

## Article

# Satellite Subgenomic Particles Are Key Regulators of Adeno-Associated Virus Life Cycle

Junping Zhang <sup>1</sup>, Xiangping Yu <sup>1</sup>, Ping Guo <sup>1</sup> , Jenni Firrman <sup>2</sup>, Derek Pouchnik <sup>3</sup>, Yong Diao <sup>1</sup>, Richard Jude Samulski <sup>4,\*</sup> and Weidong Xiao <sup>5,\*</sup>

<sup>1</sup> School of Medicine, Huaqiao University, Quanzhou 362021, China; zhanju@iu.edu (J.Z.); yxping66@hotmail.com (X.Y.); guop159@126.com (P.G.); diaoyong@hqu.edu.cn (Y.D.)

<sup>2</sup> Dairy and Functional Foods Research Unit, Eastern Regional Research Center, Agricultural Research Service, U.S. Department of Agriculture, Wyndmoor, PA 19038, USA; jenni.firrman@usda.gov

<sup>3</sup> School of Molecular Biosciences, Washington State University, Pullman, WA 99164, USA; pooch@wsu.edu

<sup>4</sup> Department of Pharmacology, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

<sup>5</sup> Herman B Wells Center for Pediatric Research, Indiana University, Indianapolis, IN 46202, USA

\* Correspondence: rjs@med.unc.edu (R.J.S.); xiaow@iu.edu (W.X.)

**Abstract:** Historically, adeno-associated virus (AAV)-defective interfering particles (DI) were known as abnormal virions arising from natural replication and encapsidation errors. Through single virion genome analysis, we revealed that a major category of DI particles contains a double-stranded DNA genome in a “snapback” configuration. The 5′-snapback genomes (SBGs) include the P5 promoters and partial *rep* gene sequences. The 3′-SBGs contains the capsid region. The molecular configuration of 5′-SBGs theoretically may allow double-stranded RNA transcription in their dimer configuration. Our studies demonstrated that 5-SBG regulated AAV *rep* expression and improved AAV packaging. In contrast, 3′-SBGs at its dimer configuration increased levels of *cap* protein. The generation and accumulation of 5′-SBGs and 3′-SBGs appears to be coordinated to balance the viral gene expression level. Therefore, the functions of 5′-SBGs and 3′-SBGs may help maximize the yield of AAV progenies. We postulate that AAV virus population behaved as a colony and utilizes its subgenomic particles to overcome the size limit of a viral genome and encodes additional essential functions.

**Keywords:** adeno-associated virus; subgenomic particles; “snapback” configuration; life cycle



**Citation:** Zhang, J.; Yu, X.; Guo, P.; Firrman, J.; Pouchnik, D.; Diao, Y.; Samulski, R.J.; Xiao, W. Satellite Subgenomic Particles Are Key Regulators of Adeno-Associated Virus Life Cycle. *Viruses* **2021**, *13*, 1185. <https://doi.org/10.3390/v13061185>

Academic Editor: Shan-Lu Liu

Received: 28 May 2021

Accepted: 17 June 2021

Published: 21 June 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

AAV is a replication-defective parvovirus which requires a helper virus to complete its life cycle [1]. The virus is best known for its small genome, which is tightly packaged inside a capsid that is 20–25 nm in diameter. Its single-stranded DNA genome contains approximately 4700 nucleotides and encodes the *rep* and *cap* genes which represent the non-structural and capsid proteins, respectively. AAV replication is mediated by the *rep* proteins, which are capable of nicking the inverted terminal repeats (ITR) and therefore initiates AAV replication. In the presence of a helper virus, AAV undergoes its lytic infection. Without a helper virus, AAV integrates into the host genome and maintains a relatively stable, latent state.

AAV infection is highly regulated in both latent and lytic infections. In latent infection, AAV genes remain silent. In the lytic life cycle, *Rep78* and *Rep68*, under control of the P5 promoter, are expressed early, followed by expression of the packaging related genes, such as *Rep52/Rep40* and capsid genes *VP1*, *VP2* and *VP3*. In the late stage of AAV lytic infection, expression from the P5 promoter is downregulated to facilitate virus encapsidation, which was also demonstrated in recombinant AAV replication and packaging [2]. A variety of host, viral, and helper virus factors are involved in the temporal and spatial regulation of viral replication and packaging.

Although the canonical AAV particle typically contains both ITRs and the entire coding region, AAV populations are often found to be heterogeneous [3–7]. Upon being processed through a CsCl gradient, the full AAV particles are present in the fraction with 1.4 g/mL density. Lighter density AAV particles (1.32 and 1.35 g/mL) are largely deemed as empty particles or defective interfering (DI) particles. DI particles contain aberrant, shorter AAV DNA genomes, and are named for their ability to inhibit AAV replication intracellularly. Template switch was proposed to explain the nonunit-length AAV genome [6]. However, their composition and molecular conformation are elusive.

In this study, we systemically sequenced the genomes of AAV viruses within a population and revealed the molecular state of AAV genomes, both full length genome and DI particles at the single virus level. Unexpectedly, we discovered that a major class of DI particles are capable of regulating important biological functions for AAV expression and are an integral part of the AAV life cycle. In contrast to the general belief that all subgenomic particles are waste-byproducts and only compete for essential resources required for wild type virus replication and packaging, our findings suggest that the AAV virus has evolved to utilize these subgenomic particles to potentially encode dsRNA that regulate rep expression and augment cap expression. This is a clever manipulation evolved by AAV to circumvent its size limitation.

## 2. Materials and Methods

### 2.1. Cell Lines and Transfections

HEK 293 cells were maintained in Dulbecco's modified essential medium (DMEM) containing 10% bovine calf serum. PolyJet™ DNA In Vitro Transfection Reagents (Signa-Gen Laboratories, Frederick, MD, USA) were used to deliver DNA into HEK 293 cells. Cells are plated into 10-cm-diameter culture dishes 18 to 24 h prior to transfection so that the monolayer cell density reaches to the optimal 70~80% confluency at the time of transfection. Complete culture medium with serum is freshly added to each plate 30 min before transfection. We prepared the PolyJet™-DNA Complex for transfection according to the ratio of 3 µL PolyJet™ to 1 µg DNA using serum-free DMEM to dilute DNA and the PolyJet™ reagent. We incubated them for 10~15 min at room temperature and added the PolyJet™/DNA mixture onto the medium. We removed the PolyJet™/DNA complex-containing medium and replaced with fresh serum-free DMEM 12~18 h post transfection.

### 2.2. Expression Plasmids

The pH29 negative expression plasmid was made by deleting a 786-bp sbf1 and BsiwI fragment from pH29 plasmid. This plasmid expresses a normal level AAV Rep in rAAV production system but do not express AAV-2 capsid gene. The pssAAV-CB-RepCap plasmid was made by cloning the 4436-bp rep&cap fragment from pCI-RepCap into the backbone of pssAAV-CB-GFP at Bgl II site. The AAV 5'-SBG dimer was generated through the intramolecular ligation of the ITR-CB-Rep fragment, which contains AAV ITR and partial CB-Rep fragment, and isolated from pssAAV-CB-RepCap by NaeI-BamHI digestion. AAV 3'-SBG monomer was made by intramolecular-ligation of BamHI-SmaI fragment isolated from psub201. AAV 3'-SBG dimer was made by intermolecular ligation of BamHI-SmaI fragment isolated from psub201.

Head-to-head dimer DNA molecules containing Gluc gene were generated by ligation of the 1587 bp fragment, which is from pssAAV-CB-Gluc plasmid digested with SmaI and ClaI. The monomer fragment includes a single CB promoter, Gluc gene and a single Poly A signal. Tail to Tail dimer DNA molecules containing Gluc gene were generated by ligation of the 1557 bp fragment, the monomer fragment is from pdsAAV-CB-Gluc digested with SmaI and MluI.

### 2.3. Virus Production

AAV viruses were produced using the triple plasmid transfection system in HEK 293 cells according to method that were described previously [8].

#### 2.4. Western Blot

To detect Rep and Cap protein expression, at 72 h after transfection, total protein was extracted. The cell pellet was washed with PBS twice, added RIPA lysis buffer plus proteinase inhibitor (PefablocSC PLUS, Roche Applied Science, Indianapolis, IN, USA), incubated for 30 min on ice, centrifuged 30 min at 4 °C 15,000 rpm, and cell supernatant was harvested. We used 10 µL of cell supernatant to quantify using the Pierce™ BCA Protein Assay Kit (Thermo Scientific, USA). The rest supernatant was heated in a 1× dilution loading buffer NuPAGE™ LDS Sample Buffer (4×, Invitrogen, Waltham, MA USA) at 70 °C for 10 min to denature sample. We loaded 30 µg of protein from each sample on 4% of concentrating gel and 8% of separating SDS-PAGE gel alongside a set of protein molecular weight markers (PageRuler™ Prestained Protein Ladder, Thermo Scientific), run 120 min at 100 V. The samples were transferred to the PVDF membrane on a semi-dry transfer apparatus (Bio Rad Trans-blot SD Semi dry transfer cell), at 16 V, 30 min. The membrane was blocked with 5% of BSA in TBS buffer. To detect Rep or Cap protein, the membrane was incubated with the mouse anti-AAV-Rep monoclonal antibody 76.3 at 1:2000 dilution (American Research Products, Inc. Waltham, MA, USA) or mouse anti-AAV-Cap monoclonal antibody 61058 (American Research Products, Inc., Waltham, MA, USA), at 4 °C for overnight. The membrane was washed three times with the TBST buffer, 10 min each time. The membrane was incubated with the secondary antibody HRP-linked with anti-mouse antibody (dilution: 1:2000; Cell Signaling Technology, Inc., Danvers, MA, USA). The antibody was detected by chemiluminescence with Amersham ECL (Sigma-Aldrich, Inc., St. Louis, MO, USA).

#### 2.5. Quantitative Real Time PCR (qPCR) Assay

Viral vectors ( $1 \times 10^{10}$  vg, 1 µL) in the DNase solution containing DNase I (1 U/mL) were incubated for 30 min at 37 °C, adding 1 µL of 0.5 M EDTA (to a final concentration of 5 mM) and subsequently heated for 10 min at 75 °C to cease DNase I activity. Control samples each received a lysis buffer containing proteinase K (40 µg/mL), then were incubated for 1 h at 56 °C and finally heated for 10 min at 95 °C. The samples intended for thermal treatment were directly heated following heat inactivation of DNase I treatment at the indicated temperatures. The copy numbers of viral genomes subsequently released were quantified by real-time PCR and expressed in vg/mL.

#### 2.6. Dexamethasone Induction

pFΔ6, pMMTV trans (Rep-cap), pssAAV-CB-GFP and wild-type defective DNA were co-transfected into HEK 293 cells, at 8 h, replaced the old medium with fresh DMEM containing final concentration  $1 \times 10^{-5}$  M of dexamethasone, and incubated for 72 h, AAV-GFP vector yield was detected by qPCR.

#### 2.7. AAV Genome Sequencing

The AAV genome was sequenced by PacBio SMRT Sequencing. Samples were prepared according to the general protocol of the SMRTbell library. DNA was extracted and purified by AMPure PB Beads. DNA damage/end repair was performed using the SMRTbell™ Damage repair kit. Adaptor ligation reaction was performed and then add ExoIII and ExoVII to remove failed ligation products. Templates were purified by AMPure PB Beads three times. Sequencing was performed by PacBio SMRT platform. PacBio raw subreads were modified with recalladapters v7.1.0 (<https://github.com/PacificBiosciences/recalladapters>, accessed on 21 September 2019) to generate high quality circular consensus sequencing reads (HiFi reads) using the SMRT link analysis ccs v4.0.0 with `-minPasses 3` and `-minPredictedAccuracy 0.99` (<https://github.com/PacificBiosciences/ccs>, accessed on 21 September 2019). Filtered HiFi reads were mapped to the reference genome using the minimap2 v2.17 with `-ax map-pb` (<https://github.com/lh3/minimap2>, accessed on 21 September 2019). Reads were subsampled in varying length ranges with seqkit tools v0.10.1- (<https://anaconda.org/bioconda/seqkit>, accessed on 23 September 2019) and

processed by doing loops of BLAST-based alignment to reference genome to categorize molecules based on identity. For each loop, we aligned unmatched fragments in the previous alignment of the same HiFi read on the left or right sides to the reference genome with `-task blastn` and `-max_hsps 1`. The comprehensive detail data of alignments were visualized in AAV configurations.

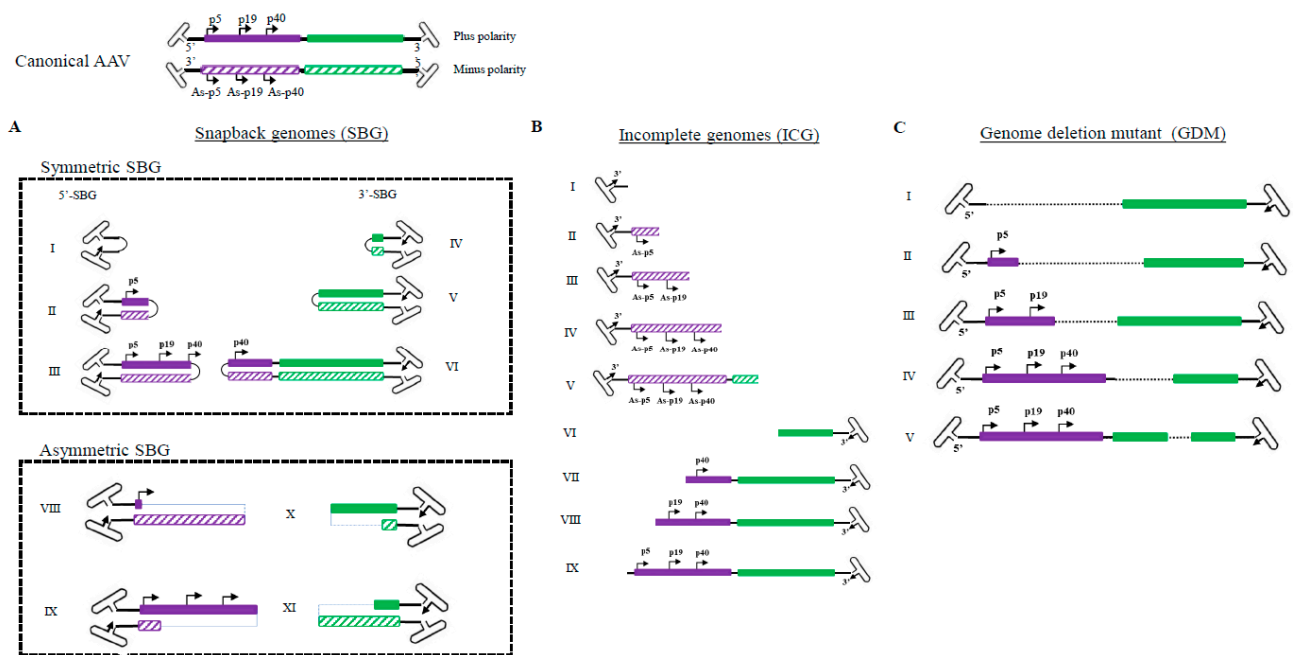
### 2.8. Statistic Analysis

All data were presented with as means  $\pm$  SD. Statistical analysis was performed by Student's unpaired t-test in SPSS software version 1.0.0.1406. A  $p$  value  $< 0.05$  was considered statistically significant.

## 3. Results

### 3.1. Molecular State of AAV Subgenome Particles at a Single-Virus Level

The high GC content and palindromic nature of the AAV ITR has been a major obstacle for analyzing AAV genomes. The full DNA genome of the AAV virus has largely been obtained through assembling multiple fragments to obtain a consensus sequence. Even with next generation sequencing, the library construction procedures often require breaking the sample DNA and reassembling the genomes. At the single-virus level, important information became lost by this maneuver since the viral genomes in a population are derivatives of the same consensus sequences. Here we utilized the Pacbio Single Molecule, Real-Time (SMRT) sequencing platform and mapped the population genomic configuration distribution of the AAV genome at a single virus level (Figure 1). There were no enzymatic or mechanic actions that altered the original viral DNA configuration. From analysis of more than 400,000 ccs reads, four major categories of molecules are found in the AAV population. Category 1: Canonical AAV genomes, which contain the full AAV genome flanked by two copies of the AAV ITR; Both positive and negative strands are encapsidated equally into AAV capsids. Category 2: Snapback genomes (SBG), which contain partial duplex AAV genomes with an ITR. Such a molecule was shown to be able to snap back and anneal to itself upon denaturing and renaturing cycle [6] SBGs are essentially a self-complementary DNA molecule with either the left moiety (5'-SBG) or the right moiety (3'-SBG) genomes. These snapback genomes were either symmetric or asymmetric. The symmetric SBG has near equal length of top strand and bottom strand when in self-complementary conformation. The asymmetric SBG has varying lengths of top strand or bottom, which leads to a single strand region as a loop. Category 3: Incomplete genomes (ICG), which contain an intact 3'ITR and a partial AAV coding sequence, but the sequences do not reach the 5' ITR. Category 4: Genome deletion mutants (GDM), in which the middle region of the AAV genomes is missing. The molecular distribution of GDM is shown Figure S1 and the ratio of ICG vs. SBG was presented in Figure S2. Any viral particles that do not have the canonical AAV genomes are referred to as subgenomic particles (SDG), which may include those containing host genomic DNA and helper DNA, which were not presented here in details.



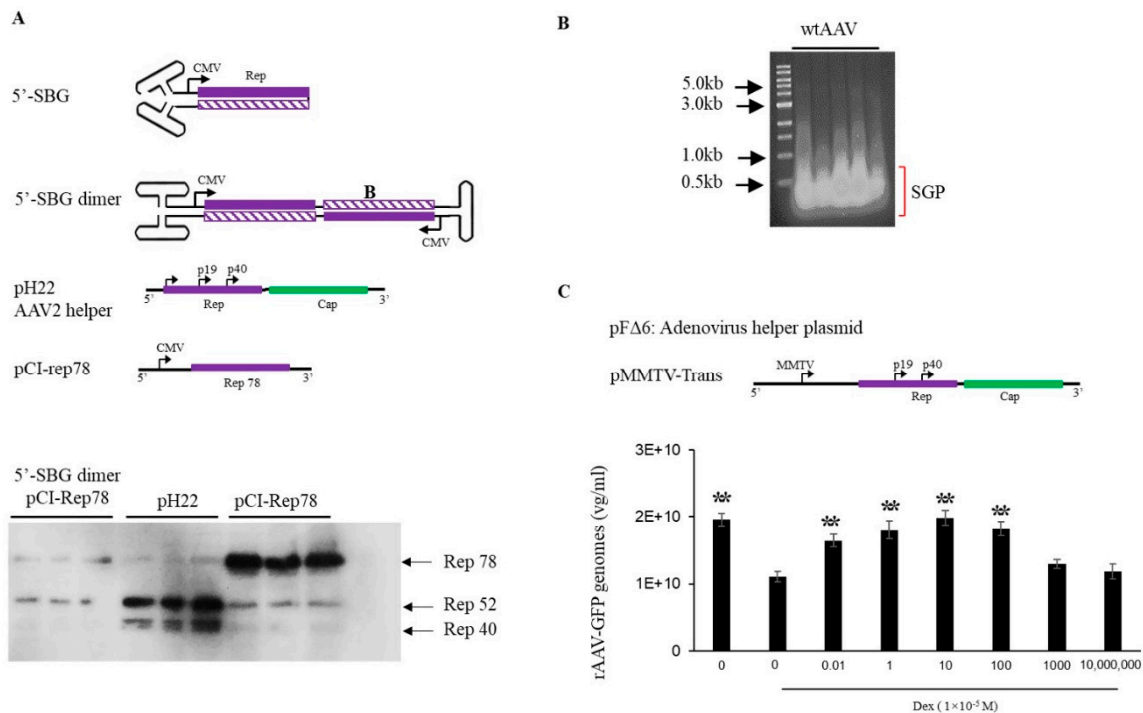
**Figure 1.** Illustration of molecular configurations for major categories of subgenomic particles in a wild type AAV population. (Summarized from 409919 sequencing reads of two PacBio sequencing runs: 220645 and 189274) Both plus and minus strands of the canonical AAV genome are shown at the top. (A). Varying lengths of snapback AAV genomes (SBG) are illustrated, including both symmetrical snapback genomes (sSBG) and asymmetrical snapback genomes (aSBG). (B). Incomplete genomes (ICG) in the AAV population are missing the 5' terminal sequences. (C). Genome deletion mutants (GDM) have both 5' and 3' ITR and miss in the mid-region of AAV. Only plus strand particles are plotted.

### 3.2. 5' Snapback Subgenomes (5'-SBG) Function as a Cis-Negative Regulator of Rep Expression

5'-SBG assumed a self-complementary configuration and was capable of replicating in the presence of Rep proteins and adenovirus helper functions, which indicated that the dimeric 5'-SBG molecules consisted of the P5 promoter followed by an inverted, head-to-head coding region. Therefore, logically dsRNA could be expressed from the dimer 5'-SBG when 5'-SBG undergoes DNA replication and becomes a dimer. Such dsDNA overlaps with P5 transcripts in AAV replication which would be able to downregulate P5 promoter expression, mainly Rep78 or Rep 68. To demonstrate its effect on rep expression, we engineered a 5'-SBG dimer with a CMV promoter, a partial rep sequence was constructed to assume a tail-to-tail dimeric configuration (Figure 2A). The RNA transcript would, therefore, assume dsRNA configuration when folding on itself. As shown in Figure 2A, the expression of Rep78 was indeed, significantly reduced in the presence of dimer 5'-SBG, which suggested that 5'-SBG, potentially functioning through dsRNA intermediates, exerted its effects against Rep gene as the trans-active regulator.

It has been previously shown that down-regulation of *Rep78* gene expression improves rAAV packaging [2] To test if 5'-SBG may exhibit such regulatory role, we applied extracted 5'-SBG DNA to a rAAV production system containing pMMTV-trans as the AAV rep and cap expression plasmids (Figure 2B). MMTV is an inducible promoter which can be activated by dexamethasone. In the absence of dexamethasone, low expression of rep expression gave rise to a higher rAAV yield. Conversely, when dexamethasone was added, the vector yield was reduced. When the 5'-SBG molecules were added to the production system in the presence of dexamethasone, rAAV vector yield was increased at low concentration of 5'-SBG genomes. However, when the amount of 5'-SBG was increased to more than 10 copies per cell, it started to exhibit an inhibitory effect, and the vector yield was reduced from the peak. Therefore, it was concluded that 5'-SBG had a regulatory function which senses the AAV genome pool. When SBG/AAV ratio is at a low level,

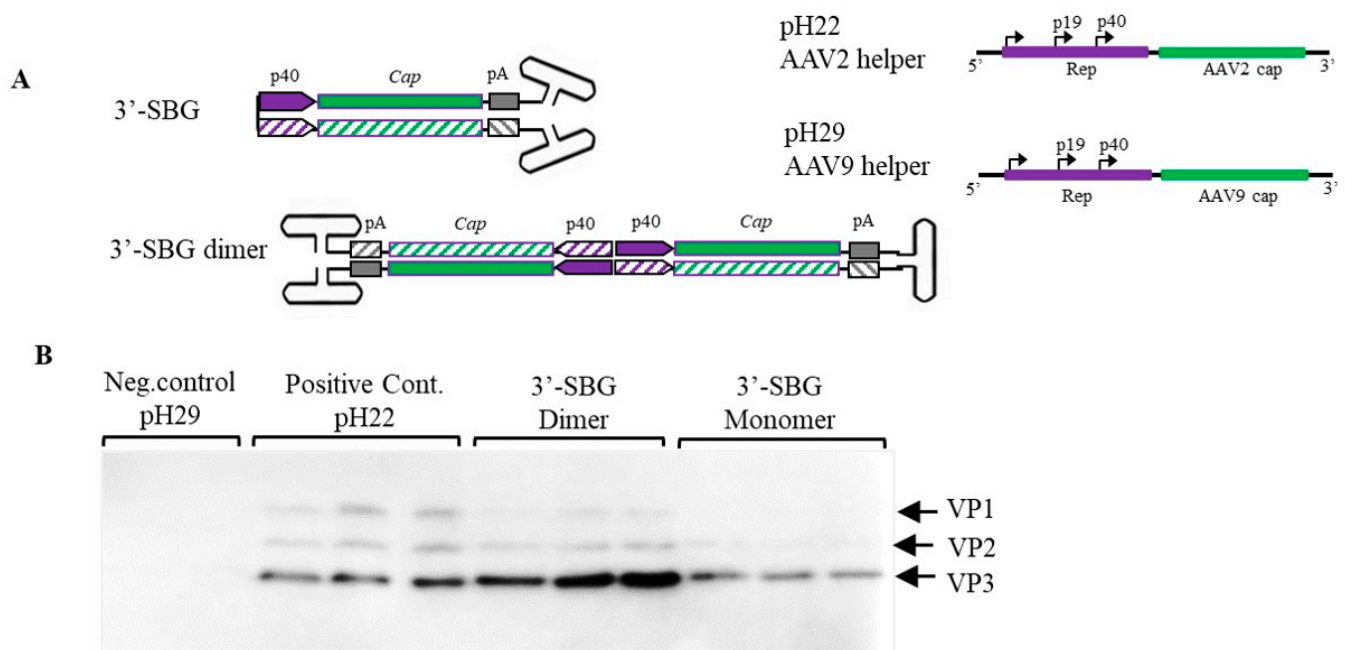
it increases AAV packaging efficacy. However, when SBG is present in excess, it had an inhibitory effect and became a true “defective interfering particles”.



**Figure 2.** Effects of AAV 5'-SBG dimers on rep gene expression. (A). Top panel, sketch of a 5'-SBG in self-complementary configuration; 5'-SBG dimer is extended form of 5'-SBG after 2nd stranded DNA synthesis; pH22 is the AAV2 helper plasmid expressing low level of Rep78; pCI-Rep78 is the high Rep78 expression plasmid. Bottom panel, Hek 293 cells were transfected with pCI-Rep78 with or without the presence of 5'-SBG dimers. 5'-SBG dimers were made as detailed in the experiment methods. The transfected cells were harvested at 24 h post-transfection and the expression of proteins were electrophoresed in 10% PAGE gel. The rep proteins were detected by western blot using mouse anti-AAV-Rep 76.3 antibody. Shown in the Figure are triplicate of each testing condition. Effects of AAV subgenomic particles (SGP) on rAAV replication and packaging. (B). AAV SGP were separated in an agarose gel and less than 1 kb SGP DNA identified in an agarose gel, isolated for the transfection. (C). The adenovirus helper plasmid pFΔ6, pMMTV-trans and pssAAV-CB-GFP were co-transfected into 293 cells with or without SGP DNA. Rep expression was under the control of a MMTV promoter in pMMTV-trans, which is inducible by Dexamethasone. ssAAV-GFP was collected at 72 h post-transfection and measured by qPCR. The X axis showed the copy number of SGP molecules used for transfection. “0” means no SGP DNA was added. The cells treated with DEX (1 × 10<sup>-5</sup> M) are indicated in the Figure. Y axis: rAAV-GFP yield expressed by vector genomes per ml. \*\* means  $p < 0.05$ , there is statistically significant as compared to 0 M Dex treatment group.

### 3.3. Bidirectional Transcription from 3' Snapback Subgenomes (3'-SBG) Boosts Capsid Protein Level to Maximize Vector Yields

Similar to 5'-SBG, 3'-SBG replicates in the presence of rep proteins and external AAV helper function. However, the dimeric configuration of 3'-SBG was elucidated as a head-to-head molecule with the P40 promoter in the center of the dimer. This means that any enhancer in the proximity of the P40 promoter would also affect the neighbor P40 promoter on the opposite strand, which would increase the strength of P40 promoter and expression of the cap protein. To demonstrate the double enhancer effects of 3'-SBG, 3'-SBG monomeric and dimeric constructs were engineered, and the expression of the capsid genes analyzed. As shown in Figure 3, capsid expression from 3'-SBG monomer containing a single P40 promoter was low, whereas expression from 3'-SBG dimer was significantly increased.



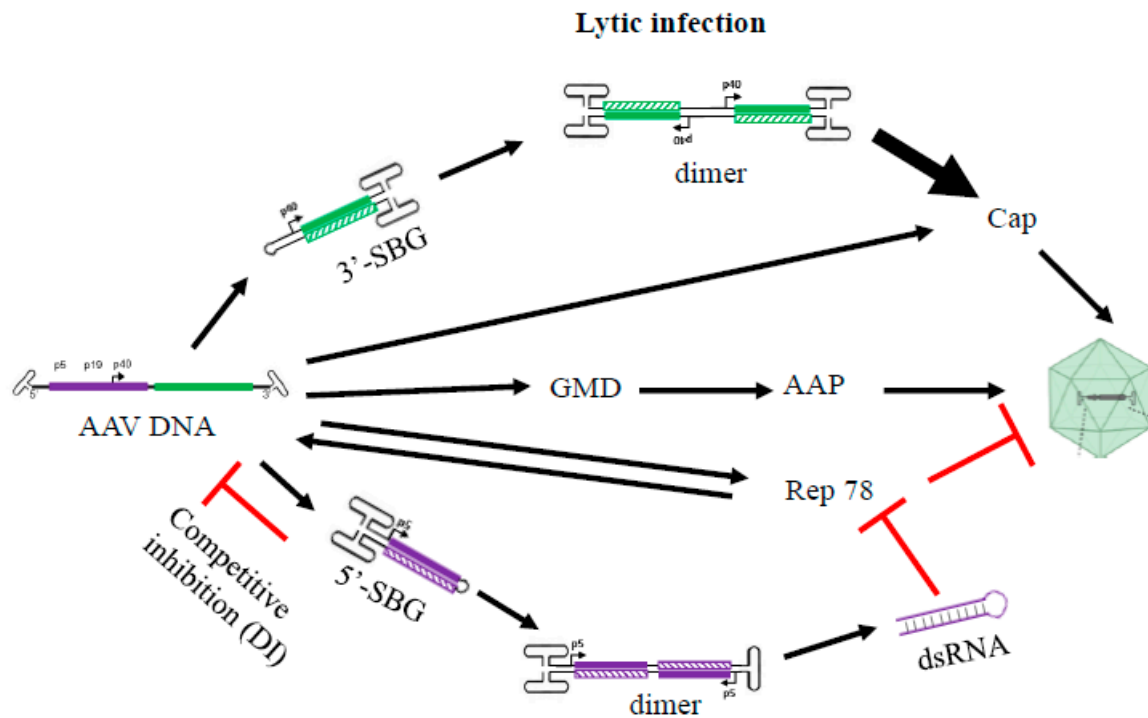
**Figure 3.** Enhancement of Cap protein expression by 3'-SBG. **(A)** 3'-SBG is shown in monomer configuration; 3'-SBG dimer is the extended form of 3'-SBG after 2nd stranded DNA synthesis or replication. **(B)** Effects of 3'-SBG on capsid protein expression. AAV 3'-SBG molecule dimer was modeled by intermolecular ligation of a *Sam*I and *Bam*HI digestion fragment from the AAV infectious clone psub201. AAV 3'-SBG monomer was modeled by an intramolecular ligation of the same *Sam*I and *Bam*HI digestion fragment. Cap protein expression was detected with mouse anti-AAV Capsid antibody by western blotting at 72 h post transfection. pH29 negative: HEK 293 cells were transfected with pFΔ6 and pH29 mutation plasmid without AAV2 capsid expression; pH22: HEK 293 cells were transfected with pFΔ6 and pH22 plasmid which express AAV capsid; AAV 3'-SBG dimer: HEK 293 cells were transfected with pFΔ6 and 3-SBG dimer molecules; AAV 3-SBG monomer: HEK 293 cells were transfected with pFΔ6 and 3-SBG monomer molecules. Each condition is shown as triplicate.

#### 4. Discussion

Previously, the heterogeneity of AAV genomes within a population was not fully characterized because the palindromic structure of the AAV ITR inhibited progression of the typical sequencing reaction, thus requiring AAV genome to be broken/interrupted in order to make compatible libraries that were compatible for the sequencing instrument. However, this maneuver causes the loss of the detailed information of individual molecules, since DNA in AAV subgenomic particles are inherently diverse and heavily rearranged from the standard genomes (Figure 1). Reported studies using PacBio to sequence scAAV only recovered special categories of AAV molecules and did not provide details on the individual genomes within the entire AAV population [9]. In studies with single-stranded DNA genomes [10], only annealed AAV genomes were captured in the library. Here we have obtained the full profile of AAV genomes within a population at the single virus level (Figure 1). Those unannealed molecules are also fully sequenced. This high-resolution analysis allowed us to identify and characterize an array of genome configurations that are present. Besides canonical AAV full-length genomes, the presence of particles with incomplete AAV genomes, starting from the 3'ITR that failed to reach the 5'-ITR, which indicated an aborted packaging process (Figure 1). Since AAV packaging initiation starts from the 3'ITR, all ICGs have intact 3'ITR and no ICGs have 5'ITR [11]. We did not observe obvious hot spots in ICG population, which suggested that incomplete packaging of the AAV genome was most likely a random event.

In addition, we revealed the divergence of genome deletion mutants (GDM) in the AAV population supplementary data Figure S1. The existence of GDMs as well as the discovery of asymmetric SBG configuration are the primary reason that we proposed that NHEJ is the mechanism of subgenomic particles formation in the AAV population (citation

of Preprint). NHEJ is the mechanism that can simultaneously explain the formation sSBG, aSBG, GDM and various forms of subgenomic particles identified in the rAAV population (including those containing foreign genetic element such as host genome DNA and helper genes). Although we did not study GDMs, some GDMs will bring AAP reading frame closer to either P5, 19 or P40 promoter, which would increase AAP expression and enhance AAV package. This possibility is outlined in Figure 4.



**Figure 4.** A hypothetical model summarizing the role of 5'-SBG and 3'-SBG in the AAV life cycle. 3'-SBG improves an AAV cap expression when its copy number is increased along with its double enhancer effects in dimer conformation. 5'-SBG may potentially express dsRNA against *Rep78*. *Rep78* in excess is detrimental to AAV DNA replication and package. Therefore, the relative abundance of AAV genomes and 3'-SBG and 5'-SBG may form a dedicated positive and negative loop that ensure an optimum level of production of the next generation progeny. This model provides a potential explanation that the excessive DNA template therefore requires another layer of gene regulation when there is dramatic amplification of AAV genomes during replication.

AAV SBG molecule is naturally occurring self-complementary DNA genome which is similar to the well-known recombinant scAAV vector which used successfully in clinical trials, but does not need a special mutated AAV ITR. In this study, our findings revealed a critical role for SBGs in the AAV life cycle. The 5'-SBG containing the P5 promoter can downregulate *Rep78* expression, its underlying mechanism may be through expressing dsRNA overlapping the rep gene transcript. dsRNA is the precursor for RNA Interference [12], which would function as an inhibitor for the P5 promoter transcripts. This is a very clever mechanism that can balance P5 expression levels when the viral template increases excessively at the end of viral replication and packaging and there is no additional need to recruit host factors downregulate the promoter. Since the expression of *Rep78* and its inhibitor is under the same transcription mechanism, the inhibition rate is determined by relative copy numbers of full-length genome and 5'-SBG, 5'-SBG therefore becomes a de facto AAV genome population sensor. In a hypothetical replication model starting with one full length particle, the P5 promoter initiates expression of *Rep 78* which replicates the AAV genome. The SBG molecules are generated when AAV genomes have accumulated to a certain level. Since SBG particles replicate faster than the full AAV genome, at a certain point, SBG replication will overtake the full AAV genome and *Rep78* expression levels will



be significantly reduced. Thus, 5-SBG is a rather effective inhibitor and can regulate the AAV genome population.

On the other hand, the snapback molecules were also formed at 3'-end. 3'-SBG cannot produce dsRNA because the P40 promoters are sitting in a head-to-head configuration. This leads to the enhancement of P40 gene expression, since the expression can be increased greatly as the copy number of 3'-SBG increases (Figure 3).

#### *A Hypothetical Model for the Essential Role of Snapback Subgenomes in the AAV Life Cycle*

Based on the detailed genomic state and molecular function of the AAV subgenomes, we propose that such molecules are not waste byproducts but play an active and critical role in the AAV life cycle. During replication and packaging, these molecules would function as "cis" and "trans" regulators of rep and cap gene expression, balancing viral gene copy number and expression levels. Without SBG molecules, AAV primarily replicates itself with reduced packaging. However, the successive replication of AAV genomes also lead to the production of SBG molecules, which may have a growth advantage over the wild type AAV genome and lead to a decrease in *Rep 78* expression and an increase in cap expression. The outcome of this replication mechanism would be an increased packaging of the AAV progenies. This hypothetical model is summarized in Figure 4.

The excessive amounts of 5-SBG in the wild type AAV population may have another implication. Based on the results of various genomic studies, it is known that, during a latent infection of host cells, only AAV fragments are found. It is likely that 5'-SBG also functions in the host cells as a suppressor of P5 expression, which may stabilize its latent infection (Figure S3). That may be a reason that it is beneficial to have subgenomic particles present in large quantities when the AAV virus is preparing its latent infection. Protein kinase R (PKR) is activated by dsRNA produced during virus replication. Adenovirus virus-associated RNA-I (VAI) is a short, noncoding transcript that functions as an RNA decoy to sequester PKR in an inactive state. VAI is an essential gene for rescuing AAV from latent infection. It further implies the complicated relationship between AAV helper virus and AAV replication.

AAV virus has a small genome of 4.7 k nucleotide in size. Yet, its genome design is a rather efficient and space conscience, with all genes overlapping each other in the same reading frame. Nevertheless, the virus developed a mechanism to use subgenomic particles, i.e., defective interfering particles, to express the negative regulator dsRNA, and to enhance expression of capsid proteins to facilitate packaging during the late stage of infection. The presence of such 5'-SBG and 3'-SBG is not a coincidence but evolved over virus-spread in the human population. The utilization of snap-back molecules is a perfect example of how extra-genomic molecules can be an integral part of AAV regulation. This is the first report describing a virus utilizing these mechanisms, which seems to be shared by all members of the parvovirus family (data not shown).

## 5. Conclusions

From analyzing the genome configurations of AAV defective interfering particles (DI), a major category of DI particles is revealed, which contains a double-stranded DNA genome in a "snapback" configuration. Such molecules may enhance the capsid protein expression and modulate rep expression. Such subgenomic particles play an important in wild type AAV life cycle.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/v13061185/s1>. Figure S1: Illustration of GDM population. A. varying GDM that misses part of AA genomes. B. All recovered GDM genomes from sequencing reaction are aligned to the canonical wtAAV genome. The 5' moiety and 3' moiety are arranged according to the length that match the canonical AAV genome. Figure S2: The ratio of SBG/ICG in the subgenomic particle population based on the two sequencing results. The top number indicates nucleotide range in the wtAAV genomes. Gray dots showed the average ratio in two sequencing reactions. (a). Distribution of the percentage of various symmetric snapback genomes (sSBG). (b). Distribution of the percentage

of incomplete genomes (ICG). Figure S3: A hypothetical model illustrating the role of 5'-SBG in the AAV latent life cycle. In the presence of 5'-SBG integration, the dsRNA may repress P5 promoter which can stabilize the latent infection.

**Author Contributions:** Conceptualization, W.X. and J.Z.; methodology, W.X. and J.Z.; software, X.Y.; validation, W.X., R.J.S. and J.Z.; formal analysis, X.Y. and D.P.; investigation, J.Z. and P.G.; resources, W.X.; data curation, W.X., X.Y. and D.P.; writing—original draft preparation, W.X. and J.Z.; writing—review and editing, W.X., R.J.S., J.F., Y.D.; visualization, J.Z., W.X. and R.J.S.; supervision, W.X.; project administration, J.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The wide type AAV datasets analyzed during the current study have been freely accessible deposited in NCBI Sequence Read Archive (SRA) under Submission ID: SUB9100196, BioProject: PRJNA702658.

**Conflicts of Interest:** W.X. holds equity in Ivygen Corporation and Nikegen LLC.

## References

1. Muzyczka, N.; Samulski, R.J.; Hermonat, P.; Srivastava, A.; Berns, K.I. The Genetics of Adeno-Associated Virus. *Adv. Exp. Med. Biol.* **1984**, *179*, 151–161. [[CrossRef](#)] [[PubMed](#)]
2. Li, J.; Samulski, R.J.; Xiao, X. Role for highly regulated rep gene expression in adeno-associated virus vector production. *J. Virol.* **1997**, *71*, 5236–5243. [[CrossRef](#)] [[PubMed](#)]
3. Carter, B.J.; Laughlin, C.A.; de la Maza, L.M.; Myers, M. Adeno-associated virus autointerference. *Virology* **1979**, *92*, 449–462. [[CrossRef](#)]
4. De la Maza, L.M.; Carter, B.J. Heavy and light particles of adeno-associated virus. *J. Virol.* **1980**, *33*, 1129–1137. [[CrossRef](#)] [[PubMed](#)]
5. Denhardt, D.T.; Eisenberg, S.; Bartok, K.; Carter, B.J. Multiple structures of adeno-associated virus DNA: Analysis of terminally labeled molecules with endonuclease R-Hae III. *J. Virol.* **1976**, *18*, 672–684. [[CrossRef](#)] [[PubMed](#)]
6. Hauswirth, W.W.; Berns, K.I. Adeno-associated virus DNA replication: Nonunit-length molecules. *Virology* **1979**, *93*, 57–68. [[CrossRef](#)]
7. Laughlin, C.A.; Myers, M.W.; Risin, D.L.; Carter, B.J. Defective-interfering particles of the human parvovirus adeno-associated virus. *Virology* **1979**, *94*, 162–174. [[CrossRef](#)]
8. Kapranov, P.; Chen, L.; Dederich, D.; Dong, B.; He, J.; Steinmann, K.E.; Moore, A.R.; Thompson, J.; Milos, P.M.; Xiao, W. Native Molecular State of Adeno-Associated Viral Vectors Revealed by Single-Molecule Sequencing. *Hum. Gene Ther.* **2012**, *23*, 46–55. [[CrossRef](#)] [[PubMed](#)]
9. Xie, J.; Mao, Q.; Tai, P.W.; He, R.; Ai, J.; Su, Q.; Zhu, Y.; Ma, H.; Li, J.; Gong, S.; et al. Short DNA Hairpins Compromise Recombinant Adeno-Associated Virus Genome Homogeneity. *Mol. Ther.* **2017**, *25*, 1363–1374. [[CrossRef](#)]
10. Tran, N.T.; Heiner, C.; Weber, K.; Weiland, M.; Wilmot, D.; Xie, J.; Wang, D.; Brown, A.; Manokaran, S.; Su, Q.; et al. AAV-Genome Population Sequencing of Vectors Packaging CRISPR Components Reveals Design-Influenced Heterogeneity. *Mol. Ther. Methods Clin. Dev.* **2020**, *18*, 639–651. [[CrossRef](#)]
11. King, J.A.; Dubielzig, R.; Grimm, D.; Kleinschmidt, J.A. DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids. *EMBO J.* **2001**, *20*, 3282–3291. [[CrossRef](#)]
12. Macrae, I.J.; Zhou, K.; Li, F.; Repic, A.; Brooks, A.N.; Cande, W.Z.; Adams, P.; Doudna, J.A. Structural Basis for Double-Stranded RNA Processing by Dicer. *Science* **2006**, *311*, 195–198. [[CrossRef](#)] [[PubMed](#)]