MOLECULAR CHARACTERIZATION OF ADENO-ASSOCIATED VIRUS IN THE NATURAL HOST

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
The Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By
Ryan L. Jensen, B.S.

*****

The Ohio State University
2008

Dissertation Committee:
Virginia M. Sanders, Adviser
Philip R. Johnson
Jeffrey S. Bartlett
Douglas M. McCarty

Approved by

______________________________
Adviser
Integrated Biomedical Science Graduate Program
ABSTRACT

Infection with wild-type adeno-associated virus (AAV) is common in humans, but very little is known about the in vivo biology of AAV. On a molecular level, it has been shown in cultured cells that AAV integrates in a site-specific manner on human chromosome 19, but this has never been demonstrated directly in infected human tissues. Based on the extensive in vitro data demonstrating site-specific integration of AAV, our hypothesis was that targeted integration would also occur in vivo. To that end, we tested 175 pediatric tissue samples for the presence of AAV DNA, and when present, examined the specific form of the viral DNA. AAV was detected in 7 of 101 tonsil-adenoid samples and in 2 of 74 other tissue samples (spleen and lung). Sequence analyses showed that 8 of the capsid sequences were AAV2-like (~98% amino acid identity), while the single spleen isolate was intermediate between serotypes 2 and 3. In these 9 samples, we were unable to detect AAV integration in the AAVS1 locus using a sensitive PCR assay designed to amplify specific viral-cellular DNA junctions. Additionally, we used a second complementary assay (LAM-PCR) to widen our search for integration events. Analysis of individual LAM-PCR products revealed that the AAV genomes were arranged predominately in a head-to-tail array, with deletions and extensive rearrangements in the inverted terminal repeat (ITR) sequences. A single AAV-cellular junction was identified from a tonsil sample and it mapped to a highly repetitive satellite
DNA element on chromosome 1. Given these data, we entertained the possibility that instead of integrated forms, AAV genomes were present as extra-chromosomal forms. We used a novel amplification assay (linear rolling circle amplification) to show that the majority of wild-type AAV DNA existed as circular double-stranded episomes in our tissues. We then go on to describe the molecular structure of several AAV molecular clones (some infectious) isolated from these pediatric tissues. DNA sequence analysis of the molecular clones revealed the ubiquitous presence of a double-D ITR structure, which suggests a mechanism by which the virus is able to maintain ITR sequence continuity and persist in the absence of host chromosome integration. To further define the molecular form of AAV persistence in a relevant animal model, we surveyed rhesus macaque tissues for the presence of wild-type AAV sequences and when detected, determined the molecular form of AAV persistence. To that end, we tested 28 monkey tissue samples from 9 animals and detected AAV sequence in 25 of the samples (89.3%). Significantly, in these 9 samples, we were unable to detect AAV integration in the macaque AAVS1 locus using a sensitive PCR assay designed to amplify viral-cellular DNA junctions. This was particularly notable given the high AAV copy number in several tissues. Given these data combined with our findings from human tissues, it seemed likely that the AAV genomes were persisting un-integrated and present predominately as extra-chromosomal forms. Finally, based on the knowledge that AAV persists as episomes, we go on to create novel AAV phenotypes with enhanced characteristics.
Dedicated to my family
ACKNOWLEDGMENTS

I wish to thank my adviser, Dr. Philip Johnson, for his guidance, support, and expertise throughout this research. His training taught me to think critically and always try to ask the right questions. I would also like to thank Drs. Jeffrey Bartlett, Douglas McCarty, and Virginia Sanders for serving on my dissertation committee and for their support, suggestions, and comments. Specifically, I acknowledge Dr. Bartlett for his work generating the computer models of AAV capsids. Additionally, I thank Dr. Allan Yates for serving on my dissertation committee for most of this research and for his dedicated service to myself and all of the students associated with the Integrated Biomedical Science Graduate Program. I wish to thank Drs. Reed Clark and Bruce Schnep for their friendship and for creating a supportive working environment. In particular, I thank Bruce Schnep for always being there as a mentor, friend, and guide. I acknowledge that his work created the foundation for the research presented herein. Bruce Schnep developed and modified many of the assays described throughout this research.

I am thankful for the constant love and support of my family. I thank my parents for their support, guidance and love throughout my life. Finally, I am most thankful for my wife Emily, who has been through it all with me. I am forever grateful for her many sacrifices, love, and support throughout graduate school.
VITA

May 1, 1977…………………………………… Born, Roy, Utah

2002……………………………………………… B.S., Biology, Utah State University

2002 – 2005………………………………… Graduate Research Associate, Integrated Biomedical Science Graduate Program, The Ohio State University

2005 – present……………………………… Research Associate, Division of Infectious Diseases, Children’s Hospital of Philadelphia

PUBLICATIONS


*Contributed equally to this work


FIELDS OF STUDY

Major Field: Integrated Biomedical Science
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>v</td>
</tr>
<tr>
<td>Vita</td>
<td>vi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xii</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>xiv</td>
</tr>
</tbody>
</table>

Chapters:

1. **Introduction** .............................. 1
   - Wild-type adeno-associated viruses
     - Overview .................................. 3
     - Serotypes ................................ 6
     - Genome structure ........................ 7
     - Latency .................................. 10
     - Site-specific integration ................ 13
   - Recombinant Adeno-Associated Virus Vectors
     - Overview .................................. 15
     - Tropism modification ...................... 16
     - Integration of rAAV vectors .............. 19
     - rAAV vector persistence and gene expression .... 20
   - Statement of hypothesis ................... 21
2. Molecular characterization of adeno-associated viruses in children

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>22</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>23</td>
</tr>
<tr>
<td>Human tissues</td>
<td>23</td>
</tr>
<tr>
<td>Adenovirus isolation from tonsil and adenoids</td>
<td>24</td>
</tr>
<tr>
<td>DNA isolation from human tissue samples</td>
<td>24</td>
</tr>
<tr>
<td>AAV capsid PCR</td>
<td>25</td>
</tr>
<tr>
<td>Adenovirus hexon PCR</td>
<td>26</td>
</tr>
<tr>
<td>Quantitative PCR</td>
<td>26</td>
</tr>
<tr>
<td>Sequence analyses</td>
<td>27</td>
</tr>
<tr>
<td>Heparin binding analysis</td>
<td>27</td>
</tr>
<tr>
<td>AAV serology</td>
<td>27</td>
</tr>
<tr>
<td>Genbank accession numbers</td>
<td>29</td>
</tr>
<tr>
<td>Results</td>
<td>29</td>
</tr>
<tr>
<td>AAV in tonsils and adenoids</td>
<td>29</td>
</tr>
<tr>
<td>AAV DNA in other tissues</td>
<td>35</td>
</tr>
<tr>
<td>AAV cap and rep sequences from human tissues</td>
<td>35</td>
</tr>
<tr>
<td>Most circulating AAV2-like viruses are predicted to bind heparin sulfate</td>
<td>39</td>
</tr>
<tr>
<td>Sequence variation in the AAV capsid maps to surface exposed</td>
<td>48</td>
</tr>
<tr>
<td>Regions</td>
<td>52</td>
</tr>
<tr>
<td>Seroreactivity to AAV2</td>
<td>52</td>
</tr>
<tr>
<td>Discussion</td>
<td>55</td>
</tr>
</tbody>
</table>

3. Characterization of adeno-associated virus genomes isolated from human tissues

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>60</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>62</td>
</tr>
<tr>
<td>Cell propagation</td>
<td>62</td>
</tr>
<tr>
<td>Human tissues and DNA isolation</td>
<td>62</td>
</tr>
<tr>
<td>Quantitative PCR and DNA hybridization</td>
<td>63</td>
</tr>
<tr>
<td>AAVS1-PCR</td>
<td>64</td>
</tr>
<tr>
<td>LAM-PCR</td>
<td>65</td>
</tr>
<tr>
<td>Linear rolling circle amplification (LRCA)</td>
<td>67</td>
</tr>
<tr>
<td>Results</td>
<td>68</td>
</tr>
<tr>
<td>Identification of AAV sequences in human tissues</td>
<td>68</td>
</tr>
<tr>
<td>Molecular characterization of AAV DNA in tissues</td>
<td>68</td>
</tr>
<tr>
<td>Analysis of integration into the AAVS1 locus</td>
<td>69</td>
</tr>
<tr>
<td>LAM-PCR analysis of AAV integration</td>
<td>69</td>
</tr>
<tr>
<td>LRCA amplification of episomal DNA</td>
<td>79</td>
</tr>
<tr>
<td>AAV genomes persist predominantly as episomes in human tissues</td>
<td>88</td>
</tr>
<tr>
<td>Discussion</td>
<td>89</td>
</tr>
</tbody>
</table>
4. Infectious molecular clones of adeno-associated virus isolated from humans

Introduction...........................................................................................................97
Materials and Methods...........................................................................................99
   Cells and viruses.................................................................................................99
   Manipulation of nucleic acids............................................................................99
   Linear rolling circle amplification (LRCA).........................................................99
   Analysis of LRCA amplified ITR products.......................................................100
   Functional analysis of AAV molecular clones.................................................101
   Western blot analysis.......................................................................................102
   Real-time PCR..................................................................................................102
Results...................................................................................................................103
   Sequence-specific linear rolling circle amplification (SSLRCA)………………103
   Direct amplification of AAV genomes from human tissues............................106
   Isolation on infectious AAV molecular clones.................................................107
   Characterization of AAV products from infectious molecular clones............113
   Double-D ITR structures present in tissues......................................................117
Discussion.............................................................................................................122

5. Episomal persistence of adeno-associated virus genomes isolated from rhesus macaques

Introduction...........................................................................................................127
Materials and Methods...........................................................................................129
   Cell propagation.................................................................................................129
   Rhesus macaque tissue procurement...............................................................129
   Functional analysis of wild-type AAV molecular clones....................................129
   Isolation and detection of AAV sequences.......................................................130
   Quantitative PCR.............................................................................................131
   Sequence analysis............................................................................................131
   AAVS1 PCR......................................................................................................131
   Cloning and sequence of rhesus AAVS1 locus.................................................132
   PS-DNase treatments.........................................................................................133
   Linear rolling circle amplification (LRCA).......................................................133
Results...................................................................................................................134
   Detection of wild-type AAV sequences on rhesus macaque DNA................134
   AAVS1 integration analysis..............................................................................135
   Presence of episomal AAV genomes in rhesus tissue....................................142
   Use of linear rolling circle amplification (LRCA) to amplify and clone AAV DNA.................................................................147
   Molecular clones of AAV isolated from LRCA are functional......................151
Discussion.............................................................................................................155
6. Directed evolution of adeno-associated virus in primary human cells

Introduction..........................................................................................159

Materials and Methods.............................................................................162
  Cell propagation and viruses.................................................................162
  Replication-competent AAV (rcAAV) plasmid construction...............162
  DNA shuffling....................................................................................163
  rcAAV particle library generation......................................................164
  Sequence analysis..............................................................................165
  AAV infections....................................................................................165
  DNA isolation.....................................................................................166
  Transductions.....................................................................................166
  Acquisition and analysis.................................................................167
  Southern blot analysis......................................................................167

Results..................................................................................................167
  Library generation..............................................................................167
  Selection of AAV library clones targeting CD34+ hematopoietic stem cells.................................................................171
  Selection of AAV library clones targeting other primary human cells........................................................................177

Discussion..............................................................................................181

7. General conclusions............................................................................186

Bibliography..........................................................................................189
<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Status of clinical trials using AAV-based vectors</td>
</tr>
<tr>
<td>2.1</td>
<td>Summary of AAV and adenovirus sequence detection in human tissues</td>
</tr>
<tr>
<td>2.2</td>
<td>AAV sequence relatedness and DNA copy number in pediatric tissues</td>
</tr>
<tr>
<td>2.3</td>
<td>Heparin sulfate column binding of AAV2 virus preparations</td>
</tr>
<tr>
<td>2.4</td>
<td>Summary of synonymous and non-synonymous nucleotide substitutions in rep and cap</td>
</tr>
<tr>
<td>3.1</td>
<td>Summary of LAM-PCR positive clones isolated from clinical samples</td>
</tr>
<tr>
<td>4.1</td>
<td>LRCA assay sensitivity</td>
</tr>
<tr>
<td>4.2</td>
<td>Summary of molecular clones isolated by LRCA</td>
</tr>
<tr>
<td>5.1</td>
<td>AAV DNA copy number in rhesus macaque tissue</td>
</tr>
<tr>
<td>5.2</td>
<td>Percentage of sequence similarity between REP78 and VP1 amino acid sequences</td>
</tr>
<tr>
<td>5.3</td>
<td>Summary of molecular clones isolated by LRCA</td>
</tr>
<tr>
<td>6.1</td>
<td>Quality control of shuffled libraries</td>
</tr>
<tr>
<td>6.2</td>
<td>Comparison of transduction efficiencies of rAAV-eGFP vectors on CD34+ cells</td>
</tr>
<tr>
<td>6.3</td>
<td>Transduction efficiencies of rAAV-eGFP vectors on SMM cells</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>A negatively stained electron micrograph of AAV and adenovirus particles</td>
<td>5</td>
</tr>
<tr>
<td>1.2</td>
<td>Neighbor-joining phylogenies of the VP1 sequence of primate AAVs</td>
<td>9</td>
</tr>
<tr>
<td>1.3</td>
<td>Secondary structure of the AAV2 ITR</td>
<td>12</td>
</tr>
<tr>
<td>1.4</td>
<td>Overview of the recombinant AAV transient transfection production system</td>
<td>18</td>
</tr>
<tr>
<td>2.1</td>
<td>PCR schematic for amplification of the complete AAV capsid coding region</td>
<td>32</td>
</tr>
<tr>
<td>2.2</td>
<td>Predicted VP1 capsid amino acid alignment of AAV2 and novel human AAVs</td>
<td>38</td>
</tr>
<tr>
<td>2.3</td>
<td>Phylogenetic analysis of VP1 capsid nucleotide sequences</td>
<td>41</td>
</tr>
<tr>
<td>2.4</td>
<td>S17 sequence homology comparison with AAV2 and AAV3</td>
<td>43</td>
</tr>
<tr>
<td>2.5</td>
<td>Surface diagrams of AAV2 trimer atomic models</td>
<td>47</td>
</tr>
<tr>
<td>2.6</td>
<td>Ribbon diagrams of atomic models of AAV2 VP3 trimers showing the location</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>of predicted amino acid substitutions in the human AAV isolates</td>
<td></td>
</tr>
<tr>
<td>2.7</td>
<td>Serology of AAV infection as a function of age</td>
<td>54</td>
</tr>
<tr>
<td>3.1</td>
<td>AAVS1 PCR on human samples</td>
<td>71</td>
</tr>
<tr>
<td>3.2</td>
<td>Schematic of linear amplification mediated PCR (LAM-PCR) to detect random</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>AAV integrants</td>
<td></td>
</tr>
<tr>
<td>3.3</td>
<td>LAM-PCR validation and sensitivity using Detroit 6 cells</td>
<td>76</td>
</tr>
<tr>
<td>3.4</td>
<td>LAM-PCR of positive human clinical samples</td>
<td>78</td>
</tr>
<tr>
<td>3.5</td>
<td>Schematic of AAV-AAV junctions isolated following LAM-PCR</td>
<td>82</td>
</tr>
</tbody>
</table>
3.6 Linear rolling circle amplification (LRCA) for the detection of AAV episomes

3.7 Wild-type AAV genomes are present predominantly as episomes

4.1 LRCA schematic for the isolation of infectious AAV molecular clones

4.2 LRCA method detects episomal AAV in human tissue DNA

4.3 Identification of infectious AAV molecular clones

4.4 Rep and cap expression of T88 molecular clones

4.5 Structure of the AAV ITR junctions from selected molecular clones

4.6 Episomal AAV contains double-D ITR structures in vivo

5.1 AAVS1 PCR validation on Detroit 6 cells

5.2 AAVS1 PCR on rhesus samples

5.3 Schematic of the PS-DNase assay for the episomal wild-type AAV Genomes

5.4 Southern blot analysis of wild-type AAV molecular clones in rhesus lymph node following PS-DNase treatment

5.5 LRCA products

5.6 Functional analysis of wild-type AAV molecular clones

6.1 Chimerism of L18B library

6.2 Schematic of linear rolling circle amplification mediated evolution

6.3 LRCA products after one round of CD34+ cell selection
## LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAV</td>
<td>Adeno-associated virus</td>
</tr>
<tr>
<td>AAVS1</td>
<td>Adeno-associated virus integration site 1</td>
</tr>
<tr>
<td>Ad</td>
<td>Adenovirus</td>
</tr>
<tr>
<td>ADA</td>
<td>Adenosine deaminase</td>
</tr>
<tr>
<td>cAAV</td>
<td>Circular adeno-associated virus</td>
</tr>
<tr>
<td>CPE</td>
<td>Cytopathic effect</td>
</tr>
<tr>
<td>DRP</td>
<td>DNase resistant particle</td>
</tr>
<tr>
<td>FKBP52</td>
<td>FK506-binding protein</td>
</tr>
<tr>
<td>HMW</td>
<td>High molecular weight</td>
</tr>
<tr>
<td>HSPG</td>
<td>Heparin sulfate proteoglycan</td>
</tr>
<tr>
<td>HSC</td>
<td>Hematopoietic stem cells</td>
</tr>
<tr>
<td>HVR</td>
<td>Hypervariable region</td>
</tr>
<tr>
<td>ITR</td>
<td>Inverted terminal repeat</td>
</tr>
<tr>
<td>IU</td>
<td>Infectious unit</td>
</tr>
<tr>
<td>LRCA</td>
<td>Linear rolling circle amplification</td>
</tr>
<tr>
<td>MOI</td>
<td>Multiplicity of infection</td>
</tr>
<tr>
<td>NHEJ</td>
<td>Non-homologous end joining</td>
</tr>
<tr>
<td>NHP</td>
<td>Non-human primates</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
</tbody>
</table>

xiv
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>rAAV</td>
<td>Recombinant adeno-associated virus</td>
</tr>
<tr>
<td>Rfd</td>
<td>Replicating dimeric form</td>
</tr>
<tr>
<td>Rfm</td>
<td>Replicating monomeric form</td>
</tr>
<tr>
<td>RBS</td>
<td>Rep biding site</td>
</tr>
<tr>
<td>SCID</td>
<td>Severe combined immunodeficiency</td>
</tr>
<tr>
<td>ssLRCA</td>
<td>Sequence-specific linear rolling circle amplification</td>
</tr>
<tr>
<td>trs</td>
<td>Terminal resolution site</td>
</tr>
<tr>
<td>VP1</td>
<td>Viral protein 1</td>
</tr>
<tr>
<td>VP2</td>
<td>Viral protein 2</td>
</tr>
<tr>
<td>VP3</td>
<td>Viral protein 3</td>
</tr>
<tr>
<td>WAS</td>
<td>Wiscott-Aldrich syndrome</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

One of the most promising virus vectors being studied for human gene transfer is derived from adeno-associated virus (AAV), a single-stranded DNA virus that is ubiquitous to humans. Wild-type AAV has characteristics that make it an attractive candidate for development as a gene transfer vector, including: (1) no known pathogenicity; (2) limited capacity to induce immune responses; (3) the ability to transduce post-mitotic cells; (4) possible advantages of site-specific integration; and, (5) a very broad tropism. Due to this promising profile, recombinant AAV (rAAV) vectors, based on the wild-type virus, are being developed for an ever-growing variety of therapeutic applications that has resulted in the initiation of several human clinical trials (Table 1.1) (25, 78, 189, 190). Currently, these clinical trials are either in phase I or in phase II. The former studies aim at determining safety and often also maximum tolerable doses of the therapeutic agent, while the latter deals with the assessment of its efficacy and have higher statistical significance to detect potential side effects. Enthusiasm for AAV is due, not only to the relative safety of these vectors, but also to advances in understanding the unique biology of this virus.
<table>
<thead>
<tr>
<th>Indication</th>
<th>Gene</th>
<th>Route of administration</th>
<th>Phase</th>
<th>Subject number</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cystic fibrosis</td>
<td>CFTR</td>
<td>Lung via aerosol</td>
<td>I</td>
<td>12</td>
<td>Complete</td>
</tr>
<tr>
<td></td>
<td>CFTR</td>
<td>Lung via aerosol</td>
<td>II</td>
<td>38</td>
<td>Complete</td>
</tr>
<tr>
<td></td>
<td>CFTR</td>
<td>Lung via aerosol</td>
<td>II</td>
<td>100</td>
<td>Complete</td>
</tr>
<tr>
<td>Hemophilia B</td>
<td>FIX</td>
<td>Intramuscular</td>
<td>I</td>
<td>9</td>
<td>Complete</td>
</tr>
<tr>
<td></td>
<td>FIX</td>
<td>Hepatic artery</td>
<td>I</td>
<td>6</td>
<td>Ended</td>
</tr>
<tr>
<td>Arthritis</td>
<td>TNFR:Fc</td>
<td>Intrarticular</td>
<td>I</td>
<td>1</td>
<td>Ongoing</td>
</tr>
<tr>
<td>Hereditary emphusema</td>
<td>AAT</td>
<td>Intramuscular</td>
<td>I</td>
<td>12</td>
<td>Ongoing</td>
</tr>
<tr>
<td>Muscular dystrophy</td>
<td>Sarcoglycan</td>
<td>Intramuscular</td>
<td>I</td>
<td>10</td>
<td>Ongoing</td>
</tr>
<tr>
<td>Parkinson's</td>
<td>GAD65, GAD67</td>
<td>Intracranial</td>
<td>I</td>
<td>12</td>
<td>Complete</td>
</tr>
<tr>
<td>Canavan's</td>
<td>AAC</td>
<td>Intracranial</td>
<td>I</td>
<td>21</td>
<td>Ongoing</td>
</tr>
<tr>
<td>Batten's</td>
<td>CLN2</td>
<td>Intracranial</td>
<td>I</td>
<td>10</td>
<td>Ongoing</td>
</tr>
<tr>
<td>Alzheimer's</td>
<td>NGF</td>
<td>Intracranial</td>
<td>I</td>
<td>6</td>
<td>Ongoing</td>
</tr>
</tbody>
</table>


**Table 1.1: Status of clinical trials using AAV-based vectors.**
Because of the interest in recombinant DNA delivery vectors based on wild-type AAV, a more thorough understanding of the basic biology of wild-type AAV infection seems warranted. In fact, the ultimate success of rAAV vectors may be dependent on understanding the mechanisms of AAV persistence after a successful infection; how it affects and is affected by various cell types, and the long- and short-term consequences for the infected cell. A better understanding of these issues will undoubtedly influence the use of rAAV vectors for gene transfer in humans.

**BACKGROUND INFORMATION**

**WILD-TYPE ADENO-ASSOCIATED VIRUSES**

**Overview.** AAV is currently the only member of the genus Dependovirus, which belongs to the Parvoviridae family. AAV was initially observed in the mid-1960s as 20-nm contaminants ‘associated’ with several adenovirus preparations (Figure 1.1). It was then determined that the contaminants were viral particles and not adenovirus degradation products (5, 82). Infectious isolates of AAV have been obtained from humans in the presence of numerous adenovirus serotypes (18, 19). Evidence of AAV infection has also been sought using molecular epidemiology tools. AAV DNA has been detected in many tissues among cold- and warm-blooded animals, ranging from snakes to humans (20, 39, 50, 58, 90, 146, 164). Furthermore, serum antibodies to AAV are extremely common in the human population (16, 18, 19, 61, 62, 141, 154). However, in spite of this seemingly ubiquitous distribution and high seroprevalence to AAV, no study has conclusively linked AAV infection to any human disease.
Figure 1.1: A negatively stained electron micrograph of AAV and adenovirus particles. In this image, the larger adenovirus (diameter 60 – 90 nm) particles are easily distinguishable from the smaller adeno-associated virus (diameter 20 nm) particles.
Figure 1.1: A negatively stained electron micrograph of AAV and adenovirus particles.
AAV has the ability to establish a latent infection in the absence of coinfection by a helper virus (typically adenovirus) (14, 99). In the presence of adenovirus coinfection, a productive infection ensues and AAV is replicated to high levels.

**Serotypes.** Several distinct serotypes of AAV have been characterized, of which many were initially isolated as contaminants of adenovirus preparations (5, 82). AAV serotype 1 (AAV1) and AAV2 were found to be contaminants of the simian adenovirus type 15 (SV15) and SV12 stocks, respectively (5). AAV2 was the first to be cloned into bacterial plasmids, therefore it is the best characterized of all the serotypes and the bulk of our understanding of AAV biology has come from AAV2 (158). AAV3 was also found to be a contaminant of the simian adenovirus stock SV15 and shares overall sequence similarity of 82% with AAV2 (5, 130). Serotypes 1, 2, and 3 are the most common, and based on seroepidemiological studies, are only acquired in association with adenovirus infection. In fact, it is estimated that 70-90% of the human population by early adolescence contain antibodies against AAV1, 2, or 3 (16, 19). Serotype 4 is most likely a monkey specific AAV as it was isolated only in African green monkeys infected with the SV15 stock, and humans do not appear to have antibodies reactive with AAV4 (30). AAV5, originally identified from a human clinical sample, is thought to be sexually transmitted in association with herpes simplex virus (9, 24). The overall nucleotide sequence is only about 60% identical to AAV2, making AAV5 the most divergent of the human AAV serotypes (29). AAV6 was isolated as a contaminant of an SV5 stock and is a recombinant between AAV1 and AAV2 (155, 202). AAV6 is not considered a distinct serotype because it is neutralized by anti-AAV1 serum (155). AAV7 and AAV8 were isolated from rhesus monkey tissues by PCR, utilizing primers to highly conserved
regions of the capsid sequence (58). AAV9 was isolated from human tissues by the same PCR method as described for the isolation of AAV7 and AAV8 (60). Furthermore, Gao et al. screened hundreds of monkey and human tissues and sequence characterization of rescued AAV DNAs demonstrated a diverse array of molecular forms that segregated into clades whose members shared functional and serological similarities (Figure 1.2) (60). In addition to human and monkey tissues, AAV has also been isolated from different species including cow, chicken, sheep, snake, lizard, and goat (20, 39, 50, 90, 146, 164).

**Genome structure.** AAV encapsidates a linear, single-stranded DNA genome of both polarities with equal frequency (152). The most extensive studies have involved AAV serotype 2 (AAV2), which contains 4,679 deoxynucleotides, but the genetic map of all of the AAV serotypes is highly conserved (75, 119, 160, 184). The AAV2 genome includes 1 copy of a 145 bp inverted terminal repeat sequence at each end, and a unique sequence region of ~4.4 kb comprising the 2 main open reading frames for the rep (replication) and cap (capsid) gene (111, 112, 179).

The inverted terminal repeat (ITR) sequences are required in cis to provide functional origins of replication (ori), as well as provide signals for DNA replication, viral genome encapsidation, and integration (12, 119, 160, 206). The first 125 bases of the ITRs form a palindromic sequence, that when folded, forms a T-shaped structure. The ITRs are divided into four sub regions denoted as A, B, C, and D. The sequence at each end of the genome is highly conserved, but the palindromic regions of each ITR can be in
Figure 1.2: Neighbor-joining phylogenies of the VP1 sequence of primate AAVs.

Major nodes with bootstrap values of <75 are indicated with an “X.” A goose parvovirus and an avian AAV (6) were used as the out-group. Clades are indicated by name and by vertical lines to the right of the taxa from which they are made. The nomenclature for the taxa is either the serotype name or a reference to the species source (hu, human; rh, rhesus macaque; cy, cynomologus macaque; bb, baboon; pi, pigtailed macaque; ch, chimpanzee), followed by a number indicating the order in which they were sequenced. Clade C was identified and positively determined to have originated through the recombination of known clades. The AAV2-AAV3 hybrid clade originated after one recombination event, and its unrooted neighbor-joining phylogeny is shown.
Figure 1.2: Neighbor-joining phylogenies of the VP1 sequence of primate AAVs.
two different orientations, designated as flip and flop. In the flip orientation, the B palindrome (Figure 1.3) is closer to the end of the genome, whereas in the flop orientation, the C palindrome is closer to the end. This secondary structure provides a free 3' hydroxyl group for the initiation of viral DNA replication via a self-priming, strand-displacement mechanism involving leading-strand synthesis and double-stranded replicative intermediates (74, 181). The ITRs flank the two viral genes rep and cap encoding nonstructural and structural proteins, respectively. The rep gene is transcribed from the two separate promoters that produce two families of Rep proteins (Rep78 [78kDa], -68, and Rep52, -40) (120). The Rep proteins participate in almost every phase of the AAV life cycle and play central role in targeted integration. In addition, under non-permissive conditions, these proteins down-regulate gene expression (88, 98, 168, 185). The cap gene is expressed from a single promoter giving rise to two 2.3 kb mRNA transcripts that are alternatively spliced (129). The majority of the mRNA encodes the capsid VP3 protein, while the other mRNA transcript is spliced into the other capsid proteins VP1 and VP2, albeit at a 10-fold lower frequency.

**Latency.** Soon after its discovery, the ability of AAV to establish latency in the absence of any helper virus was shown (80). In these early studies, the “continuous carriage” of the AAV genome was described in an epithelial cell line (Detroit 6) isolated from bone marrow of a pulmonary carcinoma patient at a Detroit hospital (81). In these experiments, the cells were infected with a high multiplicity of AAV followed by continuing passages. After seven passages no infectious AAV could be detected. However, following super-infection with a helper virus (adenovirus) the cells readily produced infectious AAV. These latently infected cells could produce infectious AAV
**Figure 1.3: Secondary structure of the AAV2 ITR.** The AAV2 ITR serves as origin of replication and is composed of two arm palindromes (B-B' and C-C') embedded in a larger stem palindrome (A-A'). The ITR can acquire two configurations (flip and flop). The flip (depicted) and flop configurations have the B-B' and the C-C' palindrome closest to the 3' end, respectively. The D sequence is present only once at each end of the genome thus remaining single-stranded. The boxed motif corresponds to the Rep-binding site (RBS) where the AAV Rep78 and Rep68 proteins bind. The RBS consists of a tetranucleotide repeat with the consensus sequence 5'-GNGC-3'. The ATP-dependent DNA helicase activities of Rep78 and Rep68 remodel the A-A' region generating a stem-loop that locates at the summit the terminal resolution site (trs) in a single-stranded form. In this configuration, the strand- and site-specific endonuclease catalytic domain of Rep78 and Rep68 introduces a nick at the trs.
Figure 1.3: Secondary structure of the AAV2 ITR.
even after >100 passages by co-infection with a helper virus. However, given the lack of molecular techniques at the time, investigators were unable to delineate the mechanisms involved in establishing this latent state. Almost 20 years later, characterization of the AAV genomes in latently infected cells established that AAV DNA had integrated into cellular DNA in tandem arrays (14, 28). Through the years, these studies have been extended to a variety of cell lines, and these tandem arrays have been the common feature of AAV proviral structures and have been the basis of much research concerning AAV replication and integration (71, 99). However, it is important to note that in contrast to some other DNA viruses, AAV does not appear to have a transforming phenotype. Latently infected cells are normal with regards to morphology, growth rate, and viability. This notion, even in vitro, is consistent with the non-pathogenic nature of AAV.

**Site-specific integration.** Since its discovery, this unique feature by which AAV integrates into the host genome has been the driving force of AAV research. To date, many latently infected cell lines have been analyzed and data clearly shows (almost 100% efficiency) that AAV DNA integrates into a site-specific site on the human chromosome 19q13.3-qter (93, 94, 161). This pre-integration site, named AAVS1, has been cloned and a region of about 10 kb has been sequenced (94). Sequence analysis revealed no extensive homologies with AAV and suggested integration via non-homologous recombination. However, microhomologies (4-10bp) were found, including a motif acting as a Rep binding site (RBS) that matched a sequence within the viral ITR. Furthermore, eight nucleotides upstream of the RBS a motif similar to sequence found within the viral ITRs was also present, later termed the terminal resolution site (trs) (22, 172). Functional analyses later confirmed that mutations within the trs or the RBS
completely abolished any integration activity suggesting both motifs were necessary for integration to occur (210). Concurrent with advances in the understanding of the role of the AAVS1 site as the target for AAV integration, research was beginning to shed light on the role of the viral proteins involved in integration. Virally encoded Rep proteins were shown to simultaneously bind the AAVS1 RBS and the AAV ITR sequence (196). This immediately suggested a mechanism for AAV site-specific integration wherein Rep protein:AAVS1 multimerization was thought to be essential to bring both elements into proximity for recombination. In the simplest integration model, the Rep protein binds to the viral ITR and to the RBS in the AAVS1 site bringing them both together. After formation of this complex, Rep introduces a nick at the AAVS1 trs site providing a primer for Rep-mediated, unidirectional DNA replication of AAVS1. Mutational analyses of the large Rep proteins (Rep78/68 proteins) confirmed that DNA binding, helicase, and site-specific endonuclease activities were all necessary for integration at the AAVS1 locus (104, 187). Following integration, AAV proviral structures are predominately found in head-to-tail concatameric arrays, and possess microhomology at the viral/cellular junctions (28, 119, 156, 206). As noted earlier, all these processes have come from work done in transformed cultured cells and have yet to be demonstrated to occur in vivo. To date, this unique property of site-specific integration has only been documented in transformed cultured cells, and has never been demonstrated in tissues taken directly from animals. Only 2 studies (both in non-human primates) have attempted to examine experimental wild-type AAV infection (1, 77). The first study established parameters for AAV infection, but did not specifically address molecular characterization of the viral genome (1). In the second study, rhesus macaques were
inoculated with wild-type AAV in the presence or absence of wild-type adenovirus (18). AAV DNA was found most readily in peripheral blood mononuclear cells in a subset of animals. Site-specific integration into the AAVS1 locus was apparently detected in a single animal using PCR amplification and dot-blot hybridization. No attempt to confirm the site of integration or the molecular structure was reported.

RECOMBINANT ADENO-ASSOCIATED VIRUS VECTORS

Overview. In 1982, Samulski et al. showed that AAV sequences cloned into prokaryotic plasmids were infectious (158). For example, when the wild-type AAV/pBR322 plasmid is transfected into human cells in the presence of adenovirus, the AAV sequences are rescued from the plasmid and a normal AAV lytic cycle ensues. This realization not only enabled the detailed genetic analyses of the virus but also provided a substrate to generate rAAV particles. The former task was advanced by the fact that the AAV ITRs contain all cis-acting factors required in rescue, replication, and packaging as previously mentioned. Additionally, since the AAV ITRs are separated from the viral encoding regions, rAAV design is based on the total gene removal or “gutless” vector design of retrovirus-based vectors. Both systems function provided the cis-acting elements involved in gene amplification and packaging are in linkage with the heterologous “transgene” sequences of interest, whereas the viral encoding sequences necessary for virion assembly are given in trans. rAAV particles are produced by transfecting producer cells with a plasmid containing a cloned rAAV genome composed of a transgene of interest flanked by AAV ITRs and a construct expressing in trans the
viral *rep* and *cap* genes (Figure 1.4). In the presence of adenovirus helper functions, the rAAV genome is enabled to undergo the lytic phase of the wild-type AAV life cycle by being rescued from the plasmid backbone, replicated, and packaged into preformed capsids as single-stranded virions.

**Tropism modification.** As previously mentioned, AAV2 was the first serotype to be cloned into bacterial plasmids; therefore, it is the best characterized of all the serotypes and the bulk of our understanding of AAV biology (including natural tropism) has come from AAV2. Ubiquitous heparin sulfate proteoglycan (HSPG) has been identified as the primary receptor for AAV2-derived viral vectors (182). This likely explains why these vectors demonstrate a broad host range and infect a wide variety of cell types. An increasingly important area in AAV vector development involves the engineering of altered cell tropisms to narrow or broaden rAAV-mediated gene delivery. Some concerns associated with rAAV vectors may be overcome by ablating endogenous undesired tropisms and retargeting them to a specific tissue. Several of these approaches rely on modification by chemical (bi-specific conjugates) or recombinant (genetically modified capsids) means (10, 63, 197, 208, 209). As other serotypes of AAV have been discovered and are being used in vector production, it has become apparent that there are distinct and often advantageous differences among vector serotypes. For example, AAV1 and AAV7 appear to be superior to other serotypes for transduction of murine muscle tissue, but AAV5 vectors also demonstrate enhancement compared with AAV2 vectors (26, 58, 79, 147). AAV8 is the most efficient serotype for murine liver transduction (58) being up to 100-fold higher than AAV2 vectors, and AAV5 appears superior in transduction of cells in the murine respiratory tract (7, 211). AAV5 generally
Figure 1.4: Overview of the recombinant AAV transient transfection production system. The generation of the first infectious clones of AAV permitted functional dissection of the virus genome. This allowed the construction of plasmids encoding rAAV genomes in which the minimal complement of wild type sequences necessary for genome replication and packaging (i.e., the AAV ITRs) frame a gene of interest (transgene) instead of the AAV rep and cap genes. When these constructs are transfected into packaging cells together with a rep and cap expression plasmid they lead to the production of rAAV particles. Helper activities required for the activation and support of the productive phase of the AAV life cycle were originally introduced by infection of the packaging cells with wild-type Ad as depicted. Current transfection-based production methods make use of recombinant DNA encoding the helper activities instead of Ad infection. Cellular DNA polymerase activities together with the Rep78 and Rep68 proteins lead to the accumulation of replicative intermediates both in the duplex monomer (DM) and duplex dimer (DD) forms. A fraction of this de novo synthesized DNA is incorporated in the single-stranded format into preformed empty capsids most likely through the catalytic activities of the Rep52 and Rep40 proteins. The resulting infectious rAAV virions are released from the producer cells together with helper Ad particles. Sequential heat treatment and buoyant density centrifugation allows the selective elimination of the helper virus from the final rAAV preparation.
Figure 1.4: Overview of the recombinant AAV transient transfection production system.
appears to be superior to AAV2 in all tissues types tested, including central nervous system (CNS), muscle, liver, and retina (7, 26, 41, 79, 127, 147). Similarly, AAV6 vectors are more efficient than AAV2 in transducing murine airway epithelial, whereas AAV3 vectors are superior in transducing smooth muscle cells (67, 68). Overall, there is a great variability among different AAV serotypes in their ability to transduce cells, tissues, or organs depending on the route of administration and the capsid type.

**Integration of rAAV vectors.** The impetus for using AAV as a gene transfer vector was the *in vitro* data showing that it was an integrating virus. This unique property would allow infected cells the ability to pass the transgene of interest to daughter cells, and re-administration of the vector would be unnecessary. It is important to note that most rAAV vectors lack all viral genes and are therefore not considered to have the potential to undergo site-specific integration. Nevertheless, under selective pressure, rAAV vectors have been shown to integrate *in vitro* in many cell types (76, 101, 119, 126, 156). In contrast, in the absence of selection, studies have analyzed that rAAV vectors integrated at a very low frequency. In fact, cultured transformed cells transduced with rAAV under no selection, have shown a loss of gene expression over time (15, 114). It was further demonstrated that the loss of gene expression directly correlated with a loss of rAAV vector DNA, suggesting that rAAV vector DNA persisted as extra-chromosomal episomes that declined in a dividing cell population (15, 114).

A somewhat analogous situation appears to exist for plasmid DNA transfected into transformed cells in culture. There is a loss of plasmid DNA copies in cells over time due to the episomal nature of this DNA (51). However, under selective conditions, plasmid DNA has been shown to readily integrate *in vitro*, predominately in head-to-tail
arrays (198). Conversely, several in vivo studies have demonstrated that plasmid DNA persists only at very low levels following intramuscular injection in mouse muscle (102, 115).

Early experiments with rAAV vectors in vivo approximated an integration efficiency of about 0.1% - 0.5%; however, recent studies that have tried to address this question have yielded results that are somewhat inconsistent (52, 123, 174, 203). Schnepp et al. were unable to detect integrated vector DNA to a sensitivity of <0.5% of total vector DNA in mouse muscle (165). Because Rep has no role in the integration of rAAV vectors, the molecular fate of the viral DNA, once in the nucleus, must rely on host cell activities. These cellular activities, that only recently have started to be identified, might depend on the type and status of the cell, and might operate through interactions with the AAV ITRs (134). In contrast, other data suggests that the AAV ITRs enhance the persistence of vector DNA episomally, by resisting degradation, promoting the formation of circular molecules and not by integration (44, 135).

rAAV vector persistence and gene expression. Despite this lack of understanding, unlike retrovirus vectors, which must integrate to express their genes, it has been clearly shown that rAAV vectors do not require integration for gene expression (55). rAAV vectors have proven remarkably efficient for long-term gene expression in vivo, most notably in differentiated target cells, which are either slowly proliferating or post-mitotic (36, 53, 54, 91, 173, 203). Generally, these studies point to the processing of the rAAV genomes by intramolecular and intermolecular recombination to form circles and concatamers, respectively. In fact, episomal forms of vector DNA have been easily documented and characterized in vivo (42, 44, 122, 188). Clark et al. previously
observed a low-molecular weight rAAV species that was characterized as a double-stranded episome (37, 42, 44, 122, 188). Additionally, other investigators using Southern blot hybridization have observed monomeric and dimeric rAAV vector forms in different tissues (2, 135, 136, 175, 176, 188).

**Statement of hypothesis.** The work presented in this dissertation focuses on the persistence of AAV and helps to define the molecular forms present in infected tissues. Multiple studies have demonstrated that rAAV vectors are able to mediate long-term gene expression in many different tissues. Based on extensive *in vitro* cell culture data demonstrating preferential site-specific integration of wild-type AAV into the AAVS1, our beginning hypothesis was that targeted integration would also occur *in vivo*. This hypothesis will be directly addressed in chapters 2-5.
CHAPTER 2

MOLECULAR CHARACTERIZATION OF
ADENO-ASSOCIATED VIRUSES IN CHILDREN

INTRODUCTION

Infection with wild-type adeno-associated virus (AAV) is common in humans and occurs without apparent pathogenic consequence (16, 18, 19, 49). The bulk of our understanding about naturally acquired AAV infection comes from observations made over 30 years ago (17-19). However, the emergence of human gene transfer vectors based on recombinant AAV (rAAV) has rekindled an interest in the biology of wild-type AAV (157). Past epidemiologic data notwithstanding, almost nothing is known about the life cycle of wild-type AAV in a permissive host like humans. For example, we lack clear information about the portal of entry, the role of helper viruses in primary infection, cellular receptors for viral attachment, sites of primary or secondary replication, sites of latency, and in vivo molecular forms of viral DNA. A better understanding of these issues will undoubtedly influence the use of rAAV vectors for gene transfer in humans.

Recently, when viral genomes of unexpected diversity were recovered from a surprising array of tissues and organs, it became clear that AAV infection of primates (human and non-human) is much more complex than previously appreciated (56, 58). The genetic diversity observed in these studies suggested that multiple genotypes (many
more than the historically accepted 5 serotypes) of AAV circulate in humans and monkeys. Furthermore, the data implied that whatever the portal of entry, AAV can become widely disseminated following primary infection. The implications of these findings for rAAV gene transfer vectors are unclear, but the immunologic and genetic influence of prior infection on rAAV-mediated gene transfer must be considered.

To extend our understanding of wild-type AAV infection in humans, we set out to characterize AAV genomes directly out of freshly acquired human tissues. Because our assumption was that many (if not all) AAV infections begin in the respiratory tract in children as they are concurrently infected with adenoviruses, we collected tonsils and adenoids from pediatric subjects undergoing surgical excision in an outpatient surgery center. We demonstrated that 7% of these samples contained wild-type AAV DNA. In a follow-on set of experiments, we obtained a range of archived, frozen normal tissues from a repository and showed that 3% of these samples also contained wild-type AAV DNA. Analysis of the cap gene sequences from all samples revealed that most of the isolates were closely related to AAV serotype 2; a single isolate shared significant homology with serotypes 2 and 3. Interestingly, none of the isolates in our study were predicted to bind heparin sulfate, suggesting that this receptor is not necessary for wild-type infection in humans.

**MATERIALS AND METHODS**

**Human tissues.** All samples were acquired after approval from the Columbus Children’s Hospital Institutional Review Board, and where required, informed consent was obtained. Fresh human tonsil and adenoid specimens (n = 101) were collected from
children aged 2 - 13 years at Columbus Children’s Hospital undergoing elective surgical excision. Additional human tissues (n = 74) representing normal liver, spleen, muscle, heart, and lung were obtained from the Cooperative Human Tissue Network based at Columbus Children’s Hospital. The ages of the individual subjects ranged from 0 – 30 years. Subjects under 6 months of age (n = 49) were analyzed but were considered unlikely to have been exposed to adenovirus and AAV infection. In addition, for 3 of the 74 samples, the age of the subjects could not be determined. Thus, there were 22 samples from individuals aged 6 months to 30 years available for analyses, and 6 of the 22 were from individuals older than 14 years.

**Adenovirus isolation from tonsils and adenoids.** Freshly collected tissue from each subject (n = 101) was minced, mixed, and divided into 3 aliquots. Two aliquots were frozen for DNA isolation (see below) while the third was further minced and sieved through a cell strainer. The sieved material was adsorbed onto A459 cells cultured in DMEM containing 20% fetal calf serum. Cultures were maintained at 37°C in 5% CO₂ for 14 days and then blind passed into fresh A549 cells and monitored for an additional 14 days. Cultures were scored positive or negative solely on the appearance of characteristic cytopathic effects.

**DNA isolation from human tissue samples.** Freshly thawed tissue (0.2 - 0.5 g) was digested for 15 hr in 3 ml of digestion buffer (10 mM Tris, pH 8; 100 mM NaCl; 0.5% SDS; 25 mM EDTA) supplemented with 2 mg/ml proteinase K at 55°C with constant agitation. DNA was extracted twice with phenol/chloroform/isoamyl alcohol (25:24:1) and RNA removed by DNase-free RNase A digestion (20 µg/ml) (Qiagen Inc.) at 37°C for 30 min. RNase A was removed by two sequential
phenol/chloroform/isoamyl alcohol extractions. A final chloroform extraction was performed, followed by DNA ethanol precipitation using 0.3 M NaOAc (pH 5.2) and 2 volumes of ethanol. The DNA pellet was air-dried and suspended in 1 ml of 10 mM Tris, pH 8.0 and DNA concentration determined by A_{260}.

**AAV capsid PCR.** Nested PCR was used to amplify a 255 bp conserved region of the AAV cap gene. An initial round of PCR was performed on genomic DNA (100 ng) using a conserved degenerate primer set (CapSS2978: 5’-GGYGCCGAYGGAGTGYYARTKCC - 3’; and, Cap 18S: 5’-GAWKCCCCARTWGGTTGTRATGAGTC - 3’). One µl of the first round PCR served as the template for the second round using nested, degenerate PCR primers (Cap 19S: 5’-GYARTKCCTCRGGWRATTTGGCA - 3’; and, CapSS3189: 5’-GATGAGTCKYTGCCAGTCWGKGG - 3’). Reaction components for both rounds were 400 nM of each primer, 400 nM dNTP, 0.5 unit SureStart Taq polymerase (Stratagene), and 1X SureStart reaction buffer in a final volume of 25 µl. PCR cycling conditions were: 1 cycle at 94°C for 12 min; 36 cycles at 94°C for 30 sec, 52.5°C for 30 sec, and 72°C for 1 min; followed by a 4 min extension step at 72°C. To confirm PCR amplicon identity, the AAV nested capsid PCR (or Ad PCR products - see below) were resolved on 0.8% agarose gels and in-gel Southern blot hybridization was performed using AAV2 cap or adenovirus hexon sequences as the [\alpha^{-32}P]dCTP labeled hybridization probes. DNA samples identified as containing AAV sequences were subjected to further PCR to isolate the complete AAV capsid coding region. Due to the length of the cap gene product, a dual PCR approach was employed whereby the 5’ half of the cap gene was amplified as a 1.8 kb PCR product, while the 3’ half was amplified as
a 1.5 kb PCR product. Two different forward 1.8 kb primers were alternatively utilized to amplify these PCR fragments depending on which primer yielded the greatest amplification efficiency (AAV2-1.8F1: 5’ – AACATGTGCGCCGTGATTGACGGG - 3’; or, AAV2-1.8F2, 5’ - GACCGGATGTTCAAATTTGAACTC - 3’). Similarly, 2 different reverse 1.5 kb primers were alternately employed for amplification (AAVCap3’Rev, 5’ - TCGTTTCAGTTGAACCTTTGGTC TCTGCG - 3’; or, AAVCap3’RevDeg: 5’ – CARWRTTYWACTGAMACGAAT - 3’). PCR reaction conditions and primer concentrations were the same as for the 255 bp conserved capsid region. Amplified PCR products were agarose gel purified and cloned into pCR2.1-TOPO vector using TOPO TA cloning kit according to manufacturer’s instructions (Invitrogen, Inc.). DNA clones were sequenced using BigDye terminator chemistry and an ABI 727 capillary electrophoresis automatic sequencer (PE Applied BioSystems Inc.) by the CCRI Core Sequencing Laboratory.

Adenovirus hexon PCR. For detection of adenovirus sequences, total cellular DNA (100 ng) was subjected to identical PCR conditions as that used for AAV capsid PCR. A primer set targeting a conserved region (300 bp) of the hexon gene was employed (48): hex1, 5’-GCCSCARTGGKCWTACATGCACATC - 3’; and, hex2, 5’ CAGCACSCCICGRATGCACATC - 3’.

Quantitative PCR. AAV genome copy number in tissue samples was quantified using real-time TaqMan PCR analysis (ABI 7700, PE Applied BioSystems). The primers and probe set were selected following alignment of 255 bp cap DNA sequences (ForCAPSS: 5’ -AACGACAACCACCTACTTTGGC - 3’ (50 nM); RevCAPSS: 5’ - AAGTGGCAGTGGAATCTGTTCG - 3’ (900 nM); probe, [6-FAM]5’ -
CTACAGCACCCCTGGGGGTATTTTGA - 3’[6-TAMRA-FAM] (270 nM). PCR conditions were: 50°C 2 min., 95°C 10 min., 40 cycles of 95°C 15 sec. and 60°C 1 min. using 250 ng of human total cellular DNA in 1X Taqman PCR master mix.

**Sequence analyses.** The DNA and putative protein sequences were aligned and analyzed using the Clustal W method implemented in MegAlgn software in DNASTAR (DNASTAR, Inc). The phylogenetic relationship of all AAV DNA sequences and corresponding putative protein sequences were carried out using Neighbor-Joining method with Kimura 2-parameter model implemented in MEGA2 package (97). Similarity percentages between AAV2 and the new AAV sequences were determined using one-pair alignment according to the Lipman-Pearson method. Recombination analysis was performed by using the Similarity Plot method as implemented in the SimPlot software (available at [http://www.med.jhu.edu/deptmed/sray/](http://www.med.jhu.edu/deptmed/sray/))(107, 110).

**Heparin binding analysis.** To assess the ability of different AAV capsids to bind heparin, preparations of infectious AAV (harboring the capsid of interest) were subjected to iodixanol density gradient purification (215) and then applied to a POROS HE-20 heparin column (1.7 ml bed volume) using a Biocad Sprint HPLC apparatus as previously described (35). Virus was eluted using a linear salt gradient (0.1 - 1M NaCl). Flow-through, wash, and eluate (1 ml fractions) were collected for AAV DNase resistant particle (DRP) determination.

**AAV serology.** Subjects with the diagnosis of cystic fibrosis were recruited from Columbus Children’s Hospital’s Cystic Fibrosis Clinic and were eligible for the study if blood was to be drawn or a centrally placed catheter accessed for a clinical indication. Consent for participation was obtained from subjects (or their legal guardian if they were
under 18 years of age) after protocol approval from the Columbus Children’s Hospital Institutional Review Board.

To detect antibodies reactive with AAV2 capsid proteins, an enzyme-linked immunosorbent assay (ELISA) was performed. Polystyrene 96-well plates (Nunc Immuno Plate/Maxisorp Surface, Nalge Nunc International) were coated with 50 ng of viral capsid protein. AAV2 capsids were affinity purified as previously described (35) from lysates of 293 cells infected with a recombinant adenovirus type 5 carrying the AAV type 2 capsid coding sequences driven by a standard human cytomegalovirus promoter. After plates were blocked (1% normal sheep serum in PBS containing 5% dry skim milk) for 2 hours at room temperature, serum samples diluted 1:100 in PBS/5% dry skim milk were added to the test wells and incubated for 2 hours at room temperature. After washing 3 times (PBS containing 0.05% Tween), sheep anti-human IgG horseradish peroxidase-linked (Sigma Chemical Co., St. Louis, MO) was added, and the plates incubated for an additional 2 hours at room temperature. After 3 washes, o-phenylenediamine dihydrochloride (Sigma) was added to each well and developed in the dark. Optical density at 450 nm (OD$_{450}$) was recorded after 30 min (HTS 7000 Bio Assay Reader; Perkin-Elmer Corp.). The OD$_{450}$ reported was the difference between capsid and mock-coated wells. To show that the AAV capsid preparation was not contaminated with adenovirus proteins, AAV2 coated wells were also developed with adenovirus type 5 immune rabbit serum, and were shown to be at background levels. The relationship between age and OD value was estimated using the Pearson correlation.

Serum neutralization activity directed toward AAV type 2 was determined as previously described (105). Briefly, dilutions of sera were incubated at 57°C for 15
minutes to inactivate complement and were then mixed for 1 hour at 37°C with 1,000 infectious units of a rAAV2 vector that expressed *E. coli* β-galactosidase (rAAV/β-gal). The samples were applied onto monolayers of the C12 cells (38) and incubated at 37°C for 4 hours. Subsequently, the monolayers were infected with adenovirus type 5 at a multiplicity of infection of 20. At 48 hours after rAAV/β-gal transduction, cell monolayers were stained for β-galactosidase activity using the substrate 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-Gal). All serum samples were screened at a dilution of 1:10, and those found to have >90% neutralization (reduction in the number of infectious units as compared to controls) were assayed again over a range of 2-fold dilutions (1:100 to 1:1600).

**Genbank accession numbers.** The nucleotide sequences described in this chapter have been submitted to Genbank. The accession numbers are: AY695370, AY695371, AY695372, AY695373, AY695374, AY695375, AY695376, AY695377, AY695378.

**RESULTS**

**AAV in tonsils and adenoids.** It is likely that many primary AAV infections occur during childhood when the incidence of primary adenovirus infection increases with age (17, 18, 21). Moreover, the probable portal of entry for most of these primary infections is the oropharynx, which serves as the gateway to the respiratory and gastrointestinal tracts. Since lymphoid tissues in the oropharynx (tonsils and adenoids) are known targets for adenovirus replication, we set out to characterize the frequency and
complexity of AAV in freshly excised tonsils and adenoids obtained from children ages 2 to 13 years.

Using standard virus culture techniques, including blind passage of cellular lysates after 2 weeks in culture, we were unable to isolate replicating adenovirus from any of the 101 tonsil and adenoid samples. Because replicating adenoviruses were not isolated, we made no further attempt to identify AAV in these cultures. Instead, we prepared total cellular DNA from the 101 samples and then screened for the presence of AAV DNA sequences by PCR. Using a previously identified region within the AAV cap gene as our target (58), we designed and validated a nested, degenerate primer combination that readily detected the published AAV serotypes (Figure 2.1A). Control reactions were performed on human cellular DNA that was spiked with plasmid DNA representing cap genes of AAV serotypes 1 - 5. Using the optimized primer sets, we achieved a sensitivity of 15 copies in a background of 100 ng of total cellular DNA (data not shown).

With this approach, we identified AAV DNA sequences in 7 of the 101 (7%) tonsil and adenoid samples (Figure 2.1B and Table 2.1). To determine the AAV copy number in these samples, quantitative PCR was also performed on the same total cellular DNA. Copy numbers ranged from 210 – 33,000 copies/µg cellular DNA (Table 2.2).

The same DNA samples were also screened for adenovirus sequences using a PCR primer pair specific for a conserved region in the adenovirus hexon gene. This hexon-based PCR strategy has been shown to detect representative serotypes of all of the known adenovirus subgenera A-F (48). Not surprisingly, 19 tonsil and adenoid samples (19%) scored positive for adenovirus hexon sequences (Table 2.1). Analysis of the
**Figure 2.1: PCR schematic for amplification of the complete AAV capsid coding region.** (A) The diagram depicts the relative location of degenerate primers (given in Materials and Methods) used to amplify the AAV cap gene. Initially samples were screened with degenerate nested primers (cap18s, cap19s, capss3189, capss2978) to two conserved regions that flank the HVR3 coding region (gray box). To amplify the complete capsid gene, another set of nested primers were constructed (AAV2-11.8F1, AAV2-11.8F2, AAVcap3'rev, AAVcap3'revdeg) that bind to 3’ regions of rep and cap and amplify 1.8 and 1.5 kb DNA amplicons. (B) Representative amplification of the 255 bp conserved AAV sequence from human tissue DNA (100 ng) following nested PCR (see Materials and Methods for reaction conditions). Asterisks indicate the samples that are positive for AAV amplification.
Figure 2.1: PCR schematic for amplification of the complete AAV capsid coding region.
Table 2.1: Summary of AAV and adenovirus sequence detection in human tissues.

<table>
<thead>
<tr>
<th>Viral DNA Found</th>
<th>Tissues Analyzed (# of samples)</th>
<th>T+A (101)(^a)</th>
<th>Liver (19)</th>
<th>Spleen (21)</th>
<th>Muscle (15)</th>
<th>Heart (3)</th>
<th>Lung (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAV Positive</td>
<td></td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>AdV Positive (^b)</td>
<td></td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AAV &amp; AdV Positive</td>
<td></td>
<td>2(^c)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\(^a\) T+A = tonsils and adenoids.

\(^b\) AdV = adenovirus

\(^c\) Two T+A samples (T17 and T32) contained both AAV and adenovirus sequences.
Table 2.2: AAV sequence relatedness and DNA copy number in pediatric tissues.

<table>
<thead>
<tr>
<th>Sample (age)</th>
<th>Amino Acid Identity&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Nucleotide Identity&lt;sup&gt;c&lt;/sup&gt;</th>
<th>AAV copies/µg&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>T17 (2.3)</td>
<td>AAV2 (98.1%)</td>
<td>AAV2 (96.0%)</td>
<td>560</td>
</tr>
<tr>
<td>T32 (5.5)</td>
<td>AAV2 (98.1%)</td>
<td>AAV2 (96.0%)</td>
<td>210</td>
</tr>
<tr>
<td>T40 (2.9)</td>
<td>AAV2 (98.0%)</td>
<td>AAV2 (96.1%)</td>
<td>6,550</td>
</tr>
<tr>
<td>T41 (2.6)</td>
<td>AAV2 (98.0%)</td>
<td>AAV2 (96.0%)</td>
<td>590</td>
</tr>
<tr>
<td>T70 (3.2)</td>
<td>AAV2 (98.4%)</td>
<td>AAV2 (96.8%)</td>
<td>7,800</td>
</tr>
<tr>
<td>T71 (5.8)</td>
<td>AAV2 (98.0%)</td>
<td>AAV2 (96.0%)</td>
<td>7,600</td>
</tr>
<tr>
<td>T88 (4.7)</td>
<td>AAV2 (98.0%)</td>
<td>AAV2 (96.5%)</td>
<td>33,000</td>
</tr>
<tr>
<td>S17 (8)</td>
<td>AAV3 (92.7%)</td>
<td>AAV2 (90.0%)</td>
<td>200</td>
</tr>
<tr>
<td>LG15 (1)</td>
<td>AAV2 (98.4%)</td>
<td>AAV2 (96.8%)</td>
<td>80</td>
</tr>
</tbody>
</table>

<sup>a</sup> T = tonsil/adenoid; S = spleen; LG = lung. Age of the subject in years is in parentheses.

<sup>b</sup> Percentage amino acid identity to the indicated AAV serotype is given in parentheses.

<sup>c</sup> Percentage nucleotide identity to the indicated AAV serotype is given in parentheses.

<sup>d</sup> AAV DNA copy number determined using Q-PCR (see Materials and Methods). Values shown are the average of 2 separate determinations.
hexon PCR amplicon (301 bp) revealed significant identity (95% - 99%) to human adenovirus serotypes 2 and 5 for the majority (17/19) of the positive samples. Two tonsil and adenoid samples (T17 and T32) contained both AAV and adenovirus sequences.

**AAV DNA in other tissues.** To extend the findings described above, we obtained 74 frozen, archived normal tissues from individuals aged 0 – 30 years. The range of tissues and number of samples analyzed are shown in Table 2.1. Subjects under 0.5 years of age (n = 49) were analyzed but were considered unlikely to have been exposed to adenovirus and AAV. In addition, for 3 of the 74 samples, the age of the subjects could not be determined. Thus, there were 22 samples from individuals aged 0.5 to 30 years available for analysis, and 16 of the 22 were from individuals between the ages of 0.5 and 14 years. Using the same PCR scheme described for tonsils and adenoids, we found AAV DNA in 2 of the 16 (12.5%) samples from children aged 0.5 to 14 years. One was lung from a 1-year old subject and the other was spleen from an 8-year old subject. DNA copy numbers were lower in these 2 samples than in the tonsils and adenoids (Table 2.2). None of the other 72 samples contained detectable AAV DNA and none of the 74 samples contained adenovirus DNA (Table 2.1).

**AAV cap and rep gene sequences from human tissues.** DNA sequence analysis of the 255 bp amplified cap gene sequence from the 9 positive samples revealed significant homology with the corresponding region of AAV2. To isolate complete AAV capsid genes, we synthesized additional degenerate primers at the 3’ ends of the rep and cap genes (Figure 2.1A). These primers were combined with the forward and reverse conserved sequence primers to amplify 1.8 kb and 1.5 kb PCR products representing the 5’ and 3’ halves of the cap gene. Individual PCR products were cloned and sequenced; a
minimum of 4 individual clones were analyzed for each 1.8 and 1.5 PCR product (intra-clone variation was ~ 0.1%). After a single contiguous sequence was assembled for each tissue isolate, nucleotide sequence alignments revealed significant identity with AAV2 (96 - 97%) for 8 sequences. The spleen sequence (S17) was the most divergent and intermediate between serotypes 2 and 3 (Table 2.2). The complete rep coding sequences were subsequently determined for 6 of the 9 isolates (T32, T40, T70, T71, T88, and S17) and all shared > 99% amino acid identity with AAV2. Three conservative mutations in rep (for all 6 isolates) were identified (T183A, V508A, F619S).

The capsid gene amino acid translation and alignment for all 9 clones is shown in Figure 2.2. Consistent with the nucleotide sequence analysis, 8 of the amino acid sequences shared 98% identity with AAV2. Moreover, the majority of the observed amino acid substitutions (relative to AAV2) found in the 7 of the tonsil sequences and the lung sequence were conserved among the individual samples. This suggests that a specific virus isolate was circulating in the local population during the time period (winter 2002 - 2003) of tissue procurement. The majority of the observed amino acid substitutions were located in previously identified hypervariable regions (HVR) 5-7, 9, and 10 (56), all of which were predicted to be exposed on the surface of the virion. Two isolates possessed identical sequences (T41 and T71), and two others (T17 and T32) were nearly identical (2 amino acid differences). The deduced phylogenetic relationship among the 9 cap gene sequences is depicted in Figure 2.3, along with previously identified clade B and C viruses (56-58). As expected, 7 of the 8 sequences clustered closely with each other within AAV2-like clade B.
Figure 2.2: Predicted VP1 capsid amino acid alignment of AAV2 and novel human AAVs. Diagram shows sequence alignment using the CLUSTAL W program. Black boxes designate amino acid substitutions compared to the AAV2 sequence. The location of previously identified HVR regions (56) are labeled (HVR 1-12), as is an additional region (HVR2') that possesses several substitutions. Several HRV regions (5-7, 9, and 10) are colorized to facilitate visualization of these regions onto the known atomic structure of AAV2, while invariant HVRs are labeled with black boxes. Location of R585S and R588T are starred, and arrows denote approximate location of nested primers used to amplify the 255 bp HVR3 fragment.
Figure 2.2: Predicted VP1 capsid amino acid alignment of AAV2 and novel human AAVs.
The spleen isolate (S17) appeared to be intermediate between AAV2 and AAV3 and shared significant identity to the recently described clade C isolates (Figure 2.3) (57). Additional sequence analyses (Simplot) revealed two discreet AAV3-like domains (Fig. 4), suggesting that sequential recombination events between AAV2 and AAV3 serotypes generated the S17 isolate. The first AAV3-like region corresponded to amino acids 180 – 235 and overlapped HVR 2' (Figure 2.2). The second AAV3-like region encompassed amino acids 466 – 673 and covered HVRs 5 – 11. S17 possessed several surface exposed amino acid clusters in these HVRs that were identical to AAV3.

Most circulating AAV2-like viruses are not predicted to bind heparin sulfate. Careful inspection of the AAV capsid sequences described above revealed that none of the 8 AAV2-like sequences retained arginine residues at positions 585 and 588. These residues have been shown to be critical for heparin sulfate proteoglycan (HSPG) receptor binding (89, 139). In each of our sequences, R585S and R588T substitutions were observed. To further experimentally define the heparin binding capacity of AAV capsids with these substitutions, we directly measured the ability of 2 AAV2 capsids derived from tonsil and adenoids (T70 and T88) to bind heparin. Infectious AAV preparations representing T70 and T88 were generated from full-length molecular clones derived directly from the tonsil and adenoid tissue. A full description of these and other infectious clones derived from tissues is presented elsewhere (Chapter 4). DNA sequences of the clones were identical to the original cap gene nucleotide sequences generated directly from tissues. When applied to a standard heparin chromatography column (HE20-POROS) under low salt conditions, 93.5% of the total DRP applied to the column were found in the flow-through or wash buffers (Table 2.3). In contrast,
Figure 2.3: Phylogenetic analysis of VP1 capsid nucleotide sequences. A neighbor-joining program with a Kimura 2-parameter setting was used to derive phylogenetic distances based on 2,200 bp of VP1 sequence. Recently described AAV Clade nomenclature (57) was adopted and organized by vertical brackets. The human isolates identified herein are designated in teal type. Due to space restrictions, only a few representative isolates from clades A, D, and E are shown. Sequence isolates are labeled with reference to the source species (bb, baboon; ch, chimpanzee; cy, cynomolgus macaque; hu, human; rh, rhesus macaque). Clade B sequences possessing R585 and R588 amino acids and predicted to bind HSPG efficiently are labeled in red type. The scale for genetic distance is indicated in the bottom left corner.
Figure 2.3: Phylogenetic analysis of VP1 capsid nucleotide sequences.
Figure 2.4: S17 sequence homology comparison with AAV2 and AAV3. Simplot analysis of similarity percentages of S17 VP1 versus AAV 2 (red) and AAV3 (blue) are shown. Data were plotted within a sliding window of 200 bp, centered on the position plotted, with a step size between data points of 20 bp. Positions containing gaps were excluded from the comparison. The bar on the top shows predicted composition of the S17 capsid gene. The corresponding position of the HVRs are labeled as magenta boxes (HVR 2' in gray).
Figure 2.4: S17 sequence homology comparison with AAV2 and AAV3.
prototype AAV2 readily bound to the column, with on average 86% of the DRP eluting at 300 mM NaCl (35). Following chromatography, the identity of each virus isolate was further confirmed by amplifying and sequencing a 600 bp portion of viral DNA that was collected from either the flow-through or peak fractions. In each case, the predicted sequence was recovered (data not shown).

In support of the experimental observations noted above, a computer generated electrostatic potential map (Figure 2.5) of the VP3 trimer was produced using the recently described atomic structure of AAV2 VP3 (205). AAV2 VP3 monomers contain large regions of strong positive charge (Figure 2.5A, circled blue regions) grouped at the threefold axis of symmetry in the VP3 trimer (17). These regions have been implicated in HSPG co-receptor binding due to the collective involvement of 5 basic amino acids (R484, R487, K532, R585, R588) that map to this region. As a consequence of the serine and threonine substitutions in our sequences at positions 585 and 588, respectively, the model predicted a significant net reduction in overall positive charge (Figure 2.5B). Thus, the electrostatic potential mapping data was consistent with failure of AAV derived from T70 and T88 to bind to the heparin column.

We also examined the ability of the S17 isolate (AAV2/3-like) to bind heparin based on the known heparin-binding properties of prototype AAV3 (89). To accomplish this analysis, we exchanged the cap coding region in the T70 molecular clone for the S17 cap gene. Infectious particles were generated and analyzed for heparin binding as described above. As observed for T70 and T88, the S17 capsid also failed to bind to a heparin column (< 4% bound) (data not shown).
Virus | Non-bound DRP $^a$ | Bound DRP $^b$ | Percent Bound
---|---|---|---
AAV2 | $2.4 \times 10^9$ | $1.5 \times 10^{10}$ | 0.86
T70-43 | $2.0 \times 10^{10}$ | $1.0 \times 10^9$ | 0.05
T88-41 | $1.7 \times 10^{10}$ | $1.4 \times 10^9$ | 0.08

$^a$ DNase resistant particle (DRP) copy number was determined using Taqman Q-PCR. Primers/probe were homologous to a conserved cap sequence (see Materials and Methods). Non-bound DRP represent total virus found in wash and flow-through material. Data shown are the average of 2 experiments.

$^b$ DRP copy number present in pooled 1 ml gradient fractions. Data shown are the average of 2 experiments.

Table 2.3: Heparin sulfate column binding of AAV2 virus preparations.
Figure 2.5: Surface diagrams of AAV2 trimer atomic models. (A) Electrostatic surface potential of the VP3 AAV2 trimer viewed down the threefold axis (yellow triangle) calculated with GRASP (138) running from negative (red) to positive (blue). Labeled arrows indicate the positions of residues implicated in HSPG binding. (B) Predicted electrostatic surface potential of AAV2 VP3 trimer as a result of R585S and R588T substitutions. Amino acid substitutions were modeled using energy minimization simulations with Quanta (Accelrys, San Diego, CA) prior to generating the electrostatic potential map in GRASP. The surface electrostatic potential scale is the same as depicted in panel A. Highlighted regions denote predicted HSPG co-receptor engagement domains in the VP3 trimer.
Figure 2.5: Surface diagrams of AAV2 trimer atomic models.
Sequence variation in the AAV capsid maps to surface exposed regions.

Given the recently described atomic ribbon structure of AAV2 (25), we were able to map the observed amino acid substitutions in the 8 AAV2-like sequences onto the capsid surface (Figures 2.6A and 2.6B). The majority of amino acid substitutions were located in areas of the capsid predicted to be surface exposed and previously identified as hypervariable regions (HVR 5-7, 9, 10) (Fig. 3). In contrast, regions predicted to encode core β-barrel domains responsible for structural integrity were almost invariant.

Similarly, the more divergent S17 capsid amino acid sequence was modeled onto the same atomic structure (Figures 2.6C and 2.6D). The majority of the S17 amino acid substitutions also mapped to surface exposed HVR regions 5-7, 9, and 10. Interestingly, several HVR 10 substitutions (purple shaded region) were located near the central pore complex that is thought to directly interact with the viral encapsidation complex.

Because of the predilection for surface substitutions, we analyzed synonymous (s) vs. non-synonymous (ns) nucleotide substitutions in the cap and rep gene coding regions for all 9 sequences (Table 2.4). The s/ns ratios were calculated for nucleotides corresponding to (i) surface exposed HVR, (ii) non-HVR, and (iii) approximately 600 bp of the 3’ end of the rep gene. Statistically significant differences in the s/ns ratios were observed when we compared the non-HVR cap region ratio (12.1 average s/ns) to either the HVR (1.6 average s/ns) or rep (2.8 average s/ns) ratios (p = 0.000002 and 0.0001, respectively). These data indicated a strong bias for conservative synonymous mutations within the β-barrel domains in non-HVR regions. This conclusion was consistent with scanning mutagenesis experiments that demonstrated severe functional constraints on particle assembly and stability in these capsid core regions (148, 169, 197). We also
Figure 2.6: Ribbon diagrams of atomic models of AAV2 VP3 trimers showing the location of predicted amino acid substitutions in the human AAV isolates. (A) Ribbon drawing viewed down the three-fold axis of symmetry of the AAV2 VP3 trimer. \( C_\alpha \)-backbones for the 3 VP3 monomers are rendered as teal ribbons. Predicted locations of the observed amino acid substitutions present within the 8 AAV2-like sequences are color coded to reflect HVR location (HVR 5-7, 9, and 10) within the primary sequence (Fig. 2). White space-filling amino acid substitutions mapped outside the known HVRs. (B) Side view of the predicted location of the observed amino acid substitution demonstrating surface display (right side of Panel). (C) Superimposition of observed S17 amino acid substitutions relative to the AAV2 VP3 trimer atomic structure viewed down the 3-fold axis. (D) Side view of the predicted location of the observed amino acid substitutions in isolate S17 (surface display oriented on right side of Panel). Images were generated in NAMD/VMD (UIUC Theoretical Biophysics Group) and rendered using Raster3D.
Figure 2.6: Ribbon diagrams of atomic models of AAV2 VP3 trimers showing the location of predicted amino acid substitutions in the human AAV isolates.
Table 2.4: Summary of synonymous and non-synonymous nucleotide substitutions in rep and cap.

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Subst. Cap HVR (400 bp)</th>
<th>Subst. Cap non-HVR (1800 bp)</th>
<th>Subst. 3' rep (600 bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>s</td>
<td>ns</td>
<td>s/ns</td>
</tr>
<tr>
<td>LG15</td>
<td>20</td>
<td>11</td>
<td>1.8</td>
</tr>
<tr>
<td>S17</td>
<td>48</td>
<td>58</td>
<td>0.8</td>
</tr>
<tr>
<td>T17</td>
<td>17</td>
<td>10</td>
<td>1.7</td>
</tr>
<tr>
<td>T32</td>
<td>17</td>
<td>10</td>
<td>1.7</td>
</tr>
<tr>
<td>T40</td>
<td>12</td>
<td>10</td>
<td>1.2</td>
</tr>
<tr>
<td>T41</td>
<td>15</td>
<td>8</td>
<td>1.9</td>
</tr>
<tr>
<td>T70</td>
<td>10</td>
<td>6</td>
<td>1.7</td>
</tr>
<tr>
<td>T71</td>
<td>15</td>
<td>8</td>
<td>1.9</td>
</tr>
<tr>
<td>T88</td>
<td>18</td>
<td>9</td>
<td>2.0</td>
</tr>
<tr>
<td>Avg ± SD</td>
<td>1.6±0.27</td>
<td>8.2±3.9</td>
<td>12.1±2.8</td>
</tr>
</tbody>
</table>

*a s is the number of observed synonymous substitutions in the indicated coding region.
*b ns is the number of observed non-synonymous substitutions in the indicated coding region.
*c s/ns is the calculated ratio.
*d Value represents the number of observed nucleotide substitutions per 100 bp of sequence.
*e Substitution rate in HVR region was greater than that observed for non-HVR and 3’ rep region. p = 0.04 and 0.02, respectively, using paired t-test.
*f Ds/Da ratio in non-HVR region was greater than that observed for HVR and 3’ rep region. p = 0.00002 and 0.0001, respectively, using paired t-test.
observed a significantly greater nucleotide substitution rate in the HVR (average of 8.2 substitutions per 100 nucleotides) relative to the non-HVR (average of 3.5 substitutions per 100 nucleotides) and rep coding regions (average of 1.8 substitutions per 100 nucleotides) (p=0.04 and 0.02, respectively). These rates likely reflect the inherent flexibility for amino acid substitutions that each region possesses in that surface exposed loop domains (that overlap HVR domains) are known to tolerate multiple amino substitutions and remain infectious, while much of the non-HVR regions and rep coding regions are functionally constrained.

**Seroreactivity to AAV2.** To further characterize AAV infection in children, we would have liked to evaluate the AAV2 serologic status of our 175 subjects. Unfortunately, we were unable to collect or acquire serum from these individuals. Instead, we screened a separate cohort of 68 individuals (ages 3 – 39) drawn from a local cystic fibrosis clinic. Sera were assayed both for binding antibodies (ELISA) to the AAV2 capsid and for the ability to neutralize AAV2 in vitro (Figure 2.7). If one restricts attention to those individuals aged 14 and younger (same ages as our tonsil and adenoid cohort), the seropositivity rate was 11.5% (3 of 26). Interestingly, 1 of the 3 positive samples had low titer (1:200) *in vitro* neutralizing activity but did not register as positive of AAV DNA found in tissues outside the oropharynx. The rates estimated from DNA detection would represent a minimum based on 2 assumptions: (i) not every infected individual would harbor AAV in the oropharynx or other organs; and (ii) there was a reasonable chance of sampling error.

After the age of 14 years, the number of seropositive samples rose dramatically to 76% (32 of 42). The correlation with age was significant (r = 0.464, p <0.001 by Pearson
Figure 2.7: Serology of AAV infection as a function of age. OD_{450} values from a standard ELISA (see Materials and Methods) are plotted versus the age of the subject. Sera were tested at a 1:100 dilution. OD values below 0.2 (thin solid line) were considered negative. The same sera were tested for neutralization activity against AAV2 (see Materials and Methods). Data points that are circled represent samples that had neutralization titers > 1:100.
Figure 2.7: Serology of AAV infection as a function of age.
correlation). Only 24% (10 of 42) of the sera from those older than 14 years possessed significant in vitro neutralizing activity. Considering all ages, only 16% (11 of 68) of the samples had significant (titer > 1:100) in vitro neutralizing activity. There were 6 samples from those over 14 years that had significant ELISA activity (O.D. > 0.4) that did not mediate significant in vitro neutralization. Such ELISA reactivity probably represented antibodies that bound to the capsid but did not neutralize the virus.

**DISCUSSION**

Renewed interest in the biology of wild-type AAV infection prompted us to pursue prospective molecular characterization of AAV infection in a cohort of children undergoing elective surgical excision of tonsils and adenoids during the winter season. Although we were unable to recover infectious adenovirus using standard in vitro culture techniques, we did identify AAV DNA sequences in 7% (7/101) of the samples analyzed. We also found adenovirus DNA in 19% of the samples, and in 2 cases, AAV and adenovirus DNA were present in the same tissue. One might speculate that these latter 2 samples were recovered from individuals recently infected with AAV and adenovirus, although there are clearly other possibilities to explain these observations.

**Molecular epidemiology of AAV infection in children.** Our discovery of AAV2 DNA in tonsils and adenoids from children was not unexpected. AAV2 appears to be the most prevalent of the human AAV serotypes (17, 18, 79, 202), and importantly, our cohort was temporally (winter) and geographically (central Ohio) restricted. Moreover, the portal of entry for AAV would be expected to parallel that of adenovirus.
Thus, the oropharynx would be a fertile field from which to harvest AAV (and adenovirus).

While the DNA sequences recovered from these children were similar to the prototype AAV2 and to each other, 2 significant observations emerged from the sequence analyses. First, the majority of the observed amino acid substitutions were located in previously defined hypervariable regions (56) predicted to be exposed on the surface of the virion (Figures 2.6A and 2.6B). This observation also extended to the more distantly related S17 (AAV2/3) capsid sequence (Figures 2.6C and 2.6D). These data, considered together with the observed nucleotide substitution patterns, suggest that the HVR domains are structurally flexible and possess the capacity to evolve, perhaps in response to host immune pressure. The fact that rep gene sequences from the same tissues were nearly invariant further supported the idea of functional constraints. The identification of the S17 AAV2/3 chimeric sequence confirmed earlier published observations (57), and suggested another mechanism (intemolecular recombination) involved in AAV evolution.

Second, and perhaps most intriguing, was the discovery that none of the 7 AAV2-like sequences derived from tonsils and adenoids were predicted to bind HSPG. This was also true for the AAV2-like sequence recovered from lung tissue. To confirm that other compensatory mutations that would restore HSPG binding in the capsid had not occurred, we generated infectious molecular clones from 2 of the tissue samples and demonstrated they indeed did not bind HSPG (Table 2.3). To extend this observation, we examined other available AAV2-like sequences (57) and noted that 77% of the 31 currently identified clade B AAV2-like sequences lack R585 and R588. Thus, these data suggest
that preponderance of AAV2-like isolates do not bind HSPG, and that HSPG is not required for natural infection in humans. This notion is consistent with the fact that most other serotypes of AAV do not bind HSPG.

**Seroepidemiology of AAV2 infection in children.** To see if the rates of AAV infection as judged by DNA isolation (7 – 12%) were consistent with serological estimates, we analyzed a set of sera from children in the same geographic locale. Ideally, we would have also analyzed sera from the cohort of children who donated tissue samples, but we were unable to obtain serum from these children. In those children 14 years and younger (same ages as our tonsil and adenoid cohort), the seropositivity rate was roughly 12% and thereby confirmed our estimates derived from AAV DNA detection in tissues.

**AAV disseminates beyond the portal of entry.** Previously published work has described the presence of AAV DNA in multiple adult human tissues (57). We have now extended these findings to children with the discovery of AAV DNA in tissues within and beyond the probable portal of entry (oronasal). In children ages 0.5 to 14 years, we found AAV DNA in 2 (lung and spleen) of the 16 non-oropharyngeal tissue samples (12.5%) available for analysis. Although the number of samples analyzed was smaller, the percentage containing AAV DNA compared favorably (12.5% vs 18%) with data from adults (57).

To extend beyond the portal entry, infectious agents generally use one of 3 pathways: direct (contiguous) spread, lymphatic, or bloodstream. In our samples, AAV could have easily reached the lung by contiguous spread from the oropharynx, either with or without adenovirus. To reach the spleen, however, the route of viral spread was
almost certainly hematogenous, in the form of free or cell-associated (perhaps leukocytes) virus.

**Biology of wild-type AAV infection.** The now emerging picture is that AAV infection of humans is more complex than previously appreciated. It has been known for decades that AAV is not associated with disease or pathology, and that infection incidence in humans generally parallels that of adenovirus. Not surprisingly, infectious AAV has been isolated from sites where adenovirus is traditionally recovered, including the gastrointestinal tract (1). More recently, it has been appreciated that AAV DNA can be found in many human tissues including liver, muscle, lymph nodes, leukocytes, kidney, and cervical tissues (56, 57, 66, 70, 183, 193), and now tonsils/adenoids.

Considered together, these data suggest the following biologic scenario. AAV most likely enters the body through the oropharynx in association with adenovirus. Replication ensues (with adenovirus help) and new AAV particles are formed and released from infected cells in the oropharynx. Secondary rounds of replication in newly infected cells follow, again creating new waves of AAV particles. Such rounds of replication would continue until the host immune system responds and blunts the infectious process. By the time replication is controlled, AAV has had the opportunity to spread to the lungs (contiguous) and through the bloodstream to more distant sites. This scenario does not address the rare AAV5 serotype, which apparently enters the body through the genital tract (9, 62). However, there remains only a single isolate of AAV5, and recent studies of human tissues have failed to find AAV5-like DNA.

While the events envisioned above are entirely plausible, many important questions regarding the *in vivo* biology of natural AAV infection remain unanswered.
For example, while AAV and adenovirus appear to be linked early in the infectious process, there appears to be an unlinking sometime during and following dissemination, allowing AAV DNA (but not adenovirus) to persist in organs and tissues outside the oropharynx. It is formally possible that AAV DNA persists by integrating in target cells on chromosome 19 (AAVS1), but this has not been demonstrated in vivo. Moreover, the target cells for AAV replication and persistence have not been identified, nor have the specific cellular receptors for viral attachment been defined.

It should be remembered that even with widespread dissemination, AAV has not been found to cause disease or pathology. Nonetheless, the effect of prior (or subsequent) wild-type AAV infection on gene transfer mediated by rAAV vectors is unknown, and a more thorough understanding of the natural AAV infectious process is needed.
CHAPTER 3

CHARACTERIZATION OF ADENO-ASSOCIATED VIRUS GENOMES ISOLATED FROM HUMAN TISSUES

INTRODUCTION

Adeno-associated viruses (AAV) are ubiquitous, non-cytopathic, replication incompetent members of the Parvoviridae family. AAV replication requires the presence of a helper virus, and this is usually one of the many serotypes of adenovirus. The epidemiology of AAV infection in humans was extensively studied after its initial description some 40 years ago (5, 6, 17, 82, 153). Two major conclusions were drawn from this work. First, many adults have antibodies reactive against one or more AAV serotypes (31, 49, 194), a finding which is entirely consistent with early and repeated exposures to AAV and adenoviruses throughout life. Second, even with this level of exposure, AAV does not cause any disease or other pathologic condition in humans.

As noted above, AAV genomes are replicated and packaged into new infectious particles only in the presence of a helper virus. In the absence of helper virus, AAV is unique among viruses in its ability to direct site-specific integration of its genome into a specific locus (AAVS1) on human chromosome 19 (94, 96, 161). A similar locus has also been identified in non-human primates, and recently in rodents (47). Site-specificity is mediated by virally encoded rep gene products via recognition and binding to similar
viral and cellular sequences. Such sequence-specific interaction ultimately results in the insertion of head-to-tail proviral AAV DNA arrays that are characterized by rearrangement of viral inverted terminal repeat and flanking cellular sequences (103, 133, 149, 206). The AAV DNA is harbored in this “latent” state until subsequent infection with a helper virus causes reactivation or "rescue" of the AAV genome, thereby resulting in renewed viral replication and production of infectious particles.

To date, this unique property of site-specific integration has only been documented in transformed cultured cells, and has never been demonstrated in tissues taken directly from humans. In fact, surprisingly little is known about the molecular events of AAV infection in vivo, either in humans or in permissive animal models. Only 2 studies have attempted to examine experimental wild-type AAV infection in non-human primates (1, 77). The first study established parameters for AAV infection and did not specifically address molecular characterization of the viral genome (1). In the second study, rhesus macaques were inoculated with wild-type AAV in the presence or absence of wild-type adenovirus (18). AAV DNA was found most readily in peripheral blood mononuclear cells in a subset of animals. Site-specific integration into the AAVS1 locus was apparently detected in a single animal using PCR amplification and dot-blot hybridization. No attempt to confirm the site of integration or the molecular structure was reported.

Because of the interest in recombinant DNA delivery vectors based on wild-type AAV, a more thorough understanding of the basic biology of wild-type AAV infection seems warranted. To that end, we have recently analyzed tissues from 175 children for the presence of AAV and adenoviral DNA (Chapter 2). In that study, we found 9
samples that harbored AAV sequences. Seven of the 9 were from tonsil-adenoid tissues removed at surgery, while the remaining 2 positives (lung, spleen) were normal samples retrieved from a tissue bank. In the present study, we undertook evaluation of the molecular forms of the AAV DNA in these tissues using 3 distinct techniques. In contrast to the prevailing hypothesis, we were unable to identify wild-type AAV integration at the AAVS1 locus using a sensitive S1-specific PCR assay. When we broadened our search for viral-host sequence junctions using a modified linear amplification-mediated PCR assay (LAM-PCR), we found a single integration event on chromosome 1, but again were unable to find any AAVS1 insertions. Finally, using a novel linear rolling circle amplification assay, we showed that the majority of the AAV genome sequences in these tissues were extra-chromosomal circles.

**MATERIALS AND METHODS**

**Cell propagation.** HeLa cells and Detroit 6 cells were maintained in DMEM supplemented with 10% fetal bovine serum (FBS) and penicillin and streptomycin. 293 human embryonic kidney cells were grown in EMEM supplemented with 10% FBS and penicillin and streptomycin. HeLa and 293 cells were purchased from the American Type Culture Collection (Rockville, MD) and Detroit 6 cells were obtained from R. Jude Samulski.

**Human tissues and DNA isolation.** All tissue samples were acquired after approval from the Columbus Children’s Hospital Institutional Review Board, and where required, informed consent was obtained. Fresh human tonsil and adenoid specimens (n = 101) were collected from children aged 2 - 13 years at Columbus Children’s Hospital.
undergoing elective surgical excision. Additional human tissues (n = 74) representing normal liver, spleen, muscle, heart, and lung were obtained from the Cooperative Human Tissue Network at Columbus Children’s Hospital. The ages of the individual subjects ranged from 0 – 30 years. Subjects under 6 months of age (n = 49) were analyzed but were considered unlikely to have been exposed to adenovirus and AAV infection. In addition, for 3 of the 74 samples, the age of the subjects could not be determined. Thus, there were 22 samples from individuals aged 6 months to 30 years available for analyses, and 6 of the 22 were from individuals older than 14 years. Samples were stored frozen at -80° C until DNA isolation.

Freshly thawed tissue (0.2 - 0.5 g) was digested for 15 hr in 3 ml of digestion buffer (10 mM Tris, pH 8.0; 100 mM NaCl; 0.5% SDS; 25 mM EDTA) supplemented with 2 mg/ml proteinase K at 55° C with constant agitation. DNA was subjected to three phenol/chloroform/isoamyl alcohol (25:24:1) extractions. A final chloroform extraction was performed, followed by DNA ethanol precipitation and resuspension in 10 mM Tris, pH 8.0.

**Quantitative PCR and DNA hybridization.** AAV genome copy number in tissue samples was quantified using real-time TaqMan PCR analysis (ABI 7700, PE Applied BioSystems). The primers and probe set were selected following alignment of 255 bp cap DNA sequences (ForCAPSS (bp 3004-3024): 5’ - AACGACAACCACACTACTTTGGC - 3’ (50 nM); RevCAPSS (bp 3074-3054): 5’ - AAGTGGCAGTGGGAATCTGTTG - 3’ (900 nM); probe (bp 3024-3050), [6-FAM]5’ - CTACAGCACCCCCCTGGGGGTATTTTGA - 3’[6-TAMRA-FAM] (270 nM). PCR
conditions were: 50°C 2 min, 95°C 10 min, 40 cycles of 95°C 15 s and 60°C 1 min using 250 ng of human total cellular DNA in 1X Taqman PCR master mix.

For hybridization analyses, DNA was fractionated on 0.8% agarose gels, dehydrated and subjected to in-gel hybridization or transferred to a nylon membrane using a vacuum manifold dot blot apparatus. DNA hybridization conditions were 65°C for 16 hr in buffer containing 6X SSC, 1X Denhardt's reagent, and 200 μg/ml sonicated herring sperm DNA. Rehydrated gels and nylon membranes were washed twice at 60°C in 2X SSC, 0.2% SDS for 30 min, and then twice at 60°C in 0.2X SSC, 0.2% SDS for 30 min.

**AAVS1-PCR.** To detect AAV integration in the AAVS1 locus, nested PCR was performed using 1 μg of total cellular DNA as template with primers specific for the human AAVS1 locus and AAV rep or cap genes (84, 186). The reaction conditions for first round PCR were as follows: 25 pmol of each primer, 200 μM dNTPs, 5 U Herculase Hotstart DNA polymerase (Stratagene) with Herculase reaction buffer. The AAV-specific primers used in first round PCR were as follows: *cap*; CAPGSP1 (bp 4320-4349), 5’-GTCTGTTAATGTGGACTTTACTGTG GACAC-3’, or *rep* (bp 479-455); REPGSP1, 5’-CAGGGGTGCCTGCTCAATCAGATTC-3’. The AAVS1-specific primer was AAVS1-1R, 5’-ATGGCTCCAGGAAATGGGGGTGTG-3’ (186). The PCR cycling conditions were as follows: denaturation at 95°C for 90 s, followed by 30 cycles of 95°C for 30 s, 55°C for 30 s, 72°C for 6 min, with a final extension of 72°C for 10 min. For nested PCR, 1 μl of the first round PCR reaction was used as template with the same reaction conditions as noted for the first round PCR using the following primers: AAV-specific primers, *cap* (bp 4357-4379); CAPGSP2, 5’-
GTGTATTTCAGAGCCTCGCCCCAT-3’, rep (bp 404-427); REPSP2, 5’-
TCCCCATTCTTTCGGCCACCCAG-3’, and AAVS1-specific primer AAVS1-2R; 5’-
CACCAGATAAGGAATCTGCC-3’ (186). To verify the integrity of the AAVS1 locus
and our ability to amplify AAVS1 sequences, we performed control PCR reactions using
specific primers spanning 2.2 kb of the AAVS1 locus that included the Rep binding site
(RBS). The control primers were: AAVS1-F4, 5’GATTTCCTGCCTGCGT3’ and
AAVS1-1R (above). PCR products were cloned into a TOPO-TA cloning vector
(Invitrogen Corp.) and sequenced using an ABI 727 capillary electrophoresis automatic
sequencer (PE Applied BioSystems) by the CCRI Sequencing Core Laboratory.

**LAM-PCR.** Linear amplification-mediated (LAM)-PCR was used to isolate
sequences contiguous with AAV genomes in samples of total cellular DNA representing
a variety of human tissues. The AAV sequence was first amplified from 1 µg of genomic
DNA using with a single, biotinylated primer corresponding to either the 5’ region of rep
(bp 545-519); REPBIO, 5’[BioTEG] CTCCGGGGCCTTACTCACRCGGCGCCA-3’, or
the 3’ region of cap (bp 4257-4286); CAPBIO, 5’-
[BioTEG]GCTGCAGAARGARAACAGCAAACGCTGGAA-3’. For the internal
human erythropoietin (epo) control, the HepoBIO primer was used: 5’-
[BioTEG]TGGTTTCAGTTCTTGTCAATGAGGTTG-3’. The reaction conditions were
as follows: 0.25 pmol of primer, 200 µM dNTPs, 5 U Herculase Hotstart DNA
polymerase (Stratagene Corp.) with the Herculase reaction buffer, and overlayed with
mineral oil. The cycling conditions were as follows: denaturation at 95°C for 90 s,
followed by 50 cycles of 95°C for 30 s, 65°C for 30 s, 72°C for 6 min, with a final
extension of 72°C for 10 min. This cycle was repeated with the addition of 5 U of fresh
Herculase Hotstart DNA polymerase for a total of 100 cycles of linear amplification. The biotinylated PCR products were captured using 200 µg of streptavidin-coated, magnetic beads (Dynabeads kilobase BINDER kit, Dynal Inc.) in conjunction with a magnetic particle concentrator (Dynal) according to the manufacturer’s protocol. After binding, the beads were washed twice with 100 µl of 10 mM Tris-HCl (pH 8.0) and incubated with 2 U of Klenow (Roche), 300 µM dNTPs (Invitrogen Corp.) and random hexanucleotide mix (Roche Applied Science) in a 20 µl reaction volume for 1 hr at 37°C to create double-stranded DNA. Following incubation, the DNA/beads were washed twice as described above and incubated for 2 hr at 37°C with 5 U of a blunt-cutting restriction endonuclease (EcoR V, Pvu II, or Stu I: New England Biolabs) and 5 U of a restriction enzyme that generates a 3’ overhang end and cuts once within the AAV genome (Sal I). After restriction enzyme digestion, the beads were washed as above, and a blunt-end, double-stranded DNA adaptor (100 pmol of the GenomeWalker Adaptor, Universal GenomeWalker Kit, Clontech) was ligated to the restricted DNA in a 10 µl volume containing 20 U of T4 DNA ligase and the accompanying NEB ligation buffer (New England Biolabs) at 16°C for 16 hr. The DNA/beads were washed as described above and used as template in a nested PCR reaction. The first round of PCR used 25 pmol of each primer, 200 µM dNTPs, 5 U of Herculase Hotstart DNA polymerase (Stratagene) with the Herculase reaction buffer using the same cycling conditions as above with the exception that only 30 cycles were performed. The primer specific to the adaptor (AP1, 5’-GTAATACGACTCACTATAGGGC-3’) was supplied in the Universal Genome Walker Kit. The AAV-specific primers used in first round PCR were as follows: cap, CAPGSP1, 5’-GTCTGTTAATGTGGACTTTACTGTGGACAC-3’; or rep,
REPGSP1, 5’-CAGGGGTGCCTGCTCAATCAGATTC-3’. The epo control primer used was HepoGSP1, 5’-GACCCCCATGAGAGCCAGAGGCCAG-3’. For nested PCR, 1 µl of first round PCR was used as the template with the same reaction conditions as for first round PCR with the following primers: Adaptor primer - AP2, 5’-ACTATAGGGGCACGCGTGGT-3’; cap, CAPGSP2, 5’-GTGTATTCAGAGCCTCGCCCCAT-3’; rep, REPSP2, 5’-TCCCATTCCTTCTCGGCC ACCCAG-3’; epo, HepoGSP2, 5’-CATCCTGTCTTCATGGGTCCCAC-3’. PCR products were cloned into TOPO-TA cloning vector (Invitrogen Corp.) for sequence analysis.

**Linear rolling circle amplification (LRCA).** To prepare genomic DNA samples for LRCA, 1 µg of genomic DNA was first digested in a 15 µl volume with a restriction enzyme that does not cut within the AAV genome (Spe I for human tissue samples; EcoRV for Detroit 6 and plasmid spike samples). The resulting linearized DNA was digested by incubation with 10 U of Plasmid-Safe ATP-Dependent DNase (Epicentre Technologies) in a 25 µl final volume for 16 hr at 37°C in 33 mM Tris (pH 7.8), 66 mM potassium acetate, 10 mM magnesium acetate, 0.5 mM DTT, and 2 mM ATP. The nuclease was then heat inactivated for 30 min at 70°C. The resulting template DNA (2.5 µl, equivalent to 100 ng) was mixed in a final volume of 15 µl in 10 mM Tris HCl, pH 8.0, with 300 pmol each of two AAV cap-specific primers: temp1 (bp 2882-2898), 5’-ATTGGCATTGCGATTCC-3’; temp2 (bp 2926-2910), 5’-TGGTGATGACTCTGTCG-3’. Each primer contained a phosphothioate linkage between the last 3’ bases to increase primer stability. The reaction was heated to 95°C for 3 min, then cooled slowly to 4°C to allow primer annealing, and mixed with 15 µl of phi29 DNA phage polymerase.
(TempliPhi, Amersham Biosciences) reaction buffer containing phi29 polymerase, and incubated for 18 hr at 30°C. Amplified products were heat inactivated at 65°C for 10 min and then digested with EcoR I for subsequent dot blot hybridization.

RESULTS

Identification of AAV sequences in human tissues. As part of a study of the molecular epidemiology of AAV infection in children (Chapter 2), we tested 175 tissues for the presence of AAV DNA by PCR. From tonsil-adenoid tissues, we identified 7 AAV positive samples (designated as T17, T32, T40, T41, T70, T71, and T88). Samples T17 and T32 contained both AAV and adenovirus sequences. Two (lung and spleen, LG15 and S17, respectively) of 74 additional tissues were also AAV positive. From these 9 AAV positive samples, entire AAV capsid gene sequences were generated and analyzed. Sequence analysis revealed that 8 of the 9 isolates were closely related to AAV2, while the other isolate shared significant homology to the recently described AAV2/3 hybrid clade (57).

Molecular characterization of AAV DNA in tissues. Having identified AAV DNA sequences within these human tissues, we turned our attention to characterization of the molecular form of the DNA within the cell. Prevailing wisdom suggested that AAV DNA would be found integrated within the AAVS1 locus on human chromosome 19 (96, 161). As a first step, we determined (using quantitative PCR) the AAV genome copy number in tissues; values ranged from 80 - 33,000 copies/µg of total DNA. Given these values, it was not surprising that attempts to characterize AAV DNA within total cellular DNA by standard restriction enzyme digestion and Southern blot analysis were
unsuccessful (data not shown). This led us to develop alternative assays to delineate the predominant viral form(s) present within these human tissues.

**Analysis of integration into the AAVS1 locus.** To directly examine AAV DNA insertion into the AAVS1 locus, we developed a set of nested PCR primers to amplify AAV/AAVS1 viral-cellular junctions. For a positive control, we used the well-characterized Detroit 6 cell line in which head-to-tail tandem integration of AAV2 into the AAVS1 locus had been previously described (Figure 3.1A) (93). Quantitative PCR revealed that Detroit 6 cells contained 5 viral genome copies per cell, and this value was used to define the sensitivity of the assay. As seen in Figure 3.1B, we readily detected a single cell equivalent (5 AAV copies) as evidenced by the amplification of the expected 1.5 kb PCR product. Importantly, identical assays of the tissue DNA samples failed to yield a PCR product (Figure 3.1C). DNA integrity was validated using a control AAVS1 PCR amplification and all samples amplified the expected 2.2 kb fragment (data not shown).

**LAM-PCR analysis of AAV integration.** Our failure to detect AAVS1 integration by direct PCR led us to consider the possibility that AAV genomes were either (i) randomly integrated elsewhere into the host genome, or, (ii) not integrated. To address these possibilities, we adapted a linear amplification mediated PCR (LAM-PCR) method that was previously used to map retroviral vector integration sites (163). The assay (outlined in Figure 3.2) used a single AAV-specific, biotinylated primer to amplify linear fragments (100 cycles) that extended into unknown flanking sequences. The linear DNA fragments were captured on streptavidin beads, converted to double-stranded DNA, and digested with restriction enzymes to create a blunt DNA end in the flanking
Figure 3.1: AAVS1 PCR on human samples. (A) Restriction map schematic of the human AAVS1 locus. The Rep binding site (RBS) and Detroit 6 (Det6) integration site are designated by vertical arrows. The location of the AAVS1-specific primers used in the nested PCR reactions are also shown (solid arrowheads). AAVS1 locus numbering follows the convention of Kotin et al. (93), whereby the Eco RI site defining base pair “0” is denoted (positive integers to the right, negative integers to the left). (B) Southern blot hybridization analysis of AAVS1-PCR products using Detroit 6 cellular DNA. Decreasing copies of Detroit 6 genomes (100,000 to 1) were spiked into 1 µg of naïve human genomic DNA and subjected to nested AAVS1 PCR. The PCR products were fractionated on an agarose gel and Southern hybridization was performed using an AAVS1-specific probe. The expected 1.5 kb AAV-AAVS1 junction fragment is readily detected by nested PCR using a single Detroit 6 cell junction equivalent. (C) Southern blot AAVS1-PCR on human clinical samples. AAVS1-PCR was performed on 1 µg of genomic DNA isolated from various human clinical samples. Southern blot hybridization using an AAVS1-specific probe failed to detect any signal (except in Detroit 6 control DNA).
Figure 3.1: AAVS1 PCR on human samples.
**Figure 3.2: Schematic of linear amplification mediated PCR (LAM-PCR) to detect random AAV integrants.** To isolate DNA sequences flanking integrated wild-type AAV genomes, linear PCR (100 cycles) was performed on total cellular DNA (1 µg) using an AAV-specific biotinylated primer homologous to a conserved region in the cap or rep genes (cap is shown). PCR products were captured on streptavidin beads and converted to double-stranded DNA by random hexanucleotide priming. The double-stranded DNA was then digested with a blunt-cutting restriction enzyme (Eco RV, Pvu II, or Stu I) to generate a substrate for ligation of a blunt linker (vertical box) to the end of the digested DNA. To reduce the level of competing internal AAV-AAV junctions, double stranded fragments were also digested with Sal I, which created an incompatible end for ligation of the blunt adaptor. The resulting DNA fragments were subjected to 2 rounds of PCR using AAV and linker-specific primers. PCR products were analyzed by Southern hybridization and cloned into an appropriate PCR vector for subsequent analysis.
Figure 3.2: Schematic of linear amplification mediated PCR (LAM-PCR) to detect random AAV integrants.
sequence. A blunt ended linker was then ligated to the free end that generated an anchoring point for nested PCR primers. Nested PCR was performed using primers specific for AAV and linker sequences, and products were cloned and sequenced. To validate the LAM-PCR assay, we mapped the known insertion site of AAV2 in Detroit 6 cells (Figure 3.1A) using a variety of blunt cutting restriction enzymes (Stu I, Pvu II, and EcoR V). Regardless of the blunt cutting restriction enzyme employed, clones possessing identical viral/cellular junctions were obtained, which matched the published insertion site (93). LAM-PCR sensitivity was determined by preparing 10-fold serial dilutions of Detroit 6 DNA (ranging from 1.6 to 160,000 copies of AAV-cellular junctions) in a background of 1 µg of genomic DNA isolated from a naïve human tonsil. We were able to amplify as few as 160 AAV-cellular junctions using either a cap (Figure 3.3) or rep biotinylated primer (data not shown). As an internal control for DNA integrity and purity, a human erythropoietin (epo) gene specific LAM-PCR primer was designed and yielded the expected 0.5 kb epo LAM-PCR product (when used with the linker specific primer) in all DNA samples assayed (Fig. 3.3, bottom panel).

LAM-PCR was performed on the 9 AAV positive human samples using both cap and rep primer sets, with an array of blunt cutting restriction enzymes (Stu I, Pvu II, and EcoR V) to maximize the chances of obtaining a blunt-cutting restriction site in close proximity to an integration site. Representative Southern blots of the resulting AAV specific LAM-PCR products (cap primer) are shown in Fig 3.4A. Multiple hybridizing PCR products of various sizes (0.5 – 5.5 kb) were present in tissues T70, T71, and T88. As expected, the Detroit 6 template DNA also yielded fragments of the predicted size based on the specific blunt cutting restriction enzyme employed (see Figure 3.1A for
**Figure 3.3: LAM-PCR validation and sensitivity using Detroit 6 cells.** Southern blot hybridization was performed on LAM-PCR products to determine assay sensitivity. Naïve human total cellular DNA (1 μg), containing various spiked copies of Detroit 6 cell DNA, was used as the template for LAM-PCR using an AAV cap-specific primer and a ligated blunt linker. Positive Southern hybridization (using an AAVS1 probe) detected amplification of as few as 160 Detroit 6 viral/cellular junctions. Shown in the bottom panel is an ethidium bromide stained gel of a human erythropoietin (epo) LAM-PCR genomic fragment that was generated using the same DNA templates used for the AAV LAM-PCR analysis. All samples amplified the expected 500 bp epo genomic fragment following LAM-PCR with an epo gene specific primer (see Materials and Methods) confirming the DNA integrity of the sample and the reproducibility of the LAM-PCR assay.
Figure 3.3: LAM-PCR validation and sensitivity using Detroit 6 cells.
Figure 3.4: LAM-PCR of positive human clinical samples. (A) Southern blot of LAM-PCR using AAV-cap primers. LAM-PCR was performed using 1 µg of total cellular DNA with either AAV rep or cap primers (cap is shown). The top panel shows LAM-PCR using EcoRV as the blunt-cutting restriction enzyme, while the middle panel shows LAM-PCR using Pvu II. The DNA was hybridized with an AAV-specific probe. Detroit 6 DNA (D6) was used as a positive control and gave the predicted size based on the restriction analysis of the AAVS1 locus (see Figure 1A). Positive hybridization of various sizes can clearly be seen in samples T70, T71, and T88. Analysis of the LAM-PCR products revealed that an approximate 5.5 kb band in sample T71 contained an AAV-cellular junction (starred). The bottom panel shows an ethidium bromide stained gel of the LAM-PCR products using a human epo gene specific primer as an internal control. All sample amplified the expected 500 bp genomic fragment, confirming the DNA integrity of the samples and the reproducibility of the LAM-PCR assay. (B) Schematic of AAV-cellular junction from tissue T71. The 3’ AAV sequence is represented on the left ranging from bp 4357 (the position of the cap LAM primer) to the junction at bp 4604 in the AAV ITR. The AAV sequence in the LAM-PCR product contained the 3’ cap sequence as well as a complete D, A, and C ITR region. There appears to be only 2 bp of homology between the breakpoint of the AAV ITR and the chromosomal sequence (dotted underline). The LAM-PCR product contained approximately 5 kb of human chromosomal DNA that mapped to position 1q31.1, composed primarily of a (GGAAT)$_n$ repeat sequence (solid underline).
Figure 3.4: LAM-PCR of positive human clinical samples.
restriction map). Moreover, the internal human epo control LAM-PCR primer amplified the expected DNA fragment, verifying DNA template quality (Figure 3.4A).

We successfully isolated 44 AAV specific LAM-PCR products from 6 of the 9 positive human samples (T17, T32, T40, T70, T71, and T88), and subsequently cloned and sequenced the fragments. From these clones, a single AAV integration event was documented within a 5.5 kb Pvu II clone from tonsil sample T71 (Figure 3.4). This clone contained 250 bp of AAV DNA followed by approximately 5 kb of chromosome 1 sequence that ended with the expected LAM-PCR blunt linker. The cellular sequence mapped to position 1q31.1 and is composed primarily of a (GGAAT)$_n$ pentameric repeat sequence (Figure 3.4B). The viral junction breakpoint consisted of a single 3’ AAV ITR that contained a complete D, A and C region, with the remainder of the ITR deleted (Figure 3.4B). There were two base pairs (AA) of microhomology between the breakpoint of the AAV ITR and the cellular sequence. The remaining LAM-PCR products (n = 43) consisted of either partial AAV sequence or AAV-AAV junctions separated by varying amounts of intervening viral ITR sequence (summarized in Table 3.1). The fine structure of several AAV-AAV junction clones (isolated from T70) are shown in Figure 3.5, the majority of which contain large ITR deletions. However, several clones (T70-11, T70-12, T70-C1) appeared to contain a more fully intact ITR sequence, characterized by the presence of a double-D sequence structure (204). We were unable to obtain complete DNA sequence beyond the D regions, which was likely due to the extreme secondary structure present within the ITR cruciform.

**LRCA amplification of episomal DNA.** Given that 2 complementary methods failed to show integration of AAV DNA into the S1 locus, and that integration elsewhere
Three of 9 positive samples (T41, S17, LG15) did not give rise to any detectable LAM-PCR products. LAM-PCR products contained AAV sequence corresponding to a portion of a single AAV genome.

Table 3.1: Summary of LAM-PCR positive clones isolated from clinical samples.

<table>
<thead>
<tr>
<th>Tissue</th>
<th>AAV-Cellular Junction</th>
<th>AAV Fragment</th>
<th>Head-to-Tail Junction</th>
<th>Head-to-Head Junction</th>
<th>Tail-to-Tail Junction</th>
<th>Total Isolates</th>
</tr>
</thead>
<tbody>
<tr>
<td>T17</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>T32</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>T40</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>T70</td>
<td>0</td>
<td>5</td>
<td>12</td>
<td>1</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>T71</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>T88</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>23</td>
<td>15</td>
<td>2</td>
<td>3</td>
<td>44</td>
</tr>
</tbody>
</table>

* Three of 9 positive samples (T41, S17, LG15) did not give rise to any detectable LAM-PCR products. 

* LAM-PCR products contained AAV sequence corresponding to a portion of a single AAV genome.
Figure 3.5: Schematic of AAV-AAV junctions isolated following LAM-PCR. A complete head-to-tail AAV ITR junction is shown at the top of the figure, along with the position of the cap and rep primers used. Shown below are various AAV-AAV junctions isolated from tissue T70. The majority of junctions analyzed were in a head-to-tail orientation; however, tail-to-tail (T70-14) and head-to-head (T70-E9) orientations were observed. The breakpoints for each junction are given, with deleted sequence designated using a solid line. For several clones (T70-C1, T70-11, T70-12), we were unable to obtain readable sequence after the D region from either side (designated as dotted lines), which was presumably due to strong secondary structure associated with the ITR structure.
Figure 3.5: Schematic of AAV-AAV junctions isolated following LAM-PCR.
seemed rare, we were left with the conclusion that the majority of the AAV DNA in the human tissues must be extra-chromosomal. To directly address this possibility, we developed a novel assay using linear rolling circle amplification (LRCA) to specifically amplify double-stranded, circular AAV genomes from total cellular DNA. A schematic of the approach is shown in Figure 3.6. Total cellular DNA was initially digested with a restriction enzyme that was not predicted to cut within the AAV genome. This generated linear DNA fragments that were substrates for degradation by digestion with a novel exonuclease (Plasmid-Safe DNase) that selectively removes double-stranded and single-stranded linear DNA, as well as single-stranded circular DNA. Importantly, Plasmid-Safe DNase does not degrade double-stranded, circular DNA molecules, thus effectively enriching the DNA sample for free circular AAV genomes while degrading integrated AAV forms. We previously demonstrated that this exonuclease is able to efficiently degrade single-strand AAV genomes and all detectable replicative forms (37, 165). After digestion, remaining intact circular AAV episomes were amplified by isothermic rolling circle amplification using phi29 phage DNA polymerase. AAV specific primers (cap gene) were used to yield AAV-specific DNA consistent with high molecular weight head-to-tail amplicons.

T88 DNA (100 ng) was subjected to LRCA, which resulted in the synthesis of high molecular weight concatameric AAV forms that upon digestion with a one-cut restriction enzyme generated the unit-length form (Figure 3.6). Control LRCA reactions using either linear or circular plasmid DNA spiked into naïve cellular DNA confirmed assay specificity (Figure 3.7). These results indicated that episomal wild-type AAV forms were present within sample T88. A minor hybridizing band with faster mobility
Figure 3.6: Linear rolling circle amplification (LRCA) for the detection of AAV episomes. Total cellular DNA was digested with a restriction enzyme that does not cut within the AAV genome. The DNA was then treated with Plasmid-Safe DNase which degrades linear fragments, but leaves circular, double-stranded DNA intact. The digestion reaction served as a template for linear rolling circle amplification (LRCA) using AAV-specific primers and phi29 phage DNA polymerase. Large, linear concatameric arrays (U, uncut LRCA DNA from tissue T88) were produced following LRCA of circular AAV episomes. The linear arrays were subsequently digested into unit-length monomers by restriction enzyme digestion with an enzyme that cleaves the AAV genome once (labeled as a "1" in figure). The unit-length fragment was then cloned into an appropriate vector for further sequence analysis.
Figure 3.6: Linear rolling circle amplification (LRCA) for the detection of AAV episomes.
Figure 3.7: Wild-type AAV genomes are present predominantly as episomes. To determine the sensitivity of the LRCA assay, 10-fold dilutions of circular (C) or linear (L) DNA plasmids containing the AAV2 rep and cap genes were spiked into 100 ng of genomic DNA from a naïve human tonsil sample. LRCA was performed using cap-specific primers and the products were analyzed by dot blot Southern hybridization using an AAV-specific probe. Positive hybridization could be detected at 1,000 circular plasmid copy spike (representing episomal DNA) whereas linear plasmid DNA (representing integrated AAV) could not be amplified to any detectable level. Total cellular DNA (100 ng) from AAV-positive clinical samples (S) was then analyzed by LRCA, and tissues T70, T71, and T88, showed positive amplification. The AAV genome copy numbers in these particular tissues were at or above the level of sensitivity of LRCA as determined from the plasmid spike experiments. Detroit 6 (Det6) cells, which have 83,000 integrated AAV genomes in 100 ng, failed to amplify the integrated AAV, further confirming LRCA assay specificity.
Figure 3.7: Wild-type AAV genomes are present predominantly as episomes.

<table>
<thead>
<tr>
<th>Plasmid Standard</th>
<th>Clinical Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copies/100 ng</td>
<td>C</td>
</tr>
<tr>
<td>$10^6$</td>
<td>56</td>
</tr>
<tr>
<td>$10^5$</td>
<td>21</td>
</tr>
<tr>
<td>$10^4$</td>
<td>655</td>
</tr>
<tr>
<td>$10^3$</td>
<td>59</td>
</tr>
<tr>
<td>$10^2$</td>
<td>780</td>
</tr>
<tr>
<td>$10^1$</td>
<td>760</td>
</tr>
<tr>
<td>$10^0$</td>
<td>3,300</td>
</tr>
<tr>
<td></td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>8.3x10$^4$</td>
</tr>
</tbody>
</table>
was also observed (Figure 3.6), which did not correspond in size to the predicted head-to-head or tail-to-tail orientation. This smaller DNA form was probably derived from deleted episomal forms, consistent with the AAV-AAV junctions isolated by LAM-PCR (Fig. 3.5). Moreover, 5 additional samples (T32, T40, T70, T71, and S17) also contained varying amounts of LRCA amplifiable AAV DNA, which indicated the presence of similar episomal forms in these tissues (data not shown). The 3 samples that did not amplify (T17, T41, and LG15) tended to have a lower copy number of AAV DNA (Figure 3.7) or lower quality of the total genomic DNA preparation.

**AAV genomes persist predominantly as episomes in human tissues.** The ability to amplify circular AAV DNA led us to consider whether the LRCA assay could be used to estimate the amount of AAV DNA that existed as extra-chromosomal circles. As noted above, the LRCA assay was validated by subjecting plasmid DNA (in supercoiled or linear form) to the LRCA assay using ten-fold serial dilutions of a 15 kb plasmid spiked into 100 ng of human genomic DNA from a naïve tonsil (Figure 3.7, “Plasmid Standard”). Amplified DNA was observed by Southern dot-blot hybridization using as few as 1,000 circular input copies, while linear plasmid DNA failed to yield amplified DNA at any input (up to $10^6$ copies), thereby confirming LRCA specificity. Furthermore, for Detroit 6 cells, which have 83,300 integrated AAV genomes in 100 ng, LRCA failed to amplify the integrated AAV genome, providing further confirmation of assay specificity.

We analyzed the AAV-positive human samples for the presence of episomal AAV genomes using LRCA in combination with Southern dot-blot hybridization to visualize the amplified product (Figure 3.7, “Clinical Samples”). Samples T70 and T71
demonstrated amplification at an intensity on the Southern blot similar to the 1,000 circular copy plasmid standard (Figure 3.7). Based on the initial AAV genome copy number in these samples (T70 had 780 AAV copies/100 ng and T71 had 760 AAV copies/100 ng), these data suggested that the majority of the T70 and T71 AAV DNA existed in a circular conformation. Likewise, tissue sample T88 (3,300 AAV copies/100 ng) showed DNA amplification at a similar level as that detected for the 10,000 plasmid copy standard, which is also consistent with the majority of T88 AAV DNA also being in a circular conformation. The lower amplification levels observed in the remaining samples (T17, T32, and S17) did not permit an estimation of episomal prevalence using the Southern dot-blot methodology.

**DISCUSSION**

The goal of this study was to define the molecular form of AAV genomes that we found in human tissue samples. Based on extensive in vitro cell culture data demonstrating preferential site-specific integration of wild-type AAV into the AAVS1 (117, 133), our belief was that targeted integration would also occur in vivo. However, we unable to show site-specific integration of AAV DNA using 2 complementary PCR-based assays. In addition, we went on to show that the major form of AAV DNA in these tissues was extra-chromosomal and circular.

Although the precise mechanism of targeted AAV genome integration remains unknown, detailed knowledge regarding cis and trans requirements for integration are well documented. The sole viral cis elements necessary for AAVS1 integration are the 145-bp ITR, although enhanced integration has been observed with the inclusion of a
portion of the p5 promoter (142). Integration is dependent on the large Rep proteins and involves dual recognition in cis of rep binding sites (RBS) within the viral ITR and the AAVS1 locus. Mutational analysis of Rep78/68 proteins indicated that DNA binding, helicase, and site-specific endonuclease activities were all necessary for AAVS1 integration (104, 187). Both the AAVS1 RBS and trs sequence elements have also been shown to be essential for viral targeting to this locus via Rep78/68-dependent replication from the AAVS1 origin of replication (210). Mechanistically, Rep protein multimerization is thought to be essential to bring both physical elements into proximity for recombination via limited DNA replication at the AAVS1 origin. Following integration, wild-type AAV proviral structures are predominately found in head-to-tail concatameric arrays, and possess microhomology at the viral/cellular junctions (28, 119, 156, 206). As noted earlier, this process has been well documented in transformed cultured cells, and may occur in vivo. In a study by Hernandez et al. (77), 9 rhesus macaques were inoculated with AAV in the presence or absence of wild-type adenovirus. Site-specific integration was apparently detected in 1 animal by PCR amplification of wt AAV-cellular DNA junctions via dot-blot hybridization. Recently, site-specific integration has been observed in an AAVS1 transgenic mouse model (149) using a recombinant adenovirus vector harboring a AAV ITR containing vector. Under conditions of inducible AAV Rep protein expression, AAV vector genome integration into AAVS1 was clearly documented with this hybrid vector. This study reaffirms the requirement for Rep protein expression and AAV ITR elements for targeted integration, but whether this occurs at similar frequencies in vivo following wt AAV infection remains to be determined in this experimental model.
At the outset, we chose to use the well-characterized latently infected Detroit 6 cell line to validate and optimize an AAVS1 integration assay. In these cells, we detected single AAV- AAVS1 integration events in Detroit 6 genomic DNA spiked into a competing background of naïve total cellular DNA. When applied to our AAV-positive human tissues, we were unable to detect AAV-AAVS1 cellular junctions in any of the 9 samples. Failure to observe AAVS1 targeted integration was unexpected, but similar results had been observed in vivo using an experimental model of AAV infection in non-human primates (77).

The absence of targeted integration did not preclude the possibility of random integration into the host genome. Several cellular pathways exist that could facilitate random integration of exogenous DNA into the host cell, including either homologous or illegitimate recombination via host cellular DNA repair enzymes (121, 200). In fact, for many common modalities of in vivo exogenous DNA transfer (i.e. naked plasmid DNA and adenovirus vectors) random integration of introduced DNA has been shown to occur (73, 140, 195, 199). Moreover, recent data suggested that the frequency of random integration is related to the amount of intranuclear exogenous input DNA (195). Increased levels of random integration of plasmid DNA into the host genome were observed following highly efficient in vivo electroporation of muscle tissue, which increased intracellular levels of plasmid DNA by 6 - 34 fold compared to standard DNA injection (195). Interestingly, random integration of rAAV vectors in cultured cells has been shown to preferentially occur at sites of double-strand chromosomal DNA breaks (DSB), possibly via interaction between viral ITR and cellular non-homologous end-joining (NHEJ) repair enzymes at regions possessing microhomology (125, 177).
Whether wild-type virus in vivo also integrates at such sites is unknown, but it is notable that the NHEJ cellular enzyme Ku86 appears capable of binding to AAV ITR in a rep-independent manner in cultured cells (212).

To directly address the question of random integration, we developed a modified LAM-PCR integration assay to identify and amplify random insertion sites. The LAM-PCR technique was previously exploited to map retroviral vector insertion sites (163). We modified the assay to minimize detection of AAV-AAV head-to-tail junctions, which are characteristically formed following wild-type infection. By simultaneously using two different restriction enzymes that yield either overhanging or blunt ends, we were able to achieve a detection sensitivity of 1 in 160 viral-cellular integration events, again using control Detroit 6 cellular DNA. The specificity of the assay was confirmed by our ability to map the identical AAV2 insertion site in Detroit 6 cells as that previously published (14, 93). Using this method, we were able to isolate a single AAV integration event in sample T71 that mapped to the q arm of chromosome 1 (1q31.1). The sequence consisted of a highly repetitive pentameric sequence (GGAAT) characteristic of satellite II/III repetitive DNA. Similar purine tract sequences have been associated with illegitimate integration sites observed in vitro (200). Also, this integrant possessed 2 base pairs of microhomology with the cellular breakpoint sequence, which was similar to the microhomologies observed in cell culture at sites of rAAV vector integration (124).

The fact that we detected a single, cell-specific (i.e., non-clonal) integration event encompassing 5 kb of flanking cellular DNA and possessing the expected blunt restriction site and 3' linker sequence supported the notion that LAM-PCR is both robust and sensitive. In addition, our ability to amplify a control gene (epo) from all the samples
suggested that the assay is reproducible and has sufficient sensitivity to detect a
significant proportion of potential in vivo integration events.

Although integration events were rare, we readily identified LAM-PCR products
that contained either AAV DNA fragments or consisted of AAV-AAV junction elements
oriented predominately in a head-to-tail manner (Table 3.1). These forms likely arose by
one of several mechanisms. First, several of the AAV-AAV junction clones were
isolated because the blunt cutting enzyme that was predicted to not cut within the AAV
genome actually was present and therefore yielded the expected internal AAV-AAV
junction product. Second, not all of the LAM-PCR products contained the expected blunt
restriction site located just 5' to the ligated linker. Hexamer primer annealing at the end
of the single-stranded amplicon during conversion to a double-stranded template might
have created such random blunt ends. An alternative possibility is that double-stranded
DNA breaks could have occurred prior to cleavage with the blunt cutting restriction
enzyme and would therefore be substrates for linker ligation. This underscored the
importance of using multiple blunt cutting enzymes in independent LAM-PCR reactions
to increase the probability of obtaining viral-cellular LAM-PCR junction products.
Regardless of the underlying cause, we identified multiple AAV-AAV junction
sequences that were predominately in a head-to-tail orientation. Sequence analysis of the
products revealed structures with deleted sequence within the viral ITR and flanking rep
and cap sequences, which are very similar to ITR rearrangements observed in vitro
following wild-type AAV infection (84, 206).

These data caused us to consider an alternative possibility, namely, that wild-type
AAV DNA persisted in vivo as extra-chromosomal elements. To explore this possibility,
we adopted a technique known as linear rolling circle amplification (LRCA) to specifically amplify circular \textit{in vivo} AAV structures. In previous work, we had characterized the activity of a novel DNA exonuclease (Plasmid Safe DNase) that readily digests double and single-stranded linear molecules, but not double-stranded circular episomes (37, 165). With this activity in mind, we used this exonuclease to enrich the AAV positive samples for episomal AAV forms, while also removing greater than 99% of the total cellular DNA (linear) by digestion with a "no-cut" (does not cleave the AAV genome) restriction enzyme. We then employed an isothermal phage polymerase dependent amplification process (originally commercialized to amplify plasmid DNA for high-throughput sequencing projects) to amplify circular templates via rolling circle replication. The assay was validated using linear or circular plasmid DNA spiked into total cellular DNA. Amplification ($10^5$-fold) of low copy circular plasmid inputs (1,000 copies) was easily detected by dot-blot hybridization, while linear DNA ($10^6$ copies) failed to amplify.

Considered together, these data suggest that AAV genomes rarely integrate into the host genome, but instead, persist as free circular forms. While we cannot rule out the possibility that the circular form(s) detected are integration intermediates, we believe this is unlikely for 2 reasons. First, most of our AAV positive samples (7 of 9) did not contain detectable adenovirus sequences, which is consistent with a “latent” AAV infection. Although it is formally possible that other viruses (e.g., herpesviruses) could have been present as helpers, AAV has never been (to our knowledge) isolated along with herpesviruses from clinical samples. Second, it seems unlikely that for all the
samples assayed, failure to detect targeted integration was due to the preferential isolation of pre-integration "intermediates" at the complete exclusion of integrated forms.

Persistence of wild-type AAV genomes as circular episomes is entirely plausible. Complete double-D ITR structures (similar to those detected by our LAM-PCR assay; Fig. 3.5) contain sufficient cis-sequence information for the complete AAV life cycle (204). The double-D ITR structure, as part of a circular plasmid, is able to undergo resolution (in a rep-independent process) in vitro to yield a linear no-end substrate. In the presence of helper virus co-infection, this no-end linear substrate undergoes a productive linear replication cycle after rep-dependent nicking at the trs site. Thus, a double-D circularized AAV form could efficiently undergo productive lytic replication if provided with complete helper functions. Interestingly, an alternative AAV replication pathway involving circular duplex monomer genomes (cAAV) was recently identified (131, 132). These circular species appear to constitute up to 10% of monomer duplex intermediates of wild-type AAV. The circularization point of cAAV also appears to yield the double-D ITR element. This cAAV form was shown to either replicate along the standard linear strand-displacement pathway following resolution of the double-D ITR domain or by a mechanism that was able to maintain the integrity of the circular conformation. These data, combined with the propensity of recombinant derivatives to assume circular monomeric or multimeric structures (36, 42, 44, 207), suggest that circularization in vivo may be the preferred result of interaction between cellular double-strand DNA repair enzymes and the incoming AAV genome. Whether in vivo circularization occurs via a single-strand or double-strand intermediate is currently unknown, but recent in vitro data suggests circularization likely occurs via a double-strand intermediate (33).
In conclusion, we have shown that following naturally acquired infection, AAV DNA can persist mainly as circular episomes in human tissues. These findings are consistent with the circular episomal forms of rAAV vectors that have been isolated and characterized from *in vivo* transduced tissues (43, 44). Since the wild-type and vector genomes share only the AAV ITR, it is tempting to conclude that the ITR is the predominant determinant of genome persistence *in vivo*, and that recombination events drive circle formation in both cases. While we were unable to demonstrate integration into the AAVS1 locus, more human samples should be studied to confirm these findings before we presume that site-specific integration directed by AAV is restricted to cells in culture. Moreover, it is now time to go beyond the straightforward identification and characterization of AAV DNA sequences in tissues, and move on to understanding which cell types harbor these persistent AAV genomes.
CHAPTER 4

INFECTIOUS MOLECULAR CLONES OF ADENO-ASSOCIATED VIRUS
ISOLATED FROM HUMANS

INTRODUCTION

Since its initial description over 40 years ago (82), adeno-associated virus (AAV) has emerged as a unique member of the parvovirus family. AAV is replication incompetent and requires a helper virus (e.g., adenovirus) to produce new infectious virions (5, 23, 85, 133, 150, 151). However, the precise role of helper viruses in the AAV replicative cycle remains incompletely defined. In the absence of helper virus, the hallmark of AAV infection in cultured cells is site-specific integration of the viral genome into a specific locus (AAVS1) on human chromosome 19 (95, 96, 162). Virally encoded gene products, through the recognition and binding of similar viral and cellular sequences, mediate this unique site-specificity. Following integration, the AAVS1 locus is characterized by integrated head-to-tail proviral AAV DNA arrays that contain rearranged viral inverted terminal repeat and flanking cellular sequences (103, 133, 206). Helper virus infection of cells harboring integrated AAV DNA results in rescue of the AAV genome leading to the production of new infectious particles. Importantly, to date, this unique property of site-specific integration has only been documented in transformed cultured cells, and has never been demonstrated in tissues taken directly from humans.
While virtually ubiquitous in nature, AAV has never been associated with any disease or pathologic condition in humans (16, 18, 19, 49). Once this fact was realized over 30 years ago, studies pursuing the in vivo biology of AAV became less frequent. Thus, many observations made about AAV biology in cultured cells have been assumed to carry over to natural infection, but have never been investigated or proven. The continued preclinical and clinical development of recombinant DNA delivery vectors based on wild-type AAV has served to highlight and emphasize our lack of knowledge about the biology of wild-type AAV infection in humans.

To extend our understanding of naturally-occurring AAV infection, we recently analyzed tissues from 175 children for the presence of AAV DNA. Most of the genomes were closely related to the prototype AAV serotype 2, but one isolate shared significant homology with serotypes 2 and 3 (Chapter 2). Interestingly, none of the isolates were predicted to bind heparin sulfate, suggesting that this receptor is not necessary for natural infection in humans. Detailed analyses of the molecular forms of the AAV DNA in these tissues revealed that the vast majority of the genome sequences in these tissues were extra-chromosomal episomes (Chapter 3). In contrast to the prevailing hypothesis (derived from in vitro studies), we were unable to identify any integrates at the human AAVS1 locus.

In the present study, we developed a novel method (sequence-specific linear rolling circle amplification, SSLRCA) that allowed us to amplify extra-chromosomal AAV genomes directly from tissue DNA. The resulting circular monomeric AAV genomes are shown to be biologically active when molecular clones are directly transfected into HeLa cells. Furthermore, the resulting clones gave rise to infectious
virus as judged by replicating AAV DNA in P1 cells. We were able to isolate infectious molecular clones from 5 of the 6 SSLRCA positive tissues. Based on ITR sequence analysis, we concluded that infectious molecular clones contained a complete double-D ITR element. Conversely, limited ITR sequence analysis of non-infectious clones suggested that deletion of internal ITR regions was responsible for the lack of infectivity since these clones expressed the expected rep and cap proteins.

MATERIALS AND METHODS

Cells and viruses. HeLa cells were purchased from the American Type Culture Collection (Rockville, Md) and maintained in DMEM supplemented with 10% fetal bovine serum (FBS) and penicillin and streptomycin. Wild-type adenovirus type 5 (Ad5) was grown and titered as described previously (Chapter 3). Ad5 at a multiplicity of infection of 20 was used in all assays involving helper virus.

Manipulation of nucleic acids. Low molecular weight DNA was isolated from cells using standard procedures as previously described (Chapter 3). For Southern blot hybridization analysis, DNA was fractionated on 0.8% agarose gels and transferred to a nylon membrane. DNA hybridization conditions were 65°C for 16 hr in a buffer containing 6X SSC, 1X Denhardt's reagent, and 200 ug/ml sonicated herring sperm DNA. Membranes were washed twice at 60°C in 2X SSC, 0.2% SDS for 30 min, and then twice at 60°C in 0.2X SSC, 0.2% SDS for 30 min.

Linear rolling circle amplification (LRCA). To detect the presence of double-stranded circular AAV genomes, 1 ug of genomic DNA was first digested in a 15 ul volume with a restriction enzyme that does not cut within the AAV genome (Spe I for
human samples; Eco RV for Detroit 6 and plasmid spike samples). Linear DNA (both double- and single-stranded) was degraded by incubation with 10 U Plasmid-Safe ATP-Dependent DNase (PS-DNase, Epicentre Technologies) in a 25 ul volume for 16 hr at 37°C in 33 mM Tris (pH 7.8), 66 mM potassium acetate, 10 mM magnesium acetate, 0.5 mM DTT, and 1 mM ATP. PS-DNase was heat inactivated for 30 min at 70°C, and 2.5 ul (equivalent to 100 ng) was used as a template for LRCA using phage phi29 DNA polymerase (TempliPhi, Amersham Biosciences). Template DNA was mixed in a final volume of 15 ul in 10 mM Tris, pH 8.0, with 300 pmol each of two AAV cap primers (temp1, 5’-ATTGGCATGGCAGATTCC-3’; temp2, 5’-TGGTGATGACTCTGTCG-3’) that contained a phosphothioation linkage between the last two 3’ nucleotides. This mixture was heated to 95°C for 3 min, cooled slowly to 4°C to allow primer annealing, and mixed with 15 ul of TempliPhi reaction buffer containing phi29 polymerase and incubated for 18 hr at 30°C. Amplified products were heat inactivated at 65°C for 10 min and then either digested with Xba I or Hind III for Southern hybridization. For cloning, LRCA amplified DNA was digested with a single-cut enzyme (Xba I or Hind III) and the unit-length DNA fragment cloned into plasmid pBlueScript KS- (Stratagene Corp.). DNA sequencing was performed by the CCRI Sequencing Core Laboratory. In addition, manual DNA sequencing of the AAV ITR region was accomplished using the SequiTherm EXCEL II DNA Sequencing kit (Epicentre) following the manufacturer’s protocol for isothermal sequencing with α-labeled 35S dATP.

**Analysis of LRCA amplified ITR products.** To examine the fidelity of the LRCA assay, circular templates were generated that contained defined ITR-ITR structures, and then subjected to the LRCA reaction. Products were examined by
sequencing and Smith-Birnstiel end-labeling of the linear amplified products (170).

Briefly, a complete ITR-ITR end-joined template was generated by digesting a pSub201 based rAAV vector plasmid (rAAV/CMV/eGFP) with Pvu II to release the intact rAAV expression cassette (including ITRs). This fragment was self-ligated and then treated with PS-DNase to remove all linear forms. A portion of this circular template (diluted 1:100) was used as template in a LRCA reaction as described above. Both self-ligated and LRCA products were linearized by Sal I enzyme digestion, treated with alkaline phosphatase, and end-labeled using polynucleotide kinase and γ\(^{32}\)P-ATP. After labeling, the products were digested with Nde I to create linear, end-labeled fragments. Selected molecular clones isolated from tonsil sample T88 also underwent LRCA and end-labeling in a similar manner as described above, but with the following minor modifications. T88 clones were released from the plasmid vector backbone by Xba I digestion prior to self-ligation and treatment with PS-DNase, and then linearized with Sna BI before treatment with alkaline phosphatase and polynucleotide kinase. Dra III was used to generate linear end-labeled fragments. To determine ITR-ITR junction heterogeneity occurring \textit{in vivo}, LRCA products were treated identically as described above for T88 derived molecular clones. DNA fragments were fractionated on a 6% denaturing polyacrylamide gel and exposed directly to film. Specific band counts were determined using a Typhoon 9400 Variable Mode Imager (Amersham Biosciences) with the ImageQuant 5.2 software.

**Functional analysis of AAV molecular clones.** To determine if LRCA generated AAV molecular clones could produce infectious virions, the pBlueScript vector backbone and cap ORF was re-created by digestion with Xba I and self-ligation to form head-to-tail circular monomeric genomes. The ligated products (1 ug) were directly
transfected into $4 \times 10^5$ HeLa cells using FuGene 6 reagent (Roche Applied Science), followed by infection with Ad5. After maximum cytopathic effect development (48 hr), the transfected cells (P0) were harvested and low molecular weight DNA prepared from half of the cells. The recovered DNA was subsequently treated with Dpn I to digest input plasmid DNA. A clarified cell lysate was generated from the other half of P0 cells by 3 freeze/thaw cycles and a 60 min incubation at 56°C to heat inactivate Ad5. The clarified lysate (1:10 dilution) was used to infect $4 \times 10^5$ HeLa cells (along with Ad5). After 48 hr, the transduced cells (P1) were harvested and low molecular weight DNA isolated.

**Western blot analysis.** To detect the presence of Rep and Cap proteins, a cell lysate was generated from transfected cells (P0) after 48 hr and electrophoresed on a 10% SDS-PAGE and subsequently transferred to a PVDF membrane (GE Healthcare). The membrane was blocked for 10 min at 25°C using TBST (10mM Tris, pH 8.0, 150 mM NaCl, 0.1% Tween20) plus 5% nonfat dry milk. Primary antibodies (from American Research Products, Inc.) were used at a 1:100 dilution (Rep proteins, 1:1 mixture of anti-AAV Rep Clone 303.9 and Clone 226.7; Cap proteins, anti-AAV Cap Clone B1) and hybridized in blocking solution for 1 hr. The blot was washed 3 times for 10 min in TBST and incubated for 30 min in TBST + 2.5% dry milk with an HRP-conjugated antimouse secondary antibody (Vector Laboratories) at a 1:10,000 dilution. The blot was washed as above and developed using the ECL plus detection system (GE Healthcare).

**Real-time PCR.** Wild-type AAV genomes and AAV plasmids were quantified by real-time PCR using the ABI Prism 7000 Sequence Detection System in a standard 25 ul TaqMan PCR reaction with the following primers and probe: ForCAPSS, 5’-AAGCACAACCACACTTTTGGC-3’, 50 nM final concentration; RevCAPSS, 5’-
AAGTGGCAGTGGAATCTGTTG-3’, 900 nM final concentration; CAPSSProbe, 5’-
[6-FAM]CTACAGCACCCTGGGGGTATTTTGA[TAMRA~6-FAM]-3’, 200 nM
final concentration.

RESULTS

Sequence-specific linear rolling circle amplification (SSLRCA). In our
previous work on characterization of AAV in human tissues, we found that the majority
of AAV genomes were extra-chromosomal and not integrated in the AAVS1 locus
(Chapter 3). As part of those efforts, we developed a novel method (sequence-specific
linear rolling circle amplification, SSLRCA) that allowed us to amplify extra-
chromosomal AAV genomes directly from tissue DNA. It occurred to us that these
amplified products might represent intact, biologically active AAV genomes. To
investigate this hypothesis, templates for SSLRCA were derived from total tissue DNA
that was treated with a unique exonuclease (Plasmid-Safe DNase) which preferentially
degraded linear DNA fragments, leaving circular, double-stranded DNA intact (Figure
4.1A). This treatment enriched the substrate DNA for circular templates and facilitated
amplification using AAV-specific primers and phi29 DNA polymerase (Figure 4.1B).
Following overnight isothermic incubation, the resulting high-molecular weight
amplicons were digested into unit-length fragments (Figure 4.1C) and cloned into a
standard plasmid cloning vector for sequence and functional analysis. Unit-length AAV
genomes were generated by releasing inserts from the vector backbone and performing an
intramolecular ligation.
Figure 4.1: LRCA schematic for the isolation of infectious AAV molecular clones.

Total cellular DNA is first digested with a restriction enzyme that does not cut within the AAV genome, but restricts the host genome. This results in the creation of linear fragments that are substrates for exonuclease digestion (Plasmid-Safe DNase; step A). PS-DNase does not degrade double-stranded, circular DNA molecules. To amplify remaining circular episomes, AAV cap gene specific primers (complementary to each strand) are annealed to the circular episome, and phi29 DNA polymerase is used to initiate rolling circle replication via strand displacement (step B). LRCA reaction products are high molecular weight linear AAV concatamers (U) that can be digested with a single-cut enzyme to produce unit-length AAV genomic fragments, which can be cloned for functional analysis (step C). The unit-length AAV genomic fragment can be released from the cloning vector, self-ligated to produce a circular monomeric AAV genome (step D), and directly transfected into HeLa cells (P0). In the presence of Ad5 helper virus infection, the AAV genome can replicate and yield progeny virus. To detect the production of infectious AAV, a crude viral lysate from P0 is passed onto HeLa cells co-infected with Ad5 (P1). Hirt DNA isolated from P0 and P1 cell populations were analyzed by Southern hybridization to detect the presence of replicating monomeric and dimeric AAV forms.
Figure 4.1: LRCA schematic for the isolation of infectious AAV molecular clones.
The resulting circular monomeric AAV genome was then directly transfected into HeLa cells for further characterization.

As an important control, substrate specificity was examined using Detroit 6 cells, a well-characterized transformed cell line that contains integrated AAV genomes (head-to-tail tandem arrays) in the AAVS1 locus on chromosome 19 (q19.1). As noted above, AAV genomes integrated within the Detroit 6 DNA would not be predicted to be suitable for SSLRCA amplification. To that end, we spiked 1,000 integrated AAV2 copies (from Detroit 6 total DNA) into 100 ng of naïve tonsil DNA and performed SSLRCA. The AAV2 genome copy number was determined by cap gene quantitative PCR before and after SSLRCA. As predicted, the average fold-amplification was 1.3 ± 1.0 over 10 independent replicate assays (data not shown). Similar results were observed using 100-fold higher (10^5 copies) input of Detroit 6 DNA.

To determine the detection (amplification) limit for circular AAV, decreasing amounts a 15 kb plasmid containing AAV2 inverted terminal repeats (ITR) and rep and cap genes were spiked into 100 ng of naïve tonsil DNA and subjected to SSLRCA. The fold-amplification over a range of input copies was determined for 4 independent replicate experiments (data not shown). This experiment showed that we could reliably amplify as few as 25 plasmid copies (10^7 fold amplification). Thus, using quantitative PCR analysis, we were able to predict which tissues might be amenable to SSLRCA.

**Direct amplification of AAV genomes from human tissue.** As noted above, we previously identified 9 human DNA samples that contained AAV sequences that ranged from 80 - 33,000 copies/µg (Chapter 2). These 9 DNA samples were subjected to SSLRCA, and 6 yielded AAV2-specific amplification. Digestion of the amplified
products with a restriction enzyme that was predicted to cut once inside the AAV genome (Hind III) yielded the expected 4.7 kb unit-length fragment (Figure 4.2). The amount of amplified AAV2 product correlated with input AAV copy number previously determined from the various samples (Table 4.1). The 3 AAV samples that failed to amplify (LG15, T17, and T41) harbored copy numbers that were at or below the limit of detection for the assay (see Table 4.1). The low level of amplification for T32 and T41 were also consistent with the limit of detection data. Importantly, the integrated AAV genomes within the Detroit 6 DNA control (100 ng) did not amplify.

**Isolation of infectious AAV molecular clones.** Having amplified AAV genomes from human tissue samples, we next asked whether SSLRCA could be used to isolate intact full-length molecular clones representing infectious wild-type AAV directly. To that end, the unit-length amplified AAV2 DNA (Figures 4.1 and 4.2), was cloned into a plasmid cloning vector using the unique Hind III restriction enzyme site located in the AAV2 cap gene. To determine whether the resulting clones would give rise to infectious virus, the AAV DNA was released from the cloning vector backbone and self-ligated to restore the cap ORF and regenerate a circular monomeric AAV genome. The ligation reaction was directly transfected into adenovirus type 5 infected HeLa cells. After adenovirus-induced cytopathic effect was observed, a cell lysate was prepared, clarified, and passed onto fresh HeLa cells (again, with adenovirus). Hirt DNA was isolated from both the transfected cells (P0) and the infected cells (P1) and analyzed by Southern blot hybridization. Analyses of 9 clones isolated from 3 individual tonsil-adenoid tissues are shown in Figure 4.3. DNA replication in P0 was observed for all clones as judged by the presence of characteristic AAV monomeric (Rfm) and dimeric
Figure 4.2: LRCA method detects episomal AAV in human tissue DNA. Human total cellular DNA (100 ng) was subjected to LRCA and the resulting product was digested with Hind III (cuts once within the AAV genome), and subjected to Southern blot hybridization using a 1 kb PCR-amplified AAV2 rep-cap DNA fragment as the probe. A 4.7 kb unit-length DNA fragment was detected in 6 of the AAV positive samples, along with high molecular weight (HMW) DNA in samples T88 and T70. No evidence of LRCA mediated amplification of integrated AAV genomes was seen using 100 ng of Detroit 6 (Det 6) DNA (corresponding to 83,000 integrated copies).
Figure 4.2: LRCA method detects episomal AAV in human tissue DNA.
Table 4.1: LRCA assay sensitivity.

<table>
<thead>
<tr>
<th>Plasmid Input</th>
<th>Replicate 1</th>
<th>Replicate 2</th>
<th>Replicate 3</th>
<th>Replicate 4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>3.1 x 10⁴</td>
<td>3.3 x 10⁵</td>
<td>3.7 x 10⁷</td>
<td>1.6 x 10⁶</td>
<td>4 of 4</td>
</tr>
<tr>
<td>75</td>
<td>4.6 x 10⁵</td>
<td>2.2 x 10⁶</td>
<td>2.7 x 10⁷</td>
<td>4.1 x 10⁶</td>
<td>4 of 4</td>
</tr>
<tr>
<td>50</td>
<td>9.5 x 10⁵</td>
<td>1.5 x 10⁶</td>
<td>7.4 x 10⁷</td>
<td>1.5 x 10⁶</td>
<td>4 of 4</td>
</tr>
<tr>
<td>25</td>
<td>0</td>
<td>2.1 x 10⁶</td>
<td>1.1 x 10⁷</td>
<td>4.6 x 10⁷</td>
<td>3 of 4</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>1.0 x 10⁷</td>
<td>0</td>
<td>0</td>
<td>1 of 4</td>
</tr>
<tr>
<td>7.5</td>
<td>0</td>
<td>2.1 x 10⁷</td>
<td>0</td>
<td>0</td>
<td>1 of 4</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0 of 4</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0 of 4</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0 of 4</td>
</tr>
</tbody>
</table>

* Copies of a circular 15 kb plasmid, containing AAV2 rep-cap genes was spiked into 100 ng of total cellular DNA isolated from a naïve human tonsil and used as a template for LRCA.

b Fold-amplification was determined by dividing the AAV cap copy number present after LRCA by the cap copy numbers present before the LRCA assay was performed.

c Number of total replicates scored as positive for the detection of circular template.
Figure 4.3: Identification of infectious AAV molecular clones. Nine individual AAV molecular clones isolated from 3 tonsil tissues (T70, T71, and T88) were analyzed for the ability to replicate their respective genomes and produce infectious particles. Shown is an autoradiograph of a Southern blot hybridization (1 kb AAV2 rep-cap DNA fragment as probe) of Hirt DNA isolated from cells at pass 0 (P0) and pass 1 (P1). Residual input plasmid in P0 Hirt DNA was digested with Dpn I to reduce the input plasmid molecular weight (seen as a 2.5 kb DNA fragment). Replicating monomeric (Rfm) and dimeric (Rfd) AAV DNA forms were observed in P0 and P1 cells, indicative of replication competency (P0) and formation of infectious virus (P1).
Figure 4.3: Identification of infectious AAV molecular clones.
(Rfd) replicating DNA forms. Also present in the P0 Hirt DNA was a distinct 2.5 kb hybridizing fragment that was derived from the transfected DNA following Dpn I enzyme digestion. Importantly, all the molecular clones gave rise to infectious virus as judged by replicating AAV DNA in P1 cells. In the end, we were able to isolate infectious molecular clones from 5 of the 6 SSLRCA positive tissues (Table 4.2). For S17, it remains formally possible that we did not isolate enough clones to find one (or more) that would generate infectious AAV.

**Characterization of AAV produced from infectious molecular clones.** To further characterize AAV produced by our clones, we serially passaged infectious virions derived from clones T70-43 and T88-41 and showed that particles were DNase I resistant and purified at the predicted density in iodixanol gradients (data not shown). The cap gene sequence for both of these isolates predicted replacement of charged arginine residues at position 585 and 588 with serine and threonine, respectively (Chapter 2). As predicted, purified virions representing both clones did not bind to a heparin sulfate matrix (Chapter 2). We also analyzed AAV Rep and Cap protein expression profiles of these and selected other molecular clones (Figure 4.4). As expected, all of the infectious molecular clones produced the 4 Rep and 3 Cap proteins. Interestingly, the Cap proteins (VP1, VP2, and VP3) derived from prototype AAV2 migrated faster than the Cap proteins from the infectious molecular clones.

Full nucleotide sequence determination of several selected molecular clones was performed. The cap sequences obtained via SSLRCA were identical to the original sequence derived by nested PCR (Chapter 2). Rep gene sequences from samples T88,
**Table 4.2: Summary of molecular clones isolated by LRCA.**

<table>
<thead>
<tr>
<th>Tissue</th>
<th>AAV copies&lt;sup&gt;a&lt;/sup&gt;</th>
<th>LRCA&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Clones analyzed&lt;sup&gt;c&lt;/sup&gt;</th>
<th>% infectious clones&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>T17</td>
<td>56</td>
<td>-</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>T32</td>
<td>21</td>
<td>+</td>
<td>1/1</td>
<td>100 (1/1)</td>
</tr>
<tr>
<td>T40</td>
<td>655</td>
<td>+</td>
<td>6/60</td>
<td>17 (1/6)</td>
</tr>
<tr>
<td>T41</td>
<td>59</td>
<td>-</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>T70</td>
<td>780</td>
<td>+</td>
<td>15/80</td>
<td>93 (14/15)</td>
</tr>
<tr>
<td>T71</td>
<td>760</td>
<td>+</td>
<td>6/26</td>
<td>67 (4/6)</td>
</tr>
<tr>
<td>T88</td>
<td>3,300</td>
<td>+</td>
<td>14/80</td>
<td>71 (10/14)</td>
</tr>
<tr>
<td>S17</td>
<td>20</td>
<td>+</td>
<td>3/3</td>
<td>0 (0/3)</td>
</tr>
<tr>
<td>LG15</td>
<td>8</td>
<td>-</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Detroit 6</td>
<td>83,000</td>
<td>-</td>
<td>0</td>
<td>NA</td>
</tr>
</tbody>
</table>

<sup>a</sup> AAV copy number in 100 ng of total cellular DNA based on cap Q-PCR.

<sup>b</sup> LRCA positive based on Southern blot detection of AAV genomes.

<sup>c</sup> The fraction represents the number of molecular clones that were analyzed for infectivity (passage) divided by the number of total clones isolated by bacterial colony hybridization.

<sup>d</sup> Since only a subset of clones were screened for passage, the percentage reflects our ability to obtain replicating monomeric and dimeric AAV genomic species following passage (P1).
Figure 4.4: Rep and Cap expression of T88 molecular clones. Individual molecular clones were analyzed for AAV Rep (Rep78, 68, 52, 40) and Cap (VP1, VP2, VP3) protein expression. Shown are western blot hybridization blots of transfected cell lysates (P0). “U” designates an untransfected cell lysate, while “wt” denotes a wt AAV2 infected lysate.
Figure 4.4: Rep and Cap expression of T88 molecular clones.
T70, T71, S17, T32, T40 were highly homologous to each other and to AAV2 (> 98% amino acid identity). Sequence analysis of the 5’ and 3’ ITR regions revealed the presence of ITR D regions from either direction, characteristic of a double-D ITR structure (Chapter 3). The ITR sequence for multiple infectious clones (T88-41, T88-79, T70-19, T70-43, T70-64, T71-1, T71-5, T71-10) was determined, and all except one contained a complete double-D ITR structure; both “flip” and “flop” forms were represented. Interestingly, the single imperfect double-D ITR infectious clone (T88-62) possessed an A region duplication with a partial deletion of sequence between the two A elements (Figure 4.5). In addition, double-D ITR structures from several non-infectious molecular clones (T88-16, T88-78B) were determined and shown to possess significant internal deletions in A, B, and C hairpin regions (Figure 4.6).

**Double-D ITR structures present in tissues.** Based on ITR sequence analysis, we concluded that infectious molecular clones contained a complete double-D ITR element. Conversely, limited ITR sequence analysis of non-infectious clones suggested that deletion of internal ITR regions was responsible for the lack of infectivity since these clones expressed the expected rep and cap proteins (data not shown). Consequently, we were interested to know if double-D ITR junction formation in vivo yielded predominately complete non-rearranged structures, or if the ITR deletions were imprecise and only a subset contained the complete double-D ITR.

To answer this question, we decided to perform SSLRCA directly from tissue genomic DNA because ITR rearrangements have been observed to occur during passage in bacteria. Preceding in vivo analysis of ITR forms, we characterized the fidelity of the SSLRCA assay using a complex ITR structure that consisted of two complete ITRs
Figure 4.5: Structure of the AAV ITR junctions from selected molecular clones.

Shown schematically are ITR structures observed in several infectious molecular clones, with the palindromic ITR sequences designated as D, A, B, B’, C, C’, A’, and D’ in accord with standard nomenclature. The complete ITR sequence was determined for clones T88-41, T88-62, and T88-79. Clones 41 and 62 contain a complete double-D ITR (in both flip and flop orientations). T88-62 contains an imperfect double-D ITR junction with two A’ regions, both of which contain small deletions. In addition, three T70 clones (19, 43, 64) and three T71 (1, 5, 10) also possessed a complete double-D structure that was identical to T88-79. All ITR sequences were identical to the published AAV2 ITR sequence.
Figure 4.5: Structure of the AAV ITR junctions from selected molecular clones.
Figure 4.6: Episomal AAV contains double-D ITR structures *in vivo*. Various circular AAV templates were subjected to LRCA. The input templates (−) and the amplified products (+) were end-labeled and analyzed by acrylamide gel electrophoresis. Schematics of the ITR structure of various templates are shown to the left of the autoradiograph, with the various ITR regions denoted as A, B, C, and D. The ITR junction is oriented with respect to the AAV Rep (R) and Cap (C) genes. Black arrowheads designate the specific input (−) and LRCA (+) bands. The ITR-ITR template contains two individual end-joined ITRs. The DD+ was derived from clone T88-62, DD was derived from clone T88-41, DD from clone T88-16, and D from clone T88-78B. In all cases, the input template was amplified correctly by LRCA, as evidenced by identical band migration before and after LRCA. When LRCA products from tonsil T88 were analyzed in this fashion, several distinct bands were identified. The most prevalent band (48% of the overall signal, white arrowhead) corresponded in size to the complete DD template. The predicted size of an end-joined ITR-ITR junction from tonsil T88 is designated with a white block arrow.
Figure 4.6: Episomal AAV contains double-D ITR structures *in vivo*.
joined at their ends (Figure 4.6). Following amplification, the ITR-ITR construct (input) and SSLRCA amplified material were end-labeled and fractionated on an acrylamide gel. The end-joined ITR-ITR template was amplified faithfully, with no deletions evident based on identical mobility through the gel (Figure 4.6, lane “ITR-ITR”, boxed bands). Additionally, when T88 molecular clones possessing 4 distinct double-D structures (Figure 4.6, lanes “DD+”, “DD”, “ADD”, and “D”; see the boxed bands in each lane) were similarly analyzed, all SSLRCA amplified material was replicated faithfully.

After the preliminary analyses of these known substrates, SSLRCA was then performed on genomic DNA isolated directly from tonsil-adenoid T88 (Figure 4.6, lane “T88”). Multiple discreet species were easily identified, and when these amplified fragments were compared on the same gel to the known tonsil T88 molecular clones, the predominant in vivo form corresponded in size to the complete double-D structure (Figure 4.6, white arrowhead). Quantitative densitometry revealed that 48% of the total signal corresponded to intact double-D ITR structures. Significantly, there was no specific band present at the predicted size of an end-joined ITR-ITR structure (Figure 4.6, white star), suggesting that in this particular sample, this was not an in vivo form. However, the presence of numerous ITR species (some larger and many smaller) indicated that double-D formation is imprecise.

**DISCUSSION**

To date the vast majority of work characterizing AAV molecular forms have used cell cultures to characterize AAV replication, gene expression and genome integration. Using such detailed molecular data, a logical life-cycle paradigm was developed that
posited site-specific virus integration as a central tenet. Integration is mediated by the virally encoded large rep proteins (Rep78 or Rep68) and involves recognition of cognate binding sites within the virus and host genome (GAGC tetra-nucleotide repeats). However, only a single study to date has looked at experimental AAV infection in the primate host (77). Difficulty in detecting the presence of the virus post-inoculation in the oropharynx and respiratory tract (even in the presence of co-infecting adenovirus) illustrates the challenges with virus recovery and sampling effects using large animal primate models. We have taken an alternative approach to understanding molecular aspects of natural AAV infection, in that we have begun to characterize the molecular form and frequency of naturally acquired persisting AAV genomes in primate tissues. As we detailed herein, the development of the LRCA method enabled the isolation of molecular clones from the natural host, from which we could begin to obtain structure function data regarding the critical ITR elements resulting in episomal persistence. This analysis led to several observations: (1) the majority of in vivo forms are extra-chromosomal circles; (2) episomal circles possess complete or partially deleted double-D ITR structures; (3) infectious forms possess complete double-D structures, and (4) the cellular mechanism for circularization is imprecise resulting in multiple double-D ITR forms within a single tissue, but preferentially yields the complete double-D element. The identification that 50% of the circular forms within a human tonsil sample are perfect double-D structures suggests a powerful selection strategy the virus utilizes to maintain ITR continuity without loss of information.

Perhaps the most significant implication from the above findings relates to the AAV life-cycle. The presence of the double-D ITR structure has been demonstrated to

123
be the sole *cis* element required for the complete AAV life-cycle (204). Thus, upon
resolution of the double-D ITR to yield a no-end substrate (and in the presence of a co-
infected helper virus), this no-end substrate could be nicked at one of the 2 available
terminal resolution sites (trs) by the large Rep proteins and proceed through leading-
strand AAV DNA replication and infectious virus formation. The direct implication is
that AAV can proceed through its life-cycle in the complete absence of an integration
intermediate. Additionally, our wild-type sequence data and that of others investigators
who have reported identical double-D ITR elements for rAAV vectors in muscle tissue,
supports the view that the AAV ITR contain all the *cis* sequence information necessary to
undergo rep-independent circularization. Therefore, these data also argue for an
additional *cis* ITR property – that of circularization.

The mechanism of ITR circularization is unknown, but data using self-
complementary vectors suggest it occurs through a double-stranded intermediate
following second–strand synthesis (116). Moreover, higher order forms (concatamers)
appear to form through the interaction of double-D ITR elements on separate circles (42,
44). In cell culture, ITR circularization is a rapid process and at high virus inputs the
circularized viral form may represent an integration intermediate. Two observations are
consistent with this supposition: (1) most proviral integrants are arranged in head-to-tail
arrays, which are thought to require a circular template and limited DNA replication to
generate the arrays, and (2) efficient AAVS1 integration is heavily dependent on the MOI
(142). Thus at low MOI (<100), integration in cell culture is a rare event and the viral
genomes are lost during subsequent passage. It is unclear why an MOI exceeding 100 IU
(1,000 – 10,000 virus genomes) must be obtained to allow for targeted integration, but
could imply that saturation of the circularization process is possible. Whether similar input viral loads are necessary in vivo for integration are unknown, but we have observed wt AAV genomes in excess of 400 copies per cell in non-human primate tissue with no detectable integration into the AAVS1 locus (Chapter 5).

Additional research using cells in culture has focused on characterizing those cellular enzymes involved in the circularization process. To date, multiple DNA double-strand break repair (DSBR) enzymes are implicated and include: ATM, Mre11, DNA PKcs, KU 70/86, and Rad 50, 51, and 52 proteins (32, 45, 167, 176). These recombination proteins could be envisioned to either promote circularization (e.g. Mre 11 and ATM) or act to inhibit integration (e.g. DNA-PKcs). The multiple sized double-D ITR structures observed in tonsil sample T88 are consistent with homologous recombination between repeat regions in two juxtaposed ITRs. Conversely the failure to detect head-to-head ITR junctions argues that recombination via non-homologous end joining (NHEJ) may not be the favored pathway of ITR circularization.

Based on these data, a logical question arises as to the function or purpose of integration in the natural AAV life-cycle. Our data showing undetectable levels of S1 integration, as well as, the observation in cell culture that approximately 0.1% of infectious AAV particles integrate into AAVS1 (142, 143) make it difficult to argue for a prominent role of virus integration in AAV natural infection. However, integration would clearly be a benefit in actively dividing cell populations and perhaps this does occur at higher frequencies under “favorable” intracellular conditions (sufficient Rep expression and high virus inputs). Answers to this question and others, such as the
natural cellular reservoirs of AAV, await the execution of controlled non-human primate infection studies.

In conclusion, our data strongly suggest that AAV is not an obligatory integrating entity, but rather persists in multiple tissues in the absence of significant levels of cellular integration as monomeric and concatameric circles. Importantly, the molecular form of these extra-chromosomal circles is compatible with all currently reported aspects of the AAV life-cycle.
CHAPTER 5

EPISOMAL PERSISTENCE OF ADENO-ASSOCIATED VIRUS GENOMES ISOLATED FROM RHESUS MACAQUES

INTRODUCTION

Adeno-associated viruses (AAV) are non-pathogenic, replication defective paroviruses that productively infect human and non-human (NHP) primates in the presence of a helper virus (usually adenovirus or HSV) (5, 13, 58, 60, 82, 133). Owing to its non-pathogenicity and site-specific integration property (identified in cell cultures), AAV was quickly adapted for use as a gene transfer vector (13, 69, 92). Recombinant AAV vectors have since demonstrated a strong pre-clinical safety profile, which has led to the initiation of multiple human clinical trials for a variety of inherited and acquired diseases, which continue to expand rapidly (25). Perhaps surprisingly, only 2 studies have attempted to establish experimental AAV infection in NHPs (1, 77). Both studies were hampered by difficulty in establishing productive AAV infection conditions using human helper adenovirus infection. Given these difficulties and the cost of primate based research, the vast majority of current data regarding natural AAV infection has been to simply survey tissues for endogenous sequences, clone these sequences and compare the phylogeny of the AAV variants (serotypes) identified (56, 58-60). Such studies have
documented a rich reservoir of divergent AAV genomes in multiple tissues at surprisingly high frequencies. These data argue that the AAV:host relationship is more heterogeneous than expected and a better understanding of the persistent forms of latent provirus in these organisms needs to be more fully defined.

To further characterize the state of the persisting AAV genome in natural AAV infection, we recently reported that in AAV positive tissue samples from children, we were unable to detect site-specific S1 integration using a sensitive S1-specific PCR assay. Rather, we went on to show that episomal forms of AAV DNA were commonly found in these same children’s tissues (Chapter 3). The caveat with these human studies was that only 7 AAV positive samples were analyzed and that the AAV copy numbers were uniformly low (≤ 0.2 copies/cell) possibly precluding detection of low-level targeted integration (Chapter 2). In the current follow-up study we have compared the structural and functional properties of latent wild-type AAV in a different primate host, the rhesus macaque, to increase both the breadth of the analysis and determine whether low-level AAV integration was occurring in tissues harboring substantially greater levels of endogenous AAV. Our current analysis identified large numbers of positive AAV tissues and at substantially increased AAV copies per nuclei in several macaque tissues. Increased AAV genome levels facilitated a more thorough genome analysis than our initial human study and further added to the collective sample size. Significantly AAV persistence was found in extra-chromosomal forms, with no evidence for targeted AAV integration observed. These data argue that genome integration is not required as part of the AAV life cycle and supports the safe use of recombinant vectors which also persist in predominately in a non-integrated state following in vivo administration.
MATERIALS AND METHODS

**Cell Propagation.** HeLa cells and Cos-7 cells were maintained in DMEM supplemented with 10% fetal bovine serum (FBS) and penicillin and streptomycin. HeLa cells and Cos-7 cells were purchased from the American Type Culture Collection (Rockville, Md).

**Rhesus macaque tissue procurement.** Tissue samples were obtained from rhesus macaques (*Macaca mulatta*) undergoing euthanasia by intravenous injection of a lethal dose of sodium pentobarbital/sodium phenytoin. Organs were rapidly harvested in the following order: brain, gonads, stomach, large intestine, small intestine, kidney, spleen, heart, lung, lymph nodes (axillary, inguinal, and mesenteric), liver, left quadriceps, and right quadriceps. Each organ was surgically removed using dedicated surgical instruments and surgical gloves were changed between each organ to prevent sample cross-contamination. Upon removal, the organ was minced and placed into cryovials that were frozen in a dry ice/ethanol bath and subsequently stored at –80°C.

**Functional analysis of wild-type AAV molecular clones.** To determine whether LRCA generated AAV molecular clones were infectious, the pBlueScript vector backbone and cap ORF were re-created by digestion with Sph I or Sal I and self-ligated to form head-to-tail circular monomeric genomes. The ligated products (~1 ug) were directly transfected into 4 x 10⁵ HeLa cells using FuGene 6 reagent (Roche Applied Science) followed by infection with Ad5 (MOI = 20). After maximum cytopathic effect development (48 hr), the transfected cells (pass zero; P0) were harvested and Hirt DNA prepared from half of the cells (subsequently restricted with Dpn I to digest input plasmid DNA). A clarified cell lysate was generated from the other half of P0 cells by three
freeze/thaw cycles and a 60 min incubation at 56°C to heat inactivate Ad5. A 1:10 dilution of the clarified lysate was passed onto 4 x 10^5 HeLa cells with Ad5 (MOI = 20), and after 48 hr, the AAV transduced cells (pass one; P1) were harvested and low molecular weight DNA isolated.

**Isolation and detection of AAV sequences.** Total cellular DNA was extracted from rhesus macaque tissues as described previously. Detection of AAV was determined by capsid gene PCR amplification. Briefly, nested degenerate PCR was performed on 100 ng of total rhesus cellular DNA using a conserved degenerate primer set (CapSS2978: 5’- GGYGCCGAYGGAGTGGGYARTKCC - 3’; and, Cap 18S: 5’- GAWKCCCCARTWGTGTTRATGAGTC - 3’). One µl of the first round PCR served as the template for the second round using nested, degenerate PCR primers (Cap 19S: 5’- GYARTKCCCTCRGGWRATTGGCA - 3’; and, CapSS3189: 5’ - GATGAGTCKYTGCCAGTCWCGKGG - 3’). Reaction components for both rounds were 400 nM of each primer, 400 nM dNTP, 0.5 unit SureStart Taq polymerase (Stratagene), 1X SureStart reaction buffer in a final volume of 25 µl. PCR cycling conditions were: 1 cycle at 94°C for 12 min; 36 cycles at 94°C for 30 sec, 52.5°C for 30 sec, and 72°C for 1 min; followed by a 5 min extension step at 72°C. To confirm PCR amplicon identity, the AAV nested capsid PCR were resolved on 0.8% agarose gels. Amplified DNA fragments were cloned into pCR 4-TOPO vector using TOPO TA cloning kit according to manufacturer’s instructions (Invitrogen Inc.). DNA plasmid clones were sequenced using BigDye terminator chemistry and an ABI 727 capillary electrophoresis automatic sequencer (PE Applied Biosystems Inc.) by the CCRI Sequencing Core Laboratory.
Quantitative taqman PCR. AAV genome copy number in rhesus DNA samples were quantitated using real-time TaqMan PCR analysis (ABI 7000 PCR machine, PE Applied Biosystems Inc.). The primers and probe set were selected following alignment of Rep gene sequences of LRCA-amplified clones. Primers were: RepFor, 5’-ATCGTCACCTCCAACACCAAC-3’, 300 nM final concentration; Rep1Rev, 5’-GCTGGTGCTCGAAGGTGGTG-3’, 300 nM final concentration; Rep probe, 5’-[6-FAM]TTCCCGTCAATCACGGCGCA[TAMRA-6-FAM]-3’, 270 nM final concentration. PCR conditions were: 50°C 2 min., 95°C 10 min., 40 cycles of 95°C 15 sec. and 60°C 1 min. using 125 ng of total cellular DNA in 1X TaqMan PCR master mix.

Sequence analysis. The DNA and corresponding protein sequences were aligned and analyzed using the Clustal X method implemented in MegAlign software in DNASTAR (DNASTAR, Inc.). Similarity of published AAV sequences and the new AAV sequences was determined using one-pair alignment according to the Lipman-Pearson method.

AAVS1 PCR. Detection of AAV integration in the rhesus AAVS1 locus was performed using 500 ng of total cellular DNA as template with 3 primer sets specific for both human and rhesus AAVS1 and AAV cap genes. The reaction conditions for both rounds of PCR were as follows: 25 pmol of each primer, 200 uM dNTPs, 5 U Herculase Hotstart DNA polymerase (Stratagene) with Herculase reaction buffer. The cycling parameters for both rounds of PCR were as follows: denaturation at 95°C for 90 sec, followed by 30 cycles of 95°C for 30 sec, 58°C for 30 sec, 72°C for 3 min, with a final extension of 72°C for 10 min. The AAV-specific primers used for the first round were as follows: set #1, NHP-AAV #1-F1, 5’-GCCACCTACAACCAACCCACC-3’; set #2, NHP-
AAV #2-F1, 5’-ATTCCACATGGMTGGCGAC-3’; set #3, NHP-AAV #3-F1, 5’-CGACAGAGTCATCACAC-3’. The AAVS1-specific primer for the first round was: Rh/Hu-AAVS1-2954-rev, 5’-GAAACGAGAGATGGCAGG-3’. The AAV-specific primers for the second round were as follows: set #1, NHP-AAV #1-F2, 5’-TTTGACTTCAACAGATTCCACTG-3’; set #2, NHP-AAV #2-F2, 5’-AGAGTYATACCACCAGC-3’; set #3, NHP-AAV #3-F2, 5’-CTGCCCCACCTACAAACCA-3’. The AAVS1-specific primer for the second round was: Rh/Hu-AAVS1-2860-rev, 5’-CGGAGAGGACCCAGA-3’. For the control experiment, 5 to 50 pg (1 to 10 AAV:AAVS1 junctions) of Detroit 6 DNA were spiked into each of the 25 AAV containing rhesus DNA samples and assayed by identical PCR conditions. To verify our ability to amplify rhesus AAVS1 sequence we ran control PCR with primers spanning 1.4 kb of the AAVS1 locus including the RBS. The control primers were as follows: Rh-AAVS1-GSP-F1, 5’-GCTGTCCAGTCGATTTCCTAACT-3’; Rh-AAVS1-GSP-R1, 5’-GAGAATGTCAGGTCAGTAAGCC-3’. The reaction conditions were as follows: 500 ng of total cellular DNA as template, 200 ng of each primer, 4% DMSO, 1X Herculase Master Mix (Stratagene Inc.). The cycling conditions were as follows: 98°C 3 min., 33 cycles of 98°C 40 sec., 58.5°C 30 sec., 72°C 1 min. 40 sec., with a final extension of 72°C for 10 min.

**Cloning and sequence of rhesus AAVS1 locus.** Initially, we obtained a plasmid containing ~500 bp of rhesus AAVS1 locus sequence containing the RBS from Dr. Terence R. Flotte. Primers were designed to obtain flanking 5’ and 3’ sequences following the Universal GenomeWalker kit (Clontech). Amplified fragments were
cloned into TOPO-TA cloning vector (Invitrogen Corp.) for sequence analysis and continued PCR extension.

**PS-DNase treatments.** Twenty ug of total cellular DNA was incubated for 16 h at 37°C in 33 mM Tris (pH 7.8), 66 mM potassium acetate, 10 mM magnesium acetate, 0.5 mM dithiothreitol, 2 mM ATP with 10 U/ug PS-DNase (Epicentre Technologies Inc.). Genomic Southern blot analysis was performed on DNA fractionated on 0.6% agarose gels, which were subsequently transferred to nylon membranes for hybridization. DNA hybridization conditions were 60°C for 16 hr in buffer containing 6X SSC, 1X Denhardt's reagent, and 200 ug/ml sonicated herring sperm DNA. Nylon membranes were washed twice at 60°C in 2X SSC, 0.2% SDS for 30 min, and then twice at 60°C in 0.2X SSC, 0.2% SDS for 30 min.

**Linear Rolling Circle Amplification (LRCA).** 1 ug of genomic DNA was first digested with two restriction enzymes that do not cut within the AAV genome (Pac I and Pme I). To eliminate single-stranded DNA and/or linear double-stranded DNA, the restricted DNA was then subjected to 10 U of PS-DNase (Epicentre Technologies) in a final volume of 25 ul for 16 hr at 37°C in 33 mM Tris (pH 7.8), 66 mM potassium acetate, 10 mM magnesium acetate, 0.5 mM DTT, and 1 mM ATP. The nuclease was then heat inactivated for 30 min at 70°C. Template DNA was mixed in a final volume of 15 ul in 10 mM Tris-HCL, pH 8, with 300 pmol of each AAV specific primer. AAV specific primers were as follows: NHP-AAV #2- Temp F1, 5’-TGGGCGACAGAGT-3’; NHP-AAV #2-Temp R2, 5’-GTGCTGGTGGTGGAT-3’; each primer contains a phosphorothioation linkage between the last two 3’ bases. The mixture was then heated to 95°C for 3 min, cooled to 4°C to allow primer annealing, then mixed with 15 ul of
TempliPhi reaction buffer containing phi29 polymerase (Amersham Inc.), and incubated for 18 h at 30°C. Amplified products were heat inactivated at 65°C for 10 min and then digested with either Sph I or Sal I for Southern hybridization. Cloning of digested AAV genomes were gel-purified, cloned into pBlueScript (Stratagene), and transformed into SURE-2 bacteria (Stratagene), to minimize unwanted ITR recombinations and/or deletions. DNA sequencing was performed by the CCRI Sequencing Core Laboratory. In addition, manual DNA sequencing of the AAV ITR region was accomplished using the SequiTherm EXCEL II DNA Sequencing kit (Epicentre) following the manufacture’s protocol for isothermal sequencing with α-labeled 35S dATP.

RESULTS

Detection of wild-type AAV sequences in rhesus macaque DNA. To characterize wild-type AAV persistence in non-human primates, we surveyed 3 tissues (mesenteric lymph node, spleen, and colon) from 10 monkeys (28 total samples) for the presence of wild-type AAV capsid gene sequences. Total cellular DNA was subjected to nested PCR amplification using degenerate oligonucleotide primers spanning a previously identified hypervariable capsid gene region, that enables the detection and discrimination of AAV serotypes 1 through 9 (56). Significantly, 25 of the 28 samples (89.3%) yielded robust PCR amplification products, with all 10 animals possessing at least one positive tissue. Capsid gene PCR amplicons were subsequently cloned and the DNA sequence determined. Sequence variation was found within the same animal, as well as, between animals. Additionally, limited sequence alignment analysis in this 225 bp region suggests that the sequences share significant homology (98-99% at the
nucleotide level) to serotypes 4, 7, or 8. Subsequently, quantitative PCR was used to
determine AAV copy number within the 25 positive tissues and ranged from below 0.01
copies/cell (detection limit) to a maximum of approximately 400 copies/cell (Table 5.1).

**AAVS1 integration analysis.** Wild-type AAV DNA has been shown to
integrate in a site-specific manner on chromosome 19 (AAVS1 locus) in cultured cells
(14, 71, 99). It has been assumed (but not shown) that it integrates similarly during the
course of natural primate infection. To confirm this assumption, we tested all monkey
tissue samples that contained AAV DNA for integration at this locus. As illustrated in
Figure 5.1 (Panel A), we developed an extremely sensitive nested PCR assay to detect
AAV/AAVS1 viral-cellular junctions. To do so, we initially cloned and sequenced 3 kb
of the rhesus macaque AAVS1 locus in order to generate sequence specific anchor
primers. For the AAVS1 PCR strategy, 1 of the 3 different AAV capsid specific primers
were employed (depending on the AAV capsid sequence obtained from the particular
tissue) and used with a single set of AAVS1 anchor primers. It is important to note (for
the Detroit 6 control experiments) that all 3 capsid specific primers would also detect
AAV2. The AAVS1 primers were designed to a region 3’ of the Rep binding site (RBS).
We first defined our level of sensitivity with these primer sets by using a well
characterized latently AAV infected cell line (Detroit 6), which contains five AAV2
genome copies/cell integrated into the AAVS1 locus in a head-to-tail configuration (14).
As seen in Figure 5.1 (Panel B), we could easily detect 1 to 10 AAV/AAVS1 junctions of
Detroit 6 DNA spiked into 1 ug of naïve HeLa DNA with all 3 primer sets. Identical
<table>
<thead>
<tr>
<th>Sample name</th>
<th>AAV copies/ug</th>
<th>AAV copies/cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>rh.386-M</td>
<td>82</td>
<td>0.0005</td>
</tr>
<tr>
<td>rh.386-S</td>
<td>Undetectable</td>
<td>0</td>
</tr>
<tr>
<td>rh.387-M (rh.75)</td>
<td>6.75x10⁷</td>
<td>405</td>
</tr>
<tr>
<td>rh.387-C (rh.79)</td>
<td>3.35x10⁷</td>
<td>200</td>
</tr>
<tr>
<td>rh.387-S</td>
<td>1.8x10⁵</td>
<td>1</td>
</tr>
<tr>
<td>rh.423-M</td>
<td>3x10⁶</td>
<td>18.5</td>
</tr>
<tr>
<td>rh.423-C</td>
<td>1.63x10⁶</td>
<td>10</td>
</tr>
<tr>
<td>rh.423-S</td>
<td>1.5x10⁵</td>
<td>1</td>
</tr>
<tr>
<td>rh.424-M</td>
<td>3.4x10⁴</td>
<td>0.02</td>
</tr>
<tr>
<td>rh.424-C</td>
<td>7.5x10⁴</td>
<td>4.5</td>
</tr>
<tr>
<td>rh.424-S</td>
<td>3.1x10⁴</td>
<td>0.19</td>
</tr>
<tr>
<td>rh.425-M</td>
<td>152</td>
<td>0.001</td>
</tr>
<tr>
<td>rh.425-S</td>
<td>208</td>
<td>0.001</td>
</tr>
<tr>
<td>rh.426-M (rh.73, rh.74)</td>
<td>7.5x10⁴</td>
<td>4.5</td>
</tr>
<tr>
<td>rh.426-C</td>
<td>3.5x10⁵</td>
<td>0.02</td>
</tr>
<tr>
<td>rh.18654-M (rh.76)</td>
<td>7.5x10⁴</td>
<td>4.5</td>
</tr>
<tr>
<td>rh.18654-C</td>
<td>4.5x10⁴</td>
<td>0.27</td>
</tr>
<tr>
<td>rh.18654-S</td>
<td>1.9x10⁵</td>
<td>0.12</td>
</tr>
<tr>
<td>rh.18655-S</td>
<td>Undetectable</td>
<td>0</td>
</tr>
<tr>
<td>rh.18679-M</td>
<td>2.3x10⁵</td>
<td>0.14</td>
</tr>
<tr>
<td>rh.18679-C</td>
<td>5.5x10⁵</td>
<td>0.032</td>
</tr>
<tr>
<td>rh.18679-S</td>
<td>2.2x10⁵</td>
<td>0.013</td>
</tr>
<tr>
<td>rh.18685-M (rh.77)</td>
<td>5.5x10⁵</td>
<td>34</td>
</tr>
<tr>
<td>rh.18685-C</td>
<td>1.2x10⁶</td>
<td>7</td>
</tr>
<tr>
<td>rh.18685-S (rh.78)</td>
<td>1.3x10⁵</td>
<td>1</td>
</tr>
</tbody>
</table>

a M is mesenteric lymph node; S is spleen, and C is colon.
b Copy number determined using Q-PCR to a conserved region in the rep gene (see Materials and Methods).
c Conversion of copies/cell based on 6 pg DNA/nucleus.
d Parentheses indicate sources of tissues for complete genome analysis.

Table 5.1: AAV DNA copy number in rhesus macaque tissue.
Figure 5.1: AAVS1 PCR validation on Detroit 6 cells. (A) Schematic of the rhesus and human AAVS1 locus. The Rep binding site (RBS) and Detroit 6 integration site are designated by vertical arrows. Sequence analysis of the rhesus and human homologs reveal 91% homology (see Materials and Methods). The location of the AAV capsid-specific and AAVS1-specific primers used in the nested PCR reactions are designated as horizontal solid arrows. (B) Southern blot hybridization results of AAVS1-PCR products using Detroit 6 cellular DNA. For each AAV capsid-specific primer set, Detroit 6 genome copies were serially diluted (100,000 to 1) into 1 ug of naïve HeLa DNA and subjected to nested PCR. The PCR products were resolved on a 0.8% agarose gel and Southern hybridization was performed using an AAV-capsid probe. The expected 2.2 kb AAV-AAVS1 junction fragment is easily detected by nested PCR down to 10 Detroit 6 AAV-AAVS1 junctions with primer sets 1 and 2, and down to 1 junction with primer set 3.
Figure 5.1: AAVS1 PCR validation on Detroit 6 cells.
assay conditions were repeated using rhesus tissue DNA. To confirm the identity of the amplified product, in-gel Southern hybridization was performed using both a AAV2 Rep/Cap probe and a rhesus AAVS1 probe. As expected the Detroit 6 DNA yielded robust hybridization signals for both AAV and AAVS1 (Figure 5.2). In contrast, only five primate tissues showed positive hybridization with the AAV2 probe, all of which were negative for cross-hybridization with the rhesus AAVS1 probe (Figure 5.2, Panels A and B). Individual PCR products were cloned and sequenced from 3 out of the 5 hybridizing fragments from the AAV2 probe; a minimum of 5 clones were analyzed for each PCR product. All of the clones contained only AAV sequence with large rearrangements and/or deletions, with no cellular sequences detected.

The lack of detectable AAVS1 integration was not completely unexpected based on our previous study using AAV positive human tissue DNA that also failed to detect AAVS1 integration (166). One key difference between the human and monkey samples herein, was that several monkey samples contained much higher AAV copy numbers (0.2 vs 400 AAV copies/cell) and were still found to lack detectable targeted AAV integration. Control PCR reactions were performed on each of the rhesus DNA samples to confirm the absence of PCR inhibitors in these samples. This was accomplished by spiking 5 to 50 pg of Detroit 6 DNA (equivalent to 1 and 10 AAV/AAVS1 junctions) into 500 ng of rhesus DNA and subjecting the samples to AAVS1 PCR. All 25 samples yielded robust amplification with 50 pg (10 junction equivalents) of Detroit 6 DNA and 12/25 amplified with 5 pg (1 junction) of Detroit 6 DNA (data not shown). Furthermore, AAVS1 DNA integrity was validated using a control AAVS1 PCR amplification,
Figure 5.2: AAVS1 PCR on rhesus samples. AAVS1-PCR was performed on 1 ug of total cellular DNA from all AAV-positive samples. Southern blot hybridization using an (A) AAV capsid-specific probe and (B) a rhesus AAVS1-specific probe, are shown here with representative samples. (A) Using the AAV capsid-specific probe fragments are detected with various sizes. Analysis of these bands yielded AAV fragments with no AAVS1 sequence (see Results). (B) Using the AAVS1-specific probe failed to detect any signal. Detroit 6 positive control DNA gave robust signal with both probes. Mesenteric lymph node (M), spleen (S), and colon (C).
Figure 5.2: AAVS1 PCR on rhesus samples.
whereby all samples amplified the expected 2.2 kb AAVS1 locus region (data not shown).

**Presence of episomal AAV genomes in rhesus tissue.** To explore the alternative possibility that the majority of AAV DNA persists as extra-chromosomal elements, we performed Southern blot hybridization on several NHP DNA samples using a novel DNA exonuclease (Plasmid-Safe DNase) to discriminate between episomal and integrated forms. PS-DNase hydrolyzes linear double-stranded DNA and linear and closed-circular single-stranded DNA but does not efficiently degrade double-stranded, closed-circular supercoiled or nicked DNA (Figure 5.3), therefore restriction enzyme digestion with a “no-cut” enzyme will fragment the genomic DNA, but will leave the AAV DNA untouched if in an episomal conformation. Accordingly, total cellular DNA from a high copy number tissue (400 AAV copies/cell, rh.387-M) was subjected to PS-DNase treatment. Digestion of rh.387-M genomic DNA with Sph I, a restriction enzyme that cuts once yielded 4 hybridizing fragments of 4.7, 2.7, 2.0, and 1 kb in size (Fig. 5.4, lane 1). The 4.7 kb fragment was consistent with the wild-type AAV genome being organized in a head-to-tail array. The unexpected 2.7 kb and 2 kb bands could represent either free linear genome fragments, and/or another species of viral DNA that the cut twice within its genome. The 1 kb fragment was likely a single-stranded form based on its presence irrespective of whether a no-cut or 1-cut enzyme was used. When these same linearized DNA fragments were incubated with PS-DNase, complete loss of signal was observed (Fig. 5.4, lane 2). When the same genomic sample was digested with Pac I, a restriction enzyme that does not cut within the AAV genome, only the 4.7 kb and 1 kb
Figure 5.3: Schematic of the PS-DNase assay for episomal wild-type AAV genomes.

The predicted outcomes of treating AAV-positive DNA with selected restriction enzymes in combination with PS-DNase are shown in this diagram. In the case of treatment with an enzyme that cuts once within the genome (Single Cutter), subsequent treatment with PS-DNase leads to degradation of all AAV genomes, regardless of their forms (episomal or integrated). In the second case, when the total DNA is cut with an enzyme that does not cut within the AAV genome (Zero Cutter), after treatment with PS-DNase, integrated forms are degraded while episomal forms are protected.
Figure 5.3: Schematic of the PS-DNase assay for episomal wild-type AAV genomes.
**Figure 5.4: Southern blot analysis of wild-type AAV molecular forms in rhesus lymph node following PS-DNase treatment.** Total DNA was isolated from mesenteric lymph node of rh.387. A southern blot hybridized with an AAV capsid-specific probe is shown. Lane 1, *Sph I*-digested DNA (Single cut [in the AAV genome]); lane 2, *Sph I*-digested DNA after PS-DNase digestion; lane 3, *Pac I and Pme I*-digested DNA (Zero cut [in AAV genome]); lane 4, *Pac I and Pme I*-digested DNA after PS-DNase digestion. PS-DNase completely degraded linearized AAV DNA (lane 2) but did not degrade the zero cut AAV DNA, even though the cellular genomic DNA was degraded (ethidium bromide stained agarose gel not shown).
Figure 5.4: Southern blot analysis of wild-type AAV molecular forms in rhesus lymph node following PS-DNase treatment.
fragments were observed (Fig. 5.4, lane 3). Importantly, after PS-DNase treatment, a resistant unit-length band was observed (Fig. 5.4, lane 4), which is consistent with this species being a double-stranded, circular episome.

**Use of Linear Rolling Circle Amplification (LRCA) to amplify and clone AAV DNA.** Our ability to visualize PS-DNase resistant AAV genomes clearly indicated that a significant fraction of the virus DNA present was unintegrated. These circular forms provided an ideal template to selectively amplify the AAV sequences for further characterization. This was accomplished using a novel assay termed linear-rolling circle amplification (LRCA) previously described by our laboratory to specifically amplify double-stranded circular AAV genomes (166). The method is predicated on the ability to amplify circular AAV genomes by isothermic rolling circle amplification using phi29 phage DNA polymerase and AAV specific primers. LRCA products are contiguous head-to-tail arrays of the circular AAV genomes. LRCA analysis of the AAV positive rhesus DNA resulted in the successful amplification of AAV episomes in 12 samples. To obtain full-length molecular clones, amplified products from 11 of the 12 samples were cut with Xba I, which is predicted to cut once in the AAV genome to yield unit-length fragments (shown in Figure 5.5). Multiple clones were obtained from 7 out of the 12 samples and complete rep and cap gene sequences were determined. Upon comparison to previously identified sequences, 4 clones were determined to be unique and named rh.73 – rh.76. Interestingly, 1 tissue sample (rh426-M) contained two divergent species (rh.73 and rh.74) of viral DNA. For the 4 samples, the predicted amino acid sequences for Rep and Cap ORFs were aligned and compared to previously published serotypes
Figure 5.5: LRCA products. LRCA products positive for amplification were Sph I-digested (single cutter [in AAV genome]), releasing unit-length head-to-tail junctions. Undigested high molecular weight (HMW) DNA can also be observed.
Figure 5.5: LRCA products.
Similarity of published AAV sequences and the new AAV sequences was determined using one-pair alignment according to the Lipman-Pearson method implemented in the MegAlign software in DNASTAR (DNASTAR Inc.). Light faced numbers represent similarity in Rep78 sequences whereas bold faced numbers represent similarity in VP1 sequences. The italicized numbers represent similarity in both.

<table>
<thead>
<tr>
<th></th>
<th>AAV 1</th>
<th>AAV 4</th>
<th>AAV 7</th>
<th>AAV 8</th>
<th>rh.73</th>
<th>rh.74</th>
<th>rh.75</th>
<th>rh.76</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAV 1</td>
<td>100</td>
<td>90</td>
<td>98</td>
<td>95</td>
<td>98</td>
<td>98</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>AAV 4</td>
<td>63</td>
<td>100</td>
<td>90</td>
<td>87</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>94</td>
</tr>
<tr>
<td>AAV 7</td>
<td>85</td>
<td>63</td>
<td>100</td>
<td>96</td>
<td>97</td>
<td>98</td>
<td>98</td>
<td>98</td>
</tr>
<tr>
<td>AAV 8</td>
<td>84</td>
<td>63</td>
<td>88</td>
<td>100</td>
<td>97</td>
<td>97</td>
<td>95</td>
<td>98</td>
</tr>
<tr>
<td>rh.73</td>
<td>79</td>
<td>61</td>
<td>83</td>
<td>80</td>
<td>100</td>
<td>99</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>rh.74</td>
<td>84</td>
<td>63</td>
<td>88</td>
<td>93</td>
<td>80</td>
<td>100</td>
<td>99</td>
<td>98</td>
</tr>
<tr>
<td>rh.75</td>
<td>65</td>
<td>82</td>
<td>82</td>
<td>64</td>
<td>62</td>
<td>64</td>
<td>100</td>
<td>99</td>
</tr>
<tr>
<td>rh.76</td>
<td>85</td>
<td>63</td>
<td>91</td>
<td>86</td>
<td>84</td>
<td>86</td>
<td>84</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 5.2: Percentage of sequence similarity between REP78 and VP1 amino acid sequences.
(Table 5.2). VP1 protein sequences were analyzed and revealed strong homology to the NHP AAV clades D, E, and AAV 4-like virus isolates (bottom portion of Table 5.2). Analysis of the REP78 (top portion of Table 5.2) ORFs revealed strong homology to AAV 1 (98-99%). Significantly, 1 independent isolate (rh.75) shares significant identity with rh.32 – 34 (99%), and therefore we propose that a new clade (Clade G) be designated to account for the AAV4-like genomes.

Molecular clones of AAV isolated from LRCA are functional. Clones isolated from 4 LRCA positive tissues were tested for fitness using a passage assay. To do so, the unit-length AAV DNA was released from the cloning vector backbone and self-ligated to restore the cap ORF and regenerate the circular monomeric AAV genome. The ligation reaction was directly transfected into adenovirus type 5 infected HeLa cells (see Figure 5.1 schematic). Upon development of adenovirus induced cytopathic effect, a portion of the clarified cell lysate (P0) was passed onto fresh HeLa cells (P1), again in the presence of helper adenovirus and low-molecular weight (Hirt) DNA isolated from the P0 and P1 HeLa cell populations. Eighteen molecular clones were assayed and 5 (27.8%) were able to replicate and produce infectious progeny (Table 5.3). Thus, we were able to isolate infectious molecular clones from 2 of the 4 AAV positive tissues analyzed. A Southern blot of Hirt DNA isolated from HeLa cells at transfection (p0) and passage (p1) from HeLa and Cos-7 cells was done to detect replicating AAV genomes (Figure 5.6). Transfected AAV DNA was shown to replicate the viral genome in p0 cells as shown by the characteristic pattern of replicating genomes (monomer, Rfm; and dimer, Rfd) present in the Hirt DNA. Higher mobility bands resulting from residual Dpn I-digested input plasmid DNA can also be detected in p0. Since these clones were isolated from monkey
Table 5.3: Summary of molecular clones isolated by LRCA.

<table>
<thead>
<tr>
<th>Tissue</th>
<th>Clones analyzed&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Percent of infectious clones&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>rh.387-M</td>
<td>5/20</td>
<td>0/5 (0%)</td>
</tr>
<tr>
<td>rh.387-C</td>
<td>0/8</td>
<td>NA</td>
</tr>
<tr>
<td>rh.426-M</td>
<td>5/30</td>
<td>3/5 (60%)</td>
</tr>
<tr>
<td>rh.18654-M</td>
<td>3/10</td>
<td>2/3 (67%)</td>
</tr>
<tr>
<td>rh.18685-M</td>
<td>5/20</td>
<td>0/5 (0%)</td>
</tr>
<tr>
<td>rh.18685-S</td>
<td>0/10</td>
<td>NA</td>
</tr>
<tr>
<td>rh.18685-C</td>
<td>0/1</td>
<td>NA</td>
</tr>
</tbody>
</table>

<sup>a</sup>The fraction represents the number of molecular clones that were analyzed for infectivity divided by the total number of clones identified by bacterial colony hybridization.

<sup>b</sup>A subset of clones were screened for passage. The percentage reflects the ability to obtain infectious AAV particles following passage.
Figure 5.6: Functional analysis of wild-type AAV molecular clones. Individual molecular clones from monkey tissues (shown here from rh.426-M) were analyzed for the ability to replicate their genomes and produce infectious particles. Shown is a Southern blot of Hirt DNA isolated from HeLa (p0 and p1a) and Cos-7 (p1b) cells at pass zero (p0) and pass 1 (p1). Replicating AAV genomes can be seen in p0 at ~4.7 kb and ~9.4 kb (replicating form monomer, Rfm; and replicating form dimmer, Rfd, respectively) as well as higher replicating forms. Residual Dpn I digested input plasmid can also be seen in p0. Hybridization in p0 indicates clones are in fact replication-competent, while hybridization in p1 indicates clones are able to produce infectious progeny.
Figure 5.6: Functional analysis of wild-type AAV molecular clones.
tissues, we tested whether in vitro tropism could play a role in detecting infectious virions in p1 by passing crude lysates from p0 onto HeLa or Cos-7 cells (NHP cell line). We detected lower levels of replicating genomes in p1-HeLa cells (Figure 5.6, lanes 8 and 11) compared in p1-Cos-7 cells (Figure 5.6, lanes 9 and 12) from a number of clones.

**DISCUSSION**

Recently, we characterized the molecular fate of AAV genomes from a limited number of AAV positive tissue samples in children (Chapters 2 and 3). We were unable to detect site-specific S1 integration in these tissues, and the majority of AAV DNA was in extra-chromosomal circles. In this study, increased sample and AAV copy number present in NHP infections was used to further define the molecular form of latent AAV genome persistence.

**Prevalence and diversity of AAV DNA in Rhesus macaque tissues.** We detected AAV DNA in 89% (25/28) of the tissues samples analyzed. However, in contrast to the low AAV genome copy numbers observed in previous human samples (≤ 0.2 copies/cell), many of the NHP tissues contained higher copy numbers of up to ~ 400 copies/cell. As seen in other tissue surveys, AAV sequences were detected beyond the probable port of entry – the oropharynx, and suggest that AAV viremia occurs following natural infection and seeds many tissues throughout the host. While the DNA sequences recovered from the human samples were most similar to AAV2, sequences identified from the rhesus tissues were most similar to previously identified AAV serotypes that circulate in the species - 4, 7 and 8. Our sequence analyses identified 4 unique capsid clones (>4 amino acids different from known variants) that we have denoted as rh.73 –
rh.76. However, this level of amino acid identity would not predict them to be serologically distinct from the currently identified serotype (1-9). While AAV capsid sequence diversity has now been well documented, little has been reported with respect to the wild-type rep sequences. Our amplification techniques yielded full-length molecular clones and analysis of rep gene sequences isolated from the animals were nearly identical to the AAV1 rep gene sequence ($\geq 99\%$). AAV1 was initially isolated from a simian adenovirus preparation and subsequently found in multiple simian adenovirus isolates, but the AAV1 cap sequence has also been isolated from human samples (5, 60). This suggests that the AAV1 rep gene may have arose in NHP populations first (present in AAV1, AAV4, AAV7, and AAV8 isolates) before expanding into the human population.

**AAVS1 integration analysis.** AAV has been shown to establish persistent infections in cell culture by targeted integration into the AAVS1 locus (93, 95). Many of the elements and trans-factors required for the targeted integration mechanism have been defined. While, it was widely assumed that the same phenomenon would exist with wild-type AAV during the course of natural infection, our lab was unable to detect site-specific S1 integration in a small number of human samples. In the current study, AAV infection was observed in macaques and that showed significantly higher copy numbers per cell. Despite rigorous controls accounting for the sequence diversity between human and rhesus macaques (both for the AAVS1 locus and AAV genomes), we were still unable to detect any AAV-AAVS1 cellular junctions at the limit of detection for the assay (1 AAV-AAVS1 junction per ~1 ug of genomic DNA). The absence of integration does not exclude the possibility of integration outside the target area of the S1 PCR assay and/or random integration into the host genome. However, since these negative findings
are consistent with our previous studies, and the episomal form identified both suggest that AAV integration is not required for this virus’s life-cycle.

**Persistence of wild-type AAV genomes as episomes.** Given the high copy number within many tissues, direct visualization (without PCR amplification) using Southern blot confirmed the presence of the proviral episomal AAV genomes. Linear rolling circle amplification (LRCA) was then used to successfully amplify circular AAV genomes. LRCA products are linear concatemeric repeats of the episomal template, which provides a powerful method to characterize the complete AAV genome (including difficult ITR sequences). This enabled the isolation of molecular clones, of which several upon transfection into adenovirus infected cells in culture, yielded infectious progeny. Significantly, an analysis of the cognate ITR sequences of infectious clones revealed the presence of a complete double-D structure flanked by rep and cap coding sequences. Significantly, this ITR element has been observed in our human isolates and has been shown to be the sole cis sequence required for a productive AAV infection (in the presence of the required helper virus). A circular episome with this structure can be resolved to yield a linear wt AAV closed-end genome that is capable of entering the standard AAV replication pathway (in the presence of a helper co-infection). Moreover, an alternative AAV replication pathway involving circular genomes (cAAV) has been documented (131, 132) and the circularization point of cAAV also appears to yield a double-D ITR element. Combined, these data suggest that circularization in vivo may be the preferred molecular structure of AAV genomes to latently persist in non-productive tissues. Interestingly, several infectious clones also showed an in vitro preference for cultured cells derived from monkey cells (Vero) compared to human cells (HeLa). This
observation suggests the possibility of some level of host adaptation for simian isolates, which may translate to increased yields of rAAV using NHP isolates when using cells of simian origin.

In conclusion, we have shown that following naturally acquired infection, AAV DNA persists as circular episomes in non-human primate tissues without detectable AAVS1 integration even in cells harboring high levels of viral genomes. The fact that in these tissues integration was not occurring by an effective and targeted integration mechanism (rep/AAVS1) even at high intracellular loads argues that the primate host has effective means to circularize (or degrade) high viral inputs (keeping in mind that in muscle and even neurons, copies per cell of 2-20 are considered very high for rAAV high dose injection). These findings are consistent with the circular episomal forms detected following in vivo administration of rAAV (43, 44). Since AAV and vector genomes share only the ITR, it is logical to conclude that the ITR drives the circularization process by interaction with the host DNA repair machinery. The ability of this ubiquitous virus to persist in a silent state within the primate host for extended periods of time (years to even decades) without apparent toxicity or pathogenesis even at relatively high intracellular loads argues that this virus is well suited for use as a preferred vector for human clinical applications.
CHAPTER 6

DIRECTED EVOLUTION OF ADENO-ASSOCIATED VIRUS
IN PRIMARY HUMAN CELLS

INTRODUCTION

Recent discoveries of numerous and diverse adeno-associated virus serotypes has led to the development of new delivery vectors with enhanced properties (59, 60, 108). Despite these advances, remaining challenges in rAAV vector development still include evasion from neutralizing antibodies and inefficient transduction of certain target cell populations. AAV works very well as a vector system in instances such as transferring genes to muscle, vasculature, liver or brain, but it has been less than effective in the hematopoietic system and in the lung (8, 52, 86, 144, 201, 203). Most, if not all, of these problems can be attributed to the viral capsid proteins that are responsible for cell entry, intra-cellular trafficking, and viral uncoating. Because these problems arise from the capsid proteins, the capsid must be modified to overcome them. Site-directed and insertional mutagenesis of the AAV2 cap genes demonstrated that the viral capsid can tolerate modifications and maintain its infectious properties (75, 148). Furthermore, peptide insertions into specific capsid regions have been successfully applied to generate rAAV2 vectors with enhanced gene transfer capabilities (64, 128, 169). However, the
molecular basis of some viral properties, such as virus-cell interactions (29, 58, 139), are unlikely to be efficiently modulated by site-directed capsid modifications even if the AAV structures are available (205). Therefore, a high-throughput approach would aid the design of vectors with enhanced functions.

The ability to successfully transfer genetic material into different cell types with rAAV vectors has created new prospects for the treatments of a variety of genetic diseases (4). Of the many genetic diseases affecting humans, those, which involve the hematopoietic system, appear to be amenable targets for gene therapy (192). Recent advances in the field of hematopoiesis have facilitated the development of methods to identify and isolate hematopoietic stem cells (HSC) (11). Thus, it is possible to target genetic diseases that affect one or more derivatives of HSCs, including the lymphocyte diseases such as SCID, ADA deficiency and WAS that exist in populations of humans worldwide. However, controversies abound with reference to the efficacy of these vectors in transducing primary HSCs. For human HSCs, the following three sets of results have been reported. First, primary HSCs are impervious to AAV-mediated transduction (3). Second, rAAV vectors can transduce these cells, but only at extremely high vector:cell ratios (72, 114, 137). Lastly, successful transduction of primary HSCs by rAAV vectors can be achieved at relatively low vector:cell ratios under specific conditions (27, 109, 144, 214).

In the lung, Phase I clinical trials with rAAV vectors have shown safe delivery of the CFTR gene into epithelial cells, albeit with very low levels of gene transfer (191). The reason for the low performance of this vector system in lungs is not fully understood. Recently, data suggested that AAV-mediated gene delivery to the lung is blocked at least
at 2 key steps. First, the receptors that the virus uses to enter cells are not expressed on
the surface of the cells that face the inside of the airway, where the virus is administered
and where the CFTR gene must be expressed to be therapeutically beneficial. Instead,
the receptors are expressed on the wrong side of the cells, so the virus does not attach to
the cells. Second, that if the virus does get into the cells, it does not make it into the
nucleus very efficiently (46).

In the present study, we set out to modify the AAV capsid proteins \textit{in vitro}, and
then select for vectors with specific properties. Our approach to modify these capsid
proteins was genetic manipulation via DNA shuffling. This powerful method for \textit{in vitro}
recombination has proven effective for directed evolution of genetic mutations to
generate novel proteins (106, 213). Shuffling \textit{cap} genes from AAV1, hu.T88, and AAV3
isolates generated three AAV cap ORF genetic libraries. The libraries were then mixed
and packaged into virions. The resulting AAV particle library was crudely purified from
cellular debris, titered, and used to directly infect three primary human cell types: (1)
bone marrow derived CD34+ hematopoietic stem cells (HSC); (2) skeletal muscle
myoblasts (SMM); and, (3) bronchial epithelial cells (BEC). Since the majority of
persistent AAV DNA becomes double-stranded and circular, we used a novel linear
rolling circle amplification (LRCA) method to amplify the viral DNA. LRCA DNA can
then be directly packaged into virions and used for another round of selection. After
several rounds of selection, the novel capsid sequences can then be cloned into a \textit{rep/cap}
helper plasmid for rAAV vector production. The approach outlined will further aid
identifying novel third generation vectors that continue to expand \textit{in vivo} treatment
opportunities.
MATERIALS AND METHODS

Cell propagation and viruses. 293 human embryonic kidney cells were grown in EMEM supplemented with 10% fetal bovine serum (FBS) and penicillin and streptomycin (pen/strep). 293 cells were purchased from the American Type Culture Collection (Rockville, MD). CD34+ hematopoietic cells (HSC) were obtained from the Human Stem Cell Core at the University of Pennsylvania under an IRB approved protocol and were maintained in IMDM supplemented with 10% FBS and pen/strep. Skeletal muscle myoblasts (SMM) were grown in Skeletal Muscle Cell Basal Medium (SkBM-2) without L-Glutamine (Clonetics). Bronchial epithelial cells (BEC) were grown in Bronchial Epithelial Cell Basal Medium (BEBM) (Clonetics). C12 cells were grown in DMEM supplemented with 10% FBS, pen/strep, and G418 (700ug/ml). Huh-7 cells were grown in 10% DMEM supplemented with 10% FBS, and pen/strep. Huh-7 cells were provided from Dr. Guang-Ping Gao (University of Pennsylvania).

Replication-competent AAV (rcAAV) plasmid construction. To accomplish this, a 2.6-kb fragment was generated from 25 ng of pAV2 (100) as template by PCR using: (AAV2Rep1860for) 5’ - CAGCCATCGACGTCAGACGC – 3’; and (AAV2Cap4526-rev) 5’ – GAATTCTCTAGATGATTAACCCGCCATGCTAC – 3’ as forward and reverse primers, respectively. Primer AAV2Cap4562rev was designed with an XbaI restriction site for subsequent cloning. Reaction components for both rounds were 400 nM of each primer, 400 nM dNTP, 1 unit Herculase polymerase (Stratagene), 1X Herculase reaction buffer in a final volume of 50 µl. PCR cycling conditions were: 1 cycle at 92°C for 2 min; 30 cycles at 92°C for 30 sec, 61.5°C for 30 sec, and 72°C for
2.5 min; followed by a 10 min extension step at 72° C. Both this fragment and the cloning vector pBluescript II ks+ (pBS) were digested with HindIII and XbaI, and the products were ligated to create a 5.6-kb plasmid named pBS-pAV2CAP. This plasmid was created so that the cap gene could be modified outside of an unstable ITR-containing plasmid. pBS-pAV2CAP was then mutated to contain unique sites: PacI and NotI, using Multi-Site Mutagenesis Kit according to manufacturer’s instruction (Stratagene, Inc.). Primers used for the mutagenesis were as follows: (AAV2Swal/PacI) 5’ – GGATGACT-GCATCTTTGAACAAATAATGTTAATTAACAGGTATGGCTGCCGATGG – 3’; (AAV2NotICapUTR) 5’ – CGTATTTCTTTCTTTATCTAGTTGCGGCGCTACGTAGATAAGTAGCATGGCGGG – 3’. This plasmid (pBS-pAV2CAP-PacI/NotI) and the pAV2 plasmid were then digested with HindIII and SnaBI and the products were ligated together to create an 8.8-kb rcAAV packaging plasmid, pAV2pn4, into which the resulting shuffled cap genes could be inserted after PacI/NotI digestion.

**DNA shuffling.** An AAV cap open reading frame (ORF) genetic library was generated by DNA shuffling of cap genes from AAV1, hu.T88, and AAV3. To do this, a 2.2-kb fragment containing the complete cap ORF from each of the three mentioned serotypes were amplified by PCR in the following way: AAV1 (AAV1SpePacCapfor) 5’ – GAATTCACTAGTTTAAATTAACAGGTATGGCTGCCGATGGTTATC – 3’; (AAV1NotCapUTRrev) 5’ – GAATTCCGGCGCAACCGATAAGATAAGAAGGAGAACGG – 3’; hu.T88 (T88SpePacCapfor) 5’ – GAATTCACTAGTTTAAATTAACAGGTATGGCTGCCGATGGTTATC – 3’; (T88NotCapUTRrev) 5’ – GAATTCCGGCGCAACCGATAAGATAAGAAGGAGAACGG – 3’; AAV3 (AAV3SpePacCapfor) 5’ – GAATTCACTAGTTTAAATTAACAGGTATGGCTGCCGATGGTTATC – 3’;
(AAV3NotcapUTRrev) 5’ – GAATTCGCGGCCGCAACAAGATAAAGATAAAGAA-GTGC – 3’ as forward and reverse primers respectively. Each forward primer included both a SpeI and PacI restriction site for cloning. Each reverse primer included a NotI site for cloning. Reaction components for both rounds were 400 nM of each primer, 400 nM dNTP, 1 unit Herculase polymerase (Stratagene), 1X Herculase reaction buffer in a final volume of 50 µl. PCR cycling conditions were: 1 cycle at 95°C for 2 min; 30 cycles at 95°C for 30 sec, 61.5°C for 30 sec, and 72°C for 2.5 min; followed by a 10 min extension step at 72°C. Each individual fragment and pBS were cut with SpeI and NotI and ligated together to create pBS-AAV1CAP, pBS-T88CAP, and pBS-AAV3CAP plasmids. Each plasmid (500µg) was then sent to Maxygen, Inc. Maxygen scientists then cut each plasmid with PacI and NotI and subjected the cap genes to their proprietary DNASHuffling process. The resulting novel genes were then ligated together with the rcAAV plasmid pAV2pn4 and shipped back to CHOP.

**rcAAV particle library generation.** Three library ligation reactions (L18B, L18C, L18F) were then cloned to generate a viral plasmid library. rcAAV particles were then generated by calcium phosphate transfection as previously described (34). Briefly, in a 80% confluent 15-cm dish of HEK 293 cells, 30 µg of pHELP (Applied Viromics), 15 µg of pBS, and 2 or 20 ng of pAV2pn4 (B, C, or F). Viral libraries were harvested and crudely purified and the AAV genomic titer was determined by extracting viral DNA followed by quantification using real-time TaqMan PCR analysis (ABI 7500 Fast, PE Applied Biosystems) as previously described (34). The primers and probe set were selected to the rep gene that would be the same for all libraries (Repfor): 5’ – ATCGTC-ACCTCCAACACCAAC – 3’ (200 nM); (Reprev): 5’ - GCTGGTGTTCAAGGTCGTT
– 3’ (200 nM); (Repprobe): [6-FAM] 5’ - TTCCCGTCAATCACGCGCA – 3’ [6-
TAMRA-FAM] (200 nM).

Sequence analysis. All sequence analyses to confirm and characterize the viral
genetic libraries was performed as follows: viral DNA was extracted using QIAmp DNA
mini-kit according to manufacturer’s instructions (Qiagen). PCR amplification of the cap
genes as follows: 5 ul of viral DNA eluant from kit was used as template and primers
were designed to vector sequence (AAV-V2152+): 5’ - CGTATCAGAAACTGTGCTAC
– 3’; (AAV-V4568-): 5’ - GCTACTTATCTACGTAAGCG – 3’. PCR cycling conditions
were: 1 cycle at 95°C for 2 min; 30 cycles at 95°C for 30 sec, 55°C for 30 sec, and 72°C
for 2.5 min; followed by a 10 min extension step at 72°C. PCR products were cloned into
a TOPO-TA cloning vector (Invitrogen Corp.) and sequenced using BigDye Terminator
v3.1 Cycle Sequencing Kit and products are analyzed on a 3730 DNA Analyzer (Applied
Biosystems) at the CHOP Sequencing Core Laboratory.

AAV infections. Purified CD34+ cells were suspended in IMDM containing 2%
FBS. 5 x 10⁵ cells were used for each MOI with the rcAAV particle library and
uninfected cell controls. The cells were mixed with virus in a 1 ml volume and
centrifuged at 500 x g for 5 mins at room temperature and then placed in a 37°C water
bath for one hour. Following this, cells were cultured in 6-well culture plates (Costar) in
2 ml of IMDM with 30% FBS for 5 days at 37°C with 5% CO₂. Cells were harvested by
centrifugation at 500 x g for 5 mins at room temperature, washed with 10 mls of HBSS,
centrifuged again and stored at -80°C until DNA isolation. 1 x 10⁶ skeletal muscle
myoblasts (SMM) or 5 x 10⁵ bronchial epithelial cells (BEC) each in T25 flasks (Costar)
were infected with the rcAAV particle library for each MOI. Virus was added to the cells
in a 2-ml volume, incubated at 37º C for 3 hrs, and then an additional 3 mls of media was added. Cells were then cultured for 5 days at 37º C, 5% CO2 for 5 days. Cells were then washed with 10 mls of HBSS, detached with 1 ml of 1% trypsin, and harvested by centrifugation at 500 x g for 5 mins. at room temperature, and stored at -80º C until DNA isolation.

**DNA isolation.** Freshly thawed cells were digested for 15 hr in 1 ml of digestion buffer (10 mM Tris, pH 8.0; 100 mM NaCl; 0.5% SDS; 25 mM EDTA) supplemented with 2 mg/ml proteinase K at 55º C with constant agitation. DNA was subjected to three phenol/chloroform/isoamyl alcohol (25:24:1) extractions. A final chloroform extraction was performed, followed by DNA ethanol precipitation and resuspension in 10 mM Tris, pH 8.0.

**Transductions.** CD34+ cells were transduced with rAAV vectors expressing eGFP at MOIs of 5e^5. Cells from individual donors were collected and cultured in X-vivo media, 2% FCS, 1% P/S supplemented with cytokines: IL3 (20 ng/ml), IL6 (50 ng/ml), and SCF (300 mg/ml) for two days. Cells were then fed down to 5e^5 cells in 100 ul in a 24-well dish and mixed with virus. Two days post transduction the surface expression of CD34 and eGFP expression was assessed by flow cytometry. Briefly, the cells were harvested into FACS tubes, washed once with PBS, and stained using anti-CD34 APC antibody (BD) for 15 minutes at 4 degrees Celsius. The cells were washed in PBS and fixed using 2% PFA (paraformaldehyde). SMM and BEC were transduced with rAAV-eGFP vectors at MOI of 1e4. 2e^5 cells/well were mixed with vector and Ad5 virus at an MOI 20 in a 12-well dish in 500 ul of media and cultured for 2 days. 2 days post
transduction cells were analyzed under a fluorescent microscope and scored for eGFP expression.

**Acquisition and analysis.** Acquisition and analysis were performed with a FACScalibur flow cytometer (Becton Dickinson) and data was analyzed using FlowJo software. The “live” cell population was gated on FSC/SSC and then analyzed for CD34 and GFP expression.

**Southern blot analysis.** For hybridization analyses, DNA was fractionated on 0.8% agarose gels, dehydrated and subjected to in-gel hybridization or transferred to a nylon membrane using a vacuum manifold dot blot apparatus. DNA hybridization conditions were 65°C for 16 hr in buffer containing 6X SSC, 1X Denhardt's reagent, and 200 ug/ml sonicated herring sperm DNA. Rehydrated gels and nylon membranes were washed twice at 60°C in 2X SSC, 0.2% SDS for 30 min, and then twice at 60°C in 0.2X SSC, 0.2% SDS for 30 min.

**RESULTS**

**Library generation.** In collaboration with Maxygen, Inc., we generated three AAV *cap* ORF genetic libraries (L18B, L18C, L18F) by shuffling *cap* genes from related, but serologically different AAV isolates, namely: AAV1, hu.T88, and AAV3. Evaluation of these libraries by sequence analysis showed good chimerism (diversity), low parental contamination (< 1%) (Figure 6.1), and a high rate of presence of the *cap* ORF (> 75%) (Table 6.1). Cloning these fragments yielded a viral plasmid library with potentially > 10^5 independent clones (transformants). A total of 20 bacterial transformants were individually analyzed by sequence and determined to be genetically
Figure 6.1: Chimerism of L18B Library. Sequence analysis of 22 random clones after DNAshuffling shows good chimerism of shuffled cap genes. Horizontal lines indicate the origin of parental sequence. Green horizontal lines represent sequence originated from the AAV1 cap gene. Purple horizontal lines represent sequence originated from the T88 cap gene. Red horizontal lines represent sequence originated from AAV3 cap gene. Red vertical lines represent homologous sequence between the 3 different parental genes that cannot be differentiated. Complete DNA ORF sequence is represented in the left side and translated VP1 protein sequence is represented on the right side.
Figure 6.1: Chimerism of L18B library.
<table>
<thead>
<tr>
<th>Library</th>
<th>Vector No Cap Insert</th>
<th>Vector + Original Cap Insert</th>
<th>ORF (DNA)</th>
<th>ORF (Protein)</th>
<th>Addition or Deletion* W/ or w/O ORF</th>
<th>5’ or 3’ Absence or No Contig</th>
<th>Frame Shift or Early Terminator</th>
</tr>
</thead>
<tbody>
<tr>
<td>L18A</td>
<td>0</td>
<td>0</td>
<td>10 / 12</td>
<td>7 / 12</td>
<td>1 / 12</td>
<td>2 / 12</td>
<td>2 / 12</td>
</tr>
<tr>
<td>L18B</td>
<td>1 / 12</td>
<td>0</td>
<td>11 / 12</td>
<td>11 / 12</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>L18C</td>
<td>0</td>
<td>0</td>
<td>9 / 12</td>
<td>8 / 12</td>
<td>1 / 12</td>
<td>2 / 12</td>
<td>1 / 12</td>
</tr>
<tr>
<td>L18D</td>
<td>0</td>
<td>1 / 12</td>
<td>10 / 12</td>
<td>7 / 12</td>
<td>1 / 12</td>
<td>1 / 12</td>
<td>2 / 12</td>
</tr>
<tr>
<td>L18E</td>
<td>0</td>
<td>1 / 12</td>
<td>10 / 12</td>
<td>6 / 12</td>
<td>1 / 12</td>
<td>1 / 12</td>
<td>3 / 12</td>
</tr>
<tr>
<td>L18F</td>
<td>0</td>
<td>0</td>
<td>10 / 12</td>
<td>9 / 12</td>
<td>1 / 12</td>
<td>1 / 12</td>
<td>1 / 12</td>
</tr>
<tr>
<td>L18G</td>
<td>0</td>
<td>0</td>
<td>11 / 12</td>
<td>8 / 12</td>
<td>0</td>
<td>1 / 12</td>
<td>3 / 12</td>
</tr>
<tr>
<td>L18H</td>
<td>0</td>
<td>0</td>
<td>7 / 12</td>
<td>7 / 12</td>
<td>0</td>
<td>4 / 12</td>
<td>1 / 12</td>
</tr>
</tbody>
</table>

Table 6.1: Quality control of shuffled libraries.
unique. The viral plasmid library was then packaged into AAV particles as described (methods). Successful packaging was confirmed by real-time TaqMan PCR with viral titers > $10^{11}$ DNase Resistant Particles (DRP) per ml. To confirm the diversity of the viral particle library, cap gene PCR was performed on packaged viral DNA. The resulting amplicons were cloned into a pCR4-TOPO vector (Invitrogen) and sequenced and a total of 30 clones were analyzed and determined to be genetically unique. The resulting AAV library is both robust (high titer) and diverse and can now be selected for any new properties and functions.

**Selection of AAV library clones targeting CD34+ hematopoietic stem cells.**

In order to determine our ability to screen for an enriched pool of selected viruses, we tested our library on CD34+ hematopoietic cells, a historically AAV resistant primary cell line. A highly enriched population of CD34+ cells was isolated from lymphocyte derived bone marrow (LDBM) cells from multiple donors by using the autoMACS separation system and pooled together. Cells were either mock infected or infected with the rcAAV library at an MOI from $1e^5$ serially diluted down to $1e^0$, following which they were maintained in cell culture for 5 days. The cells were then harvested and total DNA was extracted for further analysis. Our lab has previously demonstrated that at least a portion of AAV DNA is converted to double-stranded episomes as early as 5 days after infection (unpublished observation). Additionally, we went on to show that this DNA, upon transfection with a helper plasmid, can be rescued and produce progeny virions (Chapters 4 and 5). To use this to our advantage, we developed an assay (Figure 6.2) using linear rolling circle amplification (LRCA) to specifically amplify double-stranded, circular AAV genomes from total cellular DNA. Total cellular DNA was initially
Figure 6.2: Schematic of linear rolling circle amplification mediated evolution.

Shuffled AAV libraries are used to infect cells, tissues, or animals. Total DNA is then isolated from infected cells and AAV episomes are amplified. To do so, total cellular DNA was digested with a restriction enzyme that does not cut within the AAV genome. The DNA was then treated with Plasmid-Safe DNase which degrades linear fragments, but leaves circular, double-stranded DNA intact. The digestion reaction served as a template for linear rolling circle amplification (LRCA) using AAV-specific primers and phi29 phage DNA polymerase. Large, linear concatameric arrays are produced following LRCA of circular AAV episomes. This DNA can then be directly transfected into 293T cells with an Adenovirus helper plasmid to generate new viral particles that will be used to infect new cells and begin the cycle again. For cloning and analysis, the linear arrays were subsequently digested into unit-length monomers by restriction enzyme digestion with an enzyme that cleaves the AAV genome once. The unit-length fragment was then cloned into an appropriate vector for further sequence analysis.
Figure 6.2: Schematic of linear rolling circle amplification mediated evolution.
digested with a restriction enzyme that was not predicted to cut within any of the predicted AAV genomes. This generated linear DNA fragments that were substrates for degradation by digestion with a novel exonuclease (Plasmid-Safe DNase) that selectively removes linear double-stranded and single-stranded linear DNA, as well as single-stranded circular DNA. Importantly, Plasmid-Safe DNase does not degrade double-stranded, circular DNA molecules, thus effectively enriching the DNA sample for free circular AAV genomes while degrading any potential integrated AAV forms. We previously demonstrated that this exonuclease is able to efficiently degrade single-strand AAV genomes and all detectable replicative forms (37, 165). After digestion, remaining intact circular AAV episomes were amplified by isothermic rolling circle amplification using phi29 phage DNA polymerase. AAV specific primers (rep gene) were used to yield AAV-specific DNA consistent with high molecular weight head-to-tail amplicons.

After this initial round of infections, DNA (100 ng) from each MOI condition was subjected to LRCA, which resulted in the amplification of high molecular weight concatameric AAV forms that upon digestion with a one-cut restriction enzyme generated the unit-length form (Figure 6.3). Control LRCA reactions using circular plasmid DNA spiked into naïve cellular DNA confirmed assay specificity and sensitivity (Figure 6.3). As shown in the Figure 6.3, Southern blot analysis showed unit-length forms visible at MOI 1e^3, real-time PCR indicated the presence of AAV DNA in the MOI 1e^4 condition. The high molecular weight concatemeric AAV DNA was then transfected into 293 HEK cells to package and produce the next round of selected viruses. After two such rounds of selection, the resulting LRCA analysis showed positive amplification of AAV DNA down to an MOI of 1e^3. Further rounds of selection did not increase our
Figure 6.3: LRCA products after one round of CD34+ cell selection. LRCA products were digested with Not I (single cutter in shuffled AAV genomes and plasmid controls), releasing unit-length head-to-tail junctions. Undigested concatameric DNA can also be observed.
Figure 6.3: LRCA products after one round of CD34+ cell selection.
sensitivity. Sequence analysis of the resulting selected viral DNA revealed two dominant cap genes. These two cap genes were then sub-cloned into a packaging plasmid and used to generate CD34+ HSC-specific rAAV-eGFP vectors (rAAVcd34.3-1 and rAAVcd34.4-1). One of the vectors (rAAVcd34.4-1) repeatedly gave low yields of recombinant virus and had to be excluded from further testing.

CD34+ cells were then analyzed for eGFP expression by FACS detection after transduction of the new rAAV-eGFP vectors. AAV1, hu.T88, and AAV3 vectors expressing eGFP were made and purified at the same time for comparison. AAV2 was also used as a positive control for this experiment. For this experiment, CD34+ cells from 4 individual donors were kept separate and each rAAV-eGFP vector was tested for transduction of eGFP at a relatively high MOI of 5e^5. Among the 4 donors, AAV2 transduced eGFP gene expression could be detected between 20 to 40% of cells analyzed. All other vectors demonstrated levels much lower than AAV2 (Table 6.2). Surprisingly, rAAVcd34.3-1 failed to demonstrate any measurable eGFP gene expression. All vectors were tested for relative fitness on C12 and Huh-7 cells and demonstrated consistent levels of eGFP expression as seen before in our lab.

Selection of AAV library clones targeting other primary human cells. We next applied this approach to primary skeletal muscle myoblasts (SMM) and primary bronchial epithelial cells (BEC). To select for AAV capsids that allow higher efficiency transduction, cells were either mock infected or infected with the rcAAV library at an MOI from 1e^5 serially diluted down to 1e^0, following which they were maintained in cell culture for 5 days. Again, AAV2 was used as a positive control for these experiments. The cells were then harvested and total DNA was extracted for LRCA analysis. After the
Table 6.2: Comparison of transduction efficiencies of rAAV-eGFP vectors on CD34+ cells.

<table>
<thead>
<tr>
<th></th>
<th>Donor 1</th>
<th>Donor 2</th>
<th>Donor 3</th>
<th>Donor 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>rAAV1</td>
<td>12%</td>
<td>9%</td>
<td>4%</td>
<td>9%</td>
</tr>
<tr>
<td>rAAV2</td>
<td>22%</td>
<td>40%</td>
<td>22%</td>
<td>26%</td>
</tr>
<tr>
<td>rAAVhu.T88</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
</tr>
<tr>
<td>rAAV3</td>
<td>2%</td>
<td>4%</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
</tr>
<tr>
<td>rAAVcd34.3-1</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
</tr>
</tbody>
</table>

Percentage of cells positive for eGFP expression by FACS detection after transduction of rAAV-eGFP vectors on CD34+ cells.

Table 6.2: Comparison of transduction efficiencies of rAAV-eGFP vectors on CD34+ cells.
first round of infections, AAV DNA could be detected at an MOI of $1 \times 10^2$ for both cell types. Further rounds of screening did not increase our sensitivity with BECs. However, after 1 more round of selection with SMMs, AAV DNA could be detected at an MOI of $1 \times 10^1$. Further rounds of screening did not increase our sensitivity with SMMs. Surprisingly, AAV2 could be consistently detected at an MOI of $1 \times 10^0$ in SMMs and $1 \times 10^1$ for BECs. Sequence analysis of the resulting final round viral DNA indicated the presence of one sequence for each cell type. These two cap genes were sub-cloned into the same packaging plasmid as previously mentioned. However, to broaden our pool of potential candidate cap genes, we sub-cloned several (7 total from SMM screening and 6 total from BEC) additional viral sequences from earlier rounds of screening.

Both cell types were then analyzed for eGFP expression by observing the cells with a fluorescent microscope. AAV1, hu.T88, and AAV3 vectors expressing eGFP were made and purified at the same time for comparison. rAAV2-eGFP was also used as a positive control for this experiment. SMM cells were transduced under the same conditions with the new rAAV-eGFP vectors and controls with an MOI $1 \times 10^4$ with a co-infection of Ad5 at an MOI of 20. As seen in table 6.3, there were several vectors that demonstrated similar transduction efficiencies with comparison to the parental controls. In fact, 3 of the clones selected from SMM cells showed an increased transduction profile when compared to the parentals. Consistently, when tested on SMM cells, all clones that were derived from SMM cells showed increased transduction profiles when compared with clones derived from BEC cells.
Table 6.3: Transduction efficiencies of rAAV-eGFP vectors on SMM cells.

<table>
<thead>
<tr>
<th>SMM clones</th>
<th>MOI 1e4</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>&lt;1%</td>
</tr>
<tr>
<td>S2</td>
<td>50%</td>
</tr>
<tr>
<td>S3</td>
<td>25%</td>
</tr>
<tr>
<td>S4</td>
<td>60%</td>
</tr>
<tr>
<td>S5</td>
<td>60%</td>
</tr>
<tr>
<td>S6</td>
<td>60%</td>
</tr>
<tr>
<td>S7</td>
<td>&lt;1%</td>
</tr>
<tr>
<td>BEC clones</td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>&lt;1%</td>
</tr>
<tr>
<td>B3</td>
<td>25%</td>
</tr>
<tr>
<td>B4</td>
<td>&lt;1%</td>
</tr>
<tr>
<td>B5</td>
<td>&lt;1%</td>
</tr>
<tr>
<td>B6</td>
<td>25%</td>
</tr>
<tr>
<td>Parentals</td>
<td></td>
</tr>
<tr>
<td>AAV1</td>
<td>25%</td>
</tr>
<tr>
<td>AAV2</td>
<td>50%</td>
</tr>
<tr>
<td>T88</td>
<td>&lt;1%</td>
</tr>
<tr>
<td>AAV3</td>
<td>&lt;1%</td>
</tr>
</tbody>
</table>

Percentage of cells scored positive for eGFP expression on SMM cells.
DISCUSSION

Nature’s approach to functional heterogeneity has been the evolution of viral capsid proteins from numerous AAV serotypes. Undoubtedly, a detailed examination of these types will yield both a better understanding of basic AAV biology and gene transfer vectors with enhanced properties. However, it would be highly synergistic to develop a complementary system that can mimic the natural evolution processes that can create novel serotypes with a desired set of specific features. Directed evolution has been used to generate enzymes with novel catalytic features, antibodies with enhanced binding affinities, and viruses with new features (40, 113, 178, 180).

There has been major progress in the development of safe, efficient and clinically relevant rAAV vectors (87, 118, 159, 203). These data notwithstanding, many challenges remain, including widespread pre-existing immunity against human serotypes, targeted and efficient delivery, and infection of non-permissive cell types (18, 19, 83, 171). These problems are not surprising, since the observed diversity among wild-type AAV serotypes are products of natural evolutionary pressures that select for survival and not performance in human gene transfer. To overcome these and other problems, the viral capsid must be modified in a way that will yield better vectors. We have accordingly developed a high-throughput approach to engineer rAAV variants with improved properties.

In the present study, three large (> 10^5 clones) AAV libraries with randomly distributed capsid mutations were generated and selected for enhanced gene transfer outputs. *In vitro* DNA shuffling yielded several libraries that showed good chimerism. Detailed sequence analyses indicated the presence of AAV DNA from each of the three
parental strands randomly distributed throughout the shuffled cap genes. The generation of mutated AAV libraries has been previously reported (113, 128), but have relied upon error-prone PCR and random point mutations from a single isolate of AAV2. As previously mentioned, AAV2 was originally isolated as a cell culture contaminant of an adenovirus cell line (5). AAV2 is the most widely appreciated serotype of AAV and has been used in the bulk of human clinical trials to date. One of the best-understood features about AAV2 is that it has been clearly shown to utilize heparin sulfate proteoglycan (HSPG) as a primary receptor. Recently, several AAV2-like isolates were identified from human clinical samples that seem to function differently that the original AAV2 strain. In fact, our lab has previously reported that most circulating (in humans) AAV2-like viruses are not predicted to bind HSPG (Chapter 2). Careful inspection of the AAV capsid sequences revealed that none of the AAV2-like sequences retained arginine residues at positions 585 and 588. These residues have been shown to be critical for (HSPG) receptor binding (89, 139). In each of our sequences, R585S and R588T substitutions were observed. From these data, we decided to include a clinical AAV2-like isolate named hu.T88. Although comparison of the tropisms of the isolate hu.T88 to existing serotypes has not been performed, it is safe to assume that hu.T88 is a clinically infectious isolate. To further expand the diversity of the library we included cap sequences from AAV1 and AAV3. AAV1 appears to be superior to other serotypes for transduction of murine muscle tissue and AAV3 vectors are superior in transducing smooth muscle cells (26, 58, 68). DNAshuffling and cloning yielded a viral plasmid library with >10^5 independent clones. The viral plasmid library was the significantly diluted and then packaged into virions. One caveat of the packaging process is the
potential for the viral genome to be packaged into a viral particle that doesn’t match. In order to try and prevent this from happening, the viral plasmid library was diluted so that potentially, only 1 plasmid copy would be taken up by 1 cell.

**Primary cell selection and vector comparisons.** We then attempted to select for rAAV vectors with enhanced properties. To do so, we developed an assay to direct the natural evolutionary processes of viral capsid proteins toward a new or often difficult target. This enabled the isolation and amplification of complete AAV genomes from infected cells. This process was dependent on the fact that the majority of AAV genomes persist as episomes after infection. In CD34+ HSCs, we were unable to screen for viral capsid proteins that demonstrated an enhanced ability to express protein (eGFP) when compared to existing vectors.

As mentioned previously, CD34+ HSCs have demonstrated very inconsistent results when transduced with rAAV vectors. Several factors exist that potentially create obstacles that AAV vectors encounter in transducing CD34+ HSCs. First, the lack of cell surface expression of HSPG, the primary receptor for AAV2, has been documented to occur in several donor preparations. One study concluded that up to 50% of normal human volunteer donors lack CD34+ HSC cell surface HSPG expression (144). Although precise levels of known AAV co-receptors have not been studied in these cells, the lack of successful binding of AAV does not lead to viral or vector entry. The lack of HSPG does not preclude binding and entry via unknown receptors and possible altered pathways. Furthermore, in these studies, vectors were derived from a non-HSPG binding (hu.T88) isolate of AAV whose cellular receptor is unknown. Second, cellular FKBP52, phosphorylated at both tyrosine and/or serine/threonine residues, specifically interacts
with the single-stranded D sequence within the AAV ITRs and blocks second-strand DNA synthesis (145). It is likely that the phosphorylation status of FKBP52 in cells from various donors directly impacts on the transgene expression in HSCs. Finally, several individual donor cells were difficult to maintain their viability in cell culture suggesting mechanical or biological stress outside the host. Combined, these data suggest that it is difficult to control for variables from experiment to experiment that can lead to inconsistent results. To avoid most of these complications, cells from multiple donors were collected and pooled together to attempt to mimic a heterogeneous population.

In BEC and SMM cells we were able to select for vectors that were slightly better than the original parental viruses. However, the new vectors were unable to out-perform AAV2 based vectors. In fact, rAAV2 vectors were better in SMM cells when compared to rAAV1 vectors. This data is in direct contrast to in vivo studies that clearly show rAAV1 vectors are superior to rAAV2 in transducing muscle tissue (26, 108, 202). It is important to note that both the BEC and SMM are still primary-derived cell lines that have still adapted to growing in cell culture. It is worth re-stating that AAV2 has been carried in cell culture for many years and has undoubtedly undergone adaptations of its own to maintain its usefulness.

In spite of the many questions that remain unanswered from the previous work, it was clear that AAV has the ability to undergo dramatic changes and that there can be success in generating novel phenotypes. The problem is that you get exactly what you screen for, and it appears that if you perform the screen on in vitro targets, then you create a phenotype that may only be useful in cell culture. In order to obtain a phenotype that would demonstrate its effectiveness in vivo, the selection model would perhaps need
to be done in an animal model such as a mouse. Using this type of in vivo screening would allow for selection in multiple targets (liver, muscle, brain, spleen, etc.) that would potentially be useful in identical tissues in other animals.
CHAPTER 7
GENERAL CONCLUSIONS

The research presented in this dissertation directly questioned our beginning hypothesis that targeted integration would occur in vivo. To summarize, in chapter 2, we extended our understanding of wild-type AAV infection in humans by characterizing AAV genomes directly out of freshly acquired human tissues. Importantly, this demonstrated our ability to identify AAV sequences in vivo and begin to characterize the natural infection processes involved. These data showed that AAV genomes could be detected in multiple tissues (present in up to 7% of tissues analyzed) beyond the probable portal of entry (oronasal). In addition, while the DNA sequences recovered were similar to the prototype AAV2 and to each other, 2 significant observations emerged from the sequence analyses. First, the majority of the observed amino acid substitutions were located in previously defined hypervariable regions (56) predicted to be exposed on the surface of the virion. Second, and perhaps most intriguing, was the discovery that none of the 7 AAV2-like sequences derived from tonsils and adenoids were predicted to bind HSPG.

In chapter 3, we undertook an evaluation of the molecular forms of the AAV DNA in these tissues using 3 distinct techniques. In contrast to the prevailing hypothesis, we were unable to identify wild-type AAV integration at the AAVS1 locus using a
sensitive S1-specific PCR assay. When we broadened our search for viral-host sequence junctions using a modified linear amplification mediated PCR assay (LAM-PCR), we found a single integration event on chromosome 1, but again were unable to find any AAVS1 insertions. Finally, using a novel linear rolling circle amplification assay, we showed that the majority of the AAV genome sequences in these tissues were extrachromosomal circles. To reflect these findings, we amended our hypothesis to state that the majority of AAV genomes in vivo were persisting as extra-chromosomal double-stranded episomes.

In chapter 4, to investigate this hypothesis, we developed a novel method (sequence-specific linear rolling circle amplification, SSLRCA) that allowed us to amplify extra-chromosomal AAV genomes directly from tissue DNA. We went on to show that these amplified products represent intact, biologically active AAV genomes. Additionally, based on ITR sequence analysis, we concluded that infectious molecular clones contained a complete double-D ITR element. Conversely, limited ITR sequence analysis of non-infectious clones suggested that deletion of internal ITR regions was responsible for the lack of infectivity since these clones expressed the expected rep and cap proteins.

In chapter 5 we conducted a follow-up study where we compared the structural and functional properties of latent wild-type AAV in a different primate host (compared to humans in chapter 2 and 3), the rhesus macaque, to increase both the breadth of the analysis and determine whether low-level AAV integration was occurring in tissues harboring substantially greater levels of endogenous AAV. This analysis identified large numbers of positive AAV tissues and at substantially increased AAV copies per nuclei in
several macaque tissues. Increased AAV genome levels facilitated a more thorough genome analysis than our initial human study and further added to the collective sample size. Significantly AAV persistence was found in extra-chromosomal forms, with no evidence for targeted AAV integration observed.

In chapter 6 we did not directly address the hypothesis as outlined, but instead, set out to create novel rAAV vectors that would be useful in the clinical setting. Although not directly applied to the hypothesis, the methodology for the experiments in this chapter stemmed from the knowledge generated from previous chapters. In summary, we developed a novel method (linear rolling circle amplification mediated evolution) that allowed for the directed evolution of AAV genomes.
BIBLIOGRAPHY


