCAGCACCGGGCGCACAGACAAACAGA 176800
 S C A D W Y H P V R Y Y D N K E A T E R Q C Q G L V V A C L C V S

CAGCCACACCCGACACAGCCGATGTTGCAGACTGAGAAAGAAAGCTTTATTATGAGACATCATACATAGTATAGGCGAGGTGATGGGGCGGGAAAG 176900
 E W V R C L R M

I



II

GGTGCATGACCTCCGCCGCGGTTTTACTGGGAAAGAGCCGACGCTGAACTCCAGCATGCGGTAGTGCAGGTCCTCCGAAAGCCGGAGGCGCTGAACGTGG 179400
 - Q P F L R V S F E L M R Y H P G L F G S A S F T T

TTCTCTTGTGGTCCCGCCGCGGGCCTTCTGGCGCCTGAACTCGCCGCCAGCAGCGTTACCTCCGTTACGTTGATCctgcaacaaaaccccgatttggt 179500
 R K N T G G R A K Q R R F E G A L L T V E T V N I R

cgcgcgcacgtcgttcgggtgcctcttttcgtcctccccttcccgaaacctgacgctcgtcaggctcacCTCTTGAGCGTCTTAAAGTTATCCAAAC 179600
 K L T K E N D L G

CGTGGCTCACGTCGTAGTGGTACTCCACCAACGTGTCCACCAGCGCTTGACGTACTCCTGACGGGTACGCTCCGGCAGCGCCGGGGAACACGTGTCCCA 179700
 H S V D Y H Y E V L T D V L R K V Y E Q R T R E P L A P S C T D W

GTAGTCCCACGGGCGCGCAAAGTCGTCCCCTCGGCTTCCTCCCTGCGACACTGGCCCCACACCCGCGACGCTCAGACACAAAACAGACAGAAACACACGG 179800
 Y D W P R A F D D R E A E E R R C Q G W V A V S L C L V S L F V R

TACAACCACATGTTGCACACTGAAGGTCAACTTTATTGTTACACAGGGAAGAAATATGGACGTGGGTGGTGGTTTACGGGACTCGGACAGGTGACCG 179900
 Y L W M

III

Figure 8.7: Interpretation of splicing in HCMV UL131A by RT-PCR.

Panel I shows a version of the sequence of AD169 UL131A with the natural frameshift corrected by removal of a T residue from the region highlighted in yellow, with nucleotide coordinates on the right. The gene is oriented right to left. The amino acid sequence of UL131A is shown in red with conserved residues underlined and the hydrophobic domain boxed. The intron is shown in lower case. Panel III shows the corresponding CCMV sequence, with coordinates on the right and the intron in lower case. Panel II shows an electropherogram of the sequence of the relevant region of AD169 with the splice sites indicated by a vertical arrow.

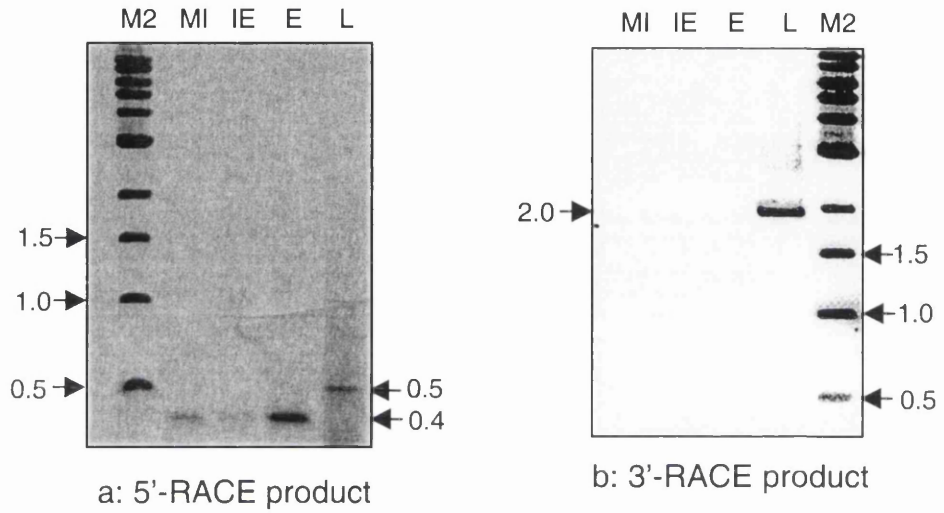


Figure 8.8: 5'- and 3'-RACE of HCMV UL131A mRNA.

EtBr-stained 1% (w/v) agarose gels showing RACE products of HCMV UL131A that were amplified from 5' or 3'-cDNA with GSPs (listed in Table 8.1) and UPM. Sizes are in kbp. (a) 5'-RACE products with primer UL131-1. (b) 3'-RACE products with primer UL131-S2. 5' or 3'-cDNA was prepared from different kinetic classes (IE, E and L) of AD169 mRNAs and from MI RNA. M2 is a 1 kbp ladder.

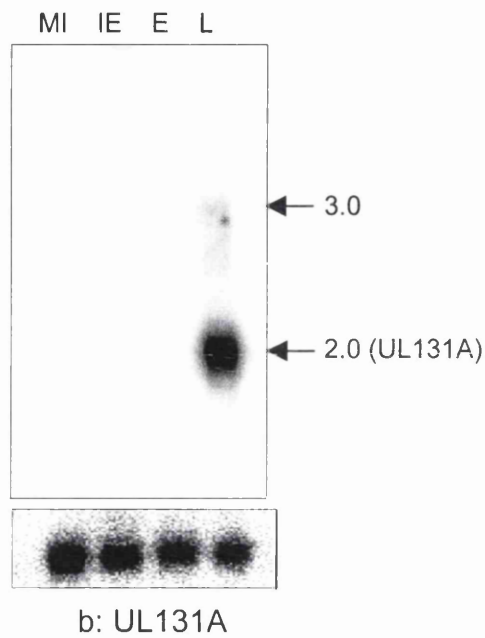


Figure 8.9: Northern blot analysis of HCMV UL131A mRNA.

Northern blot showing the expression profile of transcripts from the HCMV UL131A. MI, IE, E and L RNA were hybridised with a single-stranded specific RNA probe (top) or a GAPDH₂ probe (bottom). Transcript sizes are given in kb.

8.2.3 Northern blot analysis of UL131A transcript

A major band of 2.0 kb was detected on probing a northern blot with a UL131A probe (Fig. 8.9). Consistent with this, the 3'-RACE product and the RT-PCR product of ORF UL131A were generated from L RNA (see section 8.2.1 and 8.2.2). The 5'-end of UL131A was not detected. However, since the initiation codon of UL131A is probably at 176825 and the 3'-end is at 174842 (see section 8.2.2), the unspliced transcript is anticipated to be at least 2023 nucleotides in size and the spliced transcript 1672 nucleotides, including the splicing in UL128. Therefore, the 2.0 kb band probably corresponds to the UL131A transcript. Another faint band (3.0 kb) was also detected in this blot (Fig. 8.9) and may correspond to the UL132 transcript.

8.3 Discussion

Two novel spliced HCMV genes (UL128 and UL131A) predicted by Davison *et al.* (unpublished data) were analysed by RACE, northern blotting and RT-PCR. The transcript mapping data (summarised in Fig. 8.10) confirm and extend the predictions made by Davison *et al.* (unpublished data).

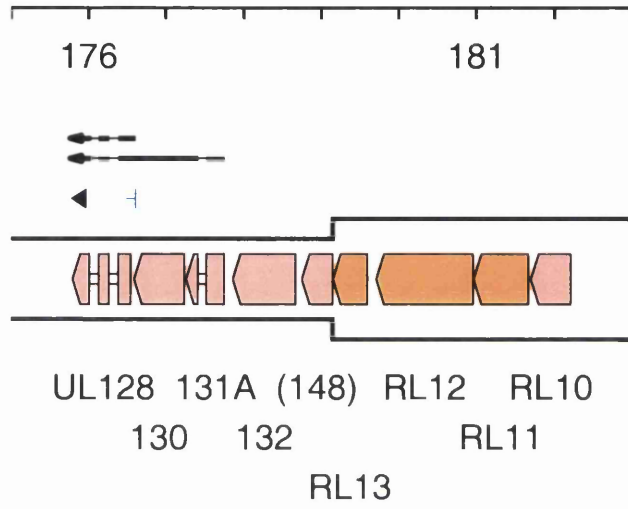
RT-PCR experiments confirmed the presence of three exons in UL128 and two exons in UL131A, and mapped the exact locations of the splice sites. The sizes of the exons and their amino acid sequences are conserved in CCMV. RACE identified a single 5'-end for UL128 and a shared 3'-end for UL128 and UL131A at 175622 and 174842, respectively. The conclusion that these genes are 3'-coterminal was confirmed by northern blotting. Moreover, the northern blot analysis also indicated that UL128, UL130, UL131A and UL132 are 3'-coterminal.

RACE, RT-PCR and northern blotting demonstrated that UL131A is an L gene. The UL128 transcript was detected in E and L RNA using RACE and RT-PCR, but predominantly in L RNA by northern blotting. Chambers *et al.* (1999) detected UL128 from E RNA using gene microarray technology. As PCR methods may not be quantitative, and the gene microarray

Figure 8.10: Transcription from HCMV novel genes, UL128 and UL131A.

Expanded version of parts of the AD169 genome (Fig. 1.9) showing the positions of 5' and 3'-ends of two HCMV novel genes (UL128 and UL131A), with the genome coordinates in kbp and transcriptional features above the relevant ORFs.

- ⊥ 5'-end
- ◀ 3'-end
- ◀ RNA (and orientation)
- exon
- intron



 *RL11 gene family*

 *other non-core genes*

system only detects transcription through the region containing the probe, UL128 is currently considered an L (or E-L) gene.

The RT, 3'-RACE and northern blotting experiments indicated that UL131A is transcribed at a moderate level. The inability to detect 5'-end of the UL131A mRNA was unexpected, and may reflect the efficiency of the primers used.

AD169 UL131A is naturally frameshifted near its 5'-end and, is unlikely to be translated into a functional protein. CCMV UL128 contains a potential frameshift error near its 5'-end and may encode a truncated protein initiated at the second ATG codon. It is interesting to note that UL128 is also non-functional in Toledo due to a local inversion.

The predicted proteins encoded by CCMV and AD169 UL131A contain a short hydrophobic domain near the N terminus (Fig. 8.9) which may act as a signal sequence for translation on membrane-bound ribosomes (McGeoch, 1985). Homologues of UL128 and UL131A are not present in protein databases. At present, the functions of UL128 and UL131A are not known.

CHAPTER 9:

Discussion

9.1 Overall discussion

9.1.1 Purpose of the study

The sequence of the highly passaged AD169 strain was described in 1990 as containing 208 ORFs that may encode proteins. These were predicted on the basis of size, extent of overlap and codon usage. However, the Toledo strain and other low passage isolates were later found to possess 19 ORFs that are absent from high passaged strains such as AD169 and Towne (Cha *et al.*, 1996; Lurain *et al.*, 1999). None of these extra ORFs have detectable counterparts in other β -herpesviruses. An additional 929 bp sequence, unrelated to the extra 15 kbp found in Toledo, was also discovered in most stocks of AD169 (Dargan *et al.*, 1997; Mocarski *et al.*, 1997). These studies provide insights into the gene content of the wild-type HCMV genome. The extra sequence found in Toledo and other low passage isolates is likely to be present in wild-type since they have undergone less laboratory passaging than AD169. Thus, wild-type HCMV has been surmised to contain over 220 genes. However, even Toledo differs from clinical isolates in that a sequence near the right end of U_L is in inverse orientation.

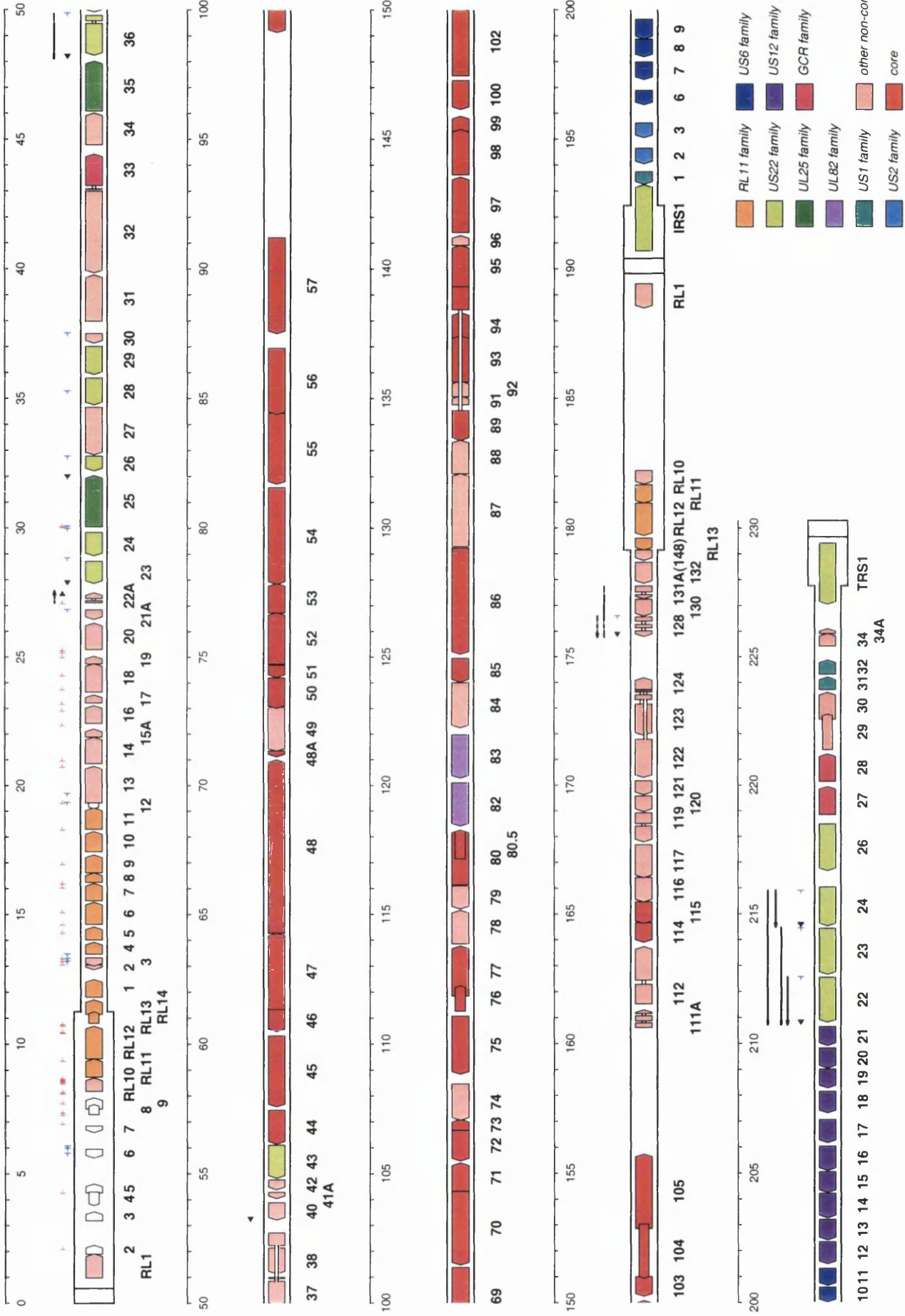
Recently, the gene content of HCMV has been re-evaluated by comparison to the CCMV sequence and the estimated number of unique protein-coding genes in AD169 is now 145 (Davison *et al.*, unpublished data; see Fig. 1.9). The best estimate for the complement of wild type HCMV is now 164-167 genes, ten (UL15A, UL21A, UL128, UL131A, UL147A, UL148A, UL148B, UL148C, UL148D and US34A) of which are novel. It anticipated that this picture of HCMV gene content will be improved further.

Although the AD169 genome was sequenced over ten years ago, transcriptional and functional data are not available for the majority of the genes, and the assignment of gene function to many genes is based only on sequence similarity to homologous genes in HSV-1. Recently, Chambers *et al.* (1999) determined the kinetic class of the majority of predicted HCMV transcripts using gene microarray technology. However, this technique is limited in the extent of information it can supply, and much work is still required to characterise transcripts fully. The purpose of this study was to evaluate transcription of a selection of HCMV genes, including several that are conserved in CCMV, some that appear

Figure 9.1: Transcription from HCMV genes.

The AD169 genome showing the positions of 5'- and 3'-ends of HCMV genes. The genome is shown in five sections, with a scale in kbp and transcriptional features above the relevant ORFs. The approximate locations of genome components are: TR_L; 11 kbp, U_L; 11-179 kbp, IR_L; 179-190 kbp, IR_S; 190-192 kbp, U_S; 192-227 kbp; TR_S; 227-229 kbp. The *a* sequence is shown at the genomic termini and at the junction of IR_L and IR_S. Protein-coding regions (TRL1-TRL14, UL1-UL132 and portion of UL148, IRL14-IRL1, IRS1, US1-US34A, TRS1) are shown as open arrows above the gene nomenclature, with prefixes omitted for clarity. ORFs predicted to be expressed as spliced mRNAs are connected by narrow horizontal bars. HCMV core genes that are common to α -, β - and γ -herpesviruses are shown in red. Gene families are shown in different colors as indicated.

- ↑ 5'-ends of ORFs that are oriented right to left
- ↓ 5'-ends of ORFs that are oriented left to right
- ◀ 3'-end
- ◀ RNA (and orientation)
- exon
- intron



unlikely to encode functional proteins because they are not conserved in CCMV, and two novel genes. Transcript mapping data are summarised in Fig 9.1. Additional work was undertaken to detect sequencing errors in published HCMV sequences.

9.1.2 Sequencing errors

On the basis of conservation of putative protein-coding regions, Davison *et al.* (unpublished data) proposed sequence errors in six ORFs (TRL13/14, UL15, UL102, UL131A, UL145 and US22). The error in UL102 and an additional error near the 3'-end of US28 were reported previously (Smith and Pari, 1995; Kuhn *et al.*, 1995). The present work did not re-examine the latter but confirmed the former. It also added errors in three other ORFs (UL15, UL145 and US22) and confirmed that two other ORFs in AD169 (UL131A and TRL13/14) are naturally frameshifted and therefore probably non-functional. As a result of this study, two novel genes (UL15A and UL131A) were introduced into the AD169 genome, and the protein-coding regions of three genes (UL102, UL145 and US22) were extended at their 5'-ends.

As originally described by Chee *et al.* (1990), the size of the US22 ORF from its first ATG codon was 530 codons, corresponding to a protein product of approximately 60 kDa. However, the error identified in the US22 sequence indicates that translation commences at an upstream ATG codon. The size of the revised US22 ORF is 576 codons, with a predicted protein product of 65 kDa. This is closer to the size of the US22 protein (76 kDa) detected in infected cells and expressed in recombinant bacteria (Mocarski *et al.*, 1988; Dal Monte *et al.*, 1998).

Since detection of these errors that were investigated depended on analysis of amino acid sequence conservation between HCMV and CCMV, it is anticipated that additional sequencing errors, especially in non-coding-regions, are likely to exist in the published AD169 and Toledo sequences.

9.1.3 Novel genes in HCMV

Drawing largely on the correspondence between the HCMV and CCMV gene arrangements, Davison *et al.* (unpublished data) proposed five novel genes in AD169: UL15A, UL21A, UL128, UL131A and US34A, of which two (UL128 and UL131A) were predicted to be spliced. The present work did not examine expression of US34A, but experimentally investigated the other four genes. The presence of UL15A was confirmed from identification of sequencing errors, but transcription of this gene was not analysed. The amino acid sequence of the predicted protein encoded by UL15A possesses a hydrophobic domain near the C terminus and is therefore likely to be membrane-bound protein. Expression of UL21A was confirmed by 5'-RACE. The UL15A and UL22A proteins lack homologues in public databases.

UL128 was predicted to consist of three exons, and UL131A of two. The splice sites were mapped by RT-PCR. HCMV UL89 is the only gene that is spliced in all of the sequenced herpesviruses. In addition, several HCMV non-core genes, including UL22A, UL33, UL36, UL37, UL111A, UL112, UL119, UL122 and UL123 are spliced, and certain of these (e.g. UL36, UL37) make use of alternative, temporally regulated splicing (Mocarski, 1996). Alternative splicing, such as that in the MIE locus (UL122 and UL123) presents an additional means of diversifying functions.

Identification of novel genes from comparisons between the HCMV and CCMV genome was dependent on substantial amino acid sequence homology. Additional spliced genes in poorly conserved regions or splice genes unique to one or other virus (such as is the case for UL111A in HCMV) may remain to be discovered. Consequently, further experimental investigation is necessary. Transcript mapping by the northern blotting, primer extension and cDNA cloning are the obvious routes.

Mapping the transcripts from UL128 and UL131A showed that they are 3'-coterminal. The 5'-end of UL128 was mapped, but the 5'-end of UL131A was not detected. The predicted protein encoded by UL131A contains a short hydrophobic domain near the N terminus

which may act as a signal sequence for secretion. However, homologues of the UL128 and UL131A predicted proteins were not found in protein databases.

The functions of the novel genes are not known. However UL131A is naturally frameshifted in AD169 and UL128 is disrupted in Toledo and CCMV. Thus, HCMV UL131A and UL128 appear to be non-essential for growth in cell culture.

9.1.4 US22 family genes

The HCMV US22 gene family includes genes UL23, UL24, UL26, UL28, UL29, UL36, UL43, IRS1, US22, US23, US24, US26 and TRS1 (Weston and Barrell, 1986; Davison *et al.*, unpublished data). In this study, transcripts were detected by northern blot for nine of these 13 genes. UL28, TRS1, IRS1 and US26 were not analysed. Of the HCMV US22 genes analysed, one is expressed with IE (UL36) and eight with E (UL23, UL24, UL26, UL28, UL29, US22, US23 and US24) kinetics. 5'- and 3'-ends were identified for eight of these genes (UL23, UL24, UL26, UL36, UL43, US22, US23 and US24). The 5'-ends of three (UL36, UL24 and US22) were further confirmed by primer extension assay. Most of the 5'-ends mapped for US22 gene family mRNAs are located 20-35 bp downstream from TATA elements, and the 3'-ends of all the ORFs analysed are located 20-24 bp upstream from AATAAA elements. The majority of TATA elements and all polyadenylation signals are conserved in CCMV.

3'-RACE implied the existence of three sets of 3'-coterminal genes containing members of the US22 family: UL23-UL24, US23-US24 and UL40-UL47. The former two sets were confirmed by northern blotting. In addition, northern blotting also indicated that UL26 is 3'-coterminal with four other genes (UL27-UL30). However, further transcript mapping of UL40-UL47 and UL26-UL30 are required to confirm these initial findings. Sets of 3'-coterminal transcripts with alternative promoters are well documented in other regions of HCMV (Adam *et al.*, 1995; Wing and Huang, 1995; Stenberg *et al.*, 1989; Welch *et al.*, 1991). Moreover, the current study identified several other sets of 3'-coterminal genes, including TRL8-TRL12, UL128-UL132, UL2-UL3, UL6-UL8, UL10-UL11, UL16-UL17 and UL18-UL19 (see section 9.1.5).

Two members of the US22 gene family were predicted to be spliced (UL36 and UL28). The splicing in UL36 shown by Kouzarides *et al.* (1988) was confirmed using RT-PCR. There is also good evidence from sequence analysis that UL28 is spliced at its 5'-end (Davison *et al.*, unpublished data). In the current study however, RACE failed to determine the 5'-ends of UL28 and UL29, and RT-PCR failed to detect splicing sites between the two ORFs or of UL28 to an ORF further upstream.

The HCMV US22 gene family has been relatively little studied and the functions it provides are for the most part unknown. pTRS1, pIRS1 and pUL36 are known to be components of the virus tegument, and pIRS1, pUL36 and pUL43 are dispensable (Romanowski *et al.*, 1997; Jones and Muzithras, 1992; Patterson and Shenk, 1999; Dargan *et al.*, 1997). There are homologues for most of the HCMV US22 genes in the β -herpesviruses HHV-6 (Gompels *et al.*, 1995), HHV-7 (Megaw *et al.*, 1998; Nicholas, 1996), MCMV (Rawlinson *et al.*, 1996), rat CMV (Vink *et al.*, 2000) and tupaïid herpesvirus (Bahr and Darai, 2001). Each of these β -herpesviruses encodes a similar number of US22 genes, suggesting that the family provides important functions during virus replication. In support of this view, Hanson *et al.* (2001) reported that the E proteins encoded by two US22 family genes in MCMV, M140 and M141 (US23 and US24 homologues, respectively) appear to act co-operatively and independently to regulate MCMV replication in a cell-type specific manner and are predicted to influence viral pathogenesis. Continued studies with HCMV mutants lacking specific US22 family genes will provide an opportunity to define the function of the gene products.

9.1.5 ORFs in TR_L and at left end of U_L

RNAs were identified for the majority of ORFs in TR_L and the adjacent region in U_L. Most 5'-ends are located 20-35 bp downstream from TATA elements. The single (i.e. not multiple) 5'-ends mapped for transcripts in TR_L and in U_L, except for UL18, are located upstream from the first ATG in the cognate ORF. The 5'-end of UL18 is located downstream from the first ATG codon in the ORF and does not have a potential TATA box upstream, whereas the first ATG codon in the UL18 ORF is preceded by a potential TATA box. Since UL18 is well conserved in CCMV, including the first ATG codon and the

TATA element upstream, it is anticipated that UL18 may specify two transcripts, and that only the abundant, shorter one was detected in this study. Consequently, further transcriptional studies of UL18 are required.

The 5'-ends mapped for certain genes (UL4, UL7, UL10 and UL27) are far upstream of the first ATG codon in each ORF, and intervening sequences contain ATG codons in other reading frames. Consequently, these ORFs are unlikely to be translated efficiently. However, since they are conserved in CCMV, they are likely to be expressed, either from the 5'-ends detected or from weakly expressed smaller transcripts that were not detected. For example, UL20 mRNA has three 5'-ends located 45-319 bp upstream from the first ATG codon, with the furthest upstream being the most abundant.

5'-RACE revealed the existence of six sets of 3'-coterminal transcripts (from TRL8-TRL12, UL2-UL3, UL6-UL8, UL10-UL11, UL16-UL17 and UL18-UL19) in the region analysed. However, the 3'-ends of these transcripts must be mapped to confirm these findings.

Except for UL36, only one gene (UL22A) in the region analysed was found to be spliced. Splice sites in UL22A mapped by Rawlinson and Barrell (1993) and Bresnahan and Shenk (2000) were confirmed using RT-PCR. The overall transcript mapping data indicate that the majority of transcripts in TR_L possess multiple start sites. It is therefore anticipated that TR_L is transcribed in a complex manner.

9.1.6 HCMV ORFs that are transcribed but may not encode functional proteins

Transcripts were detected for six (TRL2, TRL5, TRL6, TRL8, TRL9 and TRL9) of the eight ORFs (TRL2-TRL9) identified by Chee *et al.* (1990) in TR_L and one ORF (UL12) in U_L that are considered by Davison *et al.* (unpublished data) unlikely to encode functional proteins, as they are not conserved in CCMV. Davison *et al.* (unpublished data) discounted ORFs in one genome that lack positional and sequence counterparts in the other, unless they are located in the sequences that represent insertions in relation to flanking genes or

unless bioinformatic or functional data indicated otherwise. Although the HCMV and CCMV genomes exhibit limited regions of nucleotide homology in the TRL2-TRL9 (around 2-8 kbp in AD169), these regions do not align in a way that indicates a common protein-coding capacity (e.g. initiation and termination codon are not conserved, and conserved blocks of sequence are frameshifted with respect to each other).

Two other regions (UL58-UL68 at 90865-98100 and UL106-UL111 at 154950-159799) in the HCMV genome are also unlikely to encode proteins. Neither was analysed in this study. Chambers *et al.* (1999) detected the kinetics of expression of the ORFs identified in these regions (TRL2-TRL9, UL58-UL68 and UL106-UL111) except for UL58 and UL63. However, there is no evidence thus far these transcripts encode functional proteins.

It is notable that TRL2-TRL9, UL58-UL68 and UL106-UL111 regions have a higher AT content than the rest of the genome. One (UL58-UL68) contains *oriLyt* (Huang *et al.*, 1996). A high AT content has also been reported for other viral *ori* sequences, such as simian virus 40 (SV40) (Bergsma *et al.*, 1982), polyoma virus (Soeda *et al.*, 1979), EBV (Yates *et al.*, 1984) and HSV-1 (Stow and McMonagle, 1983; Weller *et al.*, 1985). An *ori* function, however, has not been mapped in either of the other two regions.

Non-protein coding RNAs are not unusual in herpesviruses. The EBER RNAs of EBV and tRNAs in MHV-68 are transcribed by RNA polymerase III (Akusjarvi *et al.*, 1980; Rosa *et al.*, 1981; Bowden *et al.*, 1997). These are smaller than 200 nucleotides in size and are neither polyadenylated nor capped. For HCMV, no such small RNAs have been reported, but a large (5 kb) partially polyadenylated RNA lacking a major ORF has been mapped in the UL106-UL111 region by Plachter *et al.* (1988).

Recently, the HCMV tegument was reported to contain a specific subset of viral transcripts, termed 'virion RNAs', originating from UL22A, UL109, TRL4 (β 2.7) and TRL7 (β 1.2) (Bresnahan and Shenk, 2000; Greijer *et al.*, 2000). It is assumed that delivery of virion RNAs to the host cell allows for their expression immediately after virus entry in the absence of new transcription. UL22A is known to encode a protein that localizes to the Golgi network, and β 2.7 is a major E transcript that probably does not encode functional

protein. It is possible that some of the virion mRNAs serve non-protein-coding functions immediately after infection. Alternatively, they could serve a structural role in organising tegument proteins.

The best parallel to non-protein-coding RNAs in HCMV may be the latency-associated transcript (LATs) of HSV-1. During latent infection in neurons, HSV-1 gene expression is restricted to a family of RNAs known as the LATs (Deatly *et al.*, 1988; Stevens *et al.*, 1987). The LATs are transcribed from a region located within TR_L and IR_L. The full-length 8.3 kb transcript (minor LAT) accumulates in infected neurons, and abundant 2.0 and 1.5 kb introns (major LATs) are processed from the full-length transcript (reviewed in Wagner and Bloom, 1997). The 2.0 kb LAT is circular. Further RNA processing occurs by which 500 bp is removed from the 2.0 kb LAT sequence, resulting in production of the 1.5 kb LAT (Alvira *et al.*, 1999). The 2.0 kb LAT is detected in productive and latent infections, whereas the 1.5 kb LAT is detectable only in latently infected neurons (Wagner *et al.*, 1988). It appears unlikely that functional proteins are encoded by the LATs.

The existence of LATs has been known for over a decade and it has been speculated that they might play a central role in viral latency. Investigation using mouse latency models has shown that the LATs are not required for the establishment or maintenance of a latent infection or for reactivation (Block *et al.*, 1990; Steiner *et al.*, 1989). However, it has been suggested that LAT mutants reactivate less efficiently from latent infection in cultured trigeminal ganglionic neurons (TG) (Block *et al.*, 1993), and that a region downstream from the LAT promoter plays a significant role in enhancing reactivation in a rabbit model (Bloom *et al.*, 1996). Recently, results supporting the view that LATs may facilitate the establishment of latent infections and also increase neuronal survival in the mouse model have been published (Thompson and Sawtell, 2001). However, the mechanism by which the LATs might facilitate establishment of latency is not known, and the role of LATs in reactivation still needs to be proven. LATs have been shown to play a potential role in apoptosis of infected TG in a rabbit model (Perng *et al.*, 2000). This conclusion, however, is difficult to reconcile with the observations from other researchers (Thompson and Sawtell, 2001), who have reported that there is no significance evidence demonstrating that LAT null virus induces more extensive apoptosis in infected neurons.

CMV LATs originate from both DNA strands in the IE1/IE2 region, and are the only viral gene products that have been associated with latent infection. Sense CLTs are oriented in the same direction as productive MIE transcripts, initiating at two unique transcription start sites in the MIE enhancer region (UL125-UL127, another region that is unlikely to encode proteins (Fig. 1.9), and are present in a subset of naturally or experimentally infected CD33⁺ myeloid progenitors (Slobedman and Mocarski, 1999; Hahn *et al.*, 1998; Kondo *et al.*, 1996). Antibodies to the ORFs in CLTs are recognised by sera from healthy CMV seropositive individuals, suggesting that they are expressed during natural infection. However, only one protein (94 amino acid residues) encoded from sense CLTs exhibits a nuclear localization pattern when expressed in mammalian cells. Genetic analysis of ORF has not revealed a role for this gene product in either productive or latent infection (White *et al.*, 2000).

Since little is known about the potential function of HCMV RNAs in the regions that are unlikely to encode proteins, it will be interesting to experimentally investigate whether these RNAs play an important role in HCMV replicative cycle or pathogenesis.

9.2 Conclusion

Determination of the complete sequence of CCMV by Davison *et al.* (unpublished data) facilitated detailed genetic comparisons between HCMV and CCMV and led to a re-evaluation of the genetic content of HCMV. Transcript mapping data shown in this study confirms several of the predictions and extend our knowledge of HCMV gene organisation. Thus the study has achieved several goals. It has improved our understanding the genetic content of HCMV, corrected certain sequencing errors in published sequences, yielded data on the kinetics of expression of US22 gene family members, generated data on the 5'- and 3'-ends of many AD169 genes, mapped the splicing sites in UL128 and UL131A, and experimentally confirmed UL21A, UL128 and UL131A as novel genes in HCMV. The strategy used in this study for transcript mapping could be extended to complete mapping of a wild-type laboratory strain of HCMV when it becomes available. This analysis thus lays the foundation for further experimental studies of gene expression in HCMV.

CHAPTER 10:

References

REFERENCES

- Adam B.L., Jervey T.Y. and Kohler C.P., Wright, G. L. Jr., Nelson, J. A. and Stenberg, R. M. (1995). The human cytomegalovirus UL98 gene transcription unit overlaps with the pp28 true late gene (UL99) and encodes a 58-kilodalton early protein. *Journal of Virology* **69**, 5304-5310.
- Ahn, J.H., Jang, W.J. and Hayward, G.S. (1999). The human cytomegalovirus IE2 and UL112-113 protein accumulate in viral DNA replication compartments that initiate from the periphery of promyocytic leukemia protein associated nuclear bodies (PODs or ND10). *Journal of Virology* **73**, 10458-10471.
- Ahn, K., Angulo, A. Ghazal, P., Peterson, P. A., Yang, Y. and Fruh, K. et al (1996). Human cytomegalovirus inhibits antigen presentation by a sequential multistep process. *Proceedings of the National Academy of Science U.S.A* **93**, 10990-10995.
- Ahn, K., Gruhler, A., Galocha, B, Jones, T. R., Wiertz, E. J., Ploegh, H. L., Peterson, P. A., Yang, Y. and Fruh, K. (1997). The ER-luminal domain of the HCMV glycoprotein US6 inhibits peptide translocation by TAP. *Immunity* **6**, 613-621.
- Akrigg, A., Wilkinson, G. W. and Oram, J. D. (1985). The structure of the major immediate early gene of human cytomegalovirus strain AD169. *Virus Research* **2**, 107-121.
- Akusjarvi, G., Mathews, M. B., Andersson, P., Vennstrom, B. and Pettersson, U. (1980). Structure of genes for virus-associated RNA and RNA of adenovirus type 2. *Proceedings of the National Academy of Science. U.S.A* **77**, 2424-2428.
- Alderete, J. P., Child, S. J. and Geballe, A. P. (2001). T-Abundant early expression of gpUL4 from a human cytomegalovirus mutant lacking a repressive upstream open reading frame. *Journal of Virology* **75**, 7188-7192.
- Alderete, J. P., Jarrahan, S. and Geballe, A. P. (1999). Translational effects of mutations and polymorphisms in a repressive upstream open reading frame of the human cytomegalovirus UL4 gene. *Journal of Virology* **73**, 8330-8337.
- Alford, C. A. and Britt, W. J. (1990). Cytomegalovirus. In: *Fields Virology* pp 1981-2010. Fields, B. N., Knipe, D. M. and Howley, P. M. (eds). Raven Press, New York.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410.
- Alvira, M. R., Goins, W. F., Cohen, J. B. and Glorioso, J. C. (1999). Genetic studies exposing the splicing events involved in herpes simplex virus type 1 latency-associated transcript production during lytic and latent infection. *Journal of Virology* **73**, 3866-3876.
- Anders, D.G. and McCue, L.A. (1996). The human cytomegalovirus genes and proteins required for DNA synthesis. *Intervirology* **39**, 378-388.

- Antonsson, A. and Johansson, P. J. (2001).** Binding of human and animal immunoglobulins to the IgG Fc receptor induced by human cytomegalovirus. *Journal of General Virology* **82**, 1137-1145
- Appereley, J. F. and Goldman, J. M. (1988).** Cytomegalovirus: biology, clinical features and methods of diagnosis. *Bone Marrow Transplantation* **3**, 253-264.
- Arnosti, D. N., Preston, C. M., Hagmann, M., Schanffner, W., Hope, R. G., Laug, G. and Luisi, B. F. (1993).** Specific transcriptional activation in vitro by the herpes simplex virus protein VP16. *Nucleic Acids Research* **21**, 5570-5576.
- Bahr, U. and Darai, G. (2001).** Analysis and characterization of the complete genome of *Tupaia* (tree shrew) herpesvirus. *Journal of Virology* **75**, 4854-4870.
- Bairoch, A. (1988).** Swiss-Prot protein sequence data bank release 8.0. Department de Biochimie Medicale, Centre Mecical Universitaire, Geneva.
- Baldick, C. J. Jr., Marchini, A., Patterson, C. E. and Shenk, T. (1997).** Human cytomegalovirus tegument protein pp71 (ppUL82) enhances the infectivity of viral DNA and accelerates the infectious cycle. *Journal of Virology* **71**, 4400-4408.
- Baldick, C. J. Jr. and Shenk, T. (1996).** Proteins associated with purified human cytomegalovirus particles. *Journal of Virology* **70**, 6097-6105.
- Barracchini, E., Glezer, E., Fish, K., Stenberg, R. M., Nelson, J. A. and Ghazal, P. (1992).** An isoform variant of the cytomegalovirus immediate-early auto repressor functions as a transcriptional activator. *Virology* **188**, 518-529.
- Bastian, F. O., Rabson, A. S., Yee, C. L. and Tralka, T. S. (1972).** Herpesvirus hominis: isolation from human trigeminal ganglion. *Science* **178**, 306-307.
- Batterson, W., Furlong, D. and 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 stage of viral reproductive cycle. *Journal of Virology* **45**, 397-407.
- Battista, M. C., Bergamini, G., Boccuni, M. C., Campanini, F., Ripalti, A. and Landini, M. P. (1999).** Expression and characterization of a novel structural protein of human cytomegalovirus, pUL25. *Journal of Virology* **73**, 3800-3809.
- Baxter, M. K. and Gibson, W. (2001).** Cytomegalovirus basic phosphoprotein (pUL32) binds to capsids in vitor through its amino one-third. *Journal of Virology* **75**, 6865-6873.
- Bear, R., Bankier, A. T., Biggin, M. D., Deininger, P. L., Farrell, P. J., Gibson, T. J., Hatfull, G., Hudson, G. S., Satchwell, S. C., Seguin, C., Tuffnell, P. S. and Barrell, B. G. (1984).** DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature* **310**, 207-211.

- Beck, S. and Barrell, B. (1991).** An HCMV reading frame which has similarity with both the V and C regions of the TCR γ chain. *DNA sequence* **2**, 33-38.
- Becroft, D. M. O. (1981).** Prenatal cytomegalovirus infection: epidemiology, pathology, and pathogenesis. *In: Perspective in Pediatric Pathology* pp 203-241. Rosenberg, H. S. and Berstein, J. (eds).
- Beisser, P. S., Laurent, L., Virelizier, J. L. and Michelson, S. (2001).** Human cytomegalovirus chemokine receptor gene US28 is transcribed in latently infected THP-1 monocytes. *Journal of Virology* **75**, 5949-5957.
- Benedict, C. A., Butrovich, K. D., Lurain, N. S., Corbeil, J., Rooney, I., Schneider, P., Tschopp, J. and Ware, C. F. (1999).** Cutting edge: A novel viral TNF receptor superfamily member in virulent strains of human cytomegalovirus. *Journal of Immunology* **162**, 6967-6970.
- Benton, M. J. (1990).** *Vertebrate Palaeontology*. HarperCollins Academic, London.
- Bergamini, G., Reschke, M., Battista, M.C., Bocconi, M. C., Campanini, F., Ripalti, A. and Landini, M. P. (1998).** The major open reading frame of the β 2.7 transcript of human cytomegalovirus: In vitro expression of a protein posttranscriptionally regulated by the 5' region. *Journal of Virology* **72**, 8425-8429.
- Bergsma, D. J., Olive, D. M., Hartzell, S. W. and Subramanian, K. N. (1982).** Territorial limits and functional anatomy of the simian virus 40 replication origin. *Proceedings of the National Academy of Science U.S.A.* **79**, 381-385.
- Bhella, D., Rixon, F. J. and Dargan, D. J. (2000).** Cryomicroscopy of human cytomegalovirus virions reveals more densely packed genomic DNA than in herpes simplex virus type 1. *Journal of Molecular Biology* **295**, 155-161.
- Biegelke, B. J. (1999).** Human cytomegalovirus US3 gene expression is regulated by a complex network of positive and negative regulators. *Virology* **261**, 155-164.
- Blanton, R. A. and Tevethia, M. J. (1981).** Immunoprecipitation of virus-specific immediate-early and early polypeptides from cells lytically infected with human cytomegalovirus strain AD169. *Virology* **112**, 262-273.
- Block, M. J., Goossens, V. J., Vanherle, S. J. V., Tacken, N., Middeldorp, J. M., Christiaans, M. H. L., van Hooff, J. P. and Bruggeman, C. A. (1998).** Diagnostic value of monitoring human cytomegalovirus late pp67 mRNA expression in renal-allograft recipients by nucleic acid sequence-based amplification. *Journal of Clinical Microbiology* **36**, 1341-1346.
- Block, T. M., Deshmane, S., Masonis, J., Maggioncalda, J., Valyi-Nagi, T. and Fraser, N. W. (1993).** An HSV LAT null mutant reactivates slowly from latent infection and makes small plaques on CV-1 monolayers. *Virology* **192**, 618-630.

Block, T. M., Spivack, J. G., Steiner, I., Deshmane, S., McIntosh, M. T., Lirette, R. P. and Fraser, N. W. (1990). A herpes simplex virus type 1 latency-associated transcript mutant reactivates with normal kinetics from latent infection. *Journal of Virology* **64**, 3417-3426.

Bloom, D. C., Devi-Rao, G. B., Hill, J. M., Stevens, J. G. and Wagner, E. K. (1994). Molecular analysis of herpes simplex virus type 1 during epinephrine-induced reactivation of latently infected rabbits in vivo. *Journal of Virology* **68**, 1283-1292.

Bloom, D. C., Hill, JM, Devi-Rao, G., Wagner, E. K., Feldman, L. T. and Stevens, J. G. (1996). A 348-base-pair region in the latency-associated transcript facilitates herpes simplex virus type 1 reactivation. *Journal of Virology* **70**, 2449-2459.

Bodaghi, B., Jones, T.R. and Zipeto, D. Vita, C., Sun, L., Laurent, L., Arenzana-Seisdedos, F., Virelizier, J. L. and Michelson, S. (1998). Chemokine sequestration by viral chemoreceptors as a novel viral escape strategy: Withdrawal of chemokines from the environment of cytomegalovirus-infected cells. *Journal of Experimental Medicine* **188**, 855-866.

Boppana, S. B., Pass, R. F., Britt, W. J., Stango, S. and Alford, C. A. (1992). Symptomatic congenital cytomegalovirus infection: neonatal morbidity and mortality. *Pediatric Infectious Disease Journal* **11**, 93-99.

Borson, N. D., Sato, W. L. and Drewes, L. R. (1992). A lock-docking oligo (dT) primer for 5' and 3' RACE PCR. *PCR methods application* **2**, 144-148.

Borst, E. M., Hahn, G., Koszinowski, U. H. and Messerle, M. (1999). Cloning of the human cytomegalovirus (HCMV) genome as an infectious bacterial artificial chromosome in escherichia coli: a new approach for construction of HCMV mutants. *Journal of Virology* **73**, 8320-8329.

Borst, E. M., Mathys, S., Wagner, M., Muranyi, W. and Messerle, M. (2001). Genetic evidence of an essential role for cytomegalovirus small capsid protein in viral growth. *Journal of Virology* **75**, 1450-1458.

Bowden, R. J., Simas, J. P., Davis, A. J. and Efstathiou, S. (1997). Murine gammaherpesvirus 68 encodes tRNA-like sequences which are expressed during latency. *Journal of General Virology* **78**, 1675-1687.

Boyle, K. A. and Compton, T. (1998). Receptor-binding properties of a soluble form of human cytomegalovirus glycoprotein B. *Journal of Virology* **72**, 1826-1833.

Breathnach, R. and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. *Annual reviews Biochemistry* **50**, 349-383.

Bresnahan, W. A., Hultman, G. E. and Shenk, T. (2000). Replication of wild-type and mutant human cytomegalovirus in life-extended human diploid fibroblasts. *Journal of Virology* **74**, 10816-10818.

- Bresnahan, W. A. and Shenk, T. (2000).** A subset of viral transcripts packaged within human cytomegalovirus particles. *Science* **288**, 2373-2376.
- Britt, W. (1996).** Human cytomegalovirus overview: the virus and its pathogenicity. *Bulletin of Clinical Infectious Disease* **3**, 307-325.
- Britt, W. and Alford, C. A. (1996).** Cytomegalovirus. In: *Fields Virology* pp 2493-2523. Fields, B. N., Knipe D. M., Howley P. M. *et al.* (eds). Lippencott Williams and Wilkins.
- Britt, W. and Mach, M. (1996).** Human cytomegalovirus glycoproteins. *Intervirology* **39**, 401-412.
- Brown, J. M., Kaneshima, H. and Mocarski, E. S. (1995).** Dramatic interstrain differences in the replication of human cytomegalovirus in SCID-Hu mice. *Journal of Infectious Diseases* **171**, 1599-1603.
- Browne, H., Churcher, M. and Minson, T. (1992).** Construction and characterisation of a human cytomegalovirus mutant with the UL18 (class I homolog) gene deleted. *Journal of Virology* **66**, 6784-6787.
- Buerger, I., Reefschlaeger, J., Bender, W., Eckenberg, P., Popp, A., Weber, O., Graeper, S., Klenk, H. D., Ruebsamen-Waigmann, H. and Hallenberger, S. (2001).** A novel nonnucleoside inhibitor specifically targets cytomegalovirus DNA mutation via the UL89 and UL65 gene products. *Journal of Virology* **75**, 9077-9086.
- Butcher, S. J., Aitken, J., Mitchell, J., Gowen, B. and Dargan, D. J. (1998).** Structure of the human cytomegalovirus B capsid by electron cryomicroscopy and image reconstruction. *Journal of Structural Biology* **124**, 70-76.
- Campadelli-Fiume, G. (1994).** Herpes simplex viruses: Viral glycoproteins. In: *Encyclopedia of Virology* Webster, R. G. and Granoff, A. (eds). Academic Press.
- Cao, J. and Geballe, A. P. (1995).** Translational inhibition by a human cytomegalovirus upstream open reading frame despite inefficient utilization of its AUG codon. *Journal of Virology* **69**, 1030-1036.
- Cao, J. and Geballe, A. P. (1996).** Inhibition of nascent-peptide release at translation termination. *Molecular and Cellular Biology* **16**, 7109-7114.
- Carbon, P., Murgo, S., Ebel, J. P., Krol, A., Tebb, G. and Mattaj, I. W. (1987).** A common octamer motif binding protein is involved in the transcription of U6 snRNA by RNA polymerase III. *Cell* **51**, 71-77.
- Cardin, R. D., Abenes, G. B., Stoddart, C. A. and Mocarski, E. S. (1995).** Murine cytomegalovirus IE2, an activator of gene expression, is dispensable for growth and latency in mice. *Virology* **209**, 236-241.

Caswell, R., Bryant, L. and Sinclair, J. (1996). Human cytomegalovirus immediate-early 2 (IE2) protein can transactivate the human hsp70 promoter by alleviation of Dr1-mediated repression. *Journal of Virology* **70**, 4028-4037.

Caswell R., Hagemeyer, C., Chiou, C. J., Hayward, G., Kouzarides, T. and Sinclair, J. (1993). The human cytomegalovirus 86K immediate early (IE) 2 protein requires the basic region of the TATA-box binding protein (TBP) for binding, and interacts with TBP and transcription factor TFIIB via regions of IE2 required for transcriptional regulation. *Journal of General Virology* **74**, 2691-2698.

Cavanaugh, V. J., Stenberg, R. M., Staley, T. L., Virgin, H. W., MacDonald, M. R., Paetzold, S., Farrell, H. E., Rawlinson, W. D. and Campbell, A. E. (1996). Murine cytomegalovirus with a deletion of genes spanning *HindIII*-J and I displays altered cell and tissue tropism. *Journal of Virology* **70**, 1365-1374.

Cha, T. A., Tom, E., Kemble, G. W., Duke, G. M., Mocarski E. S. and Spaete R. R. (1996). Human cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains. *Journal of Virology* **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. and Ghazal, P. (1999). DNA microarrays of the complex human cytomegalovirus genome: profiling kinetic class with drug sensitivity of viral gene expression. *Journal of Virology* **73**, 5757-5766.

Chan, L., Stefanac, T., Lavalee, J. F., Jin, H., Bedard, J., May, S. and Falardeau, G. (2001). Design and synthesis of new potent human cytomegalovirus (HCMV) inhibitors based on internally hydrogen-bonded 1,6-naphthyridines. *Bioorganic and Medical Chemistry Letters* **11**, 103-105.

Chandler, S. H. and McDougall, J. K. (1986). Comparison of restriction site polymorphisms among clinical isolates and laboratory strains of human cytomegalovirus. *Journal of General Virology* **67**, 2179-2192.

Chang, C. P., Vesole, D. H., Nelson, J., Oldstone, M. B. A. and Stinski, M. F. (1989). Identification and expression of a human cytomegalovirus early glycoprotein. *Journal of Virology* **63**, 3330-3337.

Chapman, T. L. and Bjorkman, P. J. (1998). Characterization of a murine cytomegalovirus class I major histocompatibility complex (MHC) homolog: comparison to MHC molecules and to the human cytomegalovirus MHC homolog. *Journal of Virology* **72**, 460-466.

Chapman, T. L., Heikeman, A. P. and Bjorkman, P. J. (1999). The inhibitory receptor LIR-1 uses a common binding interaction of recognize class I MHC molecules and the viral homologue UL18. *Immunity* **11**, 603-613.

- Chee, M. S., Bankier, A. T., Beck, S., Bohni, R., Brown, C. M., Cerny, R., Horsnell, T., Hutchison, C. A. III, Kouzarides, T., Martignetti, J. A., Preddie, E., Satchwell, S. C., Tomlinson, P., Weston, K. M. and Barrell, B. G. (1990) Analysis of protein-coding content of the sequence of human cytomegalovirus strain AD169. *Current Topics in Microbiology and Immunology*. **154**, 126-169.
- Chen, D. H., Jiang, H., Lee, M., Liu, F. and Zhou, Z. H. (1999) Three-dimensional visualization of tegument/capsid interactions in the intact human cytomegalovirus. *Virology*. **260**, 10-16.
- Chenchik, A., Moqadam, F. and Siebert, P. (1995) Marathon cDNA amplification: A new method for cloning full-length cDNAs. *CLONTECHniques X*. **1**, 5-8.
- Chenchik, A., Moqadam, F. and Siebert, P. (1996) A new method for full-length cDNA cloning by PCR. In: *A laboratory guide to RNA: Isolation, Analysis, and Synthesis*. pp 273-321. Krieg, P. A. (ed). Wiley-Liss, Inc.
- Chenchik, A., Zhu, Y., Diatchenko, L., Li, R., Hill, J. and Siebert, P. (1998) Generation and use of high-quality cDNA from small amounts of total RNA by SMART PCR: RT-PCR Methods for Gene Cloning and Analysis. In: *BioTechniques Books*. pp 305-319. Siebert, P. and Larrick, J. (eds).
- Cherrington, J. M., Khoury, E. L. and Mocarski, E. S. (1991) Human cytomegalovirus *ie2* negatively regulates α gene expression via a short target sequence near the transcription start site. *Journal of Virology*. **65**, 887-896.
- Chomczynski, P. and Sacchi, N. (1987) Single step RNA isolation method. *Annals Biochemistry*. **162**, 156-160.
- Chou, J. and Roizman, B. (1986) The terminal is sequence of the herpes simplex virus genome contains the promoter of a gene located in the repeat sequences of the L component. *Journal of Virology*. **57**, 629-637.
- Colberg-Poley, A. M. (1996) Functional roles of immediate early proteins encoded by the human cytomegalovirus UL36-38, UL115-119, TRS1/IRS1 and US3 loci. *Intervirology*. **39**, 350-360.
- Colberg-Poley, A. M., Huang, L., Soltero, V. E., Iskenderian, A. C. Schumacher, R. F. and Anders, D. G. (1998) The acidic domain of pUL37x1 and gpUL37 plays a key role in transactivation of HCMV DNA replication gene promoter constructions. *Virology*. **246**, 400-408.
- Colberg-Poley, A. M., Patel, M. B., Erez, D. P. P. and Slater, J. E. (2000) Human cytomegalovirus UL37immediate-early regulatory proteins traffic through the secretory apparatus and to mitochondria. *Journal of General Virology*. **81**, 1779-1789.

- Colberg-Poley, A. M., Santomenna, L. D., Harlow, P. P., Benfield, P. A. and Tenney, D. J. (1992). Human cytomegalovirus US3 and UL36-38 immediate-early proteins regulate gene expression. *Journal of Virology* **66**, 95-105.
- Collier, A. C., Meyers, J. D., Corey, L., Murphy, V. L., Roberts, P. L. and Handsfield, H. H. (1987). Cytomegalovirus infection in homosexual men. Relationship to sexual practices, antibody to human immunodeficiency virus, and cell-mediated immunity. *American Journal of Medicine* **23**, 593-601.
- Comps, M. and Cochenec, N. (1993). A herpes-like virus from the European oyster *ostreaedulis* L. *Journal of Invertebrate Pathology* **62**, 201-203.
- Costa, R. H., Draper, K. G., Kelly, T. J. and Wagner, E. K. (1985). An unusual spliced herpes simplex virus type 1 transcript with sequence homology to Epstein-Barr virus DNA. *Journal of Virology* **54**, 317-328.
- Dal Monte, P., Varani, S., Lazzarotto, T., Pignatelli, S. and Landini, M. P. (1998). Prokaryotic expression of human cytomegalovirus pUS22 and its reactivity with human antibody. *Archives of Virology* **143**, 2413-2419.
- Dambaugh, T. and Hennessy, K. (1986). In: *Epstein-Barr Virus* pp1. Epstein, M. A., Achong, B. G. (eds). Heineman Medical, London.
- Dargan, D. J., Jamieson, F. E., Maclean, J., Dolan, A., Addison, C. and McGeoch, D. J. (1997). The published DNA sequence of human cytomegalovirus strain AD169 lacks 929 base pairs affecting genes UL42 and UL43. *Journal of Virology* **71**, 9833-9836.
- Davis, M. G., Mar, E. C., Wu, Y. M. and Huang, E. S. (1984). Mapping and expression of a human cytomegalovirus major viral protein. *Journal of Virology* **52**, 129-135.
- Davison, A. J. (1992). Channel Catfish virus: A new type of herpesvirus. *Virology* **186**, 9-14.
- Davison, A. J. (1993). Herpesvirus Genes. *Reviews in Medical Virology* **3**, 237-244.
- Davison, A. J. (1998). The genome of salmonid herpesvirus 1 *Journal of Virology* **72**, 1974-1982.
- Davison, A. J. (2002). Evolution of the herpesviruses. *Journal of Veterinary Microbiology* (In press).
- Davison, A. J. and Clements, J.B. (1996). Herpesviruses: general properties. In: *Topley and Wilsons Microbiology and Microbiol Infections* pp 309-323. Mahy B. W. J. and Collier, L. (eds). Oxford University Press, Oxford.
- Davison, A. J. and McGeoch, D. J. (1986). Evolutionary comparisons of the S segments in the genomes of herpes simplex virus type 1 and varicella-zoster virus. *Journal of General Virology* **67**, 597-611.

- Davison, A. J., Sauerbier, E., Dolan, A., Addison, C., McKinnell, R. G. (1999).** Genomic studies of the Lucke tumor herpesvirus (RaHV-1). *Journal of Cancer Research and Clinical Oncology* **125**, 232-238.
- Davison, A. J. and Scott, J. E. (1986).** The complete DNA sequence of varicella-zoster virus. *Journal of General Virology* **67**, 1759-1816.
- Davison, A. J. and Taylor, P. (1987).** Genetic relations between varicella-zoster virus and Epstein-Barr virus. *Journal of General Virology* **68**, 1067-1079.
- Dealty, A. M., Spivack, J. G., Lavi, E., O'Boyle, D. R. and Fraser, N. W. (1988).** Latent herpes simplex virus type 1 transcripts in peripheral and central nervous system tissue of mice map to similar regions of the viral genome. *Journal of Virology* **62**, 749-756.
- 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.
- Depto, A. S. and Stenberg, R. M. (1989).** Regulated expression of the human cytomegalovirus pp65 gene: octamer sequence in the promoter is required for activation by viral gene products. *Journal of Virology* **63**, 1232-1238.
- Dieterich, D. T. and Rehmin, M. (1991).** Cytomegalovirus colitis in AIDS: presentation in 44 patients and a review of the literature. *Journal of Immunodeficiency Virus* **4**, S29-S35.
- Drew, W. L., Sweet, E. S., Miner, R. C. and Mocarski, E. S. (1984).** Multiple infections by cytomegalovirus in patients with acquired immunodeficiency syndrome: documentation by Southern blot hybridization. *Journal of Infectious Diseases* **150**, 952-953.
- Efstathiou, S., Lawrence, G. L., Brown, C. M. and Barrell, B. G. (1992).** Identification of homologues to the human cytomegalovirus US22 gene family in human herpesvirus 6. *Journal of General Virology* **73**, 1661-1671.
- Efstathiou, S., Minson, A. C., Field, H. J. and Barrell, B. G. (1986).** Detection of herpes simplex virus-specific DNA sequences in latently infected mice and in humans. *Journal of Virology* **57**, 446-455.
- Elliott, R. M., Crook, N. E., Desselberger, U., Hull, R. and McGeoch, D. J. (1991).** Some highlights of virus research in 1990. *Journal of General Virology* **72**, 1761-1779.
- Farrell, P. J., Deininger, P., Bankier, A. and Barrell, B. (1983).** Homologous upstream sequences near Epstein-Barr virus promoters. *Proceedings of the National Academy of Science U.S.A.* **80**, 1565-1569.
- Fleckenstein, B., Muller, I. And Collins, J. (1982).** Cloning of the complete human cytomegalovirus genome in cosmids. *Gene* **18**, 39-46.
- Flint, S. J., Enquist, L. W., Krug, R. M., Racaniello, V. R. and Skalka, A. M. (2000).** *In: Principles of Virology. Molecular Biology, Pathogenesis and Control.* American Society for Microbiology, Washington D. C.

- Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T. and Tumer, D. H. (1986). Improved free-energy parameters for predictions of RNA duplex stability. *Proceedings of the National Academy of Science U.S.A.* **83**, 9373-9377.
- Gallant, J. E., Moore, R. D., Richman, D. D., Keruly, J. and Chaisson, R. E. (1992). Incidence and natural history of cytomegalovirus disease in patients with advanced human immunodeficiency virus disease treated with zidovudine. *Journal of Infectious Diseases* **166**, 1223-1227.
- Gallina, A., Simoncini, L., Garbelli, S., Percivalle, E., Pedrali-Noy, G., Lee, K. S., Erikson, R. L., Plachter, B., Gerna, G. and Milanese, G. (1999). Polo-like kinase 1 as a target for human cytomegalovirus pp65 lower matrix protein. *Journal of Virology* **73**, 1468-1478.
- Gao, J. L. and Murphy, P. M. (1994). Human cytomegalovirus open reading frame US28 encodes a functional β chemokine receptor. *Journal of Biology and Chemistry* **269**, 28539-28542.
- Geballe, A. P., Leach, F. S. and Mocarski, E. S. (1986). Regulation of cytomegalovirus late gene expression: γ genes are controlled by posttranscriptional events. *Journal of Virology* **57**, 864-874.
- Geballe, A. P. and Mocarski, E. S. (1988). Translational control of cytomegalovirus gene expression is mediated by upstream AUG codons. *Journal of Virology* **62**, 3334-3340.
- Gebert, S., Schmolke, S., Sorg, G., Floss, S., Plachter, B. and Stamminger, T. (1997). The UL84 protein of human cytomegalovirus acts as a transdominant inhibitor of immediate-early-mediated transactivation that is able to prevent viral replication. *Journal of Virology* **71**, 7048-7060.
- Gelb, L. D., Adams, S. G. and Dohner, D. E. (1990). Differentiation between the oka varicella vaccine virus and American wild-type varicella-zoster virus (VZV). *Advances in Experimental and Medical Biology* **278**, 59-69.
- George, D. G. Barker, W. C. and Hunt, L. T. (1986). The protein identification source (PIR). *Nucleic Acids Research* **14**, 11-15.
- Ghazal, P., Lubon, H., Fleckenstein, B. and Hennighausen, L. (1987). Binding of transcription factors and creation of a large nucleoprotein complex on the human cytomegalovirus enhancer. *Proceedings of the National Academy of Science. USA* **84**, 3658-3662.
- Gibson, W. (1983). Protein counterparts of human and simian cytomegalovirus. *Virology* **128**, 391-406.
- Gibson, W. (1996). Structure and assembly of the virion. *Intervirology* **39**, 389-400.

- Gibson, W., Baxter, M. K. and Clopper, K. S. (1996).** S-Cytomegalovirus "missing" capsid protein identified as heat-aggregable product of human cytomegalovirus UL46. *Journal of Virology* **70**, 7454-7461.
- Gibson, W. and Irmieri, A. (1983).** Selection of particles and proteins for use as human cytomegalovirus subunit vaccines. *Birth Defects* **20**, 305-324.
- Gold, D., Bowden, R., Sixbey, J., Riggs, R., Katon, W. J., Ashley, R., Obrigewitch, R. M. and Corey, L. (1990).** Chronic fatigue. A prospective clinical and virologic study. *Journal of the American Medical Association* **264**, 48-53.
- Goldmacher, V. S, Bartle, L. M., Skaletskaya, A., Dionne, C. A., Kedersha, N. L., Vater, C. A., Han, J. W., Lutz, R. J., Watanabe, S., McFarland, E. D., Kieff, E. D., Mocarski, E. S. and Chittenden, T. (1999).** A cytomegalovirus-encoded mitochondria-localized inhibitor of apoptosis structurally unrelated to bcl-2. *Proceedings of the National Academy of Science. USA* **96**, 12536-12541.
- Gompels, U. A., Nicholas, J., Lawrence, G. Jones, M., Thomson, B. J., Martin, M. E. D., Efstathiou, S., Craxtons, M. and Macaulay, H. A. (1995).** The DNA sequence of human herpesvirus-6 structure, coding content, and genome evolution. *Virology* **209**, 29-51.
- Greenway, P. J., Oram, J. D., Downing, R. G. and Patel, K. (1982).** Human cytomegalovirus DNA: BamHI, EcoRI and PstI restriction endonuclease cleavage maps. *Gene* **18**, 355-360.
- Greenaway, P. J. and Wilkinson, G. W. G. (1987).** Nucleotide sequence of the most abundantly transcribed early gene of human cytomegalovirus strain AD169. *Virus Research* **7**, 17-31.
- Greijer, A. E., Dekkers, C. A. and Middeldorp, J. M. (2000).** Human cytomegalovirus virions differentially incorporate viral and host cell RNA during the assembly process. *Journal of Virology* **74**, 9078-9082.
- Greis, K. D., Gibson, W. and Hart, G. W. (1994).** Site-specific glycosylation of the human cytomegalovirus tegument basic phosphoprotein (UL32) at serine 921 and serine 952. *Journal of Virology* **68**, 8339-8349.
- Haarr, L. and Skulstad, S. (1994).** The herpes simplex virus type 1 particle: structure and molecular functions. *APMIS* **102**, 321-346.
- Hagemeier, C., Walker, S., Caswell, R., Kouzarides, T. and Sinclair, J. (1992).** The human cytomegalovirus 80-kilodalton but not the 72-kilodalton immediate-early protein transactivates heterologous promoters in a TATA box-dependent mechanism and interacts directly with TFIID. *Journal of Virology* **66**, 4452-4456.
- Hall, A. and Brown, R. (1985).** Human N-ras: cDNA cloning and gene structure. *Nucleic Acids Research* **13**, 5255-5268.

Handsfield, H. H., Chandler, S. H., Caine, V. A., Meyers, J. D., Corey, L., Medeiros, E. and McDougall, J. K. (1985). Cytomegalovirus infection in sex partners. Evidence for sexual transmission. *Journal of Infectious Diseases* **151**, 128-130.

Hanson, L. K., Dalton, B. L., Karabekian, Z., Ciocco-Schmitt, G. and Campbell, A. E. (2001). Products of US22 genes M140 and M141 confer efficient replication of murine cytomegalovirus in macrophages and spleen. *Journal of Virology* **75**, 6292-6302.

Hanson, L. K., Dalton, B. L., Karabekian, Z., Farrell, H. E., Rawlinson, W. D., Stenberg, R. M. and Campbell, A. E. (1999). Transcriptional analysis of the Murine cytomegalovirus HindIII-I region: Identification of a novel immediate-early gene region. *Virology* **260**, 156-164.

Harper, D. R. and Kinchington, P. R. (1998). *Molecular Virology*. BIOS Scientific Publishers Limited.

Hayashi, M. L., Blankenship, C. and Shenk, T. (2000). Human cytomegalovirus UL69 protein is required for efficient accumulation of infected cells in the G1 phase of the cell cycle. *Proceedings of the National Academy of Science. USA* **97**, 2692-2696.

Hayward, G. S., Ambinder, R., Ciuffo, D., Hayward, S. D. and LaFemina, R. L. (1984). Structural organization of human herpesvirus DNA molecules. *Journal of the Investigative Dermatology* **83**, 29-41.

Hensel, G., Meyer, H. H., Buchmann, I., Pommerehne, D., Schmolke, S., Plachter, B., Radsak, K. and Kern, H. F. (1996). Intracellular localization and expression of the human cytomegalovirus matrix phosphoprotein pp71 (ppUL82): evidence for its translocation into the nucleus. *Journal of General Virology* **77**, 3087-3097.

Hensel, G., Meyer, H. H., Gartner, S., Brand, G. and Kern, H. F. (1995). Nuclear localization of the human cytomegalovirus tegument protein pp150 (ppUL32). *Journal of General Virology* **76**, 1591-1601.

Hermiston, T. W., Malone, C. L., Witte, P. R. and Stinski, M. F. (1987). Identification and characterization of the human cytomegalovirus immediate-early region 2 gene that stimulates gene expression from an inducible promoter. *Journal of Virology* **61**, 3214-3221.

Ho, M., Suwansirikul, S., Dowling, J. N., Youngblood, L. A. and Armstrong, J. A. (1975). The transplanted kidney as a source of cytomegalovirus infections. *New England Journal of Medicine* **293**, 1109-1112.

Hobom, U., Brune, W., Messerle, M., Hahn, G. and Koszinowski, U. H. (2000). Fast screening procedures for random transposon libraries of cloned herpesvirus genomes: mutational analysis of human cytomegalovirus envelope glycoprotein genes. *Journal of Virology* **74**, 7720-7729.

Homa, F. L. and Brown, J. C. (1997). Capsid assembly and DNA packaging in herpes simplex virus. *Reviews in Medical Virology* **7**, 107-122.

- Homer, E. G., Rinaldi, A., Nicholl, M. J. and Preston, C. M. (1999).** Activation of herpesvirus gene expression by the human cytomegalovirus protein pp71. *Journal of Virology* **73**, 8512-8518.
- Honess, R. W. (1984).** Herpes simplex and "The herpes complex" Diverse observations and a unifying hypothesis. *Journal of General Virology* **65**, 2077-2107.
- Honess, R. W. and Roizman, B. (1975).** Regulation of herpesvirus macromolecular synthesis. I. Cascade regulation of the synthesis of three groups of viral proteins. *Journal of Virology* **14**, 8-19.
- Horwitz, C. A., Henle, W., Henle, G., Snover, D., Rudnick, H., Balfour, H. H. Jr., Mazur, M. H., Watson, R., Schwartz, B. and Muller, N. (1986).** Clinical and laboratory evaluation of cytomegalovirus-induced mononucleosis in previously healthy individuals. Report of 82 cases. *Medicine* **65**, 124-134.
- Huang, L. L., Zhu, Y. and Anders, D. G. (1996).** The variable 3'-ends of a human cytomegalovirus *oryLyt* transcript (SRT) overlap an essential, conserved replicator element. *Journal of Virology* **70**, 5272-5281.
- Huber, M. T. and Compton, T. (1998).** The human cytomegalovirus UL74 gene encodes the third component of the glycoprotein H-glycoprotein L-containing envelope complex. *Journal of Virology* **72**, 8191-8197.
- Hutchinson, N. I., Sondermeyer, R. T. and Tocci, M. J. (1986).** Organisation and expression of the major genes from the long inverted repeat of the human cytomegalovirus genome. *Virology* **155**, 160-171.
- Irmiere, A. and Gibson, W. (1983).** Isolation and characterization of a noninfectious virion-like particle released from cells infected with human strains of cytomegalovirus. *Virology* **130**, 118-133.
- Iskederian, A. C., Huang, L., Reilly, A., Stenberg, R. M. and Anders, D. G. (1996).** Four of eleven loci required for transient complementation of human cytomegalovirus DNA replication cooperate to activate expression of replication genes. *Journal of Virology* **70**, 383-392.
- Iwayama, S., Yamamoto, T. and Furuya, T. (1994).** Intracellular localization and DNA-binding activity of a class of viral early phosphoproteins in human fibroblasts infected with human cytomegalovirus (Towne strain). *Journal of General Virology* **75**, 3309-3318.
- Jabs, D. A., Enger, C. and Bartlett, J. G. (1989).** Cytomegalovirus retinitis and acquired immunodeficiency syndrome. *Archives Ophthalmology* **107**, 75-80.
- Jacob, R. J. Morse, L. S. and 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. *Journal of Virology* **29**, 442-457.

- Jacobson, M. A. and Mills, J. (1986).** Serious cytomegalovirus disease in the acquired immunodeficiency syndrome (AIDS). *Annals of Internal Medicine* **108**, 585-594.
- Jacobson, M. A. and Mills, J. (1988).** Cytomegalovirus infection. *Clinical Chest Medicine* **9**, 443-448.
- Jahn, G., Knust, E., Schmolla, H., Sarre, T., Nelson, J. A., McDougall, J. K. and Fleckenstein, B. (1984).** Predominant immediate-early transcripts of human cytomegalovirus AD169. *Journal of Virology* **49**, 363-370.
- Jahn, G., Scholl, B. C., Traupnee, B. and Fleckenstein, B. (1987).** The two major structural phosphoproteins (pp65 and pp150) of human cytomegalovirus and their antigenic properties. *Journal of General Virology* **68**, 1358-1367.
- Jenkins, D. E. and Mocarski, E. S. (1994).** Human cytomegalovirus late protein encoded by *ie2*: A trans-activator as well as a repressor of gene expression. *Journal of General Virology* **75**, 2337-2348.
- Jones, T. R. and Muzithras, V. P. (1991).** Fine mapping of transcripts expressed from the US6 gene family of human cytomegalovirus strain AD169. *Journal of Virology* **65**, 2024-2035.
- Jones, T. R. and Muzithras, V. P. (1992).** A cluster of dispensable genes within the human cytomegalovirus genome short component: IRS1, US1 through US5, and the US6 family. *Journal of Virology* **66**, 2541-2546.
- Jones, T. R., Wiertz, E. J., Sun, L., Fish, K. N., Nelson, J. A. and Ploegh, H. L. (1996).** Human cytomegalovirus US3 impairs transport and maturation of major histocompatibility complex class I heavy chains. *Proceedings of the National Academy of Science. USA* **93**, 11327-11333.
- Kari, B., Li W., Cooper, J., et al (1994).** The human cytomegalovirus UL100 gene encodes the gc-II glycoproteins recognised by group 2 monoclonal antibodies. *Journal of General Virology* **75**, 3081-3086.
- Kaye, J., Browne, H., Stoffel, M. and Minson, A. (1992a).** The UL16 gene of human cytomegalovirus encodes a glycoprotein that is dispensable for growth *in vitro*. *Journal of Virology* **66**, 6609-6615.
- Kaye, J., Gompels, U. and Minson, A. (1992b).** Glycoprotein H of human cytomegalovirus (HCMV) forms a stable complex with the HCMV UL115 gene product. *Journal of General Virology* **73**, 2693-2698.
- Keay, S. and Baldwin, B. (1992)** The human fibroblast receptor for gp86 of human cytomegalovirus is a phosphorylated glycoprotein. *Journal of Virology* **66**, 4834-4838.
- Kieff, E. and Liebowitz, D. (1990).** Epstein-Barr virus and its replication. *In: Virology*. pp 1889-1921. Fields, B. N., Knipe, D. M., Chanock, R. M., Hirsch, M. S., Melnick, J. L., Monath, T. P. and Roizman, B. (eds). Raven Press New York.

- Klages, S. Ruger, B. and Jahn, G. (1989)** Multiplicity dependent expression of the predominant phosphoprotein pp65 of human cytomegalovirus. *Virus Research* **12**, 159-168.
- Kledal, T. N., Rosenkilde, M. M. and Schwartz, T. W. (1998)**. Selective recognition of the membrane-bound CX3C chemokine, fractalkine, by the human cytomegalovirus-encoded broad-spectrum receptor US28. *FEBS Letters* **441**, 209-214.
- Klucher, K. M., Sommer, M., Kadonaga, J. T. and Spector, D. H. (1993)**. In vivo and in vitro analysis of transcriptional activation mediated by the human cytomegalovirus major immediate-early proteins. *Molecular Cell Biology* **13**, 1238-1250.
- Kondo, K. Kaneshima, H. and Mocarski, E. S. (1994)**. Human cytomegalovirus latent gene infection of granulocyte-macrophage progenitors. *Proceedings of the National Academy of Science. USA* **91**, 11879-11883.
- Kotenko, S. V., Saccani, S., Izotova, L. S., Mirochintchenko, O. V. and Pestka, S. (2000)**. Human cytomegalovirus harbors its own unique IL-10 homolog (cmvIL-10). *PNAS* **97**, 1695-1700.
- Kouzarides, T., Bankier, A. T., Satchwell, S. C., Preddy, E. and Barrell, B. G. (1988)**. An immediate early gene of human cytomegalovirus encodes a potential membrane glycoprotein. *Virology* **165**, 151-164.
- Kouzarides, T., Bankier, A. T., Satchwell, S. C., Weston, K., Tomlinson, P. and Barrell, B. G. (1987)**. Large-scale rearrangement of homologous regions in the genomes of HCMV and EBV. *Virology* **157**, 397-413.
- Kozak, M. (1984)**. Compilation and analysis of sequences upstream from the translation start site in eukaryotic mRNAs. *Nucleic Acids Research* **12**, 857-872.
- Kozak, M. (1996)**. Interpreting cDNA sequences: some insights from studies on translations. *Mammalian Genetics* **7**, 563-574.
- Krech, U. (1973)**. Complement-fixing antibodies against cytomegalovirus in different parts of the world. *Bulletin of the World Health Organisation* **49**, 103-106.
- Kuhn, D. E., Beall, C. J. and Kolattukudy, P. E. (1995)**. The cytomegalovirus US28 protein binds multiple CC chemokine with high affinity. *Biochemistry and Biophysics research community* **211**, 325-330.
- Kumar, S. and Hedges, S. B. (1998)**. A molecular time scale for vertebrate evolution. *Nature* **392**, 917-920.
- Kyte, I. And Doolittle, R. F. (1982)**. A simple method for displaying the hydrophobic character of a protein. *Journal of molecular Biology* **157**, 105-132.
- LaPierre, L. A. and Biegelke, B. J. (2001)**. Identification of a novel transcriptional repressor encoded by human cytomegalovirus. *Journal of Virology* **75**, 6062-6069.

- Leatham, M. P., Witte, P. R. and 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. *Journal of Virology* **65**, 6144-6153.
- Lembo, D., Gribaudo, G., Hofer, A., Riera, L., Cornaglia, M., Mondo, A., Angeretti, A., Gariglio, M., Thelander, L. and Landolfo, S. (2000).** Expression of an altered ribonucleotide reductase activity associated with the replication of murine cytomegalovirus in quiescent fibroblasts. *Journal of Virology* **74**, 11557-11565.
- Leong, C. C., Chapman, T. L., Bjorlman, P. J. (1998).** Modulation of natural killer cell cytotoxicity in human cytomegalovirus infection: The role of endogenous class I major histocompatibility complex and a viral class I homolog. *Journal of Experimental Medicine* **187**, 1681-1687.
- Levine, A. (1992).** *Viruses*. Scientific American Library, New York.
- Lewin, B. (1994a).** Eukaryotic transcription and RNA processing. *In: Genes V* pp 845-877. Oxford University Press Inc., New York, U. S. A.
- Lewin, B. (1994b).** Control of prokaryotic gene expression. *In: Genes V* pp 374-410. Oxford University Press Inc., New York, U. S. A.
- Levy, J. (1997).** Three new human herpesviruses (HHV-6, 7 and 8). *Lancet* **349**, 558-563.
- Lilley, B. N., Ploegh, H. L. and Tirabassi, R. S. (2001).** Human cytomegalovirus open reading frame TR11/IR11 encodes an immunoglobulin G Fc-binding protein. *Journal of Virology* **75**, 11218-11221.
- Lindahl, T., Adams, A., Bjursell, G., Bornkamm, G. W., Kaschka-Dierich, C. and Jehn, U. (1976).** Covalently closed circular duplex DNA of Epstein-Barr virus in a human lymphoid cell line. *Journal of Molecular Biology* **102**, 511-530.
- Lindahl, T., Klein, G., Reedman, B. N., Johansson, B. and Singh, S. (1974).** Relationship between Epstein-Barr virus (EBV) DNA and the EBV-determined nuclear antigen (EBNA) in Burkitt lymphoma biopsies and other lymphoproliferative malignancies. *International Journal of Cancer* **13**, 764.
- Liu, B. and Stinski, M. F. (1992).** Human cytomegalovirus contains a tegument protein that enhances transcription from promoters with upstream ATF and AP-1 cis-acting elements. *Journal of Virology* **66**, 4434-4444.
- Lu, M. and Shenk, T. (1999).** Human cytomegalovirus UL69 protein induces cells to accumulate in G1 phase of the cell cycle. *Journal of Virology* **73**, 676-683.
- Lunetta, J. M. and Wiedeman, J. A. (2000).** Latency-associated sense transcripts are expressed during in vitro human cytomegalovirus productive infection. *Virology* **278**, 467-476.

- Lurain, N. S., Kapell, K. S., Huang, D. D., Short, J. A., Paintsil, J., Winkfield, E., Benedict, C. A., Ware, C. F. and Bremer, J. W. (1999). Human cytomegalovirus UL144 open reading frame: Sequence hypervariability in low-passage clinical isolates. *Journal of Virology* **73**, 10040-10050.
- Mach, M., Kropff, B., DalMonte, P. and Britt, W. (2000). Complex formation by human cytomegalovirus glycoproteins M (gpUL100) and N (gpUL73). *Journal of Virology* **74**, 11881-11892.
- MacLean, C. A., Robertson, L. M. and Jamieson, F. E. (1993). Characterization of the UL10 gene product of herpes simplex virus type 1 and investigation of its role in vivo. *Journal of General Virology* **74**, 975-983.
- Maidji, E., Tuguzov, S., Abenes, G. Jones, T. and Pereira, L. (1998). A novel human cytomegalovirus glycoprotein, gpUS9, which promotes cell-to-cell spread in polarized epithelial cells, colocalizes with the cytoskeletal proteins E-cadherin and F-actin. *Journal of Virology* **72**, 5717-5727.
- Margulies, B. J., Browne, H. and 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**, 111-125.
- McGeoch, D. J. (1985). On the predictive recognition of signal peptide sequences. *Virus Research* **3**, 271-286.
- McGeoch, D. J. (1989). The genomes of the human herpesviruses: contents, relationships and evolution. *Annual Reviews Microbiology* **43**, 235-265.
- McGeoch, D. J. (1990). Protein sequence comparisons show that the 'pseudoproteases' encoded by poxviruses and certain retroviruses belong to the deoxyuridine triphosphatase family. *Nucleic Acids Research* **18**, 4105-4110.
- McGeoch, D. J. and Cook, S. (1994). Molecular phylogeny of the alphaherpesvirinae subfamily and a proposed evolutionary timescale. *Journal of Molecular Biology* **283**, 9-22.
- McGeoch, D. J. and Davison, A. J. (1999a). The molecular evolutionary history of the herpesviruses. In: *Origin and evolution of viruses* pp. 441-465. Domingo, E., Webster, R. and Holland, J. (eds). Academic press. London, England.
- McGeoch, D. J. and Davison, A. J. (1999b). The descent of human herpesvirus 8. *Seminars in Cancer Biology* **9**, 201-209.
- McGeoch, D. J., Dalrymple, M. A., Davison, A. J., Dolan, A., Frame, M. C., McNab, D., Perry, L. J., Scott, J. E. and Taylor, P. (1988). The complete DNA sequence of the long unique region in the genome of herpes simplex virus type 1. *Journal of General Virology* **69**, 1531-1574.

- McGeoch, D. J., Dolan, A. and Ralph, A. C. (2000)** Toward a comprehensive phylogeny for mammalian and avian herpesviruses. *Journal of Virology*. **74**, 10401-10406.
- McGregor, A. and Schliess, M. R. (2001)** Molecular cloning of the guinea pig cytomegalovirus (GPCMV) genomes as an infectious bacterial artificial chromosome (BAC) in *Escherichia coli*. *Molecular Genetics and Metabolism*. **72**, 15-26.
- McVoy, M. A., Nixon, D. E. and Adler, S. P. (1997)** Circularization and cleavage of guinea pig cytomegalovirus genomes. *Journal of Virology*. **71**, 4209-4217.
- Megaw, A. G., Rapaport, D., Avidor, B., Frenkel, N. and Davison, A. J. (1998)** The DNA sequence of the RK strain of human herpesvirus 7. *Virology*. **244**, 119-132.
- Meier, J. L. and Pruessner, J. A. (2000)** The human cytomegalovirus major immediate-early distal enhancer region is required for efficient viral replication and immediate-early gene expression. *Journal of Virology*. **74**, 1602-1613.
- Meier, J. L. and Stinski, M. F. (1997)** Regulation of human cytomegalovirus immediate-early gene expression. *Intervirology*. **39**, 331-342.
- Mellerick, D. M. and Fraser, N. W. (1987)** Physical state of the latent herpes simplex virus genome in mouse model system. Evidence suggesting an episomal state. *Virology*. **158**, 265-275.
- Melton, D. W., Konecki, D. S., Brennand, J. and Caskey, C. T. (1984)** Structure, expression and mutation of the hypoxanthine phosphoribosyltransferase gene. *Proceedings of the National Academy of Science. U.S.A.* **81**, 2147-2151.
- Messerle, M., Keil, G. M. and Koszinowski, U. H. (1991)** Structure and expression of murine cytomegalovirus immediate-early gene 2. *Journal of Virology*. **65**, 1638-1643.
- Meyers, J. D., Flournoy, N. and Thomas, E. D. (1982)** Nonbacterial pneumonia after allogeneic marrow transplantation: a review of ten years' experience. *Reviews in Infectious Disease*. **4**, 1119-1132.
- Meyers, J. D., Flournoy, N. and Thomas, E. D. (1986)** Risk factors for cytomegalovirus infections after human marrow transplant. *Journal of Infectious Disease*. **153**, 478-488.
- Michaels, M. G., Jenkins, F. J., St George, K., Nalesnik, M. A., Starzl, T. E. and Rinaldo, C. R. (2001)** Detection of infectious baboon cytomegalovirus after baboon-to-human liver xenotransplantation. *Journal of Virology*. **75**, 2825-2828.
- Michelson, S., Turowski, P., Picard, L., Goris, J., Landini, M. P., Topilko, A., Hemmings, B., Bessia, C., Garcia, A. and Virelizier, J. L. (1996)** Human cytomegalovirus carries serine/threonine protein phosphatases PP1 and a host-cell derived PP2A. *Journal of Virology*. **70**, 1415-1423.

- Miller, G. (1990)** Epstein-Barr virus: biology, pathogenesis and medical aspects. *In: Virology*. pp 1843-1887. Fields, B., Knipe, D. M., Chanode, R. M., Hirsch, M. S., Melnick J. L., Monath, T. P. and Roizman, B (eds). Raven Press, New York.
- Minson, A. C., Davison, A., Eberle, R., Desrosiers, R. C., Fleckenstein, B., McGeoch, D. J., Pellett, P. E., Roizman, B. and Studdert, D. M. J. (2000).** The seventh report of the International Committee on Taxonomy of Viruses. *In: Virus Taxonomy: The classification and nomenclature of Viruses*. pp. 203-225. Van Regenmortel, M. H. V., Fauquet, C. M., Bishop, D. H. L., Carstens, E. B., Estes, M. K., Lemon, S. M., Maniloff, J., Mayo, M. A., McGeoch, D. J., Pringle, C. R. and Wickner, R. B.(eds). Academic Press, San Diego.
- Mocarski, E. S. (1993)** Cytomegalovirus biology and replication. *In: The human herpesviruses*. pp 173-226. Riozman, B., Whitley, R. and Lopez, C. (eds). Raven press, New York.
- Mocarski, E. S. (1996)** Cytomegalovirus and their Replication. *In: Fields Virology*. pp 2447-2491. Fields, B. N., Knipe, D. M. and Howley, P. M. (eds). Lippincott Williams and Wilkins.
- Mocarski, E. S. and Courcelle, C. T. (2001)** Cytomegaloviruses and their replication. *In: Fields Virology*. pp 2629-2672. Fields, B. N., Knipe, D. M. and Howley, P. M. (eds). Lippincott Williams and Wilkins.
- Mocarski, E. S., Liu, A. C. and Spaete, R. R. (1987)** Structure and variability of the a sequence in the genome of human cytomegalovirus (Towne strain). *Journal of General Virology*. **68**, 2223-2230.
- Mocarski, E. S., Pereira, L. and McCormick, A. L. (1988)** Human cytomegalovirus ICP22, the product of the HWLF1 reading frame, is an early nuclear protein that is released from cells. *Journal of General Virology*. **69**, 2613-2621.
- Mocarski, E. S., Prichard, M. N., Tan, C. S. and Brown, J. M. (1997)** Reassessing the organization of the UL42-UL43 region of the human cytomegalovirus strain AD169 genome. *Virology*. **239**, 169-175.
- Mocarski, E. S. and Stinski, M. F. (1979)** Persistence of the cytomegalovirus genome in human cells. *Journal of Virology*. **31**, 761-775.
- Mori, Y., Yagi, H., Shimamoto, T., Isegawa, Y., Sunagawa, T., Inagi, R., Kondo, K., Tano, Y. and Yamanishi, K. (1998)** Analysis of human herpesvirus 6 U3 gene, which is a positional homolog of human cytomegalovirus UL24 gene. *Virology*. **249**, 129-139.
- Muralidhar, S., Doniger, J., Mendelson, E., Araujo, J. C., Kashanchi, F, Azumi, N., Brady, J. N. and Rosenthal, J. (1996)** Human cytomegalovirus mtrII oncoprotein binds to p53 and down-regulates p53-activated transcription. *Journal of Virology*. **70**, 8691-8700.

- Nicholas, J. and Martin, M. E. (1994)** Nucleotide sequence analysis of a 38.5-kilobase-pair region of the genome of human herpesvirus 6 encoding human cytomegalovirus immediate-early gene homologs and transactivating functions. *Journal of Virology*. **68**, 597-610.
- Nicholas, J. (1996)** determination and analysis of the complete nucleotide sequence of human herpesvirus 7. *Journal of virology*. **70**, 5975-5989.
- Nicholl, M. J. and Preston, C. M. (1996)** Inhibition of herpes simplex virus type 1 immediate-early gene expression by alpha interferon is not VP16 specific. *Journal of Virology*. **70**, 6336-6339.
- Novotny, J., Rigoutsos, I., Coleman, D. and Shenk, T. (2001)** In silico structural and functional analysis of the human cytomegalovirus (HHV5) genome. *Journal of Molecular Biology*. **310**, 1151-1166.
- Nowak, B., Gmeiner, A., Sarnow, P., Levine, A. J. and Fleckenstein, B. (1984)** Physical mapping of human cytomegalovirus genes: identification of DNA sequences coding for a virion phosphoprotein of 71 kDa and a viral 65-kDa polypeptide. *Virology*. **134**, 91-102.
- O'Hare, P. (1993)** The virion transactivator of herpes simplex virus. *Seminars in Virology*. **4**, 145-155.
- Oien, N. L., Thomsen, D. R., Wathen, M. W., Newcomb, W. W., Brown, J. C. and Homa, F. L. (1997)** Assembly of herpes simplex virus capsids using the human cytomegalovirus scaffold protein: critical role of the C terminus. *Journal of Virology*. **71**, 1281-1291.
- Oram, J. D., Downing, R. G., Akrigg, A., Doggleyby, C. J., Wilkinson, G. W. and Greenway, P. J. (1982)** Use of recombinant plasmids to investigate the structure of the human cytomegalovirus genome. *Journal of General Virology*. **59**, 111-129.
- Pari, G. S. and Anders, D. G. (1993)** Eleven loci encoding trans-acting factors are required for transient complementation of human cytomegalovirus ori-Lyt-dependent DNA replication. *Journal of Virology*. **67**, 6979-6988.
- Pari, G. S., Kacica, M. A. and 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. *Journal of Virology*. **67**, 2575-2582.
- Patterson, C. E. and Shenk, T. (1999)** Human cytomegalovirus UL36 protein is dispensable for viral replication in cultured cells. *Journal of Virology*. **73**, 7126-7131.
- Pearson, W. R. and Lipman, D. J. (1988)** Improved tools for biological sequence comparison. *Proceedings of the National Academy of Science. USA*. **85**, 2444-2448.

- Penfold, M. E., Dairaghi, D. J., Duke, G. M., Saederup, N., Mocarski, E. S., Kemble, G. W. and Schall, T. J. (1999) Cytomegalovirus encodes a potent alpha chemokine. *Proceedings of the National Academy of Science. USA.* **96**, 9839-9844.
- Penfold, M. E. and Mocarski, E. S. (1997) Formation of cytomegalovirus DNA replication compartments defined by localization of viral proteins and DNA synthesis. *Virology.* **239**, 46-61.
- Petti, L., Sample, J., Wang, F. and Kieff, E. (1988) A sixth Epstein-Barr virus nuclear protein (EBNA3B) is expressed in latently infected growth-transformed lymphocytes. *Journal of Virology.* **62**, 2173-2178.
- Pignatelli, S., DalMonte, P. and Landini, M. P. (2001) gpUL73 (gN) genomic variants of human cytomegalovirus isolates are clustered into four distinct genotypes. *Journal of General Virology.* **82**, 2777-2784.
- Plachter, B., Traupe, B., Albrecht, J. and Jahn, G. (1988) Abundant 5 kb RNA of Human Cytomegalovirus without a Major Translation Reading Frame. *Journal of General Virology.* **69**, 2251-2266.
- Plotkin, S. A., Furukawa, T., Zygraich, N. and Huygelen, C. (1975) Candidate cytomegalovirus strain for human vaccination. *Infection and Immunity.* **12**, 521-527.
- Prichard, M. N., Duke, G. M. and Mocarski, E. S. (1996) Human cytomegalovirus uracil DNA glycosylase is required for the normal temporal regulation of both DNA synthesis and viral replication. *Journal of Virology.* **70**, 3018-3025.
- Prichard, M. N., Gao, N., Jairath, S., Mulamba, G., Krosky, P., Coen, D. M., Parker, B. O. and Pari, G. S. (1999) A recombinant human cytomegalovirus with a large deletion in UL97 has severe replication deficiency. *Journal of Virology.* **73**, 5663-5670.
- Prichard, M. N., Jairath, S., Penfold, M. E. T., St Jeor, S., Bohlman, M. C. and Pari, G. S. (1998) Identification of persistent RNA-DNA hybrid structures within the origin of replication of human cytomegalovirus. *Journal of Virology.* **72**, 6997-7004.
- Prosch, S., Stein, J., Staak, K., Liebenthal, C., Volk, H. D. and Kruger, D. H. (1996) Inactivation of the very strong HCMV immediate early promoter by DNA CpG methylation in vitro. *Biological Chemistry Hoppe-Seyler.* **377**, 195-201.
- Proudfoot, N. R. and Brownlee, G. G. (1976) 3' non-coding region sequences in eukaryotic messenger RNA. *Nature.* **263**, 211-214.
- Quackenbush E. J., Wershil, B. K. Aguirre, V. and Gutierrez-Ramos, J. C. (1998) Eotaxin modulates myelopoiesis and mast cell development from embryonic hematopoietic progenitors. *Blood.* **92**, 1887-1897.

- Quinnan, G. V., Delery, M., Rook, A. H., Frederick, W. R., Epstein, J. S., Manischewitz, J. F., Jackson, L., Ramsey, K. M., Mittal, K., Plotkin, S. A. and Hilleman, M. R. (1984). Comparative virulence and immunogenicity of the Towne strain and a nonattenuated strain of cytomegalovirus. *Annals of Internal Medicine* **101**, 478-483.
- Rasmussen, L. E., Nelson, R. M., Kelsall, D. C. and Merigan, T. C. (1984) Murine monoclonal antibody to a single protein neutralizes the infectivity of human cytomegalovirus. *Proceedings of the National Academy of Science U.S.A* **81**, 876-880.
- Rawlinson, W. D. and Barrell, B. G. (1993). Spliced transcripts of human cytomegalovirus. *Journal of Virology* **67**, 5502-5513.
- Rawlinson, W. D., Farrell, H. E. and Barrell, B. G. (1996). Analysis of the complete DNA sequence of murine cytomegalovirus. *Journal of Virology* **70**, 8833-8849.
- Ripalti, A., Boccuni, M. C., Campanini, F. and Landini, M. P. (1995). Cytomegalovirus-mediated induction of antisense mRNA expression to UL44 inhibits virus replication in an astrocytoma cell line: identification of an essential gene. *Journal of Virology* **69**, 2047-2057.
- Rixon, F. J. (1993). Structure and assembly of herpesviruses. *Seminars in Virology* **4**, 135-144.
- Rock, D. L. (1993). The molecular basis of latent infections by alphaherpesviruses. *Seminars in Virology* **4**, 157-165.
- Rock, D. L. and Fraser, N. W. (1983). Detection of HSV-1 genome in central nervous system of healthy infected mice. *Nature* **302**, 523-525.
- Rock, D. L. and Fraser, N. W. (1985). Latent herpes simplex virus type 1 DNA contains two copies of the virion DNA joint region. *Journal of Virology* **55**, 849-852.
- Roizman, B. (1979). The structure and isomerisation of herpes simplex virus genomes. *Cell* **16**, 481-494
- Roizman, B. (1996). Herpesviridae. In: *Fields Virology*. pp 2221-2230. Knipe D. M., Howley P. M. *et al.* (eds). Raven publishers, Philadelphia.
- Roizman, B., Carmichael, L. E., Deinhardt, F., de-The, G., Nahmias, A. J., Plowright, W., Rapp, F., Sheldrick, P., Takahashi, M. and Wolf, K. (1981). Herpesviridae. Definition, provisional nomenclature, and taxonomy. *Intervirology* **16**, 201-217.
- Roizman, B., Desrosiers, R. C., Fleckenstein, B., Minson, A. C. and Studdert, M. J. (1992). The family herpesviridae: an update. *Archives of Virology* **123**, 425-449.
- Roizman, B., Desrosiers, R. C., Fleckenstein, B., Minson, A. C. and Studdert, M. J. (1995). *Herpesviridae*. In. "Virus Taxonomy". *Archives of Virology* **10**, 114-127.

- Roizman, B. and Furlong, D. (1974).** The replication of herpesvirus. *In: Comprehensive virology* pp 229-403. Fraenkel-Conrat, H. and Wagner, R. R. Plenum, New York.
- Roizman, B. and Sears, A. E. (1990).** Herpes simplex virus and their replication. *In: Virology* pp 1795-1841. Fields, B. N., Knipe, D. M. and Howley, P. M. (eds). Raven, New York.
- Roizman, B. and Sears, A. E. (1996).** Herpes simplex virus and their replication. *In: Virology* pp 2231-2295. Fields, B. N., Knipe, D. M. and Howley, P. M. (eds). Lippincott Williams and Wilkins.
- Romanowski, M. J., Garrido-Guerrero, E. and Shenk, T. (1997).** pIRS1 and pTRS1 are present in human cytomegalovirus virions. *Journal of Virology* **71**, 5703-5705.
- Romanowski, M. J. and 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. *Journal of Virology* **71**, 1485-1496.
- Rosa, M. D., Gottlieb, E., Lerner, M. R. and Steitz, J. A. (1981).** Striking similarities are exhibited by two small Epstein-Barr virus-encoded ribonucleic acids and the adenovirus-associated ribonucleic acids VAI and VAII. *Molecular and Cellular Biology* **1**, 785-796.
- Rowe, W. P., Hartley, J. W., Waterman, S., Turner, H. C. and Hueber, R. J. (1956).** Cytopathogenic agent resembling human salivary gland virus recovered from tissue cultures of human adenoids. *Journal of Virology* **64**, 5167-5172.
- Rudolph, S. A., Kuhn, J. E., Korn, K., Braun, R. W. and Jahn, G. (1990).** Prokaryotic expression of the major capsid protein of human cytomegalovirus and antigenic cross-reactions with herpes simplex virus type 1. *Journal of General Virology* **71**, 2023-2031.
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989).** Molecular cloning: A laboratory manual. Cold Spring Laboratory, Cold Spring Harbor, New York.
- Sanchez, V., Angeletti, P. C., Engler, J. A. and Britt, W. J. (1998).** Localization of human cytomegalovirus structural proteins to the nuclear matrix of infected human fibroblasts. *Journal of Virology* **72**, 3321-3329.
- Sanchez, V., greis, K. D., Sztul, E. and Britt, W. J. (2000).** Accumulation of virion tegument and envelope proteins in a stable cytoplasmic compartment during human cytomegalovirus replication: characterization of a potential site of virus assembly. *Journal of Virology* **74**, 975-986.
- Schmolke, S., Kern, H. F., Drescher, P., Jahn, G. and Plachter, B. (1995).** The dominant phosphoprotein pp65 (UL83) of human cytomegalovirus is dispensable for growth in cell culture. *Journal of Virology* **69**, 5959-5968.

- Senapathy, P., Shapiro, M. B. and Harris, N. L. (1990).** Splice junctions, branch points sites, and exons: sequence statistics, identification, and applications to genome project. *Methods in Enzymology* **183**, 252-278.
- Severi, B., Landini, M. P., Cenacchi, G., Zini, N. and Maraldi, N. M. (1992).** Human cytomegalovirus nuclear and cytoplasmic dense bodies. *Archives of Virology* **123**, 193-207.
- Sheaffer, A. K., Weinheimer, S. P. and Tenney, D. J. (1997).** The human cytomegalovirus UL98 gene encodes the conserved herpes alkaline nuclease. *Journal of General Virology* **78**, 2953-2961.
- Sinclair, J., Ballie, J., Bryant, L. and Caswell, R. (2000).** Human cytomegalovirus mediates cell cycle progression through G (1) into early S phase in terminally differentiated cells. *Journal of General Virology* **81**, 1553-1565.
- Sinzger, C. A. and Grefte A., Plachter, B., Gouw, A. S., The, T. H. and Jahn, G. (1995).** Fibroblasts, epithelial cells, endothelial cells and smooth muscle cells are major targets of human cytomegalovirus infection in lung and gastrointestinal tissues. *Journal of General Virology* **76**, 741-750.
- Sinzger, C. A. and Jahn, G. (1996).** Human cytomegalovirus cell tropism and pathogenesis. *Intervirology* **39**, 302-319.
- Sinzger, C. A., Kahl, M., Laib, K., Klingel, K., Rieger, P., Plachter, B. and Jahn, G. (2000).** Tropism of human cytomegalovirus for endothelial cells is determined by a post-entry step dependent on efficient translocation of the nucleus. *Journal of General Virology* **81**, 3021-3035.
- Skaletskaya, A., Bartle, L. M., Chittenden, T., McCormick, A. L., Mocarski, E. S. and Goldmacher, V. S. (2001).** A cytomegalovirus-encoded inhibitor of apoptosis that suppresses caspase-8 activation. *Proceedings of the National Academy of Science USA* **98**, 7829-7834.
- Slobedman, B. and Mocarski, E. S. (1999).** Quantitative analysis of latent human cytomegalovirus. *Journal of Virology* **73**, 4806-1482.
- Smiley, M. L., Wlodaver, C. G., Grossman, R. A., Barker, C. F., Perloff, L. J., Tustin, M. B., Starr, S. E., Plotkin, S. A. and Freidman, H. M. (1985).** The role of pretransplant immunity in protection from cytomegalovirus disease following renal transplantation. *Transplantation* **40**, 157-161.
- Smith, J. A. (1956).** Propagation in tissue culture of a cytopathogenic virus from human salivary gland virus disease. *Proceedings of the Society for Experimental Biology and Medicine* **92**, 424-430.
- Smith, J. A. and Pari, G. S. (1995).** Human cytomegalovirus UL102 gene. *Journal of Virology* **69**, 1734-1740.

- Soeda, E., Arrand, J. R., Smolar, N. and Griffin, B. E. (1979). Sequence from early region of polyoma virus DNA containing viral replication origin and encoding small, middle and (part of) large T antigens. *Cell* **17**, 357-370.
- Somogyi, T., Colimon, R., Bertrand, C. and Michelson, S. (1987). Conservation and map location of human cytomegalovirus strain AD169 transforming sequences in the DNA of clinical isolates. *Micobiologia* **10**, 125-131.
- Spaderna, S., Blessing, H., Bogner, E., Britt, W. and Mach, M. (2002). Identification of glycoprotein gpTRL10 as a structural component of human cytomegalovirus. *Journal of Virology* **76**, 1450-1460.
- Spaete, R. R., Perot, K., Scott, P. I., Nelson, J. A., Stinski, M. F. and Pachl, C. (1993). Coexpression of truncated human cytomegalovirus gH with the UL115 gene product or the truncated human fibroblast growth factor receptor results in transport of gH to the cell surface. *Virology* **193**, 853-861.
- Speack, S. T. and Strominger, J. L. (1989). Tumorigenic DNA viruses. In: *Advances in Viral Oncology* pp 133. Klein, G. (ed). Raven Press. New York.
- Speckner, A., Glykofrudes, D., Ohlin, M. and 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. *Journal of General Virology* **80**, 2183-2191.
- Spear, P. G. (1985). Glycoproteins specified by herpes simplex viruses. In: *Herpesviruses* pp 315-356. Roizman, B., (ed). Plenum Press, New York and London.
- Spear, P. G. (1993). Entry of alphaherpesviruses into cells. *Seminars in Virology* **9**, 143-159.
- Spector, D. H. (1996). Activation and regulation of human cytomegalovirus early genes. *Intervirology* **39**, 361-377.
- Spector, D. H., Hock, L. and Tamashiro, J. C. (1982). Cleavagemaps for human cytomegalovirus DNA strain AD169 for restriction endonucleases EcoR1, BglII, and HindIII. *Journal of Virology* **42**, 558-582.
- Spector, S. A., Hirata, K. K. and Newman, T. R. (1984). Identification of multiple cytomegalovirus strains in homosexual men with acquired immunodeficiency syndrome. *Journal of Infectious Diseases* **150**, 953-956.
- Spector, S. A., Weigeist, T., Pollard, R. B., Dieterich, D. T., Samo, T., Benson, C. A., Busch, D. F., Freeman, W. R., Monatgue, P. and Kaplan, H. J. (1993). A randomised, controlled study of intravenous ganciclovir therapy for cytomegalovirus peripheral retinitis in patients with AIDS. AIDS clinical trial group and cytomegalovirus cooperative study group. *Journal of Infectious Diseases* **168**, 557-563.

- Spector, S. A., Wong, R. and Hsia, K. (1998)** Plasma cytomegalovirus (CMV) DNA load predicts CMV disease and survival in AIDS patients. *Journal of Clinical Investigation*. **101**, 497-502.
- Spivack, J. G. and Fraser, N. W. (1987)** Deletion of herpes simplex virus type 1 transcripts during latent infection in mice. *Journal of Virology*. **61**, 3841
- Stainer, I., Spivack, J. G., Lirette, R. P., Brown, S. M., MacLean, A. R., Subak-Sharpe, J. H. and Fraser, N. W. (1989)** Herpes simplex virus type 1 latency-associated transcripts are evidently not essential for latent infection. *EMBO Journal*. **8**, 505-511.
- Stamminger, T. and Fleckenstein, B. (1990)** Immediate-early transcription regulation of human cytomegalovirus. *Current Topics in Microbiology and Immunology*. **154**, 3-20.
- Stango, S. and Cloud, G. A. (1994)** Working parents: the impact of day care and breast-feeding on cytomegalovirus infections in offspring. *Proceedings of the National Academy of Science. USA*. **91**, 2384-2389.
- Stango, S., Pass, R. F., Dworsky, M. E. and Alford, C. A. (1983)** Congenital and perinatal cytomegaloviral infections. *Seminars in Pathology*. **7**, 31-42.
- Stannard, L. M., Fuller, A. O. and Spear, P. G. (1987)** Herpes simplex virus glycoproteins associated with different morphological entities projecting from the virion envelope. *Journal of General Virology*. **68**, 715-725.
- Stannard L. M. and Hardic, D. R. (1991)** An Fc receptor for human immunoglobulin G is located within the tegument of human cytomegalovirus. *Journal of Virology*. **65**, 3411-3415.
- Staprans, S. I., Rabert, D. K. and Spector, D. H. (1988)** Identification of sequence requirements and *trans*-acting functions necessary for regulated expression of a human cytomegalovirus early gene. *Journal of Virology*. **62**, 3463-3473.
- Stasiak, P. C. and Mocarski, E. S. (1992)** Transactivation of the cytomegalovirus ICP36 gene promoter requires the α gene product TRS1 in addition to IE1 and IE2. *Journal of Virology*. **66**, 1050-1058.
- Steiner, M., Bostrum, B., Leonard, A. S. and Denher, L. P. (1989)** Undifferentiated (embryonal) sarcoma of the liver. A clinicopathology study of a survivor treated with combined technique therapy. *Cancer*. **64**, 1318-1322.
- Stenberg, R. M. (1996)** The human cytomegalovirus major immediate-early gene. *Intervirology*. **39**, 343-349.
- Stenberg, R. M., Thomsen, D. R. and Stinski, F. (1984)** Structural analysis of the major immediate early gene of human cytomegalovirus. *Journal of Virology*. **49**, 190-199.

- Stenberg, R. M., Witte, P. R. and Stinski, 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. *Journal of Virology* **56**, 665-675.
- Stenberg, R. M., Depto, A. S., Fortney, J. and Nelson, J. (1989).** Regulated expression of early and late RNAs and proteins from the human cytomegalovirus immediate-early gene region. *Journal of Virology* **63**, 2699-2708.
- Steven, A. C. and Spear, P.G. (1997).** Herpesvirus capsid assembly and envelopment. *In: Structural Biology of Viruses* pp 312-351. Chiu, W. (ed). Oxford University Press Inc., USA.
- Stevens, J. G., Haarr, L., Porter, D. D., Cook, M. L. and Wagner, E. K (1988).** Prominence of the herpes simplex virus latency-associated transcript in trigeminal ganglia from seropositive humans. *Journal of infectious diseases* **158**, 117-123.
- Stinski, M. F. (1976).** Glycoproteins associated with virions and dense bodies. *Journal of Virology* **19**, 594-609.
- Stow, N. D. and McMonagle, E. C. (1983).** Characterization of the TR_S/IR_S origin of DNA replication of herpes simplex virus type 1. *Virology* **136**, 427-438.
- Subak-Sharpe, J. H. and Dargan, D. J. (1998).** HSV molecular biology: general aspects of herpes simplex virus molecular biology. *Virus Genes* **16**, 239-251.
- Szilagyi, J. F. and Berriman, J. (1994).** Herpes simplex virus L particles contain spherical membrane-enclosed inclusion vesicles. *Journal of General Virology* **75**, 1749-1753.
- Takekoshi, M., Ihara, S., Tanaka, S., Maeda-Takekoshi, F. and Watanabe, Y. (1987).** A new human cytomegalovirus isolate has an invertible subsegment within its L component producing eight genome isomers. *Journal of General Virology* **68**, 765-776.
- Taylor, P. (1986).** A computer program for translating DNA sequences into protein. *Nucleic Acids research* **14**, 437-441.
- Taylor-Wiedman, J. Sissons, J. G., Borysiewicz, L. K. and Sinclair, J. H. (1994).** Monocytes are a major sites of persistence of human cytomegalovirus in peripheral blood mononuclear cells. *Journal of General Virology* **72**, 2059-2064.
- Thomas, S. K., Gough, G., Latchman, D. S. and Coffin, R. S. (1999).** Herpes simplex virus latency-associated transcript encodes a protein which greatly enhances virus growth, can compensate for deficiencies in immediate-early gene expression, and is likely to function during reactivation from virus latency. *Journal of Virology* **73**, 6618-6625.
- Thompson, R. L. and Sawtell, N. M. (1997).** The herpes simplex virus type 1 latency-associated transcript gene regulates the establishment of latency. *Journal of Virology* **71**, 6660-6675.

- Thompson, R. L. and Sawtell, N. M. (2001).** Herpes simplex virus type 1 latency-associated transcript gene promotes neuronal survival. *Journal of Virology* **75**, 5432-5440.
- Thomson, B. J. and Honess, R. W. (1992).** The right end of the unique region of the genome of human herpesvirus 6 U1102 contains a candidate immediate early gene enhancer and a homologue of the human cytomegalovirus US22 gene family. *Journal of General Virology* **73**, 1649-1660.
- Thomsen, D. R., Roof, L. L. and Homa, F. L. (1994).** Assembly of herpes simplex virus (HSV) intermediate capsids in insect cells infected with recombinant baculoviruses expressing HSV capsid proteins. *Journal of Virology* **68**, 2442-2457.
- Tomasec, P., Braud, V. M., Rickards, C., Powell, M. B., McSharry, B. P., Gadola, S., Cerundolo, V., Borysiewicz, L. K., McMichael, A. J. and Wilkinson, G. W. (2000).** Surface expression of HLA-E, an inhibitor of natural killer cells, enhanced by human cytomegalovirus gpUL40. *Science* **287**, 1031
- Tomita, M. and Marchesi, V. T. (1975).** Amino acid sequence and oligosaccharide attachment sites of human erythrocyte glycophorin. *Proceedings of the National Academy of Science USA* **72**, 2964-2968.
- Trus, B. L., Gibson, W., Cheng, N. and Steven, A. C. (1999).** Capsid structure of simian cytomegalovirus from cryoelectron microscopy: evidence for tegument attachment sites. *Journal of Virology* **73**, 2181-2192.
- Underwood, M. R., Harvey, R. J., Stanat, S. C., Hemphill, M. L., Miller, T., Drach, J. C., Townsend, L. B. and Birion, K. K. (1998).** Inhibition of human cytomegalovirus DNA maturation by a benzimidazole ribonucleoside is mediated through the UL89 gene product. *Journal of Virology* **72**, 717-725.
- Van der Meer, J. T. M., Drew, W. L., Bowden, R. A., Galasso, G. J., Griffiths, P. D., Douglas, A. J., Katlama, C., Spector, S. A. and Whitley, R. J. (1996).** Summary of the international consensus symposium on advances in the diagnosis, treatment, and prophylaxis of cytomegalovirus infection. *Antiviral Research* **32**, 119-140.
- Vink, C., Beuken, E. and Bruggeman, C. A. (2000).** Complete DNA sequence of the rat cytomegalovirus genome. *Journal of Virology* **74**, 7656-7665.
- Wagner, E. K. (1985).** Individual HSV transcripts. *In: Herpesviruses*. pp45-104. Roizman, B. (ed). Plenum Publishing Corp. New York.
- Wagner, E. K., Flanagan, W. M., Devi-Rao, G., Zhang, Y. F., Hill, J. M. and Anderson KP, Stevens, J. G. (1988).** The herpes simplex virus latency-associated transcript is spliced during the latent phase of infection. *Journal of Virology* **62**, 4577-4585.
- Waldman, W. J., Sneddon, J. M., Stephens, R. E. and Roberts, W. H. (1989).** Enhanced endothelial cytopathogenicity induced by a cytomegalovirus strain propagated in endothelial cells. *Journal of Medical Virology* **28**, 223-230.

- Wathen M. W., Thomsen, D. R. and Stinski, M. F. (1981) Temporal regulation of human cytomegalovirus transcription at immediate early and early times after infection. *Journal of Virology*. **38**, 446-459.
- Welch, A. R., McGregor, L. M. and Gibson, W. (1991) Cytomegalovirus homologues of cellular G protein-coupled receptor genes are transcribed. *Journal of Virology*. **65**, 3915-3918.
- Weller, T. H. (1971) The cytomegaloviruses: ubiquitous agents with protean clinical manifestations. *New England Journal of Medicine*. **4** and **5**, 203-214 and 267-274.
- Weller, T. H., Macaulay, J. C., Craig, J. M. and 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.
- Weller, S. A., Spadaro, A., Schaffer, J. E., Murray, A. W., Maxam, A. M. and Schaffer, P. A. (1985) Cloning, sequencing, and functional analysis of *ori_L*, a herpes simplex virus type 1 origin of DNA synthesis. *Molecular and Cellular Biology*. **5**, 930-942.
- Weston, K. (1988) An enhancer element in the short unique region of human cytomegalovirus regulates the production of a group of abundant immediate early transcript. *Virology*. **162**, 406-416.
- Weston, K. and Barrell, B. G. (1986) Sequence of the short unique region, short repeats and part of the long repeats of human cytomegalovirus. *Journal of Molecular Biology*. **192**, 177-208.
- Weststrate, M. W., Geelen, J. L., Wertheim, P. M. and van der Noordaa, J. (1983) Comparison of the physical maps of the DNAs of two cytomegalovirus strains. *Journal of General Virology*. **64**, 47-55.
- White, K. L., Slobedman, B. and Mocarski, E. S. (2000) Human cytomegalovirus latency-associated protein pORF94 is dispensable for productive and latent infection. *Journal of Virology*. **74**, 9333-9337
- Whitley, R. and Schlitt, M. (1991) Encephalitis caused by herpesviruses, including B virus. In: *Infections of the central nervous system*. pp 41-69. Scheld, W., Whitley, R. and Durack, D. (eds). Raven Press, New York.
- Wiertz, E. J., Jones, T. R., Sun, L., Bogoy, M., Geuze, H. J. and Ploegh, H. L. (1996) The human cytomegalovirus US11 gene product dislocates MHC class I heavy chain from the endoplasmic reticulum to the cytosol. *Cell*, **84**, 769-779.
- Wildy, P., Russel, W. C. and Horne, R. W. (1960) The morphology of herpesvirus. *Virology*. **12**, 204-222.

- Wilkinson, G. W., Akrigg, A. and Greenway, P. J. (1984). Transcription of the immediate early genes of human cytomegalovirus strain AD169. *Virus Research* 1, 101-116.
- Williamson, W. D., Demmler, G. J., Percy, A. K. and Catlin, F. I. (1992). Progressive hearing loss in infants with asymptomatic congenital cytomegalovirus infection. *Pediatrics* 90, 862-866.
- Wing, B. A. and Huang, E. S. (1995). Analysis and mapping of a family of 3'-coterminal transcripts containing coding sequences for human cytomegalovirus open reading frames UL93 through UL99. *Journal of Virology* 69, 1521-1531.
- Winkler, M., Aus Dem, S. T. and Stamminger, T. (2000). Functional interaction between pleiotropic transactivator pUL69 of human cytomegalovirus and the human homolog of yeast chromatin regulatory protein SPT6. *Journal of Virology* 74, 8053-8064.
- Winkler, M., Rice, S. A. and 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. *Journal of Virology* 68, 3943-3954.
- Winkler, M., Shmolke, S., Plachter, B. and Stamminger, T. (1995). The pUL69 protein of human cytomegalovirus (HCMV), a homologue of the herpes simplex virus ICP27, is contained within the tegument of virions and activates the major immediate-early enhancer of HCMV in synergy with the tegument protein pp71 (ppUL82). *Scandinavian Journal of Infectious Disease* 99, 8-9.
- Winkler, M. and Stamminger, T. (1996). A specific subform of the human cytomegalovirus transactivator protein pUL69 is contained within the tegument of virus particles. *Journal of Virology* 70, 8984-8987.
- Winston, D. J., Ho, W. G. and Champlin, R. E. (1990). Cytomegalovirus infection after allogenic bone marrow transplantation. *Reviews in Infectious Disease* 12, 5776-5792.
- Wright, D. A. and Spector, D. H. (1989). Posttranscriptional regulation of a class of human cytomegalovirus phosphoproteins encoded by an early transcription unit. *Journal of Virology* 63, 3117-3127.
- Wright, D. A., Staprans, S. I. and Spector, D. H. (1988). Four phosphoproteins with common amino termini are encoded by human cytomegalovirus AD169. *Journal of Virology* 62, 331-340.
- Xiao, J., Tong, T., Zhan, X., Haghjoo, E. and Liu, F. (2000). In vitro and in vivo characterization of a murine cytomegalovirus with a transposon insertional mutation at open reading frame M43. *Journal of Virology* 74, 9488-9497.
- Yurochko, A. D., Hwang, E. S., Rasmussen, L., Keay, S., Pereira, L. and Huang, E. S. (1997). The human cytomegalovirus UL55 (gB) and UL75 (gH) glycoprotein ligands initiate the rapid activation of Sp1 and NF-kappaB during infection. *Journal of Virology* 71, 5051-5059.

- Yates, J., Warren, N., Reisman, D. and Sugden, B. (1984).** A cisacting element from the Epstein-Barr virus genome that permits stable replication of recombinant plasmids in latently infected cells. *Proceedings of the National Academy of Science. USA* **81**, 3806-3810.
- Zhang, H., Al-Barazi, H. O. and Colberg-Poley, A. M. (1996).** The acidic domain of human cytomegalovirus UL37 immediate early protein is dispensable for its transactivating activity and localization but is not for its synergism. *Virology* **223**, 292-302.
- Zhang, H., Fu, S., Busch, A., Chen, F., Qin, L. and Bromberg, J. S. (2001).** Identification of TNF-alpha-sensitive sites in HCMV *ie1* promoter. *Experimental and Molecular Pathology* **71**, 106-114.
- Zhou, Y., Chandran, B. and Wood, C. (1997).** Transcriptional patterns of the pCD41 (U27) locus of human herpesvirus 6. *Journal of Virology* **71**, 3420-3430.
- Zhou, Z., Chen, D., Jakana, J., Rixon, F. and Chiu, W. (1999).** Visualisation of tegument/capsid interactions and DNA in intact herpes simplex virus type 1 virions. *Journal of Virology* **73**, 3210-3218.
- Zhou, Z., He, J., Jakana, J., Tatman, J. D., Rixon, F. J. and Chiu, W. (1995).** Assembly of VP26 in herpes simplex virus-1 inferred from structures of wild-type and recombinant capsids. *Nature Structural Biology* **2**, 1026-1030.
- Zhu, Y., Huang, L. and Anders, D. G. (1998).** Human cytomegalovirus *oriLyt* Sequence requirements. *Journal of Virology* **72**, 4989-4996.
- Zipeto, D., Bodaghi, B., Laurent, L., Virelizier, J. L. and Michelson, S. (1999).** Kinetics of transcription of human cytomegalovirus chemokine receptor US28 in different cell types. *Journal of General Virology* **80**, 543-547.
- Zabolotny, J. M., Krummenacher, C. and 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. *Journal of Virology* **71**, 4199-4208.
- Zur Hausen, H. and Schultz-Hausen, H. (1970).** Presence of EB virus nucleic acid homology in a "virus free" line of Burkitt tumor cells. *Nature* **227**, 245-248.

5'-ends detected by RACE.

^a Position of 5'-ends in the AD169 genome. ^b Total number of 5'-cDNA clones corresponding to the 5'-end. ^c Number of G residues immediately upstream to the 5'-end.

^d PCR products not originating from HCMV. 5'-ends located far upstream and downstream from the first ATG (as described by Chee *et al.*, 1990) are shown in pink and blue, respectively.

ORF	5'-ends ^a	No. of clones ^b	SMART RACE ambiguity ^c of 5'-end
TRL1	No RACE product		
TRL2	2083 2086	4 1	1
TRL3	No RACE product		
TRL4	Cellular ^d		
TRL5	4252	5	
TRL6	5800 5796 5990 5996 5999 6080	4 1 5 1 1 5	1 1
TRL7	No RACE product		
TRL8	7350 7259 7257 7252 6911	4 3 1 2 5	1 1 3
TRL9	7704 7701	6 2	2
TRL10	8140 8143 8144 8071 8074	7 2 1 3 1	4 2 1
TRL11	8452	4	

	8509	5	
	8501	1	
	8499	1	1
	8560	4	
	8567	2	2
	8605	5	2
	8600	1	1
	8662	3	
	8661	1	2
TRL12	9342	6	1
	9339	1	
TRL13	10406	4	
	10404	1	1
	10401	2	1
	10443	5	
	10677	3	1
	10680	1	1
	10736	2	
	10738	4	2
	10745	1	
TRL14	10406	4	
	10404	1	1
	10401	2	1
	10443	5	
	10677	3	1
	10680	1	1
	10736	2	
	10738	4	2
	10745	1	
UL1	No RACE product		
UL2	13186	4	
	13175	2	4
UL3	13471	4	
	13469	2	
	13468	1	
	13465	2	
	13296	3	
	13299	2	1

UL4	13290	3	1
	13288	1	1
	13182	5	
	13185	1	1
	13188	2	1
	13046	6	
	13050	1	
UL5	No RACE product		
UL6	14271	4	1
	14275	1	
	14277	1	1
	14585	7	
	14595	2	3
UL7	15068	3	
	15073	2	
UL8	16027	4	4
	16023	2	
	16019	1	1
	16180	5	
UL9	No RACE product		
UL10	16933	6	3
	16937	2	
UL11	18270	5	
	18272	1	
	18269	2	1
	18267	1	1
UL12	19699	4	
	19353	5	
	19354	1	1
UL13	19308	4	
	19311	1	1
	19317	1	1
UL14	20713	5	
	20719	1	2
	20709	2	1
	20704	1	3
	20952	3	4

UL15A	Not analysed		
UL16	22325	5	
	22324	1	2
	22312	2	
	22311	1	
UL17	22896	3	
	22894	2	1
	22890	2	3
		3	2
	23171	1	3
	23176	1	1
	23168		
UL18	23724	25	4
	23716	4	
UL19	24269	4	
	24271	1	1
	24275	2	
	24281	1	3
UL20	25253	4	
	25252	1	2
	25249	2	
	25248	1	1
	25161	3	
	25168	2	3
	25158	1	4
	24979	4	
	24977	1	2
	24971	2	2
UL21A	26858	10	
UL22A	27080	5	
UL23	28855	22	
	28858	4	1
UL24	29994	12	2
	29990	1	3
	29997	1	
	30086	15	
	30098	3	1
	30088	1	3
	30084	2	
	30082	4	1
			1
UL25	30016	5	
	30020	1	
	30024	2	

	30025	2	
	30029	1	1
	30065	3	1
UL26	32785	5	
	32787	2	1
	32791	1	1
UL27	35295	4	1
	35292	1	
	35286	2	2
UL28	No RACE product		
UL29	Cellular ^d		
UL30	37525	6	
	37526	1	1
	37521	2	
	37518	1	
UL36	49863	21	1
	49866	3	1
	49856	2	1
	49855	5	1
	49849	3	2
UL43	Cellular ^d		
US22	211626	25	
US23	213544	12	
	213545	3	1
	213547	1	
	213549	2	
	213550	1	1
	213654	15	
	213650	1	1
	213646	5	
US24	214970	14	
	214975	5	3
	214971	2	
	214966	4	2
	214965	3	
US26	Cellular ^d		
UL128	175681	5	
	175685	1	
	175677	2	
	175676	1	
UL131A	Cellular ^d		