Stony Brook University

The Adenovirus L4-22K Protein Serves as the Master Regulator to Coordinate Late Gene Expression and Viral DNA Packaging

A Dissertation Presented

by

Diana Guimet

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Genetics

Stony Brook University

December 2014

Stony Brook University

The Graduate School

Diana Guimet

We, the dissertation committee for the above candidate for the Doctor of Philosophy degree, hereby recommend acceptance of this dissertation.

Patrick Hearing, Ph.D.
Dissertation Advisor
Professor, Department of Molecular Genetics and Microbiology, Stony Brook University

Laurie Krug, Ph.D.
Chairperson of Defense
Assistant Professor, Department of Molecular Genetics and Microbiology,
Stony Brook University

Eckard Wimmer, Ph.D.
Distinguished Professor, Department of Molecular Genetics and Microbiology,
Stony Brook University

Adrian Krainer, Ph.D.
Chair, Department of Cancer and Molecular Biology, Cold Spring Harbor Laboratory

Janet Hearing, Ph.D.
Associate Professor, Department of Molecular Genetics and Microbiology,
Stony Brook University

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

The Adenovirus L4-22K Protein Serves as the Master Regulator to Coordinate Late Gene Expression and Viral DNA Packaging

by

Diana Guimet

Doctor of Philosophy

in

Genetics

Stony Brook University

2014

The adenovirus L4-22K protein is multifunctional and critical for different aspects of viral infection. Following DNA replication, increasing amounts of the L4-22K protein are observed prior to the appearance of other late viral proteins. We found that L4-22K controls viral gene expression at the post-transcriptional level and regulates the accumulation of the L4-33K protein, another critical viral regulator, at the level of alternative pre-mRNA splicing. This initiates an activation cascade of late viral proteins, while early genes are being turned off. To gain a better understanding of the mechanism of action of L4-22K, we sought to map the critical elements responsible for L4-22K-mediated splicing of the L4-33K mRNA. We show that activation of L4-33K splicing by L4-22K requires the enhancer element, a *cis*-acting sequence located at the 3' splice site of the L4-33K pre-mRNA. This element is necessary for L4-22K-mediated activation but it is not sufficient. Furthermore, we found that

regulatory sequences located in the second exon of the L4-33K pre-mRNA prevent constitutive splicing of the transcript. We propose that the repressive effect of these putative elements on L4-33K splicing is relieved by the L4-22K protein acting in cooperation with a yet unknown cellular factor. The action of L4-22K protein results in the strict temporal regulation of gene expression during the virus replication cycle. The late phase of the adenovirus life cycle is focused on producing sufficient quantities of structural proteins to assemble the capsid and package the viral genome. Packaging of the viral genome into an empty capsid absolutely requires the binding of L4-22K protein to packaging sequences in a complex with other viral proteins. We focused on the role of conserved pairs of cysteine and histidine residues in the C-terminal region of L4-22K protein. We reveal that the conserved cysteine residues are required for optimum levels of viral genome packaging, but are dispensable for the ability of the L4-22K protein to regulate viral gene expression. These results demonstrate that these two functions of L4-22K may be uncoupled. We propose that the L4-22K protein may serve as a useful target for antiviral therapy since it provides critical functions at different steps in the viral life cycle.

Table of Contents

List of Figures vii

List of Tables ix

List of Abbreviations x

Acknowledgements xiv

Chapter 1: Introduction 1

General Introduction 2

Classification 3

Ad Genome Organization 4

Attachment and Penetration 5

Early Gene Expression 6

Early Region 1A (E1A) 7

Early Region 1B (E1B) 11

Early Region 2 (E2) 12

Early Region 3 (E3) 14

Early Region 4 (E4) 25

Viral DNA Replication 31

VA RNA Genes 33

Late Gene Expression 34

Viral DNA Packaging 40

Ad Alternative pre-mRNA splicing 45

Chapter 2: Materials and Methods 49

Cells, Viruses, and Infections 50

Plasmids, Mutagenesis, and Cloning 51

Replication Assay and Quantitative PCR 55

Plaque Assays and Immunofluorescence Analysis 55

Virus Growth and Cesium Chloride Equilibrium Gradients 56

Silver Staining of Proteins in Virus Particles 58
Gel Mobility Shift Assays 58
Western Blot Analyses 59
Northern Blots and Reverse Transcriptase PCR (RT-PCR) 60
In Vivo Analysis of Alternative Splicing 62
Protein Affinity Chromatography and Analysis 63

Chapter 3: The Adenovirus L4-22K Protein Has Distinct Functions in the Posttranscriptional Regulation of Viral Gene Expression and Encapsidation of the Viral Genome 65
Abstract 66
Results 67
Discussion 99

Chapter 4: Characterization of the Critical Elements Responsible for L4-22K-mediated Splicing of the L4-33K pre-mRNA 110
Abstract 111
Results 112
Discussion 129

Chapter 5: Conclusions and Future Directions 143

References 148

Appendix: The Cre/loxP System for the Production of Helper-Dependent Adenovirus Vectors is Inefficient Due a Defect in Vector DNA Replication 160
Abstract 161
Introduction 162
Materials and Methods 166
Results and Discussion 170
References 190

List of Figures

1-1 Functional map of E1A proteins Page 8

1-2 Schematic of the E3 region Page 16

1-3 Schematic of the E4 region Page 27

1-4 Schematic of the major late transcription unit Page 36

1-5 Model of adenovirus genome packaging Page 41

3-1 The Ad5 L4-22K protein Page 68

3-2 Properties of L4-22K mutant viruses Page 71

3-3 Viral growth kinetics of L4-22K mutant viruses Page 74

3-4 The L4-22K H166Q/H170Q mutant virus does not inhibit Ad5-WT growth Page 78

3-5 Analysis of proteins in mature particles of Ad5- WT, C137S/C141S, and H166Q/H170Q mutant viruses Page 80

3-6 The L4-22K-C137S/C141S and L4-22K-H166Q/H170Q mutant proteins bind to Ad5 packaging sequences *in vitro* Page 82

3-7 Western blot analyses of viral gene expression Page 85

3-8 Analysis of Ad late mRNA levels Page 87

3-9 Analysis of E1A protein levels and mRNA levels Page 90

3-10 *In vivo* splicing assay for the L4-33K pre-mRNA Page 93

3-11 *In vivo* splicing assay for the L4-33K pre-mRNA using mutant L4-22K proteins Page 96

3-12 The L4-22K-H166Q/H170Q mutations disrupt splicing of the L4-33K pre-mRNA Page 98

4-1 Proposed model for the role of L4-22K as the master regulator of Ad gene expression Page 113

4-2 The *cis*-acting elements within the L4-33K pre-mRNA necessary for regulated splicing are within the 3' proximal intron nucleotides Page 116

4-3 The polypyrimidine tract of the L4-33K pre-mRNA is highly conserved Page 119

4-4 The 3' exon of L4-33K contains elements required for regulated splicing by the L4-22K protein Page121

4-5 Epitope-tagged L4-22K proteins are functional for splicing activity Page 125

4-6 Proteomic analysis of L4-22K and interacting proteins Page 127

A-1. The Cre-*loxP* system for the production of helper-dependent Ad vectors Page 165

A-2. Analysis of HV floxing efficiency in N52E6-Cre cells Page 173

A-3. Analysis of HDAd amplification Page 175

A-4. Analysis of replication efficiency of HDAd Page 179

A-5. A replication-defective infectious clone mimics the low efficiency of HDAd amplification Page 182

A-6. Production of HDAd FK7 Page 187

List of Tables

Table 4-1 Predicted exonic splicing silencer sequences_____ Page 139

List of Abbreviations

3RE	IIIa splicing repressor element
3VDE	IIIa virus infection-dependent splicing enhancer
aa	amino acid
AA	arachidonic acid
Ad	adenovirus
ADP	adenovirus death protein
Ad-Pol	ad DNA polymerase
BCA	bicinchoninic acid
BSA	bovine serum albumin
BPS	branch point sequence
CAR	coxsackie B virus-adenovirus receptor
CDKs	cyclin-dependent kinases
cPLA2	cytosolic phospholipase a2
CR	conserved region
CsCl	cesium chloride
CTL	cytotoxic T cell
DE	downstream elements
DBP	DNA binding protein
DD	death domain
DED	death effector domain
DISC	death inducing signaling complex

DNA-PK	DNA-dependent protein kinase
E1A	early Region 1A
E1B	early Region 1B
E2	early region 2
E3	early region 3
E4	early region 4
ER	endoplasmic reticulum
ESE	exonic splicing enhancer
ESS	exonic splicing silencer
FADD	Fas-associated death domain
FasL	Fas ligand
FFU	fluorescence focus unit
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
h	hours
hpi	hours post-infection
HATs	histone acetyl transferases
HDACs	histone deacetylases
HDAd	helper-dependent Ad vector
hnRNP	heterogeneous nuclear ribonucleoprotein
HV	helper virus
IFN	interferon
IL-1β	Interleukin-1 beta
ISE	intronic splicing enhancer

ISS	intronic splicing silencer
ITR	inverted terminal repeats
kbp	kilobase pair
kDa	kilodalton
L4P	L4 promoter
MAP	mitogen-activated protein
MIC (A & B)	MHC class I chain related proteins A and B
min	minutes
MLP	major late promoter
MLTU	major late transcription unit
MRN	Mre11-Rad50-Nbs1
NK	natural killer cell
NKG2DL	NKG2D ligand
nt	nucleotide(s)
OPTN	optineurin
P	passage
PCR	polymerase chain reaction
PBS	phosphate-buffered saline
PFU	plaque forming unit
PML	progressive multifocal leukoencephalopathy
PML-NB	PML nuclear bodies
PODs	PML- oncogenic domains
PTB	polypyrimidine tract binding protein

pTP	precursor terminal protein
PVDF	polyvinylidene difluoride
Py	polypyrimidine tract
RCA	replication competent adenovirus
RID	receptor internalization and degradation
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
snRNP	small nuclear ribonucleoprotein
ss	splice site
SR	serine/arginine-rich
sT	small T antigen
SUMO	small ubiquitin-like modifier
TBP	TATA-binding protein
TBS	Tris-buffered saline
TD	Tagmentation DNA buffer
TLR	toll-like receptor
TMD	transmembrane domain
TNFR	tumor necrosis factor receptor
TP	terminal protein
TPL	tripartite leader
TRADD	TNFa receptor associated death domain
U2AF	U2 auxiliary factor
WT	wild-type

Acknowledgements

First and foremost, I would like to thank my advisor, Dr. Patrick Hearing. His mentorship and support have allowed me to be more successful in my graduate career than I could have hoped for. During the first few months in the lab he taught me first-hand many of the basic laboratory and technical skills that helped set the stage for a multitude of successful experiments in the future. I feared that I was not as experienced as other entering graduate students, but Dr. Hearing taught me that what was important was the effort that you put into your work and the passion to thrive and succeed as a Scientist. He always acknowledged my hard work and dedication by encouraging me to write first-author manuscripts, collaborate with other lab members in second-author papers, and to attend the top conferences in the field. Being the excellent mentor that he is, he was always available to discuss my data no matter how busy he was. He magically had time to dedicate to all of his students and perform his own experiments, in addition to all of his other duties as a Professor. His support and encouragement for a healthy work-life balance significantly impacted my successful graduate career, and for that I am very grateful.

I would also like to thank my committee members, Dr. Laurie Krug, Dr. Eckard Wimmer, Dr. Janet Hearing, and Dr. Adrian Krainer for all of their support and guidance on all of my projects. I would especially like to thank Dr. Janet Hearing for all of the wonderful

letters of recommendation and for allowing us to have social gatherings at her home to celebrate all of the lab member's accomplishments, both personal and professional, always making us feel appreciated. I would also like to thank Dr. Adrian Krainer for all of his guidance through the difficult journey in the field of alternative RNA splicing, and all of the reagents that his lab so generously provided for my experiments.

I would like to thank my colleagues in the Hearing lab, for countless hours of insightful discussion, not all of which was science related. I would especially like to thank Dr. Philomena Ostapchuk for excellent scientific guidance. She has always been my second mentor in the lab, someone I could always count on, no matter what my question was. I would also like to thank Dr. Kai Wu for collaborations in manuscript writing.

Without the administrative support provided by Kate Bell, graduate school would have been a very difficult journey, and for that I am thankful. I would also like to thank Dr. Turhan Canli and Dr. Martha Furie for all of their hard work and dedication to providing us with a Genetics Graduate program that we can thrive in. I am especially grateful to Dr. Jim Bliska for allowing me the opportunity to be on the Molecular and Cell Biology Training Grant, and for all of the career-building seminars, projects, and programs that he provided us with. I will always be grateful to Nina Maung-Gaona, Toni Sperzel, and all other members of the Center for Inclusive Education for all of their financial help and emotional support and for the career building workshops and assistance in surviving graduate school.

I would like to thank my family for their moral and emotional support and for their motivation and encouragement through all these years of graduate school. Without their motivating speeches, their long nights of homework help, and their celebration of all of my accomplishments, both big and small, I would not be here today. I would like to thank my

husband for always believing in me and encouraging me to be ambitious and driven, even when it meant less quality time with him. Without the sacrifice that he made moving across the country with me, I would not have survived graduate school. Finally, I would like to thank my beautiful daughter and my loving dog. They gave me yet another reason to love life and a reason to smile every single day, even on the toughest, most stressful days of my graduate career.

Vita, Publications and/or Fields of Study

1. Guimet, D. and Hearing, P. (2015). Adenovirus Replication. Adenoviral Vectors for Gene Therapy. This is chapter 1 of this thesis. Sections on Early Region 1A (E1A), Early Region 1B (E1B), and Early Region 4 (E4) written by Patrick Hearing. All other sections written by Diana Guimet

2. Guimet, D. and Hearing, P. (2013). The Adenovirus L4-22k protein has distinct functions in post-transcriptional regulation of gene expression and encapsidation of the viral genome. Journal of Virology. 87(13): 7688-7699

3. Wu, K., **Guimet, D**., and Hearing, P. (2013). The Adenovirus L4-33K protein regulates both late gene expression patterns and viral DNA packaging. Journal of Virology. 87(12): 6739-6747

4. Wu, K., **Orozco, D**., and Hearing, P. (2012). The Adenovirus L4-22K protein is multifunctional and an integral component of crucial aspects of infection. Journal of Virology. 86: 10474-10483

Chapter 1: Introduction

General Introduction

The basic biology of adenovirus (Ad) has been studied for more than half a century, unraveling the profound range of virus-host interactions including the complexity of viral regulation of gene expression and Ad inhibition of host anti-viral activities. Human Ad was first isolated from adenoid tissue in the 1950s as viral agents associated with respiratory infections (1). Ads can establish acute and persistent infections and are generally associated with mild or subclinical diseases in immunocompetent patients. However, disease can be severe or even fatal in immunocompromised patients (2). Each Ad serotype can infect a great variety of tissues and cells, however a distinct disease pattern is observed for Ads that belong to different subgroups. For example, subgroup A and F Ads (see below) cause gastrointestinal infections, while subgroup B and C Ads cause upper respiratory tract infections, which may be accompanied by acute respiratory disease. Interestingly, some Ads of subgroup D cause a distinct disease, epidemic keratoconjunctivitis (EKC). Over 100 Ads have been identified and characterized in a wide range of vertebrate species. The virion is a non-enveloped, icosahedral capsid with a diameter of ~80-90nm, containing a linear double-stranded DNA genome of ~36 kbp for human Ads. The size of Ad genomes vary from ~20-45 kbp (3).

In the early 1960s researchers demonstrated that some human Ads cause tumors in rodents, which led to the surge in studies of the molecular biology, genetics, and physiology of Ads that continues to this day. Most Ad infections are associated with mild disease

occurring in the general population, and more severe forms observed mainly in children and U.S. military recruits. Ad is increasingly being recognized as a significant viral pathogen in immunocompromised individuals (2). Patients undergoing immunosuppressive therapy and AIDS patients, are included in this category. Additionally, recombinant Ad vectors are being utilized in ~25% of current human gene therapy trials. Of particular excitement is the use of oncolytic Ads as a new class of anti-cancer agent with great therapeutic potential (4). Extensive insight into the biology of Ads has opened the door for engineered Ads that target cancer cells. For example, Ad is extremely successful in combating innate immune responses, but Ad vectors can be made pathogenic by targeting specific viral proteins encoded by the E1 and E3 regions to fight cancerous tissue (5).

Classification

Ads belong to *Adenoviridiae* family. Currently there are five genera: *Mastadenovirus* and *Aviadenovirus* originate from mammals or birds, respectively; *Atadenovirus* and *Siadenovirus* have a broader range of hosts including birds and frog; and *Ichtadenovirus*, identified in fish. Human Ads comprise more than 60 different serotypes classified into seven subgroups, designated A–G, that are organized based on their capacity to agglutinate red blood cells of human, rat and monkey as well as on their oncogenicity in rodents. Recently, the field of DNA sequencing has taken the classification to a new level, and new sequence availability has allowed for more detailed phylogenetic analyses (3).

Ad Genome Organization

The Ad genome is flanked by inverted terminal repeats (ITR) of ~100bp which contain the origins of viral DNA replication. This sequence is followed by the viral packaging sequence at the left-end of the genome, which directs viral DNA encapsidation. Ad terminal protein (TP) is covalently linked to the 5' ends of the genome and plays an important role in the initiation of viral DNA replication. The capsid is composed of three major proteins (II, III, and IV) and five minor proteins (IIIa, IVa2, VI, VIII, and IX). Proteins V, VII, and mu are associated with the DNA and form the core within the virion. These proteins are believed to condense the Ad DNA and mediate interactions between the core and the capsid. Ad protease is required for the maturation of the assembled particle to form fully infectious virus (6).

The human Ad genome encodes ~40 proteins, which are classified as either early or late, based on their expression before or after DNA replication, respectively. The genome contains the immediate-early region E1A, four early transcription units (E1B, E2, E3, and E4), and a set of "delayed early" units encoding proteins IX and IVa2. The Ad major late promoter (MLP) directs the synthesis of a single late pre-mRNA that is alternatively polyadenylated and alternatively spliced to generate five families of late mRNAs, regions L1–L5. Two additional small late transcripts are produced, VA RNAI and VA RNAII. The transcription units of the Ad genome are transcribed from both strands of the chromosome.

Ads are an excellent example of viruses that efficiently use limited genetic space and information to maximize their protein production for optimal virus propagation. Expression

of the viral genes is temporarily regulated at many different levels to produce a stepwise, logical progression of gene expression in order to take full advantage of the cellular machinery to direct virus production. As an example, the late viral genes are not expressed to a full extent until after viral replication takes place, and, even then, late gene expression is tightly controlled at both the transcriptional and post-transcriptional level. This tight regulation of viral gene expression provides for maximal production of virus from infected cells.

Attachment and Penetration

The Ad fiber protein binds to the coxsackie-B virus-adenovirus receptor (CAR), which is the primary receptor for both Ad5 and coxsackie B virus (7, 8). Attachment to the receptor occurs in concert with the binding of the RGD peptide on the penton base to cellular integrins ($\alpha_v\beta_3$ and $\alpha_v\beta_5$). Ad5 enters cells using heparin sulfate proteoglycans as an alternative receptor, likely through interaction of Ad with blood factors such as factor IX, factor X or complement component C4-binding protein. CD46 also is used as a receptor for group B human adenoviruses (7, 8). Ad evades degradation by escaping from the early endosome and is internalized by receptor-mediated endocytosis via clathrin-coated pits (9). The endosome acidifies, which alters virus topology and the capsid components disassemble. The virion is transported through the cytoplasm to the nucleus along the microtubule network and the capsid is slowly disassembled en route (reviewed in (9)). Upon reaching the nuclear pore complex, the protein VII-wrapped Ad DNA enters the nucleus. Current published

findings suggest that it is only protein VII-wrapped DNA that enters the nucleus, escorted by histone H1 (6). Protein VII protects the viral DNA from activating the DNA damage response (10). At this point, the DNA is still highly condensed and must undergo extensive remodeling to decondense before transcription of early genes can begin (11). At the beginning of viral gene expression, histones are found bound to the viral DNA, particularly histone H3.3, along with protein VII. Remodeling continues throughout the life cycle. Viral DNA replication and assembly of progeny virions occur entirely within the nucleus of infected cells. Ad DNA associates with newly synthesized pVII and the DNA-protein complex subsequently is packaged into the Ad capsid (6). The Ad life cycle takes 24-36 h, with a single virus-infected cell producing ~10^4 daughter virions.

Early Gene Expression

When the viral genome enters the nucleus, early gene expression is directed toward achieving three main objectives. First, the host cell must be stimulated to enter S phase of the cell cycle and provide the virus with the necessary intracellular niche for virus replication (11). Second, Ads must devote a considerable part of their coding capacity to immune evasion functions that facilitate infection and maintain a state of balance for lasting infections (12). Third, viral gene products must be produced and used in concert with cellular proteins to carry out viral DNA replication. Ad encodes ~25 early gene products. The early genes are expressed in a temporal and coordinated manner.

Early Region 1A (E1A)

The first early region expressed after Ad infection is the immediate early transcription unit E1A since it requires only cellular proteins for its expression. The E1A gene products in turn activate transcription from the other early promoters. The E1A gene is comprised of two exons, and several E1A polypeptides are produced following alternative splicing of a primary RNA transcript (Fig. 1-1). The most abundant of the E1A proteins are referred to as the E1A 243 amino acid (243aa) and 289 amino acid (289aa) gene products. The E1A 243aa and 289aa proteins act as major regulators of early viral transcription as well as important modulators of host cell gene expression and proliferation (13). The E1A 243aa and 289aa proteins share two conserved regions within exon 1 referred to as CR1 and CR2 as well as another conserved region (CR4) at the C-terminus in exon 2 (Fig. 1-1). The two proteins differ only in a 46-residue internal exon segment present in the 289aa protein, referred to conserved region 3 (CR3). This region is important for the transcriptional transactivation properties of the E1A 289aa protein (13).

The E1A proteins exert their effects by interactions with numerous cellular

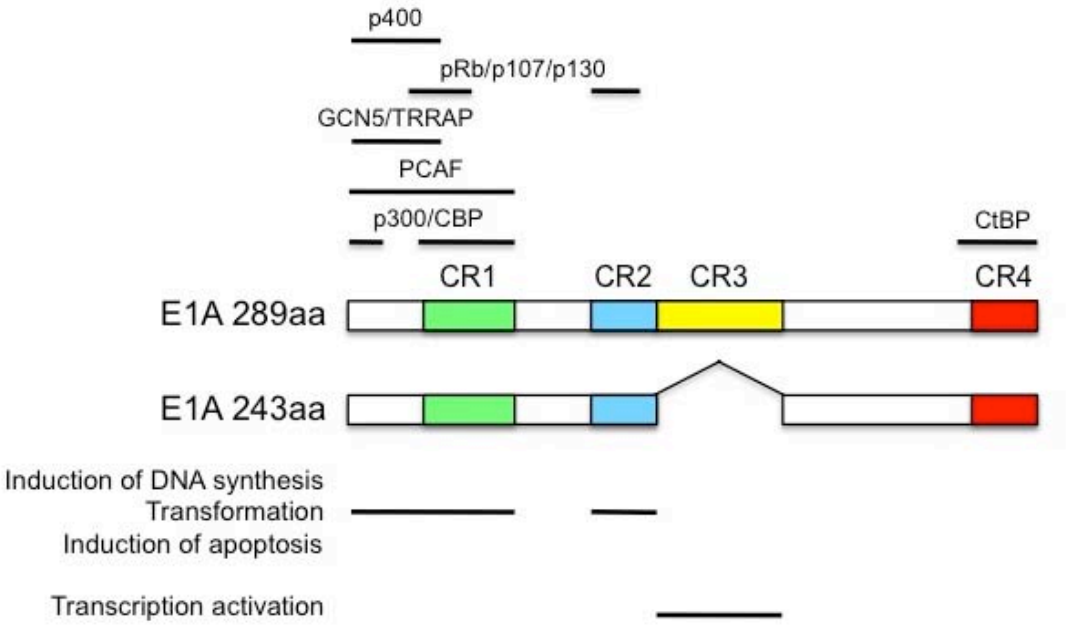


Figure 1-1. Functional map of E1A proteins. The coding sequences of the E1A 289aa and 243aa proteins are shown with conserved regions depicted (CR1, CR2, CR3, and CR4). Binding sites for cellular proteins are indicated by bars above. E1A functional activities for each domain are listed below

proteins, many of which are involved in transcriptional regulation (Fig. 1-1). The E1A proteins interact with a number of important cellular proteins including: (1) the retinoblastoma tumor suppressor family members pRb, p107, and p130 via CR1 and CR2; (2) transcriptional coactivators p300/CBP, PCAF, GCN5, TRRAP, and p400 via amino terminal sequences and CR1; (3) additional transcription factors such as TATA-binding protein (TBP), members of the ATF family, and the RNA polymerase II mediator complex via CR3; and (4) the transcriptional repressor CtBP via CR4 (13). The expression of E1A alone is sufficient to induce immortalization of primary rodent cells. E1A fully transforms such cells in conjunction with other oncogenes such as the Ad E1B proteins or activated Ras (14). The expression of E1A also is sufficient to induce S phase progression in quiescent cells (15). E1A activates gene expression via the E2F family of transcription factors. E2F transcription factors play a major role in the expression of cellular genes important for cell cycle progression. E2Fs both positively and negatively regulate gene expression (16). In general, E2F-1, -2, and -3a/b activate gene expression, while E2F-4 and -5 repress gene expression. E2F-6, -7 and -8 may function as dominant-negative effectors. Activating E2Fs recruit histone acetyl transferases (HATs) and other transcriptional activators to activate transcription. In contrast, repressing E2Fs recruit histone deacetylases (HDACs) and other transcriptional repressors to repress transcription. These latter complexes are formed via the interaction of repressing E2Fs with members of the retinoblastoma gene family (16). Rb family binding to E2Fs is controlled through phosphorylation by cyclin-dependent kinases (CDKs). The phosphorylation of Rb family proteins by CDKs in the G1 phase of the cell cycle results in their dissociation from E2Fs and derepression of E2F-responsive genes (16). The activation of E2F complexes results in the promotion of S phase of the cell cycle via the

activation of cellular genes that promote cell cycle progression. E1A acts to subvert the tight control of E2F by binding directly to Rb proteins via an LXCXE motif in E1A common to other DNA tumor virus transforming proteins (16). E1A sequesters Rb family members and frees E2F heterodimers to activate viral and cellular gene expression. Both the E1A 243aa and 289aa products direct the release of Rbs from E2Fs, and both E1A proteins promote cellular transformation, in part, via this mechanism. p300/CBP, PCAF, and GCN5 are all HATs, while TRRAP and p400 serve as scaffolding proteins to bridge interactions of HATs with other transcriptional regulators (17). The binding of E1A to these proteins promotes cellular DNA synthesis, and E1A mutants that cannot interact with these effectors are defective for transformation (17). E1A binds to p400 through the N-terminal domain. p400 is related to the yeast chromatin-modifying proteins SWI2/SNF2 and forms a complex with TRRAP. An E1A mutant that is defective for p400 binding also is defective for transformation. p400 is part of a larger HAT complex, termed TIP60, that contains TRRAP, GCN5, PCAF, TIP48, and TIP49 (17).

All the aforementioned results demonstrate the complex regulatory circuit that governs the regulation of cellular proliferation and how disruption of this carefully coordinated system by the N-terminal domain of E1A leads to profound effects on cellular life and death decisions. The C-terminal exon of the E1A proteins contains a nuclear localization signal and a binding site (CR4) for the transcriptional corepressor CtBP (18, 19). E1A exon 2 exhibits transcriptional regulatory activities that antagonize exon 1 functions. E1A exon 2 negatively regulates E1A functions in transformation and tumorigenesis (18, 19). CtBP functions as a transcriptional corepressor when tethered to a promoter region and binds to a number of HDACs including HDAC-1, -4, and -5. The C-terminal region of E1A also

binds cellular proteins Dyrk and Foxk, which may contribute to, or regulate, E1A activities (18, 19). E1A also plays a role in the induction of apoptosis in infected cells. Sustained, unregulated E2F activity triggers a cellular checkpoint and causes an increase in the level of the tumor suppressor p53 (20). Activated p53 induces gene expression by binding specific promoter sequences, which activates genes that are involved in a number of cellular processes. p53 can induce cell cycle arrest, thus inhibiting progression of cell division; p53 also can induce cell death by the induction of apoptosis (20). The activation of p53 and induction of cellular apoptosis would be deleterious to Ad replication. Therefore, Ad has evolved several proteins encoded by the E1B and E4 transcription units that repress p53 activity and inhibit apoptosis.

Early Region 1B (E1B)

E1B encodes the E1B-19K and E1B-55K proteins. The major roles of these proteins in Ad infection are to inhibit apoptosis and further modify the intracellular environment in order to make the cell more hospitable to viral protein production and viral DNA replication. Viruses with mutations in either E1B protein are significantly reduced in virus yield due to cell death by apoptosis prior to the completion of the replication cycle. The E1B-55K protein is essential for a variety of important functions in the viral life cycle including the inhibition of the induction of p53-dependent apoptosis (21). The E1B-55K protein binds to the N-terminal transactivation domain of p53 and inhibits p53-induced transcription. E1B-55K also disrupts the interaction of p53 with the HAT PCAF and interferes with p53 acetylation (21).

Interestingly, the E1B-55K protein promotes cell transformation independently of repression of p53 transcriptional activity. This may relate to the ability of E1B-55K to inhibit other apoptotic activities in the cell including the pro-apoptotic cellular activity of Daxx (21). The E1B-55K protein is modified by the small, ubiquitin-like modifier protein SUMO-1 and sumoylation of E1B-55K is required for its role in transformation (22). Finally, E1B-55K acts in a complex with another Ad early protein, E4-ORF6, to promote the proteasome-dependent degradation of p53 among other cellular protein targets (23).

The E1B-19K protein also is involved in the inhibition of apoptosis. E1B-19K acts to block apoptotic pathways that do not rely on p53, such as the TNFα and Fas ligand cell death pathways (24). E1B-19K is a functional homologue of a cellular suppressor of apoptosis, Bcl-2. E1B-19K acts in the same manner as Bcl-2 and predominantly inhibits apoptosis by binding pro-apoptotic activists Bax and Bak (24). E1B-19K also plays a role in the inhibition of TNFα-induced apoptosis by blocking the oligomerization of death-inducing complexes involving FADD (24). FADD is a protein that is activated by binding Fas via death domains, thus its name (Fas-associated death domain). The exact function of E1B 19K in FADD regulation is not well understood.

Early Region 2 (E2)

The E2 transcription unit encodes proteins that are vital to viral DNA replication: DNA binding protein (DBP), precursor terminal protein (pTP), and Ad DNA polymerase (Ad-Pol). DBP is a 529aa protein with a molecular mass of 59kDa and binds cooperatively to

single-stranded DNA to stimulate both the initiation and elongation of DNA replication (25). DBP contains two major domains: the N-terminal globular core domain consists of amino acids 1-173 and the C-terminal part which consists of aa 174-529 and harbors most of the biological functions given to DBP including nucleic acid binding and replication (25). The outermost C-terminal part of DBP, aa 510-529, is sufficient for all DNA replication functions and forms a protruding C-terminal arm and contains a hook. This hook is involved in the formation of a DBP chain where a C-terminal arm hooks into another DBP to form a multiprotein complex. This multimerization is the driving force for ATP-independent DNA unwinding by DBP during elongation (25). During the initiation of viral DNA replication, DBP also stimulates the formation of a pTP-CAT nucleotides intermediate by lowering the K_m value of the reaction, possibly via a direct interaction with the pTP-Pol complex (25). Indirectly, DBP stimulates replication initiation by increasing the binding of NFI (see below) to the origin. Other roles for DBP during DNA replication include enhancing the processivity of Ad-Pol, which is achieved by cooperative binding to the displaced strand during replication, thereby protecting it from nuclease digestion and facilitating strand displacement (25).

pTP is a 80kDa protein that binds both single-stranded and double-stranded DNA and has several functions in DNA replication (26). pTP forms a heterodimer with Ad-Pol which constitutes the pre-initiation complex for Ad DNA replication (26). The C-terminal 60kDa of pTP contains the Ad-Pol binding region, while the N-terminal portion is involved in DNA binding. pTP binds to the entrance of the primer binding groove of Ad-Pol, with its priming part located at the polymerase active site and allows for protein-primed DNA synthesis to begin (26). Besides helping to stabilize the pTP-Pol complex, other functions of the parental

pTP include attachment of the DNA to the nuclear matrix which is important for efficient transcription of Ad DNA and possibly for DNA replication (27). The presence of pTP also renders viral DNA inaccessible to 5' exonucleases and prevents binding of end-recognizing proteins, which could inhibit DNA replication (26).

Ad-Pol is a 140kDa phosphoprotein that belongs to a group of Pol-α DNA polymerases that employs a protein primer for DNA replication (26). Ad-Pol has DNA binding activity, DNA polymerase activity, and participates in the initiation of Ad DNA replication. The polymerase and exonuclease active sites of Ad-Pol are spatially distinct: the exonuclease activity of Ad-Pol is located at the N-terminal part of the protein and the polymerase activity is located at the C-terminus. The molecular architecture of Ad-Pol is similar to that of RB69 DNA polymerases, a model enzyme for the family B polymerases (26). When bound to DNA, Ad-Pol covers a region of 14 to 15 nucleotides. Compared to free Ad-Pol, the pTP-Pol complex has reduced polymerase and exonuclease activity. This is most likely due to the fact that pTP binds at the entrance of the primer binding groove of Ad-Pol, which results in a competition between pTP and the DNA located at the primer binding groove (26). After dissociation, Ad-Pol becomes a more active and processive enzyme.

Early Region 3 (E3)

The E3 region of the Ad2 and Ad5 genome contains seven expressed open reading frames, most of which encode proteins with immunomodulatory functions (Fig. 1-2). The E3

region is dispensable for viral replication in culture since E3 genes are involved in the evasion of host immune defenses (28). E3 encodes integral membrane proteins that subvert host defense mechanisms by modulating immune response mechanisms including inhibition of antigen presentation, suppression of natural killer (NK) cell activation, down-regulation of apoptosis receptors, and interference with TNFα receptor induced activities (28, 29). It is important that the infected cell remain viable during the extended period of infection, so anti-apoptotic mechanisms are critical for survival. TNFα and Fas are members of the TNFR superfamily that contain death domains, which participate in protein-protein interactions that lead to the activation of pro-apoptosis cascades. At high concentrations, TNFα can inhibit the replication of certain viruses, including Ads, by inducing lysis and/or apoptosis of infected cells (28, 29). Binding of a death ligand in the TNF family, such as TNFα, Fas ligand (FasL), and TRAIL to its cognate death receptor TNFR1, FAS, and TRAIL receptors 1 and 2 (TNFR1, TNFR2), respectively, leads to complex protein-protein interactions that result in a cascade of caspase-mediated proteloytic cleavages and the activation of transcription factors such as NF-κB and AP1. This triggers events that ultimately lead to the destruction of the cell via apoptosis. The receptors for TRAIL, Fas ligand, and TNFα play a role in killing infected cells, but Ad E3 proteins have the ability to inhibit such killing to prolong acute and persistent infections (28-30). Another viral strategy aims to inhibit antigen presentation. Cytotoxic T cells (CTL) recognize antigenic peptides presented by MHC Class I antigens on the surface of infected cells. Assembly of MHC class I antigens occurs in the ER and this process is assisted by chaperones. MHC class I molecules carry peptides that are targeted by the proteasome in the cytosol and translocate across the ER membrane by the transporter

E3 protein	Function(s)
12.5K	unknown
6.7K	inhibits TRAIL apoptosis
gp19K	blocks MHC class I presentation, inhibits CTL killing
ADP	cell killing, promotes virus release
RIDα/RIDβ	Receptor Internalization and Degradation complex, blocks TNFα, FasL, and TRAIL apoptosis
14.7K	Inhibits TNFα, FasL, and TRAIL apoptosis, blocks arachadonic acid release

Figure 1-2. Schematic of the E3 Region. The different proteins encoded by the E3 region are indicated by bars. The functions ascribed to different E3 proteins are listed below the diagram.

associated with antigen presentation (TAP) (28, 29). After recognition, CTLs release perforins and granzymes, which promote killing of the infected cell. Alternatively,CTL can induce apoptosis by the Fas pathway. By avoiding antigen presentation, the continued replication of the virus in infected host cells is possible, contributing to the ability of Ads to establish and maintain persistent infections.

E3-19K is a membrane glycoprotein localized in the ER and contains three segments: a luminal portion, a long cytoplasmic tail, and a transmembrane segment. E3-19K functions to counter the recognition of infected cells by both innate and adaptive cellular immune responses. E3-19K forms a complex with MHC class I molecules and inhibits transport of newly synthesized MHC molecules to the cell surface, preventing peptide presentation and suppressing recognition by T cells (28, 29). The function of E3-19K is based on two activities: MHC-I binding activity by use of the luminal portion of the protein, and the ability to localize to the ER where it retains the MHC-I molecules. The latter activity is mediated by two structural elements, the ER retention signal contained in the transmembrane segment and the ER retrieval signal in the cytoplasmic tail. The retrieval signal mediates retrograde transport of E3-19K and associated MHC-I from the *cis*-golgi to the ER, where it is retained. Recently an additional functional element has been characterized in the transmembrane domain (TMD) of E3-19K (31). The TMD, together with the ER retrieval function, play a role in efficient ER localization and transport inhibition of MHC-I. A potential caveat to HLA down-regulation by E3-19K is the potential to render Ad-infected cells vulnerable to NK cell recognition. NK cells are a heterogeneous population of cells expressing a wide range of activating and inhibitory receptors, including NKG2. MHC class I chain related proteins A and B (MICA and MICB) are two of seven human cellular NKG2D ligands

(NKG2DLs). Induction of NKG2DLs occurs in response to stress, such as virus infection, and has been shown to be down-regulated by E1A which renders infected cells susceptible to NK cell attack (32). However, while infection enhances synthesis of NKG2D ligands, MICA and MICB expression on the cell surface is suppressed by E3-19K (33). This is a newly discovered function of E3-19K, in which it sequesters both MICA and MICB within the ER and Ad successfully evades NK activation (31).

The knowledge of the mechanism by which E3-19K protein targets MHC class I molecules for retention in the ER is derived mainly from studies of Ad2. Interestingly, there are low levels of amino acid conservation between E3-19K proteins of different Ad serotypes, suggesting that proteins from different serotypes have distinct MHC-I binding properties. This leads to differing abilities of Ad serotypes to cause persistent infections (34). The structural requirements of E3-19K for MHC-I and MICA/B binding are well characterized including a number of conserved amino acids within the luminal domain (35).

The E3-10.4-14.5 complex, named receptor internalization and degradation, RID, modulates a selective set of plasma membrane receptors involved in apoptosis and growth control (28). E3-10.4, or the RIDα subunit, is expressed as two isoforms: in one, the signal peptide is cleaved, whereas in the other form, it remains attached and serves as a second membrane anchor. The two isoforms form a disulfide-linked dimer and associate with E3-14.5, or the RIDβ subunit. RIDβ is a type I transmembrane protein that is O-glycosylated and phosphorylated. Both viral proteins contain sequence elements in their cytoplasmic tails that include conserved transport motifs. RIDβ has three putative tyrosine motifs in the cytoplasmic tail and RIDα has a dileucine sorting motif, both of which trigger rapid internalization from the cell surface and aid in sorting to endosomal and lysosomal

compartments, as well as mediate trafficking to the trans-Golgi network (TGN) (36). The YXXϕ motif is important for endocytosis of the RID complex, while the dileucine motif plays a role in targeting the complex to a recycling pathway, diverting it away from a degradation pathway. Therefore, RID targets specific cell surface receptors for degradation, but at some point along the endocytic pathway, RID drops off its target molecule and recycles back to the cell surface. A closer look at the RIDβ subunit showed that it is phosphorylated on serine 116 and that tyrosine 124 is important for the protein-protein interaction function of the complex, including Fas- and TRAIL- induced apoptosis (see below) (36). The RID complex is localized predominantly to the plasma membrane, but neither subunit alone can reach the cell surface: RIDα alone is primarily localized in the golgi, and RIDβ alone is localized in the ER and golgi (12, 28).

E3-RID removes epidermal growth factor receptor, EGFR, from the cell surface by diverting internalized receptors to a degradation compartment and protects the infected cells from ligand-induced apoptosis (12, 28). E3-RID also plays a role in protecting human lymphocytes from apoptosis induced by ligation of Fas, a mechanism important for regulating lymphocyte populations. Upon ligand engagement with, and subsequent trimerization of Fas, Fas associates with death domain (DD) protein FADD via the DD present in both proteins. In turn, the death effector domain present in FADD and procaspase 8 interact resulting in the cleavage, and production of caspase 8, eventually leading to cellular apoptosis. E3-RID down-regulates surface Fas by a similar mechanism as EGFR, via endocytosis of the receptors into endosomes, followed by transport to and degradation within, lysosomes. This blocks events immediately downstream of Fas ligation, like Fas-FADD association, and caspase-8 cleavage (28, 37). It is possible that expression of RID

facilitates long-term infection by preventing Fas mediated deletion of persistently infected lymphocytes. The apparently overlapping functions of E3-RID can be separated at the molecular level, leading to distinct mechanisms of action. For example, the RIDα subunit has an extracellular domain that contains sequences that are important for down-regulation of EGFR, but not Fas, suggesting that RID has a specific mechanism for the down-regulation of receptors (38).

As introduced earlier, TNFR1 is a proinflammatory receptor that activates both NF-κB and AP1 transcription factors in parallel. Both transcription factors are involved in proapoptotic and proinflammatory functions, both of which hinder the survival of the virus in infected cells. E3-RID complex is sufficient to inhibit TNF signal transduction through TNFR, and the inhibitory effects extend to both the AP-1 pathway and the NF-κB pathway (30). E3-RID's inhibition of signal transduction through TNFR is accomplished by elimination of TNFR1 from the cell surface by clathrin mediated endocytosis. E3-RID down-regulates TNFR1 via an AP2 clathrin mediated pathway and associates with TNFR1 in the same complex (39). Once again, the mechanism of action for down-regulation of TNFR1 and Fas by E3-RID are different: it is likely that E3-RID associates with TNFR1 on the plasma membrane, and Fas in the endosome (similar to where E3-RID associates with EGFR). Additional studies have shown that the E3-RID complex also interferes with the activation of the NF-κB pathway by interfering with a critical phosphorylation step and preventing NF-κB from entering the nucleus and becoming active (40).

TNFα-induced apoptosis requires activation of one or more cytokines or chemokines. Chemokines play important roles where injury and infection are present, for the early phase

of inflammation. Interleukin-1 beta (IL-1β) is an example of such chemokine that is produced in response to inflammation, or induced by lipopolysaccharide, a component of the cell wall of Gram-negative bacteria. Both IL-1β and lipopolysaccharide share significant portions of the signaling pathways downstream of the respective receptors, IL-1R and TLR4, including the adaptor molecules that are required to stimulate signaling and both signaling pathways are strongly inhibited by E3-RID. However, E3-RID expression has differential inhibitory effects on IL-1R and TLFR4 signaling (41). The inhibition of TLR4 signaling by E3-RID does not involve receptor down-regulation. This suggests a new, as yet unidentified mechanism by which E3-RID disrupts chemokine expression and signal transduction without having a direct effect on the receptor. In addition, TNFα-induced apoptosis requires the activation of cytosolic phospholipase a2 (cPLA2), an enzyme responsible for the production of inflammatory mediators. In the presence of submicromolar levels of ca^{+2}, cPLA2 translocates to membranes where it cleaves arachidonic acid, AA, a potent mediator of inflammation, from membrane phospholipids. TNF-induced release of AA is inhibited by E3-RID (and E3-14.7K, discussed below), by preventing TNFα-induced translocation of cPLA2 to membranes of infected cells (28, 37).

E3-RID is one of three proteins encoded by Ads that independently inhibit TRAIL-induced apoptosis of infected cells. The other two proteins are E1B-19K and E3-14.7K (42). T lymphocytes and NK cells use TRAIL to induce apoptosis in virus-infected and tumor cells. TRAIL induces apoptosis through two receptors, TRAILR1 (or death receptor 4) and TRAIL R2 (death receptor 5) (28). E3-RID induces the internalization of TRAIL-R1 from the cell surface where it is internalized in endosomes/lysosomes and is eventually degraded in lysosomes (42). One study showed that E3-RID is necessary and sufficient for down-

regulation of TRAIL-R1 (42), while another study showed that E3-RID may need E3-6.7K (discussed below) for this function (43). Later, it was shown that both E3-RID and E3-6.7K are necessary for internalization and degradation of TRAIL-R2, whereas only E3-RID is required for TRAIL-R1 down-regulation (44).

E3-6.7K is a small integral hydrophobic integral membrane glycoprotein encoded by subgroup C Ads (28). A small portion of the E3-6.7K protein is localized on the plasma membrane and forms a complex with the E3-RIDβ protein, and this complex is sufficient to induce down-modulation of TRAIL-R1 (and possibly TRAIL-R2) from the cell surface and inhibit TRAIL mediated apoptosis (43, 44). E3-6.7K was shown to maintain ER Ca^{+2} homeostasis and inhibit the induction of apoptosis, reduce the levels of AA induced by TNFα, and protect cells against apoptosis induced through Fas and TNFα, in addition to TRAIL receptors. E3-6.7K translocates across the membrane of the ER in a posttranslational, ribosome-independent, ATP-dependent manner. In addition, it has the ability to adopt more than one membrane topology (45). It contains a single hydrophobic stretch that initiates membrane insertion and acts as the transmembrane domain, containing a signal anchor sequence. E3-6.7K has defied much of the current understanding of membrane protein conformation.

The nonmembrane E3-14.7K protein has a large proportion of aa with charged residues, giving the protein hydrophilic properties. It is localized to both the cytosol and nucleus. E3-14.7K inhibits TNFα-induced apoptosis by an independent mechanism (12, 28). TNFα-mediated apoptosis is initiated by ligand-induced recruitment of TNFα receptor-associated death domain, TRADD, Fas-associated death domain, FADD, and caspase-8 to the death domain of TNFR1, thereby establishing the death-inducing signaling complex,

DISC. E3-14.7K inhibits apoptosis by targeting key factors in this process. For example, caspase 8 cleaves key structural components of cells ultimately leading to apoptosis and cPLA2 activation. E3-14.7K can bind and inhibit the function of caspase 8, and ultimately inhibit cPLA2 levels and the release of AA, reviewed in (12, 28). A number of cellular E3-14.7K-interacting proteins, named FIPs, are known to bind E3-14.7K directly. The FIPs in turn interact with other cellular proteins that are involved in membrane trafficking and morphogenesis, the NF-κB signal transduction pathway, cell cycle control, and trafficking from and to the nucleus and cytoplasm (12, 28). New molecular mechanisms implemented by E4-14.7K to escape immunosurveillance have been discovered over the past decade. For instance E3-14.7K targets TNFR1 endocytosis and directly prevents TNFα-induced DISC formation (46). Additionally, E3-14.7K is a potent inhibitor of NF-κB transcription activity following Toll-like receptor, TLR, or TNFR signaling. E3-14.7K can inhibit transcriptional activity at a point distal to the initial signaling pathway: specifically, it prevents binding of NF-κB complexes by directly binding to the NF-κB subunit p50 (47). E3-14.7K has also been shown to independently inhibit TRAIL-induced apoptosis in certain cell types (42), and inhibit STAT1 function by preventing its phosphorylation and nuclear translocation (48).

To elucidate the molecular mechanism for E3-14.7K-mediated cell death protection, the biophysical properties of the protein have been characterized. The C-terminal two-thirds of the protein is highly structured and binds its putative cellular receptors. This C-terminal domain retains the capacity to interact with FIP-1 and the death effector domain (DED) of caspase 8, and binds zinc. Additionally, another FIP found is optineurin, OPTN (reviewed in (49)). OPTN has several functions including regulation of receptor endocytosis, vesicle

trafficking, antiviral signaling, and regulation of NF-κB pathway. Binding of OPTN to E3-14.7K potentially recruits both proteins to TNFR1 complex, but OPTN is dispensable for E3-14.7K-mediated protection against TNFα-induced cytotoxicity (50).

Recently, a new E3 protein, E3-49K protein was discovered with subgroup D Ads. Overall, E3 represents one of the most divergent regions of Ads, and some E3 genes seem to be unique to a particular subgroup. Ad19a and Ad65 contain E3-49K gene which is absent in Ads of other subgroups but is present in all subgroup D Ads examined. This protein is a glycosylated type I transmembrane protein of approximately 80-100kDa molecular weight, the largest E3 protein discovered. The protein is localized in the Golgi-trans-Golgi network, in early endosomes, and in lysosomes during the late phase of infection. E3-49K is cleaved and the large ectodomain is secreted. This finding is significant, because to date, E3-49K is the only known E3 protein that is shed or secreted and offers an excellent example of immunomodulatory activities that are expressed by a single Ad subgroup. Interestingly, the protein targets noninfected leukocytes by binding to the protein phosphatase CD45, which suppresses leukocyte activation. E3-49K suppresses the functions for both NK cells and T cells (51).

Finally, the E3-11.6K protein, now known as adenovirus death protein, ADP, plays a role in efficient cell lysis and subsequent release of virus from the infected cell. ADP is a nuclear membrane and Golgi glycoprotein that is produced in small amounts from scarce E3 mRNAs at early times post-infection, but is greatly amplified at late stages of infection (52). ADP expression at late stages of infection is regulated by the L4-22K protein, a master regulator of late gene expression (discussed below) (53). During the late stages of infection, ADP is synthesized from mRNAs that contain the Ad tripartite leader (TPL). Over the course

of infection, the abundance of ADP increases in the Golgi and ER, but ultimately ADP accumulates in the nuclear membrane late in infection. Extensive structure-function studies have identified specific domains important for protein processing, exit from the Golgi, and subcellular localization. Specifically, the luminal domain of ADP is important for protein stability and efficiency of cell lysis, while the cytoplasmic-nucleoplasmic domain is important for protein localization. In addition, ADP undergoes a complex process of N- and O-linked glycosylation and proteolytic cleavage (54). Just as it is critical for viruses to block apoptosis and evade the immune system, viruses must eventually be released from the cell in order to further amplify virus production. Mutation or deletion of ADP leads to a small plaque phenotype due to impaired virus release (55), which suggests that ADP plays an important role in killing infected cells and for virion release. ADP may cause virus release by disrupting nuclear membrane integrity. Interestingly, the requirement for ADP for virus spread can be readily compensated for by disrupting the functions of the E1B-19K protein, suggesting that the two viral proteins act as antagonists to influence viral spread (52). Both caspase-dependent and caspase-independent mechanisms of cell killing have been observed when ADP is over-expressed (56), but the mechanism is not well understood. However, over-expression of ADP has been reported to induce a pattern of cell death that exhibited some characteristics of apoptosis (57), while other studies show that ADP is involved in induction of necrosis-like cell death (58).

Early Region 4 (E4)

Whether they are expressed early or late during infection, a common theme among the Ad transcription units is that they encode multiple proteins of related functions. However, Ad early region 4 (E4) is the only transcription unit that produces proteins of relatively disparate functions. E4 encodes at least seven proteins according to analysis of open reading frames (ORF) and spliced mRNAs (Fig. 1-3). The gene products exhibit a wide range of activities. Proteins expressed from the E4 region have been shown to be important for transcriptional regulation, viral DNA replication, viral mRNA transport and splicing, shutoff of host cell protein synthesis, oncogenic transformation, and the regulation of apoptosis. Subgroup D Ads are unique in their ability to induce estrogen-dependent mammary tumors in animals (59). The primary oncogenic determinant of these viruses is the E4-ORF1 protein, rather than the E1A and E1B proteins described above (60). Based on sequence similarity, the E4-ORF1 protein appears to have evolved from a cellular dUTP pyrophosphatase gene, although E4-ORF1 does not possess this enzymatic activity (60). Rather, E4-ORF1 appears to have utilized the structural aspects of dUTP pyrophosphatases in order to form homotrimers. The tumorigenic property of E4-ORF1 depends on a C-terminal sequence motif referred to as a PDZ domain. Ad9 E4-ORF1 contains a class I PDZ-binding domain motif following the consensus sequence (S/T)-X-(V/I/L)-COOH (X is any aa) at the C-terminus of the protein (60). PDZ domains are involved in protein–protein interactions and, in the case of Ad9 E4-ORF1, mediate the binding of E4-ORF1 to cellular proteins including Dlg1, MUPP1, PATJ, MAGI-1, and ZO-2 (60). PDZ proteins function as scaffolds to target signaling complexes to specific sites at the plasma membrane, and the PDZ proteins that bind to Ad9 E4-ORF1 have tumor suppressor activities. The binding of Ad9 E4-ORF1 to PDZ-containing proteins mediates the oncogenic function of this viral gene product. MUPP1,

E4 protein	Function(s)
ORF1	binds PDZ proteins, stimulates proliferation/transformation
ORF2	unknown
ORF3	inhibits PML-NB activities, blocks the DNA damage response, inhibits p53
ORF4	regulates PP2A, induces cell killing
ORF6	forms E3 ubiquitin ligase complex and induces degradation of cellular substrates, blocks the DNA damage response, inhibits p53, host cell shutoff
ORF6/7	regulates E2Fs

Figure 1-3. Schematic of the E4 region. The different proteins encoded by the E4 region are indicated by bars. The functions ascribed to different E4 proteins are listed below the diagram.

PATJ, MAGI-1, and ZO-2 localize to tight junctions at sites of cell-cell contact with epithelial cells. In epithelial cells, Ad9 E4-ORF1 prevents localization of PATJ and ZO-2 to tight junctions, and as such, disrupts tight junctions resulting in a loss of apicobasal polarity (60). Tight junction disruption and the loss of apicobasal polarity are common features of epithelial cancers. An additional function of Ad9 E4-ORF1 is activation of PI-3 kinase and this activity also is involved in the oncogenic properties of this protein. Activation of PI-3 kinase is associated with many human cancers and the interaction of Ad9 E4-ORF1 with Dlg1 may mediate this process (60). Finally, the E4-ORF1 protein recently was shown to induce Myc activation and promote cellular anabolic glucose metabolism to promote viral replication (61).

The E4-ORF3 protein is highly conserved among different Ads and is multifunctional. The E4-ORF3 and E4-ORF6 proteins both bind to the E1B-55K product, although with different outcomes. E4-ORF6 enhances the inhibition of p53 by E1B-55K, whereas E4-ORF3 transiently relieves the repression of p53 by E1B-55K (62). E4-ORF3 has been shown to localize with discrete nuclear structures alternatively known as PML nuclear bodies (PML-NB), PML oncogenic domains (PODs), or ND10 (63). PML-NBs exist as multiprotein complexes that exhibit a discrete, punctate appearance in the nucleus of a cell. E4-ORF3 is necessary and sufficient to cause redistribution of these protein complexes into long, track structures. PML-NBs have been implicated in a number of cellular processes including transcriptional regulation, the regulation of apoptosis, DNA damage repair, protein modification, and an antiviral response (64). PML-NBs have also been shown to react to stresses such as heat shock and heavy metals as well as interferon, suggesting a role in

cellular defense mechanisms. The E4-ORF3 protein of subgroup C Ads (e.g., Ad2 and Ad5) inhibits a cellular DNA damage response. E4-ORF3 directs the reorganization of the Mre11-Rad50-Nbs1 complex (MRN complex) into PML-containing tracks (62). The MRN complex serves as a sensor of DNA damage and is recruited to the ends of damaged DNA. This serves to trigger effector cascades that lead to cell cycle arrest and the repair of the DNA damage. In the context of Ad infection, this process, if uninhibited, results in the end-to-end ligation of viral genomes, effectively inhibiting viral DNA replication. E4-ORF3 interferes with this process by sequestering MRN proteins in the nucleus and blocking their function (62). The E4-ORF3 protein also blocks p53 signaling by inducing heterochromatin formation at p53-induced promoters, thereby blocking p53 DNA binding and transactivation (65).

The E4-ORF4 protein is a multifunctional regulator. First, E4-ORF4 binds to the B55 subunit of the serine/threonine phosphatase PP2A (66, 67). By binding this subunit, the trimeric form of PP2A is activated which results in the dephosphorylation of target proteins such as mitogen-activated protein (MAP) kinases that are important in signal transduction pathways. Increased PP2A activity leads to decreased phosphorylation and inactivation of certain transcription factors, such as E4F, through direct interaction or through the inactivation of MAP kinases. E4-ORF4 expression also results in decreased E1A phosphorylation at MAP kinase consensus sites that are important for E4 transactivation (66, 67). Through decreasing the activity of E1A and E4F, E4-ORF4 regulates the expression of the E4 region itself and thus may suppress the oncogenic potential of the E1A proteins. Second, E4-ORF4 is able to induce p53-independent apoptosis in transformed cells (66, 67). Oncogenic transformation of cells sensitizes them to E4-ORF4-induced cell killing. Depending on the cell type, E4-ORF4-induced apoptosis utilizes the classical pathway

involving caspases or a non-classical pathway that is caspase-independent (66, 67). The binding to, and regulation of, PP2A by E4-ORF4 is essential for the induction of cell death. E4-ORF4-dependent apoptosis also requires modulation of Src-family kinases (66, 67). Thus, by an alternative mechanism, E4-ORF4 represses the oncogenic potential of the Ad E1 proteins. E4-ORF4 may be useful in the future as a therapeutic agent to target human cancers for apoptotic cell death.

The E4-ORF6 protein binds to and inhibits p53, providing Ad yet another defense for p53 effects within the cell (13, 59, 62). E4-ORF6 augments the transformed phenotype of Ad E1-transformed cells through the down-regulation of p53 expression (13, 59, 62). E4-ORF6 forms a direct protein complex with the E1B-55K protein. The E4-ORF6/E1B-55K complex recruits a CUL5-containing E3 ubiquitin ligase complex to target p53, and other cellular proteins including proteins involved in DNA damage repair, for poly-ubiquitination and proteasome-dependent degradation (13, 59, 62). By this mechanism, the E4- ORF6/E1B-55K complex counteracts the induction of p53 stability provided by E1A. Other targets of this virus-induced E3 ligase activity include Mre11, Rad50, and DNA ligase IV (13, 59, 62).

The E4-ORF6/7 protein is produced from a spliced mRNA that encodes the amino terminus of E4-ORF6 linked to the unique E4-ORF7 sequence. E4-ORF6/7 forms stable homodimers that contribute to viral DNA synthesis by enhancing the production of E2 products. E4-ORF6/7 binds free E2F and induces cooperative and stable binding of E2F/DP heterodimers to inverted E2F-binding sites in the Ad E2 early promoter (68). E4-ORF6/7 induces expression from the cellular E2F-1 promoter and is able to compensate functionally for E1A in Ad infection by displacing Rb family members from E2Fs (69). Further, the E4-6/7 protein alters the subcellular localization of E2F family members and directs E2F-4 from

the cytoplasm to the nucleus (70). Thus, E4-6/7 displays functional redundancy with the E1A proteins in terms of activating E2F family members.

Viral DNA Replication

Ad DNA replication is a very efficient process. Infected cells produce approximately one million new copies of viral DNA, similar to the cellular DNA content. Three viral proteins (pTP, Ad-Pol, and DBP) and three cellular transcription factors (NFI, OCT-1 and NFII) are required for efficient Ad replication. The core origin of replication, a conserved region located between nucleotides (nt) 9-18 at the ends of the viral genome, binds the pTP-Pol complex. Adjacent to the core origin is the auxiliary origin bound by NFI and OCT-1. NFI and OCT1 enhance initiation by several hundred-fold (25, 71). Between the core region and auxiliary region is an A/T-rich region, also important for replication.

NFI is part of a family of related proteins that plays an important role in transcription regulation of a large variety of cellular and viral genes (71). The conserved DNA binding domain and dimerization region of NFI are required for efficient Ad DNA replication. NFI binds as a dimer to the consensus sequence TGG(C/A)(N_5)GCCAA located at the origin of replication nt 25 to 38. DBP enhances the binding of NFI to the auxiliary origin by causing a structural change in the DNA that allows for increased flexibility of the NFI binding site (25). NFI stabilizes binding of the pTP-Pol complex to the origin by directly interacting with the pTP-Pol complex (71). NFI influences the kinetics of replication by increasing the number of active initiation complexes. Specifically, NFI facilitates formation of the pre-

initiation complex by inducing a bend of 60° in DNA upon binding to the Ad origin which, in addition to the origin binding site, requires the A/T-rich region preceding this site (72). OCT-1 belongs to the family of octamer-binding transcription factors containing a POU DNA binding domain (73). OCT-1 binds at nt 39-47 adjacent to the NFI binding sequence and interacts with pTP via its POU domain. This interaction recruits pTP to the origin and tethers it there. Both NFI and OCT1 act in concert to enhance initiation, bending the DNA, and facilitating optimal assembly of the pre-initiation complex, thereby strongly stimulating replication (72, 73).

After the preinitiation complex is formed, the stress induced by bending assists in the unwinding of the origin. Ad replication begins, and employs a protein primer for replication initiation. During initiation, pTP presents its Ser-580 residue to the Ad-Pol active site, and influences its catalytic activity. After binding with an incoming dCTP nt, Ad-Pol covalently couples the first dCTP residue with the OH group of Ser-580 (25, 72, 73). After coupling of the first dCTP, a pTP-CAT nt intermediate is formed using an integral GTA triplet at positions 4-6 as a template. DBP stimulates the formation of the pTP-CAT intermediate by lowering the K_m for the reaction by influencing the Ad-Pol active site, either via direct interaction with Ad-Pol, or by changing the template conformation (25, 73). Subsequently, the pTP-CAT intermediate jumps back three bases and becomes paired with template residues 1-3 at the beginning of the template strand. A common feature among protein-primed replication systems is replication initiation at an internal start site, rather than at the genome termini. A similar so-called sliding back mechanism was first identified in bacteriophages. This mechanism allows for error corrections during initiation that cannot yet be repaired by the proofreading ability of Ad-Pol because when bound to pTP, the

exonuclease activity of Ad-Pol is inhibited (25, 26). During, or shortly after, the three-nucleotide jump, Ad-Pol dissociates from pTP, increasing the rate of polymerization and the proofreading activity of Ad-Pol (25, 26).

During elongation, the pTP-CAT primer is efficiently and processively elongated by Ad-Pol by strand displacement. DBP enhances the rate of processivity of Ad-Pol and modifies the sensitivity to nucleotide analogues indirectly, by modifying the DNA structure (25, 72). DBP helps to unwind short stretches of double-stranded DNA in an ATP-independent manner. A third cellular factor, NFII, is necessary for elongation and ensures synthesis of genome-length DNA. NFII is a type I DNA toposisomerase (72, 73). During elongation, a new duplex genome is formed and the non-template strand is displaced. The ITRs of the displaced non-template strand may anneal together to restore the functional origins and can serve as substrates for new rounds of replication.

VA RNA Genes

Ad encodes two virus-associated (VA) RNAs, VA RNAI and VA RNAII, transcribed by cellular RNA polymerase III of about 160 nt. They are single-stranded RNA molecules that fold into a well-conserved secondary structure that can be divided into a terminal stem, an apical stem, and a central domain. The VA RNAs counter cellular anti-viral defense mechanisms in order to allow efficient synthesis of viral proteins (74). VA RNAI blocks the activity of the RNA-dependent protein kinase, PKR, an interferon (IFN)-inducible protein kinase activated in infected cells as part of the antiviral response (74). Ad infection leads to

the production of double-stranded RNA, which activates PKR. Activated PKR phosphorylates the α subunit of eukaryotic initiation factor 2 (eIF-2), which eventually leads to the shutoff of protein synthesis. Ad VA RNAI binds to PKR and blocks PKR activation to maintain protein synthesis. In addition, VA RNAI stabilizes ribosome-associated viral mRNAs, which could lead to enhanced levels of protein synthesis (74). The apical stem-loop is required for the binding to PKR and the central domain is involved in the inhibition of PKR activation (74).

Late Gene Expression

Ad replication initiates a transcription cascade that drives late gene expression from the MLP (Fig. 1-4). The MLP shows basal transcriptional activity at early times of infection, with an efficiency comparable to other early viral promoters, but it is highly active late after infection. During the replication cycle, there is a switch from early to late gene expression in which the MLP is fully activated (75). This activation requires replication initiation, and at least one viral protein, IVa2. Transcription of IVa2 is first activated as a result of viral DNA synthesis-dependent titration of a cellular transcriptional repressor that binds to the IVa2 promoter. The synthesis of IVa2 in turn leads to maximally efficient transcription from the MLP. The late-specific increase of MLP activity requires additional *cis*-acting sequence elements located downstream of the MLP start site, termed downstream elements, DE1 and DE2 (Fig. 1-4) (76). DE1 is bound by a heterodimer of IVa2 (DEF-A) while DE2 is bound by DEF-B, which is suggested to be a heterodimer of IVa2 and an unknown viral protein.

One study found that the L4-33K protein is the only DEF-A component and that IVa2 and L4-33K are the only proteins necessary and sufficient for controlling the MLP (77). However, other reports have presented evidence suggesting that L4-22K binds specifically to DE sequences and activates the MLP (78-80).

The MLP drives transcription of a the major late transcription unit, MLTU, that results in five different groups of mRNAs (L1 to L5; Fig. 1-4). Ad late proteins act as capsid structural proteins, promote assembly, direct genome packaging, and serve regulatory functions. The MLP produces all late mRNAs by alternative splicing and alternative polyadenylation of a primary transcript (Fig. 1-4). Prior to DNA replication, the MLP is active at low levels with transcription proceeding only as far as the L3 region and mRNA production restricted to the L1-52/55K and i leader proteins (75). Following DNA replication, the MLP is fully activated and transcription proceeds to the L4 and L5 regions. After polyadenylation, Ad late primary transcripts are spliced so that each mature mRNA contains the untranslated TPL composed of leaders 1, 2, and 3 (Fig. 1-4). There is a fourth exon, called the i-leader, that is sometimes inserted between leaders 2 and 3 in early L1 52/55K transcripts, but the i-leader is not present in all other late mRNAs. L1 mRNAs that contain the i-leader coding region are less stable that those that lack it (81). The TPL enhances translation of mRNAs during the late phase of Ad infection and can also increase the efficiency of mRNA export from the nucleus (82). The L4-100K protein is the first late viral protein to be synthesized and it promotes translation of TPL mRNAs by ribosome shunting and inhibits cellular protein synthesis (83). Ribosome shunting involves the loading of the 40S ribosome subunits to the 5’ end of the capped mRNA, followed by its direct translocation to the downstream initiation codon, directed by shunting elements in the TPL.

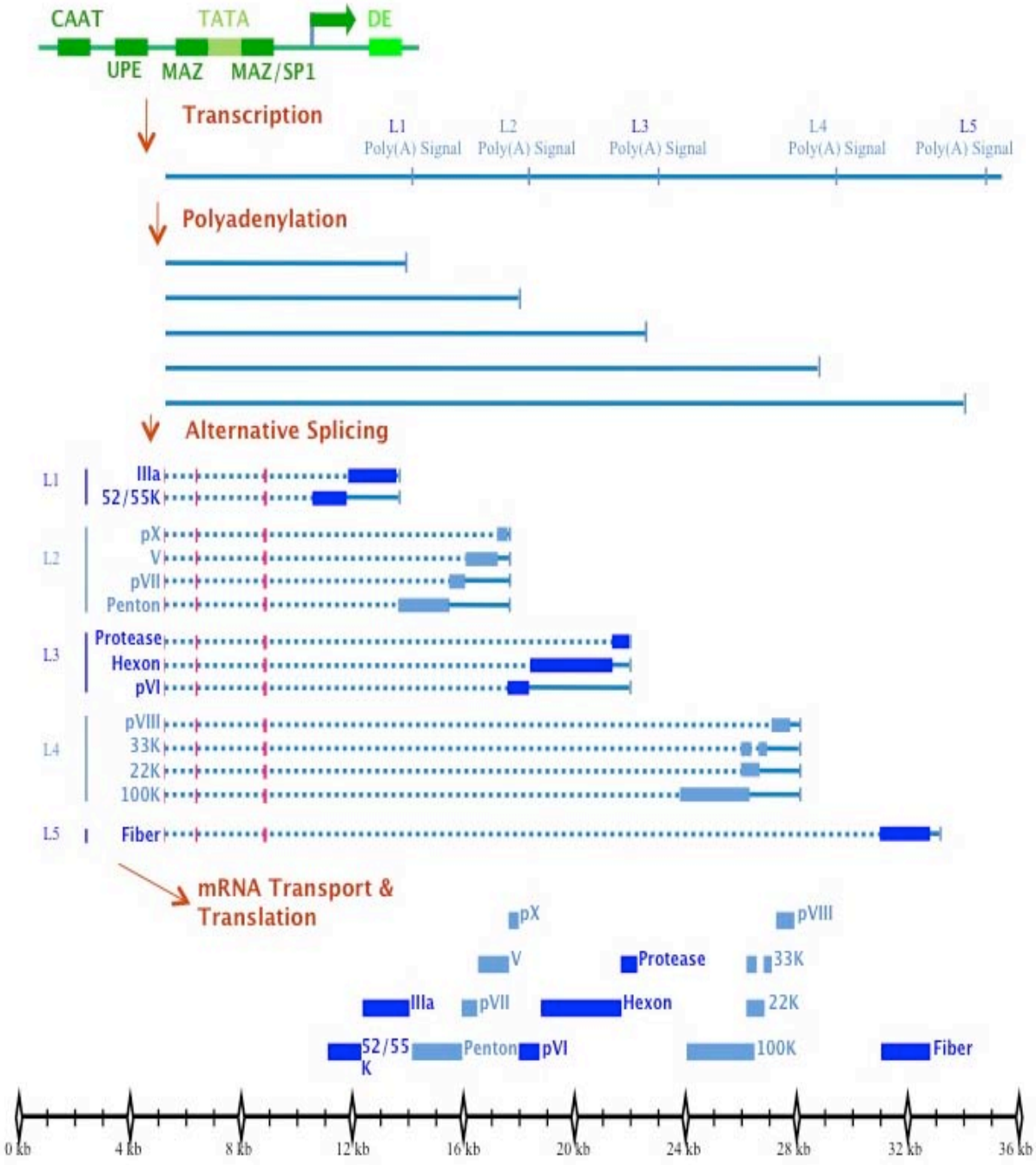

CAAT
TATA
DE
UPE
MAZ
MAZ/SP1
Transcription
L1
Poly(A) Signal
L2
Poly(A) Signal
L3
Poly(A) Signal
L4
Poly(A) Signal
L5
Poly(A) Signal
Polyadenylation
Alternative Splicing
L1
IIIa
52/55K
L2
pX
V
pVII
Penton
L3
Protease
Hexon
pVI
L4
pVIII
33K
22K
100K
L5
Fiber
mRNA Transport &
Translation
pX
pVIII
V
Protease
33K
IIIa
pVII
Hexon
22K
52/55
K
Penton
pVI
100K
Fiber
0 kb
4 kb
8 kb
12 kb
16 kb
20 kb
24 kb
28 kb
32 kb
36 kb

Figure 1-4. Schematic of the major late transcription unit. The MLP, depicted in green, drives transcription from the MLTU regions L1 to L5 producing all late mRNAs by alternative splicing and polyadenylation of the primary transcript. Following polyadenylation, the primary transcript is spliced so that each mature mRNA contains the untranslated TPL composed of leaders 1, 2, and 3, depicted as red rectangles. The TPL is then spliced to one of many alternative 3' splice sites, generating the late cytoplasmic mRNAs, depicted as light blue and dark blue rectangles. Introns are depicted as dashed lines.

The cellular initiation factor eIF4F is a protein complex composed of phosphorylated cap-binding protein eIF4E, eIF4E kinase MnK1, eIF4A, poly(A) binding protein, and eIF4G. L4-100K binds to the carboxyl terminus of eIF4G and competitively displaces MnK1 from cap initiation complexes, preventing eIF4E phosphorylation (83). Late in infection, modified L4-100K cap initiation complex associates with higher specificity to mRNAs that contain the TPL. The L4-100K-TPL complex enhances association with initiation factor eIF4G and poly(A) binding protein, and together, the complexes are recruited to Ad late mRNAs, which utilize the complex to promote TPL-directed translation by ribosome shunting (83, 84). The mechanism behind this function of L4-100K is partially understood. L4-100K is tyrosine phosphorylated, which was shown to be essential for efficient ribosome shunting and late protein synthesis, but not involved in binding eIF4G (83). Additionally, L4-100K is arginine-methylated, a modification that plays an essential role in modulating protein-protein/RNA interactions by L4-100K (85).

Following full activation of the MLP, transcription continues to the L4 and L5 regions (Fig. 1-4). L4 poly(A) site usage is prominent and it is only after an increase in expression of L4 gene products that the L2, L3, and L5 poly(A) sites are used generating the complete late viral gene expression profile (86). The L4 region encodes four proteins: L4-100K (described above), protein pVIII, a structural protein of the viral capsid, L4-33K, a viral splicing factor, and L4-22K. Besides its role in DNA encapsidation (described below) L4-22K is associated with other functions pertaining to the regulation of Ad late gene expression (53, 79, 87). The timing of L4 expression fits well with the idea that its products function to regulate late viral gene expression (79, 86). However, the essential role of L4 family proteins at this stage in the viral life cycle creates a paradox, since their expression is

achieved only as a consequence of the activation of late-phase expression. Morris *et al.* identified a novel, intermediate-phase Ad5 promoter, the L4 promoter (L4P), that directs the expression of the L4-22K and L4-33K proteins independent form the MLP, providing an apparent solution to this paradox (88). L4P is active in its natural context and the amount of L4-22K protein expressed via L4P is sufficient to induce the early-to-late transition in MLTU activity. L4P is strongly activated by the replication of the viral genome and is partly activated by E1A, E4-ORF3, and IVa2. This promoter is also significantly activated by the cellular stress response regulator, p53 (89). In addition, the two products of L4P activation, L4-22K and L4-33K, inhibit p53 activation of the promoter, suggesting the presence of a negative feedback mechanism controlling L4P.

Following polyadenylation, the TPL is spliced to one of many alternative 3' splice sites, generating the 20 or more cytoplasmic mRNAs that encode the Ad late proteins (Fig. 1-4) (75). The individual 3' acceptor splice and poly(A) sites within the main MLTU body have different efficiencies of usage that change as infection proceeds. The L4-33K protein serves as a viral splicing factor in the L1 unit, (75) and will be described in detail below. Although the role of L4-33K in alternative splicing is well understood, the role of L4-22K in the regulation of Ad late gene expression is a very young field of study, and a deeper understanding regarding the mechanisms of regulation is still needed. Morris and Leppard established that L4-22K has an important role in regulating the pattern of MLTU gene expression and this role is independent of the effect of L4-33K on late mRNA splicing (79). Analysis of the phenotypes of L4-22K mutant viruses confirmed that L4-22K is important for the transition of early-to late viral gene expression (53). In this study, we show that L4-22K acts at the level of late gene mRNA production and that this function is specific to a unique

set of mRNA transcripts, with L4-33K and pVIII identified as the primary targets for L4-22K regulation. Specifically, L4-22K is required for efficient splicing of L4-33K mRNA transcript, and subsequent expression of the protein. We propose a working model for regulated late gene expression in which L4-22K is expressed early in infection from the L4P (88) and serves as the master regulator of late gene expression. In this model L4-22K regulates accumulation of different L4 transcripts, including its own, thereby regulating the levels of L4 proteins (53, 79). As L4-22K protein levels increase, L4-33K pre-mRNA splicing is stimulated which effectively reduces the accumulation of L4-22K mRNAs. L4-33K, in turn, acts as a virus-encoded regulator of late mRNA-splicing (90). Additionally, we show that L4-22K suppresses E1A expression during the late stage of infection, although it is not clearly understood if this is also achieved by a post-transcriptional mechanism (53). Thus, L4-22K is proposed to finely tune viral gene expression patterns during the late phase of infection.

Viral DNA Packaging

The assembly of Ad particles is a multistep process that takes place in the nucleus. There are four different Ad particle forms identified by CsCl equilibrium centrifugation: empty capsids devoid of viral DNA, light assembly intermediates that contain the left end of the genome, heavy assembly intermediates that contain the full viral genome and precursor forms of certain capsid proteins, and mature virus particles that contain the full-length genome with proteolytically processed capsid proteins. Efficient packaging is required for the

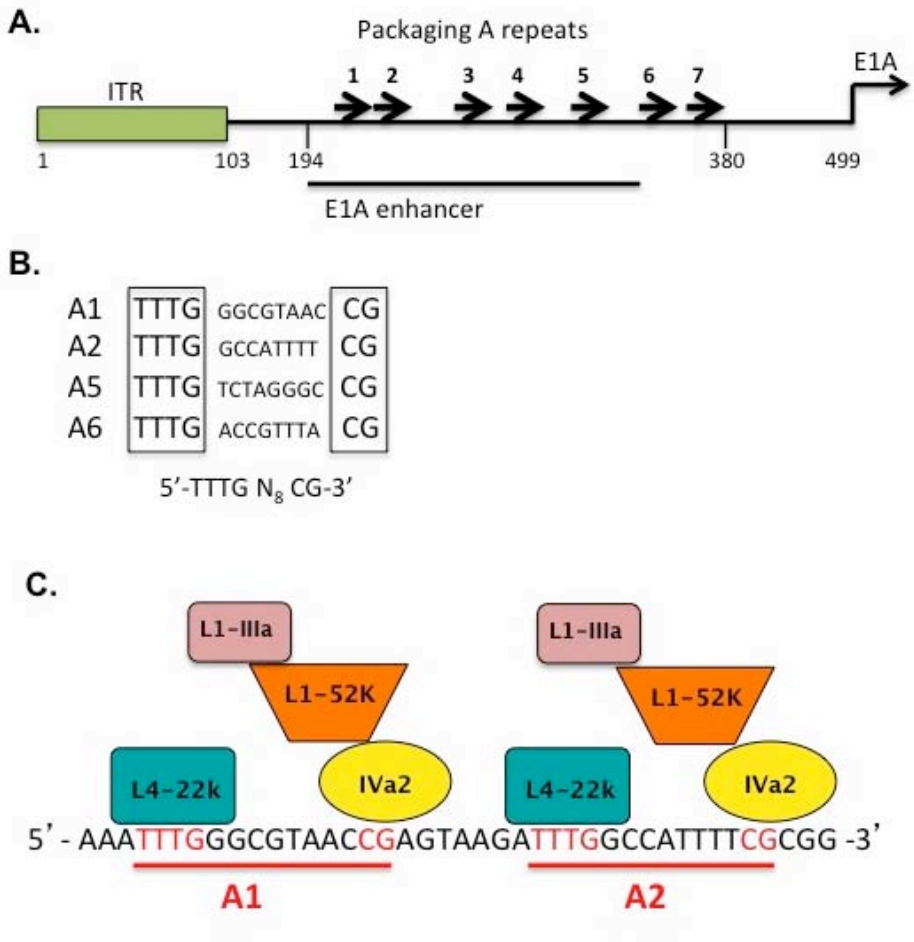


Figure 1-5. Model of Adenovirus DNA packaging.

Panel A. A schematic diagram of the left end of the Adenovirus genome including the ITR, the packaging/enhancer regions, nt 194 to 380, and the E1A 5' flanking region. The packaging repeats (A1 through A7) are represented by arrows. The E1A transcriptional start site is indicated by an arrow at nt 499.

Panel B. The A repeat consensus sequence is shown.

Panel C. Model of Ad packaging: IVa2 binds to the CG motif of the A1 repeat, and L4-22K binds to the TTTG motif and forms a complex with IVa2 on the A2 repeat. A succession of complexes is formed in this manner, where an additional IVa2 protein binds to the CG nucleotide on the A2 repeat. The exact binding of L1-52/55K and L1-IIIa is still under question, but evidence suggests that L1-52/55K interacts with IVa2, while L1-IIIa interacts with the L1-52/55K protein.

formation of mature virus particles and a complex of DNA and proteins at the packaging domain plays a key role in this process. The current model of Ad DNA packaging includes a multiprotein complex for packaging similar to the molecular motors of double-stranded DNA bacteriophages, which insert their genomes into preformed capsids driven by an ATP-hydrolyzing molecular machinery The IVa2 protein was shown to bind ATP (91), leading to the speculation that IVa2 is the ATPase providing the power stroke of the packaging machinery.

Ad DNA packaging is dependent on a *cis*-acting region located at the left-end of the viral genome (nt 230 to 380) termed the packaging domain (Fig. 1-5). To achieve optimal packaging activity, the packaging domain must be near an end of the genome (92).The role of the packaging domain is to target the genome to an immature procapsid via proteins that bind to its sequences. The packaging sequences contain seven AT-rich repeats, termed A1 thru A7, that have the consensus sequence 5'-TTTG-N_8-CG-3' (92). Repeats A1, A2, A5, and A6 are the most important repeats functionally. The sequences of the eight spacer nt and the flanking nt are important for packaging function (92, 93). The identification of the core nucleotide sequences involved in packaging allowed for the production of simplified packaging domains consisting of multiple copies of one A repeat or several A repeats, with A1 and A2 commonly used as a synthetic sequence that functions as wild-type (94). Several viral proteins have been described for their critical function in DNA packaging, including IVa2 and L4-22K. IVa2 and L4-22K mutant viruses produce empty particles containing no viral DNA (75, 102). *In vitro,* IVa2 binds to the CG motif of the A1 repeat, and L4-22K binds to the TTTG motif and forms a complex with IVa2 on the A2 repeat. A model has been proposed in which IVa2 and L4-22K bind the A1 and A2 repeats as observed *in*

vitro and an additional IVa2 protein binds to the CG sequence in the A2 repeat (Fig. 1-5C) (78, 95). *In vivo* studies showed that the L4-22K and IVa2 proteins are dependent on each other for binding to the packaging domain (53). The binding of these proteins to the packaging domain may result in a complex initiating the formation of a portal entry site in an immature virus particle that leads to DNA encapsidation.

Detailed characterization of the assembly of IVa2 and L4-22K proteins onto packaging sequences shows that L4-22K binding promotes cooperative assembly of IVa2 onto packaging sequences (96). The critical roles of both of these viral proteins in DNA encapsidation, and other aspects of infection, have prompted several structure-function studies. The Ad IVa2 protein has highly conserved sequences that resemble Walker boxes A and B associated with ATPases, and these elements are required for Ad packaging to occur (97). The putative helix-turn-helix motif at the extreme C-terminus of IVa2 is responsible for DNA binding activity specifically related to packaging, but not its activation of the MLP (98). The L4-22K protein has a unique C-terminal region that is highly conserved among human Ad serotypes (87). We postulated that the functions of the L4-22K protein are mediated by this region and used a genetic approach to identify critical functional domains. In this study we focused on a conserved pair of cysteine residues and conserved pair of histidine residues originally believed to be a putative zinc finger of the L4-22K DNA binding domain. Two different L4-22K mutant viruses were constructed and analyzed: one with amino acid sunstitutions in the cysteine residues and the other with amino acid substitutions in the histidine residues. Analyses of these mutant viruses showed that the double cysteine mutant uncouples the functions of L4-22K in Ad DNA packaging from the regulation of viral late gene expression. The cysteine residues were not required for binding of the L4-22K

protein to the packaging sequences so the observed defect in mature virus particle production was not due to lack of DNA binding. We found that the double histidine mutant had a complex phenotype with defects observed in genome packaging and expression of specific mRNA transcripts (87).

A working model for Ad DNA packaging suggests that the IVa2 and L4-22K bind directly to packaging sequences and together they recruit the L1-52/55K and L1-IIIa proteins to promote encapsidation (Fig. 1-5C). L1-52/55K is a nuclear phosphoprotein that is present in empty capsids and assembly intermediates but is not found in mature virions, suggesting a scaffolding role for this protein. In the absence of L1-52/55K only empty capsids are formed (99), further strengthening the notion that this protein is critical for virus packaging. *In vivo*, IVa2 and L1-52/55K proteins bind to the packaging sequence, but each individual protein's interaction with the packaging sequence is independent of the other (94, 100). L1-52/55K also interacts with IVa2 via the N-terminal 173 aa of L1-52/55K (Fig 1-5C) (100). Additionally, the L1-IIIa protein likely interacts with the L1-52/55K protein (Fig. 1-5C) and associates with viral packaging sequences (101), indicating that the complex of proteins on the packaging sequence is more intricate than initially believed.

Outside of the conventional model of packaging that includes the four viral proteins described above, other cellular and viral factors have been implicated and extensively studied. Several cellular DNA binding proteins that bind packaging repeats were identified, but were not found to be relevant to the packaging process. These cellular proteins include COUP-TF, OCT-1 and P complex containing CCAAT displacement protein (CDP) (92, 102). The packaging domain overlaps with the transcriptional enhancer region of E1A, so these cellular proteins as well as Ad packaging proteins may be important for E1A transcription.

Additionally, L4-33K is suggested to play a role in virus assembly (103-105) and genome packaging (106). A virus deficient in L4-33K produces only empty capsids but the protein does not bind to packaging sequences or influences the interaction of other packaging sequences, so a better understanding of this protein's role in this aspect of the viral life cycle is needed.

Ad Alternative pre-mRNA splicing

A typical mammalian gene is composed of several relatively short exons that are interrupted by much longer introns. To generate the correct mature mRNAs, the exons must be properly identified and joined together in a coordinated and efficient process. Alternative splicing pathways generate different mRNAs encoding distinct protein products, thus expanding the coding capacity of genes. The pathway of alternative splicing may act as an on-off switch for gene expression by having multiple levels of regulatory control. Most human genes express more than one mRNA by alternative splicing. The splicing reaction is carried out by the spliceosome which consists of five small nuclear ribonucleoprotein complexes (snRNP) U1, U2, U4, U5, and U6 and a large number of non-snRNP proteins (reviewed in reference (107). The spliceosome functions through a series of RNA-RNA, protein-protein, and RNA-protein interactions to correctly remove the introns and join the exons. The minimum regulatory sequences for splicing include the conserved 5' splice site (5'ss), and a branch point sequence (BPS) followed by the polypyrimidine tract (Py) and the 3' splice site (3'ss). Briefly, assembly of the spliceosome occurs via formation of the E complex, initiated by recognition of the 5'ss by U1 snRNP. U2 auxiliary factor (U2AF)

contains two subunits, $U2AF_{65}$ and $U2AF_{35}$ which binds the Py and 3'ss, respectively. The critical step in 3'ss definition is the stable recruitment of U2 snRNP to the 3'ss . The U2 snRNP base pairs with the BPS to form the A complex. U4 and U6/U5 tri-snRNP bind to the pre-mRNA resulting in the B complex. Finally, the C complex (spliceosome) is formed by remodeling of RNA-RNA and RNA-protein interactions and is activated for catalysis. Additional factors including serine/arginine-rich (SR) proteins facilitate spliceosome assembly (108).

Although many sequences within a pre-mRNA transcript resemble the consensus splice sites, most of them are not used. The arrangement of positive and negative *cis*-acting sequence elements found both within introns and exons create a higher level of regulation to ensure that the correct splice sites are utilized in time and space. These positive and negative *cis*-acting elements referred to as enhancers and silencers, can be near (50-100 nt) or far (100s-1000s nt) from the splice sites (109). Splicing enhancers promote splicing at appropriate times and at correct splice sites, are often found within exons (exon splicing enhancers; ESEs) and are generally bound by SR proteins. ESEs and ISEs (intronic splicing enhancers) play stimulatory roles in both constitutive and regulated splicing. Exon and intron splicing silencers (ESSs and ISSs, respectively) are less well characterized but seem to interact with negative regulators that often belong to the family of heterogeneous nuclear ribonucleoproteins (hnRNPs) (109). Unlike constitutive exons, alternative exons are included or skipped to produce ratios of natural mRNA isoforms. Mutations that affect either the basal level of alternative exon recognition or the binding site of a specific RNA binding protein that regulates exon usage, will alter this natural ratio. In humans, mutations in *cis*-acting

elements cause mis-splicing of genes that alter gene function and contribute to disease pathology (110, 111).

The accumulation of Ad mRNAs is temporally regulated to ensure that proteins that are needed at certain stages of the virus life cycle are produced in a timely fashion. Nuclear splicing factors are dynamically relocalized during the course of infection, reflecting the changes in gene expression activity of the infected cells (112, 113). The Ad E4-ORF4 protein induces dephosphorylation of SR proteins, resulting in the inhibition of splicing of cellular pre-mRNAs with consensus-type splicing signals and a shift to splicing of Ad pre-mRNAs (114). The pre-mRNA produced from the MLTU, for example, is alternatively spliced so that the TPL is joined to one of many alternative 3' splice sites, a process that is tightly controlled by both viral and cellular factors as infection progresses (Fig. 1-4). The L1 unit represents an example of an alternatively spliced gene with a common 5' donor splice site that can be joined to one of two alternative 3' acceptor splice sites forming the L1-52/55K or IIIa mRNAs (Fig 1-4) (75). Alternative splicing of the L1 unit is regulated at two levels: A basal level, where the two distinct 3' ss differ in their sequence; a higher order of regulation results from the presence of a ISS element (3RE) and an ISE element (3VDE) within the intron of IIIa. Early after infection, the L1-52/55K acceptor site is preferentially used, while the IIIa acceptor splice site is not active until late in infection (75). The L4-33K protein is required for the early-to-late shift in L1 alternative splicing pattern and plays a positive role in regulating L1 splice site selection (75). L4-33K binds to the splicing enhancer 3VDE. The 3VDE contains the IIIa BPS and Py. The 3RE is the IIIa splicing repressor element that binds the SR family of cellular splicing factors (75, 115) and is located immediately upstream of the IIIa BPS. Although the inhibitory effect of the cellular SR proteins is partly responsible

for the lack of IIIa splicing early after infection, L4-33K plays a positive role in regulating L1 splice site selection (75). Of relevance, L4-33K interacts with, and is phosphorylated by, DNA-dependent protein kinase (DNA-PK) (116). DNA-PK is involved in cellular transcriptional regulation and DNA repair, and also plays a role in inhibiting the shift from L1-52/55K to IIIa alternative splicing. The role of L4-33K as an alternative RNA splicing factor has been confirmed in the context of viral infection, with primary targets identified as the genes coding for the IIIa and pVI proteins, with a lesser effect on the gene coding for the fiber protein (106).

In this study, we investigated the mechanism for L4-22K regulation of L4-33K pre-mRNA splicing. We determined that the L4 pre-mRNA sequence that is necessary for this regulation is the L4-33K enhancer element located at the 3'ss of the transcript. Constitutive expression was observed with no response to L4-22K, when this element and flanking sequences were replaced with corresponding sequences from the constitutively spliced SV40 small T antigen pre-mRNA. Additionally, we suspected that the enhancer element not sufficient to regulate L4-33K splicing. Detailed analysis of the 5' and 3' exons of L4-33K revealed the presence of regulatory sequences located in the 3' exon that may act as ESSs. This regulatory sequence may act via an L4-22K- dependent or L4-22K-independent mechanism, but at the very least, acts as an additional level of regulation to that provided by the L4-22K protein.

Chapter 2: Materials and Methods

Cells, Viruses, and Infections

N52.E6-Cre cells are derived from the E1-expressing cell line N52.E6 (117), a gift from G. Schiedner and S. Kochanek, University of Ulm, Ulm, Germany. Cre recombinase was not important for these experiments. N52.E6-Cre cells were maintained in Dulbecco's modified Eagle medium supplemented with 10% Fetalclone III serum (HyClone), penicillin, and streptomycin. A549 cells (ATCC) were maintained in Dulbecco's modified Eagle medium supplemented with 10% bovine calf serum (HyClone), penicillin, and streptomycin. TetC4-22K-ΔC (53) and TetC4-33K (53) cell lines express the Ad5 L4-22K and L4-33K proteins, respectively, under the control of a Tet-inducible promoter. The TetC4-22K-ΔC cell line expresses a truncated form of L4-22K lacking the C-terminal 23 aa that are not required for activity. TetC4 cells and the TetC4 cell line derivatives were maintained in Dulbecco's modified Eagle medium supplemented with 8% fetal bovine serum, 200µg/ml Geneticin, 200µg/ml hygromycin B, 100µg/ml blasticidin S and 0.1µg/ml doxycycline. Doxycycline was removed from the culture media 36-48 h prior to conducting experiments to induce expression of L4-22K and L4-33K.

The construction of viruses containing the L4-22K point mutations and the tagged L4-22K virus are described below. The construction of viruses L4-Δ22K and L4-22K$^-$ were described elsewhere (53). The L4-22K-C137S/C141S mutant virus was propagated in N52.E6-Cre cells and the L4-22K-H166Q/H170Q and the L4-Δ22K viruses were propagated in TetC4-22K-ΔC cells. Virus particles were purified by cesium chloride (CsCl) equilibrium centrifugation as described below. All infections were carried out in N52.E6-Cre, TetC4-

22K-ΔC, or TetC4-33K cells with a multiplicity of infection of 100-200 particles/cell of each virus stock, unless otherwise noted, as previously described (118).

Plasmids, Mutagenesis, and Cloning

Expression plasmids were generated using pcDNA3 (Invitrogen Life Technologies). Expression plasmids pcDNA3-22K and pcDNA3-33K were previously described (78). Plasmid pcDNA3-22K-C137S/C141S was constructed by introducing double serine substitution mutations into the L4-22K gene by site-directed mutagenesis (Quick-Change Mutagenesis; Stratagene). The pair of cysteines was mutated to serines and an *NheI* restriction site was introduced using a forward primer 5'-GCCATAGTTGCTAGCTTGCAAGACAGTGGGGGCAACATC-3' and reverse primer 5'GATGTTGCCCCCACTGTCTTGCAAGCTAGCAACTATGGC-3' corresponding to Ad5 nt 26591- 26619. The underlined nucleotides correspond to the mutation sites. Plasmid pcDNA3-22K-H166Q/H170Q was constructed in a similar manner by introducing double glutamine substitution mutations into the L4-22K gene. The pair of histidines were mutated to glutamines and a *ScaI* restriction site was introduced using a forward primer 5'-GTAACATCCTGCAGTACTACCGTCAACTCTACAGCCCAT-3' and reverse primer 5'-ATGGGCTGTAGAGTTGACGGTAGTACTGCAGGATGTTAC-3' corresponding to Ad5 nt 26679-26717. The underlined nucleotides correspond to the mutation site. Each mutated segment was then recombined into an infectious clone, pTG3602 (119) as described (118). Each mutated viral genome was excised from the pTG3602 plasmid by *PacI* digestion and used to transfect N52.E6-Cre cells (pTG3602-22K-C137S/C141S) or TetC4-22K-ΔC cells

(pTG3602-22K-H166Q/H170Q) using Lipofectamine 2000 reagent (Invitrogen) according to the instructions provided by the manufacturer. Viruses were isolated after 7 days by plaque purification and propagated in the cell lines specified above. Particles for each virus stock were purified by two successive rounds of CsCl equilibrium centrifugation as described below. Mutant virus sequences were confirmed by DNA sequencing and restriction enzyme digestion and Southern blot analysis for the presence of the indicated mutations. For simplicity, these mutant viruses will be referred to as C137S/C141S and H166Q/H170Q indicating their aa substitution mutations. The construction of Δ22K virus was described previously (53).

Plasmid pcDNA3+5'22K-STREP-HA was generated by polymerase chain reaction (PCR) cloning using a forward primer 5'-CGCGGATCCACCATG**TGGAGCCACCCGCAGTTCGAAAAA**AGC**TATCCTTATG ACGTGCCTGACTATGCC**GCACCCAAAAAGAAGCTG- 3' (where the bold nucleotides code for STREP-tag and HA-tag in 5' to 3' order) and a reverse primer for the 3' end of the L4-22K gene. Similarly, plasmid pcDNA3+3'22K-STREP-HA was generated by PCR cloning using a forward primer for the 5' end of the L4-22K gene, and a reverse primer having the complimentary sequence as above, with an *EcorI* site. The fragments were cloned into plasmid pcDNA3 to test for protein function. The tagged L4-22K sequences were excised by *BamHI* and *NheI* digestion and cloned into plasmid pAd-CMV (118) which contains Ad5 left-end sequences nt 1-450, the CMV promoter/enhancer, and Ad5 nt 3330-5780. Each tagged L4-22K segment was excised by *Bstz17* and *ScaI* digestion and recombined into the left end of pTG3602 infectious clone as described (118). Each tagged

viral genome was excised from the infectious clone by *PacI* digestion and used to transfect N52.E6-Cre cells and viruses were isolated as described above.

Expression plasmid pcDNA3-L4 was generated by PCR cloning of the regions from Ad5 nt 26195 to 27086 into pcDNA3. Note that this region contains the coding regions for both L4-22K (Ad5 nt 26195-26785) and L4-33K (26195-27086) (Fig. 3-1A). Plasmid pcDNA3-L4-C137S/C141S was constructed by introducing double serine substitution mutations to the L4-22K gene in the background of the pcDNA3-L4 plasmid. Plasmid pcDNA3-L4-H166Q/H170Q was constructed by introducing double glutamine substitution mutations to the L4-22K gene in the background of the pcDNA3-L4 plasmid. Expression plasmid pcDNA3-L4-22K^{-} was constructed by introducing a stop codon (TAG) in the coding region of L4-22K by changing Ad5 nt 26533-26536 from GTTA to CTAG, in the background of the pcDNA3-L4 plasmid. Similarly, plasmid pcDNA3-L4-33K^{-} was constructed by introducing a stop codon in the coding region of 33K by changing Ad5 nt 26814-26816 from ACA to TAG in the background of the pcDNA3-L4 plasmid. Plasmid pcDNA3-L4-22K^{-}-C137S/C141S was constructed by introducing double serine substitution mutations to the L4-22K gene in the background of the pcDNA3-L4-22K^{-} plasmid. Similarly, plasmid pcDNA3-L4-22K^{-} H166Q/H170Q was constructed by introducing double glutamine substitution mutations to the L4-22K gene in the background of the pcDNA3-L4-22K^{-} plasmid.

Plasmids expressing chimeras of SV40/L4-33K intronic sequences were constructed by PCR cloning of specific L4-33K or SV40 sequences into pcDNA3 as follows: Plasmid pcDNA3+33K-SV40-33K contains nucleotide sequences for the 5' and 3'exons of the L4-33K pre-mRNA (nts 26195-26508 and 26714-27086) with the entire intron (nts 26509-

26713) replaced with 346 nucleotides that correspond to the intron of the small T antigen encoded by SV40, including the last three nucleotides in the 5' exon, and the first nucleotide in the 3' exon, of SV40 as to not disturb potentially critical splicing sequences. Primers were designed to contain 5' to 3' and corresponding 3' to 5' sequences expanding nucleotides at the junctions of the chimeric sections. Intron hybrid 1 contains the 1st 53 nt of the L4-33K intron fused to the last 150 nt of the SV40 intron, while hybrid 2 contains the reverse configuration (the 1st 53 nt of SV40 fused to the last 150 nt of the L4-33K intron). Hybrid 3 contains the 1st 168 nt of the SV40 intron fused to the last 35 nt of the L4-33K intron (Fig. 4-2A).

Plasmids expressing chimeras of SV40/L4-33K exonic sequences were constructed by PCR cloning of specific L4-33K or SV40 sequences into pcDNA3 as follows, with no changes to the intronic sequences of L4-33K: primers were designed to contain 5' to 3' and corresponding 3' to 5' sequences expanding nucleotides at the junctions of chimeric sections. Plasmid pcDNA3+ Exon hybrid 1 contains the first 152 nt of the 5' exon replaced by the corresponding sequences of SV40 exon 1, while plasmid pcDNA3+ Exon hybrid 3 contains the opposite configuration in the 5' exon (the last 163 nt replaced by corresponding SV40 sequences). Plasmid pcDNA3 + Exon hybrid 2 contains the 2nd half, or last 188 nt, of the 3' exon replaced by the corresponding sequences of SV40 exon 2, while plasmid pcDNA3 + Exon hybrid 4 contains the opposite configuration (where the first 185 nt of the 3' exon are replaced by the corresponding sequences of SV40 exon 2). Plasmid pcDNA3 + Exon hybrid 5 contains the entire 5' exon replaced by corresponding sequences of the SV40 exon 1, while plasmid pcDNA3 + Exon hybrid 6 contains the entire 3' exon replaced by corresponding sequences of the SV40 exon 2 (Fig. 4-3A).

Replication Assay and Quantitative PCR

Infected N52.E6-Cre cells were harvested at 2 and 24 hours post-infection (hpi). Cells were washed in phosphate-buffered saline (PBS), pelleted, resuspended in cold isotonic buffer containing NP40 (140mM NaCl, 10mM Tris [pH 7.4], 1.5mM $MgCl_2$, 0.6% NP40), and incubated on ice for 10 min. Following another round of centrifugation at 2000 x *g*, the pellet containing nuclei was resuspended in PBS and DNA was prepared using a DNeasy blood and tissue kit (Qiagen) following the manufacturer's protocol. The amount of viral DNA was quantified with a DyNAmo HS SYBR Green quantitative PCR (qPCR) kit (Finnzymes). Primer pairs surrounding the packaging domain were used to quantify viral DNA, and appropriate primers were used to quantify glyceraldehyde 3-phosphate dehydrogenase (GAPDH) copy number, as described previously (53). The genome replication was calculated as the viral genome copy number normalized by the GAPDH copy number. The value calculated for each virus at each time point was divided by that of Ad5-wild-type (WT) at 2 hpi (input) and shown as the relative viral genome copy number.

Plaque Assays and Immunofluorescence Analysis

CsCl-purified virus particles (described below) from Ad5-WT, C137S/C141S, H166Q/H170Q, and Δ22K were used to infect A549 and TetC4- 22K-ΔC cells. After 7-10

days the number of plaques were counted, and the ratio of particles (determined as described below) to plaques was calculated as the particle/plaque forming unit (PFU). Infectivity was directly measured by a fluorescence focus assay by infecting A549 cells grown on cover slips in 24-well multi-well trays with different dilutions of particles from Ad5-WT, C137S/C141S, and H166Q/H170Q. Eighteen hpi, cells were fixed with methanol and stained using an antibody against DBP (provided by Arnold Levine, Institute for Advanced Study, Princeton, NJ) and a secondary goat anti-mouse antibody conjugated with fluorescein isothiocyanate (FITC) (Invitrogen). For each coverslip, the average number of DBP-positive cells in 10 random fields was determined and expressed as fluorescent focus units (FFU) per field. The particle/FFU ratio was calculated as the number of particles used to infect cells in the well divided by the total number of FFU in the well.

To determine if the H166Q/H170Q mutant virus inhibits Ad5-WT growth, N52.E5-Cre cells were infected with 200 particles/ cell each of Ad5-WT alone, H166Q/H170Q alone, or equal amounts of Ad5-WT and H166Q/H17OQ together. Forty-eight hpi, cell lysates were prepared and serial dilutions of the lysates were used to infect A549 cells grown on cover slips. Infectivity was directly measured by a fluorescence focus assay as described above. The FFU per well was determined and plotted versus virus particles used to infect the cells. The number of particles was calculated based on the known ratio for WT lystate: 10^9 PFU/ml = $2x10^{10}$ particles/ml (118), assuming that the particle: PFU ratio for the WT and mutant viruses is the same.

Virus Growth and Cesium Chloride Equilibrium Gradients

For virus growth curves, N52.E6-Cre cells were infected with Ad5-WT, C137S/C141S and H166Q/H170Q and harvested at 6, 12, 20, 28, and 36 hpi and lysed by four cycles of freeze-thawing. A specific volume of each lysate was used to infect A549 cells grown on coverslips and processed for immunofluorescence, as described above. Virus growth was calculated as a log of the fluorescent focus units per volume of lysate used (log FFU/ml). To compare virus growth on three different cell lines, N52.E6-Cre, TetC4-22K-ΔC, and TetC4-33K cells were infected with Ad5-WT, C137S/C141S, H166Q/H170Q or mock infected. For TetC4-22K-ΔC, and TetC4-33K cells, two sets of mock infected cells were prepared, one with doxycycline removed from the media to induce expression of 22K and 33K, and one with doxycycline maintained in the media. Cells were harvested 48hpi and a portion was used for whole cell extract preparation for western blot analysis (described below) and the rest was subjected to four cycles of freeze thawing. Virus lysates were used to infect A549 cells grown on coverslips and processed for immunofluorescence as described above. Virus yield was measured as FFU/ml.

For CsCl equilibrium gradients, cells were infected with Ad5-WT, C137S/C141S, or H166Q/H170Q. At 72 to 96 hpi, cells and media were collected, the cells were pelleted, resuspended in 6 ml of Tagmentation DNA buffer (TD) buffer (25mM [pH 7.4], 137mM NaCl, 5mM KCL, 0.7mM Na_2HPO_4), and frozen and thawed for four cycles. The lysates were clarified by centrifugation, added to step gradients of 1.25g/cm^3 and 1.40g/cm^3 CsCl in TD buffer, and centrifuged at 175,587 x *g* at 15 ℃ for 1 h. Visible bands at each step were collected and mixed with 1.34g/cm^3 CsCl and spun for 20 h at 174,618 x *g* at 15℃. The visible band was collected from the equilibrium gradient and the absorbance was measured at

a wavelength of 260 nm and converted to densities using the value 1 OD_{Abs260} = 1x 10^{12} particles/ml (28).

Silver Staining of Proteins in Virus Particles

CsCl was removed from samples by precipitation and prepared for gel electrophoresis. Two volumes of water and six volumes of ethanol were added to samples to precipitate virions overnight at -20℃ and then subjected to centrifugation at 16,000 x *g* at 4℃ for 30 min. The samples were washed once with 75% ethanol and resuspended in 1x sample buffer (60mM Tris pH 6.8, 20% glycerol, and 2% SDS). Protein concentrations were determined by a bicinchoninic protein assay kit (BCA) and equivalent amounts of protein were loaded based on hexon content, approximately 5x10^9 particles per sample. The samples were separated on 12.5% SDS polyacrylamide gel electrophoresis (SDS-PAGE) and silver stained as described previously (97, 120).

Gel Mobility Shift Assays

Nuclear extracts from Ad5-WT, C137S/C141S, H166Q/H170Q, and Δ22K infected, or mock infected, N52.E6-cre cells were prepared 24 hpi as described previously (78, 92, 121). Six µg of nuclear extract was incubated with 0.5-1.25 µg of poly(dI-dC), with or without anti-22K antibody, for 15 min at room temperature. Following this incubation, 120,000-

150,000 cpm of ^{32}P-labeled DNA probe (~12-15 fmol) was added to each reaction and incubated for an additional 30 min at room temperature. Packaging sequence probes used in the assays contained A repeats 1 and 2, corresponding to Ad5 nt 236 to 282. Samples were analyzed on nondenaturing polyacrylamide gels as described previously (92).

Western Blot Analyses

N52.E6-Cre, TetC4-22K-ΔC, or TetC4-33K cells were mock- or virus- infected and whole cell extracts were prepared as described previously (53). Protein concentrations were determined by a BCA protein kit and equal amounts of protein were separated by SDS-PAGE and transferred to nitrocellulose (Whatman) membranes. The membrane was blocked for 1 h at room temperature in a solution of 1% nonfat dry milk/Tris-buffered saline (TBS) followed by an incubation with a designated antibody in 1% nonfat dry milk/TBS with 1 mM sodium azide for 1 h at room temperature, or overnight at 4°C. The membrane was washed four times for 5 min each wash in TBS with 0.5% Tween 20 and incubated with 800 IRdye800-conjugated goat anti-rabbit immunoglobulin G (Rockland Immunochemicals, Gilbertsville, PA) for 45 min followed by four more washes in TBS with 0.5% Tween 20 for 5 min each. Two final washes in TBS were done and the membrane was scanned using an Odyssey system (LiCOR, Lincoln, NE). For L4-22K detection, the SDS-PAGE gel was transferred to polyvinylidene difluoride (PVDF) (GE Healthcare) membrane, blocked with 3% Bovine serum albumin (BSA)-TBS for 1 h at room temperature, and probed with primary antibody overnight at 4°C. All washes and secondary antibody incubation were done with

5% nonfat milk-TBS containing 0.05% Tween 20. Proteins were detected using the following antibodies: rabbit polyclonal antibodies to fiber and penton (gifts from Carl Anderson, Brookhaven National Laboratory); rabbit polyclonal antibody to VII (Daniel Engel, University of Virginia); rabbit polyclonal antibody to hexon; (GeneTex, Inc., Irvine, CA); rabbit polyclonal antibodies to L4-22K and L4-33K (78); rabbit polyclonal antibody to VIII (William Wold, St. Louis University); rabbit polyclonal antibodies to IVa2 and L1-52/55K (92); rabbit polyclonal antibody to IIIa (101); rabbit polyclonal antibody to V (David Matthews, University of Bristol); rabbit polyclonal antibody to L4-100K (106); rabbit polyclonal antibody to VI (Christopher Wiethoff, Loyola University Chicago); rabbit polyclonal antibody to E1A (Santa Cruz Biotechnology); rabbit polyclonal antibody to γ-tubulin (Sigma); anti-HA rabbit polyclonal antibody (Rockland); and rabbit polyclonal antibody to Strep tag (GeneTex).

For western blot analyses of particles from CsCl centrifugation gradients, particles were prepared as described above, and equivalent amounts of protein samples were loaded based on hexon content, equivalent to approximately $5x10^9$ particles. Samples were separated on a 12.5% SDS gel, transferred to nitrocellulose overnight at 4℃ and western blots were performed as described above.

Northern Blots and Reverse Transcriptase PCR (RT-PCR)

Infected N52.E6-Cre cells or A549 cells were harvested at 12h, 24 h or 48 hpi and the cytoplasmic fraction was extracted using an RNeasy isolation kit for cytoplasmic RNA from

animal cells (Qiagen). Four to 12 µg of total cytoplasmic RNA were separated on a 1% agarose gel containing 2.2 M formaldehyde and transferred to a nylon membrane (GE Healthcare). The filter was then stained in a solution of 0.02% methylene blue and 0.3M sodium acetate, pH 5.5 to detect ribosomal RNAs. The stain was removed in 1% SDS and the filter hybridized with 0.5-1µg of a designated DNA probe ^{32}P-labeled by random priming (Random primer DNA labeling kit, Takara Bio) followed by a series of washes, as previously described (122). DNA probes were prepared by PCR using primers surrounding the co-terminal 3' end of each late gene family: Ad5 L1 nt 13025-13751, Ad5 L2 nt 16834-17452, Ad5 L3 nt 21573-22322, Ad5 L4 nt 26769-27590 (or the entire Ad5 L4 region, nt 26195-27086 when specified), and Ad5 L5 nt 31920-32465. For E1A blot, a DNA probe was prepared using primers surrounding nt 147-1334. The fragments were extended by PCR and resolved through 1% agarose followed by gel extraction (Qiagen).

Two and a half µg of RNA were mixed with 50ug/ul of oligo(dT) to generate cDNA in a 20 µl reaction using the Superscript II reverse transcriptase kit (Invitrogen). Two units of *E. coli* RNase H were added to remove RNA complimentary to the cDNA, and the samples were incubated at 37℃ for 20 min. A 1 µl portion of the RT reaction was amplified by PCR in a total volume of 12 µl for 20 cycles. Each reaction included a forward primer complementary to Ad5 tripartite leader 3 (nt 9700-9719) and a reverse primer within the intron of L4-33K (nt 26682-26703). PCR products were resolved on a 1% agarose gel and the fragments extracted for sequencing (gel extraction kit, Qiagen). Equal portions of each cDNA were used in a PCR reaction with primers GAPDH-1 (5'-ACCCAGAAGACTGTGGATGG-3') and GAPDH-2 (5'-TTCTAGACGGCAGGTCAGGT-3') to compare levels of GAPDH.

In Vivo Analysis of Alternative Splicing

N52.E6 cells seeded in 100-mm dishes were transfected with specific expression plasmids using Lipofectamine 2000 reagent (Invitrogen). Twenty µg of each expression plasmid was co-transfected with 2 µg of pEGFP-C1 expression plasmid (Clonetech) and cells were harvested 48h after transfection. A portion of cells from each transfected dish was used for whole cell extract preparation and western blot analysis while the rest of the cells were used for total cytoplasmic RNA preparation. For western blot analysis, protein levels were normalized to levels of GFP based on quantification using an Odyssey infrared imaging system (Li-COR) to account for any differences in transfection efficiencies. Proteins were separated by SDS-PAGE, transferred to a nitrocellulose membrane, and analyzed for protein levels as described above. For primary antibodies, rabbit anti-L4-22K and L4-33K antibodies (78) and rabbit anti-GFP antibodies (Abcam) were used. For mRNA analysis, cytoplasmic RNA was prepared as described above. Messenger RNA levels were normalized to levels of GFP based on quantification using PhosphorImage analysis to account for any differences in transfection efficiencies. Samples were used for Northern blotting, as described above, with a designated DNA probe ^{32}P-labeled by random priming. The DNA probe used to detect levels of L4-22K and L4-33K was prepared by PCRB amplification of Ad5 nt 26769-27590, separated on a 1% agarose gel and extracted (gel extraction kit, Qiagen). The DNA probe used to detect levels of GFP was prepared by digesting plasmid pEGFP-C1 with *NheI* and *BamHI*, followed by gel purification of the digested fragment (Qiagen).

Protein Affinity Chromatography and Analysis

A total of $1.74x10^8$ cells were infected with a virus that contains L4-22K-tagged with tandem affinity tags (Strep- and HA-tags), described above, or mock infected. Cells were harvested 20hpi, washed twice with PBS, and pelleted by centrifugation. The cell pellets was lysed in 5x cell pellet volumes of F lysis buffer (10mM TRIS, pH 7.5, 50mM NaCl, 10% glycerol, 0.5% triton X-100, 5µM ZnCl2) and a cocktail of protease and phosphatase inhibitors (apoptonin, leupeptin, pepstatin, benzamidine, PMSF, NaF, and Na-Vanadate) at 1mM concentration each, followed by sonication. The samples were cleared by centrifugation at maximum speed for 30 min and 5% of the starting lysate was saved for Western blot analysis. The lysates were precleared with 50ul of protein A-agarose beads (Roche) for 1 h and incubated with 50ul of Strep Tactin-conjugated beads (Qiagen) overnight. The beads were washed five times with F lysis buffer and proteins were eluted by 2 x 30 min incubations with elution buffer (10mM TRIS, pH 7.5, 50mM NaCl, 2.5mM desthiobition (Sigma)). A portion of the eluate was stored for western blot and silver stain analyses. The remainder of the samples were diluted in 1ml of F lysis buffer and incubated with 25ul of anti-HA-agarose beads (Sigma) overnight. The beads were washed five times with F lysis buffer and protein complexes were eluted by boiling the beads in 1X sample buffer (60mM Tris, pH 6.8, 2% SDS, and 20% glycerol). Samples were analyzed by western blotting and silver stain analyses. For silver stain analysis, equal amounts of pull-down proteins were separated by SDS-PAGE and the gel was fixed for 1 h, or overnight, in a

solution of 40% methanol and 10% acetic acid. The gel was washed three times for 20 min each wash in 30% ethanol followed by incubation for 1 min in a reducing solution of .02% sodium thiosulfate. The gel was washed 3 x with water for 30 sec each, followed by a 20 min incubation in a solution of silver nitrate (0.2% silver nitrate, 0.02% [v/v] formaldehyde. After washing 3x with water, for 30 sec each, the gel was developed in a solution of 3% sodium carbonate, 0.0005% sodium thiosulfate, and 0.05% (v/v) formaldehyde) until bands were visible. The gel was washed 3x with water, incubated in a stop solution of 5% acetic acid, and stored at 4℃.

Chapter 3: The Adenovirus L4-22K Protein Has Distinct Functions in the Posttranscriptional Regulation of Viral Gene Expression and Encapsidation of the Viral Genome

Abstract

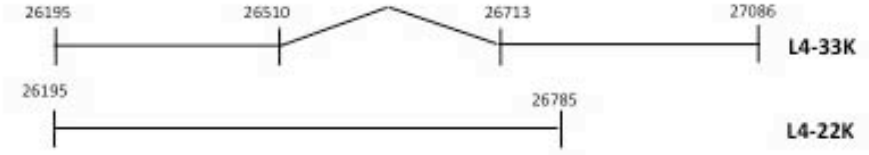
The adenovirus L4-22K protein is multifunctional and critical for different aspects of viral infection. Packaging of the viral genome into an empty capsid absolutely requires the L4-22K protein to bind to packaging sequences in cooperation with other viral proteins. Additionally, the L4-22K protein is important for the temporal switch from the early-to-late phase of infection by regulating both early and late gene expression. To better understand the molecular mechanisms of these key functions of the L4-22K protein, we focused our studies on the role of conserved pairs of cysteine and histidine residues in the C-terminal region of L4-22K. We show that mutation of the cysteine residues affected the production of infectious progeny virus but did not interfere with the ability of the L4-22K protein to regulate viral gene expression. These results demonstrate that these two functions of L4-22K may be uncoupled. Mutation of the histidine residues resulted in a mutant with a similar phenotype as a virus deficient in the L4-22K protein, where both viral genome packaging and viral gene expression patterns were disrupted. Interestingly, both mutant L4-22K proteins bound to adenovirus packaging sequences indicating that the paired cysteine and histidine residues do not function as a zinc finger DNA binding motif. Our results reveal that the L4-22K protein controls viral gene expression at the post-transcriptional level and regulates the accumulation of the L4-33K protein, another critical viral regulator, at the level of alternative pre-mRNA splicing.

Results

Generation and characterization of L4-22K mutant viruses: viral genome replication and infectivity.

The Ad5 L4-22K and L4-33K proteins are encoded by the L4 region and share a common N-terminus (Fig. 3-1A). The L4-33K protein is the product of an additional splicing event, while the L4-22K protein is translated from the unspliced version of this transcript with each protein having a unique C-terminus. Using ClustalW multiple sequence alignment (123), we found that the unique C-terminus of L4-22K, aa 116 to 175, is highly conserved amongst human Ad serotypes (Ads 4, 5, 7, 9, and 12) and many non-human primate Ads (Ads 21, 25, Y25, and 33) (Fig. 3-1B). The C-terminal 23 amino acids of Ad5 L4-22K is not conserved and is not essential for its functions (53). The conservation in the C-terminal region of L4-22K suggests a crucial and unique function of this domain. Within this region, there are two conserved cysteine residues (C137, C141) and two conserved histidine residues (H166, H170) that are clustered within 30 aa and have the general spacing requirements consistent with features of a zinc finger motif (31, 32). We postulated that these four conserved residues constituted a putative DNA binding motif involved in the DNA-protein interaction of L4-22K with the packaging domain. In order to assess the role of these conserved residues for L4-22K activity, two mutant L4-22K viruses were generated and characterized in this study: C137S/C141S, with conservative, double serine substitution mutations at aa 137 and 141, and H166Q/H170Q, with conservative, double glutamine substitution mutations at aa 166 and 170. The C137S/C141S and H166Q/H170Q mutant

A.



B.

```
Ad5-22k          KSKQPPPLAQEQQQRQGYRSWRGHKNAIVACLQDCGGNISFARRFLLYHHGVAFPRNILH 166
Ad4-22k          PTT-----QTGKKERQGYKSWRGHKNAIVSCLQACGGNISFTRRYLLFHRGVNFPRNVLH 153
Ad7-22k          PTTAS---KTGKKERQGYKSWRGHKNAIISCLHECGGNISFTRRYLLFHRGVNFPRNILH 166
Ad9-22k          --------GDGKHERQGYRSWRAHKAAIIACLQDCGGNIAFARRYLLFHRGVNIPRNVLH 130
Ad12-22k         HLGK-----NRKQARRGYCSWRAHQSNIVACFQHCGGNISFARRYLLYHDGVAIPRNVLH 158
Ad21-22k         ATTAS---KTGK-ERQGYKSWRGHKNAIISCLHECGGNISFTRRYLLFHRGVNFPRNVLH 165
Ad25-22k         PTT-----QTGKKERQGYKSWRGHKNAIVSCLQACGGNISFTRRYLLFHRGVNFPRNILH 161
AdY25-22k        PTT-----QTGKKERQGYKSWRGHKNAIVSCLQACGGNISFTRRYLLFHRGVNFPRNILH 159
Ad33-22k         PTTAS---KTGKKERQGYKSWRGHKNAIISCLHECGGNISFTRRYLLFHRGVNFPRNILH 168
                       :  *:** ***.*:  *::*:: *****:*:**:**:* ** :***:**

Ad5-22k          YYRHLYSPYCTGGSGSG---SNSSGHTEAKATG- 196
Ad4-22k          YYRHLHSPYYFEED----------------KTSS- 171
Ad7-22k          YYRHLHSPYCSQQVFATPTEKDSSDNGDQKTSS- 199
Ad9-22k          YYRHLHS---------------------------- 137
Ad12-22k         YYRHLYSPFEELDKEPT---CNSQAAH------- 182
Ad21-22k         YYRHLHSPYYSPQVSAE--KDNSSKDLQQKTSSS 197
Ad25-22k         YYRHLHSPYYFQEEAAA------AEKDQQKTSS- 188
AdY25-22k        YYRHLHSPYYFQEEAAA------AEKDQ-KTS-- 184
Ad33-22k         YYRHLHSPYCSQQAPATSAEKDSSGNGDQKTSS- 201
                 *****:*
```

C.

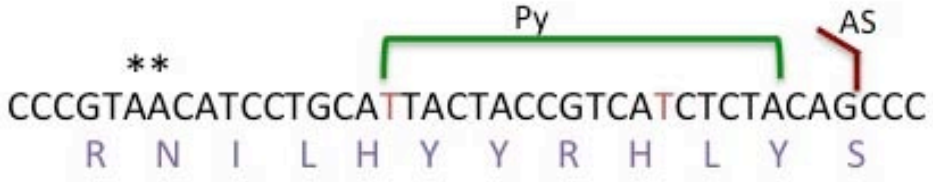


Figure 3-1 The Ad5 L4-22K protein.

Panel A. Schematic diagram of the coding sequences of the L4-22K and L4-33K proteins. Numbers correspond to Ad5 nucleotides. The intron of L4-33K is indicated by the gap and angled lines.

Panel B. ClustalW protein sequence alignment of the L4-22K carboxy-termini of L4-22K of human (Ad5, Ad4, Ad7, Ad9, Ad12), simian (Ad21, Ad25, Ad33), and chimpanzee (AdY25) Ad serotypes. The consensus symbols are as follows: the asterisk indicates positions which have a single, fully conserved residue, while the colon and period indicate conservation between groups of strongly and weakly similar properties, respectively. The cysteine and histidine residues that were mutated in this study are indicated with arrows and correspond to residues C137, C141, H166, and H170. Amino acid numbers are indicated to the right of the alignment.

Panel C. The nucleotide sequence for the L4-22K protein with the corresponding aa sequence is shown for aa 162-173. The nucleotides that were changed to generate the histidine substitutions are shown in red. This sequence is shared with the intronic 3' splice site of the L4-33K pre-mRNA as depicted by the branch point sequence (asterisks), pyrimidine tract (Py), and acceptor site (AS). Note that the histidine residues of the L4-22K protein sit in the pyrimidine tract of the L4-33K pre-mRNA.

viruses were propagated in N52.E6-Cre and TetC4-22K-ΔC cell lines, respectively. The C137S/C141S mutant virus grew poorly in N52.E6-Cre cells, but these cells were used to eliminate the possibility of recombination with the endogenous L4-22K gene in the TetC4-22K-ΔC cell line to generate a revertant virus. The H166Q/H170Q mutant virus grew poorly in both N52.E6-Cre and TetC4-22K-ΔC cell lines (discussed below), but enough purified virus was produced by the complementing cell line to conduct experiments. The TetC4-22K-ΔC cell line expresses an L4-22K protein with a truncation of the C-terminal 23 amino acids that are not required for activity (53) greatly reducing the possibility of homologous recombination to generate a revertant virus with the H166Q/H170Q mutant. Mutant virus sequences were confirmed by DNA sequencing and Southern blot analysis for the presence of the indicated mutations; these mutant virus stocks were pure from any Ad5-WT contamination (data not shown).

We evaluated viral DNA replication with the C137S/C141S and H166Q/H17 mutant viruses by qPCR and found that both mutant viruses replicated to similar levels as Ad5-WT (Fig. 3-2A). To test the infectivity of the mutant viruses, virus particles were used in a fluorescent focus assay in A549 cells (Fig. 3-2B). The particle/FFU ratio for each virus was calculated and the results showed similar infectivity between Ad5-WT and the C137S/C141S mutant virus, while a marginal 3-fold decrease in infectivity was observed with the H166Q/H170Q mutant. Infectivity was also examined using a plaque assay in A549 cells (Fig. 3-2C) and the L4-22K-ΔC complementing cell line (Fig. 3-2D) and the particle/PFU ratio for each virus was calculated. Consistent with the DNA replication and infectivity results, there was only a modest difference in the particle/PFU ratio between Ad5-WT and the C137S/C141S mutant virus in A549 or TetC4-22K-ΔC cells. In contrast, no plaques were

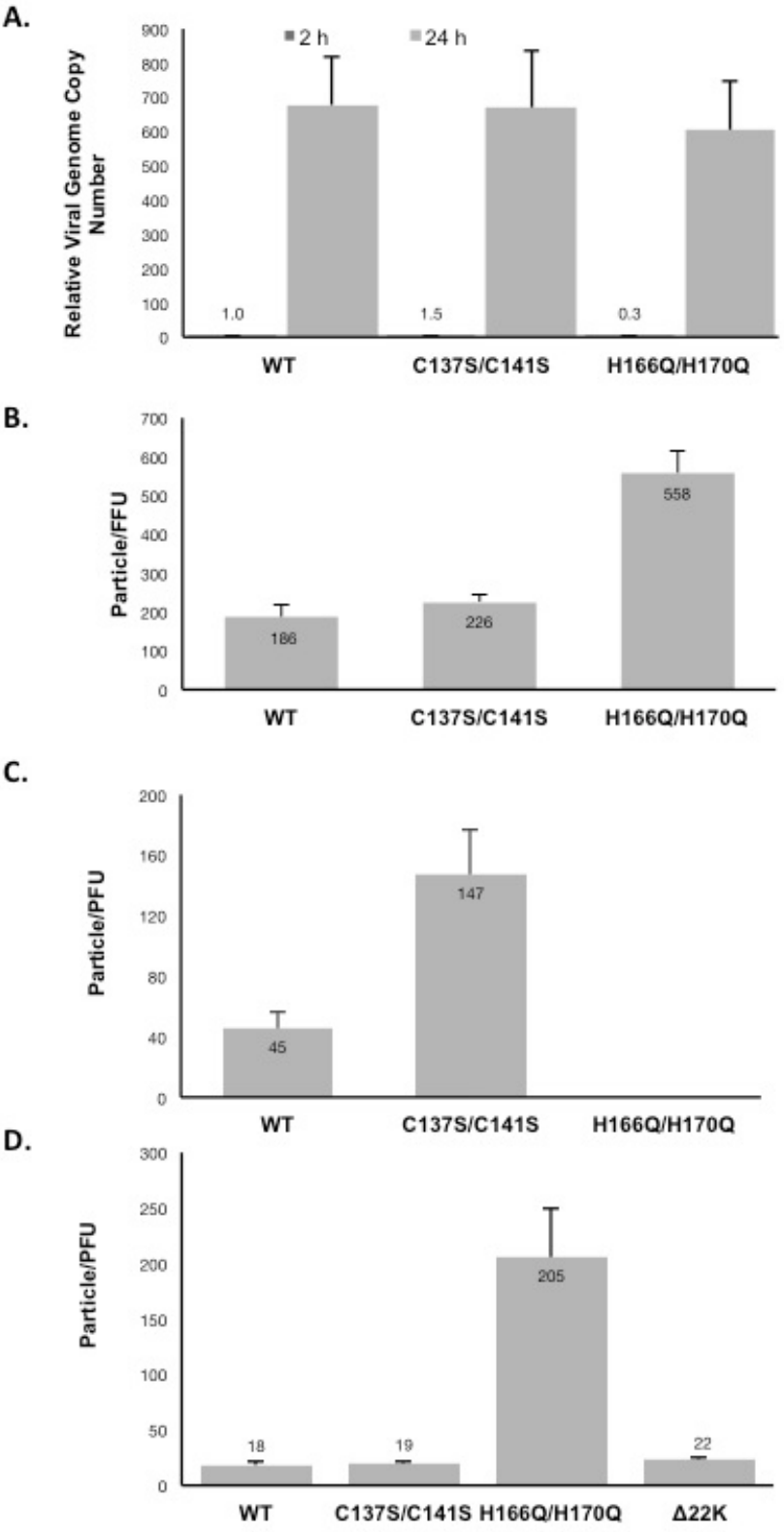

A.
Relative Viral Genome Copy Number
2 h
24 h
900
800
700
600
500
400
300
200
100
0
1.0
1.5
0.3
WT
C137S/C141S
H166Q/H170Q
B.
Particle/FFU
700
600
500
400
300
200
100
0
186
226
558
WT
C137S/C141S
H166Q/H170Q
C.
Particle/PFU
200
160
120
80
40
0
45
147
WT
C137S/C141S
H166Q/H170Q
D.
Particle/PFU
300
250
200
150
100
50
0
18
19
205
22
WT
C137S/C141S
H166Q/H170Q
Δ22K

Figure 3-2 Properties of L4-22K mutant viruses.

Panel A. Viral genome replication in Ad5-WT-, C137S/C141S-, and H166Q/H170Q- infected N52.E6-Cre cells. Values were normalized to GAPDH and compared to that of Ad5-WT at 2 hpi.

Panel B. The particle/FFU ratios of Ad5-WT, and C137S/C141S and H166Q/H170Q mutant viruses were determined in A549 cells.

Panel C. The particle/PFU ratios of Ad5-WT, and C137S/C141S and H166Q/H170Q mutant viruses were determined in A549 cells.

Panel D. The particle/PFU ratios of Ad5-WT, and C137S/C141S and H166Q/H170Q mutant viruses, were determined in TetC4-22K-ΔC cells

observed with the H166Q/H170Q mutant virus in A549 cells, even though DNA replication and infectivity were not significantly affected. To determine if providing WT L4-22K *in trans* could rescue plaque formation with the H166Q/H170Q mutant, plaque assays were performed in TetC4-22K-ΔC cells (Fig. 3-2D). We observed that the particle/PFU ratio with the H166Q/H170Q mutant virus was increased ~10-fold compared to Ad5-WT; the complementing cell line was able to efficiently rescue plaque formation with an L4-22K null mutant consistent with previous results (53). We conclude that the C137S/C141S and H166Q/H17 mutant viruses had levels of viral genome replication and infectivity comparable to Ad5-WT, while plaque formation was severely affected with the H166Q/H170Q mutant virus. The L4-22K protein *in trans* only partially rescued plaque formation with the H166Q/H170Q mutant.

The conserved cysteine residues in the C-terminal region of L4-22K are important for viral genome packaging.

To analyze the effect of the L4-22K point mutations on overall viral growth properties, single step growth curves were performed. The C137S/C141S and H166Q/H170Q mutants were reduced ~25-40-fold in the production of infectious progeny in N52.E6-Cre cells compared to Ad5-WT (Fig. 3-3A). Since L4-22K was shown to be required for genome packaging *in vivo* (53, 78) and L4-33K regulates late gene expression on the post-transcriptional level (75, 90), we wanted to examine the contribution of these two proteins to the observed growth defect phenotype with the mutant viruses. In order to examine if the defect in virus growth exhibited by the two L4-22K mutants could be restored by

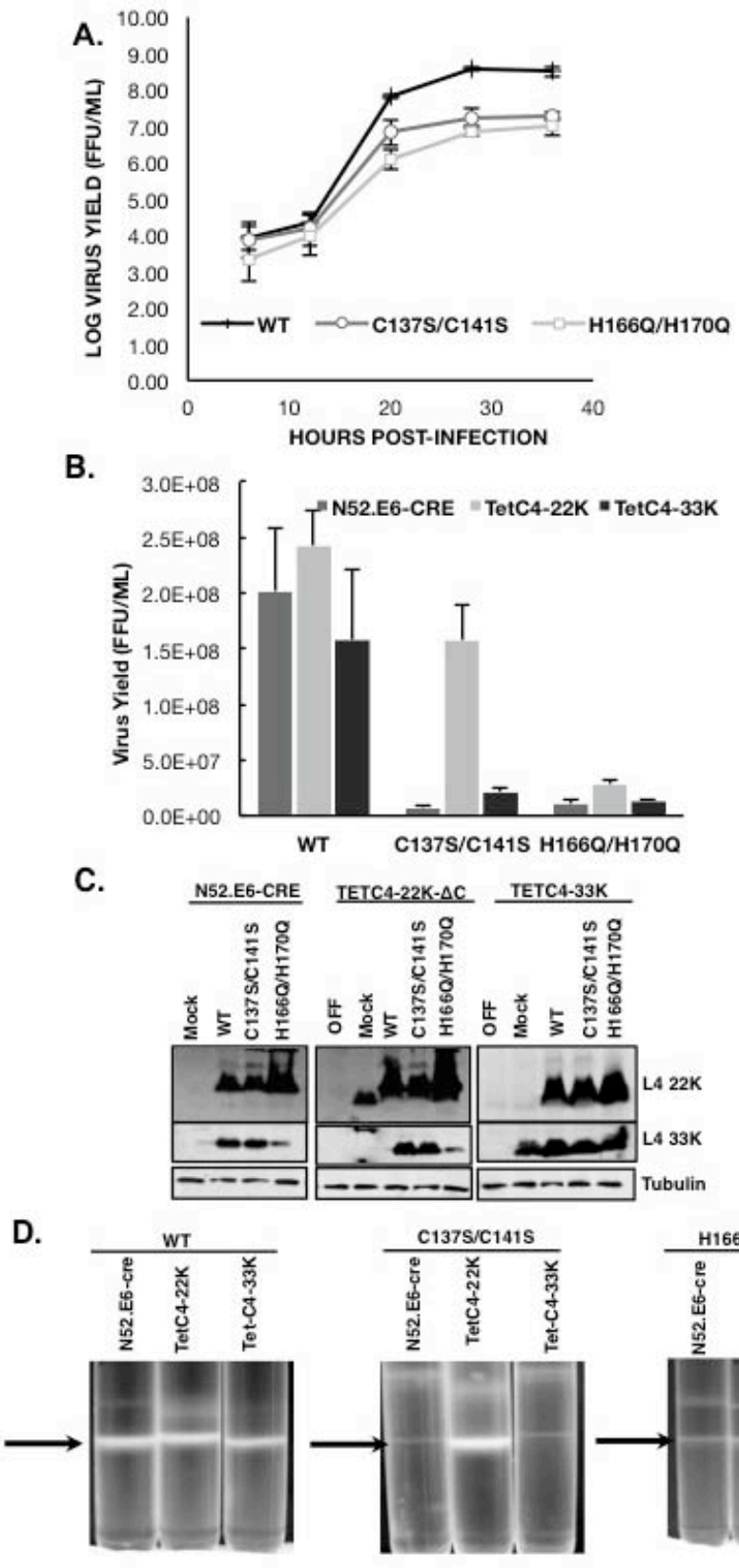

A.
LOG VIRUS YIELD (FFU/ML)
10.00
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
0
10
20
30
40
HOURS POST-INFECTION
WT
C137S/C141S
H166Q/H170Q
B.
Virus Yield (FFU/ML)
3.0E+08
2.5E+08
2.0E+08
1.5E+08
1.0E+08
5.0E+07
0.0E+00
N52.E6-CRE
TetC4-22K
TetC4-33K
WT
C137S/C141S
H166Q/H170Q
C.
N52.E6-CRE
TETC4-22K-ΔC
TETC4-33K
Mock
WT
C137S/C141S
H166Q/H170Q
OFF
Mock
WT
C137S/C141S
H166Q/H170Q
OFF
Mock
WT
C137S/C141S
H166Q/H170Q
L4 22K
L4 33K
Tubulin
D.
WT
C137S/C141S
H166Q/H170Q
N52.E6-cre
TetC4-22K
Tet-C4-33K
N52.E6-cre
TetC4-22K
Tet-C4-33K
N52.E6-cre
TetC4-22K
Tet-C4-33K

Figure 3-3 Viral growth kinetics of L4-22K mutant viruses.

Panel A. N52.E6-Cre cells were infected with Ad5-WT, or C137S/C141S and H166Q/H170Q mutant viruses, harvested at the designated time points after infection, and cell lysates were used to infect A549 cells to measure infectious virus yield (FFU/ml) using a fluorescent focus assay.

Panel B. N52.E6-Cre, TetC4-22K-ΔC, or TetC4-33K cells were infected with Ad5-WT, or C137S/C141S and H166Q/H170Q mutant viruses, harvested at 48 hpi, and cell lysates used to infect A549 cells to measure infectious virus yield (FFU/ml) using a fluorescent focus assay.

Panel C. A portion of the cell lysate from panel A was used to prepare whole-cell extracts for Western blot analysis of L4-22K and L4-33K protein levels. Tubulin is shown as a loading control.

Panel D. CsCl equilibrium density gradient profiles of virus particles produced from Ad5-WT-, C137S/C141S-, or H166Q/H170Q- infected N52-E6.Cre, TetC4-22K-ΔC, and Tet-C4-33K cells. Arrows indicate mature virus particles.

complementation with WT-L4-22K protein, the mutant viruses were grown in TetC4-22K-ΔC cells and compared to growth on N52.E6-Cre cells. Overall virus yield was analyzed using a fluorescent focus assay and calculated as total FFU/ml (Fig. 3-3B). Consistent with the single-step growth curves, we observed a 30-fold reduction in overall virus yield with the C137S/C141S mutant virus in N52.E6-Cre cells compared to Ad5-WT, a phenotype that was almost fully rescued by providing L4-22K *in trans* using the L4-22K complementing cell line. With the H166Q/H170Q mutant, virus yield was modestly increased in TetC4-22K-ΔC cells. Analysis of protein expression showed that with the C137S/C141S mutant virus, both L4-22K and L4-33K proteins were expressed to WT levels in N52.E6-Cre and TetC4-22K-ΔC cells and, therefore, decreased production of these proteins did not contribute to the decrease in infectious virus yield with this mutant (Fig. 3-3C, left and middle panels). With the H166Q/H170Q mutant virus, L4-22K protein expression was significantly increased in both N52-E6.Cre and TetC4-22K-ΔC cells, while expression of L4-33K was decreased in both cell lines. Note that the L4-22K-ΔC protein expressed in the complementing cell line runs slightly faster than WT L4-22K protein due to the C-terminal truncation. We recently showed that mutation of Ad5 L4-33K impaired the viral genome packaging process, but not virion assembly, resulting in the production of only empty capsids (106). To determine if the observed decrease in L4-33K protein levels contributed to the H166Q/H170Q mutant phenotype, we infected TetC4-33K cells with Ad5-WT, C137S/C141S, or H166Q/H170Q viruses, and analyzed virus yield (Fig. 3-3B) and L4-22K and L4-33K protein expression (Fig. 3-3C, right panel). Supplementation with WT-L4-33K *in trans* modestly increased C137S/C141S mutant virus yield. This result was not surprising since there was no indication that L4-33K protein had any contribution to the phenotype of this mutant virus. Interestingly,

although growth in the TetC4-33K cells supplemented the levels of L4-33K protein with H166Q/H170Q infected cells (Fig. 3-3C, right panel), we did not see any associated increase in virus yield with this mutant (Fig. 3-3B). The possibility that overproduction of L4-22K protein with the H166Q/H170Q mutant had any effect on production of infectious progeny was also ruled out since providing L4-33K protein during infection reduced the levels of L4-22K protein expression to that seen with Ad5-WT (Fig. 3-3C, right panel).

To determine if the reduction in infectious virus production with the mutant viruses was due to a packaging defect, we analyzed virus particle production *in vivo* using CsCl equilibrium density gradient centrifugation in the three cell lines (Fig. 3-3D). Consistent with the results presented in panel B, supplementation of L4-22K, but not L4-33K, to C137S/C141S mutant virus infections augmented production of mature virions. Note that mature particle production is consistent amongst the three cell lines with Ad5-WT infection. We conclude that the C137 and C141 residues in the C-terminal region of L4-22K play a role in viral genome packaging and mutation of these residues dramatically decreases the production of infectious virus without affecting the levels of L4-22K or L4-33K protein expression. We did not observe any significant increase in mature particle production with H166Q/H170Q-infected TetC4-22K-ΔC cells or TetC4-33K cells compared to N52.E6-Cre cells. Additionally, mutation of the H166 and H170 residues dramatically affects the levels of both L4-22K and L4-33K protein expression and, therefore, interferes with the production of infectious virus, a phenotype that is not rescued by supplementing either protein alone by the complementing cell line. This phenotype raises the possibility that the H166Q/H170Q is a dominant-negative mutant in addition to having secondary *cis*- acting effects on the L4-33K protein (discussed below).

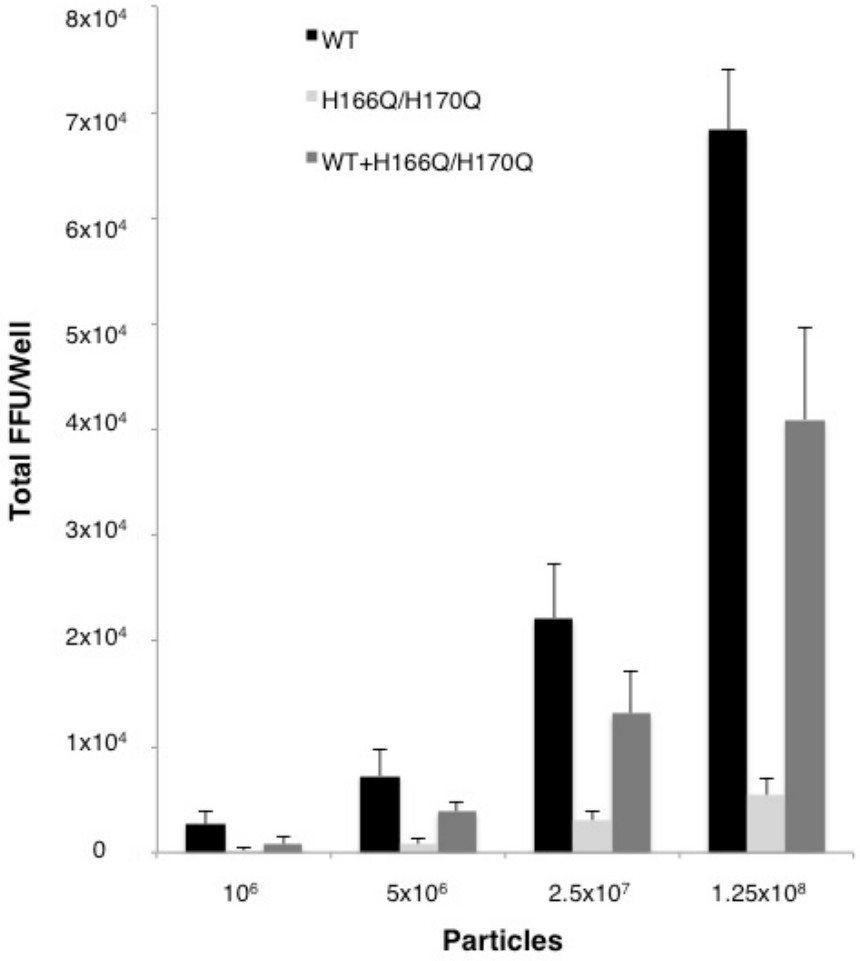


Figure 3-4 The L4-22K H166Q/H170Q mutant virus does not inhibit Ad5-WT growth. N52.E6-Cre cells were infected with equal number of particles each of Ad5-WT alone, H166Q/H170Q alone, or Ad5-WT and H166Q/H170Q combined, and harvested 48 hpi. Serial dilutions of lysates were used to measure infectivity by a fluorescence focus assay, represented as total FFU/well.

Characterization of the H166Q/H70Q virus packaging defect.

To determine if the observed decrease in packaging with the H166Q/H170Q mutant virus in N52.E6-Cre and TetC4-22K-ΔC cells was due to a dominant-negative effect, we examined if the H166Q/H170Q mutant virus inhibits Ad5-WT growth. N52.E6-Cre cells were infected with equal numbers of particles each of Ad5-WT alone, H166Q/H170Q alone, or equal numbers of both combined, and harvested at late times after infections. The lysates were used in serial dilutions to measure infectivity directly in a fluorescence focus assay and the number of DBP positive cells were calculated per well and reported as total FFU/well (Fig. 3-4). The number of infected cells was ~10 fold lower with the H166Q/H170Q mutant virus alone than with Ad5-WT. When lysates produced from a co-infection experiment were used to measure infectivity, the number of infected cells was approximately half of Ad5-WT alone and H166Q/H170Q alone, representing infected cells from both populations of viruses. This suggests that the mutant virus does not inhibit WT growth and does not appear to be a dominant-negative mutant.

To rule out the possibility that the packaging defect observed with the H166Q/H170Q mutant virus was due to an underlying defect in the assembly process or an aberrant amount of structural proteins, we isolated protein from mature particles grown on N52.E6-Cre cells by silver stain analysis (Fig. 3-5A) and western blot analysis (Fig. 3-5B). The amount of protein loaded onto the gels was normalized to the levels of hexon protein for each virus, to ensure that equal numbers of particles were represented. Consistent with the protein composition for mature particles, core proteins V and VII were present in equal proportions in samples from Ad5-WT and mutant viruses, as well as some of the other major and minor structural proteins (Fig 3-5). The results revealed that the mature particles generated from

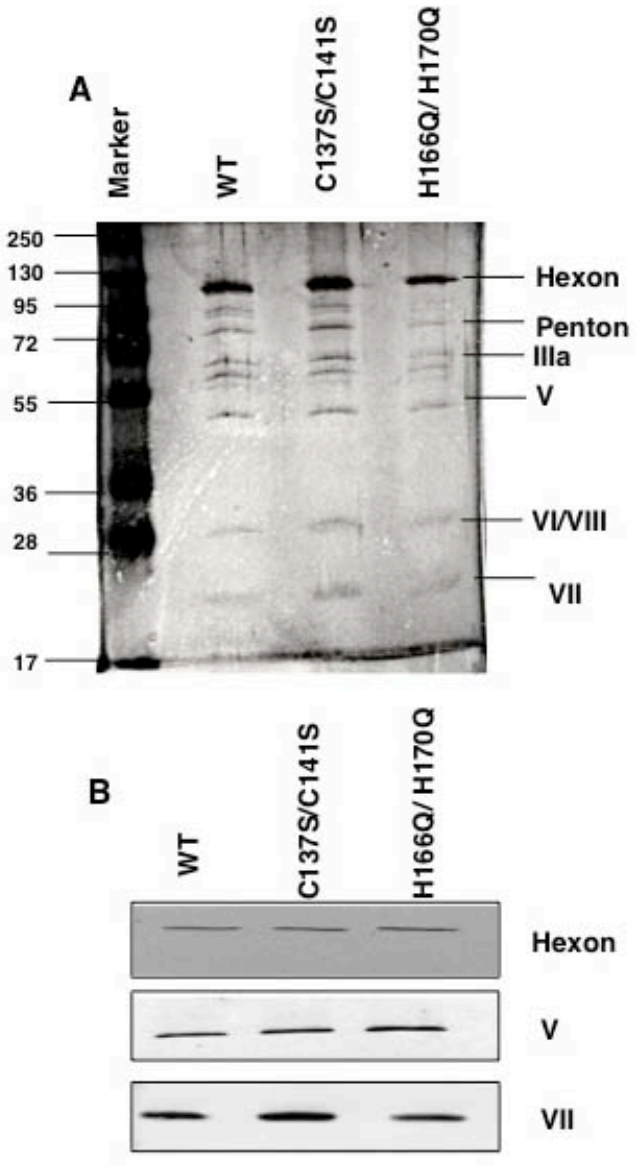


Figure 3-5 Analysis of proteins in mature particles of Ad5- WT, C137S/C141S, and H166Q/H170Q mutant viruses. Proteins from empty particles obtained from CsCl equilibrium gradients of N52.E6-Cre infected cells were analyzed by (A) silver staining and (B) SDS-PAGE. Representative examples of major and minor structural proteins are represented on the right.

both the C137S/C141S and H166Q/H170Q mutant viruses are bona-fide mature particles.

The conserved cysteine and histidine residues are not required for L4-22K protein to bind to packaging sequences.

To examine if the conserved cysteine and histidine residues constituted a DNA binding motif involved in the DNA-protein interaction of L4-22K with packaging sequences, we analyzed DNA binding by *in vitro* gel mobility shift assays. Nuclear extracts were prepared at late times after infection of cells infected with viruses Ad5-WT, C137S/C141S, or H166Q/H170Q, and the formation of complexes specific to Ad5 packaging sequences was assessed using a probe containing A repeats 1 and 2. Consistent with previous results (78), three complexes were observed with extracts form WT-infected cells (Fig. 3-6, lane 4). Complex 1 corresponds to the binding of IVa2 to the CG motif in the A1 repeat; complex 2 is formed with the L4-22K protein, in addition to IVa2 binding, to the TTTG motif of the A2 repeat; and complex 3 is due to the binding of an additional IVa2 protein to the CG sequence in the A2 repeat (78, 124, 125). Note that only complex 1 is formed with extracts from Δ22K virus-infected cells since functional L4-22K protein is not present (Fig. 3-6, lanes 10, 11). The amount of complexes 1, 2 and 3 formed relative to each other varies somewhat from experiment to experiment. With extracts from C137S/C141S mutant virus-infected cells, complexes 1 and 3 were detected (Fig. 3-6, lane 2). Although complex 2 is weak and difficult to detect in this assay, we postulate that binding to the packaging sequence is present, since addition of an antibody directed against the unique coding region of the L4-22K protein

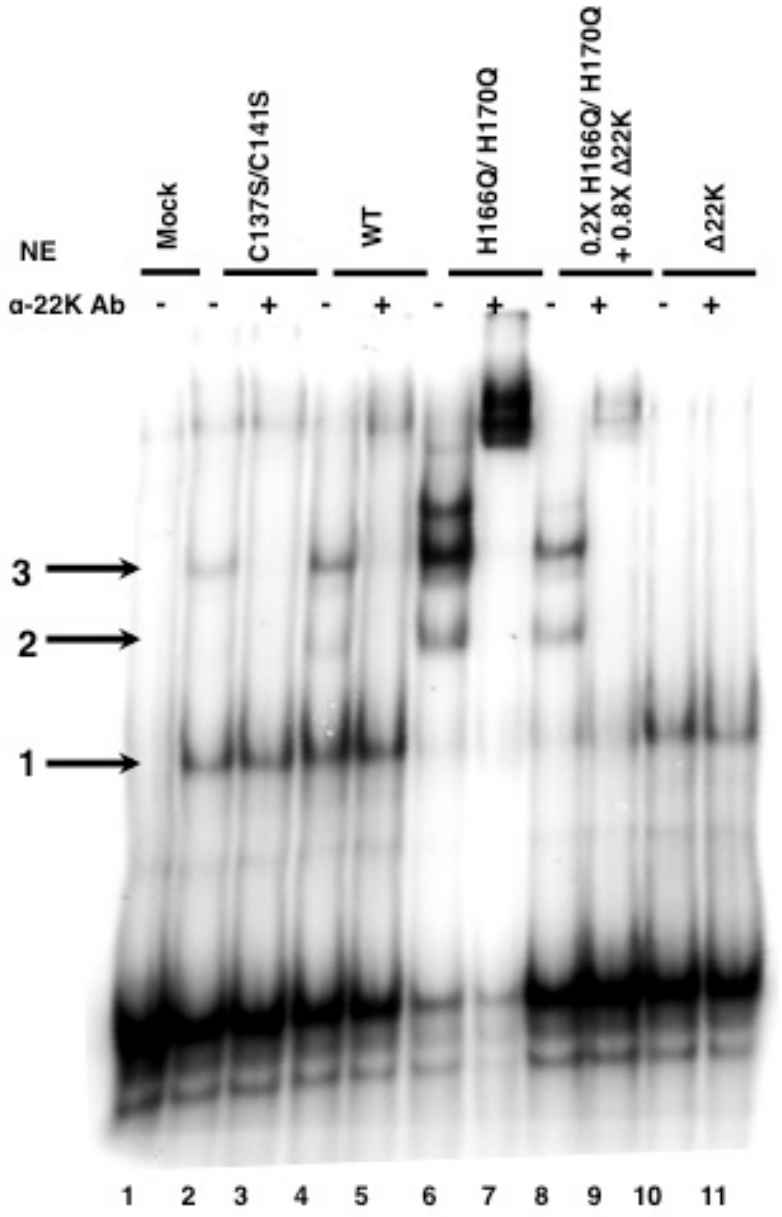


Figure 3-6. The L4-22K-C137S/C141S and L4-22K-H166Q/H170Q mutant proteins bind to Ad5 packaging sequences *in vitro*. Gel mobility shift assays were done with a ^{32}P-labeled probe containing Ad5 A repeats 1 and 2. Six µg of nuclear extract from WT, C137S/C141S, H166Q/H170Q, Δ22K, or mock-infected N52.E6-Cre cells harvested at 24hpi were used for the assays. In lanes 8 and 9, 1.2 µg (0.2X of total sample) of nuclear extract from H166Q/H170Q infected cells was mixed with 4.8 µg (0.8X of total sample) of nuclear extract from Δ22K-infected cells prior to incubation with probe. In lanes 3, 5, 7, 9, and 11, anti-L4-22K protein antibody was added to the binding reaction mixture prior to incubation with probe. The positions of complexes 1, 2, and 3 are indicated by the arrows to the left of the gel.

eliminated complex 3 (Fig. 3-6, lane 3), similar to that observed with Ad5-WT-infected extract (Fig. 3-6, lane 5).

In addition to complexes 1, 2, and 3, additional complexes formed with nuclear extracts prepared from H166Q/H170Q mutant virus-infected cells (Fig. 3-6, lane 6). As shown in Fig. 3-3C and discussed again below, L4-22K expression with the H166Q/H170Q mutant virus was elevated compared to Ad5-WT which may explain the formation of these additional complexes. Nuclear extracts from Ad5-WT, C137S/C141S, and H166Q/H170Q mutant virus- infected cells were also analyzed for L4-22K and IVa2 protein expression and the levels of both proteins were comparable to those observed with whole cell extracts (data not shown). The additional complexes observed with the H166Q/H170Q mutant contain the L4-22K protein as observed by supershifts when antibody against L4-22K protein was added to the reaction (Fig. 3-6, lane 7). Biochemical studies have provided evidence that L4-22K binding promotes cooperative assembly of IVa2 onto packaging sequences, increasing affinity for the loading of additional IVa2 monomers to complexes with an existing IVa2 and L4-22K monomer already present (33, 34). To determine if the observed complexes were due solely to elevated amounts of L4-22K protein, we normalized the levels of nuclear extract from H166Q/H170Q virus-infected cells to protein levels from Ad5-W- infected nuclear extracts (Ad5-WT L4-22K protein levels correspond to 0.2X of H166Q/H170Q mutant L4-22K protein levels, data not shown). To supplement IVa2 to the reaction, we added the remaining amount of nuclear extract (0.8X of the total amount of nuclear extract) from Δ22K virus-infected cells. As shown in lane 8, complexes 1, 2, and 3 formed and the upper complexes were reduced in intensity; the addition of an antibody against L4-22K resulted in a supershift of complexes 2 and 3 (Fig. 3-6, lane 9). We conclude that the conserved cysteine

and histidine residues in the C-terminal region of the L4-22K protein are not critical for the formation of protein-DNA complexes on the packaging sequences. We also provide further evidence of the role of L4-22K in promoting cooperative assembly of protein-protein and protein-DNA interactions onto packaging sequences as observed by the additional complexes formed with elevated levels of L4-22K protein beyond the normal levels observed during a WT infection.

L4-22K regulates Ad late gene expression and acts via a mechanism different from that of its role in viral genome packaging.

To examine if viral gene expression was affected with the L4-22K mutant viruses, N52.E6-Cre cells were infected with viruses Ad5-WT, C137S/C141S, H166Q/H170Q, or Δ22K and whole cell extracts were prepared at different times after infection and analyzed by western blot for the levels of different viral proteins. Late gene expression (regions L1 to L5) with the C137S/C141S mutant was comparable to that of WT (Fig. 3-7). Therefore, the defect in production of infectious progeny shown in Fig. 3-3 with the C137S/C141S mutant cannot be attributed to decreases in late gene products required for virus assembly and DNA packaging. Consistent with previous data (53), minor capsid proteins IIIa, VI and VIII, core proteins V and VII, and the viral splicing factor L4-33K, exhibited a decrease in expression with the Δ22K mutant compared to Ad5-WT (Fig. 3-7). All other products were only modestly delayed (hexon, penton, fiber) or not affected (L4-100K). Analysis of late proteins expressed with the H166Q/H170Q mutant showed that, similar to the Δ22K mutant, proteins IIIa, L4-33K, and VIII were decreased, while the levels of other Ad late proteins were only modestly affected. The level of L4-22K protein expression was increased with the

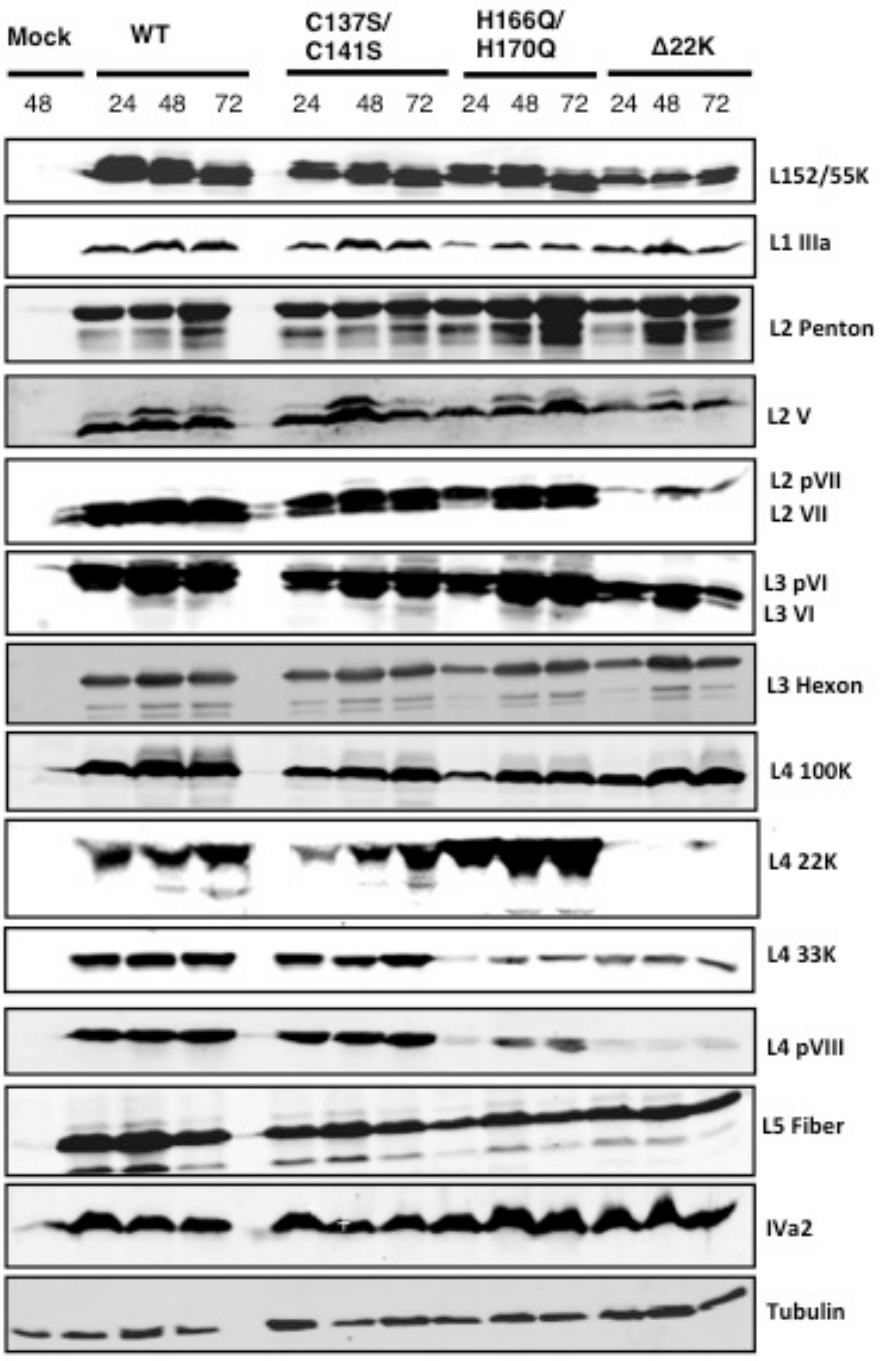


Figure 3-7 Western blot analyses of viral gene expression. N52.E6-Cre cells were infected with WT, C137S/C141S, H166Q/H170Q, Δ22K, or mock-infected and whole cell extracts were isolated at 24, 48, or 72 hpi as indicated at the top. Late proteins from regions L1-L5 and intermediate protein IVa2 are shown with designations on the right. Tubulin is presented as a loading control.

H166Q/H170Q mutant compared to Ad5-WT and the C137S/C141S mutant. Intermediate protein IVa2 was also analyzed and no significant changes were observed with the mutant viruses.

Previous studies presented evidence that L4-22K regulates gene expression and drives the early-to-late phase switch during Ad infection by acting at the level of late gene mRNA production and/or stability (53, 79). To determine whether the observed delays and decreases in late gene expression are due to L4-22K exerting its post-transcriptional effect at the level of protein or mRNA production, Northern blot analyses were performed to measure levels of Ad late transcripts (Fig. 3-8A). Consistent with the western blot analyses, late mRNA levels were not affected with the C137S/C141S mutant compared to Ad5-WT. The mRNA levels for most of the major and minor capsid proteins were decreased with the H166Q/H170Q and Δ22K mutants, suggesting that the corresponding reduction in late protein expression (Fig. 3-7A) is at the level of mRNA production and/or stability. Note that this was not a global effect and that the post-transcriptional role of L4-22K is selective for a particular set of mRNAs. The reduction in IIIa mRNA levels is consistent with the decreased levels of L4-33K protein, since previous studies have shown that L4-33K regulates alternative splicing of the L1 unit by activating the IIIa splice site during late times of infection (75, 86, 90). Since the levels of L4-22K mRNA were undetectable for Ad5-WT and the C137S/C141S mutant by Northern blot, we used RT-PCR to measure L4-22K mRNA levels (Fig. 3-8B). Consistent with the levels of protein expression (Fig. 3-7A), elevated levels of L4-22K mRNA were observed with the H166Q/H170Q mutant as early as 24 hpi compared to Ad5-WT and the C137S/C141S mutant. L4-22K mRNA levels for the H166Q/H170Q mutant were elevated enough to be visible by Northern blot (Fig. 3-8A, arrow).

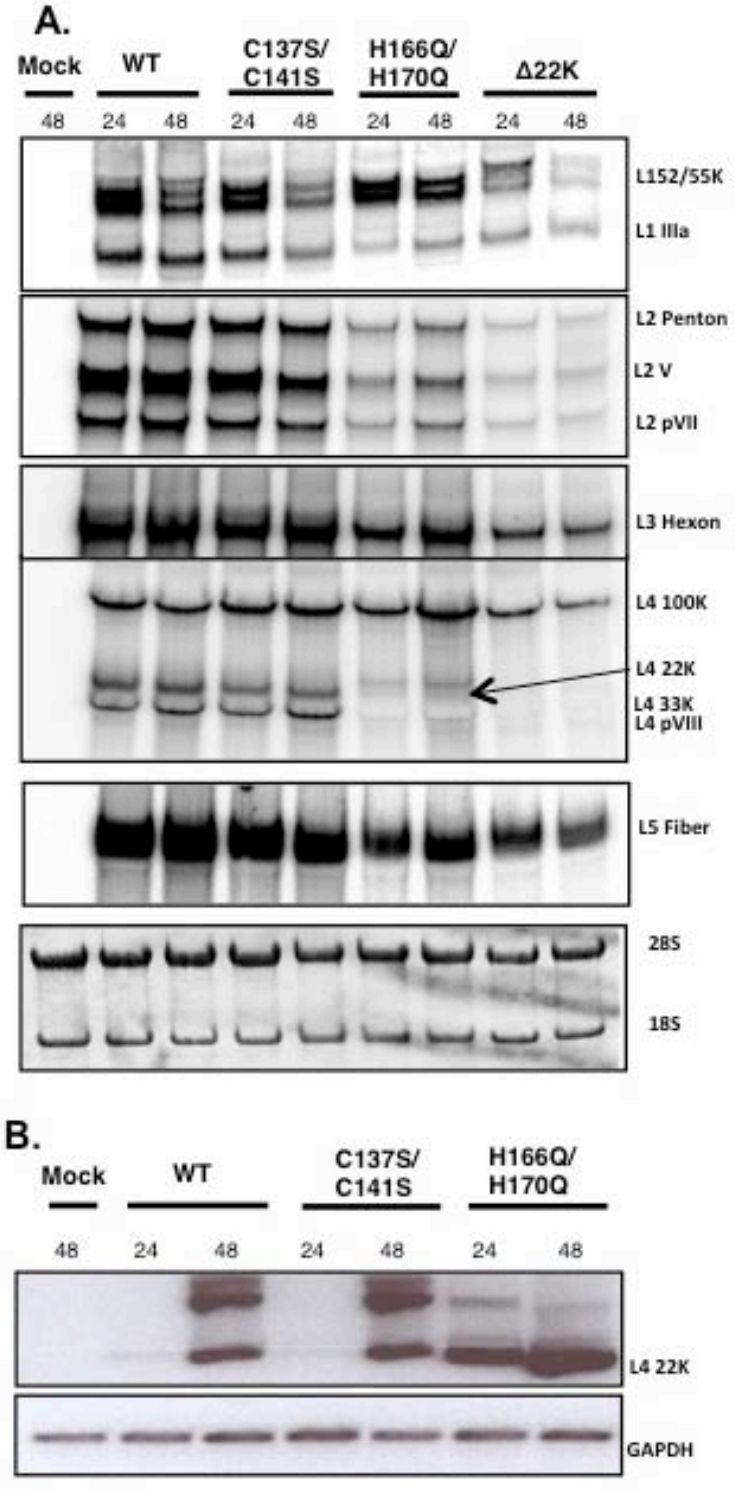

A.
Mock
WT
C137S/
C141S
H166Q/
H170Q
Δ22K
48
24
48
24
48
24
48
24
48
L152/55K
L1 IIIa
L2 Penton
L2 V
L2 pVII
L3 Hexon
L4 100K
L4 22K
L4 33K
L4 pVIII
L5 Fiber
28S
18S
B.
Mock
WT
C137S/
C141S
H166Q/
H170Q
48
24
48
24
48
24
48
L4 22K
GAPDH

Figure 3-8

Panel A. Analysis of Ad late mRNA levels. N52.E6-Cre cells were infected with Ad5-WT, C137S/C141S, H166Q/H170Q, Δ22K, or mock-infected and total cytoplasmic mRNAs isolated at 24 and 48 hpi. Cytoplasmic RNA was analyzed by Northern blotting using a ^{32}P -labeled probes designed around the co-terminal 3' ends of each mRNA family (L1-L5) with the designated mRNAs labeled to the right. The blot was stained with methylene blue as a loading control (bottom panel).

Panel B. L4-22K mRNA was detected by RT-PCR using a forward primer specific for TPL 3 and reverse primer directed to nucleotides 26682-26703 (corresponding to the intron of L4-33K). Equivalent amounts of cDNA derived from the reverse transcriptase reaction was used in a PCR mixture with primers directed against GAPDH (bottom panel).

We conclude that the observed changes in protein levels with the L4-22K mutants are reproduced at the mRNA level, confirming a post-transcriptional mechanism for L4-22K in regulating viral gene expression. In addition, the C137S/C141S mutations in the L4-22K protein had no effect on gene expression suggesting that the L4-22K protein exploits a distinct mechanism to exert its post-transcriptional role than that of genome packaging. Finally, we found that the H166Q/H170Q mutant had a similar phenotype with respect to mRNA production and changes in protein expression as an L4-22K null mutant.

L4-22K regulates E1A expression at the level of mRNA production.

In addition to late gene expression, we wanted to analyze the levels of early gene expression with the L4-22K mutants, and used E1A as a representative example. To our surprise, the Δ22K mutant virus had significantly higher E1A expression levels than Ad5-WT as infection progressed to the late phase, suggesting that L4-22K regulates E1A protein expression (Fig. 3-9A). To determine if the conserved cysteine and histidine residues in L4-22K are key to this function, we analyzed E1A expression with the C137S/C141S and H166Q/H170Q mutant viruses and compared to Ad5-WT and the Δ22K mutant (Fig. 3-9A). The H166Q/H170Q mutant virus had elevated E1A expression levels compared to both Ad5-WT and the C137S/C141S mutant as early as 6 hpi. These results indicate that the L4-22K conserved cysteine residues are not important for suppression of early gene expression, but suggest the possibility that the conserved histidine residues may be involved in this function.

To further explore the increased levels of E1A observed with L4-22K mutant viruses during the late phase of infection, we used Northern blot analysis to measure E1A mRNA levels with the C137S/H170Q and H166Q/H170Q mutant viruses (Fig. 3-9B), and compared

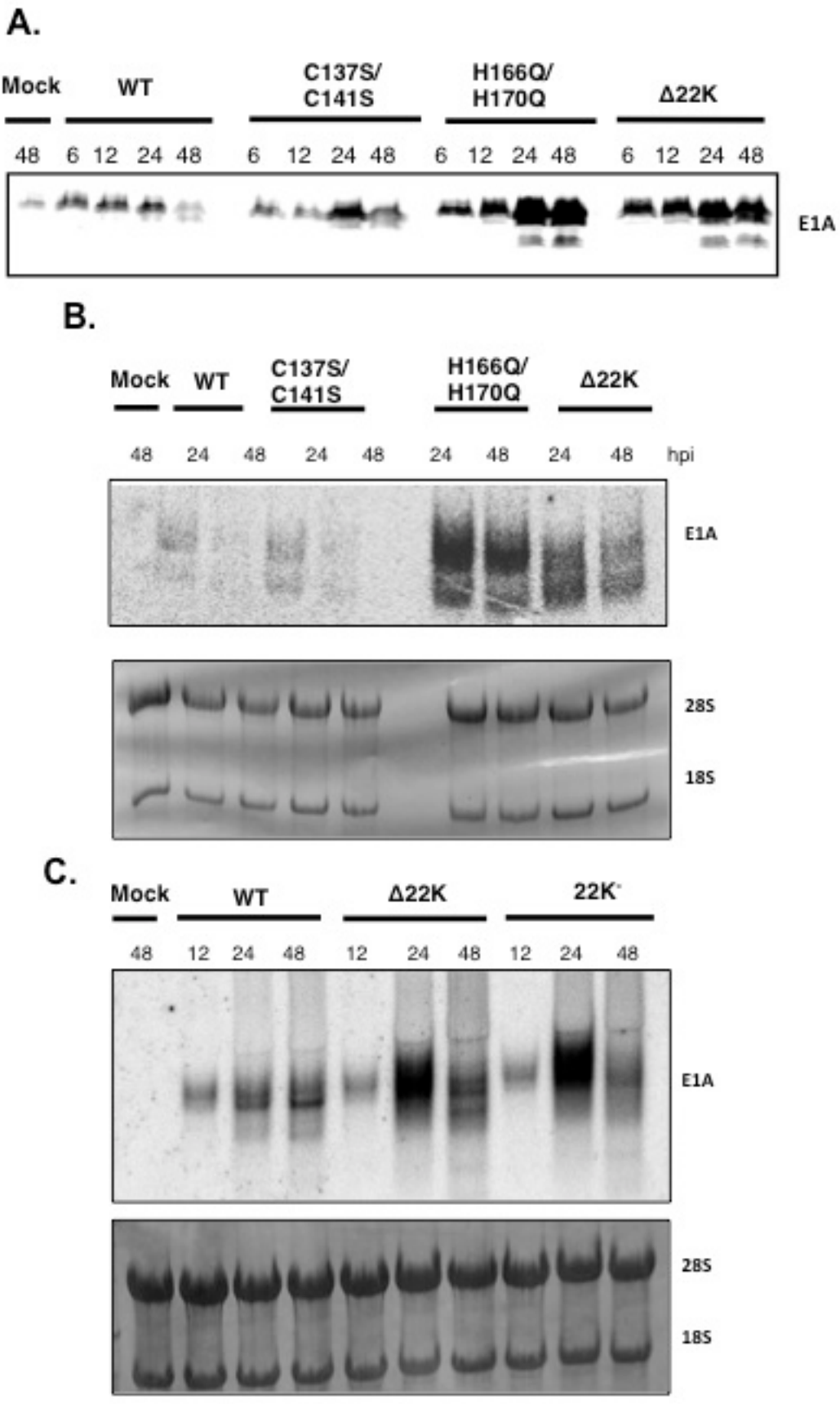

A.
Mock
WT
C137S/
C141S
H166Q/
H170Q
Δ22K
48 6 12 24 48 6 12 24 48 6 12 24 48 6 12 24 48
E1A
B.
Mock
WT
C137S/
C141S
H166Q/
H170Q
Δ22K
48 24 48 24 48 24 48 24 48 hpi
E1A
28S
18S
C.
Mock
WT
Δ22K
22K-
48 12 24 48 12 24 48 12 24 48
E1A
28S
18S

Figure 3-9 Analysis of E1Aprotein levels and mRNA levels.

Panel A. N52.E6-Cre cells were infected with Ad5-WT, C137S/C141S, H166Q/H170Q, Δ22K, or mock-infected and whole cell extracts were isolated at 24 and 48 hpi and analyzed by Western blotting.

Panel B. Total cytoplasmic mRNAs were isolated at 24 and 48hpi and analyzed by Northern blotting using a ^{32}P-labeled E1A probe from samples in panel A.

Panel C. Total cytoplasmic mRNAs isolated from Ad5-WT, Δ22K, and 22K^{-} -infected A549 cells at 12, 24, and 48hpi were analyzed by Northern blotting, as in panel B. Each blot was stained with methylene blue as a loading control and is shown below each E1A blot.

it to two different L4-22K null mutant viruses (Fig 3-9B and C). The results showed that E1A mRNA levels were significantly elevated in Δ22K, 22K^{-}, and H166Q/H170Q -infected cells at 24hpi compared to Ad5-WT and C137S/C141S-infected cells. This increase suggests that the L4-22K protein plays a role in suppressing early gene expression.

The L4-22K protein acts at the level of processing of the L4-33K mRNA.

To investigate whether the post-transcriptional role of L4-22K is at the level of alternative mRNA splicing, we analyzed the mRNA and protein levels of L4-33K in an *in vivo* splicing assay with and without L4-22K protein expression (Fig. 3-10). N52.E6-Cre cells were cotransfected with plasmids that express GFP and WT or mutant L4 proteins, and protein levels were analyzed by western blot and mRNA levels were analyzed by Northern blot. To compensate for any differences in transfection efficiencies, all samples were normalized to GFP mRNA and GFP protein levels. Expression plasmids pcDNA3-L4-22K and pcDNA3-L4-33K contain the coding sequences for the L4-22K and L4-33K proteins, respectively (Figs. 3-10A and 3-10B, lanes 2 and 3). Note that the mRNA transcribed with transfection of pcDNA3-L4-22K is not visible with the probe used in the assay (Ad5 nt 26769-27590; Fig. 3-10B, lane 2). Plasmid pcDNA3-L4 contains the genomic sequences encoding both L4-22K and L4-33K proteins including the intron interrupting the protein reading frame of L4-33K (Fig. 3-1A), hence production of the coding L4-33K mRNA requires splicing of the initial transcript (Fig. 3-10B, lane 4). When this plasmid was transfected into N52.E6-Cre cells, both L4-22K and L4-33K proteins were expressed to similar levels (Fig. 3-10A, lane 4). Plasmids pcDNA3-L4-22K^{-} and pcDNA3-L4-33K^{-} each have a stop codon in the L4-22K and L4-33K reading frames, respectively, in the background

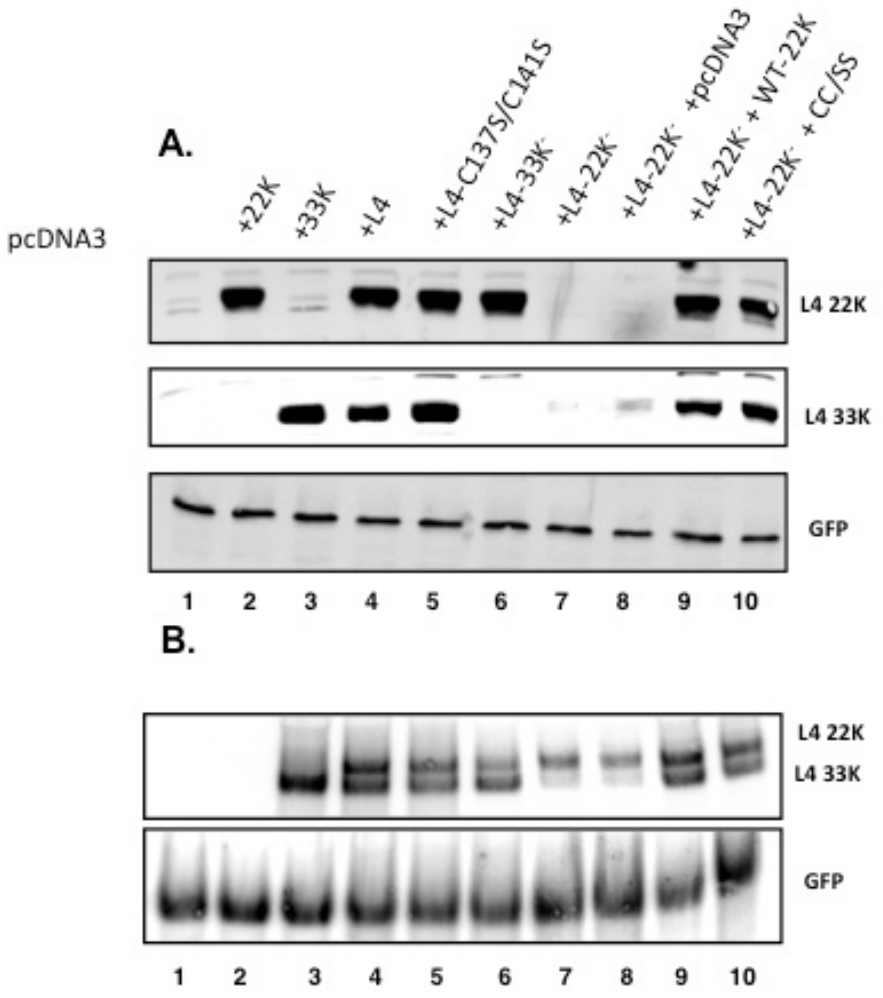


Figure 3-10 *In vivo* splicing assay for the L4-33K pre-mRNA

Panel A. N52.E6-Cre cells were co-transfected with expression plasmids for WT or mutant L4 proteins and an expression plasmid for GFP and harvested 48h after transfection. Cells were used for Western blot analysis; all samples were normalized to GFP protein or mRNA levels.

Panel B. Samples from panel A were used for Northern blot analysis; lane designations from panel A apply.

of the pcDNA3-L4 plasmid and the L4 pre-mRNA is still transcribed regardless of protein expression. Transfection of pcDNA3-L4-33K^{-} resulted in normal expression levels of the L4-22K protein, while L4-33K was not expressed (Fig. 3-10A, lane 6). Transfection of pcDNA3-L4-22K^{-} resulted in a significant decrease in L4-33K protein expression (Fig. 3-10A, compare lanes 4 and 7) due to inefficient splicing of the L4-33K pre-mRNA (Fig. 3-10B, compare lanes 4 and 7).

To further probe if the L4-22K protein is required for efficient splicing of the L4-33K transcript, we co-transfected equal amounts of pcDNA3-L4-22K^{-} and pcDNA3-L4-22K and analyzed protein expression and mRNA levels. Expression of L4-33K was complemented when the L4-22K protein was expressed in *trans* (Fig. 3-10A, lane 9) consistent with restoration of L4-33K pre-mRNA splicing (Fig. 3-10B, lane 9). This result was not observed when pcDNA3-L4-22K^{-} was cotransfected with empty vector, pcDNA3 (Figs. 3-10A and B, lane 8) demonstrating that L4-22K, and not any other sequence in the vector background, is required for efficient splicing of the L4-33K mRNA. The results from Figs. 3-7, 3-8, and 3-9 suggest that the conserved cysteine residues in the C-terminus of L4-22K are not required for the regulation of viral gene expression, so we analyzed if substitution mutations of these residues had any effect on the *in vivo* splicing of L4-33K mRNA. Transfection of plasmid pcDNA3-L4-C137S/C141S had no effect on the protein or mRNA levels of either the L4-22K or L4-33K genes (Figs. 3-10A and 3-10B, lane 5). Co-expression of plasmid pcDNA3-L4-C137S/C141S with plasmid pDNA3-L4-22K^{-} complemented the 22K^{-} plasmid to restore L4-33K protein expression, consistent with an increase in L4-33K pre-mRNA accumulation (Figs. 3-10A and 3-10B, lane 10).

We conclude that the post-transcriptional effects observed on specific Ad mRNAs

when functional L4-22K is not present during infection (Fig. 3-8 and 3-9) can be attributed to a role of the L4-22K protein in regulating alternative splicing of specific transcripts. The presence of full-length L4-22K protein in an *in vivo* splicing assay resulted in an increase in L4-33K mRNA accumulation and a subsequent increase in L4-33K protein levels. This effect was observed even with the C137S/C141S L4-22K mutant suggesting that the conserved cysteine residues are not essential for the splicing activity of L4-22K.

The conserved histidine residues are required for correct splicing of the L4-33K pre-mRNA.

Since the H166Q/H170Q mutant had the same phenotype as a null L4-22K mutant virus in terms of viral gene expression (Fig. 3-7 and Fig 3-8), we postulated that these residues were critical for the regulation of late gene expression at the level of pre-mRNA splicing. To determine the role of these residues in activating splicing of the L4-33K pre-mRNA, we analyzed mRNA and protein levels in an *in vivo* splicing assay (Fig. 3-11 and Fig. 3-12). As shown in Figure 3-10, N52.E6-Cre cells were cotransfected with plasmids that express GFP and WT or mutant L4 proteins, and protein levels were analyzed by western blot and mRNA levels were analyzed by Northern blot and normalized to GFP levels. Expression plasmids pcDNA3-22K-C137S/C141S and pcDNA3-22K-H166Q/H170Q contain the coding sequences for the L4-22K protein with the respective mutations. Note that the expression of the of L4-22K with the C137S/C141S and H166Q/H170Q mutations are comparable to those observed with mutant virus infected extracts (Fig. 3-3C), where protein expression is slightly increased with the H166Q/H170Q mutant (Fig 3-11A, lanes 4) as compared to WT L4-22K (lane 2) and C137S/C141S (lane 3). Expression of the L4-33K

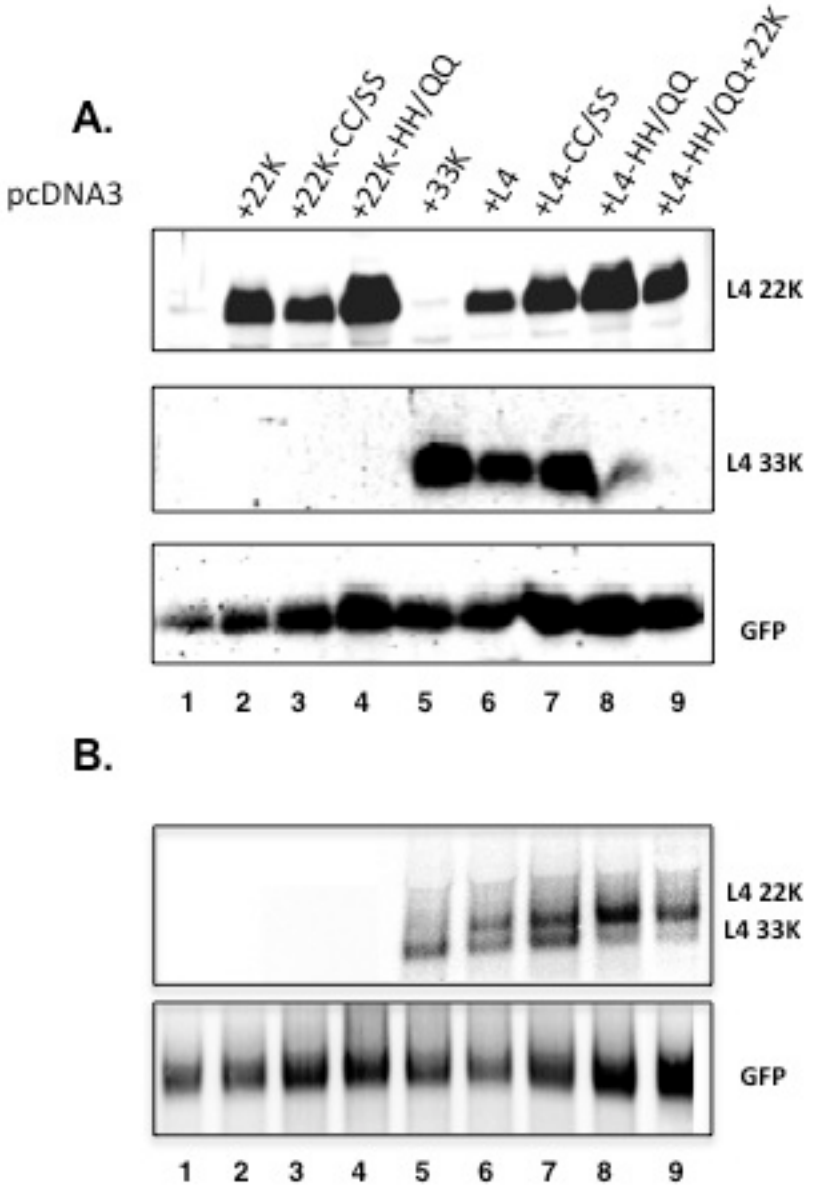


Figure 3-11 *In vivo* splicing assay for the L4-33K pre-mRNA using mutant L4-22K proteins .

Panel A. N52.E6-Cre cells were co-transfected with expression plasmids for WT or mutant L4 proteins and an expression plasmid for GFP and harvested 48h after transfection. Cells were used for Western blot analysis; all samples were normalized to GFP protein or mRNA levels.

Panel B. Samples from panel A were used for Northern blot analysis; lane designations from panel A apply

protein and mRNA with the pcDNA3-33K plasmid is shown as a control (Fig 3-11A and B, lane 5), as well as expression of the L4-22K and L4-33K proteins and mRNAs with the pcDNA3-L4 plasmid (lane 6). Plasmids pcDNA3-L4-C137S/C141S and pcDNA3-L4-H166Q/H170Q contain double serine substitution mutations or double glutamine substitution mutations in the background of the pcDNA3-L4 plasmid (described above), respectively. Therefore, correct splicing, and hence expression, of the L4-33K protein would require the mutant L4-22K protein to be functional. As recapitulated in Figure 3-10A and B, transfection of plasmid pcDNA3-L4-C137S/C141S had no effect on the protein or mRNA levels of either the L4-22K or L4-33K genes (Fig. 3-11A, 3-11B, lane 7). However, transfection of plasmid pcDNA3-L4-H166Q/H170Q resulted in a significant decrease in L4-33K mRNA splicing (Fig. 3-11B, lane 8) and L4-33K protein expression (Fig. 3-11A, lane 8). Expression of L4-33K was not rescued when the L4-22K protein was expressed in *trans* (Fig. 3-11A and B, lane 9).

To further explore the role of the histidine residues in regulating splicing of the L4-33K transcript, we co-transfected equal amounts of pcDNA3-L4-22K$^-$ with either pcDNA3-C137S/C141S or pcDNA3-H166Q/H170Q. As expected, the C137S/C141S mutant protein rescued L4-33K splicing and protein levels (Fig. 3-12A and B, lane 4), to levels similar to WT L4-22K (Fig 3-12A and B, compare lanes 1 and 2). The H166Q/H170Q mutant protein only provided partial restoration of L4-33K splicing and protein expression (Fig. 3-12A and B, lane 3). Due to the proximity of the histidine residues to plausible *cis*-acting regulatory sequences for splicing in the L4-33K gene (Fig. 3-1C), it is possible that the two point mutations introduced to change the histidine residues were having a *cis* effect on the L4-33K

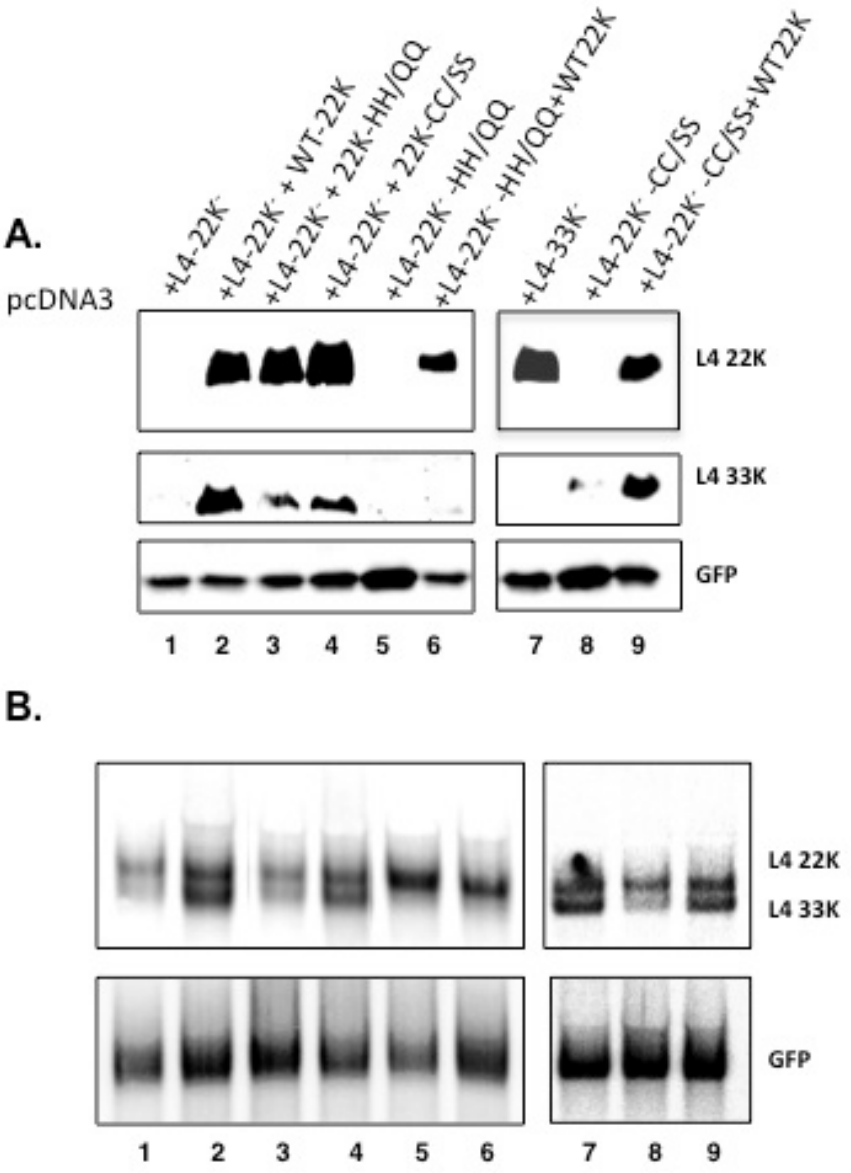


Figure 3-12 The L4-22K-H166Q/H170Q mutations disrupt splicing for the L4-33K pre-mRNA.

Panel A. N52.E6-Cre cells were co-transfected with expression plasmids for WT or mutant L4 proteins and an expression plasmid for GFP and harvested 48h after transfection. Cells were used for Western blot analysis; all samples were normalized to GFP protein or mRNA levels.

Panel B. Samples from panel A were used for Northern blot analysis; lane designations from panel A apply.

regulatory sequences, in addition to, or in place of, a *trans* effect from a mutant L4-22K protein. To test this hypothesis, we introduced double serine or double glutamine substitution mutations in the background of the pcDNA3-L4-22K$^-$ plasmid. Since functional L4-22K protein is not made, the only contributing factors to any observed phenotypes are the point mutations themselves. Transfection of plasmid pcDNA3-L4-22K$^-$-C137S/C141S abolished splicing of the L4-33K transcript (Fig. 3-12B, lane 8) and levels of L4-33K protein were reduced (Fig. 3-12A, lane 8), but levels were restored when WT L4-22K protein was supplemented *in trans* (Fig 3-12A and B, lane 9). In contrast, transfection of plasmid pcDNA3-L4-22K$^-$-H166Q/H170Q completely abolished splicing of L4-33K pre-mRNA (Fig. 3-12B, lane 5), and co-transfection with pcDNA3-22K had no restoring effect (Fig 3-12B, lane 6). In turn, L4-33K protein levels were dramatically reduced (Fig 3-12A, lanes 5 and 6). L4-22K and L4-33K gene expression levels are shown for co-transfection with the pcDNA3- L4-33K$^-$ plasmid as a control (Fig. 3-12A, B, lane 7).

We conclude that the splicing defect on the L4-33K transcript observed with the H166Q/H170Q mutant protein is caused by a double effect arising from the point mutations: first, L4-33K gene expression levels were decreased with the 22K-H166Q/H170Q mutant protein as a result of disrupting the inherent function of the L4-22K protein; and second, L4-33K gene expression levels were decreased due to the proximity of the histidine residues to the putative *cis*-acting regulatory sequences of the L4-33K pre-mRNA (Fig. 3-1, C).

Discussion

We previously demonstrated that the Ad5 L4-22K protein is an integral component at key steps in the viral infectious cycle. First, L4-22K binds to Ad packaging sequences *in vitro* and *in vivo* and is essential for viral genome packaging and, therefore, for the production of infectious progeny (53, 78). Additionally, L4-22K has a distinct function in the temporal control of viral gene expression by activating MLP transcription (80), and more importantly, by regulating the accumulation of specific Ad late transcripts (53, 79). To further explore the functions of L4-22K, we generated and characterized two new L4-22K mutant viruses containing point mutations in conserved pairs of cysteine and histidine residues in the C-terminal region of L4-22K (C137S/C141S and H166Q/H170Q, Fig. 1). We proposed that the primary sequence and configuration of these amino acids was consistent with a DNA binding motif and speculated that the DNA binding activity of L4-22K to the packaging sequences may be mediated by this putative zinc finger motif. Our results showed that the conserved cysteine residues are essential for efficient generation of infectious progeny, while no changes in gene expression patterns were observed with the C137S/C141S mutant virus. The H166Q/H170Q mutant recapitulated the phenotype of an L4-22K null mutant virus suggesting an important role for these histidine residues in both Ad genome packaging and viral gene expression, although the possibility of an indirect effect on L4 mRNA splicing cannot be ruled out (discussed below). Both mutant proteins bound to packaging sequences *in vitro* suggesting that the tandem CXXXC and HXXXH motifs do not comprise a protein domain critical for DNA binding. Our results confirm *in vivo* the essential role of L4-22K as a post-transcriptional regulator of late gene expression of specific Ad mRNAs with primary targets the L4-33K and L4-VIII mRNAs and proteins. Our findings

show that the cysteine mutations did not affect the ability of L4-22K to regulate gene expression (early or late) at the mRNA or protein levels (Figs. 3-7 and 3-8). These results uncouple the functions of L4-22K in regulating viral gene expression and genome packaging and suggest that the conserved cysteine residues are unique to the molecular mechanism behind Ad genome packaging. Finally, we confirmed the post-transcriptional effect of L4-22K on L4-33K expression and reveal that this occurs at the level of mRNA splicing.

Both the C137S/C141S and H166Q/H170Q mutant viruses had comparable infectivity and viral genome replication to Ad5-WT (Figs. 3-2A and B), demonstrating that events leading up to the late phase of infection do not account for the observed effects on late protein production and viral genome packaging. With respect to plaque formation, the C137S/C141S mutant virus behaved similarly to WT, while the H166Q/H170Q mutant virus had a severe defect in plaque formation on A549 cells (Fig. 3-2C). The apparent discrepancy in the results with the H166Q/H170Q mutant virus likely reflects the different assays used. The FFU assay (Fig. 3-2B) is a direct measure of how many cells are infected by the virus at an earlier time of infection (18 hpi), while the plaque assay (Fig. 3-2C) is a measure of virus production and spread, a late event during the infectious cycle. With the H166Q/H170Q mutant virus, infectivity (particle/FFU) was similar compared to WT, consistent with the DNA replication data. However, the H166Q/H170Q mutant virus had a significant defect in virus spread. When supplemented with L4-22K *in trans*, plaque formation with the H166Q/H170Q mutant virus was partially restored (Fig. 3-2D). A role for L4-22K in Ad-induced cell death by regulating expression of ADP was previously proposed (53). ADP is known to facilitate cell lysis, and its deletion results in a small plaque phenotype which is due to impaired virus release (126, 127). Western blot analysis showed that ADP protein

levels were significantly reduced with the H166Q/H170Q mutant, and comparable to the low levels observed with L4-Δ22K (data not shown). This may explain the severe defect in plaque formation with the H166Q/H170Q mutant in non-complementing cells. Additionally, the fact that we did not observe complete rescue in plaque formation when L4-22K was supplemented *in trans*, whereas the L4-Δ22K null mutant was rescued to WT levels, suggests that the H166Q/H170Q mutant protein may act as a dominant negative effector (discussed below).

The conserved cysteine residues of L4-22K are important for virus growth and the production of mature virions (Figs. 3-3A and 3-3D). The infectious virus yield and mature particle production with the C137S/C141S mutant was restored when the L4-22K protein was supplemented *in trans* (Figs. 3-3B and 3-3D). To determine if the reduction in genome packaging *in vivo* correlated with a defect in binding of the C137S/C141S mutant protein to packaging sequences, we analyzed DNA binding *in vitro* using a gel mobility shift assay (Fig. 3-6). The data show that the conserved cysteine residues are not critical for binding of the L4-22K protein to the packaging sequences *in vitro,* indicating that these amino acids are important for a different function of the protein in viral genome packaging. The conserved histidine residues also were not required for DNA binding indicating that it is highly unlikely that the cysteine and histidine residues are part of a zinc finger motif involved in binding of L4-22K to the packaging domain. A role for L4-22K as a transcription factor stimulating major late transcription through direct binding to DNA sequences located downstream of the MLP start site was previously proposed (78-80). We cannot rule out the possibility that the conserved cysteine and histidine residues are part of a DNA binding motif necessary for L4-22K to bind to the DE and activate the MLP. However, this seems unlikely since the

C137S/C141S mutations had no effect on expression of genes transcribed from the major late promoter (Fig. 3-7). The possibility that mutation of the conserved cysteine residues leads to a disruption in the overall protein folding also seems unlikely since there was no observed phenotype with respect to mRNA processing and viral protein expression (Figs. 3-7 and 3-8). Additionally, zinc finger motifs also mediate protein-protein interactions in addition to protein-nucleic acid interactions (128). Although no published data exists for any viral or cellular factors that directly interact with the L4-22K protein, this possibility is an open question.

The phenotype of the H166Q/H170Q mutant virus suggests that the mutations severely disrupt the major functions of the L4-22K protein. Infectious virus yield and the production of mature virions were impaired (Fig. 3-3). Alterations in both early and late gene expression were also observed with similar targets and levels as a non-functional L4-22K mutant, and the observed results correlated well with mRNA levels (Figs. 3-7 and 3-8). Of significance, the mRNA and protein expression levels of L4-33K and L4-pVIII were decreased with both the H166Q/H170Q and Δ22K mutant viruses, with a large increase in mRNA and protein levels of L4-22K with the H166Q/H170Q mutant virus as early as 24 hpi, when mRNA levels were undetectable with WT and the C137S/C141S mutant (Figs. 3-7 and 3-8). The increase of L4-22K transcripts inversely correlates with the decreased levels of L4-33K mRNA (Fig. 3-8). Both transcripts are post-transcriptionally processed from the same pre-mRNA, but the L4-33K mRNA contains an additional splice of its internal intron (Fig. 3-1A). If the additional splicing event required for appropriate translation of the L4-33K protein was disrupted, this would increase the pool of L4-22K mRNAs. Our data demonstrate that L4-22K regulates splicing of the L4-33K pre-mRNA (Fig. 3-9) and we suggest that the

conserved histidine residues are critical for this activity, with additional effects from a direct interference with the internal splice acceptor site of L4-33K, discussed below in detail.

In order to examine the contribution of reduced L4-33K gene expression to the H166Q/H170Q mutant virus phenotype, we investigated if virus yield was altered when L4-33K was supplemented *in trans.* The results showed that this did not augment overall virus yield or mature virion production (Fig. 3-3B and 3-3D), although both L4-33K and L4-22K protein expression levels were comparable to WT (Fig. 3-3C). Interestingly, supplementation of L4-22K to H166Q/H170Q mutant infections also had no significant effect on virus yield or mature virion production. This result could be due to a dominant negative effect of the H166Q/H170Q mutant protein on the activity of WT L4-22K and correlates with the inability of L4-22K to restore plaque formation with the H166Q/H170Q mutant (Fig. 2-2D). We also ruled out the possibility that the virus progeny produced by infection with the H166Q/H170Q mutant are not true mature particles (Fig. 3-5). Major core proteins V and VII and other representative major and minor structural proteins are present in mature particles produced from H166Q/H170Q- infected cells, to levels comparable to those from Ad5-WT and C137S/C141S -infected cells. To further investigate if the H166Q/H170Q mutant protein acts as a dominant-negative effector by inhibiting Ad5-WT infectivity and growth, we measured infectivity by infecting cells with a high multiplicity of infection of both Ad5-WT and the H166Q/H170Q mutant virus. We postulated that if the H166Q/H170Q mutant virus was directly inhibiting the WT function of the protein, that infectivity would be reduced to levels observed when using lysates from the H166Q/H170Q mutant infected cells. Interestingly, we found that the H166Q/H170Q mutant virus did not compete with Ad5-WT, but that the number of infected cells observed represented a mixed population of WT and mutant virus

(Fig. 3-4). If the H166Q/H170Q mutant virus is not a dominant-negative effector, then the other possibility for the inability of L4-22K to restore virus yield is that there is a secondary effect of the point mutations on the L4-33K transcript, discussed below.

Morris and Leppard have shown that the L4-22K protein is important for the temporal control of viral gene expression (79). In this report together with data published in Wu *et al* (53), we show that L4-22K activates Ad late gene expression and is involved in the transition from the early to late stages of infection (53). Previous studies have proposed a role for a late viral factor in the down-regulation of viral early gene expression (129). In a WT infection, we observe that E1A mRNA levels and protein levels (Fig. 3-9) decrease as the infectious cycle enters the late stages. However, both null L4-22K mutant viruses and the H166Q/H170Q point mutant virus showed significantly increased E1A expression and mRNA levels late after infection (Fig. 3-9). Based on our current findings, L4-22K regulates late gene expression at a post-transcriptional level, and it seems plausible it also regulates E1A expression post-transcriptionally. E1A encodes five proteins generated by differential splicing of a single primary RNA transcript. At the earliest stages of infection, transcription of E1A is controlled by a constitutive enhancer and the two larger isoforms of E1A dominate, while the smaller isoforms accumulate at late stages of infection. The accumulation of these isoforms of E1A during late time of infection is not well understood (130), so it seems plausible that L4-22K regulates differential splicing of these isoforms to drive the life cycle to late stages of infection. Additionally, the significance of suppressing E1A expression as infection progresses might be to redistribute the cellular machinery from the early genes to late genes.

Our results reveal *in vivo* one mechanism by which the L4-22K protein influences late

gene expression at the level of pre-mRNA processing (Fig. 3-10). In the absence of functional L4-22K protein, the levels of L4-33K protein expression were significantly decreased which correlated with a decrease in L4-33K mRNA levels (Fig. 3-10). When L4-22K was supplied by transient transfection, we observed an increase in L4-33K mRNA accumulation and a corresponding increase in L4-33K protein levels (Fig. 3-10). Our data demonstrate that L4-22K indeed functions at the level of pre-mRNA splicing and activates splicing of the internal intron of L4-33K. The role of L4-22K as a regulator of pre-mRNA splicing as opposed to other post-transcriptional mechanisms is not surprising given that the L4-22K mRNA is nearly identical to the L4-33K mRNA with the exception of the internal intron of L4-33K. The L4-33K protein is an RNA splicing factor that preferentially activates L1-IIIa splicing (90). We recently showed that L4-33K is required for the accumulation of L1-IIIa and pVI, but other late mRNAs are not primary targets of L4-33K in regulating late gene expression (106). It is therefore likely that L4-33K is not the only viral factor acting as an activator of pre-mRNA splicing and that L4-22K is a strong candidate due to shared sequences between the two proteins that may be critical for their function as mRNA splicing activators. It is interesting to note that the splice acceptor site for the second exon of L4-33K is common to the acceptor site of the pVIII gene as previously published (36) and as we confirmed by RT-PCR and nucleotide sequence analysis (data not shown). L4-pVIII is the other gene that was significantly decreased both at the mRNA and protein level with H166Q/H170Q and Δ22K mutant virus infections (Fig. 3-7 and 3-8). We would anticipate that expression of L4-22K in an *in vivo* splicing assay would stimulate pVIII mRNA accumulation.

Based on protein and mRNA levels of early and late transcripts (Fig. 3-7, 3-8, 3-9), we

suggest that the conserved histidine residues are critical for the function of L4-22K in regulating pre-mRNA splicing of the L4-33K transcript. However the phenotype of the H166Q/H170Q mutant is so complex in that plaque formation and production of infectious progeny are not complemented by providing either L4-22K or L4-33K *in trans,* that we suspected that the point mutations perturb the *cis*-acting regulatory elements of the internal L4-33K splice site (Fig. 3-1C). To distinguish between the two possibilities, we used different H166Q/H170Q containing L4-22K plasmids in *in vivo* splicing assays. A plasmid coding for the L4-22K protein with the double glutamine substitutions (pcDNA3-22K-H166Q/H170Q) can partially restore L4-33K splicing and protein expression when co-transfected with plasmid pcDNA3-L4-22K^{-}, but not to the same levels as WT pcDNA3-22K protein (Fig. 3-12A and B, compare lanes 1, 2, and 3). This suggests that the histidine residues play a role, but are not critical, to the L4-22K splicing regulatory function and that mutating them does not completely abolish function. To rule out any *trans*-acting effect of the L4-22K protein, we constructed double glutamine substitution mutations in the background of the pcDNA3-L4-22K^{-} plasmid (pcDNA3-L4-22K^{-} H166Q/H170Q), so any observed phenotype was directly related to the amino acid changes and not a direct effect of the protein function. Transfection of plasmid pcDNA3-L4-22K^{-} H166Q/H170Q alone, or supplemented with WT L4-22K protein, completely abolished L4-33K mRNA and protein levels (Fig. 3-12A and B, lane 5 and 5). When the double glutamine substitutions were constructed in a pcDNA3-L4 plasmid, both the L4-22K protein was mutated, as well as the L4-33K splice site, in the same transcript. The phenotype observed with this mutant was therefore more pronounced: when this transcript was expressed in N52.E6-Cre cells, splicing

of L4-33K was perturbed, and protein expression was lacking (Fig. 3-11A and B, lane 8). This phenotype was not rescued when cells were co-transfected with WT L4-22K protein (Fig. 3-11A and B, lane 9). This suggests a critical role for these RNA residues in the *cis*-regulatory sequences of the internal splice site of L4-33K. In accordance with this hypothesis, the histidine residues sit very close to the acceptor site of the L4-33K intronic sequences (Fig 3-1C), where the putative polypyrimidine tract would sit. The *cis*-acting elements of the L4-33K protein required for splicing activity will be discussed in the next chapter.

It is interesting to note that L4-22K mRNA levels were almost undetectable for WT infection as compared to the other highly abundant Ad late mRNAs (Fig. 3-8). By using an antibody against the common regions of L4-22K and L4-33K proteins, we detected significantly higher amounts of L4-33K protein compared to L4-22K protein in WT infection that proportionally corresponded to the mRNA levels (unpublished data). These observations suggest that although L4-22K is multifunctional and critical for Ad infection, it is not found in abundance as compared to the total pool of Ad late viral proteins. We propose a feedback loop, wherein L4-22K regulates accumulation of different L4 transcripts, including its own, thereby regulating the levels of L4 gene products. As L4-22K protein levels increase, L4-33K pre-mRNA splicing will be stimulated which would effectively reduce the accumulation of L4-22K mRNAs. Since L4-33K itself regulates Ad late mRNA splicing (75, 86, 90), this finely tunes the temporal switch of gene expression patterns during the late phase of infection. Finally, the results of gel mobility shift assays with extracts prepared from H166Q/H170Q mutant-infected cells demonstrates that over-expression of L4-22K alters the binding of packaging proteins to the packaging sequences (Fig. 4). We speculate that L4-22K

protein levels are low and precisely regulated to promote the appropriate formation of a packaging assembly complex on the viral packaging domain.

Chapter 4: Characterization of the Critical Elements Responsible for L4-22K-mediated Splicing of the L4-33K pre-mRNA

Abstract

The Ad major late transcription unit (MLTU) is an excellent example of an alternatively spliced gene, and serves as a useful model to study the mechanisms of regulation of alternative pre-mRNA splicing. The temporal change in MLTU alternative splicing results in activation of 3' splice sites of transcripts that are only needed during the late stages of infection; this activation is tightly controlled by both viral proteins and cellular factors in a cooperative manner. The L4-22K protein controls viral gene expression at the post-transcriptional level and regulates the accumulation of the L4-33K protein at the level of alternative pre-mRNA splicing. This places the L4-22K protein as the master regulator of late gene expression, initiating an activation cascade of late viral protein expression, while early genes are being turned off. Here we show that L4-22K activates L4-33K splicing through an enhancer element that overlaps the 3' ss of the transcript. This element is necessary, but not sufficient, for activation of L4-33K splicing in the presence of the L4-22K protein. Furthermore, we found that regulatory sequences located in the second exon of the L4-33K pre-mRNA prevent constitutive splicing of the transcript. We propose that at least two distinct ESSs are present in the 3' exon and contain binding sites for splicing repressors. We hypothesize that the L4-22K protein regulates viral mRNA production by a direct protein interaction with a cellular splicing regulator. Based on our findings, it seems likely that this splicing regulator may act as a repressor that binds to the highly conserved Py tract of the L4-33K pre-mRNA. L4-22K could relieve such repression by binding the repressor element and

displacing it from the L4-33K pre-mRNA. To this end, we developed a proteomics assay in search for L4-22K binding partners.

Results

The L4-33K enhancer element is the major element mediating L4-22K-activation of pre-mRNA splicing.

We demonstrated previously that L4-22K regulates Ad gene expression at the level of alternative splicing, with the L4-33K pre-mRNA as the major target (Fig. 4-1). This gives L4-22K a significant role as the master regulator of Ad late gene expression. To better understand its mechanism of action, we analyzed the effect of L4-22K on L4-33K pre-mRNA splicing in more detail. To map the critical element(s) responsible for the splicing activity of L4-22K, we analyzed the mRNA and protein levels of L4-33K in an *in vivo* splicing assay using chimeric transcripts where intronic sequences in the L4-33K pre-mRNA were replaced by the intronic sequences from the SV40 small T (sT) antigen (referred to as SV40) pre-mRNA (Fig. 4-2). All results were normalized to GFP mRNA and GFP protein levels, as described in Chapter 3. It should be noted that SV40 splicing is constitutive and is not enhanced by the L4-22K protein in the context of a plasmid that contains the entire SV40 sT pre-mRNA (data not shown). To confirm that SV40 splicing is not activated by the L4-22K protein, we generated a 33K-SV40-33K plasmid that replaces the entire L4-33K intron with the entire SV40 sT intron (Fig. 4-2A). Constitutive expression of L4-33K was observed with no response to L4-22K (Fig 4-2B, lanes 9 and 10).

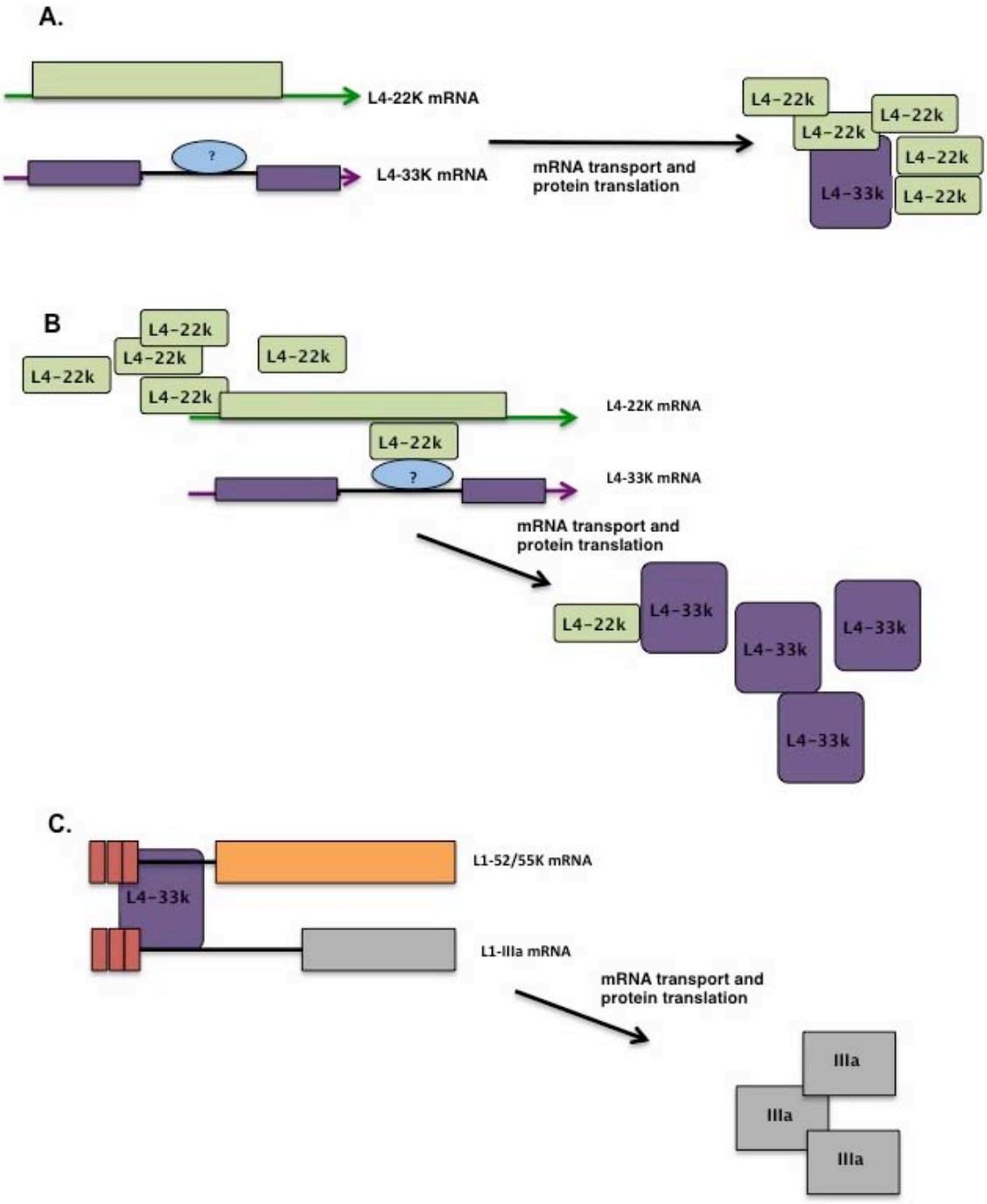

A.
L4-22K mRNA
?
L4-33K mRNA
mRNA transport and
protein translation
L4-22k
L4-22k
L4-22k
L4-22k
L4-33k
L4-22k
B
L4-22k
L4-22k
L4-22k
L4-22k
L4-22k
L4-22K mRNA
L4-22k
?
L4-33K mRNA
mRNA transport and
protein translation
L4-22k
L4-33k
L4-33k
L4-33k
L4-33k
C.
L1-52/55K mRNA
L4-33k
L1-IIIa mRNA
mRNA transport and
protein translation
IIIa
IIIa
IIIa

Figure 4-1 Proposed model for the role of L4-22K as the master regulator of gene expression.

Panel A. The L4-22K protein accumulates at the onset of major late promoter activation, before synthesis of any other viral late products. Low levels of L4-33K protein synthesized from the L4P are observed (see text).

Panel B. As L4-22K protein levels increase, L4-33K pre-mRNA splicing is stimulated, reducing the accumulation of L4-22K mRNAs.

Panel C. The L4-33K protein increases in abundance and regulates Ad mRNA splicing of the L1 transcript.

This result is in contrast to transfection of pcDNA3-L4-22K^{-} which resulted in low levels of L4-33K protein expression (Fig. 4-2B, lane 6) due to inefficient splicing of the L4-33K pre-mRNA (Fig. 4-2C, lanes 2); L4-33K expression was significantly increased when L4-22K protein was expressed *in trans* (Fig. 4-2B, lane 8, Fig. 4-2C, lane 3). As described in Chapter 3, plasmid pcDNA3-L4-22K^{-} contains the genomic sequences encoding both L4-22K and L4-33K proteins, with a stop codon in the L4-22K reading frame. Expression levels of pcDNA3-22K, pcDNA3-33K, and pcDNA3-L4 are shown as protein controls (Fig. 4-2B, lanes 2, 3, and 4). Note that transfection of plasmid 33K-SV40-33K, with or without L4-22K present, resulted in accumulation of only the L4-33K pre-mRNA at the expense of the L4-22K pre-mRNA (Fig. 4-2C, lanes 4 and 5). This data suggests that the SV40 sT intron is constitutively spliced, and that there are elements within the L4-33K intron that modulate regulated splicing by the L4-22K protein.

To determine if the putative elements within the L4-33K intron required for L4-22K-activated splicing are found near the acceptor or the donor splice site, we generated two hybrids: hybrid 1 contains the first third of the L4-33K intron (the first 53 nt) fused to the last 150 nt of the corresponding sequences in the SV40 sT intron, while hybrid 2 contains the reverse configuration (Fig. 4-2A). Note that the SV40 sequences disrupt the L4-22K open reading frame. Transfection experiments with hybrid 1 showed that replacement of the 150 nt proximal to the 3' ss of the L4-33K intron with SV40 sequences (hybrid 1) resulted in constitutive splicing of the L4-33K pre-mRNA (Fig. 4-2C, lanes 6 and 7) and expression of the L4-33K protein (Fig. 4-2B, lanes 11 and 12). In contrast, the vector that retains the native L4-33K 3' proximal sequences also retained L4-22K-regulated splicing (Fig. 4-2C, lanes 8 and 9) and L4-33K protein expression (Fig. 4-2B, lanes 13 and 14). These results

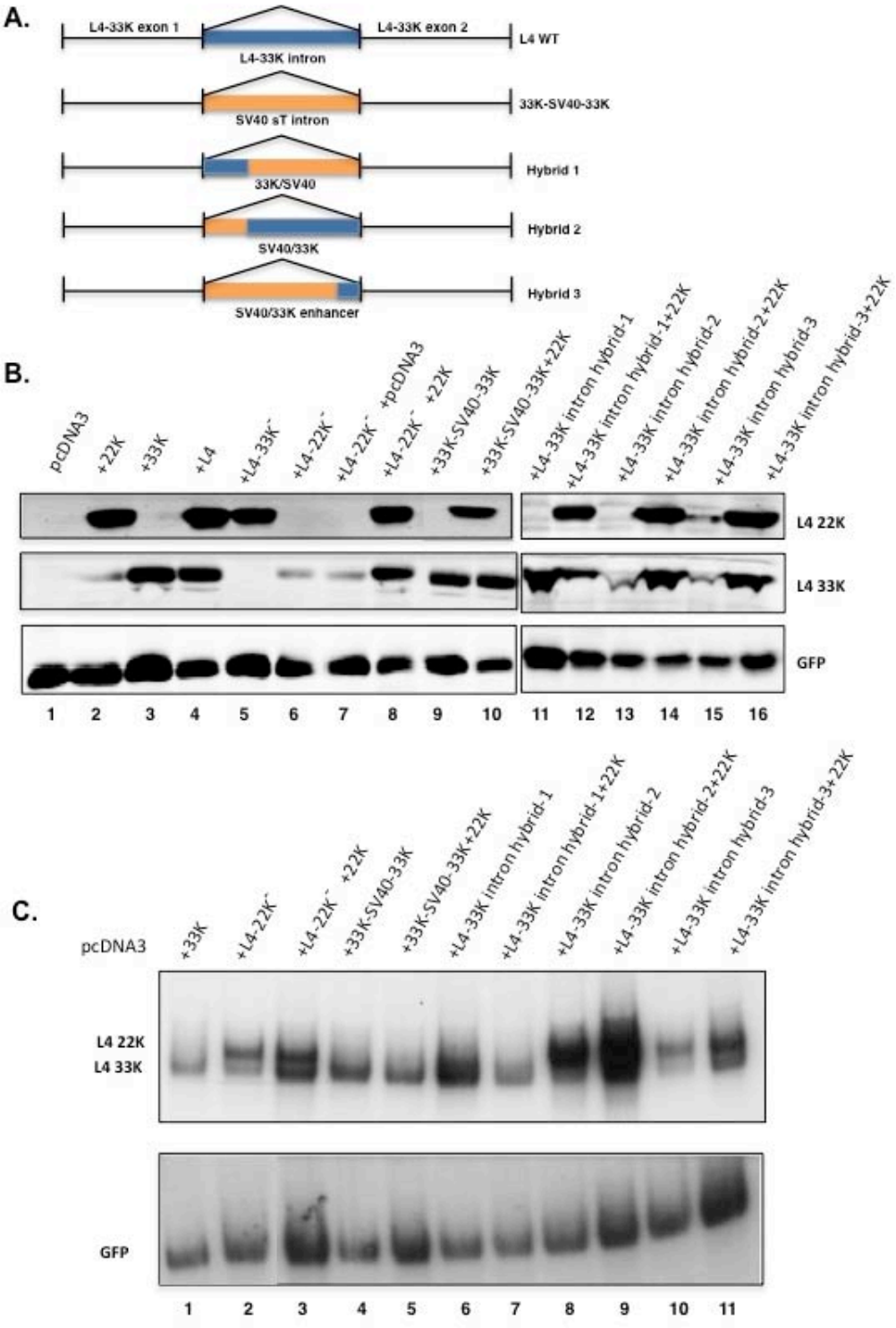

A.
L4-33K exon 1
L4-33K exon 2
L4 WT
L4-33K intron
33K-SV40-33K
SV40 sT intron
Hybrid 1
33K/SV40
Hybrid 2
SV40/33K
Hybrid 3
SV40/33K enhancer
B.
pcDNA3
+22K
+33K
+L4
+L4-33K-
+L4-22K-
+L4-22K-
+pcDNA3
+L4-22K-
+22K
+33K-SV40-33K
+33K-SV40-33K+22K
+L4-33K intron hybrid-1
+L4-33K intron hybrid-1+22K
+L4-33K intron hybrid-2
+L4-33K intron hybrid-2+22K
+L4-33K intron hybrid-3
+L4-33K intron hybrid-3+22K
L4 22K
L4 33K
GFP
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
C.
pcDNA3
+33K
+L4-22K-
+L4-22K-
+22K
+33K-SV40-33K
+33K-SV40-33K+22K
+L4-33K intron hybrid-1
+L4-33K intron hybrid-1+22K
+L4-33K intron hybrid-2
+L4-33K intron hybrid-2+22K
+L4-33K intron hybrid-3
+L4-33K intron hybrid-3+22K
L4 22K
L4 33K
GFP
1 2 3 4 5 6 7 8 9 10 11

Figure 4-2 The *cis*-acting elements within the L4-33K pre-mRNA necessary for regulated splicing are within the 3' proximal intron nucleotides.

Panel A. Depiction of the SV40/L4-33K intron chimeras. Note that the intronic and exonic sequences are not drawn to scale.

Panel B. N52.E6-Cre cells were cotransfected with expression plasmids for WT L4 proteins or plasmids expressing chimeras of SV40/L4-33K intronic sequences and an expression plasmid for GFP and harvested 48h after transfection. Cells were used for western blot analysis; all samples were normalized to GFP protein or mRNA levels.

Panel C. Samples from panel B were used for Northern blot analysis; all samples were normalized to GFP protein or mRNA levels.

show that the 3' 150 nt of the L4-33K intron provide for L4-22K-regulated splicing. To delineate further the minimal L4-33K intron sequences that are required for L4-22K-regulated splicing, we replaced most of the L4-33K intron with the SV40 intron and kept only the last 35 nt of the L4-33K intron intact (Fig 4-2A, hybrid 3). These 35 nt include the BPS, Py tract, and AG dinucleotide which we refer to as the L4-33K enhancer element. Transfection of hybrid 3 alone resulted in low levels of L4-33K protein expression (Fig. 4-2B, lane 15), which was augmented when L4-22K was provided *in trans* (Fig. 4-2B, lane 16). The subsequent increase in L4-33K protein expression when hybrid 3 was cotransfected with pcDNA3-22K directly correlates with activated splicing of the L4-33K pre-mRNA (compare Fig. 4-2C, lanes 10 and 11). We also performed a ClustalW multiple sequence alignment on the DNA sequences of the proximal 3' intronic sequences and interestingly found that this putative enhancer element is highly conserved amongst various serotypes of human Ad (Fig. 4-3). Of particular interest is the highly conserved Py tract (see Discussion). We conclude that the L4-33K enhancer element is an essential element mediating L4-22K activation of splicing *in vivo.* However, based on our current findings thus far, we cannot rule out the presence of another *cis*-acting element elsewhere within the pre-mRNA.

The 3' exon, but not the 5' exon, of L4-33K contains elements necessary for L4-22K-regulated splicing.

To test if the L4-33K enhancer element is the minimal sequence mediating L4-22K-regulated splicing of the L4-33K pre-mRNA, we generated the reverse chimeric constructs, where we replaced the entire SV40 intron (and retained the exonic sequences) with the entire L4-33K intron and tested this construct in our splicing assay, with and without L4-22K

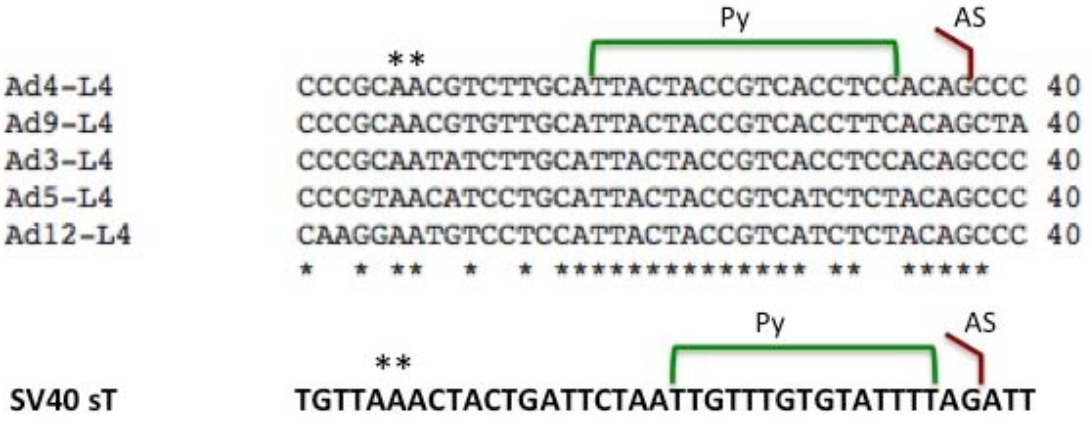


Figure 4-3 The polypyrimidine tract of the L4-33K pre-mRNA is highly conserved. ClustalW DNA sequence alignment of the proximal 3' intronic sequences of the L4-33K pre-mRNA showing different representative serotypes of various species of human Ad. (Bottom) depicts the corresponding sequences for the SV40 sT intron. Bold asteriks (top) indicate the two A residues used as the BPS in L4-33K splicing. Green lines and Py designate the putative polypyrimidine tract. Red lines and AS designate the acceptor site. The asteriks (bottom) indicate positions which have a fully conserved residue. The Py and BPS were predicted using Corvelo, et al. PLoS Computational Biology (2010).

expression. The results were inconclusive and led us to believe that there may be elements present within the exons of the L4-33K pre-mRNA that mediate L4-22K-activated splicing since constructs that lack the L4-33K exons showed aberrant splicing and protein expression (data not shown). To determine if either, or both, of the L4-33K exons contain regulatory elements that play a role in L4-33K pre-mRNA splicing, we generated a series of new chimeric constructs (Fig. 4-4). Cytoplasmic RNA was analyzed by Northern blot to examine pre-mRNA splicing. Note that these chimeras disrupt the L4-33K reading frame and therefore protein analysis is not possible (Fig. 4-4A). Exon hybrids 1 and 3 replaces the first or second half of the 5' exon with the corresponding exonic sequences from the first exon of SV40, respectively (Fig. 4-4A). Transfection of hybrid 1 or hybrid 3 alone resulted in a significant decrease in L4-33K pre-mRNA splicing (Fig. 4-4B, lanes 3 and 7, respectively), similar to that observed with transfection of plasmid pcDNA3-L4-22K^{-} (Fig. 4-4B, lane 1). Co-expression of plasmid pcDNA3-22K with hybrid 1 or hybrid 3 complemented the chimeric plasmids to rescue the accumulation of L4-33K pre-mRNA to levels similar to those observed with plasmids containing all L4 sequences (Fig. 4-4B, compare lanes 2, 4, and 8). Exon hybrids 4 and 2 replace the first or second half of the 3' exon with the corresponding sequences from SV40 (Fig. 4-4A). To our surprise, transfection of each of these vectors alone resulted in accumulation of spliced L4-33K mRNA, even in the absence of the L4-22K protein (Fig. 4-4B, lanes 5 and 9); splicing was further activated when the L4-22K protein was expressed (Fig. 4-4B, lanes 6 and 10). Interestingly, these results suggest that there may be at least two important *cis*-acting elements within the 3' exon for activated splicing (one in the first half, and one in the second half of the exon) and no regulatory

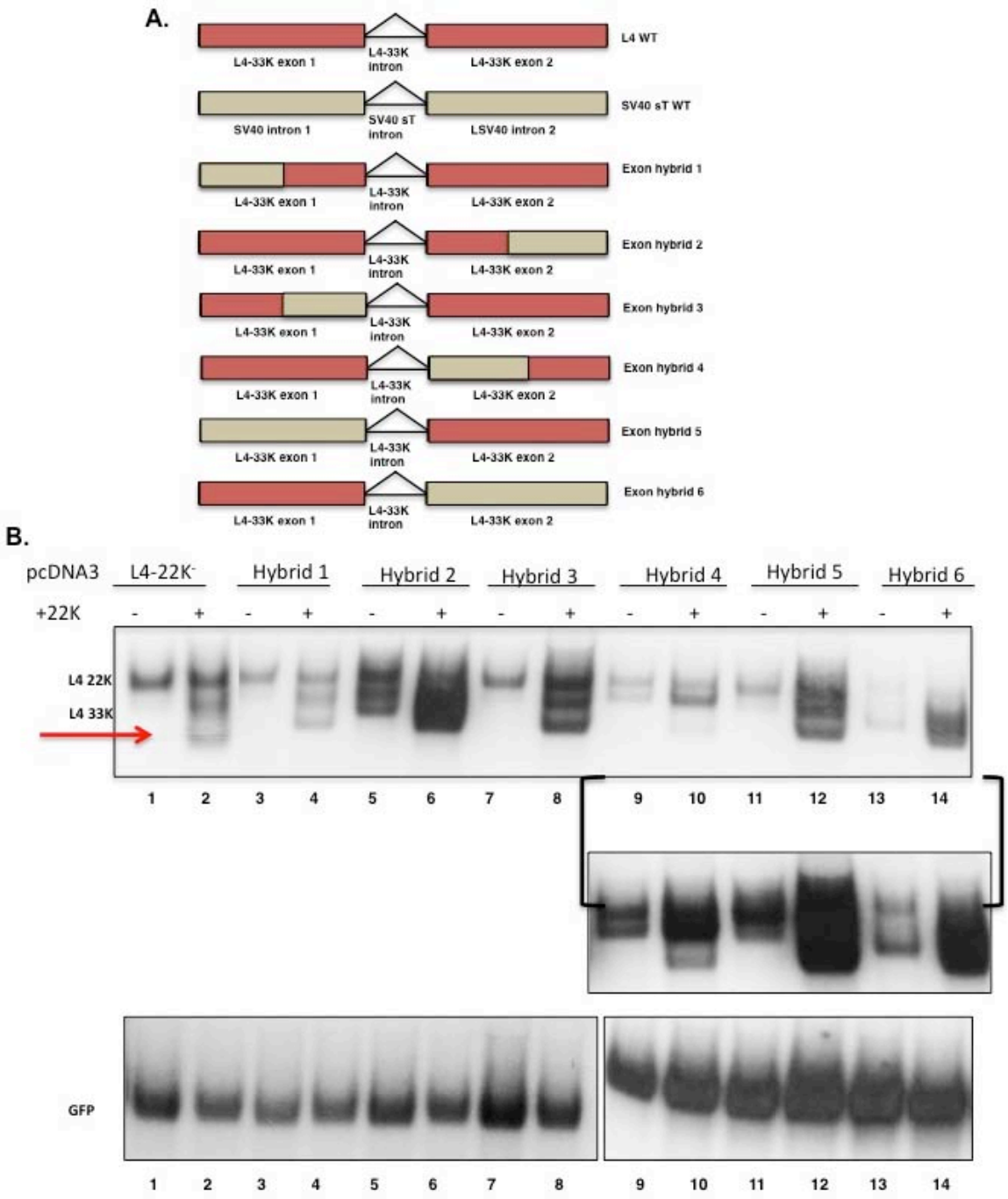

A.
L4 WT
L4-33K exon 1
L4-33K intron
L4-33K exon 2
SV40 sT WT
SV40 intron 1
SV40 sT intron
LSV40 intron 2
Exon hybrid 1
Exon hybrid 2
Exon hybrid 3
Exon hybrid 4
Exon hybrid 5
Exon hybrid 6
B.
pcDNA3
L4-22K-
Hybrid 1
Hybrid 2
Hybrid 3
Hybrid 4
Hybrid 5
Hybrid 6
+22K
L4 22K
L4 33K
GFP
1 2 3 4 5 6 7 8 9 10 11 12 13 14

Figure 4-4 The 3' exon of L4-33K contains elements required for regulated splicing by the L4-22K protein.

Panel A. Schematic of the SV40/L4-33K exon chimeras. Note that the intronic and exonic sequences are not drawn to scale.

Panel B. N52.E6-Cre cells were cotransfected with expression plasmids for WT L4 proteins or plasmids expressing chimeras of SV40/L4-33K exonic sequences and an expression plasmid for GFP and harvested 48h after transfection. Cells were used for Northern blot analysis with a ^{32}P-labeled probe designed around nt 26195-27086. The red arrow indicates the band corresponding to the L4-22K mRNA from transfection of the pcDNA3-22K plasmid. The insert represents a longer exposure of the Northern blot for the specified lanes. All samples were normalized to GFP mRNA levels and shown for lower exposure (lanes 1-8) and higher exposure (9-14).

elements present within the 5’ exon. Alternatively, there could be a putative repressor element that spans the junction of the two chimeras (see Discussion). To test this hypothesis, we generated a chimera that replaces the entire L4-33K 5’ exon or 3’ exon with exonic sequences from SV40, hybrids 5 and 6, respectively (Fig. 4-4A). Transfection of hybrid 5 alone resulted in a decrease in L4-33K pre-mRNA accumulation (Fig. 4-4B, lane 11) and a subsequent increase when the L4-22K protein was expressed (Fig. 4-4B, lane 12). In contrast, transfection of hybrid 6 alone resulted in basal levels of splicing of the L4-33K pre-mRNA (Fig. 4-4B, lane 13); an increase in L4-33K pre-mRNA splicing was observed when L4-22K was provided in *trans* (Fig. 4-4B, lane 14; see insert). We conclude that the role of the L4-22K protein in regulating L4-33K alternative pre-mRNA splicing is mediated, in part, by *cis*-acting regulatory sequences found in the 3’ exon, but not the 5’ exon. These sequences may represent a putative binding site for a repressor element, that is absent from SV40 sT exonic sequences. This repressor element, together with the L4-33K intronic enhancer element (Fig. 4-2, Fig. 4-3), is necessary for L4-22K-regulated splicing.

Proteomic analysis of L4-22K-interacting proteins.

We, and others, have characterized several critical functions of the L4-22K protein including regulation of viral gene expression and a critical role in the encapsidation of the viral genome by cooperative binding of the IVa2 and L4-22K proteins to the packaging sequences. Despite these characterized functions, little is known about the mechanisms of action of this protein. Specifically, to date, there are no known cellular proteins that interact with the L4-22K protein during the course of Ad infection. We hypothesize that L4-22K functions by specific interaction with a cellular splicing component(s) to regulate L4-33K

pre-mRNA splicing and took a proteomics approach to uncover its identity. First, we generated L4-22K expression constructs with N-terminal or C-terminal tandem Strep-HA tags (pcDNA3+5'22K-Strep-HA and pcDNA3+3'22K-Strep-HA, respectively) and tested the tagged proteins for proper function. We cotransfected each of the tagged constructs with pcDNA3-L4-22K^{-} and analyzed splicing activity using the transient splicing assay (Fig. 4-5). These tagged proteins functioned like WT L4-22K in that they expressed normal levels of L4-22K protein (Fig. 4-5 lanes 8 and 9) and they were able to rescue L4-33K expression to levels similar to WT L4-22K (Fig. 4-5 compare lanes 6, 8, and 10). Next, we used these tagged constructs to generate recombinant Ad5 viruses that express the tagged L4-22K proteins under the control of the CMV promoter/enhancer in place of E1. We utilized E1 replacement viruses under the control of the CMV promoter/enhancer for higher levels of protein expression since the L4-22K protein normally is expressed at low levels during Ad infection and this may limit our ability to detect relevant protein complexes. Note that these recombinant viruses lack E1A and E1B, but can be propagated successfully by growing the viruses in N52.E6 cells which express these Ad5 E1 proteins (117). Additionally, these viruses express both tandem tagged L4-22K proteins and WT-L4-22K protein in the context of Ad5 infection.

Both the 5' and 3' L4-22K-tagged viruses were used in a tandem affinity tag pull-down assay. The virus expressing the 5' tagged L4-22K protein gave us the most consistent results in replicate experiments (data not shown) and was used in subsequent experiments. N52.E6-Cre cells were mock-infected or infected with Ad5ΔE1-5'22K-Strep-HA virus (referred to as 5'-22K-Strep-HA). Twenty hours post-infection, cells lysates were prepared

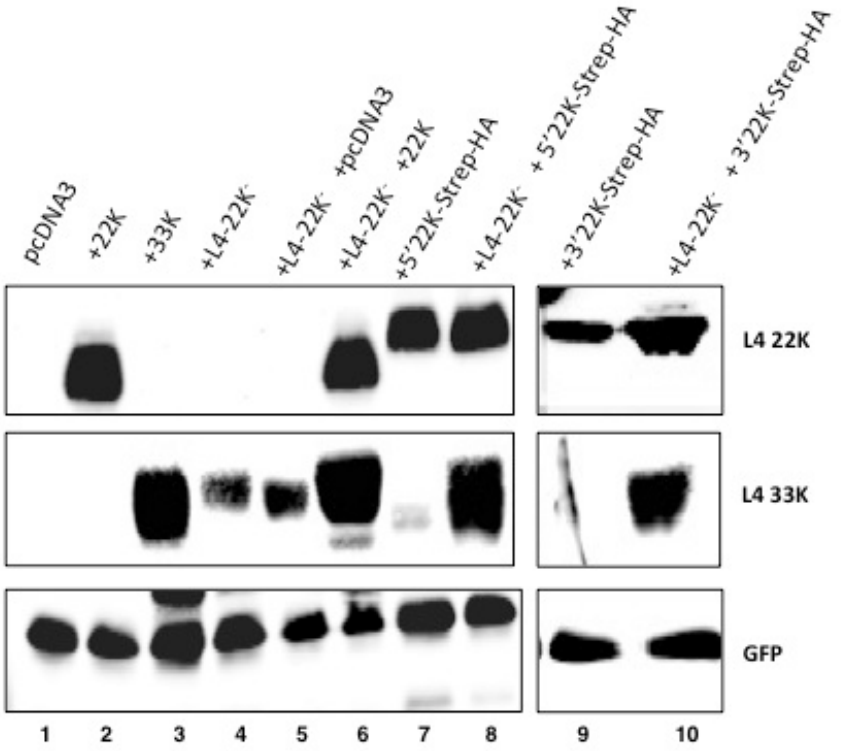


Figure 4-5 Epitope-tagged L4-22K expressing plasmids are functional for splicing activity. N52.E6-Cre cells were cotransfected with plasmids expressing WT L4 proteins or plasmids expressing epitope-tagged L4-22K proteins and an expression plasmid for GFP and harvested 48 h after transfection. Whole cell extracts were used for western blot analysis and normalized to GFP protein levels.

and incubated with Strep-tactin beads followed by elution with desbiotin. A second pull-down assay was performed with anti-HA-agarose beads, followed by boiling the samples in 2% SDS buffer. Starting lysates and proteins pulled-down in these assays were analyzed by western blot (4-6A) and silver stain analyses (4-6B). Western blot analysis showed that tagged L4-22K protein was pulled-down at both steps, as depicted by blotting with antibodies against L4-22K, Strep, and HA (Fig. 4-6A, compare lysates and IP columns). Prior to submitting samples for proteomic analysis, we visualized protein(s) interacting with L4-22K by silver stain analysis, at the first Strep pull-down step, and second HA pull-down step (Fig. 4-6B, left and right panels respectively). Note that following each step, silver stainable quantities of the L4-22K protein were present in the 5'-22K-Strep-HA-infected samples but not the mock-infected sample (Fig. 4-6B, blue arrow, left and right panels). Even though equal amounts of starting material were used in these analyses, we observed a greater number of unidentified proteins potentially interacting with the L4-22K protein during the first pull-down step, that were not present in mock-infected samples (Fig. 4-6B, left panel). Following pull-down with anti-HA-agarose beads, there were fewer proteins visible with both mock- and 5'-22K-Strep-HA-infected samples, but we were still able to distinguish at least four distinct bands evident in the 5'-22K-Strep-HA-infected samples that were not seen in mock-infected samples (Fig. 4-6B, right panel, red arrows). The identity of these proteins is currently being evaluating by SILAC analysis (131).

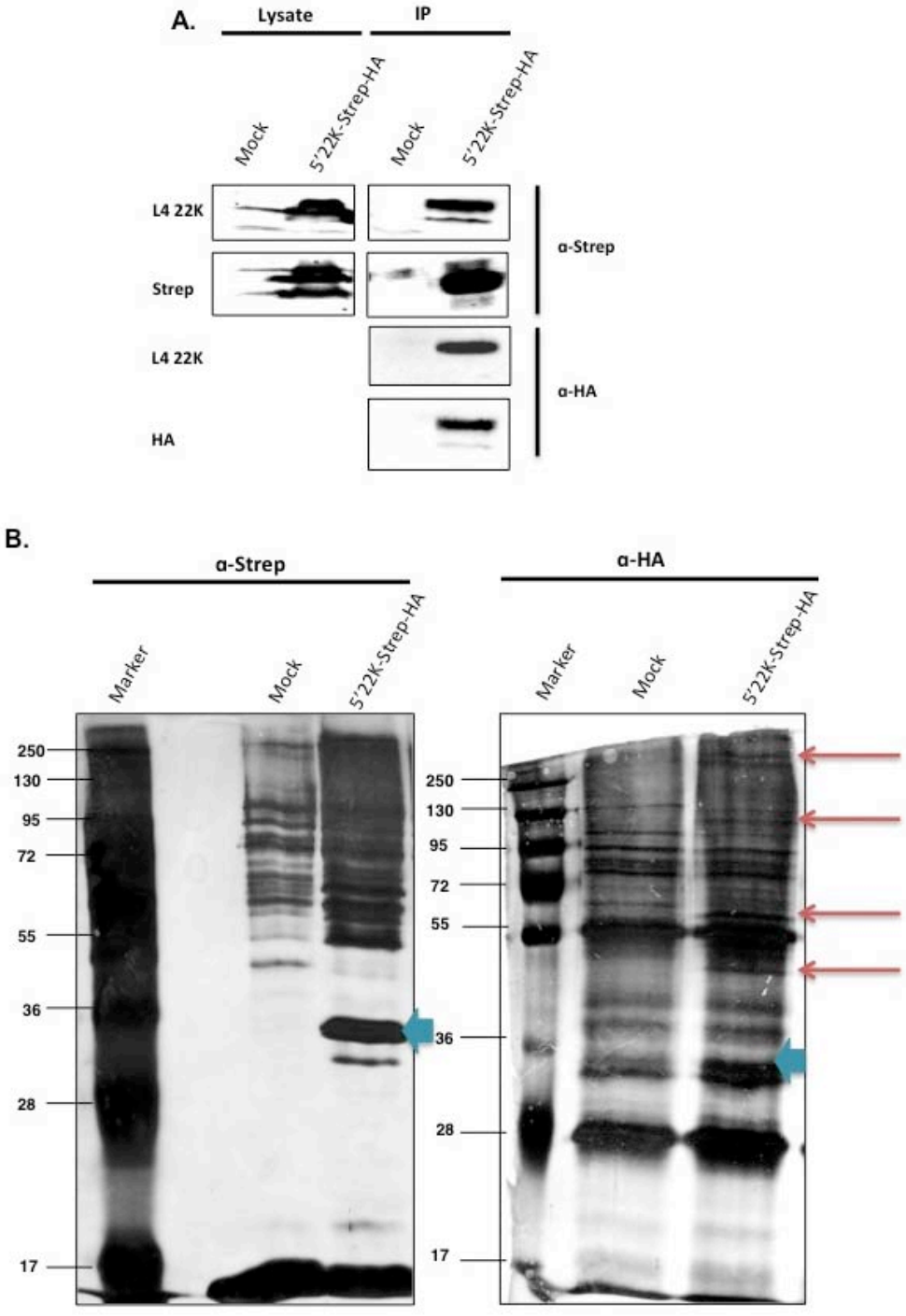

A.
Lysate
IP
Mock
5'22K-Strep-HA
Mock
5'22K-Strep-HA
L4 22K
Strep
α-Strep
L4 22K
HA
α-HA
B.
α-Strep
α-HA
Marker
Mock
5'22K-Strep-HA
Marker
Mock
5'22K-Strep-HA
250
130
95
72
55
36
28
17
250
130
95
72
55
36
28
17

Figure 4-6 Proteomic analysis of L4-22K and interacting proteins.

Panel A. N52.E6-Cre cells were infected with a Strep-HA-tagged L4-22K virus or mock-infected and harvested at 20 hpi. Total lysates were prepared and used in a two-step pull down assay using Strep-tactin conjugated beads and desthiobiotin elution followed by HA-agarose beads and boiling in 2% SDS. Total lysates and pulled-down proteins were separated on SDS-PAGE and analyzed by western blotting

Panel B. Pulled-down proteins from panel A were separated on SDS-PAGE and analyzed by silver stain analysis .The red arrows point to protein(s) pulled down with L4-22K not present in mock-infected samples. Blue arrow heads point to the L4-22K protein.

Discussion

Based on these results, and results published by others, a working model emerges for the temporal regulation of Ad late gene expression and the transition from the early to the late stages of infection, in which the L4-22K protein functions at several steps. This model assumes that the L4-22K protein plays a role in the activation of the MLTU (80). Following DNA replication, the L4P directs expression of basal levels of the L4-22K and L4-33K mRNAs, independent of the MLP (88) (Fig. 4-1A). As the levels of the L4-22K protein accumulate, we observe an increase in L4-33K protein levels as a result of L4-22K-mediated activation of splicing of the L4-33K pre-mRNA (Fig. 4-1B). Since both proteins are produced from the same transcript, an increase in translation of the L4-33K spliced transcript concomitantly leads to a decrease in L4-22K expression. Accumulation of the L4-33K protein occurs at a time during infection when all other late transcripts are being produced from the MLP. The L4-33K protein promotes the accumulation of specific late viral proteins by acting as a virus encoded splicing factor (Fig. 4-1C). At the same time, both L4-22K and L4-33K shut-off this negative feedback loop by preventing the p53-regulated activation of the L4P (89).

To gain a better understanding of the mechanism of action behind L4-22K function, we sought to map the critical element(s) responsible for L4-22K-mediated splicing of the L4-33K pre-mRNA. We found that the L4-33K enhancer element, composed of the BPS, Py, and AS, is a critical sequence mediating L4-22K-activation of L4-33K pre-mRNA splicing *in*

vivo (Fig. 4-2). Interestingly, we found the Py to be highly conserved amongst different serotypes of human Ad (Fig. 4-3) raising the possibility that this sequence serves as a putative binding site for a splicing regulator and/or the L4-22K protein. Replacing the 3' intronic sequences with corresponding sequences from the constitutively spliced SV40 sT intron, led to constitutive splicing of the L4-33K pre-mRNA (Fig. 4-2). We were puzzled to find that the opposite configuration, where SV40 sT intronic sequences were replaced with L4-33K intronic sequences, did not lead to L4-22K-regulated splicing (data not shown). In fact, the results were inconclusive which led us to believe that other elements, present in the exonic sequences of L4-33K, also mediate regulated splicing. To this end, we generated a series of chimeras replacing a portion, or all, of each of the two exons of L4-33K, with corresponding sequences from SV40 sT. We found that the 5' exon is dispensable for L4-22K-mediated activation of splicing but that sequences present in the 3' exon may represent binding sites for silencer(s) elements (Fig. 4-4). Finally, we initiated experiments to identify proteins that interact with L4-22K with the goal of determining how this viral protein functions as a regulator of splicing (Fig. 4-5). and used in a pull-down assay to find potential binding partners. We found at least four distinct proteins were pulled-down when tagged-L4-22K protein was expressed that were not present in mock-infected cell samples (Fig. 4-6).

L4-33K is a well-understood virus-encoded splicing factor and serves as an excellent model to study the similar role of the L4-22K protein. Briefly, the L1 unit contains a common 5' ss, the TPL, that can be joined to two alternative 3' ss, resulting in the formation of the 52/55K (proximal 3'ss) or IIIa (distal 3' ss) mRNAs (Fig. 1-4). Early during infection, IIIa splicing is absent because hyperphosphorylated SR proteins bind to the *cis*-acting ISS element, termed 3RE, and block U2 snRNP recruitment. Phosphorylated SR proteins are

needed for spliceosome assembly, while dephosphorylated SR proteins are required during splicing catalysis (108), so maintaining the SR proteins hyperphosphorylated maintains their activity as splicing repressors. In addition, the IIIa 3'ss contains a non-consensus type of Py tract and U2AF binds inefficiently to this element. As infection progresses, virus-induced SR protein dephosphorylation by E4-ORF4 lifts the repressive effect of SR proteins on IIIa pre-mRNA splicing. In addition, the L4-33K protein acts as a virus-induced Py binding factor to aid in the recruitment of U2 snRNP to the IIIa BPS. L4-33K acts on the 3VDE, an enhancer element, located in the 3' ss of the IIIa pre-mRNA via a U2AF-independent mechanism. The current working model does not show that L4-33K binds to RNA, so it seems likely that L4-33K acts in cooperation with a cellular splicing factor (reviewed in (75)).

We postulated that like the L1 IIIa pre-mRNA, similar silencer or enhancer elements may be located within the L4-33K pre-mRNA to direct L4-22K-regulated splicing. We sought to define these regulatory sequences by replacing them with those of the known constitutively spliced intron of SV40 small T antigen (SV40 sT). Our data shows that replacing the entire L4-33K intron with that of SV40, generates a constitutively spliced transcript (Fig. 4-2B, lanes 9 and 10, Fig. 4-2C, lanes 4 and 5). Furthermore, we showed that SV40 is constitutively spliced *in vivo*, with no effect from L4-22K (data not shown). We also found that the first 168 nt of the L4-33K intron are not required for L4-22K-activated splicing, while the last 35 nt mediate, in part, L4-22K-regulated splicing (Fig. 4-2B, lanes 15 and 16, Fig. 4-2C, lanes 10 and 11). We named this element the enhancer element of L4-33K and it has similar sequence composition as the 3VDE of L1 IIIa (115) comprising the BPS, Py tract, and AS (Fig. 4-3). Constitutively-active L1 52/55K and cellular β-globin pre-mRNAs, for example, have long Py tracts of 18 or 19 nt (90), whereas L1 IIIa and L4-33K

pre-mRNAs have short, non-consensus Py tracts (Fig. 4-3). The U2AF splicing factor is composed of two subunits, $U2AF^{65}$ and $USAF^{35}$, that bind the Py tract and the AG dinucleotide at the AS, respectively. Mühlemann et al. showed that the Py tract of L1 IIIa does not bind $USAF^{65}$ efficiently *in vitro* (132). In addition, this group found that splicing of pre-mRNAs that contain long prototypical Py tracts and bind U2AF efficiently, is repressed in Ad-infected nuclear extracts while splicing of 3' splice sites which have a weak Py sequence context are enhanced in Ad-infected nuclear extracts (133). These results suggest that transcripts that contain these non-consensus Py tracts, seen at late times of Ad infection, require a virus-encoded factor to activate their splicing. This model for a different splicing mechanism utilized in virus-encoded transcripts coincides with the idea that Ad causes a general inhibition of cellular mRNA production. One way to do this is by reducing splicing efficiency of consensus cellular pre-mRNAs. We recently showed that L4-33K is required for the accumulation of L1-IIIa and pVI mRNAs and proteins and that other Ad late mRNAs are not primary targets of L4-33K splicing activity (106). Thus, it seems logical that another viral encoded factor, L4-22K in this case, would be required to enhance splicing of some Ad late transcripts, through an $U2AF^{65}$- independent mechanism. Our data strongly suggests that the L4-22K protein acts, in part, via the enhancer element of the L4-33K pre-mRNA, which includes the BPS, Py tract, and AS (Fig. 4-3).

Interestingly, a ClustalW DNA sequence alignment of this region of different human Ad serotypes showed that the Py tract, in particular, is highly conserved (Fig. 4-3). We propose that L4-22K may act in one of two ways. First, L4-22K may replace $U2AF^{65}$ as the Py tract recognition factor and aid in the recruitment of other splicing factors to the spliceosome. This model requires binding of the pre-mRNA by L4-22K, an activity that has

not yet been reported. Alternatively, L4-22K may interact with the L4-33K pre-mRNA via a yet unidentified cellular splicing factor (see below). Second, L4-22K may relieve a splicing repressor that binds at the enhancer element. Therefore, we postulated that L4-22K acts directly via the Py tract by binding a known splicing regulator such as the polypyrimidine tract-binding protein (PTB). The PTB is an abundant and widely expressed RNA binding protein that belongs to the hnRNP family and acts as a regulatory repressor or activator of alternative splicing. These two activities of PTB are determined by the location at which PTB binds relative to target exons (134). The general trend is that splicing is repressed by PTB when PTB binds to the upstream intronic flank (3'ss) or within the 5' exon. Exons are activated for splicing by PTB when PTB binds the immediate downstream intronic flank (5'ss) between two exons. As a repressor on the Py, it acts in direct competition with $U2AF^{65}$. To determine if PTB cooperates with L4-22K to regulate L4-33K splicing, we took a targeted coimmunoprecipitation (coIP) approach. N52.E6-Cre cells were mock-infected or infected with 5'-Strep-HA-L4-22K virus and a coIP was performed using protein A agarose beads supplemented with anti-HA antibody, followed by western blot analysis. We probed for PTB but did not observe any interaction with L4-22K under our experimental conditions (data not shown). We also probed for $U2AF^{65}$ and did not observe any interaction with L4-22K (data not shown). The latter results make sense since our working model suggests that L4-22K acts on the enhancer element, possibly by direct or indirect binding to the Py tract, via a $U2AF^{65}$-independent mechanism. We cannot rule out the possibility that L4-22K relieves a repressor bound to the Py tract solely based on our results from the coIP experiment with PTB.

The idea that L4-22K binds to a repressor RNA element or repressor protein to activate L4-33K pre-mRNA splicing is consistent with the relief of splicing repression observed by replacing the L4-33K Py region with that of SV40 (Fig. 4-2). When the 3' 150 nt of the intron was replaced with those of SV40 sT, we observed constitutive splicing (Fig. 4-2B, lanes 11 and 12, Fig. 4-2C, lanes 6 and 7). Interestingly, sequence analysis of the Py tract of SV40 shows that it is also a non-consensus type of Py tract composed of short stretches of Py, similar to that of L4-33K, yet these SV40 sequences do not mediate L4-22K-activated splicing. Future experiments in which the BPS, Py tract, and AS of L4-33K are replaced with those of SV40 may allow for a more detailed model of regulated splicing of the L4-33K pre-mRNA.

Based on the intronic chimeric transcripts analyzed in this study, we can conclude that the 150 nt proximal to the 3'ss (comprising nt 54-203 of the intron) are important for L4-22K-mediated regulation of L4-33K pre-mRNA splicing. We can also conclude that the enhancer element is an essential sequence necessary for such activity. However, we cannot rule out the possibility that additional elements between nt 54 and 168 of the intron may also play a role in L4-22K-regulated splicing. The 3RE of L1 IIIa is located immediately upstream of the 3VDE and binds all tested members of SR family of splicing factors to repress splicing during early stages of Ad infection, as outlined above. Interestingly, E4-ORF4 activates L1 IIIa splicing by inducing SR protein dephosphorylation and requires the 3RE (135). This mechanism could potentially be a general trend for the basal activation of late transcript splicing since most have a 3'ss with a weak sequence context. It is plausible that a more detailed analysis of the nucleotide sequences directly upstream of the L4-33K enhancer element would reveal an ISS element that binds SR proteins and maintain low

levels of the L4-33K protein at earlier times of infection, in addition to the activation of splicing provided by the L4-22K protein.

Using the L1 unit as a model, we originally hypothesized that the L4-33K enhancer element was both necessary and sufficient for activation of L4-33K splicing in the context of L4-22K expression. However, results from some of the chimeras that we generated led us to believe otherwise. Even though we were able to successfully show that SV40 sT pre-mRNA is constitutively spliced *in vivo* and that splicing is not affected by L4-22K *in trans*, we were not able to show that replacing the SV40 intron with that of L4-33K results in L4-22K-regulated splicing (data not shown). In fact, our results were inconclusive. We therefore reasoned that other elements contained within the L4-33K exonic sequences also play a role in L4-33K pre-mRNA splicing. Regulation of alternatively spliced mRNAs requires crosstalk between multiple regulatory sites and *cis*- and *trans*-acting factors. In this report we focus on the *cis*- acting factors. There are two types of regulatory elements found in exons, as described previously. ESEs are splicing enhancers that are classically considered binding sites for SR proteins, which often function to activate splicing (but not always are evidenced by L1 IIIa-mediated repression by SR proteins). SR proteins recognize ESEs, which mediate splicing stimulation. ESSs are splicing silencers that are less well characterized, but seem to interact with negative regulators that often belong to the hnRNP family (109). In search for potential ESEs and ESSs, we replaced the first and second half of L4-33K exons 1 and 2, individually and in combination, with SV40 5' and 3' exonic sequences. We found that L4-33K exon 1 is dispensable for L4-22K-regulated splicing since a partial, or complete, replacement of this exon with the 5' exon of SV40 had no effect (Fig. 4-4, lanes 3, 4, 7, 8, 11, and 12). However, a partial replacement of exon 2 with the 3' exon of SV40 led to

constitutive splicing of the transcript (Fig. 4-4, lanes 5, and 9). This effect was more pronounced when the entire exon was replaced (Fig. 4-4, lane 13). We also found that addition of L4-22K to the splicing assay led to increased levels of L4-33K pre-mRNA accumulation, even in constitutively spliced transcripts (Fig. 4-4, lanes 6, 10, 14). The data strongly suggest the presence of one or more ESSs within exon 2 of the L4-33K pre-mRNA that prevents constitutive splicing of the transcript in the absence of L4-22K. By replacing the ESSs and flanking sequences with those of SV40, we removed the putative binding sites for repressors, and L4-33K splicing took place even in the absence of the L4-22K protein. It is interesting to note that the observed constitutive splicing in the absence of these *cis*-acting exonic sequences is not as pronounced as that observed with the intronic chimeric constructs (Fig. 4-2). Note that L4-22K and L4-33K are derived from the same parent transcript and we have previously shown that in the presence of L4-22K, both the L4-22K and the L4-33K pre-mRNAs are observed in equal amounts (Fig. 3-10) When intronic regulatory sequences were replaced with those of SV40, we observed a complete shift to 3'ss usage of the L4-33K pre-mRNA, at the expense of the L4-22K mRNA (Fig. 4-2C, lanes 4, 5, 6, 7). In contrast, replacing exonic regulatory sequences led to "normal" levels of L4-33K mRNA accumulation (Fig. 4-4C). These results suggest that the intronic enhancer element is the major L4-33K *cis*-acting element mediating L4-22K splicing and that the regulatory sequences located within exon 2 play auxiliary roles.

Based on our results it seems plausible that there are at least two ESSs located within the 3' exon of the L4-33K pre-mRNA. Splicing activators and repressors function by influencing the formation of the E and A complexes early in spliceosome assembly or regulate the secondary structure of RNA. ESEs can activate splicing by binding SR proteins

and by recruiting other factors to the 3' ss or 5' ss. Models for ESE function are based on the ability of the RS domain of SR proteins to mediate protein-protein interactions (136). For example, an SR protein bound to an ESE could activate splicing by recruiting $U2AF^{65}$ to weak Py tracts through mutual interaction of their RS domains with $U2AF^{35}$ at the AS serving as an intermediary. Splicing silencers work by interacting with negative regulators. Repression frequently requires cooperative repressor binding to multiple silencer elements located near one another (136). Cooperative repressor binding makes exons, splice sites, or enhancers inaccessible to their respective splicing factors. In addition, enhancers and silencers can act in a coordinated manner. For example, SR proteins that are bound to ESEs can promote exon definition by directly recruiting the splicing machinery through their RS domain and/or by antagonizing the action of nearby silencer elements (111). Our results suggest the presence of at least two ESSs within the L4-33K 3' exon, which possibly act in concert to repress L4-33K splicing in the absence of the L4-22K protein. Binding of a negative regulator to several sites around or within a silenced exon has previously been observed (111). Dimerization of bound negative regulators might cause relevant portions of the pre-mRNA to loop out, making them unavailable for binding by a splicing activator. It seems plausible that dimerization of the bound negative regulators becomes sterically blocked by the interaction of the L4-22K protein directly, or indirectly, to the enhancer element. Such mechanisms of cooperative regulation by both enhancers and silencers has been previously observed in many different systems (111, 136). Another possibility is that the repressive effect by the ESSs is distinct from the activation effect by the L4-22K protein. Our current findings cannot distinguish between these two models. However, we did observe increased accumulation of L4-33K pre-mRNA in the presence of L4-22K even in

constitutively spliced transcripts (Fig. 4-4) but this could be due to L4-22K interaction with the enhancer element that is left intact in these transcripts. At the very least, we can conclude that both the L4-22K protein acting on the enhancer element, and ESSs present in the 3' exon, mediate L4-33K pre-mRNA splicing and that replacing either the enhancer element, or the 3' exon with corresponding sequences of SV40, leads to constitutive expression of L4-33K. Further experiments need to be conducted to determine if these two functions are coordinated.

Based on our results, we used a series of ESSs and ISSs prediction tools to determine if there were any known regulator sequences located within the 3' exon of L4-33K (Table 1). The table shows only sequences predicted by more than one software tool since most of the computational analysis tools produced an unusually high number of predicted sites (data not shown). Interestingly, we found two major sets of predicted regulatory sequences located in exon 2. One is located in the first half of exon 2 (~nt 616-621) and the other is located in the second half of exon 2 (~nt 687-891) (Table 1, Fig. 4-4A). However, at least three groups of sequences located within L4-33K exon 1 were predicted to be ESSs, even though our data does not suggest the presence of such sequences in this exon. This finding is not surprising for several reasons. First, ESSs are variable in sequence (136). Most transcripts contain multiple ESEs/ISEs and ISSs/ESSs and are bound by different combinations of regulatory proteins that can antagonize one another directly or indirectly (108). Second, SR proteins act in a substrate-specific manner by binding to cognate ESEs, which consist of degenerate sequence motifs. Degenerate motifs allow for overlap in binding, therefore specificity results from combinatorial effects by different SR protein levels, binding affinities, and specific interactions with other proteins making it very difficult to predict binding sites based on

Table 4-1. Predicted exonic splicing silencer sequences

Exon	Position	Sequence	Source
Exon 1	77-82	AGGTTT	(1,2,5)
	78-83	GGTTTG	
	79-84	GTTTTG	
	80-85	TTTTGG	
Exon 1	81-86	TTTGGA	(4)
Exon 1	104-111	TGATGGAA	(3)
Exon 1	105-110	GATGGA	(4)
Exon 1	269-274	CCGTTC	(1,2,5)
	270-275	CGTTCG	
	289-294	AGATGG	
	290-295	GATGGG	
Exon 2	616-621	GCTTAG	(1, 2,5)
Exon 2	618-623	TTAGAA	(4)
Exon 2	685-690	GAAAAT	(4)
Exon 2	686-693	AAAATAAA	(3)
	687-694	AAATAAAA	
	688-695	AATAAAAA	
	689-696	ATAAAAAA	
	691-698	AAAAAACA	

sequence specificity alone (107). Third, many regulators of splicing can act as positive or negative regulators depending on the location within the intron and/or exon. For example, moving the 3RE of L1 IIIa from its natural position to the L1 IIIa second exon converts the 3RE from a splicing repressor element to a classical splicing enhancer element (114).

Such complex mechanisms of action are common in viruses. Alternative pre-mRNA splicing is widely utilized by DNA viruses and retroviruses yet little is known about the underlying mechanisms of regulation. HIV serves as an excellent example, as alternative 5' and 3' ss result in the synthesis of over 40 different mRNA species. Of importance is the abundance of *cis*-acting regulatory elements coordinating an intricate array of alternative splice site usage (137). Both ESSs and ESEs are present, some in multiple sequences within the same exons or spread out across exons. For example, an ESS in exon 3 of HIV-1 binds hnRNPA1 with high affinity and inhibits splicing by stimulating additional hnRNPA1 binding towards the 3'ss. This repressive effect can be lifted by binding of ASF/SF2 to the upstream ESE, an excellent example of crosstalk between different regulatory elements (137).

Based on our current findings, we hypothesize that the L4-22K protein functions by direct protein interaction with a splicing regulator(s), positive or negative in activity. Our targeted coIPs to probe for PTB and $U2AF^{65}$ binding did not show an interaction between either of these factors and L4-22K under our experimental conditions. Therefore, we conducted proteomic experiments to attempt to identify relevant cellular proteins that bind L4-22K. To this end, viruses were generated that express both Strep and HA tags on the N-terminus or C-terminus of the L4-22K protein. Experimental conditions were optimized for a two-step pull-down assay. The first purification step was done with Strep-tactin beads,

followed by desthiobiotin elution. The second IP step was done using anti-HA-agarose, followed by boiling in the presence of 2% SDS buffer. Silver stain analysis showed that at least for distinct proteins were observed that appeared specific to L4-22K protein expression (Fig. 4-6B, right panel). Further analysis of these proteins will reveal their identity (see next Chapter). A similar proteomics approach was utilized for the L4-33K protein in search of cellular factors that affect L4-33K-mediated L1 IIIa splicing activation. Instead of virus infection, the researchers utilized a GST pull-down assay using purified L4-33K protein (116). The results showed that L4-33K interacts with, and is phosphorylated by, DNA-PK. DNA-PK has been connected to transcriptional regulation and DNA repair, and was found to inhibit the shift from L1-52/55K to L1-IIIa alternative splicing. Interestingly, they found L4-22K to be a substrate for DNA-PK phosphorylation, although to a lower degree compared to L4-33K. Persson et al. (116) did not find any known splicing factors to be associated with the L4-33K protein, but this could be due to the different experimental conditions they utilized.

In this study, we uncovered at least two distinct regulatory elements found in the L4-33K pre-mRNA that regulate splicing. One of the elements was found within the 3'ss which we termed the enhancer element. This sequence is necessary for L4-22K-mediated splicing of the L4-33K pre-mRNA, but is not sufficient for this to occur. In addition, we found that sequences within the 3' exon of L4-33K may act as ESSs to prevent constitutive splicing of L4-33K. The significance of the intron and exon silencers and enhancer sequences is augmented by their ability to cause human disease when perturbed. Mutations in *cis*-acting elements cause aberrant splicing in genes, altering protein function, and contribute to disease pathology (110). Sequence variations in these elements can modify the pre-mRNA secondary structure. This can have detrimental effects by affecting the way the RNA sequence is

displayed to the incoming binding factor and disrupt splicing efficiency (109). Therefore a deeper understanding of the regulatory mechanisms of alternative splicing is critical both for the field of virology and for the relevance in human pathology.

The Adenovirus MLTU is an excellent example of an alternative spliced gene in which one 5’ ss can be spliced a large number of different 3’ ss. Mechanisms of alternative splicing are generally accompanied by different degrees of regulation and the Ad MLTU is no exception. The temporal change in MLTU alternative splicing results in activation of additional 3’ ss and as a consequence, the synthesis of additional mRNA species required as infection progresses. This is important for the virus life cycle since it allows for concomitant expression of multiple structural and regulatory proteins from a single transcription unit to maximize efficiency of virus progeny production.

Chapter 5: Conclusions and Future Directions

Adenoviruses have a complex life cycle that represents a series of events that are temporally regulated in order to optimize efficient production of progeny virus. Ads must devote a considerable part of their genome to immune evasion functions to facilitate infection and a state of persistence in the host. At the same time, Ads are excellent at utilizing the cellular machinery, when needed, to carry out regulatory functions in concert with the viral proteins. Since Ads are excellent at manipulating the host cell to promote viral infection, studies into their basic biology have contributed to the fields of cellular biology, including DNA replication, control of gene expression, and tumorigenesis. There are still numerous unanswered questions and unsolved problems regarding Ad biology. As the growing cases of serious Ad infections in immunologically compromised individuals rise, there is still no virus-specific therapy available to the general public. Additionally, with the growing interest in the development of Ads as gene therapy vectors, it is becoming more critical to gain a better understanding of the molecular mechanisms underlying the infection cycle.

In this Chapter, I highlight the most recent findings in our laboratory pertaining to the significant role of the L4-22K protein in the Ad life cycle and describe future directions for these analyses. Of particular importance is the role of the L4-22K protein in the encapsidation of the viral genome. To gain a better mechanistic understanding of this function, we took a genetic approach to map the structure-function relationship of the protein. Upon close inspection of the conserved C-terminal domain of L4-22K, we found a conserved pair of cysteine residues that upon mutation uncouple the functions of L4-22K in viral DNA packaging and the regulation of late gene expression. A virus with substitution mutations in these cysteine residues did not disrupt the ability of the L4-22K protein to

regulate late gene expression, but showed a decrease in viral genome packaging. Interestingly, this mutant virus was still able to bind to the packaging sequences, a critical step in DNA encapsidation. We postulate that these cysteine residues play a critical role in a yet unknown protein-protein interaction domain of the protein and we would like to explore this hypothesis further in future experiments. In the proteomics experiments described below, we would like to explore the role of the cysteine residues in potential interactions with relevant viral and/or cellular proteins.

Of biological significance in this study, we discovered that L4-22K is the master regulator to coordinate Ad late gene expression. We focused specifically on its role to activate alternative splicing of the L4-33K pre-mRNA transcript, initiating a series of regulatory cascades to turn on all the late genes required to drive the Ad life cycle through the late phase of infection. Besides the need for temporal control, a strict regulation of alternative splicing is required to prevent the generation of mRNAs that are unstable or code for defective or deleterious isoforms. This is a concept that holds true for many human diseases that are caused by a dysregulation in alternative splicing. This allows for the use of Ads as a model to study the mechanism of regulation behind this significant eukaryotic function. The need to regulate alternative splicing introduces an extra requirement for signals that must modulate splicing, beyond classical splicing signals. We were intrigued to find that such regulatory signals are present in the L4-33K pre-mRNA that allows for an intricate pattern of regulation that involves the L4-22K protein.

In future experiments, we wish to isolate L4-22K mutants that are defective for splicing regulation. The L4-22K protein does not contain classical RGG/RG motifs that act as RNA binding motifs for many splicing factors. However, we postulate that the N-terminal

domain of L4-22K that is shared with L4-33K (also a splicing regulator) recruits splicing factors, while the unique C-terminal domain may contain a yet unidentified RNA binding motif. We would like to further explore this hypothesis by mutating individual, and clusters, of conserved arginine residues in this region to alanines and measure splicing activity using our transient splicing assay with the goal of generating one or more L4-22K mutants that lack splicing regulator activity. A similar approach can be used to study the N-terminal domain of L4-22K and analyze functional splicing motifs.

We would also like to determine if L4-22K binds RNA. We can identify RNA binding targets by immunoprecipitation using rabbit polyclonal anti-L4-22K antibody to find any associated RNAs with our protein of interest. Immunoprecipitated samples could be validated using RT-qPCR with primer pairs that span intervals across the L4 region. These samples also could be analyzed by RNA-Seq. We also would use RNA-ChIP to examine L4-22K binding to RNA *in vivo.* Specifically, we would like to determine if L4-22K binds in a site-specific manner to the *cis*-acting sequences in the L4-33K pre-mRNA that we have identified in this study that mediate L4-22K-regulated splicing. These results will provide important information regarding how L4-22K protein functions to regulate alternative pre-mRNA splicing. There are two possible scenarios: L4-22K may bind directly to the pre-mRNA and displace a splicing repressor or recruit a splicing activator (see Chapter 4). Alternatively, L4-22K does not bind to RNA and may function by interaction with a cellular splicing factor that itself binds RNA.

Next, we would like to complete our model for L4-22K regulation of alternative splicing by uncovering the binding partners, or splicing regulators, that mediate L4-22K regulatory functions. To this end, we optimized conditions for L4-22K pull-down assays

utilizing a virus that encodes a tandem-tagged L4-22K protein (Chapter 4). Our results thus far indicate that the complexity of these samples is large, and that there appear to be numerous non-specific interactions at play. To distinguish between non-specific and specific interactions, we will utilize I-DIRT (Isotopic Differentiation of Interactions as Random or Targeted (138), a variation of SILAC (131). We plan to label uninfected cells with ^{13}C-Lys, while using cells grown in normal media for L4-22K-tagged virus infection. The samples will be used for two rounds of immunoprecipitation, and the final eluted sample will be submitted for mass spectrometry analysis to identify L4-22K-specific interactions based on the mass spec differences from the differential isotope labeling in infected and uninfected samples.

Finally, we would like to explore if L4-22K affects spliceosome assembly by using nuclear extracts prepared from cells infected with WT Ad5 or the L4-22K null mutant (described in Chapter 3). These nuclear extracts will be analyzed in spliceosome EMSA assays using ^{32}P-labeled RNA probes that correspond to the L4-33K pre-mRNA introns and adjacent flanking sequences. Additional constructs from our SV40-L4-33K chimeric constructs will be generated and analyzed for H, E, A, and B/C spliceosome complex formation (see Chapter 1).

References

1. Rowe WP, Huebner RJ, Gilmore LK, Parrot RH, Ward TG. Isolation of a cytopathogenic agent from human adenoids undergoing spontaneous degeneration in tissue culture. Proc Soc Exp Biol Med. 1953;84:570-3.

2. Echavarría M. Adenoviruses in immunocompromised hosts. Clin Microbiol Rev. 2008;21:704-15.

3. Berk AJ. Adenoviridiae: the viruses and their replication. Knipe DM, Howly PM, editors. Philadelphia, PA: Lippincott Williams & Wilkins; 2007. 2355-94.

4. Yamamoto M, Curiel DT. Current issues and future directions of oncolytic adenoviruses. Mol Ther. 2010;18(2):243-50.

5. Russell SJ, Peng KW, Bell JC. Oncolytic virotherapy. Nature Biotechnol. 2012;30(7):658-70.

6. Davison AJ, Benkö M, Harrach B. Genetic content and evolution of adenoviruses. J Gen Virol. 2003;84:2895-908.

7. Giberson AN, Davidson AR, Parks RJ. Chromatin structure of adneovirus DNA througout infection. Nucleic Acids Res. 2012;40(6):2369-76.

8. Waddington SN, McVey JH, Bhella D, Parker AL, Barker K, Atoda H, et al. Adenovirus serotype 5 hexon mediates liver gene transfer. Cell. 2008;132:397-409.

9. Russell WC. Adenoviruses: update on structure and function. J Gen Virol. 2009;90 (Pt 1):1-20.

10. Wolfrum N, Greber UF. Adenovirus signalling in entry. Cell Microbiol. 2013 Jan;15(1):53-62.

11. Karen KA, Hearing P. Adenovirus core protein VII protects the viral genome from a DNA damage response at early times after infection. J Virol. 2011;85:4135-42.

12. Berk AJ. Recent lessons in gene expression, cell cycle control, and cell biology from adenovirus. Oncogene. 2005 Nov 21;24(52):7673-85.

13. Ruley HE. Adenovirus early region 1A enables viral and cellular transforming genes to transform primary cells in culture. Nature. 1983 Aug 18-24;304(5927):602-6.

14. Frisch SM, Mymryk JS. Adenovirus-5 E1A: paradox and paradigm. Nat Rev Mol Cell Biol. 2002 Jun;3(6):441-52.

15. DeCaprio JA. How the Rb tumor suppressor structure and function was revealed by the study of Adenovirus and SV40. Virology. 2009 Feb 20;384(2):274-84.

16. Turnell AS, Mymryk JS. Roles for the coactivators CBP and p300 and the APC/C E3 ubiquitin ligase in E1A-dependent cell transformation. Br J Cancer. 2006 Sep 4;95(5):555-60.

17. Chinnadurai G. Opposing oncogenic activities of small DNA tumor virus transforming proteins. Trends Microbiol. 2011 Apr;19(4):174-83.

18. Yousef AF, Fonseca GJ, Cohen MJ, Mymryk JS. The C-terminal region of E1A: a molecular tool for cellular cartography. Biochem Cell Biol. 2012 Apr;90(2):153-63.

19. Levine AJ. The common mechanisms of transformation by the small DNA tumor viruses: The inactivation of tumor suppressor gene products: p53. Virology. 2009 Feb 20;384(2):285-93.

20. Blackford AN, Grand RJ. Adenovirus E1B 55-kilodalton protein: multiple roles in viral infection and cell transformation. J Virol. 2009 May;83(9):4000-12.

21. Wimmer P, Schreiner S, Dobner T. Human pathogens and the host cell SUMOylation system. J Virol. 2012 Jan;86(2):642-54.

22. Schreiner S, Wimmer P, Dobner T. Adenovirus degradation of cellular proteins. Future Microbiol. 2012 Feb;7(2):211-25.

23. White E. Mechanisms of apoptosis regulation by viral oncogenes in infection and tumorigenesis. Cell Death Differ. 2006 Aug;13(8):1371-7.

24. De Jong RN, Van der Vliet PC, Brenkman AB. Adenovirus DNA replication: protein priming, jumping back and the role of the DNA binding protein DBP. Curr Top Microbiol Immunol. 2003;272(187-211).

25. de Jong RN, Meijer LA, van der Vliet PC. DNA binding properties of the adenovirus DNA replication priming protein pTP. Nucleic Acids Res. 2003 Jun 15;31(12):3274-86.

26. Schaack J, Ho WY, Freimuth P, Shenk T. Adenovirus terminal protein mediates both nuclear matrix association and efficient transcription of adenovirus DNA. Genes Dev. 1990;4:1197-208.

27. Fessler SP, Delgado-Lopez F, Horwitz MS. Mechanisms of E3 modulation of immune and inflammatory responses. Curr Top Microbiol Immunol. 2004;273:113-35.

28. Lichtenstein DL, Toth K, Doronin K, Tollefson AE, Wold WS. Functions and mechanisms of action of the adenovirus E3 proteins. Int Rev Immunol. 2004;23(1-2):75-111.

29. Fessler SP, Chin YR, Horwitz MA. Inhibition of tumor necrosis factor (TNF) signal transduction by the adenovirus group C RID complex involves downregulation of surface levels of TNF receptor 1. J Virol. 2004;78(23):12113-3121.

30. Sester M, Ruszics Z, Mackley E, Burgert HG. The transmembrane domain of the adenovirus E3/19K protein acts as an endoplasmic reticulum retention signal and contributes to intracellular sequestration of major histocompatibility complex class I molecules. J Virol. 2013;87(11):6104-17.

31. Routes jM, Ryan S, Morris K, Takaki R, Cerwenka A, Lanier LL. Adenovirus serotype 5 E1A sensitizes tumor cells to NKG2D-dependent NK cell lysis and tumor rejection. J Exp Med. 2005;202:1477-82.

32. McSharry BP, Burgert HG, Owen DP, Stanton RJ, Prod'homme V, Sester M, et al. Adenovirus E3/19K promotes evasion of NK cell recognition by intracellular sequestration of the NKG2D ligands major histocompatibility complex class I chain-related proteins A and B. J Virol. 2008;82(9):4585-94.

33. Garnett CT, Talekar G, Mahr JA, Huang WM, Zhang Y, Ornelles DA, et al. Latent species C adenoviruses in human tonsil tissues. J Virol. 2009;83:2417-28.

34. Fu J, Bouvier M. Determinants of the endoplasmic reticulum (ER) lumenal-domain of the adenovirus serotype 2 E3-19K protein for association with and ER-retention of major histocompatibility complex class I molecules. Mol Immunol. 2011;48(8):532-8.

35. Lichtenstein DL, Krajcsi P, Esteban DJ, Tollefson AE, Wold WS. Adenovirus RIDbeta subunit contains a tyrosine residue that is critical for RID-mediated receptor internalization and inhibition of Fas- and TRAIL-induced apoptosis. J Virol. 2002;76(22):11329-42.

36. Burgert HG, Blusch JH. Immunomodulatory functions encoded by the E3 transcription unit of adenoviruses. Virus Genes. 2000;21(1-2):13-25.

37. Windheim M, Hilgendorf A, Burgert HG. Immune evasion by adenovirus E3 proteins: exploitation of interacellular trafficking pathways. Curr Top Microbiol Immunol. 2004;273:29-85.

38. Zanardi TA, Yei S, Lichtenstein DL, Tollefson AE, Wold WS. Distinct domains in the adenovirus E3 RIDalpha protein are required for degradation of Fas and the epidermal growth factor receptor. J Virol. 2003;77(21):11685-96.

39. Chin YR, Horwitz MS. Adenovirus RID complex enhances degradation of internalized tumour necrosis factor receptor 1 without affecting its rate of endocytosis. J Gen Virol. 2006;87(11):3161-7.

40. Friedman JM, Horwitz MA. Inhibition of tumor necrosis factor alpha-induced NF-KB activation by the adenovirus E3–10.4/14.5K complex. J Virol. 2002;76:5515-21.

41. Delgado-Lopez F, Horwitz MS. Adenovirus RIDalphabeta complex inhibits lipopolysaccharide signaling without altering TLR4 cell surface expression. J Virol. 2006;80(13):6378-86.

42. Tollefson AE, Toth K, Doronin K, Kuppuswamy M, Doronina OA, Lichtenstein DL, et al. Inhibition of TRAIL-induced apoptosis and forced internalization of TRAIL receptor 1 by adenovirus proteins. J Virol. 2001;75(19):8875-87.

43. Benedict CA, Norris PS, Prigozy TI, Bodmer JL, Mahr JA, Garnett CT, et al. Three adenovirus E3 proteins cooperate to evade apoptosis by tumor necrosis factor-related apoptosis-inducing ligand receptor-1 and -2. J Biol Chem. 2001;276(5):3270-8.

44. Lichtenstein DL, Doronin K, Toth K, Kuppuswamy M, Wold WS, Tollefson AE. Adenovirus E3-6.7K protein is required in conjunction with the E3-RID protein complex for the internalization and degradation of TRAIL receptor 2. J Gen Virol. 2004;78(22):12297-307.

45. Moise AR, Grant JR, Lippé R, Gabathuler R, Jefferies WA. The adenovirus E3-6.7K protein adopts diverse membrane topologies following posttranslational translocation. J Virol. 2004;78(1):454-63.

46. Schneider-Brachert W, Tchikov V, Merkel O, Jakob M, Hallas C, Kruse ML, et al. Inhibition of TNF receptor 1 internalization by adenovirus 14.7K as a novel immune escape mechanism. J Clin Invest. 2006;116(11):2901-13.

47. Carmody RJ, Maguschack K, Chen YH. A novel mechanism of nuclear factor-kappaB regulation by adenoviral protein 14.7K. Immunology. 2006;117(2):188-95.

48. Spurrell E, Gangeswaran R, Want P, Cao F, Gao D, Feng B, et al. STAT1 interaction with E3-14.7K in monocytes affects the efficacy of oncolytic adenovirus. J Virol. 2014;88(4):2291-300.

49. Horwitz MA. Function of adenovirus E3 proteins and their interactions with immunoregulatory cell proteins. J Gene Med. 2004;6(S1):S172-S83.

50. Klingseisen L, Ehrenschewender M, Heigl U, Wajant H, Hehigans T, Schütze S, et al. E3-14.7K is recruited to TNF-receptor 1 and blocks TNF cytolysis independent from interaction with optineurin. PLoS One. 2012;7(6):e38348.

51. Windheim M, Southcombe JH, Kremmer E, Chaplin L, Urlaub D, Falk CS, et al. A unique secreted adenovirus E3 protein binds to the leukocyte common antigen CD45 and modulates leukocyte functions. Proc Natl Acad Sci USA. 2013;110(50):E4884-93.

52. Subramanian T, Vijayalingam S, Chinnadurai G. Genetic identification of adenovirus type 5 genes that influence viral spread. J Virol. 2006;80(4):2000-12.

53. Wu K, Orozco D, Hearing P. The Adenovirus L4-22K Protein is Multifunctional and Is an Integral Component of Crucial Aspects of Infection. J Virol 2012;86(19):10474-83.

54. Tollefson AE, Scaria A, Ying B, Wold WS. Mutations within the ADP (E3-11.6K) protein alter processing and localization of ADP and the kinetics of cell lysis of adenovirus-infected cells. J Virol. 2003;77(14):7764-78.

55. Doronin K, Toth K, Kuppuswamy M, Krajcsi P, Tollefson AE, Wold WS. Overexpression of the ADP (E3-11.6K) protein increases cell lysis and spread of adenovirus. Virology. 2003;305(2):378-87.

56. Zou A, Atencio I, Huang WM, Horn M, Ramachandra M. Overexpression of adenovirus E3-11.6K protein induces cell killing by both caspase-dependent and caspase-independent mechanisms. Virology. 2004;326(2):240-9.

57. Yun CO, Kim E, Koo T, Kim H, Lee YS, Kim JH. ADP-overexpressing adenovirus elicits enhanced cytopathic effect by induction of apoptosis. Cancer Gene Ther. 2005;12(1):61-71.

58. Abou El Hassan MA, van der Meulen-Muileman I, Abaas S, Kruyt FA. Conditionally replicating adenoviruses kill tumor cells via a basic apoptotic machinery-independent mechanism that resembles necrosis-like programmed cell death. J Virol. 2004;78:12243-12251.

59. Endter C, Dobner T. Cell transformation by human adenoviruses. Curr Top Microbiol Immunol. 2004;273:163-214.

60. Javier RT, Rice AP. Emerging theme: cellular PDZ proteins as common targets of pathogenic viruses. J Virol. 2011 Nov;85(22):11544-56.

61. Thai M, Graham NA, Braas D, Nehil M, Komisopoulou E, Kurdistani SK, et al. Adenovirus E4ORF1-induced MYC activation promotes host cell anabolic glucose metabolism and virus replication. Cell metabolism. 2014 Apr 1;19(4):694-701.

62. Weitzman MD, Ornelles DA. Inactivating intracellular antiviral responses during adenovirus infection. Oncogene. 2005 Nov 21;24(52):7686-96.

63. Doucas V, Ishov AM, Romo A, Juguilon H, Weitzman MD, Evans RM, et al. Adenovirus replication is coupled with the dynamic properties of the PML nuclear structure. Genes Dev. 1996 Jan 15;10(2):196-207.

64. Geoffroy MC, Chelbi-Alix MK. Role of promyelocytic leukemia protein in host antiviral defense. J Interferon Cytokine Res. 2011 Jan;31(1):145-58.

65. Soria C, Estermann FE, Espantman KC, O'Shea CC. Heterochromatin silencing of p53 target genes by a small viral protein. Nature. 2010 Aug 26;466(7310):1076-81.

66. Branton PE, Roopchand DE. The role of adenovirus E4orf4 protein in viral replication and cell killing. Oncogene. 2001 Nov 26;20(54):7855-65.

67. Kleinberger T. Induction of transformed cell-specific apoptosis by the adenovirus E4orf4 protein. Prog Mol Subcell Biol. 2004;36:245-67.

68. Huang MM, Hearing P. 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. 1989;3:1699-710.

69. O'Connor RJ, Hearing P. The E4-6/7 protein functionally compensates for the loss of E1A expression in adenovirus infection. J Virol. 2000 Jul;74(13):5819-24.

70. Schaley JE, Polonskaia M, Hearing P. The adenovirus E4-6/7 protein directs nuclear localization of E2F-4 via an arginine-rich motif. J Virol. 2005 Feb;79(4):2301-8.

71. Van der Vliet PC. Adenovius DNA replication. Curr Top Microbiol Immunol. 1995;199:1-30.

72. Hoeben RC, Uil TG. Adenovirus DNA replication. Cold Spring Harb Perspect Biol. 2013;5(3):a013003.

73. Van der Vliet PC, Hoeben RC. Adenovirus. . DePamphilis ML, editor. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2006.

74. Ma Y, Mathews MB. Structure, function, and evolution of adenovirus-associated RNA: a phylogenetic approach. J Virol. 1996 Aug;70(8):5083-99.

75. Akusjarvi G. Temporal regulation of adenovirus major late alternative RNA splicing. Front Biosci. 2008:5006-15.

76. Young CSY. The structure and function of the adenovirus lajor late promoter. Curr Top Microbiol Immunol. 2003;272(213-250).

77. Ali H, LeRoy G, Bridge G, Flint SJ. The adenovirus L4 33-kilodalton protein binds to intragenic sequences of the major late promoter required for late phase-specific stimulation of transcription. J Virol. 2007;81(13):1327-38.

78. Ostapchuk P, Anderson ME, Chandrasekhar S, Hearing P. The L4 22-Kilodalton protein plays a role in packaging of the adenovirus genome. J Virol. 2006;80(14):6973-81.

79. Morris SJ, Leppard K, N. Adenovirus serotype 5 L4-22K and L4-33K proteins have distinct functions in regulating late gene expression. J Virol. 2009;83(7):3049-58.

80. Backström E, B. KK, Lan X, Göran A. Adenovirus L4-22K stimulates major late transcription by a mechanism requiring the intragenic late-specific transcription factor-binding site. Virus Res. 2010;151:220-8.

81. Soloway PD, Shenk T. The adenovirus type 5 i-leader open reading frame functions in cis to reduce the half-life of L1 mRNAs. J Virol. 1990;64:551-8.

82. Huang W, Flint SJ. The tripartite leader sequence of subgroup c adenovirus major late mRNAs can increase the efficiency of mRNA export. J Virol. 1998;72(1):225-35.

83. Xi Q, Cuesta R, Schneider RJ. Regulation of translation by ribosome shunting through phosphotyrosine-dependent coupling of adenovirus protein 100k viral mRNAs. J Virol. 2005;79(9):5676-83.

84. Cuesta R, Xi Q, Schneider RJ. Structural basis for competitive inhibition of eIF4G-Mnk1 interaction by the adenovirus 100-kilodalton protein. J Virol. 2004;78:7707-16.

85. Koyuncu OO, Dobner T. Arginine methylation of human adenovirus type 5 L4 100-kilodalton protein is required for efficient virus production. J Virol. 2009;83(10):4778-90.

86. Farley DC, Brown J, L., Leppard K, N. Activation of the early-late switch in adenovirus type 5 major late transcription unit expression by L4 gene products. J Virol. 2004;78(4):1782-91.

87. Guimet D, Hearing P. The adenovirus L4-22K protein has distinct functions in the posttranscriptional regulation of gene expression and encapsidation of the viral genome. J Virol. 2013;87(13):7688-99.

88. Morris SJ, Scott G, E., Leppard K, N. Adenovirus late-phase infection is controlled by a novel L4 promoter. J Virol. 2010;84(14):7096-104.

89. Wright J, Leppard KN. The human adenovirus 5 L4 promoter is activated by cellular stress response protein p53. J Virol. 2013;87(21):11617-25.

90. Törmänen H, Backström E, Carlsson A, Akusjarvi G. L4-33K, an adenovirus-encoded alternative RNA splicing factor. J Biol Chem. 2006;281(48):36510-7.

91. Ostapchuk P, Hearing P. Adenovirus IVa2 protein binds ATP. J Virol. 2008;82(20):10290-4.

92. Ostapchuk P, Hearing P. Control of adenovirus packaging. J Cell Biochem. 2005;96:25-35.

93. Ostapchuk P, Jihong Y, Auffarth E, Hearing P. Functional interaction of the adenovirus IVa2 protein with adenovirus type 5 packaging sequences. J Virol. 2005;79(5):2831-8.

94. Zhang W, Low JA, Christensen JB, Imperiale MJ. Role for the adenovirus IVa2 protein in packaging of viral DNA. J Virol. 2001;75(21):10446-54.

95. Yang T-C, Karl MN. Cooperative heteroassembly of the Adenoviral L4-22K and IVa2 proteins onto the viral packaging sequence DNA. Biochemistry. 2012;51:1357-68.

96. Ostapchuk P, Almond M, Hearing P. Characterization of empty adenovirus particles assembled in the absence of a functional adenovirus IVa2 protein. J Virol. 2011;85(11):5524-31.

97. Christensen JB, Ewing SG, Imperiale MJ. Identification and characterization of a DNA binding domain on the adenovirus IVa2 protein. Virology. 2012;433(1):124-30.

98. Gustin KE, Imperiale M. Encapsidation of viral DNA requires the adenovirus L1 52/5-kilodalton protien. J Virol. 1998;72:7860-70.

99. Perez-Romero P, Gustin KE, Imperiale MJ. Dependence of the encapsidation function of the adenovirus L1 52/55-kilodalton protein on its ability to bind the packaging sequence. J Virol. 2006;80(4):1965-71.

100. Ma H-C, Hearing P. Adenovirus structural protein IIIa is involved in the serotype specificity of viral DNA packaging. J Virol. 2011;85(15):7849-4. Epub August 2011.

101. Erturk E, Ostapchuk P, Wells SI, Yang J, Gregg K, Nepveu A, et al. Binding of CCAAT displacement protein CDP to adenovirus packaging sequences. J Virol. 2003;77(11):6255-64.

102. Fessler SP, Young CSH. The role of the L4-33K gene in adenovirus infection. Virology. 1999;263:507-16.

103. Finnen RL, Biddle JF, Flint SJ. Truncation of the human adenovirus type 5 L4-33K protein: evidence for an essential role of the carboxy-terminus in the viral infectious cycle. Virology. 2001 289:388-99.

104. Kulshreshtha V, Babiuk LA, Tikoo SK. Role of bovine adenovirus-3 33K protein in viral replication. Virology. 2004;323:59-69.

105. Wu K, Guimet D, Hearing P. The adenovirus L4-33K rpteoin regulates both late gene expression patterns and viral DNA packaging. J Virol. 2013 87:6739-347.

106. Hastings ML, Krainer A, R. Pre-mRNA splicing in the new millenium. Current Opinion in Cell Biology. 2001;13:302-309.

107. Risso G, Pelisch F, Quaglino A, Pozzi B, Srebrow A. Regulating the regulators: Serine/arginine-rich proteins under scrutiny. IUBMB Life. 2012;64(10):809-16.

108. Pagani F, Baralle F. Genomic variatns in exons and introns: identifying the splicing spoilers. Nat Rev Genet. 2004;5:389-96.

109. Singh R, Cooper T. Pre-mRNA splicing in disease and therapeutics. Trends in Molecular Medicine. 2012;18:472-82.

110. Aspegren A, Rabino C, Bridge E. Organization of splicing factors in Adenovirus-infected cells reflects changes in gene expression during the early to late phase transition. Exp Cell Res. 1998;245:203-13.

111. Bridge E, Xia D, Carmo-Fonseca M, Cardinali B, Lamond AI, Pettersson U. Dynamic organization of splicing factors in adenovirus-infected cells. J Virol. 1995;69(1):281-91.

112. Kanopka A, Mühlemann O, Akusjarvi G. Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. Nature. 1996;381:535-8.

113. Mühlemann O, Yue B-G, Petersen-Mahrt S, Akusjarvi G. A novel type of splicing enhancer regulating adenovirus pre-mRNA splicing. Mol Cell Biol. 2000;20:2317-25.

114. Persson HT, Aksaas AK, Kvissel AK, Punga T, Engström Å, Skålhegg B, et al. Two cellular protein kinases, DNA-PK and PKA, phosphorylate the adenoviral L4-33K protein and have opposite effects on L1 alternative RNA splicing. PLoS One. 2012;7(2):1-10.

115. Schiedner G, Hertel S, Kochanek S. Efficient Transformation of Primary Human Amniocytes by E1 Functions of Ad5: Generation of New Cell Lines for Adenoviral Vector Production. Hum Gene Ther. 2000;11:2105-16.

116. Evans J, Hearing P. Distinct Roles of the Adenovirus E4 ORF3 Protein in Viral DNA Replicaiton and Inhibition of Genome Concatenation. J Virol. 2003;77(9):5295-304.

117. Chartier C, Degryse E, Gantzer M, Dieterle A, Pavirani A, Mehtali M. Efficient Generation of Recombinant adenovirus vectors by homologous recombination in *Escherichia coli.* . J Virol. 1996;77:6255-64.

118. Morrissey JH. Silver Stain for proteins in polyacrylamide gels: a modified procedure with enhanced uniform sensitivity. Analytical Biochemistry. 1981;117:307-10.

119. Ostapchuk P, Anderson ME, Chandrasekhar S, Hearing P. The L4-22-Kilodalton protein plays a role in packaging of the adenovirus genome. J Virol. 2006;80(14):6973-81.

120. Brown T, Mackey K, Du T. Analysis of RNA by northern and slot blot hybridization. Curr Protoc Mol Biol. 2004;4(9):1-19.

121. Larkin M, Blackshields G, Chenna R, McGettigan P, McWilliam H, Valentin F, et al. Clutal W and ClustalX verison 2. Bioinformatics. 2007;23(21):2947-8.

122. Ewing SG, Byrd SA, Christensen JB, Tyler RE, Imperiale M. Ternary complex formation on the adenovirus packaging sequence by the IVa2 and L4 22-kilodalton proteins. J Virol. 2007;81(22):12450-7.

123. Zhang W, Imperiale M. interaction of the adenovirus IVa2 protein with viral packaging sequences. J Virol. 2000;74(6):2687-93.

124. Tollefson AE, Scaria A, Hermiston Tw, Ryerse JS, Wold LJ, Wold W, S.M. The adenovirus death protein (E3-11.6k) is required at very late stages of infection for

efficient cell lysis and release of adenvirus from infected cells. J Virol. 1996;70(4):2296-306.

125. Lichtenstein DL, Toth K, Doronin K, Tollefson AE, Wold W, S.M. Functions and mechanisms of action of the adenovirus E3 proteins. Int Rev Immunol. 2004;23:75-111.

126. Brayer KJ, Segal DJ. Keep your fingers off my DNA: Protein- protein interactions mediated by C2H2 zinc finger domains. Cell Biochem Biophys. 2008;50:111-31.

127. Fessler SP, Young CSH. Control of adenovirus early gene expression during the late phase of infection. J Virol. 1998;75(2):4049-56.

128. Miller MS, Pelka P, Fonseca GJ, Cohen MJ, Kelly JN, Barr SD, et al. Characterization of the 55-residue protein encode by the 9S E1A mRNA of species C adenovirus. J Virol. 2011;86(8):4222-33.

129. Mühlemann O, Kreivin J-P, Akusjarvi G. Enhanced splicing of nonconsensus 3' splice sites late during adenovirus infection. J Virol. 1995;69(11):7324-7.

130. Lützelberger M, Backström E, Akusjarvi G. Substrate-dependent differences in U2AF requirement for splicing in adenovirus-infected cell extracts. J Biol Chem. 2005;280(27):25478-84.

131. Kafasla P, Mickleburgh I, Llorian M, Coelho M, Gooding C, Cherny D, et al. Defining the roles and interactions of PTB. Biochemical Society Transactions. 2012;40(4):815-20.

132. Kanopka A, Mühlemann O, Petersen-Mahrt S, Estmer C, Öhrmalm C, Akusjarvi G. Regulation of adenovirus alternative RNA splicing by dephosphorylation of SR proteins. Nature. 1998;393:185-187.

133. Matlin AJ, Clark F, Smith CW. Understanding alternative splicing: towards a cellular code. Nature Rev Mol Cell Biol. 2005;6:386-98.

134. Cartegni L, Chew S, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet. 2002;3:285-98.

135. Stoltzfus CM, Madesin JM. Role of viral splicing elements and cellular RNA binding proteins in regulation of HIV-1 alternative RNA splicing. Curr HIV Res. 2006;4:43-55.

136. Tackett A, DeGrase J, Sekedat M, Oeffinger M, Rout M, Chait B. I-DIRT, a general method for distinguishing between specific and nonspecific protein interactions. J Proteome Res. 2005;4:1752-6.

137. Amanchy R, Kalume DE, Pandey A. Stable isotope labeling with amino acids in cell culture (SILAC) for studying dynamics of protein abundance and posttranslational modifications. Science's STKE. 2005;2005(267):1-20.

Appendix: The Cre/loxP System for the Production of Helper-Dependent Adenovirus Vectors is Inefficient Due a Defect in Vector DNA Replication

Abstract

Helper-dependent Ad vectors (HDAds) are attractive for gene therapy due to a multitude of advantages. Previously known as “gutless” vectors, HDAds lack viral genes allowing for a large cloning capacity and can be easily genetically manipulated. A production system of HDAds can yield highly purified virus utilized to transduce a wide array of cells and tissue types, and HDAds are able to mediate high-level, long-term transgene expression in the absence of toxicity. Despite extensive work in perfecting the HDAd system, efficient production of large quantities of high quality vector has been difficult and time consuming. This production problem poses a significant obstacle that needs to be addressed to successfully drive this technology to clinical trials. In this study, we developed an assay to generate HDAd vectors by utilizing a unique helper virus (HV) generated in our lab. This HV contains synthetic packaging sequences, flanked by loxP sites, to minimize recombination between the HDAd and HV and decrease HV contamination in HDAd preparations. We found that amplification of the HDAd, which contains the minimal left-end sequences required for packaging of the viral genome, was a slow, time-consuming process. Furthermore, we uncovered an inherent deficit in the inability of the HDAd to replicate efficiently during the production process. We propose additional methods to enhance HDAd production in our model system and summarize some of the major applications of HDAds and improvements in production over the past decade.

Introduction

Adenoviruses (Ads) have been extensively studied for their use as gene therapy vectors. Due to the very efficient nuclear entry mechanism of Ad and its low pathogenicity for humans, Ad-based vectors have been extensively studied as gene delivery vehicles. Ads can infect both resting and dividing cells of different types and of different species. As an important basis of their use in gene therapy, Ad vectors can be produced in high titers, they do not integrate into the host cell genome, and they can transduce cells *in vivo* with transgenes of more than 30kb. Of significance, the lack of any viral coding region in HDAds minimizes the host immune response, which allows for both prolonged transgene expression and a safer form of therapy (reviewed in (139)).

Prior to the production of HDAds, many other types of Ad-based vectors have been generated and applied as therapeutics. First generation E1-deleted vectors are based on the substitution of the E1 gene region by the transgene of interest and so they have a cloning capacity of ~4.7kb for foreign DNA (140). In nonpermissive cells, E1 deleted vectors are replication-defective, but under certain conditions leaky expression of viral genes can occur in cells where the vector was transduced. First-generation vectors may carry additional deletions in the E3 region, which is dispensable for viral replication in culture. This yields close to 8kb of available space to carry the transgene (reviewed in (141)). A major barrier in the use of first-generation Ads for gene therapy vectors is the strong immunological reaction to viral proteins. The immunological reaction against Ads has two phases: the initial response to the capsid proteins of the infecting virus, and immune reaction against the vector-

expressing cells resulting in short-lived transgene expression (142). The initial phase results in acute toxicity immediately after vector administration, characterized by elevated levels of proinflammatory cytokines, consistent with the activation of the innate inflammatory response (143). In vectors where most viral genes are left intact, the vectors are eliminated rapidly by cytotoxic T cells when the viral genes begin synthesizing new proteins (144). This problem led to new attempts at generating vectors that have additional deletions in the viral genome including the E2 and/or E4 genes (termed second-generation Ad vectors) to reduce viral DNA replication. Some of these vectors were successful at reducing immunogenicity (145) although the majority of viral coding sequences still remain, and therefore the potential for their expression and subsequent cellular immune response to them is evident.

The most practical way to overcome this immunological barrier to long-term gene transfer is to remove all of the viral genes from the vector. One successful system for generating vectors is the use of so called “gutless” vectors, now referred to as HDAds, that are deficient is all but the minimum *cis*-acting elements required for viral DNA replication and viral genome packaging into the virion. These elements include the ITRs and the packaging domain. An additional advantage to HDAd vectors is that they can accommodate up to 36 kb of non-viral DNA allowing for large cDNAs, tissue-specific promoters, and numerous expression cassettes to be accommodated (141). In this system, a HV provides in *trans* all of the necessary viral proteins needed for vector replication and genome packaging. These HDAds only require about ~500 bp from the ends of the genome to successfully incorporate the sequences into a fully packaged virion. The helper-dependent system that we discuss here uses a HV with a loxP-flanked packaging signal. Upon infection of a Cre-expressing cell line, recombination will delete the intervening packaging sequences,

rendering the virus unpackagable (143, 144). However, the viral DNA is able to replicate and express all the proteins required *in trans* to package HDAd containing the appropriate *cis*-acting regulatory elements. HDAd DNA containing the viral packaging signal (ψ) and inverted terminal repeats (ITRs) are transfected into Cre-expressing cells infected with the HV. The titer of the vector can be increased by serial passage through HV-infected N52.E6-Cre cells, a Cre-recombinase expressing cell line (117). Because the HV genome is prevented form packaging after Cre-mediated excision of ψ, the vector can be replicated and prepared nearly free of contaminating HV. (Fig. A-1) (adapted from (146)).

First- and second-generation Ad vectors still have clinical applications that are important today, including vaccination against infectious disease and anti-tumor therapy. These are conditions in which short-term gene expression is sufficient for therapeutic effects and the immunogenicity associated with the vectors might even be helpful (141). In contrast, HDAd vectors will be applied to conditions that require long-term gene expression and the expression of single or multiple genes. Examples include gene transfer into: liver to express secreted gene products like factor VIII for treatment of Hemophilia A; muscle tissue for treatment of Duchenne muscular dystrophy; eye to express pigment-epithelial derived growth factor; CNS, for treatment of multiple sclerosis and Huntington disease; lung tissue for therapy with cystic fibrosis; and vascular system for Hyperlipidaemia therapy (reviewed in (139)).

The lack of viral gene expression from HDAds has been shown to reduce their toxicity and immunogenicity to significant levels making all of the forms of gene therapy outlined above possible. Although the initial acute toxicity caused by first- and second-generation vectors is also present in HDAd vector systems, the latter exhibit an attenuated

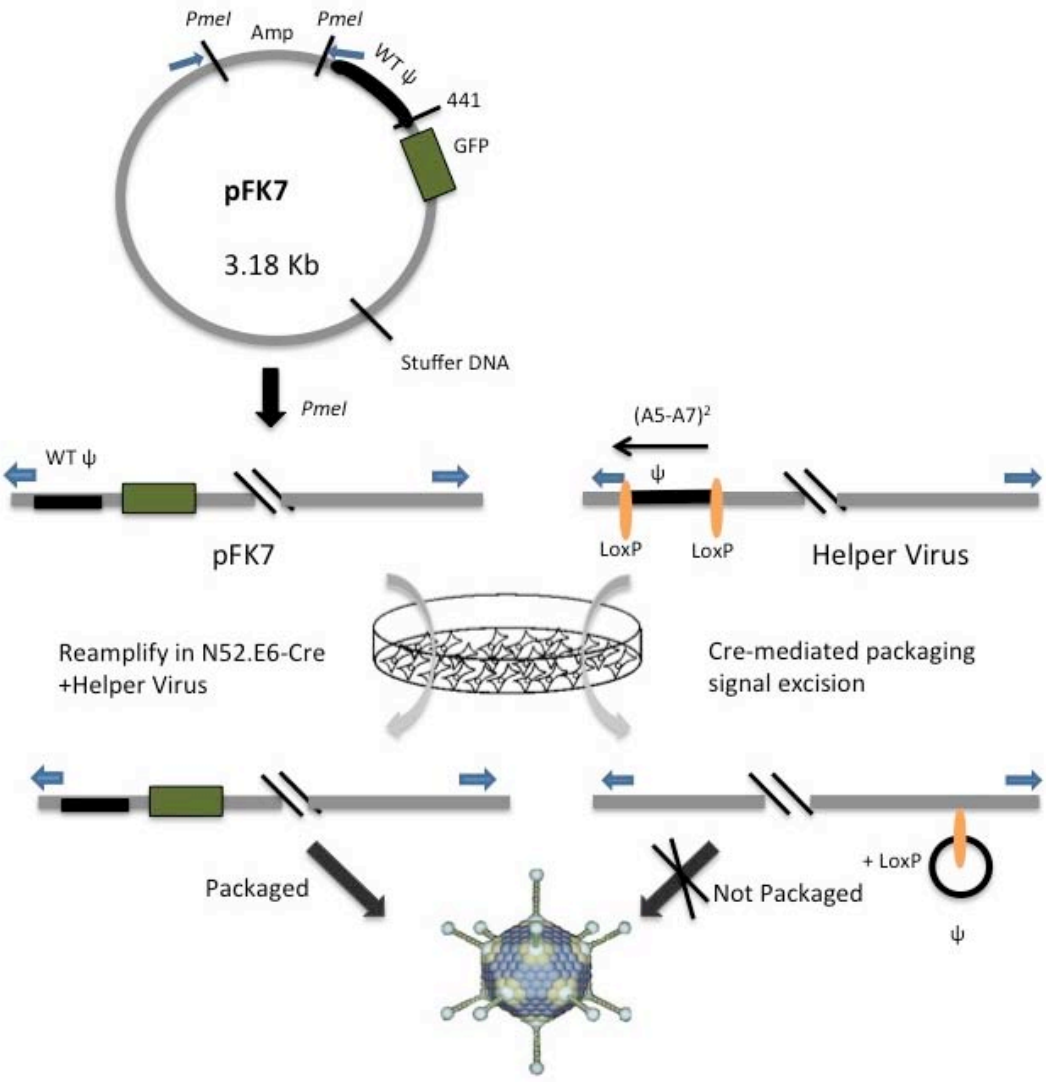


Figure A-1. The Cre-*loxP* system for the production of helper-dependent Ad vectors. pFK7 is the helper-dependent Ad (HDAd) utilized in this study and contains overlapping sequences to WT Ad from nt 1-441 including the ITRs and the packaging domain ψ. The remainder of the genome consists of a GFP expression cassette, and non-Ad stuffer DNA sequences. The pFK7 genome is separated from the plasmid by restriction digestion and transfected into N52.E6-Cre cells. To rescue the HDAd, cells are infected with a HV that contains a dimer of the A5-A7 packaging repeats, inverted, flanked by *loxP* sites. The helper virus is not packaged due to Cre-mediated excision of ψ, but provides the necessary factors required to propagate the HDAd to high titers. This figure was adapted from Palmer and Ng, Methods in Molecular Biology (2008).

adaptive immune response *in vivo* as compared to first- and second-generation Ad vectors (143). Additionally, long term transgene expression has been observed in liver cells of mice and non-human primates (147). Although recombinant Ad vectors are attractive for use in gene expression studies and therapeutic applications, the construction of these vectors remains relatively time-consuming (148). Despite modifications in vector production using the fundamental HDAd system (producer cell line, HV, HDAd backbone, and protocols) efficient production of large quantities of high quality vector has been difficult (139, 141, 148) and until very recently, most are limited to the production of modest amounts of HDAds suitable for small animal experiments or low-dose large animal studies (149). Here we describe a Cre/loxP system for generating HDAds that we adapted from previous studies. We generated our own HDAd and HV model system to try to understand why this process is inefficient. We uncovered that sufficient rescue of the HDAd does in fact require at least 4-5 serial passages for amplification, and that this is due to defective HDAd DNA replication. We suggest that there is an inherent defect in the HDAd that accounts for this defect in replication.

Materials and Methods

Cells, viruses, transfections, and infections. N52-E6-Cre cells were described in chapter 2. Cre-recombinase was turned on in the cells by supplementing with 200μg/ml Geneticin. N52.E6-Cre cells were transfected in 60-mm dishes with 5μg of vector pFK7 or other vector (described below), unless otherwise stated. Lipofectamine 2000 reagent (Invitrogen) was used for transfections according to the manufacturer's instructions. Twenty-four h posttransfection, cells were infected with 3 PFU/cell of helper virus, as previously described

in chapter 2. The helper virus utilized in this study contains a minimal packaging domain, ψ, composed of an inverted dimer of A5 and A7 repeats, flanked by *loxP* sites, virus $(A5\text{-}A7)^2$ helper virus (HV). This virus was generated by cloning the minimal packaging domain into Ad5-ψ-loxP described in chapter 2. HV stocks were derived from plaque-purified isolates and infectivity was determined by plaque assay and reported as PFU/ml.

Plasmids and amplification of HDAd. Plasmid pFK7 (Fig. A-1) was constructed as a bacterial plasmid and overlaps with Ad5 sequences nt 1-441, including the WT ψ and the ITRs. The remainder of this vector contains non-Ad stuffer DNA that is required for efficient packaging. The genome was excised from the pFK7 plasmid by *PmeI* digestion and used to transfect N52.E6-Cre cells. After 24 h, pFK7 transfected cells were infected with $(A5\text{-}A7)^2$ HV. Six hpi a portion of the cells were collected for analysis (see below). Forty-eight hpi a portion of the cells were collected for analysis, while the rest of the cells were freeze/thawed four times to prepare virus lysates, as described in chapter 2. Five-hundred µl of lysate was used to infect N52-E6-Cre supplemented with 3 PFU/Cell of $(A5\text{-}A7)^2$ HV. This process was repeated 4-5 more times to amplify sufficient quantities of HDAd for analysis by Southern blot and qPCR.

Plasmid pTG3602 was described in chapter 2. The genome was excised from the plasmid by *PacI* digestion and used to transfect N52.E6-Cre and 24 hpi the cells were infected with $(A5\text{-}A7)^2$ HV or left uninfected. Forty-eight hpi, transfected, or transfected + infected cells were harvested and used to prepare lysates for P0. Plasmid pTG3602 was amplified as described above for pFK7. pTG3602 was amplified alone, without HV, by using

pTG3602- transfected lysates from P0 to infect N52.E6-Cre for P1. This process was repeated for 2-3 more passages and amplification was analyzed by Southern blot and qPCR.

Plasmid pTG-JEI was previously described (15) and contains a 14bp deletion in E4 ORF6 and an 8bp *ClaI* insertion in E4 ORF3. The genome was excised from the plasmid by *PacI* digestion and used to transfect N52/E6-Cre cells, was amplified as described above for pFK7 and analyzed by Southern blot and qPCR.

To test for replication efficiency in amplified HDAds, prepared nuclear DNA was digested with corresponding restriction enzymes with or without *DpnI*, and analyzed by Southern blot for the presence or absence of unmethylated replicated DNA.

Southern blots and quantitative PCR. Cells at each passage of HDAd, or control plasmids were used to prepare nuclear DNA as described in chapter 2 and was digested with designated enzymes. Digested samples were separated on 0.8% Agarose gel and prepared as follows: the gel was incubated in a denaturing solution (8.8% NaCl, 2% NaOH) for 30 min followed by a 30 min incubation in a neutralization solution (8.8% NaCl, 6.1% Triz base, [pH 7.5]). The gel was transferred to a hybond-N+ membrane (GE healthcare) according to the manufacturer's instructions. The filter was hybridized in a solution of SSPE (6mM EDTA, 0.894M NaCl, 60mM sodium phosphate, [pH 7.4]; 5X Denhardst reagent (GE Healthcare); 0.5% SDS; and 100µg/ml of denatured calf thymus DNA (Sigma) at 65°C overnight with 250ng of a designated DNA probe ^{32}P-labeled by random priming (random primer DNA labeling kit; TaKaRa Bio), followed by a series of washes outlined in the GE healthcare hybond-N+ handbook. DNA probes were constructed using primers surrounding nt 1-194 of Ad5 to detect the amplified HDAd. The fragments were extended by PCR and

resolved through 1% agarose followed by gel extraction (Qiagen). To detect the DNA ladder, 100ng of a 2-log DNA ladder was labeled using the AlkPhos direct labeling system (Amersham).

The amount of helper dependent vector (FK7), or helper virus, was quantified by a DyNAmo HS SYBR green quantitative PCR (qPCR) kit (Finnzymes), as described in chapter 2. Primer pairs surrounding the loxP sites (nt 882-901 and nt 1052-1033) (Fig. A-1) were used to quantify the amount of $(A5\text{-}A7)^2$ HV viral DNA. Primer pairs surrounding nt 404-575 were used to quantify the amount of FK7 DNA. Note that these nt correspond to unique sequences in the pFK7 vector, not present in Ad5 (Fig. A-1). Samples were quantified based on a standard curve generated from known amounts of pFK7 vector to quantify FK7 DNA. A standard curve using a plasmid for Ad viral DNA (pTG3602, see chapter 2) was used to quantify the amount of HV DNA. Appropriate primers were used to quantify GAPDH copy number, and data was presented as total DNA normalized to GAPDH copy number.

Production of FK7 HDAd. HDAd pFK7 was amplified by serial passages as described above. Lysate from passage 6 was used to infect 16 x 100mm plates of N52.E6-Cre supplemented with $(A5\text{-}A7)^2$ HV. Cells and medium were collected 72 hpi, cells pelleted, and resuspended in 6 ml of TD buffer as described in chapter 2. Lysates were subjected to four cycles of freeze thaw and cleared by centrifugation. Cleared lysates were added to step gradients of 1.25g/cm^3 and 1.40g/cm^3 CsCl in TD buffer and spun at 175,587 x *g* at 15 °C for 1 h. Visible bands at each step were collected and used for a second step gradient of 1.34g/cm^3 and 1.25g/cm^3 CsCl prepared at a ratio of 2:1, respectively. The CsCl and virus mixture was spun for 3 days at 175,587 x *g*. The top band (corresponding to particles devoid

of viral DNA) and the middle band (corresponding to virus FK7) were collected taking caution to leave behind as much of the bottom band (corresponding to HV contamination) as possible. Collected bands where used in a third step gradient of 1.34g/cm^3 and 1.25g/cm^3 CsCl prepared at a ratio of 2:1, respectively and spun for 24 h at 175,587 x *g*. All centrifugation steps were done at 15 °C.

Results and Discussion

The Cre/loxP helper-dependent system is inefficient and requires multiple passages to prepare HDAd stocks.

The successful production of high titers of HDAds involves three major steps. The first step involves converting pHDAd, the plasmid form, to HDAd, the viral form by restriction enzyme digestion, followed by transfection of the excised genome and infection with HV. The second step is amplification, which involves increasing the amount of HDAd by serial coinfection with HV (referred to as passage, P). The final step is large-scale production to generate large quantities of HDAd (146). In order to better understand the underlying problem for the lack of efficient Cre/loxP HDAd production, we developed an HDAd and HV system and analyzed the production and amplification of the HDAd in serial passages. The HDAd described in this study is pFK7 and contains the Ad5 left-end and right-end ITRs and the Ad5 WT packaging domain (ψ) at the left-end followed by a reporter GFP expression cassette (Fig A-1). The remainder of the 31.8 kb vector is stuffer non-Ad DNA. Studies indicate that there is a lower limit for packaging so that vectors with genomes greater

than or equal to 27.7 kb are packaged with equal efficiencies, and those with small genomes are packaged inefficiently (140). This lower limit has been established during the recent gene therapy era, in addition to the well-defined upper packaging limit of approximately 105% of the WT genome size. Transfection of the genome requires *PmeI* digestion to release the viral DNA.

The possibility of homologous recombination between vector and HV DNA at the left-end of the genome within the packaging signal can occur during DNA replication, resulting in a HV that lacks one or both loxP sites which will not be selected against in the Cre-expressing producer cell line (144). To avoid this problem, we engineered a HV, $(A5\text{-}A7)^2$ HV, which contains a synthetic dimer of packaging repeats A5-A7 in an inverted configuration (Fig. A-1). This new HV contains a functional packaging domain but the possibility of homologous recombination to revert the HV sequences to WT is minimized since the sequences are minimal and inverted. Extensive evidence suggests that multimerized packaging elements can restore viability to a mutant virus that lacks a packaging domain (reviewed in (93)). Virus $(A5\text{-}A7)^2$ carries loxP sites flanking the synthetic packaging domain. When this virus is grown in a cell line that expresses Cre-recombinase, recombination between the loxP sites removes the packaging domain (termed floxing). The remainder of the Ad genome expresses all viral proteins necessary for successful propagation of the HDAd (Fig. A-1). Prior to performing experiments, the HV was plaque-purified, amplified, and infectious titer of the stock was determined.

In our HDAd system, we used N52.E6-Cre cells which provide E1 proteins *in trans* for successful propagation of the HDAd (117). Prior to the development of the HDAd system, classical HEK293 cells were being used for vector production. However, the

incidence of recombination between the left-terminus of first-generation Ad vector, HV, and overlapping E1 sequences in the 293 cellular genome resulted in the generation of RCAs (replication competent adenovirus), a serious safety risk (reviewed in (141)). The use of N52.E6 cell line avoids this problem since these cells have well-defined, non-overlapping E1 DNA fragments that excludes the possibility for recombination with HV (117). In order to successfully amplify the HDAd, at the expense of the HV, N52.E6-Cre should express Cre-recombinase to high levels. To determine the efficiency of floxing of the packaging domain with the HV, we infected N52.E6-Cre cells that were supplemented with geneticin, or left untreated, with $(A5\text{-}A7)^2$ HV. Nuclear DNA was prepared and digested with *BspEI* which yields characteristic restriction fragments depending on the presence or absence of the packaging domain (Fig. A-2, blue and green arrows, respectively). Results showed that N52.E6-Cre cells supplemented with geneticin (to select for the Cre-recombinase gene) successfully excised the packaging domain as shown by the nearly undetectable levels of unfloxed DNA compared to floxed DNA (Fig. A-2 compare blue, floxed, and green, unfloxed, arrows, lanes 3, 4, 7, and 8). At 6 hpi, the levels of HV DNA were undetectable in this assay, since viral replication did not begin until closer to 12 hpi (Fig. A-2, lanes 1, 2, 5, and 6).

Next, we transfected *PmeI*-digested pFK7 vector and subsequently infected with $(A5\text{-}A7)^2$ HV at 24 h post-transfection. Cells were harvested at 6 and 48 hpi and nuclear DNA prepared and viral lysates were harvested at 48 hpi. Six hpi was used to analyze viral DNA input and 48 hpi was used to measure newly replicated DNA. We subjected the viral lysates from the initial transfection to serial passage through N52.E6-Cre cells supplemented with $(A5\text{-}A7)^2$ HV infection at each passage. To assess the amplification of the FK7 virus, we analyzed

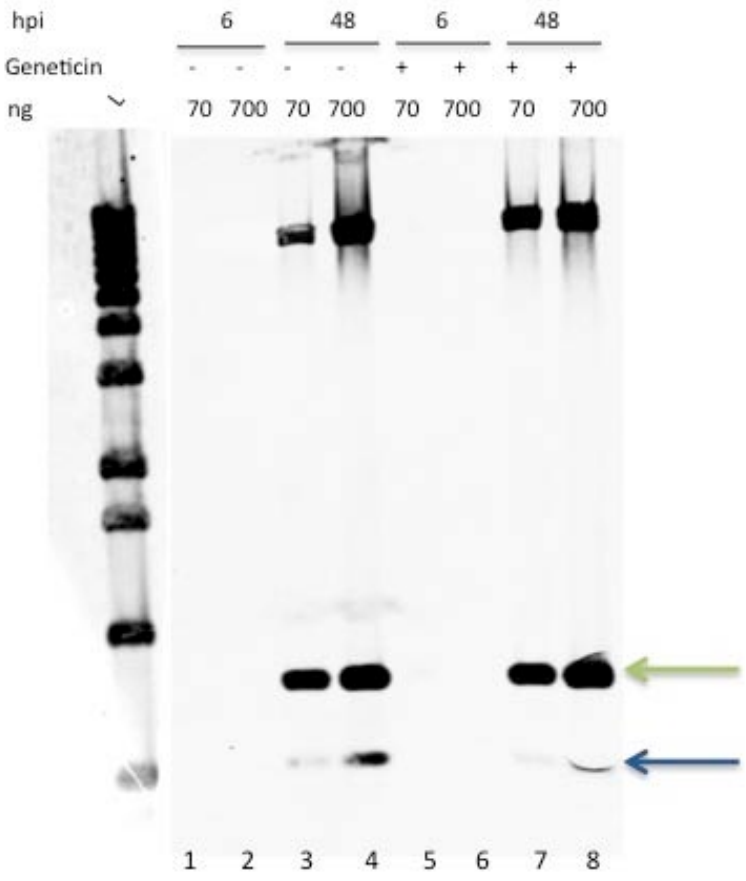
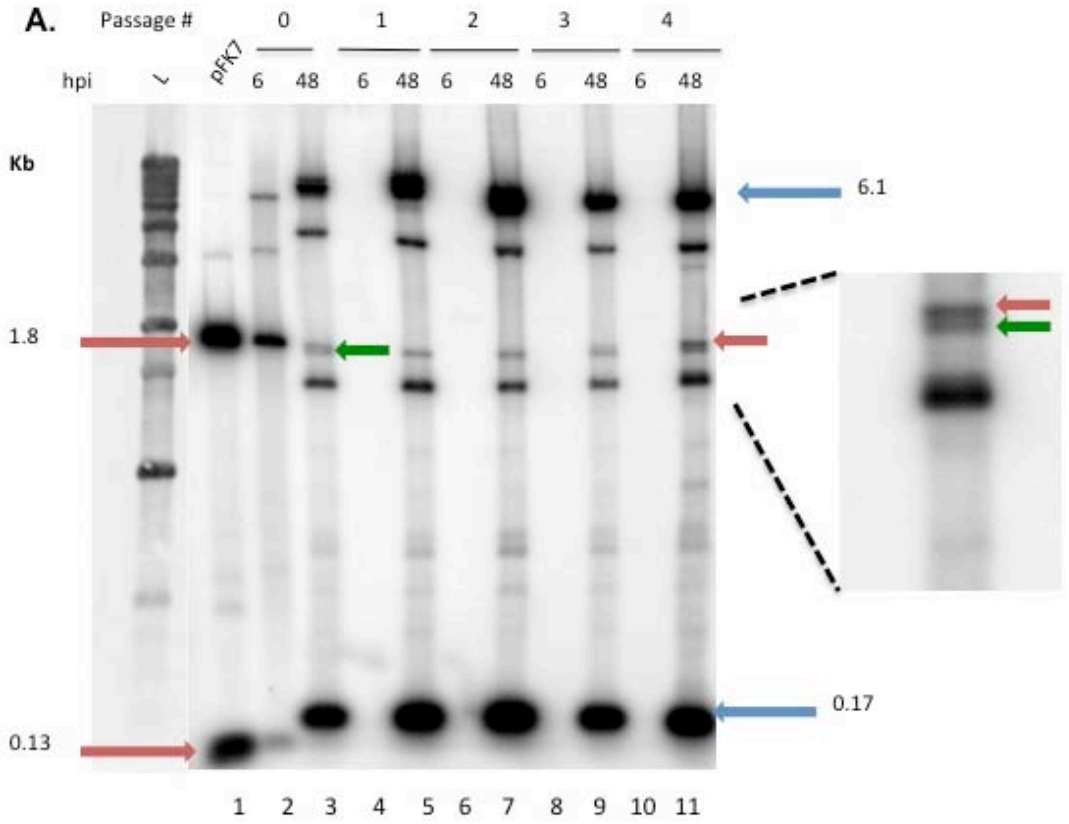
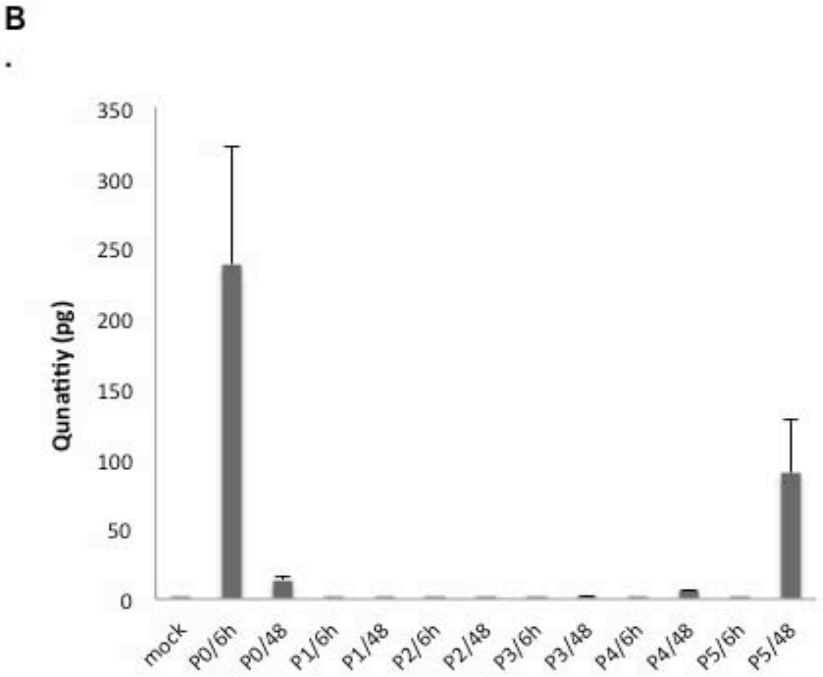


Figure A-2. Analysis of HV floxing efficiency in N52E6-Cre cells. Nuclear DNA was extracted from N52.E6-Cre cells infected with $(A5\text{-}A7)^{2}$ HV and harvested at 6 or 48 hpi and analyzed by Southern blot. Prior to infection cells were either supplemented with geneticin or left untreated. Seventy ng or 700 ng of DNA was digested with *BspEI* and hybridized with a probe to detect the marker lane (L) and a ^{32}-P labeled probe surrounding the left end of the genome, nt 1-194. Unfloxed DNA results in a 345 bp band (blue arrow) and floxed DNA results in a 826 bp band (green arrow).

fragments produced by *XbaI* and *XhoI* digestion of FK7 and HV DNA by Southern blot (Fig. A-3A) and qPCR (Figs. A-3B and A-3C). Note that for all analyses, we digested and loaded identical amounts of DNA for each sample. Analysis of vector DNA restricted with *XbaI* and *XhoI* showed that transfection efficiency was high (Fig. A-3A, lane 2) and the bands were of the expected size as shown by restriction digested pFK7 control DNA (Fig. A-3A, lane 1, red arrows). At 48 hpi during P0 (transfection passage), there was a decrease in pFK7-specific DNA in comparison to 6 hpi (Fig. A-3A, lanes 2 and 3) indicting a loss of HDAd DNA. During serial passage, the HDAd-specific bands were undetectable by Southern blot analysis at 48 hpi in passages P1 through P3, and subsequently was weakly evident at P4 (Fig A-3A, lanes 5, 7, 9 and 11, and right panel insert, upper band, red arrow). The DNA fragment running just below the HDAd fragment (red arrow) in the 48 hpi samples represents a HV-specific DNA fragment (green arrow). The probe used in our Southern blot analysis was specific to nt 1-194 of viral DNA, therefore both pFK7 and HV-specific bands were detected. Since each serial passage involves coinfection with $(A5\text{-}A7)^2$ HV, we expected to see high amounts of HV present at all passages (Fig. A-3A, lanes 3, 5, 7, 9 and 11, blue arrows) and this was observed. Note that the absence of HV at 6 hpi at each serial passage suggests that there was no, or undetectable, HV contamination carried over from the previous passage and that the HV-specific bands observed were from HV provided at each specific passage. In order to rule out that the apparent "disappearance" of FK7 viral DNA in P1-P3 was due to the lack of sensitivity of the assay, we quantified FK7 viral DNA from serial passages using qPCR (Fig. A-3B). Although the samples utilized in this assay were from an independent experiment as those from the Southern blot (Fig. A-3A), the same pattern was observed, where the quantity of FK7-specific DNA was below the levels of detection from P1-3,


A.
Passage #
0
1
2
3
4
hpi
pFK7
6
48
Kb
1.8
0.13
6.1
0.17
1 2 3 4 5 6 7 8 9 10 11


B
Qunatitiy (pg)
350
300
250
200
150
100
50
0
mock
P0/6h
P0/48
P1/6h
P1/48
P2/6h
P2/48
P3/6h
P3/48
P4/6h
P4/48
P5/6h
P5/48

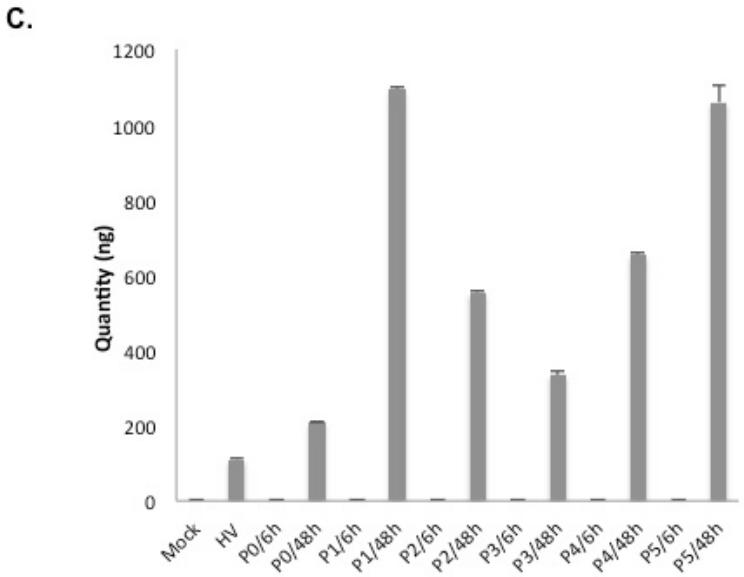


Figure A-3. Analysis of HDAd amplification. Nuclear DNA was extracted from N52.E6-Cre cells transfected with HDAd pFK7 (P0) followed by infection with $(A5\text{-}A7)^2$ HV, or coinfected with lysates obtained at each passage (P2-P5) plus HV.

Panel A. DNA was digested with *XbaI* and *XhoI* and hybridized with a probe to detect the marker lane (L) and a ^{32}-P labeled probe surrounding the left- end of the Ad5 genome, nt 1-194. Lane pFK7 represents 0.1µg of *XbaI/XhoI/PmeI*-digested plasmid pFK7 as a control. Serial passages 0-4 are shown with samples harvested at 6 and 48 hpi. Red arrows designate HDAd FK7- specific bands, while blue arrows designate $(A5\text{-}A7)^2$ HV-specific bands. Note that the HDAd specific band was undetectable at passages 1-3, and appeared again at passage 4 (right panel). The green arrow indicates a HV background band that migrates near the FK7-specific band.

Panel B Total yield of HDAd pFK7 was quantified using pFK7-specifc primers by qPCR analysis at each serial passage and represented as pg, normalized to known GAPDH quantities.

Panel C Total $(A5\text{-}A7)^2$ HV was quantified using $(A5\text{-}A7)^2$ HV-specific primers by qPCR analysis at each serial passage and represented as ng, normalized to known GAPDH quantities. Mock-infected samples shown as a control. $(A5\text{-}A7)^2$ HV-infected-only samples are shown as a control. Note the difference in scale between panel B and panel C. Southern blot and qPCR represents data from two independent experiments.

evident at low levels in P4, and detected easily at P5 (Fig. A-3B). HV DNA was readily observed in all 48 h samples (Fig. A-3C). We conclude that the production of FK7 HDAd is an inefficient process likely due to poor DNA replication during the amplification process.

The FK7 HDAd does not replicate efficiently.

Next, we wanted to understand why the process of amplification of HDAd was inefficient. We explored two possibilities: the first is that the HDAd is not able to successfully enter the cell; the second is that there is a defect in HDAd DNA replication that prevents efficient propagation during passaging. In order to examine HDAd replication, we analyzed *DpnI*-digetsted DNAs by Southern blot to distinguish between newly replicated DNA (*DpnI*-resistant) and transfected plasmid DNA (*DpnI*-sensitive). N52.E6-Cre cells were transfected with *PmeI*-linearized FK7 plasmid DNA, followed by infection with $(A5\text{-}A7)^2$ HV. Nuclear DNA was extracted and subjected to restriction digestion with *XbaI/XhoI,* plus or minus *DpnI* digestion (Fig. A-4). Transfected DNA is indicated by red arrows (no *DpnI* digestion) and red arrowheads (*DpnI*-digested). pFK7 plasmid DNA is shown as a control (Fig. A-4, lanes 1 and 2). As expected, samples collected from P0, 6 hpi, represented DNA from transfected pFK7 as shown by complete digestion of the 1.8 kb band with *DpnI* (Fig. A-4, lanes 5 and 6). To our surprise, FK7-specific bands from 48 hpi samples at P0 also showed complete digestion by *DpnI,* which strongly points to a complete, or nearly complete, lack of newly replicated FK7 DNA (Fig. A-4, lanes 7 and 8, red arrowhead) even in the presence of HV (Fig. A-4, lanes 7, 8, blue arrow indicates HV DNA). To examine this further, we utilized a WT Ad5 replication competent plasmid vector, pTG3602 (119). The Ad5 genome was excised by *PacI* digestion and used to transfect N52.E6-Cre cells. Cells were harvested

48 h post-transfection (P0) and lysates used for serial passages 1 and 2. Nuclear DNA from P1, 48 h post-transfection, was analyzed for the presence of replicated DNA by *HindIII* and *DpnI* digestion. As indicated by the presence of the 2.9kb band, much of the DNA present at 48 h post-transfection for this sample was newly replicated DNA (Fig. A-4, lanes 11 and 12, green arrow), with some DNA evident from the transfection (green arrowhead). When these lysates were used for serial passage, an increase in the intensity of the newly replicated, pTG3602-specific band was observed, (Fig. A-4, compare lanes 11, 13, 14). Note that this vector was passaged alone, without the HV present; the presence of replicated DNA was evident by the successful amplification of the virus through serial passage. From these results, we conclude that transfected WT Ad5 DNA derived from a plasmid can replicate efficiently, but that FK7 DNA derived from a plasmid does not replicate efficiently in a coinfection with HV.

To further probe why the HDAd is defective for replication, we wanted to examine another replication-defective Ad genome derived from a plasmid in conjunction with HV infection to determine if this is a general property of such vectors. We used an Ad5 vector that is defective for replication, pTG-JE1. This plasmid contains the Ad5 genome with mutations in the E4-ORF3 and E4-ORF6 genes, disrupting critical sequences in the E4 region that are required for viral replication (118). We transfected *PacI*-digested pTG-JE1 into N52.E6-Cre cells and infected with $(A5\text{-}A7)^2$ HV the next day. Lysates from P0 were collected 48 hpi and used for serial passage with HV. Six hpi and 48 hpi samples were collected and nuclear DNA was prepared. To assess the amplification of the pTG-JE1 virus, we analyzed fragments produced from *HindIII* digestion of nuclear DNA by Southern blot

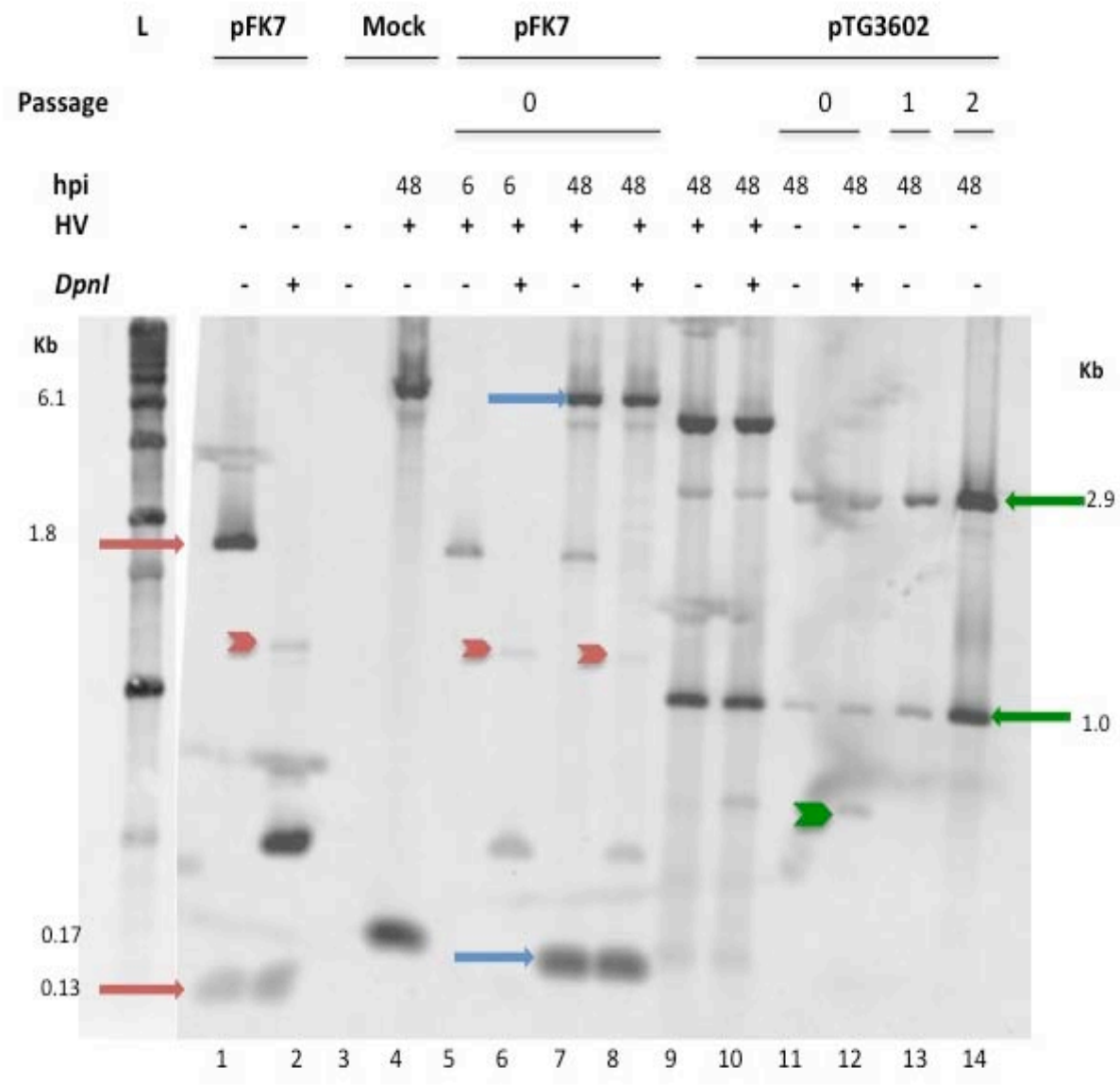
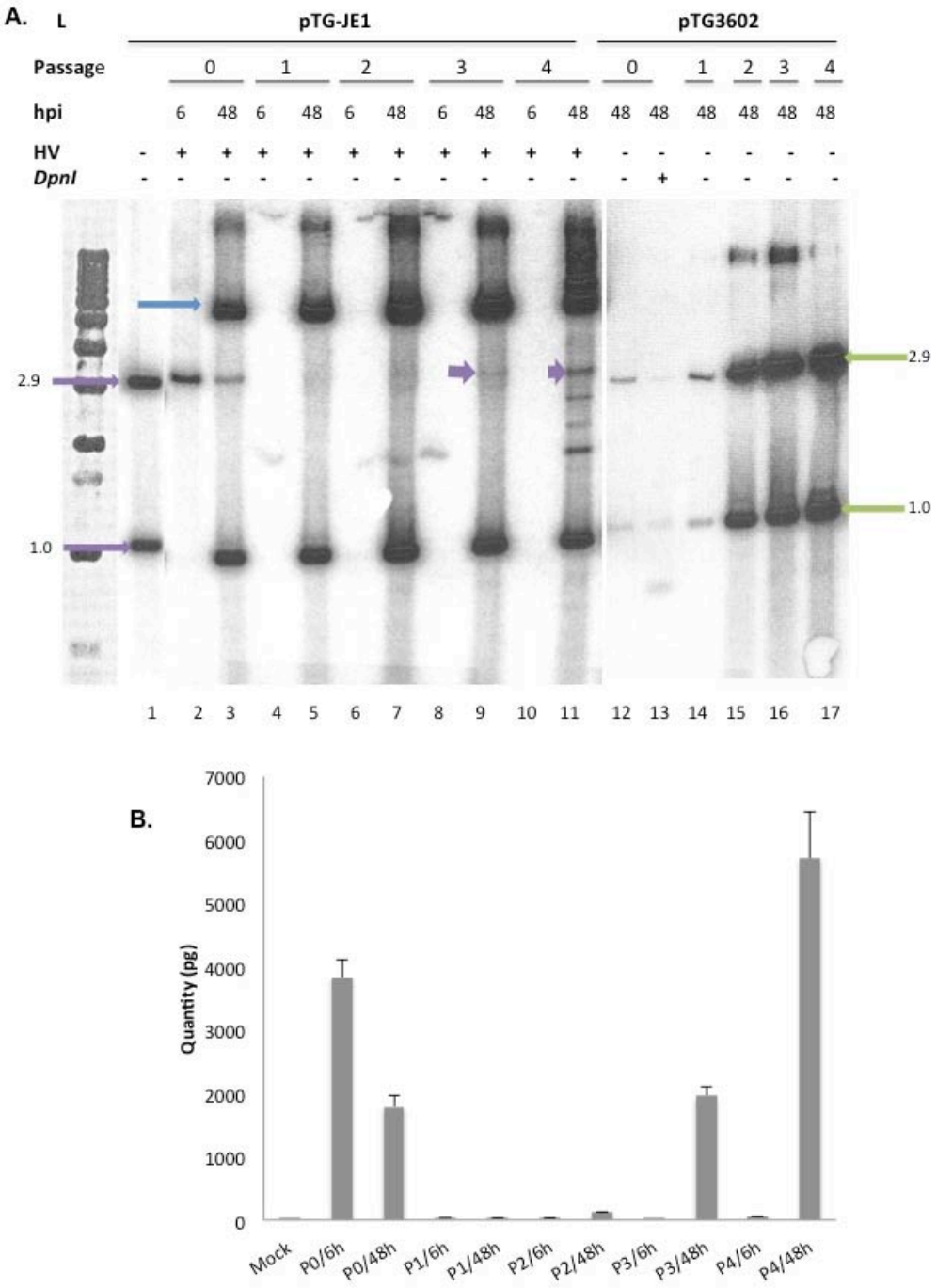

L
pFK7
Mock
pFK7
pTG3602
Passage
0
0
1
2
hpi
48
6
6
48
48
48
48
48
48
48
48
48
HV
- - - + + + + + + + - - - -
DpnI
- + - - - + - + - + - + - -
Kb
6.1
1.8
0.17
0.13
Kb
2.9
1.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14

Figure A-4. Analysis of replication efficiency of HDAd. N52.E6-Cre cells were transfected with *PmeI*-linearized pFK7 plasmid DNA and coinfected with $(A5\text{-}A7)^{2}$ HV and harvested at 6 or 48 hpi (lanes 5-8). Samples were digested with *XhoI*/XbaI, with or without *DpnI* (+ or –). N52.E6-Cre cells were transfected with *PacI*-linearized pTG3602 and harvested at 48 hpi (lanes 9-10). Samples were digested with *HindIII* with or without *DpnI* (+ or –). Lysates from pTG3602-transfected cells were subsequently passaged twice P1/P2, (lanes 13-14). Membranes were hybridized with ^{32}P labeled with a probe surrounding Ad5 nt 1-194 in a Southern blot analysis. Red arrows designate pFK7-specifc bands and red arrowheads indicate *DpnI*-digested, parent DNA. Blue arrows designate $(A5\text{-}A7)^{2}$ HV-specific bands. Green arrows designate pTG3602-specifc bands, and green arrowheads represent DpnI-digested, parent DNA. Lanes 1 and 2 represent 0.1µg of digested plasmid pFK7 with or without *DpnI*, as a control. Lane 3 is untransfected control. Lane 4 represents nuclear DNA from $(A5\text{-}A7)^{2}$ HV-only infected cells, harvest 48phi.

(Fig. A-5A) and qPCR (Fig. A-5B). Even though transfection efficiency of pTG-JE1 DNA was efficient (Fig. A-5A, lane 2), we did not detect the presence of pTG-JE1-specific bands in P1 and P2 samples; low levels were detected at P3 which became more abundant by P4 but still well below the level of HV replication (Fig. A-5A, lanes 5, 7, 9 and 11, purple arrows; HV fragment are indicated by blue arrows). We quantified the amount of pTG-JE1 viral DNA present in serial passages by qPCR and observed the same results: pTG-JE1 DNA was not evident during P1 and 2 and subsequently became evident during later passages P3 and P4. High amounts of pTG-JE1 viral DNA were observed in P0 as provided by the transfection (Fig. A-5B). Note that the samples from qPCR experiments represent independent samples from the Southern blot analysis. In contrast, vector pTG3602 passaged alone replicated very efficiently (Fig. A-5A, compare lanes 12 and 13) and amplification of the vector was observed as early as P1, even in the absence of HV (Fig. A-5A, lanes 14-17). This pattern was recapitulated when pTG3602 viral DNA was quantified by qPCR (Fig. A-5C). Note the difference in scale between Fig. A-5C (ng) and Fig. A-5B (pg). Thus, even though vector pTG-JE1, which possesses all of the viral sequences necessary for viral DNA replication (except E4 sequences which are provided by the HV), this vector replicated very inefficiently. Quantification of viral DNAs produced in these experiments showed that only about ~40-fold more viral DNA could be generated utilizing pTG-JE1 in an HDAd system, than if we utilized the gutted vector, pFK7 (compare Fig. A-3C and Fig. A-5B). This result seems marginal compared to the amount of viral DNA produced by a replication-competent vector system such as pTG3602, which generated ~4,000-fold more viral DNA than FK7 (compare Fig. A-3C and Fig. A-4C). We also examined pTG-JE1 DNA at P0 and observed some loss of the 2.9 kb pTG-JE1-specific band at 48 h after infection with HV.


A.
L
pTG-JE1
pTG3602
Passage
0 1 2 3 4 0 1 2 3 4
hpi
6 48 6 48 6 48 6 48 6 48 48 48 48 48 48 48
HV
- + + + + + + + + + + - - - - - -
DpnI
- - - - - - - - - - - - + - - - -
2.9
1.0
2.9
1.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
B.
Quantity (pg)
7000
6000
5000
4000
3000
2000
1000
0
Mock
P0/6h
P0/48h
P1/6h
P1/48h
P2/6h
P2/48h
P3/6h
P3/48h
P4/6h
P4/48h

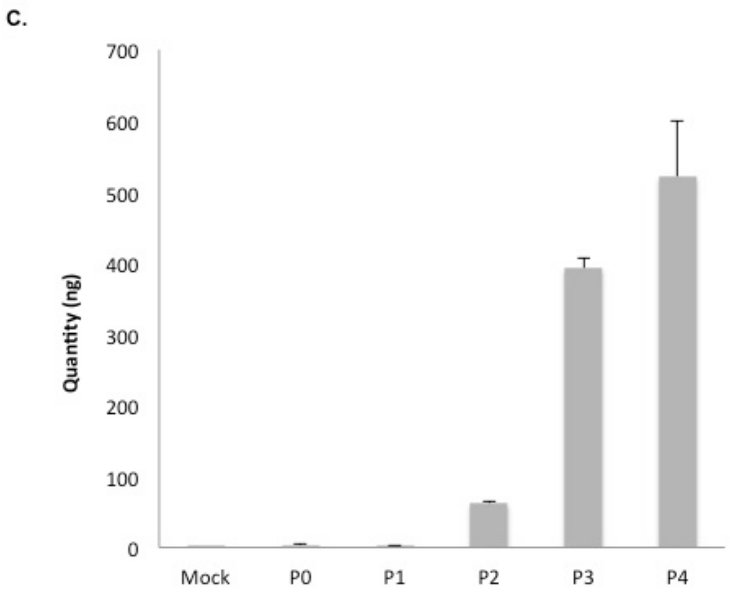


Figure A-5. A replication-defective infectious clone mimics the low efficiency of HDAd amplification.

Panel A. N52.E6-Cre cells transfected with *PacI*-linearized pTG-JE1 and coinfected with $(A5\text{-}A7)^2$ HV (left) or transfected with *PacI*-linearized pTG3602 alone, and amplified by serial passage with HV (pTG-JE1) or minus HV (pTG3602). With pTG-JE1, nuclear DNA was isolated at 6 or 48 hpi. With pTG3602, nuclear DNA was isolated at 48 hpi. Samples were digested with *HindIII* and analyzed by Southern blot using a ^{32}-P labeled probe of Ad5 nt 1-194 (A, lanes 2-11). Purple arrows designate pTG-JE1-specific bands, blue arrows designate $(A5\text{-}A7)^2$ HV, and green arrows designate pTG3602-specific bands. Lane 1 represents 0.1 µg of pTG-JE1 cut with *PacI* and *HindIIII* as a control. Lane 13 contains nuclear DNA from pTG3602-transfected cells from passage 0 digested with *HindIII* and *DpnI*.

Panel B. pTG-JE1-amplification was quantified by qPCR using primers surrounding the WT Ad5 packaging domain. Graphs represent averages of two data sets normalized to GAPDH levels. The data is represented as quantities generated using a standard curve generated by known quantities of pTG3602 plasmid.

Panel C. pTG3602 amplification was also quantified by qPCR as described for panel B. Note the difference is scale from panel B and panel C

It is possible that the defect in replication of HDAd is caused by the HV not providing a fast enough "kick-start" for replication of the vector by using the method of HDAd plasmid transfection, followed by HV coinfection. Extensive experiments have been performed to study HV function when HV DNA was introduced by viral infection or plasmid transfection (144). Hardy et al. showed that the method of Ad vector and HV plasmid cotransfection produced an Ad vector stock virtually free of HV contamination in a relatively short amount of time. In contrast, transfection of Ad vector plasmid DNA followed by infection with HV, produced Ad vector with a significant amount of HV contamination in a similar time frame (144). Earlier studies showed that Ad vector plasmid DNA replicated to 10,000 copies per cell if cotransfected with HV DNA, but replicated poorly if the HV was introduced by infection (144, 150). Together, these results of previous studies indicate that the yield of viral vectors was highest when cotransfection of the vector plasmid DNA with HV DNA was used. These published studies were performed for the production of first-generation viruses. The question still remains if cotransfection of the HDAd DNA with HV DNA would improve HDAd replication efficiency, and ultimately rapid amplification and production of the HDAd vector.

To address this question, we generated an infectious clone that contains the $(A5\text{-}A7)^2$ HV genome. We cotransfected this plasmid, after excision of the viral genome from the plasmid backbone, together with pFK7 into N52.E6-Cre cells. Lysates were harvested 72 h post-transfection and used to infect N52.E6-Cre cells, supplemented with $(A5\text{-}A7)^2$ HV infection, for serial passage. Results showed that this process was in fact less efficient than our original system (data not shown). We were not able to detect FK7 viral DNA sequences by Southern blot as late as P6. A side-by-side experiment, where pFK7 was cotransfected,

followed by HV infection, allowed for the rescue of HDAd at P4 consistent with our previous results (data not shown, Fig. A-3). One caveat to using transfection, whether of the HDAd DNA, the HV DNA, or both, is a problem with efficiency and reproducibility. We noticed in repeated experiments that rescue of the HDAd varied considerably from experiment to experiment. Sometimes HDAd rescue was observed as early as P3 in low amounts, whereas other times HDAd rescue was not observed until P5. This is not uncommon for HDAd systems (146). For the purpose of efficient HDAd production, the higher the transfection efficiency, the lower the number of passages required to achieve the maximum HDAd titer (141, 146). It is possible that if we optimize transfection efficiencies for both the HDAd and the HV in a plasmid backbone, we can improve replication efficiency and increase the production of HDAds.

Generation of HDAd FK7 using virus purified by equilibrium CsCl gradient centrifugation.

The last step in the Cre-lox-mediated system for virus production is purification of the virus vector. Even with the caveat of inefficient replication, we still observed rescue of the HDAd FK7 after P4 or P5 (Fig. A-3). To determine if we could successfully generate FK7 virus, free of HV contamination, by using our own modification of the Cre-lox system, we used FK7 lysates from P6 (data not shown) to infect N52.E6-Cre cells, coinfected with $(A5\text{-}A7)^2$ HV . When CPE was evident around 72 hpi, cells and media were collected and prepared lysates were subjected to three rounds of equilibrium CsCl gradient purification. Our results showed that we generated fully packaged FK7 virus, but that it contained HV contamination (Fig. A-6, blue arrows middle and lower bands). To our surprise, we also generated a large

amount of empty particles (Fig. A-6, upper bands). At first we were surprised to see this high amount of HV contamination, especially considering we did not observe any HV carried over to the next round of replication in our previous experiments (Fig. A-3, 6 hpi time points). There are several possible reasons for HV contamination. First, inefficient floxing of the HV viral genome is the most common. However, our Southern blot analysis to determine floxing efficiency did not reveal high levels of unfloxed, and therefore packagable, HV DNA (Fig. A-2) to the degree that we observed by CsCl gradient (Fig. A-6). Second, we used lysates from P6 because we observed the most amount of FK7 by Southern blot analysis. However, we also observed other bands present in these samples that could represent HDAd and/or HV genome rearrangements (data not shown). In future experiments, we will determine the nature of these bands. Furthermore, we should use an earlier passage, P4 or P5, since the chance of rearrangement of either the HDAd vector or HV may increase with higher passage numbers (146).

HDAds are invariably contaminated with HV, approximately 0.09 to 0.97% (143). This number is greatly reduced from first-generation Ads and is considered safe for the larger scope of determining total toxicity. For example, earlier studies showed that coinjection of HDAd with up to 10% HV contamination in mice did not reduce the duration of transgene expression or result in significantly higher toxicity compared with preparation with only 0.1 to 0.5% contamination (reviewed in (143). However, improvements need to be made to generate an efficient HDAd production system avoiding any HV contamination for the purposes of safety in human clinical trials and FDA approval. One such way is to physically block packaging of the HV by adding stuffer DNA sequences to the E3 region to render any recombinants too large to be packaged, based on the upper size limits on packaging

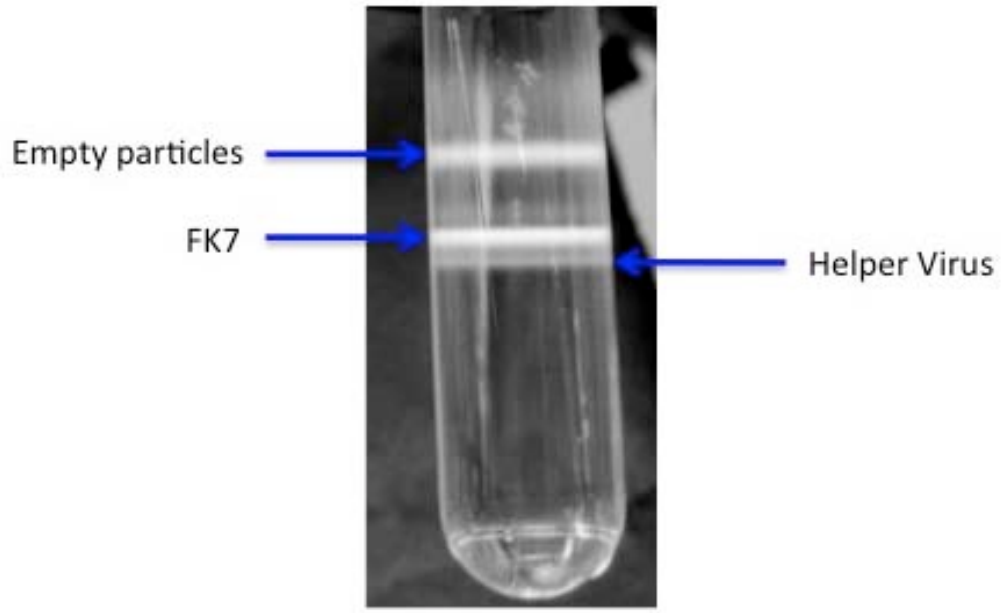


Figure A-6. Production of HDAd FK7. Plasmid pFK7 was amplified by serial passage with HV infection. Lysates from passage 6 were used to infect N52.E6-Cre cells coinfected with HV. FK7 virus stocks were generated by three subsequent step gradients of equilibrium cesium chloride centrifugation. The identity of each band is marked next to the blue arrow.

(146, 151). Our next step is to determine the culprit for the HV contamination we observed by CsCl gradient purification in the HDAd large-scale preparation (Fig. A-6).

Over the past decade the production problem of generating high-capacity HDAd vectors was addressed by utilizing a producer cell line that expressed high levels of Cre-recombinase, a HV resistant to mutations, and refined protocols from the first HDAd vector system (152). With these improvements, >1 x 10^{13} viral particles could be produced from 3 liters of cells within 2 weeks of initial vector rescue and with low HV contamination. Further improvements include a 10-chamber cell factory with adherent cells to produce HDAds on a larger scale with minimal HV contamination (153). In summary, Cre-lox is a powerful tool for genetic manipulation of Ads that allows for a HV to provide the necessary resources for viral replication and encapsidation of the HDAd, at the expense of the HV. We present an HDAd vector system in which we used a minimal packaging sequence, flanked by loxP sites, to decrease recombination between the HDAd and HV. We show that rescue of the HDAd in this system could take as long 10-12 days to amplify sufficient amounts of FK7 vector (Fig. A-3) and an additional 3-5 passages to obtain high titer cultures of FK7 that are essentially free of HV contamination (Fig. A-6). There are two factors that limit the purity of HDAd vectors produced with a HV. First is the stability and efficiency of selection against the HV. Second is the ability of the HDAd virus to be replicated and packaged efficiently by the HV. From our data it appears that neither the HV nor the HDAd performed optimally, and that this was due to a defect in HDAd DNA replication (Figs. A-4 and A-5). The E1-substituted viruses have everything they need for virus growth in E1-expressing N52.E6-Cre cells and

the HDAd geneome has the minimal sequences needed for replication and packaging of the genome, yet the system is flawed. The simplest interpretation of these results is that the HDAd is deficient in DNA replication. In our future experiments, we would like to optimize infection parameters such as MOI of the HV and HDAd, harvesting time, and cell culture characteristics to improve replication efficiency of the HDAd, at the expense of the HV (154). It would be interesting to determine if coinfection of the purified HDAd from CsCl gradients (Fig. A-6) with the HV would lead to an improved replication of the HDAd, or if the two viruses would still be in competition. Nevertheless, we expect that by providing the replication elements in a coinfection experiment, would greatly simplify production and improve consistency of our assay.

References

1. Cots D, Bosch A, Chillon M. Helper depedent adenovirus vectors: progress and future prospects. Current Gene Therapy. 2013;13(5):370-81.

2. Parks RJ, Graham SV. A Helper-Dependent system for Adenovirus vector production helps define a lower limit for efficient DNA packaging. J Virol. 1997;71(4):3293-8.

3. Volpers C, Kochanek S. Adenoviral vectors for gene transfer and therapy. J Gene Med. 2004;6:S164-S71.

4. Gilgenkrantz H, Duboc D, Juillard V, Couton D, Pavirani A, Guillet JG, et al. Transient expression of genes transduced in vivo into heart using first-generation adenoviral vectors: role of the immune response. Hum Gene Ther. 1995;6:1265-74.

5. Palmer DJ, Philip NG. Helper-depedent adenoviral vectors for gene therapy. Hum Gene Ther. 2005;16:1-16.

6. Hardy S, Kitamura M, Harris-Stansil T, Dai Y, Phipps ML. Construction of Adenovirus vectors through Cre-lox recombination J Virol. 1997;71(3):1842-9.

7. Yang Y, Nunes FA, Berencsi K, Gonczol E, Engelhardt JF, Wilson JM. Inactivation of E2a in recombinant adenovriuses improves the prospect for gene therapy in cystic fibrosis. Nat Genet. 1994;7:362-9.

8. Schiedner G, Hertel S, Kochanek S. Efficient Transformation of Primary Human Amniocytes by E1 Functions of Ad5: Generation of New Cell Lines for Adenoviral Vector Production. Hum Gene Ther. 2000;11:2105-16.

9. Palmer DJ, NG P. Methods for the Production of helper-dependent adenoviral vectors. Methods Mol Biol. 2008;1(433):33-53.

10. Schiedner G, Morral N, Parks RJ, Wu Y, Koopmans SC, Langston C, et al. Genomic DNA transfer with a high capaicty adenovirus vector results in improved in vivo gene expression and decreased toxicity. . Nat Genet. 1998;18:180-3.

11. Aoki K, Barker C, Danthinne x, Imperiale MJ, Nabel GJ. Efficient generation of recombiant adenovrial vectors by cre-lox recombination in vitro. Molecular Medicine. 1999;5:224-31.

12. Brown BD, Shi CX, Powell S, Hulburt D, Graham FL, Lillicrap D. Hlper-dependent adenoviral vectors mediate therapeuitc factor VIII expression for several months with minimal accompanying toxicity in a canine model of severe hemophilia A. Blood. 2004;103:804-10.

13. Ostapchuk P, Hearing P. Minimal *cis*-acting elements required for adenovirus genome packaging. J Virol. 2003;77(9):5127-35.

14. Chartier C, Degryse E, Gantzer M, Dieterle A, Pavirani A, Mehtali M. Efficient Generation of Recombinant adenovirus vectors by homologous recombination in *Escherichia coli.* . J Virol. 1996;77:6255-64.

15. Evans J, Hearing P. Distinct Roles of the Adenovirus E4 ORF3 Protein in Viral DNA Replicaiton and Inhibition of Genome Concatenation. J Virol. 2003;77(9):5295-304.

16. Hay RT, Stow ND, McDougall IM. Replicaiton of adenovirus mini-chromosomes. J Mol Biol. 1984;61:2555-8.

17. Bett AJ, Prevek L, Graham FL. Packaging capacity and stability of human adenovirus type 5 vectors. J Virol. 1993;67:5911-21.

18. Palmer DJ, NG P. Improved system for helper-dependent adenovirus vector production. Mol Ther. 2003;8:846-52.

19. Suzuki M, Cela R, CLarke C, Bertin TK, Mouriño S, Lee B. Large-scale production of high-quality helper-dependent Adenoviral vectors using adherent cells in cell factories. Hum Gene Ther. 2010;21:120-6.

20. Dormond E, Perrier M, A. K. From the first to the third generation adenoviral vector: what parameters are governing the production yield? Biotechnol Adv. 2009;27:133-44.