Stony Brook University

OFFICIAL COPY

The official electronic file of this thesis or dissertation is maintained by the University Libraries on behalf of The Graduate School at Stony Brook University.

© All Rights Reserved by Author.

Sequence Specific Inhibition of Adenoviral Replication by the AAV Rep78 ORF

A Dissertation Presented

By

Varsha Sitaraman

to

The Graduate School

In Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Genetics

Stony Brook University

August 2010

Stony Brook University

The Graduate School

Varsha Sitaraman

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Wadie Bahou, Professor and Vice Dean
Dissertation Advisor
School of Medicine

Patrick Hearing, Professor
Chairperson of Defense
Department of Molecular Genetics and Microbiology

Eckard Wimmer, Distinguished Professor
Department of Molecular Genetics and Microbiology

Paul Freimuth, Biochemist
Brookhaven National Laboratory

Nicholas Muzyczka, Eminent Scholar
Department of Molecular Genetics and Microbiology
University of Florida

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Sequence Specific Inhibition of Adenoviral Replication by the AAV Rep78 ORF

by

Varsha Sitaraman

Doctor of Philosophy
In

Genetics

Stony Brook University

2010

Hybrid Adenovirus/Adeno-associated viruses (Ad/AAV) combine the capacity, tropism and ease of production of Ad with AAV's ability for site-specific integration (SSI) into chromosome 19 AAVS1. The AAV Rep78 protein is required for SSI although it displays an inhibitory effect on Ad replication, particularly when co-expressed within the Ad backbone. However, strategies to construct an Ad/AAV focused on controlling Rep expression have met with limited success. We hypothesized that Rep's apparently cis-acting inhibitory effect on Ad replication could either be due to increased expression accompanying an increase in copy number, or due to a role for the sequence of the Rep ORF. To elucidate the relative contribution of the Rep ORF sequence and Rep protein levels on inhibition, we modified the 1866bp Rep nucleotide sequence *in silico* using synonymous codons. We generated two Rep coding sequences, Scrambled and Deoptimized which differed from the wild-type Rep78 nucleotide sequence by 20-30%, while encoding exactly the same protein. The Deoptimized sequence specifically uses codons in underrepresented pairs, expressing Rep78 protein at reduced levels due to codon pair bias. Codon pair bias refers to the preference for some codon pairs over other synonymous codons to encode the same pair of adjacent amino acids. Utilization of underrepresented codon pairs results in an ORF that is expressed at reduced levels, due to inefficient translation. Expression of the Scrambled, Deoptimized and wild-type Rep 78 ORFs within a first generation Adenovirus backbone (Ad/Scr, Ad/Deopt and Ad/wtRep) revealed dramatic results. Where Ad/wtRep was incapable of

replication, Ad/Scr and Ad/Deopt replicated as well as any other first generation Ad, indicating a clear role for a sequence specific signal in the inhibition of Ad replication. Modification of this signal allowed tolerance of a high level of Rep protein expression. The signal was localized to a ~135bp sequence within the Rep ORF. The identification of a sequence specific inhibitory signal for AAV Rep mediated inhibition of Ad replication explains the inconsistent and often frustrating results obtained with production of Ad/AAV over the years and paves the way for large scale production of integrating Ad/AAV

Table of Contents

List of Figures vi

List of Figures vii

Abbreviations viii

Chapter One 1

Gene Therapy 1

Viral Vectors in Gene Therapy 2

Retrovirus 3

Herpes Simplex Virus 5

Adenovirus 7

Adeno-Associated Virus 16

Dissertation Perspective 34

Chapter Two 37

Materials and Methods 37

Results 55

Discussion and Future Directions 73

References 87

Appendix 108

List of Figures

1 Organization of Ad genome and generations of Ad vectors....................................11

2 Organization of AAV genome..18

3 Model for AAV replication 23

4 Development of 911E4 derivative expressing siRNA against Rep78 39

5 Design of plasmid constructs 48

6 Scrambled/wtRep ORF sequence swaps..49

7Leaky expression from head to head orientation in tetracycline inducible cassette56

8 Flow cytometry analysis for EYFP expression from Ad/AAV-PF4FVIII 58,59

9 Analysis of modified Rep constructs..61

10 DpnI assay for viral replication 63

11 Excision assay for functional Rep expression 64

12 Localization of sequence specific signal I 67

13 Localization of sequence specific signal II...68

14 Sequence of putative sequence specific signal 69

15 Characterization of Ad/AAV capable of replication 70

16 Contribution of Rep expression to inhibition of Ad replication............................72

17 mfold mRNA secondary structure prediction 83

List of Tables

Table 1: TESS predictions for wtRep bp 1461-1596 80

List of Abbreviations

AAT	Alpha 1 Antitrypsin
AAV	Adeno Associated Virus
Ad	Adenovirus
Ad/AAV	Adenovirus/Adeno associated virus hybrid
ADA-SCID	Adenosine Deaminase –Severe Combined Immunodeficiency
ADP	Adenovirus Death Protein
ATPase	Adenosine Triphosphatase
BCA assay	Bicinchoninic Acid Assay
bp	Base Pair
CAR	Coxsackie Adenovirus Receptor
CDk21	Cyclin D Kinase 21
CFTR	Cystic Fibrosis Transmembrane Receptor
CPE	Cytopathic Effect
CREB	Cyclic AMP Responsive Element Binding Protein
CsCl	Cesium Chloride
DBP	DNA Binding Protein
Deopt	Deoptimized Rep78
dsDNA	Double Stranded Deoxyribo nucleic acid
E B/G/Y FP	Enhanced Blue/Green/Yellow Flourescent Protein
EDTA	Ethylene Diamine Tetraacetic acid
FBS	Fetal Bovine Serum
GALV	Gibbon Ape Leukemia Virus
HCC	Hepatocellular Carcinoma
HD-Ad	Helper Dependent-Adenovirus
HIV	Human Immunodeficiency Virus

HSC	Hematopoeitic stem cell
HSV	Herpes Simplex Virus
HSV-tk	Herpes Simplex Virus – thymidine kinase
IEE	Intergration Efficiency element
IL2GR	Interleukin 2 gamma chain receptor
IM	Intramuscular
IRES	Internal Ribosomal entry site
ITR	Inverted Terminal Repeats
LCA	Leber's Congenital Amaourosis
Mab	Monoclonal antibody
Min	Minute
ml	Milliliter
MLP	Major Late Promoter
MoMLV	Moloney Murine Leukemia Virus
mRNA	Messenger Ribonucleic Acid
NaOH	Sodium Hydroxide
Neo^r	Neomycin Phosphotransferase
NPC	Nuclear Pore Complex
ORF	Open Reading Frame
P19	Promoter at map unit 19 of AAV genome
P40	Promoter at map unit 40 of AAV genome
P5	Promoter at map unit 5 of the AAV genome
PBS	Phosphate Buffered Saline
pCMV	Cytomegalovirus promoter
PCR	Polymerase Chain Reaction
Pfu	Plaque Forming Units

PID	Primary immunodeficiency
PKA	cAMP dependent Protein Kinase A
PKI	cAMP dependent Protein Kinase Inhibitor
pRb	Phosphorylated Retinoblastoma protein
PRKX	Protein Kinase X
Q-PCR	Quantitative Polymerase Chain Reaction
rAAV	Recombinant Adeno Associated Virus
RBS	Rep Binding Site
RCA	Replication competent Adenovirus
RE	Restriction enzyme
RGD	Arginine-Glycine-Aspartate
RID	Receptor Internalization and Degradation protein
RMSSI	Rep mediated site specific integration
RT-PCr	Real Time-Polymerase Chain Reaction
Scr	Scrambled Rep78
SDS	Sodium Dodecyl Sulphate
SSS	Sequence specific signal
TR	Terminal Repeat
TRDD	Terminal Repeat – Double D
TRE	Tetracycline Responsive Element
Trs	Terminal Resolution Site
tTS	Tetracycline sensitive Transcriptional Silencer
Wt	Wild-type
X-SCID	X linked Severe Combined Immunodeficieny

Acknowledgements

The road to my PhD has been bumpy, frustrating at times but ultimately very rewarding. I am grateful to many, many people who have variously helped me out, supported me and kept me sane!

I am truly grateful to my advisor, Dr. Bahou, for cheerfully encouraging and supporting my science for all these years, for allowing me the freedom to follow my every hunch and for saying 'good work' even after meetings where I had only negative data to report! A big thank you to the various members of the Bahou lab – Dr. Dmitri Gnatenko who taught me so much basic science, from primer design to statistics; Dr. Valentina Schmidt who patiently answered all of my 'quick questions'; Jean Wainer for always being cheerful and being the person who fixes everyone's problems and Leslie Scudder who has all the answers and without whom T15 simply does not know how to have fun!

My committee members Dr. Patrick Hearing, Dr. Eckard Wimmer, Dr. Paul Freimuth and Dr. Nicholas Muzyczka have been wonderful. Their input has been invaluable and I realize I was a fortunate graduate student as meeting with my committee always left me motivated. I would like to especially thank Pat for patiently answering my numerous Adenovirus questions as well as for reagents, viruses and plasmid constructs through the years. I am truly grateful to Dr. Wimmer and his colleagues – Dr. Steffen Mueller, Dr. Steve Skiena and Dr. Charles B. Ward not only for their cutting edge technology which formed the basis of my thesis but also for their enthusiasm and interest this past year.

My friends and former graduate students Dr. Andrea Damon, Dr. Alanna Kennedy, Dr. Jackie Freudenberg, Dr. Sudeepti Southekal and Dr. Carmine Chiariello made graduate school so much fun! I am truly grateful to you all for being such a constant source of support and for ensuring that I never forgot that there is a light at the end of the tunnel! So many of my best memories from school are with you all!

A sincere note of thanks to Kate Bell, Dr. Peter Gergen and Dr. Gerald Thompsen from the genetics program for always being on my side.

My family and Tanuj have been wonderful and supportive through it all. I don't think I have expressed enough through the years how grateful I am to Appa, Amma and Sharanya, who have been so positive and encouraging (even when they were not quite sure whether I would really ever finish!) and for always

asking me how my viruses were doing. Your pride in my academic accomplishments really kept me going! A special thank you to my grandmothers – Kantimma and Sachumma for always understanding and encouraging me to do what I thought best.

Finally, I genuinely believe that I would not have been able to remain sane and finish my PhD without Tanuj. He always made me feel we were in this together – being equally thrilled when something worked and allowing me to vent and troubleshoot when things did not. Thank you for reassuring me when I was panicking, cheering me up when I was worried and for being such a tower of strength. Words cannot express how grateful I truly am. Thank you!

Fig 1 was reprinted with permission from Macmillan Publishers Ltd: Gene Therapy (Alba R. et.al., (2005) 12, S18–S27), copyright (2005)

Fig 2 was reprinted with permission from Macmillan Publishers Ltd: [Gene Therapy] (Merten O-W et.al., (2005) 12, S51–S61), copyright (2005)

Fig 3 was reprinted under a Creative Commons Attribution License from “Adeno-associated virus: from defective virus to effective vector” Manuel AFV Goncalves, Virology Journal, 2005, 2:43

Gene Therapy

Gene therapy is the treatment of disease with therapeutic genetic material. Though originally conceived for the treatment of monogenic disorders by the delivery of a functional copy of a faulty gene, the field has now expanded to include the treatment of polygenic disorders, cancers with the delivery of toxic genetic material and the induction of the immune response for DNA vaccines. Gene therapy can be *in vivo* where cells are treated within the patient or *ex vivo* where cells are removed from the patient, treated and re-infused. Successful gene therapy depends on the stable expression of the therapeutic gene in the target tissue for an appropriate period of time.

Since the first gene therapy clinical trial, for a primary immunodeficiency (PID) called Adenosine Deaminase - Severe combined Immundeficiency (ADA-SCID) in 1990, the field of gene therapy has seen much progress and many setbacks. The first ADA-SCID trial involved ex vivo treatment of the patients T cells with retroviral vectors (1). Though this trial was unsuccessful due to the inefficiency of gene transfer to long lasting progenitor cells, it laid the groundwork for a successful trial with hematopoietic stem cells in 2002 (2, 3). However, in the late 1990's and the early 2000's fatalities associated with two Phase I gene therapy trials which primarily evaluate safety and secondarily efficacy, delivered a setback to the progress of gene therapy, causing researchers and regulators to re-evaluate their understanding of the inherent risks of such a treatment and the risk to benefit ratio of gene therapy. In 1999, a teenager Jesse Gelsinger was injected with a high dose of a therapeutic first generation adenoviral vectors in a Phase I clinical trial for ornithine transcarbamylase deficiency. Tragically, he developed a massive immune response to the Ad capsid and died within 98 hours of treatment (4, 5). The high point of the initiation of the trial for ADA-SCID was followed by another tragic low when 5 children involved in a gene therapy trial with another retroviral vector for another PID (X-SCID) developed leukemia due to insertional mutagenesis and activation of a cellular proto-oncogene by the retroviral vector (6-8). These fatalities highlighted the fact that in spite of all previous animal studies, researchers still do not know for sure the effect of a particular treatment in humans and that one needs to weigh risk vs benefit for any such therapy.

Recently, there has been a resurgence with several positive Phase I trials for gene transfer reported. Further, the use of Adenovirus as an oncolytic vector for the treatment of cancer has gained rapid ground, proving to be safe and efficacious. Several such vectors are currently being tested in Phase IV trials and are currently available in market in China. Exciting results with Phase I/II trials

include the treatment of Lebers congenital Amaurosis (9-11),a disease characterized by progressive blindness using a recombinant Adeno Associated Virus (rAAV) vector and therapy of Adrenoleukodystrophy, a severe brain de-myelinating disease using a lentiviral vector (12) have been reported.

Viruses as Gene Delivery Vectors

Viruses are ideal gene delivery vectors as they evolved to infect cells and deliver their DNA within. Non viral gene transfer vectors including liposomes, direct injection and particle bombardment have been investigated, attractive as they do not elicit an immune response, and can be administered repeatedly. Cationic liposomes have been used in Phase I and II human trials for expression of the major histocompatibility class I gene HLA-B7 in melanoma patients as well as expression of IL-2 in patients with melanoma, sarcoma and renal cell carcinoma (13-15). Preliminary positive data with regression of proximal and distal tumor nodules was observed in some patients and the therapy was found to be safe and well tolerated. Recently, in vivo electroporation of a plasmid expressing Interleukin 12 was used in a Phase I trial for patients with metastatic melanoma (16). The procedure was tolerated well and partial tumor regression was observed in 42% of patients. Further, interesting advances such as the use of nanoparticles for delivery of tumor suppressor genes to tumors and liposomes coated with polyethylene glycol to cross the blood brain barrier and deliver genes to the brain have been reported in mice (17). However, progress with gene transfer by non-viral vectors have been stymied by the extremely low gene transfer efficiency.

The four main viruses widely used for gene therapy research are retroviruses and lentiviruses, herpes simplex virus, Adenovirus and Adeno-associated virus. The field evolves as more is known about these vectors, allowing their modification to adapt them to meet the needs of individual diseases. Each of these viruses and their applications as vectors in gene therapy are discussed below.

Retroviruses and Lentiviruses

Retroviruses are small enveloped RNA viruses, consisting of two identical single stranded positive sense RNA genomes enclosed in an enveloped capsid. They have a genome flanked by Long Terminal Repeats (LTR) and 4 main genes *gag*, *pol*, *pro* and *env*. *gag* encodes internal structure proteins, *env* encodes the surface glycoprotein and the transmembrane protein of the virion which together are responsible for interacting with the host cell receptor and *pro* encodes the viral protease. *pol* encodes the reverse transcriptase which is required for the synthesis of the DNA copy of the RNA genome and an integrase which is required for the integration of this DNA copy into the hosts genome. The integration of a DNA provirus is essential for viral replication, as mutations that block integration have been found to inhibit replication (18). Lentiviruses are a group of complex retroviruses which carry accessory gene which regulate and coordinate viral gene expression. Lentiviruses also differ from other retroviruses in their ability to infect non-dividing cells as most other retroviruses are incapable of traversing the nuclear membrane and can thus infect only dividing cells where the nuclear membrane is dissolved.

Retroviral vectors have a transgene capacity of about 10Kb. Most retroviral vectors for gene therapy are derived from the Moloney Murine Leukemia Virus (MoMLV) while HIV-1 based vectors form an important focus in the development of lentivirus vectors. A characteristic feature of all retroviruses is their ability to integrate into the host cell genome. Both Retro and Lenti viruses have been shown to prefer actively transcribed sites of the genome to integrate (19, 20) and the distribution of the sites of integration have been found to vary based on the virus used and the cell type infected(21). Retroviruses have been shown to integrate preferentially near transcription start sites, close enough for LTR mediated enhancer effect. Lentiviruses also prefer active transcription units, but have been found to not integrate downstream of the transcription start site (20). An interesting example of the effect of retrovirus mediated insertional mutagenesis is the discovery of the insertion of a Gibbon Ape Leukemia Virus (GALV)/MLV related provirus upstream of a duplicated pancreatic amylase gene in the genome of a common ancestor of great apes and Old world monkeys, resulting in the expression of amylase in the parotid gland (22). This effect of retroviral insertion can be traced to modern humans in salivary amylase expression and may have influenced the dietary preference of apes and humans (23).

The capacity for integration and long term transgene expression has made retroviruses ideal for ex vivo gene therapy. Retroviruses were the first viruses to be used in clinical gene therapy and have been used in multiple clinical trials for the treatment of primary immunodeficiencies (PID) including ADA-SCID, SCID-X1, Chronic Granulomatous disease and Wiskott Aldrich Syndrome. The two pioneering clinical trials with PID which shed light on the advantages and inherent dangers of retrovirus mediated gene transfer are discussed below. The first PID to be treated with gene corrected Hematopoietic stem cells (HSC) was Adenosine Deaminase – Severe Combined Immunodeficiency (ADA-SCID), shortly followed by treatment of SCID-X1. The lack of ADA causes an immunodeficiency characterized by impaired development of T,B and Natural Killer (NK) cells and recurrent infections. Corrected HSC infected with an GALV/MLV vector carrying the cDNA for ADA were used (3). Since corrected cells did not have a selective growth advantage, a conditioning, myeloablative treatment which reduce the number of endogenous HSCs was given prior to corrected cell infusion. 9 out of 10 original patients enrolled in Italy (3) showed a significant improvement in immune functions. No insertional mutagenesis events have been linked with gene therapy for ADA-SCID.

SCID-X1 is an X linked SCID due to mutations of the γ chain of the receptor for interleukin 2 (IL2). Patients lack T and NK cells and carry impaired B cells. The first trial for gene therapy for SCID-X1 was performed in France based on ex vivo γ chain transfer using an MLV vector to CD34+ HSCs and in London using a GALV based vector. 17/20 patients enrolled benefited from gene therapy (24, 25). However, 2 to 6 years post treatment 4 patients from the French trial and 1 patient from the English trial developed leukemia due clonal expansion of T cells following upregulation of host cellular oncogenes by the MLV LTR (6-8). These serious adverse events brought out the dangers of retroviral integration and lead to an increased interest in the use of lentiviral vectors which possess a modified integration profile. The difference in safety profile between the ADA-SCID trials and the SCID-X1 trials also reveals the possible role of transgene control in adverse effects, as IL2GR is a signaling molecule that induces cell division.

Recently, a succesful Phase I trial for Adrenoleukodystrophy using lentivirus vector was reported (12). Currently, Phase I clinical trials for Wiskott-Aldrich syndrome, an X linked disease characterized by eczema, thrombocytopenia and immune deficiency, using lentiviral vectors have been conducted by two separate groups (26) and t preliminary results of safety, lack of clonal T cell populations and efficacy reported.

Herpes Simplex Virus

HSV-1 and HSV-2 are neurotropic pathogens associated with a number of skin diseases from herpes labialis and herpes genitalis to the life threatening neonatal herpes and herpes encephalitis (27). It is an enveloped virus with a linear double stranded DNA (dsDNA) genome of 152Kb, carrying 74 separate genes. The genome consists of 2 unique sequences, one longer than the other (U_L and U_S). Each of these sequences are flanked by inverted terminal repeat sequences – with U_L flanked by Terminal Repeat ($TR_{L)}$ and Internal Repeat (IR_L) and U_S being flanked by IR_S and TR_S . Copies of an 'a' sequence carrying packaging signals lie between the two IRs and at each TR.

The advantages of using HSV as gene therapy vectors include their broad host cell range, ability to infect dividing and non-dividing cells (28, 29) and their ability to exist stably as episomes in neuronal cells (30, 31). As only half of its genes are required for replication in cell culture, replication competent viruses with a large transgene capacity can be made (32, 33). A robust immune response and cell toxicity has been found to be associated with HSV infection. However, this response has been harnessed for the production of oncolytic HSV-1 based vectors and DNA vaccines.

The two main types of HSV based vectors used are amplicon vectors and replication attenuated vectors. Amplicon vectors are plasmids made up of repeated units of the transgene, a packaging signal (pac) and an HSV origin of replication (34). When introduced into a cell along with HSV helper functions, these amplicons replicate and are packaged as head to tail concatemers into infectious HSV virions. As the only HSV genome elements these vectors carry are the pac and ori, a large transgene size (theoretically up to 152Kb) can be included. Further, based on the transgene size, multiple copies of the transgene can be packaged within one virion. Recent developments in amplicon production include the cloning of the entire HSV functions without pac signals into a Bacterial Artificial Chromosome (BAC) (35), which has reduced contamination of amplicon preps with wtHSV. Helper free amplicon systems have been used for the delivery of nerve growth factor and brain delivered neurotrophic factor, antioxidant enzymes and anti-apoptotic proteins for the treatment of brain injuries (27, 36, 37). Amplicons have also been used for targeted gene delivery of therapeutic genes in diseases of the central nervous system such as Parkinson's disease (38) and in a mouse model of Friedrich's ataxia (39). Attempts to use amplicons as DNA vaccines as they can allow presentation of antigens to MHC Class I and II molecules without the expression of HSV genes have been

documented (40). However, the clinical applications of these studies have been limited by difficulties in scaling up of amplicon vectors.

More success has been reported with the use of replication attenuated HSV vectors as oncolytic vectors. These vectors have deletions in genes (such as HSV-TK and HSV-RR) which are required for replication of the virus in non-dividing cells and are thus capable of replication only in dividing (tumor) cells. Clinical trials (Phase I) for multiple HSV-1 derived oncolytic viruses for colorectal carcinoma (41), melanoma (42), breast cancer (43)and malignant glioma (44) among others have been reported. All studies reported safety and toleration of HSV vectors. An increase in patient survival has been reported with HSV1716, but other trials reported disappointing early efficacy.

Adenovirus

Adenoviruses were originally identified in adenoid tissue in 1953 (45). Some serotypes of Ad were found to be oncogenic and capable of transforming newborn Syrian hamster cells (46, 47). However, no human Adenovirus has been associated with oncogenicity. While generally associated with a mild self limiting infection in immunocompetent patients, serious, fatal infections can occur in immunosuppressed individuals. They are dsDNA viruses belonging to the family Adenoviridae, and are found to be associated with respiratory disease, conjunctivitis and gastroenteritis. They can be divided into five clades: *Mastadenovirus* isolated from mammals and including all human adenoviruses, *Aviadenovirus* from birds, *atadenovirus* with high A-T content, *Siadenovirus* from reptiles and birds and a fifth clade with only a single member, isolated from sturgeons. Human adenoviruses have further been classified into fifty one serotypes based on their neutralization to antisera, the epitopes for which are usually the hexon and fiber domain.(48). Serotypes are further subdivided into subgroups based on their ability to agglutinate red blood cells, a property of the central shaft of the viral fiber protein. The most widely studied and used adenoviruses are Ad2 and Ad5.

Virion Structure: Adenoviral moieties are non enveloped and consist of an icosahedral capsid surrounding a DNA core. The capsid consists of three major structural proteins – the hexon, penton and the fiber. Each of the 20 facets of the capsid is made up of 12 hexon homotrimers each. At the vertices are penton complexes which have a penton base holding an antenna like fiber. The fiber consists of the shaft, which varies in length between different serotypes, ending in a globular knob domain. Attachment of Ad to its host cell surface requires the fiber knob making contact with the host cell receptor, followed by RGD motifs in the penton base binding with host cell integrins $\alpha_v\beta_3$ and $\alpha_v\beta_5$ (49). Adenovirus serotypes 40 and 41 bear two kinds of fiber proteins on their capsid, possibly increasing their tropism. This contact is followed by internalization into clathrin coated vesicles. Other minor capsid proteins include proteins II – IX and IIIa, and are required mainly for stabilizing interaction between the major proteins of the capsid. Protein VI is required for the disruption of the endosome during infection, post internalization. The viral core consists of an ~36Kb genome and five known proteins. Protein VII is the most abundant core protein. A basic, arginine rich protein, along with V and μ, it condenses with the DNA. The terminal protein is covalently attached to the 5' end of each DNA strand by formation of a phosphodiester bond between serine 562 and the terminal cytosine. It acts as a primer during DNA replication. Finally, the core contains p23 nuclease, which

serves to cleave precursors of several viral polypeptides as well as in escaping endosomes during infection.

Genome Structure (Fig.1) The complete sequence of 9 human adenoviruses has been determined. Ad2 and Ad5, the best studied serotypes, were among the first to be sequenced (50, 51). DNA sequences coding for structural proteins, DNA replication and virion assembly are largely conserved across genera, whereas sequences coding for proteins involved in host specific activities (E1, E3 and E4) are not. The genome is flanked by identical short Inverted Terminal Repeats (ITRs) that serve as origins of replication. The terminal protein is found covalently attached to the 5' cytosine residues of each ITR. Between the ITR and protein coding region E1A, extending from bp194-358, are seven repeats of the packaging domain (ψ) which is required for the encapsidation of the Ad genome into a virion (52). The location of the packaging domain with respect to the end of the genome is important for efficient packaging of the genome into the virion (53). Proteins are coded for by both strands of the genome, and coding capacity is maximized by differential splicing and the use of alternate poly(A) sites. Interestingly, the study of Ad resulted in the discovery of introns and mRNA splicing(54, 55) . The genome can be segregated based on time of expression, such that E1A, E1B, E2A, E2B, E3 ,E4 and VA are the early genes expressed before DNA replication, IX, Iva2 and E2late are delayed early and L1-L5, transcribed from a common major late promoter, are late expressing, generally expressed after the commencement of DNA replication. However, a low level of transcription from the MLP can be detected early in infection (56). Conventionally maps are drawn with E1A at the left end.

Genome expression and lytic cycle: In most cases, the infectious lytic cycle of Ad starts with the binding of the C-terminal knob domain of the fiber with the host cell receptor, which is the Coxsackie B virus and Adenovirus Receptor (CAR) receptor for Ad subgroup A, C, D, E and F and CD46 for subgroup B (57, 58). The use of CD46 as receptor allows subgroup B viruses to infect cells including hematopoietic cells which are refractory to infection by other subgroups. This contact is followed by contact between the Arginine-Glycine-Aspartate (RGD) motifs in the penton base with host $\alpha_v\beta_3$ and $\alpha_v\beta_5$ integrins on the surface of the host cell membrane. This interaction leads to the detachment of the fiber and the endoyctosis of the fiberless virion into clathrin coated vesicles (59). Subgroup F serotypes Ad40 and Ad41 alone do not have RGD motifs and do not bind to integrins for internalization (60).

Within approximately 15 minutes of docking at the host surface, viral particles escape from the endosome into the cytoplasm and are transported to the

nuclear pore complexes (NPC) on microtubules. Along the way, an ordered disassembly of the structural proteins making up the virion occurs, culminating with the cleavage of protein VI which links the Ad genome with the capsid. Interaction with the NPC is required for further uncoating of the virion (61). Finally, viral DNA and associated protein VII alone enter the nucleus.

Once inside the nucleus, the E1A gene is the first gene to be transcribed. The role of early gene expression is threefold – Induction of the S phase of the host cell cycle; suppression of host immune response and activation of genes required for DNA replication. E1A transcripts are processed into two major mRNAs encoding the 12S and 13S E1A proteins. They are differentiated by the presence of a 46 aa conserved region 3 present on 13S. They stimulate host cell entry into S phase by interacting with the Rb-family proteins (62). Preventing the interaction of pRb with E2F results in the constitutive activation of sites regulated by the E2F family of transcription factors (63, 64), including genes involved in the regulation of cell cycle. The E1A proteins and the CR3 in particular is responsible for the trans-activation of all other Adenoviral genes by interaction with cellular transcription factors. 13S protein activates transcription from all the early virus promoters (65, 66) while the 12S protein specifically transactivates the E2 early promoter. Cell cycle de-regulation by E1A results in the accumulation of tumor suppressor p53. To prevent apoptosis due to this accumulation, E1B is the next gene to be transcribed, yielding two distinct proteins, E1B19K and 55K. Both proteins are involved in the inhibition of apoptosis(67, 68).

The proteins of the E2 gene are directly involved in viral DNA synthesis. The DNA Binding Protein (DBP) is expressed from the E2A locus. It plays a role in several stages of the viral lytic cycle and binds and protects single stranded viral DNA produced during replication from nuclease action. The Ad polymerase (Ad pol) and pTP(terminal protein) are expressed from the E2B locus. They exist as a stable heterodimer. As mentioned earlier, pTP is found covalently linked to the 5’ cytosine of each ITR. Polymerization begins with the addition of the first nucleotide to pTP by Ad pol. Genome replication proceeds by the strand displacement method, with stabilization of the single stranded intermediate by DBP.

The E3 locus encodes several proteins, most of which are involved in the inhibition or evasion of the host immune response(69). These include the E3 19K protein, RID (Receptor Internalization and Degradation) protein, E3 14.7K and E311.6K. The 11.K protein, also called the Adenovirus Death Protein (ADP), promotes cell death and lysis for the release of mature virions(70, 71) .

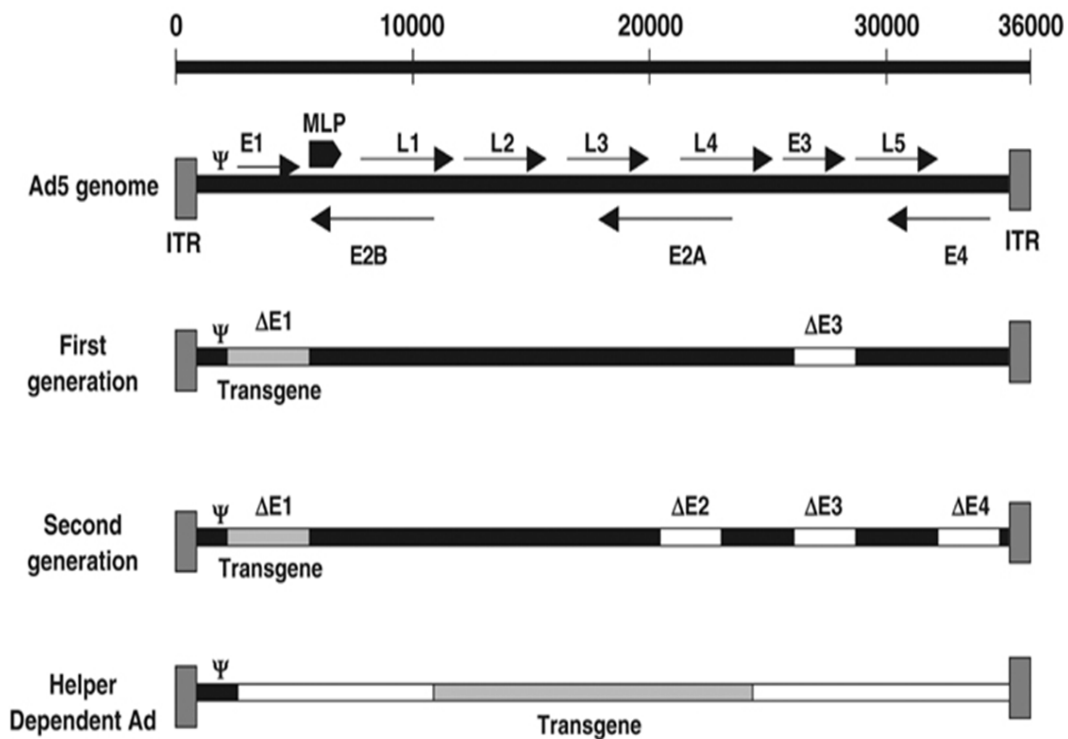
Multiple proteins with disparate functions are encoded by the E4 region (72). 7 identified ORFs give rise to proteins, E4ORF1, ORF2, ORF3, ORF3/4, ORF4, ORF6, ORF6/7. Little is known about E4ORF1, E4ORF2 and E4ORF3/4. E4ORF4 protein has been shown to down regulate expression from the E4 region, regulate mRNA splicing and induce p53 mediated apoptosis (73). E4ORF3 and E4ORF6 are the best studied E4 proteins. They mediate a number of different functions and either protein has been shown to be sufficient for replication of E4 deleted Adenoviruses in cell culture (74). Both proteins inhibit DNA repair complexes, but play opposing roles in interacting with E1B55K in the regulation of p53. Further, each protein has been shown to also possess unique non-overlapping functions (75). Finally, the E4ORF6/7 protein increases efficiency of DNA replication by enhancing production of the E2 gene products and induces cellular E2F promoters by binding directly with E2F (76, 77).

Viral DNA replication is initiated at either ITR, which serve as replication origins. Priming begins with the pTP attached to the terminal cytosine residue of the ITR, forming a pTP-dCMP linkage. Elongation proceeds by polymerization to the 3'OH of the pTP-dCMP bond and employs a 'jumping back' mechanism. The Ad template strand begins with repetitive sequence 5'GTAGTA-3'. Replication begins with the second 'GTA' repeat being copied first and the entire newly synthesized complex then jumping back to base pair at the terminus(78). *In vivo* studies indicate that replication takes place in two stages (79). First, DNA synthesis is initiated and elongation proceeds along only one of the two strands, resulting in a duplex with one parent and one daughter strand, and one displaced single stranded parent strand The single stranded template then circularizes, pairing at its complementary ITRs at each end forming a panhandle structure (80). This double stranded ITR has the same structure as the parental genome, and replication is initiated and a complementary strand synthesized.

Adenoviral late genes begin to be expressed efficiently after the onset of viral DNA replication. Most Adenoviral late genes are expressed from regions L1-L5, expressed under the MLP. A single pre-mRNA is processed by differential splicing to give rise to 15- 20 different mRNA moieties. Most of the Ad late proteins are either structural proteins or involved in assembly of the capsid. The 100K protein encoded by the L4 region binds to the hexon soon after translation, and along with pVI transports it to the nucleus, where virion assembly takes place. Evidence exists to support both conflicting theories of virion assembly - that assembly of the virion involves insertion of the genome into a preformed capsid, or formation of the capsid around the genome core (81-84). Cell lysis and release of progeny virions occurs in a process involving the E3 ADP.

Fig.1: Organization of the Adenovirus genome: The dsDNA genome is flanked by ITRs. Genes are encoded by both strands, with conventional maps drawn with E1 on the left. E1, E2, E3, E4 and VA are the early genes, expressed before DNA replication, while L1-L5 are the late genes, expressed mainly after initiation of DNA replication. Ψ: packaging signal; MLP : Major Late promoter.

Generations of Ad vectors for gene therapy are obtained by deletions of various essential Ad genes, resulting in replication deficient vectors with increasing transgene capacity and decreasing immunogenicity.



Reprinted with permission from Macmillan Publishers Ltd: Gene Therapy (Alba R. et.al., (2005) 12, S18–S27), copyright (2005)

Adenovirus as gene transfer vectors: The Adenovirus is the most commonly used vector in gene therapy trials, accounting for 24% of all clinical trials worldwide (The Journal of Gene Medicine –Clinical Trials website, www.wiley.co.uk/wileychi/genmed). A number of characteristics account for its popularity. These include its wide tropism, well characterized genome, ease of production with multiple packaging cell lines and easy to manipulate genomes, large transgene capacity and ability to be grown to high titers (85).

Tropism**:** Ad2 and Ad5, the most commonly used serotypes for Ad vectors, infect using CAR(86). CAR is widely distributed, making it difficult to target infection with Ad vectors. Simultaneously, CAR is scarcely expressed in some valuable target cells including hematopoietic stem cells. To achieve targeted infection into desired cell types, various modifications of the Ad fiber and penton base have been attempted. One approach has been the use of bispecific conjugates which re-direct receptor binding. Harari et al. directed specific targeting of vascular endothelial cells expressing E-selectin by the use of an anti-selectin E Mab conjugated with an anti-flag antibody which was bound to an Adenovirus expressing the flag epitope (87). Variations of this approach coupled with mutations of the intrinsic receptor binding domains in the fiber and penton of the vector has been shown to result in a vector with a completely modified tropism. The ability to infect a wider range of cells has been attempted by the replacement of the fiber knob domain (responsible for contact with CAR) with the corresponding domains of Adenoviruses of subgroup B which infect using CD46 (57, 88). This modification results in the modified vectors infecting cells which express CD46, including hematopoietic cells (89)

Generations of Adenoviral Vectors (Fig.1): In the development of Adenovirus as gene therapy vectors, various deletions within the genome have resulted in three generations of modified vectors with progressively decreasing immunogenicity and increasing transgene capacity. A maximum of 105% of the wild-type (wt) Ad genome size can be packaged into a recombinant vector (90).

First generation vectors: These vectors have transgene cassettes replacing the E1 and/or the E3 regions (ΔE1ΔE3 Ad), and have a total packaging capacity of up to 8.5Kb. Expression of E1, which is necessary for viral replication, is provided in *trans* by packaging cells lines such as 293 (91) or 911 (92). E3 expression is not required for in vitro viral replication (93). For viral viability, deletions in the E1 region should not affect the left ITR and ψ, or extend into the pIX region. Original strategies for recombinant virus construction involved recombination between the linear left and right ends of the virus in 293 packaging cells (94). More recently, strategies have evolved involving recombination in

bacterial cells, resulting in an isolatable recombinant infectious clone on a plasmid backbone. This construct is linearized to release the plasmid backbone and transfected into packaging cells (95). Though first generation viruses have found wide use, some drawbacks have become apparent. Recombination of the virus with E1 sequences in 293 packaging cells resulted in the production of replication competent Ad (RCA). This problem was bypassed by the construction of E1 complementing cell lines such as 911 which avoided regions of homology between the recombinant virus and the Ad sequences carried by the cell line. The other major drawback of first generation viruses is the considerable immune response it elicits (96). Cellular factors with homology to Ad E1 have been shown to be sufficient to activate a low level of Adenoviral gene expression (97). This results in the destruction of transduced cells and transient expression of transgenes.

Second generation vectors: To reduce or eliminate the low level replication seen with ΔE1 vectors, further deletions in E2 or E4 were constructed (98-101). These vectors have an increased transgene capacity of up to 14Kb and required specialized packaging cell lines which complement deletions in these necessary genes (102, 103). Deletions in E2 genes, especially Ad pol, completely eliminated replication of these genomes and resulted in sustained expression of transgene expression in immunocompetent mice (104). Association of transgene persistence with E4 deletions was less obvious (105) and was found to depend on tissue and promoter specific regulation (106, 107).

Helper dependent Adenovirus (HD-Ad): These vectors are devoid of all Adenoviral genes, and carry only the ψ and ITRs flanking the transgene. Since a genome less than 27Kb is packaged inefficiently, non coding stuffer DNA is used to increase the size of the genome (108). Necessary replicative functions are provided by a helper virus. Various strategies to minimize helper virus contamination have been attempted, including mutating the packaging signal to prevent packaging of the helper virus (109) and minimizing the size of the helper virus to enable separation of the virions (110). The most effective strategy developed thus far involves a ΔE1 helper virus with the packaging signal flanked by lox sites (111). Introduction of these viruses into 293Cre cells results in excision of the packaging signal without affecting expression of proteins necessary for Ad replication and packaging (111). A low level of <0.1% helper virus contamination is still seen with this system. The cause for this low level of contamination has been variously explained as due to recombination between the packaging signals of the two viruses (112) or inefficient expression of Cre in packaging cells (113). HD-Ads are found to be associated with sustained transgene expression. Though induction of the innate immune system due to the

Ad capsid is unavoidable with this construct, sufficient reduction in clearance of transduced cells has been seen with HD-Ad, leading to sustained transgene expression (114, 115) .

Reduction of innate immune response: Apart from the cellular immune response to the low level of viral replication, a substantial innate immune response is seen in response to the Ad capsid of all generations of vectors(116, 117). This response causes inflammation of the target and surrounding tissues, transient transgene expression, disallows retreatment with the same vector and can also lead to severe systemic shock. A tragic consequence of the induction of the innate response was seen with the death of an individual treated in a Phase I clinical trial for Ornithine Transcarbamylase deficiency, with a high dosage of a second generation Ad vector (118). Various strategies to avoid this antibody mediated vector neutralization have been attempted. These include the use of Adenoviral serotypes from other species, administration of vector with immunosuppressive drugs and coating of the vectors with inert chemicals such as Polyethylene Glycol ('PEGylation') to prevent exposure of the viral capsid (119-121).

Application of Ad vectors in gene therapy: As described, a considerable cellular and innate immune response is induced by infection with Adenoviral vectors. Though minimizing this response has been taken into account for the development of Ad vectors, this immune response still remains a sufficiently problematic barrier for the successful use of Ad for long term gene transfer. However, in the case of therapy for cancer and the development of genetic vaccines, destruction of transduced cells and the induction of an immune response can be of advantage. Currently the majority of clinical trials with Ad vectors is for the treatment of cancer. The first clinical success with an Ad vector for cancer therapy came with ONYX015 in 1997 (122). This Ad is modified in E1B, allowing the vector to replicate only in cells with defective p53 (122). Promising results were obtained with PhaseI clinical trials, with 27/30 patients displaying over 50% reduction in their tumor volumes, making it a widely studied oncolytic virus. The mode of action of the viral vector was further analyzed after Phase II trials revealed cell death in both p53 positive and negative cells (123). Lack of a clear answer to the mechanism of action has slowed the development and use of this vector.

Other strategies to utilize Adenoviral vectors against cancer involve delivery of sensitizing drugs or tumor suppressors. Suicide gene therapy which involves localized delivery of enzymes to the tumors which convert systemically administered pro-drugs into chemotherapeutic drugs is an elegant prospect. For

example, the delivery of HSV thymidine kinase gene to tumor cells will make them activate the systemically administered pro-drug gancyclovir (124). As multiple cancers are associated with loss of function mutations of the tumor suppressor gene p53 and delivery of p53 is associated with increased sensitivity to radiation, several vectors delivering wild-type p53 have been developed. Increased efficacy of these vectors is seen when in combination with radiation or chemotherapeutic drugs (as was seen with ONYX015). A current drawback in oncolytic virus delivery is the localized distribution of the virus along the path of the needle following injection. An interesting approach to increase the intra tumoral spread of the virus has been to co-express relaxin, a cell matrix degradative protein (125) or hyaluronidase (126) and has been shown to help the virus transverse the connective tissue which normally blocks virus spread within tumors. Oncolytic Ad vectors have proven safe and efficacious in a variety of trials. For instance, currently phase IV trials of gene therapy with rAd carrying p53 are ongoing for thyroid, head and neck and maxillofacial tumors, among others (source: http://clinicaltrials.gov). Further, several products including oncolytic Ad H101 similar to ONYX015 based have been approved for patient use in China and are available in the market.

The use of Ad as a genetic vaccine has not been as straightforward or as successful as their use as oncolytic vectors. Though Ad based vaccines for multiples diseases including Tuberculosis(127), malaria (128), rabies (129), influenza (130) and leishmania (131) are currently being developed, a blow to the field was the Phase II STEP trial for an HIV vaccine with an Adenovirus5 expressing conserved antigens of HIV-1 to induce T cells(132, 133). Though promising results had been obtained with early phase trials, the trial had to be stopped and unblinded due to lack of efficacy and a trend indicating an increase in infection of males with pre-existing antibodies to Ad5. The cause of this trend has not been isolated and remains a worrisome obstacle in the development of genetic vaccines.

Adeno-associated virus

The Adeno-associated virus is a small ssDNA virus belonging to the family *Parvoviridae*. It was originally isolated as a contaminating virus in preparations of Adenovirus and was found to be dependent on Ad for productive infection, and thus named. Later, several other viruses including Herpes-simplex, Epstein Barr, and cytomegalovirus were found to be capable of providing helper virus functions. AAV have been found to be widespread in their hosts which include monkeys, horses, birds, cows, sheep and humans, capable of infecting a number of different organs and organ systems. Gao et al (134) found that a hypervariable region in the AAV capsid was capable of evolution by homologous recombination during the course of a natural infection of non human primates resulting in viruses with altered tropism. This has been suggested as a possible explanation for the widespread dissemination of AAV seen in multiple tissues. AAV2 is the most widely studied AAV family member. No pathology has been convincingly linked with AAV2 infection of humans.

AAV can undergo a lytic or lysogenic life cycle. Uniquely, in cell culture in the absence of helper virus functions, the virus establishes a persistent infection by integrating site specifically into the AAVS1 site on chromosome 19 (q13.3qter) of the host genome in humans and non human primates (135-138). However, viral sequences are also found as concatemers of the AAV genome in an extrachromosomal form. Superinfection with helper virus has been shown to rescue the integrated virus *in vitro*. This rescue and replication of the AAV genome can also be triggered by stressing the host cell by exposure to genotoxic stimuli (139). It must be mentioned that only a single incidence of naturally occurring in vivo integration in 2 human testis tissue samples has ever been identified and reported (140). In every other case, AAV has been found to exist extrachromosomally..

Virion and Genome Structure (Fig.2): The AAV2 genome is a linear single stranded DNA of about 4.7Kb. Both sense and antisense strands are packaged into virions with equal frequency. The genome consists of T shaped inverted terminal repeats flanking two open reading frames (ORFs), Rep and Cap. The Rep ORF encodes Rep78 and alternately spliced Rep68 from promoter p5 and Rep52 and Rep40 from promoter 19. The p19 promoter lies within the coding sequence for the larger Rep proteins. The Cap ORF encodes three structural proteins VP1, VP2 and VP3 from the p40 promoter.

The 145bp ITRs consist of 125bp of paired complementary regions A-A', B-B' and C-C' and a 20bp unpaired D region. A characteristic T shaped

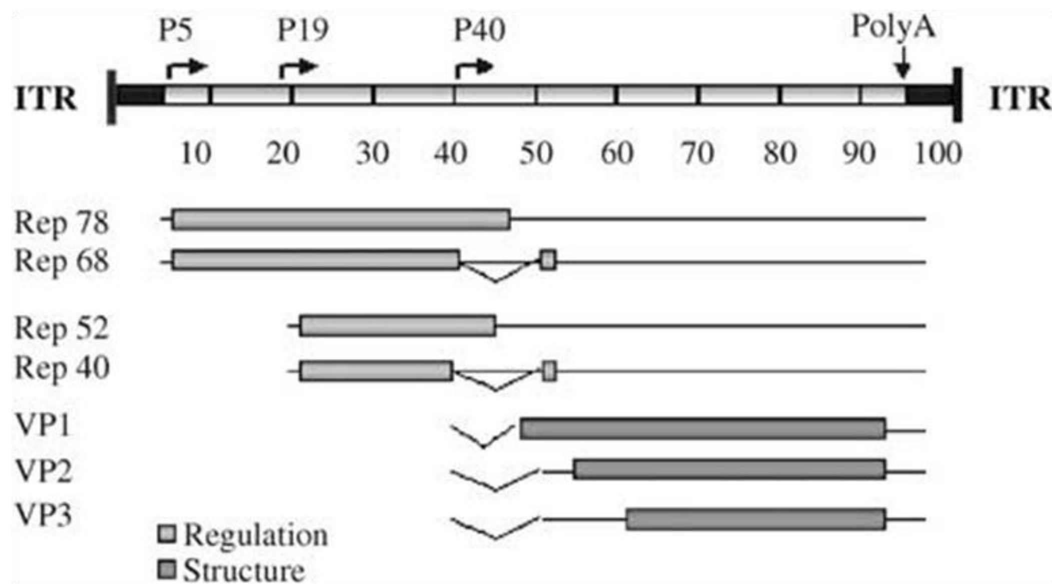
secondary structure results from the two small internal B-B', C-C' palindromes with the larger external A-A' palindrome forming the stem. The ITRs can exist in a flip or flop orientation, based on the orientation of the internal palindrome with respect to the D sequence (141). The 125 bp palindromic region carries a Rep binding element (RBE) /Rep binding site (RBS) which consists of 3 repeats of a GAGC tetramer. Only the larger Rep proteins Rep 68/78 are capable of binding with the RBE (142). Considerable variation within an RBE can be tolerated. For instance, the RBE found within the p5 promoter contains only one complete copy (GAGC) and one degenerate copy (GAGT) of the tetramer repeat (143, 144). At the junction of the D region and A/A' lies the terminal resolution site (*trs*). The ITRs are important cis acting regions, and play an important role in replication (145), regulation of gene expression (146), transcription, and site specific integration (147).

The larger Rep proteins, Rep 68 and Rep78, are produced from unspliced and spliced transcripts from the p5 promoter and are important regulatory proteins, playing significant roles in every stage of the AAV life cycle. They are capable of positively and negatively regulating AAV gene expression (148), are required for replication and site specific integration(149). The smaller Rep proteins Rep52 and Rep40, produced from unspliced and spliced transcripts respectively, from the p19 promoter, are involved in accumulation of viral DNA for packaging into virions. All four Rep proteins possess helicase and ATPase activity(150), which plays a role in the unwinding and resolution of the ITR during viral replication In addition, Rep 68 and Rep78 are capable of site specific binding to the RBS and have site specific endonuclease activity, required for separation of replicated viral genomes (150, 151) .

The products of the Cap gene (VP1, VP2 and VP3) are structural proteins making up the AAV capsid. VP1 is produced from the unspliced transcript from p40, VP2 is produced from the spliced transcript using an unusual ACG start codon and VP3 is produced from the spliced transcript from a downstream conventional AUG start codon. The AAV-2 capsid comprises 60 viral capsid proteins arranged into an icosahedral structure, with VP1, VP2 and VP3 present in a 1:1:8 molar ratio (152). A recent report identified an additional protein encoded by an alternate ORF within the cap gene, called the assembly activating protein (AAP), required for AAV capsid assembly (153).

<u>Infection and life cycle</u>: The main receptor for AAV2 infection in cell culture is heparate sulfate proteoglycan(154). Interestingly, a study of AAV-2 isolated from human subjects over a 6 month period (155) found that all of the AAV-2 capsid sequences isolated had lost the Arginine at residue 585 and 588,

Fig.2: Organization of genome of AAV. The genome is flanked by characteristic hairpin shaped ITRs and encodes 7 polypeptides from two main ORFs and three different promoters. The proteins encoded by the Rep ORF (Rep 78, Rep68, Rep52 and Rep40) are required for replication. Proteins encoded by Cap are structural proteins.



Reprinted with permission from Macmillan Publishers Ltd: [Gene Therapy] (Merten O-W et.al, (2005) 12, S51–S61), copyright (2005)

which have been shown to be critical for heparin sulfate proteoglycan binding (156), indicating that *in vivo* infection might proceed differently. Co-receptors identified include fibroblast growth factor receptor(157), hepatocyte growth factor receptor(158), and the $\alpha_v\beta_3$ and $\alpha_5\beta_1$ integrins (159, 160). Not much is definitively known about process by which the virion is transported across the cytoplasm and into the nucleus. Though heparin sulfate proteoglycan is not specifically associated with the clathrin pathway, the virions are likely endocytosed into clathrin coated pits due to interactions with $\alpha_v\beta_3$ signals through NotchI and Rac to enhance clathrin mediated endoyctosis (161-163) . Virions have been co-localized with markers of early, late and recycling endosomes as well as lysosomes (163, 164). Release from the endosome probably requires low pH as bafilomycin A1, a drug which inhibits the ATPase associated with the proton pump for endosomes, inhibits AAV infection (165). Proteasomes play an important role and treatment with proteasome inhibitors enhances AAV transduction (166). A role for a conserved phospholipase A2 motif in the N terminus of the VP1 protein in allowing the virion to escape the endosome has been reported (167). The virus is small enough (26nm) to enter the nucleus directly through the NCP. However, the mode of entry has not been definitively determined. Though AAV2 has been shown to interact with microtubule associated proteins (168) , effects of microtubule directed drugs have been unclear (161, 169, 170).

<u>Lysogenic cycle</u>: Within the nucleus, in the absence of helper virus co-infection, the cellular YY1 factor binds to two sites within the p5 promoter to silence transcription from it(171). Low levels of Rep68/78 proteins brought into the cell by covalent linkage to the AAV genome (172), or expressed, serve to inhibit further expression from the p5, p19 and p40 promoters through interactions with the p5 RBE, and mediate integration of the viral genome into the AAVS1 site. An RBE and a *trs* separated by a conserved spacer, similar to that found in the AAV ITR are found at the AAVS1 site. Multiple RBEs can be found throughout the genome, capable of being bound by Rep protein. However, an RBE with the appropriately positioned *trs* is unique to the AAVS1 site and nicking at a central di-thymidine (TT) at the *trs* is required for integration. A 33 bp chromosomal sequence carrying the RBE and *trs* with the spacer is sufficient to mediate targeted integration (173, 174). The AAVS1 site has been shown to be maintained in an open chromatin conformation (175), has multiple transcription factor binding sites and a CpG island. The RBE and *trs* lie a few bps upstream of the translation start site of the myosin binding subunit 85 gene (MBS85), a transcript found to be ubiquitously expressed in human tissues, especially in the heart (176). However, knock out of both copies of the MBS85 gene in embryonic stem cell lines was found to have no impact on the functional ability of the cells to

self renew or contribute to the development of a mouse blastocyst (177) An alternate association of AAV integration with a functional ORF was found with a muscle specific gene called slow skeletal Troponin T gene (TNNT1), which lies on 19q13.4 (5' to the AAVS1 site) (178). TNNTI is found expressed in skeletal tissues in adult humans. However, since muscle cells are multinucleated, disruption of some TNNT1 loci due to AAV integration might not result in any phenotypic manifestation

AAV factors required for integration have been found to be the Rep68/78 proteins in *trans* and an RBE in *cis (179-181)*. This RBE can be found either within the AAV Terminal Repeat or the p5 Integration Efficiency Element (p5IEE). The p5IEE overlaps the p5 promoter and has been found to be necessary and sufficient to mediate efficient site specific integration (182). Philpott et al. showed that with plasmid substrates for integration, a p5IEE functions more efficiently to mediate site specific integration (183). A functional *trs* is not required within the *AAV* moiety for integration (184). Though the mechanism of Rep-mediated site specific integration (RMSSI) has not been completely elucidated, several pieces of evidence have resulted in a widely accepted basic model. Characterization of integration into Epstein Barr Shuttle Vectors carrying the AAVS1 site (175) and naturally occurring integrants isolated from Detroit six cells (185, 186) revealed presence of multiple copies of the viral genome integrated in a head to tail orientation, with junctions formed either at the AAV ITRs or p5IEE, with integration junctions 100bp-1000bp downstream of the nicking site. A considerable level of amplification, deletions and duplications of cellular sequences at the site of integration was observed. Further, Rep expression in the absence of a suitable substrate results in a similar amplification of the AAVS1 site (187). The Rep68 protein has been shown to be capable of tethering the AAV genome to the AAVS1 site at the RBEs (188).

Linden et. al., proposed a model for integration (173) which proposed that Rep68/78 localizes an AAV genome to the AAVS1 site, binding to the RBE at the AAVS1, nicking at the cellular *trs* and binding to and stabilizing the released 5' end of the chromosome, much as it does during AAV replication. This is followed by extension of the nicked 3' end by a cellular DNA polymerase. A series of template switches from the elongating strand to the displaced chromosomal strand, and from the displaced chromosomal strand across Rep to the AAV genome, probably followed by non homologous end joining results in the imprecise insertion of the AAV genome into its integration site, downstream of the nicking site and accompanied by deletions and duplications. The tendency of AAV to form circular moieties would explain the presence of the head to tail integration events observed.

. Not much is known about cellular requirements for integration. DNA dependent protein kinase, a DNA repair enzyme has been found to inhibit AAV integration into AAVS1 (189). The other cellular factor so far found to affect RMSSI has been a TAR RNA loop binding protein 185 or TRP-185 (190). TRP-185 has previously been shown to be involved in activating HIV gene expression by preventing RNA polymerase II stalling on the RNA regulatory TAR element within the HIV genome. In the context of AAV integration, TRP-185 was found to interact directly with Rep68, prevent the binding of Rep68 to the RBE within the AAVS1 site and interestingly, influence the AAV-AAVS1 junction site to lie downstream of the RBS.

Characteristics of RMSSI: Site specific integration with wtAAV has been found to be highly specific in multiple cell lines, with one study finding 94% of all AAV positive IB3-1 cells to have site specific integration into Chr19 (191). This integration depends on expression of Rep protein and is not seen with rAAV vectors. Further, efficiency and specificity of Rep mediated site specific integration with plasmid or hybrid virus based systems has not been able to match that seen with wtAAV. To allow comparison across studies, efficiency of integration has been described in terms of integration events per infectious unit. Evaluation of earlier studies resulted in frequencies between 0.1%-0.5% of integration events per infectious genome (192). These results were supported by frequencies obtained from a PCR based quantitative PCR assay developed for analyzing integration events (193). This assay also revealed that AAV integrates with equal frequency in either orientation (194). Multiple authors have shown that integration events peak 48-96 hours post infection (195-197).

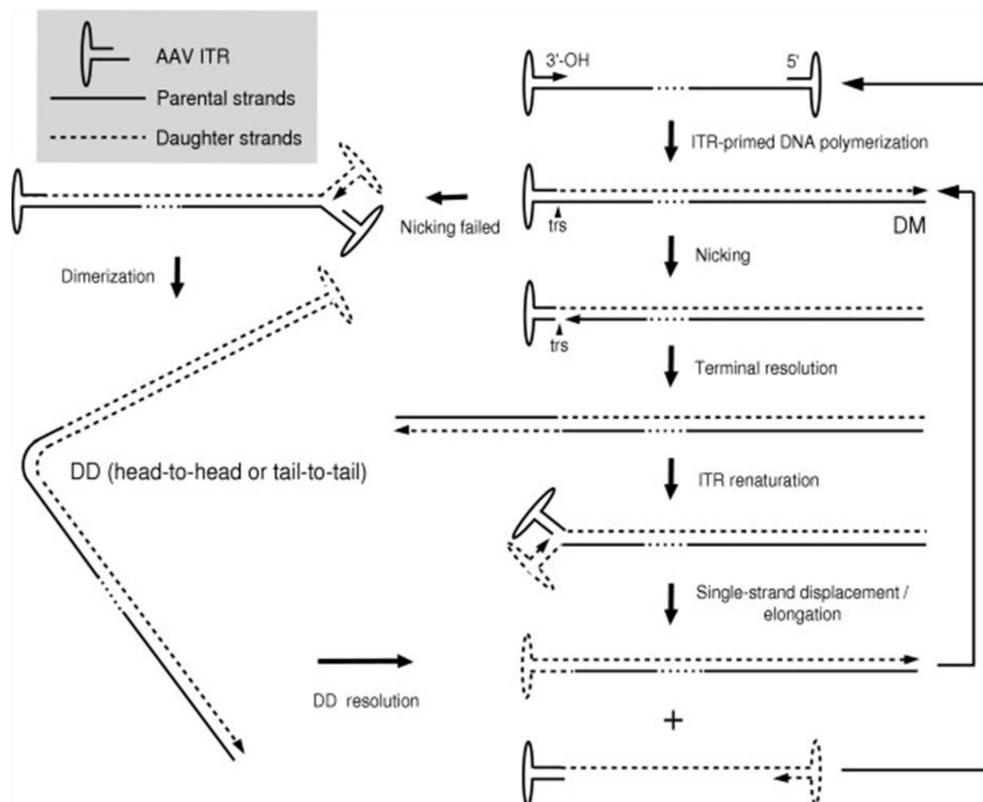
Lytic cycle: Though multiple viruses have been shown to be capable of functioning as helper viruses for AAV replication, only the functions of Adenovirus and to a lesser extent HSV, have been completely characterized. The transition into the lytic cycle involves activation of Rep genes, excision of the viral genome followed by the replication of the viral genome and expression of Cap genes. The Adenoviral proteins required have been shown to be E1A which is critical for activating AAV transcription (198), E1B which interacts with Ad E4 for transport of viral mRNA and E2A and VA RNA which increase viral mRNA stability and increase efficiency of translation (199). Ad DNA pol is not required for AAV replication. Contrastingly, the HSV proteins required to support efficient AAV replication have been found to be ICP0 to activate Rep expression, the Helicase/primase complex (UL5/8/52), the single stranded DNA Binding Protein (ICP8) and the HSV-1 DNA polymerase complex (UL30/UL42), all protein complexes involved in DNA replication (200, 201). The lack of

overlap between the viral functions required by AAV from these two helper viruses suggests that AAV might not be a true replication-deficient virus, and requires a helper virus more to establish a cellular milieu conducive to replication. This fact also explains the ability of multiple varying viruses and genotoxic agents to support AAV replication.

Ad E1A plays an important role in activating transcription from the p5 promoter by interacting with the cellular Major Late Transcription factor and YY1, allowing the expression of the Rep68/78 genes. A complex regulatory role is played by the Rep 68/78 proteins in regulating AAV gene expression, involving the RBEs present in the AAV ITR and p5 promoter. Pereira et.al, showed that Rep68/78 bind to the p5RBE to repress transcription from the p5 promoter, while activating transcription from the p19 and p40 promoters. The RBE at AAV ITR serves to increase transcription from p5. The Rep52/40 proteins serve to lift Rep68/78 mediated inhibition of p5, thus forming a feedback loop which maintains a relative steady state of Rep68/78 and Rep52/40 proteins (148).

Model for DNA Replication (Fig.3): DNA polymerase delta, along with associated replication factors including replication factor C, proliferating cell nuclear antigen, and mini-chromosome maintenance complex have been found to be the only factors required for the exclusively leading-strand replication (202, 203) of AAV. Due to the folded palindromic nature of the ITR, the free 3'OH of the A or A' region can be extended by a DNA polymerase of cellular or helper virus origin. Extension proceeds to the distal ITR, resulting in a duplex molecule cross linked at one end. The newly replicated strand can also be covalently linked to the distal ITR, resulting in a duplex which is closed at both ends. The model for DNA replication proposes that resolution of this closed dsDNA form occurs by the binding of Rep68/78 to the Rep Binding site followed by unwinding of the duplex trs and nicking at the trs (204, 205) by the helicase and endonuclease activities of Rep respectively (206). Binding of Rep to this RBS requires a correctly folded ITR (207). The location of the trs with respect to the RBS is crucial, as the addition of even 3bp results in significant reduction in replication. Following nicking, Rep68/78 binds and stabilizes the free 5'OH end (208). Since the trs is directly opposite the first phosphodiester bond formed between the ITR primer and the first nucleotide attached, cleaving at the trs results in the 3' end of the parent genome becoming the 5' end of the daughter strand, flipping the orientation of the terminal palindrome. The 3'OH end of the nicked parent strand is extended with the displaced hairpin. In the absence of immediate cleavage, the 3'ITRs of the duplex molecule can fold on themselves and act as primers for another round of extension. This results in duplex dimer molecules which have been visualized both *in vitro* and *in vivo*.

Fig 3: Model for replication of AAV: a) The 3' end of the AAV ITR is extended by host DNA polymerase to form a double stranded monomer moiety. b) Rep68/78 resolves this molecule by binding to the folded ITR and nicking at the trs, which lies directly opposite the first base of the newly synthesized daughter strand, resulting in transferring of the ITR from the parent strand to the daughter strand in the opposite orientation. c) The parent strand is then extended, using the ITR of the daughter strand as template. d) Renaturation of the ITRs is followed by extension of one strand, and displacement of the other ssDNA AAV genome. Absence of nicking at step b results in the formation of double stranded dimer molecules.



Reprinted under a Creative Commons Attribution License from "Adeno-associated virus: from defective virus to effective vector" Manuel AFV Goncalves, Virology Journal, 2005, 2:43

For the packaging of replicated AAV genomes, the VP1, VP2 and VP3 first self assemble into empty capsids in the nucleolus (209-211). Recently, an alternate ORF encoded upstream of the VP3 start site within the cap gene has been reported to be essential for AAV capsid formation and packaging (153). Spreading of the capsids from the nucleolus to the rest of the nucleus requires Rep protein expression (212). This is followed by insertion of the genomes into the capsids. Not much is known about the detailed mechanism of packaging.

Inhibition of helper virus by AAV: An interesting phenomenon is the inhibition of replication of various helper viruses including HIV(213, 214), HSV(215), SV40 (216, 217) and Ad(218) by AAV. The ability of Rep68/78 proteins to affect transcription from a wide range of heterologous promoters (219-222) is a major component in inhibition. Rep protein has been shown to inhibit expression from the HIV Long Terminal Repeat (223), inhibit HSV induced SV40 replication (215) and has been found to inhibit expression from multiple Ad genes (222, 224, 225). AAV co-infection and Rep78 protein expression has been found to inhibit expression of HSV immediate early protein ICP4 and multiple early protein including UL9, ICP8 and UL5 and late proteins VP16 and VP22, with increased expression of Rep78 resulting in enhanced inhibition (226, 227). However, with both HSV and Ad, evidence suggests that major inhibition of the helper virus is due to prevention of replication and reduction in early gene expression not causative. Blocking of replication of Adenovirus using hydroxyurea which prevents DNA replication was found to minimize the reduction in Ad early gene expression, suggesting reduction in expression levels was concomitant with reduction in genome replication and not vice versa (225). Similarly, supplementing HSV –AAV co-infections with HSV early gene expression failed to lift inhibition of HSV replication(226).

The inhibition of Adenovirus by AAV has been found to be dependent on the relative time of infection and relative MOI. In a series of papers, the group of J. Trempe characterized the effect of AAV co-infection in general and Rep expression in particular on Ad gene expression and replication. AAV co-infection was found to decrease transcription from Ad E1a, E2a and E4, with the largest decrease seen in the levels of the E4 transcripts (225). A marked decrease in Ad genome amplification as measured by the change in expression from a β-galactosidase transgene carried within the Adenoviral E3 region was reported (225). Rep protein was found to bind to and inhibit transcription from the E2A promoter as well as the Major late promoter (222, 224). Expression of Rep protein from a transfected plasmid also resulted in a decrease in Adenovirus genome replication, though to a lesser extent than with wt AAV(225). This

indicated a possible requirement for replication of AAV for complete inhibition of Adenoviral replication. Separately, two separate groups reported that Rep78 and Rept52 but not Rep68 and 40 interact with cAMP dependent protein kinase A (PKA) and its novel homologue PrKX via a putative zinc finger domain in the C terminal, inhibiting their kinase activity (228, 229). This results in reduced steady state levels of cAMP responsive element binding protein (CREB) and inhibition of CREB dependent transcriptional activation. This protein binding domain domain within Rep was found to share homology with the a cellular pseudosubstrate inhibitor of CREB – the protein kinase inhibitor (PKI) (230). Mutation of the PKI like motif resulted in an AAV moiety that could replicate but could not repress Ad replication. In the presence of Ad, this virus replicated normally in the first passages but its titers gradually reduced as it was overtaken by Adenoviral replication with further passages (231). This sheds light on an evolutionary motive for AAV's inhibition of replication of the helper virus that it requires for its own replication.

Effect of AAV on cell cycle: Infection by AAV has been shown to be associated with inhibition of cell cycle progression and induction of differentiation. AAV has been shown to inhibit Adenovirus induced transformation of Syrian Hamster cells. This oncosuppressive property has been further seen with the inhibition of H-Ras promoter (232), reduction of c-myc and c-myb oncogene products in HL60, a human leukemia cell line (233) and the inhibition of SV40 mediated cellular transformation (234).

An anti-proliferative effect of AAV has also been seen in cell culture and has mainly been associated with the Rep 78 protein. However, early work by Winocour et.al, has shown S and G2 phase cell cycle arrest of primary human and Syrian hamster fibroblasts with UV-inactivated virus (235). Similarly, Kube et.al, reported the inhibition of cell cycle progression of primary human fibroblast cells with a high MOI of both wtAAV as well as with rAAV carrying no AAV genes (236).

Of the Rep proteins, Rep 78 and to a lesser extent Rep68 have been shown to be the major effectors of cell cycle inhibition, while Rep52 and Rep40 have no effect. In NIH-3T3 cells as well as in mouse embryonic fibroblasts, Rep/78 expression was found to exert a profound effect, arresting cells in G1-S phase. This arrest was found to be due to the accumulation of active hypophosphorylated pRb protein, which negatively regulates the progression from G1-S phase. Interestingly, introduction of Adenovirus E1A which is shown to inhibit pRb was found to be capable of lifting inhibition. Rep68 expression in these cells was

associated with G1, G2 arrest and increased levels of CdK21. However, other authors have reported little or no effect in immortalized cell lines (235, 236).

AAV as a gene therapy vector: Apart from its unique ability for site specific integration, AAV has generated considerable interest as a gene therapy vector due to its ability to infect a wide range of cell types, persist extra-chromosomally and low toxicity. In fact, due to the small size of the AAV genome and concerns about the potential effects of toxic Rep expression on cellular genes, rAAV vectors used today do not carry the Rep gene and are incapable of site specific integration. The AAV ITRs are the only elements retained on these moieties.

As of 2008 more than 40 clinical trials using rAAV vectors had been approved by the FDA, the majority of them with AAV2. The earliest clinical trial involving rAAV was for Cystic Fibrosis using an AAV2 vector expressing the Cystic Fibrosis Transmembrance Conductance Regulator (rAAV2-CFTR) by sequential intranasal and endobronchial introduction to adult patients. Though DNA transfer and positive biological effects were observed for up to 60 days post treatment, expression levels decreased with the development of an antibody response to the rAAV capsid (237, 238).

The second clinical trial involving the use of rAAV2 vectors for a monogenic disorder was the development of rAAV2-FIX for the delivery of coagulation FIX to patients with Hemophilia B. This trial was similar to the CFTR trial in that it was supported by positive data in small and large animal models. The first trial involved intramuscular (IM) injection of the vector and resulted in transient expression of FIX, limited by the development of an immune response to the AAV2 capsid (239). This trial was followed by an attempt at intrahepatic delivery of the vector (240). Though correction of bleeding time by this method was superior to IM injection, effects were again transient due to the development of an anti-AAV2 capsid immune response. Data from these two trials indicated that though AAV2 was a safe vector as predicted, low expression from the vector genomes and transient expression due to induction of an immune response were obstacles to therapeutic gene delivery.

More recently, positive results with the use of rAAV vectors have been obtained for the delivery of therapeutic genes to the retina and the brain which are 'immunoprivileged'. Leber's congenital Amaurosis (LCA) is characterized by retinal degeneration and complete blindness by the age 10. 15% of LCA cases are associated with a mutation in *RPE65.* Safety and efficacy data was obtained from intra-retinal injections of rAAV2-h*RPE65v2* vectors in dogs and non human

primates(9, 241). The results of a phase I clinical trial was reported in 2009. 12 patients were enrolled, with the youngest being 8yrs. One eye was injected with the rAAV vector. All patients showed improvements in measurements of vision, with the youngest showing the most dramatic improvement. Improvement in vision for all patients has currently persisted for at least 1 year (10).

Phase I clinical trials for various approaches to treat Parkinson's disease with rAAV vectors have been conducted. Ceregene conducted a Phase I trial in 2005 with a rAAV vector expressing NTN – a neurotrophic factor which has been shown to protect against the degeneration of dopaminergic neurons associated with Parkinson's disease. This vector, called CERE-20 resulted in no reported adverse effects and an increase in the 'Unified Parkinson's Disease Rating Score' for all patients. However, a Phase II trial failed to show an appreciable difference between the CERE-120 and the control group (http://www.ceregene.com/science_parkinsons.asp).

Clinical trials with the use of pseudotyped vectors, consisting of AAV2 vectors in AAV1 capsids have also been conducted for the delivery to muscles, as AAV1 has shown to be up to 500 times more effective at delivery to muscle cells. Original studies with an IM injection of an rAAV2 vector expressing α1 antitrypsin (for α1 antitrypsin (AAT) deficiency associated lung disease) resulted in transient expression, limited due to a humoral immune response to the AAV2 capsid (242). Subsequently, the same cassette packaged into an AAV1 capsid was tested in Phase I trials. Interestingly, it was reported that even with the development of anti-AAV2 neutralizing antibodies by day 14, transgene expression of up to 0.1% persisted for one year post injection, with the highest dose (243). Independently, another group reported the development of capsid specific T cells with IM injection of rAAV1 (244).

Modifications and improvements of rAAV vectors have addressed four main areas – transgene capacity, tropism, evasion of the immune response and efficiency of infection. A novel technology to increase the transgene capacity of rAAV vectors was developed with *trans*-splicing vectors, which utilize the tendency of AAV to form circular head to tail concatemers via recombination at the ITR. The transgene cassette is split between two rAAV vectors with appropriately positioned splice donor and splice acceptor sites. Transcription from recombined AAV, followed by appropriate splicing of the resultant mRNA results in a functional mRNA molecule. Transgene cassettes up to 9Kb in size can be carried by this method (245).

The increase of tropism of rAAV vectors has been approached in three main ways. As mentioned earlier, utilization of alternate serotypes which possess different tissue specificities will expand the tropism of rAAV vectors to cell types incapable of being efficiently infected by AAV2. For instance, AAV serotypes 1, 5, 8 and 9 have been found to infect the muscle, retina, liver and heart respectively, better than AAV2 (152). This approach has allowed broad tissue tropisms. An alternate approach involves the directed evolution of AAV with altered tropisms. Here, modifications to the capsid are introduced by mixing fragments of naturally occurring serotypes or by error prone PCR (246-248). The resulting library of modified capsids is subject to sequential selective pressure (for the ability to infect the desired cell type), yielding a smaller pool of viruses with modified capsids with the desired tropism. Thirdly, the rational modification of AAV capsids either by direct targeting by the insertion of peptides and ligands specific for certain cell types into the viral capsid sequences (249, 250) or indirect targeting using bispecific antibodies (251) or biotin (252), have been used to increase the tropism of rAAV vectors.

Inefficient AAV trafficking (253) and second strand synthesis (254) have been identified as the major rate limiting steps in AAV gene expression. Self complementary AAV (scAAV) aim to bypass the limitations posed by the lag in second strand synthesis (255). scAAV vectors can fold upon themselves, thereby immediately forming double stranded DNA capable of being transcribed Though rapid transduction is observed with these vectors both *in vitro* and *in vivo*, the maximal size of the transgene is reduced to 50% the original size (256).

Production of rAAV: Required for the production of rAAV are the AAV ITR flanked transgene (rAAV sequence), AAV Rep and Cap genes and helper virus functions. Currently three main cell culture systems for rAAV generation used are : Transient transfection of mammalian cells, infection of mammalian cells and infection of insect cell lines (257).

Earliest efforts to produce rAAV involved transfection of two plasmids – one carrying the rAAV cassette (pAAV) and the other carrying the AAV gene functions (pHelper), followed by infection with wtAd (258). Titers of 10^3-10^4vg/cell were obtained by this method. An improvement was the use of replication defective E1ΔAd in 293 cells or temperature sensitive Ad mutants (259, 260). In the late 1990's a mini-Ad plasmid was constructed which provided E2, E4 and VARNA functions (199, 260, 261). Transfection of an E1 complementing cell line with the mini-Ad plasmid, pHelper and pAAV resulted in rAAV replication. This process eliminated helper virus contamination and boosted yield. However, this process was inherently inefficient due to the

requirement for a single cell to be transfected with all three plasmids for successful rAAV replication. The process was improved by the combination of pHelper and the mini-Ad plasmids within one plasmid molecule. This system thus required the co-transfection of only two plasmids. A yield of 10^{2}-10^{4} vg/cell was obtained with this method. Apart from the requirement for co-infection, scaling up of transfection based systems was found to be inefficient.

An alternative approach involves the use of mammalian packaging cell lines which express the AAV Rep-Cap proteins. Due to the effect of Ad E1 in enhancing AAV Rep protein expression from p5, Rep-Cap and E1 complementing cell lines are relatively unstable. Some cell lines expressing Rep-Cap and Ad helper genes have been constructed and used for the production of rAAV with tight control of helper gene expression (262-264). In contrast, several such cell lines based on HeLa and one A549 cell line were constructed with relative ease (265-267). Production of rAAV from these cell lines involves introduction of the rAAV cassette either by transfection or by infection with a hybrid Ad/AAV carrying the rAAV sequence followed by co-infection with wtAd, required for expression of E1 function. Yield obtained has been found to depend on the copy number of Rep-Cap sequences. Most cell lines have found to have retained 5-20 copies of Rep-Cap sequences which are found to amplify up to 100 copies/cell in an AAV ITR independent manner with the infection of wtAd (268, 269). Gao et al, used temperature sensitive E2B mutant Ad as helper virus and obtained yields as high as those obtained with classical transfection methods.

Producer cell lines are packaging cell lines which have both Rep-Cap and the rAAV sequence integrated. Thus, these cell lines merely require a single infection with wtAd or HSV-1 for rAAV production. Establishment of these cell lines are by transfection of rAAV sequences followed by selection for a resistance marker or by infection with a rAAV vector. Yields have been found to depend on Rep-Cap copy number as with packaging cell lines, but are independent of rAAV sequence copy number. Helper viruses used include wtAd, replication-defective Ad and Protease-deleted Ad (270, 271). The latter was found to result in reduced yield of rAAV.

Multiple attempts are documented to construct an Ad carrying Rep-Cap genes as infection of an rAAV sequence carrying cell line would result in efficient production. However, the few successful attempts suffered due to instability of the virus due to non homologous deletion of the Rep component and reduced titer (272, 273). HSV carrying Rep-Cap were constructed with relative ease (274) and produced high yields upon infection of 293 producer cell line(275). Booth et al., developed a single step rHSV-1 molecule which carries the rAAV sequence, Rep-

Cap sequences as well as HSV-1 helper virus functions which also resulted in efficient production of rAAV(276) .

The third approach for rAAV production is the use of baculovirus - insect cell based systems. Urabe et.al, originally demonstrated the feasibility of producing rAAV in Sf9 cells (277). In this protocol, Sf9 insect cells were infected with three baculoviral vectors - carrying the transgene cassette, Rep and Cap respectively. Infection of the cell line with these viruses was found to be sufficient for rAAV replication. The system required modification of the AAV expression systems to ensure expression of AAV proteins at the appropriate stoichiometry, but resulted in production of a high yield of rAAV of $5x10^4$ vg/cell. This system has further been shown to be scalable up to a 20L bioreactor (278).

Integration of rAAV: Though the majority of rAAV persist extra-chromosomally, a small percentage of vectors integrate into the host genome. The frequency of this integration has been calculated in HeLa cells under G418 selection to be about one integration event per 10^3 -10^4 vector genomes (279). However, integration efficiency is increased by the the stimulation of the DNA repair system (280). The hot spots for integration have been analyzed by multiple authors by sequence analysis of cellular-proviral junctions (281-283). rAAV has been found to integrate preferentially into active genes, resulting in deletions ranging from 2bp – several kb in length (284). Though integration sites share no significant sequence homology with the vector, microhomology of 1-5bp at the junctions were observed(284). The hairpin structure of the AAV ITR appears to contain hot spots of integration (282).

Worryingly, rAAV integration has been linked to development of hepatocellular carcinoma (HCC) in neonatal mice injected with a rAAV vector (285, 286). 33% of mice injected with rAAV developed HCC vs 4% in the control group. Importantly analysis of provirus junctions from 4 tumors found clustering of integration sites within a 6Kb region of the mouse genome which encoded 34 of 382 known mouse miRNA (287). The authors note that a similar region associated with tumorigenesis exists on Chr 14 in humans.

Hybrid AAV for integration: Construction of hybrid viruses carrying AAV elements on the backbone of a larger virus have potential applications not only for the production of rAAV, as described earlier, but also as potential integrating gene transfer vectors. The only difference in Rep expression in such vectors would be the requirement for both Rep68/78 and Rep52 expression for rAAV production vs only Rep68/78 expression needed for integration. Production of a rAAV possessing AAV's unique ability of Rep mediated site specific

integration is not feasible, due to the small size of its genome and the toxic effects of Rep protein on the host cell (235). Introduction of a Rep cassette into the rAAV would further reduce the viable transgene size to about 3Kb. Further, since the entire cassette flanked by the AAV ITR integrates in the presence of Rep protein, an internal Rep cassette would also be integrated. Alternatively, since the only elements required for site specific integration have been shown to be the AAV ITR or IEE in *cis* and the Rep68/78 protein in *trans*, an AAV ITR flanked transgene cassette and a Rep expression cassette outside the AAV TR's could be carried on the backbone of a larger virus such as Adenovirus or HSV, combining AAV's ability to site specifically integrate with the large transgene size of a larger virus.

Several authors have reported the construction or attempts at construction of an Ad/AAV. First described in 1996 by Fisher and colleagues (288), an Adenovirus carrying a transgene flanked by the hairpin AAV ITRs has since been reported by multiple authors and can be constructed with no apparent difficulty (289, 290). A novel vector was produced when it was discovered that within the context of a replicating Ad backbone, any inverted repeats (IR) were found to recombine, releasing a fragment flanked by the IRs carrying the Ad packaging signal, which were packaged into Ad capsids. Lieber et al., used the AAV ITRs as the inverted repeats, and produced mini-Ad/AAV moieties which only carried the Adenoviral left TR and packaging signal and an AAV ITR flanked transgene packaged into an Adenoviral capsid (289). These mini-Ad vectors are devoid of all Ad genes, and are expected to elicit a reduced immune response compared to first generation Ad vectors. Replication of these Ad/AAV vectors in the presence of Rep from Rep expressing cell lines such as C12 increased the production of these mini-Ad/AAV vectors, streamlining the process (290). However, lack of efficiency of production and reduced efficiency of infection hampered progress with these constructs.

Construction of an Ad carrying the Rep expression cassette, with or without AAV ITRs has proven to be more difficult. A few successes have been reported. Zhang et.al, reported the successful production of a first generation Ad carrying AAV Rep expressed under the endogenous p5 promoter, for use in rAAV production. However, most strategies for construction of such a virus have focused on controlling Rep expression, due to the deleterious effects of Rep protein on Ad replication. Multiple authors have documented the various failed strategies for production of first generation Ad/AAV Rep, including use of an apolipoprotein E enhancer human α1 antitrypsin promoter (AAT), an inducible heavy metal responsive promoter, promoterless Rep78 constructs (291) and crippled Rep genes (288). Successful strategies for production of a first generation

virus expressing Rep78 have included novel strategies which reconstitute the Rep expression system – for instance production of first generation Ads carrying either the 5' or 3' half of the Rep78 ORF, followed by recombination in cell culture for production of a functional ORF (the resultant virus was deleted for all Ad genes) (291) or separation of the Rep78 ORF from its promoter by stuffer DNA flanked by lox p sites during replication (292). Co-infection with a Cre recombinase expressing Ad resulted in reconstitution of the functional Rep68 cassette. However, as in the case of rAAV production, the inherent instability of such a virus and reduced titers with Rep expression has resulted in difficulties proceeding.

The generation of helper dependent Adenoviruses carrying AAV Rep has been more successful and reproducible. Recchia et al reported the production of an Ad/AAV carrying an AAV ITR flanked transgene and a tightly controlled Rep expression cassette within a single helper dependent Ad backbone. Rep78 expression in this system was controlled by a tetracycline inducible system (195). An earlier report from the same laboratory showed that both an AAT promoter and a T7 promoter could be used for the production of a helper dependent Ad expressing Rep78 (293). Wang and Lieber later reported the production of helper dependent Ad carrying the Rep78 ORF expressed under a beta globin promoter (294). This virus needed to be co-infected with a second helper dependent virus carrying the AAV ITR flanked transgene. Both reports used these viruses for site specific integration successfully in vitro and in vivo, with transgene cassettes up to 27Kb being successfully integrated into the AAVS1 site (294).

In contrast, a stable hybrid HSV amplicon vector carrying both an AAV ITR flanked transgene and a Rep expression cassette, with Rep expressed from the endogenous p5 promoter for site specific integration has been constructed by multiple authors (38, 295-298). HSV/AAV vectors for rAAV production have also been produced, and are described under rAAV production. Integrating HSV/AAV carry the HSV *ori* and *pac*, an AAV ITR flanked transgene and the Rep expression cassette outside the ITRs. A wide range in Rep dependent reductions in titer of the HSV/AAV amplicon vector from 5 fold to 2000 fold, compared to other HSV amplicons have been reported (298, 299). The position of the p5 promoter and Rep ORF with respect to the ITR seemed to play a role in virus inhibition, probably due to changes in Rep protein expression (298). Further, rescue and replication of the ITR flanked transgene cassettes as well as accumulation of high molecular weight concatemeric replication products was observed with these constructs, likely contributing to the reduced titers (300). Interestingly, initial reports of the use of HSV/AAV amplicons carrying AAV ITR flanked transgenes with or without the Rep expression cassette also reported

an increased transduction efficiency using these vectors compared to HSV amplicons (38, 296, 301). Since both HSV and HSV/AAV vectors were packaged comparably, no satisfactory explanation was available for this difference in infectivity. However, expectedly, HSV/AAV vectors carrying Rep have been shown to sustain transgene expression in a variety of cell types, up to 12 months in one report (302). Encouragingly, site specific integration of up to 80% of infected cells in the absence of selection was observed in one case (298). Utilizing the large transgene capacity of the HSV amplicon, site specific integration of a 100Kb human β-galactosidase gene was reported (303). Efficiency of site specific integration of this construct was less than those reported with smaller constructs. However, only a 5 fold reduction in titer was observed, probably due to the large size of the transgene, as every amplicon carried only one copy of the Rep expression cassette. Due to smaller transgenes, all other HSV/AAV amplicon constructs would have carried multiple copies of the transgene and the Rep expression cassette, resulting in increased expression of Rep protein. The ability of HSV/AAV vectors to replicate without the instability observed with Ad/AAV vectors, in spite of the documented inhibition of both viruses by AAV Rep is intriguing.

Dissertation perspective

In spite of multiple reported attempts and some reported successes, no reproducibly stable first generation Adenovirus carrying AAV Rep78 has so far been constructed. Most viruses either fail to grow, showing no signs of viral replication (292), grow slowly or are unstable, acquiring deletions within the Rep gene (257). In one report, analysis of two clones bearing deletions revealed no overlap of the deletions sites within the Rep ORF (272).

However, as mentioned earlier, helper dependent Ad have been constructed with tightly controlled Rep78 expression systems. Work done in this dissertation found that first generation Adenoviruses carrying Rep78 expressed under a tetracycline inducible system were incapable of growing. Previously, this Rep78 expression system had been successfully used for the production of a helper dependent Ad carrying Rep78 (195). It is interesting to note that similarly, Lieber et al., reported the failure to construct a first generation Ad carrying Rep expressed under the α1 antitrypsin promoter (291), whereas a helper dependent Ad carrying Rep78 expressed under an α1 antitrypsin promoter was reported capable of growing (293). It is to be noted that the details of the promoter used in the first generation Ad are not completely known.

Due to the undoubted inhibitory effect of Rep78 protein expression on Adenoviral replication, the absence of Ad/AAV expressing Rep to replicate has been ascribed to leaky Rep expression. It is possible that the proximity of Ad genes within a first generation Ad results in an increase in basal levels of Rep expression over that seen in helper dependent Ad deleted for all Ad genes. However, the complete absence of replication seen with these basal levels of Rep expression contrasts with the 10%-40% decrease in Ad replication reported in the presence of Rep protein expression (225). Further, robust replication of Ad can be seen in cell lines expressing Rep. For instance, as described earlier, an Ad/AAV carrying an AAV ITR flanked transgene can be made to undergo efficient recombination in the presence of Rep protein to produce mini-Ad vectors deleted of all Adenoviral genes (290). The production of the mini-Ad vectors depends on the replication of Ad/AAV in Rep expressing cell line C12 (265). Induction of Rep protein expression in the cell line C12 required infection by wtAd. Robust replication of Ad in this context necessitates the CsCl purification and separation of the mini-Ad vectors from full length Ad moieties.

Two possible explanations for this discrepancy in inhibition of Ad when Rep is carried on the backbone of a first generation Ad are possible. One possible explanation is that with an increase in template copy number due to replicating Ad, the amount of leaky Rep expression accumulates and exceeds what is tolerated by Ad. An alternative explanation is that the sequence of the Rep ORF plays a possible, inhibitory role in the inhibition of Ad. The effect of such a sequence specific inhibitory signal could be *cis* acting with respect to the Ad genome, or much like the Rep protein, could be an additive effect, increasing with increasing Rep ORF copy number.

To distinguish between these two possibilities, we applied a previously designed computer algorithm (304) to modify the 1866bp Rep78 DNA ORF using synonymous codons, generating a Scrambled (Scr) and a Deoptimized (Deopt) sequence. Due to the degeneracy of the codon table, all but 2 of the 20 amino acids can be encoded using more than one codon. These synonymous codons differ from one another often at the third base of the codon (the wobble position). Thus the same polypeptide sequence can be encoded by different nucleotide sequences which vary from one another by 30%. The Scrambled sequence has randomly chosen synonymous codons and differs from the wild type Rep78 ORF (wtRep78) by 30%, while encoding the same amino acid sequence. The Deoptimized sequence specifically uses codons in underrepresented pairs, expressing Rep78 protein at reduced levels due to codon pair bias. Codon pair bias refers to the preference, in nature, for some codon pairs over other synonymous codons to encode the same pair of adjacent amino acids. First discovered in E.coli, codon pair bias is independent of the frequency of individual codons (305). An illustrative example is given by Papamichail et.al (304). The authors found that though based on codon frequencies one would expect the amino acid pair Ala-Glu to be encoded by GCCGAA and GCAGAG about equally often, the codon pair GCCGAA was found only 1/7th as often as a GCAGAG. This was in spite of GCC being the most frequent Ala codon (304). tRNA properties have been shown to influence codon pair bias in eukaryotes, as efficient translation depends on the juxtaposition of the tRNAs at the P- and A-sites on the ribosome (306). Consequentially, utilization of sequences with a negative codon pair bias results in reduced expression due to inefficient translation in eukaryotes. Interestingly, in E.coli, highly expressed genes have been found to carry more underrepresented codon pairs (305). The Deoptimized Rep78 sequence was thus designed with not only a modified nucleotide sequence which would disrupt any sequence specific inhibitory signal, but to also express reduced levels of Rep78. Comparing the ability of Adenoviruses carrying these sequence-modified Rep genes which express different levels of Rep, to replicate,

will help tease out the extent of the contribution of the sequence of the Rep ORF vs Rep78 protein levels on the inhibition of Adenoviral replication.

Methods and Materials

Cell lines: First generation Ads which required complementation of the E1 deletion for replication were grown in 293 or N52 cells (307) (Dr. Patrick Hearing, Stony Brook University, Stony Brook). Second generation Ads with deletions in E1 and E4 were grown in 911-E4 cells (92) (Dr. Patrick Hearing, Stony Brook University, Stony Brook). 911E4 cells were maintained in DMEM, 10% FBS, 100ng/ml doxycycline, 50µg/ml hygromycin and 100µg/ml neomycin (911-E4 media). HeLa-Tet On cells which express a tetracycline transactivator were purchased from Clontech, USA. Cell lines HeLa, C12, Cos-1 and HEL were obtained from ATCC. All cell lines were maintained in DMEM ,10% FBS at 37ºC with 5% CO_2 unless otherwise mentioned. HEL cells were maintained in RPMI, 10% FBS. Cells transfected or infected with viral constructs were maintained in DMEM ,10% FBS with 50µg/ml Gentamicin.

For the production of 911-E4 cell lines expressing siRNA against Rep78, three hairpins were designed against target sites spread over the Rep78 mRNA sequence, using the Ambion siRNA target finder. These sequences, Rep-siRNA #1 (GATCCGCTGCAGCGCGACTTTCTGTTCAAGAGACAGAAAGTCGCGCTG CAGCTTA), Rep-siRNA #2 (GATCCCAGTATTT AAGCGCCTGTTTCAAGAGAAACAGGCGCTTAAATACTGTTA) and Rep-siRNA #3 (GATCCCTACCGGGAAGACCAACATTTCAAGAGAATGTTGGTCTTC CCGGTAGTTA) were synthesized as oligonucleotides with 5’BamHI and 3’ HindIII overhangs and were cloned into pSilencer 4.1CMV plasmid vector (Ambion, USA) following manufacturer’s instructions. pSilencer 4.1 CMV vectors express the cloned siRNA template as a hairpin under a modified CMV promoter, and carry a puromycin resistance gene for selection of clonal cell lines. Efficacy of hairpins to knockdown Rep78 hairpin levels was confirmed by Reverse Transcriptase-PCR (RT-PCR). Briefly HeLa cells were co-transfected with pDRep and pSilencer-Rep siRNA 1-3. 24 hours post transfection, total RNA was isolated using Trizol (Invitrogen, USA) following manufacturer’s instructions. 5µg of DNase treated RNA was used in a reverse transcription reaction with 100ng/µl random hexamers (Invitrogen, USA), .75mM dNTP, 10mMDTT and 1X RT buffer (Invitrogen,USA). 1/10th of the cDNA prep was analyzed by PCR with primers RT-PCR RepF1(AGTCCTCGGCCCAGATAGAC) and RTPCR-RepR1 (CCACCGGAAAAGTCTTTGA). Rep expression levels were compared to

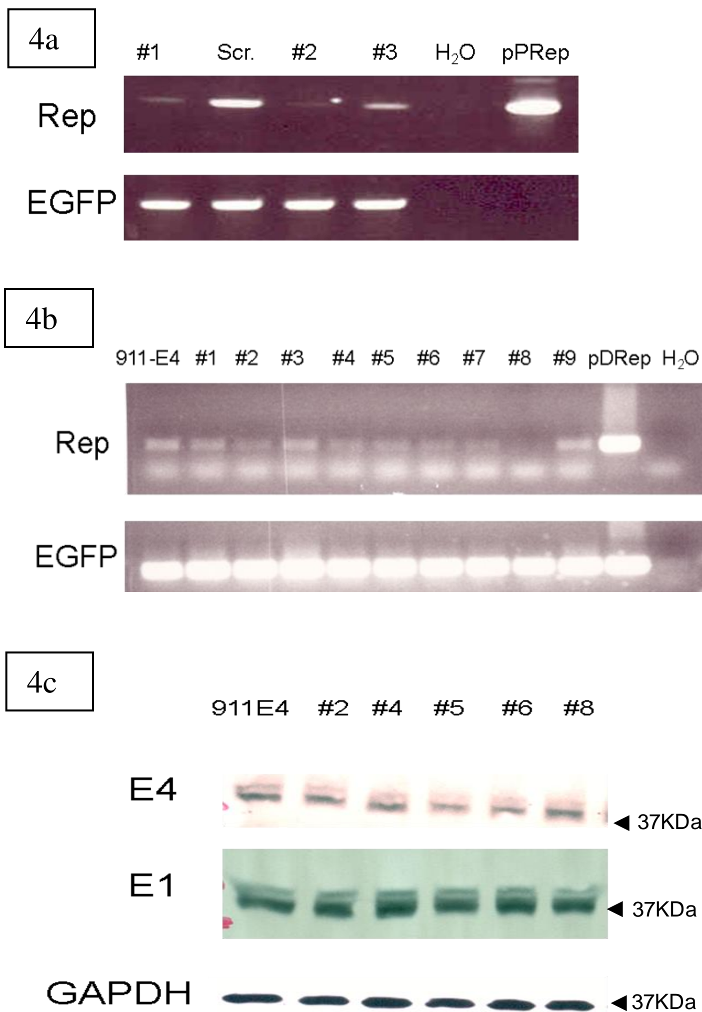
EGFP expression from the CMV-EGFP cassette in pDRep (Fig 4a). siRNA #2 was selected for clonal cell line establishment.

1 well in a 6 well plates of 911-E4 cells was transfected with 1μg of pSilencer Rep siRNA #2 . 48 hours post transfection, cells were trypsinized and 1/20th of the total volume was used to seed 2 10cm plates each. Selection was applied with 1.5μg/ml puromycin for 14 days. 9 well separated colonies were isolated using cloning cylinders and allowed to grow further in 911-E4 media containing 750ng/ml puromycin. Ability of the 9 cell lines to knockdown Rep expression was compared by RT-PCR of cells transfected with pDRep as described (Fig 4b). Expression of Ad E1 and E4 by cell lines #2, #4, #5, #6 and #8 was confirmed by western blot with antibodies anti-E1 and anti-E4ORF6 antibodies (Dr. Patrick Hearing, Stony Brook University) (Fig 4c). Cell line #8 was used for further experiments.

Plasmid construction: Plasmids and cloning (Fig 5): The first step in the construction of all viral constructs was the cloning of the required transgene into a shuttle vector. All shuttle vectors were derived from pAd/AAV-EGFP-Neo (290). pAd/AAV-EGFP-Neo was modified generating pIC59 and pD by Dr. Inderjit Chhabra (308). Both constructs retain the Ad5 sequences - Ad 1-450 upstream and Ad 3330-3950 downstream of the CMV-EGFP-Neo transgene cassettes, with a PacI restriction enzyme site directly upstream of Ad bp1. An AgeI site and unique AvrII and NheI restriction sites have been introduced between Ad450 and CMV. pIC59 carries no AAV elements, and unique BglII and XbaI restriction enzyme sites between the SV40 poly A of Neomycin Phosphotransferase and Ad3330. Downstream of the SV40 polyA, pD carries two copies of the AAV Integration Efficiency element (IEE) followed by the AAV Terminal Repeat (TR). Due to the double stranded nature of the construct, the AAV TR has an extra D element, and is sometimes referred to as TRDD. A unique ClaI restriction site lies upstream of the IEEs and a unique BglII site lies downstream of the TRDD.

All PCRs for cloning was performed using Pfu Ultra High Fidelity Polymerase (Stratagene, USA). All primers were synthesized by Operon, USA. Constructs carrying the AAV ITR were transformed into Stbl2 competent cells (*Invitrogen, USA*) to prevent loss of the hairpin due to recombination. DH5α sub-cloning competent cells (*Invitrogen, USA*) were used for all other transformations, unless specified otherwise. All ligations were performed at 16°C for 30 minutes with the Mighty mix DNA ligation kit (*Takara, USA*), unless specified otherwise.

Fig 4: Development of a packaging cell line expressing siRNA against Rep78. 4a) The ability of pSilencer- siRNA Rep #1-#3 to knockdown Rep78 mRNA levels when co-transfected into HeLa cells with pDRep was analyzed by RT-PCR. Plasmid DNA was used as a positive control for the PCR reaction. siRNA #2 was chosen for the production of clonal cell lines. 4b) Screening of clonal puromycin resistant 911-E4 cell lines for Rep78 mRNA knockdown. Cells were transfected with pDRep and mRNA levels analyzed by RT-PCR 48 hours post transfection. Transfection levels were confirmed by EGFP expression levels. Cell lines #2,4, 6, 7 and 8 were screened further. 4c) Expression of Adenovirus E1a, E1b and E4 from cell lines #2, 4, 5, 6and 8. Cell line #8 was used for further experiments.



pΔ28 (195) was a gift from Dr. Alessandra Recchia, Milan, Italy. This plasmid carries AAV Rep78 expressed under a tetracycline inducible system. The entire expression cassette can be divided into two cassettes – the Tetracycline suppressor (tTS) expressed under the CMV promoter (CMV-tTS) and Rep 78 expressed under the minimal Thymidine kinase (pTK) promoter fused with seven repeats of the Tet operator (Tet-Tk-Rep78). The Rep78 ORF is modified from the AAV Rep ORF to express only Rep78 but not Rep68, Rep52 and Rep40. For this, the ATG for Rep52 and Rep40 is mutated to GGA and the splice site required for Rep68 and Rep40 expression is mutated from GGTA to GATA.

pDRep was constructed in several steps. 2.8Kb Tet-TK-Rep78 was amplified from pΔ28 by PCR with primers *AvrIITREFwd* (CAATAATATCCTAGGGTCGAGGCCGCTCTAGTCGAG) and *AvrIIRep78PolyARev* (CAATAATAACCTAGGAGAGCTCGTCGACGATAAGCT), which introduced AvrII restriction sites. The resulting fragment was digested with AvrII (*NEB, Ipswich, MA*), cleaned with the Qiagen PCR purification kit (*Qiagen,USA*) and ligated into the XbaI site in pIC59 in a ratio of 3:1, using equal volumes of Mighty mix (*Takara, USA*). $1/15^{th}$ of the ligation reaction was transformed into DH5α cells generating pIC59-XbaI4. Putative clones were cultured in 3ml LB with 100μg/ml carbenicillin. Plasmids were isolated using Qiagen spin miniprep kit (*Qiagen, USA)* and screened by double digestion with BglII and AflII, which flank the insert. Large scale preparations from 100ml culture were performed using Qiagen Plasmid Maxi Kit (*Qiagen, USA*) and confirmed by sequencing. Sequencing of the clone revealed an additional AvrII site introduced 5' to Ad3330. Subsequently, the 2.1 Kb CMV–tTS cassette was PCR amplified with primers *BclICMVFwd* (CAATGTAAATGATCATACCGAATTCCTCGAGTCTAGC) and *BglIITSRev06* (CAATGAATAAGATCTTTGGTCGAGCTGATACTTCC), which introduced a BclI site at the 5' end of CMV and a BglII site at the 3' end of its polyA. The PCR fragment was digested with BclI (*Roche,USA*) and BglII (*Roche, USA*) and inserted into the BglII site in pIC59-XbaI4, resulting in pIC59-Rep. This construct has the CMV-tTS and TRE-Tk-Rep cassettes in a head to head orientation. pIC59-Rep was digested with PacI and BglII to release the CMV-EGFP-Neo cassette. The backbone with the Rep cassette was eluted from a 0.8% agarose gel using a gel extraction kit (*Qiagen, USA*) following manufacturers protocol. The ~4.1Kb (PacI)Ad1-450-CMV-EGFP-Neo-IEE-TRDD(BglII) was released from pD by digestion with PacI and BglII and separated from the backbone by gel elution. The eluted transgene cassette was ligated with Rep-backbone fragment from pIC59-Rep, generating ~12.5Kb

pDRep. Construct was confirmed by restriction digestion and sequencing. A derivative of pDRep, pDR-AflIIPacI (or DRAP) was constructed with a PacI site introduced downstream of Ad3950, to allow the entire transgene cassette to be excised out of the backbone by PacI digestion. The 430bp fragment from Ad3520 until Ad3950 was PCR amplified using primers *AflII3520Fwd* (GTGTGGGCGTGGCTTAAGGG) and *AflIIPacIRev* (TACACCTATCTTAAGTTAATTAAGTTCCCGGGTCAAAGAAT). The PCR fragment is flanked by a 5'AflII site which occurs naturally at Ad3520 and 3' PacI and AflII sites, both introduced by the PCR primer. The PCR fragment was digested by AflII and inserted into the AflII site at Ad3520, resulting in pDR-AflIIPacI. This construct thus has a duplication of Ad3520-3950. However, digestion by PacI would separate the duplication, resulting in a single intact sequence of Ad3330-3950 being released with the transgene.

To construct **pPRep**, the CMV-EGFP-Neo cassette was excised from pDRep by restriction digestion with AgeI and ClaI. A 15 bp fragment with 5' AgeI and 3' ClaI overhangs, flanking a NotI restriction site was ligated into the resulting backbone. This fragment was constructed by the annealing of oligonucleotides *AgeINotIClaItop* (CCGGTGCGGCCGCAT) and *AgeINotIClaIbottom* (CGATGCGGCCGCA) (pDRBB-NotI) . The 5.6Kb pPF4-FVIII-polyA cassette was released from pBSPF4FVIII (309) by digestion with NotI, separated and eluted from a 0.8% agarose gel and ligated and transformed into the NotI site of the modified pDRep, resulting in the ~15Kb pPRep. pPR-AflIIPacI (or PRAP) was constructed by introduction of the 430bp Ad3520-3950 AflII PacI PCR fragment into pDRBB-NotI, before insertion of the pPF4-FVIII cassette into the NotI site.

Plasmid P carrying just the PF4-FVIII cassette with the AAV IEE and TRDD was constructed in several steps. pIC59 was digested with AvrII and XbaI and self ligated, resulting in pIC59BB, consisting of the backbone of the shuttle vector with the Ad sequences. This construct was digested with PacI and AgeI, releasing Ad1-450 from the backbone. The PF4-FVIII-IEE-TRDD fragment was released from pPRep by PacI NgoMIV double digestion. NgoMIV cuts twice within pPRep, one in the CMV-tTSPolyA and once within the backbone, and results in sticky ends compatible with restriction enzyme AgeI. The (PacI)PF4FVIII-IEETRDD (NgoMIV) cassette was ligated into the PacI –AgeI sites of pIC59BB, resulting in pP.

To construct pBSPF4FVIII-EYFP, the ~1.5Kb IRES-EYFP-polyA cassette was PCR amplified from pIRES-EYFP (*Stratagene,USA*). The IRES in this construct is attenuated and will express reduced levels of protein in

comparison to the main ORF. The primers used for PCR amplification were *XhoIIRESEYFPFwd* (GCTATCGCTGCTCGAGCTGGAATTAATTCGCTGTCT) and *XhoINotIIRESEYFPRev* (GATTTCTAATCTCGAGCGGCCGCAGCTGGTTCTTTCCGCCTCA) . The resulting fragment was digested with XhoI (*NEB, USA*) and ligated into the XhoI site of pBSPF4FVIII, in place of its polyA. The resulting PF4-FVIII-IRES-EYFP (6.6Kb) cassette was flanked by NotI restriction sites. The fragment was released by NotI and cloned into the NotI site of pP, in place of the PF4-FVIII cassette, upstream of the AAV IEE and ITR, resulting in **pP-E.**

The **TROTS** series of plasmids have the **T**RE-TK-**R**ep **O**RF and the tetracycline **T**ranscriptional **S**ilencer. Downstream of the Rep ORF lies an E**G**FP or E**B**FP cassette expressed under an IRES.

pTROTS-wt-B (or pTRE-wtRep78) was constructed by PCR amplification and stepwise cloning of the TRE-pTK-Rep78 cassette and the pCMV-tTS cassette from plasmid pΔ28. pD was double digested with restriction enzymes AvrII and BglII releasing the entire transgene insert. The 2.8KbTRE–pTK-Rep78 cassette was PCR amplified with primers. *AvrIITREFwd* and *BglIIRepORFRev* (CTATCAAGTAAGATCTTCATTTATTGTTCAAAGATGCAGT). The resulting amplicon consisted of Tet-TK-Rep78 without the polyA and was inserted into the AvrII BglII sites of pD. The CMV-tTS cassette was then amplified from pDRep using primers *AgeICMVFwd* (CAATGTAAATACCGGTTACCGAATTCCTCGAGTCTAGC) and *AgeITSLeft* (CAATGAATATACCGGTTTGGTCGAGCTGATACTTCC), and cloned upstream of the TRE-Rep cassette into the AgeI site, in a head to tail orientation, resulting in pTROTS-wt. Next, the EBFP cassette from pEBFP (*Stratagene,USA*) was released using restriction enzymes SmaI and BsrGI, and inserted into the same sites in pIRES-EYFP (*Stratagene,USA*) replacing the EYFP cassette, resulting in pIRES-EBFP. This IRES-EBFP cassette was PCR amplified with primers *BamHIIRESEYFPFwd* (GCTATCGCTGGATCCTGGAATTAATTCGCTGTCT) and *BamHIIRESEYFPRev* (GATTTCTAATGGATCCACTCGAGCCCCAGCTGGTTC) and introduced into the BglII site in pTROTS-wt, resulting in pTROTS-wtRep78B (pTRE-wtRep78)

pTROTS-wt-G differs from pTROTS-wt-B by the replacement of EBFP with an EGFP gene and a SwaI site inserted downstream of WT Rep78 ORF. pTROTS-wt-B was digested with AvrII and BglII, to release the TRE-Rep78-

IRES-EBFP cassette. TRE-TK-Rep78 was PCR amplified from pDRep with primers *AvrIITREFwd* and *BglIISwaIRepORFR* (CCGGCCCTCTAGATCTATTTAAATCATTTATTGTTC), double digested with AvrII and BglII and inserted into the digested backbone, resulting in TROTSwaI. For the insertion of the IRES-EGFP, the EGFP cassette was excised from pD by double digestion with restriction enzymes AgeI and BsrGI, and eluted after separation from the backbone on a 0.8% agarose gel and inserted into XmaI , BsrGI digested pIRES-EYFP (*Stratagene, USA)* in place of the EYFP ORF, creating pIRES-EGFP. As EGFP and EBFP differ by 4 amino acids, the constructs were confirmed by digestion and sequencing. The IRES-EGFP cassette was PCR amplified with *BamHIIRESEYFPFwd* and *BamHIIRESEYFPRev*, digested with BamHI and inserted into the BglII site of pTROTSwaI, resulting pTROTS-wt-G.

The Scrambled Rep and Deoptimized Rep sequences were synthesized and inserted into the MCS of pUC19 by GenScript, USA (pUCScr and pUCDeopt). The plasmids were transformed into DH5α cells, screened by sequencing and large scale preparations isolated from 100ml culture. These sequences were designed with flanking unique 5’ SbfI and 3’ SwaI sites. To generate **pTROTS-Scr-G (pTRE-Scr)** and **pTROTS-Deopt-G (pTRE-Deopt)**, pTROTS-wt-B was double digested with SalI and SwaI, releasing the WT Rep78 ORF. SalI cuts once between TRETK and the WT Rep ORF and once within the WT Rep78 ORF. The Scrambled Rep and Deoptimized Rep sequences were PCR amplified using primers *XhoIRepScrFwd* (CAGATTCATTCTCGAGCCTGCAGGACCATGCCCGGA) and *XhoIDeoptFwd* (CAGATTCATTCTCGAGCCTGCAGGACCATGCCCGGG) respectively. *M13 seqprimerRev* (CAGGAACAGCTATGAC) was used as the reverse primer for both PCR reactions as both Rep sequences carried the SwaI site. As XhoI and SalI generate compatible stick ends, the PCR amplified fragments were inserted into the digested pTROTS-G backbone. Both forward primers also amplified the 5’SbfI site, resulting in the Rep ORF within the pTRE--Deopt-G and pTRE--Scr-G being flanked by a 5’ SbfI and 3’ SwaI for future swapping of cassettes.

phu1Scr and **phu1Deopt** have the Scrambled and Deoptimized Rep constructs respectively, expressed under an hu1 promoter. A left end shuttle vector carrying the hu1 promoter and SV40 poly A was constructed by digestion of the cassette from pcDNA3-hu1polyA (310) by flanking XbaI restriction sites and cloned into pD backbone, generating pITR-packaging-hu1polyA-3330.Unique BglII and HindIII restriction enzyme sites lie between the hu1 promoter and the poly A sequence. The Scrambled Rep ORF was PCR amplified

from pUCScr using *RepScrBamHIFwd* (CTGAATCATAGGATCCACCATGCCCGGATTCTACGAAA) and *M13 seqprimerRev,* and digested with BamHI and BglII (which lies within the pUC13 MCS). This construct was inserted into the BglII site in the vector, and the orientation checked by sequencing. phu1Deopt was constructed by PCR amplification of the Deopt ORF from pUCDeopt using primers *RepDeoptBglIIFwd* (CTGAATCATAAGATCTACCATGCCCGGGTTTTACGAG) and *M13 seqprimerRev* to utlize an internal HindIII site present in the MCS of pUCDeopt, 3' to the ORF. The amplified ORF was introduced into the BglII HindIII double digested vector backbone.

pCMV-WTflag, pCMV-Scrflag and pCMV-Deoptflag were constructed to allow eaysy detection of expression levels by western blotting. These constructs have a 3x C-terminal flag tag (DYKDDDDK) inserted. The backbone vector for these constructs was pCMV-3Tag-3A (*Stratagene,USA*). **pCMV-WTflag** was constructed by PCR amplification of the WT-Rep78 ORF (without the stop codon) from pDRep with primers *EcoRVRep78ORFFwd* (CAAGAAGACGATATCATGTCGGGGTTTTACGAGAT) and *EcoRVRep78ORFRev* (CTATGACATGATATCTTGTTCAAAGATGCAGTCATC) which introduced EcoRV restriction sites to either end, and insertion into the EcorV site in the MCS of the vector. Orientation of clones were confirmed by restriction digestion and sequencing. **pCMV-Scrflag** was constructed by PCR amplification of the Scr Rep ORF (without the stop codon), with primers *RepScrBamHIFwd* and *RepScrXhoIRev* (CTGAATCATACTCGAGCTGCTCGAAAATGCAGTCGT). The amplified sequence was inserted into pCMV-3Tag-3A double digested with BamHI and XhoI. **pCMV-Deoptflag** was constructed by the amplification of the Deopt cassette with primers *RepDeOptBglIIFwd* (CTGAATCATAAGATCTACCATGCCCGGGTTTTACGAG) and *RepDeOptXhoIRev* (CTGAATCATACTCGAGCTGTTCAAAAATGCAATCGTC). As BglII and BamHI generate compatible sticky ends, the digested PCR fragment was inserted into BamHI, XhoI digested vector backbone.

pTSRF has the WT-Repflag construct expressed under the TRE-TK promoter. The pCMV-tTS cassette and TRE-TK-WTflag cassettes are head to head, in the same orientation as the expression cassettes in pDRep. pBS-CMVtTS was constructed by the insertion of CMV-tTS cassette amplified *BclICMVFwd* and *BglIITSRevo06,* into the BamHI site in the MCS of pBlueScriptKS+ (*Stratagene,USA*). Next, the Thymidine kinase promoter fused to the Tet operator

(TRE-TK) was PCR amplified from pDRep using primers *TRETKNotIFwd* (CATTAATAAAGCGGCCGCGTCGAGGCCGCTCTAGTCGAG) and *TRETKEagIRev* (CATATAAATACGGCCGAGATCTGCGGCACGCTGTTGA) which introduced a 5' NotI site and a 3' EagI site. The digested PCR fragment was introduced into the NotI site in pCMV-WT-flag, in between pCMV and the WTflag ORF. The entire TRE-TK-WTflag cassette was then PCR amplified from this clone, using primers *TRETKNotIFwd* and *TagpolyANotIRev* (CATTAACATTAGCGGCCGCTCCCCAGCATGCCTGCTATT) and cloned into the NotI site in the MCS of pBS-CMVtTS, generating pTSRF.

pTROTS-WTflag was constructed by the PCR amplification of the WTRepflag ORF from pCMV-WT-flag with *SbfIRepFwd* (CATAACGACCCTGCAGGACCATGCCGGGGTTTTACGAGAT) and *SwaIFlagtagRev* (CAGTAGATATATTTAAATGGTACCGGGCCCTATTTATC). This amplicon had a 5' SbfI and 3' SwaI restriction site introduced by the primers, and was used to swap into the SbfI SwaI sites in pTROTS-ScrG in place of the Scr ORF. **pTROTS-Scrflag** and **pTROTS-DeoptFlag** were similarly constructed by PCR amplification with *XhoIRepScrFwd* or *XhoiRepDeoptFwd* respectively and *SwaIFlagtagRev*. Both *XhoIRepScrFwd* or *XhoIRepDeoptFwd* carried internal SbfI sites. The amplified sequences were digested with SbfI and SwaI and swapped into pTROTS-ScrG, in place of the Scr ORF.

For localization of the sequence signal, various sections of the Scrambled ORF were replaced with the corresponding wild-type Rep sequence in frame (Fig 6). **pTROTSWTnew** was constructed by PCR amplification of full length WT Rep from pIM45 (311, 312) with primers *SbfIRepFwd* and SwaIBglIIRepORFFRev followed by ligation into the SbfI SwaI digested pTROTS-ScrG. This ORF has is capable of expressing all 4 Rep proteins, due to an intact ATG and splice site. For **pTROTS-Scr(wt1)** bp1-660 of the wild type Rep sequence was amplified from pIM45 by PCR with primers *SbfIRep* and *AfeIRepWTRev* (CTATCAAGTAAGCGCTAGTTTTGATCTGATCACCGGC) which introduced a 3' AfeI restriction site. This 666 bp sequence was swapped into pTROTS-Scr between the SbfI and AfeI sites, replacing the corresponding 666bps of Scrambled sequence generating pTROTS-Scr(wt1). **pTROTS-Scr(wt2)** b was generated in two steps. The 648bp AfeI –BstBI wild-type Rep fragment (wt2) was PCR amplified from pTROTS-WTG using primers *AfeIRepWTFwd* (CTATCAAGTAAGCGCTAGGTACGGAGAGCTGGTCGGGT) and *BglIISwaIRepORFR*. The amplicon was digested with restriction enzymes AfeI and BstBI, the ~600bp fragment was separated and eluted from a 0.8% agarose

gel and cloned into the corresponding sites within pUCScr, making pUCScr(wt2). The entire ~1800bp Scr(wt2) Rep ORF was excised out of the backbone by digestion with SbfI and SwaI and swapped into the the SbfI SwaI site within pTROTS-ScrG, in place of the Scrambled cassette. making pTROTS-Scr(wt2). Similarly, for the construction of **pTROTS-Scr(wt3)**, pUCScr(wt3) was constructed by excision and gel elution of the BstBI – SwaI flanked 'wt3' region by restriction digestion of pTROTS-WTnew and ligation into similarly digested pUCScr, in place of the corresponding sequences within scrambled.. The entire Scr(wt3)Rep coding ORF was then swapped into pTROTS backbone, resulting in pTROTS-Scr(wt3).

pTROTS Scr(wt1,2) has the first 2/3rd of the Scrambled Rep sequence replaced with the corresponding WT sequence and was constructed in two steps. First the full length WT Rep was PCR amplified with primers *SbfIRepFwd* and *EcoRVRepORFRev*. This 1860bp amplicon was digested with SbfI and BstBI, and the 1200bp fragment was separated and agarose gel eluted, and inserted into SbfI, BstBI double digested pUCScr, generating pUCScr(wt1,2). The entire 1860 bp Scr(wt1,2) ORF was excised by digestion with SbfI and SwaI and inserted into the SbfI SwaI digested p-TROTS-G backbone. **pTROTS-Scr(wt2,3)** has the last 2/3rd of the Scrambled Rep sequence replaced with the corresponding WT Rep sequence in frame. The 600bp AfeI-BstBI fragment (corresponding to 'wt2' was excised out of pTROTS-Scr(wt2) by restriction digestion and inserted into the AfeI-BstBI sites in pUCScr(wt3) to generate pUCScr(wt2,3). The resultant ORF was excised by SbfI SwaI digestion and swapped into the pTROTS-G backbone. **pTROTS-Scr(wt1,3)** has ATG-bp660 and bp1309-stop of the Scr sequence replaced by the corresponding WT Rep sequences. ATG-bp660 of the Wild-type Rep sequence was PCR amplified from pIM45 using primers *SbfIRepFwd* and *AfeIWTRev* and ligated into SbfI, AfeI digested pTROTS-Scr(wt3) in place of the corresponding Scrambled sequence, generating pTROTS-Scr(wt1,3).

Sequences for design I, II, III and IV were designed by Dr. Charles Ward and Dr. Steve Skiena and synthesized *de novo* by GenScript, USA. Each ORF was excised out of pUC-I, pUC-II, pUC-III and pUC-IV respectively by double digestion with RE SbfI and SwaI and swapped into the pTROTS-G backbone in place of the Scr ORF. These contructs were labeled pTROTS-I-G, pTROTS-II-G, pTROTS-III-G and pTROTS-IV-G.

Virus construction and propagation: All viruses were constructed by homologous recombination in *E.coli*, between left end Ad sequences present in the shuttle vectors and a plasmid carrying the Ad genome as previously described

(95). Transgenes are inserted in the E1 locus, replacing Adenoviral sequences extending from Ad bp 450-3329.

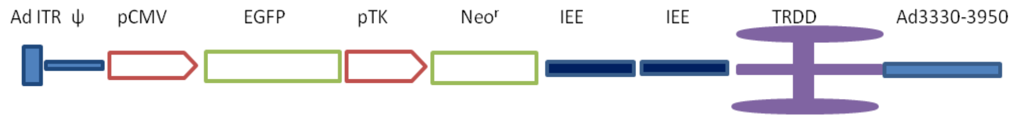
pTG3602 was modified by Dr. Patrick Hearing, SUNY Stony Brook to generate pTG3602 ΔE3ΔE4 F5/35 and pTG3602 ΔE3 F5/35 . pTG3602 ΔE3 F5/35 has the Ad5 fiber knob domain replaced by that of Ad35 and a 3Kb deletion in the E3 region (*dl7001*) (313). Deletion of E1 concomitant with the insertion of transgenes results in this virus requiring an E1 function complementing cell line such as 293 for growth. pTG3602 ΔE3ΔE4 F5/35 has an additional 3Kb deletion in the Ad E4 region (*dl366)* (314) and recombinant viruses on this backbone require an E1 and E4 complementing cell line such as 911-E4.

Ad genome vectors were transformed into dam$^-$ C2925H competent cells (NEB,USA) to remove methylation of the internal ClaI site. The viral genome vectors and shuttle vectors were each linearized by overnight digestion with restriction enzyme ClaI and PacI respectively, followed by treatment with calf alkaline phosphatase (*NEB,USA*) to prevent self ligation. 100ng of Ad genome vector and 1µg shuttle vector were co-electroporated into BJ5183 electroporation competent cells (*Stratagene, USA)* following manufacturer's directions. Putative recombinants were isolated from 3ml cultures by alkaline lysis (protocol below) and screened by PCR with primers EYFPPolyAFwd (CTGTACAAGTAAAGCGGCCCTAGA) and E2BRev (CAATCTGAACATCCAGAGTC) which span the transgene – Adenovirus junction. The name of all infectious constructs begins with a V (for virus). For all pTROTS- derived vectors, infectious clones were named after the modified Rep ORF they carried followed by G. All other infectious clones were named after the shuttle vector they were derived from.

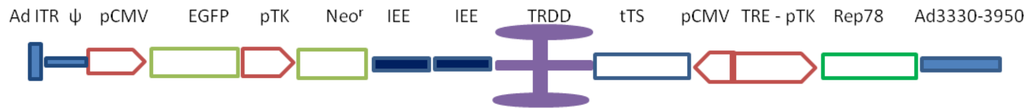
Positive clones were transformed into DH5α cells to scale up production. Viral constructs which bore the AAV hairpin (VDRep, VPRep and VPE (Ad/AAV-PF4FVIII)) were electroporated into SURE electroporation competent cells (*Stratagene, USA)* as DH5α cells were found to be unsuitable for maintenance of the intact AAV ITR. Transformed clones were confirmed by restriction digestion.

Fig.5: Design of plasmid constructs:

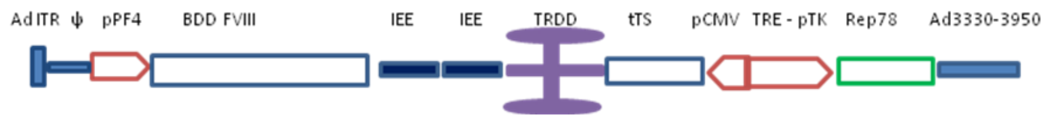
1. pD (Inderjit Chabbra, Dr. Patrick Hearing (Stony Brook Univ)).



2. pDRep

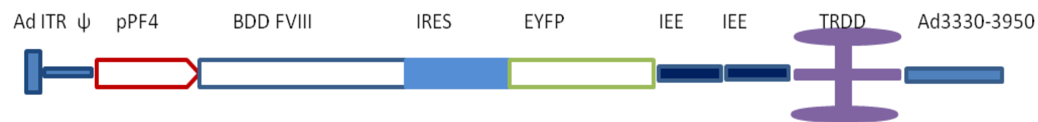


3.pPRep



3. pTROTS/pTRE-Rep78 series

4. pP-E/ (pPF4-FVIII-EYFP)



5. pTSRF

Legend:

pCMV :Cytomegalovirus promoter

EGFP :Enhanced green flourescent protein

pTK: Minimal thymidine kinase promoter

Neo^r: Neomycin Phosphotransferase gene

IEE: AAV integration efficiency element

tTS : Tetracycline controlled transcriptional silencer

TRDD: AAV Terminal Repeat

TRE: Tetracycline responsive element

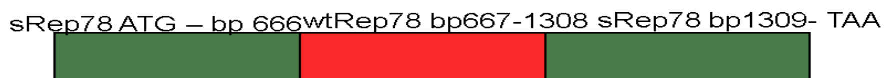
IRES: Internal ribosomal entry site

EYFP: Enhanced Yellow fluorescent protein

Fig. 6: Scrambled/wtRep ORF sequence swaps:

1: Scr(wt1)

2: Scr(wt2)



3: Scr(wt3)

4: Scr(wt1,2)

5: Scr(wt1,3)

6: Scr(wt2,3)

5µg of each of the infectious clones were digested with PacI to separate the plasmid backbone from the viral genome. Linear genomes were transfected into 80% confluent 6cm plates of HEK 293 packaging cells using Fugene 6 (*Roche, USA*), following manufacturer's directions. 10 days post transfection cells were lysed by freeze thaw and lysate used to infect fresh cells (primary infection). Cytopathic effect (CPE) was observed in all viruses capable of replication within 7 days of primary infection. At reaching CPE, cells were freeze thawed to lyse any intact cells, clarified by centrifugation and used to infect a 10cm plate of 293 cells (secondary infection). Complete CPE was consistently obtained with the secondary infection of all viruses 48 hours post infection. Viruses which did not reach CPE within the first 10 days of primary infection were serially passaged up to 50 days post transfection and observed for the development of CPE.

Concentrated recombinant viral stocks of VPE (Ad/AAV-PF4FVIII), VScr, VDeopt were obtained by infection of 10 10cm plates of 293 cells. Cells were infected with $1/20^{th}$ of the secondary infection lysate. When 100% of the cells had reached CPE, cells were harvested, spun down and resuspended in a total of 6 ml Phosphate buffered saline (PBS). Cells were lysed by 4X freeze thaw and lysate clarified by centrifugation. Purified virus was isolated by CsCl gradient purification as previously described. Viral lysate was overlayed onto a 1.25g/cc-1.4g/cc Cesium Chloride (CsCl) step gradient and centrifuged at 32000 rpm for 1 hour at 15ºC in an SW41 rotor (Beckman). Following this spin, the viral band was carefully removed using a needle and syringe, overlayed onto a 1.4g/cc CsCl cushion and topped with 1.35g/cc CsCl and centrifuged at 41000 rpm for 16 hrs at 15ºC in an SW65 rotor.

The concentrated purified viral band was carefully sucked out of the tube using a needle and syringe. To determine viral titer in particles (p)/ml , aliquot were diluted and lysed in 10mM Tris.HCl, 1mM EDTA (TE) with 0.1% SDS, centrifuged for 5 minutes to remove any precipitate formed and absorbence at 260 nm read. One OD (Optical density unit) corresponds to 10^{12} p/ml. For long term storage, virus was diluted in virion storage buffer (10mM Tris.HCl, ph8.0, 100mM NaCl, 0.1%BSA, 50% glycerol and 1mM Magnesium Chloride ($MgCl_2$)) and stored at -20ºC.

Viruses Ad5/35Pf4-EGFP-Neo and Ad5 PF4-FVIII-IEE-TRDD were obtained from Dr. Patrick Hearing (Stony Brook University).

<u>Alkaline Lysis protocol for plasmid preparation:</u> Plasmid constructs <20Kb in size were efficiently isolated using the Qiagen spin column (*Qiagen,*

USA) following manufacturer's instructions. For larger constructs, such as the infectious clones, small scale plasmid preparations needed to be done without a column. A standard protocol for plasmid isolation by alkaline lysis was followed. Briefly, 1.5ml -3ml of bacterial culture was spun down and supernatant discarded. The pellet was resuspended in 250µl of buffer resuspension buffer P1 (50Mm Glucose, 25Mm Tris.HCl pH 8.0, 10mM EDTA pH8.0). 250µl of lysis buffer P2 (0.2N NaOH, 1% SDS) was added and mixed gently, followed by addition of 350 µl neutralization buffer P3 (3M Pottasium Acetate, pH5.5). Precipitates were centrifuged at top speed for 10 min on a table top centrifuge and supernatant transferred to fresh tubes. The supernatants were extracted with an equal volume of Phenol-Chloroform-Isoamyl Alcohol(PCI) (25:24:1), followed by precipitation with 0.7 volumes of isopropanol. Pellets were washed in 1ml 70% Ethanol for 10 minutes, dried and resuspended in 30µl 10mMTris.HCl, pH8.0. About 3µl of plasmid prep was used for PCR analysis and 20µl for analysis by restriction digestion.

Infection of HEL cells and Flow cytometry analysis: ~10^5 HEL cells were incubated with VPE (Ad/AAV-PF4FVIII) at an MOI of 1000 and 10000. Ad5/35PF4-EGFP-Neo at an MOI of 1000 was used as a positive control. Uninfected cells were used as negative controls. Cells were incubated with virus in 200µl of RPMI with 2% FBS for 10 minutes at room temperature, followed by centrifugation at 3000rpm for 2 hours at room temperature in a Beckman tabletop GPR centrifuge. Post centrifugation, the cells were resuspended in 2 ml complete media (without removal of infectious media) and incubated at 37ºC, 5% Co_2 overnight in 6 well plates. 24 hours post infection, cells were washed and resuspended in fresh media. 72 hours post centrifugation, cells were spun down, resuspended in 1ml of Phosphate buffered saline (PBS) and analyzed by flow cytometry for expression of EGFP or EYFP using a filter with 488nm excitation wavelength (the EGFP excitation maximum). The excitation maximum for EYFP is 513 nm and can be detected as well using the same filter.

Hirt DNA prep: Hirt DNA prep (315) for low molecular weight DNA was performed following a modified extraction protocol. Briefly, cells were washed in PBS and the pellet was resuspended in 200µl Resuspension buffer (10mM Tris.HCl,pH8.0, 10mM EDTA). An equal volume of Lysis buffer (10mM Tris.HCl, pH8.0, 10mM EDTA, 1.2% SDS) was added, mixed gently and incubated and room temperature for 10 minutes. NaCl to a final concentration of 1M was then added and allowed to precipitate at 4ºC for 16 hours. Precipitates were centrifuged at 4ºC for 60 minutes at 13000rpm in a table top centrifuge. Supernatants were transferred to a fresh tube and digested with 2mg/ml RNase for 30 minutes at 37ºC, followed by digestion with 300µg/ml proteinase K for 30 min

at 37ºC. Digested samples were extracted with PCI and precipitated with 0.3M Sodium acetate and 1ml 100% Ethanol at -80ºC for 20 minutes. Pellets were washed with 70% Ethanol and resuspended in 50µl 10mM Tris.HCl, pH8.0.

DpnI Viral Replication Assay: A DpnI assay (316) for viral replication was performed by transfection of 293 cells in 6 well plates with 1.5µg of linearized infectious clone construct. Cells were washed 3X with PBS and fed with fresh media 24 hours post transfection. Cells were washed, spun down and frozen away at -20ºC at different time points. Samples for quantitative PCR analysis were collected day 1, day 4, day8 and day 12 post-transfection. Samples for analysis by southern blot were collected on Day 1 and Day 10. Pellets were thawed on ice and Hirt DNA isolated as described above.

For analysis of replication over time by quantitative PCR, Hirt DNA was first diluted to a 20ng/µl solution. 5µl (100ng) of DNA from this dilution was digested with 10 units of DpnI for 8 hours at 37ºC , followed by heat inactivation for 20 minutes at 80ºC. The reaction mixture was diluted five fold in water and $1/20^{th}$ was used in a quantitative PCR reaction. PCR reactions were performed with Sybr Green Master mix (*Qiagen, USA*) on the Applied Biosystems 7300 Real Time PCR system. Primers used were Ad3330qpcrL (CTGTGATGCTGGATGTGACC) and Ad3330qpcrR (TGCTTCCATCAAACGAGTTG) which flank a DpnI site between Ad3412 and Ad3536. Results were normalized to 5ng of undigested Hirt sample. Quantitation was by comparison to a standard curve generated with serial dilutions of the pHu1-Scr shuttle vector.

DpnI assays to confirm ability of viral constructs to replicate were performed on VScr(wt1), VScr(Wt2) and VScr(WT3) as well as VIG, VIIG, VIIIG and VIVG. Hirt DNA was prepared from transfected 293 cells at Day 1 and Day 10 post transfection. 2µg of DNA was digested with DpnI overnight. Digested samples were run alongside equivalent undigested samples and amount of intact DpnI resistant viral DNA was detected by Southern blotting as described below. 680bp digoxigenin labeled probes for detection of the Ad backbone were amplified from pTG3602 ΔE3 F5/35 with primers *AflII3520 Fwd* (GTGTGGGCGTGGCTTAAGGG) and *E2B Rev* and encompass Ad bp 3520-4199.

Western blots: Western immunodetection of protein levels was performed following standard procedure. 293 cells were were transfected with 1µg of plasmid DNA using Fugene6 (Roche, USA) following manufacturer's instructions. If applicable, cells were induced with 1µg/ml doxycycline 3 hours

post transfection. 48 hours post-transfection, cells were lysed in NP40 lysis media (50mM Tris.HCl pH8.0, 150mM NaCl, 1% NP-40). Concentration of protein was determined by a standard BCA assay (Bicinchoninic Acid Assay), using the BCA Protein Assay Kit (Pierce, USA). 10µg of total protein was reduced with 10% β-Mercaptoethanol (βME) and denatured by boiling and separated on a 4-15% gradient polyacrylamide gel. Separated proteins were transferred overnight onto a nitrocellulose membrane, blocked with 3% non fat milk and probed with the relevant primary antibody. Mouse monoclonal anti-flag M2 (SigmaAldrich, USA) was used for detection of Flag tagged proteins. Detection of the GAPDH protein as a loading control was performed using the mouse monoclonal anti-GAPDH MAB374 (Millipore,USA). The secondary antibody used to probe the primary antibody-substrate complexes was the ECL Anti-mouse IgG Horseradish peroxidase linked $F(ab')_2$ fragment from sheep (GE Healthcare, UK). Detection was performed using the Pierce ECL Western Blotting substrate (Thermo Scientific, USA) following standard protocol.

Excision assay for detection of functional Rep78 expression: The presence of Rep expression was assessed by an excision assay which detects the ability of Rep protein to bind to and cleave specifically at the folded AAV ITR. The Ad/AAV VPE (Ad/AAV-PF4FVIII) was used as the excision substrate. 90% confluent 293 cells in 6cm plates were co-infected with the Rep expressing virus and VPE (Ad/AAV-PF4FVIII) at an MOI of 50 each. In the case of Rep expressing plasmids, 293 cells in 6 well plates were transfected with 2µg of the plasmid using Fugene 6 (Roche,USA) following manufacturer's instructions, 48 hours before infection with VPE (Ad/AAV-PF4FVIII). Cells were induced with 1µg/ml doxycycline 3 hours post transfection/co-infection. C12 cells (265) which inducible express Rep were used as a positive controls. Cells infected with VPE (Ad/AAV-PF4FVIII) only, with no Rep source were used as a negative control. 48 hours post infection with VPE (Ad/AAV-PF4FVIII), cells were lysed and Hirt DNA prepared as described above. 3µg of Hirt DNA was analyzed by southern blotting using the DIG Easy Hybrid System (Roche,USA) with non radioactive Digoxigenin labeled probes which recognized a ~700bp sequence at the junction of pPF4- FVIII. Probes against the fVIII transgene were PCR amplified from pBSPF4fVIII with primers PF4Fwd and FVIII Rev (317), encompassing 700bp at the junction of pPF4 and FVIII.

Southern blotting: Isolated Hirt DNA was analyzed by a modified method of Southern (318), using the DIG Easy Hybrid non radioactive system.. Briefly, approximately 3µg of DNA was separated by electrophoresis on a 0.8% agarose gel in TAE (Tris-Acetate EDTA buffer 400mM Tris-Acetate, 10mM EDTA). Gels were washed in a Denaturation solution of 1N Sodium Hydroxide solution

for 40 minutes, followed by 3 washes for 20 minutes each in Neutralization buffer (1M Tris.HCl, 1.5M NaCl, pH7.0). Gels were finally washed for 20 minutes in ample 6X Saline Sodium Citrate Buffer (SSC - ~1M Sodium Chloride, 100mM Sodium Citrate, pH7.0) and transferred overnight in 6X SSC onto a positively charged nylon membrane (Roche, USA) by capillary transfer. Nylon membranes were fixed by UV exposure for 4 minutes, dried briefly, washed and hybridized with the specific DIG easy labeled probe, using the DIG Easy Hybrid protocol with some modifications. The pre-hybridization and hybridization temperature was increase to 65ºC, from the recommended 42ºC, with digoxigenin labeled probes. An extra stringency wash in 0.1XSSC, 0.1%SDS at 68ºC was also performed. All other washing and detection steps were performed in accordance with the manufacturer's protocol.

Viral titer determination: To establish a standard protocol for determination of viral titer among different viruses, 90% confluent 293 cells in 6cm plates were infected in duplicate with 50µl of secondary infection lysate. Cells were harvested when >50% of cells were floating, freeze- thawed to lyse intact cells and clarified by centrifugation. The viral titer of this lysate was determined by infecting 90% confluent 293 cells in 6 well plates with serial dilutions of the lysate. Cells were incubated with the virus for 1 hour, infection media aspirated and overlayed with complete media (DMEM, 10% FBS) containing 1% agarose. 10 days later, cells were stained with 0.03% neutral red, for identification of plaques.

Plaque forming units/µl = Number of plaques x dilution factor.
Pfu/cell = pfu/µl x total volume of lysate/ 8.8 x 10^6
Where 8.8 x 10^6 = Average number of 293 cells per 10 cm plate.

Results

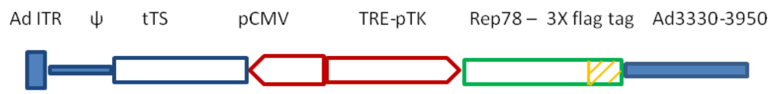
Construction and replication of second generation Ad/AAV carrying the AAV ITR and a Rep expression cassette (VDRep and VPRep): VDRep and VPRep are second generation ΔE1ΔE3ΔE4 F5/35 Ad/AAV infectious constructs bearing the transgene cassettes from pDRep and pPRep respectively. pDRep carries an EGFP reporter gene expressed under a CMV promoter and the gene for Neomycin phosphotransferase expressed under the minimal thymidine kinase promoter while pPRep carries a B domain deleted (BDD) coagulation FVIII gene expressed under the platelet factor 4 promoter. Both constructs have two copies of the AAV Integration Efficiency Element and one copy of the AAV TRDD downstream of their transgene cassettes. Downstream of the AAV elements in both constructs lies a tetracycline inducible Rep78 expression cassette. (shuttle vector layout - Fig 5). Infectious clones were linearized and transfected into 911-E4 packaging cells. Based on visualization of EGFP expression from the pCMV-EGFP expression cassette carried by VDRep under the fluorescence microscope, it was estimated that about 70% of cells were transfected. 10 days post-transfection, cells were lysed and used to infect fresh cells. However, no EGFP expression could be detected on these cells infected with lysate from VDRep transfections. Cells were lysed and passaged every ten days onto fresh cells for 5 cycles. EGFP expression beyond faint fluorescence in rare cells was never detected. Similarly, cells infected with the VPRep construct also showed no signs of viral replication over 5 passages.

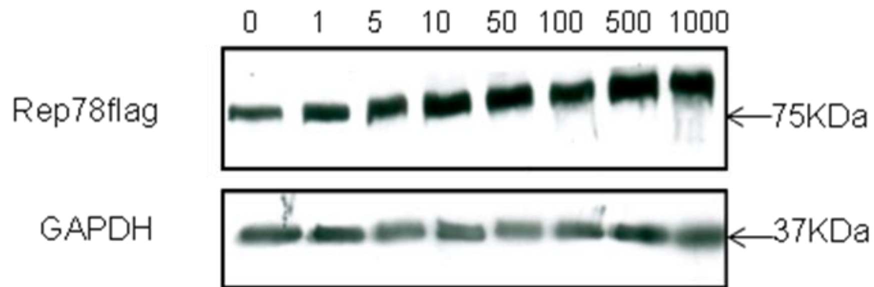
Regulation of expression of Rep from the tetracycline inducible construct was measured by immunodetection of transfected flag-tagged plasmids (pTSRF) (Fig.7). Expression of Rep from the tetracycline inducible system was found to be not tightly regulated, with considerable expression detected even in the absence of inducer. This was possibly due to the head to head orientation of the CMV promoter and Tetracycline responsive element (319). We sought to reduce Rep expression by the use of cell lines expressing siRNA against Rep. siRNA-911E4 #8 cells were transfected with linearized VDRep and VPRep, to assess the effect of knock down of Rep on replication of these constructs. However, similar to 911-E4 cells, no signs of viral replication were obtained with either construct up to 50 days post-transfection.

Several factors could have contributed to the inability of these constructs to replicate. As mentioned earlier, Rep expression from the tetracycline inducible system was leaky, with detectable levels of Rep expression even in the absence of

Fig 7: Expression of flag tagged Rep78 under the tetracycline inducible system . HeLa cells were transfected with pTSRF, and induced with different levels of doxycycline 24 hours later. 48 hours later, cell lysates were probed using an anti-flag antibody. Leaky expression is seen even in the absence of doxycycline, possibly due to the head to head orientation of promoters in the plasmid construct.

pTSRF



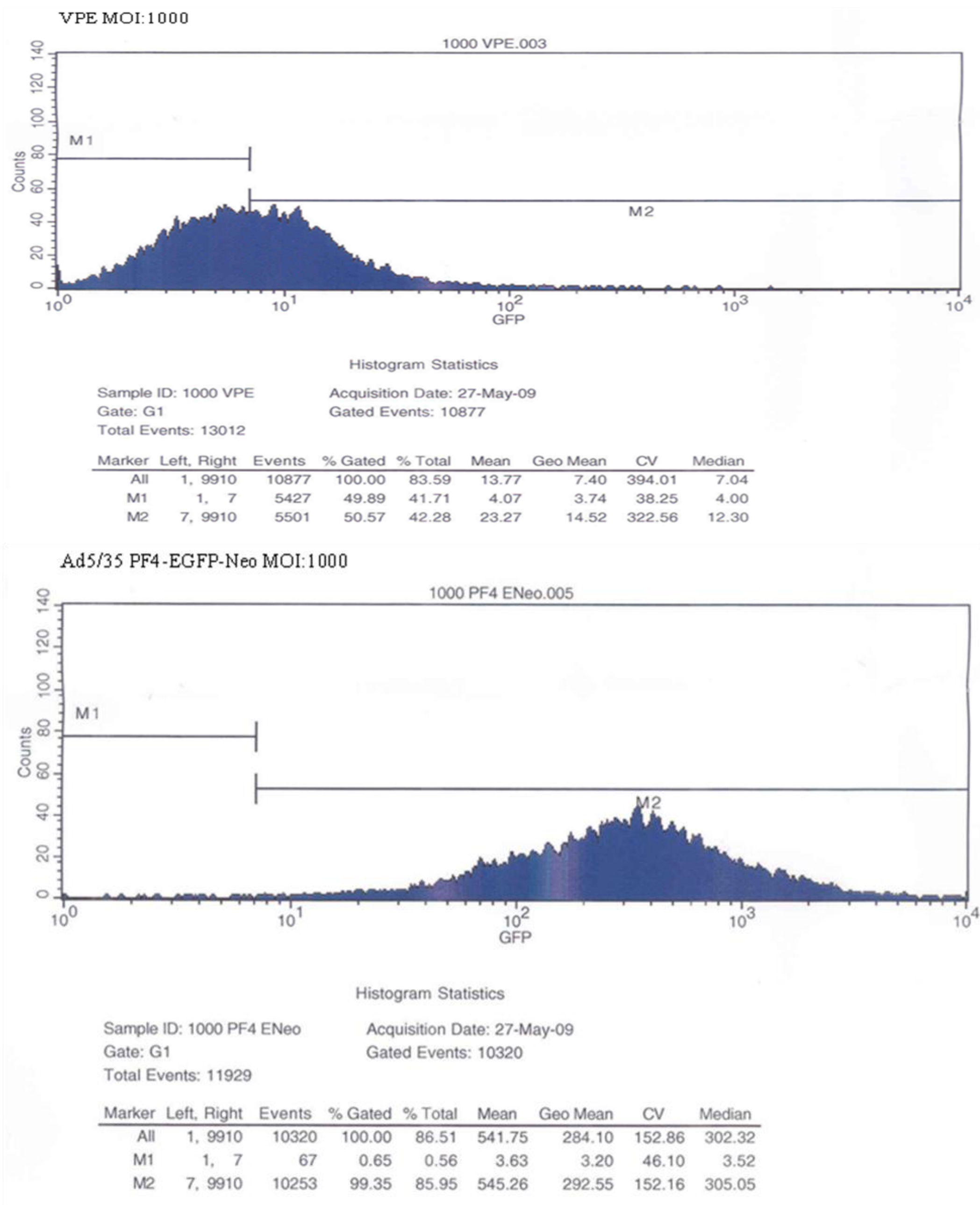


doxycycline. As Rep protein has been shown to inhibit Ad replication (218, 222, 224, 225), it was considered necessary to reduce Rep levels to allow replication. Secondly, it was noticed that pTG3602 ΔE3ΔE4 F5/35, the Ad genome vector which forms the backbone of these infectious clones, itself replicates inefficiently. This was possibly due to large deletions in the essential E4 region required to accommodate the large transgenes. To bypass both these problems, it was proposed that the vector be split into two first generation ΔE1ΔE3 F5/35 Ads, with one virus carrying the AAV ITR flanked transgene and the other carrying the Rep expression cassette cloned in a head to tail orientation.

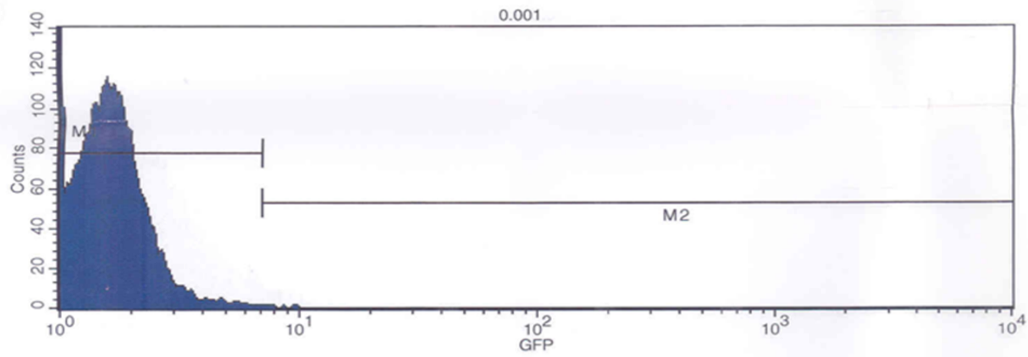
Construction and replication of VPE (Ad/AAV-PF4FVIII): VPE (Ad/AAV-PF4FVIII) was constructed on an ΔE1 ΔE3 F5/35 Ad backbone and carries a PF4-FVIII expression cassette (309) with an EYFP reporter gene expressed under an attenuated IRES, downstream of the FVIII gene (Fig 5: pP-E). Downstream of the transgene cassette are two copies of the AAV IEE followed by the AAV TRDD. VPE (Ad/AAV-PF4FVIII) was linearized by digestion with restriction enzyme PacI, transfected into low passage 293 cells and lysed to infect fresh cells in 10 days. A total of 14 days post transfection nearly 100% of cells displayed CPE. Purified concentrated viral stocks were obtained at a concentration of 1.88 x 10^{11}p/10 cm plate of 293 cells and the presence of an intact AAV ITR was confirmed by sequencing of purified viral DNA. A functional titer of 5.8 x 10^{3} pfu/cell was obtained from non-concentrated clarified lysates. A preliminary characterization of the virus was done by infection of HEL (320)cells which allow the expression of the Platelet Factor 4 promoter. A dose dependent expression of EYFP was detected and quantified by flow cytometry (Fig 8). EYFP expression could be detected in only ~50% of cells infected with VPE (Ad/AAV-PF4FVIII) at an MOI of 1000, while ~99% of cells infected with Ad5/35PF4-EGFP-Neo at the same MOI were EGFP positive. This difference is likely due to a difference in expression levels of EGFP vs EYFP as EYFP is expressed under an attenuated IRES under the PF4 promoter. Further, sensitivity of EYFP detection was reduced compared to EGFP due to settings of the flow cytometer. The difference in expression levels was represented in the Geometric mean values (VPE (Ad/AAV-PF4FVIII) – 14.52, Ad5/35PF4-EGFP-Neo – 292.55) which are a measure of the average strength of the fluorescence of positive cells. Expression of functional FVIII from VPE (Ad/AAV-PF4FVIII) has not been determined.

Construction of VRB (Ad/wtRep78): The Rep78 expression cassette was cloned in the appropriate head to tail orientation of the pCMV-tTS and TRE-pTK-Rep78 expression cassettes, and inserted into a ΔE1 ΔE3 F5/35 Ad backbone (for shuttle vector layout see fig 5 pTROTS series). This construct

Fig 8. Flow cytometry analysis of EYFP expression from VPE (Ad/AAV-PF4FVIII): Human Erythroleukemic cells (HEL cells) were infected with VPE (Ad/AAV-PF4FVIII) or control Ad5/35 PF4-EGFP-Neo (Dr. Patrick Hearing, Stony Brook University) at an MOI of 1000, by centrifugal enhancement. 72 hours post-infection, cells were analyzed for EGFP expression by flow cytometry.



Uninfected cells – Negative control



Histogram Statistics

Sample ID: 0
Gate: G1
Total Events: 10753

Acquisition Date: 27-May-09
Gated Events: 9916

Marker	Left, Right	Events	% Gated	% Total	Mean	Geo Mean	CV	Median
All	1, 9910	9916	100.00	92.22	1.73	1.61	46.75	1.58
M1	1, 7	9888	99.72	91.96	1.71	1.61	40.00	1.58
M2	7, 9910	28	0.28	0.26	9.29	8.93	33.39	8.39

bears an EBFP reporter gene expressed under an attenuated IRES downstream of the Rep78 ORF. Expression levels of Flag tagged Rep protein from this construct were too low to detect in whole cell lysates, indicating a tighter control of Rep expression than that afforded by the opposite orientation of cassettes. Infectious clones were linearized and transfected into 293 cells. However, in spite of tighter regulation of Rep control and a first generation Ad backbone which was capable of efficient replication as evinced by the replication of VPE (Ad/AAV-PF4FVIII), no signs of viral replication could be seen with this construct up to 50 days post transfection. Similarly, VWTG (Ad/wtRep) was constructed which carried the unmodified AAV Rep ORF, capable of expressing all 4 Rep proteins, expressed under the tetracycline inducible promoter. This construct was also found to show no signs of viral replication for 50 days post-transfection.

Modification of the Rep ORF: The absolute lack of viral replication seen even with tightly regulated Rep expression contrasted with evidence pointing to Adenoviral replication in the presence of Rep expression. This contrast points to an enhanced inhibitory role played by Rep when present as part of the Adenoviral genome. We reasoned that this could be purely due to increased expression of Rep by an increase in copy number due to replication of the Adenoviral genome, or could be due to a role for the DNA sequence of the Rep ORF.

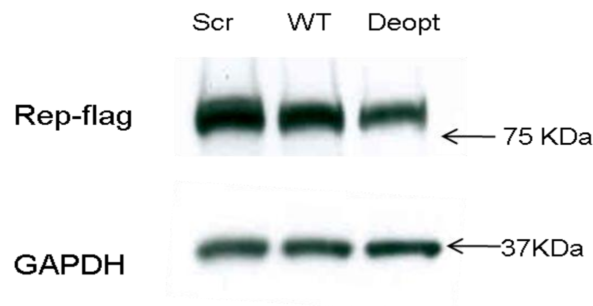
To elucidate the relative contribution of the DNA sequence of the Rep ORF and Rep protein levels on this apparently *cis*-acting inhibitory effect and to construct an Adenovirus carrying Rep, we collaborated with the laboratory of Dr. Eckard Wimmer and Dr. Steven Skiena to modify the 1866bp Rep78 nucleotide sequence *in silico*. An algorithm was utilized (304), that allowed us to modify the nucleotide sequence of Rep78 by 20-30% without affecting the amino acid sequence encoded, using synonymous codons. (Sequences and sequence alignments in Appendix). Two modified Rep sequences, Scrambled and Deoptimized, were designed and synthesized *de novo*. The Scrambled sequence randomly mixes synonymous codons, resulting in a nucleotide sequence that differs from the wild-type sequence by 30%. The protein expressed from this ORF is identical to wild-type Rep78 and is expected to be expressed at similar levels. This sequence aims to disrupt any sequence specific signal (SSS), without affecting Rep78 expression levels. Within the Deoptimized sequence, synonymous codons are specifically paired into under-utilized codon pairs. Synonymous codons can be paired in multiple ways to encode the same two adjacent amino acids. However, in nature a strong codon pair bias is found to exist, resulting in the disproportionate representation of some codon pairs over others (321). This codon pair bias is independent of codon frequency and is found to affect translation rates (304, 305). Utilization of under-represented codon pairs

Fig 9: Analysis of modified Rep78 constructs. 9a) Comparison of calculated codon pair bias scores of wild-type Rep78, Scrambled Rep 78 and Deoptimized Rep78. 9b) Western blot for expression levels from flag-tagged wtRep78, Scr Rep78 and Deopt Rep78 ORFs, expressed under a CMV promoter. Levels of GAPDH serve as loading controls. 9c) Densitometry analysis of Western blot bands for quantification of expression levels was performed using Gel Pro Analyzer 3.0. Levels of Rep expression were normalized to GAPDH levels.

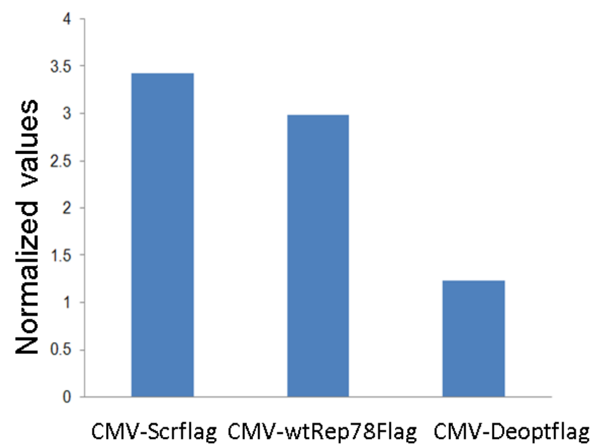
9a)

Rep78 coding ORF	Codon pair bias score
Wild-type	-0.043
Scrambled	-0.117
Deoptimized	-0.443

9b)



9c)



such as those in Deoptimized, therefore, results in an ORF with a lower codon pair bias score, that is expressed at lower levels due to inefficient translation. Thus, the Deoptimized Rep construct not only differs from the nucleotide sequence of wild-type Rep by 20%, presumably disrupting any sequence specific signal, but also further reduces levels of Rep 78 expression from the Tet-inducible promoter. Confirmation of Deoptimized Rep's reduced ability to express protein was obtained by immunoblot analysis of transfected C-terminal flag tagged constructs, expressed under pCMV. Densitometry analysis of western blots revealed that protein levels from Wild-type and Scrambled Rep were comparable to each other and roughly double that of Deoptimized (Fig 9c)

Modification of Rep ORF allows replication of Ad: The Scrambled and Deoptimized Rep constructs were cloned downstream of the tetracycline inducible promoter, in place of the wild-type Rep78 ORF, within the fiber modified first generation Adenovirus genome, generating infectious clones VScrG and VDeoptG (also referred to as Ad/sRep78 and Ad/dRep78). These constructs bear an EGFP reporter gene expressed under an attenuated IRES, downstream of the modified Rep 78 ORF. The viral constructs were linearized and transfected into HEK 293 packaging cells and observed for the development of CPE. In contrast to the lack of replication seen with VRB (Ad/wtRep78), complete CPE was observed with both VScrG and VDeoptG within a total of 15 days from transfection, within the primary infection. Replication of these viruses over a 10 day time course was compared to each other and to VPE (Ad/AAV-PF4FVIII)) using a modified DpnI viral replication assay. Restriction enzyme DpnI is capable of digesting only dam methylated DNA (of bacterial origin). Digestion with DpnI thus removes input viral DNA which is of bacterial origin. As a result, any DpnI resistant DNA detected is due to replication of the viral DNA in mammalian cells. Fold increases in DpnI resistant replicated viral DNA was analyzed by quantitative real time PCR using Adenovirus specific primers. Copy numbers of VScrG and VDeoptG steadily increased and were comparable to one another and to VPE (Ad/AAV-PF4FVIII) (Fig 10). Though no signs of viral replication were observed in 293 cells, a slower increase in replicated viral DNA from Ad/wtRep78 was also observed with the Q-PCR based assay, until Day 13 post-transfection. Production of both VScrG and VDeoptG could be scaled up with yields comparable to each other and to VPE (Ad/AAV-PF4FVIII) (Fig 15b), proving a clear role for the sequence of Rep in the inhibition of Adenoviral replication.

Noteworthy is the lack of any apparent difference in the ability of VScrG and VDeoptG to grow in spite of their differences in Rep expression levels. It indicated that at least under the control afforded by the tetracycline inducible

Fig 10: DpnI assay for viral replication. Hirt low molecular weight DNA was isolated from cells transfected with VRB (Ad/wtRep78), VScrG, VDeoptG or VPE (Ad/AAV-PF4FVIII) 24 hours, 4 days, 9 days and 13 days post transfection. Units of DpnI resistant viral DNA per unit of Hirt DNA was determined by Q-PCR. Values for each time point were the mean of twp individual experiments, each performed in duplicate. Results were represented as fold change in DpnI resistant DNA over Day 1. Error bars represent variation about mean.

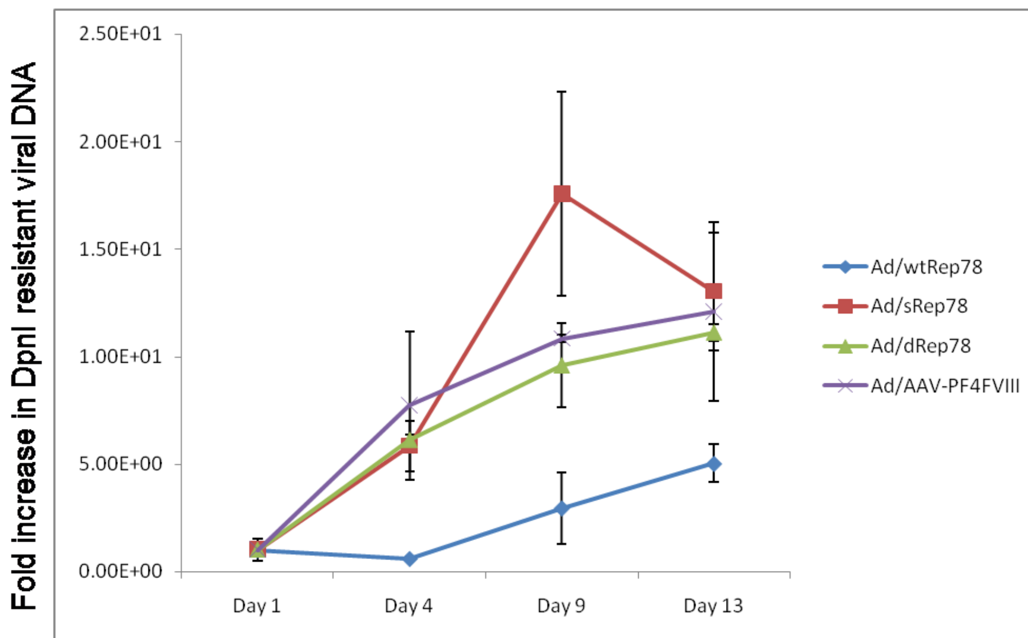
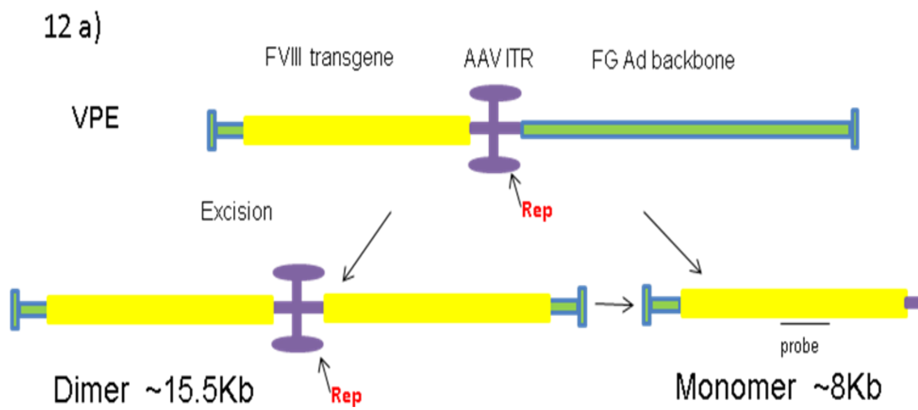
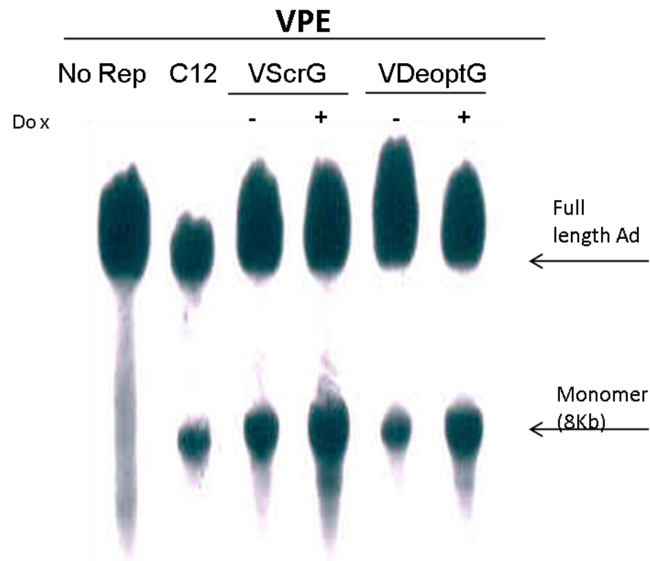


Fig 11: Doxycycline sensitive expression of functional Rep from Ad/sRep78 and Ad/dRep78 was measured by an AAV TR excision assay in 293 cells. The assay measures the ability of functional Rep to site specifically bind at the RBE within the folded AAV TR, act as a helicase to unwind and expose the *trs* and as an endonuclease to nick at the *trs*. VPE (Ad/AAV-PF4FVIII) was used as a substrate. Excision of VPE (Ad/AAV-PF4FVIII) in C12 cells (a cell line expressing Rep) co-infected with wt Ad, was used as a positive control. Cleavage at the AAV TR within VPE (Ad/AAV-PF4FVIII) results in an ~8Kb excision product, which is detected by Southern blot. 12a) Schema of assay. 12b) Southern blot with excision products.





system, the major role in inhibition of Ad replication was played by a sequence specific signal and modification of that signal alone was sufficient to completely lift inhibition.

The ability of VDeoptG and VScrG to produce functional Rep78 was confirmed by an excision assay which depends on Rep's ability to cleave at a folded AAV ITR (Fig 11). A first generation Ad/AAV carrying a single AAV ITR downstream of the FVIII (VPE) was used as a substrate for cleavage. Cleavage at the ITR by Rep would result in the release of an ~8Kb excision product. HEK-293 cells were co-infected with the substrate virus VPE (Ad/AAV-PF4FVIII) and either VDeoptG or VScrG in the presence or absence of doxycycline. Hirt DNA was prepared 48 hours post-infection and cleavage products were analyzed by southern blot with a substrate specific probe. Excision of VPE (Ad/AAV-PF4FVIII) in C12 cells, a HeLa cell line derivative that inducibly expresses Rep and Cap (265) was used as a positive control. Monomeric excision products comparable to those obtained with C12 cells, that were dependent on the presence of VScrG or VDeoptG were detected. Leaky expression resulted in some excision even in the absence of doxycycline, and a several fold increase in intensity seen with the addition of dox.

Localization of sequence specific signal (SSS): Two different strategies were employed to attempt to localize the Rep78 inhibitory sequence. In the first strategy, sections representing about 1/3rd of the entire scrambled sequence were replaced with that of WT, in frame, using unique internal restriction sites. Three different modified Scr constructs were thus made – Scr(wt1) with ATG-bp660 of the scrambled sequence replaced with that of WT, Scr(wt2) with bp 667-1308 replaced with the corresponding wt sequence, and Scr(wt3) with bp1309-TAA replaced (Fig 12). Each of these constructs were expressed under the Tetracycline inducible system cloned downstream of the tetracycline inducible promoter, in place of the wild-type Rep ORF, within the fiber modified first generation Adenovirus genome, to make VScr(wt1)G, VScr(wt2)G and VScr(wt3)G. The recombinant viral genomes were linearized and transfected into 293 cells. VScr(wt1)G and VScr(wt2)G both showed signs of viral replication shortly after primary infection and complete CPE within 15 days of transfection, comparable to replication rates of VScrG. Pfu/cell yields from clarified lysates from these constructs were also comparable to those obtained from VScrG (Table1). VScr(wt3)G showed no signs of replication for up to 40 days post-transfection, indicating that the sequence specific signal is localized within the 3' 555bps, in the region encompassing bp 1309-1866 of the wt Rep78 ORF.

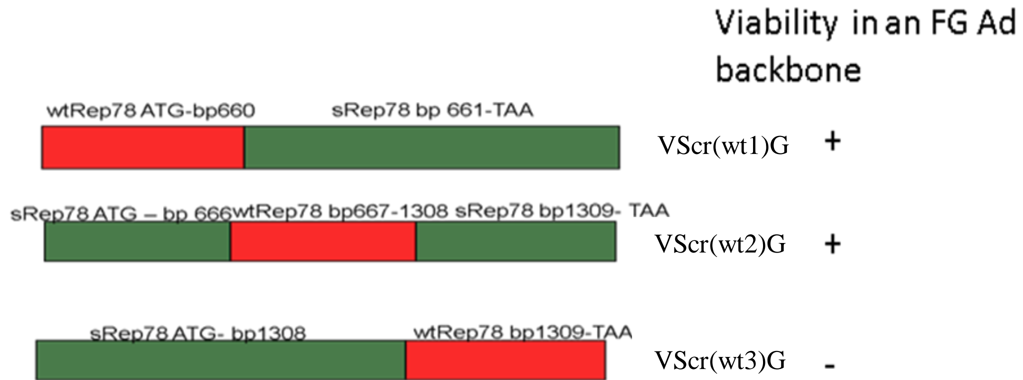
Opposite swaps with Rep expression cassettes with 1/3rd of the WT cassette replaced with the corresponding Scr sequence were also made, using the same restriction sites as the previous swaps. These viruses were Vscr(wt1,2)G, VScr(wt1,3)G and VScr(wt2,3)G. VScr(wt1,2)G which bears a Rep expression cassette with bp1-1308 from wtRep78 and only the last 555bps from Scr was capable of replication, with cells reaching CPE within 16 days of transfection, comparable to VScrG. Also, both VScr(wt1,3)G and VScr(wt2,3)G, which carry the wt3 segment showed no signs of replication for up to 50 days post-transfection. This further supports the results of the first swap, that localizes the sequence specific signal to the 3'555 bps of WT Rep78

In an alternative approach, the lab of Dr. Steven Skiena developed a novel algorithm to further narrow down cis-acting sequences that affect viral replication. The algorithm generates 4 different full length Rep encoding sequences (Design I, II, III and IV) (Sequences in appendix). Each of these sequences is sub-divided into 14 segments, consisting of corresponding sequences from either the WT or Scr sequence resulting in a 'checkered' pattern of segments. The arrangement of segments in the four full length sequences is such that, when the four sequences are lined up, every column (consisting of the corresponding segment in each of the four constructs) is unique. When cloned into an expression cassette in a viral backbone, the presence of the sequence specific inhibitory signal in a particular segment will result in all viruses carrying the WT sequence in that segment dying. Thus a unique pattern of viruses that live and die is associated with every possible location of the sequence specific signal.

The four modified sequences were synthesized *de novo* by Genscript, USA and cloned downstream of the tetracycline inducible promoter, in place of the wt Rep cassette. These cassettes were inserted into first generation Ads, generating VIG, VIIG, VIIIG and VIVG. Each of these infectious clones were linearized and transfected into 293 cells. VIIG, VIIIG and VIVG replicated as well and yielded titers comparable to VScrG (Fig 15b). VIG showed no signs of viral replication up to 50 days post transfection (fig 13). The 135bp segment implicated by the pattern of viruses capable of replication (VIIG, VIIIG and VIVG) and incapable of replication (VIG), encompassed bp 1462-1596 of wtRep78. Based on the arrangement of segments within the 4 sequences, some or all of the inhibitory signal definitely lies within this 135bp fragment. This sequence lies well within the boundaries of bp 1309 – 1866 identified by the inability of VScr(wt3)G to grow, overlaps the AAV p40 promoter and includes Rep78/68 and Rep52/40 splice site (Fig14).

Fig 12: Localization of sequence specific signal I. 12a) Schema of swaps between Scrambled and wtRep ORFs 12b) Southern blot of DpnI digested Hirt DNA of VScr(wt1), VScr(wt2) and VScr(wt3) 24 hours and 10 days post-transfection. No DpnI resistant DNA is visible 24 hours post transfection for all three viruses, though a positive signal corresponding to DpnI sensitive transfected DNA is obtained. 10 days post transfection, a DpnI resistant band representing replicated viral DNA is seen with VScr(wt1) and VScr(wt2) . However, no DpnI resistant or sensitive viral DNA is seen VScr(wt3) commiserate with no signs of replication obtained with that virus.

12 a)



12 b)

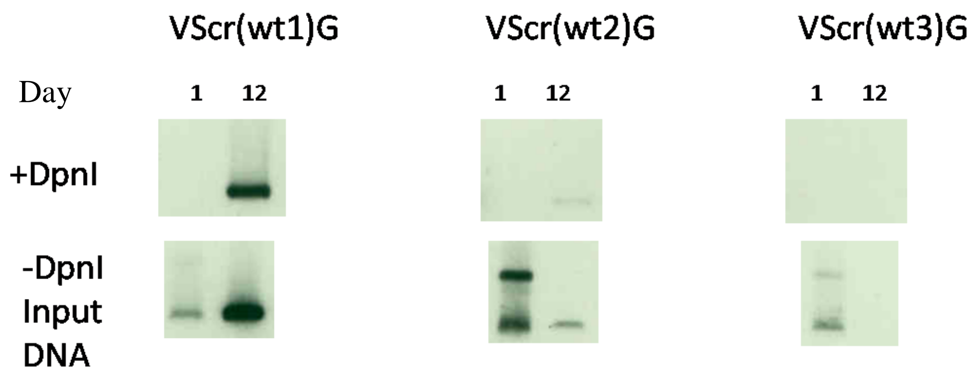
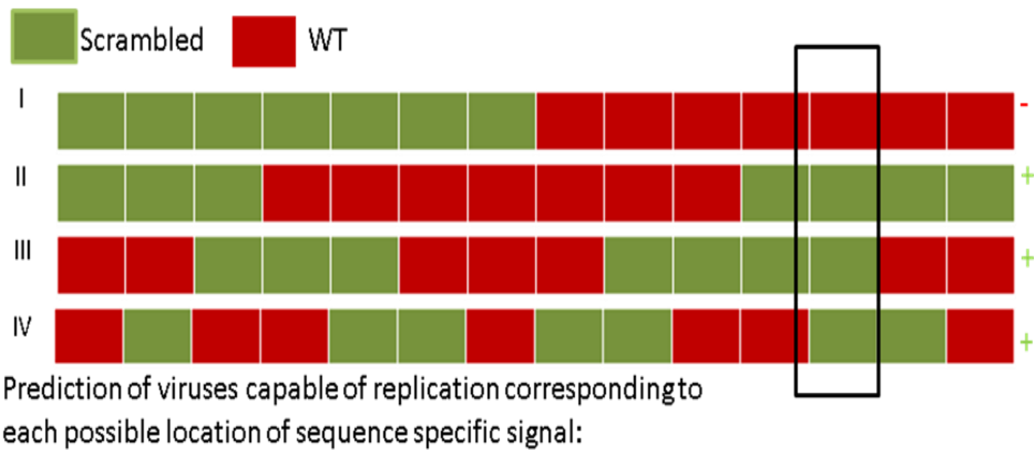


Fig 13: Localization of sequence specific signal II : 13a) Schema of design of Design I,II,III and IV Rep coding sequences . Below each column of segments is the list of viruses capable of replication if the sequence specific inhibitory signal lay within that segment. 13b) Southern blot of DpnI digested Hirt DNA isolated at Day 1 and Day 10 from cells transfected with VIG, VIIG, VIIIG and VIVG, to compare input DNA and ability to replicate of the 4 constructs. Boxed is the 135bp segment and its 5' junction identified by this strategy as being required for Rep mediated inhibition of Ad replication.

13a)



13b)

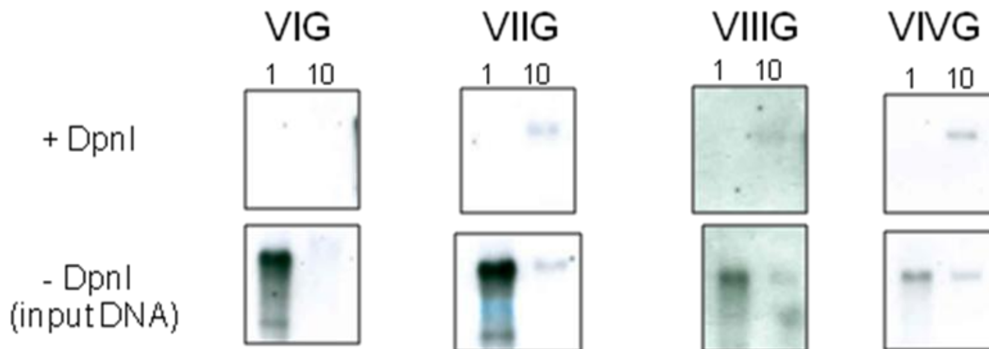


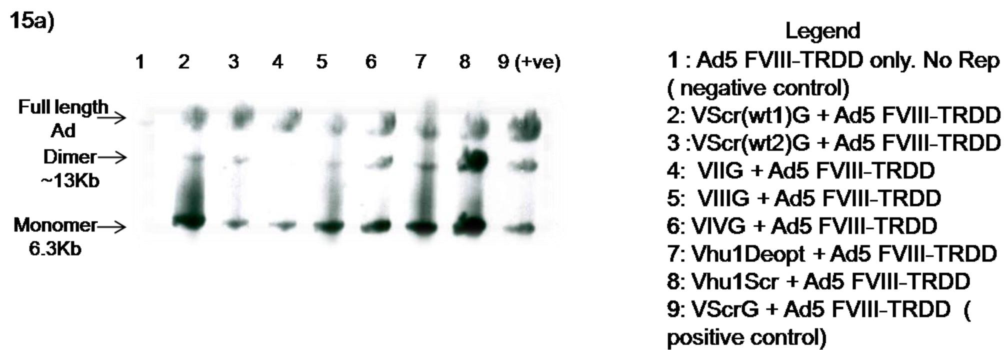
Fig 14: Sequence of the 135bp segment identified as required for Rep mediated inhibition of Ad replication is shown in red. Flanking segments which could possibly carry a part of the Rep78 inhibitory sequence are also shown. Features of interest namely the Rep68/40 splice site and sequences within the AAV p40 promoter are highlighted.

5' bp 1327
GACCGGATGTTCAAATTTGAACTCACCCGCCGTCTGGATCATGACTTT
GGGAAGGTCACCAAGCAGGAAGTCAAAGACTTTTTCCGGTGGGCAAA
GGATCACGTGGTTGAGGTGGAGCATGAATTCTACGTCAAA

AAGGGTGGAGCCAAGAAAAGACCCGCCCCCAGTGACGCAGATATAAG
TGAGCCCAAACGGGTGCGCGAGTCAGTTGCGCAGCCATCGACGTCAG
ACGCGGAAGCTTCGATCAACTACGCAGACAGGTACCAAAAC

AAATGTTCTCGTCACGTGGGCATGAATCTGATGCTGTTTCCCTGCAGA
CAATGCGAGAGAATGAATCAGAATTCAAATATCTGCTTCACTCACGGA
CAGAAAGACTGTTTAGAGTGCTTTCCCGTGTCAGAATCT 3' bp 1731

Fig 15. Characterization of all Ad/AAV carrying Rep, capable of replication. 15a) The ability of all constructs capable of replicating to express functional Rep78 was confirmed by an excision assay in 293 cells, with an Ad5 carrying an AAV TRDD flanked PF4-FVIII cassette (Dr. Patrick Hearing, Stony Brook Univ.) 15b) Infectious yields of Pfu/cell was determined from clarified lysates of secondary infections of 293 cells.



15b)

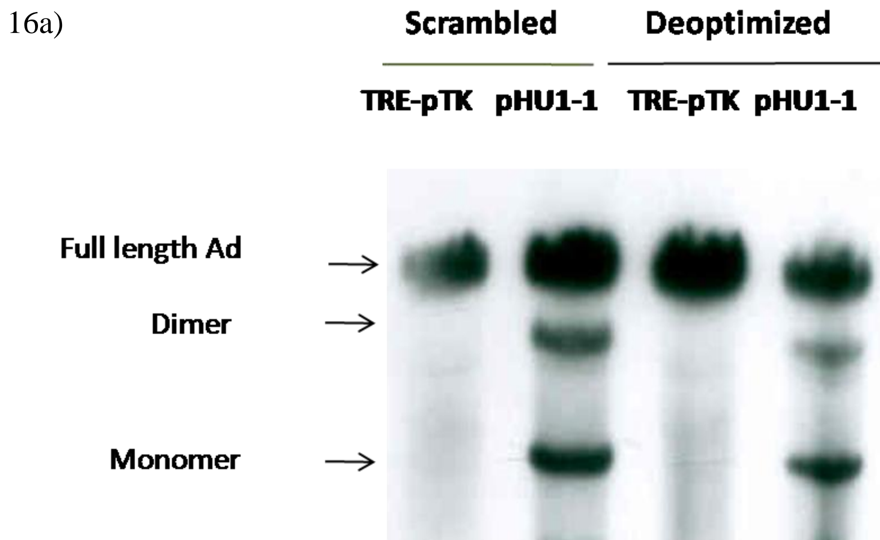
Viral Construct	Pfu/ cell
VPE (Ad/AAV-PF4FVIII)	5.8 x 10^3
VScrG	3.75 x 10^3
VDeoptG	2.73 x 10^3
VScr(wt1)G	1.81 x 10^3
VScr(wt2)G	1.63 x 10^3
VIIG	1.07 x 10^3
VIIIG	5.45 x 10^3
VIVG	3.10x 10^3
Vhu1Scr	1.48 x 10^3
Vhu1Deopt	1.23 x 10^3

The ability of all viruses capable of replication to express functional Rep was confirmed by an excision assay, as described (Fig15). Viral titers obtained from clarified lysates of secondary infections were calculated and were found to be comparable to titers from VScrG. (Fig 15b)

Relative effect of Rep protein expression on inhibition of Ad replication: The fact that VScrG and VDeoptG replicate equally efficiently in spite of VScrG expressing double the amount of Rep protein indicates that the amount of Rep protein expressed played no part in the inability of VRB (Ad/wtRep78) to replicate. However, the tetracycline inducible system has been shown to tightly regulate Rep protein expression in transfected cells and as mentioned earlier, has been previously used for the successful production of a gutless Ad carrying Rep78 (195). Further, the dependence of AAV on relative time of infection and relative copy number to inhibit Ad replication has lead authors to propose that the accumulation of Rep expression at the initial stages of infection is responsible for inhibition of Ad replication (291). Therefore, to truly understand the role of Rep78 protein expression in the inhibition of Adenoviral replication, we expressed the modified Rep ORFs under a constitutive 243bp human U1-1 small nuclear RNA promoter (pHU1-1) (310). The expression of functional Rep from these two promoters was compared by an excision assay in cells transfected with their respective shuttle vectors (Fig 16), to mimic conditions in packaging cells soon after transfection of infectious clones. In the absence of inducer, where excision due to Rep expression from the Tet system is undetectable with both pTRE-Scr and pTRE-Deopt, robust excision can be seen with expression of the Scr and Deopt ORFs under the pHU1-1 promoter, with dimer and monomer sized excision products clearly visible. In spite of expressing far higher levels of Rep78 than the tetracycline inducible system, ΔE1ΔE3 Adenoviruses carrying the hu1-Scr and hu1-Deopt Rep constructs were still capable of normal rates of replication and resulted in only ~2 fold decrease in titers (Fig 16). These results prove that a high level of Rep78 protein expression can be tolerated by replicating Adenoviruses and the dramatic inhibitory effects seen are mainly due to signals within the sequence of the Rep ORF.

Fig 16: Contribution of Rep protein expression to inhibition of Ad replication: 16a) Comparison of expression levels from Rep from the constitutive hu1 promoter vs the tetracycline inducible promoter. 293 cells were transfected with the respective shuttle vectors and infected with Ad/AAV VPE (Ad/AAV-PF4FVIII) 24 hours later. Hirt DNA was isolated 48 hours later and analyzed by southern blot for the presence of excision products. Tightly controlled expression from the TRE-pTK promoter results in no detectable excision product in the absence of dox where robust expression and excision can be detected from the pHU1-1 promoter. 16b) In spite of much higher levels of Rep protein expression, these constructs do not inhibit replication of the Ad carrying them, as shown by comparison of yield (pfu/cell). Approximately 2-fold difference in titer yields is observed.

16a)



16b)

Viral construct	Pfu/ cell
VScrG (TRE-Tk promoter)	3.75×10^3
Vhu1Scr	1.48×10^3
VDeoptG (TRE-TK promoter)	2.73×10^3
Vhu1Deopt	1.23×10^3

Discussion

Multiple authors have attempted to construct Ad/AAV hybrid viruses carrying an AAV Rep expression cassette (Ad/AAV-Rep) on the backbone of an Adenovirus, both for the production of an integrating virus with the advantages of Ad as well as for the production of an infectious virus carrying all factors necessary for efficient rAAV production (195, 272, 291, 292, 294, 322). AAV co-infection has been shown to inhibit Ad replication, and that inhibition can be mimicked by the expression of Rep protein alone (218, 222, 224, 225). As a result, most strategies for production of an Ad/AAV carrying Rep have focused on controlling expression of Rep protein. However, though successes have been reported, many attempts have been stymied due to lack of replication or instability of the virus (291, 292, 323). Work done in this dissertation identifies a sequence specific inhibitory signal present within the Rep ORF, responsible for the instability of Ad/AAV-Rep in the presence of low levels of Rep expression.

An examination of the literature reveals a higher number of reports of successful construction of helper dependent Ad carrying Rep vs first generation Adenovirus carrying Rep. Further, work done in this dissertation revealed an absence of viability of an Ad carrying Rep expressed under a tetracycline inducible promoter, which had previously been used successfully for the production of a helper dependent Ad. In fact, the regulation of Rep expression from this promoter was tight enough within the helper dependent Ad to enable the production of an Ad carrying both the AAV ITR and the Rep expression cassette, an especially difficult task given Rep's ability to cleave at the AAV ITR (142). Since helper dependent Ads are devoid of all Ad genes and depend on a helper Adenovirus for replication, it is possible that the replicative functions provided by multiple helper virus genomes in *trans* to the helper dependent virus allow replication, whereas a single genome carrying both Adenoviral genes and the Rep expression construct is unable to escape Rep's inhibitory effect. Further, the possibility exists that the surrounding promoter elements within a first generation Ad enhanced leaky expression from the tetracycline inducible promoter. However, the absolute lack of viability observed from these constructs contrasts sharply with robust Adenoviral replication seen in the presence of high levels of Rep expression produced by Rep expressing cell lines such as C12 (290). To identify if the increased inhibition of replication when Rep was carried *on* the Adenoviral moiety was due to increased expression accompanying an increase in copy number, or due to an actual role for the sequence of the Rep ORF, we modified the Rep ORF with synonymous codons, to jumble any sequence specific signal without changing the functional protein expressed. The Deoptimized Rep ORF was designed with a reduced codon pair bias score to analyze the effects of

both reduced Rep expression and modified sequence on the inhibitory effect of Rep.

The Scrambled and Deoptimized Rep sequences have codon pair bias scores of -0.117 and -.443 respectively, compared to wtRep78 with a score of -.043. The approximately 2-fold difference in scores between wtRep78 and Scr Rep 78 did not seem to affect expression levels from the two ORFs appreciably (Fig 10). However, the 10-fold difference in codon pair bias scores between Deopt and wtRep was reflected in an ~50% decrease in protein expression from the Deopt ORF, compared to wtRep78 or Scr Rep78.

When expressed under the tetracycline inducible promoter within an Adenoviral backbone, the difference between Scr, Deopt and wtRep ORFs to support Ad replication was stark. While VRB (Ad/wtRep78) showed no signs of replication in spite of 5 passages in 293 cells over 50 days from transfection, cells transfected with VDeoptG (Ad/dRep78) and VScrG (Ad/sRep78) reached CPE within 15 days, comparable to any other first generation Ad. The ability of VScrG to replicate in spite of expressing the same amount of Rep protein as VRB (Ad/wtRep78), clearly proved that the sequence of the Rep ORF plays a role in the inhibition of Ad replication. The ability of VScrG to replicate equivalent to the control first generation Ad vector carrying coagulation FVIII (VPE (Ad/AAV-PF4FVIII)) and VDeoptG underscored the utility of the tetracycline inducible system in controlling Rep expression and indicated that under the control afforded by the tetracycline inducible system, the major role in Adenoviral inhibition was played by this sequence specific signal. The comparable ability to replicate in terms of time from transfection to reach CPE (VDeoptG – 14 days, VScrG-15 days, VPE (Ad/AAV-PF4FVIII)-15 days) was confirmed by quantification of increase in vector genomes over time by a modified DpnI assay as well as by determination of Pfu/cell values (Fig 11; Fig 15b). Interestingly, the DpnI viral replication assay indicated that while no signs of viral replication in cell culture were obtained from VRB (Ad/wtRep78), a slow increase in replicated viral genomes with this construct was occurring in the first 13 days from transfection.

The expression of these modified Rep78 cassettes under a constitutive promoter phu1-1 gave unexpected results. phu1-1 is a small nuclear RNA promoter, previously used successfully for the expression of coagulation FVIII from a rAAV vector, in our lab (310). The promoter was chosen as a mid-level promoter, capable of expressing at far higher levels than the minimal inducible tetracycline inducible system (Fig 17), but less than well characterized strong promoters such as pCMV which could possibly overwhelm the cellular environment with Rep78 expression. Both viruses carrying either Scr or Deopt

Rep under the hu1-1 promoters were capable of replication and reached CPE within 15 days of transfection. Determination of infectious virus yield in terms of Pfu/cell from secondary infections, revealed only a twofold reduction in viral yield compared to those obtained from VScrG and VDeoptG respectively. This yield was not outside the range of titers obtained from various constructs expressing Rep under the tetracycline inducible system (Fig. 16b). As Rep78 protein expression has been reported to be associated with a 10%-40% decrease in Ad replication, these results seem to indicate that with modification of the sequence specific signal, a high level of Rep78 expression can be tolerated for Adenovirus replication.

<u>Localization of the sequence specific signal</u>: To narrow down the location of the sequence specific signal, two different approaches were taken.

In the first classical approach, the entire Rep ORF was divided roughly into three sections based on the availability of internal restriction enzyme sites, and the presence of the sequence specific signal within each of these sections was analyzed. For this, each of these sections from wtRep78 was swapped into the Scr Rep78 ORF, in place of the corresponding Scr sequences, in frame. Each of the resulting Scr/wtRep78 chimeras was still capable of expressing functional Rep78 protein. The ability of these constructs (Scr(wt1), Scr(wt2) and Scr(wt3)) to inhibit Ad replication under the tetracycline inducible promoter was analyzed (Fig 13). VScr(wt1)G and VScr(wt2)G were capable of replication and yielded titers comparable to VScrG. In contrast, VScr(wt3)G did not show any signs of replication for 50 days post transfection, indicating that the sequence specific signal lay completely within the 575bps of wtRep78 represented in that swap. Interestingly, in the process of VScr(wt3)G construction, a mutant infectious clone which was incapable of expressing functional Rep78 due to missense mutations introduced by PCR error was constructed, by chance. This construct was found to have 4 point mutations within the 'wt3' sequence compared to the corresponding sequence from wtRep78. 1 of the 4 mutations was silent. 3 mutations resulted in the change of $Leucine_{441}$ to Phenylalanine, $Tryptophan_{472}$ to Leucine; and $Glycine_{598}$ to Valine. When tested in a functional excision assay this construct was found to be incapable of excision (Data not shown) but capable of replication, with cells reaching CPE 20 days post transfection. Since the 4 base pair changes were not clustered, they were considered unlikely to be capable of eliminating the sequence of the inhibitory signal, allowing the virus to replicate. The ability of this construct to grow was therefore ascribed to the lack of Rep expression, indicating that expression of Rep protein is essential for complete inhibition of Ad replication. This supposition is supported by reports by Carlson et.al, and Ueno et.al, where first generation Ads carrying only partial Rep ORFs

or complete ORFs incapable of Rep expression due to sequestering of the promoter, were capable of replication (291, 292).

The ability of pHU1-1-Rep constructs to support Ad replication along with the apparent requirement for Rep expression for inhibition seems to indicate that the expression of Rep protein and maintenance of an inhibitory sequence within the Rep ORF are both essential components of Rep mediated Adenoviral inhibition. High protein expression or high copy numbers of the sequence specific signal individually do not seem to inhibit Ad replication in the absence of the other, further indicating these components inhibit Ad replication via the same mechanism. This dependence on both the sequence of the Rep ORF and expression from the Rep cassette immediately suggests two plausible mechanisms of action. An siRNA encoded by the Rep ORF would be dependent on both the coding sequence and expression from the ORF. However, the spacing out of mutations within the Rep- VScr(wt3)G mutant seems to indicate that the mechanism of action is likely not via a non coding RNA. An alternative scenario is the sequestering of proteins required for Ad replication by the sequence specific signal coupled with simultaneous inhibition of the protein by Rep protein binding either to its promoter or by direct protein-protein interaction.

The localization of the inhibitory signal within the 3' 575 bp of the wtRep78 sequence (the 'wt3' fragment) was confirmed with opposite swaps which carried 2/3rd of the wtRep78 sequence and only 1/3rd of the Scr Rep sequence, using the same restriction enzyme sites (Scr(wt1,2), Scr(wt1,3) and Scr(wt2,3)) . Both VScr(wt1,3) and VScr(wt2,3) were incapable of replication for upto 50 days post transfection. However, VScr(wt1,2) which has ~1200bps of wtRep78 sequence but not the 3' 575bps, was capable of replication with cells reaching CPE 15 days post infection. These results indicate that the sequence specific signal is completely localized within the 3' end of the wtRep78 sequence and that no inhibitory effect is associated with the 5' 1200 bps of wtRep78. Further, since the p19 promoter and the translation start site of Rep52 lies within the 5' 600bps of the Rep ORF, this paves the way for the construction of Ad/AAV expressing both Rep 78 and Rep52, required for rAAV production. However, since the splice site for Rep68 and Rep40 lies within the 3' 575bps, the Scr(wt1,2) construct will not be capable of producing those proteins.

The alternate approach to localize this signal was developed in the laboratory of Dr. Steve Skiena, Stony Brook University. As described earlier, the approach generates 4 shuffled Scr/wtRep78 chimera sequences (Design I,II III and IV), each made up of a specific pattern of 14 135-150bp segments (Fig 14). The sequence of each of these segments is the corresponding sequence from either

the Scr ORF or the wtRep ORF. The presence of the sequence specific signal in any one segment would mean that any Scr/wtRep78 chimera sequence which has wtRep78 sequences in that segment will inhibit Ad replication. The design of the four Rep coding sequences is such that the presence of the sequence specific signal within a particular segment would result in a unique pattern of constructs which allow or inhibit Ad replication. Each of these four chimera sequences were expressed under the tetracycline inducible promoter within a first generation Ad backbone, generating VIG, VIIG, VIIIG and VIVG. The inability of VIG to replicate, coupled with the ability of VIIG, VIIIG and VIVG to replicate as well as VScrG, localizes the inhibitory sequence to a 135 bp region extending from bp 1461-1596 (aa 487-532) of wtRep78. The possibility exists that the location of the sequence specific signal is not completely within the boundaries of bp 1461-1596, and extends across the junctions of the neighboring 5' segments, as those junctions are also only maintained in VIG. The wtRep78 bp1461-1596 sequence is modified by 51 mutations in the corresponding sRep78 sequence and 30 mutations in the corresponding dRep78 sequence. 25 bases in the wtRep78 sequence are modified in both sRep78 and dRep78. The codon pair bias scores for the 135bp segment alone are reflective of the entire Rep78 coding sequences, with the wtRep78 bp1461-1596 = -0.0524, sRep78 bp 1461-1596 = -0.0556 and dRep78 bp 1461-1596 = -0.681.

The 135bp sequence specific inhibitory segment also overlaps part of the p40 promoter, the transcription start site for which lies at bp1528 of the Rep78 ORF. McCarty et.al, identified three sites – GGTGGAG (GGT site) at -70 and -99 and GACCCGCCCC at -50 upstream of the transcription start site (311). Binding of Sp1 to GGT-70 and GACCCGCCC -50 have been found to be required for Rep mediated induction of the p40 promoter (324). GGT-99 was not found to play a role in Rep mediated induction (324). Both proven Sp1 binding sites lie within the 135bp fragment (bp 1461-1596) identified as required for Rep mediated inhibition of Adenoviral replication, while the upstream GGT -99 lies within the adjacent 5' segment. The apparent requirement for Rep protein expression for inhibition of Adenoviral replication raises the possibility that this previously shown interaction of the Rep DNA sequence with Rep protein via Sp1 mediated protein-protein interaction could also play a role in the inhibition of Adenoviral replication. Comparison of bp 1326-1569 from sRep78 and wtRep78 sequences reveals a modified 5bp predicted Sp1 site retained at bp1480 of sRep78, corresponding to the start site of GACCCCGCCC within wtRep78. However, the ability of this modified sequence to still bind Sp1 will need to be studied.

The 135bp sequence specific Rep78 inhibitory segment includes the splice site required for Rep40 and Rep68 production and Rep 68 encoding sequences would not include sequences 3' to the splice site. Since the Rep68 ORF alone has also been shown to inhibit Ad replication (292), it can be assumed with confidence that the sequence specific signal lies within or at the 5' junction of the segment encompassing bp1461-1596. Confirmation of the presence of the sequence specific signal completely within bp 1461-1596 is needed. The ability of a Scr/wtRep78 chimera with only bp 1461-1596 of the Scr sequence replaced by that of wtRep, to inhibit Ad replication will confirm these results.

The inhibitory effect of this sequence is mainly observable when carried on a first generation Ad, suggesting a *cis*-acting mechanism. However, it is possible that the inhibitory effect is due to an increase in copy number of the Rep ORF. When the Rep expression cassette is carried on the Ad backbone, every additional Adenovirus genome produced by replication is accompanied by an additional inhibitory element. As a result the virus is unable to outgrow the inhibitory effects of the Rep sequence. This would contrast with a helper dependent Ad, where initially multiple copies of helper virus genomes would exist for every copy of the helper dependent virus, allowing viral replication. Further, this hypothesis would predict that at higher titers, as the ratio of helper Ad: helper dependent Ad decreases, the effect of the inhibitory element would be observable. It is tempting to speculate that this mechanism of action, which merely appears to be *cis* acting, might also play a role in AAV–Ad interaction in nature, as authors have found that inhibition of Ad by AAV requires an AAV infection of the cell prior to Ad superinfection or at an increased AAV:Ad MOI.

Future directions

I. Analysis of putative sequence specific inhibitory signal: Multiple plausible mechanisms for the inhibitory action of this sequence can be proposed. Though some possibilities seem more likely than others, due to the novelty of this signal, no possibilities can be ruled out without further investigation. Modification of the DNA sequence could result in the loss of protein binding sites contained within the DNA sequence, loss of DNA secondary structure such as tandem repeats, loss of RNA secondary structure or loss of function of any non-coding regulatory RNA such as siRNA encoded.

Protein binding sites could be involved in the binding and sequestering of factors required for Ad replication including transcription factors. Concomitant with Ad replication, an increase in Rep ORF would result in an increase in the

inhibitory signal. This would likely result in a low rate of replication at early stages of replication which would cease as copy number increases and sequestering reduces available transcription factor below a threshold required for Ad replication. The requirement for Rep expression could mean that Rep protein interaction is required for sequestering of the protein as Rep has recently been shown to interact with 188 cellular proteins, including proteins involved in transcription, translation, DNA replication, DNA repair and splicing (325). Alternatively, Rep protein could act independently to reduce levels of the sequestered protein, resulting in a combined reduction in available factor below the threshold required for Ad replication. For instance, Rep protein has been shown to bind to the Ad Major Late Promoter and E2A promoter reducing their transcription (221, 222, 224). A preliminary analysis of putative transcription factor binding sites within the 135bp DNA sequence was performed using the Transcription Element Search Software (TESS) available at http://www.cbil.upenn.edu/cgi-bin/tess/tess (326). 57 putative binding sites for transcription factors were identified. Comparison of TFBDs (Transcription Factor Binding Domains) present between bp1461-1596 of wtRep78 and Scr Rep78, reduced the list to 38 binding sites which were present in wtRep78 but had been lost due to the shuffling of nucleotides in Scrambled. Only 31 of those sites were absent in Deopt as well (Table 1). Interestingly, 6 putative binding sites for the cyclic AMP Responsive Element Binding protein (CREB), delta CREB and transcription factors that interact with CREB (AP-1 and CREB) were identified within wtRep78, which were destroyed in the Scr and Deopt Rep sequences. The cAMP pathway is crucial for the replication of several viruses including Ad and HSV(327-330) Further, a site within the C terminal of the Rep78 protein has been identified which interacts with cAMP dependent Protein Kinase A (PKA) and its homolog PRKX, inhibits their kinase activity preventing phosphorylation of CREB, thereby inhibiting activation of CREB dependent pathways. Abolishment of this site has been associated with loss of inhibition of Ad replication by AAV. Identification of other cellular or Adenoviral binding partners will need to be done empirically. Electrophoretic Mobility Shift Assays which will confirm the protein binding properties of the 135bp putative sequence specific signal will be a rational starting point.

The formation of DNA secondary structure has been shown to be associated with stalled DNA polymerase and reduced replication efficiency (331). However, the ability of AAV carrying Rep to replicate under the same cellular conditions indicates that a secondary structure formed within Rep is unlikely. However, though unlikely, a possibility exists for the formation of DNA secondary structure with Adenoviral sequences flanking the Rep expression cassette. Extended nucleotide repeats have been shown to form folded RNA

hairpins which bind to and sequester cellular proteins and play a causative role in diseases such as Myotonic dystrophy (332). No DNA sequence repeats were identified within the putative 135bp sequence specific inhibitory signal (http://emboss.sourceforge.net).

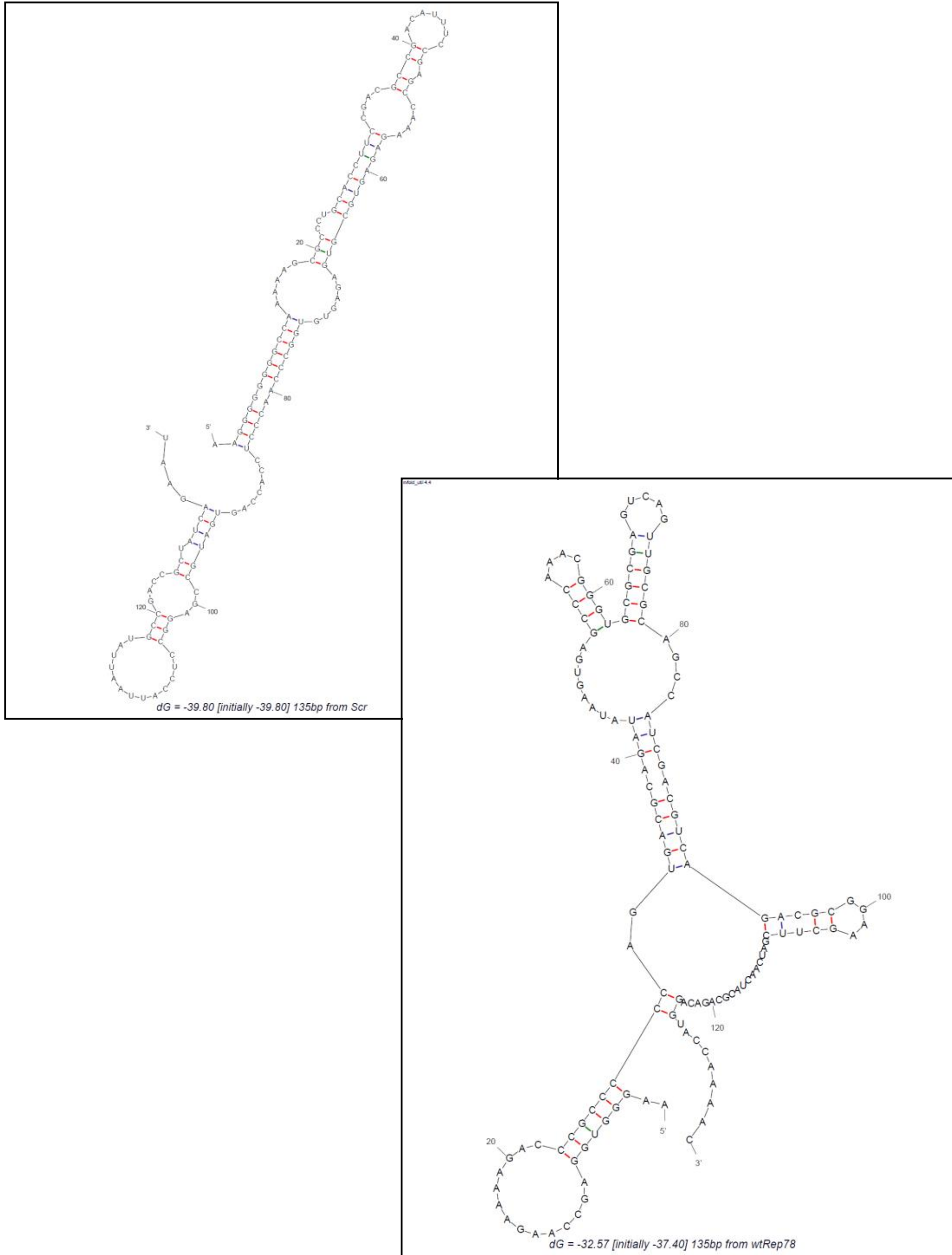
Analysis of RNA secondary structure using RNA mfold (http://mfold.bioinfo.rpi.edu) (333, 334) revealed a stable secondary structure prediction for the 135bp putative sequence specific signal from wtRep78 with a ΔG of -32.5Kcal/mol (Fig 17a). However, analysis of the equivalent 135bp (bp 1461-1596) from the Scr nucleotide sequence revealed a predicted secondary structure with a lower ΔG of -39.8Kcal/mol (Fig 18b). Though the possibility of a small non-coding RNA including siRNA or miRNA moieties encoded by the sequence offers an elegant explanation for a sequence specific inhibitory effect, results with mutated VScr(wt3) point away from non-coding RNA. As a latent, DNA virus, AAV is an ideal virus to encode miRNA for regulation. However, the AAV genome has been analyzed for the presence of miRNA coding sequences multiple times with no success. Currently, computational prediction of siRNA coding sequences is not available. An indirect method to analyze if a non coding RNA is responsible for Rep mediated Adenoviral inhibition would be construct an Adenovirus carrying both the 3' 575bp of wtRep78 inserted randomly as well as an expression cassette expressing functional Rep protein from a sequence modified ORF. The resultant hybrid Ad would be inhibited for replication if protein expression and the sequence specific effects are two separate contributors to the inhibitory effect. A non-coding RNA would require transcription from the wtRep78 sequence and would not be produced in such a vector.

Table 1: Putative transcription factor binding sites unique to wtRep78 bp1461-1596 identified by TESS (Transcription Element Search System) (326)

	Transcription Factor	Beg bp no.	Sequence
1	T00111 c-Ets-1 T00112 c-Ets-1 T00114 c-Ets-1 54 T00115 c-Ets-1 68 T00684 PEA3 T00685 PEA3 T00686 PEA3	97	SMGGAWGY
2	T00506 MEF1 T00519 Myf-3 T00524 MyoD T00525 MyoD T00526 MyoD T00527 MyoD T01128 MyoD	69	GTCAGTTG
3	_00000 ASF-1 _00000 MSN4 _00000 deltaCREB	88	ACGTCA
4	T00049 ATF T00050 atf1 T00132 c-Jun T00163 CREB T00164 CREB T00166 deltaCREB T00167 CRE-BP1 T00846 TREB-1 T00942 EivF T01095 ATF3	88	ACGTCA
5	T00163 CREB	33	TGACG
6	T00051 ATF T00052 ATF-a T00053 ATF-adelta T00054 ATF-like T00442 47-kDa CRE bind. prot. T00968 ATF-1	33	TGACGYMR
7	_00000 LRF-1	88	ACGTCA
8	_00000 ATF-1	88	ACGTCA
9	T00134 c-Jun T00893 v-Jun	66	CGAGTCAG
10	T00074 gammaCAC1 T00075 gammaCAC2 T00077 CACCC-binding factor	3	GGGTG
11	_00000 MIG1	27	CCCCAG
12	T00765 SRF (504 AA)	41	ATATA
13	_00000 RC2	18	AAGACC
14	_00000 GCN4	67	GAGTCA
15	_00000 B-factor	42	TATAAGT
16	T00386 HSTF	14	AGAAA

17	T00029 AP-1	90	GTCA
18	_00000 HBP-1	88	ACGTCA
19	T00140 c-Myc	71	CAGTTG
20	T00029 AP-1 T00123 c-Fos T00133 c-Jun T00167 CRE-BP1 T00989 CREB T01313 ATF3 T02361 CREBbeta	33	TGACGCA
21	T00182 DBF4 T00270 ETF T00530 NC1 T00794 TBP T00798 TBP T00817 TFIIA T00818 TFIIB T00820 TFIID T00835 TMF T00862 UBP-1 T02216 TFIIA-alpha/beta precursor (majorT02216 T02217 TFIIA-alpha/beta precursor (minorT02217 T02224 TFIIA-gamma	42	TATAA
22	T00074 gammaCAC1 T00075 gammaCAC2 T00077 CACCC-binding factor	59	GGGTG
23	T00321 GCN4	67	GAGTCA
24	T00029 AP-1	69	GTCA
25	T00029 AP-1	23	CCGCCCCC
26	T00878 USF2 T02115 USF2 T02377 USF2b	33	TGACGCA
27	T00422 IRF1 T00425 IRF-2	45	AAGTGA
28	T00968 ATF-1	88	ACGTCA
29	_00000 TREB-1	88	ACGTCA
30	T00354 HBP-1 T00938 HBP-1b T01393 HBP-1b(c1) T01394 HBP-1a(1) T01395 HBP-1a(c14) T02789 bZIP910	88	ACGTCA
31	T00051 ATF T00052 ATF-a T00053 ATF-adelta T00054 ATF-like T00442 47-kDa CRE bind. prot. T00968 ATF-1	86	YKRCGTCA
32	T00111 c-Ets-1 T00112 c-Ets-1 T00114 c-Ets-1 54 T00115 c-Ets-1 68 T00684 PEA3 T00685 PEA3 T00686 PEA3	97	SMGGAWGY
33	T00506 MEF1 T00519 Myf-3 T00524 MyoD T00525 MyoD T00526 MyoD T00527 MyoD T01128 MyoD	69	GTCAGTTG

Fig 17: Secondary structure predictions of mRNA encoded by bp 1461-1596 of Scr Rep78 and wtRep78 by mfold (333, 334)



II. The construction of a stable Adenovirus carrying a Rep expression cassette paves the way for production of integrating Adenovirus for gene delivery as well as helper virus carrying all replicative functions for efficient rAAV production.

Integrating Adenovirus: As described earlier, site specific integration into the AAVS1 sites *in vitro* has been shown to require an AAV ITR flanked transgene and the Rep protein in *trans* (335). The use of Ad/AAV as an integrating vector has been hampered by difficulties producing Ad expressing Rep. Successful production of Ad/AAV carrying Rep has also been associated with a reduced efficiency of site specific integration.The identification of the sequence specific inhibitory signal and the ability of VScrG and VDeoptG to be grown to high titers without loss in Rep expression, paves the way for development of a stable vector. Co-infection of VScrG with an Ad/AAV carrying the transgene flanked by the AAV TR or IEE, (such as VPE (Ad/AAV-PF4FVIII)) should enable analysis of these viruses to mediate integration. Factors to take into consideration will be:

a) Arrangement of AAV element for the transgene cassette: VPE (Ad/AAV-PF4FVIII) carries a coagulation FVIII expression cassette, immediately downstream of the Ad left ITR and ψ. Downstream of the transgene cassette are 2 copies of the integration efficiency element and the AAV TR. The reasoning behind this arrangement of AAV elements is that studies on the relative ability of the IEE and AAV TR to mediate integration indicate that the TRs do serve as integration boundaries and therefore Rep binding sites more easily than IEE, but IEE's increase efficiency of integration when used with the TR (183, 336, 337). Cleavage of the construct at the AAV TR should release a fragment carrying the transgene and the Ad left ITR, which can be tethered to the AAVS1 site by Rep at any of the 3 RBEs present. As cleavage at the TRDD would result in one RBE remaining with the Rep expression cassette and the right end Ad genome, integration of the Rep cassette should be monitored.
b) As both VScrG and VPE (Ad/AAV-PF4FVIII) are Ad F5/35, with the knob domain of the Ad5 replaced with that of Ad35, they are capable of infecting hematopoietic cells. The toxic effects of Rep in hematopoietic cells will be of concern. Since very low levels of Rep expression are associated with integration of wtAAV, leaky expression from the Tetracycline inducible promoter, in the absence of inducer, might prove sufficient for efficient integration as well as reduced cellular toxicity due to Rep protein expression.

A more elegant construct would carry both the transgene and the Rep expression cassette on the same backbone. However, leaky expression of Rep is sufficient to cleave at the AAV TRs during replication of the Ad backbone. This excision might result in instability of the viral genome, as vectors which have lost the AAV ITR would accumulate faster and be selected for.

<u>Ad/AAV for rAAV production</u>: An adenovirus carrying AAV Rep and Cap expression cassettes (Ad/RepCap) would be an efficient vector, carrying all factors necessary for rAAV replication. Currently, co-transfection with three plasmids is still used in Vector Cores for the production of rAAV, though the use of hybrid HSV and Bac systems are gaining popularity. An Adenovirus would be advantageous over HSV and Bac in their ease of production and their ability to be grown to high titers. As in integration, the introduction of an ITR flanked transgene cassette into 293 cells infected with a first generation Ad/RepCap would be sufficient for efficient high titer production of rAAV. Introduction of the rAAV sequence could be by co-infection with another Ad/AAV carrying it, or by production of a single virus carrying both the transgene cassette and the RepCap cassette. Alternatively, Ad/RepCap could be used to infect producer cell lines which have rAAV sequences integrated. The most advantageous but difficult of these strategies would be the production of a virus expressing Rep and carrying the AAV TR, as it would form a basic system that can easily manipulated to produce high titer rAAV carrying any transgene. However, as mentioned earlier, the Rep expression cassette and AAV TR on the same moiety might result in instability and loss of the TR due to excision. Replication in a Rep knockdown packaging cell line might enable increased stability.

Current VScrG and VDeoptG vectors only encode Rep78. These Rep ORFs would need to also encode Rep52/40 for rAAV production. Since the inhibitory signal lies only at the 3' 575 bp of the Rep ORF, it is possible to produce a Rep ORF carrying the p19 promoter and the ATG for Rep52. However, as the inhibitory signal overlaps with the splice site, the other two Rep proteins (Rep68 and Rep40) are incapable of being made at this point. This construct VScr(wt1,2)G replicates as well as VScrG and shows no signs of inhibition in spite of carrying 2/3rds of the ORF of wtRep78. The ability of VScr(wt1,2)G to express Rep52 has not been confirmed. Another factor to be taken into consideration would be relative production of Rep 78 vs Rep52 as studies have shown that a higher ratio of Rep52:Rep78 is required for efficient production (257). Finally, to reduce contamination due to helper virus, an Ad with the packaging signal flanked by lox P sites and an E1 complementing cell line such as 293 Cre can be considered. These systems are well characterized and have been used successfully for the production of helper dependent Ad, with very low levels

of contamination. Further, since rAAV is relatively heat resistant compared to Ad, heat treatment of viral preps should reduce helper virus contamination even further.

In conclusion, the identification of a sequence specific inhibitory signal for AAV Rep mediated inhibition of Ad replication explains the inconsistent and often frustrating results obtained with production of Ad/AAV carrying Rep over the years. Characterization of the mechanism of action of the signal might shed light on the variation observed in reports on the effect of Rep on the Ad carrying it. Finally, the ability to produce a high titer stable Ad carrying Rep by modification of this inhibitory sequence paves the way for large scale production of integrating Ad/AAV and ease of production of rAAV for gene therapy.

References

1. Bordignon C, *et al.* (1995) Gene Therapy in Peripheral Blood Lymphocytes and Bone Marrow for ADA- Immunodeficient Patients. *Science* 270(5235):470-475.
2. Aiuti A, *et al.* (2009) Gene Therapy for Immunodeficiency Due to Adenosine Deaminase Deficiency. *N Engl J Med* 360(5):447-458.
3. Aiuti A, *et al.* (2002) Correction of ADA-SCID by Stem Cell Gene Therapy Combined with Nonmyeloablative Conditioning. *Science* 296(5577):2410-2413.
4. Raper SE, *et al.* (2003) Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer. *Mol Genet Metab* 80(1-2):148-158.
5. Raper SE, *et al.* (2002) A pilot study of in vivo liver-directed gene transfer with an adenoviral vector in partial ornithine transcarbamylase deficiency. *Hum Gene Ther* 13:163-175.
6. Hacein-Bey-Abina S, *et al.* (2008) Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J Clin Invest* 118(9):3132-3142.
7. Hacein-Bey-Abina S, Von Kalle C, Schmidt M, McCormack M, & Wulffraat N (2003) LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* 302:415.
8. Howe SJ, *et al.* (2008) Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *J Clin Invest* 118(9):3143-3150.
9. Jacobson SG, *et al.* (2006) Safety of recombinant adeno-associated virus type 2-RPE65 vector delivered by ocular subretinal injection. *Mol Ther* 13(6):1074-1084.
10. Maguire AM, *et al.* (2009) Age-dependent effects of RPE65 gene therapy for Leber's congenital amaurosis: a phase 1 dose-escalation trial. *Lancet* 374(9701):1597-1605.
11. Bainbridge JWB, *et al.* (2008) Effect of Gene Therapy on Visual Function in Leber's Congenital Amaurosis. *N Engl J Med* 358(21):2231-2239.
12. Cartier N, *et al.* (2009) Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. *Science* 326(5954):818-823.
13. Nabel GJ, *et al.* (1996) Immune response in human melanoma after transfer of an allogeneic class I major histocompatibility complex gene with DNA-liposome complexes. *Proc Natl Acad Sci U S A* 93(26):15388-15393.
14. Rini BI, Selk LM, & Vogelzang NJ (1999) Phase I study of direct intralesional gene transfer of HLA-B7 into metastatic renal carcinoma lesions. *Clin Cancer Res* 5(10):2766-2772.
15. Galanis E, *et al.* (1999) Immunotherapy of advanced malignancy by direct gene transfer of an interleukin-2 DNA/DMRIE/DOPE lipid complex: phase I/II experience. *J Clin Oncol* 17(10):3313-3323.

16. Daud AI*, et al.* (2008) Phase I trial of interleukin-12 plasmid electroporation in patients with metastatic melanoma. *J Clin Oncol* 26(36):5896-5903.
17. Brasnjevic I, Steinbusch HWM, Schmitz C, & Martinez-Martinez P (2009) Delivery of peptide and protein drugs over the blood-brain barrier. *Progress in Neurobiology* 87(4):212-251.
18. Donehower LA & Varmus HE (1984) A mutant murine leukemia virus with a single missense codon in pol is defective in a function affecting integration. *Proc Natl Acad Sci U S A* 81(20):6461-6465.
19. Wu X, Li Y, Crise B, & Burgess SM (2003) Transcription start regions in the human genome are favored targets for MLV integration. *Science* 300(5626):1749-1751.
20. Schroder AR*, et al.* (2002) HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* 110(4):521-529.
21. Mitchell RS*, et al.* (2004) Retroviral DNA Integration: ASLV, HIV, and MLV Show Distinct Target Site Preferences. *PLoS Biol* 2(8):e234.
22. Samuelson LC, Wiebauer K, Snow CM, & Meisler MH (1990) Retroviral and pseudogene insertion sites reveal the lineage of human salivary and pancreatic amylase genes from a single gene during primate evolution. *Mol Cell Biol* 10(6):2513-2520.
23. Ting CN, Rosenberg MP, Snow CM, Samuelson LC, & Meisler MH (1992) Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes Dev* 6(8):1457-1465.
24. Hacein-Bey-Abina S*, et al.* (2002) Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy. *N Engl J Med* 346(16):1185-1193.
25. Gaspar HB*, et al.* (2004) Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector. *Lancet* 364(9452):2181-2187.
26. Galy A, Roncarolo M-G, & Thrasher AJ (2008) Development of lentiviral gene therapy for Wiskott Aldrich syndrome. *Expert Opinion on Biological Therapy* 8(2):181-190.
27. Watanabe D (2010) Medical application of herpes simplex virus. *Journal of Dermatological Science* 57(2):75-82.
28. D.M. K*, et al.* (1998) Development of herpes simplex virus replication-defective multigene vectors for combination gene therapy applications. *Gene Ther.* 5:1517.
29. A.R. M, M. UF, L. R, J. H, & S.M. B (1991) Herpes simplex virus type 1 deletion variants 1714 and 1716 pinpoint neurovirulence-related sequences in Glasgow strain 17+ between immediate early gene 1 and the ‘a’ sequence. *J. Gen. Virol.* 72:631.
30. R.H. L & S. E (1997) Utilization of the herpes simplex virus type 1 latency-associated regulatory region to drive stable reporter gene expression in the nervous system. *J. Virol.* 71:3197.
31. D. W*, et al.* (1999) Engineering herpes simplex virus vectors for CNS applications. *Exp. Neurol.* 159:34.

32. D.M. K*, et al.* (1998) Deletion of multiple immediate-early genes from herpes simplex virus reduces cytotoxicity and permits long-term gene expression in neurons. *Gene Ther.* 5:1593.
33. G.R. A*, et al.* (1999) Herpes simplex virus vector-mediated dystrophin gene transfer and expression in MDX mouse skeletal muscle. *J. Gene Med.* 1:280.
34. R.R. S & N. F (1982) The herpes simplex virus amplicon: a new eucaryotic defective-virus cloning-amplifying vector. *Cell* 30:295.
35. Saeki Y, Fraefel C, Ichikawa T, Breakefield XO, & Chiocca EA (2001) Improved Helper Virus-Free Packaging System for HSV Amplicon Vectors Using an ICP27-Deleted, Oversized HSV-1 DNA in a Bacterial Artificial Chromosome. *Mol Ther* 3(4):591-601.
36. Brooks AI, Cory-Slechta DA, Bowers WJ, Murg SL, & Federoff HJ (2000) Enhanced learning in mice parallels vector-mediated nerve growth factor expression in hippocampus. *Hum Gene Ther* 11(17):2341-2352.
37. Harvey BK*, et al.* (2003) HSV amplicon delivery of glial cell line-derived neurotrophic factor is neuroprotective against ischemic injury. *Exp Neurol* 183(1):47-55.
38. Costantini LC*, et al.* (1999) Gene transfer to the nigrostriatal system by hybrid herpes simplex virus/adeno-associated virus amplicon vectors. *Hum Gene Ther* 10(15):2481-2494.
39. Lim F*, et al.* (2007) Functional Recovery in a Friedreich's Ataxia Mouse Model by Frataxin Gene Transfer Using an HSV-1 Amplicon Vector. *Mol Ther* 15(6):1072-1078.
40. Santos K, Duke CM, & Dewhurst S (2006) Amplicons as vaccine vectors. *Curr Gene Ther* 6(3):383-392.
41. Kemeny N*, et al.* (2006) Phase I, open-label, dose-escalating study of a genetically engineered herpes simplex virus, NV1020, in subjects with metastatic colorectal carcinoma to the liver. *Hum Gene Ther* 17(12):1214-1224.
42. MacKie RM, Stewart B, & Brown SM (2001) Intralesional injection of herpes simplex virus 1716 in metastatic melanoma. *Lancet* 357(9255):525-526.
43. Hu JC*, et al.* (2006) A phase I study of OncoVEXGM-CSF, a second-generation oncolytic herpes simplex virus expressing granulocyte macrophage colony-stimulating factor. *Clin Cancer Res* 12(22):6737-6747.
44. Markert JM*, et al.* (2000) Conditionally replicating herpes simplex virus mutant, G207 for the treatment of malignant glioma: results of a phase I trial. *Gene Ther* 7(10):867-874.
45. Rowe WP, Huebner RJ, Gilmore LK, Parrott RH, & Ward TG (1953) Isolation of a cytopathogenic agent from human adenoids undergoing spontaneous degeneration in tissue culture. *Proc Soc Exp Biol Med* 84(3):570-573.
46. McAllister RM, Nicolson MO, Lewis AM, Jr., Macpherson I, & Huebner RJ (1969) Transformation of rat embryo cells by adenovirus type 1. *J Gen Virol* 4(1):29-36.
47. McAllister RM*, et al.* (1969) Transformation of rodent cells by adenovirus 19 and other group D adenoviruses. *J Natl Cancer Inst* 43(4):917-923.

48. Norrby E (1969) The Structural and Functional Diversity of Adenovirus Capsid Components. *Journal of General Virology* 5:221-236.
49. Wickham TJ, Mathias P, Cheresh DA, & Nemerow GR (1993) Integrins [alpha]v[beta]3 and [alpha]v[beta]5 promote adenovirus internalization but not virus attachment. *Cell* 73(2):309-319.
50. Roberts RJ, O'Neill KE, & Yen CT (1984) DNA Sequences from the Adenovirus 2 Genome. *The Journal of Biological Chemistry* 259(22):13968-13975.
51. Chroboczek J, Bieber F, & Jacrot B (1992) The sequence of the genome of adenovirus type 5 and its comparison with the genome of adenovirus type 2. *Virology* 186(1):280-285.
52. Hearing P, Samulski RJ, Wishart WL, & Shenk T (1987) Identification of a repeated sequence element required for efficient encapsidation of the adenovirus type 5 chromosome *Journal of Virology* 61(8):2555-2558.
53. Grable M & Hearing P (1992) *cis* and *trans* requirements for the Selective Packaging of the Adenovirus Type 5 DNA. *Journal of Virology* 66(2):723-731.
54. Bachenheimer S & Darnell JE (1975) Adenovirus-2 mRNA is transcribed as part of a high-molecular-weight precursor RNA. *Proceedings of the National Academy of Sciences of the United States of America* 72(11):4445-4449.
55. Chow LT, Gelinas RE, Broker TR, & Roberts RJ (1977) An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* 12(1):1-8.
56. Dery CV, Herrmann CH, & Mathews MB (1987) Response of individual adenovirus promoters to the products of the E1A gene. *Oncogene* 2(1):15-23.
57. Gaggar A, Shayakhmetov DM, & Lieber A (2003) CD46 is a cellular receptor for group B Aadenoviruses. *Nature Medicine* 9:1408-1412.
58. Bergelson J*, et al.* (1997) Isolation of a common receptor for Coxsackie B viruses and adenoviruses 2 and 5. *Science* 275(5304):1320-1323.
59. Nakano MY, Boucke K, Suomalainen M, Stidwill RP, & Greber UF (2000) The First Step of Adenovirus Type 2 Disassembly Occurs at the Cell Surface, Independently of Endocytosis and Escape to the Cytosol. *J. Virol.* 74(15):7085-7095.
60. Albinsson B & Kidd AH (1999) Adenovirus type 41 lacks an RGD [alpha]v-integrin binding motif on the penton base and undergoes delayed uptake in A549 cells. *Virus Research* 64(2):125-136.
61. Greber UF*, et al.* (1997) The role of the nuclear pore complex in adenovirus DNA entry. *EMBO J* 16(19):5998-6007.
62. Bandara LR & Thangue NBL (1991) Adenovirus E1a prevents the retinoblastoma gene product from complexing with a cellular transcription factor. *Nature* 351(494-497).
63. Bagchi S, Raychaudhuri P, & Nevins JR (1990) Adenovirus E1A proteins can dissociate heteromeric complexes involving the E2F transcription factor: A novel mechanism for E1A trans-activation. *Cell* 62(4):659-669.
64. Bagchi S, Weinmann R, & Raychaudhuri P (1991) The retinoblastoma protein copurifies with E2F-I, an E1A-regulated inhibitor of the transcription factor E2F. *Cell* 65(6):1063-1072.

65. Montell C, Courtois G, Eng C, & Berk A (1984) Complete transformation by adenovirus 2 requires both E1A proteins. *Cell* 36(4):951-961.
66. Moran E, Grodzicker T, Roberts RJ, Mathews MB, & Zerler B (1986) Lytic and transforming functions of individual products of the adenovirus E1A gene. *J. Virol.* 57(3):765-775.
67. Debbas M & White E (1993) Wild-type p53 mediates apoptosis by Ela, which is inhibited by Elb. *Genes Dev.* 7:546-554.
68. White E (1998) Regulation of Apoptosis by Adenovirus E1A and E1B Oncogenes. *Seminars in Virology* 8(6):505-513.
69. Benedict CA*, et al.* (2001) Three adenovirus E3 proteins cooperate to evade apoptosis by tumor necrosis factor-related apoptosis-inducing ligand receptor-1 and -2. *J. Biol. Chem.* 276:3270.
70. Tollefson AE*, et al.* (1996) The adenovirus death protein (E3-11.6K) is required at very late stages of infection for efficient cell lysis and release of adenovirus from infected cells. *J. Virol.* 70:2296.
71. Tollefson AE, Ryerse JS, Scaria A, Hermiston TW, & Wold WS (1996) The E3-11.6-kDa adenovirus death protein (ADP) is required for efficient cell death: Characterization of cells infected with adp mutants. *Virology* 220:152.
72. Täuber B & Dobner T (2001) Molecular regulation and biological function of adenovirus early genes: the E4 ORFs. *Gene* 278(1-2):1-23.
73. Shtrichman R, Sharf R, Barr H, Dobner T, & Kleinberger T (1999) Induction of apoptosis by adenovirus E4orf4 protein is specific to transformed cells and requires an interaction with protein phosphatase 2A. *Proc. Natl. Acad. Sci. U.S.A.* 96:10080.
74. Huang MM & Hearing P (1989) Adenovirus early region 4 encodes two gene products with redundant effects in lytic infection. *J. Virol.* 63(6):2605-2615.
75. Evans JD & Hearing P (2003) Distinct roles of the adenovirus E4 ORF3 protein in viral DNA replication and inhibition of genome concatenation. *J. Virol.* 77:5295.
76. Schaley J, O'Connor RJ, Taylor LJ, Bar-Sagi D, & Hearing P (2000) Induction of the cellular E2F-1 promoter by the adenovirus E4-6/7 protein. *J. Virol.* 74:2084.
77. O'Connor RJ & Hearing P (2000) The E4-6/7 Protein Functionally Compensates for the Loss of E1A Expression in Adenovirus Infection. *J. Virol.* 74(13):5819-5824.
78. King AJ & van der Vliet PC (1994) A precursor terminal protein-trinucleotide intermediate during initiaition of adenovirus DNA replication: regeneration of molecular ends *in vitro* by a jumping back mechansim. *EMBO J.* 13(23):5786-5792.
79. Lechner RL & Kelly TJ (1977) The structure of replicating adenovirus 2 DNA molecules. *Cell* 12(4):1007-1020.
80. Hay RT, Stow ND, & McDougall IM (1984) Replication of adenovirus mini-chromosomes. *Journal of Molecular Biology* 175(4):493-510.
81. Zhang W & Imperiale MJ (2000) Interaction of the adenovirus IVa2 protein with viral packaging sequences. *J. Virol.* 74:2687.

82. Zhang W & Imperiale MJ (2003) Requirement of the adenovirus IVa2 protein for virus assembly. *J. Virol.* 77:3586.
83. Zhang W, Low JA, Christensen JB, & Imperiale MJ (2001) Role for the adenovirus IVa2 protein in packaging of viral DNA. *J. Virol.* 75:10446.
84. Edvardsson B, Everitt E, Jornvall H, Prage L, & Philipson L (1976) Intermediates in adenovirus assembly. *J. Virol.* 19:533.
85. McConnell MJ & Imperiale MJ (2004) Biology of Adenovirus and Its Use as a Vector for Gene Therapy. *Human Gene Therapy* 15(11):1022-1033.
86. Bergelson JM*, et al.* (1997) Isolation of a common receptor for Coxsackie B viruses and adenoviruses 2 and 5. *Science* 275(5304):1320-1323.
87. Harari OA*, et al.* (1999) Targeting an adenoviral gene vector to cytokine-activated vascular endothelium via E-selectin. *Gene Ther* 6(5):801-807.
88. Segerman A*, et al.* (2003) Adenovirus type 11 uses CD46 as a cellular receptor. *J. Virol.* 77:9183.
89. Shayakhmetov DM, Thalia Papayannopoulou, Stamatoyannopoulos G, & Lieber A (2000) Efficient Gene Transfer into Human $CD34^+$ cells by a Retargeted Adenoviral Vector. *Journal of Virology* 74(6):2567 - 2473.
90. Bett AJ, Prevec L, & Graham FL (1993) Packaging capacity and stability of human adenovirus type 5 vectors. *J. Virol.* 67:5911.
91. Graham FL, Smiley J, Russell WC, & Nairn R (1977) Characterization of a human cell line transformed by DNA from human adenovirus type 5. *J. Gen. Virol.* 36:59-72.
92. Fallaux FJ*, et al.* (1998) New helper cells and matched early region 1-deleted adenovirus vectors prevent generation of replication-competent adenoviruses. *Hum. Gene Ther.* 9:1909.
93. Haj-Ahmad Y & Graham FL (1986) Development of a helper-independent human adenovirus vector and its use in the transfer of the herpes simplex virus thymidine kinase gene. *J. Virol.* 57(1):267-274.
94. Chinnadurai G, Chinnadurai S, & Brusca J (1979) Physical mapping of a large-plaque mutation of adenovirus type 2. *Journal of Virology* 32:623.
95. Chartier C*, et al.* (1996) Efficient Generation of Recombinant Adenovirus Vectors by Homologous Recombination in *Escherichia coli*. *Journal of Virology* 70(7):4805 - 4810.
96. Kaplan JM*, et al.* (1996) Humoral and cellular immune responses of nonhuman primates to long-term repeated lung exposure to Ad2/CFTR-2. (Translated from eng) *Gene Ther* 3(2):117-127 (in eng).
97. Yang Y*, et al.* (1994) Cellular immunity to viral antigens limits E1-deleted adenoviruses for gene therapy. *Proceedings of the National Academy of Science U.S.A.* 91:4407.
98. Amalfitano A*, et al.* (1998) Production and characterization of improved adenovirus vectors with the E1, E2b, and E3 genes deleted. *J. Virol.* 72:926.
99. Gorziglia MI*, et al.* (1996) Elimination of both E1 and E2 from adenovirus vectors further improves prospects for in vivo human gene therapy. *Journal of Virology* 70:4173.

100. Lusky M*, et al.* (1998) In vitro and in vivo biology of recombinant adenovirus vectors with E1, E1/E2A, or E1/E4 deleted. *J. Virol.* 72:2022.
101. Yang Y*, et al.* (1994) Inactivation of E2a in recombinant adenoviruses improves the prospect for gene therapy in cystic fibrosis. *Nat. Genet.* 7:362.
102. Amalfitano A & Chamberlain JS (1997) Isolation and characterization of packaging cell lines that coexpress the adenovirus E1, DNA polymerase, and preterminal proteins: Implications for gene therapy. *Gene Ther.* 4:258.
103. Amalfitano A, Begy CR, & Chamberlain JS (1996) Improved adenovirus packaging cell lines to support the growth of replication-defective gene-delivery vectors. *Proc. Natl. Acad. Sci. U.S.A.* 93:3352.
104. Hu H, Serra D, & Amalfitano A (1999) Persistence of an [E1-, polymerase-] adenovirus vector despite transduction of a neoantigen into immune-competent mice. *Hum. Gene Ther.* 10:355.
105. Grave L*, et al.* (2000) Differential influence of the E4 adenoviral genes on viral and cellular promoters. *J. Gene Med.* 2:433.
106. Armentano D*, et al.* (1997) Effect of the E4 region on the persistence of transgene expression from adenovirus vectors. *J. Virol.* 71:2408.
107. Brough DE*, et al.* (1997) Activation of transgene expression by early region 4 is responsible for a high level of persistent transgene expression from adenovirus vectors in vivo. *J. Virol.* 71:9206.
108. Parks RJ & Graham FL (1997) A helper-dependent system for adenovirus vector production helps define a lower limit for efficient DNA packaging. *J. Virol.* 71:3293.
109. Sandig V*, et al.* (2000) Optimization of the helper-dependent adenovirus system for production and potency invivo. *Proceedings of the National Academy of Sciences of the United States of America* 97(3):1002-1007.
110. Sargent KL, Ng P, Evelegh C, Graham FL, & Parks RJ (2004) Development of a size-restricted pIX-deleted helper virus for amplification of helper-dependent adenovirus vectors. *Gene Ther* 11(6):504-511.
111. Parks RJ*, et al.* (1996) A helper-dependent adenovirus vector system: Removal of helper virus by Cre-mediated excision of the viral packaging signal. *Proceedings of the National Academy of Sciences of the United States of America* 93(24):13565-13570.
112. Hardy S, Kitamura M, Harris-Stansil T, Dai Y, & Phipps M (1997) Construction of adenovirus vectors through Cre-lox recombination. *J. Virol.* 71(3):1842-1849.
113. Ng P, Evelegh C, Cummings D, & Graham FL (2002) Cre Levels Limit Packaging Signal Excision Efficiency in the Cre/loxP Helper-Dependent Adenoviral Vector System. *J. Virol.* 76(9):4181-4189.
114. Morsy MA*, et al.* (1998) An adenoviral vector deleted for all viral coding sequences results in enhanced safety and extended expression of a leptin transgene. *Proc. Natl. Acad. Sci. U.S.A.* 95:7866.
115. Schiedner G*, et al.* (1998) Genomic DNA transfer with a highcapacity adenovirus vector results in improved in vivo gene expression and decreased toxicity. *Nat. Genet.* 18:180.

116. Muruve DA, *et al.* (2004) Helper-dependent adenovirus vectors elicit intact innate but attenuated adaptive host immune responses in vivo. *J. Virol.* 78:5966.
117. Liu Q & Muruve DA (2003) Molecular basis of the inflammatory response to adenovirus vectors. *Gene Ther.* 10:935.
118. Raper SE, *et al.* (Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer. *Molecular Genetics and Metabolism* 80(1-2):148-158.
119. Croyle MA, Chirmule N, Zhang Y, & Wilson JM (2002) PEGylation of E1-deleted adenovirus vectors allows significant gene expression on readministration to liver. *Hum. Gene Ther.* 13:1887.
120. Mack CA, *et al.* (1997) Circumvention of anti-adenovirus neutralizing immunity by administration of an adenoviral vector of an alternate serotype. *Hum. Gene Ther.* 8:99.
121. Morral N, *et al.* (1999) Administration of helper-dependent adenoviral vectors and sequential delivery of different vector serotype for long-term liver-directed gene transfer in baboons. *Proc. Natl. Acad. Sci. U.S.A.* 96:12816.
122. Heise C (1997) Onyx-015, an E1b gene-attenuated adenovirus, causes tumor - specific cytolysis and antitumoral efficacy that can be augmented by standard chemotherapeutic agents. *Nature Med.* 3:639-645.
123. Rothmann T, Hengstermann A, Whitaker NJ, Scheffner M, & Zur Hausen H (1998) Replication of ONYX-015, a potential anticancer adenovirus, is independent of p53 status in tumor cells. *J. Virol.* 72:9470.
124. Lanuti M, *et al.* (1999) Evaluation of an E1E4-Deleted Adenovirus Expressing the Herpes Simplex Thymidine Kinase Suicide Gene in Cancer Gene Therapy. *Human Gene Therapy* 10(3):463-475.
125. Kim J-H, *et al.* (2006) Relaxin Expression From Tumor-Targeting Adenoviruses and Its Intratumoral Spread, Apoptosis Induction, and Efficacy. *J. Natl. Cancer Inst.* 98(20):1482-1493.
126. Guedan S, *et al.* (2010) Hyaluronidase Expression by an Oncolytic Adenovirus Enhances Its Intratumoral Spread and Suppresses Tumor Growth. *Mol Ther.*
127. Magalhaes I, *et al.* (2008) rBCG induces strong antigen-specific T cell responses in rhesus macaques in a prime-boost setting with an adenovirus 35 tuberculosis vaccine vector. *PLoS ONE* 3:e3790.
128. Shott JP, *et al.* (2008) Adenovirus 5 and 35 vectors expressing Plasmodium falciparum circumsporozoite surface protein elicit potent antigen-specific cellular IFN-[gamma] and antibody responses in mice. *Vaccine* 26:2818-2823.
129. Zhou D, Cun A, Li Y, Xiang Z, & Ertl HC (2006) A chimpanzee-origin adenovirus vector expressing the rabies virus glycoprotein as an oral vaccine against inhalation infection with rabies virus. *Mol Ther* 14:662-672.
130. Hoelscher MA, *et al.* (2008) A broadly protective vaccine against globally dispersed clade 1 and clade 2 H5N1 influenza viruses. *J Infect Dis* 197:1185-1188.
131. Resende DM, *et al.* (2008) Epitope mapping and protective immunity elicited by adenovirus expressing the Leishmania amastigote specific A2 antigen:

correlation with IFN-[gamma] and cytolytic activity by CD8+ T cells. *Vaccine* 26:4585-4593.

132. Buchbinder SP, *et al.* (2008) Efficacy assessment of a cell-mediated immunity HIV-1 vaccine (the Step Study): a double-blind, randomised, placebo-controlled, test-of-concept trial. *Lancet* 372:1881-1893.
133. McElrath MJ, *et al.* (2008) HIV-1 vaccine-induced immunity in the test-of-concept Step Study: a case-cohort analysis. *Lancet* 372:1894-1905.
134. Gao G, *et al.* (2003) Adeno-associated viruses undergo substantial evolution in primates during natural infections. *Proc. Natl. Acad. Sci. U.S.A.* 100:6081.
135. Surosky RT, *et al.* (1997) Adeno-Associated Virus Rep Proteins Target DNA Sequences to a Unique Locus in the Human Genome. *Journal of Virology* 71(10):7951 - 7959.
136. Kotin RM, Linden RM, & Berns KI (1992) Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination. *EMBO J.* 11:5071.
137. Kotin RM (1990) Site-specific integration by adeno-associated virus. *Proc Natl Acad Sci USA* 87:2211-2215.
138. Samulski R, *et al.* (1991) Targeted integration of adeno-associated virus (AAV) into human chromosome 19. *EMBO J* 10:3941 - 3950.
139. Yakobson B, Koch T, & Winocour E (1987) Replication of adeno-associated virus in synchronized cells without the addition of a helper virus. *J. Virol.* 61(4):972-981.
140. Mehrle S, Rohde V, & Schlehofer JR (2004) Evidence of Chromosomal Integration of AAV DNA in Human Testis Tissue. *Virus Genes* 28(1):61-69.
141. Samulski R, Berns K, Tan M, & Muzyczka N (1982) Cloning of adeno-associated virus into pBR322: rescue of intact virus from the recombinant plasmid in human cells. *Proc Natl Acad Sci USA* 79:2077 - 2081.
142. Im DS & Muzyczka N (1992) Partial purification of adeno-associated virus Rep78, Rep68, Rep52, and Rep40 and their biochemical characterization. *J. Virol.* 66:1119-1128.
143. JA C, L Y, B S, & RM K (1995) Determination of adeno-associated virus Rep68 and Rep78 binding sites by random sequence oligonucleotide selection. *J. Virol.* 69:7334.
144. Ryan J, Zolotukhin S, & Muzyczka N (1996) Sequence requirements for binding of Rep68 to the adeno-associated virus terminal repeats. *J. Virol.* 70:1542.
145. Berns KI (1990) Parvovirus Replication. *Microbiological Reviews* 54(3):316 - 331.
146. Weger S, Wistuba A, Grimm D, & Kleinschmidt JA (1997) Control of Adeno-Associated Virus Type 2 Cap Gene Expression : Relative Influence of Helper Virus, Termianl Repeats and Rep Proteins. *J. Virol* 71(11):8437-8447.
147. A.K. C, M.D. H, W.W. H, & K.I. B (1980) Integration of the adeno-associated virus genome into cellular DNA in latently infected human Detroit 6 cells. *J. Virol.* 33:739.
148. Pereira DJ, McCarty DM, & Muzyczka N (1997) The Adeno-Associated Virus (AAV) Rep Protein Acts as both a Repressor and an Activator to Regulate AAV

Transcription during a Productive Infection. *Journal of Virology* 71(2):1079 - 1088.

149. Kotin RM, *et al.* (1990) Site-specific integration by adeno-associated virus. *Proceedings of the National Academy of Science* 87:2211 - 2215.
150. Im D-S & Muzyczka N (1990) The AAV origin binding protein Rep68 is an ATP-dependent site specific endonuclease with DNA helicase activity. *Cell* 61(3):447 - 457.
151. Brister JR & Muzyczka N (2000) Mechanism of Rep-mediated adeno-associated virus origin nicking. *J. Virol.* 74:7762.
152. Daya S & Berns KI (2008) Gene Therapy Using Adeno-Associated Virus Vectors. *Clin. Microbiol. Rev.* 21(4):583-593.
153. Sonntag F, Schmidt K, & Kleinschmidt JA (2010) A viral assembly factor promotes AAV2 capsid formation in the nucleolus. *Proceedings of the National Academy of Sciences* 107(22):10220-10225.
154. Summerford C & Samulski RJ (1998) Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions. *J. Virol.* 72:1438.
155. Chen C-L, *et al.* (2005) Molecular Characterization of Adeno-Associated Viruses Infecting Children. *J. Virol.* 79(23):14781-14792.
156. Opie SR, Warrington KH, Jr., Agbandje-McKenna M, Zolotukhin S, & Muzyczka N (2003) Identification of Amino Acid Residues in the Capsid Proteins of Adeno-Associated Virus Type 2 That Contribute to Heparan Sulfate Proteoglycan Binding. *J. Virol.* 77(12):6995-7006.
157. Qing K, *et al.* (1999) Human fibroblast growth factor receptor 1 is a co-receptor for infection by adeno-associated virus 2. *Nat. Med.* 5:71.
158. Kashiwakura Y, *et al.* (2005) Hepatocyte growth factor receptor is a co-receptor for adeno-associated virus type 2 infection. *J. Virol* 79:609-614.
159. Summerford C, Bartlett JS, & Samulski RJ (1999) αv $\beta 5$ integrin: A co-receptor for adeno-associated virus type 2 infection. *Nat. Med.* 5:78.
160. Asokan A, Hamra JB, Govindaswamy L, Agbandje-Mckenna M, & Samulski R (2006) Adeno-associated virus type 2 contains an integrin $\alpha_5\beta_1$ binding domain essential for viral cell entry. *J. Virol* 80:8961-8969.
161. Sanlioglu S, *et al.* (2000) Endocytosis and nuclear trafficking of adeno-associated virus type 2 are controlled by rac1 and phosphatidylinositol-3 kinase activation. *J Virol* 74:9184-9196.
162. Duan D, *et al.* (1999) Dynamin is required for recombinant adeno-associated virus type 2 infection. *J. Virol* 73(10371-10376).
163. Bartlett J, Wilcher R, & Samulski R (2000) Infectious entry pathway of adeno-associated virus and adeno-associated virus vectors. *J Virol* 74:2777 - 2785.
164. Ding W, Zhang LN, Yeaman C, & Engelhardt JF (2006) rAAV2 traffics through both the late and the recycling endosomes in a dose-dependent fashion. *Mol Ther* 13:671-682.
165. Bartlett JS, Wilcher R, & Samulski RJ (2000) Infectious entry pathways of adeno-associated virus and adeno-associated virus vectors. *J. Virol* 74:2777-2785.

166. Yan Z*, et al.* (2002) Ubiquitination of both adeno-associated virus type 2 and 5 capsid proteins affects the transduction efficiency of recombinant vectors. *J Virol* 76:2043-2053.
167. Girod A*, et al.* (2002) The VP1 capsid protein of adeno-associated virus type 2 is carrying a phospholipase A2 domain required for virus infectivity *J. Gen. Virol.* 83:973-978.
168. Kelkar SA, Pfister KK, Crystal RG, & Leopold PL (2004) Cytoplasmic dynein mediates adenovirus binding to microtubules. *J. Virol.* 78:10122.
169. Seisenberger G*, et al.* (2001) Real-time single-molecule imaging of the infection pathway of an adeno-associated virus. *Science* 294:1929 - 1932.
170. Hirosue S*, et al.* (2007) Effect of inhibition of dynein function and microtubule-altering drugs on AAV2 transduction. *Virology* 367:10-18.
171. Shi Y, Seto E, Chang L-S, & Shenk T (1981) Transcriptional repression by YY1, a human GLI-Krüippel-related protein, and relief of repression by adenovirus E1A protein. *Cell* 67(2):377-388.
172. Prasad KM, Zhou C, & Trempe JP (1997) Characterization of the Rep78/adeno-associated virus complex. *Virology* 229(1):183-192.
173. Linden RM, Ward P, Giraud C, Winocour E, & Berns KI (1996) Site-specific integration by adeno-associated virus. *Proc. Natl. Acad. Sci. U.S.A.* 93:11288.
174. Linden RM, Winocour E, & Berns KI (1996) The recombination signals for adeno-associated virus site-specific integration. *Proc. Natl. Acad. Sci. U.S.A.* 93:7966.
175. Giraud C, Winocour E, & Berns KI (1994) Site-specific integration by adeno-associated virus is directed by a cellular DNA sequence. *Proceedings of the National Academy of Science* 91:10039 - 10043.
176. Tan I, Ng CH, Lim L, & Leung T (2001) Phosphorylation of a Novel Myosin Binding Subunit of Protein Phosphatase 1 Reveals a Conserved Mechanism in the Regulation of Actin Cytoskeleton. *Journal of Biological Chemistry* 276(24):21209-21216.
177. Henckaerts E*, et al.* (2009) Site-specific integration of adeno-associated virus involves partial duplication of the target locus. *Proceedings of the National Academy of Sciences* 106(18):7571-7576.
178. Dutheil N, Shi F, Dupressoir T, & Linden RM (2000) Adeno-associated virus site-specifically integrates into a muscle-specific DNA region. *Proc. Natl. Acad. Sci. U.S.A.* 97:4862.
179. C B, M K, & WW Z (1997) Adeno-associated virus Rep78 protein and terminal repeats enhance integration of DNA sequences into the cellular genome. *J. Virol.* 71:3299.
180. Feng D*, et al.* (2006) A 16 bp Rep Binding Element is Sufficient for Mediating Rep-dependent Integration into AAVS1. *Journal of Molecular Biology*:1 - 8.
181. Kotin RM (1990) Site-specific integration by adeno-associated virus. *Proc. Natl. Acad. Sci. USA* 87:2211-2215.
182. Philpott NJ, Gomos J, Berns KI, & Falck-Pedersen E (2002) A p5 integration efficiency element mediates Rep-dependent integration into AAVS1 at

chromosome 19. *Proceedings of the National Academy of Science* 99(19):12381 - 12386.

183. Philpott N, Gomos J, Berns K, & Falck-Pedersen E (2002) A p5 integration efficiency element mediates Rep-dependent integration into AAVS1 at chromosome 19. *Proc. Natl. Acad. Sci. USA* 99:12381.
184. Samuel M. Young J & Samulski RJ (2001) Adeno-associated virus (AAV) site-specific recombination does not require a Rep dependent origin of replication within the AAV terminal repeat. *Proceedings of the National Academy of Science* 98(24):13525 - 13530.
185. Kotin RM & Berns KI (1989) Organization of adeno-associated virus DNA in latently infected Detroit 6 cells. *Virology* 170:460-467.
186. Cheung A, Hoggan MD, Hauswirth WW, & Berns KI (1980) Integration of the adeno-associated virus genome into cellular DNA in latently infected human Detroit 6 cells. *J. Virol.* 33:739.
187. Hamilton H, Gomos J, Berns KI, & Falck-Pedersen E (2004) Adeno-Associated Virus Site-Specific Integration and AAVS1 Disruption. *Journal of Virology* 78(15):7874 - 7882.
188. Dyall J, Szabo P, & Berns KI (1999) Adeno-associated virus (AAV) site-specific integration: Formatiom of AAV-AAVS1 junctions in an *in vitro* system. *Proceedings of the National Academy of Science* 96(22):12849 - 12854.
189. Song S*, et al.* (2004) DNA-dependent PK inhibits adeno-associated virus DNA integration. *Proc. Natl. Acad. Sci. U.S.A.* 101:2112.
190. Yamamoto N*, et al.* (2007) Adeno-Associated Virus Site-Specific Integration Is Regulated by TRP-185. *J. Virol.* 81(4):1990-2001.
191. Kearns WG (1996) Recombinant adeno-associated virus (AAV-CFTR) vectors do not integrate in a site-specific fashion in an immortalized epithelial cell line. *Gene Ther.* 3:748-755.
192. McCarty DM, Samuel M. Young J, & Samulski RJ (2004) Integration of Adeno-associated Virus (AAV) and Recombinant AAV Vectors. *Annual Review of Genetics* 38:819 - 844.
193. Huser D, Weger S, & Heilbronn R (2002) Kinetics and frequency of adeno-associated virus site-specific integration into human chromosome 19 monitored by quantitative real-time PCR. *J. Virol.* 76:7554.
194. Huser D & Heilbronn R (2003) Adeno-associated virus integrates site-specifically into human chromosome 19 in either orientation and with equal kinetics and frequency. *J. Gen. Virol.* 84:133.
195. Recchia A, Perani L, Sartori D, Olgiati C, & Mavilio F (2004) Site specific integration of functional transgenes into the human genome by Adeno/AAV hybrid vectors. *Molecular Therapy* 10(No. 4):660 - 670.
196. Hu¨ser D, Weger S, & Heilbronn R (2002) Kinetics and Frequency of Adeno-Associated Virus Site-Specific Integration into Human Chromosome 19 Monitored by Quantitative Real-Time PCR. *Journal of Virology* 76(15):7554 - 7559.

197. Surosky R, Urabe M, Godwin S, McQuiston S, & Kurtzman G (1997) Adeno-associated virus Rep proteins target DNA sequences to a unique locus in the human genome. *J. Virol.* 71:7951.
198. Laughlin CA, Jones N, & Carter BJ (1982) Effect of deletions in adenovirus early region 1 genes upon replication of adeno-associated virus. *J. Virol.* 41:868.
199. Xiao X, Li J, & Samulsk RJ (1998) Production of High-Titer Recombinant Adeno-Associated Virus Vectors in the Absence of Helper Adenovirus. *Journal of Virology* 72(3):2224 - 2232.
200. Alazard-Dany N*, et al.* (2009) Definition of Herpes Simplex Virus Type 1 Helper Activities for Adeno-Associated Virus Early Replication Events. *PLoS Pathog* 5(3):e1000340.
201. Weindler FW & Heilbronn R (1991) A subset of herpes simplex virus replication genes provides helper functions for productive adeno-associated virus replication. *J. Virol.* 65(5):2476-2483.
202. Nash K, Chen W, McDonald WF, Zhou X, & Muzyczka N (2007) Purification of Host Cell Enzymes Involved in Adeno-Associated Virus DNA Replication. *J. Virol.* 81(11):5777-5787.
203. Nash K, Chen W, & Muzyczka N (2008) Complete In Vitro Reconstitution of Adeno-Associated Virus DNA Replication Requires the Minichromosome Maintenance Complex Proteins. *J. Virol.* 82(3):1458-1464.
204. Brister J & Muzyczka N (2000) Mechanism of Rep-mediated adeno-associated virus origin nicking. *J Virol* 74:7762 - 7771.
205. Brister JR & Muzyczka N (1999) Rep-Mediated Nicking of the Adeno-Associated Virus Origin Requires Two Biochemical Activities, DNA Helicase Activity and Transesterification. *J. Virol.* 73(11):9325-9336.
206. Im DS & Muzyczka N (1989) Factors that bind to adeno-associated virus terminal repeats. *J. Virol.* 63(7):3095-3104.
207. Li Z, Brister JR, Im D-S, & Muzyczka N (2003) Characterization of the adenoassociated virus Rep protein complex formed on the viral origin of DNA replication. *Virology* 313:364 -376.
208. Snyder R*, et al.* (1993) Features of the adeno-associated virus origin involved in substrate recognition by the viral Rep protein. *J Virol* 67:6096 - 6104.
209. Myers MW & Carter BJ (1980) Assembly of adeno-associated virus. *Virology* 102(1):71-82.
210. Dubielzig R, King JA, Weger S, Kern A, & Kleinschmidt JA (1999) Adeno-Associated Virus Type 2 Protein Interactions: Formation of Pre-Encapsidation Complexes. *J. Virol.* 73(11):8989-8998.
211. King J, Dubielzig R, Grimm D, & Kleinschmidt J (2001) DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids. *EMBO J* 20:3282 - 3291.
212. Wistuba A, Kern A, Weger S, Grimm D, & Kleinschmidt J (1997) Subcellular compartmentalization of adeno-associated virus type 2 assembly. *J Virol* 71:1341 - 1352.
213. Rittner K, Heilbronn R, Kleinschmidt J, & Sczakiel G (1992) Adeno-associated virus type 2-mediated inhibition of human immunodeficiency virus type 1 (HIV-

1) replication: involvement of p78rep/p68rep and the HIV-1 long terminal repeat. *J. Gen. Virol.* 73:2977.

214. NA K, AD S, C L, & PL H (1998) Involvement of protein-DNA interaction in adeno-associated virus Rep78-mediated inhibition of HIV-1. *J. Hum. Virol.* 1:441.
215. Heilbronn R, Burkle A, Stephan S, & Zur Hausen H (1990) The adeno-associated virus rep gene suppresses herpes simplex virus- induced DNA amplification. *J. Virol.* 64:3012.
216. Yang Q, Chen F, Ross J, & Trempe JP (1995) Inhibition of cellular and SV40 DNA replication by the adeno-associated virus Rep proteins. *Virology* 207:246-250.
217. Bantel-Schaal U & zur Hausen H (1988) Adeno-associated viruses inhibit SV40 DNA amplification and replication of herpes simplex virus in SV40-transformed hamster cells. (Translated from eng) *Virology* 164(1):64-74 (in eng).
218. Casto BC, Atchison RW, & Hammon MW (1967) Studies on the Relationship between Adeno-Associated virus Type I (AAV-1) and Adenoviruses I. Replication of AAV-1 in Certain Cell Cultures and Its Effect on Helper Adenovirus. *Virology* 32:52-59.
219. Horer M*, et al.* (1995) Mutational analysis of adeno-associated virus Rep protein-mediated inhibition of heterologous and homologous promoters. *J Virol* 69:5485-5496.
220. Wonderling RS, Kyostio SR, Walker SL, & Owens RA (1997) The Rep68 protein of adeno-associated virus type 2 increases RNA levels from the human cytomegalovirus major immediate early promoter. *Virology* 236(1):167-176.
221. Casper JM, Timpe JM, Dignam JD, & Trempe JP (2005) Identification of an Adeno-Associated Virus Rep PRtein Brinding Site in the Adenovirus E*2a* Promoter. *Journal of Virology* 79(1):28 - 38.
222. Needham PG*, et al.* (2006) Adeno-Associated Virus Rep Protein-Mediated Inhibition of Transcription of the Adenovirus Major Late Promoter In Vitro. *Journal of Virology* 80(13):6207 - 6217.
223. Antoni BA*, et al.* (1991) Adeno-Associated Virus Rep Protein Inhibits Human Immunodeficiency Virus Type 1 Production in Human Cells. *Journal of Virology* 65(1):396 - 495.
224. Nada S & Trempe JP (2002) Characterization of Adeno-Associated Virus Rep Protein Inhibition of Adenovirus E2a Gene Expression. *Virology* 293(2):345-355.
225. Timpe JM, Verrill KC, & Trempe JP (2006) Effects of Adeno-Associated Virus on Adenovirus Replication and Gene Expression during Coinfection. *Journal of Virology* 80(16):7807 - 7815.
226. Glauser DL*, et al.* (2010) Inhibition of herpes simplex virus type 1 replication by adeno-associated virus rep proteins depends on their combined DNA-binding and ATPase/helicase activities. *J Virol* 84(8):3808-3824.
227. Glauser DL*, et al.* (2007) Live covisualization of competing adeno-associated virus and herpes simplex virus type 1 DNA replication: molecular mechanisms of interaction. (Translated from eng) *J Virol* 81(9):4732-4743 (in eng).

228. Chiorini JA, *et al.* (1998) Inhibition of PrKX, a novel protein kinase and the cyclic AMP-dependent protein kinase PKA by the regulatory proteins of adeno-associated virus type 2. *Mol Cell Biol* 18:5921-5929.
229. Di Pasquale G & Stacey SN (1998) Adeno-Associated Virus Rep78 Protein Interacts with Protein Kinase A and Its Homolog PRKX and Inhibits CREB-Dependent Transcriptional Activation. *J. Virol.* 72(10):7916-7925.
230. Schmidt M, Chiorini JA, Afione S, & Kotin R (2002) Adeno-Associated Virus Type 2 Rep78 Inhibition of PKA and PRKX: Fine Mapping and Analysis of Mechanism. *J. Virol.* 76(3):1033-1042.
231. Di Pasquale G & Chiorini JA (2003) PKA/PrKX activity is a modulator of AAV/adenovirus interaction. *EMBO J* 22(7):1716-1724.
232. Hermonat PL (1991) Inhibition of H-ras expression by the adeno-associated virus Rep78 transformation suppressor gene product. *Cancer Res* 51:3373-3377.
233. Klein-Bauernschmitt P, zur Hausen H, & Schlehofer JR (1992) Induction of differentiation-associated changes in established human cells by infection with adeno-associated virus type 2. *J Virol* 66:4191-4200.
234. Khleif SN, Myers T, Carter BJ, & Trempe JP (1991) Inhibition of cellular transformation by the adeno-associated virus rep gene. *Virology* 181:738-741.
235. Winocour E, Callaham MF, & Huberman E (1988) Perturbation of the cell cycle by adeno-associated virus. *Virology* 167:393-399.
236. Kube DM, Ponnazhagan S, & Srivastava A (1997) Encapsidation of adeno-associated virus type 2 Rep proteins in wild- type and recombinant progeny virions: Rep-mediated growth inhibition of primary human cells. *J Virol* 71:7361-7371.
237. Flotte TR (1996) A phase I study of an adeno-associated virus CFTR gene vector in adult CF patients with mild lung disease. *Hum. Gene Ther.* 7:1145-1159.
238. Flotte TR, *et al.* (2003) Phase I trial of intranasal and endobronchial administration of a recombinant adeno-associated virus serotype 2 (rAAV2)-CFTR vector in adult cystic fibrosis patients: a two-part clinical study. *Hum Gene Ther* 14(11):1079-1088.
239. Manno CS, *et al.* (2003) AAV-mediated factor IX gene transfer to skeletal muscle in patients with severe hemophilia B. *Blood* 101(8):2963-2972.
240. Manno CS, *et al.* (2006) Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. *Nat Med* 12(3):342-347.
241. Bennicelli J, *et al.* (2008) Reversal of Blindness in Animal Models of Leber Congenital Amaurosis Using Optimized AAV2-mediated Gene Transfer. *Mol Ther* 16(3):458-465.
242. Brantly ML, *et al.* (2006) Phase I trial of intramuscular injection of a recombinant adeno-associated virus serotype 2 alphal-antitrypsin (AAT) vector in AAT-deficient adults. *Hum Gene Ther* 17(12):1177-1186.
243. Brantly ML, *et al.* (2009) Sustained transgene expression despite T lymphocyte responses in a clinical trial of rAAV1-AAT gene therapy. *Proceedings of the National Academy of Sciences* 106(38):16363-16368.

244. Mingozzi F*, et al.* (2009) AAV-1-mediated gene transfer to skeletal muscle in humans results in dose-dependent activation of capsid-specific T cells. *Blood* 114(10):2077-2086.
245. Yan Z, Zhang Y, Duan D, & Engelhardt JF (2000) Trans-splicing vectors expand the utility of adeno-associated virus for gene therapy. *Proceedings of the National Academy of Sciences of the United States of America* 97(12):6716-6721.
246. Maheshri N, Koerber JT, Kaspar BK, & Schaffer DV (2006) Directed evolution of adeno-associated virus yields enhanced gene delivery vectors. *Nat Biotechnol* 24:198-204.
247. Grimm D*, et al.* (2008) In vitro and in vivo gene therapy vector evolution via multispecies interbreeding and re-targeting of adeno-associated viruses. *J Virol* 82:5887-5911.
248. Li W*, et al.* (2008) Engineering and selection of shuffled AAV genomes: a new strategy for producing targeted biological nanoparticles. *Mol Ther* 16:1252-1260.
249. Stachler MD & Bartlett JS (2006) Mosaic vectors comprised of modified AAV1 capsid proteins for efficient vector purification and targeting to vascular endothelial cells. *Gene Ther* 13(11):926-931.
250. White SJ*, et al.* (2004) Targeted Gene Delivery to Vascular Tissue In Vivo by Tropism-Modified Adeno-Associated Virus Vectors. *Circulation* 109(4):513-519.
251. Bartlett JS, Kleinschmidt J, Boucher RC, & Samulski RJ (1999) Targeted adeno-associated virus vector transduction of nonpermissive cells mediated by a bispecific F(ab'[ggr])2 antibody. *Nat Biotech* 17(2):181-186.
252. Stachler MD, Chen I, Ting AY, & Bartlett JS (2008) Site-specific Modification of AAV Vector Particles With Biophysical Probes and Targeting Ligands Using Biotin Ligase. *Mol Ther* 16(8):1467-1473.
253. Hauck B, Zhao W, High K, & Xiao W (2004) Intracellular Viral Processing, Not Single-Stranded DNA Accumulation, Is Crucial for Recombinant Adeno-Associated Virus Transduction. *J. Virol.* 78(24):13678-13686.
254. Ferrari F, Samulski T, Shenk T, & Samulski R (1996) Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. *J. Virol.* 70(5):3227-3234.
255. McCarty DM, Monahan PE, & Samulski RJ (2001) Self-complementary recombinant adeno-associated virus (scAAV) vectors promote efficient transduction independently of DNA synthesis. *Gene Ther.* 8:1248.
256. McCarty DM (2008) Self-complementary AAV Vectors; Advances and Applications. *Mol Ther* 16(10):1648-1656.
257. Zolotukhin S (2005) Production of Recombinant Adeno-Associated Virus Vectors. *Human Gene Therapy* 16(5):551-557.
258. Samulski RJ, Chang LS, & Shenk TE (1989) Helper-free stocks of recombinant adeno-associated viruses: normal integration does not require viral gene expression. *J. Virol.* 63:3822-3828.
259. Snyder R, Spratt S, Lagarde C, Bohl D, & Kaspar B (1997) Efficient and stable adeno-associated virus-mediated transduction in the skeletal muscle of adult immunocompetent mice. *Hum. Gene Ther.* 8:1891.

260. Salvetti A*, et al.* (1998) Factors influencing recombinant adeno-associated virus production. *Hum. Gene Ther.* 9:695.
261. Grimm D, Kern A, Rittner K, & Kleinschmidt JA (1998) Novel tools for production and purification of recombinant adenoassociated virus vectors. *Hum. Gene Ther.* 9:2745.
262. Mizukami H*, et al.* (2004) Separate control of Rep and Cap expression using mutant and wild-type loxP sequences and improved packaging system for adeno-associated virus vector production. *Mol. Biotechnol.* 27:7.
263. Qiao C, Wang B, Zhu X, Li J, & Xiao X (2002) A novel gene expression control system and its use in stable, high-titer 293 cell-based adeno-associated virus packaging cell lines. *J. Virol.* 76:13015.
264. Qiao C, Li J, Skold A, Zhang X, & Xiao X (2002) Feasibility of generating adeno-associated virus packaging cell lines containing inducible adenovirus helper genes. *J. Virol.* 76:1904.
265. Clark K, Voulgaropoulou F, & Johnson P (1996) A stable cell line carrying adenovirus-inducible rep and cap genes allows for infectivity titration of adeno-associated virus vectors. *Gene Therapy* 3(12):1124 - 1132.
266. Clark KR, Voulgaropoulou F, Fraley DM, & Johnson PR (1995) Cell lines for the production of recombinant adeno-associated virus. *Hum. Gene Ther.* 6:1329-1341.
267. Farson D*, et al.* (2004) Development and characterization of a cell line for largescale, serum-free production of recombinant adeno-associated viral vectors. *J. Gene Med.* 6:1369.
268. Chadeuf G*, et al.* (2000) Efficient recombinant adeno-associated virus production by a stable rep-cap HeLa cell line correlates with adenovirus-induced amplification of the integrated rep-cap genome. *J. Gene Med.* 2:260.
269. Liu X, Voulgaropoulou F, Chen R, Johnson PR, & Clark KR (2000) Selective Rep-Cap gene amplification as a mechanism for high-titer recombinant AAV production from stable cell lines. *Mol. Ther.* 2:394.
270. Oualikene W, Lamoureux L, Weber JM, & Massie B (2000) Protease-deleted adenovirus vectors and complementing cell lines: potential applications of single-round replication mutants for vaccination and gene therapy. (Translated from eng) *Hum Gene Ther* 11(9):1341-1353 (in eng).
271. Merten OW, Geny-Fiamma C, & Douar AM (2005) Current issues in adeno-associated viral vector production. *Gene Ther* 12(S1):S51-S61.
272. Zhang HG*, et al.* (2001) Recombinant adenovirus expressing adeno-associated virus Cap and Rep proteins supports production of high-titer recombinant adeno-associated virus. *Gene Ther.* 8:704.
273. Zhang X & Li CY (2001) Generation of recombinant adenoassociated virus vectors by a complete adenovirus-mediated approach. *Mol. Ther.* 3:787.
274. J.E. C, S. Z, N. M, G.S. H, & B.J. B (1997) Recombinant adeno-associated virus type 2 replication and packaging is entirely supported by a herpes simplex virus type 1 amplicon expressing Rep and Cap. *J. Virol.* 71:8780.

275. Conway JE*, et al.* (1999) High-titer recombinant adeno-associated virus production utilizing a recombinant herpes simplex virus type I vector expressing AAV-2 Rep and Cap. *Gene Ther.* 6:986.
276. Booth MJ, Mistry A, Li X, Thrasher A, & Coffin RS (2004) Transfection-free and scalable recombinant AAV vector production using HSV/AAV hybrids. *Gene Ther.* 11:829.
277. Urabe M, Ding C, & Kotin RM (2002) Insect cells as a factory to produce adeno-associated virus type 2 vectors. *Hum. Gene Ther.* 13:1935.
278. Meghrous J*, et al.* (2005) Production of recombinant adeno-associated viral vectors using a baculovirus/insect cell suspension culture system: From shake flasks to a 20-L bioreactor. *Biotechnol. Prog.* 21:154.
279. Rutledge EA & Russell DW (1997) Adeno-associated virus vector integration junctions. (Translated from eng) *J Virol* 71(11):8429-8436 (in eng).
280. Russell DW, Alexander IE, & Miller AD (1995) DNA synthesis and topoisomerase inhibitors increase transduction by adeno-associated virus vectors. *Proc Natl Acad Sci USA* 92:5719-5723.
281. Nakai H, Iwaki Y, Kay MA, & Couto LB (1999) Isolation of recombinant adeno-associated virus vector-cellular DNA junctions from mouse liver. (Translated from eng) *J Virol* 73(7):5438-5447 (in eng).
282. Nakai H*, et al.* (2005) Large-scale molecular characterization of adeno-associated virus vector integration in mouse liver. *J Virol* 79:3606-3614.
283. Miller DG*, et al.* (2005) Large-scale analysis of adeno-associated virus vector integration sites in normal human cells. *J Virol* 79:11434-11442.
284. Nakai H*, et al.* (2003) AAV serotype 2 vectors preferentially integrate into active genes in mice. (Translated from eng) *Nat Genet* 34(3):297-302 (in eng).
285. Bell P*, et al.* (2006) Analysis of Tumors Arising in Male B6C3F1 Mice with and without AAV Vector Delivery to Liver. *Mol Ther* 14(1):34-44.
286. Donsante A*, et al.* (2001) Observed incidence of tumorigenesis in long-term rodent studies of rAAV vectors. (Translated from eng) *Gene Ther* 8(17):1343-1346 (in eng).
287. Donsante A*, et al.* (2007) AAV vector integration sites in mouse hepatocellular carcinoma. *Science* 317:477.
288. Fisher KJ, Kelley WM, Burda JF, & Wilson JM (1996) A Novel Adenovirus-Adeno-Associated Virus Hybrid Vector That Displays Efficient Rescue and Delivery of the AAV Genome. *Human Gene Therapy* 7:2079 - 2087.
289. Lieber A, Steinwaerder DS, Carlson CA, & Kay MA (1999) Integrating adenovirus-adeno-associated virus hybrid vectors devoid of all viral genes. *J Virol* 73(11):9314-9324.
290. Sandalon Z, Gnatenko D, Bahou WF, & Hearing P (2000) Adeno-associated virus (AAV) Rep protein enhances the generation of a recombinant mini-adenovirus (Ad) utilizing an Ad/AAV hybrid virus. *Journal of Virology* 74(22):10381 - 10389.
291. Carlson CA, Shayakhmetov DM, & Lieber A (2002) An Adenoviral Expression System for AAV Rep78 Using Homologous Recombination. *Molecular Therapy* 6(1):91 - 98.

292. Ueno T*, et al.* (2000) Site-Specific Integration of a Transgene Mediated by a Hybrid Adenovirus/Adeno-Associated Virus Vector Using the Cre/loxP-Expression-Switching System. *Biochemical and Biophysical Research Communications* 273(2):473-478.
293. Recchia A*, et al.* (1999) Site specific integration mediated by a hybrid adenovirus/adeno-associated virus vector. *Proceedings of the National Academy of Science* 96:2615 - 2620.
294. Wang H & Lieber A (2006) A Helper-Dependent Capsid-Modified Adenovirus Vector Expressing Adeno-Associated Virus Rep78 Mediates Site-Specific Integration of a 27-Kilobase Transgene Cassette. *Journal of Virology* 80(23):11699 - 11709.
295. Cortes ML*, et al.* (2007) Targeted Integration of Functional Human ATM cDNA Into Genome Mediated by HSV/AAV Hybrid Amplicon Vector. *Mol Ther* 16(1):81-88.
296. Fraefel C*, et al.* (1997) Gene transfer into hepatocytes mediated by helper virus-free HSV/AAV hybrid vectors. *Mol Med* 3(12):813-825.
297. Johnston KM*, et al.* (1997) HSV/AAV hybrid amplicon vectors extend transgene expression in human glioma cells. *Hum Gene Ther* 8(3):359-370.
298. Wang Y*, et al.* (2002) Herpes Simplex Virus Type 1/Adeno-Associated Virus rep+ Hybrid Amplicon Vector Improves the Stability of Transgene Expression in Human Cells by Site-Specific Integration. *J. Virol.* 76(14):7150-7162.
299. Heister T, Heid I, Ackermann M, & Fraefel C (2002) Herpes Simplex Virus Type 1/Adeno-Associated Virus Hybrid Vectors Mediate Site-Specific Integration at the Adeno-Associated Virus Preintegration Site, AAVS1, on Human Chromosome 19. *J. Virol.* 76(14):7163-7173.
300. Glauser DL, Ackermann M, Saydam O, & Fraefel C (2006) Chimeric herpes simplex virus/adeno-associated virus amplicon vectors. *Curr Gene Ther* 6(3):315-324.
301. Johnston K*, et al.* (1997) HSV/AAV hybrid amplicon vectors extend transgene expression in human glioma cells. *Hum Gene Ther* 8:359 - 370.
302. Heister T, Heid I, Ackermann M, & Fraefel C (2002) Herpes simplex virus type 1//adeno-associated virus hybrid vectors mediate site-specific integration at the adeno-associated virus preintegration site, AAVS1, on human chromosome 19. *J Virol* 76:7163-7173.
303. Oehmig A*, et al.* (2007) Integration of active human [beta]-galactosidase gene (100[thinsp]kb) into genome using HSV//AAV amplicon vector. *Gene Ther* 14(14):1078-1091.
304. Coleman JR*, et al.* (2008) Virus Attenuation by Genome-Scale Changes in Codon Pair Bias. *Science* 320(5884):1784 - 1787.
305. Gutman GA & Hatfield GW (1989) Nonrandom utilization of codon pairs in Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.* 86(10):3699-3703.
306. Buchan JR, Aucott LS, & Stansfield I (2006) tRNA properties help shape codon pair preferences in open reading frames. *Nucleic Acids Res* 34(3):1015-1027.

307. Schiedner G, Hertel S, & Kochanek S (2000) Efficient transformation of primary human amniocytes by E1 functions of Ad5: generation of new cell lines for adenoviral vector production. *Hum Gene Ther* 11(15):2105-2116.
308. Chabra I (2006) Adenovirus-AAV Hybrid Vectors for Gene Therapy. Ph.D (State University of New York at Stony Brook, Stony Brook).
309. Damon AL*, et al.* (2008) Altered bioavailability of platelet-derived factor VIII during thrombocytosis reverses phenotypic efficacy in haemophilic mice. *Thrombosis and Haemostasis* 100(6):1111 - 1122.
310. Gnatenko D*, et al.* (1999) Human factor VIII can be packaged and functionally expressed in an adeno-associated virus background: applicability to haemophilia A gene therapy. *British Journal of Haematology* 104(1):27-36.
311. McCarty DM, Christensen M, & Muzyczka N (1991) Sequences required for Coordinate Induction of Adeno-Associated Virus p19 and p40 Promoters by Rep Protein. *Journal of Virology* 65(6):2936-2945.
312. Zhou X & Muzyczka N (1998) In Vitro Packaging of Adeno-Associated Virus DNA. *Journal of Virology* 72(4):3241-3247.
313. Ranheim TS*, et al.* (1993) Characterization of mutants within the gene for the adenovirus E3 14.7-kilodalton protein which prevents cytolysis by tumor necrosis factor. *J. Virol.* 67(4):2159-2167.
314. Reichel R*, et al.* (1989) The adenovirus E4 gene, in addition to the E1A gene, is important for trans-activation of E2 transcription and for E2F activation. *J. Virol.* 63(9):3643-3650.
315. Hirt B (1967) Selective extraction of polyoma DNA from infected mouse cell cultures. *J. Mol. Biol.* 26:365.
316. Hay RT (1985) The origin of adenovirus DNA replication: minimal DNA sequence requirement in vivo. *EMBO J* 4(2):421-426.
317. Damon AL (2008) Megakaryocyte/platelet-restricted FVIII expression using a platelet factor 4 promoter. PhD (State University of New York at Stonybrook, Stony Brook).
318. Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98(3):503-517.
319. Salucci V*, et al.* (2002) Tight control of gene expression by a helper-dependent adenovirus vector carrying the rtTA2s-M2 tetracycline transactivator and repressor system. *Gene Ther.* 9(21):1415-1421.
320. Martin P & Papayannopoulou T (1982) HEL cells: a new human erythroleukemia cell line with spontaneous and induced globin expression. *Science* 216(4551):1233-1235.
321. Gutman GA & Hatfield GW (1989) Nonrandom utilization of codon pairs in Escherichia coli. *Proc Natl Acad Sci U S A* 86(10):3699-3703.
322. Philpott N, Giraud-Wali C, Dupuis C, Gomos J, & Hamilton H (2002) Efficient integration of recombinant adeno-associated virus DNA vectors requires a p5-rep sequence in cis. *J. Virol.* 76:5411.
323. Fisher K, Kelley W, Burda J, & Wilson J (1996) A novel adenovirus-adeno-associated virus hybrid vector that displays efficient rescue and delivery of the AAV genome. *Hum Gene Ther* 7:2079 - 2087.

324. Pereira D & Muzyczka N (1997) The adeno-associated virus type 2 p40 promoter requires a proximal Sp1 interaction and a p19 CArG-like element to facilitate Rep transactivation. *J. Virol.* 71(6):4300-4309.
325. Nash K, Chen W, Salganik M, & Muzyczka N (2009) Identification of Cellular Proteins That Interact with the Adeno-Associated Virus Rep Proteins. *Journal of Virology* 83(1):454 - 469.
326. Schug J ed (2003) *Using TESS to Predict Transcription Factor Binding Sites in DNA Sequence* (J. Wiley and Sons).
327. Millhouse S, Kenny J, Quinn P, Lee V, & Wigdahl B (1998) ATF/CREB elements in the herpes simplex virus type 1 latency-associated transcript promoter interact with members of the ATF/CREB and AP-1 transcription factor families. *Journal of Biomedical Science* 5(6):451-464.
328. Leza MA & Hearing P (1989) Independent cyclic AMP and E1A induction of adenovirus early region 4 expression. *J. Virol.* 63(7):3057-3064.
329. Suomalainen M, Nakano MY, Boucke K, Keller S, & Greber UF (2001) Adenovirus-activated PKA and p38/MAPK pathways boost microtubule-mediated nuclear targeting of virus. *EMBO J* 20(6):1310-1319.
330. Xia K, Knipe DM, & DeLuca NA (1996) Role of protein kinase a and the serine-rich region of herpes simplex virus type 1 ICP4 in viral replication. (Translated from English) *Journal of Virology* 70(2):1050-1060 (in English).
331. Burrow AA, Marullo A, Holder LR, & Wang Y-H (2010) Secondary structure formation and DNA instability at fragile site FRA16B. *Nucl. Acids Res.* 38(9):2865-2877.
332. Llamusi B & Artero R (2008) Molecular Effects of the CTG Repeats in Mutant Dystrophia Myotonica Protein Kinase Gene. *Curr Genomics* 9(8):509-516.
333. Mathews DH, Sabina J, Zuker M, & Turner DH (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* 288(5):911-940.
334. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucl. Acids Res.* 31(13):3406-3415.
335. McLaughlin SK, Collis P, Hermonat PL, & Muzyczka N (1988) Adeno-associated virus general transduction vectors: analysis of proviral structure. *J. Virol.* 62:1963-1973.
336. Guilbaud M, *et al.* (2008) Relative influence of the adeno-associated virus (AAV) type 2 p5 element for recombinant AAV vector site-specific integration. *J Virol* 82(5):2590-2593.
337. Philpott N, *et al.* (2002) Efficient integration of recombinant adeno-associated virus DNA vectors requires a p5-rep sequence in cis. *J Virol* 76:5411 - 5421.

Appendix

Sequence of AAV wtRep. This sequence encodes Rep78 and Rep52, and part of Rep 68 and Rep40 5' to the splice site. The Rep52 start codon and Rep68/40 splice site are underlined. The 6bp corresponding to the AfeI site and the BstBI site used for sequence swapping is in italics.

ATGCCGGGGTTTTACGAGATTGTGATTAAGGTCCCCAGCGACCTTGACGAGCATCTGC
CCGGCATTTCTGACAGCTTTGTGAACTGGGTGGCCGAGAAGGAATGGGAGTTGCCGC
CAGATTCTGACATGGATCTGAATCTGATTGAGCAGGCACCCCTGACCGTGGCCGAGA
AGCTGCAGCGCGACTTTCTGACGGAATGGCGCCGTGTGAGTAAGGCCCCGGAGGCCC
TTTTCTTTGTGCAATTTGAGAAGGGAGAGAGCTACTTCCACATGCACGTGCTCGTGGA
AACCACCGGGGTGAAATCCATGGTTTTGGGACGTTTCCTGAGTCAGATTCGCGAAAA
ACTGATTCAGAGAATTTACCGCGGGATCGAGCCGACTTTGCCAAACTGGTTCGCGGTC
ACAAAGACCAGAAATGGCGCCGGAGGCGGGAACAAGGTGGTGGATGAGTGCTACAT
CCCCAATTACTTGCTCCCCAAAACCCAGCCTGAGCTCCAGTGGGCGTGGACTAATATG
GAACAGTATTTAAGCGCCTGTTTGAATCTCACGGAGCGTAAACGGTTGGTGGCGCAG
CATCTGACGCACGTGTCGCAGACGCAGGAGCAGAACAAAGAGAATCAGAATCCCAAT
TCTGATGCGCCGGTGATCAGATCAAAAAC*TTCAGC*CAGGTACATGGAGCTGGTCGGGT
GGCTCGTGGACAAGGGGATTACCTCGGAGAAGCAGTGGATCCAGGAGGACCAGGCCT
CATACATCTCCTTCAATGCGGCCTCCAACTCGCGGTCCCAAATCAAGGCTGCCTTGGA
CAATGCGGGAAAGATTATGAGCCTGACTAAAACCGCCCCCGACTACCTGGTGGGCCA
GCAGCCCGTGGAGGACATTTCCAGCAATCGGATTTATAAAATTTTGGAACTAAACGG
GTACGATCCCCAATATGCGGCTTCCGTCTTTCTGGGATGGGCCACGAAAAAGTTCGGC
AAGAGGAACACCATCTGGCTGTTTGGGCCTGCAACTACCGGGAAGACCAACATCGCG
GAGGCCATAGCCCACACTGTGCCCTTCTACGGGTGCGTAAACTGGACCAATGAGAAC
TTTCCCTTCAACGACTGTGTCGACAAGATGGTGATCTGGTGGGAGGAGGGGAAGATG
ACCGCCAAGGTCGTGGAGTCGGCCAAAGCCATTCTCGGAGGAAGCAAGGTGCGCGTG
GACCAGAAATGCAAGTCCTCGGCCCAGATAGACCCGACTCCCGTGATCGTCACCTCC
AACACCAACATGTGCGCCGTGATTGACGGGAACTCAACGACC*TTCGAA*CACCAGCAG
CCGTTGCAAGACCGGATGTTCAAATTTGAACTCACCCGCCGTCTGGATCATGACTTTG
GGAAGGTCACCAAGCAGGAAGTCAAAGACTTTTTCCGGTGGGCAAAGGATCACGTGG
TTGAGGTGGAGCATGAATTCTACGTCAAAAAGGGTGGAGCCAAGAAAAGACCCGCCC
CCAGTGACGCAGATATAAGTGAGCCCAAACGGGTGCGCGAGTCAGTTGCGCAGCCAT
CGACGTCAGACGCGGAAGCTTCGATCAACTACGCAGACAGGTACCAAAACAAATGTT
CTCGTCACGTGGGCATGAATCTGATGCTGTTTCCCTGCAGACAATGCGAGAGAATGA
ATCAGAATTCAAATATCTGCTTCACTCACGGACAGAAAGACTGTTTAGAGTGCTTTCC
CGTGTCAGAATCTCAACCCGTTTCTGTCGTCAAAAAGGCGTATCAGAAACTGTGCTAC
ATTCATCATATCATGGGAAAGGTGCCAGACGCTTGCACTGCCTGCGATCTGGTCAATG
TGGATTTGGATGACTGCATCTTTGAACAATA

Sequence of AAV wtRep78. This sequence encodes only Rep78. The start codon for Rep52 is mutated from ATG to GGA. The splice site for Rep68 and Rep40 is mutated from GGTA to GATA.

ATGCCGGGGTTTTACGAGATTGTGATTAAGGTCCCCAGCGACCTTGACGAGCATCTGC
CCGGCATTTCTGACAGCTTTGTGAACTGGGTGGCCGAGAAGGAATGGGAGTTGCCGC
CAGATTCTGACATGGATCTGAATCTGATTGAGCAGGCACCCCTGACCGTGGCCGAGA
AGCTGCAGCGCGACTTTCTGACGGAATGGCGCCGTGTGAGTAAGGCCCCGGAGGCCC
TTTTCTTTGTGCAATTTGAGAAGGGAGAGAGCTACTTCCACATGCACGTGCTCGTGGA
AACCACCGGGGTGAAATCCATGGTTTTGGGACGTTTCCTGAGTCAGATTCGCGAAAA
ACTGATTCAGAGAATTTACCGCGGGATCGAGCCGACTTTGCCAAACTGGTTCGCGGTC
ACAAAGACCAGAAATGGCGCCGGAGGCGGGAACAAGGTGGTGGATGAGTGCTACAT
CCCCAATTACTTGCTCCCCAAAACCCAGCCTGAGCTCCAGTGGGCGTGGACTAATATG
GAACAGTATTTAAGCGCCTGTTTGAATCTCACGGAGCGTAAACGGTTGGTGGCGCAG
CATCTGACGCACGTGTCGCAGACGCAGGAGCAGAACAAAGAGAATCAGAATCCCAAT
TCTGATGCGCCGGTGATCAGATCAAAAAC*TTCAGC*CAGGTACGGAGAGCTGGTCGGG
TGGCTCGTGGACAAGGGGATTACCTCGGAGAAGCAGTGGATCCAGGAGGACCAGGCC
TCATACATCTCCTTCAATGCGGCCTCCAACTCGCGGTCCCAAATCAAGGCTGCCTTGG
ACAATGCGGGAAAGATTATGAGCCTGACTAAAACCGCCCCCGACTACCTGGTGGGCC
AGCAGCCCGTGGAGGACATTTCCAGCAATCGGATTTATAAAATTTTGGAACTAAACG
GGTACGATCCCCAATATGCGGCTTCCGTCTTTCTGGGATGGGCCACGAAAAAGTTCGG
CAAGAGGAACACCATCTGGCTGTTTGGGCCTGCAACTACCGGGAAGACCAACATCGC
GGAGGCCATAGCCCACACTGTGCCCTTCTACGGGTGCGTAAACTGGACCAATGAGAA
CTTTCCCTTCAACGACTGTGTCGACAAGATGGTGATCTGGTGGGAGGAGGGGAAGAT
GACCGCCAAGGTCGTGGAGTCGGCCAAAGCCATTCTCGGAGGAAGCAAGGTGCGCGT
GGACCAGAAATGCAAGTCCTCGGCCCAGATAGACCCGACTCCCGTGATCGTCACCTC
CAACACCAACATGTGCGCCGTGATTGACGGGAACTCAACGACC*TTCGAA*CACCAGCA
GCCGTTGCAAGACCGGATGTTCAAATTTGAACTCACCCGCCGTCTGGATCATGACTTT
GGGAAGGTCACCAAGCAGGAAGTCAAAGACTTTTTCCGGTGGGCAAAGGATCACGTG
GTTGAGGTGGAGCATGAATTCTACGTCAAAAAGGGTGGAGCCAAGAAAAGACCCGCC
CCCAGTGACGCAGATATAAGTGAGCCCAAACGGGTGCGCGAGTCAGTTGCGCAGCCA
TCGACGTCAGACGCGGAAGCTTCGATCAACTACGCAGACAGATACCAAAACAAATGT
TCTCGTCACGTGGGCATGAATCTGATGCTGTTTCCCTGCAGACAATGCGAGAGAATGA
ATCAGAATTCAAATATCTGCTTCACTCACGGACAGAAAGACTGTTTAGAGTGCTTTCC
CGTGTCAGAATCTCAACCCGTTTCTGTCGTCAAAAAGGCGTATCAGAAACTGTGCTAC
ATTCATCATATCATGGGAAAGGTGCCAGACGCTTGCACTGCCTGCGATCTGGTCAATG
TGGATTTGGATGACTGCATCTTTGAACAATA

Sequence of Scrambled Rep78. The AfeI and BstBI restriction enzyme sites used in swaps for localization are in italics.

ATGCCCGGATTCTACGAAATCGTCATCAAAGTGCCCTCTGACTTGGATGAAC
ACCTGCCGGGGATCAGCGATTCTTTCGTCAATTGGGTCGCGGAGAAAGAGTGGGAAC
TTCCCCCCGACTCGGACATGGACCTGAACTTAATCGAGCAAGCCCCGCTGACGGTGG
CGGAGAAACTGCAGCGGGACTTTCTGACCGAGTGGAGGCGCGTATCGAAAGCGCCCG
AAGCTTTGTTTTTCGTCCAGTTCGAGAAGGGGGAGTCGTACTTTCATATGCATGTGTT
GGTGGAGACTACGGGAGTGAAGAGTATGGTGCTAGGGAGGTTTCTGTCGCAAATAAG
AGAGAAGCTGATCCAGCGGATATACCGTGGCATTGAGCCCACCCTTCCCAATTGGTTT
GCCGTGACCAAAACTCGTAACGGAGCAGGGGGGGGAAATAAAGTCGTCGACGAGTG
CTATATTCCGAACTACCTCTTGCCCAAGACGCAGCCCGAATTGCAGTGGGCCTGGACC
AACATGGAGCAATACCTGTCAGCGTGCCTCAACTTGACCGAAAGAAAGAGACTCGTG
GCCCAGCACCTGACCCATGTCTCACAGACCCAGGAACAGAATAAGGAAAACCAAAAC
CCAAATAGCGACGCCCCCGTGATACGGAGCAAGACC*AGCGCT*CGCTACATGGAGTTA
GTGGGATGGTTGGTGGATAAAGGAATCACGTCTGAGAAACAATGGATTCAGGAGGAC
CAGGCGTCCTACATTAGTTTTAACGCCGCGTCAAATAGCAGATCTCAGATTAAAGCCG
CGCTCGATAACGCCGGCAAAATCATGTCGCTGACCAAGACAGCTCCCGACTACCTGG
TGGGACAGCAGCCGGTGGAGGACATCTCTTCTAACCGTATCTACAAGATCCTtGAGTT
GAATGGCTACGACCCACAGTACGCCGCCTCAGTGTTTCTGGGCTGGGCAACCAAGAA
ATTTGGGAAACGCAATACGATTTGGCTGTTCGGACCCGCCACCACTGGTAAGACTAAT
ATTGCCGAGGCGATCGCACATACCGTTCCGTTTTACGGATGCGTGAATTGGACTAACG
AAAATTTCCCCTTTAATGATTGCGTGGACAAGATGGTTATTTGGTGGGAGGAAGGAA
AGATGACTGCGAAAGTGGTGGAATCCGCTAAGGCTATCTTGGGGGGGTCGAAAGTTC
GGGTCGACCAGAAGTGCAAATCGTCCGCGCAGATTGACCCCACCCCCGTGATTGTGA
CGTCAAATACTAATATGTGTGCGGTCATCGATGGCAATAGCACCACT*TTCGAA*CATCA
GCAACCCCTCCAGGATCGTATGTTTAAGTTCGAGTTGACTCGGCGGCTGGACCACGAT
TTCGGCAAAGTGACGAAACAGGAGGTGAAGGACTTCTTTAGATGGGCCAAGGACCAC
GTGGTGGAGGTCGAGCACGAGTTTTATGTGAAGAAGGGGGGGGCCAAAAAGCGCCCT
GCACCTTCCGACGCCGACATTTCCGAGCCAAAGAGAGTGCGTGAGAGTGTGGCCCAA
CCCTCCACCAGTGATGCCGAGGCCTCCATTAATTATGCCGACCGCTATCAGAATAAGT
GCTCAAGGCATGTCGGGATGAACCTGATGCTGTTCCCATGCCGCCAGTGCGAGCGCA
TGAACCAGAACAGCAACATTTGTTTTACCCACGGGCAGAAGGATTGCCTGGAATGCT
TCCCGGTCAGCGAGTCACAGCCGGTGTCCGTGGTGAAGAAAGCCTACCAAAAGCTGT
GTTACATCCACCACATTATGGGGAAAGTCCCCGATGCCTGTACCGCATGCGACCTGGT
GAACGTTGACCTCGACGACTGCATTTTCGAGCAGTAA

Sequence of Deoptimized Rep78.

ATGCCCGGGTTTTACGAGATCGTGATTAAGGTGCCATCCGATCTCGACGAGCATCTGC
CCGGGATTAGCGATTCGTTCGTGAATTGGGTCGCCGAAAAGGAGTGGGAGTTGCCCC
CCGATAGCGATATGGACCTGAATCTGATCGAGCAGGCCCCCCTTACCGTCGCCGAGA
AACTGCAACGCGATTTCTTGACCGAGTGGAGACGCGTGAGTAAGGCCCCCGAAGCCC
TGTTTTTCGTGCAATTTGAAAAGGGCGAGTCATACTTTCATATGCACGTGTTGGTCGA
GACTACCGGCGTTAAGTCTATGGTGCTCGGACGGTTTCTGTCACAGATACGCGAAAA
ACTGATCCAGCGTATCTATCGCGGAATCGAGCCAACCCTACCGAATTGGTTCGCCGTT
ACGAAGACCCGTAACGGCGCCGGGGGGGGGAATAAGGTGGTCGACGAGTGCTATAT
CCCTAACTATCTGTTACCGAAAACGCAACCCGAGTTGCAGTGGGCCTGGACTAACAT
GGAGCAATACTTGTCCGCATGCCTGAATCTGACCGAACGCAAACGGTTGGTCGCCCA
GCATCTGACACACGTGAGTCAGACCCAGGAGCAGAATAAGGAGAATCAGAATCCGA
ACTCCGACGCCCCCGTGATACGGTCTAAGACTAGCGCTAGGTATATGGAGTTGGTGG
GGTGGTTGGTCGACAAGGGGATTACCTCCGAGAAACAGTGGATCCAGGAGGACCAGG
CGTCATACATTTCGTTTAACGCCGCATCGAACTCACGGTCACAGATTAAGGCCGCACT
CGACAACGCCGGTAAGATTATGAGTCTGACTAAGACCGCCCCCGATTACTTAGTGGG
ACAGCAACCCGTCGAGGACATTTCGAGTAATCGGATTTACAAAATCCTCGAaCTTAAC
GGATACGACCCCCAATACGCCGCTAGCGTGTTTCTGGGGTGGGCGACTAAGAAATTC
GGAAAGCGTAATACGATTTGGTTGTTCGGACCCGCTACGACCGGCAAAACGAATATC
GCCGAAGCGATCGCGCATACCGTGCCATTCTACGGGTGCGTGAATTGGACGAACGAG
AACTTTCCGTTTAACGATTGCGTCGACAAGATGGTGATTTGGTGGGAGGAGGGAAAG
ATGACCGCTAAGGTGGTCGAGTCCGCGAAAGCGATTCTGGGGGGGTCTAAGGTGAGA
GTCGACCAGAAGTGTAAGTCTTcgGCTCAGATCGATCCGACCCCCGTGATCGTGACCT
CTAACACTAACATGTGCGCCGTGATCGACGGGAATTCGACTACGTTCGAACACCAGC
AGCCATTGCAGGACCGTATGTTCAAATTTGAACTGACTAGGAGACTCGACCACGACTT
CGGAAAGGTGACTAAGCAGGAGGTGAAAGACTTTTTTCGGTGGGCGAAAGACCATGT
GGTCGAGGTCGAGCACGAGTTTTACGTGAAAAAGGGCGGAGCGAAAAAGAGACCCG
CCCCTAGCGACGCCGACATTAGCGAACCGAAACGCGTACGCGAATCCGTTGCGCAAC
CGTCAACCTCCGACGCCGAAGCGTCAATCAATTACGCCGATAGGTACCAGAATAAGT
GCTCTAGACACGTGGGGATGAATCTGATGCTGTTTCCCTGTAGACAGTGCGAGCGTAT
GAACCAGAACTCGAACATTTGCTTTACCCACGGACAGAAAGACTGTCTCGAaTGCTTT
CCCGTGTCCGAATCGCAACCCGTTAGCGTGGTGAAAAAAGCGTACCAGAAACTGTGT
TACATACACCATATTATGGGCAAAGTGCCCGACGCATGCACCGCATGCGATCTGGTG
AACGTCGACCTCGACGATTGCATTTTTGAACAGTAA

Alignments of sRep78 (**S**) and dRep78 (**D**) with wtRep78 (**W**)

Sequences corresponding to the 135bp segment identified as required for Rep mediated inhibition of Adenoviral replication are underlined.

```
    1          11         21         31
D   ATGCCCGGGT TTTACGAGAT CGTGATTAAG GTGCCATCCG
W   ATGCCGGGGT TTTACGAGAT TGTGATTAAG GTCCCCAGCG
S   ATGCCCGGAT TCTACGAAAT CGTCATCAAA GTGCCCTCTG

    41         51         61         71
D   ATCTCGACGA GCATCTGCCC GGGATTAGCG ATTCGTTCGT
W   ACCTTGACGG GCATCTGCCC GGCATTTCTG ACAGCTTTGT
S   ACTTGGATGA ACACCTGCCG GGGATCAGCG ATTCTTTCGT

    81         91         101        111
D   GAATTGGGTC GCCGAAAAGG AGTGGGAGTT GCCCCCCGAT
W   GAACTGGGTG GCCGAGAAGG AATGGGAGTT GCCGCCAGAT
S   CAATTGGGTC GCGGAGAAAG AGTGGGAACT TCCCCCCGAC

    121        131        141        151
D   AGCGATATGG ACCTGAATCT GATCGAGCAG GCCCCCCTTA
W   TCTGACATGG ATCTGAATCT GATTGAGCAG GCACCCCTGA
S   TCGGACATGG ACCTGAACTT AATCGAGCAA GCCCCGCTGA

    161        171        181        191
D   CCGTCGCCGA GAAACTGCAA CGCGATTTCT TGACCGAGTG
W   CCGTGGCCGA GAAGCTGCAG CGCGACTTTC TGACGGAATG
S   CGGTGGCGGA GAAACTGCAG CGGGACTTTC TGACCGAGTG

    201        211        221        231
D   GAGACGCGTG AGTAAGGCCC CCGAAGCCCT GTTTTTCGTG
W   GCGCCGTGTG AGTAAGGCCC CGGAGGCCCT TTTCTTTGTG
S   GAGGCGCGTA TCGAAAGCGC CCGAAGCTTT GTTTTTCGTC

    241        251        261        271
D   CAATTTGAAA AGGGCGAGTC ATACTTTCAT ATGCACGTGT
W   CAATTTGAGA AGGGAGAGAG CTACTTCCAC ATGCACGTGC
S   CAGTTCGAGA AGGGGGAGTC GTACTTTCAT ATGCATGTGT

    281        291        301        311
D   TGGTCGAGAC TACCGGCGTT AAGTCTATGG TGCTCGGACG
W   TCGTGGAAAC CACCGGGGTG AAATCCATGG TTTTGGGACG
S   TGGTGGAGAC TACGGGAGTG AAGAGTATGG TGCTAGGGAG
```

```
  321                 331                 341                 351
D G T T T C T G T C A C A G A T A C G C G A A A A A C T G A T C C A G C G T A T C
W T T T C C T G A G T C A G A T T C G C G A A A A A C T G A T T C A G A G A A T T
S G T T T C T G T C G C A A A T A A G A G A G A A G C T G A T C C A G C G G A T A
  361                 371                 381                 391
D T A T C G C G G A A T C G A G C C A A C C C T A C C G A A T T G G T T C G C C G
W T A C C G C G G G A T C G A G C C G A C T T T G C C A A A C T G G T T C G C G G
S T A C C G T G G C A T T G A G C C C A C C C T T C C C A A T T G G T T T G C C G
  401                 411                 421                 431
D T T A C G A A G A C C C G T A A C G G C G C C G G G G G G G G G A A T A A G G T
W T C A C A A A G A C C A G A A A T G G C G C C G G A G G C G G G A A C A A G G T
S T G A C C A A A A C T C G T A A C G G A G C A G G G G G G G G A A A T A A A G T
  441                 451                 461                 471
D G G T C G A C G A G T G C T A T A T C C C T A A C T A T C T G T T A C C G A A A
W G G T G G A T G A G T G C T A C A T C C C C A A T T A C T T G C T C C C C A A A
S C G T C G A C G A G T G C T A T A T T C C G A A C T A C C T C T T G C C C A A G
  481                 491                 501                 511
D A C G C A A C C C G A G T T G C A G T G G G C C T G G A C T A A C A T G G A G C
W A C C C A G C C T G A G C T C C A G T G G G C G T G G A C T A A T A T G G A A C
S A C G C A G C C C G A A T T G C A G T G G G C C T G G A C C A A C A T G G A G C
  521                 531                 541                 551
D A A T A C T T G T C C G C A T G C C T G A A T C T G A C C G A A C G C A A A C G
W A G T A T T T A A G C G C C T G T T T G A A T C T C A C G G A G C G T A A A C G
S A A T A C C T G T C A G C G T G C C T C A A C T T G A C C G A A A G A A A G A G
  561                 571                 581                 591
D G T T G G T C G C C C A G C A T C T G A C A C A C G T G A G T C A G A C C C A G
W G T T G G T G G C G C A G C A T C T G A C G C A C G T G T C G C A G A C G C A G
S A C T C G T G G C C C A G C A C C T G A C C C A T G T C T C A C A G A C C C A G
  601                 611                 621                 631
D G A G C A G A A T A A G G A G A A T C A G A A T C C G A A C T C C G A C G C C C
W G A G C A G A A C A A A G A G A A T C A G A A T C C C A A T T C T G A T G C G C
S G A A C A G A A T A A G G A A A A C C A A A A C C C A A A T A G C G A C G C C C
  641                 651                 661                 671
D C C G T G A T A C G G T C T A A G A C T A G C G C T A G G T A T A T G G A G T T
W C G G T G A T C A G A T C A A A A A C T T C A G C C A G G T A C G G A G A G C T
S C C G T G A T A C G G A G C A A G A C C A G C G C T C G C T A C A T G G A G T T
```

```
  681                 691                 701                 711
D G G T G G G G T G G T T G G T C G A C A A G G G G A T T A C C T C C G A G A A A
W G G T C G G G T G G C T C G T G G A C A A G G G G A T T A C C T C G G A G A A G
S A G T G G G A T G G T T G G T G G A T A A A G G A A T C A C G T C T G A G A A A

  721                 731                 741                 751
D C A G T G G A T C C A G G A G G A C C A G G C G T C A T A C A T T T C G T T T A
W C A G T G G A T C C A G G A G G A C C A G G C C T C A T A C A T C T C C T T C A
S C A A T G G A T T C A G G A G G A C C A G G C G T C C T A C A T T A G T T T T A

  761                 771                 781                 791
D A C G C C G C A T C G A A C T C A C G G T C A C A G A T T A A G G C C G C A C T
W A T G C G G C C T C C A A C T C G C G G T C C C A A A T C A A G G C T G C C T T
S A C G C C G C G T C A A A T A G C A G A T C T C A G A T T A A A G C C G C G C T

  801                 811                 821                 831
D C G A C A A C G C C G G T A A G A T T A T G A G T C T G A C T A A G A C C G C C
W G G A C A A T G C G G G A A A G A T T A T G A G C C T G A C T A A A A C C G C C
S C G A T A A C G C C G G C A A A A T C A T G T C G C T G A C C A A G A C A G C T

  841                 851                 861                 871
D C C C G A T T A C T T A G T G G G A C A G C A A C C C G T C G A G G A C A T T T
W C C C G A C T A C C T G G T G G G C C A G C A G C C C G T G G A G G A C A T T T
S C C C G A C T A C C T G G T G G G A C A G C A G C C G G T G G A G G A C A T C T

  881                 891                 901                 911
D C G A G T A A T C G G A T T T A C A A A A T C C T C G A A C T T A A C G G A T A
W C C A G C A A T C G G A T T T A T A A A A T T T T G G A A C T A A A C G G G T A
S C T T C T A A C C G T A T C T A C A A G A T C C T t G A G T T G A A T G G C T A

  921                 931                 941                 951
D C G A C C C C C A A T A C G C C G C T A G C G T G T T T C T G G G G T G G G C G
W C G A T C C C C A A T A T G C G G C T T C C G T C T T T C T G G G A T G G G C C
S C G A C C C A C A G T A C G C C G C C T C A G T G T T T C T G G G C T G G G C A

  961                 971                 981                 991
D A C T A A G A A A T T C G G A A A G C G T A A T A C G A T T T G G T T G T T C G
W A C G A A A A A G T T C G G C A A G A G G A A C A C C A T C T G G C T G T T T G
S A C C A A G A A A T T T G G G A A A C G C A A T A C G A T T T G G C T G T T C G
```

1001 1011 1021 1031

D G A C C C G C T A C G A C C G G C A A A A C G A A T A T C G C C G A A G C G A T

W G G C C T G C A A C T A C C G G G A A G A C C A A C A T C G C G G A G G C C A T

S G A C C C G C C A C C A C T G G T A A G A C T A A T A T T G C C G A G G C G A T

1041 1051 1061 1071

D C G C G C A T A C C G T G C C A T T C T A C G G G T G C G T G A A T T G G A C G

W A G C C C A C A C T G T G C C C T T C T A C G G G T G C G T A A A C T G G A C C

S C G C A C A T A C C G T T C C G T T T T A C G G A T G C G T G A A T T G G A C T

1081 1091 1101 1111

D A A C G A G A A C T T T C C G T T T A A C G A T T G C G T C G A C A A G A T G G

W A A T G A G A A C T T T C C C T T C A A C G A C T G T G T C G A C A A G A T G G

S A A C G A A A A T T T C C C C T T T A A T G A T T G C G T G G A C A A G A T G G

1121 1131 1141 1151

D T G A T T T G G T G G G A G G A G G G A A A G A T G A C C G C T A A G G T G G T

W T G A T C T G G T G G G A G G A G G G G A A G A T G A C C G C C A A G G T C G T

S T T A T T T G G T G G G A G G A A G G A A A G A T G A C T G C G A A A G T G G T

1161 1171 1181 1191

D C G A G T C C G C G A A A G C G A T T C T G G G G G G G T C T A A G G T G A G A

W G G A G T C G G C C A A A G C C A T T C T C G G A G G A A G C A A G G T G C G C

S G G A A T C C G C T A A G G C T A T C T T G G G G G G G T C G A A A G T T C G G

1201 1211 1221 1231

D G T C G A C C A G A A G T G T A A G T C T T C G G C T C A G A T C G A T C C G A

W G T G G A C C A G A A A T G C A A G T C C T C G G C C C A G A T A G A C C C G A

S G T C G A C C A G A A G T G C A A A T C G T C C G C G C A G A T T G A C C C C A

1241 1251 1261 1271

D C C C C C G T G A T C G T G A C C T C T A A C A C T A A C A T G T G C G C C G T

W C T C C C G T G A T C G T C A C C T C C A A C A C C A A C A T G T G C G C C G T

S C C C C C G T G A T T G T G A C G T C A A A T A C T A A T A T G T G T G C G G T

1281 1291 1301 1311

D G A T C G A C G G G A A T T C G A C T A C G T T C G A A C A C C A G C A G C C A

W G A T T G A C G G G A A C T C A A C G A C C T T C G A A C A C C A G C A G C C G

S C A T C G A T G G C A A T A G C A C C A C T T T C G A A C A T C A G C A A C C C

```
  1321                1331                1341                1351
D T T G C A G G A C C G T A T G T T C A A A T T T G A A C T G A C T A G G A G A C
W T T G C A A G A C C G G A T G T T C A A A T T T G A A C T C A C C C G C C G T C
S C T C C A G G A T C G T A T G T T T A A G T T C G A G T T G A C T C G G C G G C
  1361                1371                1381                1391
D T C G A C C A C G A C T T C G G A A A G G T G A C T A A G C A G G A G G T G A A
W T G G A T C A T G A C T T T G G G A A G G T C A C C A A G C A G G A A G T C A A
S T G G A C C A C G A T T T C G G C A A A G T G A C G A A A C A G G A G G T G A A

  1401                1411                1421                1431
D A G A C T T T T T T C G G T G G G C G A A A G A C C A T G T G G T C G A G G T C
W A G A C T T T T T C C G G T G G G C A A A G G A T C A C G T G G T T G A G G T G
S G G A C T T C T T T A G A T G G G C C A A G G A C C A C G T G G T G G A G G T C
  1441                1451                1461                1471
D G A G C A C G A G T T T T A C G T G A A A A A G G G C G G A G C G A A A A A G A
W G A G C A T G A A T T C T A C G T C A A A A A G G G T G G A G C C A A G A A A A
                                          G
S G A G C A C G A G T T T T A T G T G A A   A A G G G G G G G G C C A A A A A G C
  1481                1491                1501                1511
D G A C C C G C C C C T A G C G A C G C C G A C A T T A G C G A A C C G A A A C G
W G A C C C G C C C C C A G T G A C G C A G A T A T A A G T G A G C C C A A A C G
S G C C C T G C A C C T T C C G A C G C C G A C A T T T C C G A G C C A A A G A G
  1521                1531                1541                1551
D C G T A C G C G A A T C C G T T G C G C A A C C G T C A A C C T C C G A C G C C
W G G T G C G C G A G T C A G T T G C G C A G C C A T C G A C G T C A G A C G C G
S A G T G C G T G A G A G T G T G G C C C A A C C C T C C A C C A G T G A T G C C
  1561                1571                1581                1591
D G A A G C G T C A A T C A A T T A C G C C G A T A G G T A C C A G A A T A A G T
W G A A G C T T C G A T C A A C T A C G C A G A C A G A T A C C A A A A C A A A T
  G
S   A G G C C T C C A T T A A T T A T G C C G A C C G C T A T C A G A A T A A G T
```

```
1601                1611                1621                1631
D G C T C T A G A C A C G T G G G G A T G A A T C T G A T G C T G T T T C C C T G
W G T T C T C G T C A C G T G G G C A T G A A T C T G A T G C T G T T T T C C C T G
S G C T C A A G G C A T G T C G G G A T G A A C C T G A T G C T G T T C C C A T G

1641                1651                1661                1671
D T A G A C A G T G C G A G C G T A T G A A C C A G A A C T C G A A C A T T T G C
W C A G A C A A T G C G A G A G A A T G A A T C A G A A T T C A A A T A T C T G C
S C C G C C A G T G C G A G C G C A T G A A C C A G A A C A G C A A C A T T T G T

1681                1691                1701                1711
D T T T A C C C A C G G A C A G A A A G A C T G T C T C G A A T G C T T T C C C G
S T T C A C T C A C G G A C A G A A A G A C T G T T T A G A G T G C T T T C C C G
W T T T A C C C A C G G G C A G A A G G A T T G C C T G G A A T G C T T C C C G G

1721                1731                1741                1751
D T G T C C G A A T C G C A A C C C G T T A G C G T G G T G A A A A A G C G T A
S T G T C A G A A T C T C A A C C C G T T T C T G T C G T C A A A A A G G C G T A
W T C A G C G A G T C A C A G C C G G T G T C C G T G G T G A A G A A A G C C T A

1761                1771                1781                1791
D C C A G A A A C T G T G T T A C A T A C A C C A T A T T A T G G G C A A A G T G
S T C A G A A A C T G T G C T A C A T T C A T C A T A T C A T G G G A A A G G T G
W C C A A A A G C T G T G T T A C A T C C A C C A C A T T A T G G G G A A A G T C

1801                1811                1821                1831
D C C C G A C G C A T G C A C C G C A T G C G A T C T G G T G A A C G T C G A C C
S C C A G A C G C T T G C A C T G C C T G C G A T C T G G T C A A T G T G G A T T
W C C C G A T G C C T G T A C C G C A T G C G A C C T G G T G A A C G T T G A C C

1841                1851                1861
D T C G A C G A T T G C A T T T T T G A A C A G T A A
S T G G A T G A C T G C A T C T T T G A A C A A T A A
W T C G A C G A C T G C A T T T T C G A G C A G T A A
```

Design I ,II ,III and IV used for localization of the sequence specific signal are given below. Sequences of each segment are separated from each other. The Scr/wt pattern for each design is given above its sequence, where D represents Scrambled and W represents wtRep sequence. The 135bp putative sequence identified is in red.

Design I : DDDDDDDWWWWWWW

ATGCCCGGATTCTACGAAATCGTCATCAAAGTGCCCTCTGACTTGGATGAACACCTGC
CGGGGATCAGCGATTCTTTCGTCAATTGGGTCGCGGAGAAAGAGTGGGAACTTCCCCCCGACTC
GGACATGGAC

CTGAACTTAATCGAGCAAGCCCCGCTGACGGTGGCGGAGAAACTGCAGCGGGACTTTC
TGACCGAGTGGAGGCGCGTATCGAAAGCGCCCGAAGCTTTGTTTTTCGTCCAGTTCGAGAAGGG
GGAGTCGTAC

TTTCATATGCATGTGTTGGTGGAGACTACGGGAGTGAAGAGTATGGTGCTAGGGAGGT
TTCTGTCGCAAATAAGAGAGAAGCTGATCCAGCGGATATACCGTGGCATTGAGCCCACCCTTCC
CAATTGGTTT

GCCGTGACCAAAACTCGTAACGGAGCAGGGGGGGAAATAAAGTCGTCGACGAGTGC
TATATTCCGAACTACCTCTTGCCCAAGACGCAGCCCGAATTGCAGTGGGCCTGGACCAACATGG
AGCAATACCTG

TCAGCGTGCCTCAACTTGACCGAAAGAAAGAGACTCGTGGCCCAGCACCTGACCCATG
TCTCACAGACCCAGGAACAGAATAAGGAAAACCAAAACCCAAATAGCGACGCCCCCGTGATAC
GGAGCAAGACC

AGCGCTCGCTACATGGAGTTAGTGGGATGGTTGGTGGATAAAGGAATCACGTCTGAGA
AACAATGGATTCAGGAGGACCAGGCGTCCTACATTAGTTTTAACGCCGCGTCAAATAGCAGATC
TCAGATTAAA

GCCGCGCTCGATAACGCCGGCAAAATCATGTCGCTGACCAAGACAGCTCCCGACTACC
TGGTGGGACAGCAGCCGGTGGAGGACATCTCTTCTAACCGTATCTACAAGATCCTTGAGTTGAA
TGGCTACGAC

CCCCAATATGCGGCTTCCGTCTTTCTGGGATGGGCCACGAAAAAGTTCGGCAAGAGGA
ACACCATCTGGCTGTTTGGGCCTGCAACTACCGGGAAGACCAACATCGCGGAGGCCATAGCCCA
CACTGTGCCC

TTCTACGGGTGCGTAAACTGGACCAATGAGAACTTTCCCTTCAACGACTGTGTCGACA
AGATGGTGATCTGGTGGGAGGAGGGGAAGATGACCGCCAAGGTCGTGGAGTCGGCCAAAGCCA
TTCTCGGAGGAAGC

AAGGTGCGCGTGGACCAGAAATGCAAGTCCTCGGCCCAGATAGACCCGACTCCCGTG
ATCGTCACCTCCAACACCAACATGTGCGCCGTGATTGACGGGAACTCAACGACCTTCGAACACC
AGCAGCCGTTGCAA

GACCGGATGTTCAAATTTGAACTCACCCGCCGTCTGGATCATGACTTTGGGAAGGTCA
CCAAGCAGGAAGTCAAAGACTTTTTCCGGTGGGCAAAGGATCACGTGGTTGAGGTGGAGCATG
AATTCTACGTCAAA

AAGGGTGGAGCCAAGAAAAGACCCGCCCCCAGTGACGCAGATATAAGTGAGCCCAAA
CGGGTGCGCGAGTCAGTTGCGCAGCCATCGACGTCAGACGCGGAAGCTTCGATCAACTACGCA
GACAGGTACCAAAAC

AAATGTTCTCGTCACGTGGGCATGAATCTGATGCTGTTTCCCTGCAGACAATGCGAGA
GAATGAATCAGAATTCAAATATCTGCTTCACTCACGGACAGAAAGACTGTTTAGAGTGCTTTCC
CGTGTCAGAATCT

CAACCCGTTTCTGTCGTCAAAAAGGCGTATCAGAAACTGTGCTACATTCATCATATCAT
GGGAAAGGTGCCAGACGCTTGCACTGCCTGCGATCTGGTCAATGTGGATTTGGATGACTGCATC
TTTGAACAATAA

Design II: DDDWWWWWWWDDDD

ATGCCCGGATTCTACGAAATCGTCATCAAAGTGCCCTCTGACTTGGATGAACACCTGC
CGGGGATCAGCGATTCTTTCGTCAATTGGGTCGCGGAGAAAGAGTGGGAACTTCCCCCCGACTC
GGACATGGAC

CTGAACTTAATCGAGCAAGCCCCGCTGACGGTGGCGGAGAAACTGCAGCGGGACTTTC
TGACCGAGTGGAGGCGCGTATCGAAAGCGCCCGAAGCTTTGTTTTTCGTCCAGTTCGAGAAGGG
GGAGTCGTAC

TTTCATATGCATGTGTTGGTGGAGACTACGGGAGTGAAGAGTATGGTGCTAGGGAGGT
TTCTGTCGCAAATAAGAGAGAAGCTGATCCAGCGGATATACCGTGGCATTGAGCCCACCCTTCC
CAATTGGTTT

GCGGTCACAAAGACCAGAAATGGCGCCGGAGGCGGGAACAAGGTGGTGGATGAGTGC
TACATCCCCAATTACTTGCTCCCCAAAACCCAGCCTGAGCTCCAGTGGGCGTGGACTAATATGG
AACAGTATTTA

AGCGCCTGTTTGAATCTCACGGAGCGTAAACGGTTGGTGGCGCAGCATCTGACGCAGT
GTCGCAGACGCAGGAGCAGAACAAAGAGAATCAGAATCCCAATTCTGATGCGCCGGTGATCAG
ATCAAAAACT

TCAGCCAGGTACATGGAGCTGGTCGGGTGGCTCGTGGACAAGGGGATTACCTCGGAGA
AGCAGTGGATCCAGGAGGACCAGGCCTCATACATCTCCTTCAATGCGGCCTCCAACTCGCGGTC
CCAAATCAAG

GCTGCCTTGGACAATGCGGGAAAGATTATGAGCCTGACTAAAACCGCCCCCGACTACC
TGGTGGGCCAGCAGCCCGTGGAGGACATTTCCAGCAATCGGATTTATAAAATTTTGGAACTAAA
CGGGTACGAT

CCCCAATATGCGGCTTCCGTCTTTCTGGGATGGGCCACGAAAAAGTTCGGCAAGAGGA
ACACCATCTGGCTGTTTGGGCCTGCAACTACCGGGAAGACCAACATCGCGGAGGCCATAGCCCA
CACTGTGCCC

TTCTACGGGTGCGTAAACTGGACCAATGAGAACTTTCCCTTCAACGACTGTGTCGACA
AGATGGTGATCTGGTGGGAGGAGGGGAAGATGACCGCCAAGGTCGTGGAGTCGGCCAAAGCCA
TTCTCGGAGGAAGC

AAGGTGCGCGTGGACCAGAAATGCAAGTCCTCGGCCCAGATAGACCCGACTCCCGTG
ATCGTCACCTCCAACACCAACATGTGCGCCGTGATTGACGGGAACTCAACGACCTTCGAACACC
AGCAGCCGTTGCAA

GATCGTATGTTTAAGTTCGAGTTGACTCGGCGGCTGGACCACGATTTCGGCAAAGTGA
CGAAACAGGAGGTGAAGGACTTCTTTAGATGGGCCAAGGACCACGTGGTGGAGGTCGAGCACG
AGTTTTATGTGAAG

AAGGGGGGGGCCAAAAAGCGCCCTGCACCTTCCGACGCCGACATTTCCGAGCCAAAG
AGAGTGCGTGAGAGTGTGGCCCAACCCTCCACCAGTGATGCCGAGGCCTCCATTAATTATGCCG
ACCGCTATCAGAAT

AAGTGCTCAAGGCATGTCGGGATGAACCTGATGCTGTTCCCATGCCGCCAGTGCGAGC
GCATGAACCAGAACAGCAACATTTGTTTTACCCACGGGCAGAAGGATTGCCTGGAATGCTTCCC
GGTCAGCGAGTCA

CAGCCGGTGTCCGTGGTGAAGAAAGCCTACCAAAAGCTGTGTTACATCCACCACATTA
TGGGGAAAGTCCCCGATGCCTGTACCGCATGCGACCTGGTGAACGTTGACCTCGACGACTGCAT
TTTCGAGCAGTAA

Design III WWDDDWWWDDDDWW

ATGCCGGGGTTTTACGAGATTGTGATTAAGGTCCCCAGCGACCTTGACGAGCATCTGC
CCGGCATTTCTGACAGCTTTGTGAACTGGGTGGCCGAGAAGGAATGGGAGTTGCCGCCAGATTC
TGACATGGAT

CTGAATCTGATTGAGCAGGCACCCCTGACCGTGGCCGAGAAGCTGCAGCGCGACTTTC
TGACGGAATGGCGCCGTGTGAGTAAGGCCCCGGAGGCCCTTTTCTTTGTGCAATTTGAGAAGGG
AGAGAGCTAC

TTTCATATGCATGTGTTGGTGGAGACTACGGGAGTGAAGAGTATGGTGCTAGGGAGGT
TTCTGTCGCAAATAAGAGAGAAGCTGATCCAGCGGATATACCGTGGCATTGAGCCCACCCTTCC
CAATTGGTTT

GCCGTGACCAAAACTCGTAACGGAGCAGGGGGGGGAAATAAAGTCGTCGACGAGTGC
TATATTCCGAACTACCTCTTGCCCAAGACGCAGCCCGAATTGCAGTGGGCCTGGACCAACATGG
AGCAATACCTG

TCAGCGTGCCTCAACTTGACCGAAAGAAAGAGACTCGTGGCCCAGCACCTGACCCATG
TCTCACAGACCCAGGAACAGAATAAGGAAAACCAAAACCCAAATAGCGACGCCCCCGTGATAC
GGAGCAAGACC

TCAGCCAGGTACATGGAGCTGGTCGGGTGGCTCGTGGACAAGGGGATTACCTCGGAGA
AGCAGTGGATCCAGGAGGACCAGGCCTCATACATCTCCTTCAATGCGGCCTCCAACTCGCGGTC
CCAAATCAAG

GCTGCCTTGGACAATGCGGGAAAGATTATGAGCCTGACTAAAACCGCCCCCGACTACC
TGGTGGGCCAGCAGCCCGTGGAGGACATTTCCAGCAATCGGATTTATAAAATTTTGGAACTAAA
CGGGTACGAT

CCCCAATATGCGGCTTCCGTCTTTCTGGGATGGGCCACGAAAAAGTTCGGCAAGAGGA
ACACCATCTGGCTGTTTGGGCCTGCAACTACCGGGAAGACCAACATCGCGGAGGCCATAGCCCA
CACTGTGCCC

TTTTACGGATGCGTGAATTGGACTAACGAAAATTTCCCCTTTAATGATTGCGTGGACAA
GATGGTTATTTGGTGGGAGGAAGGAAAGATGACTGCGAAAGTGGTGGAATCCGCTAAGGCTAT
CTTGGGGGGGTCG

AAAGTTCGGGTCGACCAGAAGTGCAAATCGTCCGCGCAGATTGACCCCACCCCCGTGA
TTGTGACGTCAAATACTAATATGTGTGCGGTCATCGATGGCAATAGCACCACTTTCGAACATCA
GCAACCCCTCCAG

GATCGTATGTTTAAGTTCGAGTTGACTCGGCGGCTGGACCACGATTTCGGCAAAGTGA
CGAAACAGGAGGTGAAGGACTTCTTTAGATGGGCCAAGGACCACGTGGTGGAGGTCGAGCACG
AGTTTTATGTGAAG

AAGGGGGGGGCCAAAAAGCGCCCTGCACCTTCCGACGCCGACATTTCCGAGCCAAAG
AGAGTGCGTGAGAGTGTGGCCCAACCCTCCACCAGTGATGCCGAGGCCTCCATTAATTATGCCG
ACCGCTATCAGAAT

AAATGTTCTCGTCACGTGGGCATGAATCTGATGCTGTTTCCCTGCAGACAATGCGAGA
GAATGAATCAGAATTCAAATATCTGCTTCACTCACGGACAGAAAGACTGTTTAGAGTGCTTTCC
CGTGTCAGAATCT

CAACCCGTTTCTGTCGTCAAAAAGGCGTATCAGAAACTGTGCTACATTCATCATATCAT
GGGAAAGGTGCCAGACGCTTGCACTGCCTGCGATCTGGTCAATGTGGATTTGGATGACTGCATC
TTTGAACAATAA

Design IV: WDWWDDWDDWWDDW

ATGCCGGGGTTTTACGAGATTGTGATTAAGGTCCCCAGCGACCTTGACGAGCATCTGC
CCGGCATTTCTGACAGCTTTGTGAACTGGGTGGCCGAGAAGGAATGGGAGTTGCCGCCAGATTC
TGACATGGAT

CTGAACTTAATCGAGCAAGCCCCGCTGACGGTGGCGGAGAAACTGCAGCGGGACTTTC
TGACCGAGTGGAGGCGCGTATCGAAAGCGCCCGAAGCTTTGTTTTTCGTCCAGTTCGAGAAGGG
GGAGTCGTAC

TTCCACATGCACGTGCTCGTGGAAACCACCGGGGTGAAATCCATGGTTTTGGGACGTT
TCCTGAGTCAGATTCGCGAAAAACTGATTCAGAGAATTTACCGCGGGATCGAGCCGACTTTGCC
AAACTGGTTC

GCGGTCACAAAGACCAGAAATGGCGCCGGAGGCGGGAACAAGGTGGTGGATGAGTGC
TACATCCCCAATTACTTGCTCCCCAAAACCCAGCCTGAGCTCCAGTGGGCGTGGACTAATATGG
AACAGTATTTA

TCAGCGTGCCTCAACTTGACCGAAAGAAAGAGACTCGTGGCCCAGCACCTGACCCATG
TCTCACAGACCCAGGAACAGAATAAGGAAAACCAAAACCCAAATAGCGACGCCCCCGTGATAC
GGAGCAAGACC

AGCGCTCGCTACATGGAGTTAGTGGGATGGTTGGTGGATAAAGGAATCACGTCTGAGA
AACAATGGATTCAGGAGGACCAGGCGTCCTACATTAGTTTTAACGCCGCGTCAAATAGCAGATC
TCAGATTAAA

GCTGCCTTGGACAATGCGGGAAAGATTATGAGCCTGACTAAAACCGCCCCCGACTACC
TGGTGGGCCAGCAGCCCGTGGAGGACATTTCCAGCAATCGGATTTATAAAATTTTGGAACTAAA
CGGGTACGAT

CCACAGTACGCCGCCTCAGTGTTTCTGGGCTGGGCAACCAAGAAATTTGGGAAACGCA
ATACGATTTGGCTGTTCGGACCCGCCACCACTGGTAAGACTAATATTGCCGAGGCGATCGCACA
TACCGTTCCG

TTTTACGGATGCGTGAATTGGACTAACGAAAATTTCCCCTTTAATGATTGCGTGGACAA
GATGGTTATTTGGTGGGAGGAAGGAAAGATGACTGCGAAAGTGGTGGAATCCGCTAAGGCTAT
CTTGGGGGGGTCG

AAGGTGCGCGTGGACCAGAAATGCAAGTCCTCGGCCCAGATAGACCCGACTCCCGTG
ATCGTCACCTCCAACACCAACATGTGCGCCGTGATTGACGGGAACTCAACGACCTTCGAACACC
AGCAGCCGTTGCAA

GACCGGATGTTCAAATTTGAACTCACCCGCCGTCTGGATCATGACTTTGGGAAGGTCA
CCAAGCAGGAAGTCAAAGACTTTTTCCGGTGGGCAAAGGATCACGTGGTTGAGGTGGAGCATG
AATTCTACGTCAAA

AAGGGGGGGGCCAAAAAGCGCCCTGCACCTTCCGACGCCGACATTTCCGAGCCAAAG
AGAGTGCGTGAGAGTGTGGCCCAACCCTCCACCAGTGATGCCGAGGCCTCCATTAATTATGCCG
ACCGCTATCAGAAT

AAGTGCTCAAGGCATGTCGGGATGAACCTGATGCTGTTCCCATGCCGCCAGTGCGAGC
GCATGAACCAGAACAGCAACATTTGTTTTACCCACGGGCAGAAGGATTGCCTGGAATGCTTCCC
GGTCAGCGAGTCA

CAACCCGTTTCTGTCGTCAAAAAGGCGTATCAGAAACTGTGCTACATTCATCATATCAT
GGGAAAGGTGCCAGACGCTTGCACTGCCTGCGATCTGGTCAATGTGGATTTGGATGACTGCATC
TTTGAACAATAA