

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

A Thesis Submitted for the Degree of PhD at the University of Warwick

<http://go.warwick.ac.uk/wrap/73513>

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.

The Expression and Function of the Adenovirus Type 5 E4 region

Ian Dix

Bsc. (Hons.) (Warwick)

**A thesis presented for
the degree of PhD in the
Universtiy of Warwick**

Department of Biological Sciences

September 1994

PAGE

NUMBERING

AS ORIGINAL

Table of contents

Table of contents	i
List of figures	ix
List of tables	xiii
Acknowledgements	xiv
Declaration	xv
Abbreviations	xvi
Summary	xx
Chapter 1. Introduction	1
<u>1.1. Adenoviruses- a general introduction</u>	2
1.1.1. The adenovirus family	2
1.1.2. The structure of adenoviruses	3
1.1.3. Adenovirus replication	5
1.1.4. Adenovirus gene function	8
1.1.4.1. The E1a region	8
1.1.4.2. The E1b region	10
1.1.4.3. The E2 region	14
1.1.4.4. The E3 region	15
1.1.4.5. The E4 region	16
1.1.4.6. Late transcription units	16
1.1.4.7. VA RNA	17
<u>1.2. The regulation of gene expression in an adenovirus infection</u>	17
1.2.1. Introduction	17
1.2.2. Transcription initiation	17
1.2.3. mRNA polyadenylation	18
1.2.4. RNA splicing	20
1.2.4.1. Introduction	20
1.2.4.2. Mechanism 1: viral splicing factors	22

1.2.4.3. Mechanism 2: <i>cis</i> -acting sequences	23
1.2.4.4. Mechanism 3: general splicing factors	24
1.2.4.5. Summary	26
1.2.5. mRNA export	26
1.2.6. Translation	27
1.2.6.1. Introduction	27
1.2.6.2. VA _I RNA	28
1.2.6.3. L4-100K	29
1.2.6.4. The MLTU tripartite leader	29
<u>1.3. E1b-55K and cytoplasmic viral mRNA accumulation</u>	30
1.3.1. Introduction	30
1.3.2. RNA transport models	31
1.3.3. The E1b-55K protein - subcellular localisation	35
1.3.4. The elucidation of E1b-55K function	37
1.3.5. A model for E1b-55K function	40
<u>1.4. The E4 region</u>	44
1.4.1. E4 expression	44
1.4.2. The functions of the E4 proteins	47
1.4.2.1. Introduction	47
1.4.2.2. ORF1 and ORF2	48
1.4.2.3. ORF3 and ORF6	50
1.4.2.4. ORF3/4	57
1.4.2.5. ORF4	57
1.4.2.6. ORF5	60
1.4.2.7. ORF6/7	60
<u>1.4. Project objectives</u>	63

Chapter 2. Materials and Methods	64
<u>2.1. Materials</u>	65
2.1.1. Common buffers and solutions	65
2.1.2. Oligonucleotide primers	65
<u>2.2. Suppliers</u>	66
<u>2.3. Methods</u>	68
2.3.1. Nucleic acid manipulations	68
2.3.1.1. Ethanol precipitation of nucleic acids	68
2.3.1.2. Nucleic acid quantification	68
2.3.1.3. Phenol/chloroform extraction	68
2.3.1.4. Restriction endonuclease digestion	69
2.3.1.5. End-fill repair of 5' overhangs	69
2.3.1.6. T4 DNA polymerase removal of 3' overhangs	69
2.3.1.7. Dephosphorylation	70
2.3.1.8. Agarose gel electrophoresis	70
2.3.1.9. Purification of dsDNA fragments from agarose gels	70
2.3.1.10. Ligation of dsDNA	71
2.3.1.11. Polymerase chain reaction (PCR)	72
2.3.1.12. Site-directed mutagenesis using PCR	72
2.3.1.13. DNA sequencing	74
2.3.2. Bacteriological techniques	76
2.3.2.1. The growth and maintenance of bacteria required for routine cloning	76
2.3.2.2. The transformation of bacteria with plasmid DNA	76
2.3.2.3. Small scale plasmid DNA purification from bacteria (Mini-Prep)	78
2.3.2.4. Large scale plasmid DNA preparation from bacteria (Maxi-Prep)	78
2.3.3. M13 Phage techniques	79
2.3.3.1. Transformation procedure	79
2.3.3.2. Culturing M13 clones	80

2.3.3.3. Preparation of RF DNA from M13 cultures	80
2.3.3.4. Preparation of single stranded DNA from M13 plaques	80
2.3.4. Tissue culture and virus techniques	81
2.3.4.1. Maintenance and passage of cell lines	81
2.3.4.2. Maintenance and passage of non-transformed cells	81
2.3.4.3. Freezing cell stocks	82
2.3.4.4. Recovery of cell stocks	82
2.3.4.5. Calcium phosphate transfection for the construction of adenovirus mutants	82
2.3.4.6. Virus infection of cell monolayers	83
2.3.4.7. Rapid preparation of adenovirus DNA	83
2.3.4.8. Large scale preparation of adenovirus DNA	84
2.3.4.9. Preparation of virus stocks	85
2.3.4.10. Preparation of purified adenovirus virions	85
2.3.4.11. Titration of adenovirus by plaque assay	86
2.3.5. Analysis of viral RNA expression	86
2.3.5.1. Preparation of cytoplasmic and nuclear RNA	86
2.3.5.2. Preparation of antisense RNA probes	87
2.3.5.3. Ribonuclease protection analysis	88
2.3.5.4. RNA polyacrylamide gel electrophoresis (PAGE)	88
2.3.6. Analysis of proteins	89
2.3.6.1. SDS polyacrylamide gel electrophoresis (SDS-PAGE)	89
2.3.6.2. Coomassie-blue staining of SDS-polyacrylamide gels	89
2.3.6.3. Monochromatic silver staining	90
2.3.6.4. Preparation of SDS-polyacrylamide gels for autoradiography	90
2.3.6.5. Bradford assay	90
2.3.7. Bacterial protein expression	91
2.3.7.1. Protein expression and purification using the GST system	91
2.3.7.2. Inclusion body purification of pGEX fusion proteins	92

2.3.7.3. Protein expression and purification using the histidine tag system	93
2.3.8. Production of mono-specific polyclonal antibodies	93
2.3.9. Immunochemical techniques for the detection of proteins	94
2.3.9.1. Western blot analysis	94
2.3.9.2. Immunoprecipitation of proteins using polyclonal antibodies	95
2.3.9.3. Enzyme linked immunoadsorbent assay (ELISA)	96
2.3.10. Analysis of viral protein expression	97
2.3.10.1. ³⁵ S-methionine labelling of proteins in tissue culture	97
2.3.10.2. Quantification of ³⁵ S-methionine incorporation into protein	97
2.3.10.3. Subcellular fractionation using hypotonic buffer	98
2.3.10.4. Sucrose gradient analysis	98
2.3.11. <i>In vitro</i> protein expression	99
2.3.11.1 <i>In vitro</i> protein expression	99
2.3.11.2. Quantitation of protein produced using rabbit reticulocyte lysate	99
2.3.12. Autoradiography	99
2.3.13. Densitometry	100
Chapter 3. The Sequence of ORF1 and 2 of Ad5 early region 4	101
<u>3.1. Introduction</u>	102
<u>3.2. Cloning of the Ad5 DNA to be sequenced</u>	102
<u>3.3. Results</u>	103
<u>3.4. Discussion</u>	103
<u>3.5. Computer prediction and homology studies of ORF1 and ORF2</u>	107
3.5.1. Introduction	107
3.5.2. ORF1 structure predictions	107
3.5.3. ORF1 homology searches	112
3.5.4. ORF2 structure predictions	114

3.5.5. ORF2 homology searches	116
3.5.6. Conclusion	118
Chapter 4. The expression of the Ad5 E4 mRNAs	119
<u>4.1. Introduction</u>	120
<u>4.2. Methods</u>	122
4.2.1. Cells and viruses	122
4.2.2. The antisense RNA probes	124
<u>4.3. Results</u>	126
4.3.1. The temporal classes of E4 mRNA	126
4.3.2. The D3 to A3 intron is excised late in the infection	129
4.3.3. Relative abundance of the E4 mRNAs	129
4.3.4. Late E4 mRNAs are dependent on DNA replication for accumulation	132
4.3.5. Dependence of late E4 mRNA expression on the E1b-55K protein	132
4.3.6. E4 mRNA A/full length mRNA is dependent on E4 ORF6-34K for optimal cytoplasmic accumulation	136
4.3.7. E4 mRNA A/full length mRNA dependence on E1b-55K is at a post- processing level	136
<u>4.4. Discussion</u>	139
4.4.1. Classes of E4 mRNA	139
4.4.2. The expression of the E4 ORF products	142
4.4.3. Regulation of the expression of mRNA A and/or full length mRNA	144
Chapter 5. The construction and analysis of a virus containing a mutated E4 D1 donor site	146
<u>5.1. Introduction</u>	147
<u>5.2. Experimental strategy</u>	148
<u>5.3. Construction of the D1 and ORF6 mutant virus</u>	150
<u>5.4. Analysis of mRNA A/full length mRNA expression from the vD1 virus</u>	152

Chapter 6. The generation of a polyclonal antisera to E4 ORF1	157
<u>6.1. Introduction</u>	158
<u>6.2. Viruses and cells used</u>	158
<u>6.3. Expression of the ORF1 protein using the pGEX vector</u>	159
<u>6.4. The generation of an anti-ORF1 antiserum using the GST/ORF1 inclusion body preparation</u>	168
<u>6.5. The expression of ORF1 using the XPRESS expression system</u>	170
<u>6.6. The expression of ORF1 using the QIAexpress expression system</u>	174
<u>6.7. Attempts to detect ORF1 in infected cells</u>	176
<u>6.8. Discussion</u>	185
Chapter 7. The E4 ORF2 protein	189
<u>7.1. Introduction</u>	190
<u>7.2. Viruses and cells used</u>	190
<u>7.3. Expression of the ORF2 protein using the pGEX vector</u>	190
<u>7.4. The generation of an anti-ORF2 antiserum using the GST/ORF2 fusion protein</u>	197
<u>7.5. Expression of E4 ORF2 during a lytic infection</u>	200
7.5.1. The ORF2 protein exists in infected cells	200
7.5.2. The time course of expression of the ORF2 protein	200
7.5.3. Subcellular localisation of the ORF2 protein	204
7.5.4. Interactions of the ORF2 protein with other infected cell components	206
<u>7.6. Discussion</u>	209
Chapter 8. The function of the ORF1 and ORF2 proteins	213
<u>8.1. Introduction</u>	214
<u>8.2. The growth of adenovirus in non-transformed cells</u>	215
<u>8.3. Protein expression in ORF1 and ORF2 mutant viruses</u>	218

<u>8.4. E4 ORF1 and ORF2 mutant virions</u>	220
<u>8.5. Conclusion</u>	222
Chapter 9. General discussion	223
<u>9.1. Summary and interpretation of the results</u>	224
<u>9.2. Future work</u>	227
Chapter 10. References	231

List of figures

Fig. 1.1. Schematic representation of the adenovirus virion	4
Fig. 1.2. Transcription and translation map of adenovirus type 2	9
Fig. 1.3. The E1a region	11
Fig. 1.4. The predicted Ad5 E1b products	13
Fig. 1.5. The observed mRNA species from the E4 region with encoded ORFs	46
Fig. 1.6. The functions of the E4 ORF3-11K, E4 ORF6-34K and the E1b-55K proteins in the lytic viral cycle	58
Fig. 2.1. Diagram depicting site directed mutagenesis using PCR	73
Fig. 3.1. ORF1 and ORF2 sequence data depicting the regions of variation between the published Ad2, Ad5 and revised Ad5 sequences	104
Fig. 3.2. Comparison of the revised sequence of Ad5 E4 ORF1 with Ad5 (Chroboczek et al., 1992) and Ad2 (Roberts et al., 1984)	105
Fig. 3.3. Comparison of the revised sequence of Ad5 E4 ORF2 with Ad5 (Chroboczek et al., 1992) and Ad2 (Roberts et al., 1984)	105
Fig. 3.4. The structure of Ad5 ORF1 and ORF2	108
Fig. 3.5. Alignment of the predicted Ad5, Ad9, Ad12 and Ad34 E4 ORF1 protein sequences	110
Fig. 3.6. Predicted secondary structures of the Ad5, Ad9, Ad12 and Ad34 E4 ORF1 proteins	111
Fig. 3.7. Sequence alignment over 119 amino acids of Ad5 ORF1 and the <i>Saccharomyces cerevisiae</i> dUTPase	113
Fig. 3.8. Alignment of the predicted Ad5, Ad9, Ad12 and Ad40 E4 ORF2 protein sequences	115
Fig. 3.9. Predicted secondary structures of the Ad5, Ad9, Ad12 and Ad40 E4 ORF2 proteins	117
Fig. 4.1. E4 transcription map showing the location of the RNA probes	125
Fig. 4.2. Analysis of cytoplasmic levels of E1a, β -actin and E4 mRNAs during infection of HeLa cells by wild-type virus	127

Fig. 4.3. Analysis of cytoplasmic levels of E4 mRNAs in HeLa cells infected by wild-type dl309 or E1b-55K mutant dl338	130
Fig. 4.4. Analysis of E4 mRNAs expression in the presence and absence of an inhibitor of DNA replication	133
Fig. 4.5. Analysis of cytoplasmic levels of E1a and E4 mRNAs during infection of HeLa cells by wild-type dl309 or E1b-55K mutant virus dl338	134
Fig. 4.6. Analysis of cytoplasmic levels of E4 mRNAs in HeLa cells infected by wild-type dl309, E4 ORF6 mutant dl355, or E1b-55K/E4 ORF6 mutant dl367	137
Fig. 4.7. Analysis of cytoplasmic and nuclear levels of E4 mRNAs during infection of HeLa cells by wild-type dl309 or E1b-55K mutant virus dl338	138
Fig. 4.8. The E4 donor splice sites as compared to the eukaryotic consensus sequence of Mount (1982)	141
Fig. 5.1. The strategy for the detection of the mRNA A derived from the D1 and E4 ORF6 mutated virus vD1	149
Fig. 5.2. The sequence of the D1 site of dl309 at 35542 bp, the <i>Hind</i> III site in vD1 at 35543 bp and the mutated <i>Kpn</i> I site in the ORF6 reading frame of vD1	151
Fig. 5.3. Restriction digest analysis of the D1 mutant virus vD1 and dl309	153
Fig. 5.4. RNase protection analysis of dl309, dl355 and vD1 mRNA A/full length mRNA using a probe made from pID50	155
Fig. 6.1. The pGEX-2T expression vector	160
Fig. 6.2. PCR amplification of the ORF1 cDNA using primers IDC and IDD	160
Fig. 6.3. The construction of the pGEX-2T/ORF1 cDNA expression vector pID22	162
Fig. 6.4. Protein expression from the pGEX constructs	164
Fig. 6.5. The expression of the GST/ORF1N and GST/ORF1C fusion proteins	166
Fig. 6.6. The GST/ORF1 inclusion body preparation used in the generation	

of the first polyclonal antiserum	167
Fig. 6.7. Titration of the anti-GST/ORF1 activity in the serum samples by ELISA	169
Fig. 6.8. Western blot analysis of the GST/ORF1 polyclonal antisera	169
Fig. 6.9. The pTrcHis vector	171
Fig. 6.10. The expression of the His/ORF1 fusion at 37 °C from pID35	173
Fig. 6.11. SDS-PAGE analysis of samples of the 0.5 ml fractions from the nickel resin column used in the purification of the His/ORF1 fusion expressed from pID35	173
Fig. 6.12. The pQE-30 and pREP4 plasmids	175
Fig. 6.13. The expression of the His/ORF1 fusion at 37 °C from pID38	177
Fig. 6.14. SDS-PAGE analysis of samples of the 0.5 ml fractions from the nickel resin column used in the purification of the His/ORF1 fusion expressed from pID38	177
Fig. 6.15. Titration of the anti-His/ORF1 activity in serum samples by ELISA	179
Fig. 6.16. Western blot analysis employing sequential anti-His/ORF1 antisera to detect 400 ng of purified recombinant ORF1	179
Fig. 6.17. Detection of ORF1 in infected cells	181
Fig. 6.18. Immunoprecipitation of an <i>in vitro</i> transcribed/translated ORF1 product	182
Fig. 6.19. Western blot analysis employing the anti-His/orf1 antisera	184
Fig. 7.1. PCR amplification of the ORF2 cDNA using primers IDE and IDF	192
Fig. 7.2. The expression of the GST/ORF2 fusion protein from pID22 at 30°C and 37°C	193
Fig. 7.3. SDS-PAGE analysis of samples of the 0.5 ml fractions from the glutathione-sepharose column used for the purification of the GST/ORF2 fusion expressed from pID19	193
Fig. 7.4. SDS-PAGE analysis of 10 µg of glutathione sepharose-bound GST/ORF2 fusion protein cleaved with thrombin	195

Fig. 7.5. Purification of ORF2 protein released from the GST/ORF2 fusion by thrombin digestion	196
Fig. 7.6. Analysis of the GST/ORF antisera	198
Fig. 7.7. Western blot analysis employing sequential anti-GST/ORF2 antisera to detect 40 ng of purified recombinant ORF2	199
Fig. 7.8. Detection of ORF2 in infected cells	201
Fig. 7.9. Steady state levels of the ORF2 protein during the course of an infection	202
Fig. 7.10. Synthesis of the ORF2 protein during the course of an infection	203
Fig. 7.11. The subcellular localisation of the ORF2 protein within the infected cell	205
Fig. 7.12. Determination of the size of the ORF2 protein by sucrose gradient analysis	207
Fig. 7.13. Co-immunoprecipitation analysis to detect ORF2 protein-protein interactions	208
Fig. 7.14. Immunoprecipitation analysis of the ORF2 protein under reducing and non-reducing conditions	210
Fig. 8.1. Analysis of the protein production of wt300, dl1-3, dl327, dl351 and dl352 during the course of an infection	219
Fig. 8.2. Analysis of the protein content of dl327, dl351 and dl352 virions	221

List of tables

Table 3.1. The DNA sequence variations found in Ad5 (Chroboczek <i>et al.</i> , 1992), Ad5 (Dix and Leppard, 1992) and Ad2 (Roberts <i>et al.</i> , 1984) E4 ORF1 and ORF2	106
Table 3.2. Amino acid changes due to the nucleotide differences between Ad2 and Ad5 ORF1 and ORF2	106
Table 5.1. Classification of E4 mRNAs	131
Table 8.1. The titres of wt300 and dl1-3 virus, 4 days after infection of HeLa, MRC5, WI-38, HUV-EC-C and HISM cells	216
Table 8.2. The titres of dl327, dl351, and dl352, 4 days after infection of HeLa and WI-38 cells	217

Acknowledgements

First, I would like to thank the members of both the Adenovirus and Rotavirus groups at the University of Warwick for all their assistance during the course of this study. I am particularly grateful to Cally Caravokyri and Sue Thomas for their advice, help and support over the last three years. I would also like to thank Carol Hill for her assistance in all the animal work contained within this study. I especially want to thank Keith Leppard, my supervisor, who has given me encouragement and invaluable guidance during this period, to whom I will always be indebted.

I also would like to thank my parents for their love and support over the last 25 years and for always encouraging me to 'get an education'. However special thanks must go to Amanda Young, my wife to be, for putting up with me over the last three years. I particularly thank her for her love, understanding and support, especially when experiments 'crashed and burned'! Finally, I thank the Lord Jesus Christ for the strength and hope that He alone can give.

This work was supported by a grant from the Science and Engineering Research Council.

Declaration

All the work presented in this thesis was performed by the author in the Biological Sciences Department of the University of Warwick, under the supervision of Dr. K.N. Leppard and with the assistance of the University of Warwick technical staff in the handling of animals and the synthesis of oligonucleotides. None of this material has been previously submitted for examination at another institution. However the studies reported in chapter 4 derive from preliminary work performed by the author as part of an undergraduate honours project and submitted in accordance with the requirement of the BSc degree, 1991, University of Warwick. None of the data generated in this preliminary work is presented here.

During the course of this study two publications have been made by the author. The first was in 1992 and contained data from chapter 3:

Dix, I., & Leppard, K.N. (1992). Open reading frames 1 and 2 of adenovirus region E4 are conserved between human serotypes 2 and 5. *J. Gen. Virol.* **73**, 2975-2976.

The second was in 1993 and contained data from chapter 4:

Dix, I., & Leppard, K.N. (1993). Regulated splicing of adenovirus type 5 E4 transcripts and regulated cytoplasmic accumulation of E4 mRNA. *J. Virol.* **67**, 3226-3231.

Abbreviations

General abbreviations

³⁵S-met	³⁵ S-methionine
A₂₆₀	absorbance at 260 nm
Ab	antibody
Ad(5) (2) (9) (12) (34) (40)	adenovirus serotype (5) (2) (9) (12) (34) (40)
Ag	antigen
Amp	ampicillin
ATP	adenosine 5'-triphosphate
bp	base pair
BSA	bovine serum albumin
Ci	curie
cpm	counts per minute
CTP	cytidine 5'-triphosphate
CS	newborn calf serum
Da	Dalton
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
ddATP	dideoxyadenosine triphosphate
ddCTP	dideoxycytidine triphosphate
ddGTP	dideoxyguanosine triphosphate
ddNTP	dideoxynucleoside triphosphate
ddTTP	dideoxythymidine triphosphate
DMEM	Dulbecco's modified Eagle medium
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxynucleoside triphosphate

DOC	deoxycholate
ds	double-stranded (DNA or RNA)
dTTP	deoxythymidine triphosphate
DTT	dithiothreitol
dUMP	deoxyuridine monophosphate
dUTP	deoxyuridine triphosphate
dUTPase	deoxyuridine triphosphate nucleotidohydrolase
EDTA	ethylenediaminetetraacetic acid (disodium salt)
ELISA	enzyme-linked immunosorbent assay
FCS	fetal calf serum
GST	glutathione <i>S</i> -transferase
GTP	guanosine 5'-triphosphate
hnRNP	heterogeneous nuclear ribonucleoprotein
Ig	immunoglobulin
IPTG	isopropyl-1-thio- β -D-galactosidase
Kan	kanamycin
KLH	keyhole limpet haemocyanin
MCS	multiple cloning site
MLP	Adenovirus major late promoter
MLTU	Adenovirus major late transcription unit
MOI	multiplicity of infection
nt	nucleotides
NP40	Nonidet P40
oligo	oligonucleotide
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEG	polyethylene glycol
pfu	plaque-forming unit

PMSF	phenylmethylsulphonyl fluoride
poly(A)	polyadenylate
PP_i	pyrophosphate
RBS	ribosome binding site
RF	replicative form
RNA	ribonucleic acid (prefixes: m = messenger, r = ribosomal, sn = small nuclear, t = transfer)
RNase	ribonuclease
rpm	revolutions per minute
S	Svedberg unit
SDS	sodium dodecyl sulphate
ss	single stranded (DNA or RNA)
TCA	trichloroacetic acid
TEMED	<i>N,N,N',N'</i> -tetramethyl-ethylenediamine
Tris	Tris(hydroxymethyl)aminomethane
UTP	uridine 5'-triphosphate
UTR	untranslated region
uv	ultraviolet light
Xgal	5-bromo-4-chloro-3-indolyl- β -D-galactoside

Virus genome abbreviations

In adenovirus, genes and proteins are named according to the early region (E1a, E1b, E2, E3, E4) major late region (L1, L2, L3, L4, L5), or minor late region (IVa2, IX) of the genome from which they are expressed (Fig. 1.2.). This is normally followed by either the open reading frame number (e.g. ORF1), or the molecular weight of the of the protein (e.g. E1b-55K), or the number of amino acids in the protein (e.g. E1a-243R), or if the protein is a virion component a Roman numeral (e.g. L3 II = hexon protein). All adenovirus genome numbering in this thesis refers to the published

sequence of Chroboczek et al., 1992 (Gen Bank acc. no. M73260) amended by Dix and Leppard, 1992 (Gen Bank acc. no. D12587) unless otherwise stated.

Summary

Human adenovirus type 5 is a dsDNA virus which replicates in the nucleus of the infected cell, exploiting a number of host cell mechanisms. This close association with the eukaryotic cell has made adenovirus the target of numerous studies attempting to understand how cellular systems function. This study focuses on the E4 transcription unit, which has the potential to encode at least 7 distinct polypeptides from reading frames accessed by differential splicing of a single primary transcript. In this study, the pattern of expression of these mRNAs during lytic infection was examined, and two distinct temporal classes were defined; early and late. It had been previously shown that adenovirus mRNAs produced late in the infection depended on a virus-coded RNA transport regulator, E1b-55K, for optimal cytoplasmic accumulation. However, only one of the E4 late class mRNAs was dependent on this E1b protein for cytoplasmic accumulation leading to the hypothesis that for an E4 late mRNA to be dependent on E1b, it had to retain intact splice sites or intronic sequences. To examine this hypothesis, a virus was built lacking an important splice site of the E4 region to see if, by removing this splice site, the mRNA could leave the nucleus in the absence of the E1b complex. The results of initial experiments reported here supported this hypothesis.

Two of the E4 open reading frames (ORF1 and ORF2) identified in Ad2 were disrupted in the published Ad5 E4 sequence, but these differences were subsequently found to be sequencing artefacts. The presence of these two proteins in the infected cell had never been previously demonstrated so polyclonal antisera were generated against bacterially expressed ORF1 and ORF2. The ORF2 antiserum allowed the identification of the ORF2 protein in the cytoplasm of infected cells, from early stages of the infection. No associations of ORF2 with other infected cell components were detected. In contrast, the ORF1 antiserum only reached a low titre and no ORF1 protein was detected in infected cells. Now that ORF2 has been found in the infected cell more work needs to be undertaken to elucidate its function.

Chapter 1.

Introduction

1.1. Adenoviruses - A General Introduction

1.1.1. The adenovirus family

Adenoviruses were first isolated from tonsil and adenoid biopsies from children by Rowe *et al.* (1953) Since then adenoviruses have been shown to be the aetiological agents of a wide variety of diseases in birds and mammals including a number of acute febrile respiratory illnesses, epidemic keratoconjunctivitis, acute haemorrhagic cystitis and gastro-enteritis (reviewed in Horwitz, 1990b). Some adenoviruses have also been associated with the induction of malignant tumours in rodents although there is no evidence of these viruses having oncogenic potential in humans.

Adenoviruses are of the family *Adenoviridae* which is divided into two genera; *Mastadenoviridae* and *Aviadenoviridae*. The *Mastadenoviridae* genus contains human, simian, bovine, equine, porcine, ovine and canine viruses. The human adenoviruses have been classified into six subgroups (A-F) by a number of serological and biochemical techniques such as haemagglutination patterns of rat and rhesus monkey red blood cells, oncogenic potential in rodents, DNA base composition and homology, and the ability to transform cells in tissue culture. The antigenic determinants important in the classification of adenoviruses are found on the hexon, penton and fibre proteins, for example all human adenoviruses share a cross-reacting group antigen (α) on the hexon capsomer while the type-specific antigens (ϵ and γ) are found on the hexon and fibre proteins respectively (reviewed in Horwitz, 1990a).

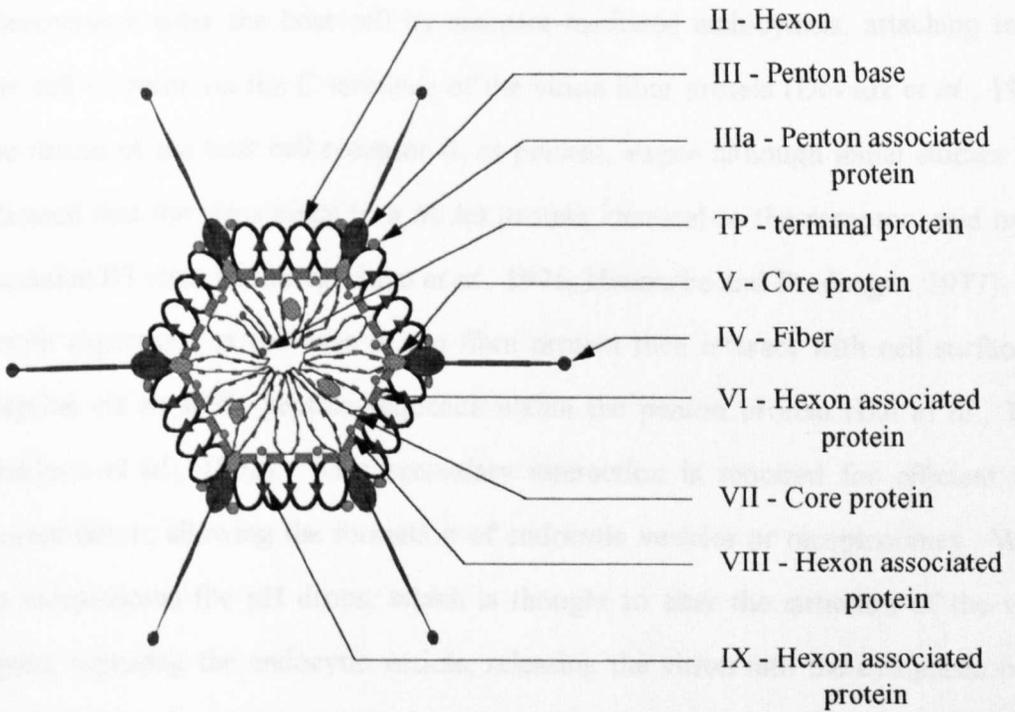
This study focuses on a human subgroup C virus, adenovirus serotype 5 (Ad5). Subgroup C viruses are characterised by their partial agglutination of rat erythrocytes, their inability to induce tumours in rodents, their ability to transform cells in tissue culture and the high G+C content (57-59 %) of their genomes. Of the subgroup C viruses isolated, serotypes 2 and 5 have been studied in greatest detail. The genomes of these viruses have been completely sequenced, Ad2 in 1984 by Roberts *et al.* and Ad5 in 1992 by Chroboczek *et al.*, amended by Dix and Leppard (1992). Sequence analysis shows 94.7 % homology between the two viruses, with virtually identical gene

organisation. The differences between the sequences of these two viruses are concentrated in regions encoding hexon, fibre and some of the E3 proteins (Kinloch *et al.*, 1984; Cladaras and Wold, 1985; Chroboczek and Jacrot, 1987), probably due to selective pressure exerted by the host immune system during infection as the hexon and fibre proteins contain major antigenic determinants. As these two serotypes are so closely related, data derived from one or other of these viruses is routinely taken to apply to both serotypes; in the introduction to this study no distinction between the two serotypes is usually made.

1.1.2. Structure of adenoviruses (Fig. 1.1.)

The structure of the adenovirus virion has been well characterised and is reviewed by Horwitz (1990a). Adenovirus virions are non-enveloped, regular icosahedrons of about 65 to 80 nm in diameter which comprise a double stranded DNA molecule of about 36,000 bp (Ad5 is 35938 bp) surrounded by a protein shell (capsid) made of at least 11 distinct structural proteins (figure 1.1). The capsid comprises 240 hexon capsomeres (3 x 120 kd polypeptide II molecules) and 12 penton capsomeres (5 x 85 kd polypeptide III molecules noncovalently attached to 3 x 62 kd polypeptide IV molecules; the fibre structure). The penton capsomeres are located at each of the 12 vertices of the capsid with the fibre structure projecting out from the virion. Each of these penton capsomeres is surrounded by five hexon capsomeres (the peripentonal hexons) associated with polypeptide IIIa (66kd). Also associated with the hexon capsomeres are 3 further polypeptides; VI (24 kd), VIII (13 kd) and IX (12 kd), which are probably involved in stabilising the capsid structure. Within the capsid shell of the virion is the core structure which contains the DNA genome covalently attached to the 55 kd terminal protein (TP) at its 5' ends and polypeptides V (48.5 kd), VII (18.5 kd) and μ (4 kd). Polypeptides V and VII are basic proteins which form irregular nucleosome-like structures along the viral genome, each spanning approximately 200 bp; the μ protein has no assigned function.

Figure 1.1. Schematic representation of the adenovirus virion (reproduced from Horowitz, 1990a). The μ protein and the probable virion protein, L3-23K protease, are not featured as their location within the virion is not known. The thread-like strand within the virion shown below is representative of the virion genome.



1.1.3. Adenovirus replication (reviewed by Horwitz, 1990a)

The adenovirus lytic cycle in cultured cells results in the production of about 10,000 virions per infected cell during the course of one virus life cycle (ca. 36 hours). The lytic life cycle of adenovirus has been divided, by convention, into two phases, early and late, separated by the onset of viral DNA replication, with distinct patterns of gene expression in each phase. This division of the virus life cycle is not rigid, early mRNAs are also made after vDNA replication has begun and some late mRNAs can be found prior to vDNA replication.

Adenoviruses enter the host cell by receptor-mediated endocytosis, attaching to the host cell receptor via the C terminus of the virion fiber protein (Devaux *et al.*, 1987). The nature of the host cell receptor is, at present, vague although initial studies have indicated that the virus binds to a 50 kd protein identical to the receptor used by the coxsackie B3 virus (Lonberg-Holm *et al.*, 1976; Hennache and Boulanger, 1977). The penton capsomers at the base of the fibre protein then interact with cell surface α_v integrins via an RGD peptide sequence within the penton protein (Bai *et al.*, 1993; Wickham *et al.*, 1993). This secondary interaction is required for efficient virus internalisation, allowing the formation of endocytic vesicles or receptosomes. Within the receptosome the pH drops, which is thought to alter the structure of the virion capsid, rupturing the endocytic vesicle, releasing the virion into the cytoplasm of the cell (Seth *et al.*, 1985). It has been postulated that the penton protein plays an important role in the rupture of the receptosome, possibly through the interaction with cellular integrins (Seth *et al.*, 1984). The penton capsomere is believed to be lost during this process (reviewed by Nemerow *et al.*, 1994).

The virion is then transported through the cytoplasm to the nucleus of the infected cell in a process which probably involves the interaction of the hexon capsomeres with the microtubules of the host cell. The viral DNA and the core proteins (V, VII, TP and μ) then enter the nucleus through the nuclear pores leaving the remainder of the virion proteins in the cytoplasm.

The viral genome then undergoes early transcription. The virus appears to rely almost exclusively upon the host cell machinery for transcription and RNA processing with transcripts being spliced, capped and polyadenylated in a manner similar to host transcripts (Ziff, 1980; Darnell, 1982; Tooze, 1982; Sharp, 1984). Transcription from the six early transcription units (E1a, E1b, E2, E3, E4 and L1) results in the synthesis of a variety of mRNAs which encode proteins required for subsequent events in the lytic cycle, such as transcriptional activation, viral DNA replication and mRNA metabolism. Early transcription from these transcription units results in the production of only low levels of viral mRNA, compared to the level of viral mRNAs late in the infection.

The late phase of the infection is marked by the onset of DNA replication, about 10-12 hours post infection. Viral DNA replication has been extensively studied, the current model incorporating both type I (duplex template) and type II (single stranded template) replication to produce progeny genomes. Replication initiates at either end of the parental duplex molecule, displacing one of the parental strands to produce a semi-conservatively replicated genome and a single stranded molecule (type I replication). The inverted terminal repeats of the genome allow the single-stranded molecule to form a panhandle structure where the double-stranded region resembles the end of a duplex genome. This is then replicated by a type II replication mechanism to produce a second daughter genome (reviewed by Stillman, 1989). *In vivo* and *in vitro* studies have demonstrated that only three viral proteins are directly involved in DNA replication; the 140 kd viral DNA polymerase, the 72 kd ssDNA binding protein, and the 80K precursor to the terminal protein (reviewed by Stillman, 1989). Recently it has been demonstrated that the ORF3 and ORF6 proteins of the E4 region are implicated in the regulation of vDNA replication (Weiden and Ginsberg, 1994, see section 1.4.2.3.). Three host proteins are also important in viral DNA replication; nuclear factors I and III, transcription factors which bind adjacent to the core origin of replication; and nuclear factor II, a type I topoisomerase, (reviewed by Stillman , 1989).

Following the onset of viral DNA replication the nuclear structure is reorganised with both DNA replication and late transcription occurring in association with the nuclear matrix in discrete viral inclusion bodies (Moyné *et al.*, 1978; Zhonghe *et al.*, 1986; Walton *et al.*, 1989; Moen *et al.*, 1990; Jiménez-García and Spector, 1993). Transcription occurring in these inclusion bodies results in the production of late mRNAs, primarily encoded by the major late transcription unit (MLTU). These late mRNAs, which mainly encode the structural proteins of the virion, can account for up to 20-40% of the total cellular mRNAs (Flint, 1986). As the structural proteins are produced, the progeny virions are assembled in the nucleus of the infected cell; initially the penton and hexon polypeptides assemble into capsomeres in the cytoplasm and then form a 600S capsid intermediate within the nucleus. At this stage the viral DNA enters the capsid along with the core proteins, by virtue of packaging signals located between 290 and 390 bp from the left end of the genome, at an opening at one of the vertices. The last stage in virion assembly is the processing of some of the proteins (pVI, pVII, pVIII, pTP) within the virion, which are assembled in precursor form, by the viral protease (L3-23 kd) to produce the mature virion. This entire process is aided by virus-encoded scaffolding proteins which are necessary during virion formation but are absent from the virion (e.g. L4-100 kd and IVa2-50 kd). The mechanism of virion release from the cell has not yet been clearly elucidated (reviewed by Horwitz *et al.*, 1990a).

Cellular macromolecular synthesis undergoes major changes during the late phase of an adenovirus infection, with both cellular DNA replication and protein synthesis being inhibited. The gradual cessation of cellular DNA replication is probably a secondary effect of the inhibition of cellular protein synthesis. The block in cellular protein synthesis has been shown to be at the level of translation and not at the level of mRNA transcription or mRNA transport into the cytoplasm as cytoplasmic levels of cellular mRNAs are still high when host protein synthesis is shut off (see section 1.2.6.)

1.1.4. Adenovirus gene function

Gene expression from the 36 kbp adenovirus genome is a very complicated process, with transcription of most of both strands of the double stranded genome as can be seen in Fig. 1.2. (reviewed by Flint *et al.*, 1986 and Horwitz, 1990a). The regulation of this expression is discussed in section 1.2..

1.1.4.1. The E1a region

The E1a transcription unit is encoded on the r strand of the genome. Five mRNAs are derived from this region by the differential splicing of the primary transcript: two predominant early mRNAs (12S and 13S); a predominant late mRNA (9S); and two minor late mRNAs (10S and 11S) (Fig. 1.3a.). The two major products from this region are encoded by the 12S and 13S mRNAs; the 243R and 289R proteins, which are identical apart from 46 internal residues. Both these proteins contain discrete regions conserved among the different adenovirus serotypes, the CR1 and CR2 regions, while only 289R contains a third conserved region, CR3. Both proteins also contain two auxiliary regions (AR1 and AR2), shown to be important in functional studies in addition to CR1-3, and a C terminal nuclear localisation signal (Fig. 1.3b).

The 243R and 289R proteins of E1a are involved in the induction of cellular DNA synthesis which is related to their role in transformation of non-permissive cells. The mechanism of action of these proteins in cellular transformation is not fully understood but it is known that they form complexes, via the CR1 and CR2 regions, with cellular proteins involved in cell cycle control, including the tumour repressor protein p105^{Rb} and other related proteins (reviewed by Dyson and Harlow, 1992).

The CR3 domain of 289R is involved in the transactivation of E1a-dependent promoters. The E1a protein activates transcription by a number of different mechanisms (reviewed by Akusjarvi, 1993): it regulates certain transcription factors (e.g. E2F, E4F and TFIIC) by phosphorylation (Hoeffler *et al.*, 1988; Bagchi *et al.*, 1989; Raychaudhuri *et al.*, 1989); it can bind upstream activating sequence (UAS)-binding transcription factors bringing its own transcriptional activation domain into

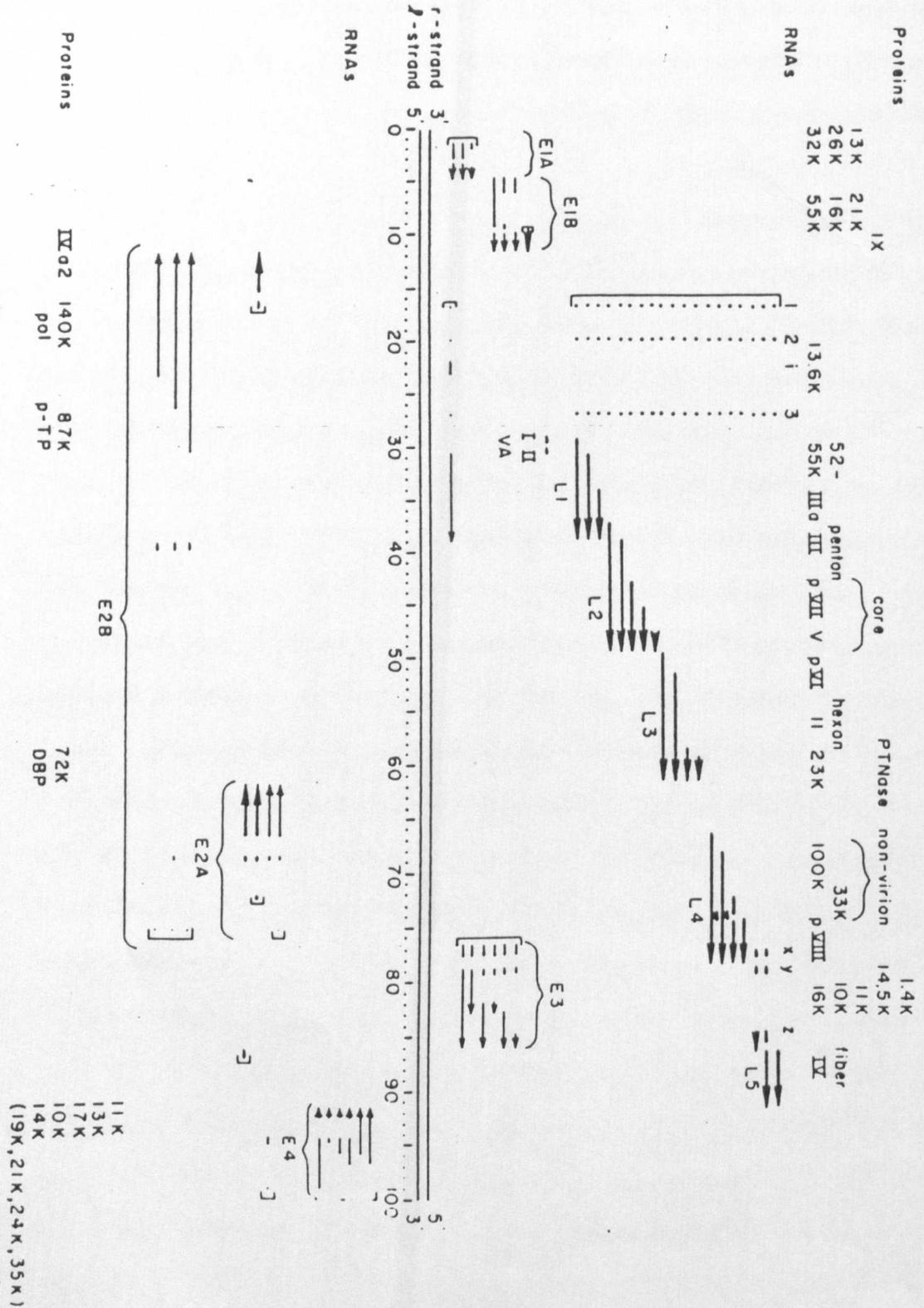


Figure 1.2. Transcription and translation map of adenovirus type 2 (reproduced from Horwitz, 1990a). Early mRNAs are represented by thin lines, late mRNAs (and early mRNAs which are transcribed late, e.g. L1 transcripts) by bold lines, polyadenylation sites by arrow heads, nonstructural gene products by their molecular mass, and virion components by Roman numerals. The three components of the major late tripartite leader are designated 1, 2 and 3. Some minor mRNA species and corresponding polypeptides are not shown at this scale.

proximity with the core promoter proteins (Liu and Green, 1990; Chatton *et al.*, 1993); E1a can also interact directly with the TBP (TATA Box binding protein) activating transcription (Horikoshi *et al.*, 1991; Lee *et al.*, 1991).

The CR1 and CR2 domains of 243R and 289R can also transactivate transcription via their interaction with tumour suppressor proteins, such as p105^{Rb}, as this interaction can cause the dissociation of transcription factors from these regulatory, inhibitor proteins (Bandara and La Thangue, 1991; Chellappan *et al.*, 1991). A well studied example of this phenomenon is the E2F-p105^{Rb} interaction, where E2F has no transactivating function when bound to p105^{Rb} but in the presence of the E1a protein it is released in an active form, as E1a interacts with p105^{Rb} displacing E2F.

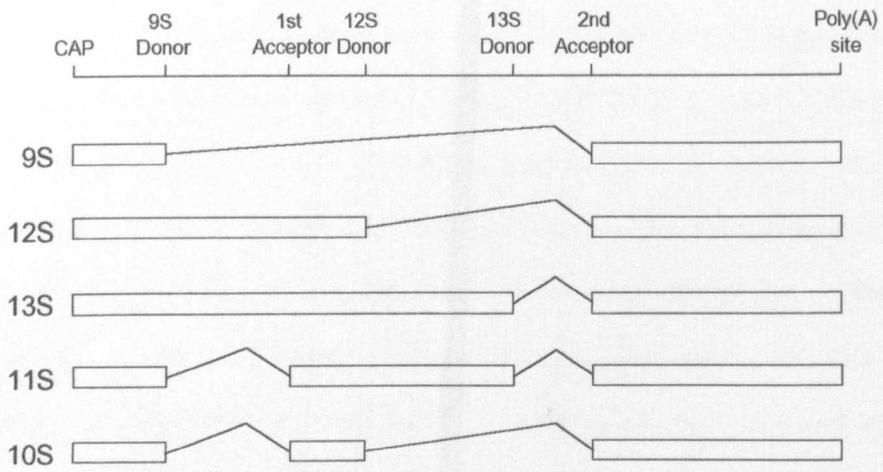
The multifunctional E1a proteins may therefore regulate one promoter in a number of different ways, e.g. the E2 early promoter. The E2 promoter consists of an activating transcription factor (ATF) binding site and two inverted E2F sites. It is activated by E1a by three distinct mechanisms: ATF can bind to the upstream ATF binding site and transactivate the promoter via E1a which associates with ATF; E1a can dissociate E2F from p105^{Rb}-E2F complexes where E2F is inactive and it can also phosphorylate E2F, so increasing its DNA binding activity and so its ability to transactivate the E2 promoter (Bagchi *et al.*, 1989; Bandara and La Thangue, 1991; Chellappan *et al.*, 1991; Chatton *et al.*, 1993).

Evidence is mounting that E1a-243R can also inhibit the activation of some promoters, a function shown to be a property of the N terminus of the protein (Berk, 1986). The E1a-243R protein, which lacks the transactivation domain CR3, is believed to repress transcription by interacting with a TFIID protein from the core of the transcription machinery and so reducing expression from certain promoters (Jelsma *et al.*, 1989; Stein *et al.*, 1990; Bautista *et al.*, 1991).

1.1.4.2. The E1b region

The E1b transcription unit is also on the r strand of the genome and encodes proteins important in virus replication and cellular transformation. At least 5 mRNAs are

A



B

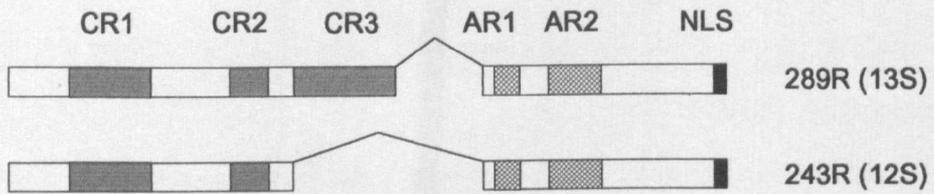


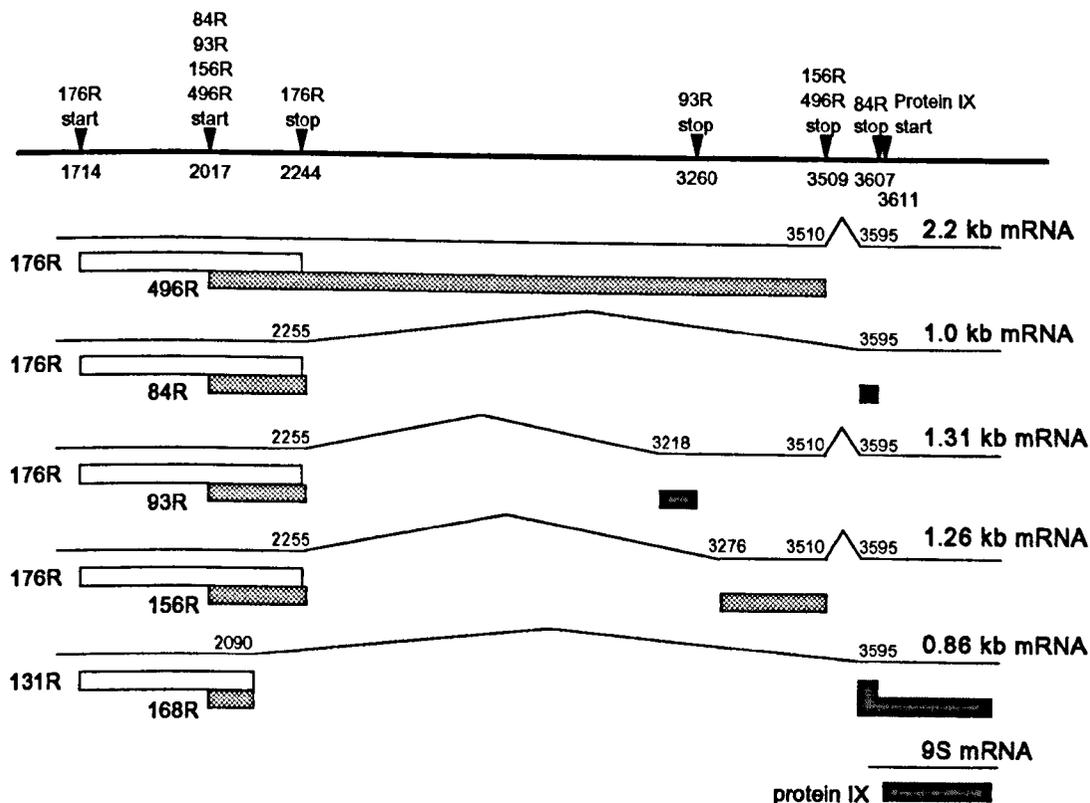
Figure 1.3. The E1a region. Panel A: The E1a mRNAs (from Harper and Manley, 1991). The open boxes represent exon regions while the lines represent intron regions. Indicated above the mRNAs are the transcriptional start sites, splice sites and polyadenylation sites. Indicated to the left of the mRNAs are the popular names given to the mRNAs. **Panel B: The multifunctional domains of the major E1a proteins, 289R and 243R.** 289R and 243R are the products of the 12 and 13S mRNAs respectively. CR1, CR2 and CR3 are regions conserved in E1a proteins of all the adenovirus serotypes. AR1 and AR2 are auxiliary regions shown to be functionally important in addition to the CR regions. NLS is the nuclear localisation signal.

derived from this region (not including the pIX mRNA) which potentially encode 7 distinct proteins (Perricaudet *et al.*, 1979; Bos *et al.*, 1981; Anderson *et al.*, 1984; Virtanen and Petterson, 1985; Lewis and Anderson, 1987; Takayesu *et al.*, 1994, see Fig. 1.4.). A 2.2kb mRNA is predominant early in the infection and is translated to produce the major E1b proteins 19K (176R) and 55K (496R), the E1b-55K protein being translated from a reading frame overlapping that of the 19K protein, initiating at a site within the 19K reading frame. Late in the infection a 1.0 kb mRNA accumulates, due to differential splicing, encoding the 19K protein and an 84 residue protein (84R) which comprises the N- and C- termini of the 55K protein. Also produced at late times are three less abundant mRNAs of 0.86 kb, 1.26 and 1.31 kb producing various products related to the major E1b products.

The 19K protein is involved in a number of different aspects of the viral infection including the protection of viral DNA, and in non-permissive cells, cellular transformation (Chinnadurai, 1983; Subramanian *et al.*, 1984; Stillman, 1986; White and Stillman, 1987). The 19K protein also protects against programmed cell death (apoptosis), induced as a result of the disruption of cell growth control pathways by E1a (White *et al.*, 1991; 1992; Rao *et al.*, 1992; Debbas and White, 1993), and cytolysis induced by the tumour necrosis factor- α (White *et al.*, 1992). The 19K protein may also have the ability to transactivate early viral genes (Yoshida *et al.*, 1987), although no transcriptional activity was found by Telling *et al.* (1994).

The 55K protein is also a multifunctional protein being involved in cellular transformation (Barker and Berk, 1987; McLorie *et al.*, 1991) and in the cytoplasmic accumulation of late mRNAs, host cell shutoff and late protein synthesis during lytic infection (Babiss and Ginsberg, 1984; Babiss *et al.*, 1985; Pilder *et al.*, 1986a,b; Williams *et al.*, 1986). The 55K protein derives some or all of its transforming activity from its interaction with the tumour suppressor protein p53 (Sarnow *et al.*, 1982; Kao *et al.*, 1990), inhibiting p53-mediated transcriptional activation (Yew and Berk, 1992). Malette *et al.* (1983) demonstrated that the 55K protein is phosphorylated at three distinct sites. Subsequent studies have shown that this

Figure 1.4. The predicted Ad5 E1b products (reproduced from Takayesu et al., 1994). The E1b mRNAs are depicted as lines, with the protein products indicated by boxes. The 3 translation reading frames are shown by different shading. Also included are the relevant nucleotides for translation start and stop sites and for splice start and donors.



phosphorylation is important in regulating 55K function, especially in transformation (Teodoro *et al.*, 1994). Although the E1b-55K protein is multifunctional, the viral replication and transforming functions are not divided into discrete functional domains, as in the E1A proteins, although the two functions can be separated by mutational analysis (Yew *et al.*, 1990). The function of the E1b-55K protein in viral replication is dealt with in detail in section 1.3. The functions of the minor E1b proteins has not yet been established.

As mentioned, both the E1a and the E1b proteins are involved in the immortalisation and transformation of primary rodent cells infected with adenovirus, however an extensive review of the literature reporting these functions is outside the scope of this thesis. In addition to the above references the subject is reviewed by Horwitz *et al.* (1990a) and Boulanger and Blair (1991).

1.1.4.3. The E2 region

The E2a region encodes a nuclear phosphoprotein with an apparent molecular weight of 72 kDa (DBP) which is essential in viral DNA replication, binding non-specifically to single-stranded and double-stranded DNA changing its structure (reviewed by Hay and Russell, 1989; Stillman 1989; Horwitz *et al.*, 1990a). It has also been demonstrated to be involved in the regulation of transcription (Nevins and Winkler, 1980; Johnston *et al.*, 1985), mRNA stability (Babich and Nevins, 1981), mRNA processing (see section 1.2.4.) and in virus assembly (Nicolas *et al.*, 1983).

The E2b region encodes two phosphoproteins; the 80 kDa pre-terminal protein (pTP) and the 140 kDa DNA polymerase. The DNA polymerase catalyses the covalent attachment of dCMP to a serine residue of the pTP protein in the presence of specific DNA residues at the origin of replication. This acts as a novel priming mechanism for DNA replication (reviewed by Hay and Russell, 1989; Stillman, 1989). Later in the infection the protein is cleaved by the L3-23K viral protease to give the 55K form (TP). The pTP/TP has a number of functions; it attaches the DNA to the nuclear

matrix so allowing efficient transcription and possibly DNA replication (Shaack *et al.*, 1990; Fredman and Engler, 1993), it protects the DNA ends from exonuclease activity (Dunsworth-Browne *et al.*, 1980), it prevents the attachment of DNA termini binding proteins which inhibit DNA replication (DeVries *et al.*, 1989) and it ensures the correct initiation site for DNA replication (reviewed by Hay and Russell, 1989; Stillman, 1989).

The viral DNA polymerase is involved in the initiation and elongation of the nascent DNA strand and is probably regulated by phosphorylation (reviewed by Hay and Russell, 1989; Stillman, 1989).

1.1.4.4. The E3 region

The E3 region is not required for viral replication in tissue culture, being involved in the virus' ability to circumvent immune surveillance in the natural host (reviewed by Wold and Gooding, 1991). The Ad5 E3 region expresses approximately 9 alternatively spliced and polyadenylated mRNAs coding for at least seven proteins; gp19K, 14.7K, 14.5K, 12.5K, 11.6K, 10.4K, and 6.7K. The 14.7K, 14.5K and the 10.4K proteins prevent tumour necrosis factor (TNF)-mediated cytolysis of the infected cell. The 10.4K and 14.5K proteins also form a complex which down-regulates the cell surface epidermal growth factor receptor levels although it is unclear whether this function is linked to the proteins' ability to prevent TNF-mediated cytolysis. Recently the 10.4K and 14.5K proteins have been associated with the down regulation of the expression of E1a gene products, at the level of translation (Zhang *et al.*, 1991, 1994). This reduction in E1a gene expression is believed to cause a reduction in cytolysis by adenovirus-specific T cells (Zhang *et al.*, 1991). The gp19K binds to the major histocompatibility complex (MHC) class I antigens and retains them in the endoplasmic reticulum so preventing their transport to the cell surface. This failure of virus-infected cells to present MHC class I antigens, and consequently virus peptide antigens, on their cell surface results in virus specific cytotoxic T-lymphocytes

of the host immune system failing to recognise and lyse virus-infected cells. No functions have yet been ascribed to the 12.5K, 11.6K or the 6.7K proteins.

1.1.4.5. E4 region

The E4 region encodes potentially seven different polypeptides; E4-ORF1, ORF2, ORF3, ORF3/4, ORF4, ORF6, and ORF6/7. The expression and functions of these proteins will be discussed in detail in section 1.4.

1.1.4.6. Late transcription units

Late in the infection the major late promoter (MLP) is strongly activated to produce 5 families of late transcripts (L1-L5) by alternative splicing and polyadenylation. Every RNA from the MLTU contains an identical 5' non-coding region of about 200 nucleotides called the tripartite leader which is believed to be important in the translation of MLTU transcripts (see section 1.2.6.). Most of the late proteins are encoded by these MLTU transcripts although the minor late polypeptides IX and IVa2 are encoded elsewhere on the genome. The expression of the late products is restricted to the period after viral DNA replication with the exception of L1 mRNA which is also expressed at low levels during the early phase. In addition to structural proteins, the MLTU encodes a number of late, nonstructural proteins such as 'scaffolding' proteins, involved in virion assembly (e.g. L1-55K, L4-100K), and other nonstructural proteins (e.g. L4-33K).

Three other late-specific transcription units are also active late in the infection, IVa2, IX and E2-L. The IVa2 and IX transcription units are often referred to as the intermediate transcription units as they are transcribed before late MLTU transcription, at the end of the early phase. The IVa2 region encodes a 50K scaffolding protein while region IX encodes a 12K virion protein. The activation of the E2-L promoter at late times results in an increase in the levels of the viral proteins required for DNA replication.

1.1.4.7. VA RNA

Adenoviruses also encode two distinct virus-associated (VA) RNAs; VA RNA_I and VA RNA_{II}, each about 160 nucleotides long, transcribed by RNA pol III (Price and Penman, 1972; Weinmann *et al.*, 1974; Mathews, 1975; Pettersson and Philipson, 1975). These RNAs are initially synthesised early in the infection but production increases at late times and has been shown to be necessary for efficient viral mRNA translation (see section 1.2.6.).

1.2. The regulation of gene expression in an adenovirus infection

1.2.1 Introduction

Adenovirus genes are expressed in a highly regulated fashion during the course of a lytic infection. This regulation of gene expression results in a clearly defined temporal expression pattern of the gene products so that they are produced at the required times and at the required levels to ensure the efficient replication of the virus. This highly regulated pattern of gene expression has been the subject of intense research, in an attempt to elucidate the mechanisms by which the virus controls gene expression. Studies have shown that adenovirus gene expression is regulated both at the level of transcription and at post-transcriptional levels. Details of these studies are summarised below.

1.2.2. Transcription initiation (reviewed by Akusjarvi, 1993)

Adenovirus genes are transcribed by two host cell, DNA-dependent RNA polymerases; RNA polymerase II which transcribes both strands of the virus producing over 99% of the viral mRNAs, and RNA polymerase III which transcribes the short, noncoding RNAs known as VA RNA (sections 1.1.4.7. and 1.2.6.2.). Transcription from the adenovirus genome occurs from five early (E1a, E1b, E2, E3 and E4) and four late (MLTU, IX, IVa2 and E2-L) promoters. These genes can be subdivided further into immediate early (E1a), delayed early (E1b, E2, E3 and E4) and intermediate (IVa2,

IX) genes according to when transcription initiates in the virus lytic cycle. This temporal regulation is not fully understood but a number of viral proteins have been demonstrated to be involved in the formation and the activity of transcription complexes within adenovirus infected cells, so regulating gene expression. Firstly, E1a transactivates E1a, E1b, E2, E3, E4, VA RNA and a limited number of cellular promoters such as *hsp70*, *c-fos*, *c-jun* and proliferating cell nuclear antigen (PCNA). This transactivation has been shown to be due to a number of indirect mechanisms (reviewed in section 1.1.4.1.). Secondly, both the E1b-19K and the E1b-55K proteins also have the potential to regulate transcription during an adenovirus infection. However, mutational analysis suggests that these proteins have no significant effects on viral promoters during lytic infection (see sections 1.1.4.2. and 1.3), but their involvement in promoter transactivation is probably limited to cellular transformation. Thirdly, the multifunctional E2-DBP protein has a specific repressive effect on the E4 promoter, although the mechanism by which this repression occurs is unclear. Fourthly, two E4 proteins have been shown to be possibly important in transcriptional regulation, the ORF6/7 protein and the ORF4 protein. The ORF6/7 protein can bind to the host transcription factor E2F and induce the co-operative binding of this factor to the two inverted E2F binding sites in the E2 early promoter so inducing E2 transcription (see section 1.4.2.7.). The importance of the E4-ORF4 protein in transcription is not yet clear; initial experiments have shown it is involved in the phosphorylation of a number of transcription factors, so regulating their activity (see section 1.4.2.5.).

1.2.3. mRNA polyadenylation

The formation of a 3' end is an essential part in the biogenesis of a mature mRNA molecule. In higher eukaryotes (and adenovirus) the 3' end is formed by endonucleolytic cleavage of a pre-mRNA molecule followed by the addition of 200 to 250 adenylate residues to the new 3' OH terminus (reviewed by Proudfoot, 1991; Wahler and Keller, 1992). Transcription units usually contain one poly(A) site but

within the adenovirus genome there are a number of 'complex' transcription units which produce multiple mRNAs from a single promoter by utilising alternative splice sites and/or polyadenylation sites (Flint *et al.*, 1986; Leff *et al.*, 1986). The utilisation of alternative polyadenylation signals can lead to the production of different gene products and in the case of the adenovirus MLTU, can be temporally regulated.

The MLTU produces five 3' co-terminal families of mRNA, L1 through to L5, each defined by a unique polyadenylation site (reviewed by Nevins and Chen-Kiang 1981; Horwitz 1990a). By the temporally regulated utilisation of these sites the virus regulates the production of the structural and nonstructural proteins encoded by this region. Early in the infection L1 through to L3 RNAs are transcribed but the utilisation of the L1 site is three times that of the L3 site. This difference is even higher in the cytoplasmic mRNA population indicating many of the L3 polyadenylation events non-productive, resulting in mRNAs which fail to be exported from the nucleus. After DNA replication all 5 polyadenylation sites are utilised with L3 RNA outnumbering the L1 RNA by 3:1 (Nevins and Darnell, 1978; Chow *et al.*, 1979; Shaw and Ziff, 1980; Akusjarvi and Persson, 1981; Nevins and Wilson, 1981). This change in polyadenylation allows the virus to produce nonstructural proteins at early times and structural proteins later in the infection from the same transcription unit.

Recently, Larsson *et al.* (1992) proposed that there is an intermediate step between the early and late patterns of MLTU polyadenylation, with mRNAs from regions L1 and L4 being selectively over expressed as compared to the L2, L3, and L5 mRNAs. They proposed that for the late pattern of polyadenylation to occur both viral DNA replication and late protein synthesis are required.

Although a number of the sequences involved in poly(A) site selection have been identified, along with some of the protein factors responsible for the recognition and processing of poly(A) sites, the selection and regulation of alternative poly(A) sites is not understood. Work carried out on the MLTU has suggested the involvement of the RNA sequence of the poly(A) sites (DeZazzo *et al.*, 1991; Prescott and Falck-Pederson, 1992, 1994; Sittler *et al.*, 1994), nuclear factors that either selectively

enhance or depress the use of a specific poly(A) site and the relative levels of general factors involved in poly(A) site usage (late in the infection the high levels of MLTU precursor RNA may sequester the factors involved in the formation of the 3' termini of mRNAs, so reducing their relative levels in the nucleus (Mann *et al.*, 1993)).

The other adenovirus transcription unit which contains more than one polyadenylation site is the E3 unit. This region contains multiple splice sites and two polyadenylation sites, E3A and E3B, allowing the regulation of gene expression from this region by alternative splicing and polyadenylation (reviewed by Wold and Gooding, 1991; Brady *et al.*, 1992).

1.2.4. RNA splicing

1.2.4.1. Introduction

Eukaryotic pre-mRNA splicing is one of the steps in the nuclear processing of precursor RNA (pre-mRNA) into mature, translatable mRNA, common to almost all eukaryotic transcripts. RNA splicing occurs within large, nuclear ribonucleoprotein complexes, known as spliceosomes, where RNA sequences (introns) are precisely excised from the pre-mRNA and the remaining RNA sequences (exons) are accurately ligated together. The mechanism by which cells accomplish RNA splicing has been studied intensively but is still not fully understood (reviewed by Smith *et al.*, 1989; Green, 1991).

All but one of the adenovirus primary transcripts (the exception is IX) are spliced, with transcripts often being differentially spliced to produce more than one mRNA from a given transcription unit. This alternative pre-mRNA splicing is not unique to adenovirus, being found in a number of cellular transcription units (e.g. α -tropomyosin, calcitonin/CGRP, immunoglobulin κ light chain). In adenovirus, primary transcripts can be differentially spliced by the use of either alternative 5' donor sites with the same 3' acceptor site (e.g. E1a), or the same 5' donor site but alternative 3' acceptor sites (e.g. MLTU), or a combination of both of these events as in E4.

There are some notable differences between cellular and viral exon-intron arrangements: first, adenoviral transcripts have relatively few introns, the maximum discovered is six (the fibre mRNAs) and these introns are relatively short (probably because the genetic material has to be kept as short as possible if it is to be compressed into the viral capsid) while cellular genes commonly have numerous large introns; second, the viral introns tend to occur in the 5' or 3' untranslated regions (UTRs) of the transcripts while cellular genes generally have introns disrupting the actual coding regions, although a number of cellular genes have now been found where differential splicing only alters the UTRs, e.g. colony stimulating factor-1 RNA (Ladner *et al.*, 1987). The difference in the positioning of the introns is probably due to a subtle difference between the function of viral introns/exons and those introns/exons found in cellular genes. Eukaryotic cells differentially splice transcripts in order to: (a) localise proteins to alternative cell compartments (e.g. immunoglobulin μ , neural cell adhesion molecules, decay-acceleration factor, all of which are generated as either membrane-bound or secreted forms (Alt *et al.*, 1980; Caras *et al.*, 1988; Gower *et al.*, 1988)); (b) modulate the function of a protein either by deleting a region (e.g. the *Drosophila* transposase involved in the germ line-dependent transposition of P elements (Laski *et al.*, 1986)) or exchanging a segment(s) of a protein for an alternative sequence to produce a isoform of the protein (e.g. troponin-T, pyruvate kinase (Medford *et al.*, 1984)); (c) produce proteins with completely different functions from the same transcript (e.g. calcitonin and the calcitonin gene-related peptide (Amara *et al.*, 1982)); (d) regulate the efficiency of mRNA stability, transport and translation (e.g. CSF-1 (Ladner *et al.*, 1987)). In comparison, the majority of alternative splicing found in adenovirus gene expression is involved in the production of several completely distinct viral proteins from a single promoter (one exception is the E1a transcription unit where transcripts are differentially spliced to produce the two protein isoforms 243R and 289R).

Alternative splicing of most eukaryotic gene transcripts is regulated either in a developmental or cell type specific manner, whilst the regulation of alternative splicing

in an adenovirus lytic infection is regulated in a temporal manner, with the early splicing pattern within a given transcription unit often being different from the pattern late in the infection. Studies of the biochemistry of temporal splicing regulation in adenovirus infections have drawn various conclusions as to the mechanisms for splicing regulation (1.2.4.2. to 1.2.4.4.). The mechanisms identified can be arranged into a number of groups but it must be noted that these are not mutually exclusive but are probably highly interlinked.

1.2.4.2. Mechanism 1: Viral splicing factors

Recent work by Nordqvist *et al.* (1994) has demonstrated that adenovirus produces functional analogues of some splicing factors so regulating the temporal splicing of MLTU transcripts. All the MLTU mRNAs contain the tripartite leader sequence at their 5' end consisting of exons 1,2 and 3. The splicing of this tripartite leader is temporally regulated with an extra exon being included at early times to give a leader of 1-2-i-3. Ohman *et al.* (1993) and Nordqvist *et al.* (1994) demonstrated that this shift was due, in part at least, to two E4 proteins: ORF3-11K and the ORF6-34K. These two E4 proteins had different effects on the splicing of the tripartite leader sequences, with the E4 ORF3-11K protein facilitating 'i' leader exon inclusion while the E4 ORF6-34K protein facilitated 'i' leader exon exclusion (Nordqvist *et al.*, 1994). They suggest that the ORF3-11K and ORF6-34K have antagonistic functions in the modulation of RNA splicing comparable to that of the ASF/SF2 and hnRNP A1/distal splicing factor (DSF) proteins respectively (Ge and Manley, 1990; Harper and Manley, 1991; Krainer *et al.*, 1991; Mayeda and Krainer, 1992). It was also proposed that these E4 proteins may actually be involved in splice site selection (Ohman *et al.*, 1993; Nordqvist *et al.*, 1994). However the homology between the E4 proteins and the ASF/SF2 and hnRNP A2/DSF proteins is poor, as are the homologies for ORF3-11K and ORF6-34K with all the other known splicing factors, although the ORF6-34K protein does have a weak homology to an RNA binding domain at its C-terminus (Nordqvist *et al.*, 1994).

Nordqvist *et al.* (1994) argued that it is excessive for the virus to produce two proteins just to regulate the expression of the 'I' leader exon, which only encodes a non-essential (in tissue culture) 16K protein. They subsequently demonstrated the possible involvement of the E4 proteins in other splicing events such as the early-to-late shift in splicing of E1b region RNA and the splicing of a chimeric β -globin transcript *in vitro*. This would suggest a more general relevance for the E4 ORF3-11K and E4 ORF6-34K proteins in the processing of viral and cellular mRNA.

Klessig and Chow (1980), Anderson and Klessig (1984) and Ross and Ziff (1992) all implicated another viral protein in the regulation of splicing: the DBP protein. They demonstrated the importance of a functional DBP in the correct splicing of fibre and E4 mRNAs in monkey cells. Ross and Ziff (1992), working with E4 transcripts, proposed that DBP interacts with cellular components, overcoming a factor which restricts splicing to the early pattern, allowing the late pattern of RNA splicing. Ross and Ziff (1994) also observed other defects in the processing of MLTU RNAs in abortive infections of monkey cells, where the DBP does not function correctly. However they could not ascertain whether the defects were at the level of splicing or RNA transport. They also noted that the complex phenotype observed may be due to secondary effects of the DBP, as the E1b-55K/E4 ORF6-34K heterodimer (see section 1.3) is not believed to be produced efficiently due to aberrant processing of the E1b and E4 transcripts (see section 1.4.1.).

1.2.4.3. Mechanism 2: *cis*-acting sequences

The actual sequence of the primary transcript has been shown to be very important in regulating the temporal shift in splicing. For example, L1 pre-mRNAs are alternatively spliced in a temporal manner with a single 5' donor site spliced to two alternative 3' acceptor sites. Early in the infection the 52/55K mRNA (proximal 3' splice site used) is predominant, while at late times the IIIa mRNA (distal 3' splice site used) is predominant (Chow *et al.*, 1979; Akusjarvi and Persson, 1981; Nevins and Wilson, 1981). It has been shown that the IIIa 3' splice site is a less efficient site than the

52/55K 3' splice site having a poor polypyrimidine tract compared to that found in the 52/55K 3' splice site (Kreivi *et al.*, 1991). This, combined with its distal position in the transcript, results in its inefficient use at early times. It has been hypothesised that the virus produces a novel splicing factor (or modifies a host splicing factor) at late times (mechanism 1) which enhances the usage of the IIIa 3' splice site. This, combined with a virus induced repression of 52/55K splicing, may lead to the production of IIIa mRNA at late times (Kreivi and Akusjarvi, 1994).

Another example of the importance of the transcript sequence in the regulation of splicing is found in the E1a region (section 1.1.4.1.). The 10S and 11S mRNAs are produced only at low levels because of an unusually long distance (>50 nucleotides as compared to the usual 18 to 40 nucleotides) between the branch point and the 3' splice site in the first intron (Gattoni *et al.*, 1988; Chebli *et al.*, 1989). Chebli *et al.* (1989) demonstrated that a cis-acting element, a hairpin structure present between the branch point and the 3' splice site, is important in allowing the excision of this intron. It has been hypothesised that this hairpin has the effect of shortening the effective distance between the branch point and the 3' splice site allowing the excision of the intron, albeit inefficiently. This inefficient splicing ensures only a low level of the 10S and 11S mRNAs are made, so regulating gene expression.

1.2.4.4. Mechanism 3: General splicing factors

Studies of the E1a transcription unit (1.1.4.1.) have suggested that the levels of ubiquitous splicing factors may be crucial in the temporal regulation of adenovirus splicing. The onset of production of the 9S mRNA at late times has been shown to be due, in part, to a change in levels of general splicing factors (Gattoni *et al.*, 1991; Larsson *et al.*, 1991). Also *in vitro* experiments have demonstrated that the concentration of splicing factors can be critical in splice site choice (Ge and Manley, 1990; Krainer *et al.*, 1990). Larsson *et al.* (1991) hypothesised that during the late phase of the infection many of the general splicing factors are efficiently sequestered by the high concentration of MLTU transcripts, so heightening the competition between

splice sites for factors and favouring the splicing of efficient sites. This would result in a different pattern of mRNAs early and late in the infection.

The results of experiments carried out by Adami and Babiss (1991) indicated that global changes in trans-acting splicing factors cannot explain totally the observed temporal changes in Ad splicing. They staggered a co-infection of two distinguishable viral genomes and observed that both early and late genomes could exist simultaneously in the nucleus; E1b mRNA was spliced in both early and late patterns depending on its genome of origin, indicating that neither genome had influence on the other. They also demonstrated that the role of potential cis-elements is not as simple as thought; they introduced an SV40 gene, which displayed no regulated splicing during an SV40 infection or SV40/adenovirus co-infection of COS cells, into the adenovirus genome. On infection with the recombinant virus the splicing of the RNA from the SV40 gene was regulated. A possible hypothesis is that RNA processing in adenovirus-infected cells is linked to the state of the DNA template or to the DNA being present within a distinct nuclear subcompartment, with trans-acting splicing factors only having a local effect. Experiments exploring the compartmentalisation of the adenovirus genome within the nucleus have indicated that it is closely associated with the nuclear matrix and that this association is important for viral transcription and replication (Bodnar *et al.*, 1989, Shaack *et al.*, 1990). During the late phase of the infection, adenovirus DNA is found within viral inclusion bodies or replication factories, which are believed to be the sites of DNA replication, late transcription and possibly RNA processing (Moyné *et al.*, 1978; Zhonghe *et al.*, 1986; Walton *et al.*, 1989; Moen *et al.*, 1990; Jiménez-García and Spector, 1993). This would imply that viral RNA production and processing occur in discrete centres in the adenovirus-infected nucleus, each of which may present different microenvironments for gene expression. This may explain how RNA transcribed from two genomes in the same cell can be processed differently. Work exploring RNA metabolism in uninfected cells has also suggested that RNA processing and transport occurs in discrete sites on the nuclear matrix (Xing *et al.*, 1993; Carter *et al.*, 1993; Zachar *et al.*, 1993; reviewed by

Rosbash and Singer, 1993; Xing and Lawrence, 1993) giving extra weight to the hypothesis that the product of individual viral genomes can be processed differently according to their microenvironment. Alternatively, the nature of the DNA template may affect the processing of its transcripts: progeny DNA may be transcribed differently from input virion DNA late in the infection, with possibly different elongation rates, pausing or even different transcriptional machinery. This change in transcription may then affect splice site choice since it has been shown that RNA transcription and processing are closely linked (Sisodia *et al.*, 1987; Beyer and Osheim, 1988; Neuberger and Williams, 1988; Moen *et al.*, 1990; Jiménez-García and Spector, 1993; Bridge *et al.*, 1993a; reviewed by Xing and Lawrence, 1993; Rosbash and Singer, 1993).

1.2.4.5. Summary

The regulation of the alternative splicing found in adenovirus-infected cells is a complex process which we are only just beginning to comprehend. The three mechanisms of regulation described above are not mutually exclusive but are highly interrelated. The balance of these regulatory systems results in the production of precisely spliced viral mRNAs at the required time during the infection and at the required level.

1.2.5. mRNA export

To function, the cell has to traffic macromolecules in and out of the nucleus in an ordered, regulated fashion. Compared to the transport of proteins into the nucleus, very little is known about the mechanism of RNA export from the nucleus. The term 'mRNA export' encompasses two distinct steps: firstly the mRNA must leave its site of production on the nuclear matrix and travel through the nuclear subcompartments to the nuclear periphery. Secondly the RNA must cross the nuclear membrane and enter the cytoplasmic pool of actively translating mRNAs.

Alterations in this pathway of RNA transport can result in the altered accumulation of cytoplasmic mRNA and so a different pattern of gene expression. This step in gene expression is targeted by certain viruses when altering the gene expression in the infected cell; examples are human immuno-deficiency virus type 1 (HIV-1), influenza A virus and adenovirus. The HIV-1 Rev protein mediates the stabilisation and the nuclear export of unspliced and partially spliced viral pre-mRNAs (reviewed by Chang and Sharp, 1990; Krug, 1993). Early in the replication of HIV only spliced viral mRNAs are exported to the cytoplasm. One of these fully spliced RNAs encodes the Rev protein which then allows the export of partially and unspliced messages to the cytoplasm, which encode the structural proteins of the virus. It is believed that the N-terminal of the Rev protein interacts with the *cis*-acting Rev response element (RRE) found in a major intron sequence of the HIV transcript while a domain in the C-terminal interacts with the transport machinery, facilitating the export of unspliced messages. The NS1 protein of influenza virus has been shown to be involved in the inhibition of splicing and a general block in the transport of mRNA from the nucleus, but as yet it is unclear as to how it functions or what role it plays in the influenza infection (Alonso-Caplen *et al.*, 1992; Fortes *et al.*, 1994; reviewed by Krug, 1993). Adenovirus regulates RNA transport via the E1b-55K and E4 ORF6-34K proteins. For a detailed discussion of the function of these two proteins see sections 1.3. and 1.4.2.3.

1.2.6. Translation

1.2.6.1 Introduction

The late phase of an adenovirus infection is marked by the inhibition of cellular protein synthesis, with the preferential translation of late viral mRNAs (late viral mRNAs constitute about 20% of the mRNA found in the cytoplasm but about 90-95% of the mRNAs found in polysomes (Tal *et al.*, 1975). This virus-induced decrease in cellular protein synthesis is believed to be due to an inhibition of cellular mRNA translation. The concurrent decrease in cellular mRNA export from the nucleus is thought to play

only a minimal role in inhibiting cellular protein synthesis for three reasons. First, Babich *et al.* (1983) observed that the shut-off of protein synthesis from selected genes was complete when the block in mRNA transport was only minimal early in the late phase. Second, Moore *et al.* (1987) noticed that mRNAs which escape the block in transport, such as *hsp70* and *hsx70*, still fail to be translated indicating a block to protein synthesis at the level of mRNA translation. Third, some drugs prevent the adenovirus shut-off of translation without affecting the block in the transport of host mRNAs (Huang and Schneider, 1990).

Other possible mechanisms have been proposed for the shutoff of protein synthesis such as the degradation of cellular mRNA, the modification of cellular mRNA so rendering it untranslatable or the swamping of the translation systems with viral transcripts so preventing the translation of cellular mRNA. These have all been discounted in the adenovirus system as host mRNAs remain in the cytoplasm through the late phase of the infection in a translatable state (Thimmappaya *et al.*, 1982). Also Tal *et al.* (1975) demonstrated that late viral mRNAs only constitute about 20% of the total cytoplasmic pool of mRNAs, not enough to prevent the almost 100% shutoff of host translation observed by a simple competition mechanism.

So how does adenovirus achieve preferential translation of viral mRNAs in the late phase? Evidence has accumulated over the past few years that adenovirus regulates translation by at least three different mechanisms mediated by the VA_I RNA, the L4-100K protein, and the MLTU tripartite leader. It is unclear whether these mechanisms represent an integrated effort by the virus to modulate translation or whether they are distinct mechanisms that act separately.

1.2.6.2. VA_I RNA (reviewed by Mathews and Shenk, 1991)

In the absence of VA_I RNA there is little or no translation of viral or cellular mRNAs within the cell (Thimmappaya *et al.*, 1982). This inhibition is due to the inactivation of the initiation factor eIF-2 through the phosphorylation of its α -subunit (Reichel *et al.*, 1985; Schneider *et al.*, 1985) by the interferon-induced protein kinase, DAI (dsRNA

activated inhibitor). DAI is activated by the presence of dsRNA which results from the symmetrical transcription of the viral genome late in the infection (Moran and Mathews, 1988). VA_I RNA prevents the phosphorylation of eIF-2 α by directly binding to DAI so inhibiting its function (Katze *et al.*, 1987; Kostura and Mathews, 1989).

O'Malley *et al.* (1989) also proposed that VA_I RNA is important in the selective translation of late viral mRNA late in the viral infection. They observed that late in the infection about 30% of cellular eIF-2 is phosphorylated, and so inactive. It is known that this level of phosphorylation completely inhibits protein synthesis in uninfected cells but in infected cells viral late mRNAs are still translated. This, and other observations, led O'Malley *et al.* to propose that viral and host mRNAs may be segregated into functionally distinct compartments in the late infected cell. Translation of host mRNAs would be inhibited because of DAI activity, but viral mRNAs would be still translated because a specific association between these RNAs and VA_I RNA (Mathews, 1980) causes inhibition of the action of the DAI kinase in the local vicinity of the viral late mRNA.

1.2.6.3. L4-100K

The L4-100K protein is a non-structural protein expressed late in the infection. It is essential for in the construction of virus particles (Cepco and Sharp, 1983), but also has a role in the regulation of translation. The L4-100K protein is found in close association with mRNA (Adam and Dreyfuss, 1987) and is believed to be involved in the initiation of translation of late viral mRNAs (Hayes *et al.*, 1990; Riley and Flint, 1993). The specificity of this function of the 100K protein is not yet clear, nor is the mechanism by which it acts.

1.2.6.4. The MLTU tripartite leader (reviewed by Zhang and Schneider, 1993)

Late in the infection the majority of viral mRNAs are transcribed from the MLTU and therefore possess an identical 200 nucleotide long 5' UTR known as the tripartite

leader (Berget *et al.*, 1977). This leader sequence is required for the translation of MLTU transcripts late in the infection (Logan and Shenk, 1984; Berkner and Sharp, 1985), when most other translation events have been inhibited.

In an uninfected cell, the 5' end of an mRNA is unwound by the eIF-4F cap-dependent helicase, so allowing translation (reviewed by Rhoads, 1988). However in the late phase of an adenovirus infection it is believed that this action of eIF-4F is inhibited by an alteration in its phosphorylation state, possibly by the activity of the DAI kinase (Huang and Schneider, 1991). This would result in the inhibition of translation of any mRNAs with secondary structures in their 5' ends which would probably include most of the cellular mRNAs. However, the 5' tripartite leader of the MLTU mRNAs lacks any secondary structure (Dolph *et al.*, 1990) and so these mRNAs are still translated efficiently in the absence of eIF-4F helicase activity.

Adenovirus is not alone in disabling the eIF-4F helicase to achieve selective translation of its proteins. Polio virus is known to degrade one of the eIF-4F components so rendering it inactive. It circumvents the requirement for eIF-4F in the translation of its mRNAs by binding ribosomes internally on the RNA (Sonenberg, 1987; Pelletier and Sonenberg, 1988).

1.3. E1b-55K and cytoplasmic viral mRNA accumulation

1.3.1. Introduction

Two adenoviral proteins are involved in the accumulation of viral mRNAs in the cytoplasm of infected cells, the E1b-55K and E4 ORF6-34K proteins. Co-immunoprecipitation studies with the E1b-55K protein have found it physically associated with the E4 ORF6-34K protein within infected cells (Sarnow *et al.*, 1984) although unassociated populations of both proteins have also been identified (see below). Genetic evidence that this complex acts as a functional unit has also been found since it was observed that E4 ORF6-34K, E1b-55K and E4 ORF6-34K/E1b-55K double mutants all displayed similar phenotypes (Halbert *et al.*, 1985; Cutt *et al.*,

1987; Samulski *et al.*, 1988; Bridge and Ketner, 1990). Thus in the following discussion of E1b-55K it is presumed that the E4 ORF6-34K is also required for the functions described.

It is believed that the E1b-55K/E4 ORF6-34K heterodimer facilitates the cytoplasmic accumulation of viral mRNAs in the cytoplasm of infected cells by altering the RNA transport machinery, although it is conceivable that it may function at the level of RNA stability. Discussed in the following section are the current models for RNA transport, E1b function and the evidence for them. A detailed discussion of the function of the E4 ORF6-34K protein can be found in section 1.4.2.3.

1.3.2. RNA transport models

It is thought that in the uninfected cell, RNA, transcribed from decondensed stretches of chromatin, travels on defined paths through RNA processing sites on route to the cytoplasm. The basis for this model of RNA transport comes from several observations. Jackson and Cook (1985), Zeitlin *et al.* (1989) and Ciejek *et al.* (1983) demonstrated the close association of mRNA transcription and subsequent maturation steps with the nuclear matrix while Ciejek *et al.* (1982) observed that pre-mRNA, in several stages of maturation, is also associated with the nuclear matrix. This suggests that the production of mature mRNA occurs in association with the nuclear matrix, probably in a controlled, regulated manner as opposed to pre-mRNAs being processed while diffusing freely in the nucleoplasm. This association with the nuclear matrix is believed to be due to the interactions of the RNA exons and ribonucleosome proteins (reviewed by Agutter, 1991) while excised introns are free to diffuse within the nucleus (Zeitlin *et al.*, 1987; Xing *et al.*, 1993). Poly(A)⁺ RNA (believed to represent most of the nuclear pre-mRNA) has been shown to accumulate within 20-50 'transcript domains' that coincide with the location of a number of proteins involved in splicing (Carter *et al.*, 1991, 1993; Xing and Lawrence, 1991) including the spliceosome assembly factor, SC-35 (Fu and Maniatis, 1990). It is believed that these 'transcript domains' are also associated with active transcription (reviewed by Spector, 1993).

Beyer and Osheim (1988) and LeMaire and Thummel (1990) also reported the close association of transcription and RNA processing. Elaborate *in situ* hybridisation experiments, targeting certain RNA species, have confirmed the above observations with RNA species being found on defined tracks within the nucleus (Lawrence *et al.*, 1989; Huang and Spector, 1991; Xing *et al.*, 1993), suggesting that pre-mRNAs are constrained from free diffusion, possibly due to the interaction of the RNA with the nuclear matrix. These tracks originate from sites of transcription and contain both intronic and exonic sequences at the end proximal to the site of transcription but only exonic sequences at the distal end (Xing *et al.*, 1993). Xing *et al.* (1993) also observed that these tracks co-localised with the 'transcript domains' and proposed that splicing occurs within these tracks in a spatially ordered manner. However, only a few of the tracks observed extend from the site of transcription to the nuclear membrane, possibly because an mRNA may exit the nucleus very rapidly after processing is completed, so making it undetectable by the methods employed.

It should be noted however that an alternative model has been proposed by Zachar *et al.* (1993) (reviewed by Kramer *et al.*, 1994). Zachar *et al.* (1993) observed that specific transcripts moved away from their parental gene to the nuclear surface, in a pattern indicative of simple diffusion. In their model they suggest that RNA processing factors also interact with the RNA by simple diffusion within a nuclear subcompartment. They explain the tracks observed in mammalian cells, believed to be complete RNAs undergoing processing, as nascent transcripts still intimately associated with their corresponding gene. Co-transcriptional splicing evidence (Beyer and Osheim, 1988; LeMaire and Thummel, 1990) explained the presence of intronic sequences only in the end of the transcript proximal to the site of transcription, as the 3' end of the transcript could still be colinear with the parental gene. Experiments by Jackson *et al.* (1993) and Wansink *et al.* (1993) also argue that nascent RNA is not associated with the 20 to 50 'transcript domains' but is in fact located in hundreds of domains though out the nucleus. This lends support to the idea that RNA processing occurs outside of the transcript domains, and that the transcript domains are simply

storage centres for the splicing machinery. It is not clear, however, if this simple diffusion model, derived from data obtained largely in *Drosophila* cells containing polytene chromosomes, can be applied to diploid mammalian cells or whether the mechanism of RNA transport is different in these cells.

As yet it is unclear which, if either, model is correct (for an extensive discussion of the subject see Rosbash and Singer, 1993; Spector, 1993; Xing and Lawrence, 1993; Kramer *et al.*, 1994).

Our understanding of the translocation of mRNA across the nuclear membrane is also limited. Electron microscopy studies of nuclear pore complexes have shown that mRNA, associated with hnRNP proteins, exits the nucleus via the nuclear pore complexes (NPCs), (Dworetzky and Feldherr, 1988; Mehlin *et al.*, 1992). The RNA/hnRNP complex associates with fibrous material extending from a pore and then passes through the pore in a 5' to 3' orientation (Mehlin *et al.*, 1992). These observations were confirmed by NPC inhibition studies where anti-pore proteins and wheat germ agglutinin, which inhibit the protein transport function of the NPC, inhibited RNA export (Featherstone *et al.*, 1988; Bataillé *et al.*, 1990; Neuman de Vegvar and Dahlberg, 1990; Dargemont and Kühn, 1992; Michaud and Goldfarb, 1992).

The mechanisms that regulate these steps in the uninfected cell are still poorly understood. RNA export is known to be a saturable and energy-dependent process implying the involvement of proteins as carriers of the RNA. (Zasloff, 1983; Dworetzky and Feldherr, 1988; Khanna-Gupta and Ware, 1989; Bataillé *et al.*, 1990; Dargemont and Kühn, 1992). Using competition experiments, Jarmolowski *et al.* (1994) suggested that the export of different classes of RNAs (i.e. RNAs from RNA pol I (rRNAs), RNA pol II (mRNAs and snRNAs) and RNA pol III (tRNAs and 5S rRNA)) is mediated by specific rather than common factors, at least for some of the steps in RNA export. The identity of these factors is unclear, but it is known that within the nucleus RNA pol II transcripts are associated with a number of nuclear proteins, collectively known as the heterogeneous nuclear ribonucleoprotein (hnRNP)

proteins (reviewed by Dreyfuss *et al.*, 1993). Further work by Piñol-Roma and Dreyfuss (1991, 1992) has suggested that these hnRNP proteins are important in mediating RNA transport as a number of them are associated with poly(A) transcripts both in the nucleus and in the cytoplasm (reviewed by Piñol-Roma and Dreyfuss, 1993).

The fact that most primary transcripts require post-transcriptional processing to achieve their functional form means that there are many RNA molecules in the nucleus that, because of incomplete processing, must not exit to the cytoplasm. This implies the existence of signals which allow the export machinery to discriminate between mature and immature transcripts. It seems unlikely that a common signal sequence lies in the primary sequence, especially for mRNAs, as different mRNAs display no obvious sequence homology. It has been proposed that signals for RNA export are located at the termini of RNA pol II transcripts; Hamm and Mattaj (1990) and Dargemont and Kühn (1992) demonstrated that the 5' mono-methylated cap structure of RNA pol II transcripts was involved in RNA export. This work was later confirmed when a 80K cap-binding protein was isolated, and found to be involved in RNA export (Izaurradle *et al.*, 1992). However the 5' m⁷G cap structure is not the sole requirement for export; working on the RNA pol II transcript of U3 RNA, Terns and Dahlberg (1994) demonstrated that RNAs could possess a 5' m⁷G cap structure but not be exported from the nucleus. This would suggest either the presence of additional signals and/or that the localisation of the transcript within the nucleus is important. Working with histone RNA, Eckner *et al.* (1991) identified another positive signal important in RNA transport; the formation of the 3' end of the transcript. However it is unclear whether this processing is important in all RNA export events or specific to histone RNA export. Another study, investigating 5S rRNA, has shown the importance of secondary structures in nuclear export (Allison *et al.*, 1993). The importance of secondary structures to RNA export may not be unique to 5S rRNA as there is growing evidence that RNA structures such as helices, loops, bulges, mismatches and pseudoknots are key elements in protein-RNA interactions (reviewed

by Draper, 1989). He *et al.* (1994) have also demonstrated the importance of the Alu sequence present in some mRNAs in mRNA transport, although the key element here may still be the secondary structure of this region.

Evidence also exists for the presence of negative signals in RNA transport. Intact introns present in the RNA molecule have been shown to act as nuclear retention signals, blocking efficient RNA export (Ciejek *et al.*, 1982; Chang and Sharp, 1989; Legrain and Rosbash, 1989; Hamm and Mattaj, 1990). It has been proposed that there is a competition effect between the proteins involved in intron identification and the transport machinery. Changes in the efficiency of splice site recognition (be it *cis*-acting elements or *trans*-acting factors) or changes in the efficiency of the transport machinery can potentially alter the levels of partially or even unspliced messages in the cytoplasm. It is unclear whether intron recognition and nuclear sequestration necessarily guarantee that a pre-mRNA will be incorporated into a complex committed to splicing or whether further nuclear interactions have to take place. However it is clear that the formation of splicing complexes on the pre-mRNA prevents the export of any nascent mRNAs that retain intronic sequences.

For a complete discussion of the regulation of RNA export see Izaurralde and Mattaj (1992), Hanover, (1992), Miller and Hanover (1992) and Elliot *et al* (1994).

1.3.3. The E1b-55K protein - subcellular localisation

The E1b-55K protein is a phosphoprotein initially expressed early in the infection. Studies have found that the E1b-55K protein is localised primarily in the nucleus of infected cells, although low levels can be detected in the cytoplasm (Sarnow *et al.*, 1982; Rowe *et al.*, 1983; Smiley *et al.*, 1990; Ornelles and Shenk, 1991). Using subnuclear fractionation and immunofluoresence, Smiley *et al.* (1990) attempted to locate precisely the intranuclear location of E1b-55K. Two distinct populations of the E1b protein were found within the nucleus: 30-50% of the total nuclear E1b-55K was released on RNase A digestion of the nuclei while the rest of the E1b-55K protein localised in the pore complex-lamina fraction. Within the pore complex-lamina, both

free E1b-55K and E1b-55K/E4 ORF6-34K complex were found, but it was noted that neither form of the protein appeared to form a stable or intimate association with the pore complexes. However, Ornelles and Shenk (1991), employing immunofluorescence and immuno-electromicroscopy to localise E1b-55K, found that the E4 ORF6-34K protein was required for association of the E1b-55K/E4 ORF6-34K complex with nuclear viral inclusion bodies, believed to be the sites of transcription and DNA replication (Moyne *et al.*, 1978; Zhonghe *et al.*, 1986; Walton *et al.*, 1989; Moen *et al.*, 1990; Jiménez-García and Spector, 1993). They found the E1b-55K protein in five different sites within the cell. In the cytoplasm it was found in a diffuse, reticular pattern and also in association with a fibrous body adjacent to the nucleus. In the nucleus it was found in a diffuse, granular pattern excluded from the nucleolus, in randomly distributed, discrete spicules and also associated with viral inclusion bodies. When the E4 ORF6-34K protein was mutated, the E1b-55K protein failed to associated efficiently with the viral inclusion bodies, while the other populations of E1b-55K were apparently unaffected. Other studies have demonstrated that such E4 mutants display defects in mRNA metabolism (Halbert *et al.*, 1985; Cutt *et al.*, 1987; Samulski *et al.*, 1988; Bridge and Ketner, 1990) suggesting that these inclusion bodies are the probable site of action of the E1b-55K protein, although the involvement of 'free' E1b-55K protein in mRNA metabolism cannot be completely discounted. The experiments of Ornelles and Shenk (1991) also indicated that a sequence within the E4 ORF6-34K protein or within the complex, but not in E1b-55K alone, is required for the association with the viral inclusion bodies.

The large number of intracellular sites found in these studies have also been observed by biochemical fractionation (Rowe *et al.*, 1983) and may reflect different functional forms of the protein (i.e. E1b-55K/p53, E1b-55K/E4 ORF6-34K etc.). Cutt *et al.* (1987) noted that the formation of the E1b-55K/E4 ORF6-34K complex was not co-translational, with complex formation taking a number of hours. This could explain why both free E1b-55K and E1b-55K/E4 ORF6-34K complex are found within the cell.

1.3.4. The elucidation of E1b-55K function

When the E1b-55K gene was mutated by Babiss and Ginsberg (1984) and Logan *et al.* (1984), an impaired growth phenotype in HeLa cells was observed with virus yields reduced to 10% or less of wild type. More detailed studies demonstrated that the E1b-55K protein was required for the cytoplasmic accumulation of viral late mRNA (the early mRNAs studied showed no dependence on the E1b-55K protein) and for the inhibition of the cytoplasmic accumulation of host cell mRNAs during a lytic infection (Babiss and Ginsberg, 1984; Babiss *et al.*, 1985; Pilder *et al.*, 1986a,b; Williams *et al.*, 1986). This observed decrease was not due to a defect in late RNA transcription as the transcription rate across the MLTU early in the late phase of the mutant infection was unchanged while there was a substantial decrease in cytoplasmic MLTU mRNA accumulation (Babiss *et al.*, 1985; Pilder *et al.*, 1986a,b). Later in the infection the transcription rate across the MLTU did decrease in the mutant infection as compared to wild type, probably due to secondary effects of altered gene expression. Nuclear post-transcriptional RNA processing events, such as splicing and polyadenylation, were also not the cause of this defect as the processing of the MLTU RNA was normal in the E1b mutant infection (Babiss *et al.*, 1985; Pilder *et al.*, 1986a,b; Williams *et al.*, 1986). These results suggested that neither the transcription nor post-transcriptional processing of late viral mRNAs was the primary site of E1b-55K action, but it was concluded that the E1b-55K protein acted at the level of mRNA transport or possibly RNA stabilisation post-maturation.

Cytoplasmic accumulation of mRNAs is dependent on two factors, the rate of nuclear export of mRNAs and the stability of mRNAs in the cytoplasm. Pilder *et al.* (1986b) observed some variation in cytoplasmic stability of L5 mRNA late in the infection but concluded that this was probably a secondary effect of the E1b-55K mutation because no difference was observed in L5 mRNA cytoplasmic stability early in the late phase while at the same time there was still a substantial decrease in L5 mRNA cytoplasmic accumulation. Williams *et al.* (1986) also explored the effect of E1b on late mRNA

cytoplasmic stability. They found some reduction in hexon mRNA cytoplasmic stability (no time post-infection was given) but still this reduction could not account for the level of hexon mRNA reduction observed in the cytoplasm. To conclude, it would appear that the primary site of E1b-55K action is the transport of mRNA from the nucleus of the infected cell while the differences observed in late mRNA transcription and cytoplasmic stability in E1b-55K mutant infections are secondary effects of the mutations.

The E1b-55K protein also prevents the continued accumulation of cellular mRNAs in the cytoplasm (Babiss and Ginsberg, 1984; Babiss *et al.*, 1985; Pilder *et al.*, 1986a,b; Williams *et al.*, 1986). This effect is coincident with the inhibition of host protein synthesis late in the infection but is not thought to be the primary cause, as cellular mRNAs remain intact in the cytoplasm throughout the late phase but are not translated. This change of synthetic activity in the infected cell is believed to be due primarily to a block at the level of translation of host cell mRNAs (see section 1.2.6.). The block to the accumulation of cellular mRNA in the cytoplasm is thought to be due to an effect on mRNA transport or mRNA cytoplasmic stability as cellular RNA transcription occurs as normal (Babiss *et al.*, 1985; Pilder *et al.*, 1986a,b). It is unclear if the reciprocal effects of E1b-55K on viral and cellular mRNA accumulation in the cytoplasm are the result of two distinct functions of the protein or whether E1b-55K directly affects one process and as a result indirectly affects the other (see below). However it appears that this block in cellular mRNA accumulation in the cytoplasm is selective as a number of distinct cellular species do continue to accumulate in the cytoplasm of infected cells, i.e. β -tubulin, *hsp70* (Moore *et al.*, 1987). Moore *et al.* (1987) proposed that the process of transcriptional induction allowed the resulting mRNA to escape the viral block to cytoplasmic accumulation.

The exact site of E1b-55K action on the transport pathway of RNA is unknown. Leppard and Shenk (1989) examined at which step in the RNA transport pathway the E1b-55K protein acted by exploiting the cold-sensitivity of a E1b-55K mutant (dl338). At 32 °C the phenotype of the E1b-55K mutation is exacerbated and the RNA

transport pathway is slowed down as compared to the same infection at 37 °C. This allowed them to study RNA metabolism in infected cells. Cells were pulse labelled *in vivo*, the label chased for varying periods and then fractionated into one cytoplasmic and four nuclear fractions (nuclear membrane, DNase I-soluble, salt soluble and nuclear matrix). The fractions were subsequently probed for various late RNA species. Newly synthesised late RNA was first found in association with the nuclear matrix (where it was transcribed), then seen to accumulate in the DNase-I and salt soluble fractions before exiting the nucleus through the nuclear membrane and entering the cytoplasmic fraction. In the E1b-55K mutant infection, late viral mRNAs left the nuclear matrix fraction and accumulated in the nuclear membrane fraction more slowly and less efficiently than late viral mRNAs from a wild type virus. These transcripts also failed to accumulate in the DNase-I soluble fraction of the nucleus indicating they were degraded before or when reaching this compartment. Early viral mRNAs displayed no difference in accumulation at any step between the mutant and wild type virus infections. From these observations it was concluded that the E1b-55K protein functions at an early step in the transport pathway for late mRNAs, possibly at the step of RNA release from the nuclear matrix. This E1b-55K functional data correlates well with the localisation studies carried out by Smiley *et al.* (1990) and Ornelles and Shenk (1991).

It has also been noted that not all viral mRNAs are dependent on the E1b-55K protein for optimal cytoplasmic accumulation. Early viral mRNAs are not dependent on the E1b-55K protein for cytoplasmic accumulation (Babiss and Ginsberg, 1984; Pilder *et al.*, 1986a, 1986b; Williams *et al.*, 1986; Leppard, 1993) including the L1 52/55K mRNA expressed early in the infection. Only RNAs expressed late in the infection have any dependence on E1b-55K for optimal cytoplasmic accumulation, including the L1 52/55K mRNA expressed late in the infection. Initial studies of late mRNAs focused on the MLTU mRNAs but Leppard (1993) demonstrated that mRNAs from the other late promoters (E2-L, IVa2 and IX) are also dependent on the E1b-55K protein for optimal cytoplasmic accumulation although to varying degrees. From the

data accumulated by Leppard (1993) it appears that the MLTU transcripts are not equally dependent on the E1b-55K protein for optimal accumulation, with the longest mRNA of a 3'-coterminal family displaying the greatest dependence.

Some non-adenovirus transcripts are also dependent on E1b-55K for optimal cytoplasmic accumulation. Samulski and Shenk (1988) demonstrated that the E1b-55K protein and E4 ORF6-34K protein were required for the optimal cytoplasmic accumulation of adeno-associated virus (AAV) - derived transcripts.

To summarise, the E1b-55K product (probably in conjunction with the E4 ORF6-34K protein) acts post-transcriptionally on RNA metabolism, facilitating the transport and cytoplasmic accumulation of late viral mRNAs to varying degrees. Conversely, cellular mRNAs are prevented from accumulating in the cytoplasm of infected cells by E1b-55K late in the infection. E1b-55K could potentially function at any of the steps in RNA transport but function and localisation studies have indicated that it acts at an early step, most probably the release of the RNA from the matrix. This action of E1b-55K allows the cytoplasmic accumulation of late viral mRNAs and so the translation of the late gene products.

1.3.5. A model for E1b-55K function

Our current working model for RNA export during an adenovirus infection is based upon the first model in section 1.3.2., in which RNA follows discrete paths within the nucleus, and encompasses the 'gene gating' hypothesis of Blobel (1985), (Moore *et al.*, 1987; Leppard and Shenk, 1989; Ornelles and Shenk, 1991; Leppard 1993). Blobel proposed that the eukaryotic genome has a complex 3-D organisation with regions actively transcribed being structurally expanded and linked, or 'gated', to nuclear pore complexes, so facilitating the exit of their transcripts to the cytoplasm. All transcripts of a particular gene would leave the nucleus via a single, linked nuclear pore; RNA export from a gene not linked to a nuclear pore would be possible but very inefficient. This arrangement was suggested to be dynamic, so that when transcription halted, the gene was removed from the site of facilitated RNA transport and vice versa. Blobel

envisioned the export of RNA from these gated genes to take the form of non-diffusable tracks along which the RNA processing occurs in agreement with the model reviewed by Xing *et al.* (1993). Further evidence for this hypothesis comes from the observation that expanded chromatin (i.e. chromatin undergoing active transcription) is localised around the nuclear periphery of the nucleus (Hutchinson and Weintraub, 1985; Krystosek and Puck, 1990; De Graaf *et al.*, 1991), possibly suggesting that active genes are close to, or organised with respect to, nuclear pores, thereby facilitating RNA export. Huang and Spector (1991) also observed that *c-fos* transcripts exit the nucleus at a very limited area in the nuclear membrane, suggesting that certain transcripts are targeted to a small number of nuclear pores, consistent with the hypothesis of Blobel.

The model suggests that there is a limited capacity in the nucleus for the export of RNA, and that the process of RNA transport can be saturated (as observed by Zasloff, 1983; Khanna-Gupta and Ware 1989; Bataillé *et al.*, 1990; Dargemont and Kühn, 1992). Applying this model to an adenovirus infection, it is proposed that, early in the infection, a viral transcription complex would occupy one of these facilitated sites, as the system would not be saturated by the transcription and subsequent transport of cellular RNAs, so allowing the transport of early mRNAs independently of the E1b-55K/E4 ORF6-34K complex. However, late in the infection, after the onset of DNA replication, there would be considerably more viral transcription complexes present within the nucleus as transcription would now be occurring from the numerous progeny viral genomes. This large number of transcription complexes would saturate the facilitated sites and, in the absence of a vacant site, these transcription complexes would either have to compete for a site already occupied by another transcription complex or be transcribed elsewhere in the nucleus in a site less favourable for RNA export. In either scenario, it is hypothesised that the E1b-55K/E4 ORF6-34K complex would circumvent this situation, which arises late in the infection, either by (a) actively displacing cellular genes from these facilitated sites allowing viral transcription complexes to occupy the sites, or by (b) recruiting transport factors from existing sites

(so rendering them inoperative) for the transport of viral mRNAs, or by (c) creating a novel, virus-specific, RNA export mechanism. Which of these mechanisms actually operates is unclear, but Jiménez-García and Spector (1993) observed the recruitment of transcription and splicing factors to sites of active adenovirus transcription (the viral inclusion bodies) late in the virus infection, suggesting that the virus facilitates viral pre-mRNA processing and transport by a recruitment mechanism, possibly by the action of E1b-55K found in association with the inclusion bodies. The identification of a potential RNA binding motif within the E1b-55K protein (K. Leppard, unpublished observations) may be important in this function. It is also of note that Moen et al. (1990) observed that Ad2 DNA, late in the infection, was localised in foci at the nuclear periphery, as expected according to this model. Only the first two mechanisms explain the block in host cell mRNA export as an indirect effect of facilitated viral RNA export, but it is possible that the E1b-55K/E4 ORF6-34K complex directly inhibits host cell mRNA transport in addition to having a direct effect on viral RNA.

This model explains a number of observations. Firstly it explains why only late gene transcripts are dependent on the E1b-55K/E4 ORF6-34K complex for optimal cytoplasmic mRNA accumulation. Secondly it accounts for the general block in the cytoplasmic cellular mRNA accumulation while transcripts of some genes, such as the E1a-induced β -tubulin, accumulate normally; Blobel hypothesised that transcriptionally induced genes would have an increased linkage to the transport machinery, so allowing them to compete effectively with viral transcription complexes. Thirdly it explains why L1 52/55K mRNA is independent of the E1b-55K/E4 ORF6-34K complex for optimal cytoplasmic accumulation early in the infection while late in the infection it is dependent. Fourthly it can account for the observation made by Leppard and Shenk (1989) where it was noticed that early and late messages were handled differently; viral early messages and the host β -actin message were strongly biased toward the cytoplasm while viral late messages were found in significant amounts within the nucleus. Finally it can account for the observations made in Ad12 by Schramayr *et al.* (1990) where it was found that E1b-55K plays a role in the decondensation of the

chromatin at a limited number of sites. This phenomenon was not observed in Ad5 but does suggest a role for the E1b-55K protein in the reorganisation of the nuclear architecture.

The model can also be adapted to explain the observations made by Leppard (1993) where it was noted that the longest mRNA species from each MLTU family was more dependent for cytoplasmic accumulation on the E1b-55K protein than the shortest member. Leppard (1993) hypothesised that it was not the length of a mRNA that made it more dependent on E1b-55K but the presence of unused splice sites; the longer mRNAs from a MLTU family also contain more splice sites and potential intronic sequences than do the short mRNAs. RNAs containing such sequences may be recognised as incompletely processed by the host RNA processing machinery and so be retained in the nucleus. In the HIV system, it has been noted that the presence of a single intron in an mRNA can cause its retention in the nucleus (Chang and Sharp, 1989). Many processed MLTU mRNAs retain some intact splice sites (due to the alternative splicing events) and these could interact with the splicing machinery or nuclear matrix so preventing the export of the mRNA. Such effects would explain the observed differences in dependence of various late mRNAs on the E1b-55K protein for optimal cytoplasmic accumulation.

It is not clear however if the observed difference in the dependence of RNAs on the E1b-55K protein for cytoplasmic accumulation is due to an indirect effect of the protein's hypothesised ability to gate genes to nuclear pores or if it is a separate, distinct function of the protein. One appealing adaptation of the simple model described above is as follows: mature late mRNA is made, but not transported efficiently when not in a facilitated site (i.e. in an E1b-55K mutant infection), as the transport mechanism is non-existent or very inefficient. In contrast, in a facilitated site (due to the action of the E1b-55K/E4 ORF6-34K complex), the transport mechanism is more efficient, so is more capable of competing with the splicing machinery for the mRNA and so facilitating mRNA export. mRNAs containing no intronic sequences or splice sites would still be dependent on the E1b-55K protein for optimal cytoplasmic

accumulation as they need to be in a facilitated site, provided by the E1b-55K protein, for efficient export. However they would not have such a dependence on the E1b-55K protein as mRNAs containing splice sites, as they do not possess these possible nuclear retention signals.

Bridge and Ketner (1990) also proposed that the E1b protein is involved in the stabilisation of late viral unprocessed messages in the nucleus. This is not inconceivable as the role of E1b-55K in RNA transport may also promote stability in viral messages. Evidence for decreased stability of late mRNAs in E1b mutant infections was also observed by Leppard and Shenk (1989), however Öhman *et al.* (1993) found no evidence that E1b-55K was involved in nuclear RNA accumulation.

1.4. The E4 Region

1.4.1. E4 expression

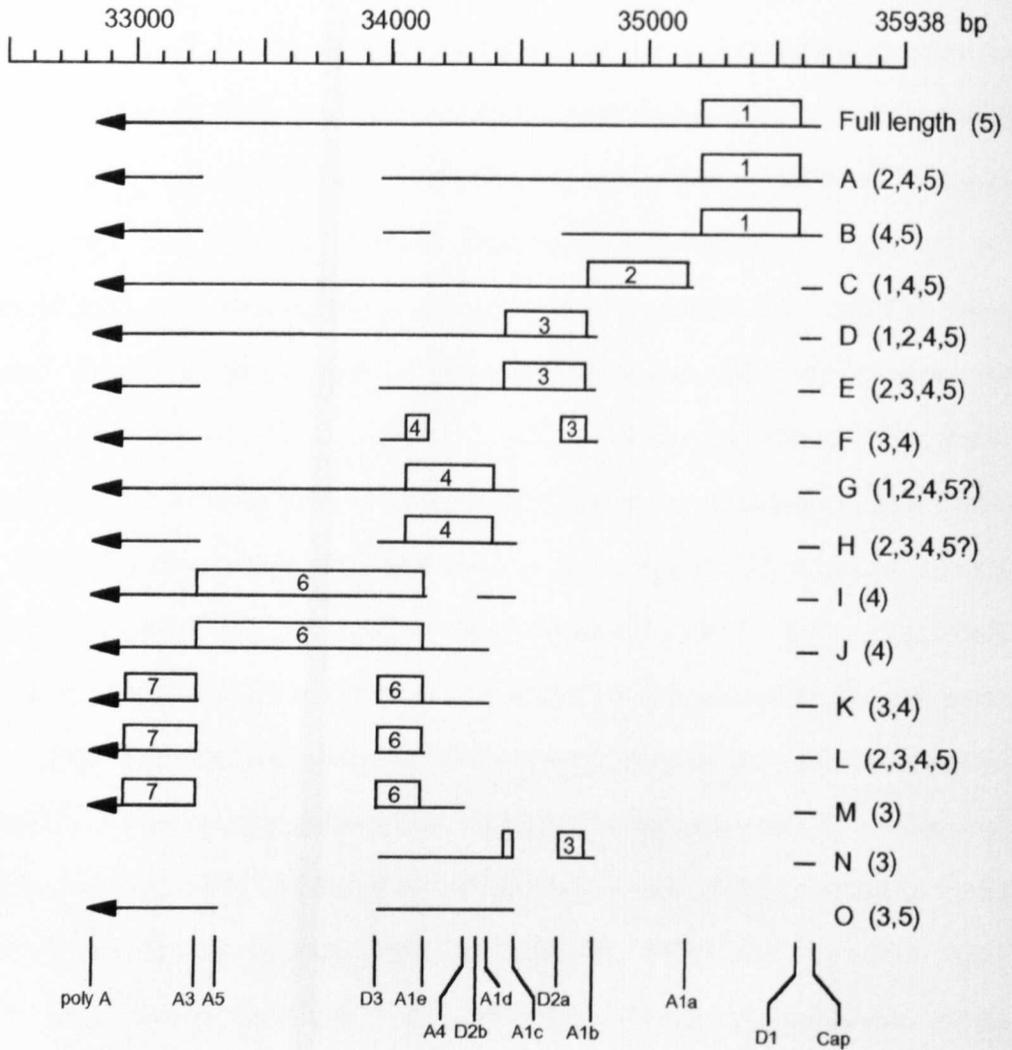
The E4 region of Ad5 is located between 32803 bp and 35614 bp on the viral genome and is transcribed from right to left. E4 transcription is induced by the indirect action of E1a, beginning in the early phase and continuing, although at a reduced rate, into the late phase of infection. Studies of the E4 promoter have identified a repeated sequence element critical for E1a-mediated E4 transcription (Lee and Green, 1987), which is known to bind a number of transcription factors including those of the activating transcription factor (ATF) family (Hai *et al.*, 1988). It is believed that the ATF proteins bind to the DNA of the promoter and recruit E1a by a CR3-specific, protein-protein interaction. This subsequently allows the E1a protein to interact with other components of the transcriptional machinery, activating transcription (Lillie and Green, 1989; Liu and Green, 1990; Martin *et al.*, 1990; reviewed by Akusjarvi, 1993). Raychaudhuri *et al.* (1987) and Rooney *et al.* (1990) have also implicated the cellular transcription factor E4F in E1a-mediated transactivation of the E4 promoter. E4F can bind to two of the ATF binding sites within the E4 promoter, with kinetics which parallel those of E4 activation (Raychaudhuri *et al.*, 1987). This E4F DNA binding

activity is dependent on phosphorylation (Raychaudhuri *et al.*, 1989), possibly regulated by the E1a protein. As yet it is unclear whether both factors (ATF and E4F) are required for E1a-mediated transactivation of the E4 promoter during an adenovirus infection (Rooney *et al.*, 1990; Jones and Lee, 1991; Bondesson *et al.*, 1992).

E4 transcription is down-regulated after the initiation of DNA replication due to the action of the DBP binding to the promoter late in the infection so sterically blocking the transcription machinery (Nevins and Winkler, 1980). *In vitro* experiments confirmed this observation, demonstrating that while E4 transcription is inhibited by the presence of DBP, the transcription of the E1a, E1b and MLTU regions is unaffected (Handa *et al.*, 1983).

There is some degree of heterogeneity in the 5' end of the E4 mRNAs, with transcripts starting within a 6 base pair region, 35614 to 35608 bp (Baker and Ziff, 1981; Hashimoto *et al.*, 1981; Virtanen *et al.*, 1984; Dix and Leppard, unpublished observations). Polyadenylation occurs at a unique site between 32803/32802 bp to produce a primary transcript of between 2805 and 2811 nucleotides in length. This is then subject to a complex pattern of differential splicing, producing up to 25 mRNAs (Berk and Sharp, 1978; Chow *et al.*, 1979; Freyer *et al.*, 1984; Tigges and Raskas, 1984; Virtanen *et al.*, 1984). Although these studies of E4-derived mRNAs come to similar conclusions they do not completely agree, possibly due to the different methods employed in the detection of the mRNAs. These studies will be discussed further in chapter 4. The nomenclature employed for the E4 mRNAs for the purposes of this thesis is the one proposed by Virtanen *et al.* (1984) with RNAs8 M, N and O identified by Freyer *et al.*, 1984 (see Fig. 1.5.).

It was observed by Tigges and Raskas (1984) that splicing of the E4 primary transcript was temporally regulated, with a 2.1 kb mRNA being the major product early in the infection while at late times a 0.8kb mRNA class predominated. It was proposed by Tigges and Raskas (1984) that this temporal change in splicing was due, in part at least, to the E2A-DBP protein, as viruses lacking this protein failed to accumulate the late 0.8kb mRNA. This was confirmed by Ross and Ziff (1992) but they also observed



1) Berk and Sharp, 1978

4) Virtanen et al., 1984

2) Chow et al., 1979

5) Tigges and Raskas, 1984

3) Freyer et al., 1984

Figure 1.5. The observed mRNA species from the E4 region with encoded ORFs shown as open rectangles (in the notation of Virtanen et al. (1984)). The right terminal portion of the Ad5 genome (bp 32500 - 35938) is shown as a line scale. The splice sites are shown below the transcription map, acceptor sites A4 and A5 are the additional sites observed by Freyer et al. (1984). mRNAs N and O are incomplete and represent the pGY4 and pGY9 cDNA clones of Freyer et al. (1984), respectively. The papers reporting the existence of each mRNA are indicated at the right in parenthesis with the key to the number code below.

that the DBP was not required for the expression of the late pattern of E4 splicing if the E4 mRNAs were expressed from a plasmid rather than a viral template. They proposed that before DNA replication a factor besides the DBP would restrict E4 splicing to the early pattern. At late times during an infection the DBP would overcome the effects of this factor and allow the expression of the late E4 splicing pattern. In transfected cells this additional factor is thought to be absent. Alternatively the difference in E4 splicing observed between an infection and a transfection may be due to the state of the DNA template. Adami and Babiss (1991) demonstrated that two identical templates can produce RNAs which are spliced differently in the same cell suggesting the state of the DNA template is very important in determining the splicing pattern of an RNA (see section 1.2.4.). It is conceivable that the DBP protein is required to produce late E4 splicing in an infection and not in a transfection, simply because the DNA templates are in different microenvironments (it is believed that late transcription and RNA processing occurs within viral inclusion bodies which are not observed early in the infection or presumably in transfections (Moyne *et al.*, 1978; Zhonghe *et al.*, 1986; Walton *et al.*, 1989; Moen *et al.*, 1990; Jiménez-García and Spector, 1993)).

Sequence analysis combined with the information obtained from mRNA mapping indicated that E4 potentially encodes at least 7 polypeptides (Tigges and Raskas (1984) suggested as many as 16), designated ORF1, ORF2, ORF3, ORF3/4, ORF4, ORF6, and ORF6/7, of which four, ORF3, ORF4, ORF6 and ORF6/7, have been identified during the course of an Ad5 infection in tissue culture (Downey *et al.*, 1983; Sarnow *et al.*, 1982, 1984; Cutt *et al.*, 1987; Bridge *et al.*, 1993b). The ORF1 protein has also been identified in an adenovirus infection but only in Ad9 (Javier, 1994).

1.4.2. The functions of the E4 proteins

1.4.2.1. Introduction

The functions of the proteins from the E4 region have been studied over the last 10 years, primarily by mutation analysis. These studies revealed that the E4 proteins were

required for the normal progression of the virus infection into the late phase of the infectious cycle. Viruses lacking the entire E4 region display a very complex phenotype, with defects in viral DNA replication, accumulation of late RNA, efficiency of late protein synthesis, particle assembly and a failure to shut off host cell protein synthesis (Halbert *et al.*, 1985; Weinberg and Ketner, 1986; Yoder and Berget, 1986; Cutt *et al.*, 1987; Falgout and Ketner, 1987; Bridge and Ketner, 1990). However, mutation of the individual E4 reading frames revealed an unexpected result, as only mutation of the E4 ORF6-34K protein compromised the virus' ability to grow in human cells and then only to a moderate extent (Halbert *et al.*, 1985). This suggested that the other ORFs were not required for growth in tissue culture, although they presumably play an important role during virus propagation in its natural host. The details of this work and subsequent studies are discussed below.

1.4.2.2. ORF1 and ORF2

To date, no evidence has been demonstrated for the presence of the ORF1-14K or ORF2-14K proteins in Ad5 or Ad2-infected cells. Mutation of the ORF1-14K or the ORF2-14K reading frames in Ad5 had no significant effect on virus propagation (Halbert *et al.*, 1985; Huang and Hearing, 1989a). Huang and Hearing (1989a), using a virus mutated in ORFs 1, 2 and 3 (dl1-3), also demonstrated that viral DNA accumulation in HeLa cells was unaffected in the absence of these proteins, as was late protein synthesis. However it should be noted that Huang and Hearing (1989a) also observed that a virus mutated in all the E4 proteins apart from ORFs 1 and 2 grew 10 fold better than a total E4 deletion (dl366) suggesting that these two ORFs do make a minor contribution to viral growth in tissue culture.

Recently Javier (1994) raised an antiserum to the Ad9 E4 ORF1-14K and found the protein in Ad9-infected tissue culture cells. Ad9 elicits oestrogen-dependent mammary tumours in female rats (Ankerst *et al.*, 1974; Ankerst and Jonsson, 1989; Javier *et al.*, 1991) while Ad26, another oncogenic subgroup D adenovirus very closely related to Ad9, displays no evidence of this, producing sarcomatous tumours at the site of

inoculation (Javier *et al.*, 1991). Javier *et al.* (1992) and Javier (1994) concluded that Ad9 encoded a novel protein that targets tumourigenesis to the mammary gland. On investigation it was established that the E4 ORF1-14K protein was responsible for this observed phenotype. It was also found that Ad9 E4 ORF1-14K protein could transform continuous rat embryo fibroblasts (CREF cells) in tissue culture (Javier, 1994). These data indicate that the Ad9 ORF1-14K protein is an oncoprotein.

The Ad9 ORF1-14K protein is probably not the sole cause of mammary tumours in female rats infected with Ad9, but probably interacts with the well known adenoviral oncogenes E1a and E1b. Javier *et al.* (1991) implicated E1a in Ad9 oncogenesis when E1a mRNA was detected in rat mammary tumours, however, no E1b mRNA was found indicating that E1b is either not involved or is active at undetectable levels. The Ad9 ORF1-14K protein can transform CREF cells without the aid of any other proteins. Surprisingly the Ad26 ORF1-14K protein can also transform CREF cells in tissue culture, although the Ad26 virus cannot elicit mammary oncogenesis in rats like Ad9. Javier (1994) proposed that this could be due to either a difference in the ORF1-14K protein in the two viruses or due to a difference between other regions of the viruses, i.e. the two ORF1-14K proteins could be identical with respect to transformation and oncogenesis but another region could dominantly repress mammary oncogenesis in Ad26 infections. Mutant analysis suggests that the E4 ORF2 and/or ORF3 regions are possible candidates for such a repressive function (it was noted that the Ad9 ORF3 protein had negative effect on mammary oncogenesis), (Javier *et al.*, 1992; Javier, 1994).

Ad5 displays no oncogenic potential for either mammary or other tissues *in vivo*, although its E1 region is able to transform CREF cells (Horwitz, 1990a). The E4 ORF1 sequences of Ad5 and Ad9 display strong similarities (see chapter 3) suggesting that they have similar functions in an adenovirus infection, but Javier (1994) found that the Ad5 E4 ORF1-14K could not transform CREF cells in tissue culture. However, the effect of the Ad5 E4-ORF1-14K protein on cell growth may be more subtle; work carried out with the group A adenovirus, Ad12 ORF1-14K protein demonstrated that

it had no transforming potential in CREF cells (Javier, 1994) but that the Ad12 E4 region was able to confer an increased potential for growth in soft agar on Ad12 E1-transformed cells (Shiroki *et al.*, 1984). This effect may have been due to the Ad12 E4 ORF1-14K protein; what effect the Ad5 ORF1-14K protein has on similar cells in soft agar is unknown.

1.4.2.3. ORF3 and ORF6

The E4 ORF3 region encodes a nuclear protein of about 11K in size (Sarnow *et al.*, 1982; Downey *et al.*, 1983) while the ORF6 protein is known to be 34K in size but migrates at 25K on SDS-polyacrylamide gels (Sarnow *et al.*, 1984). Cellular localisation studies by Cutt *et al.* (1987) have demonstrated that about 50% of the ORF6-34K protein is found complexed with the E1b-55K protein (Sarnow *et al.*, 1984) within the nucleus of the infected cell. These two E4 proteins will be discussed together as it has been observed that they have somewhat redundant functions in an adenovirus lytic infection.

Halbert *et al.* (1985) observed that a virus lacking an intact E4 ORF6-34K protein was modestly defective for growth, with the first observable defect in the life cycle occurring at the onset of viral DNA replication. It was noted that both viral DNA and protein production was perturbed in the E4 ORF6-34K mutant virus infection, and that the virus failed to shut-off host cell metabolism. Measurements of RNA production revealed that viral RNAs, although produced, failed to accumulate in the cytoplasm which suggested that the E4 ORF6-34K protein was involved in post-transcriptional RNA metabolism. This defect in RNA accumulation had also been observed by Logan *et al.* (1984) with an E1b-55K mutant, and, together with the data provided by Sarnow *et al.* (1984) documenting that the E1b-55K and E4 ORF6-34K formed a physical complex in infected cells, suggested that the E1b-55K and E4 ORF6-34K proteins acted as a functional complex. Cutt *et al.* (1987) compared late protein synthesis in three different mutant viruses (E1b-55K minus, E4 ORF6-34K minus, and a double mutant (E1b-55K/E4 ORF6-34K minus)) and found no observable difference between

the three mutant viruses, all were equally defective in late protein synthesis. The identical phenotypes of the three mutants confirmed the proposal of Halbert *et al.* (1985) that the E1b-55K and E4 ORF6-34K proteins function as a complex to affect late gene expression. Similar results were obtained by Bridge and Ketner (1990). However, Sarnow *et al.* (1984) also observed that DNA replication was delayed with the E4 ORF6-34K mutant virus but not with the E1b-55K mutant virus, suggesting that the E4 ORF6-34K protein has other functions in the virus life cycle apart from those performed in association with the E1b-55K protein.

Halbert *et al.* (1985) noted that the E4 ORF6-34K mutant virus displayed the same range of phenotypes as a virus defective in its entire E4 region, but the defects of the E4 ORF6-34K mutant virus were less severe. This suggested that the E4 region encoded an additional protein(s) important for viral growth which were functionally related to the E4 ORF6-34K protein and so able to rescue partially a defect in the E4 ORF6-34K protein.

DNA synthesis

The defect in DNA replication in E4 mutants observed by Halbert *et al.* (1985) was also seen in a number of other studies (Weinberg and Ketner, 1986; Huang and Hearing, 1989a; Weiden and Ginsberg, 1994). However, both Cutt *et al.* (1987) and Bridge and Ketner (1989) reported no difference in the level of DNA accumulation between E4 ORF6-34K mutant and wild type virus infections. This discrepancy in the results may be due to the experimental design as the results of Huang and Hearing (1989a) indicate that the defect in DNA replication is a delay in DNA accumulation in the infection. In their experiment there is a distinct difference in the levels of viral DNA between the mutant and wild type viruses at 24 hours p.i. but by 48 hours p.i. no difference in viral DNA levels can be seen. It is conceivable that this defect may be missed if growth conditions are varied (e.g. an increased MOI) and DNA levels are observed at only one time point (as in Bridge and Ketner, 1989). Alternatively, the

difference in DNA replication observed between these groups may be a reflection of the different mutants used.

Huang and Hearing (1989a) also implicated the E4 ORF3-11K protein in DNA replication. Viruses mutated in ORF3 accumulated viral DNA at levels comparable to wild type while ORF6 mutated viruses demonstrated a delayed onset of DNA replication and viruses mutated in both reading frames had a severe defect in DNA accumulation (comparable to that of an E4 minus virus). This, combined with the fact that a virus expressing only ORF3 in its E4 region demonstrated kinetics of DNA accumulation similar to that of the ORF6 mutated virus, suggested that the ORF3 and ORF6 proteins have redundant (but probably not equal) roles in DNA replication. Weiden and Ginsberg (1994) observed that DNA from viruses mutated in both ORF3 and ORF6 formed concatemers during viral DNA synthesis in KB cells while the presence of just one of these reading frames in the virus suppressed this effect. Weiden and Ginsberg (1994) tentatively suggest that these proteins are involved in resolving higher-order forms of the genome to monomers during the course of DNA replication. It is unclear how the E4 ORF3-11K and E4 ORF6-34K proteins exert an effect on DNA replication. Their role is probably regulatory or indirect as no E4 products are required in purified cell-free DNA replication systems (reviewed by Stillman, 1989). Experiments looking at the phenotypes of E1b-55K, E4 ORF3-11K and E4 ORF6-34K combination mutants suggest that facilitation of DNA synthesis may be mediated by the E1b-55K/E4 ORF6-34K complex acting in parallel with the E4 ORF3-11K protein (Bridge and Ketner, 1990). However, it should be noted that defective DNA synthesis has never been associated with a virus lacking E1b-55K unless it also contains lesions in the E4 ORF3 and/or ORF6 regions. It would appear therefore that the roles of the two proteins in the E1b-55K/E4 ORF6-34K complex are not the same.

As yet it is unclear whether the over-expression of the E2a-72K DBP in some E4 mutants (Halbert *et al.*, 1985; and inconsistently observed in some experiments by Huang and Hearing, 1989a) is significant in the defect observed in DNA replication.

Weinberg and Ketner (1986), Huang and Hearing (1989a), Sandler and Ketner (1989) and Bridge and Ketner (1990) all concluded that the defect in DNA replication did not fully account for the complicated phenotype associated with the E4 region since late protein production was defective in E4 mutant cells even when the DNA levels were normal (i.e. late in the infection, 48 hours post infection). This is a rather tenuous argument as the defect observed in DNA accumulation at earlier time points may have a role in the defect subsequently observed in protein production when DNA levels are comparable.

Late RNA and protein expression

The redundancy of the E4 ORF3-11K and E4 ORF6-34K proteins in affecting DNA replication has also been observed in late protein expression (Huang and Hearing, 1989a; Bridge and Ketner, 1989). Huang and Hearing (1989a) analysed a number of E4 mutants for late protein production and found that while lesions in just the E4 ORF3-11K or E4 ORF6-34K protein resulted in no or only a moderate effect respectively, viruses mutated in both E4 ORF3-11K and ORF6-34K were severely defective for late protein synthesis. This would indicate that these proteins regulate late protein synthesis independently. Similar results were obtained by Bridge and Ketner (1989).

A number of groups had suggested that the above defect in late protein synthesis was due to the defect in late mRNA accumulation in the cytoplasm (Halbert *et al.*, 1985; Weinberg and Ketner; 1986; Bridge and Ketner, 1989; Huang and Hearing, 1989a; Sandler and Ketner, 1989). Bridge and Ketner (1990), working with a series of E4 and E1b mutants, proposed that this defect in late mRNA accumulation in the cytoplasm was due to defects in both nuclear RNA stability and mRNA transport from the nucleus to the cytoplasm (although the transport defect may be a defect in mRNA stability in the cytoplasm). Their results indicated that the E4 ORF6-34K and the ORF3-11K protein work in parallel to confer stability on unprocessed and processed late viral RNA in the nucleus of infected cells (also noted by Bridge and Ketner, 1989;

Sandler and Ketner, 1989). The semi-quantitative data suggest that these proteins exert most of their effect on RNA stability on unprocessed RNA transcripts. The E4 ORF3-11K protein appears less efficient than the E4 ORF6-34K in conferring RNA stability indicating that they may be involved in a different pathways each leading to nuclear RNA accumulation. It was also observed that the E1b-55K has a role in nuclear RNA stability but it was noted that the E1b-55K and E4 ORF6-34K proteins of the heterodimer do not function equivalently opposite the E4 ORF3-11K protein in nuclear RNA stability with the E1b-55K apparently being less efficient in conferring stability to RNA in the nucleus. This suggests that the E4 ORF6-34K and the E1b-55K proteins may have functions independent of the complex or play non-equivalent roles within the complex. A similar conclusion was reached in transfection studies by Ohman *et al.* (1993) where it was concluded that E1 products gave no, or only a small, contribution to tripartite leader mRNA accumulation in the nucleus. These functions are in addition to the role of the E1b-55K/E4 ORF6-34K heterodimer in regulating mRNA transport as already discussed (section 1.3.). No evidence was found that suggested the involvement of the E4 ORF3-11K protein in this latter process.

Bridge *et al.* (1991) examined the substrate specificity of the E4 proteins as previously the only late mRNAs studied were those from the MLTU. Working with pIX, IVa2 and E2-L transcripts they found that only the pIX displayed any dependence on the E4 products, IVa2 and E2-L transcripts being completely independent for optimal cytoplasmic accumulation. The pIX transcript, however, was independent of the E4 proteins for nuclear RNA stability and only slightly dependent on the ORF6-34K protein for mRNA export. These results appear to conflict with the results published by Leppard (1993) where it was observed that all the late mRNAs were dependent, to varying degrees, on the E1b-55K protein (the other half of the E1b/E4 heterodimer involved in RNA transport) for optimal cytoplasmic accumulation. This could be explained by the E1b-55K protein having additional functions, within the heterodimer, in mRNA export. Alternatively it could reflect differences in experimental technique.

Bridge *et al.* (1991) explain their result that only the MLTU transcripts are strongly dependent on the E4 proteins for nuclear accumulation by suggesting that the MLTU leader sequences are important in conferring E4-dependence on a transcript. Nordqvist and Akusjarvi (1990) demonstrated that E4 post-transcriptionally stimulated accumulation of tripartite leader-containing mRNAs by a mechanism that required an intron in the 5' UTR of the nuclear precursor RNA. Öhman *et al.* (1993) and Nordqvist *et al.* (1994) subsequently demonstrated the involvement of the ORF3 and ORF6 regions in the accumulation of correctly spliced tripartite leader mRNA in the nucleus and suggested that they have a role in the regulation of pre-mRNA splicing. However the ORF3-11K and ORF6-34K proteins did not show any redundancy in their splicing activities with ORF3-11K facilitating exon inclusion and ORF6-34K facilitating exon exclusion (see section 1.2.4.). No evidence for the involvement of the E1b-55K protein in regulation of splicing was found (Ohman *et al.*, 1993).

It is unclear how the E4 proteins confer stability on late viral RNA in the nucleus. The E4 proteins could bind to the RNA directly or initiate the binding of the hnRNP proteins to the RNA so protecting the transcripts from nuclease degradation. Alternatively the microenvironment of the viral inclusion body may be altered by an effect of the E4 proteins on DNA replication so enhancing RNA stability. However Nordqvist *et al.* (1994), Ohman *et al.* (1993) and Nordqvist and Akusjarvi (1990) suggested that RNA stability is the result of the E4 proteins mediating MLTU RNA metabolism. They argued that the E4 products may enhance the MLTU RNA levels within the nucleus by facilitating spliceosome assembly which presumably would protect the RNA against nuclease activity. More work needs to be undertaken to explore the interaction of DNA replication, RNA stabilisation and RNA processing within the infected nucleus before any definite conclusions can be reached as to how these E4 proteins work at this level.

Host cell shut-off

Huang and Hearing (1989a) found redundancy in ORF3 and ORF6 function in the shut-off of host cell protein synthesis. This may be a secondary effect of the ORF3 and ORF6 proteins i.e. due to the production of MLTU derived proteins which inhibit cellular translation (see section 1.2.6.). Halbert *et al.* (1985) suggested that this defect in host cell shut-off could be due partially to a block in the accumulation of cellular mRNAs in the cytoplasm (as observed by Pilder *et al.* (1986a,b) for the E1b-55K protein) but this is believed not to be the case (see section 1.2.6.1.). As the ORF3 and ORF6 products have been shown to be involved in viral RNA stability (Bridge and Ketner, 1989, 1990; Sandler and Ketner, 1989), Sandler and Ketner (1991) studied the stability of three representative cellular mRNAs in wild-type and ORF3/ORF6 mutant infections to see if the ORF3-11K and ORF6-34K products modulate host cell shut-off by altering host cell RNA stability. However they concluded that these E4 products have only a small effect in the stability cellular mRNAs.

Summary

Most of the groups working with E4 ORF3-11K and E4 ORF6-34K conclude that they function in association with the viral inclusion bodies in the nucleus. The complex phenotype observed with mutations of these proteins may be the result of the proteins being multifunctional, acting on DNA replication, RNA splicing, RNA stability and RNA transport. However, it is more likely that these proteins affect only one or two events directly and that the complex phenotype observed is an indirect result of this. At present it is unclear at which step(s) the E4 proteins act; the delay in DNA accumulation could be due to the altered expression of MLTU mRNA, and so altered late protein synthesis, in E4 mutants, or the altered MLTU mRNA expression may be due to the reduction in DNA accumulation, or the two phenotypes may be completely separate. It would appear, however, that the E4 proteins function at least at two steps in the viral life cycle with the effect of the E4 proteins on viral DNA accumulation and RNA stability in the nucleus being distinct from the secondary function of the E4

ORF6-34K protein (along with the E1b-55K protein) in mRNA transport to the cytoplasm (Bridge *et al.*, 1990; Ohman *et al.*, 1993).

For a summary of the conclusions from the studies of the E4 ORF3-11K and ORF6-34K proteins see Fig.1.6.

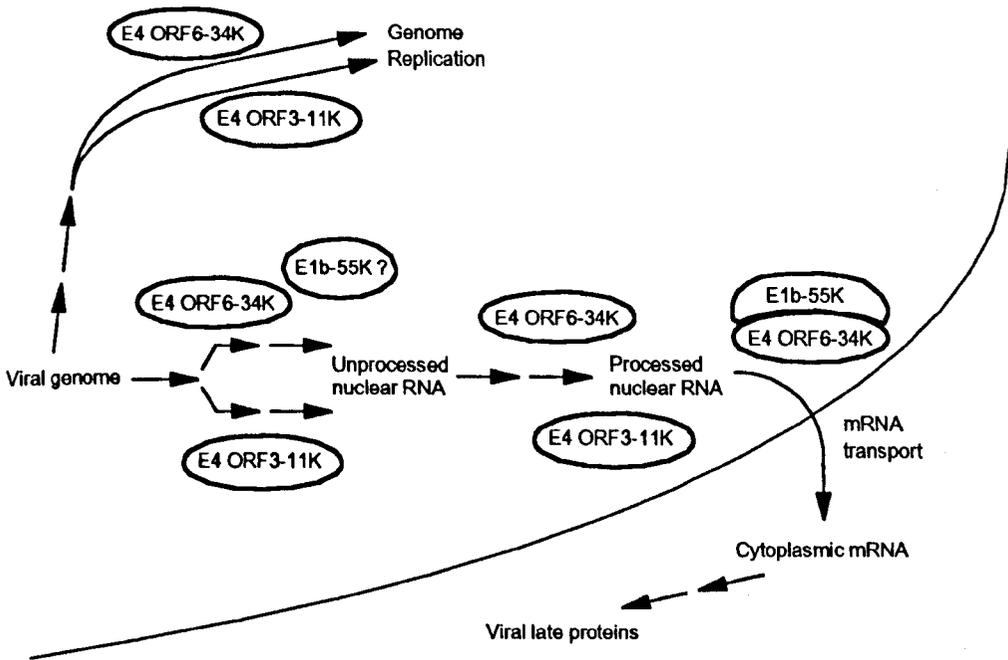
1.4.2.4. ORF3/4

No evidence for the existence of the E4 ORF3/4-7K protein in infected cells has yet been found. Mutation analysis does not reveal any phenotype other than those observed in ORF3-11K or ORF4-13K mutants.

1.4.2.5. ORF4

The E4 ORF4-13K protein was first identified in adenovirus infected cells by Bridge *et al.* in 1993b. By mutation analysis, Bridge *et al.* (1993b) also proposed that the E4 ORF4-13K protein was involved in DNA replication. As mentioned in the above sections, viruses containing gross deletions of the E4 region show a defect in DNA replication. However, when characterising a set of E4 mutants, Bridge and Ketner (1989) observed that a mutant lacking E4 ORF3 and ORF6 but retaining ORF4 displayed an even more severe defect in DNA accumulation than the total E4 deletion. Bridge *et al.* (1993b) mutated the first methionine of ORF4 in this mutant and found that this more severe defect in DNA accumulation was lost. They concluded that the E4 ORF4-13K protein was necessary for the inhibition of viral DNA replication. It would appear that the ORF4 product antagonises the effect of the ORF3 and 6 products (both involved in promoting DNA replication). The interplay of E4 ORF3, 4 and 6 products may regulate the onset and rate of DNA synthesis in adenovirus infected cells. However, it should be noted that Huang and Hearing (1989a) observed a defect in DNA replication no more severe than that observed in a gross E4 mutant for a mutant virus just expressing the E4 ORF4-13K protein in its E4 region. There are no obvious reasons why there is a discrepancy in the results from these two groups

Figure 1.6. The functions of the E4 ORF3-11K, E4 ORF6-34K and the E1b-55K proteins in the lytic viral cycle. It is unclear whether the functions of E4 ORF3-11K and E4 ORF6-34K in DNA replication, nuclear stability of viral late messages and late mRNA splicing are separate functions of these proteins. Two pathways are depicted for genome replication and unprocessed nuclear RNA accumulation as it appears that the E4 ORF3-11K and the E4 ORF6-34K proteins function by different mechanisms in these two processes. It is also unclear whether or not E1b-55K exerts an effect on unprocessed late viral mRNA stability.



concerning the ORF4 region. More studies of the E4 ORF4 region need to be undertaken before the whole picture can be obtained.

If Bridge *et al.* (1993b) are correct, the E4 ORF4-13K protein may negatively regulate viral DNA synthesis in a number of ways. Firstly it may directly interact with the proteins involved in DNA replication so reducing the rate of DNA synthesis. Secondly it may indirectly interfere with DNA replication by altering the organisation of the viral replication complex. Thirdly the E4 ORF4-13K protein may regulate DNA replication via phosphorylation events. Müller *et al.* (1992) demonstrated that the ORF4-13K protein has an activity that inhibits a cellular kinase, or activates a cellular phosphatase, which regulates E1a and c-Fos phosphorylation. Since no exhaustive search for ORF4 targets has been made, its effects may extend to phosphoproteins involved in DNA replication, such as NF-III and the viral DNA polymerase (the activity of both of these proteins in viral DNA replication is known to be regulated by phosphorylation (reviewed by Stillman, 1989). Fourthly the E4 ORF4-13K protein may alter the expression of the DNA replication proteins. This could occur either at the level of transcription or of RNA metabolism. Müller *et al.* (1992) demonstrated that ORF4-13K altered the levels of the general transcription factor, AP-1, and altered the activity of the E1a proteins which, via effects on the E2 promoters, could change the levels of the DNA replication proteins. Alternatively the ORF4-13K protein may have a role in post-transcriptional RNA metabolism. It is known that ORF4-13K antagonises the effects of the E4 ORF3-11K and E4 ORF6-34K proteins in DNA replication and both the ORF3 and 6 products are involved in the metabolism of late viral mRNA. Maybe these three proteins interact in some way to modulate the post-transcriptional processing of E2 messages.

The observations by Müller *et al.* (1992) that the ORF4-13K protein is involved in regulation of the phosphorylation of c-Fos and E1a (both are involved in the regulation of adenovirus transcription) suggest that the protein has a role in general transcriptional control during an adenovirus infection. This effect of ORF4-13K seems to be specific to c-Fos and E1a as other phosphoproteins like E1b-55K and JunB are

not affected. It is unclear if the involvement of the ORF4-13K protein in phosphorylation events in the infected cell is related to the possible role of the protein in DNA replication.

Lazinski *et al.* (1989) identified a potential RNA binding domain within the E4 ORF4-13K protein. The significance of this motif in the function of the E4 ORF4-13K protein has yet to be established.

1.4.2.6. ORF5

The E4 ORF5 reading frame was first proposed by Hérisse *et al.* (1981). This ORF had no AUG codon but was proposed to be fused onto the ORF2 reading frame by RNA splicing to create an ORF1/ORF2/ORF5 fusion protein. However subsequent analysis by Freyer *et al.* (1984), Tigges and Raskas (1984), and Virtanen *et al.* (1984) found no evidence for the existence of an mRNA that could encode such a protein. As a result no attempts have been made to identify such a protein in infected cells so this potential protein will not be discussed here.

1.4.2.7. ORF6/7

The E4 ORF6/7 region encodes a 19.5K nuclear protein by the fusion of the N-terminal part of open reading frame 6 with the open reading frame 7 (Cutt *et al.*, 1987). This protein has a half life of just 4 hours (Cutt *et al.*, 1987) and is involved in the transactivation of the E2a promoter. Kovesdi *et al.* (1986a,b) noticed that the binding of the cellular transcription factor E2F to the E2a promoter was dramatically induced in nuclear extracts from adenovirus cells as compared to extracts of non-infected cells. The viral function responsible for this induction was subsequently found to be the E4 ORF6/7-19.5K protein (Huang and Hearing, 1989b; Reichel *et al.*, 1989; Marton *et al.*, 1990; Neill *et al.*, 1990; Raychaudhuri *et al.*, 1990).

There are two binding sites for E2F within the E2 early promoter (Siva Raman and Thimmappaya, 1987; Yee *et al.*, 1987). The ORF6/7-19.5K protein has been shown to bind to the transcription factor E2F and to induce the co-operative binding of two E2F

molecules to these sites (two-site complex), thereby activating transcription (Huang and Hearing, 1989b; Marton *et al.*, 1990; Neill *et al.*, 1990; Raychaudhuri *et al.*, 1990). Studies by Obert *et al.* (1994) indicate that this 'two-site complex' contains two E2F molecules and at least two ORF6/7-19.5K proteins. It is believed that the E4 ORF6/7-19.5K does not cause the association of E2F molecules before binding to the promoter, but provides a direct dimerisation function at the promoter or induces a structural change in E2F bound to its target site which facilitates the binding of a second E2F molecule.

Mutation analysis of the E4 ORF6/7 region by O'Conner and Hearing (1991) found that only the C-terminal 70 amino acids of the 150 amino acid ORF6/7 protein are required to complex with E2F and stabilise its binding at the E2a promoter, so activating transcription. Similar results were obtained by Neill and Nevins (1991) and Obert *et al.* (1994) suggesting that the N-terminus, containing ORF6 sequences, is completely dispensable and is required solely for the purposes of expression. The E4 ORF6/7-19.5K C-terminus can be divided into three separate regions, two regions required for interaction with E2F, which flank a third region required for stable two-site complex formation (Neill and Nevins, 1991). Neill and Nevins (1991) observed that mutations located in this third region abolish the formation of the two-site complex but do not impair the interaction of ORF6/7-19.5K with E2F. They suggested that a helix-loop-helix domain within this third region is important in E4-mediated E2F dimerisation, although Obert *et al.* (1994) suggest this motif is not the key element in dimerisation. It was also noted that ORF6/7-19.5K mutants which failed to promote E2F dimerisation still could transactivate, although not as efficiently, the E2a promoter (Neill and Nevins, 1991). This would suggest that they can activate E2a transcription in the absence of stable E2F complex formation. Neill and Nevins (1991) suggested that the ORF6/7-19.5K protein can transactivate the E2a promoter by two mechanisms; firstly through the formation of a stable two-site E2F/ORF6/7-19.5K complex and secondly through a one site E2F/ORF6/7-19.5K complex. In the second instance the ORF6/7-19.5K protein may act as a transcriptional activation

domain for E2F, however it should be noted that E2F can activate transcription from a number of promoters independently of ORF6/7-19.5K.

The E4 ORF6/7-19.5K protein is not essential for viral propagation in HeLa cells (Halbert *et al.*, 1986, Bridge and Ketner, 1989; Hemström *et al.*, 1991), since ORF6/7 mutant viruses grow to levels comparable to wild type. This suggests that E4-assisted activation of the E2a promoter is not critical for E2 expression in tissue culture. So why has the virus developed this mechanism of activation? The E4 ORF6/7-19.5K association with E2F may be important in viral infections of the natural host. 'Adenovirus mainly infects quiescent cells that have escaped from the cell cycle; to be able to replicate efficiently in these cells, the virus probably needs to activate certain components of the replication machinery of its host' (Hemström *et al.*, 1991). This is known to be accomplished in part by the E1a proteins which have been shown to release transcription factors important in cell growth, such as E2F, from inactive complexes. It is possible that the ORF6/7 interaction with E2F is important in regulating virus induced cellular proliferation, as Mudryj *et al.* (1990) have demonstrated that E2F is involved in the activation of a number of genes involved in cell proliferation. However, none of the promoters of these genes contains two inverted E2F binding sites characteristic of the E2a promoter. Nevertheless the E4 protein may be involved in the activation of transcription of a number of cellular promoters which contain only one E2F site (Mudryj *et al.*, 1990) through a one site complex as described above.

1.5. Project objectives.

It is evident from these studies that E4 expression is highly complex and probably highly regulated. It is also evident that very little is known about the function of some of the E4 proteins such as the ORF1, ORF2 and ORF3/4 proteins of Ad5. As a result the objectives of this study were to:

- (1) Characterise the expression of the E4 mRNAs over the course of an adenovirus infection and examine the regulation of this expression, in particular the role of the E1b-55K protein in E4 mRNA expression.

- (2) Identify and characterise the remaining unidentified E4 proteins, predicted to be encoded by the E4 mRNAs (i.e. E4 ORF1, ORF2 and ORF3/4), in infected cells and if possible attribute functions to these proteins.

Chapter 2.

Materials and methods

2.1. Materials

2.1.1. Common Buffers and Solutions

2YT: 1.6% (w/v) bacto-tryptone, 1% (w/v) yeast extract, 86 mM NaCl

Chloroform/amyl alcohol: 96% (v/v) chloroform, 4% (v/v) amyl alcohol

Hypotonic buffer: 10 mM Tris pH7.5, 10 mM NaCl, 1.5 mM MgCl₂.

Isotonic buffer (protein extraction): 150 mM NaCl, 50mM Tris pH8.0, 1% NP40.

LB: 1% (w/v) bacto-tryptone, 0.5% (w/v) yeast extract, 173 mM NaCl

LB-agar: 1% (w/v) bacto-tryptone, 0.5% (w/v) yeast extract, 173 mM NaCl, 1.5% (w/v) bacto-agar

PBS: 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na₂HPO₄.7H₂O, 1.4 mM KH₂PO₄

Phenol/chloroform: 50% (v/v) phenol/TNE , 50% (v/v) chloroform/amyl alcohol

Phenol/TNE: liquified phenol (Fisons, Tris-buffered) equilibrated with excess TNE

RIPA: 150mM NaCl, 50mM Tris pH8.0, 1% NonidetP40, 0.5% sodium deoxycholate, and 0.1% SDS.

SOC: 2% (w/v) bacto-tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM glucose

TD: 25 mM Tris base, 137 mM NaCl, 5 mM KCl, 0.7 mM Na₂HPO₄

TE: 10 mM Tris.HCl pH 8.0, 1 mM EDTA pH 8.0

TBE: 89 mM Tris base, 89 mM boric acid, 1 mM EDTA pH 8.0

TNE: 10 mM Tris.HCl pH 8.0, 100 mM NaCl, 1 mM EDTA pH 8.0

TS: 25 mM Tris base, 137 mM NaCl, 5 mM KCl, 0.7 mM Na₂HPO₄, 0.9 mM CaCl₂, 1 mM MgCl₂

10 x Taq Buffer: 500 mM KCl, 100 mM Tris.HCl pH 8.4, 1 mg/ml gelatin

2.1.2. Oligonucleotide primers (see appendix 2)

All primers are in 5' to 3' orientation

IDB: ATATTGTCGTTAGAACGCGG

IDC: TGATGGATCCACTGTTATGGCTGCCG

IDD: AGGTGAATTCTCAAACATTAGAAGCCTGTC

IDE: TGATGGATCCGCAGACATGTTTGAGAGA

IDF: AGGTGAATTCCCTCAAGCAGCGAATC

IDG: AGCAAGTATATAGCATGGCC

IDH: GGAGCTGCATGTGTCAGAGG

IDJ: ACAGCTCCTCGGTCATGTCC

IDK: TTTTACTAAGCTTGCTGACTGTTATGGC

IDL: TAGCCCTGGGGAATACATAC

IDM: TTCATCCACGGTGCCTGACT

2.2. Suppliers

The chemicals used in this study were supplied by Sigma Chemical Company Ltd (Poole, Dorset), BDH Laboratory Supplies (Merck Ltd, Poole, Dorset) or Fisons Scientific Equipment (Loughborough, Leics.), unless otherwise stated below.

PBS, DMEM, neutral red and sterile distilled water were supplied by the media preparation staff of the Biological Sciences Department, University of Warwick.

Amersham International PLC (Amersham, Bucks.)

³²P-UTP, ³⁵S-Methionine, ³⁵S-dATP, Amplify, Hybond C membrane, antibody conjugates, SDS-PAGE Rainbow markers

Bio-Rad (Hemel Hempstead, Herts.)

Acrylamide, N,N'-methylene-bis-acrylamide, ammonium persulphate, TEMED

Boehringer Mannheim UK (Lewes, East Sussex)

Calf intestinal phosphatase, Proteinase K

Dako (High Wycombe, Bucks.)

1,2-phenylenediamene tablets

Difco Laboratories (Basingstoke, Hants.)

Noble agar, bacto-agar, bacto-tryptone, minimal agar, yeast extract

Dynatech Laboratories Inc. (Billingshurst, West Sussex)

ELISA plates

Gibco-BRL Life Technologies Ltd. (Renfrewshire, Scotland)

DNA modification enzymes, restriction enzymes, T4 DNA ligase, nucleases, immuno-precipitin, SDS-PAGE standards

Invitrogen Corporation (San Diego, California, USA)

Probond nickel resin

New England Biolabs (CP Laboratories, Bishop Stortford, Herts.)

DNA modification enzymes, restriction enzymes, T4 DNA ligase, nucleases

Northumbria Biologicals Ltd. (Cramlington, Northumbria)

Calf serum, foetal calf serum

Pharmacia Ltd (Milton Keynes, Bucks.)

Glutathione sepharose 4B beads, dNTPs, pGEX vector

Promega Corporation (Madison, Wisconsin, USA)

Taq DNA polymerase, TnT Coupled Reticulocyte Lysate System

Stratagene Ltd (Cambridge)

Nucrap columns

United States Bioscience (Cambridge Bioscience, Cambridge)

Sequenase DNA-dependent DNA polymerase

Whatman International (Maidstone, Kent)

3MM chromatography paper, GF/C glass microfibre filters, DE52 resin

Qiagen GmbH (Hilden, Germany)

pQE vector

All chemicals used were of molecular biology or analytical grade.

2.3 Methods

For a general review of most of the methods employed in this study see Sambrook *et al.* (1989).

2.3.1. Nucleic Acid Manipulations

2.3.1.1. Ethanol precipitation of Nucleic Acids

Nucleic acid was precipitated out of aqueous solution (minimum volume of 50 μ l) by adding salt to a final concentration of 0.3 M (sodium acetate) or 0.1 M (sodium chloride) and 2.5 volumes of ethanol. The solution was placed at -20°C for 30 min. or 15 min. at -80°C , followed by a 15 min. centrifugation at 16,000 x g. The resulting pellet was then washed using 70% (v/v) ethanol at -20°C and re-centrifuged at 16,000 x g for 10 min.. The washed pellet was vacuum dried and resuspended in the required volume of water or TE. Variations on this method are described in detail where they arise.

2.3.1.2. Nucleic acid quantification

The concentration of nucleic acid in a solution was quantified by spectrophotometry. The absorbance of the solution (diluted to give a reading between 0.05 and 0.1 (A_{260})) was read at A_{260} and A_{280} . Nucleic acid concentrations were then calculated from the A_{260} reading assuming $A_{260} = 1$ for 50 $\mu\text{g/ml}$ ds DNA, 40 $\mu\text{g/ml}$ for ssRNA or 30 $\mu\text{g/ml}$ for ssDNA. The A_{280} reading allows the calculation of the R value (A_{260}/A_{280}), an indication of the purity of the solution. Values below 1.8 for DNA or 2.0 for RNA indicate contaminants in the solution, making the estimated concentration derived from the A_{260} value inaccurate.

2.3.1.3. Phenol/chloroform extraction

An equal volume of phenol/chloroform was added to the DNA solution and mixed (for most procedures this involved vortexing but for higher molecular weight DNA (~10

kbp plus) repeated inversion was used). The heavier phenol/chloroform fraction was separated from the lighter aqueous phase by centrifugation at 15,000 x g for 5 min.. The aqueous phase was then removed and the phenol/chloroform fraction discarded. An equal volume of chloroform was then added to the aqueous phase, mixed and centrifuged as before. Again the lighter aqueous phase was removed to a fresh tube and the chloroform phase discarded. Variations on this method are described in detail where they arise.

2.3.1.4. Restriction endonuclease digestion

Restriction enzyme digests were carried out according to the supplier's instructions; the reaction buffers used were normally those supplied (by the supplier) as 10 x concentrates. Typical reactions (i.e. less than 1 µg of DNA) were carried out in 20µl while larger volumes were used when more DNA was cut. This ensured that the amount of enzyme added was less than 10% of the volume (enzyme dilutions of less than 1/10 may cause the glycerol concentration (from the storage buffer) to inhibit enzyme activity). Where possible, multiple restriction digests were carried simultaneously in a buffer that allowed activity of all the enzymes involved.

2.3.1.5. End-fill repair of 5' overhangs

The DNA polymerase I large fragment (Klenow fragment) was used to fill in the 5' overhanging ends of dsDNA (resulting from some restriction enzyme digests) so allowing blunt ended cloning. The reaction conditions used were 50 mM Tris. HCl pH 8.0, 10 mM MgCl₂, 1 mM DTT, 50 µg/ml BSA, 25 µM of each dNTP and 1 U of enzyme per µg of DNA) in 25 µl, incubated at 37°C for 30 min.. The reaction was stopped by phenol /chloroform extraction.

2.3.1.6. T4 DNA polymerase removal of 3' overhangs

3' overhangs were removed, to create blunt ends for cloning by T4 DNA polymerase (BRL). Reactions of 50 µl containing 50 mM Tris. HCl pH 8.0, 5 mM MgCl₂, 5 mM

DTT, 50 µg/ml BSA, 100 µM of each dNTP and 5 U of enzyme per µg of DNA were incubated at 11 °C for 30 min.. The reactions were stopped by phenol/chloroform extraction.

2.3.1.7. Dephosphorylation

To prevent the unwanted ligation of dsDNA termini in ligation reactions, the 5' phosphate groups were removed using calf intestinal alkaline phosphatase. Reactions were carried out in a 50 µl volume containing 20 mM Tris. HCl pH 8.0, 1 mM MgCl₂, 1 mM ZnCl₂, 1 mM spermadine and 0.1 U of enzyme for every 10 pmol of DNA termini present (1 µg of 3 kbp linear DNA contains ~1 pmol of 5' termini). 5' overhanging termini phosphate groups were removed by incubation at 37°C for 30 min.. Blunt or 5' recessed ends were dephosphorylated at 37°C with 0.1 U of enzyme/10 pmol of DNA termini for 15 min. followed by the addition of a second aliquot of enzyme (0.1 U/10 pmol of DNA termini) and incubation at 55°C for 45 min.. All the reactions were stopped by phenol/chloroform extraction.

2.3.1.8. Agarose gel electrophoresis

dsDNA fragments were separated in horizontal agarose gels containing 0.6 to 2.0% type 1 agarose and 0.5 µg/ml ethidium bromide, in 0.5 x TBE. 1/5 the sample volume of 5 x loading buffer (50% glycerol, 5 x TBE, 0.02% bromophenol blue, 0.02% xylene cyanol) was added to each sample prior to loading. Electrophoresis was carried out at 150V in 0.5 x TBE buffer until the required bands had been resolved. The fragments were visualised on a ultraviolet transilluminator and photographed using Polaroid 665 or 667 film.

2.3.1.9. Purification of dsDNA fragments from agarose gels

Fragments of under 10 kbp:

A 2-3 mm slice, the width of the fragment band, was cut out of the gel immediately in front of the band in question. The amount of running buffer in the tank was reduced so

that it just covered the gel, ensuring the hole in front of the band was then filled with buffer. The gel was then run for between 1-3 min. (according to the size of the fragment and the agarose concentration of the gel) so that the fragment ran into the hole. The buffer in the hole, now containing the fragment, was removed and extracted with phenol/chloroform. The DNA was subsequently ethanol precipitated, dried and resuspended in water or TE.

b) Fragments of over 10 kbp:

The required fragment was cut out of the gel and placed in a short section of dialysis tubing filled with 0.5 x TBE. The tubing was sealed using clips and placed in a gel tank containing 0.5 x TBE. The DNA was electroeluted from the gel into the buffer in the dialysis tubing for 30 min. at 150V. Just prior to removing the sample from the gel tank, the current was reversed for 30 seconds to remove the DNA from the tubing wall. The buffer inside the dialysis tubing was then removed and added to a TNE-equilibrated DE52 (diethylaminoethyl cellulose, Whatman) anion exchange column (constructed in a Pasteur pipette). The column was washed with 6 ml TNE and the DNA eluted from the column with 1 ml of 1M NaCl in TNE. It was then extracted with phenol/chloroform, ethanol precipitated, dried and resuspended in water or TE.

2.3.1.10. Ligation of dsDNA

Ligations between DNA fragments with complementary ends (blunt and overhanging) were carried out using T4 DNA ligase (BRL) in a volume of 20 to 50 μ l. The reaction conditions used were 40 mM Tris. HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 0.5 mM ATP, 50 μ g/ml BSA and 1 U of enzyme (if the ligation was between blunt ended fragments then 10 U of enzyme was used). A typical ligation of insert DNA and vector was done at a molar ration of about 5:1, using 50-200 ng of vector DNA. Ligations between cohesive ends were carried out 15°C for about 16 hours, to encourage the annealing of the DNA, while blunt- ended ligations, where annealing was not a factor, were carried out at room temperature for 4 hours or more.

2.3.1.11. Polymerase chain reaction (PCR)

DNA was amplified to a high level using the PCR. Two DNA oligonucleotides were created which complemented either end of the section of DNA which was to be amplified. 200 ng of each primer was mixed with 200ng template DNA, 1mM dNTPs, 1 x Taq buffer, 2.5 mM MgCl₂ and 3 units Taq polymerase, in a total volume of 25 µl. A drop of paraffin was placed on top of the mixture to prevent evaporation during the thermal cycling step. Using a DNA Thermal Cycler (Perkin Elmer Cetus) the reaction mixtures were cycled through 17 cycles of 94 °C for 1min., 55 °C for 1 min. and 72 °C for 3 min. followed by 3 cycles of 94 °C for 2 min., 55 °C for 2 min. and 72 °C for 10 min. to ensure that the ends of the PCR product were uniform. The aqueous phase was then withdrawn and extracted with phenol/chloroform. The efficiency of the PCR was estimated by running a sample of the product on an agarose gel, this also confirmed that the correct length of DNA was amplified.

2.3.1.12. Site-directed mutagenesis using PCR

This method is taken from Ito et al (1991). 4 primers were constructed IDJ, IDK, IDL and IDM (section 2.1.2. and see Fig. 2.1.), IDJ and IDM (about 20 nucleotides long) matched the DNA to be amplified, while IDK and IDL (about 30 nucleotides long) had mismatches to the DNA to be amplified. Primer IDK had mismatches across the region which was to be mutagenised (A) while primer IDL had a mismatch which destroyed the restriction site RS1 when used in the PCR. The first 2 PCRs carried out used the primers IDJ and IDK in one reaction and IDL and IDM in the other, and were set up as described above (2.3.1.11). The reactions were then cycled through 20 cycles of 94 °C for 1 min., 55 °C for 1 min. and 72 °C for 3 min.. The aqueous phase of the reactions were then withdrawn from the reaction tubes and 2.5 µl of each were run on an agarose gel (2.3.1.8) to test the reaction products. 2.5 µl of these samples (P1 and P2) were then denatured and annealed in the presence of the primers IDJ and IDM in the following reaction mixture by heating to 85 °C for 2 min. then allowing the mixture to cool slowly to room temperature:

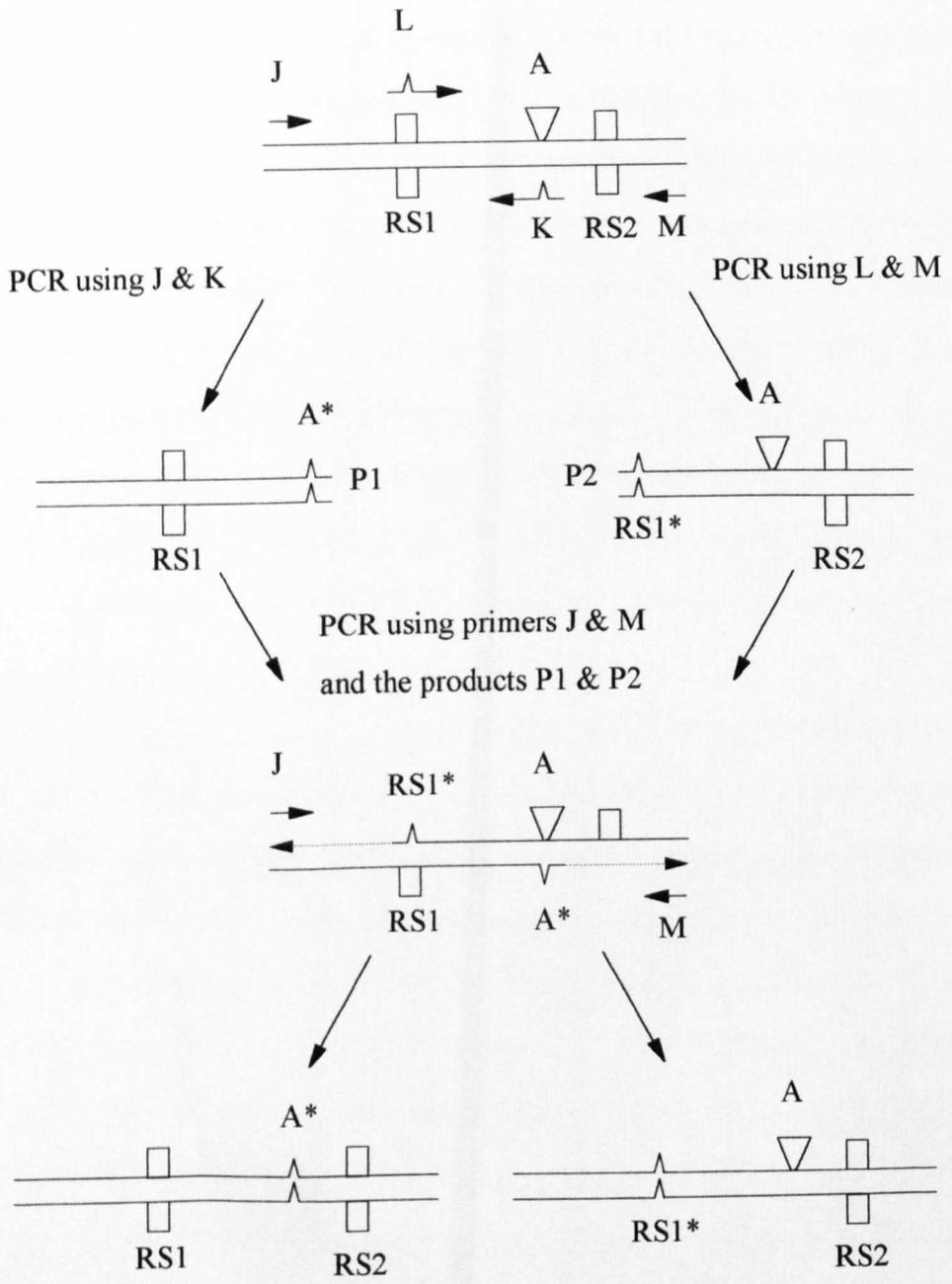


Figure 2.1. Diagram depicting site directed mutagenesis using PCR. Primers IDJ, IDK, IDL, and IDM are indicated by short arrows; A represents the site to be mutagenised; A* and RS1* represent sites after being mutagenised; P1 and P2 represent the two products produced after the first round of PCR; RS1 and RS2 represent the two available restriction sites utilised in the cloning of PCR fragments into a suitable vector and in doing so selecting for mutated clones. Only clones mutated at A will have intact RS1 and RS2 so by cloning with these two enzymes the population of correctly mutated cDNA clones will be enhanced.

P1	2.5 μ l
P2	2.5 μ l
Primer IDJ	200 ng
Primer IDK	200 ng
10 x Taq buffer (Promega)	2.5 μ l
25 mM MgCl ₂	2.5 μ l
Water	to 13 μ l

Then 8 μ l of 0.125 mM dNTPs, 0.5 μ l Taq polymerase (5 U/ μ l, Promega), water to 25 μ l and one drop of paraffin were added. The reaction mixture was cycled through 20 cycles of 94 °C for 1 min., 55 °C for 1 min., 72 °C for 3 min. and 1 cycle of 94 °C for 1 min., 55 °C for 1 min., 72 °C for 7 min. using a DNA Thermal Cycler (Perkin Elmer Cetus). The aqueous phase was withdrawn and 2.5 μ l run on an agarose gel (2.3.1.8) to check the quantity of the DNA present. The remainder of the reaction mixture was extracted with phenol/chloroform and digested with the restriction enzymes RS1 and RS2 (Fig. 2.1.). Only PCR fragments which had a mutagenised site (A*) had both restriction sites intact so allowing the enhancement of positive clones when the fragment is subcloned into a suitable vector. Prospective positive clones (ie. those which have PCR fragments cloned into them using RS1 and RS2) were screened by DNA sequencing (2.3.1.13).

2.3.1.13. DNA sequencing

Denaturation of dsDNA

400 ng of dsDNA was denatured in a volume of 20 μ l containing 200 μ M EDTA and 200 mM NaOH for 15 min. at 42 °C. The reaction was terminated by the addition of ammonium acetate pH 4.5 to a concentration of 182 mM and the denatured DNA precipitated by the addition of 2.5 volumes of 95% ethanol and placed at -20 °C for 30 min.. The DNA precipitate was pelleted by centrifugation at 16,000 x g for 15 min.

and subsequently washed with 70% ethanol and centrifuged again at 16,000 x g for 15 min.. The DNA pellet was then dried in a vacuum desiccater.

Annealing reaction

25 ng of primer DNA was annealed to the denatured DNA sample or single stranded M13 DNA (2.3.3.3.). The reaction conditions used were 40 mM Tris.HCl pH 7.5, 10 mM MgCl₂, 50 mM NaCl in a volume of 10 µl for two min. at 65 °C. The reaction was then allowed to cool to room temperature for at least 10 min..

Sequencing reaction

The annealing mixture was made up to 15.5 µl to give a reaction mixture which contained 26 mM Tris.HCl pH 7.5, 6.5 mM MgCl₂, 32 mM NaCl, 1.5 µM dG/dC/dTTP (the concentration of dNTPs was decreased if sequence data proximal to the primer was required, or increased if data more distal from the primer was required), 6.5 mM DTT, 3.25U Sequenase (DNA-dependent DNA polymerase) and 5 µCi α-³⁵S dATP. The reaction was incubated on ice for 5 min. and then divided in to four 3.5 µl aliquots. Each aliquot was added to one of four 2.5 µl ddNTP mixtures (80 µM dNTP, 8 µM ddA/ddG/ddC/ddTTP). These were incubated at 42 °C for 5 min. when 4 µl of stop reaction buffer was added (96% formamide, 20 mM EDTA pH 8.0, 0.05% (v/v) xylene cyanol FF, 0.05% (v/v) bromophenol blue). The resulting 4 reaction mixtures were subsequently heated to 95 °C for 2 min. and placed on ice.

Denaturing polyacrylamide gel electrophoresis using a 6% gel

Using Bio-Rad SequiGen gel apparatus, a 6% polyacrylamide denaturing gel was prepared in 1 x TBE containing 5.7% acrylamide, 0.3% *N,N*-methylene-bis-acrylamide, 46% (w/v) urea, 1 µg/ml ammonium persulphate and 0.25 µl/ml TEMED. 2 µl of each sequencing reaction was loaded onto a gel (which had been pre-run at 45W for 30 min.) and electrophoresised at 45 W (ca. 50 °C) for 2-3 hours depending on whether the required sequence data were proximal or distal to the primer. The gel

was then soaked in 10% methanol and 10% acetic acid for 30 min. and dried at 60 °C under vacuum for 1-2 hours onto a piece of paper. The sequence data was visualised by autoradiography (2.3.12.).

2.3.2. Bacteriological Techniques

2.3.2.1. The growth and maintenance of bacteria required for routine cloning.

Escherichia coli (*E. coli*) DH1 (*recA1 endA1 thi-1 hsdR17 supE44 gyrA96 (Nal^r) relA1*) (Hanahan, 1983) were used in all the routine cloning procedures, details of other strains are given as they arise. Bacteria were grown either in liquid culture form in LB medium at 37 °C in shaking incubators or on LB-agar plates at 37 °C. When transformed bacteria were grown, plasmids encoding ampicillin-resistance genes were maintained in the culture by supplementing the media with ampicillin at 50 µg/ml. Cultures were maintained for up to 1 month on agar plates at 4 °C. For long-term storage, 0.5 ml of a stationary phase culture was mixed with 0.5 ml 80% glycerol (v/v) and rapidly frozen in a dry ice/ethanol slurry and stored at -70 °C.

2.3.2.2. The transformation of bacteria with plasmid DNA.

a) Calcium chloride method

Preparation of competent cells:

A single colony of bacterial cells was grown overnight in 4 ml of LB media at 37 °C shaking at 250 rpm. The next day this culture was added to 400 ml of LB media and shaken at 37 °C until the A₆₀₀ had reached 0.4. The cells were then placed on ice for 10 min. and centrifuged at 2,000 x g for 7 min. at 4 °C. The resulting pellet was gently resuspended in 10 ml of 60 mM CaCl₂ at 4°C and centrifuged again (as above). The pellet was again resuspended in 10 ml 60 mM CaCl₂ at 4 °C and then left on ice for 30 min.. The cells were then re-centrifuged (as above) and resuspended in 2-3 ml of 60 mM CaCl₂ with 15% (v/v) glycerol and frozen at -70 °C in 200 µl aliquots. Variations in this method are given as they arise.

Transformation Procedure

DNA for transformation was made up to 25 μ ls with distilled water and incubated with 100 μ ls of competent cells on ice for 15 min. in a round bottom tube. The mixture was placed at 42 °C for 2 min. exactly and immediately afterwards placed on ice again for another 15 min.. One ml of SOC media was then added and the mixture shaken at 37 °C for 1 hour. 2% and 20% of the mixture was subsequently plated onto LB-agar plates containing a suitable antibiotic and incubated overnight at 37 °C. Individual colonies were picked the next day for small scale preparation of DNA (2.3.2.3.)

b) The electroporation procedure

Preparation of competent cells

A single colony of bacterial cells was grown in 5 ml of LB medium overnight at 37 °C at 250 rpm. The next day 1 ml of the overnight culture was used to inoculate 100 ml of LB. This was shaken at 37 °C until the A_{600} reached about 0.6 when it was chilled on ice for 15 min.. The culture was then centrifuged at 2,000 x g for 10 min. at 4 °C, the pellet resuspended in 100 ml of water and centrifuged a second time. The pellet was resuspended in 50 ml of water and centrifuged a third time. The pellet was then resuspended in 2 ml of 10% (v/v) glycerol and centrifuged for the last time. The pellet was resuspended in a volume of 2-300 μ l of 10% (v/v) glycerol and stored at -70 °C in 40 μ l aliquots.

Electroporation

The DNA for electroporation, purified and desalted as required was resuspended in 20 μ l of water. 10 μ l (typically 100 ng) of this was mixed with a 40 μ l aliquot of competent cells (thawed briefly on ice) and incubated on ice for 1 min. before placing it into the electroporation cuvette. The DNA was then electroporated into the bacteria using a Bio-Rad Gene Pulser with a voltage pulse at 2.5 kV, at a capacitance of 25 μ F (pulse controller at 200 Ω s). Immediately after electroporation the cells were mixed with 1 ml of SOC media and shaken for 1 hour at 37 °C. 2% and 20% of the mixture

was then plated onto LB-agar plates containing a suitable antibiotic and incubated overnight at 37 °C. Individual colonies were picked the next day for small scale preparation of DNA (2.3.2.3.).

2.3.2.3. Small scale plasmid DNA purification from bacteria (Mini Prep)

This method was published by Zhou *et al.* (1990). A single colony of transformed bacterial cells was grown overnight in 5 ml of LB medium supplemented with the appropriate antibiotic. The following day 1.5 ml of the culture was centrifuged at 16,000 x g for 20 seconds. The supernatant was decanted off leaving 50 - 100 µl of media covering the pellet. This was then vortexed for 5 seconds to resuspend the bacterial pellet and 300 µl of TENS (0.1 M NaOH, 0.5% (w/v) SDS in TE buffer pH 8.0) added. The solution was vortexed for 5 seconds, 150 µl of 3 M NaOAc pH 5.2 added and vortexed for a third time for 5 seconds. The solution was then centrifuged at 16,000 x g for 2 min. and the supernatant transferred to a fresh tube. 1 ml of ethanol at -20 °C was added to the supernatant and the mixture centrifuged at 16,000 x g for 2 min.. The resulting pellet was washed twice in 70% (v/v) ethanol, dried in a vacuum desiccator and resuspended in 20 µl of water. 5 µl was used per subsequent restriction digest analysis.

2.3.2.4. Large scale plasmid DNA preparation from bacteria (Maxi Prep)

500 ml of LB-Amp was inoculated with 100 µl of an overnight culture and shaken at 37 °C for 6 hours at which point the cultures were supplemented with 50 µg/ml chloramphenicol and left to incubate overnight at 37 °C. The following day the cells were pelleted at 2,500 x g for 15 min. at 4 °C and resuspended in 4 ml of GTE (50 mM glucose, 25 mM Tris.HCl pH 8.0, 10 mM EDTA). 1 ml of GTEL (50 mM glucose, 25 mM Tris.HCl pH 8.0, 10 mM EDTA, 25 mg/ml Lysozyme) was added and left at room temperature for 10 min.. 10 ml of 0.2 M NaOH and 1% (v/v) SDS solution was then added and left a further 10 min. on ice. Afterwards, 7.5 ml of 3M KOAc, 1.3M formic acid (pH adjusted to 5.5 with acetic acid) was added to the lysed

cells, allowed to stand on ice for 10 min. and centrifuged at 16,000 x g for 10 min.. 0.6 volumes of propan-2-ol was added to the supernatant and centrifuged at 2,000 x g for 15 min.. The resulting pellet was washed with 2 ml of 70% (v/v) ethanol and dried in a vacuum desiccator. The pellet was resuspended in 4.3 ml of TNE, 4.85 g of CsCl and 320 µl of 10 mg/ml ethidium bromide. This solution was centrifuged at 300,000 x g for 18 - 20 hours at 20 °C, and the purified plasmid band recovered from the gradient with the aid of a hypodermic needle and syringe. The ethidium bromide was removed from the sample by repeated extraction with CsCl-saturated propan-2-ol. The DNA was then precipitated by diluting the final aqueous phase with three volumes of TNE and 10 volumes of ethanol and placing it at -70 °C for 1 hour. The DNA precipitate was subsequently pelleted by centrifugation at 10,000 x g and the resulting pellet resuspended in 0.5 - 1.0 ml of TE buffer or water.

2.3.3. M13 Phage Techniques

The bacterial strain used as a host strain for M13 phage vectors (Messing *et al.*, 1977) was the F⁺ bacterial strain TG1 (Gibson, 1984; Sambrook *et al.*, 1989).

2.3.3.1. Transformation procedure

Preparation of competent E.coli

TG1s (*supE hsdΔ5 thiΔ(lac-proAB) F'[traD36 proAB⁺ lacI^q lacZΔM15]*) were made competent as per the method described in 2.3.2.2.

Transformation

DNA (<25 µl) was mixed with 150 µl of competent TG1 bacteria in a round bottom 15 ml tube and left on ice for 30 min.. The mixture of DNA and competent cells was then placed at 42 °C for exactly 2 min. and subsequently placed back on ice for a further 30 min.. Afterwards, 89%, 10%, and 1% volume aliquots of the suspension were mixed with 3 ml of top agar at 46 °C supplemented with 20 µl of 100 mM IPTG,

30 μ l 2% (v/v) Xgal and 200 μ l of TG1 cells grown to saturation. This was then poured on to a pre-warmed (about 40 °C) LB plate and, when set, incubated at 37 °C overnight. Plaques were picked the next day for either small scale replicative form (RF) DNA purification or single strand preparations for sequencing.

2.3.3.2. Culturing M13 clones

3 ml of a TG1 culture in stationary phase was used to seed 50 ml of 2YT medium. This culture was incubated at 37 °C, 250 rpm for one hour after which 3 ml aliquots were placed in universals. Positive plaques were identified in the TG1 lawn, following transformation, by the absence of the blue colouration indicative of β -galactosidase activity. These plaques were picked using a sterile tooth pick and inoculated into the TG1 aliquots. These TG1/M13 cultures were then incubated for either 5 hours (for ssDNA preparations) or 7 hours (for RF preparations) at 37 °C, 300 rpm.

2.3.3.3. Preparation of RF DNA from M13 cultures

RF M13 DNA was prepared as in 2.3.2.3. except 1.5 ml of a TG1/M13 culture was used as starting material.

The supernatant from the first centrifugation step (2.3.2.3.) was recentrifuged at 16,000 x g for 10 min. to clear any residual bacterial cells. The pellet from this second centrifugation was discarded and the supernatant stored at 4 °C as a recombinant M13 phage stock.

2.3.3.4. Preparation of single strand DNA from M13 plaques

Single stranded M13 DNA was prepared by centrifuging 1.2 ml of TG1/M13 culture at 16,000 x g for 5 min. and removing the supernatant to a clean tube. The supernatant was then centrifuged again as before and transferred to a clean tube. 300 μ l of PEG/NaCl solution (20% polyethylene glycol 600, 2.5 M NaCl) was added to the supernatant and inverted 3 times to mix. The solution was then left to stand at room temperature. After 15 min. the solution was centrifuged at 16,000 x g for 5 min. to

pellet the PEG-phage complexes which had formed. The supernatant was discarded and the pellet recentrifuged as before. Again, any supernatant was discarded and the pellet resuspended in 120 µl of TE. 50 µl of phenol/TNE was added and vortexed for 30 seconds. The resulting two phases were separated by centrifugation at 16,000 x g for 5 min. and 100 µl of the aqueous phase removed to a clean tube. To this was added 50 µl of chloroform/isoamyl alcohol and the two phases mixed by vortexing for 30 seconds. The solution was then centrifuged at 16,000 x g for 5 min. and 80 µl of the aqueous phase placed in a clean tube. The single stranded M13 DNA was ethanol precipitated, dried (2.3.1.1.) and resuspended in 30 µl TE. 5 µl of this solution was used for each subsequent sequencing reaction.

2.3.4. Tissue culture and virus techniques

2.3.4.1. Maintenance and passage of cell lines

HeLa, W162 (Weinberg and Ketner, 1983) and HEK-293 (Graham *et al.*, 1977) cell lines were passaged routinely in 90 mm tissue culture dishes containing DMEM/10% CS. All cells were grown at 37 °C in a 5% CO₂ humidified atmosphere. When passaged, cells were washed with TD and then soaked in TDE/trypsin (TD supplemented with EDTA at a final concentration of 0.5 mM and trypsin at a final concentration of 0.25 g/l) for 1 to 2 min.. Excess CS was then added to 20% volume and the cells centrifuged at 2,000 x g for 3 min.. The resulting pellet was resuspended in DMEM/10% CS and placed into tissue culture dishes at the required density. HeLa and W162 cells were passaged at a ratio of 1:10 while HEK-293 cells were split at a ratio of 1:4.

2.3.4.2. Maintenance and passage of non-transformed cells

The non-transformed cell lines used in this work were MRC5 (gift of C.R. Pringle), HISM, WI-38 and HUV-EC-C (American Type Tissue Culture collection). The MRC5, HISM, and WI-38 cell lines were passaged as in 2.3.4.1. except FCS was used

instead of CS and the maximum ratio of subcultivation was 1:4. The HUV-EC-C cell line was grown in F12 K medium supplemented with 10% FCS, 100 µg/ml of heparin, and endothelial cell growth supplement at 10 µg/ml. The cells were passaged as above with a maximum subcultivation ratio of 1:3.

2.3.4.3. Freezing cell stocks

Cells were prepared for long-term storage by suspending the cells from one 90 mm dish in 8% (v/v) dimethylsulphoxide/ 92% FCS and placed in freezing vials suitable for storage in liquid nitrogen. The vials were slowly cooled to -70 °C and then placed in liquid nitrogen.

2.3.4.4. Recovery of cell stocks

The cells were removed from the liquid nitrogen and rapidly thawed in a 37 °C waterbath. Once thawed the cells were placed in a 90 mm dish containing prewarmed growth medium, which was changed the following day.

2.3.4.5. Calcium phosphate transfection for the construction of adenovirus mutants

W162 cells from one 90 mm dish were passaged and divided between 3 x 60 mm dishes 24 hours before the transfection. Three hours before the transfection the medium was replaced with DMEM/10% FCS. Prior to the transfection the DNA/calcium phosphate precipitates were prepared; 10 µl of 1.25 M calcium chloride was added slowly to 50 µl of 2 x HEPES buffer (16 g/l NaCl, 0.74 g/l KCl, 2 g/l glucose, 10 g/l HEPES), 3 µl 100 x phosphate solution (9.9 g/l Na₂HPO₄), 10 µl DNA (1 to 10 µg) and 27 µl of water and mixed by passing air through the solution. The precipitate was allowed to form at room temperature for 30 min. after which it was added to 900 µl of DMEM/10% FCS. The medium was removed from the dishes of W162 cells and the cells overlaid with the 1 ml of precipitate. The cells were subsequently incubated at 37 °C/5% CO₂ for 3 hours with occasional rocking to ensure the cells did not dry out. The media was then removed and 1 ml of 20% (v/v)

glycerol in TS buffer was added to each dish for 1 min.. The dishes were washed twice with 2 ml of TS and overlaid with DMEM/10% CS for 30 min.. The cells were then overlaid with semi-solid media (4% (v/v) CS, 0.375% (w/v) NaHCO₃, 1% (w/v) noble agar, in 1 x DMEM). After 3-4 days another semi-solid overlay was added (1% (v/v) CS, 0.1875% (w/v) NaHCO₃, 1% (w/v) noble agar, in 1 x DMEM). When plaques were visible (normally 6 days post-infection onwards) a third semi-solid overlay was added (1% (v/v) CS, 0.1875% (w/v) NaHCO₃, 0.01% (w/v) neutral red, 1% (w/v) noble agar, in 1 x DMEM) and the plaques counted the next day. The neutral red stained any live cells present facilitating the identification of virus plaques. Any plaques visible were picked using a Pasteur pipette and stored at -70 °C in 0.5 ml TS/2% (v/v) CS.

2.3.4.6. Virus infection of cells monolayers

The medium overlaying the cell monolayer was aspirated off and TS/2% (v/v) CS containing virus at a concentration to achieve the required multiplicity of infection was added. For a 90 mm culture dish the volume of TS/2% (v/v) CS containing the virus added was 0.5 ml; this was adjusted for other culture dishes according to their size. The infected monolayer was incubated in a humidified incubator at 37 °C, 5% CO₂ and occasionally rocked to ensure the monolayer did not dry out. After approximately 1 hour prewarmed growth medium was added (10 ml for a 90 mm plate) and the infected monolayer returned to the incubator.

2.3.4.7. Rapid preparation of adenovirus DNA

The cells from an infected monolayer in a 90 mm dish were harvested when cytopathic effect could be seen. The cells were centrifuged at 2,000 x g for 10 min., the supernatant removed and the pellet resuspended in 0.4 ml TE (pH 9.0)/10 mM spermine tetrahydrochloride. To this suspension 0.4 ml DOC lysis buffer (20% (v/v) ethanol, 100 mM Tris.HCl pH 9.0, 0.4% (w/v) sodium deoxycholate) was added and the resulting cell debris pelleted by centrifugation at 16,000 x g for 5 min.. The

supernatant was collected and 60 μ l of 10% (w/v) SDS, 20 μ l 0.5 M EDTA, and 20 μ l 20 mg/ml proteinase K were added. This was mixed by inversion and incubated at 37 °C for 1 hour. The solution was then extracted with phenol/chloroform and the DNA precipitated by the addition of 30 μ l 5 M NaCl and 0.6 ml of propan-2-ol. This was mixed by inversion and left for 10 min. at room temperature, after which the DNA was pelleted by centrifugation at 16,000 x g for 15 min.. The pellet was washed in ice-cold 70% (w/v) ethanol and then dried in a dessicator. The viral DNA pellet was resuspended in 30 μ l of distilled water.

2.3.4.8. Large-scale preparation of adenovirus DNA

Infected cells were harvested on day 3 of the infection when the monolayer was starting to lift from the surface of the dish. The cells were scraped from 10 infected 90 mm dishes and centrifuged at 2,000 x g for 10 min.. The cell pellet was resuspended in 10 ml of 0.1 M Tris.HCl pH 8.0 and sonicated on ice for 2 sets of 10 x 1s pulses separated by 30s using a Jencons GE375 ultrasonic processor and a 3 mm microtip. The sonicated cells were then centrifuged at 3,750 x g for 10 min. at 4 °C and the supernatant placed over a CsCl step gradient (5 x 90mm dishes per gradient) consisting of 3 ml of CsCl in TD (density 1.25 g/ml) over 2ml of CsCl in TD (density 1.4 g/ml). The gradients were centrifuged at 150,000 x g for 1 hour at 15 °C in a Beckman SW41 rotor. The resulting virus band was collected by puncturing the base of the tube and collecting the required drip fraction. The collected virus was centrifuged again but in CsCl/TD (density 1.35 g/ml) at 150,000 x g for 16 hours at 15 °C in a Beckman SW50.1 rotor to purify the virus further. The virus was collected as before, diluted with two volumes of water and then the virions precipitated with two volumes of ethanol at -70 °C for 15 min.. The virions were then pelleted by centrifugation at 5,000 x g for 20 min. at 4 °C. The virion pellet was dried and resuspended in 2 ml of TNE, 120 μ l 10% (w/v) SDS, 40 μ l 0.25 M EDTA, and 20 μ l 10 mg/ml proteinase K. This suspension was incubated at 37 °C for 1-2 hours then extracted with 2 ml phenol/TNE. The tube was gently inverted repeatedly over one minute when 1 ml of

chloroform/ethyl alcohol was added, and the solution mixed gently again for a further minute. The two phases were separated by centrifugation at 2,000 x g for 10 min. at 4 °C. The aqueous phase was then phenol/chloroform extracted as before. 0.1 vol of 3 M sodium acetate and 2 volumes of ethanol were then added to the resulting aqueous phase and the solution placed at -70 °C for 30 min. to precipitate the viral DNA. The DNA was pelleted by centrifugation at 5,000 x g for 30 min. at 4 °C after which the pellet was washed with 70% (w/v) ethanol. The pellet was then resuspended overnight at 4 °C in 0.5 ml of TE.

2.3.4.9. Preparation of virus stocks

Infected cells were harvested when viral cytopathic effect could be clearly seen in the monolayer. The cells were centrifuged at 2,000 x g for 10 min. and half of the supernatant removed (so as to concentrate the virus stock). The cells were then resuspended and subjected to four freeze-thaw cycles to lyse the cells and release attached virions. The samples were then centrifuged again at 2,000 x g for 10 min. to remove the cell debris and the supernatant stored in 2 to 3 ml aliquots at -70 °C.

2.3.4.10. Preparation of purified adenovirus virions

The protocol used for the purification of adenovirus virions was identical to that of the large scale purification of viral DNA (2.3.4.8) upto the point when the virions are purified from the second CsCl gradient. The purified virions were collected from the gradient by the puncture of the base of the tube and then placed on ice. The concentration of the virions was established by measuring the A_{260} value of virions diluted in TNE/0.1% SDS (1 A_{260} unit is equivalent to 10^{12} virions/ml (1 pfu is equal to 20-50 virions (Green et al, 1967))). The virion stock was stabilized by diluting 5-fold with stabilizing buffer (0.1% BSA, 50% glycerol, 10 mM Tris (pH 8.0), 100 mM NaCl, and 2 mM $MgCl_2$) and stored at -70 °C.

2.3.4.11. Titration of adenovirus by plaque assay

A serial dilution of the virus stock was made from 10^{-1} to 10^{-8} in TS/2% CS. 200 μ l of the appropriate dilution was used to infect duplicate HEK-293 cell monolayers in 35 mm wells in 6 well trays. The cells were then overlaid with semi-solid medium as described in 2.3.4.5. The number of plaques were recorded after the staining of the monolayer with neutral red (2.3.4.5.), allowing the calculation of the concentration of the virus stock in pfu/ml.

2.3.5. Analysis of viral RNA expression

2.3.5.1. Preparation of cytoplasmic and nuclear RNA

The cells from a 150 mm dish were washed twice with 10 ml of complete PBS (PBS supplemented with 0.1 g/l CaCl_2 and 0.1 g/l MgCl_2) and centrifuged at $2,000 \times g$ for 3 min.. The cell pellet was resuspended in 2 ml isotonic buffer (150 mM NaCl, 10 mM Tris.HCl pH 7.6, 1.5 mM MgCl_2) and vortexed with 200 μ l of 10% (v/v) Nonidet P40 (NP40). The suspension was then left to stand for 10 min. and the nuclei pelleted by centrifugation at $2,000 \times g$ for 3 min.. The supernatant was removed and the pellet washed with 1 ml of isotonic buffer and centrifuged as before. The resulting supernatants were combined and 60 μ l of 10% (w/v) SDS and 120 μ l 0.25 M EDTA added. The pellet was resuspended in RSB (10mM Tris.HCl pH 7.6, 3 mM MgCl_2 , 10 mM NaCl) and 2 μ l of 10 mg/ml DNase I (RNase-free) was added. The resuspended nuclei were left on ice for 30 min. when 20 μ l of 10% (w/v) SDS and 40 μ l of 0.25 mM EDTA were added. Both the nuclear RNA fraction (the lysed pellet) and the cytoplasmic RNA fraction (the supernatant) were then extracted with one volume of phenol/TNE and one volume of chloroform/amyl alcohol and centrifuged at $2,000 \times g$ for 10 min.. This was then repeated with one volume of phenol/TNE and two volumes of chloroform/amyl alcohol. The aqueous phase was extracted repeatedly with chloroform/amyl alcohol until it was clear. The RNA in the resulting aqueous phase was precipitated with ethanol ($-70 \text{ }^\circ\text{C}$) for one hour and pelleted by centrifugation at

8,000 x g for 30 min.. The RNA pellet was dried and resuspended in 300 µl of water. (All manipulations were carried out at 4 °C using RNase-free solutions).

2.3.5.2. Preparation of antisense RNA probes

1 µg of linearized template was prepared by the appropriate digestion at the 3' end of the sequence which was to be transcribed in the relevant pGEM subclone. This template DNA was purified by phenol/chloroform extraction and ethanol precipitation. The purified, linear template was then resuspended in water at a concentration of 100 ng/µl. The *in vitro* transcription reactions were then prepared in the order as follows:-

	Volume (µl)
10 µCi/µl α - ³² P UTP (800 Ci/mmol)	2.5
5 x Sp6 or T7 buffer	1.25
10 x nucleotide mix (0.5 mM ATP, CTP, GTP and 0.1 mM UTP)	0.625
200 mM DTT	0.3125
Water	0.3125
Human placental RNase inhibitor (15 U/µl)	0.25
Template DNA (100ng/µl)	0.7
Sp6 or T7 RNA polymerase	0.5

The mixture was incubated for 1 hour at 37 °C (for T7 polymerase) or 40 °C (for Sp6 polymerase). The reaction was then diluted to 50 µl with water, 2 µl of 200 mM vanadyl ribonucleotide complex and 1 µl of 10 mg/ml RNase-free DNase added. The reaction was incubated at 37 °C for 10 min. when 0.5 µl 10% (w/v) SDS was added and the solution extracted with phenol/chloroform. The radioactive RNA probe was purified using a Stratagene Nucltrap push column and stored at -20 °C.

2.3.5.3. Ribonuclease protection analysis

All the ribonuclease protection assays were carried out according to the method of Melton *et al.* (1984). 5 µg of cellular RNA and 5 µl of antisense probe were co-precipitated by increasing the volume to 40 µl with water and adding 1 µl of 5M NaCl, 100 µl of ethanol and placing at -20 °C for 30 min.. The resulting RNA pellet was dried and resuspended in 20 µl of hybridisation buffer (80% (v/v) deionised formamide, 40 mM PIPES pH 6.7, 0.4 M NaCl, 1 mM EDTA). This was heated to 80 °C for 5 min. and then allowed to cool slowly overnight to 56 °C. The hybridised RNA solution was then diluted to 200 µl in RNase solution (10 mM Tris.HCl pH 7.5, 5 mM EDTA, 300 mM NaCl, 100 units/ml RNase T1, and 5 µg/ml pancreatic RNase A) and incubated at room temperature for 30 min.. To this 10 µl 10% (v/v) SDS and 1 µl of 20 mg/ml proteinase K was added and the solutions vortexed vigorously and incubated at room temperature for 15 min. to ensure the complete degradation of the RNases present. The solution was then extracted with phenol/chloroform, and the protected RNA precipitated with 3 volumes of ethanol for 30 min. at -20 °C. The RNA was pelleted by centrifugation at 16,000 x g for 15 min., dried and resuspended in 6 µl of formamide dye mix (80% (v/v) deionised formamide, 10 mM EDTA (pH 8.0), 0.05% (v/v) xylene cyanol and 0.05% (v/v) bromophenol blue). The sample was heated to 90 °C for 4 min. and then placed immediately on ice prior loading onto a 5% polyacrylamide gel (2.3.5.4).

2.3.5.4. RNA Polyacrylamide Gel Electrophoresis (PAGE)

Using Bio-Rad SequiGen gel apparatus, a 5% polyacrylamide denaturing gel was prepared in 1 x TBE containing 4.75% acrylamide, 0.25% N,N' -methylene-bis-acrylamide, 46% (w/v) urea, 0.8 µg/ml ammonium persulphate and 0.83 µl/ml TEMED. 2 µl of each RNA sample was loaded onto a prewarmed gel and electrophoresised at 45 W (ca. 50 °C) for 2-3 hours depending on whether the protected fragments were large or small. The gel was then immediately dried at 60 °C

under vacuum for 1-2 hours onto a piece of paper. The protected fragments were visualised by autoradiography (2.3.12.).

2.3.6. Analysis of proteins

2.3.6.1. SDS polyacrylamide gel electrophoresis (SDS-PAGE)

SDS polyacrylamide gel electrophoresis was carried out according to Laemmli (1970). Protein samples in 100 mM DTT, 25 mM Tris.HCl pH 6.8, 10% (v/v) glycerol, 2% (w/v) SDS and 56 µg/ml bromophenol blue were heated to 95 °C for 3 min. prior to being loaded onto the discontinuous SDS-polyacrylamide gel. The resolving gel contained 0.43 M Tris.HCl pH 8.8, 10 to 15% (w/v) acrylamide, 0.27 to 0.4% (w/v) N,N'-methylene-bis-acrylamide, 0.095% (w/v) SDS, 0.067% to 0.033% (w/v) ammonium persulphate and 0.2 µl/ml TEMED. The amount of acrylamide in the resolving gel was varied according to the size of the protein to be resolved, low percentage acrylamide gels were run for the resolution of large proteins and *vice versa*. The stacking gel contained 0.12 M Tris.HCl pH 6.8, 4.57% (w/v) acrylamide, 0.12% (w/v) N,N'-methylene-bis-acrylamide, 0.1% (w/v) SDS, 0.095% (w/v) ammonium persulphate and 0.48 µl/ml TEMED. The gel was run in the BRL vertical gel electrophoresis system at 240 V for about 3 hours in a buffer containing 30.3 g/l Tris, 144.2 g/l glycine, and 10.0 g/l SDS.

2.3.6.2. Coomassie-blue staining of SDS-polyacrylamide gels

To observe proteins present in a gel, the gel was stained and fixed for at least 1 hour in 7% (v/v) acetic acid, 25% (v/v) methanol and 0.1 mg/ml Coomassie blue. It was then destained overnight using 7% (v/v) acetic acid and 25% (v/v) methanol and dried for 1-2 hours at 70 °C under vacuum.

2.3.6.3. Monochromatic silver staining

If a more sensitive staining procedure was required than coomassie blue staining the PAGE-SDS gels with the diamine silver stain of Wray *et al.*, (1981). The gel was soaked in 50% methanol for at least one hour and then soaked in freshly prepared silver stain for 15 minutes with gentle agitation. The silver stain was prepared by adding, dropwise, 4 mls of a 0.2g/ml silver nitrate solution to 22.4 ml of 0.3475% NaOH and 0.925M ammonium hydroxide, and then made up to 100 mls with deionised water. The gel was then washed 3 times for 5 minutes in deionised water and developed in 500 mls of fresh developing solution (0.005% citric acid, 0.0185% formaldehyde). When the stain was of the required intensity then the gel was washed in deionised water and the development stopped by placing it in 50% methanol.

2.3.6.4. Preparation of SDS-polyacrylamide gels for autoradiography

To visualise any radioactive proteins within a gel, the gel was fixed for 1 hour in 7% (v/v) acetic acid, 25% (v/v) methanol, treated for 20 min. with Amplify (Amersham) for fluorographic enhancement of radioactive emissions and then dried for 1-2 hours at 70 °C under vacuum. Any radioactive proteins present were observed by autoradiography (2.3.12.)

2.3.6.5. Bradford Assay (Bradford, 1978)

The amount of protein in a sample was quantified by performing a Bradford Assay on a portion of the sample. 5 to 10 µl of the protein sample was mixed with 0.15 M NaCl to bring the volume up to 100 µl. This was added to 1 ml of Coomassie G-250 Brilliant Blue Solution (100 µg/ml Coomassie Brilliant Blue (G-250), 0.05% (v/v) ethanol, 8.5% (v/v) phosphoric acid) and vortexed. The colour was allowed to develop for 2 min. and the A_{595} was measured. The amount of protein present in the sample was deduced by comparing the result of the unknown to the equivalent A_{595} reading on a BSA standard curve ranging from 0 to 10 µg (protein samples were diluted to give an A_{595} reading within this range).

2.3.7. Bacterial protein expression

2.3.7.1. Protein expression and purification using the Glutathione-S-Transferase system

Small scale protein purification

E. coli, containing the pGEX-2T plasmid with the gene to be expressed cloned in frame in the short polylinker after the GST gene, were used to seed 2 ml of LB-Amp. The culture was allowed to grow, at 37 °C/250 rpm, to a density of 0.5 A₆₀₀ when IPTG was added to a concentration of 0.1 mM to induce the transcription of the GST/gene fusion. The culture was then grown for a further 2-6 hours and then centrifuged at 16,000 x g for 10 seconds to pellet the cells. The cells were resuspended in 1 ml of ice-cold PBS and sonicated on ice for 30 seconds or until the solution became translucent, using a Jencons GE375 ultrasonic processor and a 3mm microtip. The solution was subsequently centrifuged at 16,000 x g for 5 min., to remove the insoluble fraction (P), and 50 µl of a 50% (w/v) slurry of glutathione sepharose 4B beads (Pharmacia) added to the supernatant (S/N). The mixture was rocked gently for 10 min. and the beads pelleted by centrifugation at 16,000 x g for 10 seconds. The beads were then washed 3 times with PBS and resuspended in an equal volume of 2 x SDS/sample buffer (200 mM DTT, 50 mM Tris.HCl pH 6.8, 20% (v/v) glycerol, 4% (w/v) SDS and 112 µg/ml bromophenol blue) if the samples were to be analysed by SDS polyacrylamide gel electrophoresis.

Large scale protein purification.

A 500 ml culture (as above) was allowed to grow, at 37 °C/250 rpm, to a density of 0.5 A₆₀₀ when IPTG was added to 0.1 mM. The culture was then placed at 30 °C/250rpm, left overnight and then centrifuged at 2,000 x g for 15 min. to pellet the cells. The cell pellet was resuspended in 10 ml of ice-cold PBS and broken using a French Cell Pressure Press (American Instruments Company) at 1010 psi. The lysed cells were centrifuged at 32,500 x g for 60 min. to remove the bacterial debris and the

supernatant removed. The supernatant was diluted to 500 ml with PBS and 2.5 ml of a 50% (w/v) slurry of glutathione sepharose 4B beads added. This mixture was rocked gently at room temperature for 10 min. and the beads subsequently pelleted by centrifugation at 2,000 x g for 3 min.. The beads were washed three times with 100 ml of PBS/1% (v/v) Triton X100, and then immobilised on a small column. The GST/gene fusion protein was eluted off the column with 50 mM Tris.HCl pH 8.0/20 mM reduced glutathione and collected in 0.5 ml fractions. The protein content of these fractions was assayed by SDS-PAGE (2.3.6.1.) and Bradford analysis (2.3.6.5.). The fractions containing the most pure protein were pooled and stored at -70 °C.

Thrombin cleavage of the GST-fusion protein

If the GST fusion protein was to be cleaved to release the pure recombinant protein then the PBS washed beads, with the recombinant protein attached were washed with 150 mM NaCl, 50 mM Tris pH7.5, and then with 150 mM NaCl, 50 mM Tris pH7.5, 2.5 mM CaCl₂. The GST fusion protein was then cleaved by digestion with the required amount of thrombin at 30 °C in 150 mM NaCl, 50 mM Tris pH7.5, 2.5 mM CaCl₂ (40 µl of buffer/10 µl of 50% (w/v) glutathione sepharose beads).

2.3.7.2. Inclusion body purification of pGEX fusion proteins

The recombinant protein was expressed in a 2 ml culture under standard conditions (2.3.7.1). The cells were then pelleted by centrifugation at 16,000 x g for 10 seconds and resuspended in 200 µl of TEN (50 mM Tris.HCl pH 7.5, 0.5 mM EDTA, 0.3M NaCl, 1 mg/ml lysozyme). This was incubated on ice for 15 min. when NP40 was added to 0.2% and left for a further 10 min.. 300 µl of 1.5M NaCl, 12 mM MgCl₂, 3.3 µg/ml DNase was added and the mixture left at 4 °C overnight. The mixture was then centrifuged at 1,000 x g for 5 min. to remove any intact cells, the supernatant removed and centrifuged again at 5,000 x g for 20 min. to pellet the inclusion bodies. These were then washed three times with 100 µl 6M urea and then three times with 100 µl TEN (without the lysozyme). The method was scaled up when required.

2.3.7.3. Protein expression and purification using the histidine tag system

The pTrcHis expression system (Invitrogen)

Expression and purification of recombinant proteins using this system was accomplished as per the instructions provided by the manufacturer.

The pQE expression system (Qiagen)

E. coli containing the pQE-30 (Qiagen) plasmid, with the gene to be expressed cloned in frame in the short polylinker after the 6 histidine residues, were used to seed 50 ml of LB containing suitable antibiotics. The culture was allowed to grow, at 37 °C/250 rpm, to a density of 0.8 A₆₀₀ when IPTG was added to 2mM. The culture was then incubated for a further 5 hours at 37 °C and centrifuged at 2,000 x g for 15 min. to pellet the cells. The cells were then lysed in 10 ml of 6M guanidine.HCl, 0.1 M NaH₂PO₄, 0.01 M Tris.HCl, pH 8.0 and left for 1 hr at room temperature. The resulting lysate was centrifuged at 10,000 x g for 15 min. at 4 °C and mixed with 4 ml of nickel-resin (ProBond, Invitrogen), previously equilibrated with 6M guanidine.HCl, 0.1 M NaH₂PO₄, 0.01 M Tris.HCl, pH 8.0, for 1 hour at room temperature. The resin was then immobilised on a small column and washed with 10 volumes of 6M guanidine.HCl, 0.1 M NaH₂PO₄, 0.01 M Tris.HCl, pH 8.0 and then 5 volumes of 8M urea, 0.1 M NaH₂PO₄, 0.01 M Tris.HCl, pH 8.0 until the eluate had an A₂₈₀ of less than 0.01. The column was then washed with a high stringency wash (8M urea, 0.1 M Na₂PO₄, 0.01 M Tris.HCl, pH 6.8) until the eluate had an A₂₈₀ of less than 0.01. The recombinant protein was eluted from the column with 8M urea, 0.1 M Na₂PO₄, 0.01 M Tris.HCl, pH 4.5) in 0.5 ml fractions. The protein content of the fractions was determined by SDS-PAGE (2.3.6.1.) and Bradford analysis (2.3.6.5.).

2.3.8. Production of mono-specific polyclonal antibodies

500 µl of purified recombinant fusion protein (from either the GST or Histidine tag system, 200 to 500 µg/ml) was mixed with an equal volume of Freund's adjuvant (complete adjuvant was used in the first injection, subsequent boosts used incomplete

adjuvant) by vortexing for 30 min. or until the mixture was glutinous. The Freund's adjuvant prevents the rapid catabolism of the antigen when in the animal by forming a deposit and also stimulates the animals immune response non-specifically by increasing the lymphokine level of the animal so stimulating the activity of the antigen processing cells. The mixture was then inoculated into a virgin rabbit by subcutaneous injection into the scruf of the neck. The rabbit was boosted with antigen every 2-3 weeks. Blood samples were taken by bleeding the rabbit from the marginal ear vein at 6 weeks after the first inoculation and then every 2-3 weeks. The blood was allowed to clot overnight at 4 °C. The clot was then detached from the container and allowed to shrink for 24 hours at 4 °C. The clot was then removed by two 15 min. centrifugations at 2,000 x g and the supernatant (the serum) stored as 500 µl aliquots at -20 °C.

The immunisation and bleeding of rabbits was kindly performed by Mrs. Carol Hill.

2.3.9. Immunochemical techniques for the detection of proteins

2.3.9.1. Western blot analysis (Towbin *et al.*, 1979)

Protein transfer

The SDS-polyacrylamide gel was soaked in transfer buffer (3.03 g/l of Tris base, 14.4 g/l of glycine, 20% (v/v) methanol) for 30 min. along with 6 pieces of Whatman 3MM filter paper, 2 nylon pads and one piece of nitrocellulose (Hybond C, Amersham). The pads, gel, filter paper and nitrocellulose were then sandwiched together, placed in a BioRad Trans-Blot Cell and a potential difference of 70V placed across the sandwich so that the negatively charged proteins were transferred onto the nitrocellulose. The proteins were allowed to transfer for 3-4 hours at 4 °C. To see if the proteins had transferred, the nitrocellulose was soaked in Ponceau S solution (0.5% (w/v) Ponceau S, 1% (v/v) acetic acid) for 5 min., and subsequently washed with water to allow the visualisation of the transferred proteins.

Detection of antigen present on the nitrocellulose

The nitrocellulose was soaked for at least 1 hour in PBS/5% (w/v) low fat dried milk (Marvel), and then washed for 5 x 5 min. with PBS/0.1% (v/v) Tween 20. The primary antibody (diluted to the required concentration) was then added to the nitrocellulose in PBS/5% (w/v) marvel for 2 hours at room temperature. The nitrocellulose was then washed for 5 x 5 min. with PBS/0.1% (v/v) Tween 20 and the secondary antibody conjugate added (anti-rabbit immunoglobulin, biotinylated species-specific whole antibody, Amersham) at a dilution of 1 in 400 in PBS/1% (w/v) Marvel. This was incubated with the nitrocellulose for 1 hour at room temperature when the nitrocellulose was washed for 5 x 5 min. with PBS/0.1% (v/v) Tween 20. Then a streptavidin-biotinylated horseradish peroxidase complex (Amersham) was added to the nitrocellulose at a dilution of 1 in 300 in PBS/1% (w/v) Marvel and left at room temperature for 25 min.. The nitrocellulose was subsequently washed for 5 x 5 min. with PBS/0.1% (v/v) Tween 20 and then twice with PBS. Any antigen present was visualised by the addition of the development solution (16.7% (v/v) methanol, 0.5 mg/ml 4-chloro-1-naphthol, 0.015% hydrogen peroxide in PBS). The developed nitrocellulose membrane was then air-dried on Whatman 3MM paper at 37 °C in the dark.

2.3.9.2. Immunoprecipitation of proteins using polyclonal antibodies

Most immunoprecipitations were carried out with 1 ml RIPA buffer (150 mM NaCl, 50 mM Tris.HCl pH 8.0, 1% NP40, 0.5% sodium deoxycholate, 0.1% SDS, 100 µM PMSF) lysates of 10^7 HeLa cells. Sometimes the buffer conditions and the amount of antigen present were different, in which case the amount of antibody used was varied accordingly; the wash buffer used was always of the same composition as the lysis buffer. If the lysates were labelled with ^{35}S -met then an equal number of incorporated counts (see section 2.3.10.2.) was assayed ensuring that the same amount of labelled protein entered the assay.

5 µl of preimmune sera was added to the lysate and incubated at room temperature for 15 min. when 50 µl of Immuno-precipitin (formalin-fixed *Staphylococcus aureus* cells (10% w/v)) were added for a further 15 min. to pre-adsorb any background protein. The mixture was then centrifuged twice at 16,000 x g for 30 seconds to remove the *S. aureus* cells. To the supernatant 2-3 µl of antibody was added and the mixture allowed to shake gently at 4 °C overnight. To this 25 µl Immuno-precipitin was added and shaken gently for a further 30 min.. The Immuno-precipitin/antibody complex was then pelleted by centrifugation at 16,000 x g for 30 seconds and the pellet washed thoroughly 3 times to remove any nonspecifically bound proteins. The pellet was then resuspended in 30 µl of SDS gel loading buffer (100 mM DTT, 25 mM Tris.HCl pH 6.8, 10% (v/v) glycerol, 2% (w/v) SDS and 56 µg/ml bromophenol blue) and the sample run on a SDS-polyacrylamide gel (2.3.6.1)

2.3.9.3. Enzyme linked immunoadsorbent assay (ELISA)

0.5 µg of antigen in 100 µl of coating buffer (0.1 M Na₂CO₃, 0.1 M NaHCO₃) was placed into the required wells of an Immulon 2 ELISA plate and left overnight at 4 °C. The liquid was then removed and the wells washed three times with PBS/0.05% (v/v) Tween 20. 200 µl of 5% (w/v) Marvel was added to each well, left for 1 hour at 37 °C and again washed three times with PBS/0.05% (v/v) Tween 20. The test antisera were then added in 100 µl of 0.1% (w/v) BSA/0.05% (v/v) Tween 20/PBS as a dilution series. This was left for 1 hour at 37 °C when the wells were washed three times with PBS/0.05% (v/v) Tween 20. Bound antibody was then detected by the addition of 100 µl of a 1 in 400 dilution of a biotinylated species-specific whole antibody preparation of anti-rabbit immunoglobulin to each well in 0.1% (w/v) BSA/0.05% (v/v) Tween 20/PBS. After one hour at 37 °C the wells were washed three times with PBS/0.05% (v/v) Tween 20 and 100 µl of a 1 in 3,000 dilution of a streptavidin-biotinylated horseradish peroxidase complex added in 0.1% (w/v) BSA/0.05% (v/v) Tween 20/PBS. This was left for 15 min. at room temperature and washed three times with PBS/0.05% (v/v) Tween 20. 100 µl of substrate (35 mM

citric acid, 67 mM Na₂HPO₄, 670 mg/ml 1,2-phenylenediamine (OPD, Dako) and 1.25 x 10⁻⁴% (v/v) H₂O₂) was then added and left for 5 to 10 min.. The resulting colour change was stabilized by the addition of 50 µl of 2.5 M sulphuric acid. The A₄₉₂ was then measured for each well using an ELISA plate reader (Titertek Multiskan Plus MkII). The titre of a sample was taken as the dilution required to give an A₄₉₂ 50% of that of the plateau A₄₉₂.

2.3.10. Analysis of viral protein expression

2.3.10.1. ³⁵S-methionine labelling of proteins in tissue culture

Cell monolayers on 90 mm dishes (10⁷ HeLa cells, labelling experiments carried out using different size dishes used amounts of reagents that were scaled accordingly) were incubated with 2 ml of methionine-free GMEM for 30 min.. This was replaced with 1 ml of methionine free GMEM supplemented with 100 µCi of ³⁵S methionine (specific activity >800 µCi/mmol, Amersham) and the dish incubated for a further 1 to 3 hours and then placed on ice. The cells were then washed with ice cold TD and lysed in the required lysis buffer. In most instances the cells were lysed in RIPA buffer (150 mM NaCl, 10 mM Tris.HCl pH 7.6, 1% (v/v) Triton X-100, 1% (w/v) sodium deoxycholate, 0.1% (w/v) SDS, 100 µM PMSF) but sometimes the cells were fractionated by lysis in alternative buffers (2.3.10.3, 2.3.10.4). RIPA buffer lysates to be immunoprecipitated were then sonicated on ice, at low power, using a Jencons GE375 ultrasonic processor and a 3 mm microtip to disipate the released DNA. The samples were then centrifuged at 16,000 x g for 10 min. to remove any debris, quantified (2.3.10.2.) and the supernatant immunoprecipitated (2.3.9.2).

2.3.10.2. Quantification of ³⁵S-methionine incorporation into protein

10 µl of labelled cell extract was diluted to 0.5 ml with ice cold water and mixed with 100 µl of 100% (v/v) TCA in water. This was allowed to stand for 15 min. and the resulting precipitate collected on a 2.5 cm Whatman GF/C glass microfibre filter. The

filter was then washed twice with 1 ml of 5% (v/v) TCA (4 °C) and twice with 1 ml of ethanol. The filter was then air-dried and placed in 5 ml of Optiphase Safe scintillation fluid. The radioactivity present on the filter was measured using a LKB Wallac 1219 Rackbeta scintillation counter.

2.3.10.3. Subcellular fractionation using hypotonic buffer

The washed tissue culture cells were scraped off the 90 mm dish into 5 ml of TD. They were then centrifuged at 2,000 x g for 5 min., resuspended in 1 ml of hypotonic buffer (10 mM Tris.HCl pH 7.5, 10 mM NaCl, 1.5 mM MgCl₂, 100 μM PMSF), and allowed to stand on ice for 20 min.. The swollen cells were lysed with 25 strokes in a Dounce homogeniser and the lysate was centrifuged at 500 x g for 5 min. to pellet the nuclei (P_{0.5}, >236600 Svedberg Units). The supernatant (S_{0.5}) was then either fractionated further or made up to RIPA conditions for analysis. To fractionate the supernatant further it was centrifuged at 50,000 x g for 20 min. to produce a pellet (P₅₀, >1,113 Svedberg Units (cell membrane fraction)) and a supernatant (S₅₀). The supernatant was then centrifuged at 200,000 x g for 60 min. to produce a pellet (P₂₀₀, >59 Svedberg Units) which contained polysomes, ribosomes and smaller microsomes found within the cytoplasm. The resulting supernatant (S₂₀₀) contained proteins which were in complexes of less than 59 Svedberg Units. This final supernatant was made up to RIPA conditions and the pellets were resuspended in 1ml of RIPA buffer. All the fractions were then sonicated on ice for 10 seconds, at low power, using a Jencons GE375 ultrasonic processor and a 3 mm microtip. The fractions were then centrifuged at 16,000 x g for 10 min. to remove any debris and analysed by immunoprecipitation (2.3.9.2)

2.3.10.4. Sucrose gradient analysis

The monolayer from a 90 mm dish of cells (10⁷ cells) was lysed by hypotonic lysis

2.3.10.3. The supernatant fraction (S_{0.5}) was then placed on a 5 ml sucrose step gradient (5 to 20% (wt/wt) gradient in 6 steps, with a 60% fraction at the base, all in

hypotonic buffer (10 mM Tris.HCl pH 7.5, 10 mM NaCl, 1.5 mM MgCl₂, 100 μM PMSF)) which had previously been allowed to stand for 2 hours at 4 °C. The gradient was then centrifuged at 240,000 x g for 6 hours at 6 °C. 0.5 ml fractions were then collected from the gradient by bottom puncture. The protein content of each fraction was examined by immunoprecipitation.

2.3.11. *In vitro* protein expression

2.3.11.1. *In vitro* protein expression

The expression of proteins *in vitro* was accomplished using the Promega TnT Coupled Transcription-Translation Reticulocyte Lysate System and used according to the manufacturers instructions.

2.3.11.2. Quantitation of protein produced using rabbit reticulocyte lysate

2 μl of a 50 μl translation reaction (2.3.11.1.) was mixed with 98 μl of 1 M NaOH/2% (v/v) H₂O₂ and incubated for 10 min. at 37 °C. Subsequently 900 μl of ice cold 25% (v/v) TCA/2% (w/v) casamino acids was added and allowed to stand on ice for 30 min.. The resulting precipitate was collected on a 2.5 cm Whatman GF/C glass microfibre filter and processed as 2.3.10.2.. Total counts were measured by adding 5 μl of the TCA reaction directly onto a filter and measuring the radioactivity present. This allowed the determination of the percent incorporation into high molecular weight material in the reaction.

2.3.12 Autoradiography

Radiolabelled protein, RNA or DNA on dried, radioactive, polyacrylamide gels was visualised by placing the gel in close contact with preflashed Fuji Rx X-ray film in a light-fast autoradiography cassette at -70 °C. The appropriate preflash distance was determined by measuring the absorbance of test film at 545 nm and comparing it to

that of unflashed film. The flash distance was set so that the ratio of absorbance of preflashed to unflashed film was in the order of 0.1 to 0.2. This allowed the film to respond in a linear fashion to the radioactivity in the gel. The film was then developed after the required period of time (normally the next day).

2.3.13. Densitometry

The bands on autoradiographs which corresponded to radioactivity in the gels were quantified on a Molecular Dynamics laser densitometer, using Image Quant 3 analysis software. Only autoradiographs which were 'pre-flashed' (2.3.14.) and not over exposed were quantified by densitometry..

Chapter 3.

The Sequence of ORF1 and 2 of Ad5 early region 4

3.1 Introduction

ORF1 and ORF2 of Ad5 E4 region were originally sequenced by Steenbergh and Sussenbach (1979) using the Maxim-Gilbert chemical degradation technique; this sequence was subsequently used by Chroboczek et al. (1992) in compiling the complete Ad5 genome sequence. However, comparison of the published Ad2 and Ad5 sequences (Roberts et al., 1984, Chroboczek et al., 1992) showed a number of differences in the E4 region, including disrupted ORFs 1 and 2 in Ad5. The high ratio of frameshift to point mutations in these putative coding regions, together with the high degree of sequence conservation as compared with adjacent non-coding regions, suggested that the frame-shifts might be sequencing artefacts. To determine whether or not Ad5 could potentially encode E4 ORF1 and ORF2 products, the region of Ad5 between nucleotides 34790 and 35615 was resequenced.

3.2 Cloning of the Ad5 DNA to be sequenced

All adenovirus numbering in this chapter refers to the published sequence of Chroboczek et al., 1992 (Gen Bank acc. no. M73260) amended by Dix and Leppard, 1992 (Gen Bank acc. no. D12587) unless otherwise stated.

The relevant pieces of the Ad5 strain wt300 *KpnI*-G genomic fragment were subcloned into plasmids and then the M13 vectors mp18 and mp19: the *SmaI* (35354) - *HaeIII* (35615) fragment was cloned into the *SmaI* site of pGEM4, to give the subclone pID13, and then into the polylinker of mp18 and 19 using the *HindIII* and *EcoRI* sites of pGEM4; the *HindIII* (34930) - *SmaI* (35354) fragment was cloned into the polylinker of pGEM4, to give the subclone pID11, and then into the polylinker of mp18 and 19 using the *HindIII* and *EcoRI* sites of pGEM4. Single stranded DNA was prepared from the M13 clones and sequenced with the -40 M13 universal primer. The sequence from 34790 to 34930 was determined by sequencing directly a plasmid

subclone (pID9) containing the Ad5 wt300 *Bgl*III (34115) to *Hind*III (34930) fragment using an SP6 specific primer; IDB (see section 2.1.2.).

3.3 Results

In each of the four regions where the published Ad5 sequence contained frameshifts in ORF1 and ORF2 (positions 34934, 35320/21, 35509/10 and 35522/23), the sequence obtained showed the ORFs to be intact, as in Ad2 (see fig. 3.1., and table 3.1.). All except one of the other differences between Ad2 and Ad5 expected on the basis of previously published data were confirmed (see table 3.1.) discounting the possibility that the DNA sequenced was of Ad2 origin. The exception was the ORF2 missense mutation at nucleotide 34860 (Ad5) where this data indicates that the Ad5 wt300 strain ORF2 encodes the same amino acid as that of Ad2 at this position (met78). The data presented here would suggest a revised length of 35938 bp for the Ad5 strain wt300 genome as opposed to the previously published length of 35935 bp (Chroboczek et al., 1992). The revised sequences of ORF1 and ORF2, together with the differences as compared to Ad2 (Roberts et al., 1984) and Ad5 (Chroboczek et al., 1992), are displayed in Fig. 3.2 and 3.3, respectively.

3.4. Discussion

The above results demonstrate that both ORF1 and ORF2 are intact in Ad5. The Ad5 nucleotide sequences of both ORF1 and ORF2 are over 99% identical with their Ad2 counterparts, with 3 nucleotides varying in each ORF (see Table 3.1.). However, while the predicted 128 amino acid Ad5 ORF1 is identical in protein sequence with the Ad2 ORF1 sequence, the predicted 130 amino acid Ad5 ORF2 has a single amino acid difference compared to the Ad2 protein (Ad5 = Ser30, Ad2 = Ala30) (Table 3.2.).

Figure 3.1. ORF1 and ORF2 sequence data depicting regions of variation between published Ad2, Ad5 and revised Ad5 sequences. Ad5 wt300 DNA was sequenced using the dideoxynucleotide chain termination method (2.3.1.13.). The arrow indicates the 5' to 3' direction of the L-strand while an * indicates a nucleotide different to that found in either the Ad2 (Roberts et al., 1984) and/or the published Ad5 (Chroboczek et al., 1992) sequence (see Fig. 3.2 and 3.3 and table 3.1.). All gels are loaded G, A, T and C.

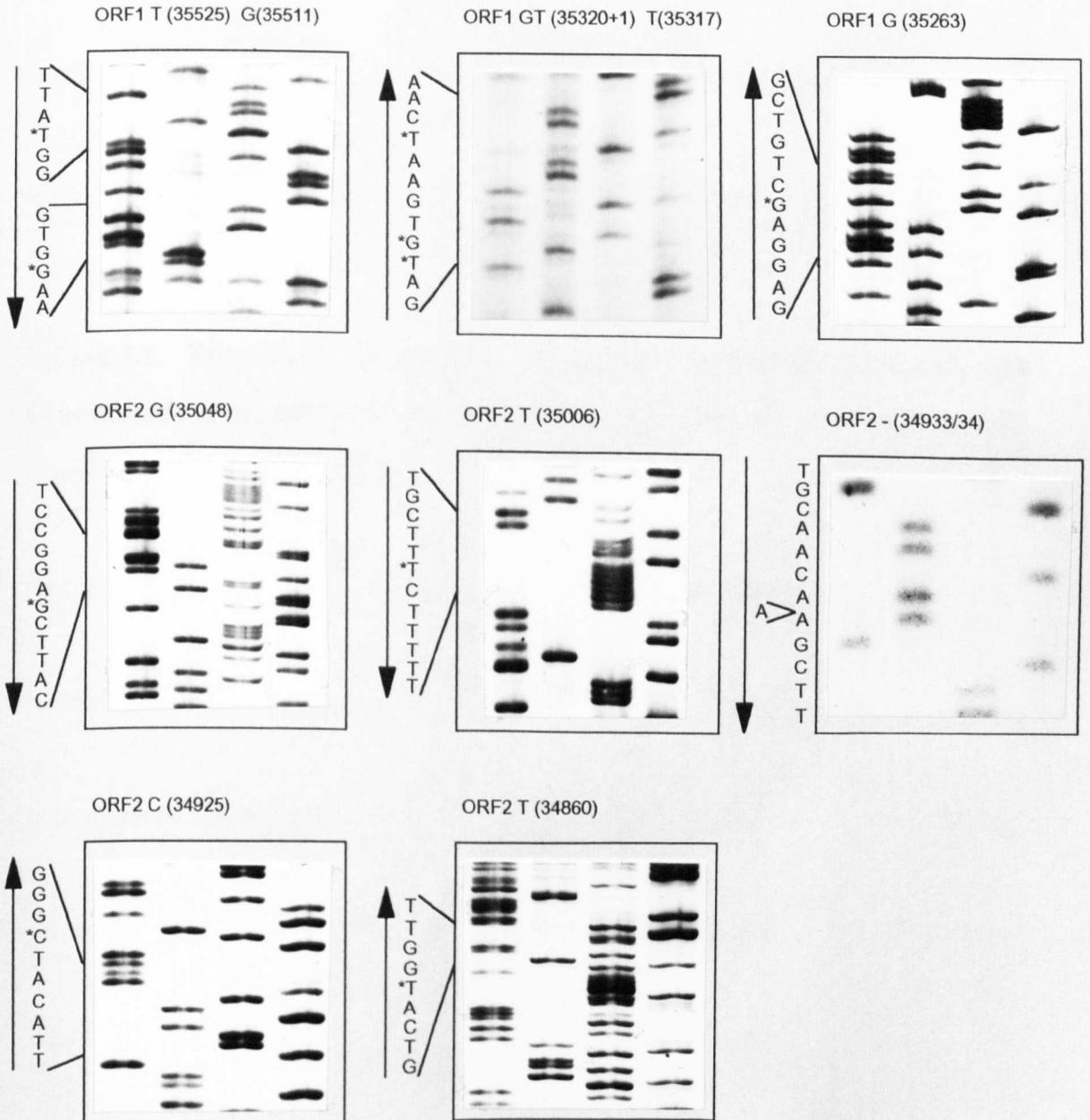


Figure 3.2. The comparison of the revised sequence of Ad5 E4 ORF1 with Ad5 (Chroboczek et al., 1992) and Ad2 (Roberts et al., 1984) 35526-35143 (L-strand).

'-' indicates an absent nucleotide.

ATGGCTGCCGCTGTGGAAGCGCTGTATGTTGTTCTGGAGCGGGAGGGTGCTATTTTGCCTAGGCAGGAGGGTTTTTCAGGTGTTTA

TGTGTTTTTCTCTCCTATTAATTTTGTATACCTCCTATGGGGGCTGTAATGTTGCTCTACGCCTGCGGGTATGTATTCCCCCGG

GCTATTTCCGGTCGCTTTTTAGCACTGACCGATGTGAATCAACCTGATGTGTTACCGAGTCTTACATTATGACTCCGGACATGACC

--
T C

GAGGAGCTGTCGGTGGTGCTTTTTAATCACGGTGACCAGTTTTTTTTACGGTCACGCCGGCATGGCCGTAGTCCGTCTTATGCTTAT

A

AAGGGTTGTTTTCTGTTGTAAGACAGGCTTCTAATGTT Ad5 (Dix and Leppard)
Ad5 (Chroboczek et al.)
Ad2 (Roberts et al.)

Figure 3.3. The comparison of the revised sequence of Ad5 E4 ORF2 with Ad5 (Chroboczek et al., 1992) and Ad2 (Roberts et al., 1984) 35092-34704 (L-strand).

'-' indicates an absent nucleotide.

ATGTTTGAGAGAAAAATGGTGTCTTTTTCTGTGGTGGTTCCGGAGCTTACCTGCCTTTATCTGCATGAGCATGACTACGATGTGCT

A

TTCTTTTTGCGCGAGGCTTGCCTGATTTTTGAGCAGCACCTTGCATTTTATATCGCCGCCCATGCAACAA-GCTTACATCGGG

G

A

A

GCTACGCTGGTTAGCATAGCTCCGAGTATGCGTGTGTCATAATCAGTGTGGGTTCTTTTGTGATGGTTCTTGGCGGGGAAGTGCCGC

A

GCTGGTCCGTGCAGACCTGCACGATTATGTTTCAGCTGGCCCTGCGAAGGGACCTACGGGATCGCGGTATTTTTGTTAATGTCCGC

TTTTGAATCTTATACAGGTCTGTGAGGAACCTGAATTTTTGCAATCA Ad5 (Dix and Leppard)
Ad5 (Chroboczek et al.)
Ad2 (Roberts et al.)

Table 3.1. The DNA sequence variations of the E4 ORF1 and ORF2 regions of Ad5 (Dix and Leppard, 1992), Ad5 (Chroboczek et al., 1992) and Ad2 (Roberts et al., 1992) (sequence changes indicated are to the L-strand)

Ad5 (Chroboczek)		Ad5 (Dix and Leppard)		Ad2 (Roberts)	
34860	A	34860	T	34863	T
34925	C	34925	C	34928	A
34934	A	34933/34	-	34936/37	-
35005	T	35006	T	35008	G
35049	G	35048	G	35051	A
35264	G	35263	G	35269	A
35318	T	35317	T	35323	C
35320/21	--	35320+21	TG	35326+7	TT
35509/10	-	35511	G	35517	G
35522/23	-	35525	T	35531	T

Table 3.2. Amino acid changes due to the nucleotide differences between Ad2 and Ad5 ORF1 and ORF2. Ad5 sequence is from Dix and Leppard (1992) while the Ad2 sequence is from Roberts et al. (1984), sequence is taken from the L-strand.

Adenovirus 5 (Dix and Leppard)		Adenovirus 2 (Roberts)	
Nucleotide	Amino acid	Nucleotide	Amino acid
34925 C	ORF2 Ile 56	34928 A	ORF2 Ile 56
35006 T	ORF2 Ser 30	35008 G	ORF2 Ala 30
35048 G	ORF2 Glu 15	35051 A	ORF2 Glu 15
35263 G	ORF1 Glu 88	35269 A	ORF1 Glu 88
35317 T	ORF1 Asn 69	35323 C	ORF1 Asn 69
35320 G	ORF1 Val 68	35326 T	ORF1 Val 68

3.5. Computer prediction and homology studies of ORF1 and ORF2

3.5.1. Introduction

The Ad5 ORF1 and ORF2 protein sequences deduced from the DNA sequences have been subject to computer analysis to predict protein structure, using both the University of Wisconsin Genetics Computer Group sequence analysis software package (GCG) and the Microgenie sequence analysis program (Queen and Korn, 1984). Two different algorithms were used for secondary structure predictions, that of Garnier et al. (1978) and Chou and Fasman (1974, 1977), and two algorithms were also used to determine the hydrophobicity/ hydrophilicity of the proteins, that of Hopp and Woods (1981) and Kyte and Doolittle (1982). Where possible, the consensus of the results of alternative prediction methods was used in subsequent discussions.

Homology searches were also carried out for these proteins, screening the EMBL data base and the SwissProt data base, using the GCG Fasta and Tfasta programs at default settings.

In the following sections the results from these analyses are discussed.

3.5.2. ORF1 structure predictions

Analysis of the amino acid content of the 128 residue Ad5 E4 ORF1 protein indicates that it is hydrophobic (see figs 3.4 and 3.6) and has a negative overall charge at neutral pH with a pI of 5.12.

The sequences of a number of different adenovirus ORF1 proteins have been published. Using the Pileup and Pretty programs on the GCG package, ORF1 proteins representative of 4 adenovirus subgroups were analysed for homology: the subgroup A virus Ad12, 52% identity to Ad5 (Sprengel et al., 1994); the subgroup B virus Ad34, 29% identity to Ad5 (Chen and Horwitz, 1990); the subgroup C virus Ad5 (Dix and Leppard, 1992), the subgroup D viruses Ad9, 45% identity to Ad5 (Javier and Shenk, 1992), (see Fig. 3.5.). The E4 sequence of Ad40, a subgroup F

Figure 3.4a. Ad5 ORF1

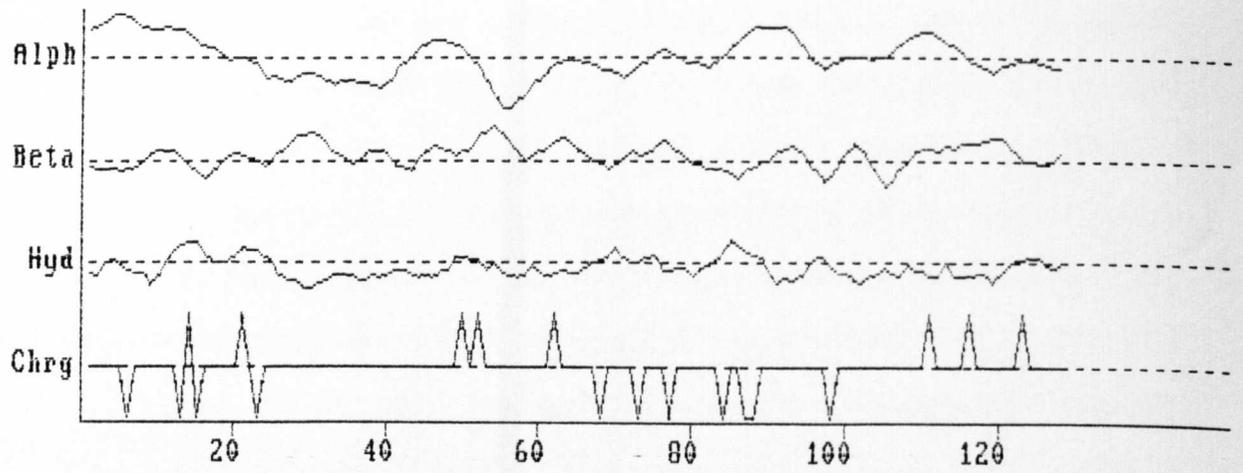


Figure 3.4b. AD5 ORF2

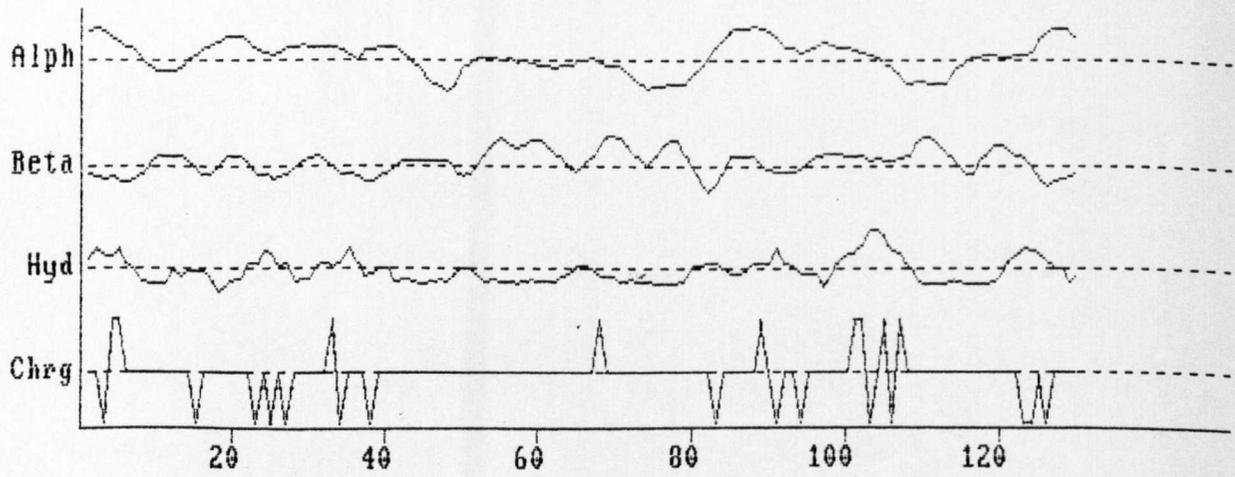


Figure 3.4. The structure of Ad5 ORF1 and ORF2 as predicted by the Microgenie Sequence Analysis Program (Queen and Korn, 1984). Displayed are one dimensional, multi-paneled plots with the protein residues on the x axis and the attributes represented as continuous curves on several different panels. All structure was determined using the algorithm of Garnier et al. (1978), while hydrophobicity was determined using the algorithm of Hopp and Woods (1981). Values above the dashed line indicate an increased probability of the protein containing either an alpha helix (Alpha), beta sheet (Beta) and/or a hydrophilic region (Hyd) at this point and vice versa for values below the dashed line. 'Charg' indicates the predicted net charge at this point; values above the norm indicate a positive charge while values below the norm indicate a negative charge.

virus, has been published by Davison et al. (1993) and does not contain any region capable of encoding an ORF1-related protein. This would indicate that Ad40, which causes infantile gastro-enteritis, does not require the ORF1 product for efficient replication within the gut. The other viruses sequenced, which all retain the ORF1 protein, are not primarily associated with gut infections, suggesting the protein is necessary for viral replication only in certain cells of the host.

From Fig. 3.5 it can be seen that there is one distinct region of homology between all of the ORF1 proteins, between residues 53 and 68 (Ad5); VxIPxGYxGxFLxLxD. This may signify a functionally important region of the protein. As a result of this sequence conservation, the predicted secondary structure of the region, a turn - β -strand - turn, is conserved in all of the ORF1 proteins at this point. The lack of any other distinctive regions of homology between all four ORF1 proteins can be attributed to the divergence of the Ad34 ORF1 protein. On closer inspection of the Ad34 right end DNA sequence, a good homology to the other Ad ORF1's from about residue 70 onwards to the C terminus was identified in the +1 reading frame, suggesting that there has been a frame-shifting sequencing error in this region of E4 in the data of Chen and Horwitz (1990), (Figure 3.5.). Presuming this to be the case, the Ad34 protein is probably about 124 amino acids in length as opposed to the 108 amino acids of the current sequence.

Assuming that the C terminus of the Ad34 ORF1 protein has the alternative sequence discussed previously, all the Ad ORF1 proteins have a second conserved region between residues 88 to 96 (Ad5); ELxVxLFNH. This region is also close to a potential N-linked glycosylation site in the Ad9 (NHT, residues 92 to 94) and Ad12 (NHS, residues 94 to 96) ORF1 proteins. However this site is not conserved in the Ad5 and Ad34 sequences suggesting that it may not be of significance. Apart from conserved sequence motifs, the ORF1 proteins also show similarities in their predicted secondary structure and distribution of hydrophobic residues that are not readily detectable from the primary sequence. A distinct hydrophilic

	1					50
ad5orf1	MAAAVEALYV	VLEREGAILP	RQEGFSGVYV	FFSPINFVIP	PMGAVMLSLR	
ad9orf1	MA...ESLYA	FIDSPGGIAP	VQEGTSNRYT	FFCPESFHIP	PHGVVLLHLK	
ad12orf1	MA.AFETLYV	YFTGPGAML	KQEGDSNAYV	LFSAANFVIP	PHGVVLLYLH	
ad34orf1	M...IEALYV	YLEGPGATLP	EQQQ.RNNYI	FYSPVPFTLY	SRGVALLYLR	

Conserved (4)	M----E-LY-	-----G---P	-Q-----Y-	**-----F---	--G--*L-L-	
Conserved (3)	M*---E-LYV	Y**--PGA-LP	-QEG-SN-Y*	FFSP--F-IP	P-GVVLL-L-	

	51					100
ad5orf1	LRVCIPPGYF	GRFLALTDVN	QPDVFTESYI	MTPDMTEELS	VVLFNHGDQF	
ad9orf1	VSVLVPTGYQ	GRFMALNDYH	ARDILTQSDV	IFAGRRQELT	VLLFNHTDRF	
ad12orf1	IAVDIPPGYL	GTLFSLCDMN	ARGVFGAET	LYPGSRMELS	VLLFNHSCVF	
ad34orf1	LSIIIPKGYV	GCFFSLTDAN	SLDCMLHQEL	FMLVTARNFR	CCCLITMTGF	
ad34orf1 (alt)			LYASSRI	IHAGHREELS	VLLFNHHDHF	

Conserved (4)	---*P-GY-	G**-L-D--	-----*-----	-----*-	---*-----F	
Conserved (3)	--V-IP-GY-	G-F*-L-D-N	--D**-----*	-----EL-	V-LFNH---F	

	101			128
ad5orf1	FYGHAGMAVV	RLMLIRVVFP	VVRQASNV	
ad12orf1	CDVRAKQPVA	RLLLSRVVFP	PVCQASLI	
ad9orf1	LYVRKGHVPV	TLLLERVIFP	SVKIATLV	
ad34orf1	..TRSSWRPC	GLVL.....	
Ad34orf1 (alt)	YEVELETLWA	CLVMERLIYP	PVRQATMI	

Conserved (4)	-----	-L-L-----	-----	
Conserved (3)	---R----V-	-L*L-RV*FP	-V--A--*	

Figure 3.5. Alignment of the predicted Ad5, Ad9, Ad12 and Ad34 E4 ORF1 protein sequences, prepared using the GCG PileUp and Pretty programs. Displayed are the points of identity in the sequences and any conserved amino acids present in all four of the sequences (Conserved (4)) and three of the sequences (Conserved (3)). An * denotes where there is a conservative constitution in all 4 of the proteins (in 3 of the proteins with Conserved (3)). The Ad5 , Ad9, Ad12 and Ad34 E4 ORF1 proteins are 128, 125, 127 and 108 amino acids respectively. The Ad34ORF1 (alt) is the alternative Ad34 ORF1 C-terminal sequence, corrected for the presumed sequence error in the DNA (see text).

Figure 3.6. Predicted secondary structures of the Ad5, Ad9, Ad12 and Ad34 ORF1 proteins as determined using the Plotstructure and Peptidestructure commands of the Genetics Computer Group Sequence Analysis Software (version 3.1, 1991). KD Hydrophobicity = hydrophobicity as determined using the algorithm of Kyte and Doolittle (1982). CF = secondary protein structure as determined by the algorithm of Chou and Fasman (1978). GOR = secondary protein structure as predicted by the algorithm of Garnier et al. (1978). See page 111a for an extended legend.

Figure 3.6a. Ad5 ORF1

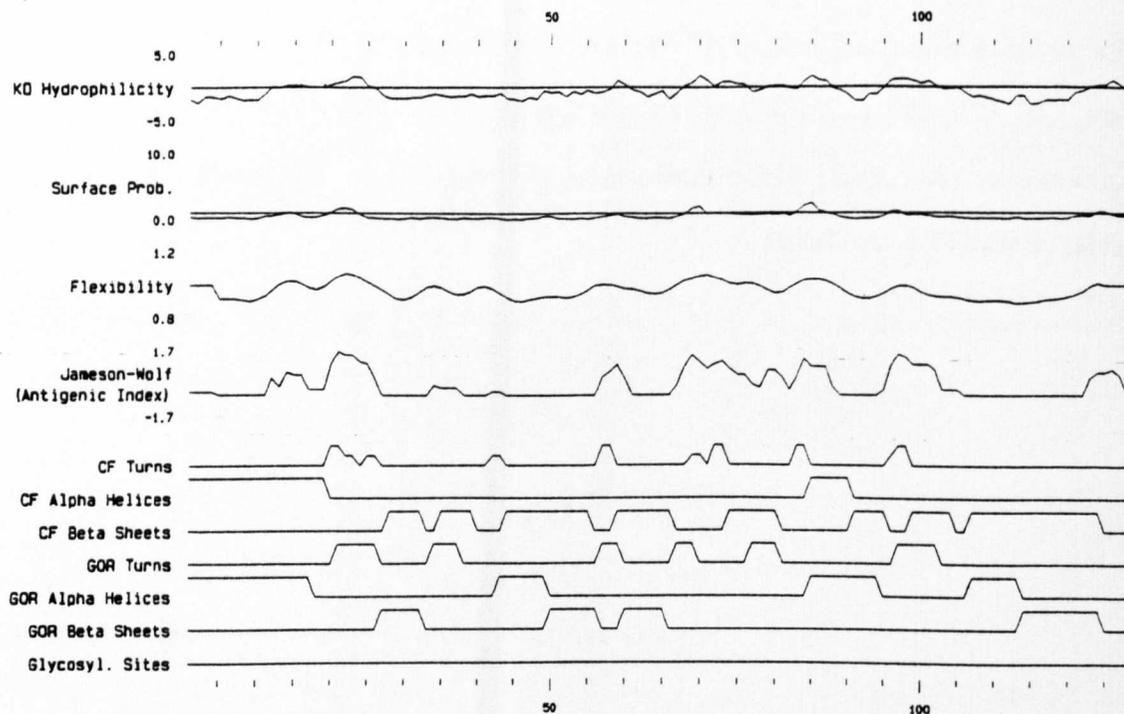


Figure 3.6b. Ad9 ORF1

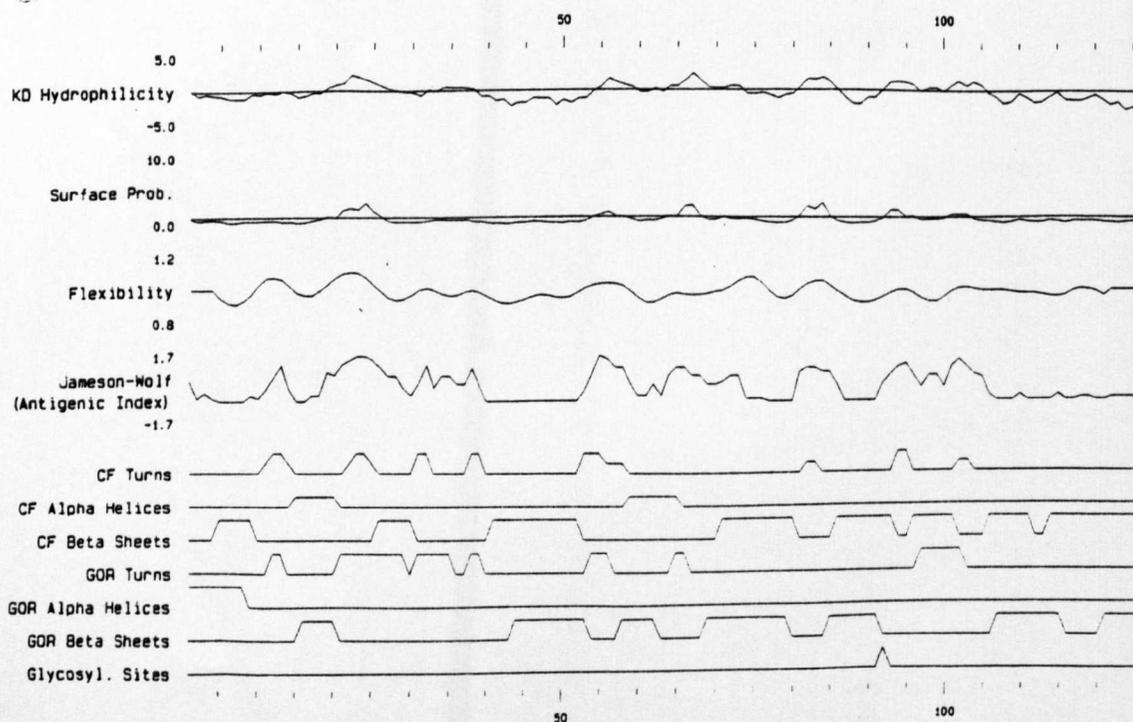


Figure 3.6c. Ad12 ORF1

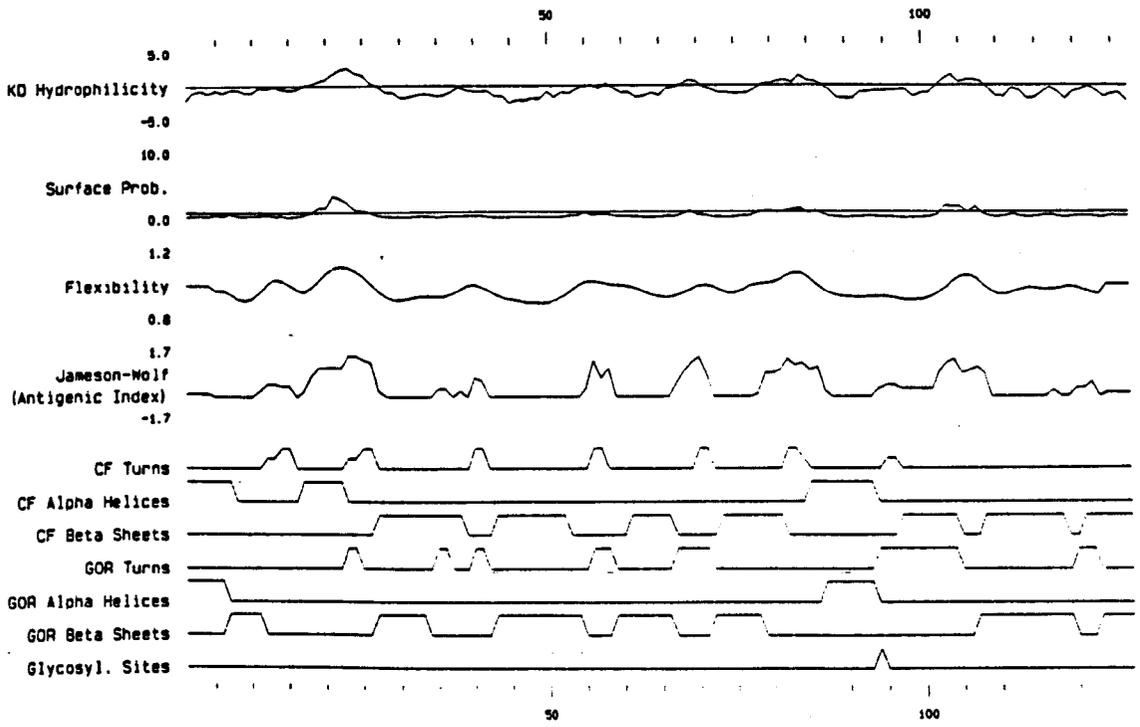
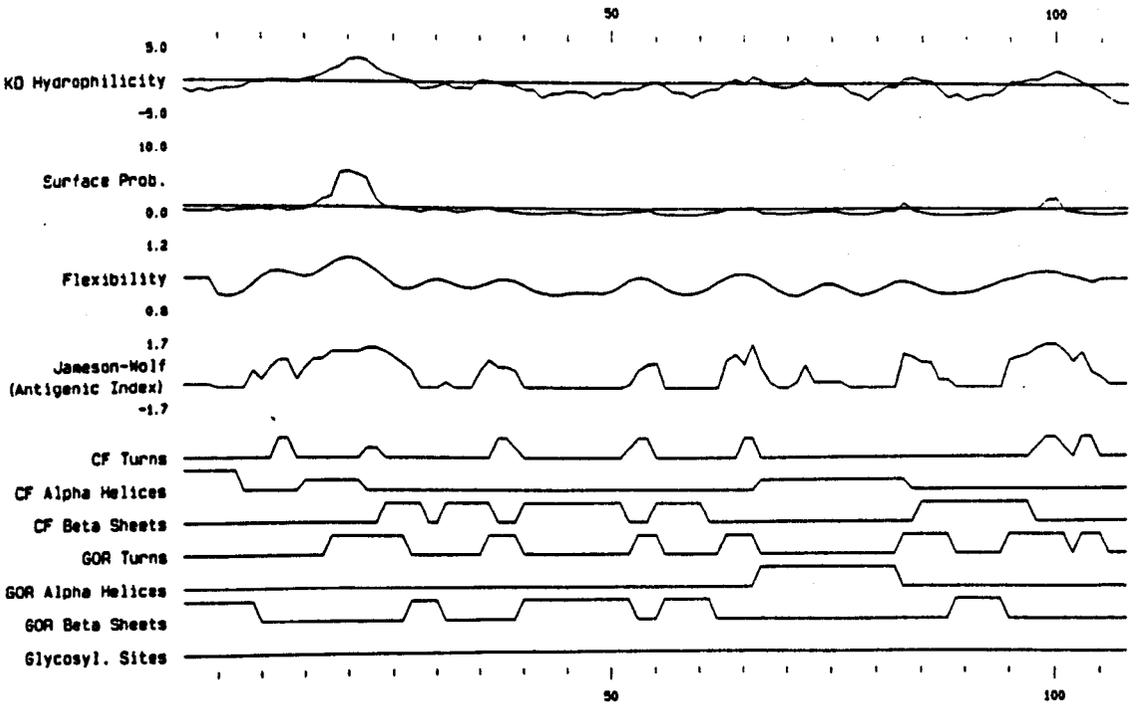


Figure 3.6d. Ad34 ORF1



Extended legend for Figures 3.6 and 3.9. Figures 3.6a-d and 3.9a-d are one dimensional, multi-paneled plots with the protein residues numbered on the x-axis and the attributes represented as continuous curves in each of several different panels. Values above the horizontal line across the KD Hydrophilicity panel indicate hydrophilic residues at this point in the protein while values below the line indicate hydrophobic residues. The horizontal line across the surface probability (surface prob.) panel at position 1.0 on the y-axis indicates the expected surface probability calculated for a random sequence. Values above this line indicate an increased probability of this region being found on the protein surface as determined by Emini *et al.* (1985). Flexibility is an indication of the flexibility of the 'backbone' of the protein with high values indicating a high flexibility. The antigenic index (AI) is a measure of the probability that a region is antigenic. It is calculated by summing several weighted measures of secondary structure (Jameson and Wolf, 1988). The higher the value the on the y-axis the higher the probability that the region is antigenic. The secondary structure motifs (turns, alpha helices, beta sheets) were predicted using either 1) the method of Chou and Fasman (1978) modified by Nishikawa (1983) to resolve overlapping regions of alpha helices and beta sheets and to locate turns that are not in conflict with other secondary structures (CF) or 2) the method of Garnier *et al.* (1978) (GOR). The higher the value on the y-axis the higher the probability of the relevant secondary structure at this point. Glycosylation sites are predicted for sites where the residues have the composition NxT or NxS.

region at about residue 23 and other, less distinct hydrophilic regions at about residues 70 and 85 are present in all four proteins. Turns in the predicted structure can be identified in all of the sequences at about residues 23, 57 and 97. Predicted alpha helix at the amino-terminus and β -sheet structures between residues 45 to 55 and from about residue 110 to 125 are also seen in all the ORF1 protein sequences.

3.5.3 ORF1 homology searches

A data base search (EMBL and SwissProt) for homology of Ad5 ORF1 to any other known protein sequence, using the GCG Fasta and Tfasta programs, revealed that the ORF1 protein has homology to a 147 amino acid deoxyuridine 5'-triphosphate pyrophosphatase protein (dUTPase) found in *Saccharomyces cerevisiae* (SwissProt acc.no. P33317), (see Fig. 3.7.). This protein has 26.9% sequence identity with ORF1 over 119 amino acids and the significance of this homology is strengthened by the observation that 15 of the 30 amino acids conserved between all four adenovirus ORF1 proteins are conserved between Ad5 ORF1 and the dUTPase protein (Fig. 3.7). The homology is particularly evident over the conserved motif ELxVxLFNH at residues 88 to 96 in the ORF1 protein (Ad5). ORF1 was also found to have homology to a similar dUTPase protein of Vaccinia virus (SwissProt acc.no. P17374) although there is only 17.5% identity over 120 amino acids. However, ORF1 only has poor homology with other viral dUTPase enzymes (McGeoch, 1990) and it does not have strong homology with the 5 recognised protein motifs characteristic of dUTPases (Fig. 3.7.) (McGeoch, 1990; McIntosh et al., 1992; Gadsden et al., 1993). ORF1 is actually truncated compared to most dUTPases, lacking the C-terminus 'P-loop' structure thought to be important in the binding of nucleotides (McIntosh et al., 1992). dUTPases ensure the cellular dUTP level is kept low by catalysing dUTP hydrolysis, creating dUMP, and thereby effectively excludes dUTP as a substrate for DNA polymerase. Extensive uracil incorporation into DNA is potentially lethal for the cell as it induces extensive excision repair which is believed to lead to DNA fragmentation.

A more restricted data base search (SwissProt), for homologies with the two ORF1 regions which are conserved in all the adenovirus ORF1 proteins (residue 53 to 68 and residue 88 to 96 (Ad5)), was also carried out using the GCG Fasta program. The VxIPxGYxGxFLxLxD motif (Ad5 ORF1, residues 53 to 68), was found to have 60% identity with a sequence found in the Marburg virus RNA dependent RNA polymerase (SwissProt acc.no. P31352). The ELxVxLFNH motif (Ad5 ORF1, residues 88 to 96), which was found to be conserved with the yeast dUTPase, has good homology (86% identity) to the BMRF1 protein of the early antigen-D complex of Epstein-Barr Virus (EBV). Only Glu88 of ORF1 was not identical with the BMRF1 motif having another acidic amino acid, aspartate at this point. Even the less well conserved amino acids of this motif in ORF1, i.e. Ser90 and Val90, are conserved, with BMRF1 having serine and isoleucine respectively at these positions. The 50kDa BMRF1 protein is thought to be a activator of transcription in virus-infected cells (reviewed by Kieff and Liebowitz, 1990; Miller, 1990) and also to be involved in DNA synthesis, interacting with the EBV DNA polymerase. What role the VxIPxGYxGxFLxLxD and ELxVxLFNH motifs play in the function of these proteins is as yet unclear.

A search for known protein motifs within the ORF1 protein sequence, using the GCG Motif program, was unsuccessful.

3.5.4. ORF2 structure predictions

Analysis of the amino acid content of the 130 residue Ad5 E4 ORF2 protein indicates that it is hydrophobic (Figs. 3.4 and 3.9) and has a negative overall charge at neutral pH with a pI of 4.91.

The sequence of a number of other adenovirus ORF2 sequences have been published. Using the Pileup and Pretty programs on the GCG package, ORF2 proteins representative of 4 adenovirus subgroups were analysed for homology: the subgroup A virus Ad12, 48% identity to Ad5 ORF2 (Sprengel et al., 1994), the subgroup C virus Ad5 (Dix and Leppard, 1992), the subgroup D virus Ad9, 31% identity to Ad5

	1					50
ad5orf2	MFERKMVSFS	VVVPELTCLY	LHEHDYDVLS	FLREALPDFL	SSTLHFISPP	
ad9orf2	MLQRRGVSYH	IVVPGVLVTY	LE..DFSITQ	MIKEKLPRFI	THILEGITGD	
ad12orf2	MFQRSLVHYS	VLFPESLRNY	LHGLDFEVVT	FLKDVLPFW	LLVMHYLTPP	
ad40orf2	MYLLRPVFIS	VVVPEPLMFY	LNNMDISVVD	FLKTNMSEFL	LHILHCVTPV	

Conserved (4)	M*-----V---	*--P-----Y	L---D---*--	-----*--F-	---*-----	
Conserved (3)	M*-R--V-*S	VVVPE-*--Y	L---D*-V--	FLK*-LP*F-	--*LH-*TP-	

	51					100
ad5orf2	MQQAYIGATL	VSIA.PSMRV	IISVGSFVMV	PGGEVAALVR	ADLHDYVQLA	
ad9orf2	TKRAYSSMQF	LGASFGALKF	SLTLASPTLS	PGSEL SAVVA	QDLSDFLQLT	
ad12orf2	MRDVYVGATL	TNMG.PFVQV	VCSVGTPELV	PGGELSLLLA	SDLYDFIQLA	
ad40orf2	LQAAYCGFTL	TSLA.PTLEM	ICCVASPELL	PDSELTELIC	NLDLDFVQLM	

Conserved (4)	-----Y-----*	-----	-----*--	P--E-----	-DL-D*-QL-	
Conserved (3)	*--AY-G-TL	-----P-*--	*--V-SP-L-	PG-EL--L*-	-DL-DF*QL-	

	101					133
ad5orf2	LRRDLRDRG.	IFVNVPLLNL	IQVCEEPEFL	QS.		
ad9orf2	LRRELRAEGR	TLLNLVVLNT	LQVVEQQDLF	LL.		
ad12orf2	LRCQLRDQG.	VEPNVNLLNL	LQVFEDPDDF	QQI		
ad40orf2	LRMEIRDRG.	FEPNVNLLNE	LQVTREPDFL	QP.		

Conserved (4)	LR---R--G-	---N---LN-	-QV-----*-*	---		
Conserved (3)	LR-ELRD-G-	---NV-LLN-	LQV-E*PD-F	Q--		

Figure 3.8. Alignment of the predicted Ad5, Ad9, Ad12 and Ad40 E4 ORF2 protein sequences, prepared using the GCG PileUp and Pretty programs. Displayed are the points of identity in the sequences and any conserved amino acids present in all four of the sequences (Conserved (4)) and three of the sequences (Conserved (3)). An * denotes where there is a conservative substitution in all 4 of the proteins (in 3 of the proteins with Conserved (3)). The Ad5, Ad9, Ad12 and Ad40 E4 ORF2 proteins are 130, 130, 131 and 130 amino acids respectively.

ORF2 (Javier and Shenk, 1992), and the subgroup F virus Ad40, 42% identity to Ad5 ORF2 (Davison et al., 1993), (Fig. 3.8). As yet it is unclear whether Ad34 has an ORF2 reading frame as only a small region of the Ad34 genome has been sequenced. The Ad5, 9 and 12 ORF2 proteins lack any potential N-linked glycosylation sites; the Ad40 ORF2 protein however has one such site at residue 35. This difference in potential glycosylation coincides with a different predicted pattern in hydrophobicity/hydrophilicity in the N terminal half of the proteins. The Ad5, 9 and 12 ORF2 proteins have a hydrophilic peak around residue 50 while Ad40 ORF2 lacks this peak but has an alternative peak at residue 35, the predicted glycosylation site. All the ORF2 proteins have a well conserved hydrophilic peak between residues 100 and 110. The ORF2 proteins have few regions where conserved residues are concentrated, one that does stand out is the DLxDxxQLxLR motif between residues 91 and 102. This motif of charged amino acids has no distinct predicted secondary structure element associated with it, but its conservation is suggestive that it is important in function. A search for known protein motifs, using the GCG Motif program, within the ORF2 protein was unsuccessful.

3.5.5 ORF2 homology searches

When the protein sequence data base (EMBL and SwissProt) was searched for homologies to Ad5 ORF2 no distinctive homology to any other known protein was found, apart from other adenovirus E4 ORF2 proteins. However, a search for the conserved ORF2 motif of DLxDxxQLxLR at residues 92 to 102 found good homologies in three other proteins: a rat peptidylarginine deiminase (SwissProt acc.no. P20217), a putative adenylate cyclase regulatory protein (SwissProt acc.no. P23799) and a diphtheria toxin repressor protein (SwissProt acc.no. P33120). The significance of these homologies, if any, in the function of the ORF2 protein is unclear.

Figure 3.9. Predicted secondary structures of the Ad5, Ad9, Ad12 and Ad40 ORF2 proteins as determined using the Plotstructure and Peptidestructure commands of the Genetics Computer Group Sequence Analysis Software (version 3.1, 1991). KD Hydrophobicity = hydrophobicity as determined using the algorithm of Kyte and Doolittle (1982). CF = secondary protein structure as determined by the algorithm of Chou and Fasman (1978). GOR = secondary protein structure as predicted by the algorithm of Garnier et al. (1978). See page 111a for an extended legend.

Figure 3.9a. Ad5 ORF2

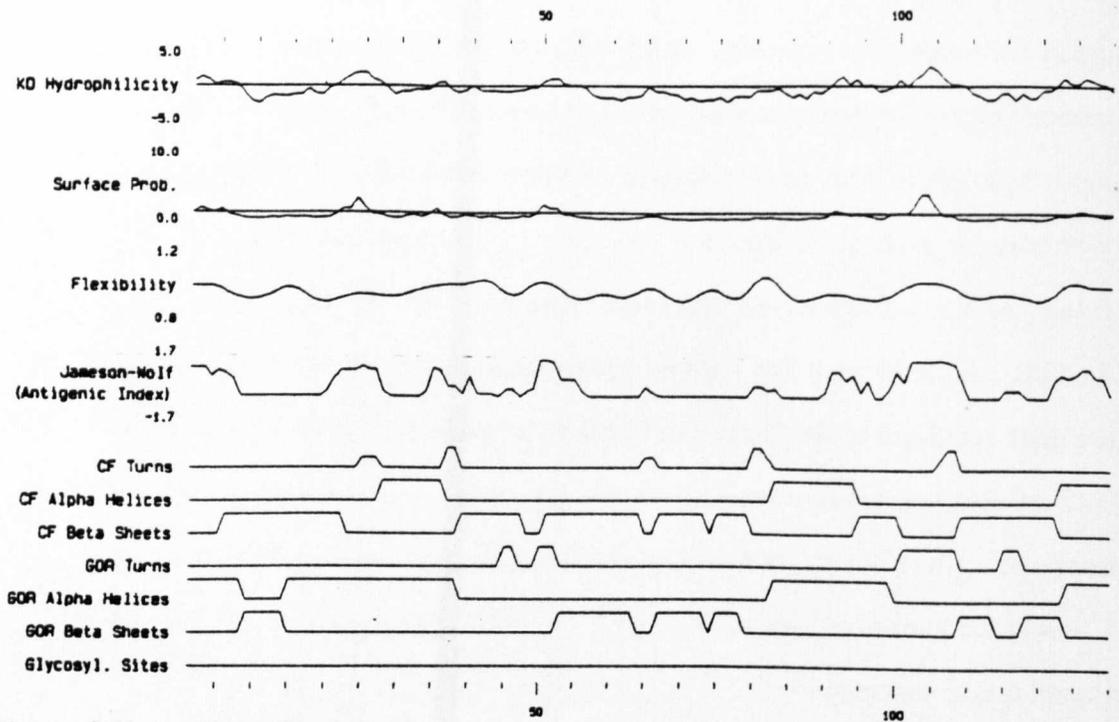


Figure 3.9b. Ad9 ORF2

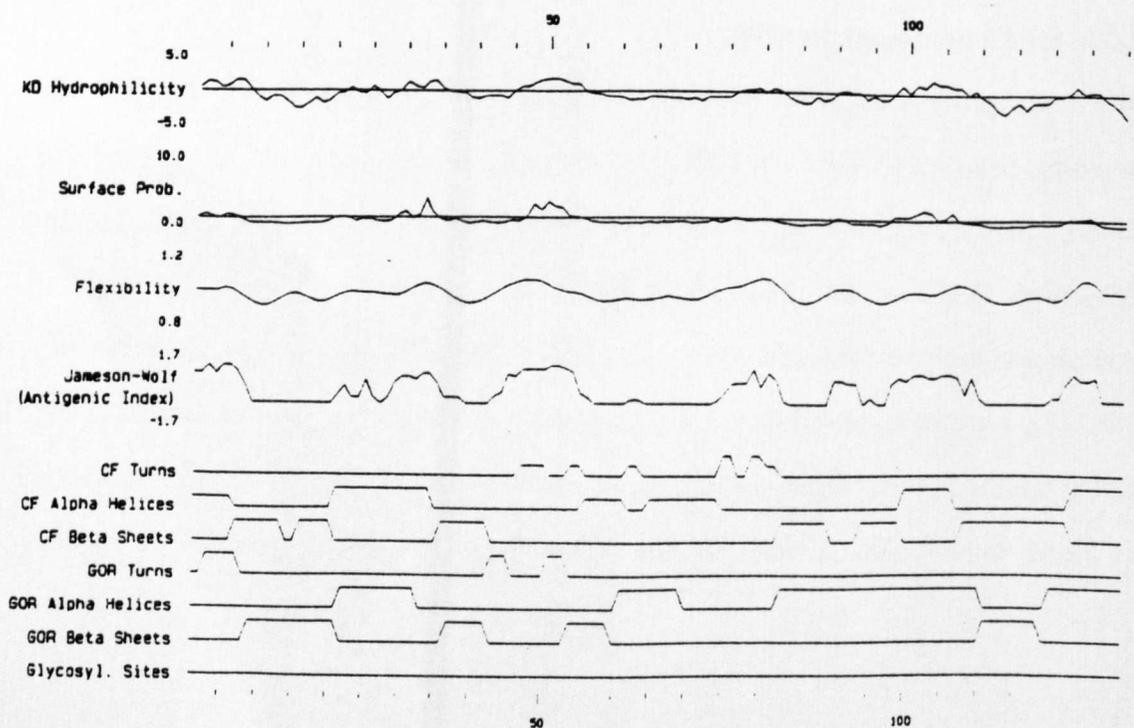


Figure 3.9c. Ad12 ORF2

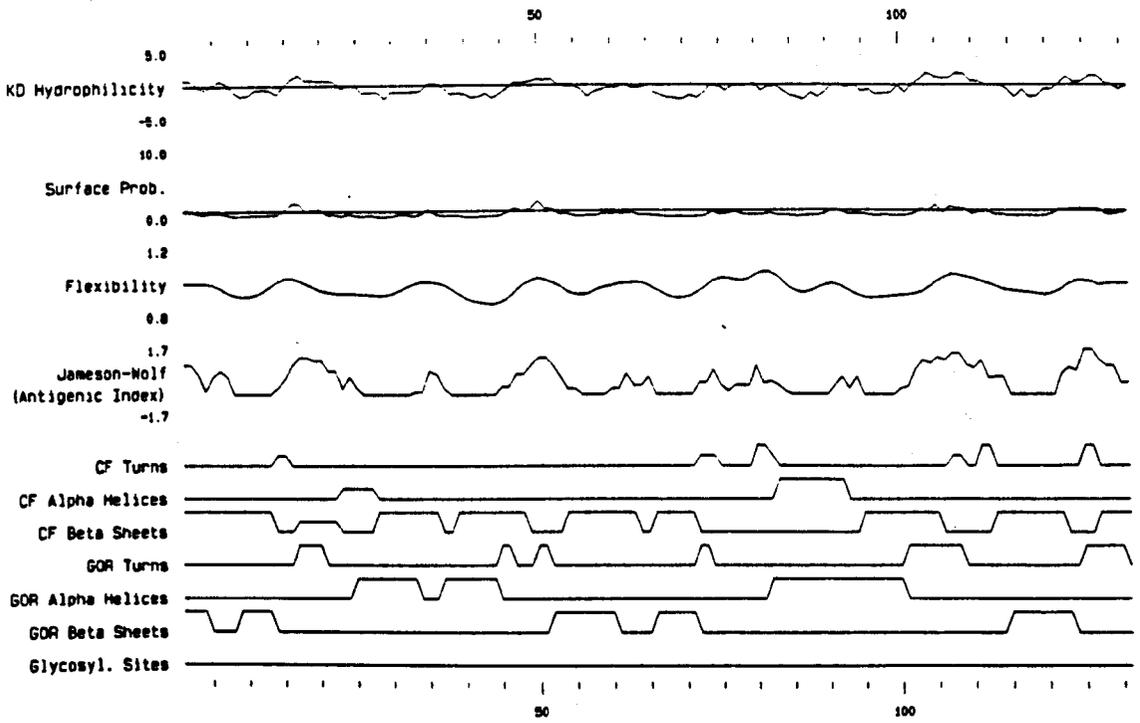
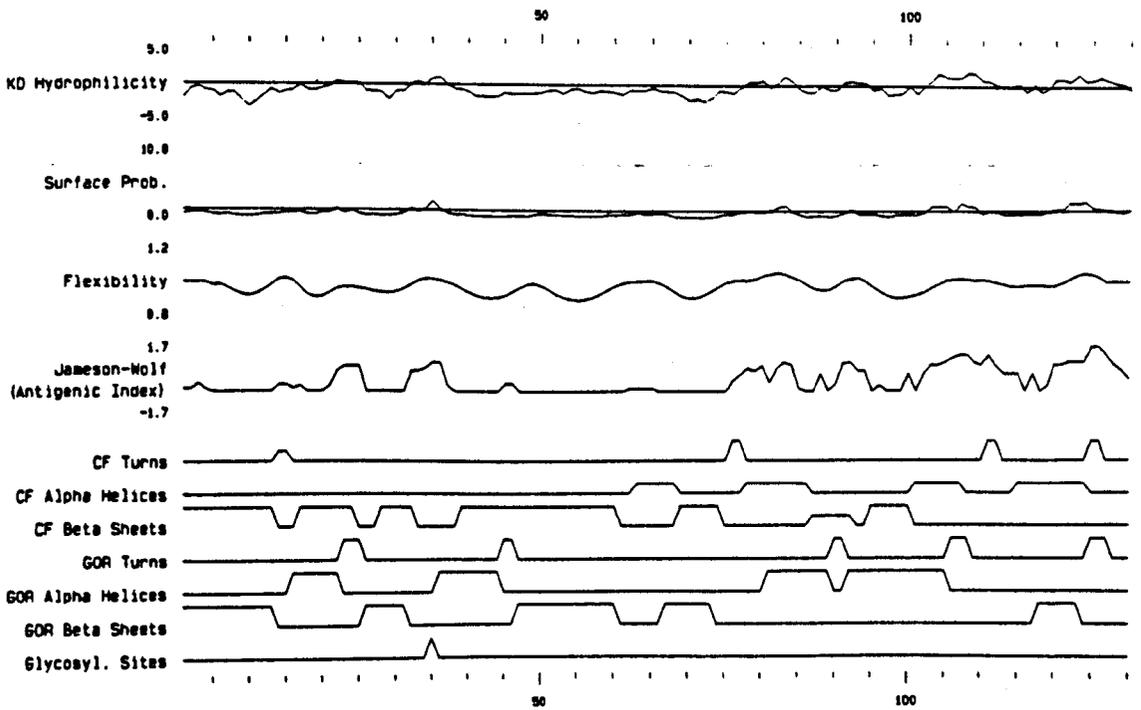


Figure 3.9d. Ad40 ORF2



3.5.6. Conclusion

The analyses presented here do not allow the prediction, with any certainty, the functions for either ORF1 or ORF2. However, the ORF1 protein displays some homology to a number of different proteins involved in nucleic acid metabolism. The fact that ORF1 has homology over almost its entire region with a dUTPase, together with the fact that many of the conserved ORF1 residues are also present in the yeast dUTPase, would suggest that ORF1 has dUTPase activity. However it should be noted that Ad5 ORF1 does not have homology with the conserved motifs found in dUTPases indicating that, while ORF1 may not be a classical dUTPase, it may have a related function. Despite this, it may still prove productive to examine the role of ORF1 in dUMP and dUTP metabolism in infected cells. How this possible function in nucleic acid metabolism is related to the phenotype associated with the ORF1 protein of Ad9, i.e. that of cellular transformation (see section 1.4.2.2.), is unclear.

Despite the absence of an obvious function for these proteins, they clearly are required for the virus infection of its host organism as they are well conserved across widely divergent adenoviruses. Viruses, in general, do not carry genetic information that is not important in their propagation, as carrying extra, nonessential nucleic acid is a drain on the virus's resources, making replication less efficient. On this premise one can conclude that both ORF1 and ORF2, found in most serotypes of adenovirus, are necessary for viral replication in the host organism, even if they have no discernible function in tissue culture.

Chapter 4.

The Expression of the Ad5 E4 mRNAs.

4.1 Introduction

As discussed in the introduction, the lytic infectious cycle of Ad5 is divided by convention into early and late phases, separated by the initiation of viral DNA replication. The E4 region is designated an early region as it is first expressed before the initiation of DNA replication. However like many of the other early regions, the expression of E4 is not confined to the early phase, with expression continuing into the late phase of the infection, although E4 transcription is reduced late in the infection.

A number of studies have explored the pattern of Ad2 E4 mRNA expression. In this section E4 mRNAs A-L are as referred to in Virtanen et al. (1984), with mRNAs M, N and O inferred from the cDNA clones (pGY7, pGY4 and pGY9 respectively) of Freyer et al. (1984). Initial studies of the E4 region identified 3 differentially spliced E4 mRNAs which had identical 5' and 3' termini (mRNAs C, D and G, see Fig. 4.1.), (Berk and Sharp, 1978). Berk and Sharp (1978) identified these RNAs by performing endonuclease S1 and exonuclease VII assays on cytoplasmic RNA preparations of cytosine arabinoside-treated cells, 8 hours post-infection. Chow et al. (1979), using electron microscopy of heteroduplexes of viral mRNA and single-stranded Ad2 DNA, identified 6 E4 mRNAs, believed to be the A,D,E,G,H and L mRNAs of Virtanen et al. (1984). Virtanen et al. (1984) used a number of different techniques to examine the E4 transcripts. As well as using northern blot and S1 nuclease analysis they constructed a cDNA library from mRNA prepared from Ad2 infected cells treated with cycloheximide, at 2 to 10 hours post-infection. They identified 11 mRNAs in the cytoplasm of cycloheximide treated cells, mRNAs A-L. From these studies they concluded that there were 4 splice donor sites (D1, D2a, D2b and D3) and 6 splice acceptor sites (A1a, A1b, A1c, A1d, A1e, and A3), (Fig 4.1.). Two other studies conducted at the same time as the Virtanen study (Freyer et al., 1984; Tigges and Raskas, 1984) largely confirmed these results, but with some detailed differences. Freyer et al. (1984) constructed a cDNA library from mRNA prepared from anisomycin treated cells at 7 or 8 hours post infection. They identified 5 intact cDNA

clones, corresponding to mRNAs D,F,K and L of Virtanen et al. (1984) and a novel mRNA referred to here as mRNA M. mRNA M is constructed using a novel splice acceptor site referred to as A4. Freyer et al. (1984) also isolated a number of partial cDNA clones, two of which implied the existence of novel mRNAs, referred to here as mRNAs N and O. mRNA O is formed by the utilisation of a 3' acceptor site close to the A3 site of Virtanen et al. (1984) referred to as A5 in this study. It is unclear if the use of this alternative splice site A5 affects an ORF or if it just alters an UTR. Tigges and Raskas (1984) used S1 nuclease and exonuclease VII analysis to identify E4 transcripts in the cytoplasm of cycloheximide- and cytosine arabinoside-treated cells at 5 h and 7 h post-infection respectively. They identified a novel full length, unspliced transcript and a number of mRNAs akin to mRNA L identified by Virtanen et al (1984). These mRNA L-like mRNAs were formed by the utilisation of alternative D3 sites and A3 sites, including the A5 site identified by Freyer et al. (1984). These mRNA L-like mRNAs are not included in Fig. 4.1. (except mRNA O also identified by Freyer et al. (1984)) as no evidence for the use of these novel splice sites has been indicated in any other studies. To conclude, the primary transcript from the E4 region is differentially spliced, producing a set of mRNAs which have a common 5' terminus (see 1.4.1.) and polyadenylation site (Fig. 4.1.).

It should be noted that only in the study of Chow et al. (1979) (and a limited study included in Tigges and Raskas,(1984)) were virus infections conducted in the absence of translation or DNA replication inhibitors. Chow et al. (1979) demonstrated that the inhibition of translation or DNA replication during an infection, which prevents the early to late transition, increases the relative abundance of the early mRNAs, including the E4 mRNAs, in the cytoplasm. This increased abundance of E4 mRNAs facilitated the studies of the pattern of E4 expression. However, it cannot be discounted that some mRNAs observed by studies which utilised inhibitors, may be artefacts of the experiments since it is unclear how the inhibition of protein synthesis or DNA replication affects the splicing of E4 transcripts.

Chow et al. (1979) and Tigges and Raskas (1984) both conducted limited experiments examining the pattern of expression of the E4 mRNAs over the course of the infection, in the absence of inhibitors. Both groups concluded that early in the infection a long mRNA predominated, probably mRNA D, while after DNA replication a short mRNA accumulates, probably mRNA L. These observations were confirmed by Ross and Ziff (1992) while comparing E4 expression in abortive and non-abortive infections. All of these experiments suggest that the pattern of mRNA expression from the E4 region is temporally regulated.

The results of all these experiments has led to some confusion as to the precise expression pattern for the E4 mRNAs over the course of a lytic infection in the absence of translation or DNA replication inhibitors. To characterise the pattern of expression of the known and putative E4 products in more detail, a systematic study of the levels of individual mRNAs during the Ad5 infectious cycle was undertaken. In parallel with this study, the role of the E1b-55K and E4 ORF6-34K proteins in regulating the cytoplasmic accumulation of the E4 mRNAs was also examined, as it had been demonstrated that cytoplasmic accumulation of late viral messages was dependent on the E1b/E4 heterodimer (see introduction). The approach used to investigate the expression of the E4 mRNAs was to probe cytoplasmic and nuclear RNA samples, prepared from infected cells at different times post infection, for the E4 mRNAs by RNase protection analysis. This allowed the quantification of splice site utilisation in the production of the E4 mRNAs over the course of the infection.

4.2. Methods

4.2.1. Cells and Viruses.

dl309 was used as a wild-type virus for the purposes of these studies (dl309 was generated from wt300 and retains only one of the four *Xba*I cleavage sites (Jones and Shenk, 1978)). It displays a wild-type phenotype and is the parental virus of all the mutant viruses employed in these experiments. Mutant viruses were used in this study

to examine the expression of the E4 region in the absence of the E1b-55K protein and/or E4 ORF6-34K proteins. dl338 carries a deletion of 524 bp in the E1b coding region located between 2805 and 3329 bp, preventing the synthesis of the E1b-55K and E1B-17K (156R, Fig. 1.4.) proteins (Logan and Shenk, 1984). However, Pilder et al. (1986b) demonstrated that the phenotype associated with this mutant, i.e. the reduction in the accumulation of late viral mRNAs in the cytoplasm of the infected cell, was due to the lack of the E1b-55K protein; Pilder et al. constructed a variant, pm380, which has a base pair change at position 3275, which destroys the 3' acceptor site required for the synthesis of the E1B-17K mRNA. pm380 grew as a wild-type virus suggesting that the E1b-17K protein has no effect in virus replication in tissue culture and that the phenotype associated with the dl338 virus is due to the absence of the E1b-55K protein. A virus only lacking the intact E1b-55K protein (pm381), was also constructed by Leppard et al. (1987). This virus displayed identical growth properties to the dl338 suggesting that the observed phenotype is due exclusively to the E1b-55K protein. The data of Montell et al. (1984) supported this conclusion, as they observed that a virus mutated in the 5' donor site of the E1B-17K mRNA also grew as wild-type. dl355 (Halbert et al., 1985) has a 14 base pair deletion in the E4 ORF6-34K reading frame, resulting potentially in the production of a 188 aa truncated protein but this was not detectable in infected cells (Cutt et al., 1987). Finally, dl367 (Cutt et al., 1987), contains the deletions present in both dl338 and dl355 in a single virus.

Virus stocks were grown on HeLa (dl309), HEK-293 (dl338, dl367), and W162 cells (dl355). HEK-293 cells are a human embryonic kidney cell line transformed with a DNA fragment carrying the left 11% of the Ad5 genome (Graham et al., 1977) and support the growth of E1b-deficient viruses such as dl338. The W162 cell line is Vero line carrying an integrated copy of the Ad5 E4 region (Weinberg and Ketner, 1983) and support the growth of E4 defective viruses such as dl355. Virus was quantified either by plaque assay on 293 cells or, as purified virus, by the measurement of optical density of lysed virions, at 260 nm, using a conversion factor of $1 A_{260} \text{ unit} = 10^{12}$

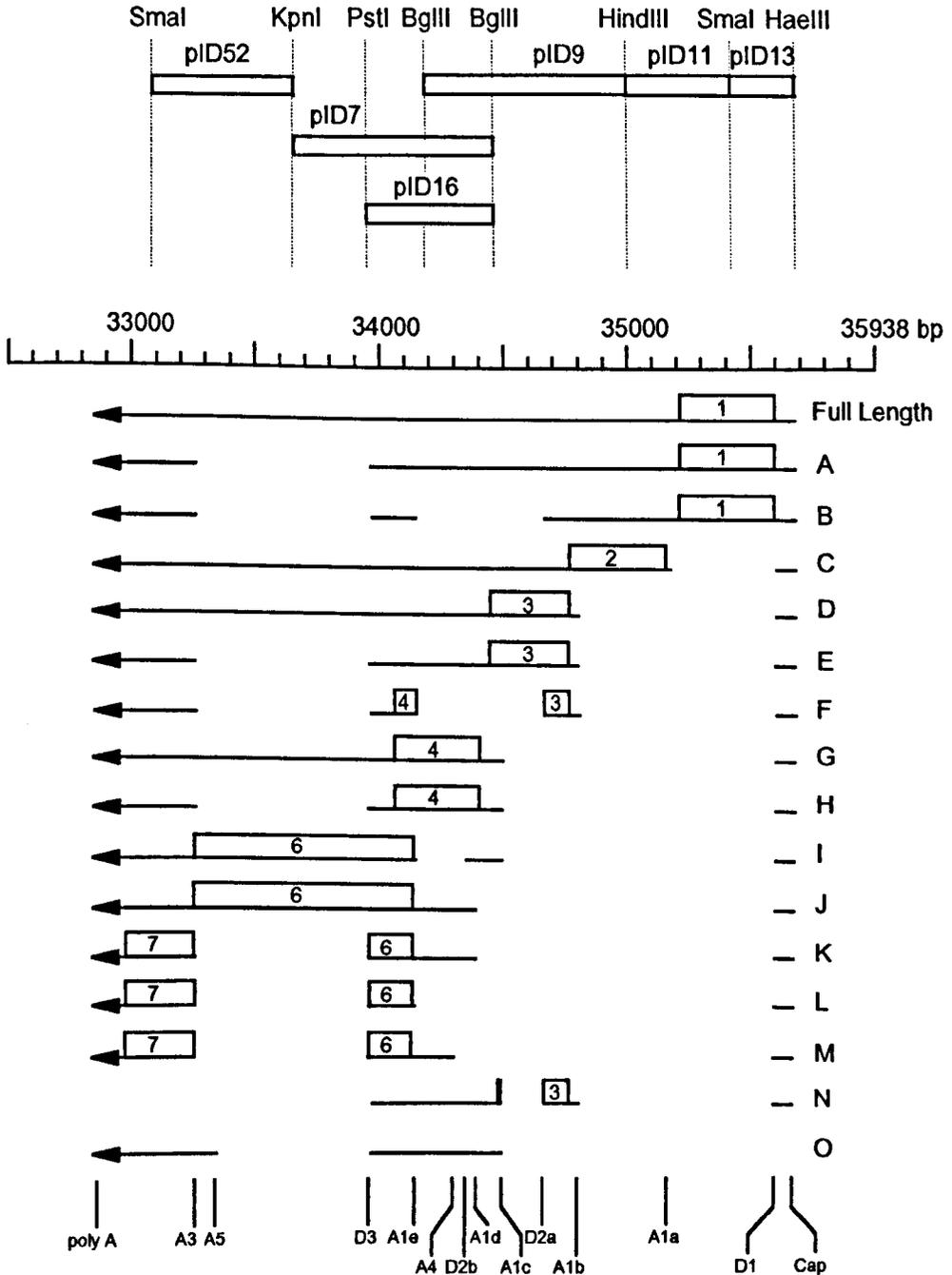
virions/ml. All infections were carried out using HeLa cells at a MOI of 10 pfu/cell or 500 virions/cell.

4.2.2. The antisense RNA probes (see appendix 3)

A series of antisense E4 RNA probes was constructed for use in quantitative RNase protection assays (Figure 4.1). These probes were designed to allow the identification and quantitation of each of the E4 mRNAs as described previously. They were constructed by cloning the relevant pieces of dl309 *KpnI*-G genomic fragment into either pGEM3 or pGEM4. pID7 was constructed by first cloning the *KpnI* (33594) - *BglII* (34115) fragment into the modified pUC19 MCS (the *PstI* site is substituted for a *BglII* site) of pEXP5 (a gift of K. Leppard) using the *KpnI* and *BglII* sites of the MCS. The *BglII* (34115) - *BglII* (34387) fragment was then inserted into this plasmid and the orientation checked to ensure the Ad5 sequence was contiguous across both fragments using *HinfI* digestion. The *KpnI* (33594) - *BglII* (34387) fragment was then moved into pGEM3 using *HindIII* and *EcoRI* sites of the MCS to create pID7. pID9 was constructed in a similar manner, beginning with the *BglII* (34387) - *HindIII* (34930) fragment. pID11 was constructed by subcloning the *HindIII* (34930) - *SmaI* (35354) fragment into pGEM4 between the *HindIII* and *SmaI* sites. pID13 was constructed by subcloning the *SmaI* (35354) - *HaeIII* (35616) fragment into the *SmaI* site of pGEM4. A clone was then selected which allowed the production of anti-sense E4 RNA when using the T7 promoter. pID16 was constructed by excising the *KpnI* (33594) - *PstI* (33875) fragment from pID7. pID52 was constructed by subcloning the *SmaI* (33091) - *KpnI* (33594) fragment into pGEM4 cut with *KpnI* and *HincII*. The two other probes used in the course of this study, an E1a probe (*PvuII* (455) - *PstI* (1838)) and a β -actin probe, were both gifts of K. Leppard.

All the above constructs were purified by large scale plasmid DNA preparations and then linearised so allowing the optimal production of antisense E4 RNA; pID7, -9, and -16 were linearised with *HindIII*, while pID11, -13, -52 were linearised with

Figure 4.1. The E4 transcription map showing the location of the RNA probes. The right terminal portion of the Ad5 genome (bp 32500 to 32598) is shown as a line scale. Below the line is the E4 transcription map inferred from studies of the closely related virus, Ad2. The positions on the genome map of the cloned fragments from which the RNA probes were derived are shown at the top of the diagram (pID7, -9, -11, etc.).



EcoRI. These linearised plasmids were then transcribed using either T7 polymerase or SP6 polymerase to produce radiolabelled, antisense RNA probes for use in RNase protection assays.

4.3. Results

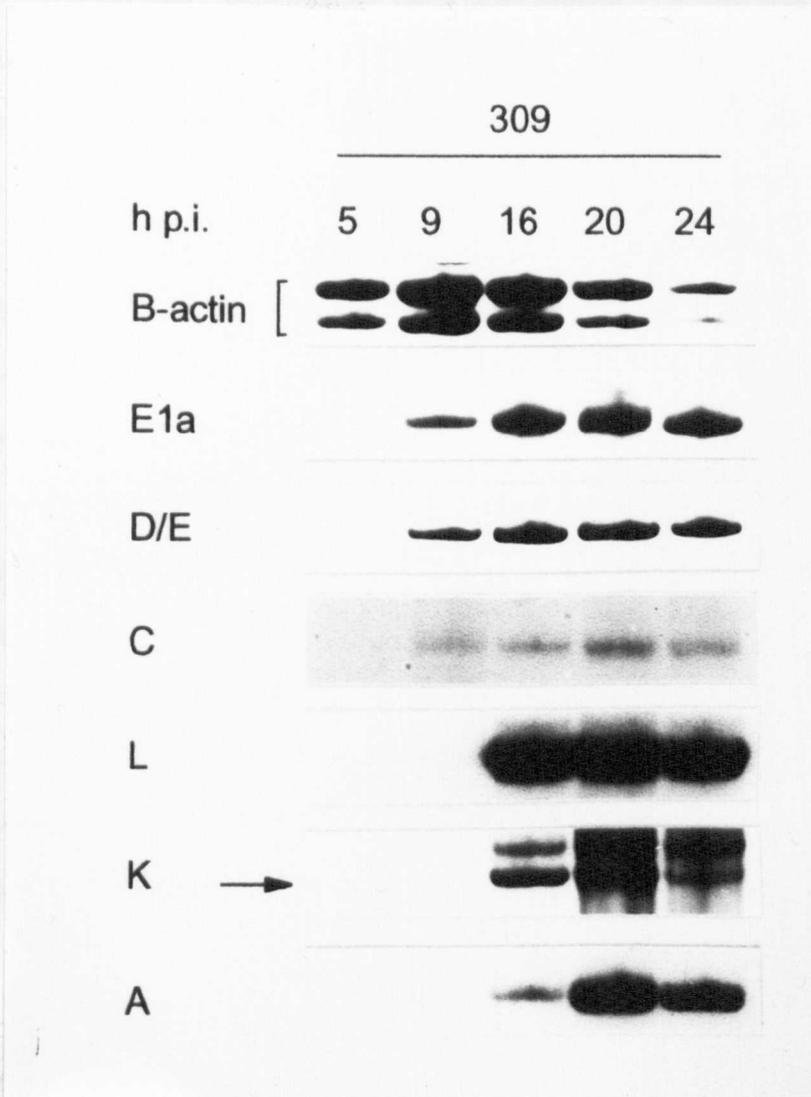
4.3.1. The temporal classes of E4 mRNA

To examine the temporal control of Ad5 E4 splicing, the cytoplasmic level of each E4 mRNA was determined at various times after infection with wild-type Ad5 (dl309). Each RNA probe used (see 4.2.2.) gave protected fragments of characteristic length, allowing the identification and quantitation of individual mRNAs, or in some cases of pairs of mRNAs predicted to encode the same translation product. The levels of E1a and β -actin mRNA were also quantified in parallel with the E4 mRNA to control for differences in either infection multiplicities or RNA inputs into the assays.

Results of representative assays are shown in Fig. 4.2. The probes used to identify mRNAs A, B and full length (pID11 and pID13) did not distinguish between these mRNAs. However, experiments conducted using pID9 showed no protection by the A1b to D2a fragment, characteristic of mRNAs B and N, suggesting that these mRNAs were absent in the infection. The pID11/13 signal is thus attributed to mRNA A and full length mRNA. Similarly, the pID9 probe did not distinguish between mRNAs D and E, nor between mRNAs G and H. However, from results with pID7, which protected a fragment from mRNAs A, E, and H, it could be deduced that both mRNAs E and H were significantly less abundant than mRNAs D and G, respectively, and were expressed with different kinetics (see below). The major component of the protected fragments D/E and G/H shown in Fig. 4.2 is therefore attributable to mRNAs D and G respectively.

The quantification of β -actin confirmed that the levels of RNA put into the experiment were comparable (the RNA was quantified using a spectrophotometer) as β -actin mRNA is an abundant cellular message. The decrease observed in the 24

Figure 4.2. Analysis of cytoplasmic levels of E1a, β -actin and E4 mRNAs during infection of HeLa cells by wild-type virus, dl309, by RNase protection assay (2.3.5.). HeLa cells were infected at an MOI of 10 and total cytoplasmic RNA was isolated at the time p.i. indicated above each lane. Five micrograms of RNA was probed for specific E4, E1a 3' common exon and β -actin E4 mRNAs, indicated to the left. A doublet was routinely isolated using the β -actin probe. Fragments shown (the predicted size from the Ad5 sequence is shown in brackets): E1a, 404 nucleotides; β -actin, (~532 nucleotides); D/E, pID9 (~382 nucleotides); C, pID11 (174 nucleotides); L, pID7 (178 nucleotides); K, pID7 (426 nucleotides, arrow indicates lower band); A, pID13 (258 nucleotides). It should be noted that any fragment indicated as mRNA A could also be due to protection of full length mRNA.



hour sample probably reflects the increase of MLTU messages in the cytoplasm so reducing the relative level of β -actin message.

Two distinct classes of E4 mRNAs were identified by this analysis, early and late. mRNAs belonging to the early class (mRNAs C, D, G and J (C and D are shown in Fig. 4.2)) were first detected in the cytoplasm 9 hours post-infection with levels remaining almost constant throughout the infection. The low levels of E1a mRNA detected at 5 h p.i. in this experiment (this is not observable in figure 4.2. due to the poor reproduction of the photograph) may explain why little or no E4 mRNA of any type was detected at this time point as the E1a products are involved in the activation of the E4 promoter (see section 1.4.1.). The small increase in the level of mRNAs D and E that was observed between 9 and 16 h p.i. is consistent with estimates of mRNA E synthesis from pID7 experiments (data not shown). No increase in the levels of mRNAs G and H was observed between 9 and 16 h p.i. possibly indicating that mRNA H is not present (or present at low levels) late in the infection (data not shown). mRNAs of the late class (A, E/H, K, L) were first detectable at significant levels at 16 h p.i.. However, mRNA A/full length mRNA was only observed in low amounts at 16h p.i., with its levels not peaking until 20 h p.i. while the other mRNAs with late kinetics (i.e. mRNA K and L in Fig. 4.2) reached maximum, or near-maximum, levels at 16 hours p.i..

mRNA-probe duplexes covering the region around position 34250 consistently showed sensitivity to RNase that resulted in 50 to 80% cleavage of the probe. Fragment lengths were in each case consistent with an mRNA-probe RNA discontinuity at this position, although no splice donor or acceptor sequence was present. Inspection of the sequence revealed a stretch of 11 A residues in the E4 mRNA. Stuttering of RNA polymerase on this sequence, either in the cell or, in the antisense during probe synthesis in vitro, could explain this observation.

4.3.2. The D3 to A3 intron is excised late in the infection.

The expression pattern of E4 mRNA, reveals two distinct temporal classes. Correlation of these data with the transcription map reveals that members of the early class all retain sequences distal to the between splice site D3 in the 3' half of the E4 unit, while this donor site is utilised late in the infection, presumptively in combination with the A3 or A5 acceptor sites, to produce a family of shorter mRNAs. Using pID52, these acceptor sites were found to be utilised at 24 h p.i. but not 9 h p.i., confirming that the D3-A3 intron is only excised late in the infection and indicating that the D3-A5 intron is also only excised late in the infection (Fig. 4.3.). The efficiency of usage of the A3 splice site is 40 fold greater than the A5 site. There is no evidence for the use of other alternative splice sites proposed by Tigges and Raskas (1984) to produce the family of 0.8kb mRNAs encoding the series of ORF6/ORF7 fusions. The only splice sites utilised in this region, in these experiments, were the D3 donor site (Freyer et al., 1984; Tigges and Raskas, 1984; Virtanen et al., 1984) and the A3 and A5 acceptor sites (Freyer et al., 1984; Tigges and Raskas, 1984; Virtanen et al., 1984).

4.3.3. Relative abundance of E4 mRNAs

The accurate determination of the relative amounts of the different E4 mRNAs is complicated by the need to assume that different probe RNAs were labelled to the same specific activity. This assumption is reasonable only for probes prepared in parallel from the same mixture of α -³²P-labelled and unlabelled UTP. For such probe sets, protected fragments were quantified using laser scanning densitometry and the data corrected to take into account the number of uridine residues in the protected fragment to give a measure of relative mRNA levels. A full data set was obtained by linking the results for groups of mRNAs having one or more members in common. These estimates of relative E4 mRNA levels are shown in Table 4.1. with mRNA L being the most abundant RNA species detected. mRNAs E and H have not been identified individually; however neither is found in significant amounts at 9 h p.i., and

Figure 4.3. Analysis of cytoplasmic levels of E4 mRNAs in HeLa cells infected by wild type dl309 or E1b-55K mutant dl338 at the times p.i. indicated. 5 μ g of cytoplasmic RNA was probed for the use of the A3 and A5 by RNase protection analysis (2.3.5.3.) using pID52. The predicted protected fragments are : 503 nt - unspliced at A3 or A5; A5 (189nt) - spliced at A5; A3 (99nt) - spliced at A3. The 547nt fragment is probably due to protection of undigested template DNA, including the pGEM sequences.

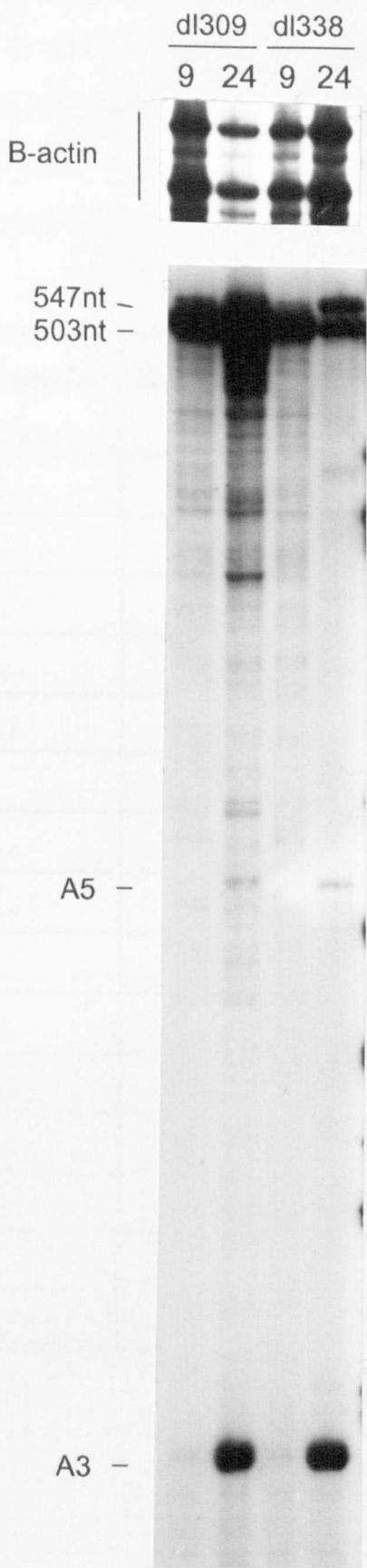


Table 4.1. Classification of E4 mRNAs. mRNAs were quantified using laser scanning densitometry and the data corrected to take into account the number of uridine residues in each protected fragment. The semi-quantitative amounts displayed are all relative to the levels of mRNA L at late times. * = The amounts shown for mRNAs E and H are shown collectively. See appendix 4.

E4 RNA	Relative amount		Kinetic Class
	9 h p.i.	20 h p.i.	
A/full length	-	++	Late
B	-	-	
C	+	+	Early
D	+++	+++	Early
E*	-	++	Late
F	-	-	
G	++	++	Early
H*	-	++	Late
I	-	-	
J	+	+	Early
K	-	++	Late
L	(+)	++++	Late
M	-	+	Late
N	-	-	
O	-	+	Late

their collective abundance is less than that of mRNA D. mRNAs B, F, I, and N were not detected in any experiment.

4.3.4. Late E4 mRNAs are dependent on DNA replication for accumulation.

To determine whether expression of the late class of E4 mRNAs was dependent on viral DNA synthesis, their levels in cells infected in the presence of the DNA synthesis inhibitor hydroxyurea (10mM) were assayed. As shown in Fig. 4.4., expression of mRNA D (early class) in a wild-type infection was unaffected by the inhibitor, whereas the accumulation of the late E4 mRNAs L and A/full length mRNA was largely and completely prevented, respectively (the low level of mRNA L at 24 h p.i. in the presence of HU is not visible due to the poor reproduction of the photo in Fig. 4.4.). Thus, these late E4 mRNAs, like other viral late mRNAs, depend on DNA replication for their expression. The dependence of the expression of the minor late mRNAs (i.e. mRNAs E/H, K and M) was not determined, however their expression pattern was identical to that of mRNA L, in the absence of any DNA replication inhibitors (see Fig. 4.2.), possibly suggesting that these mRNAs are also dependent on DNA replication.

4.3.5. Dependence of late E4 mRNA expression on the E1b-55K protein.

As mentioned in section 1.3., cytoplasmic mRNA accumulation for genes expressed from the Ad5 genome at late times is dependent on a function provided by the E1b-55K protein. The possible dependence of the late class of E4 mRNAs, as defined above, on this E1b function was therefore examined using a mutant virus, dl338, which is unable to express the E1b-55K protein.

The E4 mRNAs present during a dl338 infection were quantified in a comparative analysis with wild-type virus. Representative analyses are shown in Fig. 4.5a. with quantitation of these data shown in Fig. 4.5b. The E1a mRNA levels were comparable in the two infections demonstrating that the multiplicity of infection of the two viruses were similar. As expected, the early class of E4 mRNA (represented

Figure 4.4. Analysis of late E4 mRNA expression in the presence and absence of an inhibitor of DNA replication. HeLa cells infected by wild-type virus, dl309 at the times p.i. indicated, in either the absence (-) or presence (+) of hydroxyurea, added at 10 mM to the culture medium. 5 μ g of unselected, cytoplasmic RNA was probed for specific mRNAs indicated at the left; other details as Fig. 4.2.

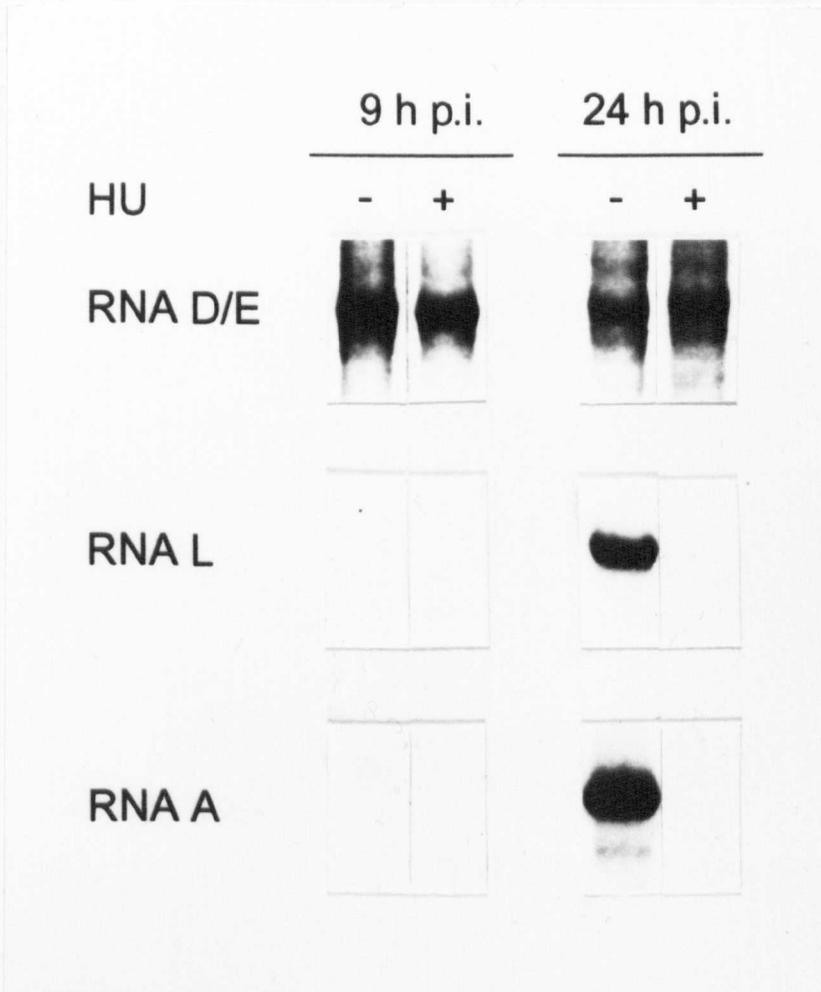


Figure 4.5b. Quantitation of data shown in Fig. 4.5a. Results of laser scanning densitometry (2.3.12., 2.3.13) are shown in arbitrary units plotted against time p.i. for dl309 (◆) and dl338 (■). These data are not corrected for protected fragment length and specific activity differences; therefore, amounts are not comparable between panels.

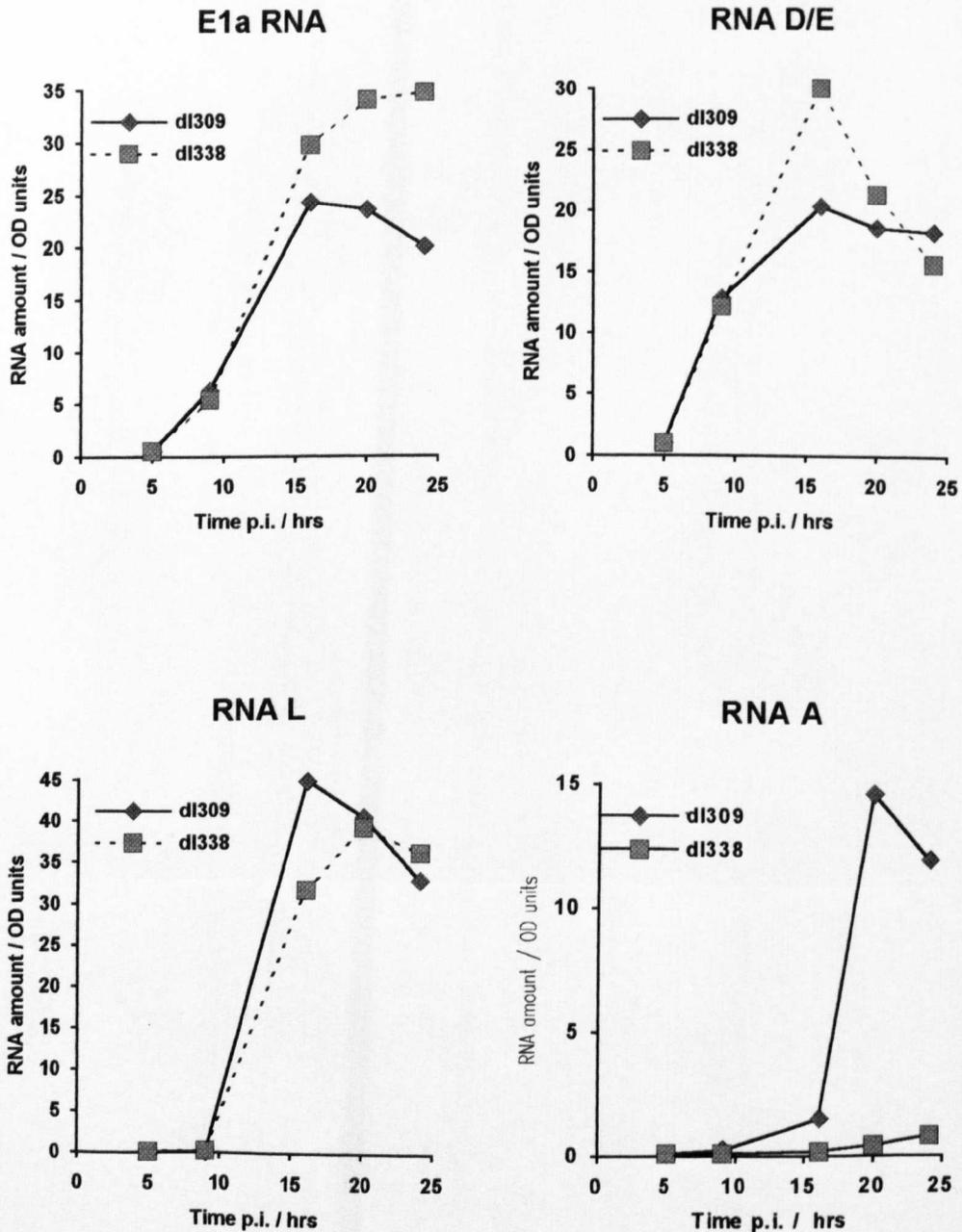
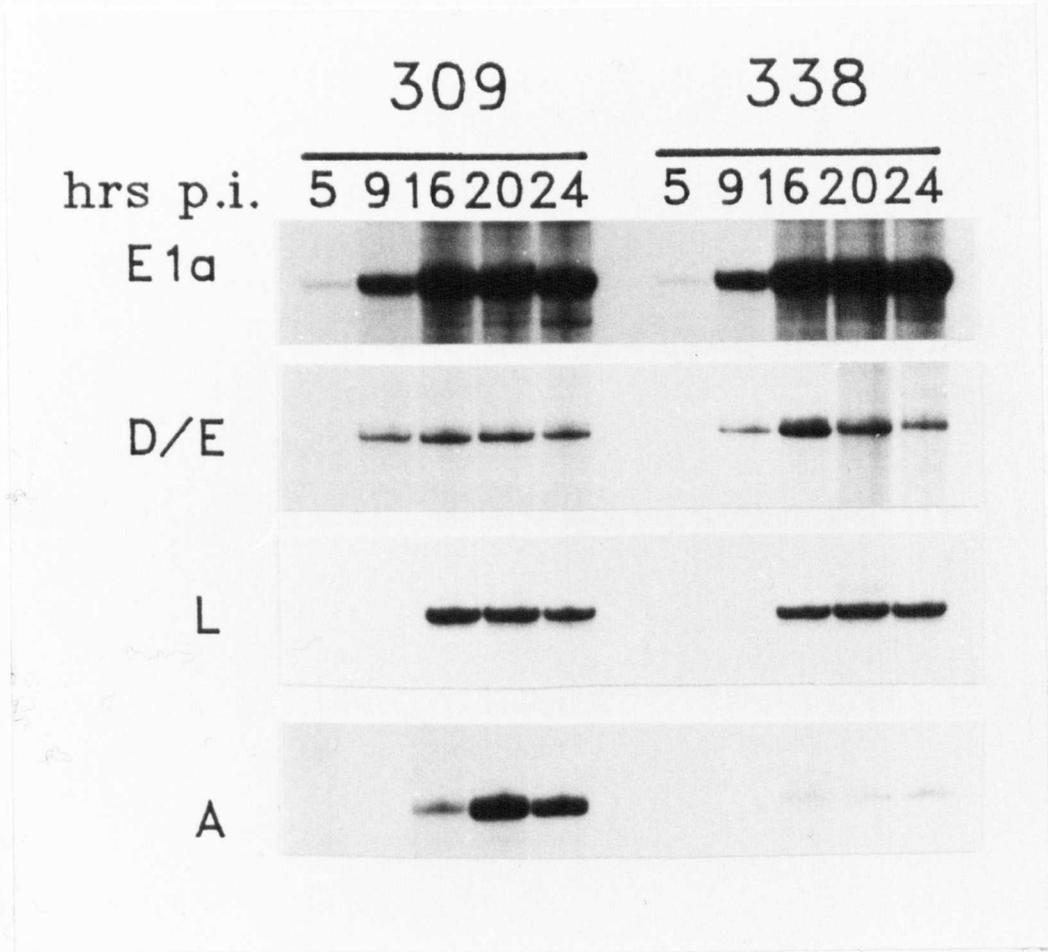


Figure 4.5a. Analysis of cytoplasmic levels of E1a and E4 mRNAs during infection of HeLa cells by wild-type virus dl309 or E1b-55K mutant virus dl338, as indicated. Details are as for Fig. 4.2.



by mRNA D) accumulated in the cytoplasm of wild-type- and dl338-infected cells with identical kinetics. The peak observed for this mRNA in dl338 infection is an artefact of the experiment as it was not reproducible. More surprisingly, most of the late class of E4 mRNAs (represented by mRNA L) displayed only minimal dependence of E1b-55K protein function. The one exception was mRNA A/full length mRNA, which had shown a unique delayed late kinetic profile in a wild-type infection (Fig. 4.2.). Accumulation of this mRNA was found to be strongly dependent on E1b function, with cytoplasmic levels in dl338-infected cells only 5 to 10% of the wild-type level at 20 to 24 h p.i.

It is also of note that the levels of mRNAs utilising the A3 site (and A5 site) are comparable in the cytoplasm of wild-type- and dl338- infected cells at 24 h p.i. (Fig. 4.3.). One would have expected that if mRNA A is dependent on the E1b-55K protein then the level of the A3-*Sma*I fragment would be lower in the dl338 fraction. It may be that the other mRNAs (E,H,K,L and M) dilute the difference so making it impossible to observe in the infections using this probe. However there is a distinct difference in the level of full length protected mRNA (using pID52) between the wild-type-infected 24 h p.i. RNA sample and the dl338-infected 24h p.i. sample (Fig. 4.3) suggesting that the full length mRNA exists and is dependent on the E1b-55K mRNA for optimal cytoplasmic accumulation. However the existence of the full length mRNA cannot however be proved conclusively using the probes shown in Fig. 4.1., only inferred, as no one probe protects full length mRNA in such a way to yield a fragment of a unique size. The additional 547 nt band present in all the lanes, especially in the dl338/24 hr p.i. lane, can only be the complete probe with the pGEM4 MCS sequences intact. This cannot be protected by the viral mRNA and is probably an artefact of the probe, possibly due to secondary structure, although it is unclear why it is not even in each sample.

4.3.6. E4 mRNA A/full length mRNA is dependent on E4 ORF6-34K for optimal cytoplasmic accumulation.

As mentioned in section 1.3., the E1b-55K protein exists as a molecular complex with the E4 ORF6-34K protein in infected cells. In previous studies, virus dl355 (lacking E4 ORF6) and dl367 (lacking E4 ORF6 and E1b-55K) showed phenotypes very similar to that of the E1b-55K mutant dl338 in respect to the cytoplasmic accumulation of late viral mRNA (Halbert et al., 1985; Cutt et al., 1987). To confirm that the effect of the E1b mutation on the levels of E4 mRNA A/full length mRNA was due directly to the absence of the E1b mRNA transport regulatory function, the levels of mRNA A/full length mRNA in E4 ORF6 mutant infections were examined (Fig. 4.6.). Cytoplasmic levels of mRNA D (early class) were unaffected at 9 h p.i. in these infections and only marginally at 24 h p.i.; mRNA L showed a similar pattern at 24 h p.i. These marginal reductions at late times may be due to the slight reduction in the rate of DNA replication previously noted for these mutants (Halbert et al., 1985; Cutt et al., 1987). In contrast, cytoplasmic levels of mRNA A/full length mRNA were severely depressed in both E4 mutant infections, as they were in dl338 infections.

4.3.7. E4 mRNA A/full length mRNA dependence on E1b-55K is at a post processing level.

Since all E4 mRNAs derive from the same primary transcript, the selective depression of cytoplasmic mRNA A/full length mRNA levels seen in the dl338 infection must be due to a post-transcriptional effect. mRNA A/full length mRNA results from the non-usage of the E4 5' splice donor site D1 (Fig 4.1.). To determine whether the absence of E1b-55K protein resulted in increased usage of this site, the relative abundance of RNA spliced and unspliced at this site in nuclear RNA was assessed for cells infected with either wild-type or dl338 virus. The results of this analysis are shown in Fig. 4.7. Levels of a control E4 mRNA (mRNA L) in both nucleus and cytoplasm were unaffected by the absence of E1b-55K protein. Levels of mRNA A/full length mRNA in the cytoplasm were, as before, greatly reduced in the mutant infection.

Figure 4.6. Analysis of cytoplasmic levels of E4 mRNAs in HeLa cells infected by wild-type dl309, E4 ORF6 mutant dl355, or E1b-55K/E4 ORF6 mutant dl367 at the times p.i. indicated. Virus stocks were CsCl gradient-purified particle preparations used at 500 particles per cell. 5 μ g of unselected RNA was probed for specific mRNAs indicated at the left; other details are as for Fig. 4.2.

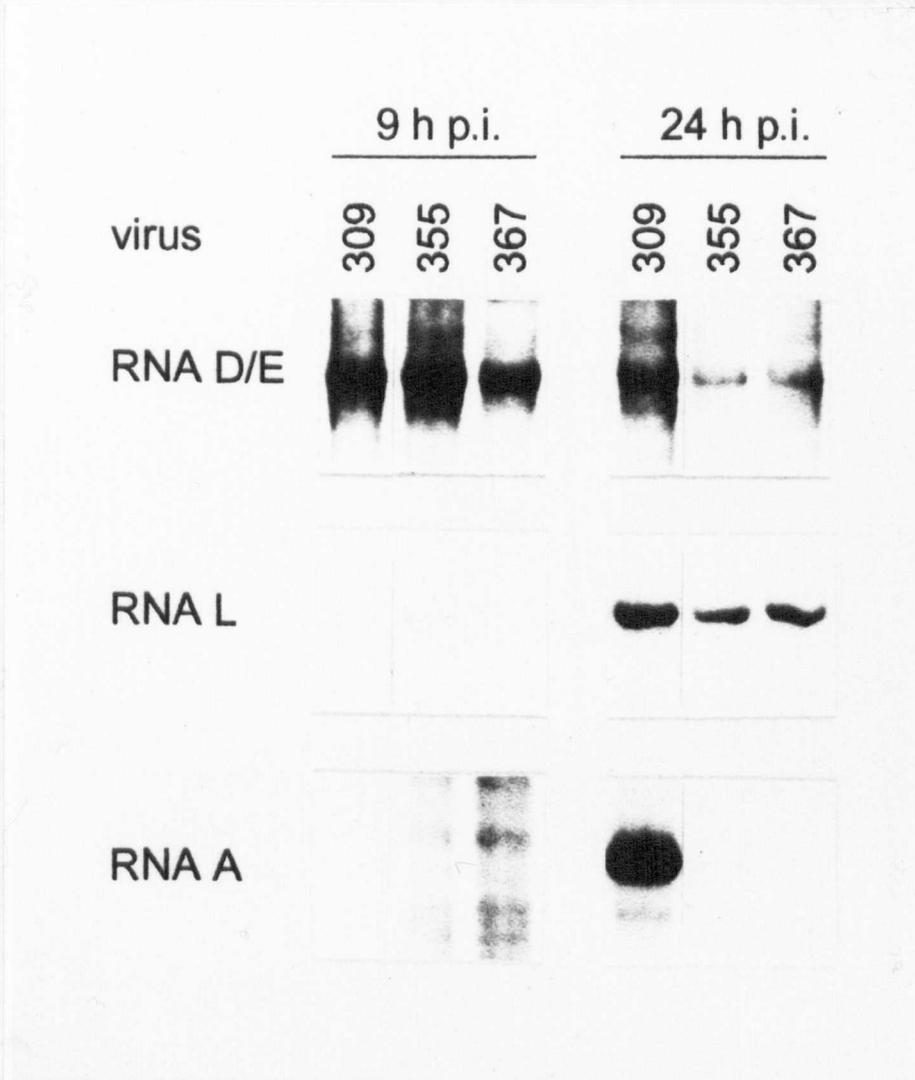
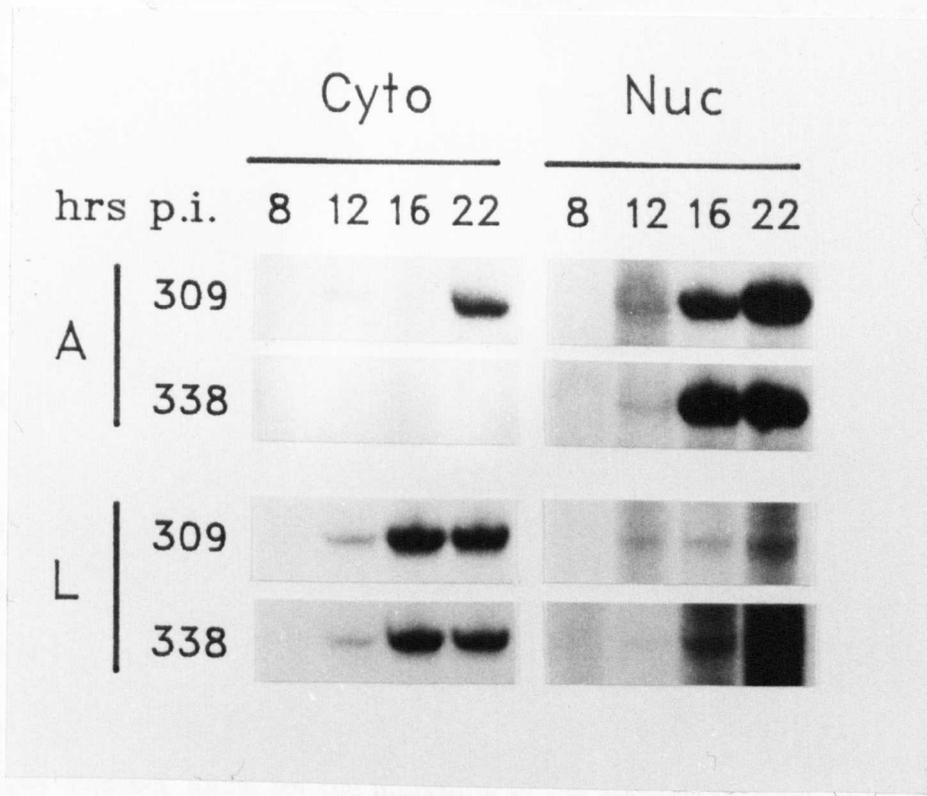


Figure 4.7. Analysis of cytoplasmic and nuclear levels of E4 mRNAs during infection of HeLa cells by wild-type dl309 or E1b-55K mutant virus dl338. Total cytoplasmic (Cyto) and total nuclear (Nuc) RNAs were isolated at the times p.i. indicated, and 5 μ g samples analysed. The virus used and E4 RNA detected are indicated at the left. Other details are as for Fig.4.2.



However, no similar reduction in nuclear levels of mRNA A/full length mRNA was seen. This result shows that, as for other mRNAs which depend on E1b-55K protein for accumulation, E4 mRNA A/full length mRNA requires this function for its efficient movement out of the nucleus or for cytoplasmic stabilisation rather than to modulate its splicing. No probe fragment protected by mRNAs spliced at site D1 could be detected in either wild-type- or dl338-infected cell nuclear RNA. Also the cytoplasmic/nuclear ratio for mRNA L was very much greater than that for mRNA A/full length mRNA. These findings suggest that, once spliced at the D1 site, E4 mRNA is rapidly exported to the cytoplasm. A similar conclusion was reached previously regarding E2 mRNA expression (Leppard and Shenk, 1989).

4.4. Discussion.

4.4.1. Classes of E4 mRNA

The various Ad5 E4 mRNAs have been divided into two temporal classes based on a detailed analysis of RNA levels over the time course of a wild-type infection. mRNAs C, D, G, and J form an early class, while mRNAs A/full length, E, H, K, L, M, and O mRNA form a late class, dependent on viral DNA replication for cytoplasmic expression. Previously described mRNAs B, F, I, and N could not be detected, nor was any evidence found for the utilisation of the splicing sites for the 0.8kb mRNA family described by Tigges and Raskas (1984). Within the late class, mRNA A and/or full length mRNA are unique, both in the late time p.i. at which their accumulation peaks and in their dependence on the E1b-55K and E4 ORF6-34K proteins for optimal cytoplasmic accumulation.

When members of these E4 classes are compared with one another, structural relationships can be identified. All the mRNAs expressed at early times retain the D3 to A3/A5 intron suggesting that the usage of these sites is temporally regulated during the course of the Ad5 infection. The regulation may be due to the action of the E2a DBP as described by Ross and Ziff (1992). Like other examples of regulated

adenovirus splicing, such as in the expression of the E1a, E1b, and L1 genes (Spector et al., 1978; Esche et al., 1980; Lewis and Mathews, 1980; Akusjarvi and Persson, 1981; Montell et al., 1984), processing in E4 moves toward more heavily spliced, shorter mRNAs in the late phase of the infection. These results are in agreement with the observations of Chow et al. (1979) and Tigges and Raskas (1984) who also observed the predominance of shorter mRNA late in the infection. The exception is the full length mRNA which possibly appears very late in the infection, with kinetics distinct from the other late mRNAs (apart from possibly mRNA A), which is not spliced (Fig. 4.2. and Fig. 4.3.).

The undetected mRNAs B, F, I, and N are also related, being formed by the use of donor sites D2a or D2b in conjunction with an acceptor site. The lack of usage of these sites may be sequence related, as they are more divergent from the consensus sequence (Mount, 1982) than are sites D1 and D3, where splicing was observed (Fig. 4.8.). Given the detection of mRNAs spliced at these sites by other workers, their absence in these experiments may indicate that usage of D2a and D2b is regulated and apparent only under certain circumstances, such as in the presence of translation inhibitors, or only in the Ad2 infection. Similar analysis of the predicted splice sites of the 0.8kb family of mRNA found by Tigges and Raskas (1984) was prevented by the imprecise location of the splice sites given in the paper.

Why the late E4 mRNAs are dependent on DNA replication for expression is not clear. As mentioned in section 1.2.4., the shift in the splicing of other adenovirus transcripts late in the infection has been attributed to a number of factors. Ross and Ziff (1992) examined E4 expression in abortive CV-1 infections where DNA replication is delayed and E4 and MLTU RNA processing is defective (see 1.4.1.). They did not attribute the failure of the early to late shift in E4 mRNA splicing simply to the absence of DNA replication as they noted that at 50 h p.i., when DNA replication was near normal, the early E4 splicing pattern was retained. Instead they attributed the defect in E4 RNA processing in abortive infections, to the absence of a functional DBP since a missense mutation in this protein could alleviate the splicing defect. The failure to produce late

Figure 4.8. The E4 donor splice sites as compared to the eukaryotic consensus sequence of Mount (1982). The percentages given are the percent occurrence of the nucleotide shown at that site. The slash indicates the position of RNA cleavage. The bold nucleotides of the E4 donor sequences (D1, D2a, D2b, D3) indicate bases present which are different to those found in the consensus sequence

<u>Donor Sequence</u>	
Percentage	$\begin{array}{ccccccc} & 43 & & & 62 & & \\ 4 & 3 & 7 & 100 & 2 & 8 & 3 \\ & 3 & 7 & 1 & 2 & 8 & 3 \end{array}$
Consensus	$\begin{array}{ccccccc} & C & & & A & & \\ & A & G & / & G & T & A & A & G & T \\ & A & & & G & & & & & \end{array}$
<u>Site</u>	
D1	C T G/GTA A AG G
D2a	AAG/GT G C G
D2b	T AG/GT T G C
D3	A A C/GTGAGT

E4 mRNAs in infections of HeLa cells in the absence of DNA replication may also be due to the effect of the DBP; late in the infection, after DNA replication, the E2-L promoter is active resulting in the production of high levels of DBP, the absence of this increased level of DBP may restrict E4 expression to the early pattern. So, while it would appear that the microenvironment of the DNA template is not directly involved in the late shift in E4 splicing, it may be involved indirectly, regulating the activation of the E2-L promoter and so production of the DBP.

4.4.2. The expression of the E4 ORF products

The observed temporal pattern of mRNA expression from the E4 region suggests that ORF2, ORF3, ORF4, and ORF6 products are synthesised initially during the early phase of infection, while the products of ORF1 and ORF6/7 are synthesised later in the infection.

The ORF1-14K protein, involved in the formation of mammary tumours of rats in Ad9 (see section 1.4.2.2.), is encoded by mRNA A and full length mRNA, which are possibly expressed with late kinetics (it is unclear if both or just one of these mRNAs is expressed). The role of the ORF1 product in the Ad5 infection is unknown (see chapters 3 and 6) but the data from these studies suggests that it is required late in the infection. These results are in agreement with the observations of Javier (1994) who found Ad9 E4 ORF1-14K in infected A549 cells (a human lung carcinoma cell line permissive for viral replication) 20 h p.i. No function has yet been assigned to the ORF2-14K protein although this data would suggest that it is expressed early in the infection, as mRNA C (encoding ORF2) is present early in the infection.

The early expression of ORF3 and ORF6 mRNA is in agreement with the role of their products in facilitating viral gene expression from the onset of the late phase (see section 1.4.2.3.) (data for mRNA D (ORF3) can be seen in Fig. 4.2.; the data for mRNA J (ORF6) is not shown due to difficulty with photography, however mRNA J was first seen in the cytoplasm at 9 h p.i. and remained at a constant level until at least 24 h p.i.). It is not clear how translation of the ORF6 from mRNA J occurs (mRNA I

was not found in these experiments) as three AUG codons exist upstream of the initial AUG of ORF6 (although Virtanen et al. (1984) noted that only one of these upstream, potential translational start sites is favourable according to the rules of Kozak (1981)). The obvious mRNA to encode the ORF6 protein is an mRNA lacking the D1-A1e intron but retaining the potential D3-A3 intron, but this mRNA has never been identified, a result confirmed in this study. The early expression of the E4 ORF4 mRNA is also in agreement with previous data as the E4 ORF4-13K protein was first observed to mediate the down-regulation of E1a phosphorylation in HeLa cells early in the infection (Muller et al., 1992).

The absence of mRNAs F and N in these studies suggests that the ORF3/4-7K and the ORF3/12aa fusion products do not exist in the infected cell, in the absence of cycloheximide. Similarly no evidence was found for the use of potential splice sites between ORFs1 and 2 or a potential donor site within ORF2 as predicted by Hérissé et al. (1981) confirming the results of Freyer et al. (1984), Tigges and Raskas (1984) and Virtanen et al. (1984).

The late expression of the ORF6/7 mRNA is surprising, given its role in the activation of the E2 early promoter (see section 1.4.2.6.). Hardy et al. (1989) observed that the formation of an infection-specific, E4-dependent complex on an E2 promoter fragment first occurred in HeLa cells at 6 h p.i. This complex has been shown subsequently to contain the E4 ORF6/7 protein (Huang and Hearing, 1989b; Reichel et al., 1989; Marton et al., 1990; Neill et al., 1990; Raychaudhuri et al., 1990). However, amounts of this complex increased dramatically between 6 and 12 h p.i. in these experiments. Given the differences in experimental protocol used (multiplicity of infection and cell type), this delayed-early increase in activity observed by Hardy et al. (1989), now attributed to the E4 ORF6/7, is in reasonable agreement the expression of ORF6/7 mRNA. Similar experiments by Marton et al. (1990) are harder to rationalise with the kinetics of the ORF6/7 mRNA expression in this study as they observed that E4 ORF6/7 dependent E2F complexes formed on E2 promoter fragments in nuclear fractions from 3 h p.i. while extracts prepared late in the infection (18h and 35h) failed

to produce infection-specific complexes. All the ORF6/7 mRNAs (K, L, and M) displayed late kinetics, first being observed 16 h p.i. suggesting that the protein does not occur until late in the infection.

4.4.3. Regulation of the expression of mRNA A and/or full length mRNA

Of all the E4 mRNAs detected in this study, only mRNA A and/or full length mRNA are dependent on the E1b-55K/E4 ORF6-34K complex for optimal cytoplasmic accumulation. Additionally, these two mRNAs are expressed with later kinetics than the other mRNAs, reaching maximal cytoplasmic levels at 20 h p.i. Among the E4 mRNAs detected in this study, these mRNAs are the only mRNAs to retain splice donor site D1 unused. Furthermore, these mRNAs also retain a number of acceptor sites (A1a to A1e) that can be used, with D1, to form other detectable E4 mRNAs. Thus, E4 mRNA A and full length mRNA are incompletely spliced mRNAs which contain a readily removable intron. Although other E4 mRNAs are similarly incompletely spliced, through retention of either the D2a/A1e, D2b/A1e or D3/A3 intron, these mRNAs either are synthesised in the early phase of infection or else carry introns whose removal was not detected in our experiments. In a previous study, strong dependence on E1b function for accumulation of late viral mRNA was shown to correlate with the presence in the mRNA of unused splice acceptor sites and/or intron sequences (Leppard, 1993). It was argued that this result was due to viral mRNAs, in the absence of the E1b-55K protein, being held on the nuclear matrix by host cell mechanisms whose function was to prevent the efflux of immature RNA from the nucleus. Therefore the presence of such sequences in mRNA A and full length mRNA may explain the unique dependence of these RNAs, among the E4 population, on the E1b-55K protein for optimal cytoplasmic accumulation.

Alternatively, the accumulation of mRNA A and/or full length mRNA in the cytoplasm of wild-type infected cells may be a secondary effect of the E1b-55K/E4 ORF6-34K protein complex. This would explain the late kinetics observed for mRNA A and/or full length mRNA. Fig. 4.7. demonstrates that mRNA A/full length mRNA is present

at a significant level in the nucleus of the wild-type infected cell at 16 h p.i. but is not observed in the cytoplasm at this time point implying it is restricted to the nucleus. However other studies have shown that E1b is active in the export of MLTU messages at 16 h p.i. (Pilder et al., 1986a,b); E4 mRNA A/full length mRNA does not accumulate in the cytoplasm until 22 h p.i. It is conceivable that partially and/or unspliced E4 mRNA, which is normally retained in the nucleus, is released into the cytoplasm in a wild-type infection because of the general restructuring of the viral nucleus observed late in the infection (Zhonghe et al., 1986) due to the activities of the late viral proteins encoded by the MLTU which can only be expressed in the presence of the E1b/E4 complex and not because of the direct action of E1b and E4 proteins.

Chapter 5.

The construction and analysis of a virus containing a mutated E4 D1 donor site.

5.1. Introduction

The work described in chapter 4 demonstrates that of all the E4 mRNAs only mRNA A and/or full length mRNA is dependent on the E1b-55K/E4 ORF6-34K complex for optimal cytoplasmic accumulation. These data, combined with the results of Leppard (1993), led to the hypothesis that these mRNAs displayed a strong dependence on the E1b-55K/E4 ORF6-34K complex because of the presence of intact splice sites/introns within their sequence (see section 1.3.5.). Intact introns present in an RNA molecule have been shown to act as nuclear retention signals, blocking efficient RNA export (Ciejek, 1982; Legrain and Rosbash, 1989; Chang and Sharp, 1989; Hamm and Mattaj, 1990). If this hypothesis is correct the action of the E1b-55K/E4 ORF6-34K complex is similar to that of the HIV1 Rev protein in overcoming the retention of incompletely spliced mRNAs in the nucleus of infected cells.

As mentioned, mRNA A and full length mRNA differ from the other late mRNAs of the E4 region in that the D1-A1a/A1b/A1c/A1d/A1e intron is intact in these mRNAs. As late mRNAs identical to mRNA A but lacking this intron, are independent of the E1b-55K/E4 ORF6-34K complex, it is possible that these splice sites or intronic sequences are the reason for the dependence of mRNA A and/or full length mRNA on the complex for optimal cytoplasmic accumulation. However since mRNA A and full length mRNA have not been distinguished, a role for the D3/A3 intron, removed from all late E4 RNAs except 'full length' cannot be excluded. As an initial step in evaluating these possibilities, the role of the major donor site, D1, in conferring E1b-55K/E4 ORF6-34K dependence was explored: According to this hypothesis E4 transcripts lacking the D1 site will not be so tightly associated with the nuclear matrix and/or splicing machinery and therefore not so dependent on the E1b-55K/E4 ORF6-34K complex for release into the cytoplasm.

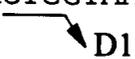
5.2. Experimental strategy

In order to explore the role of the D1 donor site in conferring E1b-55K/E4 ORF6-34K dependence on an mRNA, a virus was built which lacked the D1 donor site. If this site were important then the E4 mRNA A and/or full length mRNA, containing the mutated D1 site, would be able to accumulate in the cytoplasm in the absence of the E1b-55K and/or the E4 ORF6-34K proteins. If it were unimportant then mutant mRNA A/full length mRNA would fail to accumulate in the absence of these proteins. To achieve a comparison of mRNA A/full length mRNA accumulation from the D1 mutant in the presence and absence of RNA transport functions, D1 was mutated in an RNA transport⁻ background, and E1b-55K/E4 ORF6 functions supplied in *trans* from a co-infected, wild-type virus (or as a control, an E4 ORF6⁻ virus). Since the deletion of the D1 site would probably disrupt the expression of the whole E4 region, it was decided that it would be easier to build the D1 mutation in the context of an E4 ORF6 lesion, rather than an E1b-55K lesion. An alternative strategy was the construction of two mutant viruses, both with mutated D1 sites but with only one lacking the E4 ORF6-34K protein. However, this strategy was not pursued as a number of attempts to construct a virus lacking just the D1 site failed.

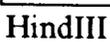
The strategy behind the analysis of the mRNAs produced from this mutated E4 region involved the co-infection of HeLa cells with the mutant virus and either wild-type (so providing the E4 ORF6-34K protein in *trans*) or dl355 (lacking the E4 ORF6-34K protein), the purification of the cytoplasmic RNA and analysis of the RNA by RNase protection. The antisense RNA probe, to be used in the RNase protection assays, was designed so that it could distinguish between mRNA A/full length mRNA from the D1 mutated virus and the D1 wild-type virus, being complementary to the mutant mRNA from the *Sma*I site (35354) to the *Hind*III site, created in the mutagenesis of the D1 site, at 35537 bp (see Fig. 5.1.). Wild-type D1 mRNA A/full length mRNA would therefore protect a fragment of 175 nt while the D1 mutant mRNA A/full length mRNA would protect a fragment of 182 nt.

Figure 5.1. The strategy for the detection of the mRNA A derived from the D1 and E4 ORF6 mutated virus vD1. In vD1 the E4 D1 splice site is mutated to a *Hind*III site. Therefore, using a probe transcribed from pID50, E4 mRNA derived from vD1 will protect a fragment 7 nucleotides larger than mRNA from dl309 (wild-type).

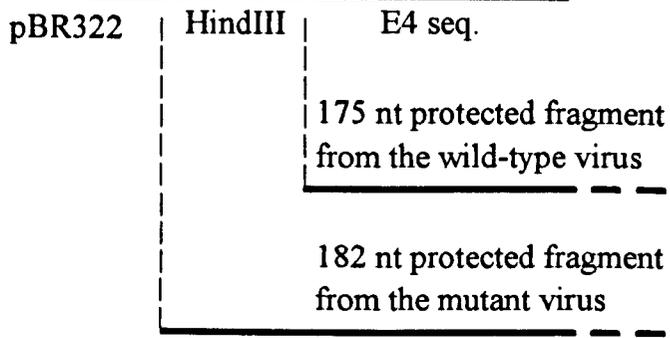
dl309 RNA

...CTTTTTTACTGGTAAGGCTGACTGTTATG...


mutant vD1 RNA

...CTTTTTTACTAAGCTTGCTGACTGTTATG...


pID50 anti-sense RNA probe

...AGTAGCGCTATTCGAAACGACTGACAATAC...


pBR322

HindIII

E4 seq.

175 nt protected fragment
from the wild-type virus

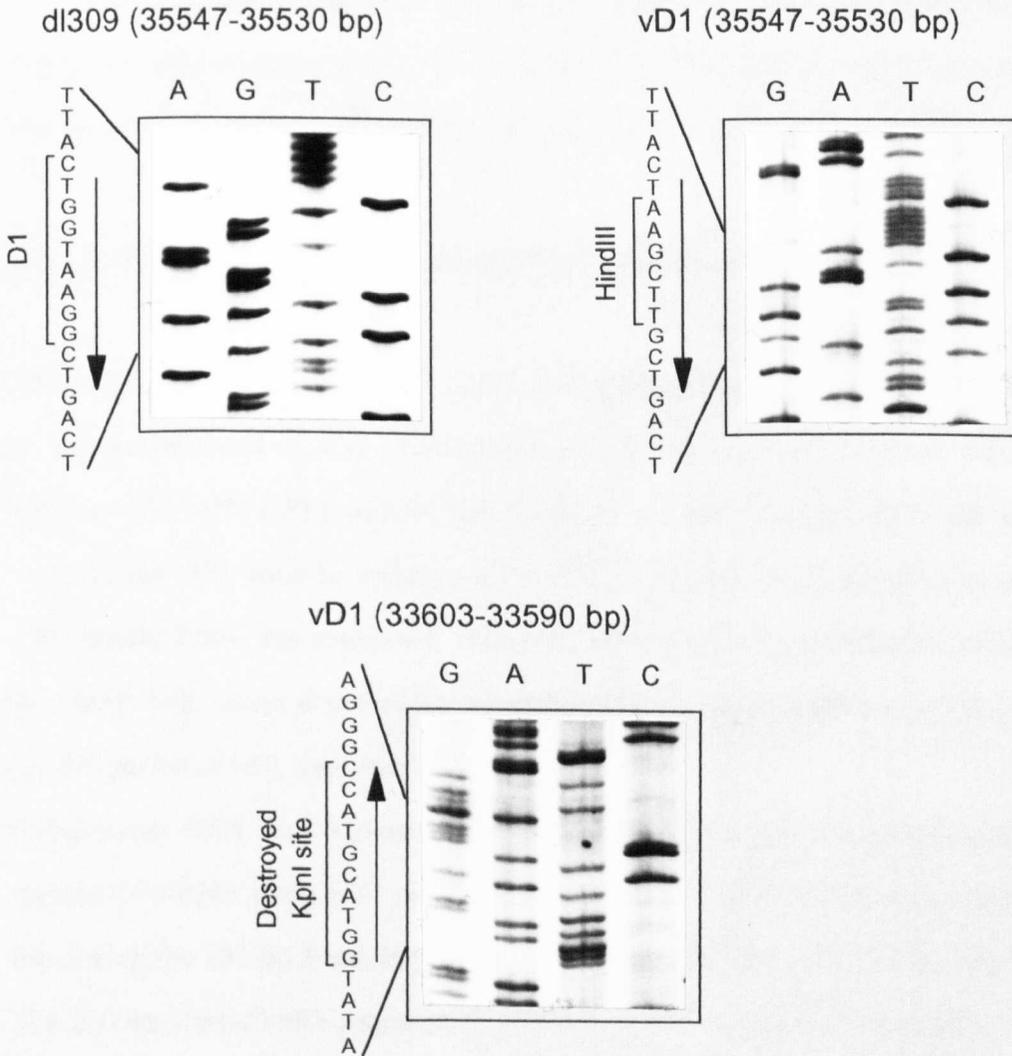
182 nt protected fragment
from the mutant virus

5.3. Construction of the D1 and ORF6 mutant virus.

The central plasmid involved in the construction of the E4 mutant virus was pRC6. pRC6 is a pBR322 based plasmid containing the righthand 8,607 bp of dl309, from the *EcoR*I site at 27331 to the end at 35938, including all the E4 region. The 2348 bp *EcoR*V fragment was removed from pRC6 and cloned into pBR322C (pBR322 with the *Cla*I site removed by cutting with *Cla*I, end filling the 5' overhang and ligating) to create pID26. Using pRC6 as a template, the D1 site was mutagenised, using a PCR protocol (see section 2.3.12., RS1 = *Sma*I (pRC6, 8023), RS2 = *Cla*I (pRC6, 8609), the primers used were J, K, L, and M (2.1.2.)), creating a *Hind*III site and destroying the D1 donor site sequence. The resulting PCR cDNA was subcloned into pID26 using *Cla*I and *Sma*I and the resulting clones screened for the presence of the *Hind*III site at the E4 D1 site by *Hind*III digestion. The region amplified during the PCR (the *Sma*I (pRC6, 8023) to *Cla*I (pRC6, 8609) fragment) was then subcloned, from a number of the positive clones, into the *Sma*I site of M13 mp19 and sequenced using the -40 M13 universal primer. A number of the positive clones (i.e. having a mutated D1 site) had point mutations in the amplified region in addition to the D1 mutation, probably due to the PCR, but one was selected, pID30, which was wild-type in sequence apart from the D1 mutation (this mutation at the D1 site in pID30 can be seen in Fig. 5.2.). The mutated 2348 bp *EcoR*V fragment from pID30 was then returned to pRC6 to create pRC6/D1x.

The E4 ORF6-34K reading frame was mutated as follows. The 4637 bp *Nhe*I fragment was removed from pRC6/D1x and subcloned into pBR322 to create pID39. pID39 was subsequently linearised with *Acc*65I, which cuts within the E4 ORF6-34K reading frame at the *Kpn*I site, Klenow repaired and self-ligated to create pID41. These manipulations resulted in the addition of 4 bp in the ORF6-34K reading frame causing a shift to the -1 reading frame so destroying the ability to produce intact ORF6-34K protein. dl355 (Halbert et al., 1985), which is phenotypically ORF6-34K minus, carries a similar frame shift at this position. The 4644 bp *Nhe*I fragment from pID41 was then returned to pRC6/D1x to create pRC6/D1x/ORF6x.

Figure 5.2. The sequence of the D1 site of dl309 at 35542 bp, the *Hind*III site in vD1 at 35543 bp (the location of the D1 site in dl309) and the mutated *Kpn*I site in the ORF6 reading frame of vD1 at 35599 bp (sequence numbered as Ad5 wt300). The DNA was sequenced using the dideoxynucleotide chain termination method (2.3.1.13.). The arrow indicates the 5' to 3' direction of the L strand. The samples were loaded as indicated.



The *EcoRI-ClaI* fragment of pRC6/D1x/ORF6x, comprising of the righthand end of the Ad5 genome, was then purified and ligated to the lefthand end *EcoRI* fragment of the Ad5 genome from dl309. This ligated product was subsequently transfected into W162 cells (Weinberg and Ketner, 1983), which are able to support the growth of E4 mutant viruses. Six potential mutant plaques were identified (the plaquing efficiency of W162s is 100 fold lower than HEK-293 cells (Halbert et al., 1984)), however only two of these contained virus, of which one (D) was mutated in the E4 region when the viral DNA was analysed by digestion with *KpnI* and *HindIII*. This virus was subsequently plaque-purified to create the E4 mutant virus vD1 which lacked both the D1 splice site and an intact E4 ORF6 reading frame, as analysed by restriction mapping (Fig. 5.3.).

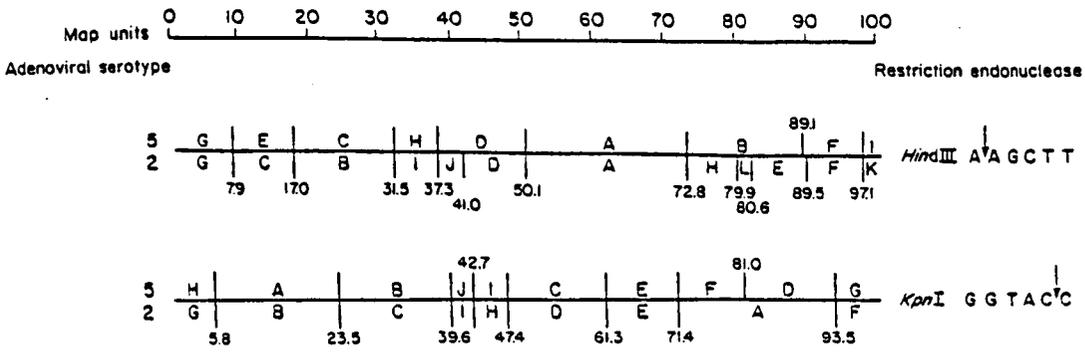
5.4. Analysis of mRNA A/full length RNA expression from the vD1 virus.

vD1 virus particles were purified on a CsCl gradient and used to co-infect HeLa cells at 500 particles/cell of vD1 virus along with 100 particles/cell of either dl309 (wild-type) or dl355 (E4 ORF6 mutant virus). dl309 provided the E4 ORF6-34K protein *in trans* to the vD1 virus to simulate a E4 ORF6-34K wild-type infection in which E4 vD1 mutant RNA was expressed, while the dl355 plus vD1 co-infection provided the E4 ORF6-34K minus situation for comparison. In parallel dl309 and dl355 infections (at 100 particles/cell) were carried out as controls.

Cytoplasmic RNA was harvested at 24 hr p.i. and analysed for the presence of the mutant E4 mRNA using a T7 probe generated from pID50. pID50 was constructed by subcloning the 181 bp *SmaI-HindIII* fragment of pID30 into pBR322 to create pID48. The 260 bp *NheI-EcoRI* fragment of pID48 was then subcloned between the *XbaI* and *EcoRI* sites of pGEM4. The resulting T7 transcribed RNA probe contained 182 contiguous nucleotides complementary to the mutant E4 mRNA and 175 contiguous nucleotides complementary to the wild-type E4 mRNA (with an intact D1 site). This 7 nucleotide difference in the length of the protected fragments in an

Figure 5.3. Restriction digest analysis of the D1 mutant virus vD1 and dl309 as analysed on a 0.8% agarose gel. The purified virus DNA was digested with either *KpnI* or *HindIII* as indicated above the gel. Also shown are the fragments which vary in size between the dl309 and vD1 viruses due to the removal of a restriction site (*KpnI*, vD1) or the addition of a restriction site (*HindIII*, vD1). All figures are in base pairs. The marker shown (in bp) is the '1 kb ladder' (Gibco BRL).

Indicated below are the restriction maps of Ad5 wt300 virus as reviewed by Tooze (1982). It should be noted that dl309 lacks the 72.8 map unit *HindIII* restriction site, fusing the *HindIII* fragments A and B together. Also the *KpnI* F fragment is slightly smaller in dl309 than in wt300 due to a substitution in the E3 region. The 1008 bp *HindIII* I fragment is cleaved by *HindIII* in vD1 as the splice donor site D1 was mutated to a *HindIII* site. Also in vD1 the *KpnI* 4811 bp D and 2340 bp G fragments are fused due to the mutation of the *KpnI* site at 93.5 map units in the ORF6 reading frame to create a 7151 bp fragment.

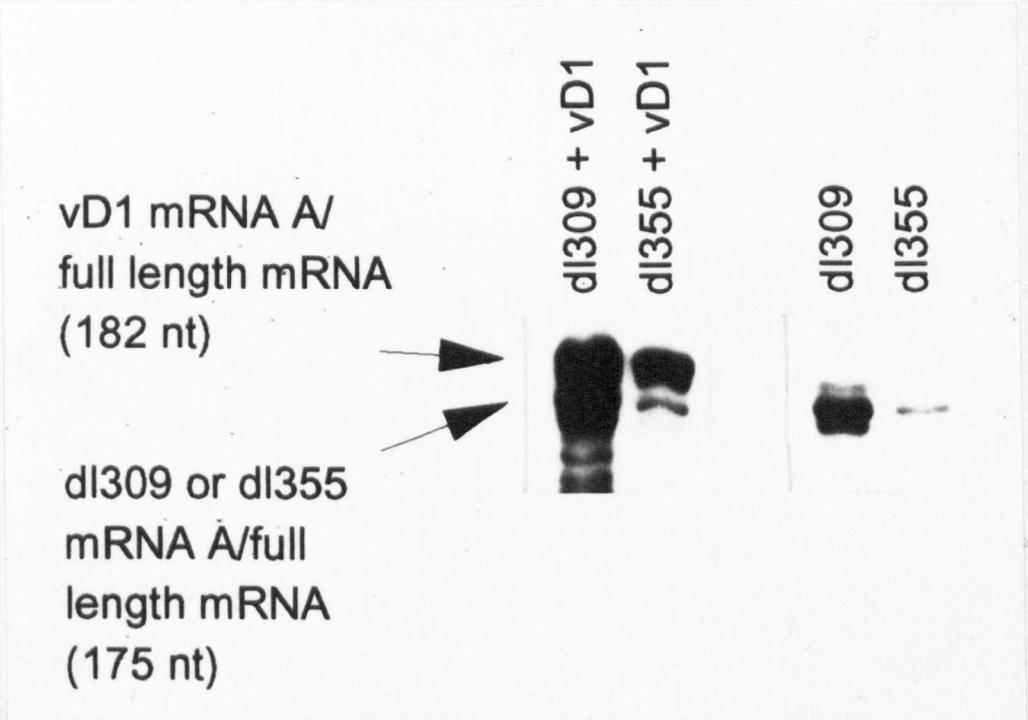


RNase protection assay was sufficient to allow the resolution of the RNA species on a 6% acrylamide gel.

The results of such RNase protection analysis can be seen in Fig. 5.4. mRNA A/full length mRNA from the mutant vD1 virus accumulates to levels in the cytoplasm in the dl355 co-infection akin to those found in the dl309 co-infection while wild-type mRNA A/full length mRNA does not accumulate to such high levels in the cytoplasm in the dl355 co-infection. This would suggest that the D1 site is involved in conferring dependence on the RNA on the E4 ORF6-34K protein, and probably the E1b-55K protein, for optimal cytoplasmic accumulation. However, the levels of mutant mRNA A/full length mRNA in the dl355 infection are still lower than those in the dl309 infection indicating that, although the removal of the D1 site does alleviate some of the dependence of mRNA A/full length mRNA on the E1b-55K/E4 ORF6-34K complex it does not alleviate it completely. This may indicate that sequences in the mRNA other than D1 contribute to the E1b-55K/E4 ORF6-34K dependence of an mRNA such as the other splice sites unique to these two late mRNAs (A1a and, for full length only, D3, A3 and A5) or undefined intronic regions.

It should be noted, however, that the results obtained when conducting this series of RNase protection assays were not as expected. For some unexplained reason the difference in the level of mRNA A/full length mRNA in wild-type and E4 ORF6 mutant infections was not as great as observed in multiple experiments conducted before (see chapter 4). Also the overall levels of the E4 mRNA detected were low in all experiments and the RNA not of good quality, resulting in a lot of degradation in the assays. To evaluate conclusively the role of the D1 donor site in E1b-55K/E4 ORF6-34K dependence more studies need to be conducted with this virus.

Figure 5.4. RNase protection analysis of dl309, dl355, and vD1 mRNA A/full length mRNA using a probe made from pID50 (2.3.5). HeLa cells were infected with 500 particles/cell of vD1 and/or 100 particles/cell of dl309 or dl355. 5 µg of unselected cytoplasmic RNA was analysed 24 hr p.i. for the presence of mRNA A/full length mRNA and β-actin (not shown).



Chapter 6.

The generation of a polyclonal antiserum to E4 ORF1

6.1. Introduction

The sequencing of Ad5 E4 DNA (see chapter 3) indicated that the predicted ORF1 protein is a hydrophobic protein of 128 aa. Three mRNAs potentially encode the ORF1 protein, mRNAs A, B and full length mRNA. Studies of the E4 mRNAs (chapter 4) demonstrated that while mRNA B does not exist in the cytoplasm of infected cells in the absence of translation inhibitors, mRNA A/full length mRNA has late expression kinetics, first appearing in the cytoplasm at 16 hr p.i. This result suggests that the ORF1 protein is made, and is presumably required, late in the infection, after DNA replication. However the Ad5 ORF1 protein has not yet been shown to exist in infected cells and no role for it has been established in the Ad5 lytic cycle, although recently, Javier (1994) found the Ad9 ORF1 protein and demonstrated that it was involved in the induction of mammary tumours in rats (section 1.4.2.2.).

This chapter describes attempts to identify the ORF1 protein in Ad5 infected cells. The strategy employed involved the expression and purification of ORF1-related proteins and the subsequent generation of mono-specific polyclonal antisera to the purified protein in out-bred half-lop rabbits. These animals were chosen as they yield a large amount of serum through the course of an immunisation regime and because their out-bred character makes it more likely that they will respond effectively to the immunogen. The resulting polyclonal antisera were then used to search for the ORF1 protein in infected cells using a number of different immunochemical techniques.

6.2. Viruses and cells used.

Four different viruses were used in this series of experiments; wt300, dl1-3, dl327 and dl351. wt300 is a plaque-purified stock of wild-type virus (Jones and Shenk, 1978). The dl1-3 mutant virus carries a deletion in the E4 region from the *Sma*I (35354) to the *Bgl*III (34387) in a wt300 background which results in the loss of ORFs 1, 2 and 3 (Huang and Hearing, 1989a). dl327 lacks the 78.5- to 84.3-map-unit *Xba*I fragment

from early region 3 and is phenotypically wild-type for growth in tissue culture. This deletion was described originally in virus dl324 (Thimmappaya et al., 1982). dl351 contains a 5 bp insertion at the *Bst*EII site (35222) causing the disruption of the E4 ORF1 reading frame in a dl327 background (Halbert et al., 1985). HeLa cells were used for virus growth throughout this study.

6.3. Expression of the ORF1 protein using the pGEX vector

The commercially available pGEX system (Pharmacia) was selected to express the ORF1-14K protein as it had been reported that it allowed the simple, high level expression and purification of the recombinant proteins in *E. coli* (Smith and Johnson, 1988). The foreign polypeptide is expressed as a fusion with the glutathione *S*-transferase (GST) protein, which can be easily purified from lysed cells by affinity chromatography with glutathione-sepharose beads, followed by elution in the presence of free glutathione. The pGEX vectors contain the *ptac* promoter which promotes the transcription of the GST open reading frame, followed by a MCS containing three unique restriction sites; *Bam*HI, *Sma*I, and *Eco*RI, followed in turn by termination codons in all three frames (see Fig. 6.1.). The vector which was selected, pGEX-2T (Smith and Johnson, 1988), also encoded a thrombin cleavage site between the GST gene and the MCS, into which the foreign polypeptide can be cloned, allowing the cleavage of the fusion protein to release the foreign polypeptide. The GST moiety can then be removed using glutathione sepharose beads leaving the purified heterologous polypeptide in solution.

A cDNA fragment suitable for the expression of the ORF1 protein was generated using PCR. The template DNA used was the Ad5 wt300 genomic DNA while the two primers used were IDC and IDD (see section 2.1.2.). The 3' terminus of IDC contained 16 nucleotides complementary to the DNA of the 5' terminus of ORF1 and DNA immediately upstream. In the 5' terminus of IDC was a *Bam*HI site positioned so that when the resulting ORF1 cDNA was cloned into the pGEX-2T vector using

Figure 6.1. The pGEX-2T expression vector

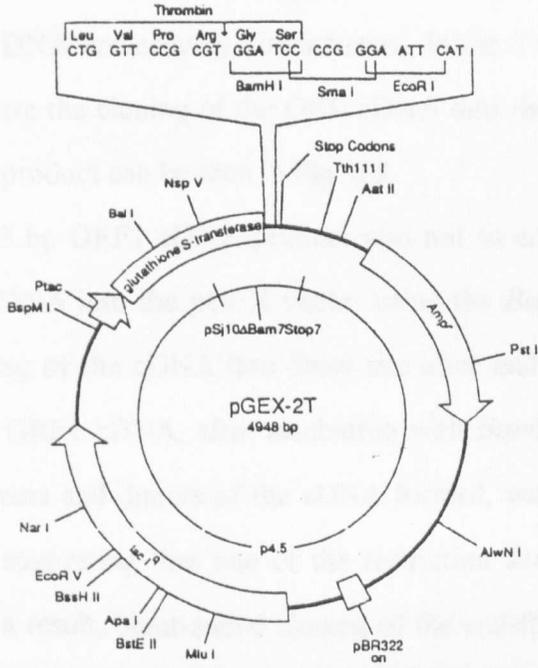
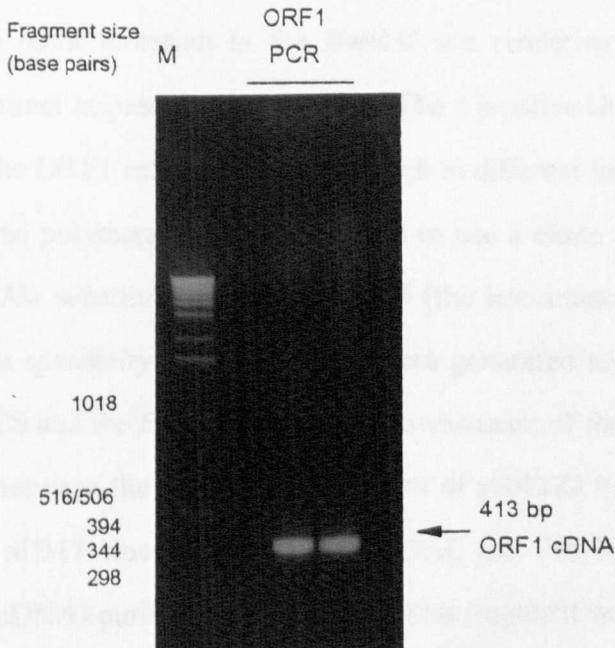


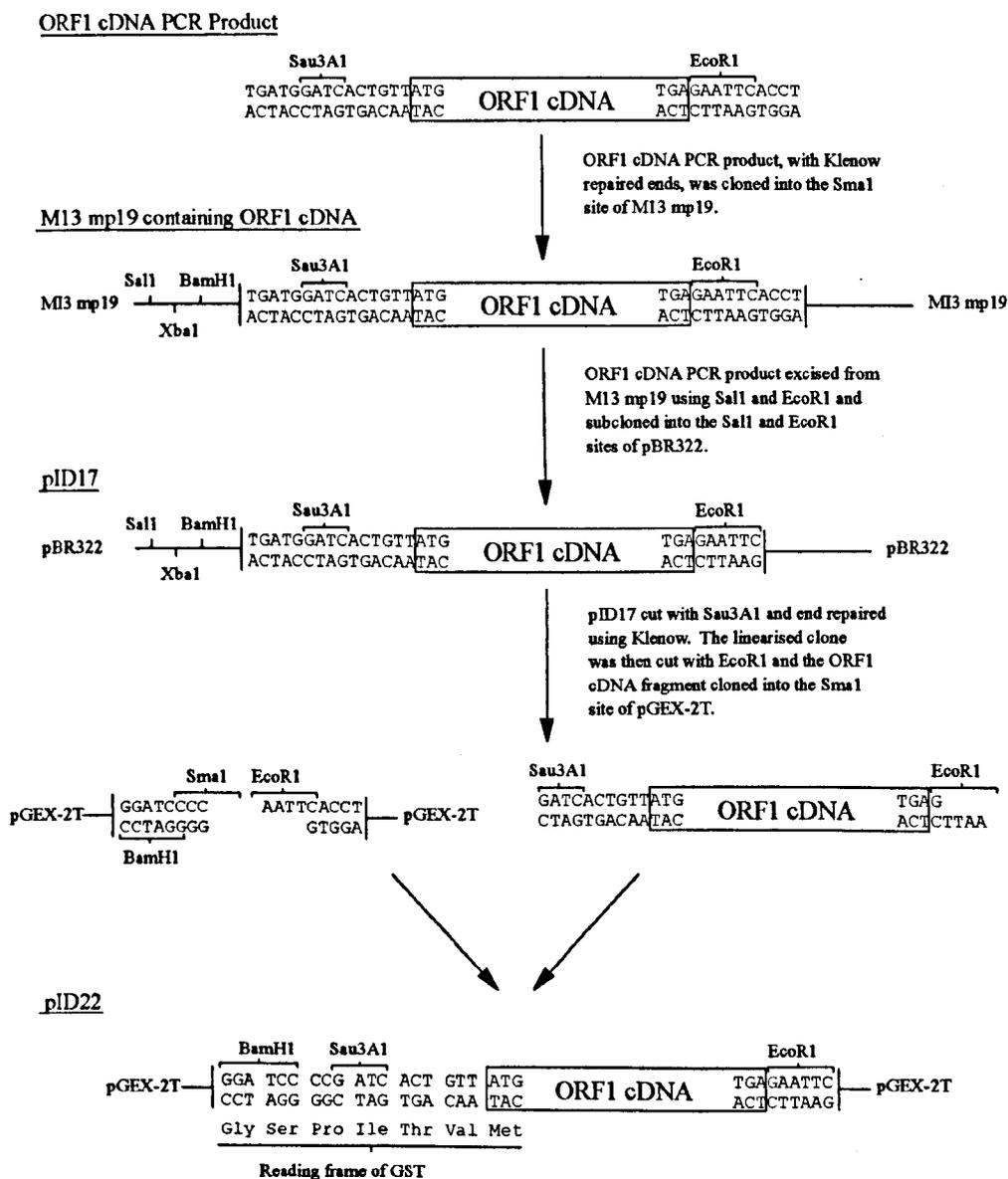
Figure 6.2. PCR amplification of the ORF1 cDNA using primers IDC and IDD (2.3.1.11.). All of the 413bp PCR product was run on a 1.2% agarose gel. See Fig. 5.3. for marker details.



the *Bam*HI site it was in frame with the GST gene reading frame. Similarly the 3' terminus of IDD contained 18 nucleotides complementary to the DNA of the 3' terminus of ORF1 and DNA immediately downstream. In the 5' terminus of IDD was an *Eco*RI site to facilitate the cloning of the ORF cDNA into the pGEX vector. The resulting ORF1 cDNA product can be seen in Fig. 6.2.

The cloning of this 413 bp ORF1 cDNA product was not as easy as expected. The direct cloning of the cDNA into the pGEX vector using the *Bam*HI and *Eco*RI sites failed, as did the cloning of the cDNA into *Sma*I site after end repair of the cDNA. On self ligation of the ORF1 cDNA, after incubation with *Bam*HI and *Eco*RI, it was noted that only monomers and dimers of the cDNA formed, with no evidence of any higher order moieties suggesting that one of the restriction sites was non-functional (data not shown). As a result, blunt-ended cloning of the end-filled ORF1 cDNA into a series of different vectors was attempted. Eventually 6 positive plaques were identified after blunt-end ligation of the ORF1 cDNA into *Sma*I site of M13 mp19. These six clones were sequenced using the M13 -40 universal primer; two were false positives while the other four contained the ORF1 sequence. All four positive clones contained the same point mutation in the *Bam*HI site rendering it non-functional, indicating that the primer sequence was incorrect. The 4 positive clones also contained point mutations in the ORF1 reading frame (although in different locations) due to the high error rate of Taq polymerase. It was decided to use a clone with just one point mutation, a Val to Ala substitution at position 128 (the last amino acid), as this was unlikely to affect the specificity of polyclonal antisera generated to it. Using the *Sal*I site of the mp19 MCS and the *Eco*RI site directly downstream of the ORF1 cDNA, the cDNA was cloned between the *Eco*RI and *Sal*I sites of pBR322 to create pID17, for ease of handling. pID17 was then cut with *Sau*3AI, the 714 bp fragment (which contains the ORF1 cDNA) purified and end-filled. This fragment was subsequently cut with *Eco*RI and the resulting ORF1 fragment cloned into the *Sma*I and *Eco*RI sites of pGEX-2T (pID22), ensuring ORF1 was in the same translational reading frame as that of GST (see Fig. 6.3.).

Fig 6.3. Construction of the pGEX-2T/ORF1 cDNA expression vector pID22.



To establish if pID22 was able to express the expected 358aa fusion protein, cells containing the construct were harvested and analysed after various periods of induction at 37°C with IPTG. At each time point, 2 ml of sample was prepared as section 2.3.7.1., and amounts of the insoluble fraction and supernatant equivalent to 0.1 ml of the culture were loaded onto an SDS-PAGE gel (Fig. 6.4a.). In parallel, the expression GST from cells containing pGEX-2T was examined as a positive control (Fig. 6.4b.) Fig. 6.4a. demonstrates that a 40K protein, presumptively GST/ORF1 (GST is a 26K protein (Smith et al., 1986) while ORF1 is predicted to be 14K in size, to give a 40K fusion protein), is induced in the cultures by IPTG but is largely insoluble. Also shown in Fig. 6.4a. (6 hr, pure), attempts to purify the GST/ORF1 protein from the 6 hr induced supernatant were not successful. In comparison, the induced GST protein in Fig. 6.4b. is present largely in the soluble fraction and was easily purified on the affinity matrix.

Although insoluble fusions can be made soluble in denaturing conditions (such as 8M urea) they cannot be purified using the GST interaction with glutathione as this interaction is dependent on the native conformation of the GST protein. A number of attempts were therefore made to produce soluble GST/ORF1 fusion protein. Initially the lysis conditions of the bacteria were altered, adding 0.03% SDS, 1% Tween 20 or 1% Triton-X100 after sonication, as it had been reported that purification using glutathione was still possible after such detergent treatment (Smith and Corcoran, 1992), but purification still failed (data not shown). The growth conditions of the bacteria were also altered in an attempt to produce soluble fusion protein: the bacteria were induced at a lower temperature and for longer (30°C for 16 hours) as it had been reported by Schein and Noteborn (1988) that a number recombinant proteins, that were insoluble in *E. coli* grown at 37°C, were soluble when the bacteria were grown at 30°C. However this was not the case for the GST/ORF1 protein (Fig. 6.4a. and data not shown). Blackwell and Horgan (1991) also demonstrated that the growth of bacteria in LB supplemented with 660 mM sorbitol and 2.5 mM betaine facilitated the production of soluble proteins. However when pID22 was induced at either 30 °C or

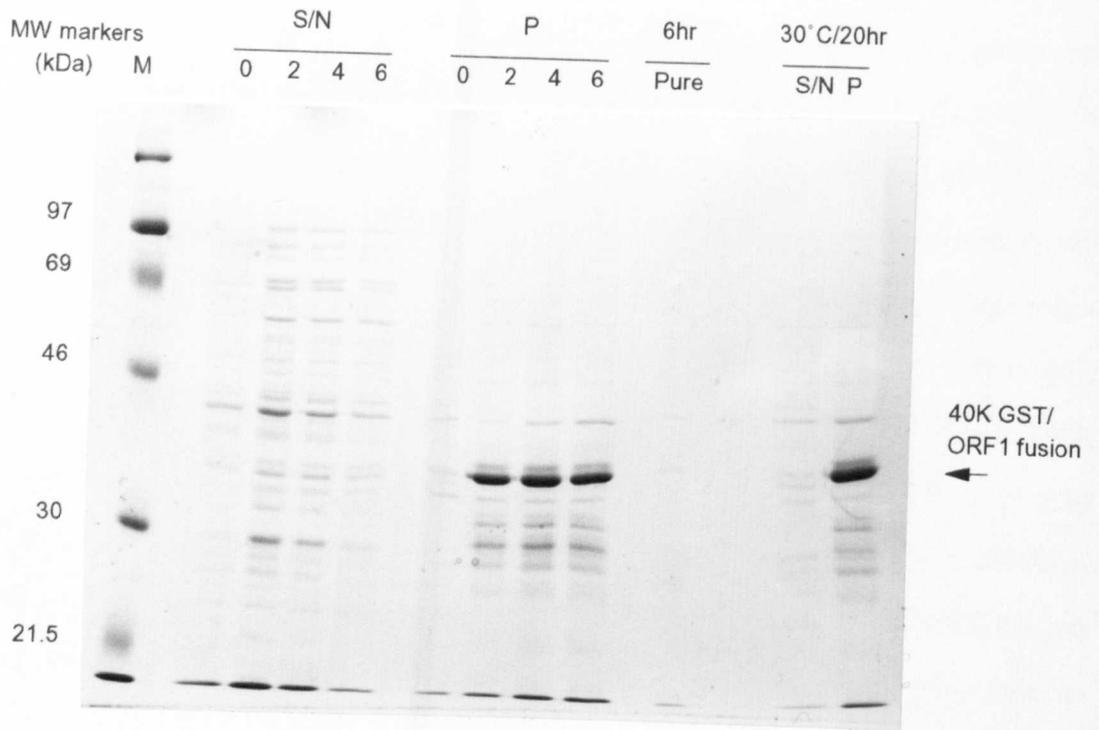
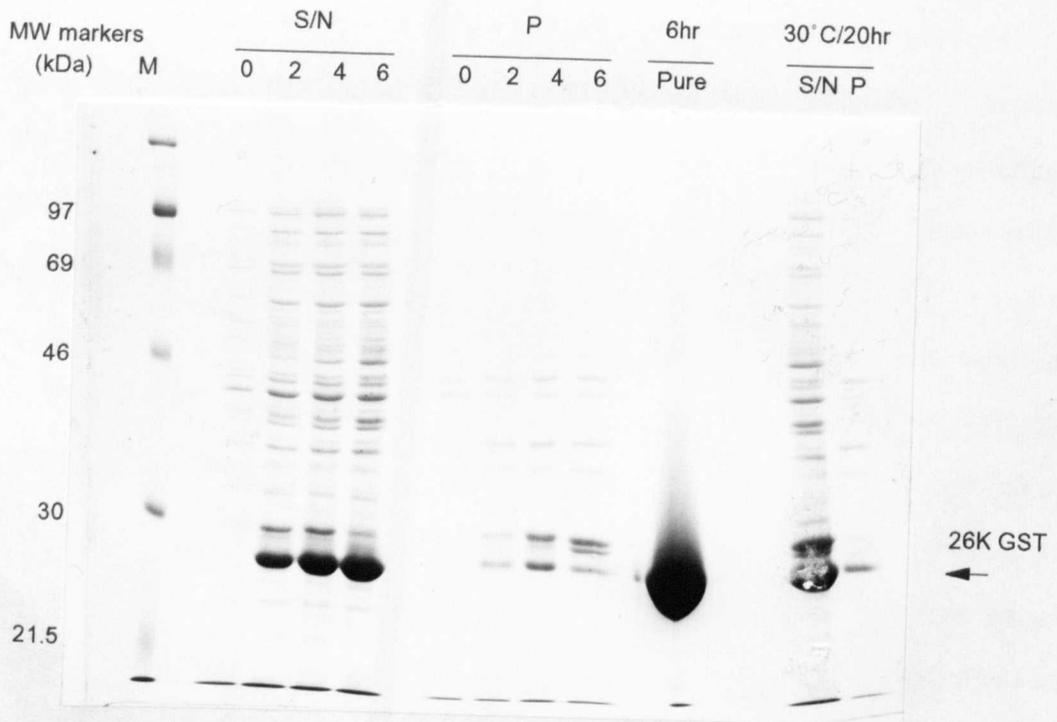
A**B**

Figure 6.4. Protein expression from the pGEX constructs. Panel A: The expression of the GST/ORF1 fusion protein from pID22 at 30°C and 37°C assayed by SDS-PAGE. The times post induction (hrs) and the sample fraction are indicated above the gel. All samples are the equivalent of 0.1 ml of the culture, except the purified fraction which is equivalent to 2 mls of culture. Visualised by coomassie blue staining. S/N=supernatant; P=pellet; Pure=glutathione sepharose beads. Figure 6.4 Panel B: The expression of the GST protein from pGEX-2T at 30°C and 37°C assayed by SDS-PAGE. All other details as Fig.6.4a.

Markers (Amersham): 97 kDa : Phosphorylase B

69 kDa : Bovine Serum Albumin

46 kDa : Ovalbumin

30 kDa : Carbonic Anhydrase

21.5 kDa : Trypsin Inhibitor

14.3 kDa : Lysozyme

37°C in this supplemented LB medium, little or no increase in solubility or efficiency of purification was observed (data not shown).

Finally, the pGEX-2T/ORF1 construct was modified to produce separate GST fusions with N and C terminal fragments of ORF1 in an attempt to allow the purification of at least part of the ORF1 protein. pID22 was cut with *Sma*I and *Eco*RI to remove DNA encoding the C terminus of ORF1 and re-ligated to produce pID22N. pID22 was also cut with *Bam*HI and *Sma*I to remove DNA encoding the N terminus of ORF1 and re-ligated to produce pID22C. Both of these new constructs were sequenced across the GST/ORF1 junction, using the primer IDG (section 2.1.2.), to ensure the reading frame was intact. After induction at 37 °C for 4 hours both the resulting fusion proteins, the 291aa GST/ORF1N and the 297aa GST/ORF1C, could be readily identified in the bacterial total cell protein samples. However, neither could be purified using glutathione sepharose (Fig.6.5.).

Since native, soluble GST/ORF1 fusion could not be produced and so purified, it was decided to purify the fusion protein as an insoluble inclusion body preparation. GST/ORF1 expression was induced from pID22 for 3 hours at 37 °C in a 1 l culture and the GST/ORF1 inclusion bodies purified as in section 2.3.7.2. This preparation gave a yield of about 6 mg/l (the GST/ORF1 band intensity on a SDS-PAGE gel was compared to that of known amounts of BSA to give an estimation of the amount of protein present). This preparation was resuspended in 6 ml of PBS, and 10 µl analysed by SDS-PAGE (Fig. 6.6.). Although the preparation was not 100% pure it was decided to use it for the production of a polyclonal antisera as the contaminating proteins were of bacterial origin and of a relatively low concentration compared to the GST/ORF1 fusion.

Figure 6.5. The expression of the GST/ORF1N and GST/ORF1C fusion proteins from 3 clones of pID22N and pID22C respectively at 37°C for 4 hours assayed by SDS-PAGE. The total samples are the equivalent of 0.1 ml of culture while the purified samples (2.3.7.1.) are equivalent to 2 mls of culture. Visualised by coomassie blue staining. See Fig. 6.4. for the details of the markers.

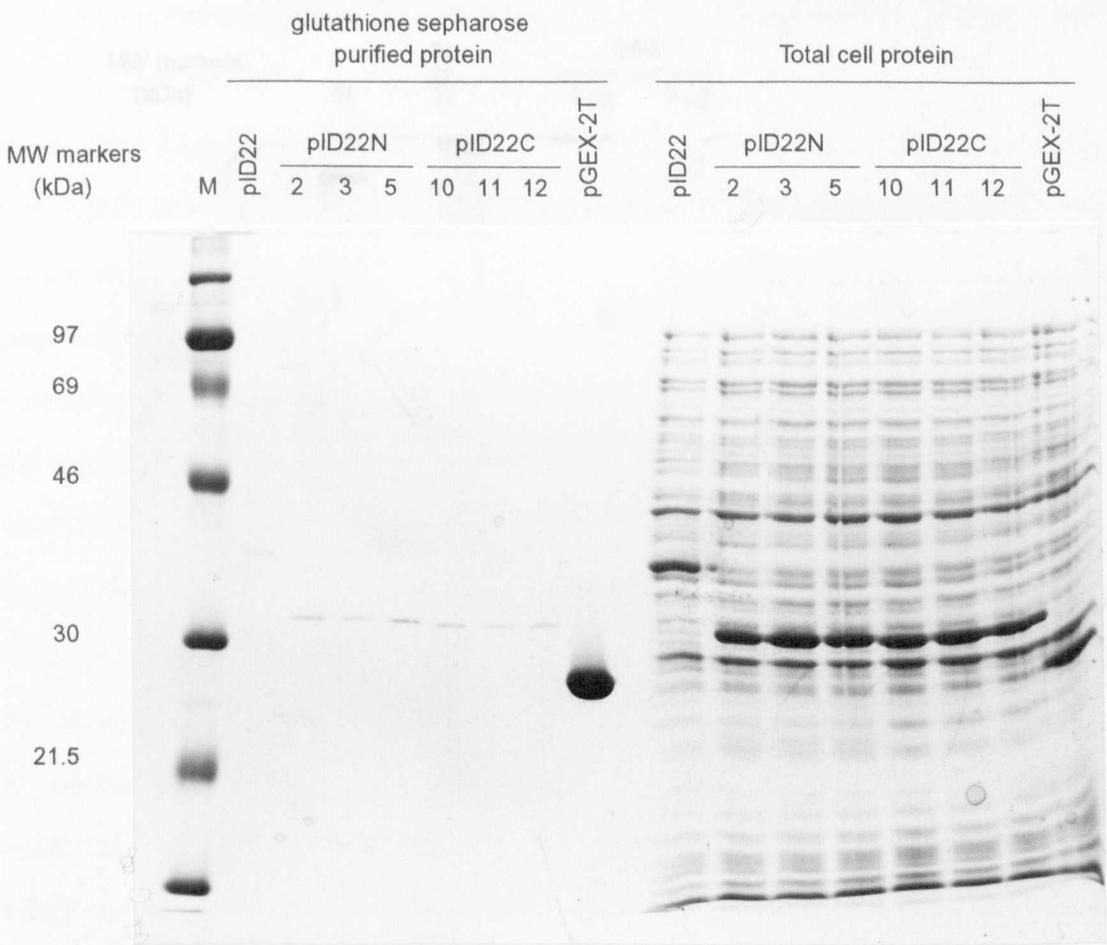
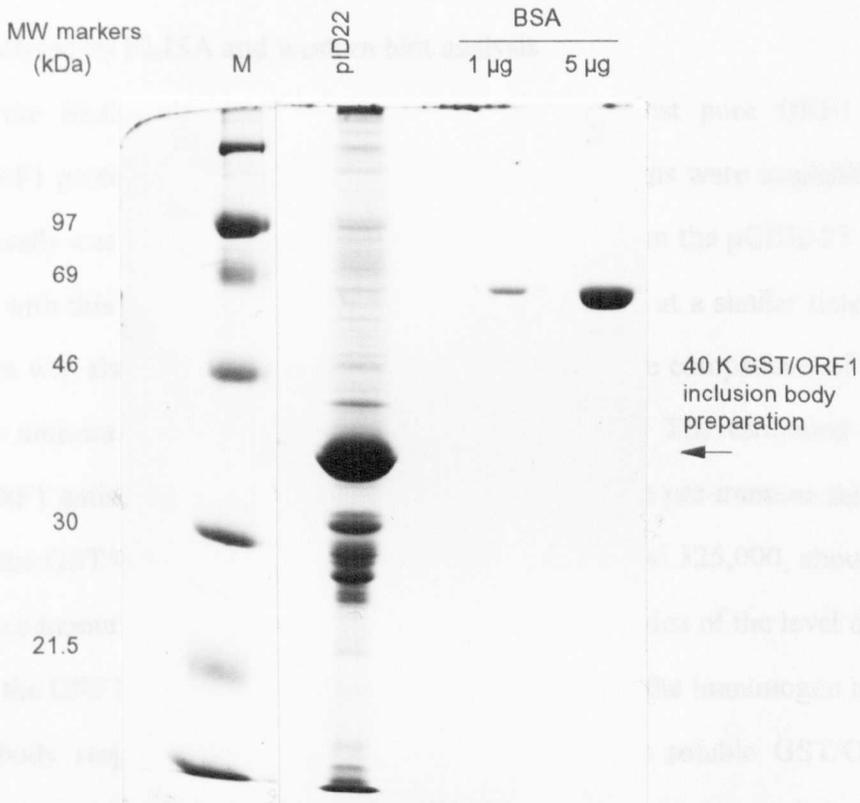


Figure 6.6. The GST/ORF1 inclusion body preparation used in the generation of the first ORF1 polyclonal antiserum assayed by SDS-PAGE. 10 μ l of the GST/ORF1 preparation was assayed on the gel along with 1 and 5 μ g of BSA. Visualised by coomassie blue staining. See text for details of the preparation. See Fig. 6.4. for marker details.



6.4. The generation of an ORF1 antiserum using the GST/ORF1 inclusion body preparation.

A half-lop rabbit was inoculated subcutaneously three times with 500 µg of the ORF1 inclusion body preparation and Freund's adjuvant (1:1 volume ratio was vortexed for 30 minutes so the emulsion was viscous), over 65 days. The rabbit was bled three times, 42, 56, and 81 days post first inoculation, giving serum samples A, B and C respectively, before it had to be put down due to self-inflicted injury. The three sera were analysed by ELISA and western blot analysis.

Ideally the ELISA should have been carried out against pure ORF1 protein or GST/ORF1 protein but as no soluble forms of these proteins were available, GST (0.5 µg per well) was used (expressed as in section 2.3.7.1. from the pGEX-2T vector). In parallel with this analysis a GST/ORF2 antisera generated at a similar time post-initial inoculum was also tested (see chapter 7). This allowed the comparison of the titres of the two antisera. The results can be seen in Fig. 6.7. The estimated titre of the GST/ORF1 antisera is 1,800, about 40 fold higher than the pre-immune sera, while the titre of the GST/ORF2 antisera is about 100 times greater at 325,000, about 4,000 fold above pre-immune. Although this assay does not give an idea of the level of antibodies against the ORF1 protein it is a guide as to how effective the immunogen is at eliciting an antibody response. It is evident from this that the soluble GST/ORF2 fusion generated a high titre of antibodies in the rabbit against GST (a good titre against the ORF2 moiety was also generated, see section 7.4.) while the insoluble GST/ORF1 inoculum only elicited a low antibody titre against GST, and presumably therefore against ORF1 also. The titre however was rising when the animal had to be put down. Western blot analysis was carried out using the anti GST/ORF1-C sera collected 81 days post the first inoculation. About 0.5 µg of the bacterially expressed GST/ORF1 fusion inclusion body protein was loaded onto a SDS-PAGE gel and transferred for probing with this serum (Fig. 6.8.). This western blot clearly demonstrates that the

Figure 6.7.

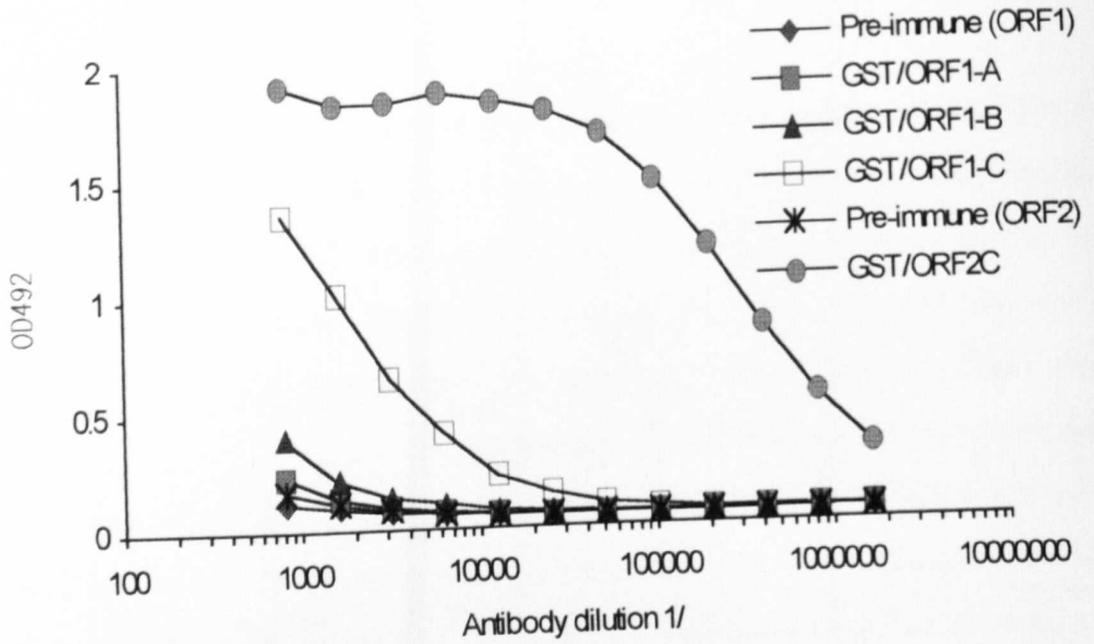


Figure 6.8.

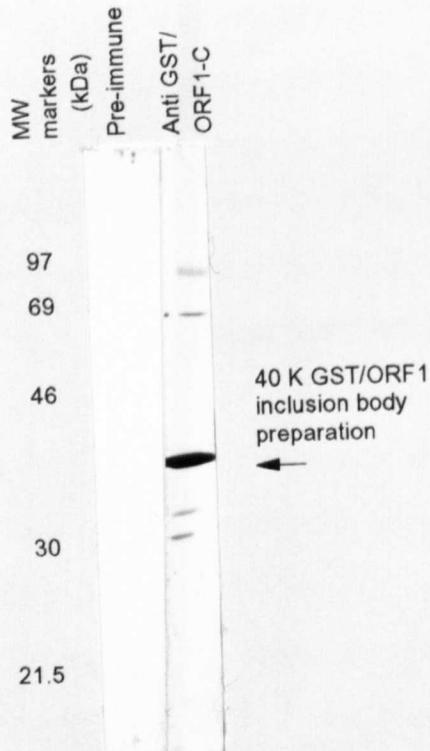


Figure 6.7. Titration of the anti-GST/ORF1 activity in serum samples by ELISA (2.3.9.3.). Serial dilutions of anti-GST/ORF1 sera taken in successive bleeds from a rabbit inoculated with the GST/ORF1 inclusion body preparation were tested against 0.5 μ g of GST protein. See text for details of the GST/ORF1 samples. Also shown is the serial dilution of the GST/ORF2 antisera from an equivalent time post inoculation (77 days) as the last GST/ORF1 antisera (see section 7.4.).

Figure 6.8. Western blot analysis of the GST/ORF1 polyclonal antisera (2.3.9.1.). A 1/500 dilution of the sera sample GST/ORF1-C was used in the detection of 0.25 μ l of the inoculum (\sim 0.5 μ g). See Fig. 6.4. for marker details.

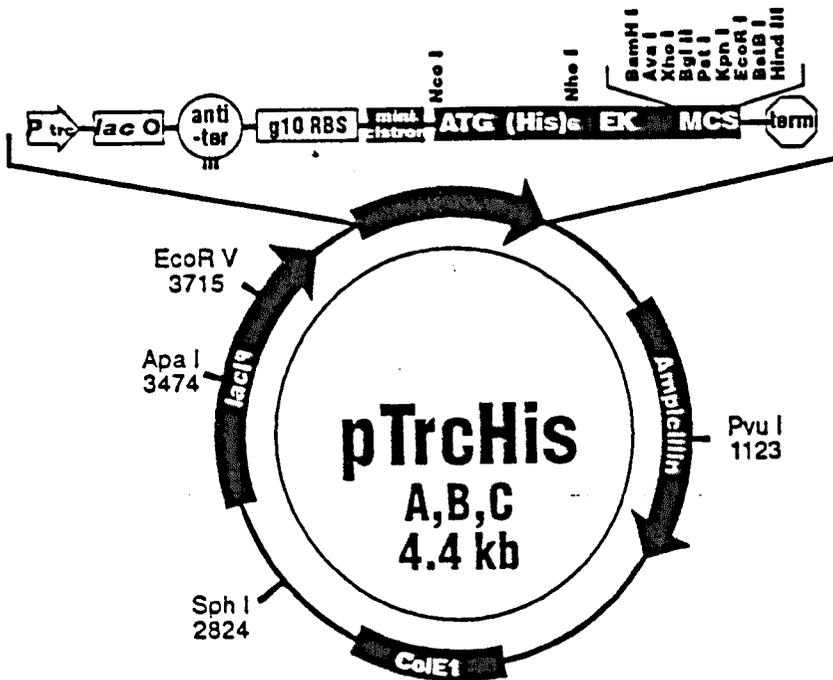
antiserum has activity against the fusion protein as it recognises the 40K fusion protein while the pre-immune serum fails to do so. It is unclear if this activity was against just the GST moiety or to both the GST and ORF1 components of the fusion.

Attempts to identify the ORF1 protein using this antiserum in infected HeLa cells by immunoprecipitation assay and western blot analysis failed; no proteins present in the wild-type infected cell but absent in an ORF1 mutant (dl351 or dl1-3) infected cell could be identified (data not shown). The most likely reason for this failure to identify the ORF1 protein is the low titre of the antisera. As a result, a fresh attempt to generate a high titre ORF1 antiserum was made.

6.5. The expression of ORF1 using the XPRESS expression system.

The XPRESS system (Invitrogen) allows the purification of bacterially expressed proteins by exploiting the metal binding activity of six contiguous histidine residues. The pUC-derived pTrcHisA vector (see Fig. 6.9) contains an IPTG inducible promoter (pTrc) which is repressed by the *LacI^q* gene product, encoded by the vector, ensuring that expression of the recombinant protein only occurs upon induction. Downstream of the promoter is a transcription anti-termination region (*E. coli rrnB*) followed by the bacteriophage T7, gene 10 RBS. Translation of the recombinant protein is enhanced by the presence of an 8 aa mini-cistron following the RBS which efficiently restarts translation at the following ATG. Immediately following the mini-cistron is the ATG translation initiation site for the recombinant protein followed by a region that encodes 6 contiguous histidine residues and a MCS enabling the fusion of the histidine residues to the N terminus of the recombinant protein. This N-terminal metal binding domain (the 6 histidine residues) allows the simple one-step purification of recombinant proteins by immobilised metal affinity chromatography using the nickel ion resin ProBond (Invitrogen). Positioned between the histidine residues and the MCS is an enterokinase cleavage site allowing the removal of the histidine residues from the recombinant protein after expression and purification. The

Figure 6.9. The pTrcHis vector of the Invitrogen Xpress bacterial expression system (from commercial literature of Invitrogen).



P_{trc} -35 (*trpB*), -10 (*lacUV5*) hybrid promoter for high level transcription initiation.

lac O The *lac* repressor (encoded by *lacI^q* gene on vector) binding site for repression of transcription. High level transcription can be induced with IPTG.

anti-ter The Sequence from *E. coli rrmB* (ribosomal RNA gene) which reduces the level of premature transcription termination.

g10 RBS The ribosome binding site from bacteriophage T7 gene 10 which has been shown to optimize translation initiation.

mini-cistron A short open reading frame containing nucleotide sequences that are efficiently translated in prokaryotic cells. A ribosome binding site (RBS) is present within the coding sequence 5' to the translation termination codon. This RBS and termination codon are positioned in frame and three nucleotides upstream from the translation initiation codon to be used to express the fusion protein of interest. Following translation of the open reading frame of the mini cistron, ribosomes efficiently reinitiate translation at the second initiation site.

ATG The translation initiation site for the fusion protein of interest.

(His)₆ The sequence encoding 6 tandem histidine residues which have a high affinity for Ni²⁺ charged ProBond™ resin.

EK The sequence encoding an enterokinase specific cleavage site (asp-asp-asp-lys-asp).

MCS The multiple cloning site, positioned in three different reading frames in versions A, B, and C relative to the poly-histidine containing leader peptide open reading frame.

term Transcription termination sequences from *E. coli rrmB* gene.

advantage of this system over the pGEX system is that the recombinant protein can be isolated under strong denaturing conditions as the interaction of the histidine residues with the nickel ions is not dependent on protein secondary structure. Insoluble proteins can therefore be solubilised in 8M urea or 6M guanidine hydrochloride and still be purified.

The ORF1 cDNA from pID22 was cloned into the pTrcHisA vector at the *Bam*HI and *Eco*RI sites. A positive clone (pID35) was then expressed as per the manufacturers instructions. 1/10 of the resulting pellet and supernatant fractions (from 2 ml of culture) were then assayed on an SDS-PAGE gel for the presence of the His/ORF1 fusion protein (Fig. 6.10.). A faint band of about 18K (estimated size of ORF1 with the histidine residues attached), which increased in intensity as the length of induction increased, was observed in the insoluble pellet fraction. This is probably the His/ORF1 fusion.

The His/ORF1 fusion protein was subsequently purified from a 50 ml culture induced as above for 6 hours. The insoluble protein was solubilised in 6M guanidine hydrochloride and the fusion protein purified under denaturing conditions using the ProBond nickel resin and denaturing elution buffer (see manufacturers instructions) at pH3.8 to elute the purified protein from the column in 0.5ml fractions. 1/25 th of each of these fractions was then assayed on an SDS-PAGE gel for the presence of the His/ORF1 fusion protein (Fig. 6.11.). The His-ORF1 protein was successfully purified using this system, eluting off the column in fractions 3-5. However the estimated purity (by SDS-PAGE analysis) of the His/ORF1 fusion was only 75% as the fractions also contained contaminating higher molecular weight bands. By quantitating the amount of protein present in each sample using the Bradford assay it was estimated that this 50 ml culture was only producing about 50 µg of His-ORF1 protein, i.e. a yield of 1 mg/l. Neither yield nor the purity of the sample could be improved in subsequent purification experiments. As a result, this system was not exploited further.

Figure 6.10.

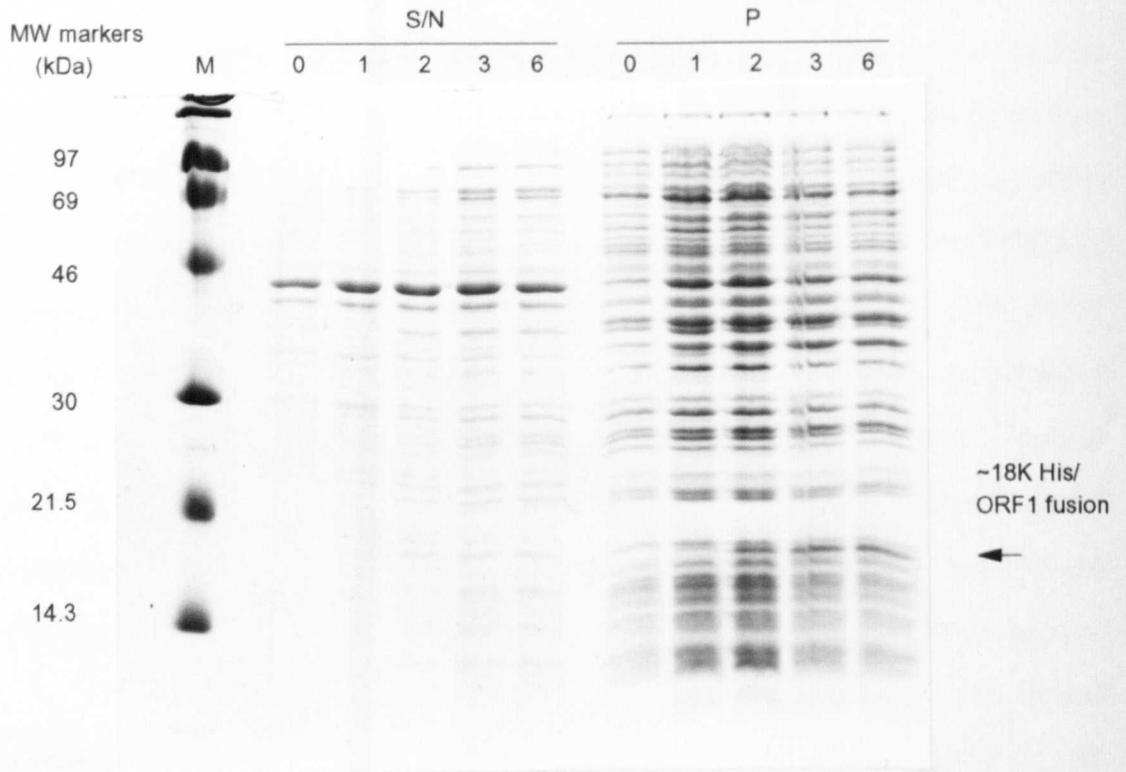


Figure 6.11.

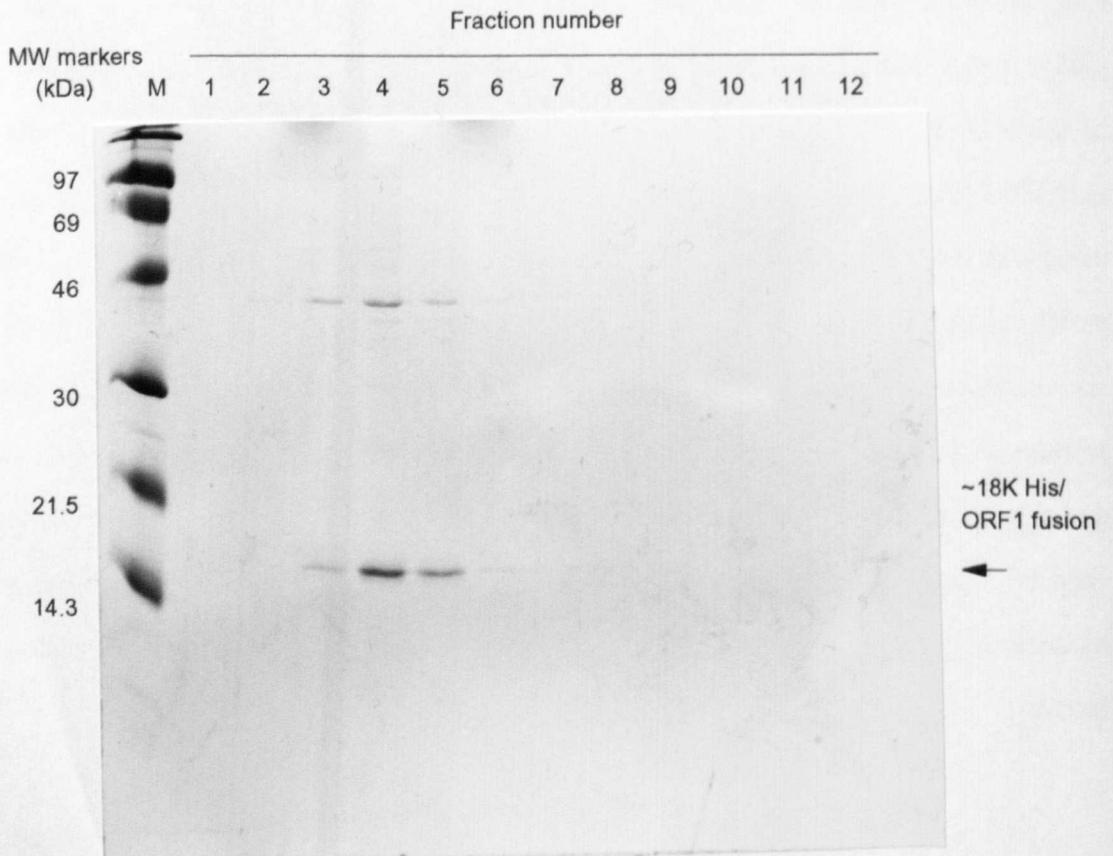


Figure 6.10. The expression of the His/ORF1 fusion at 37°C from pID35 assayed by SDS-PAGE. A 25 ml culture of bacteria containing pID35 was induced at A_{600} 0.5 with 1 mM IPTG and samples taken at the times post induction (hr) indicated above the gel. Samples were lysed in PBS by three rounds of freeze/thawing and centrifuged at 16,000 x g for 10 minutes to separate the insoluble (P) and soluble (S/N) fractions. All samples are equivalent to 0.2 ml of culture. Visualised by coomassie blue staining. S/N = supernatant; P = pellet. See Fig. 6.4. for marker details.

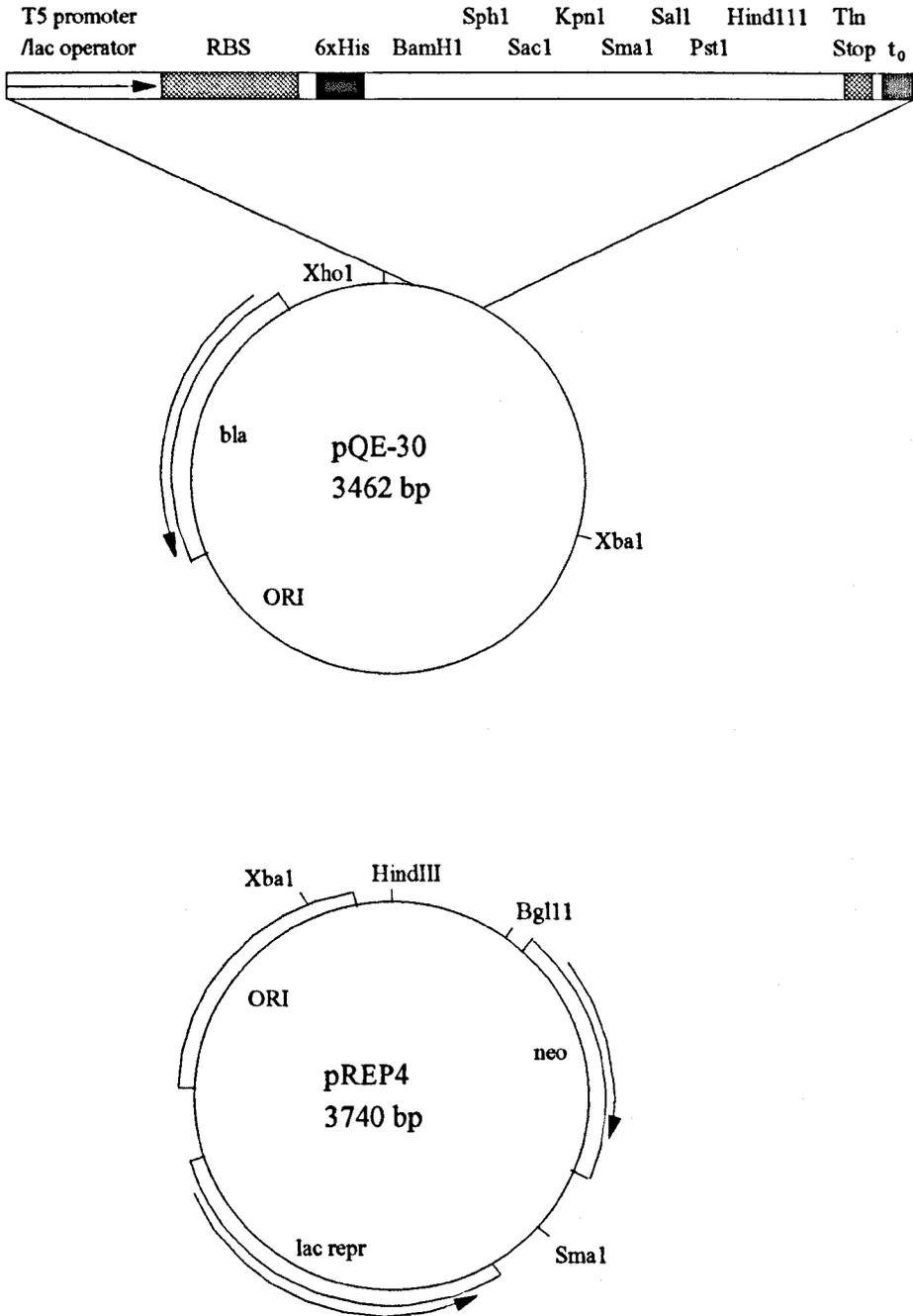
Figure 6.11. SDS-PAGE analysis of samples of 0.5 ml fractions from the nickel resin column used in the purification of the His/ORF1 fusion expressed from pID35. 1/25 th of each fraction was assayed on the gel. Visualised by coomassie blue staining. See Fig. 6.4. for marker details.

6.6. The expression of ORF1 using the QIAexpress expression system.

The QIAexpress expression system (Qiagen), like the Xpress system of Invitrogen, allows the purification of bacterially expressed proteins by exploiting the metal binding activity of six contiguous histidines, and like the Xpress system allows the purification of recombinant proteins under strong denaturing conditions. However, the plasmids used in the QIAexpress system are considerably different from the pTrcHis series of plasmids of Invitrogen. In the QIAexpress system two plasmids are used, pREP4 and pQE. The pQE plasmid (selected by ampicillin) is the vector which contains the expression cassette for the recombinant protein while the pREP4 plasmid (selected by kanamycin) encodes the *lac* repressor (see Fig. 6.12). The expression cassette of the pQE-30 plasmid contains: the *E. coli* phage T5 promoter and two *lac* operator sequences, ensuring a high level of protein expression on IPTG induction; a synthetic RBS; a 6 histidine affinity tag coding sequence; a MCS; translation stop codons in all three reading frames and the transcription terminator 't₀' from phage lambda. The presence of the strong phage T5 promoter requires the presence of high levels of the *lac* repressor protein to repress unwanted transcription prior to induction. Therefore the second, high copy number, plasmid pREP4 is used, which carries the *lacI* gene ensuring high levels of the *lac* repressor in the cell. The presence of a strong promoter made the pQE vector attractive for the expression of the ORF1 protein as it was clear from the Xpress system that reasonable purification was possible if adequate expression could be achieved. Also, if the levels of protein expression were to be increased by using the QIAexpress system then the purified heterologous protein product should be of a higher purity, due to the increased concentration of the ORF1 fusion in the cell lysate.

The ORF1 cDNA from pID22 was cloned into the pQE-30 vector using the *Bam*HI and *Eco*RI sites, ensuring the ORF1 sequence was in the same reading frame as the 6xHis peptide, to produce pID38. Bacteria containing pID38 and pREP4 were then grown in a 25 ml culture until A₆₀₀ 0.5, induced with 2 mM IPTG and 1 ml

Figure 6.12. The pQE-30 and pREP4 plasmids of the Qiagen QIAexpress bacterial expression system.



samples taken after 0, 1, 2 and 4 hours of induction at 37 °C. The 1 ml samples were processed and assayed as in section 6.5. The results of this assay can be seen in Fig. 6.13. It can be clearly seen from Fig. 6.11 and 6.13 that the expression of the His/ORF1 fusion protein is more efficient in the QIAexpress system with a distinct band of about 15K being visible in the induced samples.

The expression was then scaled up to 50mls and the His/ORF1 protein purified as in section 2.3.7.3. 1/100th of the 0.5 ml fractions eluted from the column were then assayed on an SDS-PAGE gel for the presence of the His/ORF1 fusion (Fig. 6.14). The results shown in Fig. 6.14 indicates that the His/ORF1 fusion protein eluted from the column in fractions 3-6 with a purity of greater than 90 % (estimated from the SDS-PAGE gel). The total amount of His/ORF1 protein purified, as quantified by a Bradford assay, was estimated at 1.6 mg, giving a yield of 32 mg/l. The QIAexpress system therefore allowed the purification of the His/ORF1 protein to a higher purity and in larger amounts than the Xpress system.

This preparation was subsequently used as an antigen to produce a mono-specific polyclonal antiserum in rabbits. The concentration of the His/ORF1 protein was 800 µg/ml in 8M urea. This was diluted 2 fold to give a 4M urea solution containing 400 µg His/ORF1 fusion/ml. 500 µl of this was then mixed with either complete or incomplete Freund's adjuvant (see section 2.3.8) and used as the inoculum for the generation of the polyclonal sera.

All subsequent studies using His/ORF1 fusion used this preparation of protein.

6.7. Attempts to detect ORF1 in infected cells.

The rabbit was inoculated with antigen at 3 week intervals until 112 days the first inoculation (see section 6.4.) and bled at 0, 43, 58, 84, 103, 117, 133 and 152 days post-first inoculum, to give serum samples pre, A, B, C, D, E, F and G respectively. The specific ORF1 reactivity of the serum samples was analysed by ELISA, immunofluorescence, immunoprecipitation and western blot analysis.

Figure 6.13.

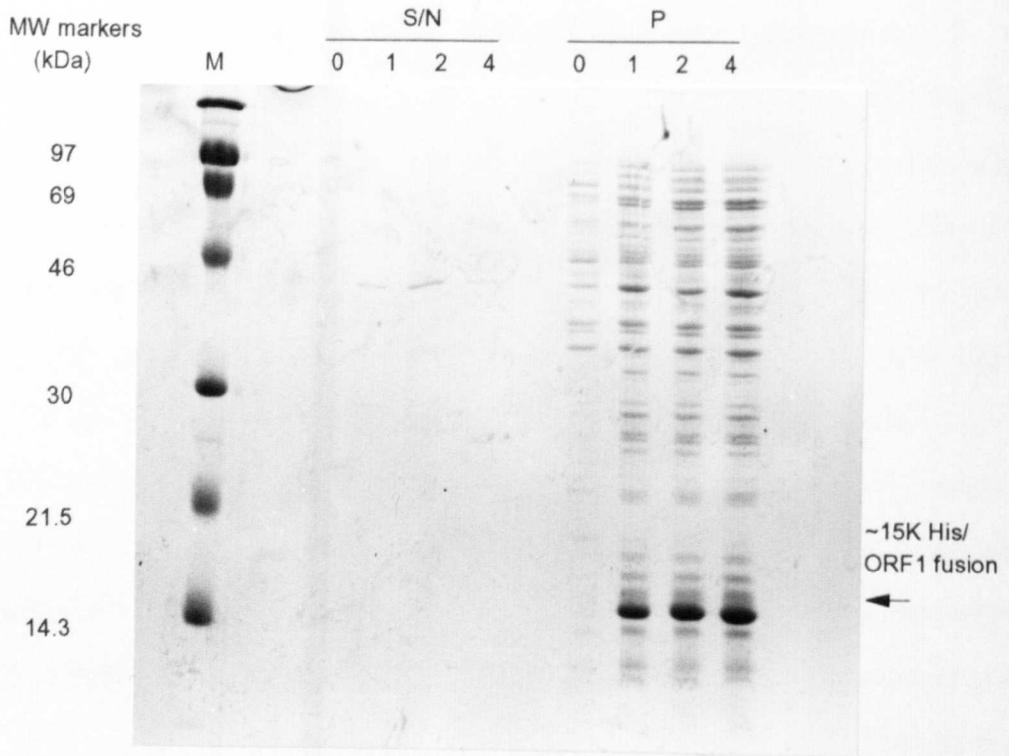


Figure 6.14.

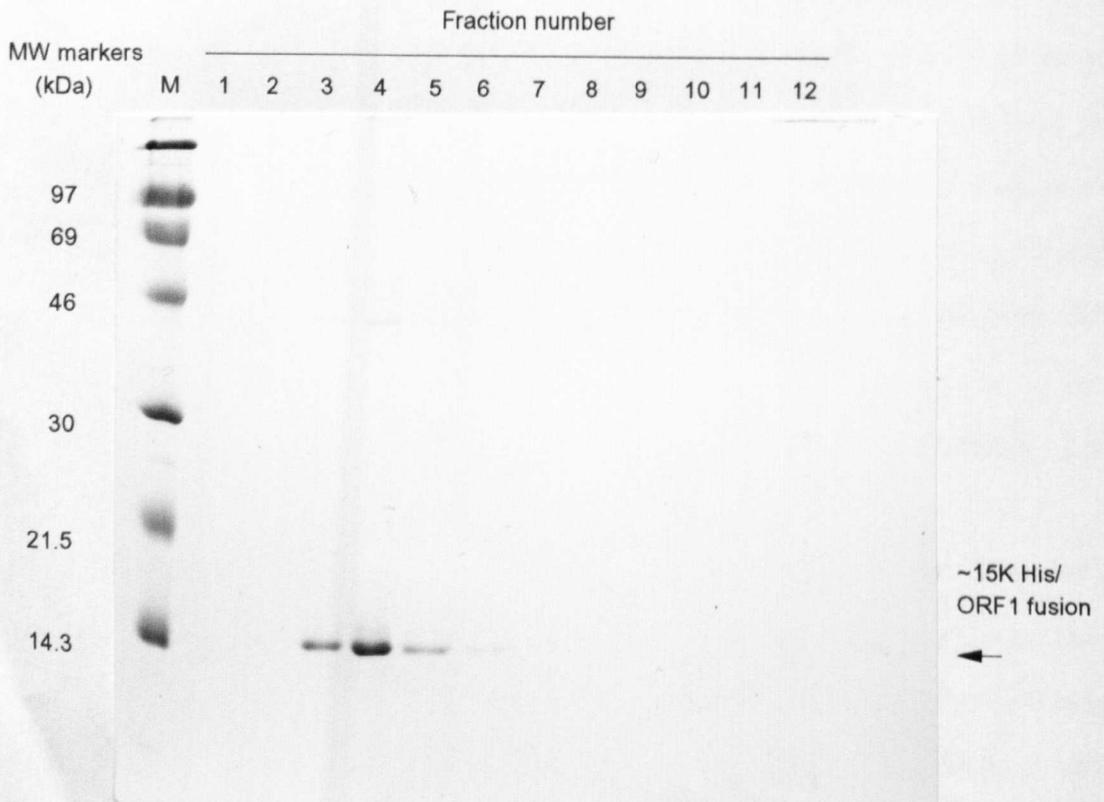


Figure 6.13. The expression of the His/ORF1 fusion at 37°C from pID38 assayed by SDS-PAGE. A 25 ml culture of bacteria containing pID38 and pREP4 was induced at A₆₀₀ 0.5 with 2 mM IPTG and samples taken at the times post induction (hr) indicated above the gel. Samples were lysed in PBS by three rounds of freeze/thawing and centrifuged at 16,000 x g for 10 minutes to separate the insoluble (P) and soluble (S/N) fractions. All samples are equivalent to 0.2 ml of culture. Visualised by coomassie blue staining. S/N = supernatant; P = pellet. See Fig. 6.4. for marker details.

Figure 6.14. SDS-PAGE analysis of samples of 0.5 ml fractions from the nickel resin column used in the purification of the His/ORF1 fusion expressed from pID38 (2.3.7.3). 1/100 th of each fraction was assayed on the gel. Visualised by coomassie blue staining. See Fig. 6.4. for marker details.

The results of the ELISA assays, using 0.5 µg of purified His/ORF1 protein per well, can be seen in Fig. 6.15. These results suggest that the specific antibody titres achieved are relatively low, with the peak specific titre being only 200 fold over the pre-immune titre (as compared to the GST/ORF2 antisera titre which is 4,000 fold greater than the pre-immune, chapter 7).

The Western blot analysis was carried out using 1/1000 dilution of each serum sample against an estimated (by Bradford analysis) 400 ng of the bacterially expressed His/ORF1 protein. The results of this analysis can be seen in Fig. 6.16. This analysis of the sera confirms the results of the ELISA; the titre increases over the first two samples and then reaches a plateau. The western blot analysis also indicates that the titre of the antibody is relatively low; similar analysis with an anti-ORF2 polyclonal serum only required 1/10th of the amount of antigen to produce the same intensity of band when carried out in parallel with this experiment (Fig. 7.7.).

No evidence of the ORF1 protein in virally infected HeLa cells could be found by immunoprecipitation analysis (Fig. 6.17). wt300- or dl1-3-infected HeLa cells (MOI of 10) were labelled for 3 hours (100 µCi ³⁵S-Met/10⁷ cells) and harvested at 24 hr p.i. The resulting RIPA buffer lysates were then immunoprecipitated using 4 µl of antiserum in the overnight step. This assay revealed a protein slightly smaller than 14K which was present in the wt300 infected cells and absent in the E4 mutant virus, dl1-3 (lacks ORFs 1-3). Further studies (data not shown) also suggested that this protein was made with late kinetics as expected for the ORF1 protein (the mRNA that encodes ORF1 is expressed with late kinetics (see chapter 4)). However this band was also present in pre-immune immunoprecipitation analysis (Fig. 6.17) and in cells infected with the ORF1 mutant dl351 (the band is faint in dl327 and dl351 as the labelling conditions were altered, see below) (Fig. 6.17.) indicating that this protein is not ORF1. It is unknown why this protein does not appear in the dl1-3 mutant.

Similar negative results were obtained for western blot analysis and indirect immunofluorescence assays with virally infected cells; no difference could be

Figure 6.15.

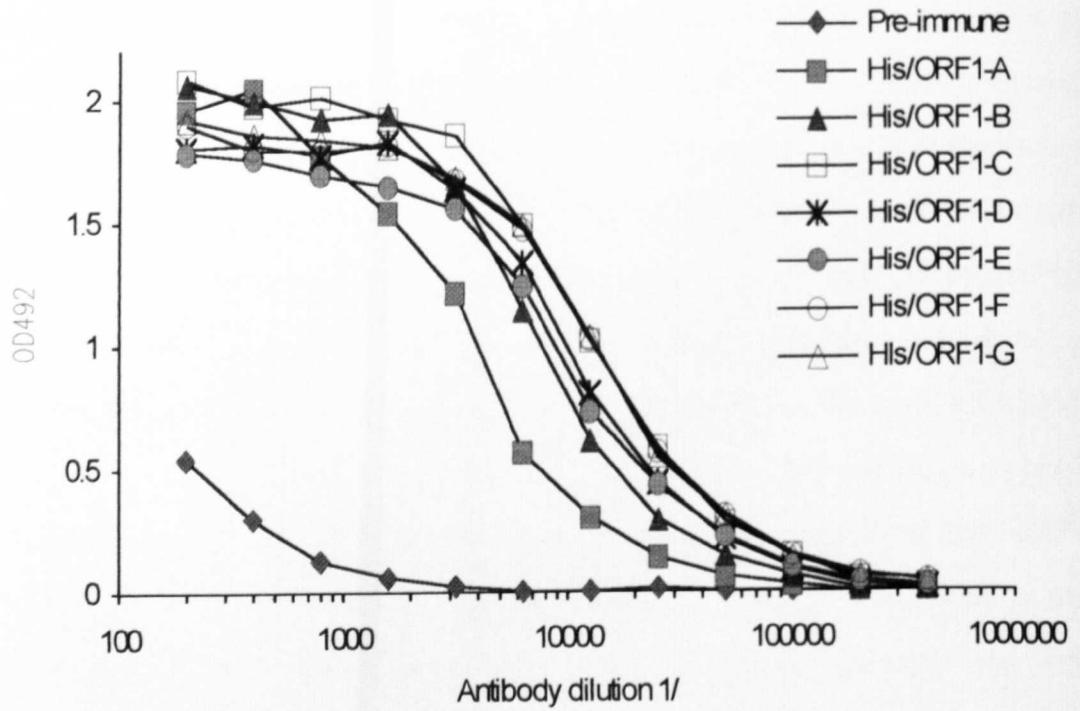


Figure 6.16.

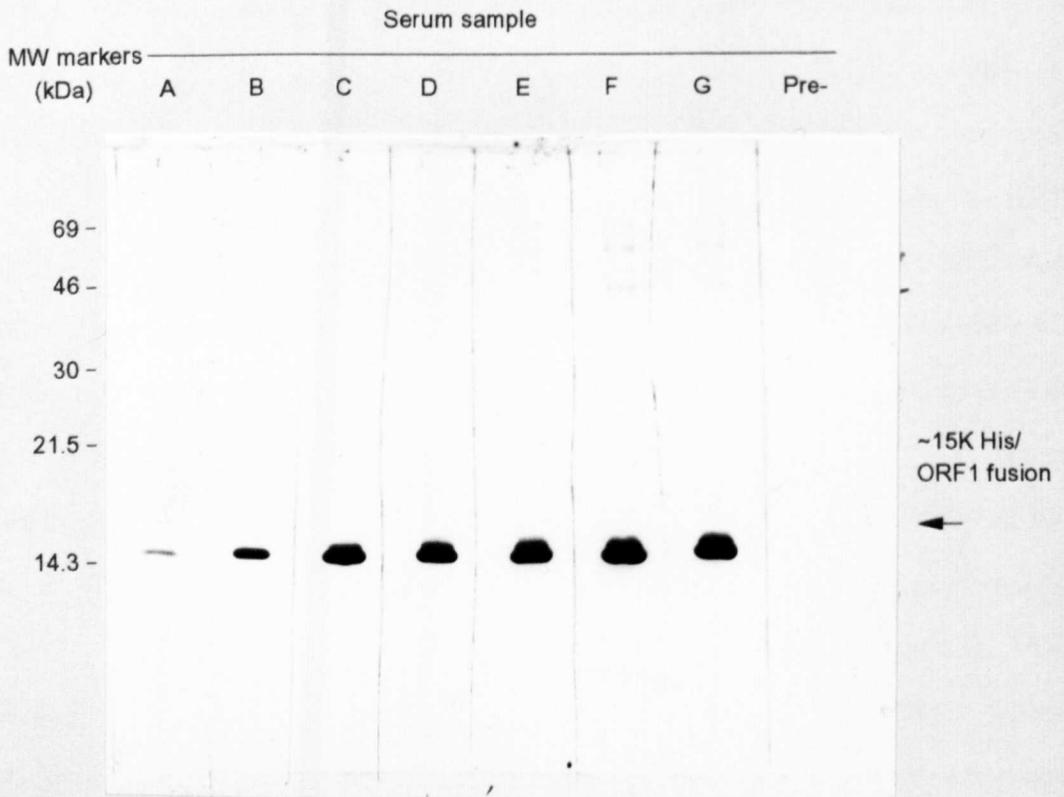


Figure 6.15. Titration of the anti-His/ORF1 activity in serum samples by ELISA (2.3.9.3). Serial dilutions of anti-His/ORF1 sera taken from successive bleeds from a rabbit inoculated with the His/ORF1 fusion protein, were tested against 0.5 μ g of His/ORF1 fusion protein. See text for details of serum samples A-G.

Figure 6.16. Western blot analysis employing sequential anti-His/ORF1 antisera to detect 400 ng of purified recombinant ORF1 (2.3.9.1). A 1/1000 dilution of each antiserum (A-G, pre-immune) was used.

ascertained between wild-type and dl1-3 infected cells or between pre-immune or immune antisera. Western blots were set up using 2.5×10^6 HeLa cells (infected and non-infected) lysed in RIPA buffer. A number of different detection conditions and antiserum concentrations were used in an attempt to detect the ORF1 protein but these were all unsuccessful (data not shown). Similarly, a number of different fixation techniques (acetone/methanol, formaldehyde/TX100, methanol) were used for immunofluorescence analysis, as were a number of different antibody combinations but no specific fluorescence was detected. Control experiments conducted in parallel, using antibodies specific for other antigens, for both these sets of experiments were successful indicating that the techniques employed were not fundamentally flawed.

There are a number of possible reasons why the ORF1 protein was not detected in these analyses. First, the antiserum may not be reactive against the antigen when presented in certain types of analysis. However this was not the case for western blot analysis (see Fig. 6.19.) and the evidence in Fig. 6.18. suggests that the His/ORF1 antiserum can successfully immunoprecipitate an *in vitro* transcribed/translated ORF1 protein. This product was produced by cloning the ORF1 cDNA of pID22 (as a *Sau3A1* - *EcoR1* fragment) into the pT74(pA) vector (a modified pGEM4 vector containing 30 adenylate residues at the SP6 end of the MCS), to create pID46. Radio-labelled ORF1 protein was expressed from 1 μ g of this clone using the Promega TnT coupled transcription-translation reticulocyte lysate system in the presence of ^{35}S -met. 1/33rd of the total *in vitro* translation was then immunoprecipitated with 2 μ l of either pre-immune serum or His/ORF1-D antiserum and assayed on an SDS-PAGE gel (Fig. 6.18.). It has not been demonstrated that the antiserum can recognise antigen in an immunofluorescence assay.

Second, the ORF1 protein may be post-translationally modified in such a way that the antiserum generated against the bacterially expressed ORF1 protein fails to recognise it. Although this possibility was not fully addressed by the use of modification inhibitors, when infected HeLa cells were labelled for only 20 min (400 μCi ^{35}S -met/ 10^7 cells) no ORF1 protein could be immunoprecipitated (Fig. 6.17).

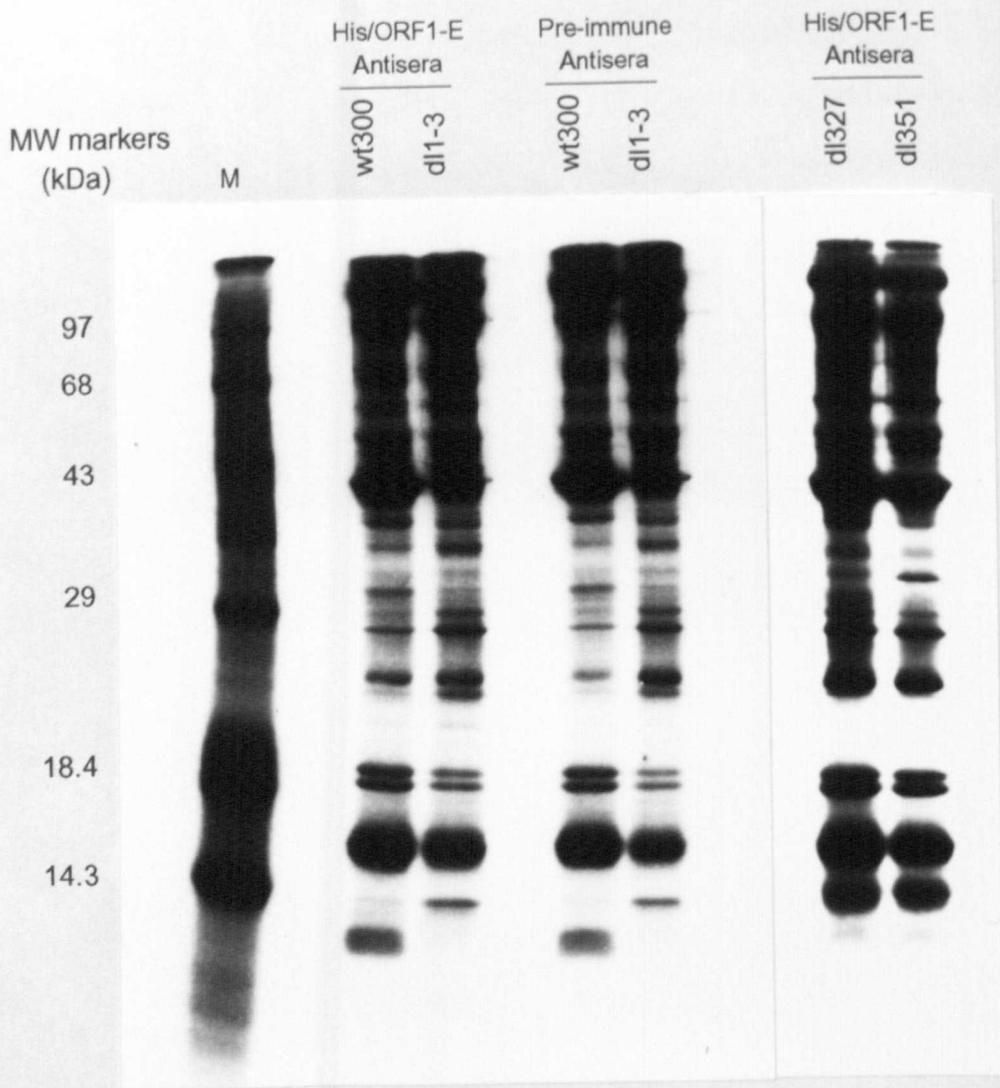
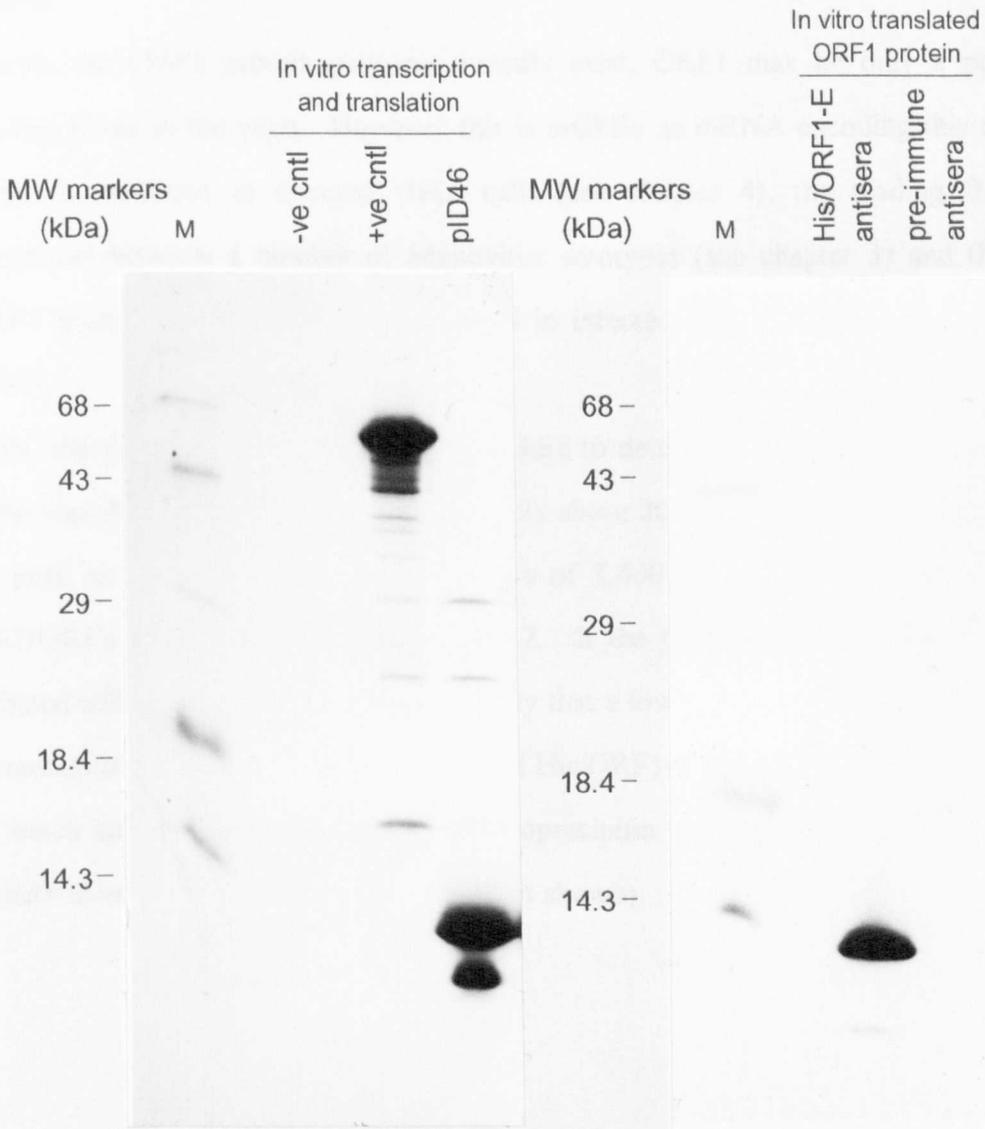


Figure 6.17. Detection of ORF1 in infected cells. wt300- and dl1-3-infected HeLa cells (MOI = 10) were labelled for 3 hours (100 μCi ^{35}S -met/ 10^7 cells) while dl327- and dl351-infected cells (MOI = 10) were labelled for 20 min. (400 μCi ^{35}S -met/ 10^7 cells) (2.3.10.1). Both were harvested into RIPA buffer 24 hours p.i. The lysates were subsequently analysed for the presence of the ORF1 protein by immunoprecipitation using 2 μl of the GST/ORF1-E or pre-immune sera (2.3.9.2.).

Markers (Gibco BRL):

- 97 kDa : Phosphorylase B
- 68 kDa : Serum Albumin
- 43 kDa : Ovalbumin
- 29 kDa : Carbonic Anhydrase
- 18.4 kDa : β -lactoglobulin
- 14.3 kDa : Lysozyme

Figure 6.18. Immunoprecipitation of an in vitro transcribed/translated ORF1 product. 1/50th of an in vitro transcription/translation of 1 μg of pID46 in the presence of ^{35}S -met using the Promega TnT in vitro transcription/translation system (2.3.11.1) was loaded onto a SDS-PAGE gel. Also shown is a negative control (no DNA template) and a positive control (the TnT control template, the luciferase gene). 1/33rd of the in vitro transcribed/translated ORF1 (from pID46) was immunoprecipitated using 2 μl of His/ORF1-E or pre-immune sera and also analysed by SDS-PAGE (2.3.9.2.). See Fig. 6.17. for marker details.

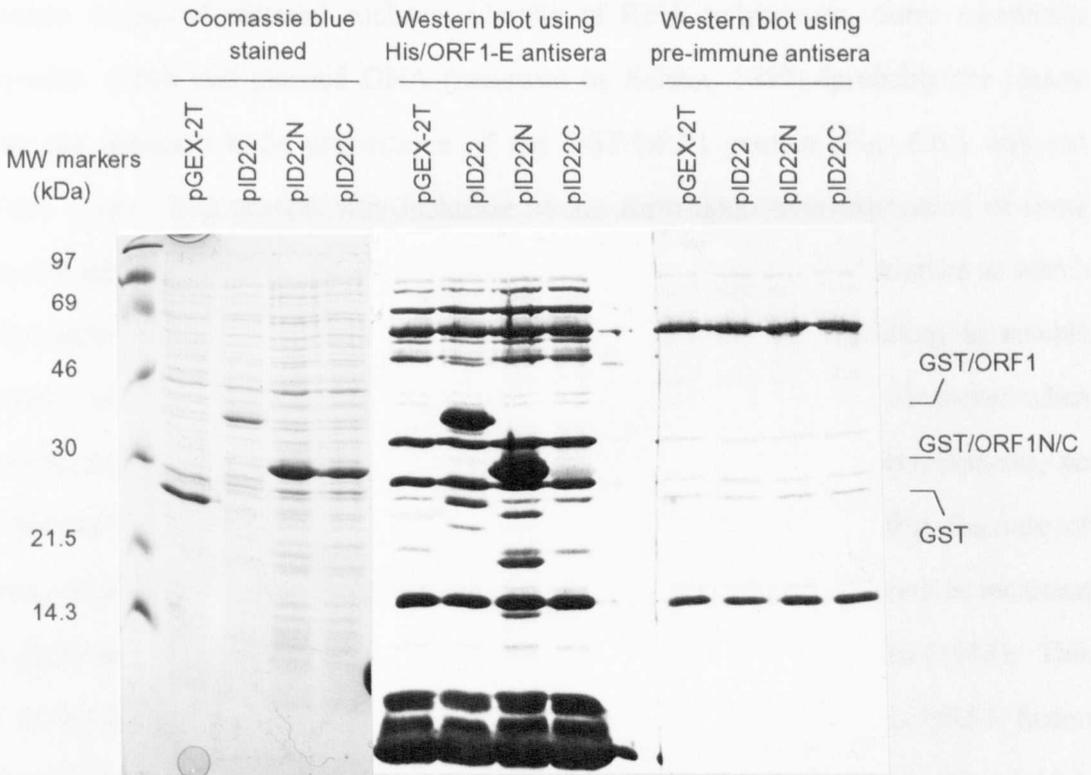


Third, the C terminal amino acid change in the recombinant ORF1 protein may be in the dominant antigenic site of the protein and so while the recombinant protein would be recognised by the antiserum, the wild-type viral ORF1 protein would not. However, western blot analysis using the His/ORF1 antiserum to detect the GST/N terminal half of ORF1 and C terminal half of ORF1, expressed separately, indicated that the immune response of the rabbit was directed largely against the N terminal half of the ORF1 protein (Fig. 6.19). This suggests that the C terminal amino acid change makes no difference in the ability of the antiserum to recognise the wild-type ORF1 protein.

Fourth, the ORF1 protein may not actually exist; ORF1 may be only a potential reading frame in the virus. However this is unlikely as mRNA encoding this reading frame is produced in infected HeLa cells (see chapter 4), the reading frame is conserved between a number of adenovirus serotypes (see chapter 3) and the Ad9 ORF1 protein has recently been identified in infected cells in tissue culture (Javier, 1994).

Fifth, and the most likely reason for the failure to detect the ORF1 protein in infected cells, was the low titre of the antiserum, only about 200-fold above the pre-immune at its peak as compared to an estimated titre of 3,400 above the pre-immune for the GST/ORF2 antisera produced in chapter 7. If the ORF1 protein is present in the infected cells at a low level then it is unlikely that a low titre antiserum would detect it. Immunoprecipitation analysis using 25 μ l of His/ORF1-E polyclonal antisera (~ 10 fold as much as usual) and 150 μ l of immunoprecipitin still failed to detect the ORF1 protein in adenovirus infected cells (data not shown).

Figure 6.19. Western blot analysis employing the His/ORF1 antisera. Bacteria containing pGEX-2T, pID22, pID22N and pID22C were induced at A_{600} 0.5 at 37°C for 3 hours. A sample of the culture was then lysed by repeated freeze/thawing and the equivalent of 50 μ l of the culture loaded onto an SDS-PAGE gel in triplicate. One set was stained with coomassie blue to estimate of the amount of protein loaded, while the other two sets were blotted and assayed with either a 1/1000 dilution of pre-immune serum or the His/ORF1-E antiserum (2.3.9.1.). See Fig. 6.4. for marker details.



6.8. Discussion

Attempts to identify the ORF1 protein in adenovirus infected cells failed, probably due to the poor titre of the polyclonal antibody obtained in both attempts to stimulate an immune response to bacterially expressed ORF1 protein.

Attempts to purify the ORF1 protein as a GST/ORF1 fusion protein using glutathione sepharose failed as the protein was insoluble, probably forming inclusion bodies (although this was never formally confirmed). Inclusion bodies are electron-dense particles that consist primarily of the recombinant protein. They vary in size and can contain additional material such as subunits of RNA polymerase, outer membrane proteins, rRNA and plasmid DNA (reviewed by Schein, 1989) (probably the reason why the inclusion body preparation of the GST/ORF1 protein (Fig. 6.6.) was not 100% pure). It is unclear why inclusion bodies form upon over-expression of some recombinant proteins in bacteria. It has been observed that the temperature at which expression occurs is important, with expression at 30°C often resulting in soluble protein, whereas at 37°C is insoluble (Schein and Noteborn, 1988). However when the expression of the pGEX/ORF1 fusion was attempted at lower temperatures, no difference in solubility was observed. It has also been postulated that the rate of production and the concentration of the recombinant protein is important in inclusion body formation, however this was not observed by Schein and Noteborn (1988). This is probably not the case for the ORF1 protein either, with the His/ORF1 fusion expressed at relatively low levels in the Xpress system still being insoluble (Fig. 6.10.). Schein (1989) also noted that the primary sequence of the expressed protein was possibly important in determining the solubility of a protein; stretches of acidic amino acids, low proline content and the ability to bind metal ions all correlate with the solubility of a protein. She also notes that stretches of hydrophobic amino acids could lead to increased aggregation and so to inclusion bodies (ORF1 is a hydrophobic protein (chapter 3)). Other plausible reasons for the formation of inclusion bodies in *E. coli* are first the lack of cellular compartmentation which causes the protein to be

produced in a reducing environment, preventing the formation of disulphide bonds, second the lack of mammalian post-translational modifying enzymes, and third the lack of proper foldases (molecular chaperones) during production since it is believed that the inclusion body forms when the heterologous protein cannot fold properly and so forms an insoluble complex with other proteins of the cell (reviewed by Schein, 1989). Using a protocol to purify inclusion bodies, the GST/ORF1 protein was purified to a degree and this was used to generate the first polyclonal antiserum. However the rabbit had to be put down before adequate titres of specific reactivity were achieved. It was decided not to inoculate another rabbit with the same immunogen as the immunogen contained a number of contaminating proteins.

The ORF1 protein was then successfully purified using the QIAexpress system. The resulting denatured His/ORF1 protein was used to generate a second polyclonal antiserum. However the titre of this antiserum was again relatively low and this also failed to identify the Ad5 ORF1 protein in infected cells.

It is not clear why the titre of the antibody generated against the His/ORF1 fusion was so low. For a protein to elicit a good immune response it must contain epitopes recognised by B and helper T cells. The binding of an antigen to a cell surface antibody of a virgin B cell is critical to the immune response as this will determine the binding site of the secreted antibodies produced by the differentiated B cell. One possible reason for the failure to generate a good immune response to the ORF1 protein is the previous development of B cell (or T cell, see below) tolerance to the antigen during the development of self tolerance. This may well be the case with the ORF1 protein as the ORF1 protein has good homology to a yeast protein (dUTPase) although the homology to a mammalian homologue (human) was weaker (see chapter 3). The appropriate B cells (and/or T cells) involved in the generation of an immune response to such a protein may have been eliminated during the development of self-tolerance, so rendering the animal unable to elicit an immune response to the homologous ORF1 protein.

The second property of the molecule that allows it to be a good immunogen is the ability to promote cell-to-cell communication between helper T cells and B cells (and helper T cells and antigen-presenting cells). The antigen, attached to the surface antibody of the B cell, is internalised and degraded within the B cell for presentation on the cell surface as peptides in conjunction with the MHC class II proteins. These B cell surface complexes are recognised by helper T cells via the interaction of T cell receptors and T cell epitopes on the peptides. If the antigen cannot be degraded within the B cell or the antigen does not contain a helper T cell epitope (possibly because of the elimination of the appropriate helper T cells during development) then no immune response can be generated.

Therefore, for a protein to be a good immunogen it must contain a B cell and T cell epitopes and must be degradable within a B cell. It has been suggested that for a protein to be a good immunogen it must be of more than 5kDa in size otherwise the chance of it containing both epitopes is low (Harlow and Lane, 1988). The His/ORF1 protein is only about 15kDa in size and it is conceivable that it may not contain a suitable T cell epitope. It is noted that the successful generation of the ORF2 antiserum (Chapter 7) was carried out using the much larger GST/ORF2 fusion, so overcoming any possible problems posed by the size of the 14K ORF2 protein. Similarly the titre of the antiserum raised against GST/ORF1 was increasing when the rabbit died. It may have been wise, in retrospect, to couple the ORF1 protein to a carrier protein which contained a good helper T cell epitope, such as KLH. Another possible reason why the animal failed to generate a good immune response to the His/ORF1 fusion could be the failure of MHC class II molecules to bind to the antigen. The presence of urea in the inoculated sample may have affected the immune response of the animal, although no visible effects were observed on the health of the rabbit. It was decided to inoculate the rabbit with protein in the presence of urea as other groups had experienced no difficulty with this procedure and had produced effective antisera to their proteins. Also only a small sample of protein was available, and so the risks of protein loss during dialysis due to precipitation seemed too great.

Chapter 7.

The E4 ORF2 protein

7.1. Introduction

The predicted 130 aa ORF2 protein is potentially encoded by just one E4 mRNA, mRNA C (Berk and Sharp, 1978; Tigges and Raskas, 1984; Virtanen et al., 1984). Studies of the expression of the E4 region indicate that mRNA C is produced early in the infection but only at a relatively low level (chapter 4). This would imply that the ORF2 protein is required initially early in the infection, before the onset of DNA replication. However attempts to identify a phenotype associated with a lack of the ORF2 protein have failed as virus lacking just the E4 ORF2 protein grows as well as wild-type virus in tissue culture (Halbert et al., 1985).

This chapter describes the generation of a mono-specific polyclonal antiserum to the ORF2 protein and the subsequent identification of the protein in the infected cell. The strategy employed in the generation of the polyclonal antiserum was identical to that used in the production of the initial ORF1 polyclonal antiserum, using the pGEX expression system and out-bred half-lop rabbits (chapter 6).

7.2. Viruses and cells used.

Four different viruses were used in this series of experiments; wt300, dl1-3, dl327 and dl352. For a description of wt300, dl1-3 and dl327 see section 6.2. dl352 contains a 4 bp insertion at the HindIII site (34930) causing the disruption of the E4 ORF2 reading frame; this lesion is present in the dl327 background. All infections were carried out at an MOI of 10. HeLa cells were used throughout this study.

7.3. Expression of the ORF2 protein using the pGEX vector

The pGEX expression system was selected for the bacterial expression of the ORF2 protein for the reasons indicated in chapter 6. A cDNA fragment suitable for the expression of the ORF2 protein in the pGEX-2T vector (Fig. 6.1.) was generated using

PCR. The template DNA used was the Ad5 wt300 genomic DNA and the two primers used were IDE and IDF (see section 2.1.2.). The 3' terminus of IDE contained 20 nucleotides complementary to the DNA at the 5' terminus of ORF2 reading frame and DNA immediately upstream. In the 5' terminus of IDE a *Bam*HI site was positioned so that when the resulting ORF2 cDNA was cloned into the pGEX-2T vector using the *Bam*HI site it was in frame with the GST gene reading frame. Similarly, the 3' terminus of IDF contained 16 nucleotides complementary to the DNA immediately downstream of the ORF2 reading frame. In the 5' terminus of IDF was an *Eco*RI site to facilitate the cloning of the ORF cDNA into the pGEX vector. The resulting 433bp ORF2 cDNA product (Fig. 7.1.) was cloned into the pGEX-2T vector using the *Bam*HI and *Eco*RI sites to create pID19. The sequence of the ORF2 cDNA and the ORF2 cDNA/GST junction was verified by directly sequencing pID19 using primers IDG (upstream of the MCS) and IDH (downstream of the MCS)(see section 2.1.2.).

To establish if pID19 was able to express the 358aa fusion protein, an expression study was undertaken where the length of induction at 37°C was varied and the expression at 30°C was examined. A 25 ml culture of pID19 was induced with IPTG and placed either at 30°C or 37°C. 2 ml samples were removed at various times post induction and prepared as in section 2.3.7.1.. The resulting pellet and supernatant fractions were then assayed for the presence of the GST/ORF2 fusion protein by SDS-PAGE (Fig. 7.2.). Also assayed on the gel was the purified GST/ORF2 product, from 2 ml of culture (purified from the 6 hour, 37°C, induction sample). In parallel, the expression and purification of GST (from pGEX-2T) was examined as a positive control for the expression of the fusion (Fig. 6.4b.) This experiment demonstrated that the GST/ORF2 fusion was relatively insoluble as compared with the GST protein at 37°C, although, unlike the GST/ORF1 fusion, a substantial amount of protein was purified from the soluble fraction on glutathione sepharose. However, expression for 20 hours at 30°C routinely resulted in the increased solubility of the GST/ORF2 protein (Schein and Noteborn, 1988). As a result, when scaling up the expression of

Figure 7.1. PCR amplification of the ORF2 cDNA using primers IDE and IDF (2.3.1.11.). All of the 433bp PCR product was run on a 1.2% agarose gel. See Fig. 5.3. for marker details

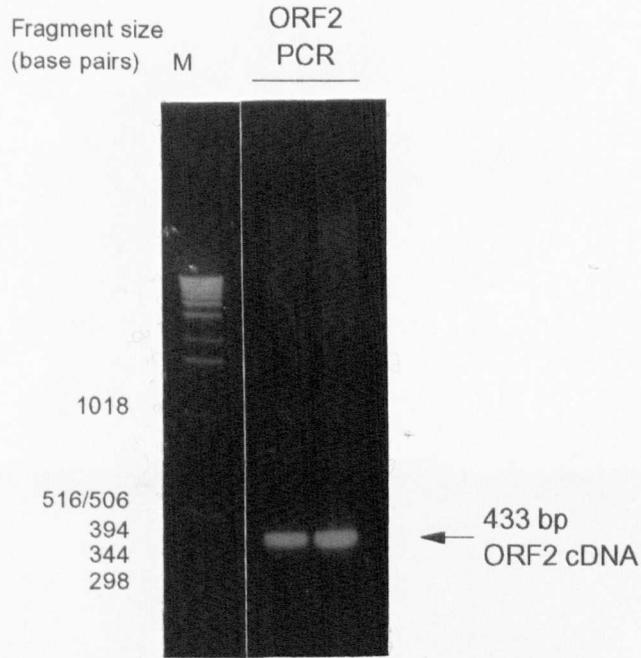


Figure 7.2. The expression of the GST/ORF2 fusion protein from pID22 at 30°C and 37°C (2.3.7.1.) assayed by SDS-PAGE. The times post induction (hrs) and the sample fraction are indicated above the gel. All samples are the equivalent of 0.1ml of the culture, except the purified fraction which is equivalent to 2 mls of culture. Proteins visualised using coomassie blue. S/N = supernatant; P = pellet; Pure = glutathione sepharose purified. See Fig. 6.4. for marker details.

Figure 7.3. SDS-PAGE analysis of samples of the 0.5 ml fractions from the glutathione-sepharose column used for the purification of the GST/ORF2 fusion expressed from pID19 (2.3.7.1.). 1/50th of each fraction and 1/50th of the glutathione-sepharose column was assayed on the gel. Proteins visualised with coomassie blue. See Fig. 6.4. for marker details.

GST/ORF2 to 500ml the expression conditions used were 30°C for 20 hours. This resulted in the expression of large amounts of GST/ORF2 protein that could be easily purified using glutathione sepharose. However only about 50% of the protein initially bound to the glutathione sepharose could be eluted off using 20 mM reduced glutathione (the 0.5 ml fractions 2 and 3 (Fig. 7.3.)). The overall yield from the purification was 1 mg of the GST/ORF2 protein, as estimated by Bradford assay (i.e. 2mg/l).

The 40K GST/ORF2 fusion protein, attached to the glutathione sepharose, could be cleaved into the 26K GST and 14K ORF2 proteins, using the protease thrombin (see section 2.3.7.1.). A range of thrombin concentrations were used (from 0.2% to 4.0% total protein) to cleave 10 µg of the fusion at 30°C for one hour (Fig. 7.4.); it was found that the cleavage of the fusion protein was equally efficient at both 30°C and 37°C (data not shown). For the complete cleavage of the fusion protein, at both temperatures, large quantities of thrombin were required, but if the incubation was left for 8 hr at 30 °C the fusion protein was cleaved to completion in the presence of just 0.5% thrombin (Fig. 7.5.). In this experiment the released ORF2 protein was further purified from GST and uncleaved fusion protein by separating the glutathione sepharose, with its bound protein, from the protein free in the supernatant by centrifugation at 16,000 x g for 5 min. (compare S/N and P fractions, Fig. 7.5.). It was found that only about 70% of the ORF2 protein was purified in the supernatant, the rest still being associated with the pellet. Subsequent washes of this pellet with PBS (W1-W3) failed to release any substantial amount of further ORF2 protein. The S/N fraction, which contained the bulk of the purified ORF2 protein was also contaminated with thrombin and uncut GST/ORF2, so it was decided to produce an ORF2 polyclonal antiserum using the uncleaved GST/ORF2 protein. This larger protein also had more chance of containing both B and helper T cell epitopes, important for immune recognition, than the ORF2 protein alone (see section 6.8.).

Figure 7.4. SDS-PAGE analysis of 10 µg of glutathione sepharose-bound GST/ORF2 fusion protein cleaved with thrombin (2.3.7.1). The percentage (wt/wt) thrombin used is indicated above the gel. All digests were carried out at 30°C for 1 hour. Proteins visualised with coomassie blue. For marker details see Fig. 6.4.

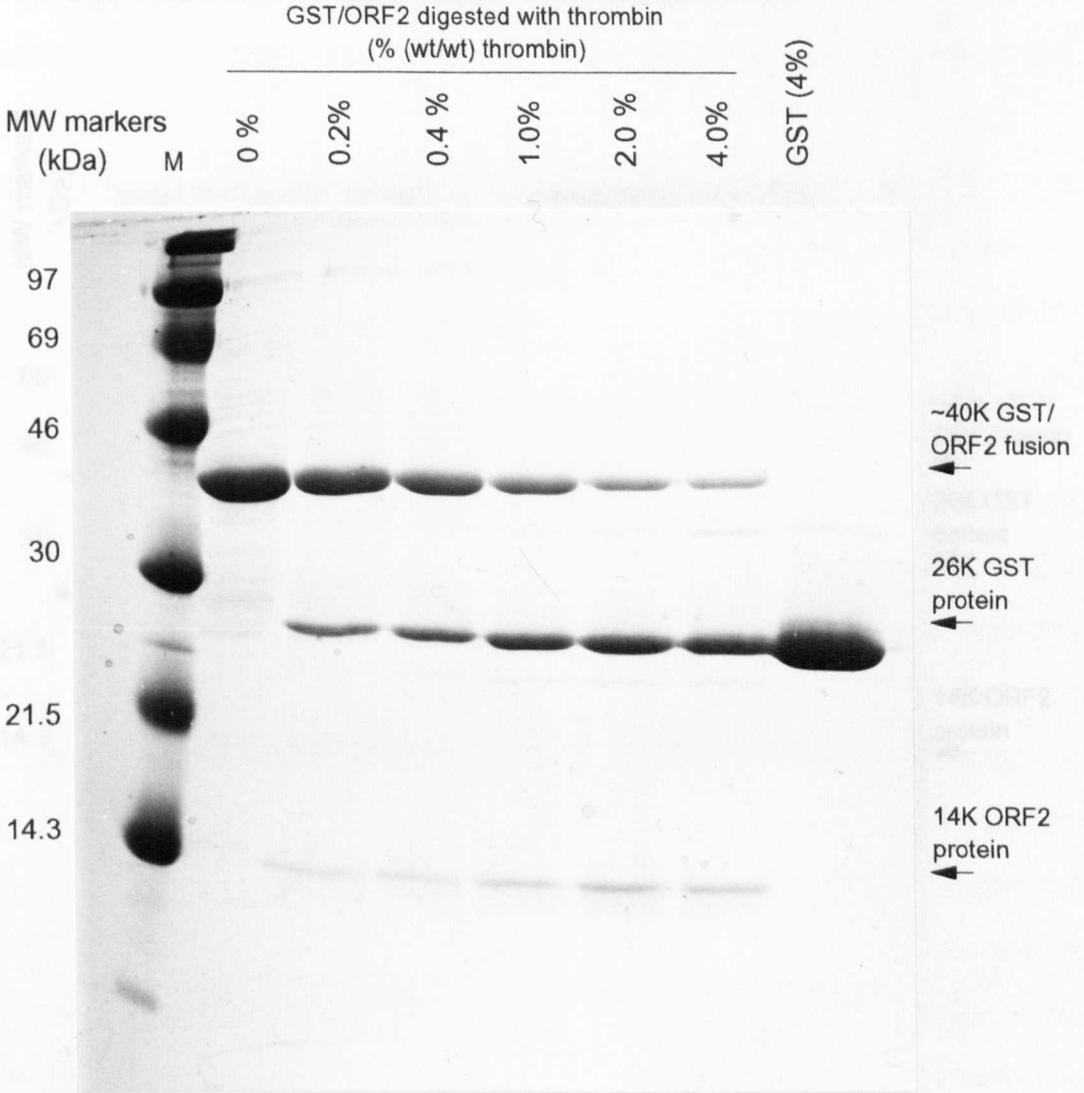
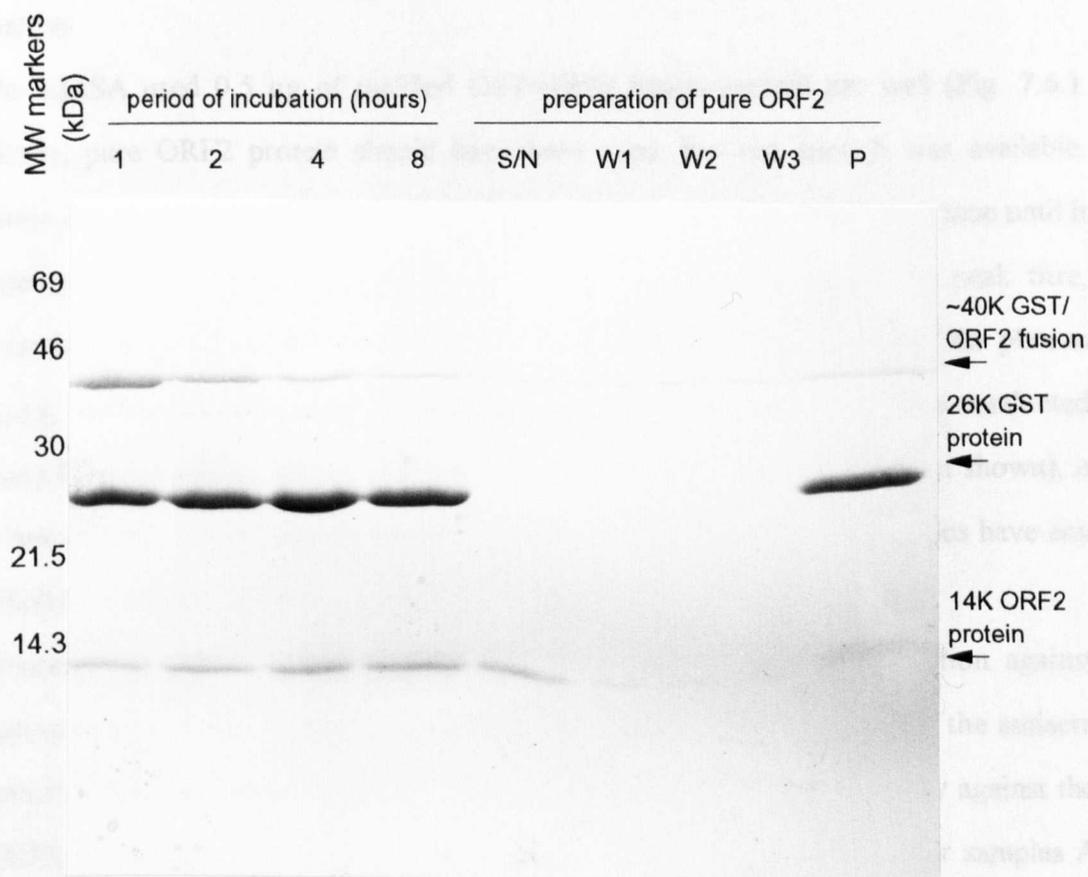


Figure 7.5. Purification of the ORF2 protein cleaved from the GST/ORF2 fusion by thrombin digestion. 10 μ g of glutathione sepharose-bound GST/ORF2 fusion protein cleaved with thrombin and assayed by SDS-PAGE analysis (2.3.7.1.). The length of incubation with thrombin is indicated above the gel. All digests were carried out at 30°C with 0.5% (wt/wt) thrombin. Also indicated are the supernatant fraction (S/N), the three PBS washes (W1-W3) and the resulting pellet fraction (P), all from the 8 hr digest (see text). Proteins visualised with coomassie blue.



7.4. The generation of an anti-ORF2 antiserum using the GST/ORF2 fusion protein.

A half-lop rabbit was subcutaneously inoculated four times with 200 µg of the GST/ORF2 fusion preparation in PBS and Freund's adjuvant over 58 days. The rabbit was bled seven times, 34, 58, 77, 94, 108, 128, and 124 days after the first inoculation, giving serum samples A, B, C, D, E, F, and G respectively which were stored as 0.5 ml aliquots at -20°C. These sera were analysed initially by ELISA and western blot analysis.

The ELISA used 0.5 µg of purified GST/ORF2 fusion protein per well (Fig. 7.6.). Ideally, pure ORF2 protein should have been used, but not enough was available. These data demonstrated that the GST/ORF2 antibody titre increased over time until it reached a maximum at about 110 days after the first inoculum. The peak titre, estimated by noting the dilution of the antibody when the A₄₉₂ was 50% the plateau A₄₉₂, was 3,400 times higher than the pre-immune titre. Since ELISAs conducted using GST as antigen gave a similar titre of antibody (Fig. 6.7. and data not shown), it is not clear from this study if any of the antibodies within the serum samples have any affinity for the ORF2 portion of the fusion protein.

Western blot analysis using all of the serum samples at a 1/1000 dilution against approximately 40 ng of the purified, cleaved ORF2 product, revealed that the antisera generated against the GST/ORF2 fusion protein did have specific reactivity against the ORF2 protein (Fig. 7.7.). The intensity of the ORF2 band increased over samples A and B but then reached a plateau, while the ELISA results indicated that the titre of the antibody carried on increasing until sample D. The difference in these results is probably due to the lack of sensitivity of the western blot with the ORF2 antibodies being in excess in samples B to G.

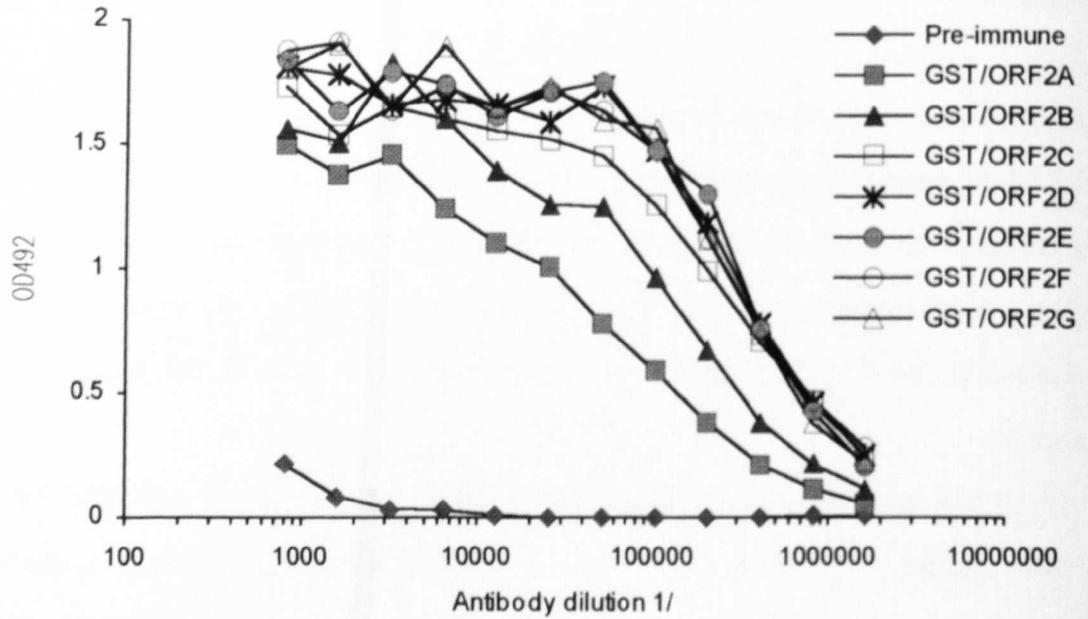
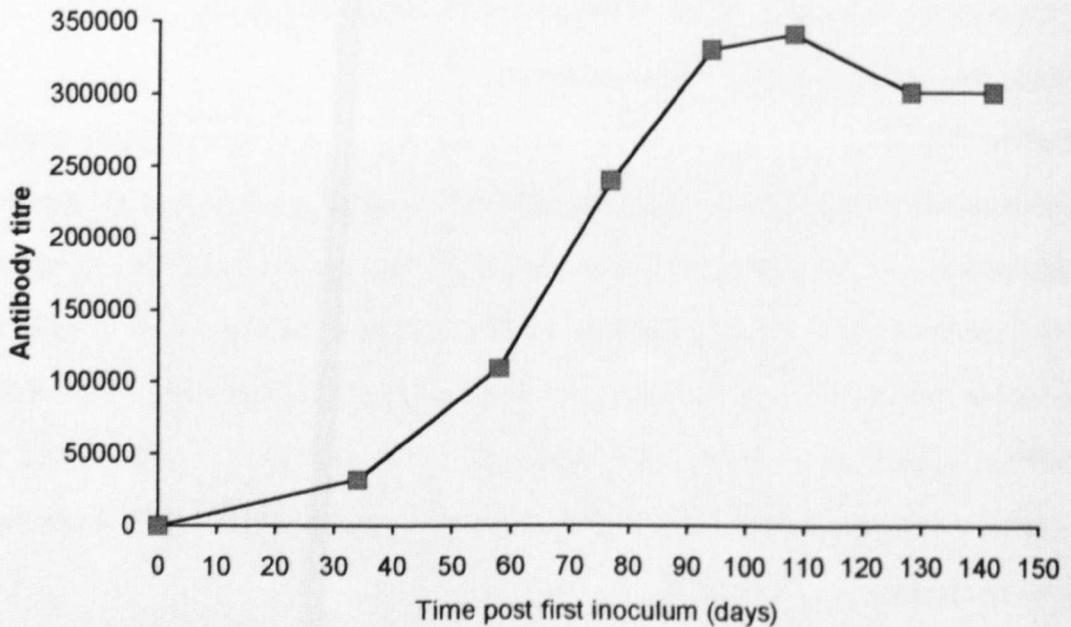
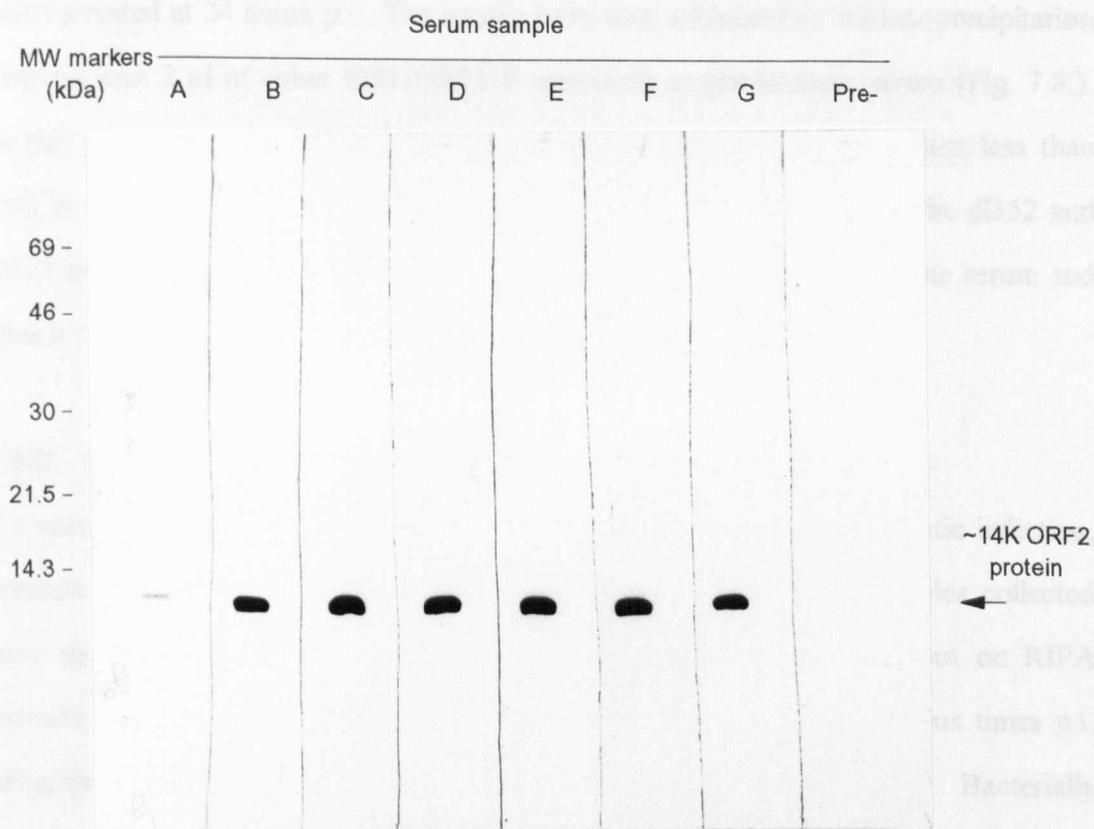
A**B**

Figure 7.6 Analysis of the GST/ORF2 antisera. Panel A. Titration of the anti-GST/ORF2 activity in serum samples by ELISA (2.3.9.3.). Serial dilutions of anti-GST/ORF2 sera taken in successive bleeds from a rabbit inoculated with the GST/ORF2 fusion preparation were tested against 0.5 μg of GST/ORF2 protein. See text for the details of the GST/ORF2 samples. **Panel B. Specific anti-GST/ORF2 antibody titres in serum, samples, as estimated by ELISA (Fig. 7.6a.).** The titre for a sample was taken as the dilution required to reduce the A_{492} to 50% of the plateau A_{492} .

Figure 7.7. Western blot analysis employing sequential anti-GST/ORF2 antisera to detect 40 ng of purified recombinant ORF2 (2.3.9.1). A 1/1000 dilution of each antiserum (A-G, pre-immune) was used.



...less than
 ...352 and
 ...serum and
 ...on RIPA
 ...times p.i.
 ...Bacterially
 ...positive control for the
 ... (Fig. 7.5).
 ...In this analysis, ORF2 protein was present at 9 hours p.i. in the
 ...cells, although at very low levels, and antibody subsequently increased
 ...immunoprecipitation
 ...hours of labelled WC00-
 ...cells revealed that the
 ...amounts (this is not visible
 ...between 16 and 24
 ...). No ORF2 protein was seen at 32 hours p.i., in contrast to the
 ...Western blot analysis

7.5. Expression of E4 ORF2 during a lytic infection.

7.5.1. The ORF2 protein exists in infected cells.

To show whether the ORF2 protein was expressed during a virus infection HeLa cells, infected with either wt300, dl352 or dl1-3 virus, were labelled with ^{35}S -methionine and harvested at 24 hours p.i. The lysates were then subjected to immunoprecipitation analysis with 2 μl of either GST/ORF2-E antiserum or pre-immune serum (Fig. 7.8.). In this experiment the GST/ORF2 antiserum precipitated a protein of just less than 14K in size that was only present in the wt300 extract, being absent in the dl352 and dl1-3 extracts. This ~14K protein was not precipitated by the pre-immune serum and thus it was concluded to be the E4 ORF2 protein.

7.5.2. The time course of expression of the ORF2 protein.

To determine the pattern of expression of the ORF2 protein during a lytic infection, western blot and immunoprecipitation analyses was performed on samples collected over the course of an infection. Western blot analysis was carried out on RIPA extracts of $\sim 2.5 \times 10^6$ wt300- or dl1-3-infected cells harvested at various times p.i. using the GST/ORF2-C antiserum and the pre-immune serum as a control. Bacterially expressed, purified ORF2 protein was also analysed as a positive control for the experiment (Fig. 7.9.). In this analysis, ORF2 protein was present at 9 hours p.i. in the wt300 infected cells, although at very low levels, and amounts subsequently increased to an approximately constant level from 16 hours p.i. onwards. Immunoprecipitation analysis, using 2 μl of the GST/ORF2-C antiserum, on RIPA lysates of labelled wt300- or dl352-infected HeLa cells harvested over the course of infection revealed that the ORF2 protein was synthesised at 8 hours p.i., albeit in small amounts (this is not visible in Fig. 7.10 as the band was very faint) with production peaking between 16 and 24 hours p.i. (Fig. 7.10.). No ORF2 protein was made at 32 hours p.i., in contrast to the levels of steady-state ORF2 protein as detected by Western blot analysis.

Figure 7.8. Detection of ORF2 in infected cells. wt300-, dl352- and dl1-3-infected HeLa cells (MOI = 10) were labelled for 3 hours (100 μ Ci 35 S-met/90mm dish) and then harvested in RIPA buffer at 24 hours p.i. (2.3.10.1.). The lysates were subsequently analysed for the presence of the ORF2 protein by immunoprecipitation using 2 μ l of the GST/ORF2-E or pre-immune antiserum (2.3.9.2. (N.B. the lysates were not cleared with pre-immune serum before immunoprecipitation)). See Fig. 6.17. for marker details.

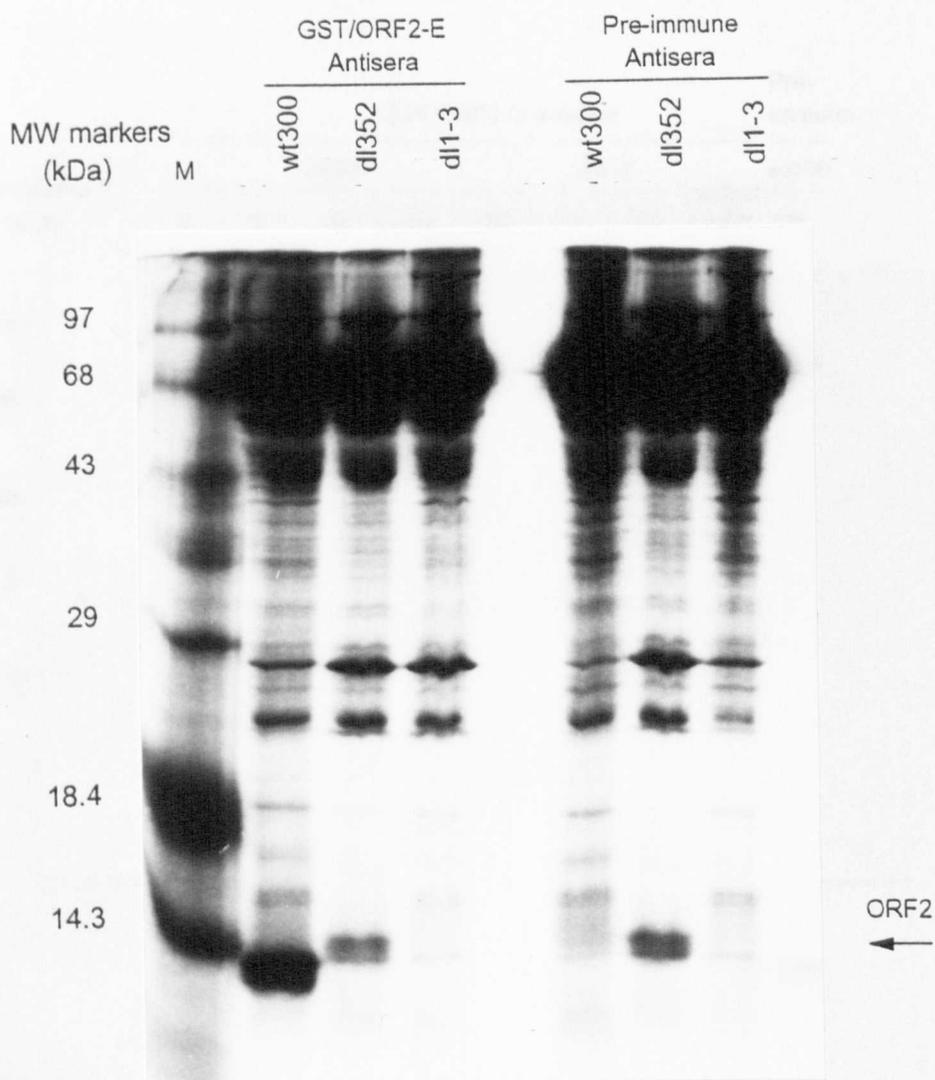


Figure 7.9. Steady state levels of the ORF2 protein during the course of an infection. The amount of the ORF2 protein in wt300- and dl1-3-infected cells was analysed by western blotting using a 1/1000 dilution of the GST/ORF2-C antiserum or pre-immune serum (2.3.9.1.). RIPA extract from 2.5×10^6 infected HeLa cells (MOI = 10) at each time point (indicated in hours p.i. above the blot) was loaded onto the SDS-PAGE gel. As a positive control, 40 ng of the purified ORF2 protein was also analysed (see Fig. 7.5.).

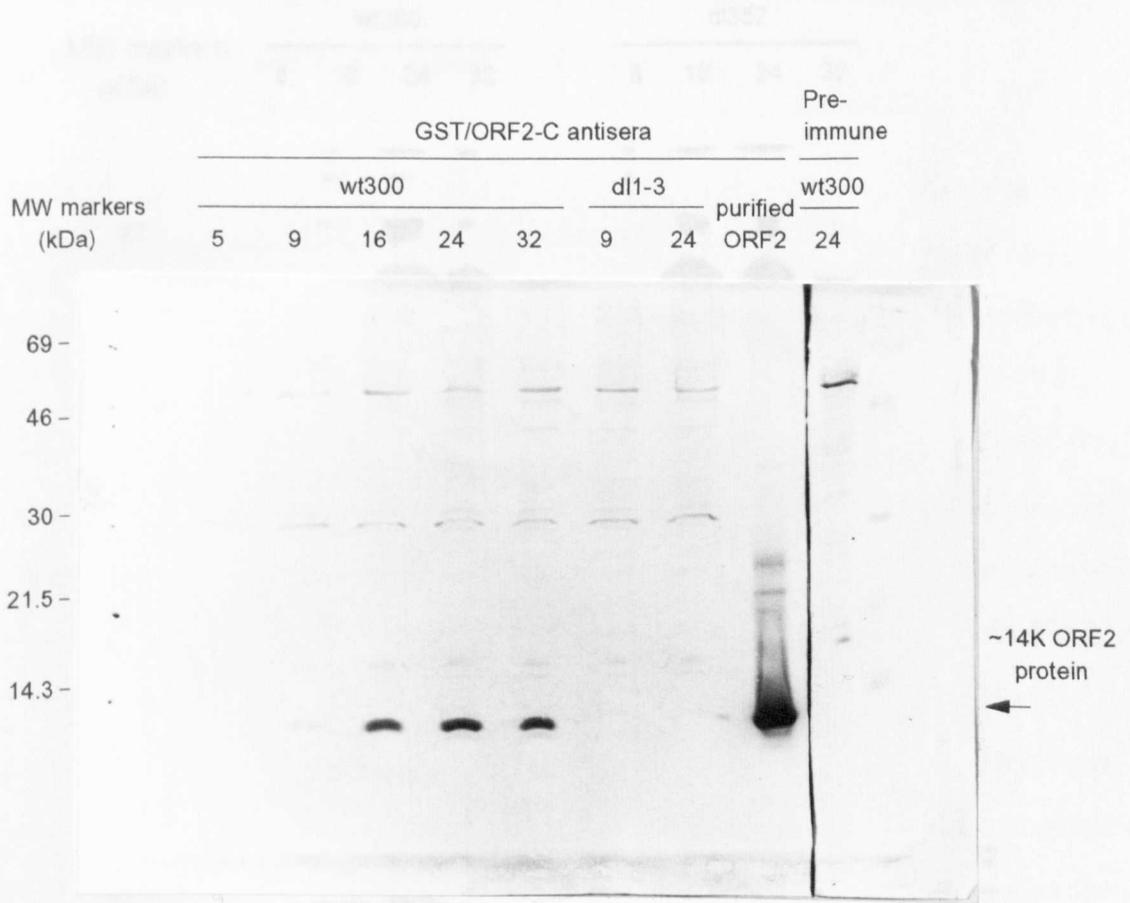
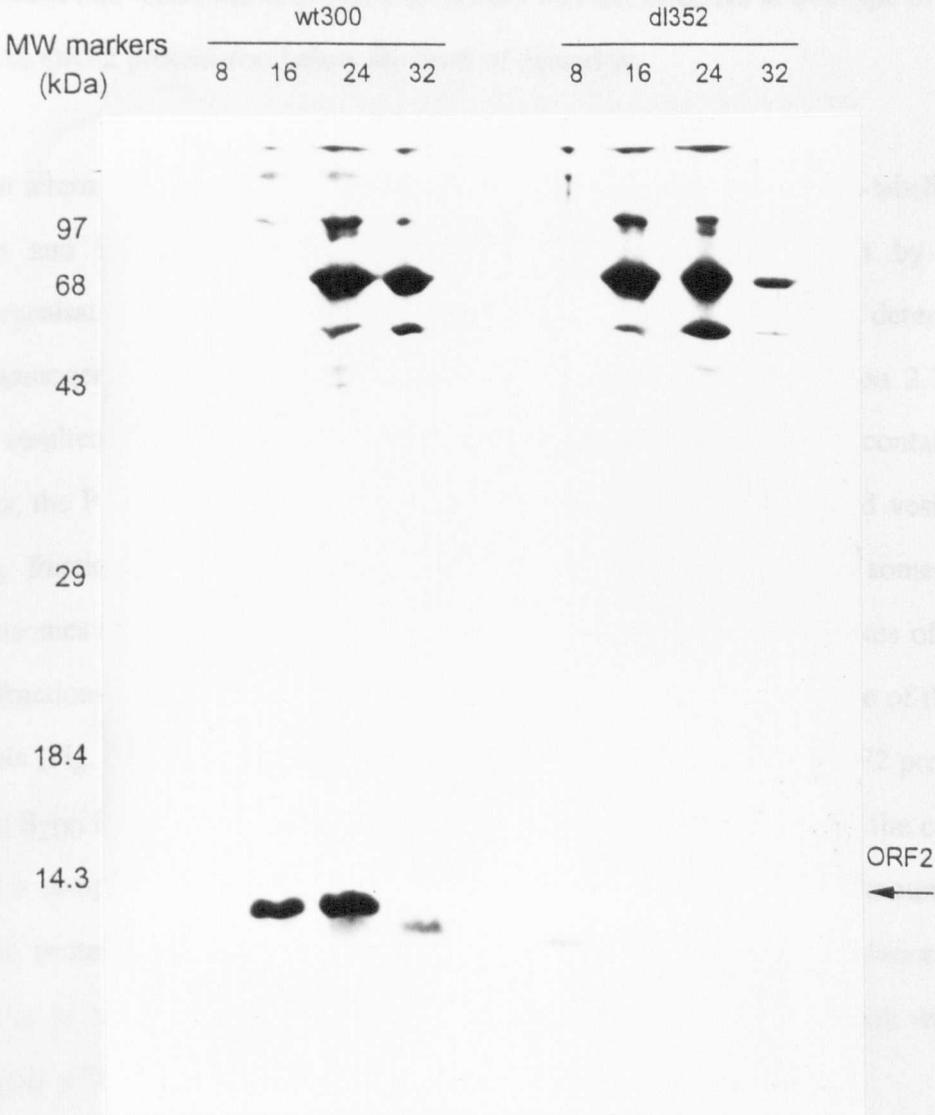


Figure 7.10. Synthesis of the ORF2 protein during the course of an infection.

wt300- and dl352-infected HeLa cells (moi = 10) were labelled for 1hr (100 μ Ci- 35 S-met/ 10^7 cells) (2.3.10.1.) and labelled ORF2 protein detected in RIPA buffer extracts prepared at various times p.i. (indicated in hours above the gel) by immunoprecipitation analysis, using 2 μ l of GST/ORF2-C antiserum (2.3.9.2.). See Fig. 6.17. for marker details.

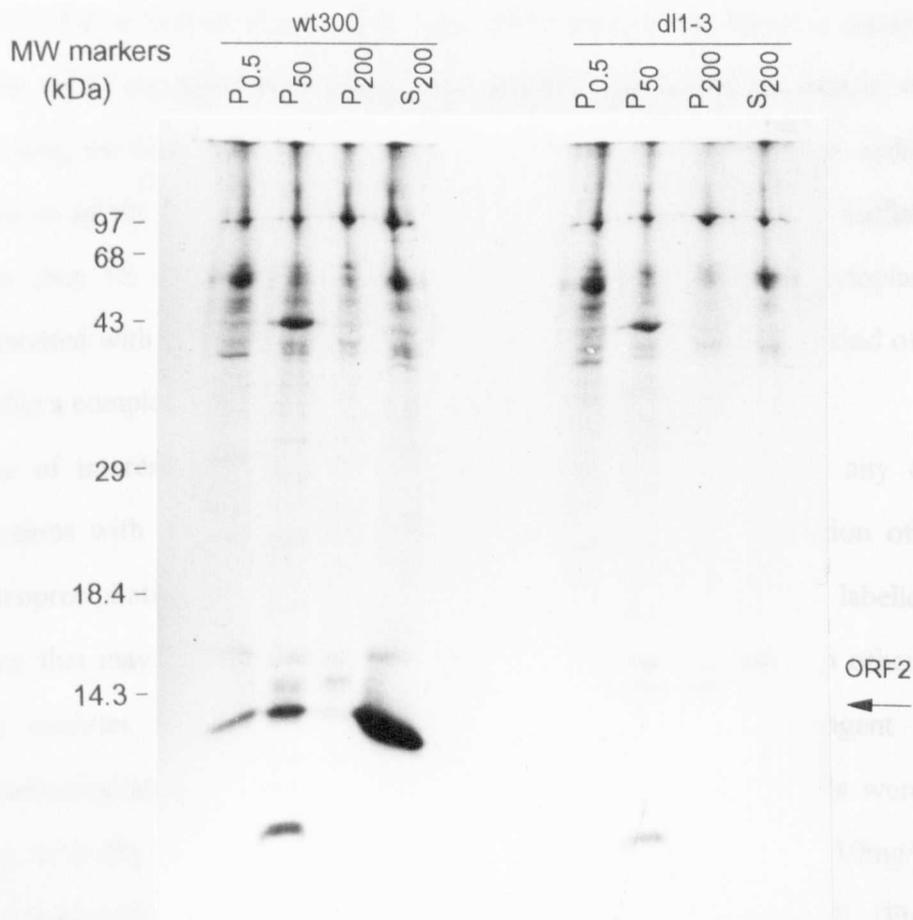


7.5.3. Subcellular localisation of the ORF2 protein

To determine which compartments of the infected cell contained the ORF2 protein, indirect immunofluorescence was attempted. However, in a series of experiments employing various fixation and cell permeabilisation techniques and different antibody sandwiches for detection, no specific fluorescence was detected (data not shown). Since controls conducted in parallel, using monoclonal antibodies against the E1b-55K protein and E2a-72K DBP, were successful in detecting their cognate antigens, it was concluded that either the anti-ORF2 antiserum was not effective in this type of assay or that the ORF2 protein was below the level of detection.

As an alternative approach, wt300- and dl1-3-infected cells were radio-labelled for 4 hours and harvested at 18 hr p.i. The cells were then broken by Dounce homogenisation following swelling in hypotonic buffer in the absence of detergent and the homogenates fractionated by sequential centrifugations (see section 2.3.10.3).. This resulted in a series of fractions; the P_{0.5} fraction (>236,600 S) containing the nuclei; the P₅₀ fraction (>1,113 S) containing membrane fragments and vesicles; the P₂₀₀ fraction (>59 S) consisting of polysomes, ribosomes and some smaller microsomes and the S₂₀₀ soluble fraction (also containing any complexes of <59 S). The fractions were then assayed by immunoprecipitation for the presence of the ORF2 protein (Fig. 7.11.). This experiment demonstrated that most of the ORF2 protein was in the S₂₀₀ fraction indicating that the ORF2 protein is either soluble in the cytoplasm or in a complex of less than 59 S. It is not clear whether the small amounts of the ORF2 protein in the P_{0.5} and P₅₀ fractions represent distinct populations of the protein or whether they result from contamination of these fractions with small amounts of the S₂₀₀ fraction.

Figure 7.11. The subcellular localisation of the ORF2 protein within the infected cell. wt300- and dl1-3-infected HeLa cells (MOI = 10) were labelled for 4 hours ($100 \mu\text{Ci } ^{35}\text{S-met}/10^7$ cells) and harvested at 18 hours p.i. The cells were lysed by Dounce homogenisation and fractionated by sequential centrifugation (2.3.10.3.). The ORF2 protein was then immunoprecipitated with the GST/ORF2-E antiserum (2.3.9.2.). P = pellet; S = supernatant. 0.5 = 500 x g, 50 = 50,000 x g, 200 = 200,000 x g. See Fig. 6.17. for marker details.



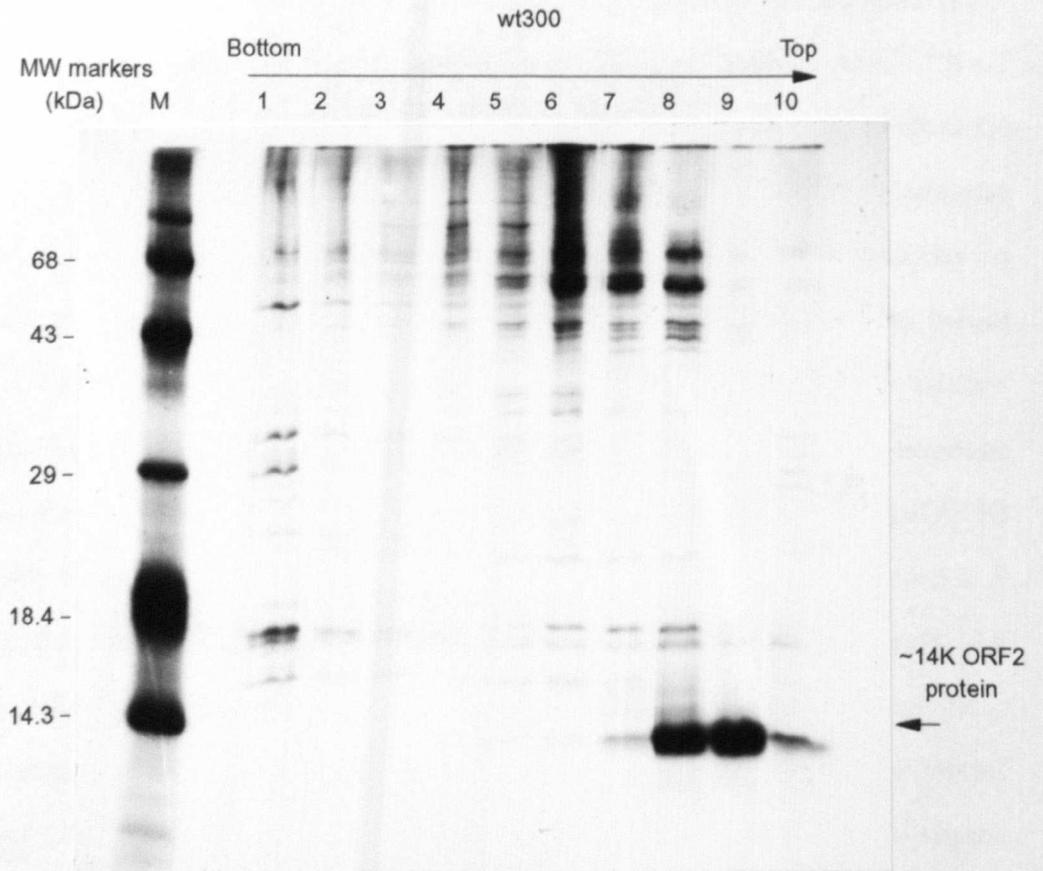
7.5.4. Interactions of the ORF2 protein with other infected cell components.

To determine whether the ORF2 protein was part of a complex of proteins when released from cells, radio-labelled wt300- and dl1-3-infected cell extracts were subjected to sucrose gradient analysis (2.3.10.4.). $S_{0.5}$ fractions prepared at 16 hours p.i. (similar analysis later in the infection is confused by a background 14K band) were loaded onto a 5 to 20% (wt/wt) sucrose gradient. The gradients were centrifuged at 240,000 x g for 6 hours, to allow the resolution of proteins from 2.6S to 16.7S, and collected as a series of 0.5 ml fractions which were then immunoprecipitated with GST/ORF2-E antiserum (Fig. 7.12.). The ORF2 protein was found in fractions 8 and 9, at the top of the gradient (fraction 10 consisted of the buffer the sample was loaded in). Using the tables and equations reviewed by Griffith (1975), the sedimentation coefficient of the ORF2 protein, immunoprecipitated in a hypotonic buffer, was no greater than 4S suggesting that it is present in the infected cell cytoplasm either unassociated with any other proteins (which could be detected by this kind of analysis) or within a complex of less than 4S in size.

It was of interest to determine whether the ORF2 protein made any detectable interactions with other infected cell components. However inspection of repeated immunoprecipitation analyses revealed no specifically co-precipitated, labelled protein species, that may have indicated molecular interactions of ORF2 with other proteins. These analyses were extended to include the use of less stringent lysis and immunoprecipitation buffers than the standard RIPA conditions; cells were lysed in hypotonic buffer, isotonic buffer or NP40 buffer (PBS, 2% NP40, 10mg/ml PMSF (immunoprecipitations washed in 0.5M LiCl, 0.1M Tris pH8.5)) (Fig. 7.13.). Similarly, no co-sedimentation of labelled species with the ORF2 protein was observed during sucrose gradient analysis (Fig. 7.12.).

Other analyses have been conducted with the ORF2 protein in an attempt to identify any protein-protein interactions it may be involved in the infected cell. One method was used previously to detect protein-protein interactions of the Rb¹⁰⁵ protein

Panel A



Panel B

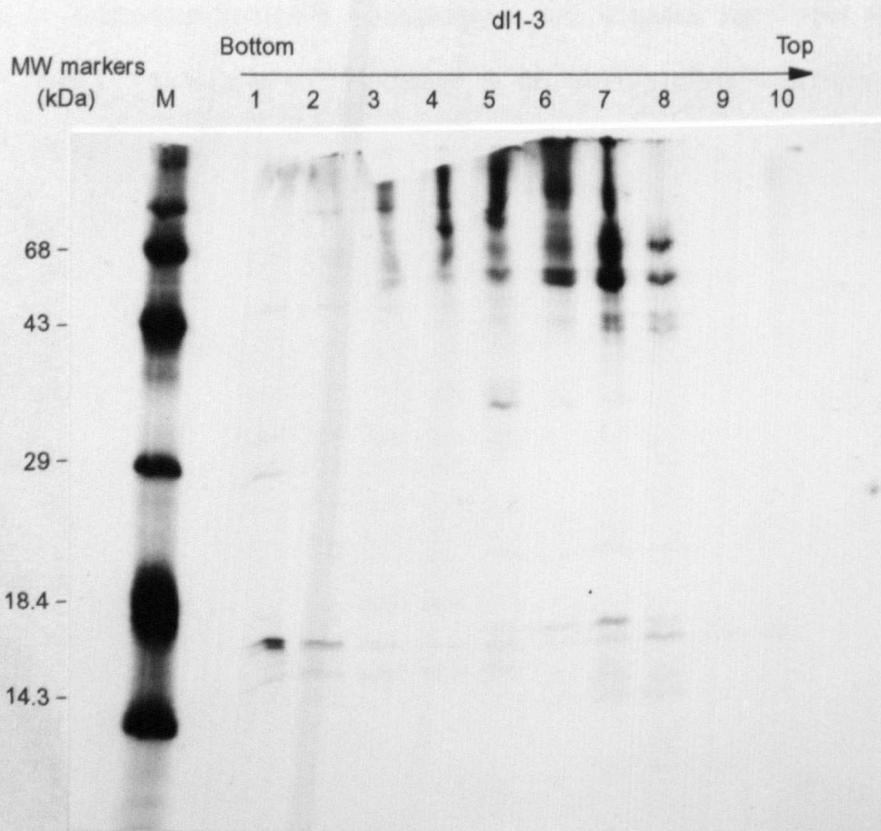
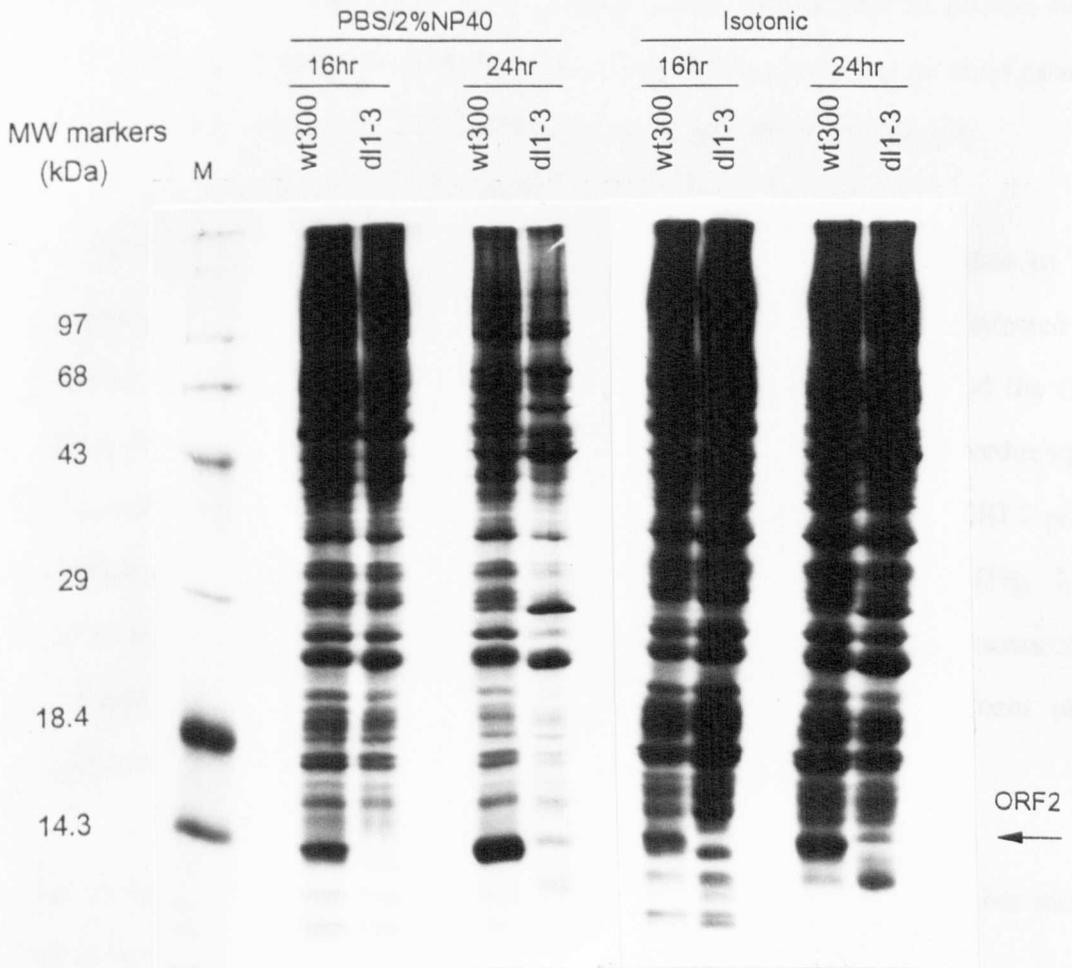


Figure 7.12. Determination of the size of the ORF2 protein by sucrose gradient analysis (2.3.10.4.). wt300 (Panel A)- and dl1-3 (Panel B)-infected HeLa cells (MOI = 10) were labelled for 3 hours (100 μ Ci 35 S-met/90mm dish) and harvested at 16 hours p.i. The cells were lysed by Dounce homogenisation and the supernatant of a 500 x g centrifugation loaded onto a 5 to 20% (wt/wt), 5 ml sucrose gradient formed in hypotonic lysis buffer. 0.5 ml fractions were then taken, and analysed for the presence of the ORF2 protein by immunoprecipitation with 2 μ l of the GST/ORF2-E antiserum (2.3.9.2.). See Fig. 6.17. for marker details.

Figure 7.13. Co-immunoprecipitation analysis to detect ORF2 protein-protein interaction. wt300- and dl1-3-infected HeLa cells were labelled for 3 hours (100 μ Ci 35 S-met/90 mm dish) and lysed with either isotonic buffer or PBS/2% NP40 at 16 and 24 hours p.i. (indicated above the gel) (2.3.10.1.). ORF2 protein and any associated proteins were immunoprecipitated using 2 μ l of the GST/ORF2-E antiserum, and only 2 washes with the appropriate buffer (see text)(2.3.9.2.). See Fig. 6.17. for marker details.



(Kaelin et al., 1991). This involved GST/ORF2 bound to glutathione sepharose as an affinity matrix for binding components of a labelled cell extract. The column was subsequently washed and the bound proteins resolved by SDS-PAGE. However, while Kaelin et al. (1991) managed to get relatively clean results, experiments with the GST/ORF2 fusion, conducted as reported, resulted in a high background of proteins that were present in both GST- and GST/ORF2-bound fractions, making it difficult to identify any protein-protein interactions with certainty (data not shown). Attempts to change the labelling conditions, wash conditions, and the amount of protein on the column made no difference to the high background observed. No proteins could be identified in these analyses which specifically bound the GST/ORF2 fusion.

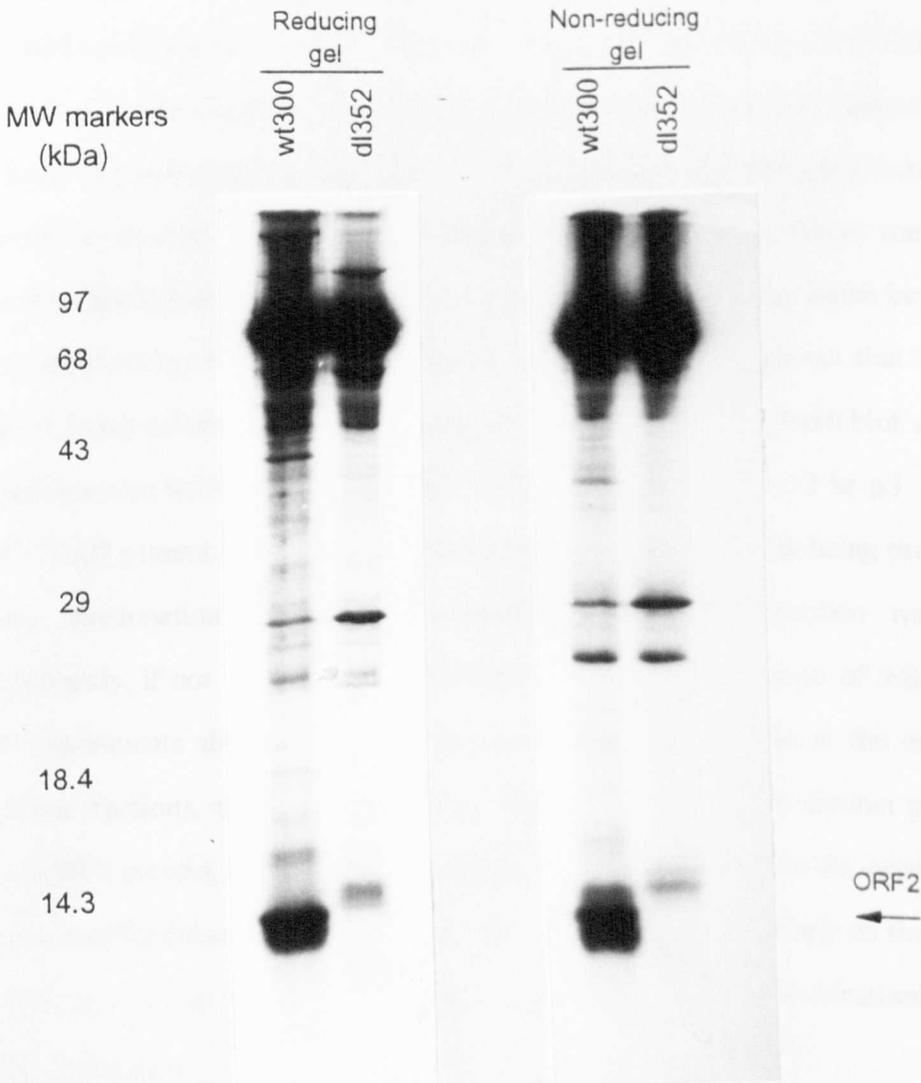
The ORF2 protein contains 2 cysteine residues which have the potential to form disulphide bridges with either each other or other proteins within the infected cell. However when labelled cells were lysed in isotonic buffer (18hrs p.i.), and the ORF2 protein immunoprecipitated with GST/ORF2-E, and assayed on a non-reducing gel (DTT omitted from sample buffer) no difference in the migration of the ORF2 protein was observed, suggesting that it has no disulphide bridge interactions (Fig. 7.14.). This result is not surprising as data from cellular fractionation experiments above suggested that the ORF2 protein is present in the reducing environment of the cytoplasm where the formation of disulphide bonds is unfavourable.

These results suggest that the ORF2 protein does not interact with any other infected cell components, at least in a way which is detectable by these analyses.

7.6. Discussion

An ORF2 antiserum was successfully generated using the GST/ORF2 fusion protein which was purified using one-step affinity chromatography with glutathione

Figure 7.14. Immunoprecipitation analysis of the ORF2 protein under reducing and non-reducing conditions. wt300- and dl352-infected cells (MOI = 10) were labelled for 3 hours (100 μ Ci 35 S-met/90mm dish) and harvested at 18 hours p.i. (2.3.10.1.). The ORF2 protein was then immunoprecipitated with 2 μ l of the GST/ORF2-E antiserum (2.3.9.2.). The samples were either boiled in reducing buffer prior to loading on the SDS-PAGE gel as normal or loaded onto the gel in non-reducing buffer (i.e. no DTT) without being boiled (2.3.6.1.). See Fig. 6.17. for marker details.



sepharose. The titre of this antiserum increased to a plateau over 110 days following the first immunisation. This antiserum recognised the ORF2 protein both in the bacterially expressed, recombinant form and in virus infected cell extracts.

The experiments reported here using this antiserum demonstrate the presence of a novel gene product, E4 ORF2, in Ad5-infected cells that had been recognised previously only as a potentially translatable open reading frame present in E4 mRNA C. The ORF1/ORF2 or ORF1/ORF2/ORF5 fusion proteins predicted by Hérissé et al. (1981) were not detected, confirming the results of the E4 mRNA studies (Freyer et al., 1984; Tigges and Raskas, 1984; Virtanen et al., 1984) that these proteins do not exist in adenovirus-infected cells. The time course of expression of the ORF2 protein correlates with that observed previously for E4 mRNA C (chapter 4), with the protein appearing with early kinetics (Fig. 7.9. and 7.10.). However, while the protein is being expressed at 16 and 24 hours p.i., as assessed by immunoprecipitation analysis (Fig. 7.10.), the steady-state levels of the ORF2 protein do not increase much beyond 16hr p.i., as assessed by western blot analysis (Fig. 7.9.). This suggests that the ORF2 protein is being turned over, or sequestered from detection by western blot analysis, at a significant rate between 16 hr p.i. and 24 hr p.i. However at 32 hr p.i. the ORF2 protein is still present in the infected cell although no new protein is being produced.

Cellular fractionation experiments revealed that the ORF2 protein was located predominantly, if not exclusively, in the soluble cytoplasmic fraction of infected cells. These experiments also revealed low levels of the ORF2 protein in the nuclear and membrane fractions, although it is unclear whether these represent distinct populations of the ORF2 protein. If distinct populations of the ORF2 protein do exist they may reflect a multifunctional aspect of the protein, although this is unlikely as the protein is only 14K in size. Alternatively they may reflect functional and non-functional forms of the ORF2 protein.

As mentioned, no function has yet been ascribed to the ORF2 protein. In an attempt to elucidate a function, it was of interest to identify any proteins within the infected cell, whether viral or cellular, which interacted with the ORF2 protein. However, no such

proteins were detected suggesting that either the ORF2 protein does not closely interact, or interacts only transiently, with any other infected cell components or the methods employed were not sensitive enough. It may be that only a small subset of the cytoplasmic ORF2 may be active and making interactions, putting these interactions below detection limits. Alternatively the anti-ORF2 polyclonal antisera may have affinity for only a limited number of epitopes on the ORF2 protein so possibly rendering the ORF2 protein within a complex undetectable as these epitopes could be masked. If this were the case only ORF2 protein not in a complex would be detected. However, this is unlikely as the antiserum generated was polyclonal and should not be directed to just one epitope of the ORF2 protein, although the polyclonal antiserum generated to the ORF1 protein, which is of a similar size to ORF2, was directed predominantly against the N-terminus of the protein suggesting that, for the ORF1 antisera at least, the antisera may not be truly polyclonal.

One possible explanation of these results is that the ORF2 protein acts as a cytoplasmic enzyme, interacting with its substrate (if it is a protein) for only a short period of time or only loosely interacting with it.

Chapter 8.

The function of the ORF1 and ORF2 proteins

8.1. Introduction.

No function has yet been assigned to either the ORF1 or ORF2 proteins of Ad5. Mutant strains of Ad5 unable to make ORF1 and/or ORF2 grow to virtually wild-type levels in HeLa cells: the dl351 and dl352 mutants, constructed by Halbert et al. (1985) using an E3-deleted virus background, which lack intact ORF1 and ORF2 respectively, are not compromised in their growth in HeLa cells. Similar results were obtained by Bridge and Ketner (1989) and Huang and Hearing (1989a) using viruses lacking the entire ORF1 and ORF2 reading frames (dl1005 and dl1-3 respectively). Both of these mutants displayed normal late protein synthesis and viral DNA accumulation and the dl1005 mutant also displayed a normal plaquing efficiency (the plaquing efficiency of dl1-3 was not investigated).

However, as discussed in chapter 3, these proteins, being conserved amongst the adenovirus serotypes (with the exception of ORF1 in Ad40), almost certainly have a function in the adenovirus infection. Adenovirus 5 normally infects quiescent, non-transformed epithelial cells of the upper respiratory tract and other tissues and not transformed, aneuploid HeLa cells, derived from human cervical carcinoma tissue, which have been used in all the tissue culture studies of the Ad5 ORF1 and ORF2 mutants. The ORF1 and ORF2 proteins may therefore have important functions in the growth of the virus in the host which are not observed in HeLa cells in tissue culture. As examining wild-type and E4 mutant adenovirus 5 replication in the natural host (human) was impossible, it was decided to explore the role of these proteins in infections of non-transformed human epithelial cells in tissue culture. This and other studies of the E4 mutant viruses will be discussed in this chapter.

8.2. The growth of adenovirus in non-transformed cells

Four different human non-transformed cell types were selected to observe the growth of adenovirus; MRC5 (human embryonic lung cells obtained from Prof. C.R. Pringle), WI-38 (human diploid lung cells), HUV-EC-C (human umbilical cord, endothelial cells) and HISM (human intestinal smooth muscle cells), all obtained from the American Type Tissue Culture Collection. Each cell type, along with HeLa cells as a control, was cultured on either 60 mm or 90 mm dishes and infected at an MOI of 5 with either wt300 or dl1-3, at either 30% confluence or after reaching 100% confluence for 24 hours. Although not a rigorous control of cell growth, it was presumed that the cells in the 30% confluent monolayer were actively dividing while the cells of the confluent monolayer had ceased dividing due to cell contact inhibition. The cells and tissue culture fluid were harvested at 4 days p.i. and the virus titre from each dish determined by performing a plaque assay on HEK-293 cells. The results from this analysis can be seen in Table 8.1.

Of all the cell types examined, the only cell type to support a reproducible variation in yield between mutant and wild-type virus was WI-38, and only when under-confluent. However, the result was not what was expected with the mutant dl1-3 virus growing to a 12-fold higher titre in under-confluent WI-38 than the wt300 virus. Thus, ORF1, ORF2 and/or ORF3 appear to inhibit rather than to facilitate virus growth under these conditions. It is unclear why the mutant grows better than the wild-type in under-confluent WI-38 cells. It may be due to one of the deleted ORFs (1,2 or 3) repressing virus replication in the wild-type virus or it may be a *cis*-effect due to the deletion of this part of the viral genome, perhaps on the levels of expression of the other E4 gene products.

To determine if the higher titres reached in the dl1-3 mutant virus (as compared to wt300) in under-confluent WI-38 cells were due to the ORF1 or ORF2 proteins, the titres of dl327, dl351 and dl352 were compared after 4 days growth in under-confluent WI-38 and HeLa cells (Table 8.2.). This experiment revealed that dl351 and dl352

Table 8.1. The titres of wt300 and dl1-3 virus, 4 days after infection of HeLa, MRC5, WI-38, HUV-EC-C and HISM cells at an MOI of 5. The virus titres are not comparable between cell types as the experiments were conducted at different times. However, HeLa cells were infected in parallel with all the experiments conducted with the non-transformed cell types. All experiments were conducted in duplicate (Mean values displayed).

Cell type	Virus	Yield pfu/cell	Yield Ratio (dl1-3/wt300)
HeLa (confluent)	wt300	600	
	dl1-3	434	0.72
HeLa (1/3 confluent)	wt300	1250	
	dl1-3	1450	1.2
MRC5 (confluent)	wt300	584	
	dl1-3	638	1.1
MRC5 (1/3 confluent)	wt300	428	
	dl1-3	786	1.8
WI-38 (confluent)	wt300	1812	
	dl1-3	4125	2.3
WI-38 (1/3 confluent)	wt300	200	
	dl1-3	2500	12.5
HUV-EC-C (confluent)	wt300	1087	
	dl1-3	1837	1.7
HUV-EC-C (1/3 confluent)	wt300	1500	
	dl1-3	3250	2.2
HISM (confluent)	wt300	1175	
	dl1-3	1063	0.9
HISM (1/3 confluent)	wt300	525	
	dl1-3	1000	1.9

Table 8.2. The titres of dl327, dl351, and dl352, 4 days after infection of HeLa and WI-38 cells at a MOI of 5.

Cell type	virus	Yield pfu/cell	Yield Ratio (dl351or2/dl327)
HeLa (confluent)	dl327	104	
	dl351	132	1.3
	dl352	126	1.2
HeLa (1/3 confluent)	dl327	97	
	dl351	83	0.9
	dl352	112	1.2
WI-38 (confluent)	dl327	29	
	dl351	35	1.2
	dl352	46	1.6
WI-38 (1/3 confluent)	dl327	69	
	dl351	64	0.9
	dl352	66	1.0

reach titres comparable to their wild-type parent, dl327, in confluent and under-confluent WI-38 cells. This indicates that the absence of the ORF1 and ORF2 proteins in the dl1-3 virus was not the reason for the increased titres observed when grown in under-confluent WI-38 cells and points rather to a role for ORF3 in this process.

The conclusion of these studies is that the deletion of E4 ORF1 and ORF2 has no observable effect on the ability of the virus to grow in non-transformed cells, either quiescent or dividing. However this study was not exhaustive, no attempt was made to observe the virus yields in infections with a multiplicity below 1; in the host it is conceivable that the multiplicity of infection is low and maybe under such conditions the ORF1 and/or 2 proteins are important. Also experiments could be undertaken ensuring the quiescence of the cells under study, possibly by limiting growth factors in the growth medium. This was not undertaken here as the cells were limited in supply and could not be repeatedly passaged.

8.3. Protein expression in ORF1 and ORF2 mutant viruses.

As mentioned, both Bridge and Ketner (1989) and Huang and Hearing (1989a) examined the expression of the late proteins in E4 ORF1 and ORF2 mutant viruses and did not observe any difference compared to wild-type. However, in their studies only protein synthesis at either 26 or 24 hr p.i. was examined while in the study presented here protein expression was examined over the course of the infection. dl327-, dl352-, and dl351-infected cells were labelled for 1 hour with 30 μCi ^{35}S -met/ 3.3×10^6 cells and then harvested into RIPA buffer at 8, 16, 24 or 32 hr p.i. The amount of incorporated label in each fraction was measured and 50,000 cpm aliquots were analysed by SDS-PAGE (Fig. 8.1.). From these results, it was observed that late protein production in both dl351 and dl352 infections occurs at the same rate and to the same level as for the wild-type dl327 virus, and host cell protein shut-off

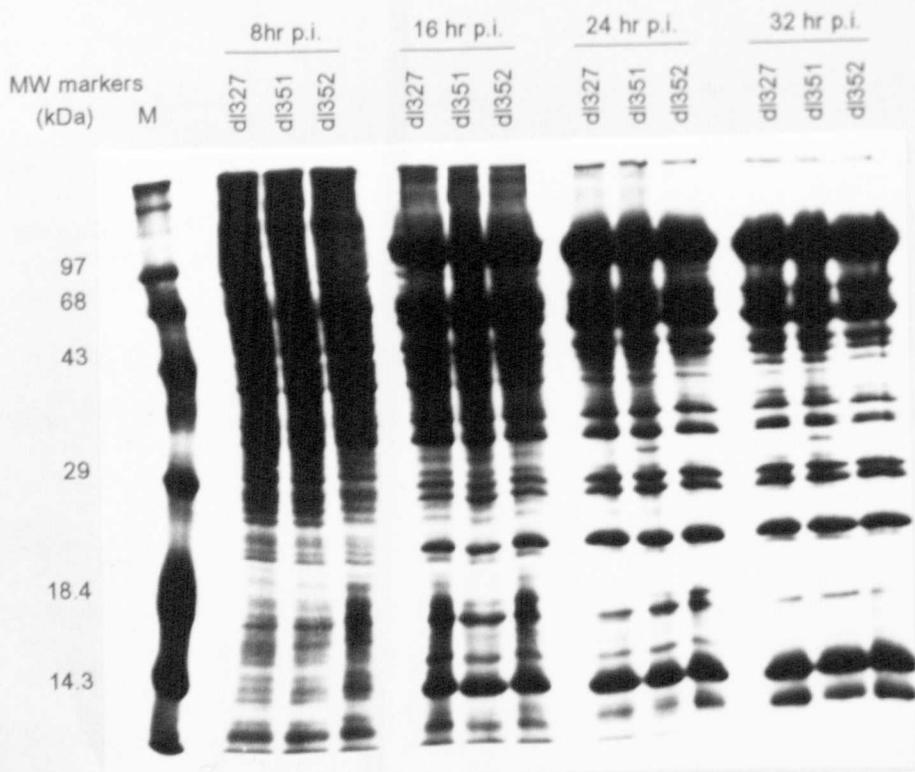
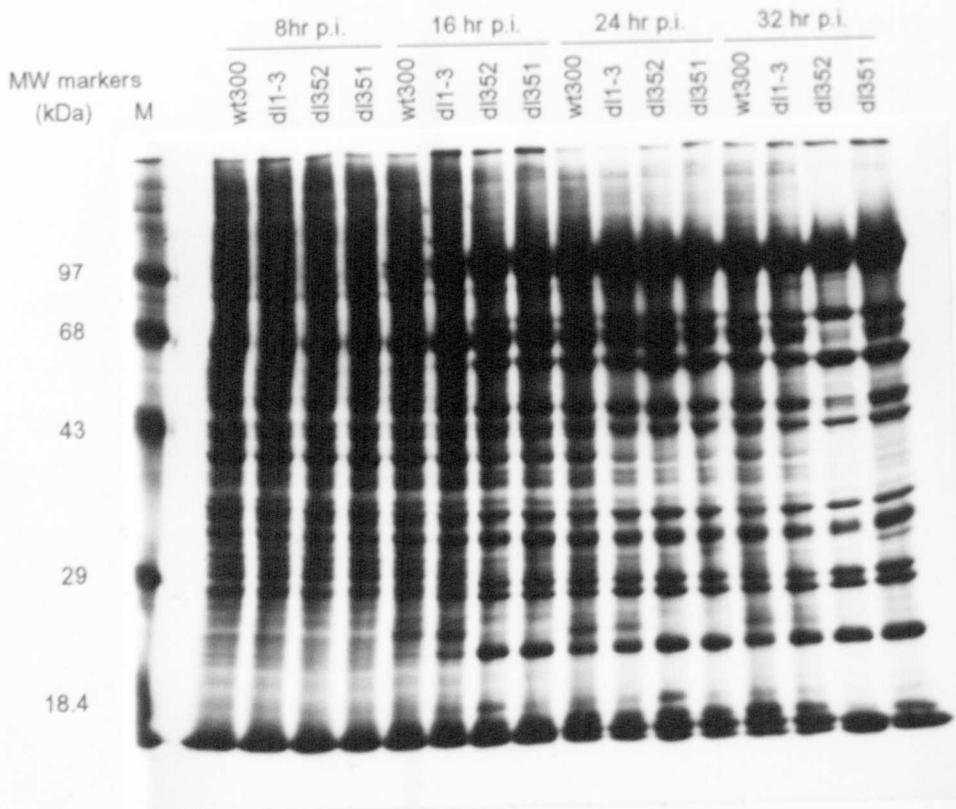


Figure 8.1. Analysis of the protein production of wt300, dl1-3, dl327, dl351 and dl352 during the course of an infection. Infected HeLa cells (MOI = 10) were labelled for 1 hour (30 μ Ci 35 S-met/ 3.3×10^6 cells) and harvested into RIPA buffer at the times indicated (2.3.10.1.). The amount of incorporated label was measured and 50,000 cpm were loaded in each lane (2.3.10.2.). The arrows indicate the 32 kDa protein unique to the dl351 infection at 24 and 32 h p.i.. See Fig. 6.17. for marker details.

occurs as normal, confirming the results of Bridge and Ketner (1989) and Huang and Hearing (1989a) working with the mutants dl1005 and dl1-3 respectively. In this experiment a protein of about 32K was produced in the dl351 virus infection at 24 and 32 hr p.i. which was not produced in the dl327 virus or dl352 virus infection. Experiments conducted in parallel demonstrated that this protein was not produced (or not produced to similar levels) in the dl1-3 infection (data not shown) suggesting that the appearance of this protein is not linked with the absence of the ORF1 protein in dl351 infection. Its appearance may be due to a *cis* effect of the dl1-3 mutation, for example altering E4 expression, or the deletion of ORF2 and ORF3 in dl1-3 might affect the phenotype associated with ORF1 deletion. Alternatively the presence of this extra protein may be due to the absence of the ORF1 protein in conjunction with the absence of the E3 region (dl351), but when an intact E3 region is present the phenotype is lost (i.e. dl1-3). As the protein was of about 32K in size it was thought possible that it was E4 ORF6-34K being over-expressed but attempts to precipitate the protein using an anti-ORF6 monoclonal antibody failed. No reproducible difference was observed between the protein profiles of wild-type and dl352 except for the over-production of an 18K protein at 16 and 24 hr p.i. Again the identity of this protein is unknown.

8.4. E4 ORF1 and ORF2 mutant virions.

The protein content of dl327, dl351 and dl352 virions was compared to see if the absence of ORF1 or ORF2 had any effect on the composition of the virions. This was especially relevant to ORF1, which is expected to be produced with late kinetics (chapter 4). Virions were prepared from virus-infected HeLa cells and 5×10^{10} virions were loaded onto an SDS-PAGE gel and proteins detected by silver staining (Fig 8.2.). No reproducible difference between the ORF1 or ORF2 mutant virions and the wild-type dl327 preparation was observed suggesting that neither the ORF1

MW markers (kDa) Mh MI d1327 d1351 d1352 Virion proteins

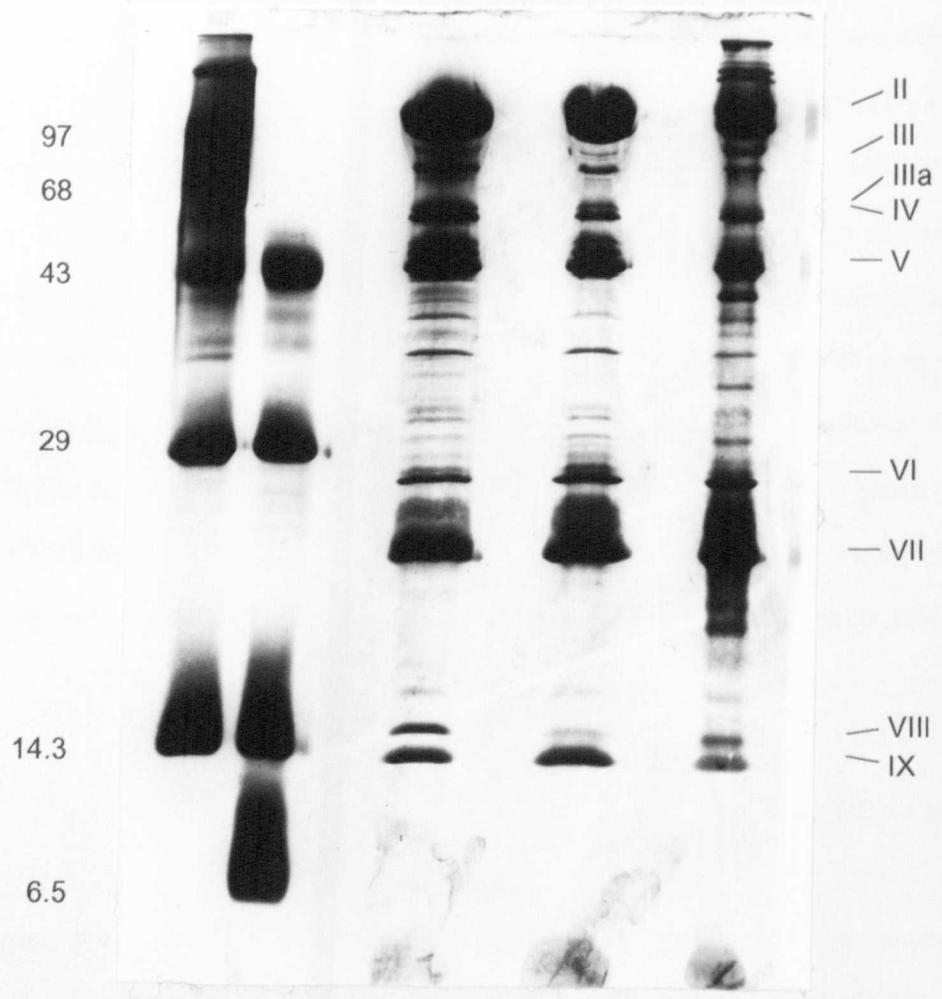


Figure 8.2. Analysis of the protein content of dl327, dl351 and dl352 virions. 5×10^{10} purified virions were loaded onto an SDS-PAGE gel as indicated (2.3.4.10.). The proteins were visualised by monochromatic silver staining 2.3.6.3.. It should be noted that the 18.4K protein of the molecular weight markers failed to stain under the conditions used. Indicated to the right of the gel are the virion proteins (see Fig. 1.1.). See Fig. 6.4. for details of the high molecular weight markers (Mh). The low molecular weight markers (Ml, Amersham) are as follows:

46 kDa : Ovalbumin

30 kDa : Carbonic Anhydrase

21.5 kDa : Trypsin Inhibitor

14.3 kDa : Lysozyme

6.5 kDa : Aprotonin

3.4 kDa : Insulin chain B

2.35 kDa : Insulin chain A

nor ORF2 protein is present in the virion at detectable levels and that neither plays a role in virion construction in HeLa cells.

8.5. Conclusion

The experiments above revealed little about the function of either the ORF1 protein or the ORF2 protein. These proteins do not appear to be required for the growth of the virus in non-transformed cells in tissue culture and ORF1 and ORF2 mutant viruses produce viral late proteins and shut off host cell protein synthesis as normal. Similarly the mutant virion protein profiles appear as wild-type.

Chapter 9.

General Discussion

9.1. Summary and interpretation of the results.

In this study I have examined the production of mRNA and protein from the E4 region of Ad5. Experiments were conducted to elucidate the temporal pattern of expression of the differentially spliced E4 mRNAs, both in the presence and in the absence of the E1b-55K/E4 ORF6-34K complex which is known to be involved in the optimal cytoplasmic accumulation of late adenoviral mRNAs. The results of these studies demonstrated that the expression of the E4 mRNAs is regulated in a temporal manner due to a change in the pattern of splicing. Early in the infection, long mRNAs are produced while after DNA replication the pattern of splicing changes and shorter mRNAs are produced, lacking the D3-A3 intron (Fig. 1.5.). However, there is an exception to this general rule with the expression of mRNA A/full length mRNA with kinetics later than all the other late mRNAs. From the data presented, it cannot be determined if both or just one of these mRNAs exists but it is clear that the expression of these mRNAs is distinct from the other E4 mRNAs. In addition to displaying a unique late expression pattern, only these two mRNAs displayed significant dependence on the E1b-55K/E4 ORF6-34K complex for optimal cytoplasmic accumulation.

It is unclear why E4 mRNA A/full length mRNA is solely dependent on the E1b-55K/E4 ORF6-34K complex for optimal cytoplasmic accumulation while the other late E4 mRNAs show no dependence. However, this unique dependence allowed the investigation of the mechanism by which the E1b-55K/E4 ORF6-34K complex functions. As late E4 mRNAs which show no dependence on the E1b-55K/E4 ORF6-34K complex differ in structure in only a few regions (D1-A1b and possibly D3-A3) as compared to mRNA A/full length mRNA it was concluded that these regions potentially hold the key to conferring E1b-55K/E4 ORF6-34K dependence on an mRNA. Similar conclusions were made by Leppard (1993) when observing the expression of MLTU transcripts. In an initial attempt to prove the importance of these regions, a virus was constructed which lacked an intact E4 D1 splice donor site. On

examination of the expression of mRNA A/full length mRNA from this mutant virus it was found that this mRNA was no longer so strongly dependent on the E4 ORF6-34K protein, and so presumably on the E1b-55K protein, for optimal cytoplasmic accumulation. Therefore, the presence of intact splice sites/potential intron sequences in E4 late mRNAs appears to confer dependence on the E1b-55K/E4 ORF6-34K protein complex for cytoplasmic accumulation, supporting the hypothesis proposed by Leppard (1993) (see section 1.3.5.).

This function of the E1b-55K/E4 ORF6-34K protein in the export of incompletely spliced viral mRNAs from the nucleus of infected cells is akin to that of the HIV-1 Rev protein in regulating HIV mRNA accumulation. However there is a distinct difference between the two systems; incompletely spliced Rev-dependent mRNAs all contain a Rev response element (RRE) with which the Rev protein interacts. No such element has been detected in E1b-55K/E4 ORF6-34K dependent mRNAs and no RNA binding activity has yet been demonstrated for the E1b-55K/E4 ORF6-34K complex. However, a potential RNA binding site has been identified in the E1b-55K sequence (Leppard, unpublished observations). Experiments conducted by Williams (1993) demonstrated that the Rev/RRE system for exporting unspliced or partially spliced mRNAs out of the nucleus could function in an E1b-deficient adenovirus, partially rescuing the E1b function, so confirming the similarities in function of these two proteins. However, the differences discussed above and the fact that no distinct regions of homology can be detected between the two proteins, suggests that, although both are involved in the export of partially spliced viral mRNAs from the nucleus, they achieve this end result by different mechanisms.

It is also of note that Rev has been observed to be involved in regulating splicing and mRNA translation. While neither the E1b-55K, nor the E4 ORF6-34K protein is thought to be involved in mRNA translation, it has been recently observed that the E4 ORF6-34K protein does play a role in the regulation of splicing (Nordqvist *et al.*, 1994), although it is unclear if this is due to E4 ORF6-34K within the E1b/E4 complex.

From the E4 mRNA expression data accumulated in previous studies (Berk and Sharp, 1978; Chow *et al.*, 1979; Freyer *et al.*, 1984; Tigges and Raskas, 1984; Virtanen *et al.*, 1984) it was found that the Ad2 E4 region encoded at least 7 potential proteins; ORFs 1, 2, 3, 3/4, 4, 6, 6/7. In this study of Ad5 it was concluded that the mRNAs encoding the ORF3/4 protein and a minor ORF3 variant, ORF3/12aa, did not exist in wild type infections in the absence of inhibitors. Of the six E4 proteins therefore expected in an Ad5-infected cell, only two had not been identified in infected cells in previous studies; ORFs 1 and 2. On examination of the recently published Ad5 E4 sequence it was discovered that the predicted ORF1 and 2 reading frames, although largely conserved, contained a number of frame shift mutations. However when this E4 region was resequenced it was found that these apparent mutations were sequencing errors and that the ORF1 and 2 reading frames were intact.

In this study I attempted to identify the two proteins potentially encoded by ORF1 and ORF2 reading frames in infected cells by generating mono-specific polyclonal antisera to these proteins expressed in bacteria. However, as discussed in chapter 6, both the antisera specific to the ORF1 protein failed to identify the ORF1 protein within infected cells, probably due to their low titres. It is of note that Javier (1994) recently produced a polyclonal antiserum to the Ad9 ORF1 protein, by intradermally inoculating a New Zealand White rabbit, with which the Ad9 ORF1 protein was identified in Ad9-infected cells. The difference in the ORF1 sequences (45% identity between Ad9 ORF1 and Ad5 ORF1), the route of inoculation and/or the rabbits used may of all contributed to the successful generation of this Ad9 ORF1 antisera.

The ORF2 antiserum was more successful allowing the identification of the ORF2 protein in the infected cells. It was found that the ORF2 protein is initially expressed early in the infection but steady-state levels of ORF2 remain high in the infected cell until at least 32 hr p.i. These data correlate well with the expression of mRNA C which is predicted to encode the ORF2 protein.

As yet no function has been ascribed to either the Ad5 ORF1 or ORF2 protein although it has been demonstrated that the Ad9 E4 ORF1 protein is involved in cellular

transformation (Javier, 1994). Homology searches demonstrated that both proteins are conserved in a number of different adenovirus serotypes, indicating significant roles for the proteins in infections of the host organism. The ORF1 protein was also found to have homology with dUTPase of *S. cerevisiae* and Vaccinia virus. However, the lack of homology of ORF1 with the conserved motifs found in dUTPases, and the fact that the Vaccinia dUTPase protein is produced early in the infection (Slabaugh and Roseman, 1989) while ORF1 is presumably produced late in the infection (see chapter 4), would suggest that ORF1 may only have a related function. Homology searches with the ORF2 protein were less revealing. However the successful generation of an antiserum to the ORF2 protein allowed an initial characterisation of the ORF2 protein *in vivo*. The ORF2 protein was identified in the cytoplasm of the infected cell and, as no ORF2 protein-protein interactions could be identified, it was concluded that the protein was either soluble or present in the cytoplasm in a complex of less than 4S. Unfortunately this data does not allow any speculation as to the function of the ORF2 protein, beyond the possibility that it may be a cytoplasmic enzyme. The fact that in Ad5 related functions are grouped into transcription units, suggests that the ORF2 protein may be involved in the control of gene expression or DNA replication at some level.

9.2. Future work

The investigation of the E4 mRNAs presented here was undertaken using RNase protection analysis as this technique was sensitive enough to identify even the low levels of some of the E4 mRNAs. However this technique did not allow the identification of complete E4 mRNAs and so the E4 mRNAs were deduced by mapping the protected fragments produced from a given probe, to known mRNAs. Therefore this form of analysis is dependent on the published structures of the E4 mRNAs being correct. Traditional methods of mRNA detection were deemed impractical or not sensitive enough to address the question of E4 mRNA expression in

infected cells in the absence of inhibitors. However, the relatively recently refined technique of reverse-transcription PCR may be useful. To test this idea, primers were designed that, when used in RT-PCR, would result in the production of E4 cDNAs of unique lengths, according to the mRNA from which they were amplified, so allowing the unambiguous identification of all E4 mRNAs. Initial experiments by S. White (undergraduate project student) identified mRNA L (the most abundant E4 mRNA) and possibly some of the other larger E4 mRNAs by this approach. However, this work needs to be continued and the technique optimised to allow the simultaneous identification and quantitation of all the E4 mRNAs.

Further work also needs to be conducted investigating the role of splice sites and potential intron regions of the E4 region in conferring E1b-55K/E4 ORF6-34K dependence on an mRNA. This work would be a continuation of the work described in chapter 5 with the further characterisation of E4 RNA expression from the vD1 virus (i.e. examination of nuclear levels of E4 mRNA from vD1 and the use of potential cryptic splice sites). The role of the other potentially important splice sites (i.e. A1b, D3, A3) in mediating E1b-55K/E4 ORF6-34K dependence also needs to be investigated, and so a series of mutants lacking either one, or a combination, of these splice sites needs to be constructed. Other virus constructs could also be made where the mRNA sequences unique to E4 mRNA A/full length mRNA are duplicated at another site in the adenovirus genome, to determine whether they can confer E1b-55K/E4 ORF6-34K dependence on the expression of a heterologous gene.

As the E4 region is relatively small, it is feasible to examine the existence of any protein-E4 RNA interactions, especially involving the E1b-55K protein, in the light of the identification of the potential RNA binding site within E1b-55K. Such experiments would involve the production of labelled, sense transcripts of fragments of the E4 region and the investigation of the binding, *in vitro*, of any proteins from infected and uninfected cell extracts by band-shift analysis. The involvement of the E1b-55K or E4 ORF6-34K proteins in any specific complexes detected could be examined by using

specific monoclonal antibodies in such experiments to cause a 'super-shift' of the labelled RNA-protein complex on the gel.

As the production of the an ORF1 antiserum was unsuccessful, there remains a large number of uncompleted experiments involved in the identification and characterisation of the Ad5 ORF1 protein. A successful ORF1 antiserum needs to be produced. As antigen produced from bacteria in this study only produced a poor immune response in the rabbit, an alternative source of antigen should be used, such as protein expressed in yeast or from Baculovirus. These eukaryotic systems may allow the production and purification of native ORF1 protein with the correct, if any, post-translational modifications (although this is less likely in the yeast system) increasing the possibility of producing antigen capable of stimulating a useful immune response. Alternatively, synthetic ORF1 oligopeptides coupled to a carrier protein such as KLH or BSA, could be used to generate an immune response. Such oligopeptides would be taken from regions of the ORF1 protein with a high probability of being surface regions, such as the region around residue 25 and be from 10 to 15 residues in length (Harlow and Lane, 1988). Whatever the antigen used, in any future attempts to generate an immune response to the ORF1 protein, two rabbits should be used to eliminate any animal-specific effects. When a successful Ad5 ORF1 antiserum is generated, similar experiments to those carried out with the ORF2 antiserum shall be conducted.

No investigation of the potential function of the ORF1 protein as a viral dUTPase has been undertaken in this study as the sequence of the *S. cerevisiae* dUTPase was only published recently. Such an investigation would involve the development of an assay for dUTPase. Such an assay would exploit the catalytic activity of dUTPase ($dUTP \Rightarrow dUMP + PP_i$), possibly by radiolabelling the substrate and assaying the production of the products by chromatography, such as TLC. Any dUTPase activity of ORF1 could be assayed in a variety of ways; in vitro transcribed/translated ORF1 (presently available) could be assayed for dUTPase activity by incubating it with labelled dUTP in a suitable buffer, or alternatively extracts from uninfected, wild-type and ORF1

mutant-infected cells could be assayed in parallel. Also, when an antiserum to ORF1 is available, immunoprecipitates could be assayed for dUTPase activity.

Other work to be carried out is the construction of further ORF1 and ORF2 specific virus mutants. At present, the mutants available are dl1-3, dl351 and dl352. dl1-3 is limited in its use as, although in a wt300 background, it lacks three reading frames of the E4 region. The specific ORF1 and ORF2 mutants, dl351 and dl352, also have shortcomings, since they lack the most of the E3 region. This may result in an altered phenotype if the ORF1 and 2 proteins are involved in interactions with E3 proteins. Attempts were made to build the dl351 and dl352 lesions into a wt300 background in the course of this study, but they failed. Such mutants would be useful in the continued search for a phenotype for the ORF1 and ORF2 proteins. Experiments would include the growth of virus in non-transformed cells, as mentioned in chapter 8, but under different conditions (e.g. serum-starved cells to induce quiescence or varying MOI). Also, comparisons of viral protein expression and rate of viral DNA synthesis in such infected cells may be of interest.

In conclusion, the E4 region of Ad5 remains of considerable interest. The further, and more complete, understanding of its expression and role in the lytic infection are important challenges for the future.

Chapter 10.

References

- Adam, S.A., & Dreyfuss, G. (1987). Adenovirus proteins associated with mRNA and hnRNA in infected HeLa cells. *J. Virol.* **61**, 3276-3283.
- Adami, G., & Babiss, L.E. (1991). DNA template effect on RNA splicing: two copies of the same gene in the same nucleus are processed differently. *EMBO J.* **10**, 3457-3465.
- Agutter, P.S. (1991). Role of the cytoskeleton in nucleocytoplasmic RNA and protein distributions. *Biochem. Soc. Trans.* **19**, 1094-1098.
- Akusjärvi, G., & Persson, H. (1981). Controls of RNA splicing and termination in the major late adenovirus transcription unit. *Nature* **292**, 420-426.
- Akusjärvi, G. (1993). Proteins with transcription regulatory properties encoded by human adenoviruses. *Trends Microbiol.* **1**, 163-170.
- Allison, L.A., North, M.T., Murdoch, K.J., Romaniuk, P.J., Deschamps, S. & LeMaire, M. (1993). Structural requirements of 5S rRNA for nuclear transport, 7S ribonucleoprotein particle assembly, and 60S ribosomal subunit assembly in *Xenopus* oocytes. *Mol. Cell Biol.* **13**, 6819-6831.
- Alonso-Caplen, F.V., Nemeroff, M.E., Qiu, Y., & Krug, R.M. (1992). Nucleocytoplasmic transport: the influenza virus NS1 Protein regulates the transport of spliced NS1 mRNA and its precursor NS1 mRNA. *Genes Dev.* **6**, 255-267.
- Alt, F.W., Bothwell, A.L.M., Knapp, M., Siden, E., Mather, E., Koshland, M., & Baltimore, D. (1980). Synthesis of secreted and membrane bound immunoglobulin m heavy chains is directed by mRNAs that differ at their 3' ends. *Cell* **20**, 293-302.
- Amara, S.G., Jonas, V., Rosenfeld, M.G., Ong, E.S., & Evans, R. (1982). Alternative RNA processing in calcitonin gene expression generates mRNAs encoding different polypeptide products. *Nature* **298**, 240-244.
- Anderson, K.P., & Klessig, D.F. (1984). Altered mRNA splicing in monkey cells abortively infected with human adenovirus may be responsible for inefficient synthesis of the virion fiber polypeptide. *Proc. Natl. Acad. Sci. USA* **81**, 4023-4027.
- Anderson, C.W., Schmitt, R.C., Smart, J.E., & Lewis, J.B. (1984). Early region 1B of adenovirus 2 encodes two coterminal proteins of 495 and 155 amino acid residues. *J. Virol.* **50**, 387-396.
- Ankerst, J., & Jonsson, N. (1989). Adenovirus type 9-induced tumorigenesis in the rat mammary gland related to sex hormonal state. *J. Natl. Cancer Inst.* **81**, 294-298.
- Ankerst, J., Jonsson, N., Kjellen, L., Norrby, E., & Sjogren, H.O. (1974). Induction of mammary fibroadenomas in rats by adenovirus type 9. *Int. J. Cancer* **13**, 286-290.

- Babich, A., & Nevins, J.R. (1981). The stability of early adenovirus mRNA is controlled by the 72 kd DNA-binding protein. *Cell* **26**, 371-379.
- Babich, A., Feldman, L.T., Nevins, J.R., Darnell, J.E., & Weinberger, C. (1983). Effects of adenovirus on metabolism of specific host mRNAs: transport control and specific translational discrimination. *Mol Cell Biol.* **3**, 1212-1221.
- Babiss, L.E., & Ginsberg, H.S. (1984). Adenovirus type 5 early region 1b gene product is required for efficient shutoff of host protein synthesis. *J. Virol.* **50**, 202-212.
- Babiss, L.E., Ginsberg, H.S., & Darnell, J.E. (1985). Adenovirus E1B proteins are required for accumulation of late viral mRNA and for effects on cellular mRNA translation and transport. *Mol. Cell. Biol.* **5**, 2552-2558.
- Bagchi, S., Raychaudhuri, P., & Nevins, J.R. (1989). Phosphorylation-dependent activation of the adenovirus-inducible E2F transcription factor in a cell free system. *Proc. Natl. Acad. Sci. USA* **86**, 4352-4356.
- Bai, M., Harfe, B., & Freimuth, P. (1993). Mutations that alter an Arg-Gly-Asp (RGD) sequence in the adenovirus type-2 penton base protein abolish its cell-rounding activity and delay virus reproduction in flat cells. *J. Virol.* **67**, 5198-5205.
- Baker, C.D., & Ziff, E.B. (1981). Promoters and heterogeneous 5' termini of the messenger RNAs of adenovirus serotype 2. *J. Mol. Biol.* **149**, 189-221.
- Bandara, L.R., & La Thangue, N.B. (1991). Adenovirus E1a prevents the retinoblastoma gene product from complexing with a cellular transcription factor. *Nature* **351**, 494-497.
- Barker, D.D., & Berk, A.J. (1987). Adenovirus proteins from both E1B reading frames are required for transformation of rodent cells by viral infection and DNA transfection. *Virology* **156**, 107-121.
- Bataillé, N., Helser, T., & Fried, H.M. (1990). Cytoplasmic transport of ribosomal subunits microinjected into the *Xenopus laevis* oocyte nucleus: a generalized, facilitated process. *J. Cell Biol.* **111**, 1571-1582.
- Bautista, D.S., Hitt, M., McGrory, J., & Graham, F.L. (1991). Isolation and characterisation of insertion mutants in E1A of adenovirus type 5. *Virology* **182**, 578-596.
- Berget, S.M., Moore, C., & Sharp, P. (1977). Spliced segments at the 5' terminus of Ad2 late mRNA. *Proc. Natl. Acad. Sci. USA* **74**, 3171-3175.
- Berk, A.J. (1986). Adenovirus promoters and E1A transactivation. *Annu. Rev. Genet.* **20**, 45-79.

- Berk, A.J., & Sharp, P.A. (1978). Structure of the adenovirus 2 early mRNAs. *Cell* **14**, 695-711.
- Berkner, K.E., & Sharp, P.A. (1985). Effect of tripartite leader on synthesis of a non-viral protein in a adenovirus 5' recombinant. *Nucleic Acids Res.* **13**, 841-857.
- Beyer, A.L., & Osheim, Y.N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* **2**, 754-765.
- Blackwell, J.R., & Horgan, R. (1991). A novel strategy for production of a highly expressed recombinant protein in an active form. *FEBS* **295**, 10-12.
- Blobel, G. (1985). Gene gating: a hypothesis. *Proc. Natl. Acad. Sci. USA* **82**, 8527-8529.
- Bodnar, J.W., Hanson, P.I., Polvino-Bodnar, M.m Zempsky, W., & Ward, D.C. (1989). The terminal regions of adenovirus and minute virus of mice DNAs are preferentially associated with the nuclear matrix in infected cells. *J. Virol.* **63**, 4344-4353.
- Bondesson, M., Svensson, C., Linder, S., & Akusjarvi, G. (1992). The carboxy-terminal exon of the adenovirus E1A protein is required for E4F-dependent transcription activation. *EMBO J.* **11**, 3347-3354.
- Bos, J., Polder, L.J., Bernards, R., Schrier, P.I., van den Elsen, P.J., van der Eb, A.J., & van Ormondt, H. (1981). The 2.2 kb E1b mRNA of human Ad12 and Ad5 codes for two tumor antigens starting at different AUG triplets. *Cell* **27**, 121-131.
- Boulanger, P.A., & Blair, G.E. (1991). Expression and interactions of human adenovirus oncoproteins. *Biochem. J.* **275**, 281-299.
- Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding. *Anal. Biochem.* **72**, 248-254.
- Brady, H.A., Scaria, A., & Wold, W.S.M. (1992). Map of *cis*-acting sequences that determine alternative pre-mRNA processing in the E3 complex transcription unit of adenovirus. *J. Virol.* **66**, 5914-5923.
- Bridge, E., & Ketner, G. (1989). Redundant control of adenovirus late gene expression by early region 4. *J. Virol.* **63**, 631-638.
- Bridge, E., & Ketner, G. (1990). Interaction of adenoviral E4 and E1b products in late gene expression. *Virology* **174**, 345-353.
- Bridge, E., Hemström, C., & Pettersson, U. (1991). Differential regulation of adenovirus late transcriptional units by the products of early region 4. *Virology* **183**, 260-266.

- Bridge, E., Carmo-Fonseca, M., Lamond, A., & Pettersson, U. (1993a). Nuclear organization of splicing small nuclear ribonucleoproteins in adenovirus infected cells. *J. Virol.* **67**, 5792-5802.
- Bridge, E., Medghalchi, S., Ubol, S., Leesong, M., & Ketner, G. (1993b). Adenovirus early region 4 and viral DNA synthesis. *Virology* **193**, 794-801.
- Caras, I.W., Davitz, M.A., Rhee, L., Weddell, G., Martin, D.W., & Nussenzweig, V. (1988). Cloning of decay-accelerating factor suggests novel use of splicing to generate two proteins. *Nature* **325**, 545-549.
- Carter, K.C., Taneja, K.L., & Lawrence, J.B. (1991). discrete nuclear domains of poly(A) RNA and their relationship to the functional organization of the nucleus. *J. Cell Biol.* **115**, 1191-1202.
- Carter, K.C., Bowman, D., Carrington, W., Fogarty, K., McNeil, J.A., Fay, F.S., & Lawrence, J.B. (1993). A three-dimensional view of precursor messenger RNA metabolism within the mammalian nucleus. *Science* **259**, 1330-1335.
- Cepco, C.L., & Sharp, P.A. (1983). Assembly of major adenovirus capsid protein is mediated by a nonvirion protein. *Virology* **129**, 137-154.
- Chang, D.D., & Sharp, P.A. (1989). Regulation by HIV Rev depends upon recognition of splice sites. *Cell* **59**, 789-795.
- Chang, D.D., & Sharp, P.A. (1990). Messenger RNA transport and HIV *rev* regulation. *Science* **249**, 614-615.
- Chatton, B., Bocco, J.L., Gaire, M., Hauss, C., Reimund, B., Goetz, J., & Kedinger, C. (1993). Transcriptional activation by the adenovirus larger E1A product is mediated by members of the cellular transcription factor ATF family which can indirectly associate with E1A. *Mol. Cell. Biol.* **13**, 561-570.
- Chebli, K., Gattoni, R., Schmitt, P., Hildwein, G., & Stevenin, J. (1989). The 216-nucleotide intron of the E1A pre-mRNA contains a hairpin structure that permits utilization of unusually distant acceptors. *Mol. Cell. Biol.* **9**, 4852-4861.
- Chellappan, S.P., Hiebert, S., Mudryj, M., Horowitz, J.M., & Nevins, J.R. (1991). The E2F transcription factor is a target for the RB protein. *Cell* **65**, 1053-1062.
- Chen, M. and Horwitz, M.S. (1990). Replication of an adenovirus type 34 mutant. DNA containing tandem reiterations of the inverted terminal repeat. *Virology* **179**, 567-575.
- Chinnadurai, G. (1983). Adenovirus 2 *lp*⁺ locus encodes for a 19kd tumor antigen that plays an essential role in the transformation. *Cell* **33**, 759-766.

- Chou, P.Y., & Fasman, G.D. (1974). Prediction of protein conformation. *Biochem.* **13**, 222-245.
- Chou, P.Y., & Fasman, G.D. (1977). β -Turns in proteins. *J. Mol. Biol.* **115**, 135-175.
- Chou, P.Y., & Fasman, G.D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol.* **47**, 45-148.
- Chow, L.T., Broker, T.R., & Lewis, J.B. (1979). Complex splicing patterns of RNAs from the early regions of adenovirus 2. *J. Mol. Biol.* **134**, 265-279.
- Chroboczek, J., & Jacrot, B. (1987). The sequence of adenovirus fiber: similarities and differences between serotypes 2 and 5. *Virology* **161**, 549-554.
- Chroboczek, J., Bieber, F., & Jacrot, B. (1992). The sequence of the genome of adenovirus type 5 and its comparison with the genome of adenovirus type 2. *Virology* **186**, 280-285.
- Ciejek, E.M., Nordstrom, J.L., Tsai, M.-J., & O'Malley, B.W. (1982). Ribonucleic acid precursors are associated with the chick oviduct nuclear matrix. *Biochemistry* **21**, 4945-4953.
- Ciejek, E.M., Tsai, M.-J., & O'Malley, B.W. (1983). Actively transcribed genes are associated with the nuclear matrix. *Nature* **306**, 607-609.
- Cladaras, C., & Wold, W.S.M. (1985). DNA sequence of the early region E3 transcription unit of adenovirus 5. *Virology* **140**, 28-43.
- Cutt, J.R., Shenk, T., & Hearing, P. (1987). Analysis of adenovirus early region 4-encoded polypeptides synthesized in productively infected cells. *J. Virol* **61**, 543-552.
- Dargemont, C., & Kuhn, L.C. (1992). Export of mRNA from microinjected nuclei of *Xenopus laevis* oocytes. *J. Cell Biol.* **118**, 1-9.
- Darnell, J.E.Jr. (1982). Variety in the level of gene control in eukaryotic cells. *Nature* **297**, 365-371.
- Davison, A.J., Telford, E.A.R., Watson, M.S., McBride, K., & Mautner, V. (1993). The DNA sequence of adenovirus type 40. *J. Mol. Biol.* **234**, 1308-1326.
- Debbas, M. & White, E. (1993). Wild-type p53 mediates apoptosis by E1A which is inhibited by E1B. *Genes Dev.* **7**, 546-554.
- DeGraaf, A., Van Hemert, F., Linnemans, W.A.M., Brakenhoff, G.J., DeJong, L., Van Renswoude, J., & Van Driel, R. (1990). Three-dimensional distribution of DNase I-sensitive chromatin regions in the interphase nuclei of embryonal carcinoma cells. *Eur. J. Cell Biol.* **52**, 613-619.

- Devaux, C., Caillet-Boudin, M-L., Jacrot, B., & Boulanger, P. (1987). Crystallization, enzymatic cleavage, and the polarity of the adenovirus type-2 fiber. *Viol.* **161**, 121-128.
- DeVries, E.W., Van Driel, W., Bergsma, W.G., Arnberg, A.C. & Van der Vleit, P.C. (1989). HeLa nuclear protein recognising DNA termini and translocating on DNA forming a regular DNA multimeric protein complex. *J. Mol. Biol.* **208**, 65-78.
- DeZazzo, J.D., Falck-Pedersen, E., & Imperiale, M.J. (1991). Sequences regulating temporal poly(A) site switching in the adenovirus major late transcription unit *Mol. Cell Biol.* **11**, 5977-5984.
- Dix, I., & Leppard, K.N. (1992). Open reading frames 1 and 2 of adenovirus region E4 are conserved between human serotypes 2 and 5. *J. Gen. Virol.* **73**, 2975-2976.
- Dix, I., & Leppard, K.N. (1993). Regulated splicing of adenovirus type 5 E4 transcripts and regulated cytoplasmic accumulation of E4 mRNA. *J. Virol.* **67**, 3226-3231.
- Dolph, P.J., Racaniello, V., Villamarin, A., Palladino, F., & Schneider, R.J. (1988). The adenovirus tripartite leader eliminates the requirement for cap binding protein during translation initiation. *J. Virol.* **62**, 2059-2066.
- Downey, J.F., Rowe, D.T., Bacchetti, S., Graham, F.L., & Bayley, S.T. (1983). Mapping of a 14,000-dalton antigen to early region 4 of the human adenovirus5 genome. *J. Virol.* **45**, 514-523.
- Draper, E. (1989). How do proteins recognize specific RNA sites? *Trends Cell Biol.* **14**, 335-338.
- Dreyfuss, G., Matunis, M.J., Piñol-Roma, S., & Burd, C.G. (1993). hnRNP proteins and the biogenesis of mRNA. *Annu. Rev. Biochem.* **62**, 289-321.
- Dunsworth-Browne, M., Schell, R.E., & Berk, A.J. (1980). Adenovirus terminal protein protects single stranded DNA from digestion by a cellular exonuclease. *Nucleic Acids Res.* **8**, 543-554.
- Dworetzky, S.I., & Feldherr, C.M. (1988). Translocation of RNA-coated gold particles through the nuclear pores of oocytes. *J. Cell Biol.* **106**, 575-584.
- Dyson, N., & Harlow, E. (1992). Adenovirus E1A targets key regulators of cell proliferation. In: Cancer surveys volume 12: Tumour Suppressor Genes, the Cell Cycle and Cancer. A.J. Levine, ed. (Cold Spring Harbor Laboratory Press) pp 161-195.
- Eckner, R., Ellmeier, W., & Birnstiel, M.L. (1991). Mature mRNA 3' end formation stimulates RNA export from the nucleus. *EMBO J.* **10**, 3513-3522.

- Elliot, D.J., Stutz, F., Lescure, A., & Rosbash, M. (1994). mRNA nuclear export. *Curr. Op. Gen. Dev.* **4**, 305-309.
- Emini, E.A., Hughes, J.V., Perlow, D.S., & Boger, J. (1985). Induction of hepatitis A virus neutralizing antibody by a virus specific synthetic peptide. *J. Virol.* **55**, 836-839.
- Esche, H., Mathews, M.B., & Lewis, J.B. (1980). Proteins and mRNAs of the transforming region of wild-type and mutant adenoviruses. *J. Mol. Biol.* **142**, 399-417.
- Falgout, B., & Ketner, G. (1987). Adenovirus early region 4 is required for efficient virus particle assembly. *J. Virol.* **61**, 3759-3768.
- Featherstone, C., Darby, M., & Gerace, L. (1988). A monoclonal antibody against the nuclear pore complex inhibits nucleocytoplasmic transport of protein and RNA in vivo. *J. Cell Biol.* **107**, 1289-1297.
- Fey, E.G., Krochmalnic, G., & Penman, S. (1986). The non-chromatin substructures of the nucleus: the ribonucleoprotein (RNP)-containing and RNP-depleted matrices analysed by sequential fractionation and resinless section electron microscopy. *J. Cell Biol.* **102**, 1654-1665.
- Flint, S.J. (1986). Regulation of adenovirus mRNA formation. *Adv. Virus. Res.* **31**, 169-228.
- Fortes, P., Beloso, A., & Ortin, J. (1994). Influenza virus NS1 protein inhibits pre-mRNA splicing and blocks mRNA nucleocytoplasmic transport. *EMBO J.* **13**, 704-712.
- Fredamn, J.N., & Engler, J.A. (1993). Adenovirus precursor to terminal protein interacts with the nuclear matrix in vivo and in vitro. *J. Virol.* **67**, 3384-3395.
- Freyer, G.A., Katoh, Y., & Roberts, R.J. (1984). Characterization of the major mRNAs from adenovirus 2 early region 4 by cDNA cloning and sequencing. *Nucleic Acids Res.* **12**, 3503-3519.
- Fu, X.-D., & Maniatis, T. (1990). Factor required for mammalian spliceosome assembly is localised to discrete regions in the nucleus. *Nature* **343**, 437-441.
- Gadsden, M.H., McIntosh, E.M., Game, J.C., Wilson, P.J., & Haynes, R.H. (1993). dUTP pyrophosphatase is an essential enzyme in *Saccharomyces cerevisiae* *EMBO J.* **12**, 4425-4431.
- Garnier, J., Osguthorpe, D.J., Robson, B. (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* **120**, 97-120.

- Gattoni, R., Schmitt, P., & Stevenin, J. (1988). In vitro splicing of adenovirus E1A transcripts: characterization of novel reactions and of multiple branch points far from the 3' splice site. *Nucleic Acids Res.* **16**, 2389-2409.
- Gattoni, R., Chebli, J.R., Himmelsbach, M., & Stevenin, J. (1991). Modulation of alternative splicing of adenoviral E1A transcripts: factors involved in the early-to-late transition. *Genes Dev.* **5**, 1847-1858.
- Ge, H., & Manley, J.L. (1990). A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. *Cell* **62**, 25-34.
- Gibson (1984) Studies on the Epstein-Barr virus genome. PhD thesis, Cambridge University, England.
- Gower, H.J., Barton, C.H., Elsom, V.L., Thompson, J., Moore, S.E., Dickson, G., & Walsh, F.S. (1988). Alternative splicing generates a secreted form of N-CAM in muscle and brain. *Cell* **55**, 955-964.
- Graham, F.L., Smiley, J., Russell, W.C., & Nairn, R. (1977). Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J. Gen. Virol.* **36**, 59-72.
- Green, M., Piña, M., & Kimes, R.C. (1967). Biochemical studies on adenovirus multiplication XII. Plaquing efficiencies of purified human adenoviruses. *Virology* **31**, 562-565.
- Green, M.R. (1991). Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. *Annu. Rev. Cell Biol.* **7**, 559-599.
- Griffith, O.M. (1975). Techniques of preparative, zonal, and continuous flow ultracentrifugation. Beckman Instruments, Inc.
- Hai, T-Y., Liu, F., Allegretto, E.A., Karin, M., & Green, M.R. (1988). A family of immunologically related transcription factors that includes multiple forms of ATF and AP-1. *Genes Dev.* **2**, 1216-1226.
- Halbert, D.N., Catt, J.R., & Shenks, T. (1985). Adenovirus early region 4 encodes functions required for efficient DNA replication, late gene expression, and host cell shutoff. *J. Virol.* **56**, 250-257.
- Hamm, J., & Mattaj, I.W. (1990). Monomethylated cap structures facilitate RNA export from the nucleus. *Cell* **63**, 109-118.
- Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**, 557-580.

- Handa, H., Kingston, R.E., & Sharp, P.A. (1983). Inhibition of adenovirus early region IV transcription *in vitro* by a purified viral DNA binding protein. *Nature* **302**, 545-547.
- Hanover, J.A. (1992). The nuclear pore: at the cross roads. *FASEB J.* **6**, 2288-2295.
- Hardy, S., Engel, D.A., & Shenk, T. (1989). An adenovirus early region-4 gene product is required for the induction-specific form of cellular E2F activity. *Genes Dev.* **3**, 1062-1074.
- Harlow, E., & Lane, D. (1988). Antibodies - A laboratory manual. (Cold Spring Harbor Laboratory).
- Harper, J.E., & Manley, J.L. (1991). A novel protein factor is required for use of distal alternative 5' splice sites *in vitro*. *Mol. Cell Biol.* **11**, 5945-5953.
- Hashimoto, S., Pursley, M.H., & Green, M. (1981). Nucleotide sequences and mapping of novel heterogenous 5'-termini of adenovirus 2 early region 4 mRNA. *Nucleic Acids Res.* **9**, 1675-1689.
- Hay, R.T., & Russell, W.C. (1989). Recognition mechanisms in the synthesis of animal virus DNA. *Biochem J.* **258**, 3-16.
- Hayes, B.W., Telling, G.C., Myat, M.M., Williams, J.F., & Flint S.J. (1990). The adenovirus L4 100 kilodalton protein is necessary for efficient translation of viral late mRNA species. *J. Virol.* **64**, 2732-2742.
- He, X-P., Bataillé, N., & Fried, H.M. (1994). Nuclear export of signal recognition particle RNA is a facilitated process that involves the *Alu* sequence domain. *J. Cell Biol.* **107**, 903-912.
- Hemström, C., Virtanen, A., Bridge, E., Ketner, G., & Pettersson, U. (1991). Adenovirus E4-dependent activation of the early E2 promoter is insufficient to promote the early-to-late-phase transition. *J. Virol.* **65**, 1440-1449.
- Hennache, B., & Boulanger, P. (1977). A biochemical study of KB cell receptor for adenovirus. *Biochem. J.* **166**, 237-247.
- Herisse, J., Rigolet, M., Du Pont de Dinechin, S., & Galibert, F. (1981). Nucleotide sequence of adenovirus 2 DNA fragment encoding the carboxylic region of the fiber protein and the entire E4 region. *Nuc. Acids Res.* **9**, 4023-4042.
- Hopp, T.P. & Woods, K.R. (1981). Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA* **78**, 3824-3828.
- Hoeffler, W.K., Kovelman, R., & Roeder, R.G. (1989). Activation of transcription factor IIIc by the adenovirus E1A protein. *Cell* **53**, 907-920.

Horikoshi, N., Maguire, K., Kralli, A., Maldonado, E., Reinberg, D., & Weinmann, R. (1991). Direct initiation between adenovirus E1A protein and the TATA box binding transcription factor IID. *Proc. Natl. Acad. Sci.* **88**, 5124-5128.

Horwitz, M.S. (1990a). Adenoviridae and their replication. In: *Virology*, second edition. Fields, B.N., Knipe, D.M., Chanock, R.M., Hirsch, M.S., Melnick, J.L., & Monath, T.P., eds. (Raven Press Ltd., New York) pp. 1679-1721.

Horwitz, M.S. (1990b). Adenoviruses. In: *Virology*, second edition. Fields, B.N., Knipe, D.M., Chanock, R.M., Hirsch, M.S., Melnick, J.L., & Monath, T.P., eds. (Raven Press Ltd., New York) pp. 1723-1740.

Huang, M-M., & Hearing, P. (1989a). Adenovirus early region 4 encodes two gene products with redundant effects in lytic infection. *J. Virol.* **63**, 2605-2615.

Huang, M-M., & Hearing, P. (1989b). The adenovirus early region 4 open reading frame 6/7 protein regulates the DNA binding activity of the cellular transcription factor, E2F, through a direct complex. *Genes Dev.* **3**, 1699-1710.

Huang, J., & Schneider, R.J. (1990). Adenovirus inhibition of cellular protein synthesis is prevented by the drug 2-aminopurine. *Proc. Natl. Acad. Sci. USA* **87**, 7115-7119.

Huang, J., & Schneider, R.J. (1991). Adenovirus inhibition of cellular protein synthesis involves inactivation of cap binding protein. *Cell* **65**, 271-280.

Huang, S., & Spector, D.L. (1991). Nascent pre-mRNA transcripts are associated with nuclear regions enriched in splicing factors. *Genes Dev.* **5**, 2288-2302.

Hutchinson, N., & Weintraub, H. (1985). Localization of DNAAase I-sensitive sequences to specific regions of interphase nuclei. *Cell* **43**, 471-482.

Ito, W., Ishiguro, H., & Kurosawa, Y. (1991). A general method for introducing a series of mutations into cloned DNA using the polymerase chain reaction. *Gene* **102**, 67-70.

Izaurralde, E., & Mattaj, I.W. (1992). Transport of RNA between nucleus and cytoplasm. *Sem. Cell Biol.* **3**, 279-288.

Izaurralde, E., Stepinski, J., Darzynkiewicz, E., & Mattaj, I.W. (1992). A cap binding protein that may mediate nuclear export of RNA polymerase-II transcribed RNAs. *J. Cell Biol.* **118**, 1287-1295.

Jackson, D.A., & Cook, P.R. (1985). Transcription occurs at a nucleoskeleton. *EMBO J.* **4**, 919-925.

Jackson, D.A., Hassan, A.B., Errington, R.J., & Cook, P.R. (1993). Visualization of focal sites of transcription within human nuclei. *EMBO J.* **12**, 1059-1065.

- Jameson, B.A. & Wolf, H. (1988). The antigenic index:- a novel algorithm for predicting antigenic determinants. *CABIOS* 4, 181-186.
- Jarmolowski, A., Boelens, W.C., Izaurralde, E., & Mattaj, I.W. (1994). Nuclear export of different classes of RNA is mediated by specific factors. *J. Cell Biol.* 124, 627-635.
- Javier, R.T., (1994). Adenovirus type 9 E4 open reading frame 1 encodes a transforming protein required for the production of mammary tumors in rats. *J. Virol.* 68, 3917-3924.
- Javier, R., Raska, K., Macdonald, G.J., & Shenk, T. (1991). Human adenovirus type 9-induced rat mammary tumors. *J. Virol.* 65, 3192-3202.
- Javier, R.T., Raska, K., & Shenk, T. (1992). Requirement for the adenovirus type 9 E4 region in production of mammary tumours. *Science* 257, 1267-1270.
- Jelsma, T.N., Howe, J.A., Mymryk, J.S., Eveleigh, C.M., Cunniff, N.F.A., & Bayley, S.T. (1989). Sequences in E1A proteins of human adenovirus 5 required for cell transformation, repression of a transcriptional enhancer, and induction of proliferating cell nuclear antigen. *Virology* 171, 120-130.
- Jiminez-Garcia, L.F., & Spector, D.L. (1993). In vivo evidence that transcription and splicing are coordinated by a recruiting mechanism. *Cell* 73, 47-59.
- Johnston, J.M., Anderson, K.P., & Klessig, D.F. (1985). Partial block to transcription of human adenovirus type 2 late genes in abortively infected monkey cells. *J. Virol.* 56, 378-385.
- Jones, C., & Lee, K.A.W. (1991). E1A-mediated activation of the adenovirus E4 promoter can occur independently of cellular transcription factor E4F. *Mol. Cell Biol.* 11, 4297-4305.
- Jones, N., & Shenk, T. (1978). Isolation of deletion and substitution mutants of adenovirus type 5. *Cell* 13, 181-186.
- Jones, N., & Shenk, T. (1979). Isolation of adenovirus type 5 host range deletion mutants defective for transformation of rat embryo cells. *Cell* 17, 683-689.
- Kaelin, W.G.Jr., Pallas, D.C., DeCaprio, J.A., Kaye, F.J. & Livingston, D.M. (1991). Identification of cellular proteins that can interact specifically with the T/E1A-binding region of the retinoblastoma gene product. *Cell*, 64, 521-532.
- Kao, C.C., Yew, P.R., & Berk, A.J. (1990). Domains required for *in vitro* association between cellular p53 and the adenovirus 2 E1B 55K protein. *Virology* 179, 806-814.

- Katze, M.G., DeCorato, D., Safer, B., Galabru, J., & Hovanessian, A.G. (1987). Adenovirus VAI RNA complexes with the 68,000 M_r protein kinase to regulate its autophosphorylation and activity. *EMBO J.* **6**, 689-697.
- Khanna-Gupta, A. & Ware, V.C. (1989). Nucleocytoplasmic transport of ribosomes in a eukaryotic system: Is there a facilitated transport process? *Proc. Natl. Acad. Sci. USA* **86**, 1791-1795.
- Kieff, E., & Liebowitz, D. (1990). Epstein-Barr virus and its replicaiton. In: Virology, second edition. Fields, B.N., Knipe, D.M., Chanock, R.M., Hirsch, M.S., Melnick, J.L., & Monath, T.P., eds. (Raven Press Ltd., New York) pp. 1889-1920.
- Kinloch, R., Mackay, M., & Mautner, V. (1984). Adenovirus hexon, sequence comparison of subgroup C serotypes 2 and 5. *J. Biol. Chem.* **259**, 6431-6436.
- Klessig, D.F., & Chow, L.T. (1980). Incomplete splicing and deficient accumulation of the fiber messenger RNA in monkey cells infected by human adenovirus type 2. *J. Mol. Biol.* **139**, 221-242.
- Kostura, M., & Mathews, M.B. (1989). Purification and activation of the double-stranded RNA-dependent eIF-2 kinase DAI. *Mol. Cell Biol.* **9**,1576-1586.
- Kovesdi, I., Reichel, R., & Nevins, J.R. (1986a). E1A transcription induction: enhanced binding of a factor to upstream promoter sequences. *Science* **231**, 719-722.
- Kovesdi, I., Reichel, R., & Nevins, J.R. (1986b). Identification of a cellular transcription factor invovled in E1A trans-activation. *Cell* **45**, 219-228.
- Kozak, M. (1981). Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. *Nucleic Acids Res.* **9**, 5233-5252.
- Krainer, A.R., Conway, G.C., & Kozak, D. (1990). The essential pre-mRNA splicing factor SF2 influences splice site selection by activating proximal sites. *Cell* **62**, 35-42.
- Krainer, A.R., Mayeda, A., Kozak, D., & Binns, G. (1991). Functional expression of cloned human splicing factor SF2: homology to RNA-binding proteins, U1 70K, and Drosophila splicing regulators. *Cell* **66**, 383-394.
- Kramer, J., Zachar, Z., & Bingham, P.M. (1994). Nuclear pre-mRNA metabolism:channels and tracks. *Trends Cell Biol.* **4**, 35-37.
- Kreivi, J-P., Zerivitz, K., & Akusjarvi, G. (1991). Sequences involved in the control of adenovirus L1 alternative RNA splicing. *Nucleic Acids Res.* **19**, 2379-2386.
- Kreivi, J-P., & Akusjarvi, G. (1994). Regulation of alternative RNA splicing at the level of commitment complex formation. *Nucleic Acids Res.* **22**, 332-337.

- Krug, R.M. (1993). The regulation of export of mRNA from nucleus to cytoplasm. *Curr. Op. Cell Biol.* **5**, 944-949.
- Krystosek, A., & Puck, T.T. (1990). The spatial distribution of exposed nuclear DNA in normal, cancer, and reverse transformed cells. *Proc. Natl. Acad. Sci. USA* **87**, 6560-6564.
- Kyte, J. & Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105-132.
- Ladner, M.B., Martin, G.A., Noble, J.A., Nikoloff, D.M., Tal, R., Kawasaki, E.S., & White, T.J. (1987). Human CSF-1: gene structure and alternative splicing of mRNA precursors. *EMBO J.* **6**, 2693-2698.
- Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.
- Larsson, S., Kreivi, J-P., & Akusjarvi, G. (1991). Control of adenovirus RNA splicing: effect of viral DNA replication on RNA splice site choice. *Gene* **107**, 219-227.
- Larsson, S., Svensson, C., & Akusjarvi, G. (1992). Control of adenovirus major late gene expression at multiple levels. *J. Mol. Biol.* **225**, 287-298.
- Laski, F.A., Rio, D.C., & Rubin, G.M. (1986). Tissue specificity of Drosophila P element transposition is regulated at the level of mRNA splicing. *Cell* **44**, 7-19.
- Lawrence, J.B., Singer, R.H., & Marselle, L.M. (1989). Highly localized tracks of specific transcripts within interphase nuclei visualised by in situ hybridization. *Cell* **57**, 493-502.
- Lazinski, D., Grzadzielska, E., & Das, A. (1989). Sequence-specific recognition of RNA hairpins by bacteriophage antiterminators requires a conserved arginine-rich motif. *Cell* **59**, 207-218.
- Le, S-Y., Malim, M.H., Cullen, B.R., & Maizel, J.V. (1990). A highly conserved RNA folding region coincident with the Rev response element of primate immunodeficiency viruses. *Nucleic Acids Res.* **18**, 1613-1623.
- Lee, K.A.W., & Green, M.R. (1987). A cellular transcription factor E4f1 interacts with an E1A-inducible enhancer and mediates constitutive enhancer function in vitro. *EMBO J.* **6**, 1345-1353.
- Lee, W.S., Kao, C.C., Bryant, G.O. Liu, X. & Berk, A.J. (1991). Adenovirus E1A activation domain binds the basic repeat in the TATA box transcription factor. *Cell* **67**, 365-376.

- Leff, S.E., & Rosenfield, M.G. (1986). Complex transcriptional units: diversity in gene expression by alternative RNA processing. *Annu. Rev. Biochem.* **55**, 1091-1117.
- Lechner, R.L., & Kelly, T.J. (1977). The structure of replicating adenovirus 2 DNA molecules. *Cell* **12**, 1007-1020.
- Legrain, P., & Rosbash, M. (1989). Some *cis*- and *trans*-acting mutants for splicing target pre-mRNA to the cytoplasm. *Cell* **57**, 573-583.
- LeMaire, M.F., & Thummel, C.S. (1990). Splicing precedes polyadenylation during *Drosophila E74A* transcription. *Mol. Cell Biol.* **10**, 6059-6063.
- Leppard, K.N. (1993). Selective effects on adenovirus late gene expression of deleting the E1b 55K protein. *J. Gen. Virol.* **74**, 575-582.
- Leppard, K.N., Pilder, S., Moore, M., Logan, J., & Shenk, T. (1987). An adenovirus oncogene post-transcriptionally modulates mRNA accumulation. In: *Viral Carcinogenesis, Alfred Benzon Symposium 24*. Editors: Kjelgaard, N.O., & Forchhammer, J. Munksgaard, Copenhagen. p 196-208.
- Leppard, K.N., & Shenk, T. (1989). The adenovirus E1B 55kd protein influences mRNA transport via an intranuclear effect on RNA metabolism. *EMBO J.* **8**, 2329-2336.
- Lewis, J.B., & Anderson, C.W. (1987). Identification of adenovirus type 2 early region 1B proteins that share the same amino terminus as do the 495R and 155R proteins. *J. Virol.* **61**, 3879-3888.
- Lewis, J.B., & Mathews, M.B. (1980). Control of adenovirus early gene expression: a class of immediate early gene products. *Cell* **21**, 303-313.
- Lillie, J.W., & Green, M.R. (1989). Transcription activation by the adenovirus E1a protein. *Nature* **338**, 39-44.
- Liu, F., & Green, M.R. (1990). A specific member of the ATF transcription factor family can mediate transcription activation by the adenovirus E1a protein. *Cell* **61**, 1217-1224.
- Logan, J., & Shenk, T. (1984). Adenovirus tripartite leader sequence enhances translation of mRNAs late after infection. *Proc. Natl. Acad. Sci. USA* **81**, 3655-3659.
- Lonberg-Holm, K., Crowell, R.L., & Philipson, L. (1976). Unrelated animal viruses share receptors. *Nature* **259**, 679-681.
- 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 Res.* **18**, 4105-4110.

- McIntosh, E.M., Ager, D.D., Gadsden, M.H., & Haynes, R.H. (1992). Human dUTP pyrophosphatase: cDNA sequence and potential biological importance of the enzyme. *Proc. Natl. Acad. Sci. USA* **89**, 8020-8024.
- McLorie, W., McGlade, C.J., Takayesu, D., & Branton, P.E. (1991). Individual adenovirus E1B proteins induce transformation independently but by additive pathways. *J. Gen. Virol.* **72**, 1467-1471.
- Malette, P., Yee, S-P., & Branton, P.E. (1983). Studies on the phosphorylation of the 58 000 dalton early region 1B protein of human adenovirus type 5. *J. Gen. Virol.* **64**, 1069-1078.
- Malim, M.H., & Cullen, B.R. (1993). Rev and the fate of pre-mRNA in the nucleus: implications for the regulation of RNA processing in eukaryotes. *Mol. Cell Biol.* **13**, 6180-6189.
- Mann, K.P., Weiss, E.A., & Nevins, J.R. (1993). Alternative poly(A) site utilization during adenovirus infection coincides with a decrease in the activity of a poly(A) site processing factor. *Mol. Cell. Biol.* **13**, 2411-2419.
- Maran, A., & Mathews, M.B. (1988). Characterization of the double-stranded RNA implicated in the inhibition of protein synthesis in cells infected with a mutant adenovirus defective for VA RNA_I. *Virology* **164**, 106-113.
- Mariman, E.C.M., van Eekelen, C.A.G., Reinders, R.J., Berns, A.J.M., & van Venrooij (1982). Adenoviral heterogenous nuclear RNA is associated with the host nuclear matrix during splicing. *J. Mol. Biol.* **154**, 103-119.
- Martin, K.J., Lillie, J.W., & Green, M.R. (1990). Evidence for interaction of different eukaryotic transcriptional activators with distinct cellular targets. *Nature* **346**, 147-152.
- Marton, M.J., Baim, S.B., Ornelles, D.A., Shenk, T. (1990). The adenovirus E4 17-kilodalton protein complexes with the cellular transcription factor E2F, altering its DNA-binding properties and stimulating E1A-independent accumulation of E2 mRNA. *J. Virol.* **64**, 2345-2359.
- Mathews, M.B. (1975). Genes for VA-RNA in Ad2. *Cell* **6**, 223-229.
- Mathews, M.B. (1980) Binding of adenovirus VA RNA to mRNA: a possible role in splicing. *Nature* **285**, 575-577.
- Mathews, M.B. & Shenk, T. (1991). Adenovirus virus-associated RNA and translation control. *J. Virol.* **65**, 5657-5662.
- Mayeda, A., & Krainer, A.R. (1992). Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* **68**, 365-375.

- Medford, R.M., Nguyen, H.T., Destree, A.T., Summers, E., & Nadal-Ginard, B. (1984). A novel mechanism of alternative splicing for the developmentally regulated generation of troponin-T isoforms from a single gene. *Cell* **38**, 409-421.
- Mehlin, H., Daneholt, B., & Skoglund, U. (1992). Translocation of a specific pre-messenger ribonucleoprotein particle through the nuclear pore studied with electron microscope tomography. *Cell* **69**, 605-613.
- Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K., & Green, M.R. (1984). Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res* **12**, 7035-7056.
- Messing, J., Gronenborn, B., Müller-Hill, B., & Hofschneider, P. H. (1977). Filamentous coliphage M13 as a cloning vehicle: Insertion of a *Hind*II fragment of the *lac* regulatory region in M13 replicative form *in vitro*. *Proc. Natl. Acad. Sci. USA* **74**, 3642-3646.
- Michaud, N., & Goldfarb, D. (1992). Microinjected U snRNAs are imported to oocyte nuclei via the nuclear pore complex by three distinguishable targeting pathways. *J. Cell Biol.* **116**, 851-861.
- Miller, G. (1990). Epstein-Barr virus: Biology, Pathogenesis, and Medical aspects. In: *Virology*, second edition. Fields, B.N., Knipe, D.M., Chanock, R.M., Hirsch, M.S., Melnick, J.L., & Monath, T.P., eds. (Raven Press Ltd., New York) pp. 1921-1958.
- Miller, M.W., & Hanover, J.A. (1992). Regulation of Macromolecular traffic mediated by the nuclear pore complex. *Cell Biol. Int. Rep.* **16**, 791-798.
- Moen, P.T. Jr., Fox, E., & Bodnar, J.W. (1990). Adenovirus and minute virus of mice DNAs are localized at the nuclear periphery. *Nucleic Acids Res.* **18**, 513-520.
- Montell, C., Fisher, E.F., Caruthers, M.H., & Berk, A.J. (1984). Control of adenovirus E1B mRNA synthesis by a shift in the activities of RNA splice sites. *Mol. Cell. Biol.* **4**, 966-972.
- Moore, M., Schaack, J., Baim, S.B., Morimoto, R.I., & Shenk, T. (1987). Induced heat shock mRNAs escape the nucleocytoplasmic transport block in adenovirus-infected HeLa cells. *Mol. Cell. Biol.* **7**, 4502-4512.
- Moran, E., & Mathews, M.B. (1987). Multiple functional domains in the adenovirus E1A gene. *Cell* **48**, 177-178.
- Mount, S.M. (1982). A catalogue of splice junction sequences. *Nucleic Acids Res.* **10**, 459-472.

- Moyne, G., Pichard, E., & Bernhard, W. (1978). Localization of simian adenovirus 7 (SA 7) transcription and replication in lytic infection. An ultracytochemical and autoradiographical study. *J. Gen. Virol.* **40**, 77-92.
- Mudryj, M., Hiebert, S.W., & Nevins, J.R. (1990). A role for the adenovirus inducible E2F transcription factor in a proliferation dependent signal transduction pathway. *EMBO J.* **9**, 2179-2184.
- Müller, U., Kleinberger, T., & Shenk, T. (1992). Adenovirus E4orf4 protein reduces phosphorylation of c-Fos and E1A proteins while simultaneously reducing the level of AP-1. *J. Virol.* **66**, 5867-5878.
- Neill, S.D., Hemstrom, C., Virtanen, A., & Nevins, J.R. (1990). An adenovirus E4 gene product trans-activates E2 transcription and stimulates stable E2F binding through a direct association with E2F. *Proc. Natl. Acad. Sci. USA* **87**, 2008-2012.
- Neill, S.D., & Nevins, J.R. (1991). Genetic analysis of the adenovirus E4 6/7 *trans* activator: interaction with E2F and induction of a stable DNA-protein complex are critical for activity. *J. Virol.* **65**, 5364-5373.
- Nemerov, G.R., Cheresch, D.A., & Wickham, T.J. (1994). Adenovirus entry into host cells: a role for α_v integrins. *Trends Cell Biol.* **4**, 52-55.
- Neuberger, M.S., & Williams, G.T. (1988). The intron requirement for immunoglobulin gene expression is dependent upon the promoter. *Nucleic Acids Res.* **16**, 6713-6724.
- Neuman de Vegvar, H., & Dahlberg, J.E. (1990). Nucleocytoplasmic transport and processing of small nuclear RNA precursors. *Mol. Cell Biol.* **10**, 3365-3375.
- Nevins, J.R., & Chen-Kiang, S. (1981). Processing of adenovirus nuclear RNA to mRNA. *Adv. Virus Res.* **26**, 1-35.
- Nevins, J.R., & Darnell, J.E. (1978). Steps in the processing of Ad2 mRNA: poly(A)+ nuclear sequences are conserved and poly(A) addition precedes splicing. *Cell* **15**, 1477-1493.
- Nevins, J.R., & Wilson, M.C. (1981). Regulation of adenovirus-2 gene expression at the level of transcriptional termination and RNA processing. *Nature* **290**, 113-118.
- Nevins, J.R., & Winkler, J.J. (1980). Regulation of early adenovirus transcription: a protein product of early region 2 specifically represses region 4 transcription. *Proc. Natl. Acad. Sci. USA* **77**, 1893-1897.
- Nicolas, J.C., Sarnow, P., Girard, M., & Levine, A.J. (1983). Host-range, temperature-conditional mutants in the adenovirus DNA-binding protein are defective in the assembly of infection virus. *Virology* **126**, 228-239.

- Nishikawa, K. (1983). Assessment of secondary-structure prediction of proteins-comparison of computerized Chou-Fasman methods with others. *Biochem. Biophys. Acta* **748**, 285-299.
- Nordqvist, K. Ohman, K., & Akusjarvi, G. (1994). Human adenovirus encodes two proteins which have opposite effects on accumulation of alternatively spliced mRNAs. *Mol. Cell Biol.* **14**, 437-445.
- Nordqvist, K., & Akusjarvi, G. (1990). Adenovirus early region 4 stimulates mRNA accumulation via 5' introns. *Proc. Natl. Acad. Sci. USA* **87**, 9543-9547.
- O'Connor, R.J., & Hearing, P. (1991). The C-terminal 70 amino acids of the adenovirus E4-ORF6/7 protein are essential and sufficient for E2F complex formation. *Nucleic Acids Res.* **19**, 6579-6586.
- O'Malley, R.P., Duncan, R.F., Hershey, J.W.B., & Mathews, M.B. (1989). Modification of the protein synthesis initiation factors and shut-off of host protein synthesis in adenovirus-infected cells. *Virology* **168**, 112-118.
- Obert, S., O'Connor, R.J., Schmid, S., & Hearing, P. (1994). The adenovirus E4-6/7 protein transactivates the E2 promoter by inducing dimerization of a Heteromeric E2F complex. *Mol. Cell Biol.* **14**, 1333-1346.
- Öhman, K., Nordqvist, K., & Akusjärvi, G. (1993). Two adenovirus proteins with redundant activities in virus growth facilitates tripartite leader mRNA accumulation. *Virology* **194**, 50-58.
- Ornelles, D.A., & Shenk, T. (1991). Localization of the adenovirus early region 1B 55-kilodalton protein during lytic infection: association with nuclear viral inclusions requires the early region 4 34-kilodalton protein. *J. Virol.* **65**, 424-439.
- Pelletier, J., & Sonenberg, N. (1988). Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* **334**, 320-325.
- Perricaudet, M., Akusjarvi, G., Virtanen, A., & Pettersson, U. (1979). Structure of two spliced mRNAs from the transforming region of human subgroup C adenoviruses. *Nature* **281**, 694-696.
- Pettersson, U., & Philipson, L. (1975). Location of sequences on the adenovirus genome coding for the 5.5S RNA. *Cell* **6**, 1-4.
- Philipson, L., Lonberg-Holm, K., & Pettersson, U. (1968). Virus-receptor interaction in an adenovirus system. *J. Virol.* **2**, 1064-1075.
- Pilder, S., Moore, M., Logan, J., & Shenk, T. (1986a). The adenovirus E1B-55K transforming polypeptide modulates transport or cytoplasmic stabilisation of viral and host cell mRNAs. *Mol. Cell. Biol.* **6**, 470-476.

- Pilder, S., Leppard, K., Logan, J., & Shenk, T. (1986b). Functional analysis of the adenovirus E1b-55K polypeptide. *Cancer cells* 4, 285-290.
- Piñol-Roma, S., & Dreyfuss, G. (1991). Transcription-dependent and transcription independent nuclear transport of hnRNP proteins. *Science* 253, 312-314.
- Piñol-Roma, S., & Dreyfuss, G. (1992). Pre-mRNA binding proteins shuttle between the nucleus and the cytoplasm. *Nature* 355, 730-732.
- Piñol-Roma, S., & Dreyfuss, G. (1993). hnRNP proteins: localization and transport between the nucleus and the cytoplasm. *Trends Cell Biol.* 3, 151-155.
- Prescott, J., & Falck-Pederson, E. (1992). Varied poly(A) site efficiency in the adenovirus major late transcription unit. *J. Bio. Chem.* 267, 8175-8181.
- Prescott, J., & Falck-Pederson, E. (1994). Sequence elements upstream of the 3' cleavage site confer substrate strength to the adenovirus L1 and L3 polyadenylation sites. *Mol. Cell. Biol.* 14, 4682-4693.
- Price, R., & Penman, S. (1972). Trnascrption of the adenovirus genome by an α -amanitin-sensitive RNA polymerase in HeLa cells. *J. Virol.* 9, 621-626.
- Proudfoot, N. (1991). Poly(A) signals. *Cell* 64, 671-674.
- Queen, C. and Korn, L.J. (1984). A comprehensive sequence analysis program for the IBM Personal Computer. *Nucleic Acids Res.* 12, 581-599.
- Rao. L., Debbas. M., Sabbatini, P., Hockenbery, D., Korsmeyer, S., & White, E. (1992). The adenovirus E1a proteins induce apoptosis, which is inhibited by the E1b 19-kDa and bcl-2 proteins. *Proc. Natl. Acad. Sci. USA* 89, 7742-7746.
- Raychaudhuri, P., Bagchi, S., Neill, S., Nevins, J.R. (1990). Activation of the E2F transcription factor in adenovirus-infected cells involves E1A-dependent stimulation of DNA-binding activity and induction of cooperative binding mediated by an E4 gene product. *J. Virol.* 64, 2702-2710.
- Raychaudhuri, P., Bagchi, S., & Nevins, J.R. (1989). DNA-binding activity of the adenovirus-induced E4F transcription factor is regulated by phosphorylation. *Genes Dev.* 3, 620-627.
- Raychaudhuri, P., Rooney, R., & Nevins, J.R. (1987). Identification of an E1A-inducible cellular factor that interacts with the regulatory sequences within the adenovirus E4 promoter. *EMBO J.* 6, 4073-4081.
- Reichel, P.A., Merrick, W.C., Siekierka, & mathes, M.B. (1985). Adenovirus VA RNA_I regulates the activity of a protein synthesis initiation factor. *Nature* 313, 196-200.

- Reichel, R., Neill, S.D., Kovessi, I., Simon, M.C., Raychaudhuri, P., & Nevins, J.R. (1989). The adenovirus E4 gene, in addition to the E1A gene, is important for *trans*-activation of E2 transcription and E2F activation. *J. Virol.* **63**, 3643-3650.
- Rhoads, R.E. (1988). Cap recognition and the entry of mRNA into the protein synthesis initiation cycle. *Trends Biochem. Sci.* **13**, 52-56.
- Riley, S.J., & Flint, D. (1993). RNA-binding properties of a translational activator, the adenovirus L4 100-kilodalton protein. *J. Virol.* **67**, 3586-3595.
- Roberts, R.J., O'Neill, K.E., & Yen, C.T. (1984). DNA sequences from the adenovirus 2 genome. *J. Biol. Chem.* **259**, 13968-13985.
- Rooney, R.J., Raychaudhuri, P., & Nevins, J.R. (1990). E4F and ATF, two transcription factors that recognize the same site, can be distinguished both physically and functionally: a role for E4F in E1A *trans* activation. *Mol. Cell Biol.* **10**, 5138-5149.
- Rosbash, M., & Singer, R.H. (1993). RNA travel: tracks from DNA to cytoplasm. *Cell* **75**, 399-401.
- Rosenfeld, P.J., O'Neill, E.A., Wides, R.J., & Kelly, T.J. (1987). Sequence-specific interactions between cellular DNA-binding proteins and the adenovirus origin of DNA replication. *Mol. Cell Biol.* **7**, 875-886.
- Ross, D., & Ziff, E. (1992). Defective synthesis of early region 4 mRNAs during abortive adenovirus infections in monkey cells. *J. Virol.* **66**, 3110-3317.
- Ross, D., & Ziff, E. (1994). Defective processing of human adenovirus 2 late transcription unit mRNAs during abortive infections in monkey cells. *Virology* **202**, 107-115.
- Rowe, D.T., Graham, F.L., & Branton, P.E. (1983). Intracellular localization of adenovirus type 5 tumor antigens in productively infected cells. *Virology* **129**, 456-468.
- Rowe, W.P., Huebner, R.J., Gilmore, L.K., Parrott, R.H., & Ward, T.G. (1953). Isolation of a cytopathogenic agent from human adenoids undergoing spontaneous degeneration in tissue culture. *Proc. Soc. Exp. Biol. Med.* **84**, 570-573.
- Sambrook, J., Fritsch, E.F., & Maniatis, T. (1989). Molecular cloning, a laboratory manual, second edition. (Cold Spring Harbor Laboratory Press).
- Samulski, R.J., & Shenk, T. (1988). Adenovirus E1B 55- M_r polypeptide facilitates timely cytoplasmic accumulation of adeno-associated virus mRNAs. *J. Virol.* **62**, 206-210.

- Sandler, A.B., & Ketner, G. (1989). Adenovirus early region 4 is essential for normal stability of late nuclear RNAs. *J. Virol.* **63**, 624-630.
- Sandler, A.B., & Ketner, G. (1991). The metabolism of host RNAs in cells infected by an adenovirus E4 mutant. *Virology* **191**, 319-326.
- Sarnow, P., Ho, Y.S., Williams, J., & Levine, A.J. (1982) Adenovirus E1b-58Kd tumor antigen and SV40 large tumor antigen are physically associated with the same 54Kd cellular protein in transformed cells. *Virology* **120**, 510-517.
- Sarnow, P., Hearing, P., Anderson, C.W., Halbert, D.N., Shenk, T., & Levine, A.J. (1984). Adenovirus early region 1B 58,000-dalton tumor antigen is physically associated with an early region 4 25,000-dalton protein in productively-infected cells. *J. Virol.* **49**, 692-700.
- Schein, C.H. (1989). Production of soluble recombinant proteins in bacteria. *Bio/technology* **7**, 1141-1149.
- Shein, C.H., & Noteborn M.H.M. (1988). Formation of soluble recombinant proteins in *Escherichia coli* is favoured by lower growth temperature. *Bio/technology* **6**, 291-294
- Schneider, R.J., Safer, B., Munemitsu, S.M., Samuel, C.E., & Shenk, T. (1985). Adenovirus VAI RNA prevents phosphorylation of the eukaryotic initiation factor 2 alpha subunit subsequent to infection. *Proc. Natl. Acad. Sci. USA* **82**, 4321-4324.
- Schramayr, S., Caporossi, D., Mak, I., Jelinek, T., & Bacchetti, S. (1990). Chromosomal damage induced by human adenovirus type 12 requires expression of the E1b 55-kilodalton viral protein. *J. Virol.* **64**, 2020-2095.
- Schröder, H.C., Bachmann, M., Diehl-Seifert, B., & Müller, W.G. (1987). Transport of mRNA from nucleus to cytoplasm. *Prog. Nucleic Acid Res.* **34**, 89-142.
- Seth, P., Fitzgerald, D., Ginsberg, H., Willingham, M., & Pastan, I. (1984). Evidence that the penton base of adenovirus is involved in potentiation of toxicity of *Pseudomonas* exotoxin conjugated to epidermal growth factor. *Mol. Cell Biol.* **4**, 1528-1533.
- Seth, P., Pastan, I., & Willingham, M.C. (1985). Adenovirus-dependent increase in cell membrane permeability. *J. Biol. Chem.* **260**, 9598-9602.
- Shaack, J., Ho, W.Y-W., Freimuth, P., & Shenk, T. (1990). Adenovirus terminal protein mediates both nuclear matrix association and efficient transcription of adenovirus DNA. *Genes Dev.* **4**, 1197-1208.
- Sharp, P.A. (1984). Adenovirus transcription. In *The Adenoviruses*, H.S. Ginsberg, ed. (New York: Plenum Press), pp. 173-204.

- Shaw, G., & Kamen, R. (1986). A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell* **46**, 659-667.
- Shaw, A.R., & Ziff, E.B. (1980). Transcripts from the adenovirus-2 major late promoter yield a single early family of 3' coterminal mRNAs and five later families. *Cell* **22**, 905-916.
- Shiroki, K., Hashimoto, S., Saito, I., Fukui, Y., Fukui, Y., Kato, H., & Shimojo, H. (1984). Expression of the E4 gene is required for establishment of soft-agar colony-forming rat cell lines transformed by the adenovirus 12 E1 gene. *J. Virol.* **50**, 854-863.
- Sisodia, S.S., Sollner-Webb, B., & Cleveland, D.W. (1987). Specificity of RNA maturation pathways: RNAs transcribed by RNA polymerase III are not substrates for splicing of polyadenylation. *Mol. Cell Biol.* **7**, 3602-3612.
- Sittler, A., Gallinaro, H., & Jacob, M. (1994). Upstream and downstream *cis*-acting elements for cleavage at the L4 polyadenylation site of adenovirus 2. *Nucleic Acids Res.* **22**, 222-231.
- SivaRaman, L., & Thimmappaya, B. (1987). Two promoter-specific host factors interact with adjacent sequences in an E1A-inducible adenovirus promoter. *Proc. Natl. Acad. Sci. USA* **84**, 6112-6116.
- Slabaugh, M.B., & Roseman, N.A. (1989). Retroviral protease-like gene in the Vaccinia virus genome. *Proc. Natl. Acad. Sci. USA* **86**, 4152-4155.
- Smiley, J.K., Young, M.A., & Flint, S.J. (1990). Intranuclear location of the adenovirus type 5 E1B 55-kilodalton protein. *J. Virol.* **64**, 4558-4564.
- Smith, C.W.J., Patton, G.P., & Bernardo, N.-G. (1989). Alternative splicing in the control of gene expression. *Annu. Rev. Genet.* **23**, 527-577.
- Smith, D.B., Davern, K.M., Board, P.G., Tiu, W.U. Garcia, E.G., & Mitchell, G.F. (1986). M_r 26,000 antigen of *Schistosoma japonicum* recognized by resistant WEHI 129/J mice is a parasite glutathione S-transferase. *Proc. Natl. Acad. Sci. USA* **83**, 8703-8707.
- Smith, D.B., & Johnson, K.S. (1988). Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene* **67**, 31-40.
- Smith, D.B., & Corcoran, L.M. (1992). Supplement 19. Unit 16.7: Expression and purification of glutathione-S-transferase fusion proteins. In: Current protocols in Molecular Biology, supplement 19. Editors: Ausubel, F.M., Brent, R., Kingston, R.E. Moore, D.D., Seidman, J.G., Smith, J.A., & Struhl, K. pp 16.7.1-16.7.8
- Smith, D.H., Kegler, D.M., & Ziff, E.B. (1985). Vector expression of adenovirus type 5 E1a proteins: evidence for E1a autoregulation. *Mol. Cell Biol.* **5**, 2684-2696.

- Sonenberg, N. (1987). Regulation of translation by poliovirus. *Adv. Virus. Res.* **33**, 175-204.
- Spector, D.J., McGrogan, M., & Raskas, H.J. (1978). Regulation of appearance of cytoplasmic RNAs from region 1 of the adenovirus 2 genome. *J. Mol. Biol.* **126**, 395-414.
- Spector, D.L. (1993). Nuclear organization of pre-mRNA processing. *Curr. Op. Cell Biol.* **5**, 442-448.
- Sprengel, M., Schmitz, B., Heuss-Neitzel, D., Zock, C. and Doerfler, W. (1994). Nucleotide sequence of human adenovirus type 12 DNA: comparative functional analysis. *J. Virol.* **68**, 379-389.
- Steenbergh, P.H., & Sussenbach, J.S. (1979). The nucleotide sequence of the right-hand terminus of adenovirus type 5 DNA: implications for the mechanism of DNA replication. *Gene* **6**, 307-318.
- Stein, R.W., Corrigan, M., Yacuik, P., Whelan, J., & Moran, E. (1990). Analysis of E1A-mediated growth regulation functions: Binding of the 300-kilodalton cellular product correlates with E1A enhancer repression function and DNA synthesis inducing activity. *J. Virol.* **64**, 4421-4427.
- Stillman, B. (1986). Functions of the adenovirus E1b tumor-antigens. *Cancer surveys* **5**, 389-404.
- Stillman, B. (1989). Initiation of eukaryotic DNA replication in vitro. *Annu. Rev. Cell Biol.* **5**, 197-245.
- Subramanian, T., Kuppaswamy, M., Gysbers, J., Mak, S., & Chinnadurai, G. (1984). 19-kDa tumor-antigen coded by early region E1b of adenovirus-2 is required for efficient synthesis and for protection of viral-DNA. *J. Biol. Chem.* **259**, 1777-1783.
- Takayesu, D., Teodoro, J.G., Whalen, S.G., Branton, P.E. (1994). Characterisation of the 55K adenovirus type 5 E1B product and related proteins. *J. Gen. Virol.* **75**, 789-798.
- Tal, J.T., Craig, E.A., & Raskas, H.J. (1975). Sequence relationships between adenovirus 2 early mRNA and vira RNA size classes synthesized at 18 hours after infection. *J. Virol.* **15**, 137-144.
- Tan, R., Chen, L., Buettner, J.A., Hudson, D., & Frankel, A.D. (1993). RNA recognition by an isolated α -helix. *Cell* **73**, 1031-1040.
- Telling, G.C., Perera, S., Szatkowski-Ozers, M., & Williams, J. (1994). Absence of an essential regulatory influence of the adenovirus E1B 19-kilodalton protein on viral

growth and early gene expression in human diploid WI38, HeLa, and A549 cells. *J. Virol.* **68**, 541-547.

Teodoro, J.G., Halliday, T., Whalen, S.G., Takayesku, D., Graham, F.L., & Branton, P.E. (1994) Phosphorylation at the carboxy terminus of the 55-kilodalton adenovirus type 5 E1B protein regulates transforming activity. *J. Virol.* **68**, 776-786.

Terns, M.P. & Dahlberg, J.E. (1994). Retention and 5' Cap Trimethylation of U3 snRNA in the nucleus. *Science* **264**, 959-961.

Thimmappaya, B., Weinberger, C., Schneider, R.J., & Shenk, T. (1982). Adenovirus VA1 RNA is required for efficient translation of viral mRNA at late times after infection. *Cell* **31**, 543-551.

Tigges, M., & Raskas, H. (1984). Splice junctions in adenovirus 2 early region 4 mRNAs: multiple splice sites produce 18 to 24 RNAs. *J. Virol.* **50**, 106-117.

Tollefson, A.E., Scaria, A., Saha, S., & Wold, W.S.M. (1992). The 11,600-M_w protein encoded by region E3 of adenovirus is expressed early but is greatly amplified at late stages of infection. *J. Virol.* **66**, 3633-3642.

Tooze, J. (1982). The molecular biology of tumour viruses, Volume 2, DNA tumour viruses, Second edition (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Towbin, H., Staehelin, T., & Gordon, J. (1979). Electrophoretic transfer of protein from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. USA* **76**, 4350-4354.

Virtanen, A., Gilardi, P., Näslund, A., LeMoullec, J.M., Pettersson, U., & Perricaudet, M. (1984). mRNAs from human adenovirus 2 early region 4. *J. Virol.* **51**, 822-831.

Virtanen, A. & Pettersson, U. (1985). Organization of early region 1B of human adenovirus type 2: identification of four differentially spliced mRNAs. *J. Virol.* **54**, 383-391.

Wahler, E., & Keller, W. (1992). The biochemistry of 3'-end cleavage and polyadenylation of messenger RNA precursors. *Annu. Rev. Biochem.* **61**, 419-440.

Walton, T.H., Moen, P.T. Jr., Fox, E., & Bodnar, J.W. (1989). Interactions of minute virus of mice and adenovirus with host cell nuclei. *J. Virol.* **63**, 3651-3660.

Wansink, D.G., Schul, W., van der Kraan, I., van Steensel, B., van Driel, R., & de Jong, L. (1993). Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase-II in domains scattered throughout the nucleus. *J. Cell Biol.* **122**, 283-294.

Weiden, M.D., & Ginsberg, H.S. (1994). Deletion of the E4 region of the genome produces adenovirus DNA concatomers. *Proc. Natl. Acad. Sci. USA* **91**, 153-157.

Weinberg, D.H., & Ketner, G. (1983). A cell line that supports the growth of a defective early region 4 deletion mutant of human adenovirus type 2. *Proc. Natl. Acad. Sci. USA* **80**, 5383-5386.

Weinberg, D.H., & Ketner, G. (1986). Adenoviral early region 4 is required for efficient viral DNA replication and for late gene expression. *J. Virol.* **57**, 833-838.

Weinmann, R., Raskas, H.J., & Roeder, R.G. (1974). Role of DNA-dependent RNA polymerases II and III in transcription of the adenovirus genome late in productive infection. *Proc. Natl. Acad. Sci. USA* **71**, 3426-3420.

White, E., Cipriani, R., Sabbatini, P., & Denton, A. (1991). Adenovirus E1B 19-kilodalton protein overcomes the cytotoxicity of E1A proteins. *J. Virol.* **65**, 2968-2978.

White, E., Sabbatini, P., Debbas, M., Wold, W.S.M., Kusher, D.I., & Gooding, L.R. (1992). The 19-kilodalton adenovirus E1B transforming protein inhibits programmed cell death and prevents cytolysis by tumor necrosis factor- α . *Mol. Cell. Biol.* **12**, 2570-2580.

Wickham, T.J., Mathias, P., Cheresch, D.A., & Nemerow, G.R. (1993). Integrins $\alpha_v\beta_3$ and $\alpha_v\beta_5$ promote adenovirus internalization but not virus attachment. *Cell* **73**, 309-319.

Williams, J., Karger, B.D., Ho, Y.S., Castiglia, C.L., Mann, T., & Flint, S.J. (1986). The adenovirus E1B 495R protein plays a role in regulating the transport and stability of the viral late messages. *Cancer Cells* **4**, 275-284.

Williams, R.D. (1993). Construction and analysis of adenovirus/HIV-1 *rev* recombinants. University of Warwick.

Winberg, G., & Shenk, T. (1984). Dissection of overlapping functions within the adenovirus type 5 E1A gene. *EMBO J.* **3**, 1907-1912.

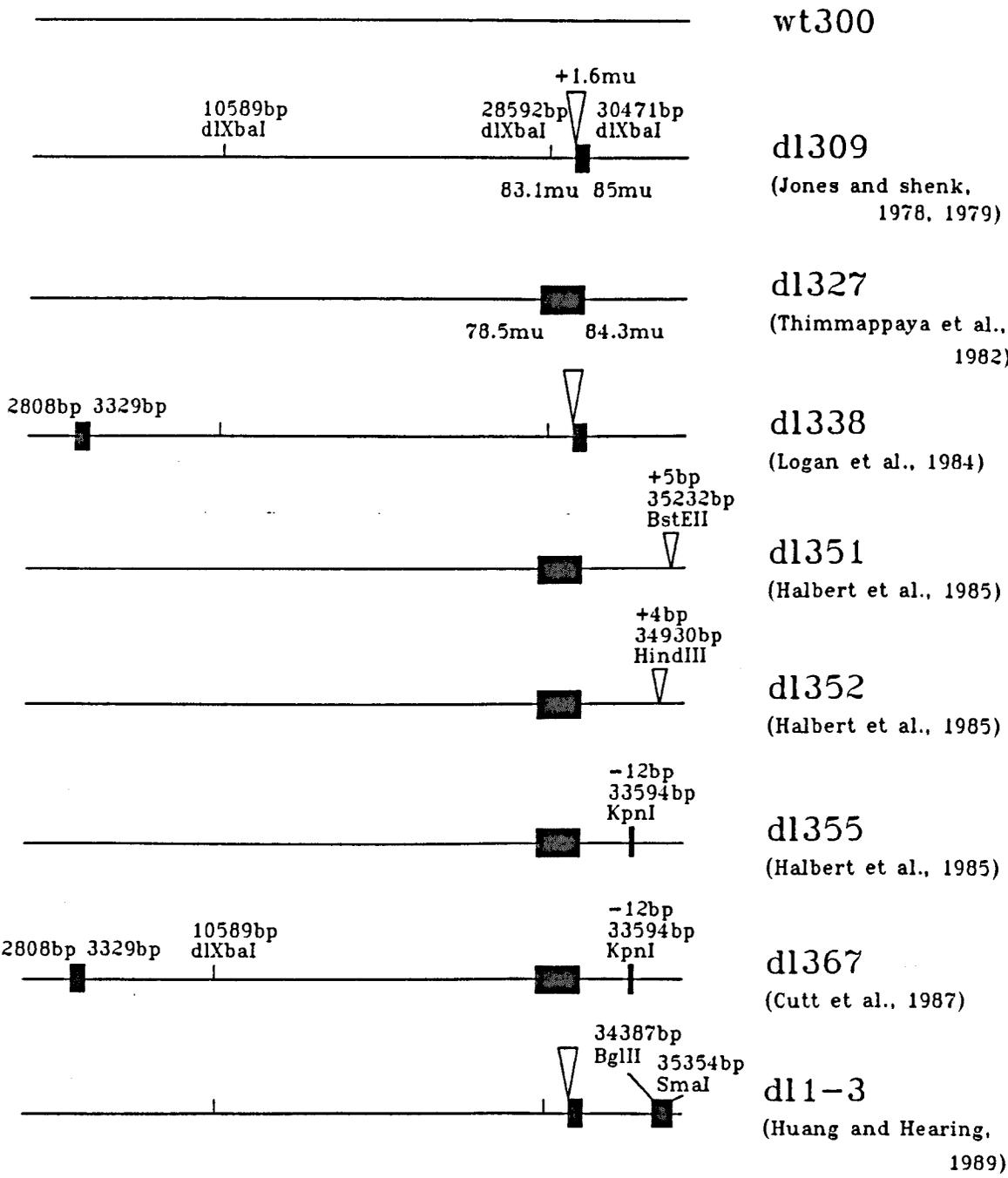
Wold, W.S.M., & Gooding, L.R. (1991). Region E3 of adenovirus: a cassette of genes involved in host immunosurveillance and virus-cell interactions. *Virology* **184**, 1-8.

Wray, W., Boulikas, T., Wray, V.P., & Hancock, R. (1981). Silver staining of proteins in polyacrylamide gels. *Anal. Biochem.* **118**, 197-203.

Xing, Y., & Lawrence, J.B. (1991). Preservation of in vivo RNA distribution within the nuclear matrix demonstrated by in situ hybridization coupled with biochemical fractionation. *J. Cell Biol.* **112**, 1055-1064.

Xing, Y., & Lawrence, J.B. (1993). Nuclear RNA tracks: structural basis for transcription and splicing. *Trends Cell Biol.* **3**, 346-353.

- Xing, Y., Johnson, C.V., Dobner, P.R., & Lawrence, J.B. (1993). Higher level organization of individual gene transcription and RNA splicing. *Science* **259**, 1326-1330.
- Yee, A.S., Reichel, R., Kovetski, I., & Nevins, J.R. (1987). Promoter interaction of the E1A-inducible factor E2F and its potential role in the formation of a multi-component complex. *EMBO J.* **6**, 2061-2068.
- Yew, P.R., Kao, C.C., & Berk, A.J. (1990). Dissection of functional domains in the adenovirus-2 early 1B-55K-polypeptide by suppression linker insertional mutagenesis. *Virology* **179**, 795-805.
- Yew, P.R., & Berk, A.J. (1992). Inhibition of p53 transactivation required for transformation by adenovirus early 1B protein. *Nature* **357**, 82-85.
- Yoder, S.S., & Berget, S.M. (1986). Role of adenovirus type 2 early region 4 in the early-to-late switch during productive infection. *J. Virol.* **60**, 779-781.
- Yoshida, K., Venkatesh, L., Kuppuswamy, M., & Chinnadurai, G. (1987). Adenovirus transforming 19-Kd T antigen has an enhancer-dependent *trans*-activation function and relieves enhancer repression mediated by viral and cellular genes. *Genes Dev.* **1**, 645-658.
- Zachar, Z., Kramer, J., Mims, I.P., & Bingham, P.M. (1993). Evidence for channeled diffusion of pre-mRNAs during nuclear RNA transport in metazoans. *J. Cell Biol.* **121**, 729-742.
- Zasloff, M. (1983). tRNA transport from the nucleus in a eukaryotic cell: carrier-mediated translocation process. *Proc. Natl. Acad. Sci. USA* **80**, 6436-6440.
- Zeitlin, S., Parent, A., Silverstein, S., & Efstratiadis, A. (1987). Pre-messenger RNA splicing and the nuclear matrix. *Mol. Cell Biol.* **7**, 111-120.
- Zeitlin, S., Wilson, R.C., & Efstratiadis, A. (1989). Autonomous splicing and complementation of *in vivo*-assembled spliceosomes. *J. Cell Biol.* **108**, 765-777.
- Zhang, Y., & Schneider, R.J. (1993). Adenovirus inhibition of cellular protein synthesis and the specific translation of late viral mRNAs. *Sem. Virol.* **4**, 229-236.
- Zhang, X.L., Bellett, A.J.D., Tha Hla, R., Braithwaite, A.W., & Mullbacher, A. (1991). Adenovirus type 5 E3 gene products interfere with the expression of the cytolytic T cell immunodominant E1a antigen. *Virology* **180**, 199-206.
- Zhang, X.L., Bellett, A.J.D., Tha Hla, R., Voss, T., Mullbacher, A., & Braithwaite, A.W. (1994). Down-regulation of human adenovirus E1a by E3 gene products: evidence for translational control of E1a by E3 14.5K and/or E3 10.4K products.



■ Deletion

▽ Insertion

Appendix 1

The figure on the facing page depicts the Ad5 physical maps of the viruses used in this study. All the viruses were based on either dl309 (dl338, dl1-3) or dl327 (dl351, dl352, dl355), (dl367 contains mutations from both viruses). Deleted regions are depicted as filled rectangles while inserted DNA is shown as open triangles. The deleted XbaI sites of dl309 are depicted (dlXbaI) as lines (the XbaI site at 1339 bp is intact in dl309 and the dl309 derived viruses shown).

Appendix 2

Details of the oligonucleotides used in this study. All oligonucleotides are written in the 5' to 3' direction.

IDB: ATATTGTCGTTAGAACGCGG		Homology to the SP6 primer region of the pGEM plasmids
IDC: TGATGGATCCACTGTTATGGCTGCCG	<u>Homology to Ad5</u> BamHI	Homology: 35533-35517bp
IDD: AGGTGAATTCTCAAACATTAGAAGCCTGTC	<u>Homology to Ad5</u> EcoRI	Homology: 35143-35159bp
IDE: TGATGGATCCGCAGACATGTTTGAGAGA	<u>Homology to Ad5</u> - BamHI	Homology: 35098-35083bp
IDF: AGGTGAATTCCTCAAGCAGCGAATC	<u>Homology to Ad5</u> EcoRI	Homology: 34686-34701bp
IDG: AGCAAGTATATAGCATGGCC		Homology to the region upstream of the MCS of pGEX-2T
IDH: GGAGCTGCATGTGTCAGAGG		Homology to the region downstream of the MCS of pGEX-2T
IDJ: ACAGCTCCTCGGTCATGTCC	<u>Homology to Ad5</u>	Homology: 35259-35278bp
IDK: TTTTACTAAGCTTGCTGACTGTTATGGC	<u>Homology to Ad5</u> HindIII	Homology: 35550-35543bp Homology: 35536-35522bp
IDL: TAGCCCTGGGGAATACATAC	<u>Homology to Ad5</u> <u>Homology to Ad5</u>	Homology: 35351-35356bp Homology: 35358-35370bp
IDM: TTCATCCACGGTGCCTGACT		Homology to pBR322

Appendix 3

A list of the important sites in this study within the Ad5 E4 region.

<u>Site</u>	<u>Location (bp)</u>
<i>Hae</i> III	35616
Transcription start	36614-35609
D1	35542
ORF1 start	35526
<i>Sma</i> I	35354
ORF1 end	35144
A1a	35104
ORF2 start	35092
<i>Hind</i> III	35930
A1b	34732
ORF2 end	34703
ORF3 start	34703
D2a	34602
A1c	34432
<i>Bg</i> II	34387
ORF3 end	34356
ORF4 start	34342
A1d	34329
D2b	34289
A4	34241
<i>Bg</i> II	34115
A1e	34082
ORF6 start	34077
ORF4 end	34001
D3	33905
<i>Pst</i> I	33875
<i>Kpn</i> I	33594
A5	33283
A3	33192
<i>Sma</i> I	33091
ORF6 end	32917
pA	33803

Appendix 4 - Quantitation of the E4 mRNAs

To determine the relative levels of the E4 mRNAs in infected HeLa cells, RNA from such cells was assayed by RNase protection analysis (see section 2.3.5. and Chapter 4.). The signals from the protected fragments characteristic of each mRNA in question, present on an autoradiograph, were quantified using laser scanning densitometry and the Imagequant image analysis software. This allowed the determination of the density of each signal and so the amount of each protected fragment present in the gel. This data had to be adjusted to take into account the number of uridine residues in the protected fragment as this was the only labelled nucleotide in the probe RNA; a protected fragment which contained a lot of uridine residues due to the nature of the probe will give a stronger signal than an equimolar protected fragment with less uridine residues. This allowed the comparison of the levels of different protected fragments from different or the same probe. However this was only applicable for data produced using probes made in parallel, i.e. which contain the same ratio of labelled to unlabelled uridine residues. Probes made using the same protocol but on different occasions potentially have a different ratio of labelled to unlabelled uridine residues due to small errors in pipetting and variations in the specific activity of the labelled uridine used. A full data set was obtained by linking the results for groups of mRNAs having one or more members in common. It was decided to present this processed data semi-quantitatively (Table 4.1.) and not as fully quantitative figures as the adjustment of the primary data to take into account the uridine residues and the linking of mRNA data sets potentially introduced inaccuracies making fully quantitative data possibly misleading.

Example:

Displayed below are the unadjusted densitometry readings followed by their adjustment according to their uridine (U) content to give comparable figures:

RNA:	A	C	D	L	M
Primary data:					
OD 9h p.i.	40	217	3680	30	0
OD 20h p.i.	2020	362	5000	11310	452
U residues/ protected frag.	79	34	79	45	85
Adjusted figures:					
9 h p.i.	0.5	6.3	46.6	0.7	0
20h p.i.	25.6	10.6	63.3	251.3	5.3

The 9 hour data less than 0 was recorded as a negative in Table 4.1. as no protected fragment was visible on the autoradiographs quantitated and the values shown represent background measurements (a (+) was displayed for mRNA L at 9 hours as a band was just visible on over exposed autoradiographs). The value of 251.3 units for mRNA L at 24 hours was the highest value of all the mRNAs and recorded as ++++. All other were recorded as relative to this, e.g. mRNA C (6.3 and 10.6 units) = +; mRNA A (25.6 units) = ++; mRNA D/E (63.3 units) = +++.

Open reading frames 1 and 2 of adenovirus region E4 are conserved between human serotypes 2 and 5

Ian Dix and Keith N. Leppard*

Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, U.K.

The E4 region of human adenovirus type 2 is predicted to encode seven proteins as judged from its nucleotide sequence and the pattern of differential splicing of its transcript. Two of the open reading frames (ORFs), ORF1 and ORF2, had been identified as being

disrupted in the recently published sequence of the related serotype 5 virus. These ORFs were resequenced and found to be intact in the wt300 strain of adenovirus type 5.

The molecular genetics of human adenovirus type 5 (Ad5) have been studied extensively. This work has been underpinned by sequence data derived from the Ad5 genome supplemented by data from the closely related Ad2, for which a complete sequence has been available for a number of years (Roberts *et al.*, 1984). The completion of a composite Ad5 sequence (Chroboczek *et al.*, 1992; GenBank accession no. M73260) has permitted those studies of Ad5 involving previously unsequenced regions to proceed on a more systematic basis.

The E4 region of Ad5 was one of those for which, until recently, only a partial sequence was available. The Ad2 sequence, together with RNA mapping studies in Ad2 (Virtanen *et al.*, 1984; Tigges & Raskas, 1984), predicted the expression of seven distinct proteins from the Ad2 E4 region and, by analogy, from the Ad5 E4 region also. Three of these proteins, the products of open reading frames (ORFs) 3, 6 and 6/7, have been detected in infected cells (Downey *et al.*, 1983; Sarnow *et al.*, 1984; Cutt *et al.*, 1987). The remainder, the products of ORF1, ORF2, ORF3/4 and ORF4, appear to perform no essential function during infections of cultured HeLa cells because mutants unable to express these proteins grow with virtually wild-type kinetics (Halbert *et al.*, 1985; Huang & Hearing, 1989).

Comparison of the published Ad2 and Ad5 E4 region sequences showed two nucleotide substitutions and four deletions in Ad5 ORF1, and four substitutions and one insertion in Ad5 ORF2; these cause frameshifts in both ORF1 and ORF2, and two missense changes in ORF2. However, the high ratio of frameshift to point mutations

in these putative coding regions, together with the high degree of sequence conservation as compared with adjacent non-coding regions, suggested that the frameshifts might be sequencing artefacts. To determine whether Ad5 could encode E4 ORF1 and ORF2 products, the region of Ad5 DNA between nucleotides 34790 and 35615 was resequenced.

Relevant pieces of the Ad5 strain wt300 *KpnI* G genomic DNA fragment were subcloned into plasmids and then M13 vectors mp18 and mp19. Sequencing was carried out by the dideoxynucleotide chain termination method using the Sequenase II system (U.S. Biochemicals). As shown in Fig. 1, in each of the four regions where the published Ad5 sequence contained frameshifts in ORF1 or ORF2, the sequence obtained showed the ORFs to be open as in Ad2, with the exception of a silent base change at one of the insertion positions. All except one of the other expected differences between Ad2 and Ad5, including all those in the non-coding regions sequenced, were confirmed by our data, discounting the possibility that the DNA sequenced was of Ad2 origin. The exception was the ORF2 missense mutation at Ad5 nucleotide 34860; our data indicate that strain wt300 ORF2 encodes the same amino acid as that of Ad2 at this position. These changes to the database sequence of Ad5 are summarized in Table 1.

ORF1 and ORF2 of the Ad5 E4 region were originally sequenced by Steenberg & Sussenbach (1979) using the chemical degradation technique; this sequence was used by Chroboczek *et al.* (1992) when compiling the complete Ad5 sequence. Our data suggest a revised length of 35938 bp for the Ad5 strain wt300 genome with potentially functional E4 ORF1 and ORF2 regions encoding products identical to, or differing in only one amino acid from, their Ad2 counterparts.

The nucleotide sequence data reported here will appear in the DDBJ, EMBL and GenBank databases under accession number D12587.

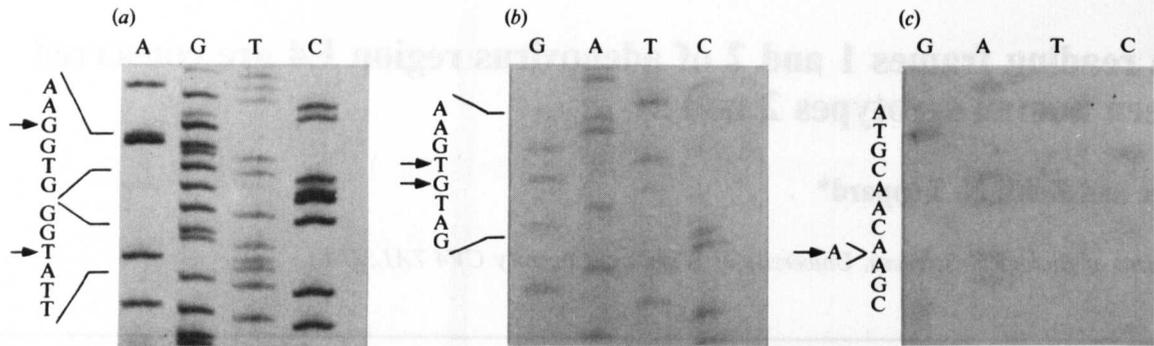


Fig. 1. Sequence data for regions of difference between published Ad5 and Ad2 sequences. The regions around (a) nucleotides 35523/22 and 35511/10, (b) nucleotides 35321/20 and (c) nucleotide 34936 are shown. The sequence of the rightward strand deduced from these data is shown to the left of each panel, with the differences from the published Ad5 sequence indicated by arrows.

Table 1. Summary of changes to the published Ad5 sequence

Published Ad5 nucleotide position	Published Ad5 sequence*	Revised Ad5 sequence*
34860	T	A
34936	T	-
35320/21	-	CA
35510/11	-	C
35522/23	-	A

* Sequence changes indicated are to the leftward strand.

The conservation of ORF1 and ORF2 in two, albeit closely related, serotypes supports the argument that they encode proteins having some function in Ad infections. Since they are not needed for successful infection of HeLa cells, they may be important only in infection of certain cell types or alternatively may be necessary only for infections of the whole organism. Very few of the potential coding regions of Ad5 lack an assigned function; furthering the molecular understanding of this virus by determining those remaining functions is an interesting and important objective for the future.

Ian Dix is the recipient of an SERC research studentship. This work was supported in part by a grant from the MRC.

References

- CHROBOCZEK, J., BIEBER, F. & JACROT, B. (1992). The sequence of the genome of adenovirus type 5 and its comparison with the genome of adenovirus type 2. *Virology* **186**, 280-285.
- CUTT, J. R., SHENK, T. & HEARING, P. (1987). Analysis of adenovirus early region 4-encoded polypeptides synthesized in productively infected cells. *Journal of Virology* **61**, 543-552.
- DOWNEY, J. F., ROWE, D. T., BACCHETTI, S., GRAHAM, F. L. & BAYLEY, S. T. (1983). Mapping of a 14000-dalton antigen to early region 4 of the human adenovirus 5 genome. *Journal of Virology* **45**, 514-523.
- HALBERT, D. N., CUTT, J. R. & SHENK, T. (1985). Adenovirus early region 4 encodes functions required for efficient DNA replication, late gene expression, and host cell shutoff. *Journal of Virology* **56**, 250-257.
- HUANG, M.-M. & HEARING, P. (1989). Adenovirus early region 4 encodes two gene products with redundant effects in lytic infection. *Journal of Virology* **63**, 2605-2615.
- ROBERTS, R. J., O'NEILL, K. E. & YEN, C. T. (1984). DNA sequences from the adenovirus 2 genome. *Journal of Biological Chemistry* **259**, 13968-13985.
- SARNOW, P., HEARING, P., ANDERSON, C. W., HALBERT, D. N., SHENK, T. & LEVINE, A. J. (1984). Adenovirus early region 1B 58,000-dalton tumor antigen is physically associated with an early region 4 25,000-dalton protein in productively infected cells. *Journal of Virology* **49**, 692-700.
- STEENBERGH, P. H. & SUSSENBAUGH, J. S. (1979). The nucleotide sequence of the right-hand terminus of adenovirus type 5 DNA: implications for the mechanism of DNA replication. *Gene* **6**, 307-318.
- TIGGES, M. & RASKAS, H. (1984). Splice junctions in adenovirus 2 early region 4 mRNAs: multiple splice sites produce 18 to 24 RNAs. *Journal of Virology* **50**, 106-117.
- VIRTANEN, A., GILARDI, P., NASLUND, A., LEMOULLEC, J. M., PETERSSON, U. & PERRICAUDET, M. (1984). mRNAs from human adenovirus 2 early region 4. *Journal of Virology* **51**, 822-831.

(Received 24 June 1992; Accepted 21 July 1992)

Regulated Splicing of Adenovirus Type 5 E4 Transcripts and Regulated Cytoplasmic Accumulation of E4 mRNA

IAN DIX AND KEITH N. LEPPARD*

Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, England

Received 12 November 1992/Accepted 10 March 1993

The E4 gene of human type C adenoviruses has been shown previously to give rise to an array of mRNAs via differential splicing. In this study, the pattern of expression of these mRNAs during lytic infection was examined, and two distinct temporal classes were defined. mRNAs of the early class were distinguished from those of the late class by the presence, in the early class, of a sequence in the 3' half of the mRNA that was removed as an intron in the late class. A single mRNA of the late class was found to show a strong dependence on the presence of the 55-kDa protein from region E1b and the open reading frame 6 protein from region E4 for its normal cytoplasmic accumulation. One feature of this mRNA that distinguishes it from other E4 mRNAs expressed at late times is the retention within it of an intron from the 5' half of E4; it may therefore be recognized as incompletely spliced by the host cell and retained in the nucleus. It is proposed that the E1b 55-kDa/E4 open reading frame 6 protein complex facilitates accumulation of this mRNA by overcoming this retention mechanism.

The lytic infectious cycle of human adenovirus type 5 (Ad5) is divided by convention into early and late phases, separated by the initiation of viral DNA replication. The products of genes expressed during the early phase of infection are generally involved in the regulation of gene expression and viral DNA replication, whereas the products of genes expressed after DNA replication function in the structure or assembly of the particle (reviewed in reference 14). However, expression of most of the viral genes is not strictly confined to only one phase of the infection: early genes are still transcribed well into the late phase of infection, and restricted expression from the major late promoter occurs at early times (reviewed in reference 9). Most Ad5 primary transcripts are differentially spliced to allow expression of the virus's many open reading frames (ORFs) from a limited number of promoters, and the pattern of differential splicing observed changes during the course of infection. This is exemplified by expression of the L1 segment of the major late transcription unit, during which only large mRNAs are produced at early times, specifying the 52- and 55-kDa proteins, and predominantly a distinct smaller species at late times, specifying protein IIIa (1, 19).

The E4 region of Ad5 is located between 91.3 and 96.8 map units on the viral genome and is transcribed from right to left. E4 transcription initiates at a unique site, beginning in the early phase and continuing into the late phase of infection. RNA transcribed from the equivalent region of Ad2 is subject to a complex pattern of differential splicing, producing a set of mRNAs that share a common poly(A) site (Fig. 1) (10, 13, 32, 33). These mRNAs potentially encode at least seven different polypeptides, products of ORF1, ORF2, ORF3, ORF3/4, ORF4, ORF6, and ORF6/7, of which three, ORF3, ORF6, and ORF6/7, have been identified during the course of infection of HeLa cells (5, 7, 29, 30). Studies of appropriate mutants have implicated these proteins in regulating late protein synthesis, nuclear and cytoplasmic accumulation of late messages, viral protein synthesis at late times, and the activation of the E2a promoter (3, 11, 15, 28,

34). Recently, the ORF4 product has been shown to regulate the phosphorylation of certain viral and cellular proteins (24).

Previous studies of E4 mRNAs indicated some degree of temporal regulation of the pattern of splicing (27, 32), although individual mRNA species were not identified. To characterize the pattern of expression of the known and putative E4 products in more detail, a systematic study of the levels of individual mRNAs during the infectious cycle was undertaken, revealing distinct early and late classes of E4 mRNA. Given the established role of the E1b 55-kDa and E4 ORF6 proteins in regulating cytoplasmic levels of the major late mRNAs (2, 26, 35), the effect on E4 mRNA levels of deleting these functions was examined. Accumulation of one E4 mRNA species was strongly dependent on both E1b 55-kDa and E4 ORF6 protein functions. The structural features and pattern of expression of this mRNA support the idea that the presence of intact splice donor and/or acceptor signals within an mRNA is a determinant of E1b dependence for efficient mRNA efflux from the late adenovirus-infected cell nucleus.

MATERIALS AND METHODS

The origins of Ad5 *dl309*, *dl338*, *dl355*, and *dl367* and methods for their growth in cell culture have been described previously (5, 11, 26). Infections were at a multiplicity of 10 PFU per cell except as noted. Cytoplasmic and nuclear fractions of infected HeLa cells were prepared by Nonidet P-40 lysis and subsequent low-speed centrifugation, and total RNA was prepared from each fraction by phenol-chloroform extraction in the presence of sodium dodecyl sulfate and EDTA, all as previously described (18).

Quantitative RNase protection assays were carried out as described by Melton et al. (21), using unselected RNA and an excess of [α - 32 P]UTP-labeled antisense probe RNA transcribed in vitro from cloned fragments of Ad5 genomic DNA in pGEM vectors (Promega). The positions on the Ad5 genome of the E4 probes generated for this study are shown in Fig. 1. The E1a probe has been described previously (18) and was the generous gift of J. Schaack. Protected fragments

* Corresponding author.

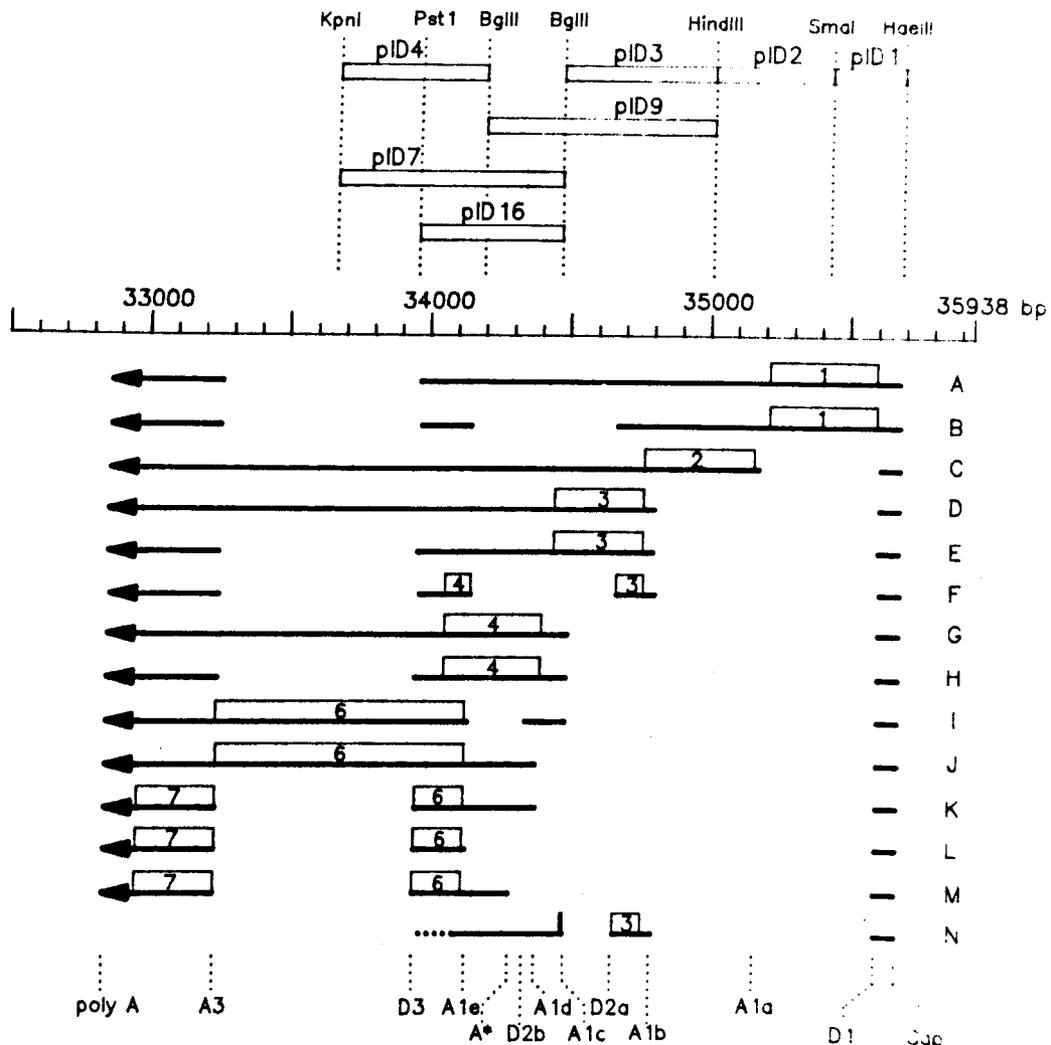


FIG. 1. Constructs used. The right terminal portion of the Ad5 genome (bp 32500 to 35938) is shown as a line scale. Below the line is a transcription map inferred from studies of the closely related virus, Ad2. mRNAs A to L were described by Virtanen et al. (33), and mRNAs M and N are additional species described by Freyer et al. (10). Splice sites are indicated below the transcription map in the notation of Virtanen et al. (33). The additional splice acceptor site inferred from the detection of mRNA M is labeled A*. Minor splice acceptor sites in the 3' portion of E4 and the mRNAs deriving from their use (10, 32) are not shown; the study reported here did not distinguish these mRNA species. The positions on the genome map of the cloned fragments from which RNA probes were derived are shown at the top (pID1, -2, -3, etc).

were separated on 5% polyacrylamide-7 M urea gels and detected by exposure to preflashed Fuji RX film at -70°C. Protected fragments were quantified by laser scanning densitometry (Molecular Dynamics Corp.), using suitable film exposures. To estimate relative levels of different RNA species, these data were corrected to take account of the number of uridine residues in the protected fragment, using information from the Ad5 sequence (4, 6).

RESULTS

Temporal classes of E4 mRNA. The E4 region primary transcript is spliced to produce at least 14 distinct mRNAs, as indicated in Fig. 1 (10, 31, 32). To examine further the temporal control of E4 RNA splicing, the cytoplasmic level of each E4 mRNA was determined at various times after infection with wild-type Ad5. Each RNA probe used gave protected fragments of characteristic length, allowing iden-

tification and quantitation of individual mRNAs, or in some cases of pairs of mRNAs predicted to encode the same translation product. mRNA-probe RNA duplexes covering the region around position 34350 consistently showed sensitivity to RNase that resulted in 50 to 80% cleavage of the probe. Fragment lengths were in each case consistent with an mRNA-probe RNA discontinuity at this position, although no splice donor or acceptor sequence was expected. Inspection of the sequence revealed a stretch of 11 A residues in the E4 mRNA. Stuttering of RNA polymerase on this sequence, either in the cell or, in the antisense, during probe synthesis in vitro, could explain this observation.

Results of representative assays are shown in Fig. 2. The probe used to detect mRNA A did not distinguish mRNAs A and B. However, other probes showed that mRNA B was not present in these experiments. Similarly, the probe for mRNA D did not distinguish mRNAs D and E. However, from results with other probes detecting mRNAs G and H, I

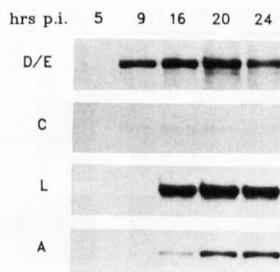


FIG. 2. Analysis of cytoplasmic levels of E4 mRNAs during infection of HeLa cells by wild-type virus, *dl309*, by RNase protection assay. Total cytoplasmic RNA was isolated at the time p.i. indicated above each lane. Ten micrograms of RNA was probed for specific E4 mRNAs, indicated at the left. Fragments shown: D/E, pID3 (345 nucleotides); C, pID2 (174 nucleotides); L, pID7 (178 nucleotides); A, pID1 (258 nucleotides).

and J, and E and H, it could be deduced that mRNA E was significantly less abundant than mRNA D and was expressed with different kinetics (see below). The major component of the protection shown for probe D+E is therefore due to mRNA D. A similar conclusion could be reached for levels of mRNA G in relation to mRNA H (data not shown).

Two distinct classes of E4 mRNAs were identified by this analysis. mRNAs belonging to the early class (mRNA D; Fig. 2) were first detected in cytoplasmic RNA 5 h postinfection (p.i.). Amounts of these mRNAs rose rapidly to 9 h p.i. and then remained constant or declined slightly during the late phase of infection. The small increase in the level of mRNAs D and E that was observed between 9 and 16 h p.i. is consistent with estimates of mRNA E synthesis from other experiments. mRNAs of the late class were only barely detectable at 9 h p.i., and levels increased dramatically thereafter. mRNA A was observed in low amounts at 16 h p.i., but its levels did not peak until 20 h p.i. Other RNAs with late kinetics (mRNA L; Fig. 2) reached maximum or near-maximum levels at 16 h p.i. The temporal classification of each E4 mRNA detected is shown in Table 1.

Relative abundance of E4 mRNAs. Accurate determination of the relative amounts of the different mRNAs is complicated by the need to assume that different probe RNAs were labeled to the same specific activity. This assumption is reasonable only for probes prepared in parallel from the

TABLE 1. Classification of E4 mRNAs

E4 RNA	Relative amt		Kinetic class
	9 h p.i.	20 h p.i.	
A	—	++	Late
B	—	—	
C	+	+	Early
D	+++	+++	Early
E ^a	—	++	Late
F	—	—	
G	++	++	Early
H ^a	—	++	Late
I	—	—	
J	+	+	Early
K	—	++	Late
L	(+)	++++	Late
M	—	+	Late
N	—	—	

^a Amounts shown are for mRNAs E and H collectively.

same mixture of α -³²P-labeled and unlabeled UTP. For such probe sets, protected fragments were quantified, and length-adjusted data were taken as a measure of relative mRNA levels. A full data set was obtained by comparing the results for groups of mRNAs having one or more members in common. These estimates of relative E4 mRNA levels, which are in reasonable agreement with previous data for Ad2 (32), are shown in Table 1. mRNAs E and H have not been identified individually; however, neither is found in significant amounts at 9 h p.i., and their collective abundance is significantly less than that of mRNA D. mRNAs B, F, I, and N were not detected in any experiment.

Late E4 mRNAs depend on DNA replication for accumulation. Transcription of Ad5 late genes is severely restricted prior to the onset of viral DNA replication. To determine whether expression of the late class of E4 mRNAs was similarly dependent on viral DNA synthesis, their levels in cells infected in the presence of the DNA synthesis inhibitor hydroxyurea were assayed. As shown in Fig. 4, expression of mRNA D (early class) in a wild-type infection was unaffected by the inhibitor, whereas accumulation of the late E4 mRNAs L and A was largely and completely prevented, respectively. Thus, these late E4 mRNAs, like other viral late mRNAs, depend on DNA replication for their expression.

Dependence of late E4 expression on E1b 55-kDa protein function. Previous work has shown that cytoplasmic mRNA accumulation for genes expressed from the Ad5 genome at late times is dependent on a function provided by the E1b 55-kDa protein product (2, 18, 26, 34). The possible dependence for proper accumulation of the late class of E4 mRNAs, as defined above, on this E1b function was therefore examined by using a mutant virus, *dl338* (26), which is unable to express the E1b 55-kDa protein.

E4 mRNAs present during a *dl338* infection were quantified in a comparative analysis with wild-type virus (*dl309*). Representative analyses are shown in Fig. 3. Results with the control E1a probe confirmed that the multiplicities of infection of the two viruses were similar. Because E1a products are involved in the activation of the E4 promoter, E1a mRNA appears in the cytoplasm before E4 mRNA. The low levels of E1a mRNA detected at 5 h p.i. in this experiment may explain why little or no E4 mRNA was detected at this time point. Other experiments have shown the presence of early E4 mRNAs at 5 h p.i.

As expected, the early class of E4 mRNA (mRNA D; Fig. 3) accumulated in the cytoplasm of wild-type- and *dl338*-infected cells with identical kinetics. More surprisingly, most of the late class of E4 mRNAs showed only minimal dependence on E1b 55-kDa protein function for normal accumulation (mRNA L; Fig. 3). The one exception was mRNA A, which had shown the latest kinetic profile in a wild-type infection. Accumulation of this mRNA was found to be strongly dependent on E1b function, with cytoplasmic levels in *dl338*-infected cells only 5 to 10% of the wild-type level at 20 to 24 h p.i.

E4 ORF6 is necessary for E4 mRNA A accumulation. E1b 55-kDa protein exists as a molecular complex with the E4 ORF6 protein in infected cells (28), and in previous studies, virus *dl355* (lacking E4 ORF6) and *dl367* (lacking E4 ORF6 and E1b 55-kDa protein) showed phenotypes very similar to that of E1b 55-kDa mutant *dl338* in respect of late gene expression (5, 11). To confirm that the effect of the E1b mutation on levels of E4 mRNA A was due directly to absence of the E1b mRNA transport regulatory function, the levels of mRNA A in E4 ORF6 mutant infections were

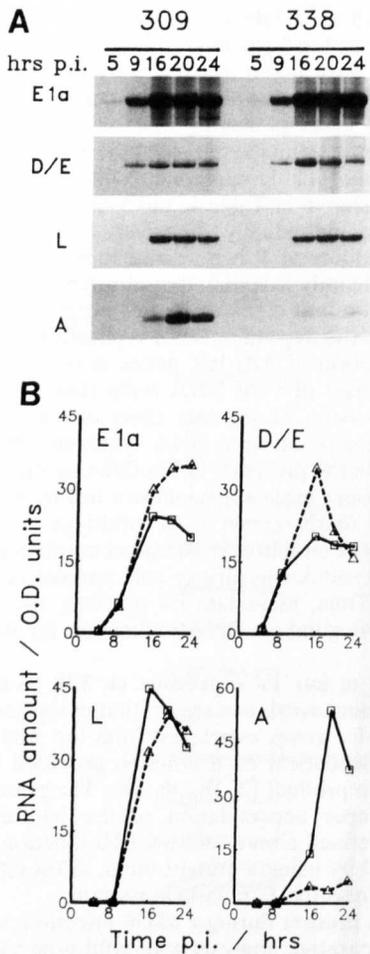


FIG. 3. (A) Analysis of cytoplasmic levels of E1a and E4 mRNAs during infection of HeLa cells by wild-type virus *dl309* or E1b 55-kDa mutant virus *dl338*, as indicated. The E1a fragment shown is protected by the 3' common exon (404 nucleotides). Other details are as for Fig. 2. (B) Quantification of data shown in panel A. Results of laser scanning densitometry are shown in arbitrary units plotted against time p.i. for *dl309* (□) and *dl338* (△). These data are not corrected for protected fragment length and specific activity differences; therefore, amounts are not comparable between panels. O.D., optical density.

examined (Fig. 4). Levels of mRNA D (early class) were unaffected at 9 h p.i. in these infections and only marginally affected at 24 h p.i.; mRNA L showed a similar pattern at 24 h p.i. These marginal reductions at late times may be due to the slight reduction in the rate of DNA replication previously noted for these mutants (5, 11). In contrast, levels of mRNA A were severely depressed in both E4 mutant infections, as they were in *dl338* infections.

E4 mRNA A dependence on E1b is at a postprocessing level. Since all E4 mRNAs derive from the same primary transcript, the selective depression of mRNA A levels seen in *dl338* infection must be due to a posttranscriptional effect. mRNA A results from nonusage of the E4 5' splice donor site D1 (Fig. 1). To determine whether the absence of E1b 55-kDa protein resulted in increased usage of this site, the relative abundance of RNA spliced and unspliced at this site in nuclear RNA was assessed for cells infected with either

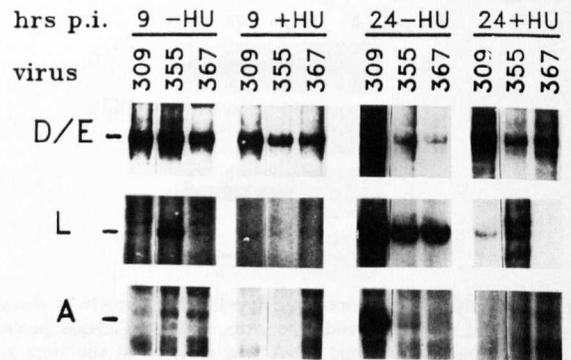


FIG. 4. Analysis of cytoplasmic levels of E4 mRNAs in HeLa cells infected by wild-type *dl309*, E4 ORF6 mutant *dl355*, or E1b 55-kDa/E4 ORF6 mutant *dl367* at the times p.i. indicated, in either the absence (-HU) or presence (+HU) of hydroxyurea, added at 10 mM to the culture medium. Virus stocks were CsCl gradient-purified particle preparations and used at 500 particles per cell. Five micrograms of unselected RNA was probed for specific mRNAs indicated at the left; other details are as for Fig. 2.

wild-type or *dl338* virus. The results of this analysis are shown in Fig. 5.

Levels of a control E4 RNA (mRNA L) in both nucleus and cytoplasm were unaffected by the absence of E1b 55-kDa protein. Levels of mRNA A in the cytoplasm were, as before, greatly reduced in the mutant infection. However, no similar reduction in nuclear levels of mRNA A was seen. This result shows that, as for other mRNAs which depend on E1b 55-kDa protein function for accumulation, E4 mRNA A requires this function for its efficient movement out of the nucleus or for stabilization rather than to modulate its splicing. No probe fragment protected by mRNAs spliced at site D1 could be detected in either wild-type- or *dl338*-infected cell nuclear RNA. Also, the cytoplasmic/nuclear ratio for mRNA L was very much greater than that for mRNA A. These findings suggest that, once spliced at the D1 site, E4 mRNA is rapidly exported to the cytoplasm. A similar conclusion was reached previously regarding E2 mRNA expression (18).

DISCUSSION

The various Ad5 E4 mRNAs have been divided into two temporal classes based on a detailed analysis of RNA levels

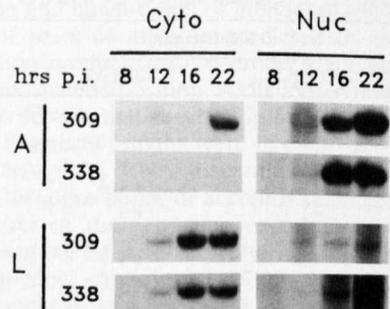


FIG. 5. Analysis of cytoplasmic and nuclear levels of E4 mRNAs during infection of HeLa cells by wild-type virus *dl309* or E1b 55-kDa mutant virus *dl338*. Total cytoplasmic (Cyto) and total nuclear (Nuc) RNAs were isolated at the times p.i. indicated, and 5- μ g samples were analyzed. The virus used and E4 RNA detected are indicated at the left. Other details are as for Fig. 2.

over the time course of wild-type infection. mRNAs C, D, G, and J form an early class, while mRNAs A, E, H, K, L, and M form a late class, dependent on viral DNA replication for expression. Previously described mRNAs B, F, I, and N could not be detected. Within the late class, mRNA A is unique, both in the late time p.i. at which its accumulation peaks and in its dependence on E1b 55-kDa protein and E4 ORF6 functions for proper cytoplasmic accumulation.

When members of these E4 mRNA classes are compared with one another, structural relationships can be identified. Members of the early class all retain the sequences between splice sites D3 and A3 in the 3' half of the E4 unit. This donor-acceptor combination is utilized in mRNAs of the late class to produce a family of shorter mRNAs. The usage of these sites is therefore temporally regulated during the course of Ad5 infection; recent studies of E4 expression in abortively infected monkey cells suggested that the E2a DNA-binding protein was important for this regulation (27). As with other examples of regulated adenovirus splicing, in expression of the E1a, E1b, and L1 genes (1, 8, 19, 22, 31), processing in E4 moves toward more heavily spliced, shorter mRNAs in the late phase of infection.

The undetected E4 mRNAs are also related, being formed by the use of donor site D2a or D2b. The lack of usage of these sites may be sequence related, as they are more divergent from the consensus sequence (23) than are sites D1 and D3, where splicing was observed. Given the detection of mRNAs spliced at these sites by other workers, their absence in our experiments may indicate that usage of D2a or D2b is regulated and apparent only under certain circumstances.

The observed temporal pattern of RNA expression from the E4 region suggests that ORF2, ORF3, ORF4, and ORF6 products are synthesized initially during the early phase of infection, while the products of ORF1 and ORF6/7 are synthesized later. The early expression of ORF3 and ORF6 is in agreement with the role of their products in facilitating viral gene expression from the onset of the late phase (3, 11, 15, 28, 34). ORF4-mediated down-regulation of E1a phosphorylation in HeLa cells (24) was observed from early in the infection, in agreement with our data. The delay in the expression of ORF6/7 is surprising given its role in the activation of the E2 early promoter (16, 20, 25). Formation of an infection-specific, E4-dependent complex on an E2 promoter fragment has been observed in extracts prepared from HeLa cells from 6 h p.i. (12). However, amounts of this complex increased dramatically between 6 and 12 h p.i. in these experiments. Given the differences in experimental protocol used (multiplicity of infection and cell type), this delayed-early increase in activity, now attributed to E4 ORF6/7 protein, is in reasonable agreement with our data on expression of ORF6/7 mRNA. No function has yet been assigned to the ORF1 or ORF2 product in infections of HeLa cells (11), although their conservation among human adenoviruses implies that they are important to the infectious process in some circumstances. Recently, Javier et al. (17) showed that the ORF1/2/3 segment of E4 from Ad9 was essential for mammary tumor induction by this virus in rats. Our data suggest that Ad5 ORF1 functions late in infection, while ORF2 is required earlier.

Among the E4 mRNAs detected in this study, mRNA A is the only one to retain splice donor site D1 unused. Furthermore, this mRNA also retains a number of acceptor sites (A1a to A1e) that are used, with D1, to form other detectable E4 mRNAs. Thus, mRNA A is an incompletely spliced mRNA which contains a readily removable intron. Although

other E4 mRNAs are similarly incompletely spliced, through retention of either the D2a/A1e, D2b/A1e or D3/A3 intron, these mRNAs either are synthesized only in the early phase of infection or else carry introns whose removal was not detected in our experiments. In a previous study, strong dependence on E1b function for accumulation of late viral mRNA was shown to correlate with the presence in the mRNA of unused splice acceptor sites (17a). It was argued that this result was due to viral mRNAs, in the absence of E1b 55-kDa protein, being held on the nuclear matrix by host cell mechanisms whose function was to prevent the efflux of immature RNA from the nucleus. The finding that the partially spliced E4 mRNA A is, among the many E4 mRNAs detected, uniquely dependent on E1b 55-kDa protein function is further evidence that this function is necessary for the efficient release from the nuclear matrix of mRNAs that are otherwise partially retained. The implication that residual splicing sites in an mRNA are a cause of E1b dependence for accumulation suggests that nuclear proteins which recognize intron sequences in heterogeneous nuclear RNA may be targets for the action of E1b 55-kDa protein.

ACKNOWLEDGMENTS

This work was supported in part by a grant from the Medical Research Council. Ian Dix is the recipient of an SERC studentship.

REFERENCES

1. Akusjarvi, G., and H. Persson. 1981. Controls of RNA splicing and termination in the major late adenovirus transcription unit. *Nature (London)* **292**:420-426.
2. Babiss, L. E., H. S. Ginsberg, and J. E. Darnell. 1985. Adenovirus E1B proteins are required for accumulation of late viral mRNA and for effects on cellular mRNA translation and transport. *Mol. Cell. Biol.* **5**:2552-2558.
3. Bridge, E., and G. Ketner. 1989. Redundant control of adenovirus late gene expression by early region 4. *J. Virol.* **63**:631-638.
4. Chroboczek, J., F. Bieber, and B. Jacrot. 1992. The sequence of the genome of adenovirus type 5 and its comparison with the genome of adenovirus type 2. *Virology* **186**:280-285.
5. Cutt, J. R., T. Shenk, and P. Hearing. 1987. Analysis of adenovirus early region 4-encoded polypeptides synthesized in productively infected cells. *J. Virol.* **61**:543-552.
6. Dix, I., and K. N. Leppard. 1992. Open reading frames 1 and 2 of adenovirus region E4 are conserved between human serotypes 2 and 5. *J. Gen. Virol.* **73**:2975-2976.
7. Downey, J. F., D. T. Rowe, S. Bacchetti, F. L. Graham, and S. T. Bayley. 1983. Mapping of a 14,000-dalton antigen to early region 4 of the human adenovirus 5 genome. *J. Virol.* **45**:514-523.
8. Esche, H., M. B. Mathews, and J. B. Lewis. 1980. Proteins and mRNAs of the transforming region of wild-type and mutant adenoviruses. *J. Mol. Biol.* **142**:399-417.
9. Flint, S. J. 1986. Regulation of adenovirus mRNA formation. *Adv. Virus Res.* **31**:169-228.
10. Freyer, G. A., Y. Katoh, and R. J. Roberts. 1984. Characterization of the major mRNAs from adenovirus 2 early region 4 by cDNA cloning and sequencing. *Nucleic Acids Res.* **12**:3503-3519.
11. Halbert, D. N., J. R. Cutt, and T. Shenk. 1985. Adenovirus early region 4 encodes functions required for efficient DNA replication, late gene expression, and host cell shutoff. *J. Virol.* **56**:250-257.
12. Hardy, S., D. A. Engel, and T. Shenk. 1989. An adenovirus early region 4 gene product is required for induction of the infection-specific form of cellular E2F activity. *Genes Dev.* **3**:1062-1074.
13. Herisse, J., M. Rigolet, S. Du Pont de Dinechin, and F. Galibert. 1981. Nucleotide sequence of adenovirus 2 DNA fragment encoding the carboxylic region of the fiber protein and the entire E4 region. *Nucleic Acids Res.* **9**:4023-4042.

14. Horwitz, M. S. 1990. Adenoviridae and their replication, p. 1679-1721. In B. N. Fields, D. M. Knipe, et al. (ed.), *Virology*, 2nd ed. Raven Press, New York.
15. Huang, M.-M., and P. Hearing. 1989. Adenovirus early region 4 encodes two gene products with redundant effects in lytic infection. *J. Virol.* **63**:2605-2615.
16. Huang, M.-M., and P. Hearing. 1989. The adenovirus early region 4 open reading frame 6/7 protein regulates the DNA binding activity of the cellular transcription factor, E2F, through a direct complex. *Genes Dev.* **3**:1699-1710.
17. Javier, R., K. Raska, Jr., and T. Shenk. 1992. Requirement for the adenovirus type 9 E4 region in production of mammary tumors. *Science* **257**:1267-1270.
- 17a. Leppard, K. N. 1993. Selective effects on adenovirus late gene expression of deleting the E1b 55kd protein. *J. Gen. Virol.* **74**:575-582.
18. Leppard, K. N., and T. Shenk. 1989. The adenovirus E1B 55 kd protein influences mRNA transport via an intranuclear effect on RNA metabolism. *EMBO J.* **8**:2329-2336.
19. Lewis, J. B., and M. B. Mathews. 1980. Control of adenovirus early gene expression: a class of immediate early gene products. *Cell* **21**:303-313.
20. Marton, M. J., S. B. Baim, D. A. Ornelles, and T. Shenk. 1990. The adenovirus E4 17-kilodalton protein complexes with the cellular transcription factor E2F, altering its DNA-binding properties and stimulating E1A-independent accumulation of E2 mRNA. *J. Virol.* **64**:2345-2359.
21. Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green. 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* **12**:7035-7056.
22. Montell, C., E. F. Fischer, M. H. Caruthers, and A. J. Berk. 1984. Control of adenovirus E1B mRNA synthesis by a shift in the activities of RNA splice sites. *Mol. Cell. Biol.* **4**:966-972.
23. Mount, S. M. 1982. A catalogue of splice junction sequences. *Nucleic Acids Res.* **10**:459-472.
24. Muller, U., T. Kleinberger, and T. Shenk. 1992. Adenovirus E4orf4 protein reduces phosphorylation of c-Fos and E1A proteins while simultaneously reducing the level of AP-1. *J. Virol.* **66**:5867-5878.
25. Neill, S. D., C. Hemstrom, A. Virtanen, and J. R. Nevins. 1990. An adenovirus E4 gene product trans-activates E2 transcription and stimulates stable E2F binding through a direct association with E2F. *Proc. Natl. Acad. Sci. USA* **87**:2008-2012.
26. Pilder, S., M. Moore, J. Logan, and T. Shenk. 1986. The adenovirus E1B-55K transforming polypeptide modulates transport or cytoplasmic stabilization of viral and host cell mRNAs. *Mol. Cell. Biol.* **6**:470-476.
27. Ross, D., and E. Ziff. 1992. Defective synthesis of early region 4 mRNAs during abortive adenovirus infections in monkey cells. *J. Virol.* **66**:3110-3117.
28. Sandler, A. B., and G. Ketner. 1989. Adenovirus early region 4 is essential for normal stability of late nuclear RNAs. *J. Virol.* **63**:624-630.
29. Sarnow, P., P. Hearing, C. W. Anderson, N. Reich, and A. J. Levine. 1982. Identification and characterization of an immunologically conserved adenovirus early region 11,000 M_r protein and its association with the nuclear matrix. *J. Mol. Biol.* **162**:565-583.
30. Sarnow, P., P. Hearing, C. W. Anderson, D. N. Halbert, T. Shenk, and A. J. Levine. 1984. Adenovirus early region 1B 58,000-dalton tumor antigen is physically associated with an early region 4 25,000-dalton protein in productively infected cells. *J. Virol.* **49**:692-700.
31. Spector, D. J., M. McGrogan, and H. J. Raskas. 1978. Regulation of appearance of cytoplasmic RNAs from region 1 of the adenovirus 2 genome. *J. Mol. Biol.* **126**:395-414.
32. Tigges, M., and H. Raskas. 1984. Splice junctions in adenovirus 2 early region 4 mRNAs: multiple splice sites produce 18 to 24 RNAs. *J. Virol.* **50**:106-117.
33. Virtanen, A., P. Gilardi, A. Naslund, J. M. LeMoulllec, U. Pettersson, and M. Perricaudet. 1984. mRNAs from human adenovirus 2 early region 4. *J. Virol.* **51**:822-831.
34. Weinberg, D. H., and G. Ketner. 1986. Adenoviral early region 4 is required for efficient viral DNA replication and for late gene expression. *J. Virol.* **57**:833-838.
35. Williams, J., B. D. Karger, Y. S. Ho, C. L. Castiglia, T. Mann, and S. J. Flint. 1986. The adenovirus E1B 495R protein plays a role in regulating the transport and stability of the viral late messages. *Cancer Cells* **4**:275-284.

Zhonghe, Z., Nickerson, J.A., Krochmalnic, G., & Penman, S. (1986). Alterations in nuclear matrix structure after adenovirus infection. *J. Virol.* **61**, 1007-1018.

Zhou, C., Yang, Y. & Jong, A.Y. (1990). Mini-prep in ten minutes. *Biotechniques* **8**, 172-173.

Ziff, E.B. (1980). Transcription and RNA processing by the DNA tumour viruses. *Nature* **287**, 491-499.