Maintenance of the Flip sequence orientation of the ears in the parvoviral left-end hairpin is a non-essential consequence of the critical asymmetry in the hairpin stem.

Lei Li¹, Susan F. Cotmore¹ and Peter Tattersall¹,²

Departments of Laboratory Medicine¹ and Genetics²,

Yale University Medical School, 333 Cedar Street, New Haven, CT 06510

* Corresponding author

Phone: (203) 785-4586   FAX (203) 688-7340

e-mail: peter.tattersall@yale.edu

Running Title: Conserving the left-end hairpin orientation of MVM = 50
18 Word count in abstract: 250
19
Parvoviral terminal hairpins are essential for viral DNA amplification, but are also implicated in multiple additional steps in the viral life cycle. The palindromes at the two ends of the MVM genome are dissimilar, and are processed by different resolution mechanisms that selectively direct encapsidation of predominantly negative sense progeny genomes, and conserve a single “Flip” sequence orientation at the 3’ (left-end) of such progeny. The sequence and predicted structure of these 3’ hairpins are highly conserved within the genus Parovirus, exemplified by the 121 nucleotide left-end sequence of MVM, which folds into a Y-shaped hairpin containing small internal palindromes that form the “ears” of the Y. To explore the potential role(s) of this hairpin in the viral lifecycle, we constructed infectious clones with the ear sequences either inverted, to give the anti-parallel “Flop” orientation, or with multiple transversions, conserving their base composition, but changing their sequence. These were compared with a “bubble” mutant designed to activate the normally silent origin in the inboard arm of the hairpin, thus potentially rendering symmetric the otherwise asymmetric junction resolution mechanism that drives maintenance of Flip. This mutant exhibited a major defect in viral duplex and single-strand DNA replication, characterized by the accumulation of covalently-closed turn-around forms of the left-end, and was rapidly supplanted by revertants that restored asymmetry. In contrast, both sequence and orientation changes in the hairpin ears were tolerated, suggesting that maintaining the Flip orientation of these structures is a consequence of, but not the reason for, asymmetric left-end processing.
Introduction

Members of the family Paroviridae have a linear single-stranded DNA genome of about 5kb, terminating in small palindromic telomeres that can fold into self-priming hairpins at each end of the viral chromosome. These hairpins, together with a few adjacent nucleotides, contain all of the cis-acting information required to mediate the virus' rolling hairpin replication strategy (24). In several paroviral genera, the terminal palindromes form part of inverted terminal repeats (ITRs), so that the same DNA origin sequence occurs at the two ends of the genome. For these viruses both termini are resolved by the same mechanism, called "terminal resolution", which occurs with equal efficiency at both ends of the genome, and ultimately results in the release and packaging of infectious progeny single strands of both polarities. One consequence of using this mechanism is that the sequence of the hairpins are inverted with each round of synthesis, creating termini in two, inverted complementary, sequence orientations, dubbed "Flip" and "Flop". However, species belonging to the Parovirus, Amdovirus and Bocavirus genera are heterotelomeric, that is, their genomic termini differ from each other in size, sequence and predicted secondary structure (39), and in MVM, type species of the genus Parovirus, the two termini are known to be processed by different mechanisms and at different rates. This property confers upon their replication a marked asymmetry, which drives the encapsidation of predominantly negative-sense progeny single strands (20). While studies on MVM have revealed the
mechanism underlying this asymmetry in some detail, precisely what advantages accrue from using this more-complex strategy remain uncertain.

One conspicuous result of asymmetric resolution is that a single sequence orientation, Flip, is conserved in left-end hairpins of the negative-sense progeny strands of all heterotelomeric parvoviruses examined to date (2, 3, 15, 39). The left-end of MVM, and likely other heterotelomeric viruses, is excised and replicated from palindromic dimer replicative-form (RF) intermediates, by a complex mechanism, termed “junction resolution” (24), which generates both a closed hairpin terminus and an extended palindromic terminus, as shown in Fig 1A (21). Since it is not clear why this elaborate asymmetric mechanism is invoked, it is possible that it exists simply to conserve specific feature(s) of the Flip orientation that are essential for virus viability. This hypothesis has proven difficult to test directly, for example, by transfecting synthetic genomes with hairpins in the Flop orientation, because these are simply converted back to the Flip orientation by the resolution mechanism (10). However, since the sequences that generate the asymmetry reside in the origin of DNA replication embedded within the hairpin stem (8, 9), we can effectively manipulate more distal elements in the hairpin. In the present study we have used a reverse genetic strategy, diagrammed in Fig 1A, to dissect these telomeric sequences and explore the importance of the distal elements.

As shown in Fig 1B, the unique 121 nucleotide left-end sequence of the negative sense MVM genome is predicted to fold into a Y-shaped hairpin, containing small internal palindromes that form these distal structures, dubbed the hairpin “ears”. The
43 base-pair duplex stem region is interrupted by a mismatched “bubble” sequence, where the triplet 5’-GAA-3’ on the inboard arm of the hairpin in virion DNA is opposed by the doublet 5’-GA-3’ on the outboard arm. Viral DNA synthesis is initially primed from the 3’ nucleotide of this hairpin, to generate a duplex molecule in which the two strands are linked to one another through the hairpin. However, in contrast to origins that are resolved by terminal resolution, potential origin sequences in the MVM left-end cannot be nicked by NS1 in this hairpin configuration. Instead, rolling-hairpin displacement synthesis must proceed, so that the hairpin is unfolded and copied to create the fully base-paired palindromic junction, spanning adjacent genomes in dimer RF, before the active origin is generated. Within this duplex structure the sequence from the outboard arm, surrounding the GA bubble dinucleotide, creates the active origin, OriL_{TC}, while the equivalent sequence from the inboard arm, OriL_{GAA}, containing the bubble trinucleotide, is inactive. The minimum linear origin sequence is approximately 50 basepairs in length, extending from two 5’-ACGT-3’ motifs spaced 5 nucleotides apart at one end, to a position some seven base pairs beyond the nick site (18). The two 5’-ACGT-3’ sequences serve as half-sites that co-operatively bind a heterodimeric cellular transcription factor called PIF, for parvovirus initiation factor (13, 14), also known as glucocorticoid modulating element-binding protein, or GMEB (30), which is abundantly expressed in many cell types. The position of the proximal PIF half-site relative to the NS1 binding site is absolutely critical for allowing PIF to activate NS1, since the single additional intervening bubble nucleotide of the inboard arm
prevents PIF from stabilizing the binding of NS1, which is therefore unable to nick oriLGAA (12).

While the actual sequence of the oriLTC doublet is relatively unimportant, we have shown that insertion of any third nucleotide here inactivates the origin, thus indicating that the bubble is a critical spacer, rather than a recognition element in its own right (18). Subsequently, we reported the use of an oligonucleotide-based reverse genetic approach to disrupt this asymmetry in isolated copies of the viral genome, and showed that genomes containing either opposing doublets or triplets in the hairpin bubble did not give rise to plaques (8). However, plaque-forming mutants were isolated at low frequency, likely derived from imperfections in the synthetic oligonucleotide pool, and were found to contain second site mutations that restored the asymmetry, either by changing the spacing or crippling one PIF binding site. These mutations either inactivated an active inboard arm, or activated an inactive outboard form of OriL, a polarity that strongly suggested that, at least in the genus Parovirus, an active inboard OriL is highly detrimental to growth. However, this reverse genetic approach did not generate a cloned form of the mutant DNA that could be used to analyze exactly which aspects of the infectious cycle were impaired. In the present study, we use an alternate strategy, diagrammed in Figs 1A and B, in which the mutation is first cloned into a modified infectious plasmid form of MVM, to re-examine the effects of introducing an active inboard OriL. These studies suggest that gene expression proceeds relatively normally for such viruses, but that the DNA amplification mechanism is severely defective.
Although there is remarkable variation in the hairpin telomeres from different genera of the Paroviridae (39), within members of the genus Parovirus the nucleotide sequence and predicted structure of the left-end hairpin ears are highly conserved, suggesting that they maybe involved in essential interactions during the virus lifecycle. For example, the junction region between the ears and stem of the hairpin has been reported to interact with the capsid, as indicated in Fig 1B (41), suggesting that these sequences might potentially mediate important interactions such as the packaging process. We have used the same plasmid-based reverse genetic approach to explore the phenotypes of two mutants in which this conserved sequence arrangement has been disrupted, either by inverting the ears, as detailed in Figure 1C, or by modifying the hairpin branch region by sequence transversion, as shown in Fig 1D, while conserving its overall nucleotide composition.
Materials and Methods

Cells and viruses. The fibrotropic prototype strain of MVM (MVMp, GenBank accession number J02275), derived by transfection of the infectious plasmid clone pdBMVp (31), was grown in monolayer cultures of A9 ouabr11 cells, in Dulbecco’s modified Eagle’s medium (DMEM) containing 5% fetal bovine serum (FBS) and antibiotics.

Mutant plasmids. Infectious clones of MVMp with mutant left end hairpins were generated using a modified vector, pCLIP-distal, as previously described (35). Briefly, mutation-bearing oligonucleotides were cloned into the recipient vector, which had been predigested with SapI, leaving non-complementary overhangs within authentic MVM hairpin sequence that facilitate directional cloning of substitute hairpin sequences. To generate pGAGA, pFiFo, and pEMu, the following pairs of synthetic oligonucleotides were annealed, ligated into SapI pre-digested pClip-distal, and the mixture transformed into the Sure-2™ strain of E. coli (Agilent Technologies, Wilmington, DE).

pGAGA: top strand oligo: 5’- TCACGTAAAGTGACGTCAGTGACGCAGCTAGGCTACTAACGT T
CTGCGCGCGCTGCTTCCGGAACGTACACGTCACCTAACGGGAAGTTA AAGCGGTTTCAGG
TACATGGA -3’; bottom strand oligo: 5’- GTTTA AACTCCCTGAAACCGCT
TATCATTTTAGAACATGACCCACCATTGTGACACGTAAGTG

10
To overcome problems caused by polymerase slippage when sequencing through complete hairpins, palindromic termini in cloned DNA were digested with BssHII, which cleaves between the hairpin ears, before being subjected to DNA sequencing.

**Virus stocks.** To generate virus stocks, subconfluent A9 monolayers were transfected with 5μg of infectious plasmid DNA using Superfect (Qiagen, Valencia, CA) and
cultured overnight. Next day cells were sub-cultured 1:4 on 10cm plates, which were then incubated until they showed cytopathic effects (typically 72h). Cells and medium were harvested, virus purified on iodixanol gradients as previously described (16), and quantitated on alkaline agarose gels as described below.

**Protein expression and western transfers.** A9 cells were seeded as monolayer cultures at 25% confluence and infected with 10,000 viral genomes per cell (vg/cell). Plates were rocked every 30 min for 4h, viral inocula removed, and the cultures incubated for a further 2h with fresh medium containing 0.04 units per ml of neuraminidase (Clostridium perfringens, Type V, Sigma, St. Louis, MO) to remove surface virus, after which the medium was changed once more, again to medium containing neuraminidase to prevent re-infection. Cells were harvested 48h post infection (p.i.) by scraping, collected by centrifugation in aliquots, resuspended in phosphate buffered saline containing EDTA-free Complete™ protease inhibitor cocktail (Roche, Branchburg, NJ) and frozen. Proteins were separated on discontinuous polyacrylamide gels in the presence of sodium dodecyl-sulfate (SDS-PAGE), transferred electrophoretically to Immun-Blot PVDF Membrane (Bio-Rad Laboratories, Hercules, CA), and probed with rabbit antisera as indicated. Blots were developed using HRP-conjugated goat anti-rabbit IgG, and bands detected by ECL according to standard procedures.
Expansion assays. Productive infection was determined using a virus expansion assay (27). Briefly, A9 cells were seeded onto Teflon coated spot slides (Cell-Line Associates, Inc., Newfield, NJ) at 20% confluence and infected at 30, 300, or 3000 vg/cell for 3 h at 37°C. After removal of the inoculum, cells were cultured for 24, 48 or 72 h, before being fixed with 2.5% paraformaldehyde, permeabilized with 0.1% Triton X-100, and stained for NS1 by indirect immunofluorescence using the murine monoclonal antibody CE10 (42) and counterstained with DAPI. The percentage of cells with NS1 positive nuclei was scored using a Nikon OptiPhot epifluorescence microscope fitted with a Kodak digital camera driven by MDS 290 software. Multiple images were quantified by single-blinded analysis in Adobe Photoshop.

DNA replication and Southern transfers. Cells were seeded and infected as for protein expression, including culture in neuraminidase from the point, 2 h p.i., when virus inocula were removed. Typically, cells were harvested at 24 h and 48 h p.i. by scraping into the medium, pelleted by centrifugation, and both cells and medium stored frozen. Where indicated, cells from a single plate were divided into two equal aliquots to allow differential processing for DNA replication and packaging analysis.

Total viral DNA was carried analyzed as described previously (21). Where indicated, samples were digested with EcoRI before analysis by electrophoresis through native agarose gels. Two-dimensional (neutral/denaturing) gel analysis was performed as described previously (17). Briefly, samples were first separated by electrophoresis through a native gel, the gel lane then excised, turned through 90°, and inserted into an
extended well, positioned across the top of an alkaline denaturing gel. For analysis of encapsidated DNA, cell pellets were resuspended in TE8.7 (50 mM Tris-HCl, 0.5 mM EDTA, pH 8.7), virus released by 3 cycles of freezing and thawing, and clarified by centrifugation. Resulting cell extracts, and equivalent samples of culture medium, were digested with micrococcal nuclease and analyzed on denaturing agarose gels, as described previously (27).

For Southern blotting, DNA was transferred to Zeta probe membranes (Bio-Rad) according to standard procedures. Blots were generally hybridized with 32P-labeled, random-primed oligonucleotide probes generated using the full MVM genome as template. To assess strand specificity in the packaging assay, blots were probed with the strand-specific 32P-labeled oligonucleotides NScPOS and NScMIN, which hybridize to positive and negative-sense DNA between MVM nucleotides 1165-1190, respectively (20), and quantitated using a Typhoon Trio Variable Mode phosphoimager (GE Healthcare) with ImageQuant Software.

Recovery of left-end hairpin sequences from mutant RF. Replicative form (RF) DNA of mutant viruses was isolated from infected A9 cells by a modified Hirt procedure (38), and monomer RF was gel purified and digested with BsrGI and BsaAI. The 1,223 basepair fragment was further gel purified and cloned into the BsaAI to BsrGI backbone fragment of pCATCH, a truncated version of the MVMP infectious clone with a single BsaAI site in the left-end palindrome (35).
Virus competition assays. A9 cells were seeded at 5x10^5 per 60 mm dish and infected with 40vg/cell of an equimolar mixture of wild-type virus (carrying the “variant II” probe sequence) and one of the GAGA, FiFo or EMu mutants (with the “standard” probe sequence) for 2h. Cells and medium were harvested after 24h, 48h or 72h. Cells pellets were resuspended in 0.2ml TE8.7, virus released by freezing and thawing (x3) and extracts clarified by centrifugation, before storage at -20°C. For analysis, cell-equivalent amounts of cell extract and culture medium were pooled, digested with micrococcal nuclease to remove all non-packaged DNA, the reaction stopped with EGTA, and standard and variant genome concentrations assayed over a range of dilutions by differential qPCR.

Differential real-time quantitative PCR. The multiplex TaqMan® RT-Q-PCR assay was described previously (35). Briefly, the “variant II” wildtype virus incorporates silent mutations that allow it to bind the fluorescent probe 5′-VIC-T A G G T T C C A G T A G C G A A C T C A T C G C C A-TAMRA-3′, while each of the ‘standard’ viruses bind the fluorescent probe 5′-FAM-T A A G T G C C T G T G G C A A A T T C G T C C C C T-TAMRA-3′. PCR amplification was accomplished with primers GF-UP-TAQ 5′-C A C A A C A A A T C A C A T T C G C T C A G A A A-3′ and GF-DOWN-TAQ 5′-T T G C C C A C G T G T G T G A G T T T-3′. Samples were analyzed using an Applied Biosystems PRISM® 7700 Sequence Detection System instrument and software.
Results

The three mutants are viable but show different levels of fitness. Since parvoviral terminal palindromes are difficult to manipulate, we have developed a method for the directional cloning of synthetic oligonucleotides carrying left-end mutations directly into a deleted plasmid form of the viral sequence, such that they reconstitute the left-end terminal palindrome of the infectious genome (35), as described in the Methods section. Using this approach, we re-assessed the effects of removing a single nucleotide from the bubble asymmetry in the inboard arm of the left-end hairpin, generating a mutant plasmid, pGAGA (opposing GA doublets, Fig 1B), which should be nicked on both sides of the palindromic junction sequence present in dimer RF. Similarly, the significance of the asymmetric ears of this hairpin was probed using two mutants: pFiFo (Flip stem with Flop ears, Fig 1C), in which the stem of the hairpin remains in its original Flip orientation but the ears are inverted, and pEMu (Ears Mutated), in which the nucleotide composition of the putative capsid-binding domain at the base of the ears is conserved, but its sequence replaced by a series of transversions (Fig 1D).

After confirming their DNA sequences, mutant plasmids were transfected into sub-confluent monolayers of A9 cells, which were then cultured for several days until they showed cytopathic effects (CPE). In cells transfected with wildtype and the two hairpin ear mutant plasmids, pFiFo and pEMu, CPE was apparent between 72 and 96 hours post transfection, and rapidly led to extensive cell lysis. By 120 hours, we also observed CPE in cells transfected with the hairpin stem mutant pGAGA, indicating that
this mutant is viable, albeit substantially impaired relative to wildtype. This was unexpected, since we had previously shown that a similar mutant, constructed as a ligated synthetic hairpin molecule, was unable to generate plaques following transfection (8).

Full virus particles, containing the viral genome, were extracted from cells transfected with each mutant plasmid, purified by sedimentation to equilibrium in iodixanol step gradients, and viral genomes quantified by denaturing gel electrophoresis and Southern transfer (27). The ability of these titered virion stocks to support the expression of each of the viral proteins in a single round infection was then examined. As shown in Fig 2A, each mutant was able to express all of the viral proteins (NS1, NS2, VP1 and VP2), but whereas FiFo achieved protein levels that were similar to wildtype at this time and input multiplicity, viruses bearing the EMu and GAGA mutations were somewhat impaired (2- to 4-fold) for expression of each protein.

We then examined the kinetics of viral spread in cultures infected at input multiplicities of 3,000, 300 or 30 vg/cell. About 60% of the cells receiving 3,000 vg/cell of wildtype virus expressed nuclear NS1 by 48 hours p.i. (Fig 2B). In contrast, at this time point ~40% of cells infected with the FiFo mutant and only 10 to 20% of cells infected with the EMu or GAGA viruses were NS1 positive. Thus, while viable, the mutants showed variably impaired overall fitness relative to the wildtype virus. By 72 hours p.i. at this input multiplicity, all infections had progressed and essentially all cells were NS1 positive. In contrast, by 72 hours at lower input multiplicities wildtype virus gave 80% NS1 positive cells at 300 vg/cell and 40% at 30 vg/cell, while the level of
infection supported by each mutant was substantially impaired, so that it remained minimal at 30 vg/cell and only affected a subpopulation of cells at 300 vg/cell (~40% for FiFo, 20% for EMu and <10% for GAGA). Thus, while inversion of the hairpin ears in the FiFo mutant did not prevent progeny virion production, it did have a marked effect on the efficiency with which it spread through the culture, while simple sequence transversion at the base of both ears had a much more pronounced effect on viral fitness. Similarly, deleting a single nucleotide from the bubble sequence on the inboard arm of the left-end palindrome, in the GAGA mutant, had a profoundly negative effect on viral fitness, but did allow some virus expansion.

To verify these differences under optimally-matched culture conditions, pairwise co-infections of wildtype and each mutant were assayed using differential quantitative PCR, as described previously (35). For this assay, a wildtype virus marked with a series of synonymous mutations in the coat protein gene was paired with each hairpin mutant in matched, low multiplicity, co-infections that were allowed to develop over several cycles of infection, so that exactly the same conditions held for both viruses. At 24 hour intervals, progeny virus from each co-infection was quantified by PCR, using differentially-tagged probes that distinguished between the two initiating genomes. As seen in Fig 2C, in control cells co-infected with a mix of variant and standard wildtype viruses, the ratio of the initiating genomes remained relatively constant over time. However, when each of the three mutant viruses was mixed with wildtype virus, the mutants proved variably defective. While FiFo and EMu were reduced by two- and five-fold, respectively, over the 72 hours of coinfection, GAGA
was reduced by more than ten-fold over the same period, indicating that this mutant
was significantly more impaired than the other two.

The hairpin ear mutants are stable, whereas the stem mutant is not.

Since we had previously found that the mutation in pGAGA abrogated plaque
formation and we expected the sequence re-arrangements introduced into the left-end
hairpin of pFiFo and pEMu to affect viability more drastically than observed, we
considered the possibility that the mutant genomes had rapidly acquired compensating
mutations or had reverted to wildtype as a consequence of the junction resolution
reaction. To test for either of these possibilities, we used a plasmid capture system,
pCATCH, to isolate the left-end sequences of individual monomer RF molecules, as
described in the Methods. As seen in Fig 3A, all of the left-end termini isolated from
FiFo- or EMu-infected cells maintained the mutant sequence, indicating that these
mutant forms of the terminus could support replication, and that the phenotypes
observed in these infections were due to behavior intrinsic to each mutant. However,
the result obtained for the GAGA mutant was quite different. We had great difficulty in
obtaining clones from GAGA monomer RF isolated at 24 hours p.i., but the two that we
were able to isolate and sequence were identical to the mutant. We then isolated DNA
from infected cells at 48 hours, and obtained a further 12 clones, whose left-end
sequences are summarized in Fig 3B. Of these, only three maintained the original
GAGA left-end terminal sequence, and thus retained the active mutant inboard (GA)
origin. One clone (#6) had an A inserted into the inboard bubble, thus restoring the
inactive wildtype sequence, and the remaining 8 had sustained single nucleotide substitutions. Notably, these substitutions always occurred in the vicinity of the bubble sequence, and six of the single nucleotide mutations (Fig 3B, #’s 7-12) substituted a C for the G of the invariant CpG central dinucleotide of the proximal PIF/GMEB binding half-site. Such mutations eliminate PIF binding to this site (7, 14) and thus would block establishment of the NS1 nicking complex on the inboard arm and effectively inactivate the engineered inboard origin. Of the remaining two second-site mutants, one (#13) substitutes a G for the T immediately beyond the bubble, while the other (#14) substitutes a G for the A in the proximal PIF ACGT half-site. Both of these mutations therefore increase the size of the bubble in the hairpin stem and change the G-C content of the nicking template, while the latter also impairs PIF binding to critical NS1-proximal half-site. Accordingly, the single true revertant and all of the second-site mutations would be expected to suppress the activity of the mutant inboard origin, effectively reverting the replication phenotype of the GAGA mutation. Importantly, since the recovered sequences only modify the inboard arm, the observed changes would not be expected to impair normal viral DNA replication.

Mutation of the hairpin stem, but not its ears, negatively affects viral DNA replication and packaging.

Next we examined the types and quantities of replication intermediates produced by each virus in high multiplicity single-round infections. As seen in Fig 4A, both monomer and dimer replicative form (RF) DNA accumulated to similar levels in
wildtype, FiFo and EMu infections by 24 or 48 hours p.i. However, there was a profound deficiency in RF accumulation in cells infected with the GAGA mutant, so that it was minimal at 24 hours, while by 48 hours it approximated the level seen at 24 hours with the other viruses. Since the GAGA mutant initiated infection and expressed all viral proteins with reasonable efficiency (Fig 2A), it therefore appears likely that its primary defect involves replication initiation or the DNA amplification mechanism.

While both wildtype and FiFo also accumulated significant levels of intracellular progeny single-stranded DNA by 24 hours, as seen in Fig 4A, this had diminished by 48 hours, at least in part due to virus release (see below), while the levels for EMu always appeared lower and for GAGA were undetectable. However, total DNA extraction followed by native agarose gel electrophoresis is a sub-optimal method for quantifying single-stranded DNA because the released single-strands undergo variable intra- and inter-molecular annealing with other partially single-stranded DNAs in the mixture and, rather than reliably migrating as denatured single strands, run as a smear down the gel ending at the position of the denatured strands. To determine packaging efficiency under more rigorous conditions a nuclease protection packaging assay (27) was used in which infected cell extracts were digested with micrococcal nuclease to remove unprotected DNA, the enzyme inactivated, and virion DNA extracted and analyzed by electrophoresis through denaturing alkaline gels. Packaged DNA was quantified by Southern blotting using two different, strand-specific probes, as described in the Methods. Single-round infections were carried out in the presence of neuraminidase, which prevents progeny virions from initiating a second round of
infection, and samples from the final culture medium were included in the packaging assay to analyze virus released from the cell during infection. The results of such an analysis are shown in Figs 4B & C. No packaged input virus was seen at 6 hours p.i. for any of the viruses, and the great majority of packaged viral DNA at 24 hours was found in the intracellular fraction (Fig 4B), while by 48 hours, approximately 50% of the total virions had been released from the cell (Fig 4C). Accumulation of packaged DNA by the EMu mutant appeared slightly retarded compared to that of either FiFo or wildtype, while intracellular levels suggested that packaging of GAGA single-strand DNA was conspicuously restricted. However, GAGA virions were released from the cell with normal kinetics, so that by 48 hours equal accumulations were present in cells and medium. When added together, total packaged GAGA DNA at 48 hours did approximately correspond to the defect in RF accumulation discussed previously. We conclude that none of the mutants have a substantial packaging defect per se, and similarly, that none of them exhibit a conspicuous virion-release defect. Importantly, at all time points, and for both intracellular and extracellular virus, all three mutants paralleled the wildtype in packaging predominantly the negative sense strand, such that any trace of packaged plus strand was observed only at the highest levels of viral DNA input per gel lane. Since strand selection reflects the relative efficiency at which left and right-end genomic termini are resolved (20), this indicates that merely activating a second nick site in the duplex GAGA dimer junction sequence failed to enhance the resolution efficiency of this telomere.
In order to explore the underlying defect(s) in mutant viral DNA replication, especially for the GAGA mutant, the RF pool DNA was extracted from infected cells at 48 hours p.i. and digested with EcoRI, which cuts the predominantly monomeric duplex RF twice. The left and right termini of monomer RF molecules exist in two conformations, either as a covalently closed “turn-around” form, in which both strands of the genome are linked through the hairpin, or as an “extended” form, containing an open-ended duplex copy of the entire terminal palindrome (23), as diagrammed in Fig 5A. Thus, EcoRI digestion yields one homogeneous internal fragment and two terminal fragments that each exist in two forms, which can be separated on neutral agarose gels to give the doublets shown in Fig 5B. Despite the previously discussed differences in overall DNA synthesis observed for these four viruses, the patterns of turn-around versus extended forms of the right-end telomere were very similar at both time points, although at 24 hours GAGA forms were difficult to see and required a longer exposure (Fig 5B). Specifically, at 24 hours, extended forms of the right-end vastly predominated, while by 48 hours turn-around forms had begun to accumulate, and constituted around 50% of the total termini. While a similar situation prevailed for the left-end fragments of wildtype, FiFo and EMu at these two time points, by 24 hours the left-end terminus of GAGA was represented by approximately equal numbers of extended and turn-around forms, while by 48 hours the larger extended-forms were almost absent, and the single dominant form co-migrated with the turn-around fragment of other viruses.

To confirm that the observed GAGA left-end fragment did represent a terminus in the turn-around configuration, we compared wildtype and GAGA EcoRI digests by
2D electrophoresis (17), during which fragments separated in the neutral dimension are rotated 90° and re-electrophoresed under denaturing conditions. For wildtype DNA, shown in Fig 5C, this resulted in clear separation of the extended form, $L_{ext}$, which migrated with the same molecular length in both dimensions, and the turn-around form, $L_{ta}$, which migrated in the alkaline dimension at a position indicating that it was twice the length that it displayed in the neutral gel. In support of the interpretation that the GAGA mutant accumulated RF with predominantly left-end turn-around termini, Fig 5D shows that the double length form, $L_{ta}$, comprised the bulk of the signal in the 2D gel for this terminus, in contrast to the approximately equal distribution between the two forms shown in Fig 5C for wildtype virus. Thus, a mutation that in vitro effectively activates the nick site derived from the inboard arm of the hairpin (data not shown), and which might thus be expected to double the rate at which extended-forms were generated, in fact suppressed its formation throughout infection, drastically curtailing RF amplification. This suggests that duplex junction resolution involves an additional mechanistic constraint that is not fully envisaged in the current heterocruciform resolution model (25).

Since packaged GAGA DNA was released from infected cells, it is likely that most of the GAGA extended-forms generated by the 48 hour time point had already been released as progeny virus. Remaining RF DNA existed as a covalent duplex, which appeared increasingly difficult to resolve as the infection proceeded, making the culture susceptible to hijack by mutants that did not share the GAGA defect, as
observed (Fig 3). The pCATCH technique used here to capture termini, can only sample telomeres in the extended configuration, which presumably explain why the GAGA fragment was difficult to clone from DNA extracted at 24h p.i., and why those obtained at 48 hours predominantly contained second site mutations.
Discussion

Roles(s) of structure versus sequence in the hairpin ears

In this study we show that the Flip sequence orientation of the hairpin ears at the left-end of the MVM genome is not absolutely required for any step in the viral life cycle, even though this orientation is strictly conserved by the viral replication mechanism. Since the MVM genome is packaged in a 3' to 5' direction (19), it is highly likely that the left-end hairpin is in some way involved in the early stages of packaging. However, we found little evidence for any sequence or orientation specificity in the packaging efficiency of the FiFo and EMu mutants. Instead, recovery of virions following transfection of these mutant plasmids corresponded approximately to that of wildtype, both in kinetics and yield, and infection with the resulting virions gave rise to approximately equivalent levels of viral duplex RF DNA, although such viruses were somewhat less fit than wildtype when used to co-infect the same cell population or when allowed to expand through multiple rounds of infection. This indicates that while the wildtype sequence may be preferred, this sequence is not required to be specific, and thus suggests that it may be the terminal DNA structure, rather than its sequence, that is critical. This finding parallels early observations with AAV2, where terminal substitutions that maintained the forked nature of the ITR were tolerated, suggesting that structure was more important than sequence (5). However, at present precise structural information is not available for any parvoviral termini. In silico polynucleotide folding generates various 2D arrangements of the wildtype, FiFo and EMu hairpins that are closely related to those presented in Fig 1. These are
characterized by relatively small differences in $\Delta G$, suggesting that the structure of the terminus might be quite flexible, and could therefore be trapped in particular 3D configurations in vivo by specific physical interactions. If so, then the transversions present in EMu, or the sequence rearrangement present in FiFo, do not radically alter its potential to form such a structure, since, as shown in Fig 2C, we observed that both mutants are almost as fit as wildtype.

Whether or not the tertiary structure of the terminal hairpin mediates specific interactions with the capsid, it is not yet clear when this might binding occur, and for what step in the viral life cycle it might be required. Since packaging occurs in the 3' to 5' direction, one possibility would be that such an interaction locks the 3' end of progeny strands on to the capsid as a first step in the encapsidation process. This idea fits with the fact that the 3' end of each strand of AAV2, both of which are packaged, would be predicted to form the same structure, since these viruses are homotelomeric. This model would also fit with the predominantly negative strand encapsidation seen for members of the genus Parvovirus, for which the 3' end of the packaged strand comprises the unique Flip sequence of the left-end hairpin. However, there are exceptions to the negative strand packaging rule in this heterotelomeric genus, such as the virus LuIII, which encapsidates both strands (4), even though the structure of the 3' hairpin of the positive strand is predicted to be quite different from that of the negative strand. Significantly, a single base insertion in the OriR sequence of MVM leads to the displacement of approximately equal numbers of each strand, both of which are
encapsidated with similar efficiencies (20), making it hard to envisage how such a left-end packaging interaction might function.

An alternative role for docking the left-end hairpin with the capsid might be to anchor the newly uncoated genome to its particle at the start of infection, since uncoating also proceeds in a 3' to 5' direction (16, 22). This model suggests that the capsid might be involved in targeting the genome to a specific nuclear site or sub-compartment, and that association of the newly emerged duplex template with the capsid might be necessary to anchor it in a location favorable for its active transcription. Such a model would imply that MVM particles containing aberrantly packaged positive strands and MVM genomes with structurally defective left-end hairpins would have a significantly lower infectivity than their normal counterparts. In support of this idea, we have recently been able to package MVM genomes with single-ear hairpins, and find that they enter cells normally, but are unable to initiate non-structural gene expression (Li, Cotmore and Tattersall, unpublished).

A second active OriL perturbs junction resolution, and is rapidly lost in vivo.

In contrast to the ear mutations, removing the asymmetry in the stem sequence that controls asymmetric resolution of this palindrome did prove highly deleterious. The wildtype left-end is arranged so that a potential inboard origin sequence is inactive due to the presence of an additional nucleotide within a spacer region, dubbed the bubble, and in a previous study we had been unable to generate infectious foci from genomes with mutations that activated this origin (8). While we originally concluded
that such mutants were non-viable, the results presented here for the GAGA mutant, using a reverse genetic technique that allows construction of hairpin mutants in plasmid form, reveal that this mutant can, in fact, be recovered as a burst of packaged, infectious virus that carries the original mutation. However, it appears that having two active origins is so deleterious to virus growth that the original mutant is rapidly supplanted by second-site mutants, in which the function of the inboard origin is disrupted. This was achieved either by restoring the additional inserted nucleotide, by disrupting the duplex across the bubble or by mutating the PuCGPy sequence of the bubble-proximal half-site for binding PIF/GMEB. This conclusion is consistent with the polarity observed in our previous analysis, in which second-site mutations restored viral viability either by inactivating the second active inboard origin, or, for mutants in which both origins were defective, by re-activating the mutated outboard origin (8).

Sequential left-end cloning from GAGA-infected cells also indicated that second-site mutations accumulated as a function of viral expansion. That this was occurring within a few cycles of infection is not unexpected, since mutation rates for the paroviruses are equivalent to those seen for RNA viruses, rather than for the more genetically stable double-stranded DNA viruses (34). However, our ability to isolate intact replicating GAGA genomes from early rounds of infection showed that the rate of appearance of compensating mutations was slow enough that meaningful short term experiments could be performed with stocks derived by mass transfection. These revealed that viral gene expression proceeds relatively normally in cells infected with the GAGA mutant, despite the severe limitations in the accumulation of potential
duplex DNA templates.

Thus, the GAGA mutation is not lethal because the assembly or activity of nicking complexes on the inboard arm of the hairpin competes with or otherwise disrupts binding of upstream control elements essential for activation of the P4 promoter (26, 36), thus suppressing early viral transcription, as previously hypothesized (8). Rather, the GAGA mutant appears to be profoundly impaired in its ability to synthesize extended forms of the left-hand end of the genome, whereas wildtype junction resolution generates a distinct 50:50 ratio of extended and turnaround termini, both in vitro and in vivo (17, 21). This result is somewhat counter-intuitive, since one might predict that the ability to nick simultaneously on both sides of the dimer junction would allow subsequent synthesis to resolve the junction by producing extended versions of both resulting termini. However, the current model for left-end junction resolution is based on evidence for a heterocruciform intermediate (25), and for its resolution by a mechanism that initiates with the melting, and hairpin rearrangement, of the extended duplex arm of this cruciform, allowing the 3' nucleotide to prime synthesis back along its parental negative-sense strand and thus creating a telomere in the turn-around configuration. Since the inboard origin in wildtype junctions cannot be nicked, the residual duplex structure then is then believed to await resolution by a rate-limiting mechanism, which we have suggested may involve the periodic melting of the inboard nick site, followed by single-strand cleavage, although compelling experimental evidence for this final step remains elusive. The present data suggest that creating a second active origin in the junction may not totally disrupt the
hetero-cruciform resolution mechanism early in infection, since the ratio of extended to
turn-around left-ends appears to be normal. However, the process goes awry later in
infection, perhaps because attempting to form initiation complexes simultaneously at
both active origins in the same junction is detrimental to the resolution process, and the
frequency with which this occurs escalates as the intra-nuclear concentration of NS1
molecules increases later in infection. Further studies will be needed to probe the
mechanism underlying this defect, but the initial observation does suggest that dimer
resolution is a complex and necessary process that cannot be short-circuited by
providing two active origins at the left-end.

Despite the mutation’s negative effect on DNA amplification and progeny
production, we were able to obtain packaged minus strand mutant genomes from cells
transfected with the cloned form of GAGA. These maintained their mutant left-end
through at least a few rounds of replication, thus it would appear that the mechanism
producing turnaround left-ends from GAGA cannot be absolute. The fact that we could
clone intact mutant GAGA left-ends using the pCATCH approach supports the
conclusion that these extended forms are produced from mutant copies of the viral
DNA, and not from emerging revertant genomes.

Influence of asymmetric junction resolution on the hairpin ears.

While a significant diversity of terminal hairpin sizes and predicted structures
exists across the family Paroviridae, these properties are remarkable conserved within
individual genera, despite the fact that the phylogenetic clustering of paroviral species
uses coding sequences and ignores those nucleotides that constitute the terminal palindromes (39). Likewise, all of the members of each genus are either homo- or hetero-telomeric, properties that appear to have profound consequences for the genetic strategy and natural history of each genus. Taken with the observation that the terminal sequences are under evolutionary selection to remain conserved during speciation within the genus itself, this suggests that the hairpins play an important roles or roles in establishing and maintaining the differences in lifestyle adopted by each genus.

While these roles mostly remain enigmatic, there is at least one well-characterized function ascribed to the hairpin ears in a member of the genus Dependovirus. Early methylation interference and nuclease protection studies showed that proteins isolated from infected cells make contact with one of the ears of the AAV2 ITR (1, 29). This interaction was shown to occur between the major replicator protein, Rep68/72, and a sequence of five bases, called the Rep Binding Element’ (RBE’), located at the tip of one of the hairpin ears (37). In the absence of the RBE’, origin firing by Rep is quite inefficient, and cannot be enhanced by providing the RBE’ in trans (11). The cis-acting nature of this interaction has led to the suggestion that it alters the structural stability of the Rep-Ori complex (28), perhaps by rendering the cleavage site single-stranded (6).

In contrast, the OriL of MVM is highly active as an isolated 60 basepair duplex sequence, embedded in a circular plasmid, in the complete absence of hairpin ear sequence elements (18), and the stabilization and activation of the MVM NS1 nickase at OriL is achieved by formation of a precise ternary complex with the host factor.
PIF/GMEB (12). Likewise, although efficient replication of the AAV2 genome requires
the Rep/RBE' interaction, this requirement cannot be absolute, since Rep also cleaves
AAVS1, its recognition site for integration on human chromosome 19, in the absence of
any adjacent RBE'-like sequence (32, 33).

While the results presented here do not define a particular role for the hairpin
ears, or for the maintenance of the Flip orientation in the wildtype genome, they do
underscore properties of the parvovirus genome for which they are not responsible. On
one hand, neither the orientation or sequences of the ears contribute to the negative
strand selection process employed by members of the genus Parvovirus. Indeed, our
finding that such progeny genomes as the GAGA mutant is able to package, are as
predominantly negative strand as the wildtype, indicates that the asymmetry of dimer
junction resolution plays no role in strand sense selection.

On the other hand, our results show that, although the wildtype version of the
left-end hairpin is observably preferred in head-to-head competition, the Flip
orientation and specific sequence of the left-end ears are by no means essential for viral
viability. Since disruption of the arrangement of active and inactive origins in the
hairpin stem is so strongly selected against, we can conclude that the asymmetry of
dimer junction resolution is critical for some aspect of the viral lifecycle, and that
conserving the Flip orientation of the hairpin ears is merely a consequence of this
mechanism, rather than its raison d'être.
Oligonucleotides used in this project were synthesized in the HHMI Biopolymer Laboratory and W. M. Keck Foundation Biotechnology Resource Laboratory at Yale University. This work was supported by Public Health Service grant AI026109 from the National Institutes of Allergy and Infectious Diseases.


Figure Legends

Fig 1: Arrangement of sequences in the extended left-hand end of MVM.

Panel A shows the extended left-hand end of monomer RF generated during genomic replication, in which the internal palindromes that form the hairpin ears are represented by hatched rectangles. The locations of the BsaAI, Pmel and BsrGI sites mentioned in the text are indicated, as is the 1,223 basepair BsaAI to BsrGI fragment that is captured from infected cell DNA by pCATCH. The middle section diagrams pCLIP-distal, in which a 1.4 kb ΦX174 'filler' DNA sequence was engineered to include terminal SapI sites, inserted between the BsaAI and Pmel sites. SapI cleaves outside of the insert into genomic sequence, allowing the direct insertion of annealed oligonucleotides incorporating the desired mutation(s), as shown in the lower section.

Panel B shows the features of the MVM LHE hairpin. The shaded region was reported to be protected from hydroxyl radicals by capsid interactions (41). A cellular heterodimer, PIF, binds to spaced proximal and distal 5'-ACGT-3" half sites. The BsaAI site that overlaps the proximal PIF half-site in the duplex form of the "outboard" strand is indicated by box with dashed lines. The nucleotide missing in the pGAGA mutant is indicated in bold.

Panels C and D show, in bold, the reorganized sequences of the hairpins ears in the pFiFo and pEMu infectious clones, respectively.
Fig 2. Comparison of mutant infection parameters with those of wildtype virus.

Panel A. Western blot analysis of intracellular viral protein expression at 48 hours p.i. Viral proteins were detected using polyclonal antibodies against the common amino-terminus of the nonstructural proteins NS1 and NS2 (top and upper middle panel), and the structural proteins VP1 and VP2 with polyclonal antibodies against denatured capsid protein (lower middle panel). Reactivity to antibody against β-actin was used as a loading control (bottom panel).

Panel B. Viral expansion kinetics in multiple-round, single virus infections of A9 cells at 3,000 (.), 300 (○) or 30 (□) vg/cell of wildtype or each of the three mutants, measured by NS1 expression.

Panel C. The fitness of each mutant relative to wildtype, as determined by differential qPCR in multiple-round co-infections of A9 cells with matched input multiplicities (40 vg/cell) of “standard” mutant and “variant” wildtype virions, harvested at the indicated times p.i.

Fig 3. Isolation and characterization of revertants.

Panel A. Total numbers of sequenced pCATCH clones isolated from duplex viral RF DNA, characterized as unchanged, revertants, or second-site revertants.

Panel B. Catalog of sequences for the 14 left-end clones isolated from GAGA mutant virus-infected cells, as described in the text.
Fig 4. Analysis of mutant and wildtype viral DNA replication and packaging.

Panel A. Agarose gel and Southern blot analysis of total viral DNA extracted from cells 24 and 48 hrs p.i. with 10,000vg/cell input under single cycle infection conditions. Migration positions of dimer replication form (dRF), monomer (mRF) and progeny single-strands (ss) are indicated.

Panels B and C. Equivalent aliquots of cell extract (panel B) and culture medium (panel C) were digested sequentially with micrococcal nuclease and proteinase K, as described in the Methods, then electrophoresed through an alkaline agarose gel, blotted and probed with labeled oligonucleotides specific for the negative (-) or positive (+) strands.

Fig 5. Analysis of left- and right-end forms of replicating wildtype and mutant DNA.

Panel A. Cartoon depicting the structures of the two forms of each end, as previously determined (23).

Panel B. Total DNA from single-cycle infections were digested with EcoRI, electrophoresed through a neutral agarose gel, then blotted and probed as described in the Methods. The lane indicated in bold is a long exposure of a DNA sample extracted from GAGA-infected cells at 24 hr p.i.

Panels C and D. Two dimensional agarose gel electrophoresis of total DNA extracted from wildtype (panel C) and GAGA (panel D) virus-infected cells. DNA was first run
in a non-denaturing neutral gel, turned through 90° and run into a denaturing alkaline gel, followed by transfer and probing as described in the Methods.
A.

- Outboard arm
- Inboard arm
- BsaAI
- 30
- 1223
- Pmel
- 108
- BsrGI

B.

- PIF binding
- Distal
- Proximal
- BsaAI
- NS1 footprint
- Big ear
- Little ear
- SapI
- 1.4 kb
- Stuffer
- Mutant MVM
- Hairpin
- pGAGA

C.

- pFiFo
- Little ear [anti-parallel]
- Big ear [anti-parallel]

D.

- pEMu
- Big ear
- Little ear
A

<table>
<thead>
<tr>
<th>pCATCH</th>
<th>GAGA</th>
<th>FiFo</th>
<th>EMu</th>
</tr>
</thead>
<tbody>
<tr>
<td>sequenced clones</td>
<td>14</td>
<td>14</td>
<td>8</td>
</tr>
<tr>
<td>same as input</td>
<td>5</td>
<td>14</td>
<td>8</td>
</tr>
<tr>
<td>mutant to wildtype</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>other changes</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

B

wildtype

<table>
<thead>
<tr>
<th>outboard strand</th>
<th>wildtype</th>
<th>NS1 binding</th>
</tr>
</thead>
<tbody>
<tr>
<td>5' - ACACGTCACTTTACCAGTACGTG - 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3' - TGCAGTGAATGCAAGTGTACCAACCA - 5'</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

pGAGA:

<table>
<thead>
<tr>
<th>inboard strand</th>
<th>pGAGA:</th>
</tr>
</thead>
<tbody>
<tr>
<td>3' - TGCAGTGAATGCAACCAACCA - 5'</td>
<td></td>
</tr>
</tbody>
</table>

pCatch GAGA:

| 1 - 5 | 3' - TGCAGTGAATGCAACCAACCA - 5' |
| 6    | 3' - TGCAGTGAATGCAACCAACCA - 5' |
| 7-12  | 3' - TGCAGTGAATGCAACCAACCA - 5' |
| 13    | 3' - TGCAGTGAATGCAACCAACCA - 5' |
| 14    | 3' - TGCAGTGAATGCAACCAACCA - 5' |
A - intracellular viral DNA

- wildtype
- GAGA
- FiFo
- EMu

M - hours post infection

6  24  48

B - cell-associated virions

- M

C - released virions
A

1085  2434  1630

L_{ext}  M  R_{ext}

L_{ta}  M  R_{ta}

B

hours post infection

<table>
<thead>
<tr>
<th>Time (hours)</th>
<th>WT</th>
<th>GAGA</th>
<th>FiFo</th>
<th>EMu</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C - wildtype

2nd dimension - denaturing

D - GAGA

2nd dimension - denaturing