

MOLECULAR CHARACTERIZATION OF THE AVIAN ADENO-ASSOCIATED VIRUS
AND ITS USE FOR THE DEVELOPMENT OF A VIRAL VECTOR SYSTEM

by

CARLOS ESTEVEZ

(Under the direction of Pedro Villegas)

ABSTRACT

The usefulness of viral-vectored delivery of genetic information to cells and tissues, both *in vivo* and *in vitro*, has been well documented. Advances in recombinant DNA technologies and cellular biology have made possible the use of recombinant viruses to correct underlying genetic defects and to express immunogenic peptides to induce protective immune responses in animals. The objective of this work is the use of a non-pathogenic, replication defective avian parvovirus, the avian adeno-associated virus (AAAV), to generate a virus-based system for gene delivery in poultry. Aiming at this purpose we have cloned and sequenced the two known strains of the AAAV (VR-865 and DA-1). Complete infective viral particles of both strains were rescued from these clones, using a previously described system that includes the use of the HEK 293 cell line and a plasmid that encode for some of the immediate early genes of the human adenovirus type 5. These infectious clones obtained were used to generate a plasmid-based system for the

production of recombinant AAV particles coding for the *LacZ* gene as a reporter. Recombinant viral populations obtained from these plasmids were used to infect primary chicken embryo cell cultures and embryonating chicken eggs. Expression of the reporter gene was observed on both systems. The recombinant plasmids obtained were also used to assess the role of the viral inverted terminal repeats and non-structural proteins on the level and duration of transgene expression *in vitro*. Results showed that both the inverted terminal repeats and expression of the non-structural proteins of the virus significantly increase the level and duration of transgene expression in cell cultures.

INDEX WORDS: Avian adeno-associated virus, Recombinant DNA, Virus vector, Gene delivery, Gene expression.

MOLECULAR CHARACTERIZATION OF THE AVIAN ADENO-ASSOCIATED VIRUS
AND ITS USE FOR THE DEVELOPMENT OF A VIRAL VECTOR SYSTEM

by

CARLOS ESTEVEZ

DVM, Universidad Nacional Pedro Henriques Ureña,

Santo Domingo, Dominican Republic, 1988

M.S., The University of Georgia, 2001

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial

Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2004

© 2004

Carlos Estevez

All Rights Reserved

MOLECULAR CHARACTERIZATION OF THE AVIAN ADENO-ASSOCIATED VIRUS
AND ITS USE FOR THE DEVELOPMENT OF A VIRAL VECTOR SYSTEM

by

CARLOS ESTEVEZ

Major Professor: Pedro Villegas

Committee: Mark Jackwood
Liliana Jaso-Friedman
John Glisson
Daniel J. King

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2004

DEDICATION

I dedicate this work to my family...

Thanks for all the love and support you have always given me.

Los quiero mucho

Carlos

ACKNOWLEDGEMENTS

To my daughter Maria Alejandra: the future is yours and you are the master of it. Make of your life a journey of discovery and always remember that you have a lot of potential to exploit, and that I love you dearly.

To my Mom and Dad: thanks for teaching me the value of family and how to be a good person. I'm proud of you guys and I love you with all my heart.

To my brother and sister: our bond goes beyond being siblings... you'll be ever-present in my thoughts and heart.

To Luciana: thanks for putting up with me during all this time... I love you Nushka.

To my labmates Alejandro John Ivan and Linda: thanks for all your support during these years.

To Dr Villegas: last, but in no way least.....thanks forgiving me the best gift a person can give to another person, helping me reach my full potential. I wish I could explain how much this means to me. Thanks.

Carlos Estevez

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	v
LIST OF TABLES	viii
LIST OF FIGURES	ix
 CHAPTER	
I INTRODUCTION	1
Purpose of this Study	1
Objective and Originality	4
II LITERATURE REVIEW	6
The Parvoviruses	6
The Adeno-Associated Virus	12
The Avian Adeno-Associated Virus	18
Adeno Associated-Virus as Viral Vectors	20
References	27
III SEQUENCE ANALYSIS, VIRAL RESCUE FROM INFECTIOUS CLONES AND GENERATION OF RECOMBINANT VIRIONS OF THE AVIAN ADENO- ASSOCIATED VIRUS	34
Introduction	36
Materials and Methods	39
Results	46

Discussion	50
References	53
IV RECOMBINANT AVIAN ADENO-ASSOCIATED VIRUS: TRANSGENE EXPRESSION IN VIVO AND ENHANCEMENT OF EXPRESSION IN VITRO.....	81
Introduction	83
Materials and Methods	85
Results	93
Discussion	95
References	97
V CONCLUSIONS AND DISCUSSION	107
References	112

LIST OF TABLES

	Page
Table 1: Transfection Protocol of Cell Lines.....	101
Table 2: Transgene Detection Method and Monolayer Usage.	102
Table 3: Chemiluminescence Readings (RLU) in Embryos and Chicks.....	103
Table 4: Chemiluminescence Readings (RLU) in Cell Cultures.	104
Table 5: Chemiluminescence Readings (RLU) in Cell Cultures (cont.).	105

LIST OF FIGURES

	Page
Figure 1: <i>Eco</i> RI Digestion of Plasmids Containing the AAV genomes	57
Figure 2: Genomic Map Distribution of the Avian Adeno-Associated Virus	58
Figure 3: T-Shaped Structure of the Viral Inverted Terminal Repeats.....	59
Figure 4: Nucleotide Alignment of the AAV and AAV strains.....	61
Figure 5: Phylogenetic Trees and Sequence Distances of the AAV and AAV strains.....	71
Figure 6: Amino Acid Sequence of the Rep 78 Protein.....	73
Figure 7: Rescued AAV from infectious clone.....	75
Figure 8: PCR of Viruses Rescued in HEK 293 Cell Cultures.....	76
Figure 9: Plasmids pEVA3V-RC-CMV and pEVA3V-LacZ-ITR.....	77
Figure 10: Electron Micrograph of rAAV Coding for the <i>LacZ</i> Reporter Gene	79
Figure 11: Expression of B-galactosidase in Primary Chicken Embryo Cell Cultures	80
Figure 12: Direct Staining of Cell Cultures at Passage #3.....	106

CHAPTER I

INTRODUCTION

Purpose of this Study

The classical approach to prevention of viral infectious diseases in poultry medicine has relied, almost exclusively, in the use of vaccines to elicit protective immune responses against viral pathogens. Many of these vaccines are composed of live attenuated viruses that cause a “controlled” outbreak of disease when applied to the birds. These controlled outbreaks elicit the production of antibodies and the activation of cell mediated responses that are capable of eliminating the vaccinal organism and, at the same time, prevent disease in future exposures to pathogens via immunogenic memory responses. The problem with the use of live vaccines is that the vaccinal virus replication in the population may become “uncontrolled”, which translates into the occurrence of vaccine reactions, disease, and heavy economic losses due to poor performance, opportunistic secondary infections and increased mortality. Furthermore, some viruses are too pathogenic to be used as live vaccines in the field. As an alternative to the use of live vaccines, viruses can be chemically inactivated to generate killed vaccines. These vaccines do not have some of the deleterious effects of live vaccines, namely uncontrolled viral replication and vaccine reactions, but they usually fail to induce strong cell mediated immunity, and a previous application of a live vaccine is generally necessary to elicit a strong immune response to the inactivated vaccine.

Recent advances in molecular biology, in the understanding of the molecular basis of pathogenesis in the context of viral and bacterial infections, and in the mechanisms involved in the generation of protective immune responses have opened new avenues for the treatment and prevention of infectious diseases. Currently, one of the most promising techniques in infectious disease prevention is the delivery of genes encoding for immunogenic peptides derived from pathogenic organisms. This technique is safe, is not associated with the deleterious consequences of vaccine reactions, and can be used to induce immunity even against viruses that are too virulent to be used as vaccines. Gene delivery has been achieved by the inoculation of plasmids encoding for immunogenic peptides via gene gun and injection. Although protection against disease has been obtained by this method in experimental settings, the cost constrains and the requirement for mass immunization in the poultry industry has made development of practical DNA vaccines difficult. Another approach to gene delivery has been the use of recombinant viruses encoding for heterologous immunogenic peptide genes derived from viruses of economic interest. In the poultry industry, different viruses such as members of the poxvirus and herpesvirus families have been used as vectors to accommodate these genes of interest that are to be delivered into the hosts. Advantages of the viral-vectored systems for gene delivery include higher efficiencies in the transfection and expression of the gene of interest, increased immune response and ease of manipulation and application in comparison to the naked DNA approach. Still, the viral-vectored approach to gene delivery is not without drawbacks. First, the generation of a recombinant virus stock using the adeno and herpesvirus relies on homologous recombination between a plasmid harboring the gene of interest and the genome of the virus used as vector. This procedure is somewhat inefficient and requires successive rounds of selective growth and purification of the recombinant virus, which is time consuming and technically

difficult. Second, the generation of the immune response against the gene of interest can be curtailed if there are circulating specific antibodies against the viral vector. Third, the expression of the viral vector-specific genes trigger a strong cell mediated response that rapidly eliminates the infected cells out of the cell pool, which diminishes the time for expression of recombinant genes, and may reduce the time for immunologic processing and response.

In the last decade, extensive research work has been performed aiming at the use of non-pathogenic, replication defective parvoviruses for the purpose of gene delivery: the adeno associated viruses. These viruses are infective for a wide variety of tissues both *in vivo* and *in vitro* and are capable of establishing latent infections by site-specifically integrating into the host's genome. Furthermore, these viruses can accommodate relatively large pieces of genetic information due to the fact that most of the genome of the virus is expendable for recombinant virus generation. Research has also proven that the absence of production of viral vector-derived proteins (since no virus-derived coding region are usually included in recombinant viruses) effectively diminishes the host's cell mediated immune response and apoptotic pathways that usually eliminate cells transduced by other types of viral vectors, thus extending the time of expression of the gene of interest for many years in some animal models. Another advantage afforded by this viral vector system is that of site specific integration. This site specific integration occurs without insertional activation of surrounding genes, which is one of the main disadvantages of gene delivery using retroviral-based delivery systems. Lastly, one of the major advantages is that these viruses are non-pathogenic, which greatly increases the safety of the procedure of gene delivery. The adeno associated viruses belong to the Dependovirus genus of the *Parvovirinae* subfamily of the parvovirus family of viruses. As mentioned above, these are non-pathogenic, replication defective ssDNA viruses that require the co-infection with a helper

virus (adeno or herpesvirus) to complete their own replication cycle. These viruses are widely distributed in vertebrate species, including mammals and avians. Specifically, in the case of the domestic chicken, two strains of this virus have been described. The avian adeno associated (AAAV) VR-865 strain was the first strain described by Yates *et al* in the 1970's, and was found as a contaminant of adenovirus isolates obtained from quails diagnosed with quail bronchitis. A second strain of the virus, the AAAV DA-1, was described by Hess *et al* in later years. These strains of the virus present genomic differences, as demonstrated by the different restriction patterns observed when the viral DNA's are digested with restriction enzymes. In spite of these genomic differences, the viruses are regarded as belonging to the same serotype, due to the high degree of cross neutralization observed by virus strain-specific antibodies in virus neutralization experiments.

Objectives and Originality

The overall objective of this research work is the generation and use of an avian adeno associated virus-based system for gene delivery. Ultimately, this recombinant viral-vectored system is to be used for the purpose of delivering genes coding for immunogenic peptides derived from economically important pathogenic avian viruses, aiming at the induction of protective immune responses. Justification for the use of this novel technology in the context of the poultry industry can be derived from the research done in the development of gene delivery systems for humans. The extensive body of research performed has demonstrated that this family of viruses is exquisitely suited for this task. It has been demonstrated, for example, that genes delivered by recombinant adeno associated viruses are expressed in high levels, and for long

periods of time. It has also been demonstrated that these viruses have none of the disadvantages attributed to other viral-based systems.

The use of the avian adeno-associated virus is a novel approach for the generation of a recombinant system for gene delivery in poultry that may offer distinct advantages over existing recombinant technologies currently used in the industry.

CHAPTER II

LITERATURE REVIEW

The Parvoviruses

The Parvoviruses are among the smallest of the DNA animal viruses. This family of viruses is composed of two subfamilies: the *Parvovirinae*, which infects vertebrates, and the *Densovirinae*, which infects insects. Each of these subfamilies has three genera: the *Parvovirus*, *Erythrovirus* and the *Dependovirus* are in the *Parvovirinae* subfamily, while the *Densovirus*, *Contravirus* and *Iteravirus* are in the *Densovirinae* subfamily (Muzyczka and Berns, 2001). The *Parvovirinae* have a wide distribution in warm-blooded animals, ranging from fowls to humans, and the Dependoviruses are unique among animal viruses because, except under special conditions, they require a co-infection with an unrelated helper virus, namely an adenovirus or herpesvirus (Muzyczka and Berns, 2001). The parvovirus virions have a relatively simple structure, composed of a single-stranded DNA molecule encapsidated in a mixture of structural proteins, named VP1, VP2 and VP3, with an icosahedral symmetry of a T=1 arrangement. The major structural protein is the VP2 (90% of viral particle), which main structural motif is an eight-stranded antiparallel β -barrels, connected by large loops of amino acid chains. These loops form most of the capsid's surface and have important biological functions, such as host species and tissue tropism, receptor binding and antigenic properties (Muzyczka and Berns, 2001).

The viral particle has a molecular weight (MW) of 5.5 to 6.2 x 10⁶ daltons. Approximately 50% of the mass is protein, and the remainder is DNA. Because of the relatively high DNA-to-protein ratio, the buoyant density of the intact virion in cesium chloride (CsCl) is

1.39 to 1.42 g/cm³ (Berns et al., 1994). The heavy buoyant density in CsCl permits the ready separation of AAV from helper adenovirus (Ad) in co-infections. Both heavier and lighter forms of the virion occur (Berns et al., 1994). The latter are particles containing DNA molecules with significant deletions, and these particles can function as defective interfering (DI) particles. The exact role of the heavier-than-normal particles in infections is unknown. The encapsidated DNA molecules are indistinguishable from those of normal-density particles and so, presumably, some of the coat protein molecules are missing. Whether the missing proteins constitute a specific set is unknown. Finally, the sedimentation coefficient of the virion in neutral sucrose gradients is 110 to 122 (Wistuba et al., 1995).

Possibly as a consequence of its structural simplicity, the virion is extremely resistant to inactivation. It is stable between pH 3 and 9 and at 56°C for 60 min. The virus can be inactivated by formalin, β -propiolactone, hydroxylamine, and oxidizing agents (Berns et al., 1994).

Genetic map: The autonomous parvovirus genome contains two large open reading frames (ORFs). The first covers much of the left half of the genome and encodes two non structural (NS) proteins, NS1 and NS2, from alternately spliced mRNAs. Mutations within NS1 block viral replication and gene expression. Only NS1 is absolutely required for DNA replication. NS2-specific mutants are defective for capsid synthesis, gene expression, and DNA replication in murine cells but show variable phenotypes in human cells (Li and Rhode, 1991; Li and Rhode, 1993; Muzyczka and Berns, 2001). The second large ORF occupies much of the right half of the genome and encodes the capsid proteins (Rhode, 1985; Rhode and Paradiso, 1983). Up to three capsid proteins have been detected in the virion. In the case of the minute virus of mice (MVM), the smallest capsid protein, VP3, is generated in the intact capsid by proteolytic cleavage of VP2. The amino acid sequences of VP1 and VP2 are identical except for additional amino acids at the

N terminus of VP1, and they are synthesized from two alternatively spliced messages. The VP1 message contains a second (smaller) ORF at the 5' end of the major right-hand ORF. Mutants altered in either the NS or coat protein ORFs can be complemented *in trans*. However, the palindromic sequences at both termini are required in cis for DNA replication and packaging (Muzyczka and Berns, 2001).

The *dependovirus* genetic map has been derived primarily from studies of the adeno-associated virus (AAV) type 2 but is highly conserved among all of the AAV serotypes (Hermonat et al., 1984; Muzyczka and Berns, 2001). There is a large ORF (cap) on the right side of the genome, which encodes the three coat proteins of the virus. Frameshift and deletion mutants within this region do not block DNA replication, but the accumulation of progeny single strands is inhibited, presumably because this requires encapsidation (Hermonat and Muzyczka, 1984). Mutations in the N-terminal region of the cap ORF, which affect VP1 exclusively, package DNA but produce virus particles that have lower infectivity. There is also a large ORF in the left half of the genome, which has been called the rep region because any frameshift mutation or significant deletion within the region blocks DNA replication (Hermonat and Muzyczka, 1984). These two ORFs are flanked by inverted terminal repeats (TR) that are palindromic in nature. At least four proteins have been detected (Rep78, Rep68, Rep52, and Rep40), which correspond to the four mRNAs that have been mapped to this region. In a productive infection, six 3' co terminal polyadenylated and capped mRNAs can be detected by Northern blotting. These mRNAs are synthesized by the host's RNA polymerase II from three different promoters designated p5, p19 and p40. The two largest transcripts are derived from p5 (4.2 and 3.9 kb in length), the next two largest transcripts are synthesized from p19 (3.6 and 3.3 kb), while the two smallest RNAs (2.6 and 2.3 kb) are produced from p40 (Muzyczka and Berns,

2001). It is possible to selectively eliminate Rep78 and 68 (from the p5 transcripts), and this type of mutation completely blocks all AAV-directed transcript accumulation that can be detected by Northern analysis, and it completely eliminates DNA replication (Hermonat et al., 1984; Hermonat and Muzyczka, 1984).

The larger Rep proteins play a critical regulatory role in every phase of the AAV life cycle. Under non-permissive conditions (absence of a helper virus) Rep68/78 negatively regulates AAV gene expression and DNA replication and is required for site-specific integration in the host cell genome to establish a latent infection (Muzyczka and Berns, 2001). In the presence of helper, Rep68/78 is a transactivator of AAV gene expression (Labow et al., 1986) and is essential for DNA replication and rescue of the viral genome from the integrated state (Hermonat et al., 1984; Labow et al., 1986). Mutants that are defective for the synthesis of the two smaller Rep proteins (Rep52 and 40 from the p19 promoter) have been made by changing the initiator AUG for these proteins to GGG (Muzyczka and Berns, 2001). The resulting mutant is able to replicate DNA, but no mature single-stranded DNA is encapsidated. The phenotype is thus similar to a capsid protein mutant. Other smaller ORFs exist near the middle of the genome, but the production of proteins corresponding to these regions is less clear.

Mutations in either the rep or cap ORFs can be complemented *in trans*. In contrast, the inverted TRs of 145 bases are required in cis for both DNA replication and transcription (Beaton et al., 1989; Senapathy et al., 1984). In addition to these functions, the TR is required for encapsidation (McLaughlin et al., 1988), integration of the genome during the establishment of a latent infection (Yang et al., 1997), and rescue of the genome from the integrated state (Samulski et al., 1983).

Capsid Proteins: Both autonomous parvovirus and replication defective virions contain three capsid proteins, with the exception of Aleutian disease virus (ADV), B19, and simian parvovirus, which contain only two coat proteins (Berns et al., 1994). For different species of autonomous parvoviruses, the proteins have approximate MWs of 80,000 to 86,000 (VP1), 64,000 to 75,000 (VP2), and 60,000 to 62,000 (VP3). Those of lapine parvovirus are significantly larger: 96,000 (VP1), 85,000 (VP2), and 75,000 (VP3). Except for ADV, VP3 is the major coat protein, representing 80% to 90% of the total mass. Some virus preparations are lacking in VP3, and its abundance appears to depend on the time during infection when the virions are isolated. All AAV preparations characterized to date have three coat proteins: VP1, 87,000; VP2, 73,000; and VP3, 62,000 (Muzyczka and Berns, 2001). In neither genus is there any evidence for glycosylation of any of the coat proteins; however, all parvovirus proteins appear to be phosphorylated. The N terminus is blocked in the case of the AAV VP3 protein (Muzyczka and Berns, 2001).

In addition to the fact that all three parvovirus capsid proteins appear to be coded for by overlapping in-frame DNA sequences, a further complexity has been reported. In the cases of both AAV1 and AAV2, the coat proteins VP1 and VP3 can be further subdivided by polyacrylamide gel electrophoresis into several subspecies (McPherson et al., 1982). The molecular basis for the difference in mobility is unknown.

DNA Replication: AAV DNA replication occurs via a single-strand displacement mechanism (Hauswirth and Berns, 1977). No evidence has been found for RNA primers or for the equivalent of Okazaki fragments (the presence of which would be indicative of lagging strand synthesis). Because all known DNA polymerases have a requirement for a primer with an available 3' OH, in addition to a template, linear DNA genomes have had to evolve specialized

terminal sequences to allow them to maintain the 5' terminal sequences intact when the primer structure is resolved during replication. The basic model for parvovirus DNA replication (often called the rolling hairpin model) was first described by Straus et al. (Straus et al., 1976). As described earlier, the two ends of AAV are identical in sequence and palindromic. The palindromic inverted TR at the 3' end of either strand can form a hairpin to serve as a primer to initiate synthesis of the complementary strand (Lusby et al., 1980). This produces a linear duplex molecule in which the original 3' end is covalently closed in the hairpin configuration, also called a monomer turnaround form. The hairpin is nicked at a site near the end of the terminal palindrome (nucleotide 124 of the AAV sequence), and this produces a new 3'-hydroxyl primer that allows repair synthesis of the hairpin. This process is called terminal resolution (or strand transfer) and the site-specific and strand-specific nicking site is called the terminal resolution site (trs). The net result of terminal resolution is the first complete duplex molecule in which both ends are extended, the monomer extended form. Either end can then denature and reanneal to form a double hairpinned structure, or rabbit ears (a process sometimes called reinitiation), and the new hairpin formed at the 3' end initiates leading-strand displacement synthesis to generate a single-stranded genome (which is packaged) and a duplex genome covalently closed at one end. The process of terminal resolution and reinitiation is then repeated. Each time the cycle is completed, a new single-stranded progeny strand is generated, and the strand transfer process that occurs during terminal resolution inverts the TR. If a molecule covalently closed at one end is not resolved before completion of strand displacement from the other end, then a dimer molecule can be formed in which two genomes are present in an inverted orientation (head to head or tail to tail) with a single TR at their junction (Muzyczka and Berns, 2001).

The Adeno Associated Virus

One of the outstanding features of AAV replication in cell culture has been the requirement for co-infection of the cell by an unrelated helper virus. Either an adeno or a herpesvirus can supply complete helper functions for fully permissive AAV infection (Muzyczka and Berns, 2001). In the absence of helper virus, AAV can establish a latent infection in cell culture that involves integration of the viral genome into a unique site on human chromosome 19 (Cheung et al., 1980). Superinfection of cell lines carrying a provirus with Ad or herpes will rescue the integrated genome and initiate a fully productive infection (Cheung et al., 1980). In light of this, it has been assumed that AAV uses latency in chromosome 19 as a strategy to survive in the absence of a helper.

Natural Infections: The situation during natural infections is not as clear. *In vivo* (i.e., in human infections), AAV has typically been found as a contaminant of Ad isolates. It is generally assumed, therefore, that natural AAV infections occur via the respiratory or gastrointestinal route as is the case for Ad. However, as yet, it is not clear what tissue or organ is a preferred site of latency in humans. Efforts by several groups to demonstrate latency in human lung samples have failed to detect AAV. Adeno-associated virus has been recovered from a small percentage (1% to 5%) of hematopoietic cells and from the genital tract of female patients suspected of herpes infections (Muzyczka and Berns, 2001). One report suggested that AAV was present in some first-trimester abortion material, but this has not been confirmed (Muzyczka and Berns, 2001). Recently, Dutheil et al. (Dutheil et al., 2000) demonstrated that the chromosome 19 integration site that is preferred by AAV is linked to the slow skeletal muscle-specific gene troponin T1 (TNNT1), and the cardiac troponin I gene (TNNI3). Several proviral cell lines established *in*

vitro were shown to contain rearrangements of the TNNT1 gene due to integration. These observations have led to the hypothesis that the natural site of AAV latency in humans might be skeletal muscle. Skeletal muscle is known to be resistant to Ad and herpes infection when these viruses are administered intravascularly, thus muscle could be a reservoir that is protected from rescue by AAV helper viruses. Because muscle fibers are multinucleated, integration of AAV and the possible disruption of the TNNT1 locus would produce minimal effects on the host. However, so far, no one has demonstrated that AAV is integrated into the chromosome 19 target in human samples.

The role of the Ad helper functions seems in part to be the induction of the appropriate cellular milieu required for AAV DNA replication. The notion that AAV replication is possible once the appropriate cellular genes have been induced suggests that the appropriate cellular milieu could be established in the absence of helper virus. Indeed, several groups have reported that pretreatment of transformed or tumor cell lines with a variety of genotoxic agents can render the cells semi permissive for AAV replication in the absence of helper virus co-infection (Muzyczka and Berns, 2001). Agents that have been successfully used include ultraviolet (UV) irradiation, cycloheximide, hydroxyurea, aphidicolin, topoisomerase inhibitors, and several chemical carcinogens.

Helper Functions: Adenoviruses (Ad), herpes simplex virus (HSV) types I and II, cytomegalovirus (CMV), and pseudorabies virus all serve as complete helpers for AAV replication (Muzyczka and Berns, 2001). The AAV host range is identical to the normal host range for the helper virus. Genetic analysis of helper functions has been most extensive for Ad. Many of the identified Ad early functions serve as helper functions for AAV replication, but no Ad late functions have been found to be necessary. Four Ad proteins have been shown to be

required for complete helper function: the early region IA (EIA) transactivator protein, the EIB 55-kd protein, the E4 34-kd protein, and the E2A DNA-binding protein. Additionally, synthesis of the Ad virus-associated (VA) RNAs is required. Ad EIA function is required for the other Ad early regions to be transcribed (Muzyczka, 1992). Similarly, an EIA function is required for AAV transcripts to be detected by Northern blotting (Janik et al., 1981). Two EIA proteins with overlapping sequences of 289 and 243 amino acids, respectively, have been identified. The former can both activate and inhibit gene expression *in trans*, whereas the latter primarily inhibits gene expression. The 289-amino-acid EIA protein is responsible for transactivation of AAV gene expression (Chang et al., 1989). EIA, in concert with other Ad early genes, also induces a variety of changes in cellular gene expression. The most important may be that EIA induces cells to enter S phase and to synthesize cellular DNA replication proteins that are likely to be needed for AAV DNA replication (Muzyczka, 1992).

E4 was originally identified as encoding a helper function specifically required for AAV DNA replication. The E4 product involved is a 34-kd protein that is the product of the E4 ORF 6 gene and it can form a complex with the 55-kd EIB protein during a productive Ad infection. Both the 55-kd and the 34-kd proteins have been shown to regulate the expression of AAV genes, possibly at the level of transport of mRNA to the cytoplasm. In a general sense, EIB has also been identified as being required for AAV DNA replication to occur, but it does not seem to have a consistent effect on AAV transcript accumulation (Muzyczka, 1992; Muzyczka and Berns, 2001). More recently, studies of transduction mediated by recombinant AAVs (rAAVs) that carry marker genes have suggested that the E4 ORF 6 protein enhances the conversion of single-stranded input viral genomes to a duplex form by promoting second-strand synthesis (Ferrari et al., 1996).

E2A encodes a 72-kd single-strand DNA binding protein (DBP) that is required for Ad DNA synthesis but that does not appear to be required for AAV DNA replication, although particle formation is greatly inhibited by certain E2A gene mutations. DBP stimulates transcription from AAV promoters (Muzyczka and Berns, 2001) and may be involved in mRNA transport and stability. In spite of the absence of genetic evidence that Ad DBP is required for AAV DNA replication, recent *in vitro* studies of AAV DNA replication using cell-free extracts suggest that Ad DBP is preferentially used for AAV DNA replication in place of the human single-strand DBP, replication protein A (RPA). In contrast, the E2B region, which produces the Ad terminal protein, and the Ad DNA polymerase, both of which are directly involved in the process of Ad DNA replication, is not required for AAV replication. Finally, Ad VAI RNA has been reported to facilitate the initiation of AAV protein synthesis by preventing the interferon-induced host-cell shutoff of translation (Muzyczka and Berns, 2001).

Two detailed studies on the HSV-1 genes required for productive AAV infection have been reported, although the conclusions were somewhat different. The first study identified the following HSV proteins as providing helper functions for AAV: ICP4 transactivator, DNA polymerase, ICP8 single-strand DBP, the origin-binding protein, and two of the three subunits of the helicase–primase complex (UL5, UL8). The second study identified ICP8 and all three components of the helicase–primase complex as helper functions. It should be noted that the latter study used a heterologous promoter; normally, ICP4 function is required for expression of HSV replication genes. The former study found that maximal AAV replication required all the genes identified but that only ICP4 was essential for AAV replication per se. It is not clear why with Ad, most of the viral DNA replication genes are not required, but with herpes, most of the helper functions are replication enzymes (Muzyczka and Berns, 2001).

Latent Infection: In the absence of a helper virus co-infection, the AAV genome is often integrated into cellular DNA to establish a latent infection that can be activated by a subsequent helper virus infection. AAV latent infection was discovered by Hoggan et al. in the course of a project to screen primary cell lots intended for vaccine production. Although there was no immunologic evidence for AAV, upon challenge by Ad infection, 20% of the lots of African green monkey kidney cells and 1% to 2% of the lots of human embryonic kidney cells tested produced AAV (Muzyczka and Berns, 2001). Thus, AAV latent infection *in vivo* appeared to be common. Although most human transmission of AAV appears to be horizontal, evidence for vertical germline transmission of avian AAV has been reported in chickens (Muzyczka and Berns, 2001). Human cells in culture can be latently infected simply by infecting with a high multiplicity of AAV (multiplicity of infection of 10 to 1,000) in the absence of helper virus (Hermonat and Muzyczka, 1984). The frequency of positive clones (estimated either by isolation of single-cell colonies or by selection with recombinant vectors carrying a *rep* gene and a selectable marker) suggests that up to 10% of the infected cells become latently infected (Hermonat and Muzyczka, 1984). These figures exceed those seen with other DNA animal viruses and approach the efficiency of lambda bacteriophage lysogeny and rescue. Characterization of AAV DNA in latently infected clones shows that the viral genome is in most cases integrated into cellular DNA (Cheung et al., 1980). The junction with cellular sequences is usually within the TR, but the site of integration within the viral TR is not unique. Recombination junctions occur throughout the TR sequences and the p5 promoter region, and portions of the TR are deleted. The viral DNA in most latently infected cell lines is integrated as a tandem (head-to-tail) array of several genome equivalents. Inverted (tail-to-tail or head-to-head) arrays have also been seen but occur at a much lower frequency. The presence of TR

deletions raises the question of how the AAV provirus is rescued. Presumably this occurs by excision of an internal tandemly repeated genome that is intact (Muzyczka and Berns, 2001).

It is not known whether deletions of the TR occur during integration or during subsequent passage of the cells. In one clone of latently infected cells that was followed for over 100 passages, the restriction patterns obtained, using enzymes that cut in the unique sequences, were unchanged, but significant changes were noted using a restriction enzyme that cuts exclusively in the TR (Cheung et al., 1980). This raises the possibility that the TR is an unstable sequence. This notion was supported by the observation that after 100 passages, free copies of the AAV genome were present, even though all of the detectable sequences had been integrated into high-MW DNA in earlier passages.

Genomic Integration: Initially, it was concluded that integration was random with respect to cellular sequences because restriction analysis found that junction fragments of viral and cellular DNA differed in size in every independently derived clone carrying a provirus. It is now clear that integration occurs at a specific site on human chromosome 19q13.3-qter. Integration is by nonhomologous recombination, although there are four to five base homologies at the site of recombination, and integration is often associated with rearrangements and inversions of both viral DNA and cellular sequences. Although integration is site specific, it is not specific at the individual nucleotide level but occurs within a region of several hundred nucleotides. Additionally, only about 70% of the integration events appear to occur on chromosome 19 (Muzyczka and Berns, 2001).

Recombinant AAV (rAAV) vectors carrying a selectable marker flanked by two TRs, but deleted for the *rep* gene, do not integrate into chromosome 19 (Kearns et al., 1996), whereas those that contain the *rep* gene integrate site specifically. Plasmids carrying an rAAV genome,

when co-transfected with a Rep-expressing plasmid (or Rep protein) into human cells, will excise the rAAV genome and integrate it into chromosome 19 (Shelling and Smith, 1994). The same occurs when human cells are infected with herpes, Ad, or baculovirus vectors carrying an rAAV genome, provided Rep is expressed (Recchia et al., 1999). Most recently, site-specific integration has been demonstrated in nonhuman cells. Mice do not have the AAVS1 preintegration sequence. However, when cells from transgenic mice carrying the AAVS1 locus are infected with *rep*⁺ AAV, integration occurs in the transgenic S1 locus (Rizzuto et al., 1999).

The Avian Adeno Associated Virus

The Avian Adeno-Associated Virus (AAAV) was first isolated as a contaminant of isolates of the Olson strain of quail bronchitis virus (Yates et al., 1973). As with every member of the *dependovirus* genus, this viruses require a coinfection with a helper virus (an adeno or herpesvirus) to complete its own replication (Muzyczka and Berns, 2001). It has been postulated that the AAAV can exist as a latent infection in the germ line of chickens. In that study, chicken kidney cells derived from the eggs of white leghorn chickens that had serological evidence of prior exposure to both adenovirus and the AAAV produced AAAV antigenic proteins, upon challenge with purified adenovirus (Sadasiv et al., 1989). Similar results were obtained using cells prepared from specific-pathogen-free chicks with no previous exposure to adenovirus or AAAV (Sadasiv et al., 1989). Another study that confirms these findings was performed by Dawson and Yates in 1982. It was demonstrated that AAAV could be consistently recovered from limiting dilutions of purified and unpurified avian Ad stocks propagated in embryonating chicken eggs derived from two independently raised flocks of White Leghorn (WL) chickens but not when these Ad stocks were propagated in duck cells. From these observations it was

concluded that AAV is a latent endogenous virus of at least some flocks of WL chickens (Dawson et al., 1982)

Interestingly, the AAV has been found to induce a dose-dependent inhibition of the mortality pattern caused by the Tipton strain of avian adenovirus in young birds, when the two viruses are inoculated simultaneously in these hosts (Pronovost et al., 1978). There is also evidence that the AAV is able to reduce the growth of the Marek's disease virus when co infected at high multiplicity of infection in cell culture (Bauer and Monreal, 1988).

Studies have demonstrated that there is genomic variability in AAV isolates. Two strains of AAV with very distinct DNA patterns upon restriction enzymes analysis have been found. Although the DNAs of both viruses characterized in this study had the same size, most restriction enzymes produced different restriction patterns. The different restriction sites were mainly located in the middle and right part of the genomes and the effect of these differences on the structure proteins was shown by western blot analysis. The immunofluorescence and immunodiffusion test showed that the two *dependoviruses* were serologically indistinguishable and therefore were regarded as two different strains of the same virus. These two strains were named as AAV VR-865 (which was the first described isolate) and isolated AAV DA-1 (Hess et al., 1995).

Of utmost interest is the fact that antibodies against these viruses have been found in humans. In a study, six of 100 serum samples from an unselected adult population were positive for antibody to AAV by agar gel precipitation (AGP), and 10 of 64 were positive by virus neutralization (VN). Three of 21 samples from poultry workers, from industry and research institutions, were positive for AAV antibody by AGP and 14 of 21 were positive by VN. Also, no cross reaction was noted by AGP when antiserum to avian AAV was reacted against

primate antigens of serotypes 1-4 or when antiserum to AAV serotypes 1-4 were reacted against AAV antigen. Antiserum prepared against primate AAV serotypes 1-4 did not neutralize the AAV. These results suggest that AAV infections are not restricted to avian species but are found in the human adult population (Yates et al., 1981). These observations also suggest the possible role of the AAV as a vector for human use.

To this date, no DNA sequence data, receptor specificity or viral particle distribution in the infected host information has been obtained for this virus.

Adeno-Associated Virus as Viral Vectors

So far, eight serotypes of AAV have been found in primates (AAV-1 to AAV-8). Type 2 is the best-characterized primate serotype, and it was the first AAV used for the development of vectors for gene transfer (McLaughlin et al., 1988). Most of the vectors currently in use are derived from AAV-2, but vectors based on AAV-1 have been recently described (Xiao et al., 1999).

In contrast to retro and lentiviral vectors, in which vector development not only resulted in different versions but also achieved significant improvements of transduction efficacy and biosafety, the original design of AAV vectors has not changed over the last decade. The AAV2-based vectors contain only the left and right ITRs, and 139 or 45 (McLaughlin et al., 1988) nucleotides of non-repeated AAV sequences adjacent to the right terminal repeat, respectively. Samulski et al. (Samulski et al., 1983) were able to show that all the cis-acting AAV functions required for replication and virion production are located within the ITR and the immediately adjacent 45 nucleotides. In the AAV-based vectors, the two viral open reading frames (ORFs), which code for the capsid proteins (VP1, 2, and 3) and the four nonstructural Rep proteins

(Rep78/68 and Rep52/40) are replaced with the transgene of interest and its promoter and then transfected into the producer cells, where the viral genes necessary for virus production and packaging of the vector genome are provided in trans by packaging plasmids and helper viruses/plasmids.

One disadvantage of AAV-based vectors is their size limitation: The optimal size for AAV vectors lies between 4.1 and 4.9 kb (Dong et al., 1996). Although larger genomes can be packaged, the packaging efficacy is sharply reduced and the maximal size of the vector genome (including the two ITRs) is 5.2 kb (Dong et al., 1996). Because of this size limitation, a dual vector, trans-splicing AAV vector system was developed (Nakai et al., 2000), which allows the transduction of target cells and tissues with two AAV vectors that contain a split expression cassette. This approach to AAV-based gene transfer is based on the finding that episomal recombinant AAV (rAAV) genomes form circular multimers that concatamerize in a head-to-tail orientation. The dual-vector strategy functions as follows: After co-infection of the target cells, the two AAV vectors that carry the split genome form head-to-tail heterodimers through sequence homology of the ITRs, thereby rejoining the split gene into one continuous DNA molecule. Although such rejoining leads to the disruption of the gene-coding sequence by the AAV ITR, this problem can be overcome by inclusion of eukaryotic splicing signals to remove the ITR sequences during RNA maturation. With this novel dual-vector approach, the *E. coli lacZ* transgene, a luciferase reporter gene, and the human erythropoietin gene have been successfully delivered *in vitro* and *in vivo* (to muscle and liver of the mouse) (Muzyczka and Berns, 2001). Although the efficacy of gene delivery in the dual-vector system was consistently reported to be lower than in the unsplit, conventional AAV vector system, the individual efficacies vary between 16% and 70% as compared to unsplit vectors (Sun et al., 2000).

AAV-Vector Packaging Strategies: The basic approach of the AAV packaging system is the same as in other vectors: The vector plasmid contains the required cis-acting viral sequences necessary for packaging and integration and the transgene, while the packaging plasmids provide the viral components required for virus production in *Trans*. However, because AAV is a dependovirus, its propagation depends on the functions provided by a helper virus (adenovirus or herpesvirus). Although replication has been described to occur in a helper-independent fashion if the host cells are stressed with genotoxic agents such as irradiation, adenovirus infection is still the most efficient way to induce AAV replication (Muzyczka and Berns, 2001).

The initial studies with AAV vectors revealed that both the AAV *rep* and *cap* genes can be provided in trans in producer cells, and this resulted in the first generation of packaging systems (McLaughlin et al., 1988). In this system, the packaging plasmid, named pAAV/Ad (Samulski et al., 1989), encodes the *rep* and *cap* genes, and the essential helper functions are provided by the adenovirus. Transfection of the AAV packaging construct together with the vector plasmid into adenovirus-infected packaging/helper cells results in excision of the ITR-flanked transgene from the vector plasmid, followed by its amplification and packaging into AAV capsids. This first generation of AAV packaging yielded up to 10² rAAV particles per cell, but there is a major disadvantage: the co-production of wild-type adenovirus along with the AAV vector particles. Although these AAV preparations can be purified by centrifugation in multiple cesium chloride (CsCl) gradients and/or the adenovirus contamination can be inactivated by heating the preparations for 30 minutes at 56°C, a complete removal of adenoviral capsid components is not always possible. In addition, this may lead to a loss in the activity of the AAV particles. Given the fact that viral proteins derived from inactivated adenovirus preparations

elicit a cytotoxic T lymphocyte (CTL) immune response at the site of injection, it is clear that this packaging system is suboptimal, at least for clinical use (Kafri et al., 1998).

To overcome this problem, a packaging system was developed that is free of adenovirus (Xiao et al., 1998). The adenoviral genes that provide essential helper functions during the productive life cycle of AAV are the E1A, E1B, E2A, E4, and VA RNA genes (Muzyczka, 1992).

E1A functions as a transactivator of AAV *rep* and *cap* transcription. Interaction of E1B with E4 stabilizes AAV mRNAs and/or facilitates the mRNA transport to the cytoplasm. A similar role has been described for E2A and VA RNA. By cloning of the E2A, E4, and VA RNA genes into a high-copy plasmid, Xiao et al. (Xiao et al., 1998) generated a helper plasmid (pXX6) that contains all the essential helper genes but lacks the adenovirus structural and replication genes and the adenoviral terminal repeats. The missing E1 gene products are provided by using 293 producer cells (Graham et al., 1977). Surprisingly, supplying the adenoviral helper functions from a plasmid that resulted from deleting most of the adenoviral genome significantly increased AAV vector production (Xiao et al., 1998). This is presumably the result of the lack of competition between the AAV and adenovirus for adenoviral gene products in the packaging cells, in addition to the removal of the lytic action of the helper virus.

A further, major improvement of AAV vector production was the development of a second generation of packaging constructs: Because *rep* expression has been shown to cause a paradoxical inhibition of rAAV production if the Rep68/78 proteins are overexpressed in the producer cells, the ATG translation start codon of the *rep* gene was mutated into an inefficient ACG codon to achieve attenuation of Rep68/78 synthesis. Because the p5 promoter exerts positive effects on p19 and p40 but negatively regulates its own transcription, Xiao et al. (Xiao et

al., 1998) reasoned that inclusion of a second copy of the p5 promoter/enhancer element 3' of the Cap gene might have positive effects on AAV vector production. Indeed, these changes increase vector yields approximately 15-fold over the conventional packaging plasmid pAAV/Ad. Thus, co-transfection of 293 cells with this second-generation packaging system, containing three plasmids (the vector, the pXX2 packaging plasmid, and the pXX6 helper construct) results in viral titers of up to 10^{10} /10-cm plate (Xiao et al., 1998).

A problem with this packaging system is that the adenoviral terminal repeats present in pAAV/Ad and pXX2 have been shown to be involved in recombination with the AAV ITRs. This recombination between the AAV vector and the packaging plasmids can lead to the production of wild-type AAV-2 particles in the range of 4% of the total physical titer. However, removal of the distal 10 nucleotides of the AAV vector ITRs reduced the amount of wild-type AAV contaminants to below the detection level, when a packaging plasmid was used that lacked adenoviral ITRs (Wang et al., 1998). The elimination of replication-competent wild-type AAV and helper virus (e.g., adenovirus) contaminations from the rAAV preparations is an important step toward the generation of clinical-grade vectors.

Packaging Cell Lines for AAV: Adeno-associated-virus vector production in stable cell lines has to overcome several problems associated with the toxicity of viral proteins and the viral life cycle. The Rep proteins exert an antiproliferative effect on mammalian cells, and they have a negative effect on heterologous gene expression (Muzyczka and Berns, 2001). Therefore, it was not surprising that initial attempts to establish a stable cell line containing AAV-2 *rep* and *cap* were not very successful. The initial AAV cell lines gave titers of 30 IU/cell, whereas transient transfection of the second generation of packaging/helper system yields up to 1,000 IU/cell (Xiao et al., 1998). Recently, a number of promising inducible cell lines have been developed, in which

the expression of Rep and Cap proteins is regulated by the tetracycline-inducible system.

Infection of the cell line with adenovirus after the addition of tetracycline or doxycycline to the medium resulted in AAV vector production with titers 10 times higher (i.e., 10,000 IU/cell) than those obtained by transient transfection. However, after 3 to 4 days in doxycycline, the number of particles per cell dropped dramatically (Inoue and Russell, 1998).

A major drawback of the presently available packaging cell lines is the requirement for the infection with a helper virus, and an extensive purification procedure has to be applied to minimize the amount of helper virus contamination. In addition, an important problem with stable cell lines for AAV vector production is that expression of the helper viral genes leads to the excision of the vector genome from the integration site, and thus prolonged expression of helper virus genes might disturb the genomic stability of the packaging/producer cells (Muzyczka and Berns, 2001).

Integration versus Episomal Persistence of rAAV: Site-specific integration would be a highly desirable feature of a gene therapy vector, as it would ensure long-term (theoretically even life-long) expression of the transgene and minimize the mutagenic and carcinogenic potential—major drawbacks of random integrating vectors based on retroviruses. Although the ITRs are the only genomic elements necessary for integration, efficient integration and site specificity require the presence of the viral Rep protein (Weitzman et al., 1994). Because AAV vectors lack the Rep expression cassette, it comes as no surprise that AAV vectors integrate with low efficacy and low specificity into the host genome. Integration of rAAV has been observed in dividing HeLa cells (Yang et al., 1997), as well as in nondividing human neuronal cell lines *in vitro* (Wu et al., 1998).

Apart from integration into the host genome, the presence of episomal forms of rAAV has been demonstrated *in vivo*. Several studies demonstrated the presence of AAV vector DNA in an episomal circular form in muscle and brain tissue transduced with AAV vectors. Because this form of episomal AAV DNA persists for up to 9 months, the episomal AAV DNA might be a major contributor to the long-term expression of transgenes delivered with AAV vectors (Duan et al., 1998). The characteristic feature of these circular genomes is that they contain two viral ITRs in a head-to-tail orientation, which appears to confer an increased persistence of episomal DNA at least in certain cell types (e.g., HeLa cells) (Duan et al., 1998). Presently, not much is known about the exact mechanisms that lead to the formation of these concatamers, but it appears to involve recombination events between separate viral genomes. This inherent ability of AAV to form concatamers fusing the 5' and 3' ends of two genomes of two viruses is the basis for the development of trans-splicing, dual-vector systems (Duan et al., 2000).

In conclusion, completion of second-strand synthesis appears to be the rate-limiting step of rAAV infection (Monahan and Samulski, 2000), and it is associated with an increase in transgene expression, which can take several weeks after administration *in vivo*. Long-term (i.e., several months) transgene expression could be achieved by integrated concatamers, or by stable concatameric episomes, or both (Muzyczka and Berns, 2001).

References

- Bauer, H.J. and Monreal, G. (1988) Avian adeno-associated parvovirus and Marek's disease virus: studies of viral interactions in chicken embryo fibroblasts. Brief report. Arch. Virol. 98(3-4), 271-7.
- Beaton, A., Palumbo, P. and Berns, K.I. (1989) Expression from the adeno-associated virus p5 and p19 promoters is negatively regulated in trans by the rep protein. J. Virol. 63(10), 4450-4.
- Berns, K.I., Bergoin, M. and Bloom, M. (1994) Parvoviridae. VIth report of International Committee on Taxonomy of Viruses. In: F.A. Murphy, C.M. Fauquet and D.H.L. Bishop (Eds), Virus Taxonomy, pp. 166-185. Springer-Verlag, Vienna.
- Chang, L.S., Shi, Y. and Shenk, T. (1989) Adeno-associated virus P5 promoter contains an adenovirus E1A-inducible element and a binding site for the major late transcription factor. J. Virol. 63(8), 3479-88.
- Cheung, A.K., Hoggan, M.D., Hauswirth, W.W. and Berns, K.I. (1980) Integration of the adeno-associated virus genome into cellular DNA in latently infected human Detroit 6 cells. J. Virol. 33(2), 739-48.
- Dawson, G.J., Yates, V.J., Chang, P.W. and Oprandy, J.J. (1982) Is avian adeno-associated virus an endogenous virus of chicken cells? Nature 298(5874), 580-2.
- Dong, J.Y., Fan, P.D. and Frizzell, R.A. (1996) Quantitative analysis of the packaging capacity of recombinant adeno-associated virus. Hum. Gene. Ther. 7(17), 2101-12.
- Duan, D., Sharma, P., Yang, J., Yue, Y., Dudus, L., Zhang, Y., Fisher, K.J. and Engelhardt, J.F. (1998) Circular intermediates of recombinant adeno-associated virus have defined

- structural characteristics responsible for long-term episomal persistence in muscle tissue. J. Virol. 72(11), 8568-77.
- Duan, D., Yue, Y., Yan, Z. and Engelhardt, J.F. (2000) A new dual-vector approach to enhance recombinant adeno-associated virus-mediated gene expression through intermolecular cis activation. Nat. Med. 6(5), 595-8.
- Dutheil, N., Shi, F., Dupressoir, T. and Linden, R.M. (2000) Adeno-associated virus site-specifically integrates into a muscle-specific DNA region. Proc. Natl. Acad. Sci. USA 97(9), 4862-6.
- Ferrari, F.K., Samulski, T., Shenk, T. and Samulski, R.J. (1996) Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. J. Virol. 70(5), 3227-34.
- Graham, F.L., Smiley, J., Russell, W.C. and Nairn, R. (1977) Characteristics of a human cell line transformed by DNA from human adenovirus type 5. J. Gen. Virol. 36(1), 59-74.
- Hauswirth, W.W. and Berns, K.I. (1977) Origin and termination of adeno-associated virus DNA replication. Virology 78(2), 488-99.
- Hermonat, P.L., Labow, M.A., Wright, R., Berns, K.I. and Muzyczka, N. (1984) Genetics of adeno-associated virus: isolation and preliminary characterization of adeno-associated virus type 2 mutants. J. Virol. 51(2), 329-39.
- Hermonat, P.L. and Muzyczka, N. (1984) Use of adeno-associated virus as a mammalian DNA cloning vector: transduction of neomycin resistance into mammalian tissue culture cells. Proc. Natl. Acad. Sci. USA 81(20), 6466-70.
- Hess, M., Paul, G., Kling, S. and Monreal, G. (1995) Molecular characterization of two strains of the avian adeno-associated virus (AAAV). Arch. Virol. 140(3), 591-8.

- Inoue, N. and Russell, D.W. (1998) Packaging cells based on inducible gene amplification for the production of adeno-associated virus vectors. *J. Virol.* 72(9), 7024-31.
- Janik, J.E., Huston, M.M. and Rose, J.A. (1981) Locations of adenovirus genes required for the replication of adenovirus-associated virus. *Proc. Natl. Acad. Sci. USA* 78(3), 1925-9.
- Kafri, T., Morgan, D., Krah, T., Sarvetnick, N., Sherman, L. and Verma, I. (1998) Cellular immune response to adenoviral vector infected cells does not require de novo viral gene expression: implications for gene therapy. *Proc. Natl. Acad. Sci. USA* 95(19), 11377-82.
- Kearns, W.G., Afione, S.A., Fulmer, S.B., Pang, M.C., Erikson, D., Egan, M., Landrum, M.J., Flotte, T.R. and Cutting, G.R. (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(9), 748-55.
- Labow, M.A., Hermonat, P.L. and Berns, K.I. (1986) Positive and negative autoregulation of the adeno-associated virus type 2 genome. *J. Virol.* 60(1), 251-8.
- Li, X. and Rhode, S.L., 3rd. (1991) Nonstructural protein NS2 of parvovirus H-1 is required for efficient viral protein synthesis and virus production in rat cells in vivo and in vitro. *Virology* 184(1), 117-30.
- Li, X. and Rhode, S.L., 3rd. (1993) The parvovirus H-1 NS2 protein affects viral gene expression through sequences in the 3' untranslated region. *Virology* 194(1), 10-9.
- Lusby, E., Fife, K.H. and Berns, K.I. (1980) Nucleotide sequence of the inverted terminal repetition in adeno-associated virus DNA. *J. Virol.* 34(2), 402-9.
- McLaughlin, S.K., Collis, P., Hermonat, P.L. and Muzyczka, N. (1988) Adeno-associated virus general transduction vectors: analysis of proviral structures. *J. Virol.* 62(6), 1963-73.

- McPherson, R.A., Ginsberg, H.S. and Rose, J.A. (1982) Adeno-associated virus helper activity of adenovirus DNA binding protein. *J. Virol.* 44(2), 666-73.
- Monahan, P.E. and Samulski, R.J. (2000) AAV vectors: is clinical success on the horizon? *Gene Ther.* 7(1), 24-30.
- Muzyczka, N. (1992) Use of adeno-associated virus as a general transduction vector for mammalian cells. *Curr Top Microbiol. Immunol.* 158, 97-129.
- Muzyczka, N. and Berns, K.I. (2001) *Parvoviridae: The Viruses and Their Replication*. In: D.M. Knipe and P.M. Howley (Eds), *Fields Virology*, pp. 2327-2379. Vol. 2. 2 vols. Lippincott Williams & Wilkins, Philadelphia.
- Nakai, H., Storm, T.A. and Kay, M.A. (2000) Increasing the size of rAAV-mediated expression cassettes in vivo by intermolecular joining of two complementary vectors. *Nat. Biotechnol.* 18(5), 527-32.
- Pronovost, A.D., Yates, V.J. and Fry, D.E. (1978) Effect of avian adeno-associated virus on pathogenicity of Tipton virus in chicks. *Avian Dis.* 22(2), 354-7.
- Recchia, A., Parks, R.J., Lamartina, S., Toniatti, C., Pieroni, L., Palombo, F., Ciliberto, G., Graham, F.L., Cortese, R., La Monica, N. and Colloca, S. (1999) Site-specific integration mediated by a hybrid adenovirus/adeno-associated virus vector. *Proc. Natl. Acad. Sci. USA* 96(6), 2615-20.
- Rhode, S.L., 3rd. (1985) Nucleotide sequence of the coat protein gene of canine parvovirus. *J. Virol.* 54(2), 630-3.
- Rhode, S.L., 3rd and Paradiso, P.R. (1983) Parvovirus genome: nucleotide sequence of H-1 and mapping of its genes by hybrid-arrested translation. *J. Virol.* 45(1), 173-84.

- Rizzuto, G., Gorgoni, B., Cappelletti, M., Lazzaro, D., Gloaguen, I., Poli, V., Sgura, A., Cimini, D., Ciliberto, G., Cortese, R., Fattori, E. and La Monica, N. (1999) Development of animal models for adeno-associated virus site-specific integration. *J. Virol.* 73(3), 2517-26.
- Sadasiv, E.C., Piela, T.H. and Chang, P.W. (1989) Detection of latent avian adenovirus-associated virus proteins in chicken cells. *Avian Dis.* 33(1), 125-33.
- Samulski, R.J., Chang, L.S. and Shenk, T. (1989) Helper-free stocks of recombinant adeno-associated viruses: normal integration does not require viral gene expression. *J. Virol.* 63(9), 3822-8.
- Samulski, R.J., Srivastava, A., Berns, K.I. and Muzyczka, N. (1983) Rescue of adeno-associated virus from recombinant plasmids: gene correction within the terminal repeats of AAV. *Cell* 33(1), 135-43.
- Senapathy, P., Tratschin, J.D. and Carter, B.J. (1984) Replication of adeno-associated virus DNA. Complementation of naturally occurring rep- mutants by a wild-type genome or an ori- mutant and correction of terminal palindrome deletions. *J. Mol. Biol.* 179(1), 1-20.
- Shelling, A.N. and Smith, M.G. (1994) Targeted integration of transfected and infected adeno-associated virus vectors containing the neomycin resistance gene. *Gene. Ther.* 1(3), 165-9.
- Straus, S.E., Sebring, E.D. and Rose, J.A. (1976) Concatemers of alternating plus and minus strands are intermediates in adenovirus-associated virus DNA synthesis. *Proc. Natl. Acad. Sci. USA* 73(3), 742-6.
- Sun, L., Li, J. and Xiao, X. (2000) Overcoming adeno-associated virus vector size limitation through viral DNA heterodimerization. *Nat. Med.* 6(5), 599-602.

- Wang, X.S., Khuntirat, B., Qing, K., Ponnazhagan, S., Kube, D.M., Zhou, S., Dwarki, V.J. and Srivastava, A. (1998) Characterization of wild-type adeno-associated virus type 2-like particles generated during recombinant viral vector production and strategies for their elimination. *J. Virol.* 72(7), 5472-80.
- Weitzman, M.D., Kyostio, S.R., Kotin, R.M. and Owens, R.A. (1994) Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. *Proc. Natl. Acad. Sci. U S A* 91(13), 5808-12.
- Wistuba, A., Weger, S., Kern, A. and Kleinschmidt, J.A. (1995) Intermediates of adeno-associated virus type 2 assembly: identification of soluble complexes containing Rep and Cap proteins. *J. Virol.* 69(9), 5311-9.
- Wu, P., Phillips, M.I., Bui, J. and Terwilliger, E.F. (1998) Adeno-associated virus vector-mediated transgene integration into neurons and other nondividing cell targets. *J. Virol.* 72(7), 5919-26.
- Xiao, W., Chirmule, N., Berta, S.C., McCullough, B., Gao, G. and Wilson, J.M. (1999) Gene therapy vectors based on adeno-associated virus type 1. *J. Virol.* 73(5), 3994-4003.
- Xiao, X., Li, J. and Samulski, R.J. (1998) Production of High-Titer Recombinant Adeno-Associated Virus Vectors in the Absence of Helper Adenovirus. *J. Virol.* 72(3), 2224-2232.
- Yang, C.C., Xiao, X., Zhu, X., Ansardi, D.C., Epstein, N.D., Frey, M.R., Matera, A.G. and Samulski, R.J. (1997) Cellular recombination pathways and viral terminal repeat hairpin structures are sufficient for adeno-associated virus integration in vivo and in vitro. *J. Virol.* 71(12), 9231-47.

- Yates, V.J., Dawson, G.J. and Pronovost, A.D. (1981) Serologic evidence of avian adeno-associated virus infection in an unselected human population and among poultry workers. *Am. J. Epidemiol.* 113(5), 542-5.
- Yates, V.J., el-Mishad, A.M., McCormick, K.J. and Trentin, J.J. (1973) Isolation and characterization of an Avian adenovirus-associated virus. *Infect. Immun* 7(6), 973-80.

CHAPTER III

SEQUENCE ANALYSIS, VIRAL RESCUE FROM INFECTIOUS CLONES AND GENERATION OF RECOMBINANT VIRIONS OF THE AVIAN ADENO-ASSOCIATED VIRUS¹

¹C. Estevez and Villegas, P: Submitted to Virus Research, 2004.

SUMMARY: Aiming at the generation of a viral-vectored system for gene delivery and vaccination in poultry, the entire genomes of the VR-865 and DA-1 strains of the avian adeno-associated virus have been cloned and sequenced. Sequence analysis of the clones showed that the genomic distribution of the structural and non-structural protein-coding genes of these viruses is conserved and in agreement with what has been previously described for the primate adeno associated viruses. Amino acid differences between the avian adeno associated viruses and the primate adeno associated viruses are more evident in the genes that code for the non-structural (Rep) proteins of the virus, while the Cap region amino acid sequence was found to be more conserved. Since all the regulatory and coding sequences of the virus were present in the plasmids obtained, complete infectious viral particles were rescued from these clones, and these rescued viral populations were amplified by co-infecting primary embryo liver cells with the rescued virus and the CELO strain of the avian adenovirus type 1. As a proof of concept of the validity of this system for the purpose of gene delivery, recombinant viruses encoding for the *LacZ* gene as a reporter system were also generated. These recombinant viruses were used to express beta galactosidase activity in primary chicken embryo cell cultures.

Key words: recombinant viruses, vaccination, gene delivery, adeno associated virus, avian adeno associated virus.

INTRODUCTION:

Poultry production methods used today in the poultry industry have increased bird densities in the poultry houses and subsequently increased both the risk of spreading old pathogens and encouraging the emergence of new ones (Silva, 2002). Prevention of viral infectious diseases in the poultry industry is mainly achieved by the use of live attenuated and killed vaccines. Although the use of those vaccines is for the most part effective, it is not without shortcomings. Disadvantages of live vaccines include reversion to more pathogenic states, induction of vaccine reactions and limitation in the combination of vaccines that may be used simultaneously due to interference of replication between the organisms (Bermudez and Stewart-Brown, 2003). Limitations in the use of killed vaccines include the large amount of antigen needed for induction of the immune response, need of an adjuvant, limited cell-mediated immune response and need of a prior administration of a live vaccine for priming (Bermudez and Stewart-Brown, 2003; Silva, 2002). To avoid some of these limitations, the poultry industry has resorted to the use of recombinant vaccines. These vaccines are based in the use of fowlpox and herpesviruses as vectors, and they code for genes of the Newcastle, avian influenza and Marek's disease viruses (Bermudez and Stewart-Brown, 2003; Silva, 2002). Still these vaccines are of difficult application to the birds, expensive and hard to produce. In recent years, an extensive body of work has been performed for the characterization and use of replication defective parvoviruses for the purpose of gene delivery. These replication defective parvoviruses, the adeno associated viruses (AAV), are non-pathogenic viruses that belong to the Dependovirus genus of the *Parvovirinae* subfamily, in the *Parvoviridae* family, capable of accommodating relatively long pieces of DNA, and capable of infecting a wide variety of cell types (both in the resting and replicating phase). One of the outstanding features of AAVs replication is the

requirement for co-infection in the same cell by an unrelated helper virus. Either an adeno or a herpesvirus can supply complete helper functions for fully permissive AAV replication (Muzyczka and Berns, 2001). These helper functions have been best characterized for adenoviruses. It's now known that the expression of the E1A, E1B, E2A, E4, and VA RNA genes of adenoviruses are essential for the completion of AAV replication (Muzyczka and Berns, 2001). The genomic organization of the virus is relatively simple. It is composed of two major open reading frames (ORF): the Rep ORF that codes for non-structural proteins (called Rep proteins) in charge of directing viral DNA replication, integration into the host's genome and repression of viral gene transcription in the absence of a helper virus, and the Cap ORF that codes for the structural (capsid) proteins of the virus (Muzyczka and Berns, 2001). These ORF are flanked by palindromic inverted terminal repeats (ITR), which are important elements for initiation of viral DNA replication and integration into the host genome, as well as for generation of recombinant viruses (Muzyczka and Berns, 2001). Interestingly, the ITR sequences are the only signals necessary *in cis* for encapsidation of DNA by the structural proteins of the virus (Muzyczka and Berns, 2001; Xiao et al., 1998). In primates, most of the research work has been focused on the AAV serotype 2 for the purpose of viral biology studies and generation of viral-vectored gene delivery systems (Carter, 1992). Serotype 2 AAV-based gene delivery systems are currently being used for correction of genetic defects in animal models and humans. Genetic defects that have been corrected in experimental settings include cystic fibrosis (Colledge, 1994; Flotte, 1993; Flotte et al., 1993; O'Neal and Beaudet, 1994), hemophilia (Herzog et al., 2001; Schwaab and Oldenburg, 2001), rheumatoid arthritis (Ghivizzani et al., 2001) and diabetes (Goudy et al., 2001), among others. The system has also been used in gene delivery for cancer treatment (Fukui et al., 2001; Grifman et al., 2001; Harrington et al., 2001; Ponnazhagan et al.,

2001; Rochlitz, 2001) and expression of immunogenic peptides and antibodies for protection against viruses (Lewis et al., 2002; Xin et al., 2001). Procedures for generating helper virus-free recombinant AAV stocks *in vitro* have been well characterized (Smith-Arica and Bartlett, 2001; Snyder, 1999; Tamayose et al., 1996; Xiao et al., 1998). The most widely used procedure relies on the co-transfection of human embryonic kidney (HEK) 293 cell line cultures (which express the E1A and E1B immediate early genes of the human adenovirus type 5) (Graham et al., 1977) with three plasmid constructs: one coding for the Rep and Cap protein genes of the AAV to provide the non-structural regulatory proteins and the capsid proteins, a second plasmid containing the heterologous gene of interest flanked by the AAV ITRs, and a third plasmid that provide *in trans* the rest of the adenoviral helper functions needed to drive the assembly of recombinant viruses, derived from the human adenovirus type 5 genome (genes E2A, E4, and VA RNA) (Muzyczka and Berns, 2001; Xiao et al., 1998).

The adeno-associated viruses have a wide distribution in vertebrates and have been isolated in many species, including avians (Muzyczka and Berns, 2001). The Avian Adeno-Associated Virus (AAAV) was first isolated as a contaminant of isolates of the Olson strain of quail bronchitis virus (Yates et al., 1973). As with every member of the *Dependovirus* genus, these viruses require a coinfection with a helper virus (an adeno or herpesvirus) to complete its own replication (Yates et al., 1973). Because very little is known regarding the avian member of this family of viruses, and because it may hold similar attributes for the purpose of cell transduction and expression of heterologous genes, when compared with its primate counterparts, the objective of this work has been the initial characterization of the AAAV and generation of infectious clones and recombinant viral particles expressing a reporter gene, as the

first steps towards the development of an AAV-based gene delivery system and recombinant vaccines for the poultry industry.

MATERIALS AND METHODS

Viral stock propagation and purification: The AAV strain VR-865, obtained from the American Type Culture Collection (ATCC VR-865), was propagated in 9 day-old specific pathogen-free (SPF) embryonating chicken eggs (SPAFAS, Wilmington, MA) by co-infection with the CELO strain of the avian adenovirus type I (obtained from the Poultry Diagnostic and Research Center, University of Georgia, Athens, GA) via the allantoic sac. The allantoic fluid (AF) was collected when embryo mortality due to the avian adenovirus infection reached 10%. The AF was then clarified from cellular debris by centrifugation at 8,000 g for 20 min. The viral particles contained in the clarified AF were concentrated by ultracentrifugation on a cesium chloride (CsCl) cushion (60% w/v CsCl in TE buffer: Tris Hcl 0.05M, EDTA 0.005M, pH 7.5) at 100,000 g for 2 hr, and separated from the contaminating avian adenovirus by ultracentrifugation on a CsCl continuous gradient (1.2-1.5 gm/ml CsCl in TE buffer) at 100,000 g for 24 hr as described elsewhere (Yates et al., 1973). After centrifugation, the concentrated virus bands were collected through the side of the tube with a syringe equipped with an 18 gauge needle. Excess CsCl salt was eliminated from the viral samples by dialysis against phosphate buffered saline (PBS) solution at 4 C under agitation overnight.

Cell cultures: The human embryonic kidney 293 cell line was obtained from ATCC (CRL-1573) and grown in 75 cm² flasks in Dulbecco's modified minimal essential medium (Gibco, Carlsbad, CA) supplemented with 10% fetal bovine serum (Sigma, St. Louis, MO) and antibiotics. Cell cultures were maintained at 37 C under a 5% CO₂ atmosphere. Primary chicken

embryo fibroblasts, kidney and liver cells were obtained from 9 day-old (fibroblasts), 14 day-old (liver) and 18 day -old (kidney) SPF embryonating chicken eggs. These cell cultures were maintained in F10-M199 medium (Sigma, St. Louis, MO) supplemented with 10% fetal bovine serum and antibiotics, and under incubation at 37 C in a 5% CO₂ atmosphere.

Viral DNA cloning and sequencing: DNA from the purified AAV samples was obtained by the sodium dodecyl sulfate (SDS)-proteinase K-phenol/chloroform extraction method as described elsewhere (Hess et al., 1995). Since the AAV package ssDNA molecules of both senses (positive and negative sense ssDNA molecules) in roughly equimolar amounts (Hess et al., 1995), the extracted DNA was heated at 95 C for 10 min and allowed to cool slowly to room temperature, thus allowing the plus and minus strands of ssDNA to anneal, resulting in the generation of viral genomic dsDNA molecules. A single dATP was added to the 3' ends of the dsDNA molecules by incubation at 72 C for 20 min, in a 50 µl reaction containing *Taq* polymerase, dNTPs and recommended buffers (Invitrogen Corp. Carlsbad, CA). The viral DNA samples obtained by this procedure were run in a 1% agarose gel containing crystal violet for DNA visualization (40 µl of a 2 mg/ml crystal violet solution per 50 ml of 1% agarose. Invitrogen Corp. Carlsbad, CA), and gel purified with the aid of the Qiagen gel extraction kit reagents and recommended procedures (Qiagen Inc. Valencia, CA). The gel-purified DNA was cloned by using the TOPO-XL PCR cloning kit and recommended procedures (Invitrogen Corp. Carlsbad, CA). Plasmids obtained from the cloning procedure were transformed in chemically competent *E. coli* (Invitrogen Corp. Carlsbad, CA), which were grown overnight on Luria-Bertani agar plates supplemented with kanamycin for selection of plasmid-containing bacterial colonies. Ligation of the viral genome into the plasmid backbone was confirmed by restriction endonuclease (RE) digestion, using the *EcoRI* RE (New England Biolabs, Inc. Beverly, MA).

The inserts of clones positive by RE analysis were sequenced in their entirety by the dideoxy-chain termination procedure. All sequence analyses were performed with the aid of the Lasergene computerized sequence analysis software (Dnastar, Inc. Madison, WI). Determination of putative promoter and regulatory sequences was performed with the aid of the EMBOSS application programs package, which is freely available on the internet (<http://www.hgmp.mrc.ac.uk>). Regulatory and promoter sequences were mapped with the aid of the MatInspector V2.2 software for regulatory DNA sequences prediction (Quandt et al., 1995). Nucleotide sequences in the positive sense strand of the viral DNA with core similarity values of one, and matrix similarity values equal or above 0.950 were used to determine the regulatory elements of the genome. Prediction of RNA splicing donor and acceptor sequences was performed with the aid of the NetGene2 V2.4 software (Pertea et al., 2001).

Rescue of complete infective viral particles from the infectious clones: Rescue of complete viral particles from the plasmids containing the viral genome was performed by co-transfection of human embryonic kidney 293 cell line cultures (grown in 75 cm² flasks) with the recombinant plasmids containing the whole viral genome and a plasmid that provided *in trans* the E2A, E4 and VA RNA helper function genes derived from the human adenovirus type 5 (pHelper plasmid, Stratagene, La Jolla, CA). Transfection was performed with the aid of the Lipofectamine 2000 reagent and recommended procedures. Briefly, 600 µl of the Lipofectamine 2000 reagent were added to a tube containing 15 ml of DMEM media without serum or antibiotics. In another tube, 100 µg of each of the plasmids described were brought up to 15 ml by addition of DMEM medium, also without serum or antibiotics. After incubation at room temperature for 5 min, the content of the tube holding the diluted Lipofectamine reagent was transferred to the tube containing the plasmid DNA, and the mixture was incubated for 20 min at

room temperature. The amount of Lipofectamine-DNA complex solution prepared was sufficient to transfect 10 HEK 293 cell culture flasks (inoculums of 3 ml per flask). Recombinant AAV particles were collected from the tissue cultures 48 hr post transfection by following the protocol described by Xiao *et al* (Xiao et al., 1998). Briefly, cell culture flasks were scraped with a sterile cell culture policeman and the resulting suspensions were transferred to 50 ml tubes. The cells were frozen (at -80 C) and thawed two times, and the cellular debris was eliminated by centrifugation at 3000 g for 15 min. The cell supernatant collected after centrifugation was concentrated by ultracentrifugation on a CsCl cushion, and the rescued virus was further purified by ultracentrifugation on a CsCl linear gradient, as described for the initial isolation of the wild type virus from allantoic fluid. Viral particles were visualized by electron microscopy (EM), and their ability to replicate was assessed by co-infection of primary chicken embryo liver cell cultures (CELiC) with the rescued viruses and the CELO strain of the avian adenovirus type I. Confirmation of viral replication was performed by polymerase chain reaction (PCR), using specific primers designed to amplify a 451 bp fragment from the AAV Cap-coding region (Forward primer: AAV 2F 5' GCTCGGTCTTGGCCTTCTTCTTTT 3'. Reverse primer: AAV2R 5' TGCTGCGGCTGATTTCTACCATT 3'). To confirm that the PCR amplicon obtained corresponded to encapsidated genomic viral DNA and not as a result of spurious amplification of the plasmids used for viral rescue, a PCR control reaction was performed using specific primers designed to amplify a 437 bp fragment from backbone of the plasmid containing the entire viral genomes (infectious clones) (Forward primer: A3V backbone 1F 5' CTAGTAACGGCCGCCAGTGTGCT 3'. Reverse primer: A3V backbone 1R 5' TCCCCCTCCCCAGATTTCATTA 3'). To verify that the adenoviral stock used to co-infect

the cell culture for viral rescue was not contaminated with wtAAAV, DNA was extracted from the stock and reacted with the AAVV specific primers AAVV2F-R mentioned above.

Generation of plasmid constructs to produce recombinant AAVV particles coding the for *LacZ* reporter gene: To generate a plasmid-based platform to produce recombinant AAVV virus in cell culture, the complete Rep and Cap coding regions of the DA-1 strain of the virus were amplified from the DA-1 infectious clone by PCR using 5' phosphorylated primers designed to span the initiation codon (underlined) of the Rep proteins (forward primer F-3.6 RC ATG-Stop. 5'phos ATGAGGTCGTACTACGAGGTC-3') and the stop codon (underlined) of the capsid proteins coding region (reverse primer R-3.6 RC ATG-Stop. 5'phos-TTTACAGAGGTTTGGTAAGTACC-3'). The PCR reaction was performed by using a proof reading polymerase (Deep Vent DNA polymerase, New England Biolabs, Inc. Beverly, MA) in 100 µl reactions (polymerase: 1 µl; 10x buffer: 10 µl; primer: 1 µl each; MgSO₄: 0.5 µl of a 25 mM stock solution; template: 0.25 µl and H₂O to 100 µl). The thermocycler was programmed as follows: initial denaturation: 95 C for 5 min; amplification (35 cycles): denaturation 95 C for 40 s, annealing 58 C for 30 s and extension at 72 C for 4 min; final extension at 72 C for 8 min. The PCR product obtained was cloned by blunt-end ligation into an eukaryote expression vector under the major late promoter sequence of the cytomegalovirus (CMV), with the simian virus (SV) 40 polyadenylation signal located at the 3' end of the plasmid multi cloning site (pCMV-MCS, Stratagene. La Jolla, CA). The procedure was performed by digesting the plasmid backbone (pCMV-MCS) with *Hind*III, followed by end-repair of the linearized plasmid with the aid of the End-It DNA end repair kit and recommended procedures (Epicentre Technologies. Madison, WI). Ten µg of the linearized, blunt-end, plasmid backbone were dephosphorylated by incubation with calf intestinal phosphatase for 30 min at 37 C (1 unit of enzyme per µg of

linearized plasmid, New England Biolabs Inc. Beverly, MA). To minimize background during ligation, the linearized plasmid was run in a 1% agarose gel and later gel purified by using the Qiagen Gel Purification kit (Qiagen Inc. Valencia, CA). Ligation was performed by using the Fast Link DNA ligation kit and recommended procedures (Epicentre Technologies. Madison, WI) by reacting the amplified PCR products (Rep-Cap coding genes of the DA-1 strain of the virus) and the gel purified plasmid backbone (1:1 molar ratio) in a 10 µl reaction. After ligation, recombinant plasmids were transformed into chemically competent *E. coli*, which were then seeded on Luria-Bertani plates supplemented with ampicillin and incubated overnight at 37 C. Plasmids were isolated from selected colonies, grown overnight in liquid Luria-Bertani media, by using the Qiagen plasmid miniprep kit and recommended procedures (Qiagen Inc. Valencia, CA). Assessment of ligation and orientation of the PCR product with respect to the CMV promoter was performed via RE analysis, by digesting the plasmids obtained with *Bam*HI. To generate a plasmid containing the *LacZ* reporter gene flanked by the AAV DA-1 strain ITRs, the *LacZ* gene was obtained by digesting plasmid pAAV-LacZ (Stratagene, La Jolla, CA) with the *Not*I RE. The fragment obtained from the *Not*I digestion contained the *LacZ* gene, downstream from a copy of the CMV major late promoter sequence, and a copy of the SV40 polyadenylation signal downstream of the *LacZ* coding sequence. This fragment was ligated into a linearized backbone plasmid derived from the DA-1 infectious clone, which was obtained by digesting this infectious clone in successive reactions with the *Bsm*BI and *Nru*I RE (New England Biolabs, Inc. Beverly, MA), followed by gel purification, using the methods already described. These linearized plasmids contained the regulatory elements for plasmid replication in *E. coli*, a kanamycin resistance gene as a selection marker (plasmid backbone), and the sequence of both inverted terminal repeats of the DA-1 strain of the AAV. The ends of both DNA

populations (the excised *LacZ* gene and the linearized plasmid containing the AAV ITRs) were repaired as described above, and used as substrate for blunt-end ligation, also as described. The recombinant plasmids thus obtained contained the *LacZ* coding gene flanked by the AAV ITRs. These plasmids were transformed into chemically competent *E. coli* and grown on Luria-Bertani agar plates supplemented with kanamycin overnight. Positive ligation of the *LacZ* sequence into the backbone plasmids was assessed by PCR with primers design to amplify a 201 bp segment of the CMV promoter region contained in the DNA insert (CMV forward: 5'GGCGGAGTTGTTACGACAT3'; CMV reverse: 5'GGGACTTTCCTACTTGGCA3') (Rohr et al., 2002).

Production of recombinant AAV virus stock expressing the *LacZ* gene: AAV recombinant viral particles coding for the *LacZ* gene were obtained by co-transfection of HEK 293 cell line cultures with 10 µg each (per 72 cm² flask) of the plasmid containing the *LacZ* gene flanked by the AAV ITRs, the plasmid containing the AAV Rep and Cap coding regions and the helper plasmid expressing the E2A, E4 and VA-RNA. Transfection of the cells was performed in 10 flasks of HEK 293 cell cultures, as described above for viral rescue from the infectious clones, also with the aid of the Lipofectamine 2000 reagent. Purification of the recombinant viral population was performed by concentration of a CsCl cushion followed by a purification step by ultracentrifugation in a linear CsCl gradient, as previously described.

Expression of the *LacZ* protein in tissue culture by recombinant AAV-vectored gene delivery: The isolated recombinant AAV particles encoding for the *LacZ* gene were used for expression of *LacZ* activity in primary chicken embryo cell cultures. For this purpose, primary chicken embryo liver, fibroblast and kidney cells were grown in 24-well plates and inoculated with 0.3 ml of a serial log base 2 dilution of the virus (dilutions from 1:2 to 1:64).

Since plasmid DNA used to generate the recombinant viral particles might still be present in the concentrated viral stock, a DNase I treatment of the inoculums was performed before cell inoculation. DNase treatment was done by adding 10 μ l of a 20 mg/ml stock solution of DNase I (Roche Diagnostics, Mannheim, Germany) per 0.5 ml of concentrated viral stock, followed by incubation at 37 C for 30 min. Expression of *LacZ* activity was assessed by staining of the infected cells with the aid of the *in situ* β -galactosidase staining kit and recommended procedures (Specialty Media, Phillipsburg, NJ). Briefly, cell culture media was discarded and the monolayers were washed twice with phosphate buffered saline solution (PBS, Specialty Media, Phillipsburg, NJ), 48 hr post inoculation. The cells were then fixed by addition of 0.5 ml of a 0.2% glutaraldehyde solution for 5 min, washed again twice with PBS, and stained by adding 0.5 ml of complete β -gal stain solution (X-gal, potassium ferrocyanide, potassium ferricyanide and PBS). The cells were stained for a minimum of 2 hr, and expression of β -gal activity was assessed by light microscopy using an inverted microscope.

RESULTS

DNA cloning and sequence analysis: DNA isolated from the purified AAV stock obtained from the initial viral propagation in embryonating chicken eggs was successfully cloned. RE analysis of the recombinant plasmids revealed two distinct viral DNA populations (Fig. 1). Closer analysis of the restriction fragment patterns showed that we cloned the genomes of the original virus obtained from ATCC (strain VR-865) and the DA-1 strain of the virus, previously described by Hess *et al* (Hess et al., 1995). Sequences of the whole genome of both viruses were obtained. Analysis of these sequences showed that the genomic mapping of the major open reading frames (Rep and Cap coding regions) is conserved and in agreement with

what has been described for the primate counterparts of this family of viruses (Fig. 2). As is the case with the AAV, the AAV Rep and Cap coding regions are also flanked by palindromic inverted terminal repeats, capable of adopting the T shape structure characteristic of this family of viruses (Fig. 3). Regulatory sequences necessary for interaction of the Rep proteins, initiation of viral DNA replication and integration into the host's genome were found to be highly conserved in the terminal repeats of both AAV strains in comparison to the primate strains of the virus (Fig. 3). As mentioned above, the genomic map of the AAV is conserved and is comprised of two major open reading frames. The Rep ORF spans nucleotides 244 to 2232 and 224 to 2235 of the VA-856 and DA-1 strains of the virus, respectively, while the Cap coding region is located between nucleotides 2250 to 4481 and 2253 to 4469 of the same strains, respectively. The genomic sizes of these viruses are 4694 bp-long (VR-865), and 4682 bp-long (DA-1), the 12 nucleotides difference being due to small deletions in the sequences of VP2 and VP3 of the DA-1 strain, with respect to the VR-865 strain sequence (Fig. 4). Putative regulatory sequences were determined as described in materials and methods, and are indicated in bold and underlined in the viral sequences in figure 4.

The phylogenetic analysis performed showed a 68% identity between the complete nucleotide sequences of the two AAV viral strains. Comparison of the AAV sequences with those of the primate AAV (strains 1, 2, 3, 4 and 6) showed percentages of identity of 35.8 to 45.1% among the primate strains and the AAV DA-1, and 40.2 to 45.5% identity among the AAV VR-865 strain and the primate strains (Fig. 5A). Comparison of the amino acid sequences of the Rep and cap coding regions of the two AAV strains showed percentages of identity of 96.1 and 90.3% for each amino acid sequence, respectively. The percentages of identity for the amino acid sequences of the Rep coding region among the AAV DA-1 strain

and its primate counterparts ranged from 53.8 to 54.8%, while the same analysis for the VR-865 strain showed percentages ranging from 55 to 55.4% (Fig. 5B). The amino acid sequences of the Cap proteins of the AAV strains were less conserved than those of the Rep proteins, with a percentage identity of 90.3%. Comparison of the Cap protein amino acid sequences of the avian vs. the primate strains of the virus showed a 54.8 to 59.3% of identity among the AAV DA-1 strain in comparison with the primate strains of the virus, while the AAV VR-865 strain showed percentages ranging from 54.4 to 58.5 in comparison with the primate strains (Fig. 5C). Analysis of the Rep coding region amino acid sequences showed a high degree of conservation in amino acid residues involved in the nuclease and helicase activity of the rep 78/68 proteins, which are indicated in figure 6.

Rescue of complete viral particles from infectious clones: Infectious viral particles were obtained by co-transfection of HEK 293 cell cultures with the plasmid containing the complete AAV genome and a helper plasmid containing genes E2A, E4 and VA RNA, derived from the human adenovirus type 5. These viral populations were visualized by electron microscopy (Fig. 7). To further determine if these viral particles were infectious, the rescued viruses and the CELO strain of the avian adeno virus type I were used to co-infect CELiC cultures as described. The rescued virus was able to replicate in the presence of the helper adenovirus, which was confirmed by PCR, using DNA extracted from the cell culture supernatant and primers specific for the Cap-coding region of the AAV (Fig. 8).

Generation of plasmid constructs to produce recombinant AAV particles coding for the *LacZ* gene: Two plasmid constructs were generated for the production of recombinant AAV particles as described in materials and methods. Plasmid pEVA3V-RC contains the complete, PCR-amplified, Rep-Cap coding regions of the DA-1 strain of AAV under the major

late promoter of the cytomegalovirus. This plasmid was used to provide *in trans* the AAV proteins necessary for viral particle generation and encapsidation of the *LacZ* coding sequence flanked by the inverted terminal repeats of the virus (Fig. 9A). Plasmid pEVA3V-LacZ-ITR contains a copy of the *LacZ* gene flanked by the AAV ITR. (Fig. 9B). The orientation of the AAV Rep and Cap coding region with respect to the CMV promoter in plasmid pEVA3V-RC, and confirmation of ligation of the *LacZ* gene in plasmid pEVA3V-LacZ-ITR, was assessed by restriction enzyme digestion (data not shown).

Generation of recombinant AAV particles coding for the *LacZ* gene: Generation of recombinant particles coding for the *LacZ* gene was performed in HEK 293 cell cultures as described. Visible viral bands, after ultracentrifugation on CsCl gradients, were collected and visualized by EM (Fig. 10).

Expression of *LacZ* in tissue culture by rAAV-based gene delivery: DNase I-treated purified recombinant viral stock expressing the *LacZ* gene was used to inoculate primary chicken embryo liver, kidney and fibroblasts cell cultures as described. Strong activity of the transgene was observed in all cell cultures, regardless of cell type. The highest level of expression was observed in primary fibroblasts, as manifested by the intensity and amount of stained cells observed by light microscopy. The second highest level observed corresponded to the liver cell cultures, with the lowest levels of expression observed in kidney cells (Fig. 11). The highest dilutions in which positive staining was observed were 1:64 for the primary fibroblast cultures, 1:32 for the liver cell cultures and 1:32 for the kidney cells.

DISCUSSION:

In this study infectious clones of the two known strains of the avian adeno associated virus have been successfully produced. Also, a plasmid-based platform to produce rAAAVs expressing a reporter gene (*LacZ*), which are able to transduce a variety of primary cells *in vitro*, have been generated. The rescue of the DA-1 strain of the virus was a serendipitous event, which we can only speculate as being the result of contamination of the initial ATCC stock of AAV used to obtain the viral DNA, contamination of the adenoviral stock used to provide the helper function during initial virus propagation, or rescue from a genome-integrated virus from the SPF eggs initially used in viral propagation. Interestingly, we were able to rescue both complete infective viral particles, and recombinant viruses, by using a system that was initially developed for the propagation of adeno-associated viruses isolated from primates. This is probably due to the high degree of homology observed between the nucleotide sequences of the terminal repeats of the AAV, as compared to the primate AAV. For instance, the primate AAV tandem nucleotide repeats of GAGY₄ (known as the Rep protein Binding Element-RBE), and the putative terminal resolution site (TRS) of nucleotide sequence TGGCCA were both present in the AAV ITRs (Fig. 3). Another previously described primate AAV nucleotide sequence necessary for the interaction of the Rep proteins with the inverted terminal repeats of the genome (the RBE') was also found in the ITRs of the AAV, with a single substitution from the consensus sequence CTTTG to CTTCG. This nucleotide sequence (CTTCG) is located at the tip of one of the arms of the T-shaped secondary structure of the inverted terminal repeats (Fig. 3), and is thought to facilitate the binding of the Rep proteins to the ITRs (Muzyczka and Berns, 2001). The nucleotide sequence and secondary structure of the AAV ITRs suggest that these viruses undergo the same type of single-strand displacement DNA replication mechanism

described for the AAVs (Muzyczka and Berns, 2001). When the amino acid sequences of each of the major coding regions were compared, the highest degree of divergence was found in the amino acid sequence of the Rep protein in comparison to the same region of the AAV strains analyzed. This was somewhat surprising at first, because the DNA sequence with which this protein initially interacts (the inverted terminal repeats) is highly conserved among the dependovirus sequences analyzed. However, a closer analysis of the protein revealed that the amino acid residues responsible for this initial interaction with the terminal repeats are highly conserved. Using site directed mutagenesis, researchers have demonstrated that residues Glu 6, Glu 83 and Glu 164 are necessary to coordinate the divalent metal ion binding (mainly magnesium and manganese ions) exhibited by the Rep protein, which is mediated by His residues at positions 90 and 92. These researches have also demonstrated that metal binding is essential for the endonuclease activity of the protein, which is catalyzed by the Tyr residue at position 155 (Fig. 6) (Yoon-Robarts and Linden, 2003). We found that these residues are highly conserved in the sequence of both AAV strains in comparison with the primate AAVs. Investigators have also identified the residues that mediate the helicase activity of the protein, which is essential for the initiation of DNA replication (Walker et al., 1997). Sequence analysis has shown that the parvoviruses non structural protein helicase belong to the superfamily III of helicases, which have three conserved sequence motifs (shown as A, B-B' and C in figure 6), that are packed in an approximately 100-amino acid domain (Walker et al., 1997). These sequence motifs were also found to be conserved in the avian and primate dependoviruses. The analyses indicate that the main functions of these proteins, namely the endonuclease and helicase activities, have been retained through evolution in this family of viruses. Regarding the structural proteins of the virus, the exact amino acid residues involved in cell attachment are less evident.

Previous research indicates that parvovirus cell tropisms and host range, as well as evolution, are controlled mainly by raised structural elements around the threefold axes of symmetry of the capsid of the virions (Hueffer and Parrish, 2003). Atomic modeling of the AAV serotype 2 capsid indicates that this structural element is also present, although composed of different parts of the primary sequence, as compared with other parvoviruses (Hueffer and Parrish, 2003). Many of the AAVs have a broad host range and the capsid can infect cells using a variety of different receptors; which might also be the case with the avian adeno associated viruses, as hinted by the ability of the recombinant virus generated in this study to readily penetrate and transduce different avian tissues *in vitro*.

The usefulness of the adeno-associated viruses as vectors for gene delivery has been well-established in the scientific literature for the viral strains capable of infecting primates. The ability of the avian adeno-associated virus to efficiently transduce cells in primary cultures and the levels of transgene expression observed in these systems indicate that these viruses could be used for gene delivery and the generation of recombinant vaccines to be used in the poultry industry.

REFERENCES

- Bermudez, A.J. and Stewart-Brown, B. (2003) Principles of Disease Prevention: Diagnosis and Control. In: Y.M. Saif (Ed), Diseases of poultry, 11th Edition ed., pp. 3-55. Iowa State Press, Ames, Iowa.
- Carter, B.J. (1992) Adeno-associated virus vectors. *Curr Opin Biotechnol* 3(5), 533-9.
- Colledge, W.H. (1994) Cystic fibrosis gene therapy. *Curr Opin Genet Dev* 4(3), 466-71.
- Flotte, T.R. (1993) Prospects for virus-based gene therapy for cystic fibrosis. *J Bioenerg Biomembr* 25(1), 37-42.
- Flotte, T.R., Afione, S.A., Conrad, C., McGrath, S.A., Solow, R., Oka, H., Zeitlin, P.L., Guggino, W.B. and Carter, B.J. (1993) Stable in vivo expression of the cystic fibrosis transmembrane conductance regulator with an adeno-associated virus vector. *Proc Natl Acad Sci U S A* 90(22), 10613-7.
- Fukui, T., Hayashi, Y., Kagami, H., Yamamoto, N., Fukuhara, H., Tohnai, I., Ueda, M., Mizuno, M. and Yoshida, J. (2001) Suicide gene therapy for human oral squamous cell carcinoma cell lines with adeno-associated virus vector. *Oral Oncol* 37(3), 211-5.
- Ghivizzani, S.C., Oligino, T.J., Glorioso, J.C., Robbins, P.D. and Evans, C.H. (2001) Direct gene delivery strategies for the treatment of rheumatoid arthritis. *Drug Discov Today* 6(5), 259-267.
- Goudy, K., Song, S., Wasserfall, C., Zhang, Y.C., Kapturczak, M., Muir, A., Powers, M., Scott-Jorgensen, M., Campbell-Thompson, M., Crawford, J.M., Ellis, T.M., Flotte, T.R. and Atkinson, M.A. (2001) Adeno-associated virus vector-mediated IL-10 gene delivery prevents type 1 diabetes in NOD mice. *Proc Natl Acad Sci U S A* 98(24), 13913-8.

- Graham, F.L., Smiley, J., Russell, W.C. and Nairn, R. (1977) Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J Gen Virol* 36(1), 59-74.
- Grifman, M., Trepel, M., Speece, P., Gilbert, L.B., Arap, W., Pasqualini, R. and Weitzman, M.D. (2001) Incorporation of tumor-targeting peptides into recombinant adeno-associated virus capsids. *Mol Ther* 3(6), 964-75.
- Harrington, K.J., Spitzweg, C., Bateman, A.R., Morris, J.C. and Vile, R.G. (2001) Gene therapy for prostate cancer: current status and future prospects. *J Urol* 166(4), 1220-33.
- Herzog, R.W., Mount, J.D., Arruda, V.R., High, K.A. and Lothrop, C.D., Jr. (2001) Muscle-directed gene transfer and transient immune suppression result in sustained partial correction of canine hemophilia B caused by a null mutation. *Mol Ther* 4(3), 192-200.
- Hess, M., Paul, G., Kling, S. and Monreal, G. (1995) Molecular characterization of two strains of the avian adeno-associated virus (AAAV). *Arch Virol* 140(3), 591-8.
- Hueffer, K. and Parrish, C.R. (2003) Parvovirus host range, cell tropism and evolution. *Current Opinion in Microbiology* 6(4), 392-398.
- Lewis, A.D., Chen, R., Montefiori, D.C., Johnson, P.R. and Clark, K.R. (2002) Generation of neutralizing activity against human immunodeficiency virus type 1 in serum by antibody gene transfer. *J Virol* 76(17), 8769-75.
- Muzyczka, N. and Berns, K.I. (2001) *Parvoviridae: The Viruses and Their Replication*. In: D.M. Knipe and P.M. Howley (Eds), *Fields Virology*, pp. 2327-2379. Vol. 2. 2 vols. Lippincott Williams & Wilkins, Philadelphia.
- O'Neal, W.K. and Beaudet, A.L. (1994) Somatic gene therapy for cystic fibrosis. *Hum Mol Genet* 3 Spec No, 1497-502.

- Pertea, M., Lin, X. and Salzberg, S.L. (2001) GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res* 29(5), 1185-90.
- Ponnazhagan, S., Curiel, D.T., Shaw, D.R., Alvarez, R.D. and Siegal, G.P. (2001) Adeno-associated virus for cancer gene therapy. *Cancer Res* 61(17), 6313-21.
- Quandt, K., Frech, K., Karas, H., Wingender, E. and Werner, T. (1995) MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res* 23(23), 4878-84.
- Rochlitz, C.F. (2001) Gene therapy of cancer. *Swiss Med Wkly* 131(1-2), 4-9.
- Rohr, U.P., Wulf, M.A., Stahn, S., Steidl, U., Haas, R. and Kronenwett, R. (2002) Fast and reliable titration of recombinant adeno-associated virus type-2 using quantitative real-time PCR. *J Virol Methods* 106(1), 81-8.
- Schwaab, R. and Oldenburg, J. (2001) Gene therapy of hemophilia. *Semin Thromb Hemost* 27(4), 417-24.
- Silva, R.F. (2002) Recombinant Vaccines and DNA Vaccines: Unfulfilled Promises. *Poultry Vaccines and Vaccination Practices Symposium*. 45th meeting of the American Association of Avian Pathologists, Nashville, Tennessee, 2002.
- Smith-Arica, J.R. and Bartlett, J.S. (2001) Gene therapy: recombinant adeno-associated virus vectors. *Curr Cardiol Rep* 3(1), 43-9.
- Snyder, R.O. (1999) Adeno-associated virus-mediated gene delivery. *J Gene Med* 1(3), 166-75.
- Tamayose, K., Hirai, Y. and Shimada, T. (1996) A new strategy for large-scale preparation of high-titer recombinant adeno-associated virus vectors by using packaging cell lines and sulfonated cellulose column chromatography. *Hum Gene Ther* 7(4), 507-13.

- Walker, S.L., Wonderling, R.S. and Owens, R.A. (1997) Mutational analysis of the adeno-associated virus type 2 Rep68 protein helicase motifs. *J Virol* 71(9), 6996-7004.
- Xiao, X., Li, J. and Samulski, R.J. (1998) Production of High-Titer Recombinant Adeno-Associated Virus Vectors in the Absence of Helper Adenovirus. *J. Virol.* 72(3), 2224-2232.
- Xin, K.Q., Urabe, M., Yang, J., Nomiyama, K., Mizukami, H., Hamajima, K., Nomiyama, H., Saito, T., Imai, M., Monahan, J., Okuda, K. and Ozawa, K. (2001) A novel recombinant adeno-associated virus vaccine induces a long-term humoral immune response to human immunodeficiency virus. *Hum Gene Ther* 12(9), 1047-61.
- Yates, V.J., el-Mishad, A.M., McCormick, K.J. and Trentin, J.J. (1973) Isolation and characterization of an Avian adenovirus-associated virus. *Infect Immun* 7(6), 973-80.
- Yoon-Robarts, M. and Linden, R.M. (2003) Identification of active site residues of the adeno-associated virus type 2 Rep endonuclease. *J Biol Chem* 278(7), 4912-8.

Acknowledgements:

Thanks to Dr. Mark Jackwood, in the Poultry Diagnostic and Research Center, for reviewing this manuscript and for his valuable commentaries and suggestions.

This work has been funded with Veterinary Medicine Agricultural Research funds.

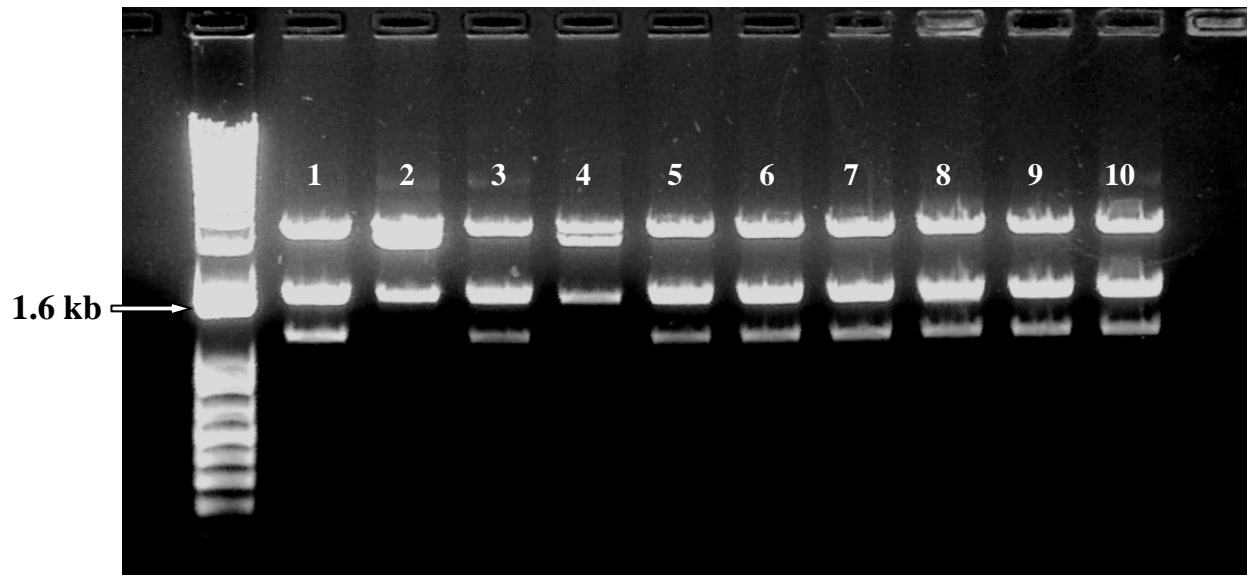


Fig. 1: *Eco*RI RE digestion of plasmids containing the entire genome of the DA-1 and VR-685 strains of the AAV. Restriction patterns in lanes 2 and 4 correspond to the VR-865 strain of the AAV. The rest of the patterns correspond to the DA-1 strain of the virus.

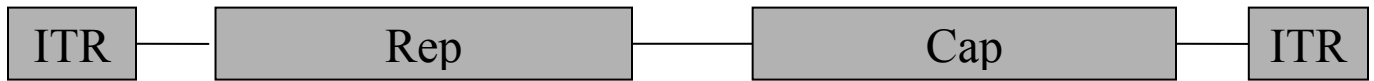


Fig. 2: Genomic map of the avian adeno associated virus.

Diagram illustrating a DNA sequence with a 10 bp deletion. The sequence is shown as a double-stranded DNA molecule. The top strand (5' to 3') contains the TRS (Target Recognition Sequence) and the RBE (Right Border Element). The bottom strand (3' to 5') contains the ACCGGT sequence and the RBE' (Right Border Element'). The 10 bp deletion is indicated by a box around the RBE region, showing the sequence GCTCGCTCGCTCACTC on the top strand and CGAGCGAGCGAGTGAG on the bottom strand. The deletion is located between the TRS and the RBE. The sequence is flanked by A and A on the top, and A and A on the bottom. The sequence is also flanked by C and T on the bottom right.

5' GCATTACCTCACCGGAGTGCT**TGGCCA**AGTTTCCAAGACAG**GCTCGCTCGCTCACTC**GGGCC 3' T

3' **ACCGGT**CAAAGGTTCTGTCC**CGAGCGAGCGAGTGAG**CCCCGG 5' A

TRS RBE

RBE'

Diagram illustrating a DNA double helix structure with a 10bp deletion. The top strand (5' to 3') is GCATTACCTCACC GGAGTGCT TGGCCA GTTTCCAAAACAG GCTCGCTCGCTCACTCGGGCC. The bottom strand (3' to 5') is ACCGGTCAAAGGTTTGTCCGAGCGAGCGAGTGAGCCCCG. A 10bp deletion is shown in the top strand between TGGCCA and GCTCGCT, where the sequence TTTCCAAAAC is missing. The TRS (Top Replication Strand) is labeled above the top strand, and the RBE (Right Branch Entry) is labeled above the deletion site. The RBE' (Right Branch Entry') is labeled below the bottom strand, with a box highlighting the sequence CTTTCCAAAAC in the bottom strand.

Fig. 3: T-shaped secondary structure of the 3' end terminal repeat of strains VR-865 (A) and DA-1 (B) of the AAV. Nucleotides in bold correspond to the terminal resolution site. Boxed nucleotides correspond to the putative Rep protein binding element.

RBE		
AAAV DA-1	-TGGCCAGTTTC-CAAAACAGGC <u>TCGCTCGCTCACTCG</u> GGGCCCCGGCCCCAAGGGGGCCGGTAGCGACGC--	68
AAAV VR 865	-TGGCCAGTTTC-CAAGACAGGC <u>TCGCTCGCTCACTCG</u> GGGCCCCGGCCCCAAAGGGGGCCCTAGCGACCG--	68
AAV 1	TTGCCCACTCCCTCTCTGCGCGCTCGCTCGTTCGGTGGGGGCTGCGGACCAAGGT-CCGCAGACGGCAGAG	71
AAV 2	TTGCCCACTCCCTCTCTGCGCGCTCGCTCGTCACTGAGGCCGGGGCGACCAAGGT-CGCCCGACGCCCGGG	71
AAV 3	-TGGCCACTCCCTCTATGCGCACTCGCTCGTTCGGTGGGGGCTGGCGACCAAGGT-CGCCAGACGGACGTG	70
AAV 4	TTGCCCACTCCCTCTATGCGCGCTCGCTCACTCACTCGGCCCTGGAGACCAAGGT-CTCCAGACTGCCGGC	71
AAV 6	TTGCCCACTCCCTCTCTGCGCGCTCGCTCGTCACTGAGGCCGGGGCGACCAAGGT-CGCCCGACGCCCGGG	71
RBE' RBE TRS		
AAAV DA-1	<u>CTTCG</u> GCGTCGCGGCCC <u>GAGTGAGCGAGCGAGC</u> CTGTTTTG-GAAAC <u>TGGCCA</u> GCAC-----TCCGG-----	129
AAAV VR 865	<u>CTTCG</u> CGGTGCGGGCCC <u>GAGTGAGCGAGCGAGC</u> CTGTCTTG-GAAAC <u>TGGCCA</u> GCAC-----TCCGG-----	129
AAV 1	CTCTGCTCTGCGGGCCCCACCGAGCGAGCGAGCGCGCAGAGAGGGAGTGGGCAACTCCATCA-CTAGGG---	139
AAV 2	CTTTGCCCCGGGCGGCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGGCAACTCCATCA-CTAGGG---	139
AAV 3	CTTTGACGTCGGGCCCCACCGAGCGAGCGAGTGCGCATAGAGGGAGTGGGCAACTCCATCA-CTAGAG---	138
AAV 4	CTCTGGCCGGCAGGGCCGAGTGAGTGAGCGAGCGCGCATAGAGGGAGTGGGCAACTCCATCATCTAGGTTTG	143
AAV 6	CTTTGCCCCGGGCGGCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGGCAACTCCATCA-CTAGGG---	139
USF SPI		
AAAV DA-1	-----TGAGGTAA---TGCC <u>GTCACGTGGT</u> CGGGAA-----CGG <u>GAACGGGAA</u> -----ATCT	173
AAAV VR 865	-----TGAGGTAA---TGCC <u>GTCACGTGGT</u> CGGGAA-----TGG <u>GAACGGGAA</u> -----ATCT	173
AAV 1	-----GTAATCGCGA-----AGCGCC-----TCCCACGCTGCC-GCGTCAGCG-----CTGA	180
AAV 2	-----GT--TCCTGG-----AGGG-----GTGGAGTC-GTGACGTGA-----ATTA	172
AAV 3	-----GT--AT-----G-----GCAGTGAC-GTAACGCGA-----AGCG	164
AAV 4	CCCACTGACGTCAATGTGACGTCTTAGGGGTAGGGAGGTCCTGTATTAGCAGTCACGTGAGTGTGCTATTT	215
AAV 6	-----GT--TCCTGG-----AGGG-----GTGGAGTC-GTGACGTGA-----ATTA	172
TATA box p5		
AAAV DA-1	CGCGAG-----AACGTAAACAA-----A <u>TATAA</u> GA-----CGA	201
AAAV VR 865	CGCGAG-----AACGTAAACAA-----A <u>TATAA</u> GA-----CGG	201
AAV 1	CGTAAATTACGTCATAG-GGGAGTGGTCTGT-AT-TAGCTGTACGTGAGTGCTTTTGCGACATTTTGCGA	249
AAV 2	CGTCA--TAGGGT-TAG-GGAGGTC--C-TGT-AT-TAGAGGTCACGTGAGTG-TTTTGCGACATTTTGCGA	234
AAV 3	CGCGAAGCGAGAC-CAC-GCCTACCAGC-TGC-GT-CAGCAGTCAGGTGACCC-TTTTGCGACAGTTTGCGA	230
AAV 4	CGCGGAGCGTAGCGGAGCGCATACCAAGCTGCCACGTACAGCCACGTG-GTCCGTTTGCGACAGTTTGCGA	286
AAV 6	CGTCA--TAGGGT-TAG-GGAGGTC--C-TGT-AT-TAGAGGTCACGTGAGTG-TTTTGCGACATTTTGCGA	234
p5 Rep78/68		
AAAV DA-1	CGCCACACGG---CGCTGCGT-CATACGCGCGCGCGCACCGGCGAG <u>ATG</u> -----	247
AAAV VR 865	CGCCACACGG---CGCTGCGT-CATACGCGCGCGCGCACCGGCGAG <u>ATG</u> -----	247
AAV 1	CACCACGTGGCCATTTAGGGTATATATGGCCGAGTGAGCGAGC-AGGATCTCCATTTTGAC-CGCGAAATTT	319
AAV 2	CACCATGTGGTCACGCTGGGTATTTAAGCCCGAGTGAGCACGC-AGGGTCTCCATTTTGAAGCGGGAGGTTT	305
AAV 3	CACCACGTGGCCGCTGAGGGTATATATCTCGAGTGAGCGAACCAGGAGCTCCATTTTGAC-CGCGAAATTT	301
AAV 4	CACCATGTGGTCA-GGAGGGTATATAACCCGAGTGAGCCAGCGAGGAGCTCCATTTTGCC-CGCGAATTTT	356
AAV 6	CACCATGTGGTCACGCTGGGTATTTAAGCCCGAGTGAGCACGC-AGGGTCTCCATTTTGAAGCGGGAGGTTT	305
AAV DA-1 AAV VR 865 AAV 1 AAV 2 AAV 3 AAV 4 AAV 6		
AAAV DA-1	-AG-----GTCGTAC-----TACGAGGTCATCGTTCAACTGCCGAACGACGTCGAGAGTCAGGTA	300
AAAV VR 865	-AG-----GTCGTAC-----TACGAGGTCATCGTTCACTGCTGCCAACGACGTCGAGAGTCAGGTA	300
AAV 1	GAACGAGCAGCAGCCATGCCGGGCTTCTACGAGATCGTGATCAAGGTGCCGAGCGACCTGGACGAGCACCTG	391
AAV 2	GAACGCGCAGCCGATGCCGGGGTTTACGAGATTGTGATTAAGGTCCCCAGCGACCTTGACGAGCATCTG	377
AAV 3	GAACGAGCAGCAGCCATGCCGGGGTTCTACGAGATTGTCTGAAGGTCCCCAGTGACCTGGACGAGCACCTG	373
AAV 4	GAACGAGCAGCAGCCATGCCGGGGTTCTACGAGATCGTGCTGAAGGTGCCGAGCGACCTGGACGAGCACCTG	428
AAV 6	GAACGCGCAGC-GCCATGCCGGGGTTTACGAGATTGTGATTAAGGTCCCCAGCGACCTTGACGAGCATCTG	376

AAAV DA-1	CCTGGAATCTCCGATTTCATTCGTCAACTGGATTACGTGCGGAGAATGGACGTTGCCGGAAGACGCCGATTGG	372
AAAV VR 865	CCTGGAATCTCCGATTTCGTTTCGTCAACTGGATTACGTGCGGAGAATGGACGTTGCCGAGGACGCCGATTGG	372
AAV 1	CCGGGCATTTCTGACTCGTTTGTGAGCTGGGTGGCCGAGAAGGAATGGGAGCTGCCCGCGATTCTGCACATG	463
AAV 2	CCCGGCATTTCTGACAGCTTTGTGAACTGGGTGGCCGAGAAGGAATGGGAGTTGCCGCGAGATTCTGCACATG	449
AAV 3	CCGGGCATTTCTAACTCGTTTGTAACTGGGTGGCCGAGAAGGAATGGGAGCTGCCCGCGATTCTGCACATG	445
AAV 4	CCCGGCATTTCTGACTCTTTTGTGAGCTGGGTGGCCGAGAAGGAATGGGAGCTGCCCGCGATTCTGCACATG	500
AAV 6	CCCGGCATTTCTGACAGCTTTGTGAACTGGGTGGCCGAGAAGGAATGGGAGTTGCCGCGAGATTCTGCACATG	448
AAAV DA-1	GATTTGGACCAGGTCGATCAAGTTCAACTGACGCTCGGCGACAAAAATCCAACGGGAGATTGGAATCACTGG	444
AAAV VR 865	GATTTGGACCAGGTCGATCAAGTTCAACTGACGCTCGGCGACAAAAATCCAACGGGAGATTGGAATCACTGG	444
AAV 1	GATCTGAATCTGATTGAGCAGGCACCCCTGACCGTGGCCGAGAAGCTGCAGCGCGACTTCTGGTCCAATGG	535
AAV 2	GATCTGAATCTGATTGAGCAGGCACCCCTGACCGTGGCCGAGAAGCTGCAGCGCGACTTCTGACGGAATGG	521
AAV 3	GATCCGAATCTGATTGAGCAGGCACCCCTGACCGTGGCCGAAAAGCTTCAGCGCGAGTTCTGGTGGAGTGG	517
AAV 4	GACTTGAATCTGATTGAGCAGGCACCCCTGACCGTGGCCGAAAAGCTGCACGCGAGTTCTGTTCGAGTGG	572
AAV 6	GATCTGAATCTGATTGAGCAGGCACCCCTGACCGTGGCCGAGAAGCTGCAGCGCGACTTCTGGTCCAAGTGG	520
AAAV DA-1	GGGACGATGGCGAAAGAACCAGGACTTTCATAATTTTATCCAACCTGGAACAAGGTGAGGTGTTCTTTTCATTTG	516
AAAV VR 865	GGGACGATGGCGAAAGAACCAGGACTTTCATAATTTTATCCAACCTGGAACAAGGTGAGGTGTTCTTTTCATTTA	516
AAV 1	CGCCGCGTGAGTAAGGCCCGGAGGCCCTCTTCTTGTTCAGTTTCGAGAAGGGCGAGTCCCTACTTCCACCTC	607
AAV 2	CGCCGCTGTGAGTAAGGCCCGGAGGCCCTTTTCTTTGTGCAATTTGAGAAGGGGAGAGACTACTTCCACATG	593
AAV 3	CGCCGCGTGAGTAAGGCCCGGAGGCCCTCTTTTGTTCAGTTTCGAAAAGGGGAGACCTACTTCCACCTG	589
AAV 4	CGCCGCGTGAGTAAGGCCCGGAGGCCCTCTTCTTTGTTCAGTTTCGAGAAGGGGACAGCTACTTCCACCTG	644
AAV 6	CGCCGCGTGAGTAAGGCCCGGAGGCCCTCTTCTTGTTCAGTTTCGAGAAGGGGCGAGTCCCTACTTCCACCTC	592
AAAV DA-1	CACGTCTGCTGGAACCGTGTTCCTGTAAGCCGATGGTACTCGGAAGATATATTTCGACATATTCAACAAAAA	588
AAAV VR 865	CACGTCTGCTGGAACCGTGTTCCTGTAAGCCGATGGTACTCGGAAGATATATCCGACATATTCAACAAAAA	588
AAV 1	CATATTCTGGTGGAGACCACGGGGGTCAAATCCATGGTGTGCGCCGCTTCTGAGTCAGATTAGGGACAAG	679
AAV 2	CACGTCTGCTGGAACCCACCGGGGTGAAATCCATGGTTTGGGACGTTTCTGAGTCAGATTTCGCGAAAAA	665
AAV 3	CACGTCTGATTGAGACCATCGGGGTCAAATCCATGGTGGTGGCCGCTACGTGAGCCAGATTAAAGAGAAG	661
AAV 4	CACATCTGCTGGTGGAGACCGTGGGCGTCAAATCCATGGTGGTGGCCGCTACGTGAGCCAGATTAAAGAGAAG	716
AAV 6	CATATTCTGGTGGAGACCACGGGGGTCAAATCCATGGTGTGCGCCGCTTCTGAGTCAGATTAGGGACAAG	664
AP-1		
AAAV DA-1	ATTGTGAGTAAAGTCTACTGCGGCCACGAGCCCGCATGGAAGGATGGATGCG- <u>GGTGACCAA</u> GACCAAAAA	659
AAAV VR 865	ATTGTGAGTAAAGTCTACTGCG-CCACGAGCCTACGATGGAAGGATGGATGCGT <u>GGTGACCAA</u> GACCAAAAA	659
AAV 1	CTGGTGAGACCATCTACCGCGGGATCGAGCCGACCTGCCCAACTGGTT-CGCGGTGACCAAGACGCGTAA	750
AAV 2	CTGATTGAGAGAATTTACCGCGGGATCGAGCCGACTTTGCCAACTGGTT-CGCGGTGACCAAGACCGAGAAA	736
AAV 3	CTGGTGACCCGCATCTACCGCGGGGTGAGCCGCGAGCTTCCGAACCTGGTT-CGCGGTGACCAAAACGCGAAA	732
AAV 4	CTGGTGACCCGCATCTACCGCGGGGTGAGCCGCGAGCTTCCGAACCTGGTT-CGCGGTGACCAAGACGCGTAA	787
AAV 6	CTGGTGAGACCATCTACCGCGGGATCGAGCCGACCTGCCCAACTGGTT-CGCGGTGACCAAGACGCGTAA	735
Est-1		
AAAV DA-1	TT---TCGGGGGCGCGAACAAGGTCCGGGGCGAGTCGTATATTCCCGCTACCTGATTCCGAAACAG <u>CAACC</u>	728
AAAV VR 865	TT---TCGGGGGCGCGAACAAGGTCCGGGGCGAGTCGTATATTCCCGCTACCTGATCCCGAAACAG <u>CAACC</u>	728
AAV 1	TGGCGCGGAGGGGGGAACAAGGTGGTGGACGAGTGCTACATCCCCAACTACCTCCTGCCAAGACTCAGCC	822
AAV 2	TGGCGCGGAGGGCGGAACAAGGTGGTGGATGAGTGCTACATCCCCAATTACTTGTGCCCAAAACCCAGCC	808
AAV 3	TGGCGCGGGGGGGCGGAACAAGGTGGTGGACGACTGCTACATCCCCAACTACCTGCTGCCCAAGACCCAGCC	804
AAV 4	TGGCGCGGAGGGCGGAACAAGGTGGTGGACGACTGCTACATCCCCAACTACCTGCTGCCCAAGACCCAGCC	859
AAV 6	TGGCGCGGAGGGGGGAACAAGGTGGTGGACGAGTGCTACATCCCCAACTACCTCCTGCCAAGACTCAGCC	807
TATA box p19		
AAAV DA-1	<u>GGAAGT</u> GCACTGGGCGTGGAATAACGTGCCCGAGTA <u>TATAAAA</u> CGCTGCCTGCACCGAGAAGCTGCCAG	800
AAAV VR 865	<u>GGAAGT</u> GCACTGGGCGTGGAATAACGTGCCCGAGTA <u>TATAAAA</u> CGCTGCCTGCACCGAGAAGCTGCCAG	800
AAV 1	CGAGCTGCAGTGGGCGTGGAATAACATGGAGGAGTATATAAGCGCCTGTTTGAACCTGGCCGAGCGCAACG	894
AAV 2	TGAGCTCCAGTGGGCGTGGAATAATGGAACAGTATTTAAGCGCCTGTTTGAATCTCACGGAGCGTAAACG	880
AAV 3	CGAGCTCCAGTGGGCGTGGAATAACATGGACAGTATTTAAGCGCCTGTTTGAATCTCGCGGAGCGTAAACG	876
AAV 4	CGAGCTCCAGTGGGCGTGGAATAACATGGACAGTATATAAGCGCCTGTTTGAATCTCGCGGAGCGTAAACG	931
AAV 6	CGAGCTGCAGTGGGCGTGGAATAACATGGAGGAGTATATAAGCGCCTGTTTAAACCTGGCCGAGCGCAACG	879
AAAV DA-1	TCTCGCGCAGCTTCACTTCGAGGAGGCGGGGCTCTCGCAATC-CAAGGAAAATCTCGCGAGAAGCTGCAGAC-	870
AAAV VR 865	TCTCGCGCAGCTTCACTTCGAGGAGGCGGGGCTCTCGCAATC-CAAGGAAAATCTCGCGAGAAGCTGCAGAC-	870
AAV 1	GCTCGTGGCGCAGCAGCTGACCCACGTGACCCAGAGCCAGGAGCAGAACAAGGAGAATCTGAACCCCAATTC	966
AAV 2	GTTGGTGGCGCAGCATCTGACGCACGTGTCGAGACGAGGAGCAGAACAAGAGAATCAGAATCCCAATTC	952
AAV 3	GCTGGTGGCGCAGCATCTGACGCACGTGTCGAGACGAGGAGCAGAACAAGAGAATCAGAATCCCAATTC	948
AAV 4	GCTGGTGGCGCAGCATCTGACGCACGTGTCGAGACGAGGAGCAGAACAAGGAAAACAGAATCCCAATTC	1003
AAV 6	GCTCGTGGCGCAGCAGCTGACCCACGTGACCCAGAGCCAGGAGCAGAACAAGGAGAATCTGAACCCCAATTC	951

Rep52/40

AAAV DA-1	-GGCGCTCCCGTGATCGCGACCCGCGTCAGCAAACGCTACATGAGAGCTCGTGGATTGGCTCGTGGAGAAGGG	941
AAAV VR 865	-GGCGCTCCCGTGATGCCGACCCGCGTCAGCAAACGCTACATGAGAGCTCGTGGATTGGCTCGTGGAGAAGGG	941
AAV 1	TGACGCGCCTGTATCCGGTCAAAAACCTCCGCGCGCTACATGGAGCTGGTCGGGTGGCTGGTGGACCGGGG	1038
AAV 2	TGATGCGCCGGTGATCAGATCAAAAACCTTCAGCCAGGTACATGGAGCTGGTCGGGTGGCTCGTGGACAAGGG	1024
AAV 3	TGACGCGCCGGTGATCAGGTCAAAAACCTTCAGCCAGGTACATGGAGCTGGTCGGGTGGCTGGTGGACCGCGG	1020
AAV 4	TGACGCGCCGGTGATCAGGTCAAAAACCTCCGCGAGGTACATGGAGCTGGTCGGGTGGCTGGTGGACCGCGG	1075
AAV 6	TGACGCGCCTGTATCCGGTCAAAAACCTCCGCGACGTACATGGAGCTGGTCGGGTGGCTGGTGGACCGGGG	1023
AAAV DA-1	GATCACCACCAGAGAAGGAATGGCTGCTGGAAAACAGAGAAAGCTTTTCGGAGCTTCCAG-GCCTCGAGCAACT	1012
AAAV VR 865	GATCACCACCAGAGAAGGAATGGCTGCTGGAAAACAGAGAAAGCTTTTCGGAGCTTTCAG-GCCTCGAGCAACT	1012
AAV 1	CATCACCTCCGAGAAGCAGTGGATCCAGGAGGACC-AGGCCTCGTACATCTCCTTCAACGCCGCTTCCAAC	1109
AAV 2	GATTACCTCGGAGAAGCAGTGGATCCAGGAGGACC-AGGCCTCATACATCTCCTTCAATGCGGCCCTCCAAC	1095
AAV 3	GATCACGTGAGAAAAGCAATGGATTCCAGGAGGACC-AGGCCTCGTACATCTCCTTCAACGCCGCTTCCAAC	1091
AAV 4	GATCACGTGAGAAAAGCAATGGATCCAGGAGGACC-AGGCCTCCTACATCTCCTTCAACGCCGCTTCCAAC	1146
AAV 6	CATCACCTCCGAGAAGCAGTGGATCCAGGAGGACC-AGGCCTCGTACATCTCCTTCAACGCCGCTTCCAAC	1094
AAAV DA-1	CGGCGCGTCAGATCAAGACGCGCCCTGCAAGGCGCCATCCAGGAGATGCTTCTGACCAAGACGGCGGAGGACT	1084
AAAV VR 865	CGGCGCGTCAGATCAAGACGCGCCCTGCAAGGCGCCATTCAGGAGATGCTTCTGACCAAGACGGCGGAGGACT	1084
AAV 1	CGCGGTCCAGATCAAGGCCGCTCTGGACAATGCCGGCAAGATCATGGCGCTGACCAAATCCGCGCCCGACT	1181
AAV 2	CGCGGTCCCAAATCAAGGCTGCCTTGGACAATGCCGGGAAAGATTATGAGCCTGACTAAAACCGCCCCGACT	1167
AAV 3	CGCGGTCCAGATCAAGGCCGCGCTGGACAATGCCCTCAAGATCATGAGCCTGACAAAGACGGCTCCGAGACT	1163
AAV 4	CGCGGTACAAATCAAGGCCGCGCTGGACAATGCCCTCAAAATCATGAGCCTGACAAAGACGGCTCCGAGACT	1218
AAV 6	CGCGGTCCAGATCAAGGCCGCTCTGGACAATGCCGGCAAGATCATGGCGCTGACCAAATCCGCGCCCGACT	1166
AAAV DA-1	ACCTCGTCGGAAAAGAACCCGTCTCGGACGACGAGATTCTGTCAGAACCCGCATCTACAAGATTTTGGAACTGA	1156
AAAV VR 865	ACCTCGTCGGAAAAGGATCCCGTCTCGGACGACGACATCCGTCAGAACCCGCATCTACAAGATTTTGGAACTGA	1156
AAV 1	ACCTGGTAGGCCC-CGCTCCGCC-----CGCGGACATTAACCAACCCGCATCTACCGCATCTCTGGAGCTGA	1247
AAV 2	ACCTGGTGGGCCAGCAGCCCGTG-----GAGGACATTTCCAGCAATCGGATTTATAAAATTTTGGAACTAA	1233
AAV 3	ACCTGGTGGGC-AGCAACCCGCC-----GGAGGACATTACCAAAAATCGGATCTACCAAAATCTCTGGAGCTGA	1229
AAV 4	ACCTGGTGGGCCAG-AACCCGCC-----GGAGGACATTTCCAGCAACCCGCATCTACCGAATCTCTGAGATGA	1284
AAV 6	ACCTGGTAGGCCC-CGCTCCGCC-----CGCGGACATTAACCAACCCGCATTTACCGCATCTCTGGAGCTGA	1232
AAAV DA-1	ACCACTACGACCCAGCGTACGTGGGGAGTATTTTGGTCGGGTGGTGCCAGAAGAAATGGGGCAAGCGAAACA	1228
AAAV VR 865	ACCACTACGACCCAGCGTACGTGGGGAGTATTTTGGTCGGGTGGTGCCAGAAGAAATGGGGCAAGCGAAACA	1228
AAV 1	ACGGGTACGAACCTGCCTACGCCGGCTCCGTCTTTCTCGGCTGGGCCAGAAAAGTTTCGGGAAGCGCAACA	1319
AAV 2	ACGGGTACGATCCCCAATATGCGGCTTCCGTCTTTCTGGGATGGGCCACGAAAAGTTTCGGGAAGAGGAACA	1305
AAV 3	ACGGGTACGATCCGCAGTACGCGGCCTCCGTCTTCTGCGGTGGGCCGAAAAGAAAGTTTCGGGAAGAGGAACA	1301
AAV 4	ACGGGTACGATCCGCAGTACGCGGCCTCCGTCTTCTGCGGTGGGCCGAAAAGAAAGTTTCGGGAAGAGGAACA	1356
AAV 6	ACGGGTACGAACCTGCCTACGCCGGCTCCGTCTTTCTCGGCTGGGCCAGAAAAGTTTCGGAAAACGCAACA	1304
AAAV DA-1	CGCTGTGGCTGTTTCGGACCTGCGACACCAGGCAAGACCAACATCGCGGAGGCTATTGCCCATGCTGTGCCGT	1300
AAAV VR 865	CGCTGTGGCTGTTTCGGACATGCGACACCAGGCAAGACCAACATCGCGGAGGCTATTGCCCATGCTGTGCCGT	1300
AAV 1	CCATCTGGCTGTTTGGGCCGGCCACCACGGGCAAGACCAACATCGCGGAAGCCATCGCCCACGCCGTGCCCT	1391
AAV 2	CCATCTGGCTGTTTGGGCCCTGCAACTACCGGGAAGACCAACATCGCGGAGGCCATAGCCCACACTGTGCCCT	1377
AAV 3	CCATCTGGCTGTTTGGGCCGGCCACGACGGGTAAACCAACATCGCGGAAGCCATCGCCCACGCCGTGCCCT	1373
AAV 4	CCATCTGGCTGTTTGGGCCGGCCACGACGGGTAAACCAACATCGCGGAAGCCATCGCCCACGCCGTGCCCT	1428
AAV 6	CCATCTGGCTGTTTGGGCCGGCCACCACGGGCAAGACCAACATCGCGGAAGCCATCGCCCACGCCGTGCCCT	1376
AAAV DA-1	TCTATGGATGCGTTAACTGGACCAACGAGAACTTTCCGTTCAACGACTGCGTCGAAAAGATGATTATCTGGT	1372
AAAV VR 865	TCTATGGATGCGTTAACTGGACCAACGAGAACTTTCCGTTCAACGACTGCGTCGAAAAGATGATTATCTGGT	1372
AAV 1	TCTACGGCTGCGTCAACTGGACCAATGAGAACTTTCCCTTCAATGATTGCGTCGACAAGATGGTGATCTGGT	1463
AAV 2	TCTACGGGTGCGTAACTGGACCAATGAGAACTTTCCCTTCAACGACTGTGTCGACAAGATGGTGATCTGGT	1449
AAV 3	TCTACGGCTGCGTAACTGGACCAATGAGAACTTTCCCTTCAACGATTGCGTCGACAAGATGGTGATCTGGT	1445
AAV 4	TCTACGGCTGCGTAACTGGACCAATGAGAACTTTCCCTTCAACGATTGCGTCGACAAGATGGTGATCTGGT	1500
AAV 6	TCTACGGCTGCGTCAACTGGACCAATGAGAACTTTCCCTTCAACGATTGCGTCGACAAGATGGTGATCTGGT	1448
AAAV DA-1	GGGAAGAGGGCAAAATGACCGCCAAGGTGGTGGAAACGGCTAAAGCGATTCTGGGAGGATCTCGGGTGAGAG	1444
AAAV VR 865	GGGAGGAGGGCAAAATGACCGCCAAGGTGGTGGAAACAGCCAAGGCGATTCTGGGAGGATCTCGGGTGAGAG	1444
AAV 1	GGGAGGAGGGCAAGATGACCGCCAAGGTTCGTGGAGTCCGCCAAGGCCATTCTCGGCGGCAGCAAGGTGCGCG	1535
AAV 2	GGGAGGAGGGGAAGATGACCGCCAAGGTTCGTGGAGTCCGCCAAGGCCATTCTCGGAGGAAGCAAGGTGCGCG	1521
AAV 3	GGGAGGAGGGCAAGATGACCGCCAAGGTTCGTGGAGAGCGCCAAGGCCATTCTGGGCGGAAGCAAGGTGCGCG	1517
AAV 4	GGGAGGAGGGCAAGATGACCGCCAAGGTTCGTAGAGAGCGCCAAGGCCATTCTGGGCGGAAGCAAGGTGCGCG	1572
AAV 6	GGGAGGAGGGCAAGATGACCGCCAAGGTTCGTGGAGTCCGCCAAGGCCATTCTCGGCGGCAGCAAGGTGCGCG	1520

AAAV DA-1	TGGACCAAAATGCAAAAGCTTCGGTTCCGATCGAACCGACGCCGGTCAATTATTACCAGTAACACCAACATGT	1516
AAAV VR 865	TGGACCAAAATGCAAAAGCTTCGGTTCCGATCGAACCGACGCCGGTCAATTATTACCAGTAACACCAACATGT	1516
AAV 1	TGGACCAAAAGTGCAAGTCGTCCGCCAGATCGACCCACCCCGTGATCGTCACCTCCAACACCAACATGT	1607
AAV 2	TGGACCAAGAAATGCAAGTCCTCGGCCAGATAGACCCGACTCCCGTGATCGTCACCTCCAACACCAACATGT	1593
AAV 3	TGGACCAAAAGTGCAAGTCATCGGCCAGATCGAACCCACTCCCGTGATCGTCACCTCCAACACCAACATGT	1589
AAV 4	TGGACCAAAAGTGCAAGTCATCGGCCAGATCGAACCCACTCCCGTGATCGTCACCTCCAACACCAACATGT	1644
AAV 6	TGGACCAAAAGTGCAAGTCGTCCGCCAGATCGATCCCACCCCGTGATCGTCACCTCCAACACCAACATGT	1592

AAAV DA-1	GTTATGTCATCGACGGGAACACGACTACGTTTCGAGCATAAGCAGCCGTTGGAGGACAGGATGTTTAAAGCTCG	1588
AAAV VR 865	GTTATGTCATCGACGGGAACACGACTACGTTTCGAGCATAAGCAGCCGTTGGAGGACAGGATGTTTAAAGCTCG	1588
AAV 1	GCGCCGTGATTGACGGGAACAGCACCACCTTCGAGCACCAGCAGCCGTTGCAGGACCGGATGTTCAAATTTG	1679
AAV 2	GCGCCGTGATTGACGGGAACCAACGACCTTCGAACACCAGCAGCCGTTGCAAGACCGGATGTTCAAATTTG	1665
AAV 3	GCGCCGTGATTGACGGGAACAGCACCACCTTCGAGCATCAGCAGCCGTTGCAGGACCGGATGTTTAAATTTG	1661
AAV 4	GCGCCGTGATCGACGGAACTCGACCACTTCGAGCACCACCAACCACTCCAGGACCGGATGTTTCAAGTTGG	1716
AAV 6	GCGCCGTGATTGACGGGAACAGCACCACCTTCGAGCACCAGCAGCCGTTGCAGGACCGGATGTTCAAATTTG	1664

AP-1

AAAV DA-1	AGCTGTTGACTCGTTTACCCGATGACTTTGGTAAAGGTGACCAACAGGAGGTGCGTCAATTCTTCAGGTGGT	1660
AAAV VR 865	AATTGCTGACTCGGTTGCCTGATGACTTTGGTAAAGGTGACCAACAGGAGGTGCGTCAATTCTTCAGGTGGT	1660
AAV 1	AACTACCCCGCGTCTGGAGCATGACTTTGGCAAGGTGACAAAGCAGGAAGTCAAGAGTCTTCCGCTGGG	1751
AAV 2	AACTACCCCGCGTCTGGATCATGACTTTGGGAAGGTACCAAGCAGGAAGTCAAGACTTTTCCGCTGGG	1737
AAV 3	AACTTACCCCGCGTTTGGACCATGACTTTGGGAAGGTACCAAAACAGGAAGTAAAGGACTTTTCCGCTGGG	1733
AAV 4	AGCTCACCAGCGCTGGAGCAGACTTTGGCAAGGTACCAAGCAGGAAGTCAAGACTTTTCCGCTGGG	1788
AAV 6	AACTACCCCGCGTCTGGAGCATGACTTTGGCAAGGTGACAAAGCAGGAAGTCAAGAGTCTTCCGCTGGG	1736

EF-1

AAAV DA-1	CCCAGGATCACCTGACCCCTGTGATCCCAGAATTCTTGGTGCAGGAGCGGAGTCTCGCAAAAGACCCGCC	1732
AAAV VR 865	CTCAGGATCACCTGACCCCTGTGATCCCAGAATTCTTAGTGCAGGAGCGGAGTCTCGCAAAAGACCCGCC	1732
AAV 1	CGCAGGATCACGTGACCGAGGTGGCGCATGAGTTCTACGTGAGAAAGGTTGGAGCCAACAAAAGACCCGCC	1823
AAV 2	CAAAGGATCAGGTGTTGAGGTGGAGCATGAATTCTACGTCAAAAGGTTGGAGCCAAGAAAAGACCCGCC	1809
AAV 3	CTTCCGATCAGTGACTGACGTGGCTCATGAGTTCTACGTGAGAAAGGTTGGAGCTAAGAAACGCCCGCCT	1805
AAV 4	CGTCAGATCAGTGACCGAGGTGACTCACGAGTTTACGTGAGAAAGGTTGGAGCTAGAAAGAGGCCGCC	1860
AAV 6	CGCAGGATCACGTGACCGAGGTGGCGCATGAGTTCTACGTGAGAAAGGTTGGAGCCAACAGAGACCCGCC	1808

TATA box p40

p40

AAAV DA-1	CCTCCGGGGAAGGC-TATATATAAGCCCGA-CAAAGCGGCCCG-CGCTCGCAGAGC---AGCAGCAGGCGTCG	1797
AAAV VR 865	CTTCCGGGGAAGGC-TATATATAAGCCCGA-CAAAGCGGCCCG-CGCTCGCAGAGC---AGCAGCAGGCGTCG	1797
AAV 1	CCGATGACGCGGATAAAAGCGAGCCCAAGCGGCCCTGCCCTCAGTCGCGGATCCATCGACGTGACAGCGCG	1895
AAV 2	CCAGTGACGCAGATATAAGTGAGCCCAACGGGTGCGCGAGTCAGTTGCGCAGCCATCGACGTGACAGCGCG	1881
AAV 3	CCAATGACGCGGATGTAAGCGAGCCAAAACGGCAGTGACGTCAGTTGCGCAGCCGACAAAGTCAGACGCGG	1877
AAV 4	CCAATGACGCAGATATAAGTGAGCCCAAGCGGCCCTGTCCGTGAGTTGCGCAGCCATCGACGTGACAGCGCG	1932
AAV 6	CCGATGACGCGGATAAAAGCGAGCCCAAGCGGCCCTGCCCTCAGTCGCGGATCCATCGACGTGACAGCGCG	1880

Splice donor

AAAV DA-1	GAG-AGCGCGGAGCCGGTTC-CCACCAGGTATCGTATCAAAATGCTCGAAACATTGCGGGATGGATAAAATGT	1867
AAAV VR 865	GAG-AGCGCGGAGCCGGTTC-CCACCAGGTATCGTATCAAAATGCTCGAAACATTGCGGGATGGATAAAATGT	1867
AAV 1	AAGGAGCTCCGGTGGACTTTGCCGACAGGTACCAAAACAAATGTTCTCGTCACGCGGGCATGCTTCAGATGC	1967
AAV 2	AAG---CTTCGATCAACTACGAGACAGGTACCAAAACAAATGTTCTCGTCACGTGGGCATGAATCTGATGC	1950
AAV 3	AAGCA---CCGGCGGACTACGCGGACAGGTACCAAAACAAATGTTCTCGTCACGTGGGCATGAATCTGATGC	1946
AAV 4	A---AGCTCCGGTGGACTACGCGGACAGGTACCAAAACAAATGTTCTCGTCACGTGGGTATGAATCTGATGC	2001
AAV 6	AAGGAGCTCCGGTGGACTTTGCCGACAGGTACCAAAACAAATGTTCTCGTCACGCGGGCATGCTTCAGATGC	1952

AAAV DA-1	TGTTTCCTTGCCAAATTTGTGAATCGATGAACAGAAATATTAATATATGTGCGATTATATAAAACACGGAAT	1939
AAAV VR 865	TGTTTCCTTGCCAAATTTGTGAATCGATGAACAGAGATATTAATATTTGTGCTATTATATAAAACGACCGACT	1939
AAV 1	TGTTTCCTTGCAAGACATGCGAGAGAATGAATCAGAATTTCAACATTTGCTTCACGCACGGGACGAGAGACT	2039
AAV 2	TGTTTCCTTGACAGCAATGCGAGAGAATGAATCAGAATTTCAATATCTGCTTCACGCACGAGAGAAAGACT	2022
AAV 3	TTTTTCCTTGTAACATGCGAGAGAATGAATCAAAATTTCAATGTCTGTTTTACGCATGGTCAAAGAGACT	2018
AAV 4	TTTTTCCTTGCGCAATGCGAGAGAATGAATCAGAATTTGGACATTTGCTTCACGCACGGGGTATGGACT	2073
AAV 6	TGTTTCCTTGCAAAACATGCGAGAGAATGAATCAGAATTTCAACATTTGCTTCACGCACGGGACGAGAGACT	2024

AAAV DA-1	GTAAGAATGTTTTCCGGAATACGGGGATAAGGACACTGTGCCAGAACTACCCCTGTACAGAACATAACG	2011
AAAV VR 865	GTAAGAGTGTTTCCCGGACTACGGGGATAAAGATGATGT---AGAACTACCCCTGTACAGAACACAACG	2008
AAV 1	GTTTCAGAGTGCTTCCCGGC-----GTGTCAGAAT-----CT-----CAACC	2076
AAV 2	GTTTAGAGTGCTTCCCG-----TGTCAGAAT-----CT-----CAACC	2056
AAV 3	GTGGGAATGCTTCCCTGGA-----ATGTCAGAAT-----CT-----CAACC	2055
AAV 4	GTGCCGAGTGCTTCCCG-----TGTCAGAAT-----CT-----CAACC	2107
AAV 6	GTTTCAGAATGTTTCCCGGC-----GTGTCAGAAT-----CT-----CAACC	2061

AAAV DA-1	TGTCTCGTTGTTATCAATGTCATTTCGGGCGAATTGTATCGCGTGACTTCGGACTCTGACGAGAAACCCGCC	2083
AAAV VR 865	TGTCTCGTTGTTATCAATGTCATTTCGGGCGAATTGTATCGCGTGACTTCGGACTCTGACGAGAAACCTGCC	2080
AAV 1	GGT-----CGTCAGAA-----AGAGGA--CGTATC-----GGAAACTC----	2107
AAV 2	CGTTTC--TGTCGTCAGAA-----AGGCG-----TATC-----AGAAACTG----	2090
AAV 3	CGTTTC--TGTCGTCAGAA-----AGAAGA--CTTATC-----AGAAACTG----	2092
AAV 4	CGTGTCT--TGTCGTCAGAA-----AGCGGA--CGTATC-----AGAAACTG----	2144
AAV 6	GGTC-----GTCAGAAAGA-----GGACG-----TATC-----GGAAACTC----	2092
AAAV DA-1	CCGAGAGTGATGAAGGCACTGAGCCATCCTATGCTCCCTGCACGATTACCATTTGATGGGCAAGAGTCGCG	2155
AAAV VR 865	CCGAGAGTGATGAAGGCACCGAGCCATCCTATGCTCCCTGCACGATTACCACCTGATGGGCAAGAGTCACG	2152
AAV 1	-----TGTCGCATTCATCATCTGCTGGGGCGGGCTCCCG	2141
AAV 2	-----TGCTACATTCATCATATCATGGGAAAGGTGCCAG	2124
AAV 3	-----TGTCGAATTCATCATATCCTGGGAAGGGCACCCG	2126
AAV 4	-----TGTCGATTCATCATATCATGGGAGGGCGCCCG	2178
AAV 6	-----TGTCGCATTCATCATCTGCTGGGGCGGGCTCCCG	2126
AAAV DA-1	GGTTAGTGTCTGTGCTGCATGTCGTTTGAAAAATAGTACGTTGCATGATGACTTGGATGACGGTGATCTTG	2227
AAAV VR 865	GGTTAGTCACTTGCGCGCGGTGTCGGTTTGAAAAATAGTACGTTGCATGATGACTTGGATGACGGTGATCTCG	2224
AAV 1	AG---ATTGCTTGCTCGGCCTGCGATCTGGTCAACG-----TG--GACCTGGATGACTGTGTTTCTG	2198
AAV 2	A-----CGCTGCACTGCCTGCGATCTGGTCAATG-----TG--GATTTGGATGACTGCATCTTTG	2178
AAV 3	AG---ATTGCTGTTTCGGCCTGCGATTTGGCCAATG-----TG--GACTTGGATGACTGTGTTTCTG	2183
AAV 4	AG---GTGGCCTGCTCGGCCTGCGAAGTGGCCAATG-----TG--GACTTGGATGACTGTGACATGG	2235
AAV 6	AG---ATTGCTTGCTCGGCCTGCGATCTGGTCAACG-----TG--GATCTGGATGACTGTGTTTCTG	2183
<div> <div>Minor splice</div> <div> <div>Stop Rep78</div> <div>VP1</div> <div>Splice acceptor</div> </div> </div>		
AAAV DA-1	AACAATTAATGATTGAAATATAGCCATGCTCTCTCATTTCTGATGCGATTCCAGATTGGTTGGAGCGGTTGGT	2299
AAAV VR 865	AACAATTAATGATTGAAATGATAGCCATGCTCTCTCATTTCTGATGCGATTCCAGATTGGTTGGAGCGGTTGGT	2296
AAV 1	AGCAATAAATGACTTAAAC-CAGGTATGGCTGCCGATGGTTATC---TTCCAGATTGGCTCGAGGACAACCT	2266
AAV 2	AACAATAAATGATTAAAT-CAGGTATGGCTGCCGATGGTTATC---TTCCAGATTGGCTCGAGGACAACCT	2246
AAV 3	AGCAATAAATGACTTAAAC-CAGGTATGGCTGCTGACGGTTATC---TTCCAGATTGGCTCGAGGACAACCT	2251
AAV 4	AACAATAAATGACTCAAAC-CAGATATGACTG---ACGGTTACC---TTCCAGATTGGCTAGAGGACAACCT	2300
AAV 6	AGCAATAAATGACTTAAAC-CAGGTATGGCTGCCGATGGTTATC---TTCCAGATTGGCTCGAGGACAACCT	2251
<div> <div>Stop Rep68</div> </div>		
AAAV DA-1	CAAAAAGGGAGTGAATGCTGCGGCTGATTTTCTACCATTTGGAAAGCGGTCTCCTCATCCTAAGGCAAATCA	2371
AAAV VR 865	CAAAAAGGGAGTGAATGCTGCAGCTGATTTTCTACCATTTGGAAAGCGGTCTCCTCGTCTAAGGCAAATCA	2368
AAV 1	CTCTGAGGGCAT-----TCGCGAGTGGT---GGGACTTGAAACCTGGAGCCCCGAAGCCAAAGCCAACCA	2329
AAV 2	CTCTGAAGGAAT-----AAGACAGTGGT---GGAAGCTCAAACCTGGCCACCACCACCAAGCCCGCAGA	2309
AAV 3	TTCTGAAGGCAT-----TCGTGAGTGGT---GGGCTCTGAAACCTGGAGTCCCTCAACCCAAAGCGAACCA	2314
AAV 4	CTCTGAAGGCGT-----TCGAGAGTGGT---GGGCGCTGCAACCTGGAGCCCCTAAACCCAAAGGCAAATCA	2363
AAV 6	CTCTGAGGGCAT-----TCGCGAGTGGT---GGGACTTGAAACCTGGAGCCCCGAAGCCAAAGCCAACCA	2314
AAAV DA-1	GCAAACCTCAAGAATCTCCTGAAAAGGACGATTTCGAGAGGTCTCGTGTTCCCGGGTTACAAGTATCTAGGCC	2443
AAAV VR 865	GCAAACCTCAAGAATCTCTTGAAAAGGACGATTTCGAGAGGTCTCGTGTTCCAGGCTACAATTATCTAGGCC	2440
AAV 1	GCAAA-----AGCAGGACGACGCGCGGGGTCTGGTGCTTCTCGGCTACAAGTACCTCGGACC	2386
AAV 2	GCGGC-----ATAAGGACGACAGCGGGGTCTTGTGCTTCTCGGCTACAAGTACCTCGGACC	2366
AAV 3	ACAAC-----ACCAGGACAACCGTCGGGGTCTTGTGCTTCCGGGTACAATAACCTCGGACC	2371
AAV 4	ACAAC-----ATCAGGACAACGCTCGGGGTCTTGTGCTTCCGGGTACAATAACCTCGGACC	2420
AAV 6	GCAAA-----AGCAGGACGACGCGCGGGGTCTGGTGCTTCTCGGCTACAAGTACCTCGGACC	2371
AAAV DA-1	TTTCAACGGTCTAGATAAAGGAAAACCCGTCACAGAGGCAGACGCTGCCGCCCTTAGAGCAGCACAAGGCTTA	2515
AAAV VR 865	TTTCAACGGTCTAGATAAAGGAGAAACCCGTCACAGAGGCAGACGCTGCCGCCCTTAGAACACGACAAGGCTTA	2512
AAV 1	CTTCAACGGACTCGACAAGGGGAGCCCGTCAACGCGCGGACGACGCGGCCCTCGAGCAGCACAAGGCCCTA	2458
AAV 2	CTTCAACGGACTCGACAAGGGAGAGCCGGTCAACAGAGGCAGACGCGCGGCCCTCGAGCAGCACAAGGCCCTA	2438
AAV 3	CGGTAACGGACTCGACAAGGAGAGCCGGTCAACAGAGGCAGCGGCGGCGGCCCTCGAACACGACAAGGCTTA	2443
AAV 4	CGGCAACGGACTCGACAAGGGGAAACCCGTCACAGCAGCGGACGCGGCCCTCGAGCAGCACAAGGCCCTA	2492
AAV 6	CTTCAACGGACTCGACAAGGGGAGCCCGTCAACGCGCGGATGCAGCGGCCCTCGAGCAGCACAAGGCCCTA	2443
AAAV DA-1	CGACCTCGAACTCAAGGACGGGC-ACAACCCGTACTTTGAGTACAACGAGGCCGATAGACGTTTCCAGGAAC	2586
AAAV VR 865	CGACCTCGAACTCAAGGACGGGC-ACAACCCGTACTTTGAGTACAACGAGGCCGACAGCGTTTCCAGGAAC	2583
AAV 1	CGACACGACGCTCAA-AGCGGGTGACAATCCGTACCTGCGGTATAACCACGCCGACGCCGAGTTTTCAGGAGC	2529
AAV 2	CGACCGGCAGCTCGACAGCGGA-GACAACCCGTACCTCAAGTACAACCACGCCGACGCCGAGTTTTCAGGAGC	2509
AAV 3	CGACACGACGCTCAA-GGCCGGTGACAACCCGTACCTCAAGTACAACCACGCCGACGCCGAGTTTTCAGGAGC	2514
AAV 4	CGACACGACGCTCAA-GGCCGGTGACAACCCGTACCTCAAGTACAACCACGCCGACGCCGAGTTTTCAGGAGC	2563
AAV 6	CGACACGACGCTCAA-AGCGGGTGACAATCCGTACCTGCGGTATAACCACGCCGACGCCGAGTTTTCAGGAGC	2514

AAAV DA-1	GTCTCAAAGACGATACCTCTTTTCGGAGGGAATCTCGTTAAAGCCATCTTTTCAGGCCAAAAAGAGGGTCTCTCG	2658
AAAV VR 865	GTCTCAAAGACGATACCTCTTTTCGGAGGCAATTTAGGTAAAGCCATCTTTTCAGGCCAAAAAGAGGGTCTCTCG	2655
AAV 1	GTCTGCAAGAAGATACGTCTTTTGGGGGCAACCTCGGGCGAGCAGTCTTCCAGGCCAAAAAGAGGGTCTCTCG	2601
AAV 2	GCCTTAAAGAAGATACGTCTTTTGGGGGCAACCTCGGACGAGCAGTCTTCCAGGCCAAAAAGAGGGTCTCTTG	2581
AAV 3	GTCTTCAAGAAGATACGTCTTTTGGGGGCAACCTTGGCAGAGCAGTCTTCCAGGCCAAAAAGAGGATCCTTG	2586
AAV 4	GGCTTCAGGGCGACACATCGTTTGGGGGCAACCTCGGCAGAGCAGTCTTCCAGGCCAAAAAGAGGGTCTCTTG	2635
AAV 6	GTCTGCAAGAAGATACGTCTTTTGGGGGCAACCTCGGGCGAGCAGTCTTCCAGGCCAAAAAGAGGGTCTCTCG	2586

VP2

AAAV DA-1	AGCCCTTTGGTCTGATTGAACAACCCG--ACAAC <u>ACG</u> ----GCCGG-GACCGG-GGAGAAGCGTCC-CGAAC	2721
AAAV VR 865	AACCCCTTTGGTCTGGTGGA--GACTCA--AAG <u>ACG</u> GCTCCGACCGGAGACAAGCGGAAAGGCGAAGACGAAC	2724
AAV 1	AACCTCTCGGTCTGGTTGAGGAAGCGCTAAGACGGCTC---CTGGAAAGAAACGTCCGGTAGAGCAGTCGC	2670
AAV 2	AACCTCTGGGCTGGTTGAGGAACCTGTTAAGACGGCTC---CGGGAAAAAGAGGCCGGTAGAGCACTCTC	2650
AAV 3	AGCCTCTGGTCTGGTTGAGGAAGCAGCTAAAACGGCTC---CTGGAAAGAAGAGGCCCTGTAGATCAGTCTC	2655
AAV 4	AACCTCTTGGTCTGGTTGAGCAAGCGGGTGAACGGCTC---CTGGAAAGAAGAGACCGTTGATTGAATCCC	2704
AAV 6	AACCTTTTGGTCTGGTTGAGGAAGGTGCTAAGACGGCTC---CTGGAAAGAAACGTCCGGTAGAGCAGTCGC	2655

AAAV DA-1	G-CGTCGAC--GACTTTTCCCGAAAAAG---AAGAAGGCCAAGA--CCGAGCAAGGCCAAAGCCCCCTGCTCA	2785
AAAV VR 865	CTCGTTTGGCCGACACTTCTTCACAGACTCCCAAGAAAAACAAGAAGCCTCGCAAGG-AAAGACCTT--CCG	2793
AAV 1	CACAAGAGCCAGACTCCTCCTCGGGCATCGGCAAGACAGGCCAGCAGCCCCGCTAAAAAGAGACTCAATTTTG	2742
AAV 2	CTGTGGAGCCAGACTCCTCCTCGGGAACCGGAAAGGCGGGCCAGCAGCCTGCAAGAAAAAGATTGAATTTTG	2722
AAV 3	CTCAGGAACCGGACTCATCATCTGGTTGGCAAAATCGGGCAAAACAGCCTGCCAGAAAAAGACTAAATTTTCG	2727
AAV 4	CCCAGCAGCCGACTCCTCCACGGGTATCGGCAAAAAGGCAAGCAGCCGGCTAAAAAGAAGCTCGTTTTCG	2776
AAV 6	CACAAGAGCCAGACTCCTCCTCGGGCATTGGCAAGACAGGCCAGCAGCCCCGCTAAAAAGAGACTCAATTTTG	2727

AAAV DA-1	AACGGGC-GAAGACCCCGGAGAAGGAACCT-CTTCCAAC--GCTGGATCA-----AGCGCCC-CCTCTAGTG	2847
AAAV VR 865	GCGGGGCAGAGATCCGGGCGAAGGCACCT-CTTCCAAC--GCTGGAGCAGC---AGCACCCGCCCTCTAGTG	2859
AAV 1	GTCAGACTGGCGACTCAGAGTCAGTCCCCGATCCACAACCTCTCGGAGAACCTCCAGCAACCCCCGCTGCTG	2814
AAV 2	TGGGACCTACTACAATGGCTTCAGGCGGTGGCGCACCAATGGCAGACAATAACGAAGGCGCCGACGGAGTGG	2794
AAV 3	GTCAGACTGGCGACTCAGAGTCAGTCCCAGACCTCAACCTCTCGGAGAACCACCAGCAGCCCCCACAAGTT	2799
AAV 4	A--AGAC--GAACT--GGAGCAGGCACGGACCC---CCTGAGGGATCAACTTCCGGAGCCATGTCTGAT-	2838
AAV 6	GTCAGACTGGCGACTCAGAGTCAGTCCCCGACCCACAACCTCTCGGAGAACCTCCAGCAACCCCCGCTGCTG	2799

VP3

AAAV DA-1	TGGGATCATCTGTCA <u>ATG</u> GCTGAAGGAGGTGGCGGTCCAATGGGCGATGCAGGCCAAGGTGCCGACGGAGTGG	2919
AAAV VR 865	TGGGATCATCTATCA <u>ATG</u> GCTGAAGGAGGTGGCGGCCAGTGGGCGATGCAGGCCAGGGTGCCGATGGAGTGG	2931
AAV 1	TGGGACCTACTACAATGGCTTCAGGCGGTGGCGCACCAATGGCAGACAATAACGAAGGCGCCGACGGAGTGG	2886
AAV 2	TGGGAATAATAACGATGGCTACAGGCAGTGGCGCACCAATGGCAGACAATAACGAGGGCGCCGACGGAGTGG	2866
AAV 3	TGGGATCTAATAACAATGGCTTCAGGCGGTGGCGCACCAATGGCAGACAATAACGAGGGTGCCGATGGAGTGG	2871
AAV 4	---GAC--AGTGAGATGCGTGCAGCAGTGGCGGAGCTGCAGTCGAGGGCGGACAAGGTGCCGATGGAGTGG	2905
AAV 6	TGGGACCTACTACAATGGCTTCAGGCGGTGGCGCACCAATGGCAGACAATAACGAAGGCGCCGACGGAGTGG	2871

AAAV DA-1	GCAATTCTCTCGGGAATTGGCATTGCGATTCCCAATGGCTGGACAACGGAGTCGTTACCCGAACCACTCGAA	2991
AAAV VR 865	GCAATTCTCTCGGAAATTGGCATTGCGATTCCCAATGGCTGGAAAACGGAGTCGTCACCTCGAACCACCCGAA	3003
AAV 1	GTAATGCCTCAGGAAATTGGCATTGCGATTCCACATGGCTGGGCGACAGAGTCATCACCACCAGCACCCGCA	2958
AAV 2	GTAATTCCTCTCGGGAATTGGCATTGCGATTCCACATGGATGGGCGACAGAGTCATCACCACCAGCACCCGAA	2938
AAV 3	GTAATTCCTCAGGAAATTGGCATTGCGATTCCCAATGGCTGGGCGACAGAGTCATCACCACCAGCACCCGAA	2943
AAV 4	GTAATGCCTCGGGTATTGGCATTGCGATTCCACCTGGTCTGAGGGCCACGTACGACCACCAGCACCCGAA	2977
AAV 6	GTAATGCCTCAGGAAATTGGCATTGCGATTCCACATGGCTGGGCGACAGAGTCATCACCACCAGCACCCGAA	2943

AAAV DA-1	CCTGGGTCTTGCCACGCTACAACAACCACCTTGTAACAAGCGGATC-CAAGGACCCGGA--GGAACCGACCCCA	3060
AAAV VR 865	CCTGGGTCTTGCCACGCTACAACAACCACCTGTACAACGAATC-CAAGGACCCAGC--GGAGGCGACAACA	3072
AAV 1	CCTGGGCCTTGCCACCTACAATAACCACCTCTACAAGCAAATCTCCAGTGCTTCAACGGGGGCCAGCAACG	3030
AAV 2	CCTGGGCCCTTGCCACCTACAACAACCACCTCTACAACAAATTTCCAGC---CAATCAGGAGCCTCGAACG	3007
AAV 3	CCTGGGCCCTTGCCACTTACAACAACCACCTCTACAAGCAAATCTCCAGC---CAATCAGGAGCTTCAACG	3012
AAV 4	CCTGGTCTTGCCACCTACAACAACCACCTCTACAAGCGA-----CTCGGA-GAGAGCTCTGCAATC	3038
AAV 6	CATGGGCCTTGCCACCTATAACAACCACCTCTACAAGCAAATCTCCAGTGCTTCAACGGGGGCCAGCAACG	3015

AAAV DA-1	ACAATAAATTCTTTGGATTACGACCCCCCTGGGGGTACTTTGACTACAACCGATTCCACTGCCACTTCTCCC	3132
AAAV VR 865	ACAACAATAATTCTTTGGATTACGACCCCCCTGGGGTACTTTGACTACAATCGATTCCACTGCCACTTTTCCC	3144
AAV 1	ACAACCACTACTTCGGCTACAGCACCCCCCTGGGGGTATTTTGATTTCACAGATTCCACTGCCACTTTTTCAC	3102
AAV 2	ACAATCACTACTTTGGCTACAGCACCCCCCTGGGGGTATTTTGACTTCAACAGATTCCACTGCCACTTTTTCAC	3079
AAV 3	ACAACCACTACTTTGGCTACAGCACCCCCCTGGGGGTATTTTGACTTCAACAGATTCCACTGCCACTTCTCAC	3084
AAV 4	-CAACCACTACAACGATTCTCCACCCCCCTGGGGTACTTTGACTTCAACCGCTTCCACTGCCACTTCTCAC	3109
AAV 6	ACAACCACTACTTCGGCTACAGCACCCCCCTGGGGGTATTTTGATTTCACAGATTCCACTGCCACTTCTCAC	3087

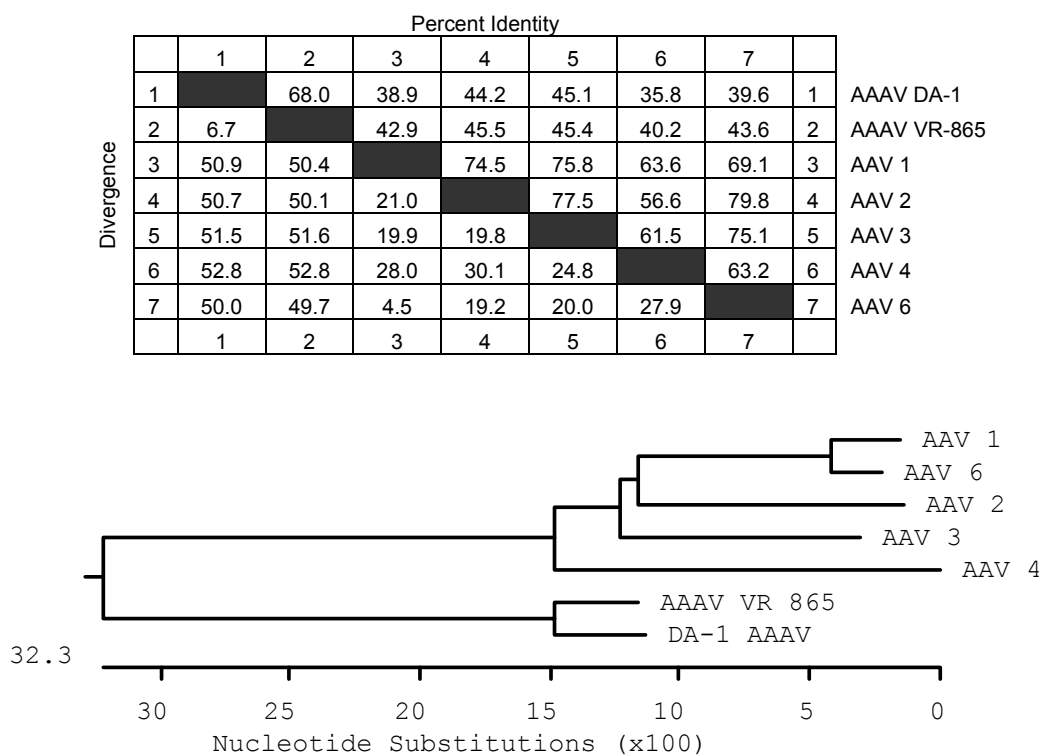
AAAV DA-1	CCCAGACTGGCAACGACTCATCAACAACAACCTGGGGCATCCGACCCAAAGCGATGCGCTTTAGACTCTTTA	3204
AAAV VR 865	CGCGAGACTGGCAACGACTCATCAACAACAACCTGGGGCATCCGTCCCAAGCGATGCGCTTTAGACTCTTTA	3216
AAV 1	CACGTGACTGGCAGCGACTCATCAACAACAATTGGGGATTCCGGCCCAAGAGACTCAACTTCAAACTCTTCA	3174
AAV 2	CACGTGACTGGCAAGACTCATCAACAACAACCTGGGGATTCCGACCCAAAGAGACTCAACTTCAAGTCTTTA	3151
AAV 3	CACGTGACTGGCAGCGACTCATTAACAACAACCTGGGGATTCCGGCCCAAGAAACTCAGCTTCAAGTCTTCA	3156
AAV 4	CACGTGACTGGCAGCGACTCATCAACAACAACCTGGGGCATGCGACCCAAAGCCATGCGGGTCAAAATCTTCA	3181
AAV 6	CACGTGACTGGCAGCGACTCATCAACAACAATTGGGGATTCCGGCCCAAGAGACTCAACTTCAAGCTCTTCA	3159
AAAV DA-1	ACATCCAGGTTAAAGAAGTCACTGTCCAAGACTCCAACACCACCATCGCCAACAACCTCACCAGCACGGTCC	3276
AAAV VR 865	ACATCCAGGTTAAAGAAGGTACGGTCCAAGACTTCAACACCACCATCGGCAACAACCTCACCAGTACGGTCC	3288
AAV 1	ACATCCAAGTCAAGGAGGTACGACGAATGATGGCGTCACAACCATCGCTAATAACCTTACCAGCACGGTTC	3246
AAV 2	ACATTCAAGTCAAAGAGGTACGCGAGAATGACGGTACGACGACGATTGCCAATAACCTTACCAGCACGGTTC	3223
AAV 3	ACATCCAAGTTAAAGAGGTACGCGAGAATGATGGCAGCAGACTATTGCCAATAACCTTACCAGCACGGTTC	3228
AAV 4	ACATCCAGGTCAAGGAGGTACGACGTGACCGGCGAGACAACGGTGGCTAATAACCTTACCAGCACGGTTC	3253
AAV 6	ACATCCAAGTCAAGGAGGTACGACGAATGATGGCGTCACGACCATCGCTAATAACCTTACCAGCACGGTTC	3231
AAAV DA-1	AAGTCTTTGCGGACAAGGACTACCAGCTGCCGTACGTCTCTCGGATCGGCTACAGAGGGCACCTTCCCGCCGT	3348
AAAV VR 865	AGGTCTTTGCGGACAAGGACTACCAACTGCCGTACGTCTCTCGGATCGGCTACCGAAGGCACCTTCCCGCCGT	3360
AAV 1	AAGTCTTCTCGGACTCGGAGTACCAGCTTCCGTACGTCTCTCGGCTCTGCGCACCAGGGCTGCCTCCCTCCGT	3318
AAV 2	AGGTGTTTACTGACTCGGAGTACCAGCTCCCGTACGTCTCTCGGCTCGGCGCATCAAGGATGCCTCCCGCCGT	3295
AAV 3	AAGTGTTTACGAGCTCGGAGTATCAGTCTCCCGTACGTGCTCGGGTCCGCGCAACAGGCTGTCTCCCGCCGT	3300
AAV 4	AGATCTTTGCGGACTCGTCTGTACGAAGTCCCGTACGTGATGGATGCGGGTCAAGAGGGCAGCCTGCCTCCTT	3325
AAV 6	AAGTCTTCTCGGACTCGGAGTACCAGTTGCCGTACGTCTCTCGGCTCTGCGCACCAGGGCTGCCTCCCTCCGT	3303
AAAV DA-1	TCCCAGCGGATATCTACACGATCCCCGAGTATGGTTACTGCACGCTAAACTACAACAA----CGAGGCGGT	3415
AAAV VR 865	TCCCAGCGGATATCTACACGATCCCCGAGTACGGGTACTGCACGCTAAACTACAACAA----CGAGGCGGT	3427
AAV 1	TCCCGGCGGACGTGTTTCATGATTCCGCAATACGGCTACCTGACGCTCAACAATGGCAGC----CA-AGCCGT	3385
AAV 2	TCCCAGCAGAGCTCTTCATGGTGCCACAGTATGGATACCTCACCTGAACAACGGGAGT----CA-GGCAGT	3362
AAV 3	TTCCAGCGGACGTCTTCATGGTCCCTCAGTATGGATACCTCACCTGAACAACGGGAAGT----CA-AGCCGT	3367
AAV 4	TTCCCAACGAGCTCTTTATGGTGCCCGAGTACGGCTACTGTGGACTGGTGACCGGCAACACTTTCGACGAAC	3397
AAV 6	TCCCGGCGGACGTGTTTCATGATTCCGCGAGTACGGCTACCTAACGCTCAACAATGGCAGC----CA-GGCAGT	3370
AAAV DA-1	GGAT---CGTTCG-GCCTTCTACTGTCTAGACTATTTCCCTCAGACATGCTGCGGACAGGAAATAACTTTG	3483
AAAV VR 865	GGAT---CGTTCG-GCCTTCTACTGTCTGGACTACTTTCCCTCAGACATGCTGCGGACAGGAAATAACTTTG	3495
AAV 1	GG---GACGTTTCAT-CCTTTTACTGCTGGAAATATTTCCCTTCTCAGATGCTGAGAACGGGCAACAACCTTTA	3453
AAV 2	AG---GACGCTCTT-CATTTTACTGCCTGGAGTACTTTCCCTTCTCAGATGCTGCGTACCGGAAACAACCTTTA	3430
AAV 3	GG---GACGCTCAT-CCTTTTACTGCTGGAGTACTTTCCCTTCGACATGCTAAGGACTGGAAATAACTTCC	3435
AAV 4	AGACTGACAGAAATGCCTTCTACTGCCTGGAGTACTTTCCCTTCGACATGCTGCGGACTGGCAACAACCTTTG	3469
AAV 6	GG---GACGGTCAT-CCTTTTACTGCCTGGAATATTTCCCATCGCAGATGCTGAGAACGGGCAATAACCTTTA	3438
AAAV DA-1	AATTCACTTACACGTTTCGAGGACGTTCCCTTTCCATAGCATGTTTGCTCACAACCAGACGCTAGACCGGCTGA	3555
AAAV VR 865	AGTTTACTTACACCTTCGAGGACGTTTCCTTTCCATAGCATGTTTGCCACAACCAGACGCTAGACCGGCTGA	3567
AAV 1	CCTTCAGCTACACCTTTGAGGAAGTGCCCTTTCCACAGCAGTACGCGCACAGCCAGAGCCTGGACCGGCTGA	3525
AAV 2	CCTTCAGCTACACTTTTGGAGACGTTTCCTTTCCACAGCAGTACGCTCACAGCCAGAGTCTGGACCGTCTCA	3502
AAV 3	AATTCACTTACCTTCGAGGATGTACCTTTTTCACAGCAGTACGCTCACAGCCAGAGTTTGGATCGCTTGA	3507
AAV 4	AAATTACGTACAGTTTGGAGAAGGTGCTTTCCACTCGATGACGCGCACAGCCAGAGCCTGGACCGGCTGA	3541
AAV 6	CCTTCAGCTACACCTTCGAGGACGTTGCCCTTTCCACAGCAGTACGCGCACAGCCAGAGCCTGGACCGGCTGA	3510
AAAV DA-1	TGAATCCTCTCGTCGATCAGTACCTGTGG-GCTTTTCAGT----TCCGTCAGCCAAAC-----AGGCTCGTCT	3617
AAAV VR 865	TGAATCCCCTCGTGGATCAGTACCTCTGG-GCTTTTCAGC----TCCGTCAGCCAAAGC-----AGGCTCATCT	3629
AAV 1	TGAATCCTCTCATCGACCAATACCTGTATTACCTG-AACAGAACTCAA-AATCAGTC--CGGAAGTGCC-CA	3592
AAV 2	TGAATCCTCTCATCGACCAAGTACCTGTATTACTTG-AGCAGAACAAA-CACTCCA--AGTGAACACCACCG	3570
AAV 3	TGAATCCTCTTATTGATCAGTATCTGTACTACCTG-AACAGAACGCAAGGAACAACCTCTGGAACAACCAAC	3578
AAV 4	TGAACCTCTCATCGACCAAGTACCTGTGGGAGTGAATCGACACCACCGG-AACCACCTGAATGCGCG--	3610
AAV 6	TGAATCCTCTCATCGACCAAGTACCTGTATTACCTG-AACAGAACTCA-GAATCAG--TCCGGAAGTGCC-CA	3577
AAAV DA-1	GGACGGGCACTCAATTATTAC-CGCGGACCAA-AACCAATATGGCAACCCAGTACAGAACTGGTTACCTG	3687
AAAV VR 865	GGACGAGCTCTTCACTACTCGC-GGGCGACTAA-AACCAACATGGCGGCTCAATATAGGAAGTGGTTACCTG	3699
AAV 1	AAACAGGACTTGCTGTTTAGCCGTGGGTCT-CCAGCTGGCATGTCTGTTTCAGCCCAAAACTGGCTACCTG	3663
AAV 2	CAGTCAAGGCTT-CAGTTTTCTCAGGCCGGA-GCGAGTGACATTCGGGACCAGTCTAGGAAGTGGCTTCCTG	3640
AAV 3	CAATCACGGCT-GCTTTTTAGCCAGGCTGGG-CCTCAGTCTATGTCTTTGAGGCCAGAAATTGGCTACCTG	3648
AAV 4	GGACTGCCACCACCAACTTTTACCAAGCTGCGGCCTACCAACTTTTCCAACTTTAAAAAGAACTGGCTGCCCG	3682
AAV 6	AAACAGGACTTGCTGTTTAGCCGGGGGTCT-CCAGCTGGCATGTCTGTTTCAGCCCAAAACTGGCTACCTG	3648

AAAV DA-1	GACCCCTTCGTCGGGATCAGCAAATCTTTACGGGGGCTAGCAACATCACCCAAAACAACGTGTTCAACGTTT	3759
AAAV VR 865	GGCCTTTCTTCCGTGATCAGCAAATCTTTACGGGGGCTAGCAACATCACTAAAAATAACGTCTTTAGCGTTT	3771
AAV 1	GACCCGTGTTATCGGCAGCAGCGCGTTTCTAAAACAAAAACAGACAAACAACAGCAATTTTACCTGGACTG	3735
AAV 2	GACCCGTGTTACCGCCAGCAGCGAGTATCAAAGACATCTGCGGATAACAACAACAGTGAATACTCGTGGACTG	3712
AAV 3	GGCCCTGCTACCGGCAACAGAGACTTTCAAAGACTGCTAACGACAAACAACAGTAACCTTTCCTTGGACAG	3720
AAV 4	GGCCTTCAATCAAGCAGCAGGCTTCTCAAAGACTGCCAATCAAACTACAAGATCCCTGCCACCGGGTCAG	3754
AAV 6	GACCCGTGTACCGGCAGCAGCGCGTTTCTAAAACAAAAACAGACAAACAACAACAGCAACTTTACCTGGACTG	3720
AAAV DA-1	GGGATAAAGGCAAGCAGTGGGTGATAGACAAT-CGGATCAAT-ATGATGCAGCCCGGCCCTGCAGCAGCGAC	3829
AAAV VR 865	GGGAAAAAGGCAAGCAATGGGAACCTCGACAAT-CGGACCAAC-CTAATGCAGCCCGGTCTGCAGCAGCGAC	3841
AAV 1	GTG---CTTCAAAATATAACCTCAATGGG--CGTGAATCCATCATCAA---CCCTGGCACTGCTATGGCCTC	3799
AAV 2	GAG---CTACCAAGTACCACCTCAATGGC--AGAGACTCTCTGGTGAA---TCCGGGCCCCGCCATGGCAAG	3776
AAV 3	CGG---CCAGCAAATATCATCTCAATGGC--CGCGACTCGCTGGTGAA---TCCAGGACCAGCTATGGCCAG	3784
AAV 4	ACAGTCTCATCAATACGAGACGCACAGCACTCTGGACGGGAATGGAGTGCCCTGACCCCCGGAGCTCCAA	3826
AAV 6	GTG---CTTCAAAATATAACCTTAATGGG--CGTGAATCTATAATCAA---CCCTGGCACTGCTATGGCCTC	3784
AAAV DA-1	CACCTTTAGCGGAGAACCCGA-CCGTCAAGCCATGCAAAACACGCTGGCCTTTAGTCGGACGGTCTACGACC	3900
AAAV VR 865	CACCTTTAGCGGAGAACCTGA-CCGTCAAGCCATGCAAAACACGCTGGCCTTTAGCAGGACCGTCTACGATC	3912
AAV 1	ACACAAAGACGACGAAGACAAGTT-CTTTCCCATGAGCGGTGTCTATGATTTTG-GAAAAGAGAGCGCCGGA	3869
AAV 2	CCACAAGGACGATGAAGAAAAGTT-TTTTCTCAGAGCGGGGTTCTCATCTTTG-GGAAGCAAGGCTCAGAG	3846
AAV 3	TCACAAGGACGATGAAGAAAATT-TTTCCTATGCACGGGAATCTAATATTTG-GCAAAGAAGGACCAACG	3854
AAV 4	TGGCCACGGCTGGACCTGCGGACAGCAAGTTCAGCAACAGCCAGCTCATCTTTCGCGGGCCTAAACAGAACG	3898
AAV 6	ACACAAAGACGACAAAGACAAGTT-CTTTCCCATGAGCGGTGTCTATGATTTTG-GAAAGGAGAGCGCCGGA	3854
AAAV DA-1	AGACAACCACTACGACCGATCGTAACCAGTTGCTCATTACCAACGAAGATGAAATCAGACCCACCAACTCGG	3972
AAAV VR 865	AAACGACCGCCACGACCGATCGTAACCAGATACTCATCACCACGAAGACGAAATCAGACCCACCAACTCGG	3984
AAV 1	GCTTCA-AACACTG-C--ATTG-GACAATGTCATGATTACAGACGAAGAGGAAATTAAAGCCACTAACCCTG	3936
AAV 2	AAAACA-AATGTGGAC--ATTG-AA-AAGGTCATGATTACAGACGAAGAGGAAATCAGGACAACCAATCCCG	3913
AAV 3	GCAAGT-AACGCAGA--ATTA-GATAATGTAATGATTACGGATGAAGAAGAGATTCGTACCACCAATCCTG	3921
AAV 4	GCAACACGGCCACCGT--ACCC-GGGACTCTGATCTTCACCTCTGAGGAGGAGCTGGCAGCCACCAACGCCA	3967
AAV 6	GCTTCA-AACACTG-C--ATTG-GACAATGTCATGATCAGACGAAGAGGAAATCAAAGCCACTAACCCCG	3921
AAAV DA-1	TCGGCATCGACACGTGGGGAGTAGTTCCCAACAACAACAGTCCAAGGTGACCGCCGGCACTCGCGCGGCCA	4044
AAAV VR 865	TCGGTATCGACGCGTGGGGAGCAGTTCCCAACAACAACAGTTCGATCGTGACCCCCGGCACTCGCGCGGCCG	4056
AAV 1	TGGCCACCGAAAGATTGGGACCGTGGCAGTCAATTCCAGACGACGACACAGACCTGCGACCCGAGATG	4008
AAV 2	TGGTCTACGGAGCAGTATGGTTCTGTATCTACCAACCTCCAGAGAGGCAACAGACAAGCAGTACCCGAGATG	3985
AAV 3	TGGCAACAGAGCAGTATGGAATGTGGCAAATAACTTGCAGAGCTCAAATACAGCTCCCACGACTAGAATG	3993
AAV 4	CCGATACGGACATGTGGGGCAACCTACCTGGCGGTGACCAGAGCAACAGCAACCTGCCGACCGTGGACAGAC	4039
AAV 6	TGGCCACCGAAAGATTGGGACTGTGGCAGTCAATCTCCAGAGCAGCAGACAGACCCCTGCGACCCGGAGATG	3993
AAAV DA-1	TCAACAACCAAGGGGCGCTTCCCGGGATGGTGTGGCAAAACAGAGACATTTACCTC-CAAGGACCCATTTGG	4115
AAAV VR 865	TCAACATCAAGGGGCGCTTCCCGGGATGGTGTGGCAAAACAGAGACATTTACCTTACAGGACCCATTTGG	4128
AAV 1	TGCATGCTATGGGAGCATTACCTGGCATGGTGTGGCAAGATAGAGACGTGTACCTG-CAGGGTCCCATTGG	4079
AAV 2	TCAACACACAAGGCGTCTTCCAGGATGGTGTGGCAGGACAGAGATGTGTACCTT-CAGGGGCCATCTGG	4056
AAV 3	TCAATGATCAGGGGGCCTTACCTGGCATGGTGTGGCAAGATCGTGACGTGTACCTT-CAAGGACCTATCTGG	4064
AAV 4	TGACAGCCTTGGGAGCGGTGCTGGAATGGTGTGGCAAAACAGAGACATTTACTAC-CAGGGTCCATTGG	4110
AAV 6	TGCATGTTATGGGAGCCTTACCTGGAATGGTGTGGCAAGACAGAGACGTATACCTG-CAGGGTCCATTGG	4064
AAAV DA-1	GCCAAAATCCCGACACAGACAATCACTTCCATCCGTCCCCGCTTATTGGCGGGTTTGGCTGCAAGCATCCC	4187
AAAV VR 865	-CCAAAATCCCGACACTGACAATCACTTCCATCCGTCCCCGCTTATTGGGCGGTTTGGCTGCAAGCATCCC	4199
AAV 1	GCCAAAATTCCTCACACAGATGGACACTTTACCCGCTCTCCTCTTATGGGCGGTTTGGACTCAAGAACCCG	4151
AAV 2	GCAAAGATTCCACACACGGACGGACATTTTACCCCTCTCCCTCATGGGTGGATTTCGGACTTAAACACCTT	4128
AAV 3	GCAAAGATTCTCACACGGATGGACACTTTATCCTTCTCCTCTGATGGGAGGCTTGGACTGAAACATCCG	4136
AAV 4	GCCAAAGATTCTATACCGATGGACACTTTACCCCTCACCCGCTGATTGGTGGGTTTGGGCTGAAACACCCG	4182
AAV 6	GCCAAAATTCCTCACACGGATGGACACTTTACCCGCTCTCCTCTCATGGGCGGTTTGGACTTAAACACCCG	4136
AAAV DA-1	CCTCCCCAGATTTTCATTAATAAACACACCCGTCCTCGGCCAACCCCTTCGGAACGTTCCAGACGGCCAAAGTG	4259
AAAV VR 865	CCTCCCCAGATTTTCATTAATAAACACACCCGTCCTCGGCCAACCCCTTCGGAACGTTCCAGACGGCCAAAGTG	4271
AAV 1	CCTCCTCAGATCCTCATCAAAAACACGCCTGTCTCTGCGAATCCTCCGGCGGAGTTTTCAGCTACAAGTTT	4223
AAV 2	CCTCCACAGATTCTCATCAAGAACACCCCGTACCTGCGAATCCTTCGACCACCTTCAGTGGCGGCAAGTTT	4200
AAV 3	CCTCCTCAAAATCATGATCAAAAATACCTCCGCTACCGGCAAAATCCTCCGACGACTTTCAGCCCGGCCAAGTTT	4208
AAV 4	CCTCCTCAAAATTTTATCAAGAACACCCCGTACCTGCGAATCCTGCAACGACCTTCAGCTCTACTCCGGTA	4254
AAV 6	CCTCCTCAGATCCTCATCAAAAACACGCCTGTCTCTGCGAATCCTCCGGCAGAGTTTTCGGCTACAAGTTT	4208

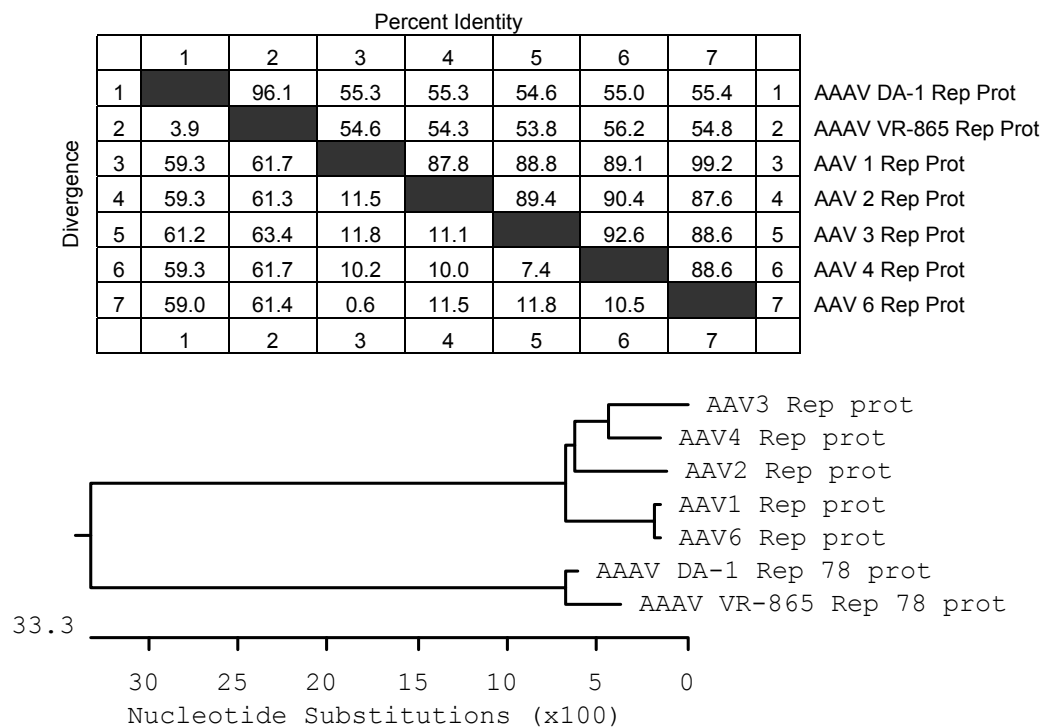
AAAV DA-1	GCCTCCTTCATCAACCAGTACTCGACCGGACAGTGCACCGTCGAAATCTTTTGGGAACTCAAGAAGGAAACC	4331
AAAV VR 865	GCCTCCTTCATCAACCAGTACTCGACCGGACAGTGCACCGTCGAAATCTTTTGGGAACTCAAGAAGGAAACC	4343
AAV 1	GCTTCATTCATCACCAATACTCCACAGGACAAGTGAGTGTGGAAATTGAATGGGAGCTGCAGAAAGAAAAC	4295
AAV 2	GCTTCCTTCATCACACAGTACTCCACGGGACAGGTGAGCGTGGAGATCGAGTGGGAGCTGCAGAAAGAAAAC	4272
AAV 3	GCTTCATTTATCACTCAGTACTCCACTGGACAGGTGAGCGTGGAAATTGAGTGGGAGCTACAGAAAGAAAAC	4280
AAV 4	AACTCCTTCATTACTCAGTACAGCACTGGCCAGGTGTCGGTGACAGATTGACTGGGAGATCCAGAAGGAGCGG	4326
AAV 6	GCTTCATTCATCACCCAGTATTCCACAGGACAAGTGAGCGTGGAGATTGAATGGGAGCTGCAGAAAGAAAAC	4280
AAAV DA-1	TCCAAGCGCTGGAACCCCGAAATCCAGTTACCTCCAACTTTGGCAACGCGGCCGACATCCAGTTTGCTGTG	4403
AAAV VR 865	TCCAAGCGCTGGAACCCCGAAATCCAGTTACCTCCAACTTTGGCAACGCGGCCGACATCCAGTTTGCCGTG	4415
AAV 1	AGCAAGCGCTGGAATCCCGAAGTGCAGTACACATCCAATTATGCAAAATCTGCCAACGTTGATTTTACTGTG	4367
AAV 2	AGCAAACGCTGGAATCCCGAATTCAGTACACTTCCAACACAAAGTCTGTTAATGTGGACTTTACTGTG	4344
AAV 3	AGCAAACGTTGGAATCCAGAGATTGAGTACACTTCCAACACAAAGTCTGTTAATGTGGACTTTACTGTA	4352
AAV 4	TCCAACGCTGGAACCCCGAGGTCCAGTTTACCTCCAACACGACAGCAAACTCTCTGTGTGGGCTCCC	4398
AAV 6	AGCAAACGCTGGAATCCCGAAGTGCAGTATACATCTAACTATGCAAAATCTGCCAACGTTGATTTTACTGTG	4352
Stop all VP		
AAAV DA-1	TCCGACACCGGATCCTATTCCGAACCTCGTCCCATCGGTACCCGGTACCTTACCAAACCTCTGTAAAT-TAA	4474
AAAV VR 865	TCCGACACGGGATCCTATTCCGAACCTCGTCCCATCGGTACCCGGTACCTTACCAAACCTCTGTAAAT-TAA	4486
AAV 1	GACAACAATGGACTTTATACTGAGCCTCGCCCCATTGGCACCCGTTACCTTACCCGTCCCCTGTAATTACGT	4439
AAV 2	GACACTAATGGCGTGTATTGAGAGCCTCGCCCCATTGGCACAGATACCTGACTCGTAATCTGTAATTGCTT	4416
AAV 3	GACACTAATGGTGTATTAGTGAACCTCGCCCTATGGAAACCCGTTATCTCACACGAAACTGTATATC-CTG	4423
AAV 4	GATGCGGCTGGGAAATACACTGAGCCTAGGGCTATCGGTACCCGCTACCTACCCACCACCTGTAATAACCT	4470
AAV 6	GACAACAATGGACTTTATACTGAGCCTCGCCCCATTGGCACCCGTTACCTTACCCGTCCCCTGTAATTGTGT	4424
Poly-A signal		
AAAV DA-1	ACCCCTTC <u>AAATAAA</u> CCGTTT-ATGCGT---AACTGTACTTCCGTCTCCTGTGCG-----TTATTCAGTCA	4533
AAAV VR 865	ACCCCTTC <u>AAATAAA</u> CCGTTT-ATGCGT---AACTGTATTTCCGTCTCCTGTGCG-----TTATTCAGTCA	4545
AAV 1	GTTAATCAATAAACCGGTTGATTTCGTTTCAGTTGAACTTTGGTCTCCTGTGCT-CTTCTTATCTTATC-GGTTA	4509
AAV 2	GTTAATCAATAAACCGTTTAATTTCGTTTCAGTTGAACTTTGGTCTCCTGTGCTA-TTCTTTCTTATCTAGTTT	4487
AAV 3	GTTAATCAATAAACCGGTTTAATTTCGTTTCAGTTGAACTTTGGTCTTGTGCACTTCTTATCTTATCTTGT	4495
AAV 4	GTTAATCAATAAACCGGTTTATTCGTTTCAGTTGAACTTTGGTCTCCGTGTCCTTCTTATCTTATCTCGTTT	4542
AAV 6	GTTAATCAATAAACCGGTTAATTTCGTTTCAGTTGAACTTTGGTCTCATGTGCTG-TTATTATCTTATCTGGTCA	4495
AAAV DA-1	C-ATGAT-----ACGGCA-----	4545
AAAV VR 865	C-ATGAT-----GCGGCA-----	4557
AAV 1	CCATGGTTATAGC---TTACACATTAAC---GCTTGG---TTGCGCTTCGCGATAA-----AAGAC	4562
AAV 2	CCATGGCTAC---GTAGATAAGTAGCA--TGGCGGG--TTAA---TCATTAACACAA-----GGA-	4538
AAV 3	CCATGGCTACTGCGTAGATAAGCAGCGGCCCTGCGGCGCTTTCGCGCTTCGCGGTTTACAACCTGCTGGTTAATAT	4567
AAV 4	CCATGGCTACTGCGTACATAAGCAGCGGCCCTGCGGCGCTTTCGCGCTTCGCGGTTTACAACCTGCCGGTTAATCA	4614
AAV 6	CCATAGCAAC---CGGTTACACA--T--TAACTGC--TTAG---TTGCGCTTCGCGA-----ATA-	4543
AAAV DA-1	TTACCTCACCGG-----AGTG-----CTGGCCAGTTTC-CAAAACAGGCTCGCTCGCTCACTCGGGC	4601
AAAV VR 865	TTACCTCACCGG-----AGTG-----CTGGCCAGTTTC-CAAGACAGGCTCGCTCGCTCACTCGGGC	4613
AAV 1	TTACGTCATCGGGTTACCCCTAGTGATGGAGTTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGGC	4634
AAV 2	-----ACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGC	4595
AAV 3	TTAACCTC-TGCCATACCTCTAGTGATGGAGTTGGCCACTCCCTCTATGCGCACTCGCTCGCTCGGTGGGGC	4638
AAV 4	GTAACCT-CTGGCAAACC--AGATGATGGAGTTGGCCACATTAGCTATGCGCGCTCGCTCACTCACTCGGCC	4683
AAV 6	-----CCCTAGTGATGGAGTTGCCCACTCCCTCTATGCGCGCTCGCTCGCTCGGTGGGGC	4599
AAAV DA-1	CGCGACGCCAAAGGCGTCGCTACCGGCCCCCTTGGGGCCGG--GGCCCCAGTGAGCGAGCGAGCCTGTTTTG	4671
AAAV VR 865	CGGGGGCCCCAAAGGGGCCCTAGCGACCGCTTTCGCGGTGCG--GGCCCCAGTGAGCGAGCGAGCCTGTCTTG	4683
AAV 1	CTGCGGACCAAAGGT-CCGCAGACGGCAGAGCTCTGCTCTGCCGGCCCCACCGAGCGAGCGAGCGCGCAGAG	4705
AAV 2	CGGGCGACCAAAGGT-CGCCGACGCGCCGGGCTTTGCCCGGGCGGCCCTCAGTGAGCGAGCGAGCGCGCAGAG	4666
AAV 3	CGGACGTGCAAAGCA-CGTCCGTCTGGCGACCTTTGGTCCGACGGCCCCACCGAGCGAGCGAGTGGCGCATAG	4709
AAV 4	CTGGAGACCAAAGGT-CTCCAGACTGCGGCGCTTGGCCAGGCGGAGTGAGTGAGCGAGCGCGCATAG	4754
AAV 6	CGGCAGAGCAGAGCT-CTGCCGTCTGCGGACCTTTGGTCCGACGGCCCCACCGAGCGAGCGAGCGCGCATAG	4670
AAAV DA-1	-GAAACTGGCCA	4682
AAAV VR 865	-GAAACTGGCCA	4694
AAV 1	AGGGAGTGGGCA	4717
AAV 2	AGGGAGTGGCCA	4678
AAV 3	AGGGAGTGGCCA	4721
AAV 4	AGGGAGTGGCCA	4766
AAV 6	AGGGAGTGGGCA	4682

Fig. 4: Nucleotide sequence alignment of AAV strains VR-856 and DA-1 and AAV strains 1, 2, 3, 4 and 6. The nucleotide sequence of the inverted terminal repeats are indicated in italics. Putative promoter, initiation, termination and splice sequences are indicated in bold.

A



B



C

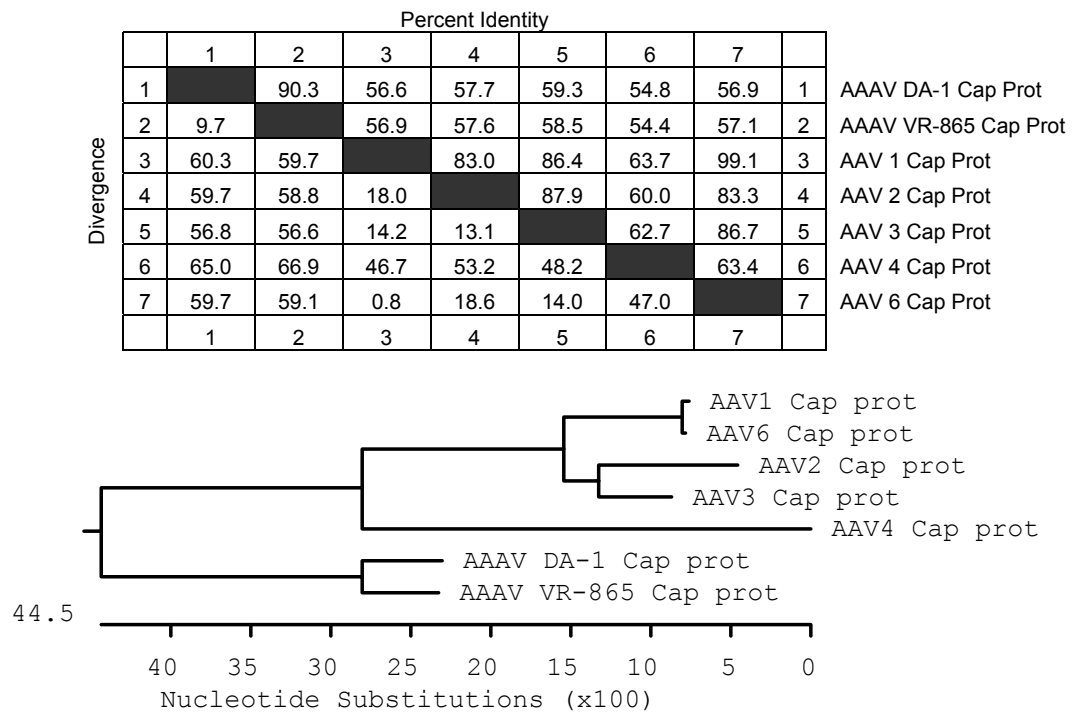


Fig. 5: Phylogenetic trees and sequence distances of the AAV strains in comparison to the AAV strains. Panel A: phylogenetic tree and nucleotide sequence distances. Panel B: phylogenetic tree and distances of the amino acid sequences of the Rep proteins. Panel C: phylogenetic tree and distances of the amino acid sequences of the Cap proteins.

E6			
AAAV DA-1 Rep prot	MRSYY E VIVQLPNDVESQVPGISDSFVNWITSREWTLPEDADWDLQVDQVQLTLGDKIQREIRN	65	
AAAV VR-865 Rep prot	MRSYY E VIVQLPNDVESQVPGISDSFVNWITSREWTLPEDADWDLQVDQVQLTLGDKIQREIRT	65	
AAV1 Rep prot	MPGFY E IVIKVPSDLDEHLPGISDSFVSWVAEKEWELPPDSMDLNLIEQAPLTVAEKLQRDFLV	65	
AAV2 Rep prot	MPGFY E IVIKVPSDLDEHLPGISDSFVNWVAEKEWELPPDSMDLNLIEQAPLTVAEKLQRDFLT	65	
AAV3 Rep prot	MPGFY E IVLKVPSDLDEHLPGISNSFVNWVAEKEWELPPDSMDPNLIEQAPLTVAEKLQREFLV	65	
AAV4 Rep prot	MPGFY E IVLKVPSDLDEHLPGISDSFVSWVAEKEWELPPDSMDLNLIEQAPLTVAEKLQREFLV	65	
AAV6 Rep prot	MPGFY E IVIKVPSDLDEHLPGISDSFVNWVAEKEWELPPDSMDLNLIEQAPLTVAEKLQRDFLV	65	
E83 * *			
AAAV DA-1 Rep prot	HWGTMAKEPDFHYFIQ L EQGEVFF HL HVLLETCSVKPMVLGRYIRHIQQKIVSKVYCGHEPAMEG	130	
AAAV VR-865 Rep prot	HWGTMAKEPDFHYFIQ L EQGEVFF HL HVLLETCSVKPMVLGRYIRHIQQKIVSKVYCATSLRWKD	130	
AAV1 Rep prot	QWRRVSKAPEALFFVQ F EKGESY F HLHILVETTVGKSMVLGRFLSQIRDKLVTIYRGIEPTLPN	130	
AAV2 Rep prot	EWRRVSKAPEALFFVQ F EKGESY F HMHLVETTVGKSMVLGRFLSQIREKLIQRIYRGIEPTLPN	130	
AAV3 Rep prot	EWRRVSKAPEALFFVQ F EKGESY F HLHVLLETIGVKSMMVGRYV S QIKEKLVTIRYRGVEPQLPN	130	
AAV4 Rep prot	EWRRVSKAPEALFFVQ F EKGDSY F HLHILVETTVGKSMVGRYV S QIKEKLVTIRYRGVEPQLPN	130	
AAV6 Rep prot	QWRRVSKAPEALFFVQ F EKGESY F HLHILVETTVGKSMVLGRFLSQIRDKLVTIYRGIEPTLPN	130	
Y156 E164			
AAAV DA-1 Rep prot	WMRVTKTKN-FGGANKVRAESYIPAY L LIPKQ Q PEVQAWTNVPEYIKACLHRELRLASLARLHFEE	194	
AAAV VR-865 Rep prot	GCVVTKTKN-FGGANKVRAESYIPAY L LIPKQ Q PEVQAWTNVPEYIKACLHRELRLASLARLHFEE	194	
AAV1 Rep prot	WFAVTKTRNGAGGKNKVDDCYIPN L LLPKTQ P ELQAWTNMEEYISACLNLAERKRLVAQHLLTH	195	
AAV2 Rep prot	WFAVTKTRNGAGGKNKVDDCYIPN L LLPKTQ P ELQAWTNMEEYISACLNLTERRKRLVAQHLLTH	195	
AAV3 Rep prot	WFAVTKTRNGAGGKNKVDDCYIPN L LLPKTQ P ELQAWTNMDQYLSACLNLAERKRLVAQHLLTH	195	
AAV4 Rep prot	WFAVTKTRNGAGGKNKVDDCYIPN L LLPKTQ P ELQAWTNMDQYISACLNLAERKRLVAQHLLTH	195	
AAV6 Rep prot	WFAVTKTRNGAGGKNKVDDCYIPN L LLPKTQ P ELQAWTNMEEYISACLNLAERKRLVAHDLTH	195	
AGLS--QSKENLARTADGAPVIATRVSKRYMELVDWLVEKGITTEKEWLLENRESFRSFQASSNS			
AAAV DA-1 Rep prot	AGLS--QSKENLARTADGAPVMPTRVSKRYMELVDWLVEKGITTEKEWLLENRESFRSFQASSNS	257	
AAAV VR-865 Rep prot	AGVS--QSKENLARTADGAPVMPTRVSKRYMELVDWLVEKGITTEKEWLLENRESFRSFQASSNS	257	
AAV1 Rep prot	VSQTQEQNKENLNPN S D-APVIRSKTSARYMELVGWLVD R GITSEKQW I QEDQASYISFNAASNS	259	
AAV2 Rep prot	VSQTQEQNKENQNPN S D-APVIRSKTSARYMELVGWLVD K ITSEKQW I QEDQASYISFNAASNS	259	
AAV3 Rep prot	VSQTQEQNKENQNPN S D-APVIRSKTSARYMELVGWLVD R GITSEKQW I QEDQASYISFNAASNS	259	
AAV4 Rep prot	VSQTQEQNKENQNPN S D-APVIRSKTSARYMELVGWLVD R GITSEKQW I QEDQASYISFNAASNS	259	
AAV6 Rep prot	VSQTQEQNKENLNPN S D-APVIRSKTSARYMELVGWLVD R GITSEKQW I QEDQASYISFNAASNS	259	
ARQIKTALQGAIQEMLLTKTAEDYLVGKEFPVSDDEIRQNRIYKILELNHYDPAYVGSILVGWCQK			
AAAV DA-1 Rep prot	ARQIKTALQGAIQEMLLTKTAEDYLVGKDPVSDDDIRQNRIYKILELNHYDPAYVGSILVGWCQK	322	
AAAV VR-865 Rep prot	ARQIKTALQGAIQEMLLTKTAEDYLVGKDPVSDDDIRQNRIYKILELNHYDPAYVGSILVGWCQK	322	
AAV1 Rep prot	RSQIKAALDNAGKIMALT S APDYLVGPAPPAD--IKTNRIYRILELNGYEPAYAGSVFLGWAQK	322	
AAV2 Rep prot	RSQIKAALDNAGKIMSLTKTAPDYLVGQ P VED--ISSNRIYKILELNGYDPQY A ASVFLGWA T K	322	
AAV3 Rep prot	RSQIKAALDNASKIMSLTKTAPDYLVGSNP P ED--ITKNRIYQILELNGYDPQY A ASVFLGWAQK	322	
AAV4 Rep prot	RSQIKAALDNASKIMSLTKTAPDYLVGQNP P ED--ISSNRIYRILEMNGYDPQY A ASVFLGWAQK	322	
AAV6 Rep prot	RSQIKAALDNAGKIMALT S APDYLVGPAPPAD--IKTNRIYRILELNGYDPAYAGSVFLGWAQK	322	
A B			
AAAV DA-1 Rep prot	KWGKRNT L W L FGPATT G KT N IAEAIAHAVPFYGCNVNWTNENFPFNDCVE K MI I W W EEGKMTAKVV	387	
AAAV VR-865 Rep prot	KWGKRNT L W L FGHATT G KT N IAEAIAHAVPFYGCNVNWTNENFPFNDCVE K MI I W W EEGKMTAKVV	387	
AAV1 Rep prot	RFGKRNT I W L FGPATT G KT N IAEAIAHAVPFYGCNVNWTNENFPFNDCVD K MI I W W EEGKMTAKVV	387	
AAV2 Rep prot	KFGKRNT I W L FGPATT G KT N IAEAIAHTVPFYGCNVNWTNENFPFNDCVD K MI I W W EEGKMTAKVV	387	
AAV3 Rep prot	KFGKRNT I W L FGPATT G KT N IAEAIAHAVPFYGCNVNWTNENFPFNDCVD K MI I W W EEGKMTAKVV	387	
AAV4 Rep prot	KFGKRNT I W L FGPATT G KT N IAEAIAHAVPFYGCNVNWTNENFPFNDCVD K MI I W W EEGKMTAKVV	387	
AAV6 Rep prot	RFGKRNT I W L FGPATT G KT N IAEAIAHAVPFYGCNVNWTNENFPFNDCVD K MI I W W EEGKMTAKVV	387	
B' C			
AAAV DA-1 Rep prot	ETA K AILGGSRVRVD Q KCKASVPIEPT P VI I TS N TNMCYVIDGN T TT F EHKQPLEDRMF K LELLT	452	
AAAV VR-865 Rep prot	ETA K AILGGSRVRVD Q KCKASVPIEPT P VI I TS N TNMCYVIDGN T TT F EHKQPLEDRMF K LELLT	452	
AAV1 Rep prot	ESA K AILGGSKVRVD Q KCKSSAQIDPT P VIV I TS N TNMCVIDGN S TT F EHQ Q PLQDRMF K FELTR	452	
AAV2 Rep prot	ESA K AILGGSKVRVD Q KCKSSAQIDPT P VIV I TS N TNMCVIDGN S TT F EHQ Q PLQDRMF K FELTR	452	
AAV3 Rep prot	ESA K AILGGSKVRVD Q KCKSSAQIEPT P VIV I TS N TNMCVIDGN S TT F EHQ Q PLQDRMF K FELTR	452	
AAV4 Rep prot	ESA K AILGGSKVRVD Q KCKSSAQIDPT P VIV I TS N TNMCVIDGN S TT F EHQ Q PLQDRMF K FELTK	452	
AAV6 Rep prot	ESA K AILGGSKVRVD Q KCKSSAQIDPT P VIV I TS N TNMCVIDGN S TT F EHQ Q PLQDRMF K FELTR	452	
RLPDDFGKVTKQEVQRFFRWSQDHLTPVIPLEFLVRKAESRKRPASSGEGYISPTKRPALAEQQQA			
AAAV DA-1 Rep prot	RLPDDFGKVTKQEVQRFFRWSQDHLTPVIPLEFLVRKAESRKRPA S SGEGYISPTKRPALAEQQQA	517	
AAAV VR-865 Rep prot	RLPDDFGKVTKQEVQRFFRWSQDHLTPVIPLEFLVRKAESRKRPA S SGEGYISPTKRPALAEQQQA	517	
AAV1 Rep prot	RLEHDFGKVTKQEVKEFFRWAQDHVTEVAHEFYVRKGGANKRPA P D-DADKSEPKRACPSVADPS	516	
AAV2 Rep prot	RLDHDFGKVTKQEVKDFFRWADHVTEVAHEFYVRKGGAKKRPA S N-DADISEPKRVRESVAQPS	516	
AAV3 Rep prot	RLDHDFGKVTKQEVKDFFRWADHVTEVAHEFYVRKGGAKKRPA S N-DADVSEPKRQCTSLAQPT	516	
AAV4 Rep prot	RLEHDFGKVTKQEVKDFFRWADHVTEVAHEFYVRKGGAKKRPA P N-DADISEPKRACPSVAQPS	516	
AAV6 Rep prot	RLEHDFGKVTKQEVKEFFRWAQDHVTEVAHEFYVRKGGANKRPA P D-DADKSEPKRACPSVADPS	516	

AAAV DA-1 Rep prot	SESAEPVPT----RYRIKCSKHCGMDKMLFPCQICESMNRNINICAIHKTECKECFPYGDKDT	578
AAAV VR-865 Rep prot	SESADPVPT----RYRIKCSKHCGMDKMLFPCQICESMNRDINICAIHKTTDCKECFPDYGDKDD	578
AAV1 Rep prot	TSDAEGAPVDFADRYQNKCSRHAGMLQMLFPCQTCERMNQNFNICFTHGTRDCSECFPGVSESQP	581
AAV2 Rep prot	TSDAE-ASINYADRYQNKCSRHVGMNLMFLPCRQTCERMNQNSNICFTHGQKDCLECFP-VSESQP	579
AAV3 Rep prot	TSDAE-APADYADRYQNKCSRHVGMNLMFLPCRQTCERMNQISNVCFTHGQRDCGECFPGMSESQP	580
AAV4 Rep prot	TSDAE-APVDYADRYQNKCSRHVGMNLMFLPCRQTCERMNQNVDICFTHGVMDCAECFP-VSESQP	579
AAV6 Rep prot	TSDAEGAPVDFADRYQNKCSRHAGMLQMLFPCQTCERMNQNFNICFTHGTRDCSECFPGVSESQP	581
AAAV DA-1 Rep prot	VPELPPCTEHNVSRCYQCHSGELYRVTSDSDEKPAPESDEGTEPSYAPCTIHHLMGKSRGLVSCA	643
AAAV VR-865 Rep prot	V-ELPPCTEHNVSRCYQCHSGELYRVTSDSDEKPAPESDEGTEPSYAPCTIHHLMGKSHGLVTCA	642
AAV1 Rep prot	-----VVRKRTYRK-----LCAIHHLLGRAP-EIACS	607
AAV2 Rep prot	V-----SVVKK-AYQK-----LCYIHHIMGKVP-D-ACT	605
AAV3 Rep prot	V-----SVVKKITYQK-----LCPIHHILGRAP-EIACS	608
AAV4 Rep prot	V-----SVVRKRTYQK-----LCPIHHIMGRAP-EVACS	607
AAV6 Rep prot	-----VVRKRTYRK-----LCAIHHLLGRAP-EIACS	607
AAAV DA-1 Rep prot	ACRLKNSTLHDDLDDGDLEQ.	664
AAAV VR-865 Rep prot	ACRLKNSTLHDDLDDGDLEQ.	663
AAV1 Rep prot	ACDLVN---VDLDDCVSEQ.	624
AAV2 Rep prot	ACDLVN---VDLDDCIFEQ.	622
AAV3 Rep prot	ACDLAN---VDLDDCVSEQ.	625
AAV4 Rep prot	ACELAN---VDLDDCDMEQ.	624
AAV6 Rep prot	ACDLVN---VDLDDCVSEQ.	624

Fig. 6: Amino acid sequence of the Rep 78 protein. Indicated in bold are the residues necessary for initiation of DNA replication at the inverted terminal repeats of the viral genome. Residues Glu 6, Glu 83 and Glu 164 are essential for the metal ion binding mediated by residues His 90 and 92 (indicated with asterisks and in bold). Metal binding is necessary for nicking at the terminal resolution site, which is mediated by residue Tyr 155. Indicated by the letters A, B, B' and C are the sequences necessary for the helicase activity of the protein.

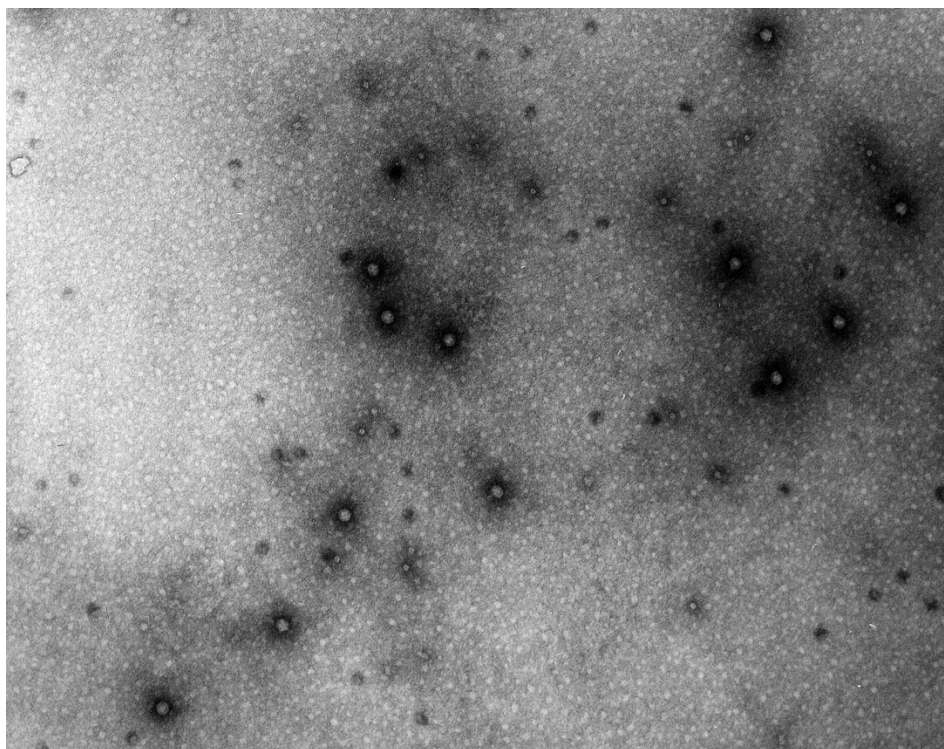


Fig. 7: Rescued AAV from infectious clone by co-transfection of HEK 293 cell cultures with pHelper plasmid and the plasmid containing the whole viral genome (50,000 X magnification).

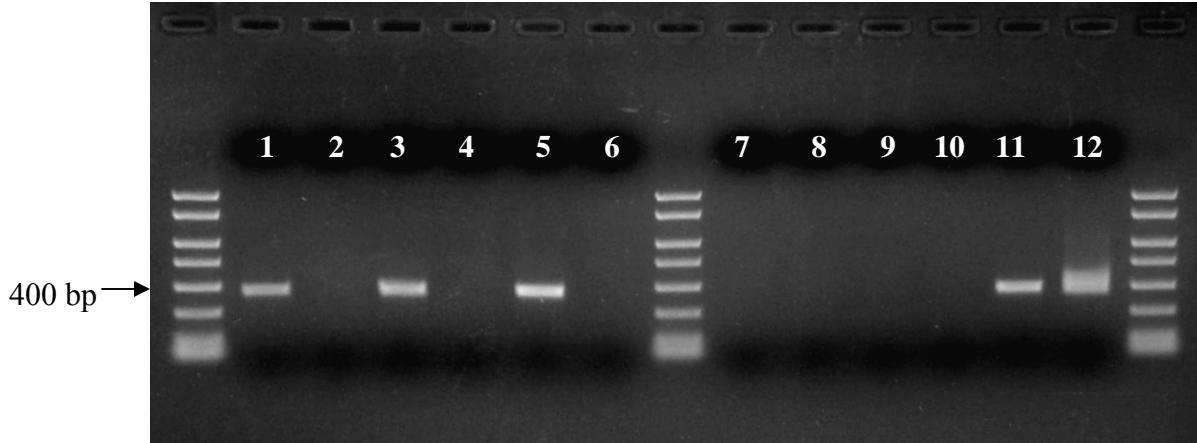
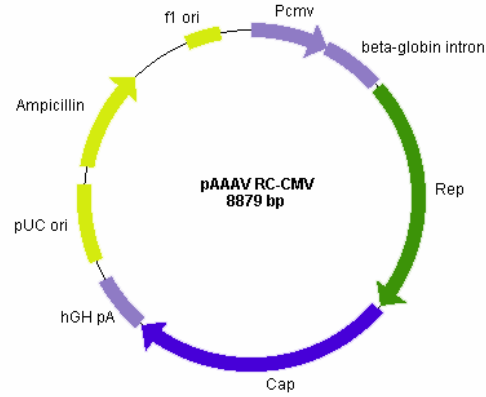


Fig. 8: PCR of viruses rescued in HEK 293 cell cultures from infectious clones of strains AV-865 and DA-1 with primers AAV 2F and AAV 2R (specific for the cap coding region of the viruses) and primers A3V backbone 1F and 1R (specific form the plasmid backbone). Lanes 1, 3 and 5: rescued strains VR-865 and DA-1, and wtAAAV control (respectively) reacted with primers AAV 2F and AAV 2R. Lanes 2, 4 and 6: rescued strains VR-865 and DA-1, and wtAAAV control (respectively) reacted with primers A3V backbone 1F and 1R. Lanes 7 and 8: DNA from adenoviral stock used to provide helper functions for AAV replication reacted with primers AAV 2F-AAV 2R and A3V backbone 1F-1R, respectively. Lanes 9 and 10: DNA extracted from mock transfected HEK 293 cells reacted with primers AAV 2F-AAV 2R and A3V backbone 1F-1R, respectively. Lanes 11 and 12: positive PCR control reactions.

A



B

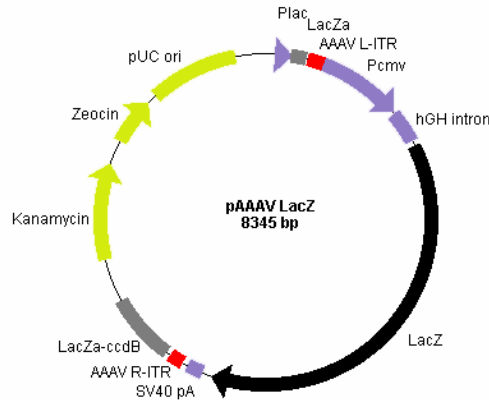


Fig. 9: A) Plasmid pEVA3V-RC-CMV. B) Plasmid pEVA3V-LacZ-ITR. pLac- LacZa: *LacZ* promoter/operon region; AAV L-ITR: avian adeno associated virus left inverted terminal repeat; pCMV: major late promoter of cytomegalovirus; hGH intron: human growth hormone intron; LacZ: *LacZ* coding sequence; SV40 pA: simian virus 40 polyadenylation signal; AAV

R-ITR: avian adeno associated virus right inverted terminal repeat; LacZ-*ccdB*: *ccdB* lethal gene sequence; Kanamycin: kanamycin resistance gene; Zeocin: zeocin resistance gene; pUC: plasmid origin of replication; Rep: avian adeno associated virus non-structural protein genes; Cap: avian adeno associated virus structural (capsid) genes; hGH pA: human growth hormone polyadenylation signal; Ampicillin: ampicillin resistance gene; f1 ori: f1 origin of replication.

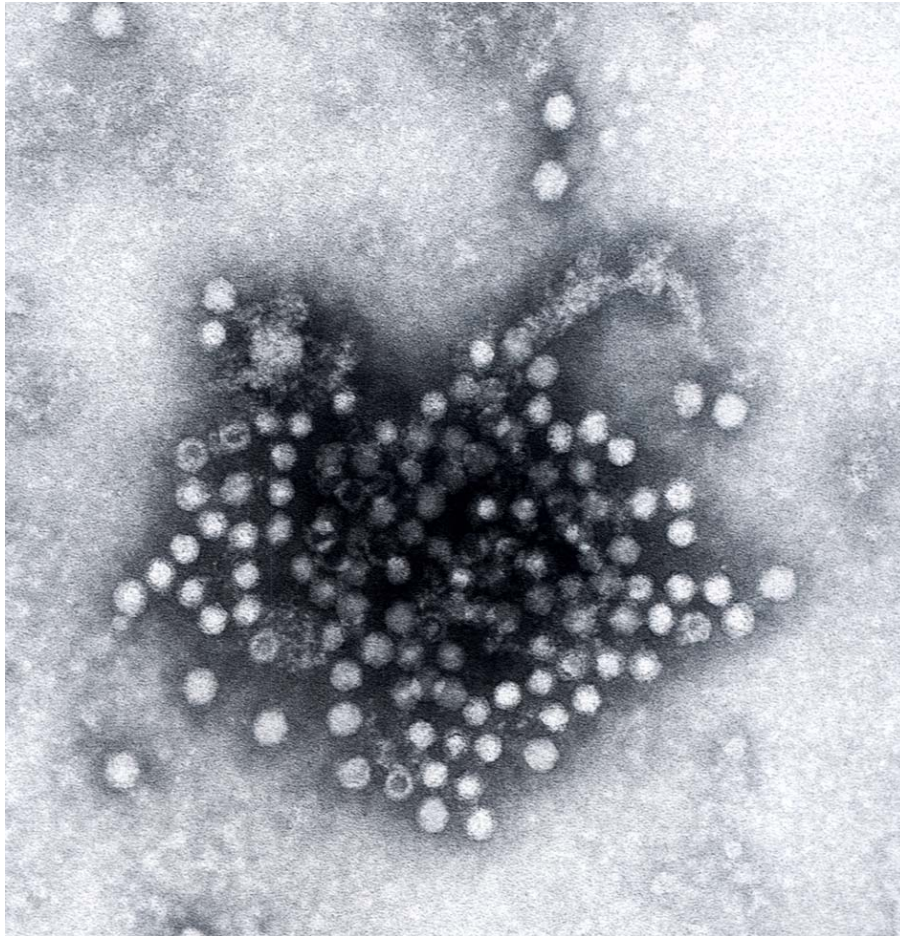


Fig. 10: Electron micrograph of rAAV coding for the *LacZ* reporter gene (100K magnification).

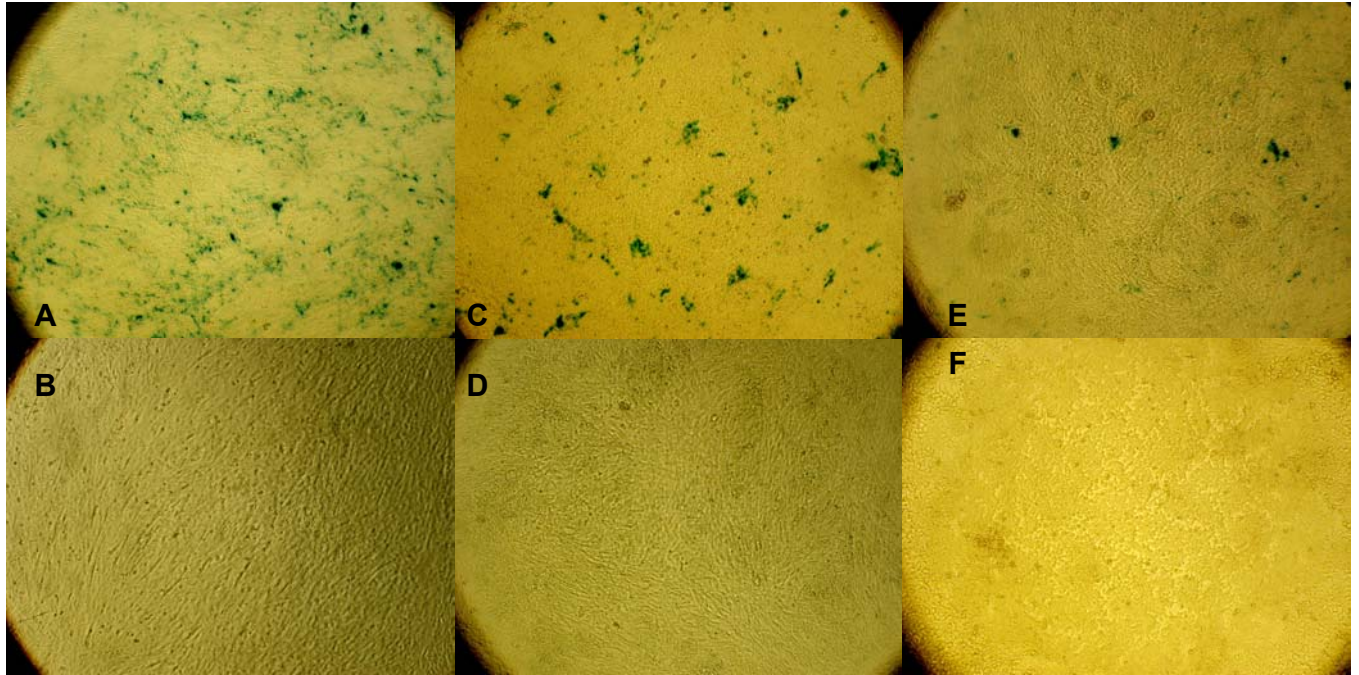


Fig. 11: Expression of β -galactosidase in primary cell culture after infection with recombinant AAV coding for the LacZ gene. Panels A, C and E: transduced chicken embryo fibroblasts, liver and kidney cell cultures, respectively. Panels B, D and F: Mock infected chicken embryo fibroblasts, liver and kidney cell cultures, respectively.

CHAPTER IV

RECOMBINANT AVIAN ADENO-ASSOCIATED VIRUS: TRANSGENE EXPRESSION IN VIVO AND ENHANCEMENT OF EXPRESSION IN VITRO¹

¹Estevez, C. and P. Villegas. Submitted to Avian Diseases, 2004.

SUMMARY:

Recombinant avian adeno-associated virus coding for the *LacZ* gene were used to inoculate embryonating chicken eggs, to assess the usefulness of the system for the expression of a transgene *in vivo*. The results obtained indicate significantly higher levels of expression of the reporter gene at various time intervals in the embryos inoculated with the recombinant virus in comparison with the mock inoculated controls. In a second experiment, different cell line cultures were transfected with plasmids encoding for a reporter gene flanked by the avian adeno-associated virus inverted terminal repeats in the presence or absence of the major non structural proteins of the virus (Rep 78/68) to assess the ability of this proteins to enhance gene expression. Results indicate that the co-expression of the Rep proteins significantly enhance the expression of the transgene in all cell lines tested.

Key words: recombinant virus, avian adeno associated virus, genomic integration, transgene expression.

INTRODUCTION:

Our current understanding of molecular biology and immunology has generated new and safer techniques for disease prevention through immunization. At the forefront of these techniques is the delivery of genes coding for immunogenic peptides via DNA vaccines or recombinant viruses. The advantages conferred by these recombinant systems include the induction of good levels of protection through cell mediated and humoral responses, without the potential risk of vaccine reactions and disease associated with live vaccines, and without the lack of cell mediated immune response of killed vaccines (13). In recent years the adeno associated viruses (AAV), which are replication-defective and non pathogenic members of the *Dependovirus* subfamily of the *Parvoviridae* virus family, have been the focus of great research effort aiming at the generation of recombinant virus-based and non viral-based systems for the purpose of gene delivery and immunization (10, 12, 14). One of the characteristic of this family of viruses is the need for a co-infection with a helper virus (adenovirus or herpesvirus) to generate complete infective viral particles. In the absence of a helper virus, AAVs latently infect cells by site specifically integrating the viral DNA into the host genome. The process of genomic integration is mediated by DNA sequences present at both ends of the viral genome called the inverted terminal repeats (ITRs), and the major non-structural proteins of the virus (Rep 78/68). The ITRs are also important for initiation of viral DNA replication and encapsidation of the genome (5). Previous research works have shown that integration of genes flanked by the AAV ITRs of the virus can occur in the absence of expression of the AAV Rep proteins, albeit in a less efficient and random manner (3,

16). Also, the viral ITRs have been shown to enhance CMV-dependent up-regulation of transgene expression and immunogenicity when used in DNA vaccines (18). Other researchers have observed enhanced levels of transgene expression, and site-specific integration, when a reporter gene flanked by the AAV ITRs is co-expressed with the Rep proteins of the virus (8).

Recombinant systems based on the AAV offer many advantages in comparison with other vector systems, e.g. its ability to transduce both dividing and non-dividing cells, their apparent lack of pathogenicity, induction of high and prolonged levels of expression of the transgene, and the ability of integrating into the host genome (4, 6, 7). Since this family of viruses offers many promising characteristics for the generation of recombinant systems, which could be used for vaccination in poultry, we have cloned the complete genomes of the two known strains, DA-1 and VR 865, of the avian adeno associated virus (AAAV). These clones were used to perform sequence analyses of the virus, and to produce a plasmid-based system for generation of recombinant viruses encoding the *LacZ* reporter gene, flanked by the inverted terminal repeats of the virus.

The AAAV was first reported in chickens by Yates *et al* in the 1970s, and is a replication-defective dependovirus thus far not associated with any disease in poultry (20). Sequence analysis of the AAAV genome performed at our lab showed that the genomic map distribution of the main coding regions (Rep and Cap genes) is conserved and in agreement with previous works describing the primates AAVs. Furthermore, all the sequences known to be necessary for viral DNA replication, and possibly integration into the host genome, are conserved in the ITRs of this viruses (1). Recombinant AAAV particles coding for the *LacZ* gene, obtained from our plasmid-based system, were

successfully used to express the β -galactosidase gene in primary chicken embryo cell cultures (1).

The purpose of the present study is to further validate the ability of this viral based system to produce gene expression *in vivo*, by inoculating embryonating chicken eggs with recombinant AAV encoding for the *LacZ* gene, and assessing the production of β -galactosidase in these embryos at various time intervals. Also, we have investigated the ability of the AAV Rep proteins to enhance gene expression *in vitro* by transfecting different cell lines with plasmids containing the *LacZ* gene flanked by the viral ITRs in the presence and absence of the Rep proteins of the virus.

MATERIALS AND METHODS:

Cells cultures: The human embryonic kidney (HEK) 293, LMH and DF-1 cell lines were obtained from the ATCC (catalog numbers CRL-1573, CRL-2117, CRL 12203, respectively, ATCC, Manassas, VA). HEK 293 and DF-1 cells were grown initially in 75 cm² flasks, and later sub-cultured in 6-well plates, using Dulbecco's modified minimal essential medium (Gibco, Carlsbad, CA) supplemented with 10% fetal bovine serum (Sigma, St. Louis, MO) and antibiotics. These cell cultures were maintained at 37 C and 39 C (respectively), under a 5% CO₂ atmosphere. LMH cell cultures were grown in the same type of flasks and plates but supplemented with Waymouth's MB 752/1 medium (Gibco, Carlsbad, CA), and were maintained at 37 C under a 5% CO₂ atmosphere.

Plasmid constructs: Plasmid pEVA3V-LacZ-ITR contains the *LacZ* gene under the major late promoter of the cytomegalovirus (CMV) and with the simian virus 40

(SV40) polyadenylation signal downstream of the *LacZ* coding region, flanked by the AAV ITR. Plasmid pEVA3V RC-CMV contains the Rep and Cap coding regions of the AAV under the CMV promoter sequence, and with the human growth hormone polyadenylation signal downstream of the described coding regions. These plasmids were obtained from the cloned AAV genomes previously generated in our laboratory, as described elsewhere (1). Plasmid pHelper is commercially available (Stratagene, La Jolla, CA) and contains the E2A, E4 and VA-RNA genes derived from the human adenovirus type 5. Plasmid Sgal (kindly provided by Dr. Mark Jackwood, from the poultry Diagnostic and Research Center, University of Georgia) contains the *LacZ* gene, without the AAV ITR, under the CMV promoter and the SV40 polyadenylation signal downstream of the coding region.

Production of recombinant AAV virus stock expressing the *LacZ* gene:

Recombinant viruses coding for the *LacZ* gene flanked by the AAV ITR (rAAV-*LacZ*) were obtained following the methodology describe by Xiao *et al* (17). Briefly, HEK 293 cell line cultures (grown in 72 cm² flasks) were co-transfected with 10 µg each of the plasmid pEVA3V-*LacZ*-ITR, plasmid pEVA3V-RC-CMV and the helper plasmid (pHelper) already described. Transfection of the cells was performed in 10 flasks of HEK 293 cell cultures with the aid of the Lipofectamine 2000 reagent (Invitrogen, Carlsbad, CA), as per manufacturer's recommendations. The HEK 293 cell line expresses the E1A and E1B immediate early genes from the human adenovirus type 5, which complement the expression of the remaining immediate early genes of the human adenovirus type 5, encoded in plasmid pHelper. This complementation allows the rescue of recombinant viral particles from plasmids pEVA3V-*LacZ*-ITR and pEVA3V-RC-CMV. Purification

of the recombinant viral population was performed 48 hr post transfection, by concentration of a CsCl cushion followed by a purification step by ultracentrifugation in a linear CsCl gradient, as described elsewhere (17).

Titration of recombinant AAV viral stock: The recombinant AAV viral stock was titrated by quantitative real time-PCR, using the LightCycler thermocycler (Roche Diagnostics, Mannheim, Germany), as described by Rohr *et al* (11). Briefly, 200 µl of the purified rAAV particles obtained were treated with 10 µl of a 20 mg/ml stock solution of DNase I (Roche Molecular Biochemicals, Mannheim, Germany), and incubated for 60 min at 37 C in a water bath. This treatment was performed to eliminate the contaminant plasmid DNA that might be carried over during purification from the recombinant viral production in the 293 cell cultures. After DNase treatment, the viral DNA was extracted by the SDS-phenol-chloroform method followed by precipitation with isopropanol. The resulting DNA pellet was resuspended in 50 µl of TE buffer. Two µl of the DNA samples were used for the real time PCR amplification in 20 µl reactions. All PCR reactions were assembled with the use of the LightCycler DNA Master SYBR green I kit reagents, following the manufacturer instructions (Roche Molecular Biochemicals, Mannheim, Germany). The primers used in the reaction were designed to amplify a 201 bp segment of the CMV promoter region contained in the recombinant viral DNA, which was used to drive the transcription of the *LacZ* gene, and which is present only once in the DNA sequence of these recombinant viruses (CMV forward: 5'GGCGGAGTTGTTACGACAT3'; CMV reverse: 5'GGGACTTTCCTACTTGGCA3') (11). Quantification was performed by the use of the second derivative maximum algorithm available in the LightCycler computer

software. For this purpose, control PCR reactions for the generation of the amplification curve, used as standard, were performed with serial dilutions of a control plasmid containing one copy of the CMV promoter sequence. These dilutions were set to contain 2 pg to 20 ng of the control plasmid DNA, in 1 log₁₀ increments, per every 2 µl of sample. Viral titers are reported as genomic copies of DNase resistant particles (DRP) per ml.

Chicken embryo inoculation with rAAAV coding for the *LacZ* transgene: To assess the ability of the rAAAV to infect cells and express the *LacZ* gene *in vivo*, forty 12 day old SPF embryonating chicken eggs were obtained (SPAFAS, Wilmington, MA) and divided in two groups: Twenty four of the eggs were inoculated at 14 days of age with 1 x 10⁹ DRP of the previously obtained rAAAV per egg, while the remaining 16 eggs were inoculated with 0.2 ml of sterile phosphate buffered saline (PBS) solution, serving as controls for the experiment. Six of the rAAAV-inoculated embryos, and 3 of the PBS inoculated embryos, were collected at 48 and 120 hr post inoculation and examined for cell expression of the *LacZ* gene. The remaining embryos were allowed to hatch, and were placed in Horsfall isolation units in groups of rAAAV-inoculated and mock-inoculated chicks. As with the embryonating eggs, six of the recombinant virus-inoculated and 3 of the mock-inoculated chicks were collected at 7 days and 14 days of age to assess the cell expression of the reporter gene.

Assessment of *LacZ* activity in cells *in vivo*: In both, embryos and hatched chicks-derived samples, the β-galactosidase activity was assessed by chemiluminescence, with the aid of the B-gal Reporter Gene Assay (chemiluminescent) kit reagents and recommended procedures (Roche Diagnostics, Mannheim, Germany). This system allows

for quantitative determination of β -galactosidase activity in cell lysates from cell cultures and tissues by reacting this enzyme with a chemiluminescent substrate. The light emissions, product of the enzymatic reaction, were reported as relative light units (RLU), which were determined with a FB 12 model Berthold luminometer (Berthold Detection Systems, Pforzheim, Germany). For this procedure, the chicken embryos' and hatched chicks' livers, kidneys, spleens, lungs and part of the pectoral muscles were collected and pooled for each individual (each embryo and hatched chick was treated as an individual sample, and its organs were pooled to make a single cell suspension sample for the chemiluminescence reaction). The pooled organs samples obtained from each of the embryos or chicks were washed 3 times with sterile PBS, minced with scissors, and submitted to trypsin digestion with a 0.25% trypsin solution in PBS containing 0.02% EDTA to obtain cell suspensions suitable for the chemiluminescent reaction. After trypsin digestion, cells were collected by centrifugation at 300 g at 4 °C for 15 min. The cell pellets were resuspended in PBS, and the cell number in each sample was quantified by the use of a Neubauer cell counting chamber. After counting the cells, the cell suspensions were centrifuged again and resuspended in the lysis buffer solution provided with the chemiluminescence kit. The amount of buffer added was calculated to provide a final concentration of 1×10^7 cells per ml of solution. This lysis step was allowed to continue for 30 minutes, at which time 1 ml of each sample was centrifuged at maximum speed in a table top microcentrifuge to eliminate the cellular debris. After centrifugation, the samples were heated at 50 °C for an hour to quench any endogenous β -galactosidase activity that may be present in the cell lysates. After heat inactivation, 50 μ l of the samples were transferred to a clean microcentrifuge tube, and 100 μ l of a substrate

reagent, provided with the kit, were added. This mixture was allowed to incubate at room temperature for 20 min. After incubation, 50 μ l of an initiation reagent (also provided) was added and the samples were immediately used to read their chemiluminescence in the luminometer. To quantify the amount of the β -galactosidase enzyme in the samples, a standard curve was generated by using serial dilutions of known amounts of a β -galactosidase positive control solution provided with the kit. This positive control was diluted in 1 log₁₀ increments from 20 ng to 20 fg of β -galactosidase per control sample. The RLU readings obtained from these control samples were used for linear regression calculations to generate a standard formula for quantification of the amount of the β -galactosidase enzyme in the unknown samples.

Transfection of cell cultures: HEK 293, LMH and DF-1 cells were grown in 6-well plates to 90% confluence before transfection. Transfection was performed with the aid of the Lipofectamine reagent and recommended procedures, using 6 μ g of each of the plasmid DNAs chosen for the experiment. The experiment was designed to obtain cell cultures transfected with a plasmid containing the *LacZ* gene flanked by the AAV ITRs (pEVA3V-LacZ-ITR) in the presence and absence of the Rep protein of the virus, expressed *in trans* from a second plasmid construct (pEVA3V-RC-CMV). As an additional control, DF-1 cell cultures were transfected with a plasmid containing the *LacZ* gene without the viral ITR (plasmid Sgal), also with or without the co-expression of the Rep proteins from pEVA3V-RC-CMV. A summary of the transfection protocol is described in table 1. From each cell line, 10 monolayers were used for each plasmid transfection treatment. Six of these were used for chemiluminescence reactions for β -galactosidase detection; one of them was used for direct staining for visual detection of β -

galactosidase; and the rest were used to obtain a subsequent passage of the cells transfected for the purpose of determining the stability of transgene expression. A total of five passages were made, in which the same number of monolayers described above were used for chemiluminescence and direct staining detection of the transgene. Six additional cell monolayers of each of the cell lines used were grown in each cell passage. Three of them were used as negative control for the chemiluminescence reaction; one for control of direct staining and two of them for sub-culturing to generate the next set of controls. An additional set of DF-1 cell cultures (20 wells in total) were used to assess the influence of the presence of the AAV ITR in transgene expression. The β -galactosidase expression assessment and sub-culturing of these cell monolayers was performed as described above. A summary of the experimental protocol is presented in table 2.

Chemiluminescence detection of *LacZ* activity in cell cultures: Detection of *LacZ* expression in cell cultures was performed with the aid of the B-gal Reporter Gene Assay (chemiluminescent) kit reagents and recommended procedures (Roche Diagnostics, Mannheim, Germany). After initial transfection, and in each of the passages thereafter, 6 of the monolayers transfected of each cell line, and 3 of the controls, were used for the procedure. Briefly, the culture media was removed from the wells and the monolayers were washed once with cold (4 C) phosphate buffered saline (PBS). One ml of the lysis solution provided in the kit was added to each of the monolayers, and the cells were detached from the plate with the aid of a cell culture policeman. The resulting suspensions were transferred to a clean microfuge tube and the lysis was allowed to continue for 30 minutes; then, samples were centrifuged to eliminate cellular debris. As for the *in vivo* cell suspensions, the cell lysates were incubated for 1 hr at 50 C to

eliminate any cellular endogenous β -galactosidase activity. Subsequent addition of reagents and incubations were performed exactly as described for the cell lysates obtained from the inoculated embryos. The RLU readings were obtained with the Berthold luminometer, also as described.

Direct staining for visual detection of *LacZ* activity in cell cultures: Direct staining for *LacZ* activity in cell cultures was performed with the aid of the Cell Culture Beta-Galactosidase Staining Kit reagents and recommended procedures (Specialty Media, Phillipsburg, NJ). For the procedure, cell culture monolayers were washed with PBS containing added calcium and magnesium (Specialty media, Phillipsburg, NJ), fixed with the addition of 2 ml of fixative reagent (provided with the kit) followed by incubation for 5 min, and washed again three times with PBS. After washing, the cells were reacted with 2 ml of complete beta gal staining reagent (also provided) and incubated at 37 C until visible staining was observed in the transfected cell monolayers, but not in the negative controls. After staining became visible in the transfected cells, the staining solution was removed and the monolayers were washed 3 times with PBS, which was followed by the addition of holding solution (provided in the kit). Visualization of the staining was performed with the aid of an inverted microscope.

Statistical analysis of results: Comparisons of results obtained from the chemiluminescence detection of the transgene activity were performed with the aid of the JMP IN statistical analysis software (SAS Institute Inc. Cary, NC). Determination of differences between groups was performed using one way analysis of variance (ANOVA) calculations, and the significance of the differences of all pairs in the groups was established by the Tukey-Kramer method.

RESULTS:

Production and quantification of rAAV-LacZ stock: Viral particles coding for the *LacZ* gene flanked by the AAV ITR were successfully produced by transfection of the HEK 293 cell cultures as described. Quantification of DNase resistant particles was done also as described. After virus purification, 13 ml of a viral suspension, with a titer of 4.8×10^9 DRP per ml, was obtained.

Chicken embryo inoculation: SPF embryonating chicken eggs were inoculated with 200 μ l of the recombinant viral stock obtained. The inoculums corresponded to a dose of 1×10^9 DRP per embryo. No mortality attributable to the virus inoculation was observed.

Assessment of *LacZ* activity *in vivo*: β -galactosidase (Bgal) activity in cell lysates from embryos and hatched chicks was assessed as described. No statistically significant differences in Bgal activity, as assessed by chemiluminescence, were found between the rAAV-LacZ inoculated and mock inoculated embryos at 48 hr post inoculation (PI) (Table 3). Significant differences ($p < 0.001$) were found between rAAV-LacZ inoculated and PBS inoculated embryos at 120 hr PI, at 7 days of age and at 14 days of age. The RLU readings obtained from samples of the 7 days of age group were significantly higher ($p < 0.001$) than those from embryos 120 hr PI and 14 days of age. No significant difference was found between the RLU values obtained at 120 hr PI and 14 days of age.

Chemiluminescence detection of LacZ activity in cell cultures: The Bgal activity of cell lysates obtained from the cell cultures was assessed as described. Statistical analyses of variance and pair comparisons were performed within each cell passage and within each cell line group. In all 5 passages and cell lines, monolayers transfected with the plasmid containing the *LacZ* gene flanked by the AAV inverted terminal repeats in the presence of the Rep protein of the virus showed RLU values significantly higher ($p < 0.001$) than the homologous cell controls (Table 4). At the first passage, HEK 293 and LMC cells transfected with the lac-rc treatment showed significantly higher chemiluminescence readings than those transfected with the lac treatment alone. These patterns of differences between the treatment groups remained unchanged through the 4th cell passages performed. At the 5th cell passage, LMH cells transfected with the lac treatment showed no significant difference from the homologous cell control group. The statistical differences among the HEK 293 group remained unchanged at this passage level.

At the first passage, DF-1 cells transfected with the lac-rc treatment showed readings significantly higher than those transfected with the Sgal-rc treatment (Table 5). These two groups had RLU values that were significantly higher than the DF-1 cells transfected with the lac treatment, which in turn, showed RLU values significantly higher than the DF-1 cells transfected with the Sgal treatment. All transfected DF-1 cell groups showed RLU values significantly higher than the homologous control group. By the second passage, DF-1 cells transfected with the lac-rc treatment still showed RLU values significantly higher than those transfected with the Sgal-rc treatment, but this Sgal-rc treated group showed no significant difference from the DF-1 cells transfected with the

lac treatment. This pattern of differences remained unchanged until the 4th passage, in which the DF-1 cells transfected with the lac-rc treatment showed RLU values significantly higher than the cells transfected with the Sgal-rc treatment, which were, at this passage level, significantly higher than the DF-1 cells transfected with the lac and Sgal treatment. There was no significant difference between these last two treatments in comparison with the homologous cell controls. The same patterns of differences were observed in the groups of treatment of this cell line in the 5th passage.

Direct staining of cell cultures for visual detection of the *LacZ* activity: Cell monolayers presented different degrees of Bgal staining, which correlated with the levels of enzymatic activity detected by chemiluminescence at the various passages. Stained monolayers are showed in figure 1.

DISCUSSION:

Recombinant systems based on members of the *Dependovirus* subfamily of parvoviruses are currently the subject of intense research work, due to the promising characteristics that they exhibit for the purpose of transgene delivery and expression. In this study, the ability of the AAV to serve as a recombinant virus-based platform to achieved transgene expression in poultry *in vivo* has been demonstrated. Chemiluminescence proved to be a suitable tool for detection of Bgal expression due to its quantitative nature, which allowed for statistical analysis of the data obtained. Detectable gene expression was present in the experimental groups from 120 hr post transfection through the end of the experiment. Our *in vitro* results suggest that the presence of the inverted terminal repeats of the virus alone can enhance gene expression,

when a transgene is flanked by these AAV DNA sequences, as shown by the differences of Bgal activity observed among the groups of DF-1 cells transfected with the *LacZ* gene with and without the viral ITR. These results corroborate what has been observed in similar experiments performed with AAV-based plasmid constructs used for HIV vaccination experiments in mice (18). The loss of expression by the 5th passage in the LMH cells may be the result of the dilution effect ensued by the passage level on the amount of integrated or episomal *LacZ* transgene. The ITR-mediated enhancement may be the result of two characteristics of this DNA structure: first, previous studies have shown that AAV ITRs and the host's cellular recombination pathways are sufficient to promote integration transgenes flanked by these structures (19). Stable integration of the transgene may translate into sustained levels of expression for longer periods of time. Second, studies in zebra fish and mice have demonstrated that the viral ITRs can up-regulate the activity of strong eukaryotic promoters such as CMV, α -actin and β -actin (2, 18). The exact mechanism by which this enhancement occurs is not entirely understood, but these researchers suggest that the ITRs seem to dissect the promoter activity from the influence of cellular transcriptional regulatory elements during gene expression (2). Another finding of the *in vitro* experiments of this work was that the co-expression of the Rep protein of the virus can significantly enhance the expression of a transgene in the presence, and absence, of the viral ITRs. This is in accordance with previous research works performed with the AAV-based systems that indicate that co-expression of the Rep protein promoted site-specific genomic integration of transgenes flanked by the viral ITRs from both viral based and non-viral based (plasmid) systems (8, 15). Steigerwald *et al* (15) have also reported that Rep can mediate integration of genes encoded in plasmids

that lack the ITRs of the virus. Integration of these constructs lacking the viral ITRs is though to occur due to the presence of sequences in the plasmid backbone, like ColE1, that can serve as target for interactions with the rep protein. Another mechanism by which the Rep protein can mediate enhancement of expression has been demonstrated by Qui *et al* (9). These researchers found that AAV Rep proteins can act to increase the ratio of spliced versus un-spliced mRNA during transcription. In their experiment they found that the protein functions as a *trans* regulator of RNA processing by influencing the composition of factors associated with the carboxy-terminal domain of the RNA polymerase II.

In summary, we have demonstrated that recombinant systems based on the AAV can achieve expression of transgenes encoded either in a recombinant viral particle or a recombinant plasmid. Although extensive work remains to be performed to assess the ability of this recombinant system to induce protection in vaccination trials, the findings of this research corroborate that the avian member of this family of viruses exhibit some of the same characteristics that have encourage researches in the use of primate's dependoviruses for gene delivery and therapy. These characteristics may confer to AAV based systems (both virus-based and plasmid-based) important advantages over the current recombinant systems used for poultry vaccination.

REFERENCES

1. Estevez, C., and P. Villegas. Sequence Analysis, Viral Rescue from Infectious Clones and Generation of Recombinant Virions of the Avian Adeno Associated Virus. Submitted to Virus Res. 2004.

2. Hsiao, C. D., F. J. Hsieh, and H. J. Tsai. Enhanced expression and stable transmission of transgenes flanked by inverted terminal repeats from adeno-associated virus in zebrafish. *Dev. Dyn.* 220:323-336. 2001.
3. Kearns, W. G., S. A. Afione, S. B. Fulmer, M. C. Pang, D. Erikson, M. Egan, M. J. Landrum, T. R. Flotte, and G. R. Cutting. 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. 1996.
4. Muzyczka, N. Adeno-associated virus (AAV) vectors: will they work? *J. Clin. Invest.* 94:1351. 1994.
5. Muzyczka, N., and K. I. Berns. *Parvoviridae: The Viruses and Their Replication*. In: *Fields Virology*, D. M. Knipe, and P. M. Howley, vol. 2. Lippincott Williams & Wilkins, Philadelphia. pp. 2327-2379. 2001.
6. Ni, Y. H. In vivo hepatic gene therapy. *Acta Paediatr. Taiwan.* 42:191-200. 2001.
7. Pfeifer, A., and I. M. Verma. Gene therapy: promises and problems. *Annu. Rev. Genomics Hum. Genet.* 2:177-211. 2001.
8. Pieroni, L., C. Fipaldini, A. Monciotti, D. Cimini, A. Sgura, E. Fattori, O. Epifano, R. Cortese, F. Palombo, and N. La Monica. Targeted integration of adeno-associated virus-derived plasmids in transfected human cells. *Virology.* 249:249-259. 1998.
9. Qiu, J., and D. J. Pintel. The adeno-associated virus type 2 Rep protein regulates RNA processing via interaction with the transcription template. *Mol. Cell. Biol.* 22:3639-3652. 2002.
10. Robbins, P. D., and S. C. Ghivizzani. Viral vectors for gene therapy. *Pharmacol. Ther.* 80:35-47. 1998.

11. Rohr, U. P., M. A. Wulf, S. Stahn, U. Steidl, R. Haas, and R. Kronenwett. Fast and reliable titration of recombinant adeno-associated virus type-2 using quantitative real-time PCR. *J. Virol. Methods.* 106:81-88. 2002.
12. Schwaab, R., and J. Oldenburg. Gene therapy of hemophilia. *Semin. Thromb. Hemost.* 27:417-424. 2001.
13. Silva, R. F. 2002. Recombinant Vaccines and DNA Vaccines: Unfulfilled Promises. *Poultry Vaccines and Vaccination Practices Symposium.* 45th meeting of the American Association of Avian Pathologists, Nashville, Tennessee, 2002.
14. Smith-Arica, J. R., and J. S. Bartlett. Gene therapy: recombinant adeno-associated virus vectors. *Curr. Cardiol. Rep.* 3:43-49. 2001.
15. Steigerwald, R., C. Rabe, V. Schmitz, I. G. Schmidt-Wolf, M. Alt, and W. H. Caselmann. Requirements for adeno-associated virus-derived non-viral vectors to achieve stable and site-specific integration of plasmid DNA in liver carcinoma cells. *Digestion.* 68:13-23. 2003.
16. Tsunoda, H., T. Hayakawa, N. Sakuragawa, and H. Koyama. Site-specific integration of adeno-associated virus-based plasmid vectors in lipofected HeLa cells. *Virology.* 268:391-401. 2000.
17. Xiao, X., J. Li, and R. J. Samulski. Production of High-Titer Recombinant Adeno-Associated Virus Vectors in the Absence of Helper Adenovirus. *J. Virol.* 72:2224-2232. 1998.
18. Xin, K. Q., T. Ooki, N. Jounai, H. Mizukami, K. Hamajima, Y. Kojima, K. Ohba, Y. Toda, S. Hirai, D. M. Klinman, K. Ozawa, and K. Okuda. A DNA vaccine containing

inverted terminal repeats from adeno-associated virus increases immunity to HIV. *J. Gene Med.* 5:438-445. 2003.

19. Yang, C. C., X. Xiao, X. Zhu, D. C. Ansardi, N. D. Epstein, M. R. Frey, A. G. Matera, and R. J. Samulski. Cellular recombination pathways and viral terminal repeat hairpin structures are sufficient for adeno-associated virus integration in vivo and in vitro. *J. Virol.* 71:9231-9247. 1997.

20. Yates, V. J., A. M. el-Mishad, K. J. McCormick, and J. J. Trentin. Isolation and characterization of an Avian adenovirus-associated virus. *Infect. Immun.* 7:973-980. 1973.

Transfection Protocol of Cell Lines

Cell Line	pEVA3V-LacZ-ITR	pEVA3V-RC-CMV	pSgal	Cell Control	# Wells
LMH	++	--	--	--	10
LMH	++	++	--	--	10
LMH	--	--	--	++	6
HEK 293	++	--	--	--	10
HEK 293	++	++	--	--	10
HEK 293	--	--	--	++	6
DF-1	++	--	--	--	10
DF-1	++	++	--	--	10
DF-1	--	--	--	++	6
DF-1	--	--	++	--	10
DF-1	--	++	++	--	10

Table 1: Transfection protocol used for *in vitro* determination of transgene enhancement mediated by AAV Rep proteins. Six micrograms of each of the plasmid indicated by the ++ symbols were used for the transfection of the cell monolayers. The monolayers were grown in 6-well plates, and the total amount of wells for each transfection treatment is indicated in column at the far right. As an additional control, DF-1 cell monolayers were transfected with a plasmid containing the *LacZ* gene without the AAV ITR (pSgal), with or without the co-expression of the AAV Rep protein from plasmid pEVA3V-RC-CMV.

Transgene Detection Method and Monolayer Usage

Treatment	Cell line	#Wells	Chemiluminescence	Direct Staining	Sub-cultured
lac ^a	DF/293/LMH	10	6	1	3
lac-rc ^b	DF/293/LMH	10	6	1	3
Sgal ^c	DF	10	6	1	3
Sgal-rc ^d	DF	10	6	1	3
Cell Control	DF/293/LMH	6	3	1	2

Table 2: Details on the use of monolayers for transgene detection and sub-culturing for subsequent passages. A total of 10 wells in each passage, for each of the cell lines, were used for each plasmid treatment. Six monolayers of each of the cell lines were used as controls. The amount of monolayers used in each detection method is described. The sub-cultured column indicates the number of monolayers used for cell passage. A total of 5 passages were made.

^a Cells transfected with plasmid pEVA3V-LacZ-ITR alone.

^b Cells transfected with plasmid pEVA3V-LacZ-ITR and pEVA3V-RC-CMV.

^c Cells transfected with plasmid pSgal alone.

^d Cells transfected with plasmid pSgal and pEVA3V-RC-CMV.

Chemiluminescence Readings (in RLU) in Embryos and Chicks

48 hr PI				120 hr PI		
Inoculum	RLU ave	STD error	Bgal (pg)	RLU ave	STD error	Bgal (pg)
rAAAV-LacZ	1387.17 ^a	115.02	74.70	3333.33 ^b	143.52	87.31
PBS	997.33 ^a	162.66	72.17	868.00 ^a	202.97	71.34
7 days of age				14 days of age		
Inoculum	RLU ave	STD error	Bgal (pg)	RLU ave	STD error	Bgal (pg)
rAAAV-LacZ	4130.83 ^c	209.01	92.47	3202.50 ^b	156.62	86.46
PBS	969.33 ^a	295.58	71.99	1015.33 ^a	221.49	72.29

Table 3: Results of chemiluminescence readings obtained from tissue samples of SPF embryonating chicken eggs and chicks inoculated with rAAAV-LacZ and mock inoculated. Relative light units (RLU) results are given as the average of six samples readings (rAAAV-LacZ inoculated) or 3 samples readings (PBS inoculated) from the embryos of hatched chicks. The standard error (STD error) for each group of readings is also presented. Different superscripts indicate statistically different ($p < 0.001$) values at each sampling time. Bgal (pg) column indicates the predicted activity of β -galactosidase present in the sample, as calculated by the linear regression formula obtained for the control reactions.

Chemiluminescence Readings (in RLU) of Cell Cultures

Cell treatment	1st Passage	2nd Passage	3rd Passage	4th Passage	5th Passage
LMH Lac	4804.17 ^a	4142.55 ^a	1899.36 ^a	1020.25 ^a	714.17 ^a
LMH Lac-rc	19835.83 ^b	17170.23 ^b	7870.88 ^b	3541.90 ^b	1062.57 ^b
LMH cc	538.33 ^c	673.33 ^c	706.66 ^c	449.33 ^c	467.67 ^a
HEK 293 Lac	20432.67 ^a	18180.83 ^a	8335.91 ^a	3751.16 ^a	1125.35 ^a
HEK 293 Lac-rc	44971.83 ^b	38862.18 ^b	17818.31 ^b	8018.24 ^b	2004.56 ^b
HEK 293 cc	746.33 ^c	895.03 ^c	821.67 ^c	491.33 ^c	529.33 ^c

Table 4: Chemiluminescence readings of cell cultures transfected with the different plasmid treatments described. Shown are the results of readings obtained at from LMH and HEK 293 cell lines. Results are the average of RLU readings of the 6 cell culture replicas of transfected cells and 3 replicas of the control monolayers. RLU readings with the same superscript within each column are not significantly different ($p < 0.001$). Statistical comparisons were made within each cell passage and within each cell line.

Cell treatment	1st Passage	2nd Passage	3rd Passage	4th Passage	5th Passage
DF1 Lac	5503.50 ^c	4740.35 ^b	2173.45 ^b	978.05 ^c	635.73 ^c
DF1 Lac-rc	15015.00 ^a	13072.84 ^a	5993.89 ^a	2697.25 ^a	1753.21 ^a
DF1 Sgal	3230.33 ^d	1562.50 ^c	622.17 ^c	601.17 ^c	577.50 ^c
DF1 Sgal-rc	8618.50 ^b	4026.67 ^b	2419.00 ^b	1922.33 ^b	1231.67 ^b
DF1 cc	727.00 ^d	856.79 ^c	656.67 ^c	467.17 ^c	439.33 ^c

Table 5: Chemiluminescence readings of cell cultures transfected with the different plasmid treatments described. Shown are the results of readings obtained at from DF-1 cell line. Results are the average of RLU readings of the 6 cell culture replicas of transfected cells and 3 replicas of the control monolayers. RLU readings with the same superscript within each column are not significantly different ($p < 0.001$). Statistical comparisons were made within each cell passage and among all groups.

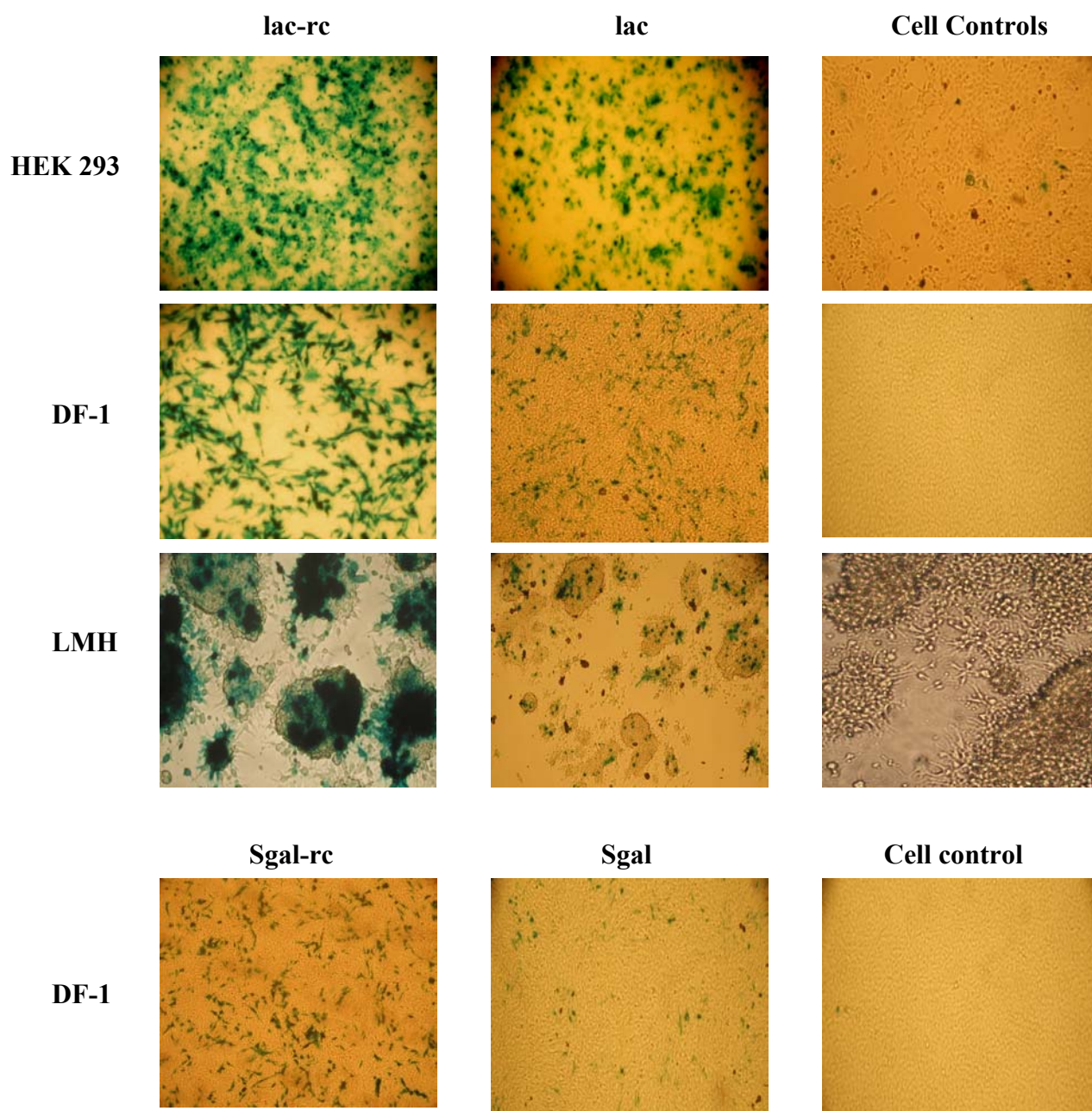


Figure 1: Direct staining of cell cultures at passage #3. Indicated at the top of each column are the plasmid transfection treatments (lac-rc, lac, Sgal-rc, Sgal and cell controls). Cell line types are indicated at the left of each row.

CHAPTER V

CONCLUSION AND DISCUSSION

Infectious clones of the two known strains of the avian adeno-associated virus have been obtained. Sequence analysis of these clones showed that these AAV strains share an overall percentage of identity of 94% in their nucleotide sequences. Detailed analysis of the two coding regions (Rep and Cap) showed identity values of 96% and 90%, respectively. When compared to the primate strains of the AAV, the overall nucleotide sequence homologies observed varied from 38% to 45% in strain DA-1, and 40% to 45% for strain VR 865. Mapping of the regulatory sequences in the viral genomes showed that regions in charge of initiation of viral DNA replication, through interactions with the non-structural (Rep) proteins of the virus, are conserved with respect to the primate member of this family of viruses. Furthermore, DNA regions with high probability of serving as target for host's cell regulatory proteins were found throughout the viral genomes, which is also in accordance with studies that indicate that the host's cellular machinery play an essential role in the replication cycle of these viruses (Muzyczka and Berns, 2001). The nucleotide sequence of the Rep coding region of the AAVs was found to be less conserved, when compared to the AAVs sequences, than the Cap coding region. However, closer analysis of the putative amino acid sequence of the viruses showed that motifs necessary for important functions of the protein, namely the helicase and nuclease activities are highly conserved. This finding suggest that the

AAAVs' uses the same mechanisms of replication that the primate counterparts of the *dependovirus* subfamily of parvoviruses.

Genomic site-specific integration, one of the hallmarks of the biology of these viruses, is known to occur in the absence of a helper virus during AAV infections. This process is mediated by the virus ITR in conjunction with the major Rep proteins (Rep 78/68) (Pieroni et al., 1998; Steigerwald et al., 2003). Although integration has been well characterized in the primates' dependoviruses, little is known about the process in the avian dependoviruses. Analysis of our clones showed that the structures necessary for the process to occur are conserved in the genome of the AAHV. The inverted terminal repeat structures found in the AAHV share common motifs with the AAV ITR, known to mediate interactions with the Rep protein of the virus. These structures, the Rep binding element, Rep binding element prime and terminal resolution site, mediate Rep protein binding of the viral DNA, nicking of the DNA molecule for initiation of DNA replication and integration into the host genome (Muzyczka and Berns, 2001). Although it is reasonable to assume that the AAHV share common mechanisms of DNA replication, and probably integration in the host's genome in the absence of a helper virus, further studies are warranted to elucidate the specifics of these viruses replication cycles. One important aspect in this regard is that integration is known to occur specifically in a DNA sequence called AAVS1 in the slow skeletal muscle troponin T1 gene of humans, which is located in chromosome 19. The existence of a homolog sequence in the chicken genome is not known, so the target for genomic integration is one of the questions that need to be elucidated.

One of the interesting findings of this work is the fact that we could generate wtAAAV particles from the cloned viral DNA, by co-infection of HEK 293 cell line cultures (which express the E1A and E1B immediate early gene of the human adenovirus 5) with these viral clones and a plasmid coding for the E2A, E4 and VA RNA genes from the human adenovirus type 5. Expression of these immediate early genes of the adenovirus is essential to drive a productive replication cycle of the AAV. These early genes do not seem to interact directly with the viral Rep proteins or DNA, but rather set the stage for productive replication of the AAV, by inducing the host cell to enter into S phase and produce proteins necessary to complete the virus replication. This seems to be confirmed by the fact that AAV could be rescued from this *in vitro* system using a human adenovirus, and also by the fact that genes homologous to these immediate early genes have not been found in avian adenoviruses, but still these avian adenoviruses are capable to induce similar host cell changes that are conducive to AAV replication. The use of the AAV infectious clones obtained in this work could be a powerful tool in the elucidation of the avian adenovirus genes homolog to the human adenovirus immediate early genes; if studies are performed using these clones as a reporter system in experiments of expression and mutagenesis of avian adenovirus genes.

From the cloned AAVs DNA, a plasmid-based system for generation of recombinant viral particles has been obtained. From this plasmid system we have produced rAAAV coding for the *LacZ* reporter gene, which was used to express β -galactosidase *in vitro* and *in vivo*. Strong levels of activity of the transgene were observed in primary cell cultures (chicken embryo fibroblasts, liver and kidney cells) regardless of cell type. The highest levels of expression were observed in primary fibroblasts, and the

second and third highest levels were observed in liver and kidney cell cultures, respectively.

In the *in vivo* part of the study, expression of the transgene was detected and sustained in chicks up to 14 days of age, at which time the experiment was terminated. Although more experimental work is necessary to assess the ability of the system to generate protection by expression of immunogenic peptides *in vivo*, these preliminary experiments showed that indeed these recombinant viruses are capable of inducing transgene expression that was long lasting.

Transgene expression was found to be enhanced *in vitro*, when the *LacZ* reporter gene was expressed in the presence of the Rep proteins of the virus. The β -galactosidase activity measured in cell line cultures transfected with plasmids harboring the reporter gene in the presence or absence of the Rep protein was significantly higher when these non-structural proteins of the virus were expressed. Furthermore, we found that the ITR alone can enhance gene expression, as demonstrated by the β -galactosidase activity measured in DF-1 cell cultures transfected with plasmids containing the *LacZ* gene with and without the viral ITR flanking it. Enhancement of gene expression in recombinant AAV-based systems has been previously reported (Hsiao et al., 2001). The mechanism of ITR mediated enhancement of expression is thought to be the result of two characteristic of this DNA sequence: The first is that the ITR can mediate integration into the host genome via cellular recombination pathways (Xin et al., 2003). This integration results in longer and sustained periods of expression. The second characteristic of these DNA structures is that they can up-regulate the activity of strong eukaryotic promoters such as CMV, α actin and β actin (Hsiao et al., 2001). This is thought to occur because these

sequences serve to dissect the promoter activity from the influence of cellular transcription regulatory elements.

The Rep-mediated enhancement of gene expression is believed to be the result of site-specific integration of genes flanked by the viral ITR that is ensued by this protein, which increases the longevity of expression. Also, this protein has been demonstrated to increase the ratio of spliced versus un-spliced mRNA during transcription. Researchers have found that this protein functions as a *trans* regulator of RNA processing by influencing the composition of factors associated with the carboxy-terminal domain of the RNA polymerase II (Qiu and Pintel, 2002).

In summary, the findings of the experiments conducted with the recombinant AAV-based system indicate that the virus could be used for the purpose of gene delivery and immunization in poultry and that these viruses offer distinctive advantages, such as up-regulated gene expression, that could be very useful in the prevention and control of infectious diseases in the poultry industry.

REFERENCES

- Hsiao, C.D., Hsieh, F.J. and Tsai, H.J. (2001) Enhanced expression and stable transmission of transgenes flanked by inverted terminal repeats from adeno-associated virus in zebrafish. *Dev Dyn* 220(4), 323-36.
- Muzyczka, N. and Berns, K.I. (2001) *Parvoviridae: The Viruses and Their Replication*. In: D.M. Knipe and P.M. Howley (Eds), *Fields Virology*, pp. 2327-2379. Vol. 2. 2 vols. Lippincott Williams & Wilkins, Philadelphia.
- Pieroni, L., Fipaldini, C., Monciotti, A., Cimini, D., Sgura, A., Fattori, E., Epifano, O., Cortese, R., Palombo, F. and La Monica, N. (1998) Targeted integration of adeno-associated virus-derived plasmids in transfected human cells. *Virology* 249(2), 249-59.
- Qiu, J. and Pintel, D.J. (2002) The adeno-associated virus type 2 Rep protein regulates RNA processing via interaction with the transcription template. *Mol Cell Biol* 22(11), 3639-52.
- Steigerwald, R., Rabe, C., Schmitz, V., Schmidt-Wolf, I.G., Alt, M. and Caselmann, W.H. (2003) Requirements for adeno-associated virus-derived non-viral vectors to achieve stable and site-specific integration of plasmid DNA in liver carcinoma cells. *Digestion* 68(1), 13-23.
- Xin, K.Q., Ooki, T., Jounai, N., Mizukami, H., Hamajima, K., Kojima, Y., Ohba, K., Toda, Y., Hirai, S., Klinman, D.M., Ozawa, K. and Okuda, K. (2003) A DNA

vaccine containing inverted terminal repeats from adeno-associated virus
increases immunity to HIV. J Gene Med 5(5), 438-45.