29

Parvovirus

Susan F. Cotmore[1] and Peter Tattersall[1, 2]

Departments of [1]Laboratory Medicine and [2]Genetics
Yale University Medical School
New Haven, Connecticut 06510

THE *PARVOVIRIDAE* (*PARVUS* FROM LATIN, MEANING SMALL) are the only known virus family with linear single-stranded DNA genomes. Their chromosomes range from 4 to 6 kb in length and consist of a relatively long (<5.8 kb) single-stranded coding region, bracketed by short imperfect terminal palindromes that fold back on themselves to form duplex hairpin telomeres. Together with a few adjacent nucleotides, these hairpin termini provide all of the *cis*-acting information required for both viral DNA replication and progeny genome encapsidation, serving both as origins of replication and as critical hinges that allow quasi-circular amplification of the linear chromosome through a series of duplex intermediates. Parvoviruses infect a broad range of invertebrate and vertebrate hosts, from arthropods to man, and although some, such as feline panleukopenia virus (FPV), can be lethal, many others are conspicuously apathogenic. They have rugged T = 1 icosahedral protein capsids, 18–26 nm in diameter, capable of shielding the genome during its sometimes extended extracellular existence, and ultimately transporting it directly into the host cell nucleus. Here we focus on viruses that infect mammals, from the subfamily *Parvovirinae*, which comprises five genera, of which at least two, the *Dependoviruses* and the *Erythroviruses,* contain important members that infect human hosts (Tattersall et al. 2005). These illustrate two very different lifestyles adopted by members of the *Parvoviridae*, for whereas the human *Dependoviruses*, generally referred to as the adeno-associated viruses (AAVs), only replicate productively in cells coinfected with a helper virus, most commonly an adenovirus or herpesvirus, but potentially also a papillomavirus, the *Erythroviruses* repli-

DNA Replication and Human Disease 0-87969-766-0

cate autonomously but are restricted to actively dividing cell populations, in most cases in the bone marrow. To date, replication strategies have been best explored for AAV2, an apathogenic defective *Dependovirus*, and for Minute Virus of Mice (MVM), an autonomously replicating species from the genus *Parvovirus*. Many of the basic mechanisms used to replicate these genomes were described in detail in the 1996 edition of this monograph (http://dnareplication.cshl.edu) and are not reiterated here, in order to devote space to broader perspectives and new developments. Structure, biology, and pathogenicity of these viruses, and their potential application as gene therapy vectors, are reviewed in detail elsewhere (Muzyczka and Berns 2001; Tattersall and Cotmore 2005; Kerr et al. 2006).

Single strands are more flexible than duplex DNA, so parvoviruses can condense a remarkable amount of genetic information into their small capsids. However, this simplicity inevitably limits their parasitic potential, as incoming virions lack accessory proteins, chromatin, or even a duplex transcription template. They therefore remain silent within their host cell nucleus until the cellular synthetic machinery manufactures a complementary DNA strand, which creates a transcription template. For the autonomously replicating viruses this typically occurs when the cell enters S phase, of its own volition, and is rapidly followed by expression of the first viral transcripts driven from an S-phase-responsive viral promoter. Little is known about the mechanism(s) of persistence during G_0/G_1, and despite convincing serologic and PCR evidence for long-term persistence by many parvoviruses, and their frequent re-emergence following immunosuppression, essentially nothing is known of the mechanisms underlying this type of latency.

In contrast, three modes of persistence have been documented for the AAVs. First, AAV2 and some other serotypes integrate their genomes site-specifically into a 4-kb locus, designated AAVS1, on human chromosome 19q13-qter. This occurs via a replication-dependent mechanism, requires low-level expression of the AAV replication proteins, and depends on sequences in AAVS1 that can function as an AAV origin. The second mode involves nonspecific integration and is primarily seen with recombinant AAV (rAAV) genomes, in which foreign genes, flanked by AAV termini and packaged within AAV capsids, are used for gene transduction. Integration of rAAVs occurs at multiple positions throughout the host genome, with a bias toward actively transcribed loci, and most efficiently at sites of naturally occurring double-strand breaks (Miller et al. 2004). A third form of persistence involves the establishment of monomeric, and later concatemeric, circular duplex episomes, following vector delivery at high copy number to nondividing cell populations, such as muscle (Duan

et al. 1998). Similar episomes have also been detected in vivo in latently infected human tissues, although whether they are created by annealing of complementary strands or by cellular repair mechanisms is currently unresolved.

BRIEF OVERVIEW OF ROLLING-HAIRPIN REPLICATION

The parvoviral chromosome comprises a non-permuted, single-stranded, linear DNA molecule, flanked by small, palindromic telomeres that are capable of adopting a self-priming hairpin configuration, as depicted for MVM and AAV2 in Figure 1. These termini, which range from 120 to 420 bases in size, can be remarkably disparate in sequence and secondary structure, both within individual genomes and between species. However, they all contain at least two essential replication elements: a DNA replication origin that allows the viral initiator protein to nick the DNA, effectively separating unit-length genomes and creating a base-paired DNA primer that supports the establishment of new replication forks; and a hinge element involving apical asymmetries, which lowers the free energy of hairpin extension, coupled with an appropriately spaced duplex DNA-binding site for the viral initiator protein that allows it to destabilize both duplex extended-form and hairpin forms of the telomere (Cotmore and Tattersall 2006). Some genera have inverted terminal repeat (ITR) sequences, so that equivalent palindromes are present at each end of the genome, and the replication and resolution processes associated with both termini are essentially similar, as illustrated here by AAV2. In contrast, others, such as MVM, have two physically and functionally disparate terminal palindromes and employ different resolution strategies at the two sites, although the molecular principles that underlie both strategies are very similar. Mammalian parvoviruses encode all of their known proteins from a single-sense DNA strand, and are unusual in encapsidating, in separate virions, strands that are either positive- or negative-sense with regard to transcription. Some parvoviruses, such as AAV2, package equimolar copies of both strands, in separate virions, whereas others, such as MVM, selectively encapsidate strands that are minus-sense. In general, the parvoviruses possess two gene cassettes, with transcripts from one-half of the genome, by convention the "right-hand" side, programming synthesis of an overlapping set of capsid polypeptides, whereas the left half gives rise to the viral replication proteins (Fig. 1). In AAV2, transcriptional promoters at map units 5 (P5) and 19 (P19) access a single open reading frame in the *rep* gene, and an intron at the 3′ end of the gene allows expression of either a spliced or unspliced version

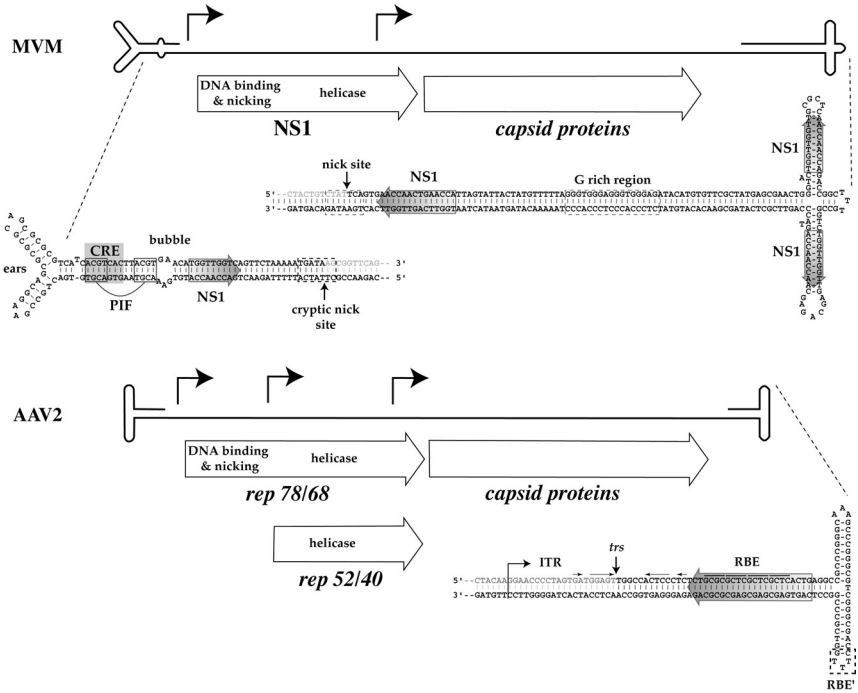

Figure 1. Comparison of genome strategies across the *Parvovirinae*. Diagrams of representative genomes from MVM and AAV, with their terminal hairpin structures shown in detail. The positions of promoters are indicated by arrows, and their major gene blocks by open arrows indicating the amino- to carboxy-terminal direction, in which their major functional domains are delineated. For AAV2, the whole ITR sequence is shown for the right-hand-end hairpin in the "flip" configuration, and the inboard end of the ITR is marked by a rightward arrow. Binding sites for the replicator protein, NS1 or Rep, are indicated by the shaded arrow boxes, and their sites of nucleolytic cleavage are represented by vertical arrows. For MVM, the left end is not a nicking substrate in the hairpin configuration, and this site is denoted as cryptic.

of each transcript. This generates two large proteins, Rep 78 and Rep 68, which can both act as initiator nucleases, and two small proteins, Rep 52 and Rep 40, which are essential for progeny genome encapsidation. MVM, in contrast, expresses only full-length forms of its initiator, NS1, as well as a series of alternatively spliced, polyfunctional NS2 molecules that are dispensable for replication in certain transformed cell types. The initiator proteins all carry two separate enzymatic cores: an amino-terminal site-specific single-strand nuclease domain (Hickman et al. 2004), bearing active-site protein motifs common to all rolling-circle initiators, which rec-

ognizes and nicks the replication origins; and a carboxy-terminal superfamily III helicase domain, which belongs to a group of viral 3′-to-5′ helicases evolutionarily related to the extensive AAA^+ family of cellular ATPases (James et al. 2004). Genome usage in all parvoviruses is remarkably efficient but, nevertheless, the viruses' limited genetic capacity dictates that they predominantly rely on the replication machinery of the host cell, using RP-A to support their extensive helicase-mediated strand-displacement activity and to recruit DNA pol-δ-based, leading-strand replication forks (Bashir et al. 2000; Christensen and Tattersall 2002).

Parvoviral genomes are amplified via a unidirectional, strand-displacement mechanism called rolling-hairpin replication (RHR), which appears to be a linear adaptation of the ancient rolling-circle replication (RCR) mechanism, practiced by some bacteriophages and bacterial plasmids, and by eukaryotic Geminiviruses and Circoviruses. RCR replicons employ distinctive duplex replication origins that contain small palindromic DNA regions, which switch between inter- and intra-strand base-pairing at different phases of the replication process. In the parvoviruses, this use of palindromic rearrangement has been expanded, so that there are two viral origins, generally with limited internal symmetries, each embedded in a larger, hinge-like, hairpin telomere that can unfold to be copied, then refold to allow continuous amplification of the linear template. Accordingly, viral replication proceeds in two distinct phases; an initial amplification phase that generates high-molecular-weight duplex replicative form (RF) DNAs, followed by a progeny genome displacement phase, in which individual single-strand genomes are excised and displaced for encapsidation.

As detailed previously (Fig. 1 in Cotmore and Tattersall 1996), the 3′ nucleotide of incoming virion DNA creates a DNA primer, pairing with an internal base to allow a host polymerase to initiate synthesis of a complementary DNA strand. This generates a monomer-length, duplex intermediate in which the two strands are covalently cross-linked at one end via a single copy of the viral 3′ telomere. In viruses with ITRs, such as AAV, this cross-linked 3′ structure creates a replication origin that can be activated by Rep 68/78 in a process called terminal resolution, whereas in viruses such as MVM this (left-end) telomere is refractory to initiation. In either case, all further DNA amplification proceeds by unidirectional strand displacement, with the initiator protein also serving as the 3′-to-5′ replicative helicase, but with all other functions expropriated from the host cell. The replication process creates a series of palindromic duplex dimeric and tetrameric concatemers in which the unit-length genomes are fused in left-end:left-end and right-end:right-end combina-

tions. Successive rounds of replication are then initiated from these concatemers, and progeny single strands are excised, and ultimately displaced, by site-specific single-strand nicks introduced into the duplex telomeric origins by the viral initiator nucleases Rep 68/78 or NS1. As for all RCR initiators, nicking involves a *trans*-esterification reaction that liberates a base-paired 3′ nucleotide and leaves the nuclease covalently attached to the 5′ nucleotide at the nick, via a phosphotyrosine bond. Other RCR initiators carry out sequential nicking and rejoining reactions, but in parvoviruses only nicking occurs, so that initiator proteins remain covalently attached to the 5′ end of progeny genomes, from which position they are available to support additional functions. Subsequent accumulation of displaced progeny single strands appears to be entirely dependent on the availability of preformed capsids, and packaging itself is driven by ongoing viral DNA synthesis.

DEVELOPMENTS IN OUR UNDERSTANDING OF VIRAL ORIGIN STRUCTURE

Parvoviral initiation sites are presented as duplex DNA, which must be melted before they can be cleaved by the viral nuclease. To achieve this, Rep 68/78 and NS1 bind site-specifically to a duplex DNA-binding site positioned near the initiation site and employ their helicase activity to melt the latter, often assisted by the ability of the separated strands to reconfigure into stem-loop structures with the nick site presented on an exposed loop. Although parvoviral origins vary in size and structure, they all contain three distinct DNA recognition elements: the duplex DNA-binding site, the nick site, and a third element that stabilizes interactions between the initiator and the DNA, although the importance and nature of the third element vary widely between origins. DNA–protein interactions at these sites are often assisted by cellular DNA-bending proteins from the HMG1/2 family, creating cleavage complexes with precise ternary structure.

Interactions of the AAV Rep Nuclease with the Viral Hairpin Origins

The AAV2 ITRs contain imperfect terminal palindromes of 125 nucleotides that assume a T-shaped hairpin when folded to maximize base-pairing and serve as origins when presented in this form at the end of a duplex RF DNA intermediate (Fig. 2 in Cotmore and Tattersall 1996). Within these hairpins just two sequence elements are absolutely required for origin function. The first of these is a 22-bp Rep-binding element

(RBE), comprising a reiterated tetranucleotide repeat, 5′-GAGC-3′ (shaded arrow in Fig. 1), that positions Rep 68/78 complexes on the DNA in the correct orientation and position for cleavage. The second is a nicking or terminal resolution site (*trs*), positioned 16 nucleotides inboard of the RBE. A third element, 5′-GTTTC-3′, called the RBE′, which promotes but is not essential for nicking, lies at the tip of the hairpin arm opposite the nick site. Although the hairpin exists in either of two inverted complementary forms, called "flip" and "flop," at both ends of the genome, the RBE′ is present in the same position relative to the nick site in each form. Since the juxtaposition of the RBEs to the Rep complex is determined by the hairpin configuration of the telomere, linear origin sequences, found at the end of many duplex replicative-form DNA intermediates in vivo, present suboptimal substrates for Rep. Thus, although not absolutely required for nicking, the RBE′ enhances origin efficiency by about 20-fold.

The recently determined atomic structures of the nuclease domain of AAV5 and the helicase domain of AAV2 Rep proteins (Hickman et al. 2004; James et al. 2004) inform our understanding of how these molecules interact with viral DNA. Structures of the AAV5 Rep nuclease domain separately complexed with DNA substrates representing the RBE and the RBE′ indicate that five Rep monomers bind five tetranucleotide direct repeats at the RBE. Each repeat is recognized by two Rep monomers from opposing faces of the DNA, and each monomer makes contact with two nucleotides from each of two adjacent direct repeats, whereas the RBE′ is contacted by a single monomer from the Rep complex (Hickman et al. 2004). The *trs* has been defined as a seven-nucleotide sequence flanked by short palindromic sequences (horizontal arrows in Fig. 1) that stabilize exposure of the critical phosphodiester bond by folding into a stem-loop structure (Brister and Muzyczka 1999). This also effectively moves the cleavage site toward the active site of the RBE-bound Rep complex, and initiation efficiency at this origin can be enhanced by, but is not dependent on, the cellular DNA-bending protein, HMG1 (Costello et al. 1997). Within this structure, the active site of the Rep nuclease assumes a novel fold, being located at the bottom of a positively charged cleft formed by the spatial convergence of a divalent metal ion and the conserved active-site tyrosine motifs that define the rolling-circle-replication superfamily. Thus, in AAV all three DNA recognition sites in the origin are contacted by elements in the amino-terminal domain of Rep 68/78, which exhibits an unexpected structural relationship to other viral origin-binding proteins such as the papillomavirus E1 protein and the SV40 T antigen (Hickman et al. 2004). Binding of Rep to the RBE

also appears to be essential for both unfolding and refolding the duplex during amplification of RF DNA and hairpin transfer. Rep helicase is activated upon binding to an RBE and likely serves to unwind the duplex, whereas the hairpin ears provide the necessary destabilizing hinge element.

Structure of the Cleavage Complex at the Right-end Hairpin of MVM

Whereas the minimal sequence requirements for nicking at AAV origins occupy less than 40 bp, the minimal MVM right-end origin is 125 bp long (~256 nucleotides) and involves the entire hairpin (Cotmore et al. 2000). This is because a third recognition element, an additional NS1-binding site located about 120 bp from the nick site, immediately next to the axis of the hairpin, is also absolutely required. However, like its AAV counterparts, this telomere is resolved in its hairpin configuration by a terminal resolution reaction, generating progeny in both "flip" and "flop" sequence orientations. This palindrome forms an almost perfect duplex, with just three unpaired bases at the axis and a single mismatched region in the stem. Here a 3-nucleotide insertion (AGA or TCT) on one strand separates opposing pairs of NS1-binding sites, creating a 36-bp palindrome that can assume an alternate cruciform configuration with little change in free energy (as shown in Fig. 1). This cruciform rearrangement could contribute to the instability of the duplex, facilitating its use as a hinge. Unfolding and refolding the hairpin during RHR requires both the helicase and site-specific binding activity of NS1, and requires the NS1-binding sites near the tip of the hairpin, rather than the nick-site-proximal sequence (Willwand et al. 2002).

Although the consensus MVM nick-site sequence, 5′-CTWWTCA-3′, cannot be deleted or inverted, the importance of its individual nucleotides remains unknown. However, insertion of an additional AT dinucleotide immediately 3′ to the nick site in the closely related virus LuIII reduces the efficiency of this otherwise dominant origin, and, as discussed below, promotes displacement and packaging of both positive- and negative-sense DNA strands.

NS1 binds site-specifically to duplex DNA at reiterations of the tetranucleotide motif, 5′-ACCA-3′, present in both the viral replication origins and repeated at multiple sites throughout the genome (Cotmore et al. 1995). This suggests a role for NS1 in the formation of viral chromatin and explains why nicking is more tightly controlled for MVM than for AAV. Oligonucleotide selection studies indicate that optimal forms of

the site contain at least three tandem copies of the ACCA repeat, but that 50% degeneracy in one of these motifs or insertion of a single base between them has little effect on affinity (Cotmore et al. 2000). This suggests that, like Rep (Hickman et al. 2004), NS1 binding is modular, with each tetranucleotide motif being recognized independently by separate molecules in the NS1 complex. In the presence of ATP, which promotes assembly of NS1 into higher-order multimers, NS1 binds asymmetrically over these motifs, protecting a 41–43-bp sequence that approximately aligns with the 3′ end of the $(ACCA)_3$ sequence but extends about 20 bp 5′ of the consensus. At the nick site, this 5′ extension positions NS1 to nick the origin, whereas in the inverted axial NS1-binding sites that create the potential cruciform structure, they should overlap. DNase I footprints show that the inverted juxtaposition of these two sites increases the binding affinity of NS1 to each, without shifting the footprints from their expected positions, suggesting that a more stable rearranged or higher-order NS1 complex is formed.

Interactions between NS1 molecules bound to these elements are potentiated by a DNA-bending, HMG1/2 family member, which is absolutely required for formation of the cleavage complex (Cotmore et al. 2000). Footprinting studies show that addition of HMG1 induces only very slight shifts in the sequences protected by NS1, but the conformation of the intervening DNA changes, folding into a double-helical loop that extends for about 30 bp through a characteristic G-rich element in the hairpin stem. Mutant origins that cannot reconfigure into a double-helical loop upon addition of HMG1 fail to nick, indicating that the loop allows the terminus to adopt the correct three-dimensional structure needed to activate the nickase.

The OriL of MVM, Its Cellular Cofactor, PIF, and Heterocruciform Dimer Resolution

As detailed previously (Fig. 3 in Cotmore and Tattersall 1996), the left-end telomere of MVM cannot function as a replication origin in its hairpin configuration. Instead, it generates a single competent origin when the hairpin is extended and copied during replication to form the fully base-paired palindromic junction sequence that spans adjacent genomes in dimer RF. Resolution of these left-end:left-end junctions proceeds by an asymmetric process termed junction resolution that results in the generation of left-end termini in a single sequence orientation, "flip." This process relies predominantly on the same sequential extension and refolding reactions seen during RHR amplification, but uses an asymmet-

ric template in which the nick site on one hairpin arm is nonfunctional. This ability to segregate one inactive arm of the palindrome appears essential for the virus, because mutated genomes with active origins on both arms are not viable.

The minimal linear origin is about 50 bp long, extending from two 5′-ACGT-3′ motifs spaced 5 nucleotides apart at one end, to a position some 7 bp beyond the nick site, and spanning an asymmetry that, in the hairpin stem, creates the "bubble" mismatch, where a GAA triplet on the inboard arm opposes a GA doublet in the outboard strand (Fig. 1). Only the sequence from the outboard arm of the stem, containing the GA doublet, serves as an origin, called $OriL_{TC}$. The actual sequence of the bubble dinucleotide is relatively unimportant, whereas insertion of any third nucleotide within it, such as in $OriL_{GAA}$, inactivates the origin, indicating that the bubble is a critical spacer, rather than a recognition element in its own right. Within the origin, the three essential recognition sequences are the NS1-binding site, the consensus nick site, 5′-CTWWTCA-3′, positioned 17 nucleotides downstream, and the two ACGT motifs. The outboard ACGT motif also forms part of a cAMP-responsive element that modulates gene expression from the P4 promoter, which we initially suggested to be an essential origin element. We now know that this is incorrect, and that the critical interaction involves a cellular factor variously called PIF, for parvovirus initiation factor, or GMEB, for glucocorticoid modulating element binding protein (Cotmore and Tattersall 2006). PIF is a heterodimeric site-specific DNA-binding factor comprising p96 and p79 subunits, which functions as a transcriptional modulator and binds DNA via a novel "KDWK" fold, or SAND domain (Surdo et al. 2003). Its binding specificity is unusual because it recognizes two ACGT half-sites spaced from 1 to 9 nucleotides apart, with an optimal spacing of 6 nucleotides (Burnett et al. 2001). This factor contacts NS1 over the bubble sequence to stabilize the binding of NS1 on the active form of the left-end origin, $OriL_{TC}$, but not on the inactive form, $OriL_{GAA}$, due to the extra nucleotide (Christensen et al. 2001).

Given the precise location of the active origin, $OriL_{TC}$, in the dimer junction, synthesis of new copies of the left-end hairpin in its unique "flip" sequence orientation is clearly not a straightforward process, since replication forks progressing from this site through the linear bridge structure would be expected to synthesize new DNA in the opposite "flop" orientation. However, this process occurs via a cruciform intermediate that allows synthesis of new DNA in the single, correct, sequence orientation, and creates an unstable intermediate structure that can ultimately be resolved by the NS1 nickase. According to this "heterocruci-

form model" (Cotmore and Tattersall 2003), resolution is driven by the helicase activity of NS1 and depends on the inherent instability of the duplex palindrome, a property that allows it to switch between a linear and a cruciform structure. A preliminary form of this model (Fig. 4 in Cotmore and Tattersall 1996), postulated that the final step in the resolution process might be mediated by a cellular recombinase, but experimental data now suggest a different mechanism. Again based on the ability of the 5′-linked NS1 helicase to destabilize a duplex, allowing both palindromes to fold back on themselves, but this time coupled with the known ability of NS1 to nick single-stranded forms of its left origin without the help of its usual stabilizing cofactor, PIF (Nuesch et al. 2001), the current model suggests that the helicase generates a dynamic intermediate structure in which the nick site in the normally inactive $OriL_{GAA}$ arm is transiently but repeatedly exposed in a single-stranded form, allowing cleavage to occur.

PROGENY SINGLE-STRAND SYNTHESIS AND ENCAPSIDATION

Whereas MVM encapsidates predominantly negative-sense DNA, its close relative LuIII, which has an impaired right-end origin, packages approximately equal numbers of plus and minus strands. This suggests that strand selection for encapsidation does not involve specific packaging signals, but rather that the relative efficiency with which the two genomic termini are resolved and replicated ultimately determines the polarity with which single strands are excised from replicating RF, and that these are then packaged with equal efficiency. Experimental evidence supporting this model was obtained in vivo using forms of the LuIII virus that differ by just the two-nucleotide nick-site insertion, and which consequently do, or do not, package positive-sense DNA (Cotmore and Tattersall 2005b). These kinetic analyses indicate that duplex RF forms of the genome do not serve directly as packaging substrates, and suggest a single-strand displacement model (recapitulated in Fig. 2) in which the substrate for encapsidation is a newly released single strand. Functional asymmetries between the right and left hairpins restrict release of positive-sense strands from replicating RF during the packaging phase of infection, creating a situation in which these are only released if the right-end nick site is suboptimal, forcing displaced negative-sense strands to cycle through a necessary dimer duplex intermediate.

Genetic disruption of the capsid gene shows that single-stranded DNA only accumulates in cells that are assembling competent viral particles, whereas inhibitor studies suggest that displacement of single-

stranded progeny genomes occurs predominantly or exclusively during active DNA replication. How empty capsids are recruited to newly displaced single strands remains unknown, but foreign DNA sequences can

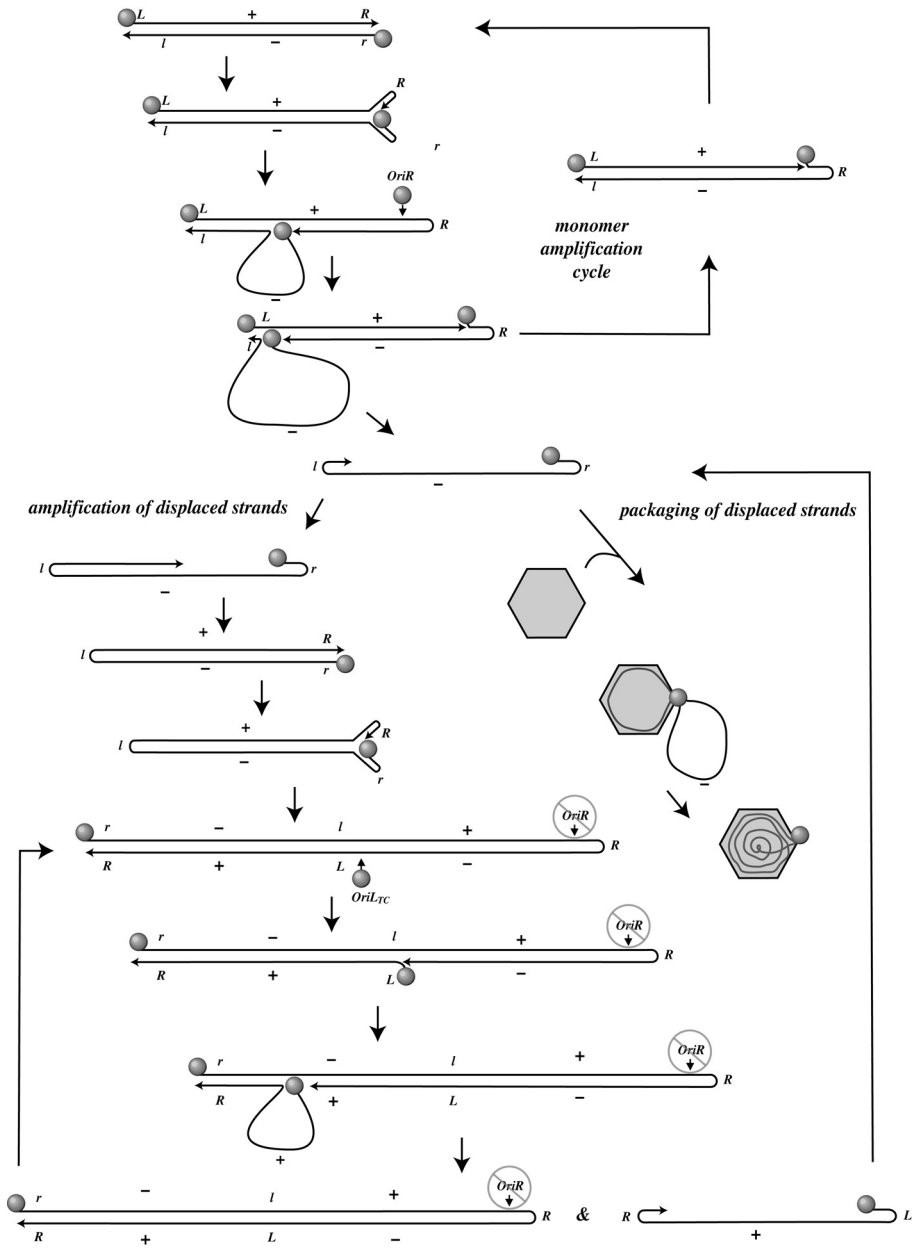

Figure 2. (*See facing page for legend.*)

be packaged during AAV2 infections, albeit somewhat inefficiently, if they simply contain sequences that function as minimal viral nick sites. Although such minimal and disparate initiation sites could also, theoretically, contain "packaging" sequences, a more likely interpretation is that this interaction is mediated by the viral initiator protein, left covalently attached to the 5′ end of the nicked DNA. Thus, genomes may be marked for packaging, at least in part, by their association with Rep 68/78 proteins, which bind efficiently to intact empty capsids.

Genomes are translocated into preformed capsid particles via a reaction that, for AAV2, has been shown to require the helicase activity of the Rep 40/52 proteins (King et al. 2001). DNase protection studies suggest that insertion of progeny strands into viral particles proceeds from the 3′ end, which correlates with the 3′-to-5′ processivity of this helicase (King et al. 2001; Cotmore and Tattersall 2005a). Thus, Rep 40/52 can function as a molecular motor, associating with a single capsid portal and effectively pumping the DNA into the virion (King et al. 2001). By analogy with SV40 T antigen, Rep 40 can be modeled as a hexameric ring (James et al. 2004) in which peptide loops carrying residues K404 and K406, known to be essential for single-strand DNA binding, project into a central pore, 18 Å in diameter, through which the single strand could pass during the process of DNA unwinding or strand translocation (Yoon-Robarts et al. 2004). These two lysine residues have also been shown to be essential for packaging, but at present we have little concrete information about the structure of the motor or how it interacts with the capsid. Although this role can also be fulfilled, albeit much less efficiently, by the large Rep 68 and Rep 78 molecules, in this case the extreme 5′

Figure 2. Encapsidation strategy for MVM. The MVM genome is represented by a continuous line, with its 3′ end indicated by a small arrowhead. The gray sphere represents an NS1 molecule. The letters L and R represent left-end and right-end palindromic sequences, respectively. Upper- and lowercase letters represent "flip" and "flop" versions of these sequences, respectively, which are inverted complements of one another. Displaced strands are packaged, in a 3′-to-5′ direction, into preassembled capsids, represented by shaded hexagons. Removal of displaced strands by encapsidation limits synthesis to negative-sense DNA (*right pathway*), but replication of the displaced strands, during the amplification phase of the reaction, generates dimer intermediates that can support displacement of both positive- and negative-sense DNA (*right pathway*). Normally, processing at OriR is many times more efficient than at OriL, promoting the preferential synthesis and displacement of negative-sense DNA. The slashed circles indicate the OriR defect in LuIII that equalizes the efficiency of the two origins and thus enhances the pathway that generates positive strands.

ends of the genomes are rarely internalized. Functionally equivalent forms of the NS polypeptides that are capable of mediating strand translocation have yet to be identified for MVM.

PARVOVIRUS REPLICATION GENERATES DIVERSITY

Parvoviral genomes are highly heterogeneous in vivo and frequently occur in infected tissues as quasispecies, even where infection was initiated with genetically homogeneous virus (Lopez-Bueno et al. 2003). This suggests that viral replication mechanisms strongly promote the generation of diversity, apparently reflecting high rates of both recombination and nucleotide substitution. Since viral genomes are replicated through a series of palindromic, concatemeric duplex intermediates, the possibility for efficient intra- and intermolecular recombination is high, and strand-displacement synthesis appears to be highly susceptible to template-strand switching. However, these forks also appear error-prone, despite the fact that they are assembled from a subset of the host cell's DNA synthetic machinery, including the high-fidelity host DNA pol-δ, for which mutation rates would be expected to be low. For example, during the evolution of canine parvovirus (CPV) from feline panleukopenia virus (FPV) since about 1978, the emergent branch for CPV had a substitution rate similar to that seen in rapidly evolving RNA viruses, such as HIV-1 and human influenza A (Shackelton et al. 2005). Possibly this occurs because unidirectional, strand-displacement mechanisms proceed with a fidelity that is different from that associated with the bidirectional replication of cellular genes, but viral genomes are also small, and successive rounds of replication are rapidly completed, so that opportunities for editing or mismatch repair may be restricted. It also seems very likely that the DNA structures elaborated during parvoviral replication would elicit host damage responses that the virus may suppress, thus adding to the error-proneness of the process. Indeed, several lines of evidence suggest that parvoviral infection activates double-strand-break repair mechanisms, and this aspect of the virus:host interaction will likely reward future research (Weitzman et al. 2004). Finally, these viruses are transmitted as a single strand, and so lack templates for the correction of environmentally induced changes, leaving them vulnerable to mutational processes such as deamination. Thus, parvoviral genomes appear to be under intense selection, with their mode of replication promoting diversity, but their constrained size, and hence their genetic complexity, severely restricting the types of modifications that can be tolerated.

ACKNOWLEDGMENTS

The authors were supported by U.S. Public Health Service grants AI26109 and CA29303 from the National Institutes of Health.

REFERENCES

Bashir T., Horlein R., Rommelaere J., and Willwand K. 2000. Cyclin A activates the DNA polymerase delta-dependent elongation machinery in vitro: A parvovirus DNA replication model. *Proc. Natl. Acad. Sci.* **97:** 5522–5527.

Brister J.R. and Muzyczka N. 1999. Rep-mediated nicking of the adeno-associated virus origin requires two biochemical activities, DNA helicase activity and transesterification. *J. Virol.* **73:** 9325–9336.

Burnett E., Christensen J., and Tattersall P. 2001. A consensus DNA recognition motif for two KDWK transcription factors identifies flexible-length, CpG-methylation sensitive cognate binding sites in the majority of human promoters. *J. Mol. Biol.* **314:** 1029–1039.

Christensen J. and Tattersall P. 2002. Parvovirus initiator protein NS1 and RPA coordinate replication fork progression in a reconstituted DNA replication system. *J. Virol.* **76:** 6518–6531.

Christensen J., Cotmore S.F., and Tattersall P. 2001. Minute virus of mice initiator protein NS1 and a host KDWK family transcription factor must form a precise ternary complex with origin DNA for nicking to occur. *J. Virol.* **75:** 7009–7017.

Costello E., Saudan P., Winocour E., Pizer L., and Beard P. 1997. High mobility group chromosomal protein 1 binds to the adeno-associated virus replication protein (Rep) and promotes Rep-mediated site-specific cleavage of DNA, ATPase activity and transcriptional repression. *EMBO J.* **16:** 5943–5954.

Cotmore S.F. and Tattersall P. 1996. Parvovirus DNA replication. In *DNA replication in eukaryotic cells* (ed. M.L. DePamphilis), pp. 799–813. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

———. 2003. Resolution of parvovirus dimer junctions proceeds through a novel heterocruciform intermediate. *J. Virol.* **77:** 6245–6254.

———. 2005a. Encapsidation of minute virus of mice DNA: Aspects of the translocation mechanism revealed by the structure of partially-packaged genomes. *Virology* **336:** 100–112.

———. 2005b. Packaging sense is controlled by the efficiency of the nick site in the right-end replication origin of parvoviruses MVM and LuIII. *J. Virol.* **79:** 2287–2300.

———. 2006. A rolling-hairpin strategy: Basic mechanisms of DNA replication in the parvoviruses. In *Parvoviruses* (ed. J.R. Kerr et al.), chap. 14, pp. 171–188. Edward Arnold, London, United Kingdom.

Cotmore S.F., Christensen J., and Tattersall P. 2000. Two widely spaced initiator binding sites create an HMG1-dependent parvovirus rolling-hairpin replication origin. *J. Virol.* **74:** 1332–1341.

Cotmore S.F., Christensen J., Nuesch J.P., and Tattersall P. 1995. The NS1 polypeptide of the murine parvovirus minute virus of mice binds to DNA sequences containing the motif [ACCA]2-3. *J. Virol.* **69:** 1652–1660.

Duan D., Sharma P., Yang J., Yue Y., Dudus L., Zhang Y., Fisher K.J., and Engelhardt J.F.

1998. Circular intermediates of recombinant adeno-associated virus have defined structural characteristics responsible for long-term episomal persistence in muscle tissue. *J. Virol.* **72:** 8568–8577.

Hickman A.B., Ronning D.R., Perez Z.N., Kotin R.M., and Dyda F. 2004. The nuclease domain of adeno-associated virus rep coordinates replication initiation using two distinct DNA recognition interfaces. *Mol. Cell* **13:** 403–414.

James J.A., Aggarwal A.K., Linden R.M., and Escalante C.R. 2004. Structure of adeno-associated virus type 2 Rep40-ADP complex: Insight into nucleotide recognition and catalysis by superfamily 3 helicases. *Proc. Natl. Acad. Sci.* **101:** 12455–12460.

Kerr J.R., Cotmore S.F., Bloom M.E., Linden R.M., and Parrish C.R., Eds. 2006. *Parvoviruses*. Edward Arnold, London, United Kingdom.

King J.A., Dubielzig R., Grimm D., and Kleinschmidt J.A. 2001. DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids. *EMBO J.* **20:** 3282–3291.

Lopez-Bueno A., Mateu M.G., and Almendral J.M. 2003. High mutant frequency in populations of a DNA virus allows evasion from antibody therapy in an immunodeficient host. *J. Virol.* **77:** 2701–2708.

Miller D.G., Petek L.M., and Russell D.W. 2004. Adeno-associated virus vectors integrate at chromosome breakage sites. *Nat. Genet.* **36:** 767–773.

Muzyczka N. and Berns K.I. 2001. *Parvoviridae:* The viruses and their replication. In *Fields virology,* 4th edition (ed. D.M. Knipe and P.M. Howley), pp. 2327–2360. Lippincott, Williams and Wilkins, Philadelphia, Pennsylvania.

Nuesch J.P., Christensen J., and Rommelaere J. 2001. Initiation of minute virus of mice DNA replication is regulated at the level of origin unwinding by atypical protein kinase C phosphorylation of NS1. *J. Virol.* **75:** 5730–5739.

Shackelton L.A., Parrish C.R., Truyen U., and Holmes E.C. 2005. High rate of viral evolution associated with the emergence of carnivore parvovirus. *Proc. Natl. Acad. Sci.* **102:** 379–384.

Surdo P.L., Bottomley M.J., Sattler M., and Scheffzek K. 2003. Crystal structure and nuclear magnetic resonance analyses of the SAND domain from glucocorticoid modulatory element binding protein-1 reveals deoxyribonucleic acid and zinc binding regions. *Mol. Endocrinol.* **17:** 1283–1295.

Tattersall P. and Cotmore S.F. 2005. Parvoviruses. In *Topley and Wilson's microbiology and microbial infections*, 10th edition (ed. B. Mahy and V. ter Meulen), vol. 1, pp. 407–438. Edward Arnold, London, United Kingdom.

Tattersall P., Bergoin M., Bloom M.E., Brown K.E., Linden R.M., Muzyczka N., Parrish C.R., and Tijssen P. 2005. *Parvoviridae*. In *Virus taxonomy,* 8th Report of the International Committee on Taxonomy of Viruses (ed. C.M. Fauquet et al.), pp. 353–369. Elsevier/Academic Press, London, United Kingdom.

Weitzman M.D., Carson C.T., Schwartz R.A., and Lilley C.E. 2004. Interactions of viruses with the cellular DNA repair machinery. *DNA Repair* **3:** 1165–1173.

Willwand K., Moroianu A., Horlein R., Stremmel W., and Rommelaere J. 2002. Specific interaction of the nonstructural protein NS1 of minute virus of mice (MVM) with $[ACCA]_2$ motifs in the centre of the right-end MVM DNA palindrome induces hairpin-primed viral DNA replication. *J. Gen. Virol.* **83:** 1659–1664.

Yoon-Robarts M., Blouin A.G., Bleker S., Kleinschmidt J.A., Aggarwal A.K., Escalante C.R., and Linden R.M. 2004. Residues within the B′ motif are critical for DNA binding by the SF3 helicase Rep40 of adeno-associated virus type 2. *J. Biol. Chem.* **279:** 50472–50481.