A Capillary Electrophoresis Sequencing Method for the Identification of Mutations in the Inverted Terminal Repeats of Adeno-Associated Virus

Cameron Mroske,¹ Hector Rivera,² Taihra Ul-Hasan,³ Saswati Chatterjee,³ and K.K. Wong¹

Abstract

Inverted terminal repeat (ITR) integrity is critical for the replication, packaging, and transduction of recombinant adeno-associated virus (rAAV), a promising gene therapy vector. Because AAV ITRs possess 70% GC content and are palindromic, they are notoriously difficult to sequence. The purpose of this work was to develop a reliable ITR sequencing method. The ITRs of two molecular clones of AAV2, pTZAAV and pAV2, were (1) sequenced directly from plasmid DNA in the presence of denaturant (direct sequencing method, DSM) or (2) first amplified in a reaction in which 7-deaza-dGTP was substituted for dGTP and the resultant amplification product sequenced (amplification sequencing method, ASM). The DSM and ASM techniques yielded clear chromatograms, read through the ITR hairpin, and revealed hitherto unreported mutations in each ITR. pTZAAV and pAV2 possess identical mutations at the upstream *MscI* site of the 5' ITR (T>G, nt 2) and the downstream *MscI* site of the 3' ITR (del. nt 4672–4679). The chromatograms for pAV2 also revealed that the ITRs of this construct were arranged in a FLOP/FLOP orientation. In addition, the DSM was successfully employed to recover ITR–chromosomal junction sequences from a variety of rAAV-transduced tissue types. Both the DSM and ASM can be employed to sequence through the AAV ITR hairpin, and both techniques reliably detect mutations in the ITR. Because the DSM and ASM offer a way to verify ITR integrity, they constitute powerful tools for the process development of rAAV gene therapy.

Introduction

THE CLINICAL SUCCESS of adeno-associated virus (AAV)-mediated gene therapy for the treatment of hemophilia B demonstrates that AAV-based vectors have the potential to be a transformative medicine (Nathwani *et al.*, 2011). AAV is a small, nonenveloped single-stranded DNA virus that is considered attractive for therapeutic gene transfer because it is nonpathogenic and occurs naturally in a large number of serotypes that may confer selective tropism to specific tissues (Michelfelder and Trepel, 2009). In addition to hemophilia B, AAV gene therapy vectors have been used successfully to treat Leber's congenital amaurosis (LCA) (Bainbridge *et al.*, 2008; Cideciyan *et al.*, 2008; Hauswirth *et al.*, 2008; Maguire *et al.*, 2008; Simonelli *et al.*, 2010) and Parkinson's disease (Feigin *et al.*, 2007; Kaplitt *et al.*, 2007; LeWitt *et al.*, 2011). In all, there are currently 38 AAV protocols approved by the Recombinant DNA Advisory Committee (National Institutes of Health, Bethesda, MD) and U.S. Food and Drug Adminis-

tration (Silver Spring, MD) for the treatment of diseases ranging from cystic fibrosis to rheumatoid arthritis (Daya and Berns, 2008; Mingozzi and High, 2011).

Central to the biology of AAV is a 145-base structure located at each end of the viral genome: the inverted terminal repeat or ITR. Thermodynamically, the AAV ITR is extremely stable due to a GC content of 70% and a palindrome made up of 125 nucleotides that folds back to form a T-shaped hairpin (Fig. 1A). The ITRs can adopt orientations at the ends of the AAV genome that are described as either FLIP/FLOP, FLOP/FLIP, FLIP/FLIP, or FLOP/FLOP. Figure 1B depicts the ITRs in a FLIP/FLOP orientation whereby the BC sequence motifs form mirror images of one another across a horizontal line of symmetry. According to the current model of AAV replication, the ITR plays a key role in virus production because it serves not only as the primer for second-strand synthesis and DNA replication, but also contains binding sites for the AAV Rep protein (RBE and RBE'), and possesses a terminal resolution site (*trs*) (Fig. 1A).

¹Division of Hematology and Division of Stem Cell Transplantation, City of Hope, Duarte, CA 91010.

²DNA Sequencing/Solexa Core, Beckman Research Institute, City of Hope, Duarte, CA 91010.

³AAV Laboratory, Division of Virology, Beckman Research Institute, City of Hope, Duarte, CA 91010.

FIG. 1. Inverted terminal repeat (ITR) structure. **(A)** Secondary structure of the AAV-2 3' ITR (GenBank accession number NC_001401). Shown are the Rep-binding elements RBE (GAGCGAGC GAGCGCGC) and RBE' (CTTTG), as well as the terminal resolution site (*trs*, GTTGG). **(B)** Representation of the AAV-2 genome (diagram not according to scale). In the case of wild-type AAV-2 (GenBank accession number NC_001401), the 5' ITR shown at right adopts a FLIP configuration whereas the 3' ITR at left adopts a FLOP orientation, so that the BC sequence motifs, as well as the *Sma*I restriction sites, mirror each other across a horizontal line of symmetry (red line). **(C)** Schematic of a typical rAAV cassette containing the elements necessary to drive transgene expression.

Importantly, rAAV vectors used for gene therapy do not contain the endogenous *rep* and *cap* genes but instead possess a therapeutic transgene expression cassette, located between the ITRs, which act as *cis* signals for packaging and transduction (Fig. 1C). Because the presence of intact ITRs is critical for the production of high-titer, fully functional rAAV, and given the fact that it is not uncommon for each ITR to mutate when present in plasmid form, it is critically important to develop a rapid, low-cost, high-resolution protocol to verify ITR sequence content.

Traditionally, ITR integrity and orientation have been verified through restriction analysis; however, this technique has several drawbacks. Restriction mapping provides only limited information regarding the nature of the mutation, and interpretation of the restriction pattern, which is often equivocal, can be time-consuming. Thus, without accurate knowledge of the nature and extent of sequence change (point mutation vs. microdeletion), it may be difficult to make an informed decision as to whether a particular clone will be suitable for vector production. For these reasons, one or more methods to sequence through the AAV ITR would be advantageous. It has been asserted that the only way to sequence through the hairpin structure of the AAV ITR is to

perform a two-step protocol whereby the hairpin is amplified in the presence of 7-deaza-dGTP before sequencing (Kieleczawa, 2006). Not only did we want to substantiate this claim, because of the paucity of ITR sequencing chromatograms in the literature, we also wished to have a more direct method for sequencing the AAV ITR. Moreover, our objective has been to develop a rapid, consistent, and accurate method to sequence each AAV ITR, using contemporary fluorescence capillary electrophoresis sequencing technology.

By taking advantage of developments in amplification and sequencing chemistry for high GC content DNA (Henke *et al.*, 1997; Kieleczawa, 2006; Musso *et al.*, 2006), we have been able to develop/adapt two independent methods to sequence the AAV ITR. Herein we show that direct sequencing of each ITR on AAV plasmids (direct sequencing method, or DSM), and sequencing of an amplification product containing a complete 7-deaza-dGTP substitution for dGTP (amplification sequencing method, or ASM), produce noise-free chromatograms that reveal mutations in concordance with the respective restriction patterns. In contrast to a previous report (Kieleczawa, 2006), we find that the DSM reads through the ITR hairpin region as well as the ASM. Using both the DSM and ASM, we demonstrate that

the infectious clone pTZAAV (Chatterjee, 1992), as well as clone pAV2 from the American Type Culture Collection (ATCC, Manassas, VA) (Laughlin *et al.*, 1983), possess identical mutations in the upstream *MscI* site of the 5' ITR (T>G, nt 2) and the downstream *MscI* site of the 3' ITR (Del. GTGGC/A, nt 4672–4679). In contrast to pTZAAV, we demonstrate that the ITRs of pAV2 are arranged in a FLOP/FLOP orientation. As well, we have successfully employed the DSM to retrieve chromosomal junction sequences from mouse liver transduced with a novel rAAV vector, AAV-HSC15 (Aravind *et al.*, 2012). We provide an example in which the DSM clearly shows that rAAV integrated into murine chromosome 2 at position 134373802 in the minus orientation. Future experiments will be necessary to determine how the particular ITR mutations identified in this study affect AAV packaging efficiency. Nevertheless, the direct sequencing and amplification sequencing methods are accurate and consistent in verifying ITR integrity and should prove useful in the process of large-scale rAAV vector production for gene therapy.

Materials and Methods

Plasmid DNA

Plasmid DNA for construct pTZAAV was isolated from a 200-ml overnight *Escherichia coli* DH5 α culture, using a Qiagen plasmid maxi kit (cat. no. 12163; Qiagen, Valencia, CA). Construct pAV2 was purchased from the ATCC (no. 37216) as a lyophilized *E. coli* HB101 pellet and reconstituted according to the manufacturer's instructions. In brief, the bacterial pellet was dissolved in 5 ml of Luria broth containing ampicillin (50 μ g/ml) and grown overnight at 37°C. The next day, 250 μ l of overnight culture was seeded into 200 ml of fresh Luria broth–ampicillin and grown for 20 hr at 37°C in a shaker incubator. Plasmid DNA was then isolated according to the Qiagen maxiprep protocol. For both pTZAAV and pAV2, isolated plasmid DNA was dissolved in sterile Tris–EDTA buffer.

Primers

Primers used for PCR amplification and cycle sequencing are listed in Table 1. DNA primers were chemically synthesized by Integrated DNA Technologies (San Diego, CA) and purified by standard desalting.

Restriction analysis

Restriction enzymes were purchased from New England BioLabs (Ipswich, MA). One microgram of plasmid DNA was digested with *MscI* at 37°C, or with *MscI* at 37°C followed by *SmaI* at 25°C. After digestion, about 400 ng of enzyme-treated DNA was resolved on a 4% agarose–0.5 $\times$ TBE gel (TBE: 45 mM Tris, 45 mM borate, 1 mM EDTA) and imaged by staining in 3 $\times$ GelRed solution (cat. no. 41002; Biotium, Hayward, CA) for 1 hr.

PCR amplification: ASM

ASM templates were generated by amplifying the 5' and 3' ITRs of the AAV2 genome carried on plasmids pTZAAV and pAV2 in a manner such that 7-deaza-dGTP was completely substituted for dGTP. 7-Deaza-dGTP solution was purchased (cat. no. 10988537001; Roche Applied Science, Man-

nheim, Germany) and mixed at an equimolar concentration with dATP, dTTP, and dCTP in order to constitute a 7-deaza-dGTP–dNTP stock. Each amplification reaction consisted of 20 ng of plasmid template, 1 $\times$ cloned *Pfu* (*Pyrococcus furiosus*) reaction buffer, forward and reverse primers at a final concentration 0.4 μ M, 7-deaza-dGTP–dNTP at a final concentration of 250 μ M, and 2.5 U of *PfuTurbo* DNA polymerase (cat. no. 600250; Stratagene/Agilent Technologies, La Jolla, CA). Reactions were performed in a total volume of 25 μ l on a Bio-Rad MyCycler thermocycler (cat. no. 170-9701; Bio-Rad, Hercules, CA) according to the following program: 96°C for 2 min (1 cycle)/96°C for 45 sec, 50°C for 45 sec, 72°C for 1 min (30 cycles)/72°C for 10 min (1 cycle)/4°C (hold). Resultant amplicons were purified with a Qiagen QIAquick PCR purification kit (cat. no. 28104). DNA was eluted in 10 mM Tris, pH 8.5, and quantified on a NanoDrop 1000 spectrophotometer (Thermo Scientific, Wilmington, DE).

Cycle sequencing: ASM and DSM

Before cycle sequencing, 2 ng of purified amplification product (ASM) or 200 ng of plasmid DNA (DSM) was diluted to a volume of 7 μ l in 10 mM Tris, pH 8.0, and denatured at 98°C for 10 min before being immediately snap-cooled on ice. The following reagents were then added to each denatured sample for a total reaction volume of 14.5 μ l per sample: 1 μ l of 5 μ M sequencing primer, 2 μ l of BigDye Terminator version 3.1, 1 μ l of BigDye Terminator dGTP version 3.0, 1.5 μ l of 5 $\times$ DNA sequencing buffer (Applied Biosystems/Life Technologies, Foster City, CA), and 2 μ l of 5 M betaine solution (Sigma-Aldrich, St. Louis, MO). Samples were then subjected to cycle sequencing in a Veriti thermocycler (Applied Biosystems/Life Technologies) according to the following program: 96°C for 1 min (1 cycle)/96°C for 10 sec, 50°C for 5 sec, 60°C for 75 sec (25 cycles)/4°C (hold). After thermocycling, samples were purified with a BigDye Xterminator purification kit and electrophoresed on a 3730 DNA analyzer (Applied Biosystems/Life Technologies) equipped with a 50-cm capillary array filled with POP-7 polymer and 1 $\times$ capillary electrophoresis buffer. Purified samples were injected electrokinetically at 1.2 kV and electrophoresed at 8.5 kV for 1.56 hr with a current stability of 3.0 μ A and an oven temperature of 60°C.

Determination of AAV–chromosomal junction sequences

Genomic DNA was isolated from the livers of mice that had been injected with the novel rAAV vector, AAV-HSC15. Liver tissue was suspended in digestion buffer (10 mM NaCl, 10 mM Tris–HCl [pH 8.0], 25 mM EDTA [pH 8.0]) along with RNase (Roche, Indianapolis, IN) and incubated on a rotating rack for 2 hr at 37°C. Sodium dodecyl sulfate (SDS) and proteinase K were then added to final concentrations of 0.5% and 0.1 mg/ml, respectively, whereupon samples were further incubated for 16 hr at 56°C. Genomic DNA was isolated from the reactions through phenol–chloroform/ethanol purification and digested with either *RsaI* or *DraI* before undergoing an additional round of purification. Using T4 ligase, GenomeWalker adaptors (Clontech, Mountain View, CA) were ligated to the digested product for 16 hr at 16°C and integration junctions were amplified through nested PCR using the following primer pairs: round 1 amplification,

TABLE 1. PRIMERS USED FOR DIRECT SEQUENCING METHOD AND AMPLIFICATION SEQUENCING METHOD

Name	Sequence	Distance to ITR (base pairs)	Prime-binding site	Technique
5'-ITR-pTZAACV-For	GCGATTAAGTTGGGTAACGCCAG	88	pTZAACV plasmid sequence	DSM ASM
5'-ITR-AAV2-Rev	CTTAAATACCCAGCGTGACCACATG	92	AAV-2 genome	DSM ASM
3'-ITR-AAV2-For	CGTTTCAGTTGAACTTTGGTCTCTGC	69	AAV-2 genome	ASM
3'-ITR-pTZAACV-Rev	CGGATAACAATTTCACACAGGAAACAG	68	pTZAACV plasmid sequence	ASM
nu-3'-ITR-AAV2-For	CGTGTATTCAGAGCCTCGCCCCATTGGC	151	AAV-2 genome	DSM
nu-3'-ITR-pTZAACV-Rev	GAAAGCGGGCAGTGAGCGCAACGC	179	pTZAACV plasmid sequence	DSM
pAV2-5'-ITR-Blac-For	GCGTATCACGAGGCCCTTTCGTCTTC	76	pAV2 plasmid sequence	DSM ASM
pAV2-3'-ITR-Rev	CGATCTTCCCCATCGGTGATGTCGGC	69	pAV2 plasmid sequence	DSM
pAV2-3'-ITR-Rev3	CGAAACAAGCGCTCATGAGCCCGAAGTG	102	pAV2 plasmid sequence	ASM
M13F	GTAAAACGACGGCCAG	NA	pcDNA2.1-TOPO plasmid sequence	DSM
M13R	CAGGAAACAGCTATGAC	NA	pcDNA2.1-TOPO plasmid sequence	DSM

ASM, amplification sequencing method; DSM, direct sequencing method; ITR, inverted terminal repeat; NA, not applicable.

AAV-specific primer AGGAACCCCTAGTGATGGAGTTGGCCAC or TACGTAGATAAGTAGCATGGCGGGTTA and adaptor-specific primer GTAATACGACTCACTATAGGGC; round 2 amplification, AAV-specific primer TGATGAGATTGGCCACTCCCTCTCTG or AGTAGCATGGCGGGTTAATCATTAAT and adaptor-specific primer ACTATAGGGCACGCGTGGT. Amplicons were then subjected to agarose gel electrophoresis and the DNA corresponding to various bands was purified with the QIAquick gel purification kit (Qiagen). Purified, nested PCR amplicons were then subcloned into pcDNA2.1-TOPO, using a TOPO TA cloning kit (Life Technologies, Carlsbad, CA). DNA was isolated from overnight cultures according to a standard miniprep protocol and sequenced by the DSM with primer M13F (TGTAACGACGGCCAGT) and primer M13R (GGAACAGCTATGACCATGAT). AAV-chromosomal junction sites were identified with the aid of the following bioinformatic tools: BLAT (University of California Santa Cruz, Santa Cruz, CA), BLAST (National Center for Biotechnology Information, Bethesda, MD), and Ensembl (EMBL-EBI/Wellcome Trust Sanger Institute, Hinxton, UK).

Sequence analysis

Chromatograms generated by the ASM and DSM were aligned against AAV2 reference sequence NC_001401 (GenBank), using Sequencher version 4.8 (Gene Codes, Ann Arbor, MI).

Results

Restriction analysis

Restriction analysis is commonly used to verify rAAV ITR integrity before the packaging of transducing particles. Be-

cause each ITR contains two *MscI* sites and two *SmaI* sites, a construct with intact ITRs should release a 117-bp fragment when digested with *MscI* and should generate fragments 64 and 42 bp in length when codigested with *MscI* and *SmaI* (Fig. 2A). During a routine screening of several constructs in our laboratory, we discovered that two molecular clones of AAV-2, pTZAACV and pAV2, produced non-wild-type patterns when subjected to restriction analysis. Neither construct released the 117-bp fragment when digested with *MscI*; however, pTZAACV did release the expected 64- and 42-bp fragments when codigested with *MscI* and *SmaI* (Fig. 2B, lane 2). In contrast, when pAV2 was codigested with *MscI* and *SmaI*, only a single fragment was generated, resolving at a position indicating a fragment approximately 64 bp in length (Fig. 2B, lane 4). These data suggested that (1) each construct possessed at least one mutation at an *MscI* site in each ITR, and (2) that the ITR configurations of pTZAACV and pAV2 were different.

ITR sequencing: DSM and ASM

Because restriction analysis suggested that the ITRs of constructs pTZAACV and pAV2 not only contained *MscI* site mutations but were also configured differently, we sought to investigate the extent of sequence change in these clones in order to ascertain whether the mutations were significant enough to pose a problem for AAV replication. To accomplish this, we developed/adapted two methods to determine the DNA base sequence of the 5' and 3' ITRs of AAV-2. In the first method, which we term DSM (direct sequencing method), primers were designed to bind plasmid template in the range of 60 to 200 bp upstream and downstream of each ITR. Cycle sequencing reactions were then performed using a blend of ABI sequencing reagents BigDye Terminator

FIG. 2. Restriction analysis. **(A)** Depiction of a construct possessing the full-length, wild-type AAV-2 genome. Shown are the 5' and 3' ITRs in a FLIP/FLOP orientation. Numbers indicate the size of fragments (in base pairs) generated by restriction digestion. Note that the 11-bp fragment generated by *SmaI* digestion is too small to be resolved on an agarose gel. **(B)** *MscI* and *SmaI* digestion of constructs pTZAAV and pAV2. Restriction product was resolved on a 4% agarose-0.5×TBE gel. When codigested with *MscI* and *SmaI*, pTZAAV (lane 2) resolves into two bands that migrate at 64 and 42 bp, whereas pAV2 (lane 4) resolves as a single band that migrates at approximately 64 bp. When digested with *MscI* alone, neither construct yields the expected 117-bp band (lanes 1 and 3).

version 3.1 and BigDye Terminator dGTP version 3.0 in the presence of the denaturant betaine. As can be seen in the upper chromatograms shown in Figs. 3 and 4, the DSM generates clear peaks with negligible background. Moreover, the technique clearly reveals the specific sequence changes at the upstream *MscI* site of the 5' ITR (T>G, nt 2; Figs. 3A and 4A, red arrows) and the downstream *MscI* site of the 3' ITR (Del. GTGGC/A, nt 4672–4679; Figs. 3B and 4B, red arrows). Both of these changes are present on pTZAAV and pAV2. Although it has never been officially reported, the deletion at the downstream *MscI* site of the 3' ITR has been commonly known in the field of rAAV-based gene therapy (A. Srivastava, personal communication, 2012). In addition, the DSM also reveals that the BC arm of the 5' ITR of construct pAV2 is arranged in a FLOP configuration (Fig. 4A, center panel, blue bracket).

Because the highly structured, strongly base-paired configuration of the AAV ITR poses a formidable sequencing challenge, we wished to have more than one method available to sequence this structure. Therefore, we adapted the ASM. For this method, the ITR, along with a small amount of flanking sequence, was first amplified by a proofreading polymerase (*PfuTurbo*) in a reaction in which the nucleotide analog 7-deaza-dGTP was completely substituted for dGTP. The resultant amplification product was then purified and served as a template for subsequent sequencing reactions. The same oligonucleotides used for amplification were used to prime the individual cycle sequencing reactions. Figure 5 illustrates the amplification product for the ASM used to sequence the 5' and 3' ITRs of pTZAAV. Although the inclusion of 7-deaza-dGTP reduced the yield, more than enough PCR product was obtained to perform multiple sequencing reactions. The lower chromatograms shown in Figs. 3 and 4 were generated by sequencing the purified

amplification product that spanned the 5' and 3' ITRs of constructs pTZAAV and pAV2. Importantly, not only are the ASM chromatograms noise-free, they also reveal the same mutations in constructs pTZAAV and pAV2 as the chromatograms generated by the DSM.

Usefulness of the DSM

Direct sequencing of the 28-base AAV ITR hairpin on plasmid DNA is problematic. Although it has been reported that the only way to read through this structure is to amplify it in the presence of 7-deaza-dGTP before sequencing (Kieleczawa, 2006), studies have successfully employed Maxam-Gilbert sequencing as well as bisulfite modification before fluorescence capillary electrophoresis Sanger sequencing to read through the hairpin (Lusby *et al.*, 1980; Inagaki *et al.*, 2007). Nevertheless, we wished to have a method to sequence this structure that did not rely on chemical modification or amplification before Sanger sequencing and that would be compatible with contemporary high-throughput sequencing equipment.

As exemplified in Fig. 6A, by directly sequencing plasmid DNA we can consistently read through not only the stem of the ITR, but also the highly structured hammerhead portion (BB'CC'; Fig. 1A) that contains the two *SmaI* restriction sites.

Importantly, we have successfully employed the DSM to map random rAAV integration into the genomes of transduced mouse cells as well as human CD34⁺ cells. Figure 6B shows an example of one such integration site identified by the DSM in rAAV-transduced mouse liver. The DSM unequivocally demonstrates that rAAV inserted into murine chromosome 2 at position 134373802 in the minus orientation. Furthermore, the technique clearly shows that the junction between the 3' ITR and chromosome occurs at the AA' portion of the ITR stem.

FIG. 3. pTZAAB ITR sequence. **(A)** Base sequence of the 5' AAV-2 ITR obtained by the direct sequencing method (DSM; upper chromatogram) and the amplification sequencing method (ASM; lower chromatogram). The left-hand panel reveals that the upstream MscI site is mutated (T>G, nucleotide 2, red arrows); in contrast, the downstream MscI site is intact (right-hand panel). **(B)** Sequence of the 3' AAV-2 ITR obtained by the DSM (upper chromatogram) and ASM (lower chromatogram). The left-hand panel reveals that the upstream MscI site is intact; however, at the downstream MscI site, nucleotides 4672–4679 are deleted (right-hand panel, red arrows). Chromatograms are aligned against the wild-type AAV-2 reference sequence NC_01401 (uppermost text sequence). Color images available online at www.liebertpub.com/hgfb

Discussion

In this study, we describe for the first time a reliable, accurate, and straightforward method to screen the sequence integrity of the palindromic, high GC-containing AAV ITRs. Sequencing of each of the two ITRs on either plasmid DNA (DSM), or on amplified product containing a complete 7-deaza-dGTP substitution for dGTP (ASM), produced noise-free chromatograms that revealed previously unreported mutations. As well, contrary to a previous report, we found that the DSM reads through the ITR hairpin structure with the same facility as the ASM (Kieleczawa, 2006). Using both the DSM and ASM, we demonstrate that the infectious clone pTZAAB, as well the antecedent construct pAV2, possess

identical mutations in the upstream MscI site of the 5' ITR (T>G, nt 2) and the downstream MscI site of the 3' ITR (Del. GTGGC/A, nt 4672–4679). These data support the reproducibility and accuracy of our methodology because both constructs were derived from the same recombinant AAV2 genomic material and are expected to possess identical mutations.* In addition, the sequence information obtained here reveals that the ITRs of pAV2 are arranged in a FLOP/FLOP orientation.

*pTZAAB was derived from the molecular AAV2 clone, pAV1. Although pAV1 and pAV2 contain the same AAV2 genomic substrate, pAV2 is not a derivative of pAV1 (Laughlin *et al.*, 1983).

FIG. 4. pAV2 ITR sequence. **(A)** Base sequence of the 5' AAV-2 ITR obtained by the DSM (upper chromatogram) and the ASM (lower chromatogram). The left-hand panel reveals that the upstream *MscI* site is mutated (T>G, nucleotide 2, red arrow). The center panel (blue bracket) illustrates how the BC sequence motif is arranged in a manner such that the pAV2 5' ITR adopts a FLOP orientation. The downstream 5' ITR *MscI* site is intact (right-hand panel). **(B)** Base sequence of the 3' AAV-2 ITR obtained by the DSM (upper chromatogram) and the ASM (lower chromatogram). The panel at left reveals that the upstream *MscI* site is intact; however, nucleotides 4672–4679, which encompass the downstream *MscI* site, are deleted (right-hand panel, red arrow). Chromatograms are aligned against wild-type reference sequence NC_001401 (uppermost text sequence). Color images available online at www.liebertpub.com/hgtb

More importantly, as screening tools, the DSM and ASM offer advantages over restriction analysis in that they reveal the exact sequence content of mutations and not simply their presence. Such information is important because microdeletions and other forms of ITR rearrangement may have implications for the replication, packaging, and integration of rAAV. In addition, the novel ITR sequencing methods described here are more cost-effective than restriction enzyme analysis because they provide a greater amount of biological information, which is important for gene transfer function.

On average, the DSM and ASM generated similar read lengths. However, this attribute could be influenced by specific sequence flanking the ITR. For example, in several instances the DSM was able to read through a 300-bp insert cloned immediately upstream of an ITR, and continued on to read well past the *SmaI* sites of the BB' portion of the ITR. Even though our data indicate that the DSM and ASM can be used interchangeably, Sanger sequencing can often be idiosyncratic. In the event that primer design is constrained, or that a particular sequence context causes one of the methods to underperform, having the redundancy of both the DSM

FIG. 5. Generation of sequencing template for the ASM. **(A)** The 5' ITR of pTZAAG, including the flanking sequence, was amplified and 2 μ l of a 25- μ l reaction was resolved on a 2% agarose gel. Sample order: lane 1, no-template control; lane 2, 5' ITR. The expected product size is 373 bp. **(B)** The 3' ITR of pTZAAG, including the flanking sequence, was amplified and 2 μ l of a 25- μ l reaction was resolved on a 2% agarose gel. Sample order: lane 1, no-template control; lane 2, 3' ITR. The expected product size is 335 bp.

FIG. 6. Usefulness of the DSM. **(A)** The DSM reads through the AAV ITR hairpin. Shown is the alignment and corresponding chromatogram of a single read through the stem and hammerhead portion of the 5' ITR of a molecular clone derived from pTZAAV. The stem (A') and *Sma*I site-containing hammerhead (BB'CC') structures are indicated. **(B)** The DSM identifies rAAV–chromosomal junction sequences. Shown is an example of an rAAV–chromosomal junction sequence recovered from rAAV-transduced mouse liver. *Top*: The recovered sequence includes a portion of internal rAAV vector, the 3' ITR, and murine chromosome 2 (134373802–134373890, minus strand). *Bottom*: The chromatogram is provided; various sequence structures are indicated by brackets and the red arrow denotes the ITR–chromosomal junction point. For clarity, the relevant reads generated by the M13 forward (M13F) and M13 reverse (M13R) primers are presented in contiguous fashion. Color images available online at www.liebertpub.com/hgtb

and ASM should ensure the successful acquisition of ITR sequence content.

Applications

To advance clinical gene therapy, it is necessary to develop a process for rapidly producing highly titered, pure batches of rAAV. Although AAV possesses an ITR self-correction mechanism when rescued from plasmid DNA, deletions can nevertheless have a negative impact on AAV production (Samulski *et al.*, 1983; Wang *et al.*, 1996). Clearly, the DSM and ASM offer a way to identify constructs possessing such mutations, thereby providing an efficient means of screening out defective material before the time consuming and labor intensive process of rAAV vector production.

As shown in Fig. 6B, the DSM allows for the successful acquisition of chromosomal-ITR junction sequences. At present, we are employing the DSM for this purpose in a number of ongoing studies and have reproducibly obtained ITR junction sequences exceeding 500 nucleotides in length. It is possible that the chemistry adaptations presented herein

could be applied to the massively parallel sequencing efforts being undertaken to map random rAAV integration on a genome-wide scale.

In conclusion, the direct sequencing and amplification sequencing methods provide additional tools with which to verify the sequences of ITR DNA and should prove useful for the development of large-scale rAAV production because DSM and ASM offer a way to screen out defective material before the time consuming and labor intensive processes of rAAV packaging and purification.

Acknowledgments

The authors thank John A. Zaia and Stephen J. Forman, City of Hope (COH), for support; Stephanie Gipson, of the Sequencing Core, for technical assistance; Bonnie Notthoff, Division of Hematology/Stem Cell Transplantation, COH, for secretarial help; and members of the laboratory who provided assistance. This research was supported in part by NHLBI grant R01HL087285. The COH Cancer Center Cores used for this study included the

Sequencing Core, Oligonucleotide Synthesis Core, and the Cloning Laboratory. Core facilities were supported by NCI grant P30CA33572.

Author Disclosure Statement

The authors have no competing financial interests.

References

- Aravind, A., Schaffer, D.V., and Samulski, R.J. (2012). The AAV vector toolkit: Poised at the clinical crossroads. *Mol. Ther.* 20, 699–708.
- Bainbridge, J.W., Smith, A.J., Barker, S.S., *et al.* (2008). Effect of gene therapy on visual function in Leber's congenital amaurosis. *N. Engl. J. Med.* 358, 2231–2239.
- Chatterjee, S., Johnson, P.R., and Wong, K.K., Jr. (1992). Dual-target inhibition of HIV-1 *in vitro* by means of an adeno-associated virus antisense vector. *Science* 258, 1485–1488.
- Cideciyan, A.V., Aleman, T.S., Boye, S.L., *et al.* (2008). Human gene therapy for RPE65 isomerase deficiency activates the retinoid cycle of vision but with slow rod kinetics. *Proc. Natl. Acad. Sci. U.S.A.* 105, 15112–15117.
- Daya, S., and Berns, K.I. (2008). Gene therapy using adeno-associated virus vectors. *Clin. Microbiol. Rev.* 21, 583–593.
- Feigin, A., Kaplitt, M.G., Tang, C., *et al.* (2007). Modulation of metabolic brain networks after subthalamic gene therapy for Parkinson's disease. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19559–19564.
- Hauswirth, W.W., Aleman, T.S., Kaushal, S., *et al.* (2008). Treatment of Leber congenital amaurosis due to RPE65 mutations by ocular subretinal injection of adeno-associated virus gene vector: Short-term results of a phase I trial. *Hum. Gene Ther.* 19, 979–990.
- Henke, W., Herdel, K., Jung, K., *et al.* (1997). Betaine improves the PCR amplification of GC-rich DNA sequences. *Nucleic Acids Res.* 25, 3957–3958.
- Inagaki, K., Ma, C., Storm, T.A., *et al.* (2007). The role of DNA-PKcs and Artemis in opening viral DNA hairpin termini in various tissues in mice. *J. Virol.* 81, 11304–11321.
- Kaplitt, M.G., Feigin, A., Tang, C., *et al.* (2007). Safety and tolerability of gene therapy with an adeno-associated virus (AAV) borne GAD gene for Parkinson's disease: An open label, phase I trial. *Lancet* 369, 2097–2105.
- Kieleczawa, J. (2006). Fundamentals of sequencing of difficult templates—an overview. *J. Biomol. Tech.* 17, 207–217.
- Laughlin, C.A., Tratschin, J.D., Coon, H., and Carter, B.J. (1983). Cloning of infectious adeno-associated virus genomes in bacterial plasmids. *Gene* 23, 65–73.
- LeWitt, P., Rezai, A., Leehey, M., *et al.* (2011). AAV2-GAD gene therapy for advanced Parkinson's disease: A double-blind, sham-surgery controlled, randomised trial. *Lancet Neurol.* 10, 309–319.
- Lusby, E., Fife, K.H., and Berns, K.I. (1980). Nucleotide sequence of the inverted terminal repetition in adeno-associated virus DNA. *J. Virol.* 34, 402–409.
- Maguire, A.M., Simonelli, F., Pierce, E.A., *et al.* (2008). Safety and efficacy of gene transfer for Leber's congenital amaurosis. *N. Engl. J. Med.* 358, 2240–2248.
- Michelfelder, S., and Trepel, M. (2009). Adeno-associated viral vectors and their redirection to cell-type specific receptors. *Adv. Genet.* 67, 29–60.
- Mingozzi, F., and High, K. (2011). Therapeutic *in vivo* gene transfer for genetic disease using AAV: Progress and challenges. *Nat. Rev. Genet.* 12, 341–355.
- Musso, M., Boccardi, R., Parodi, S., *et al.* (2006). Betaine, dimethyl sulfoxide, and 7-deaza-dGTP, a powerful mixture for amplification of GC-rich DNA sequences. *J. Mol. Diagn.* 8, 544–550.
- Nathwani, A.C., Tuddenham, E.G., Rangarajan, S., *et al.* (2011). Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. *N. Engl. J. Med.* 365, 2357–2365.
- Samulski, R.J., Srivastava, A., Berns, K.I., and Muzyczka, N. (1983). Rescue of adeno-associated virus from recombinant plasmids: Gene correction within the terminal repeats of AAV. *Cell* 33, 135–143.
- Simonelli, F., Maguire, A.M., Testa, F., *et al.* (2010). Gene therapy for Leber's congenital amaurosis is safe and effective through 1.5 years after vector administration. *Mol. Ther.* 18, 643–650.
- Wang, X.S., Ponnazhagan, S., and Srivastava, A. (1996). Rescue and replication of adeno-associated virus type 2 as well as vector DNA sequences from recombinant plasmids containing deletions in the viral inverted terminal repeats: Selective encapsidation of viral genomes in progeny virions. *J. Virol.* 70, 1668–1677.

Address correspondence to:

Dr. K.K. Wong, Jr.

Divisions of Hematology/Stem Cell Transplantation

1500 E. Duarte Road

Beckman Research Institute, City of Hope

Duarte, CA 91010

E-mail: kkwong@coh.org

Received for publication January 5, 2012;

accepted after revision April 10, 2012.

Published online: April 19, 2012.