The EMBO Journal vol.13 no.17 pp.4145–4152, 1994

An asymmetric nucleotide in the parvoviral 3′ hairpin directs segregation of a single active origin of DNA replication

Susan F.Cotmore[1] and Peter Tattersall[1,2,3]

Departments of [1]Laboratory Medicine and [2]Genetics, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510, USA

[3]Corresponding author

Communicated by B.Hirt

The 3′ telomere of the linear single-stranded DNA genome of minute virus of mice (MVM), a murine parvovirus, can assume a complex hairpin structure: This contains a stem in which there is a mismatched 'bubble' sequence where a GA doublet opposes a GAA triplet. During replication, this hairpin is copied to form an imperfect palindrome which bridges adjacent genomes in a dimer duplex intermediate, leaving the two 'bubble' sequences embedded in potential replication origins on either side of the axis of symmetry. Such junctions are resolved asymmetrically *in vitro* in a DNA synthetic reaction which requires the viral initiator protein NS1. We show that the sequence surrounding the doublet is a potent origin, but the analogous region containing the triplet is completely inactive. The active origin is ~50 bp long, extending from an Activated Transcription Factor binding site at one end to a position some 7 bp beyond the major initiation site, to which NS1 ultimately becomes covalently attached. The actual sequence of the GA doublet is unimportant, but insertion of any third nucleotide here inactivates the origin, indicating that it represents a critical spacer element. Segregation of this asymmetry, therefore, allows the virus to confine replication initiation to one particular telomeric configuration.

Key words: ATF/eukaryotic replication origin/origin recognition/parvovirus/rolling-circle DNA replication

Introduction

Parvoviruses are unique among all known viruses in having a linear single-stranded DNA genome. Minute virus of mice (MVM), a widely disseminated murine parvovirus, encapsidates a single, negative-sense DNA strand containing a relatively long (4.8 kb) coding region bracketed by short, unique, palindromic termini which are capable of folding into hairpin duplexes (Cotmore and Tattersall, 1987). In order to replicate, this unusual structure interacts with the synthetic machinery of the host cell in some novel ways, using cellular proteins and a pleiotropic virally coded protein, NS1, to establish a type of unidirectional, single-strand synthesis (Tattersall and Ward, 1976), which in many ways parallels the rolling-circle DNA replication mechanisms of single-stranded coliphages and some bacterial plasmids (Baas and Jansz, 1988; Inamoto *et al.*, 1991; Ilyina and Koonin, 1992; Wang *et al.*, 1993). This process has been dubbed 'rolling hairpin' replication because, instead of progressing continuously around a circular template, the replication fork is flipped back and forward along the linear genome by sequential synthesis and rearrangement of the palindromic viral termini.

Complex palindromic intermediates are generated during rolling hairpin replication, starting with a dimer form in which duplex copies of the unit-length genome are arranged in a 3′-to-3′ configuration, and progressing to a tetramer form in which these dimers are arranged as a 5′-to-5′ palindrome. Junction fragments from MVM replication intermediates have been isolated and cloned into bacterial plasmids (Cotmore and Tattersall, 1992). In this form, the bridge structures serve as replication origins *in vivo* provided NS1 is supplied in *trans*. These observations thus confirmed earlier studies (Faust and Ward, 1979) which showed that sequences from the viral termini, comprising <10% of the total genome, contain all of the *cis*-acting information essential for DNA replication.

Figure 1A illustrates how the sequence of the 3′ telomere becomes rearranged in the dimer junction fragment, with particular reference to internal palindromic sequences designated the 'ears', and an asymmetric mismatch 'bubble' within the stem. This bubble contains the triplet GAA on the inboard arm of the palindrome opposed to the doublet GA on the outboard strand. In the bridge arrangement, these nucleotides are base paired to their complementary sequences TTC and TC, respectively. When plasmids containing these dimer junction sequences are added to a standard HeLa replication extract, in the presence of NS1 expressed from a recombinant vaccinia virus, they are resolved and replicated asymmetrically (Cotmore *et al.*, 1993). Each junction fragment is resolved into two viral telomeres, as shown in Figure 1B. One of these, attached to sequences from the original inboard arm of the genomic palindrome, is present predominantly in the 'extended' configuration, i.e. it contains a duplex copy of the entire terminal palindrome. One of the strands of this palindrome is newly synthesized, whereas its template, the other strand, is now covalently attached, through its 5′ end, to an NS1 molecule. The other resolution product, derived from the outboard arm of virion DNA, is present predominantly in a 'turn-around' configuration, in which the two strands of the duplex are covalently continuous at the axis of the palindrome. This asymmetric process presumably underlies the mechanism by which the virus conserves a single DNA sequence in its 3′ telomere (Astell *et al.*, 1985), since with each round of resolution and replication, only one DNA strand, the 'flip' sequence, is synthesized. Although the details of the resolution process remain to be clarified, in this paper we describe what we believe to be the first critical step: the selection of a single initiation site. Since replication from this site is

unidirectional and leading strand specific (S.F.Cotmore and P.Tattersall, in preparation), such selection inevitably confines synthesis to a single DNA strand.

By defining the minimum functional 3′ origin, we are able to show that positive recognition of this sequence does not involve those characteristic secondary structure elements which are found exclusively in the 'turn-around' configuration of the viral telomere (cf. Im and Muzyczka, 1990). Instead, the active origin is a stretch of perfectly base-paired duplex DNA ~50 nucleotides long. Deletion and mutagenesis within this sequence shows that recognition is at least bipartite, and suggests that members of the Activated Transcription Factor (ATF) family of cellular transcription factors may be involved. Most strikingly, we show that there are critical steric constraints between separate recognition elements within the initiation complex. This arrangement appears to be exploited by the virus, since certain forms of the viral telomere are prevented from serving as substrates for replication by the insertion of a single nucleotide between these elements. Since the sequences which control origin activity also contain binding sites for other cellular factors involved in upregulating transcription from the viral P_4 promoter (Faisst *et al.*, 1994; Z.Gu, S.Plaza, M.Perros, C.Cziepluch, J.Rommelaere and J.J.Cornelis, submitted), it seems likely that the virus uses this sequence modulation to orchestrate transcription with replication.

Results

A replication origin is only present on one side of the 3′:3′ palindromic junction

In the dimer junction sequence, there are two candidate nick sites for NS1, one on each arm of the palindrome on opposite DNA strands (Figure 1B). However, in order to initiate asymmetric resolution it seemed likely that only one of these would function as an origin. To investigate this possibility, we constructed plasmid clones containing a single arm of the junction palindrome (Figure 2A). These each contained an intact stem sequence together with some of the adjacent nucleotides from the asymmetric ear structures, and were designated pGAA and pTC according to the nucleotide sequence in the bubble region of the strand containing the potential nick site. When these clones were used as substrates in an NS1-supplemented *in vitro* replication assay, there was a dramatic difference in their activity. The pGAA clone was unable to support any NS1-mediated DNA replication, although there was a low level of repair synthesis, comparable to that seen when vector DNA alone was used as the substrate (Figure 2B, lanes 1 and 2). In contrast, the insert in the pTC clone acted as a potent replication origin (Figure 2B, lane 3). Label predominantly appears in species which either co-migrate, or are larger than, the linearized substrate plasmid. Analysis of these products on alkaline gels (not shown) indicates that this type of synthesis results in an increase in the length of at least one of the template strands. DNA synthesized in this reaction could be almost quantitatively immunoprecipitated with anti-NS1 serum after heating in SDS under conditions which would disrupt non-covalent DNA–protein interactions without denaturing double-stranded DNA (Figure 2B, lane 6). This indicates that duplexes containing newly replicated DNA

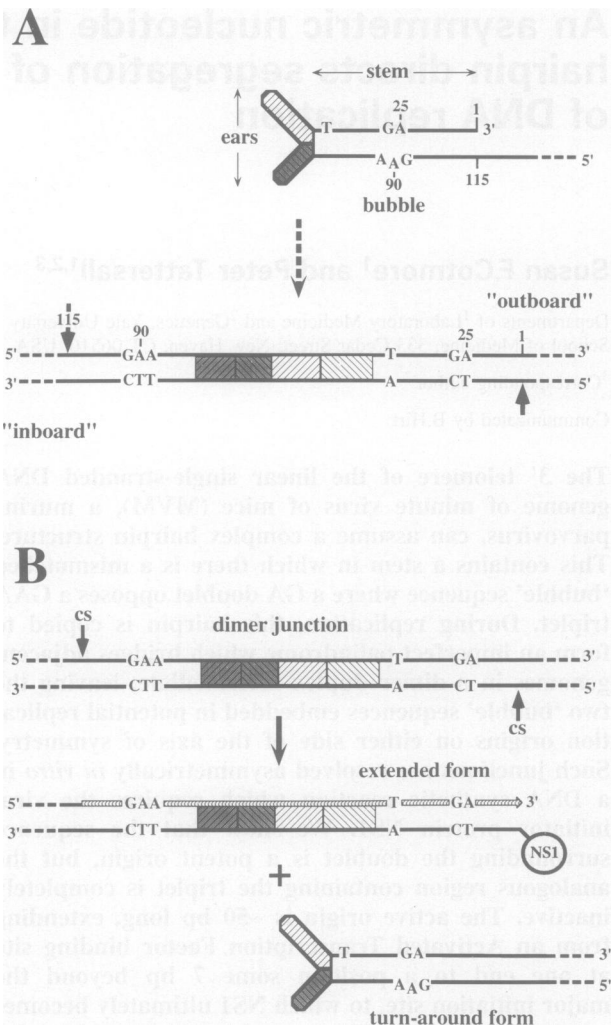

Fig. 1. (**A**) Schematic representation of the 3′ end of the single-stranded MVM DNA molecule, showing its three major palindromic elements folded into a 'stem' plus 'ears' structure, in which the stem contains the mismatch 'bubble' discussed in the text. Replication initially proceeds rightward from the terminal 3′ hydroxyl of the hairpin stem. The hairpin is then copied by a leftward-moving replication fork, producing the dimer junction configuration, in which the mismatched doublet and triplet are now base paired, and located on the 'outboard' and 'inboard' sides of the axis of symmetry. (**B**) NS1-mediated resolution of the dimer junction *in vitro* results in two major products. The 'inboard' arm now terminates in an extended form telomere, with NS1 covalently attached to its 5′ end. The symbol ⟶ represents a newly synthesized DNA strand. The 'outboard' arm terminates in a turn-around structure, with the same sequence arrangement as the original hairpin. The potential nick sites for the NS1 endonuclease are marked cs.

are covalently attached to NS1, the initiating endonuclease.

The 3′ origin is derived exclusively from sequences in the stem of the telomere

There are multiple sequence differences between the viral inserts in pTC and pGAA, derived from both the ear and stem regions of the palindrome. To ask whether the stem sequences alone could support replication, we designed a pair of oligonucleotide primers (designated L1 and L2) which would allow us to amplify these regions specifically (Figure 2A). By cloning polymerase chain reaction (PCR)

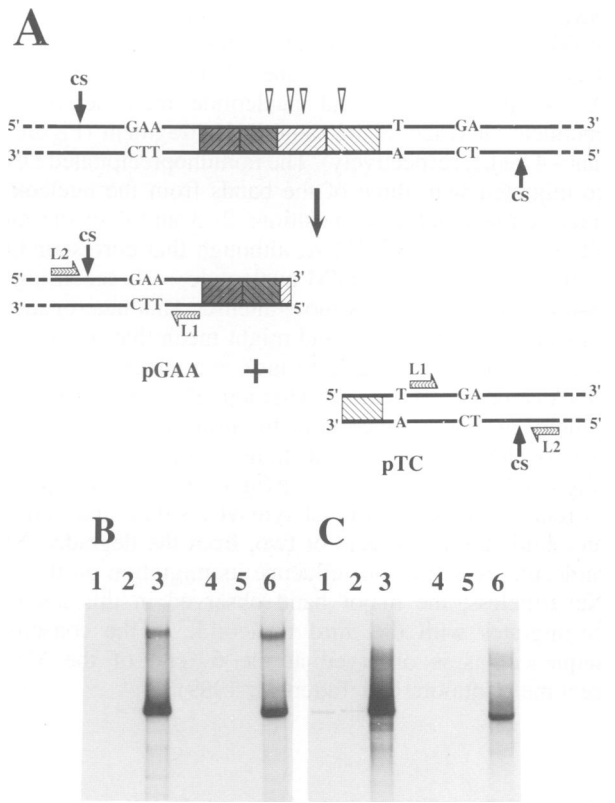

Fig. 2. **(A)** Derivation of 'single-arm' clones from the dimer junction. Vertical carets indicate *Bss*HII sites within the hairpin 'ear' sequences which were used to separate the pGAA and pTC sequences. The symbol represents the 12-mer PCR primers used to amplify and clone subregions of each plasmid insert. Primer L1 has the sequence 5′-ACGTCACTTACG-3′, which lies between nucleotides 39 and 28 on the viral strand, as shown in Figure 1A, and primer L2 has the sequence 5′-TTCGCGCCTTGG-3′, which lies between nucleotides 151 and 140. The potential NS1 cleavage sites within these DNA molecules are marked cs. **(B)** Agarose gel electrophoresis of the ^{32}P-labeled replication products of a vector plasmid pUC 19, and the 'single-arm' clones pGAA and pTC, before (lanes 1, 2 and 3, respectively), and after (lanes 4, 5 and 6, respectively) immunoprecipitation with anti-NS1 antiserum. All samples were digested with *Hin*dIII prior to electrophoresis. *Hin*dIII cuts once in each substrate plasmid to produce a linear 2.6 kb fragment, which co-migrates with the fastest major band of incorporated label. **(C)** Agarose gel electrophoresis of replication products of pCRII, pL1-2GAA and pL1-2TC (lanes 1, 2 and 3, respectively), and of these same products immunoprecipitated with anti-NS1 antiserum (lanes 4, 5 and 6, respectively). As before, all samples were digested with *Hin*dIII prior to electrophoresis. *Hin*dIII cuts once in each substrate plasmid to produce a linear 4.0 kb fragment, which co-migrates with the fastest major band of incorporated label.

products into the plasmid vector pCRII, we obtained two plasmids, pL1-2GAA and pL1-2TC, which only contained sequences from the stem region, beginning five nucleotides from the point at which the stem and ears diverge, and stretching through the telomere into the main body of the genome. These two clones contained viral inserts of 76 and 75 bp, respectively, and were identical except for the sequences in the bubble region, from which they derived their names. When these clones were used as replication substrates *in vitro* (Figure 2C), the GAA clone was totally inactive, but the TC clone contained a competent NS1-dependent replication origin. Quantitation of dAMP incorporation into newly synthesized DNA indicated that, on a molar basis, the pL1-2TC clone elicited ~30% more DNA synthesis *in vitro* than its more complex parent clone pTC. The basis for this enhanced activity is not clear at present. These assays were carried out at limiting DNA and NS1 concentrations, so that differences between various substrates could be readily identified. However, it should be noted that the pGAA and pL1-2GAA clones remained completely inactive over a wide range of DNA and NS1 concentrations.

This is a somewhat surprising and paradoxical result, since we have previously shown that *in vitro* resolution of the 3′:3′ junction under similar conditions yields final products in which the newly synthesized strand in the extended form DNA is ultimately associated with the 'inboard', or GAA arm, as shown in Figure 1B. This finding initially suggested that the active origin would be associated with the cut site (cs) upstream of this sequence, rather than with the cut site on the 'outboard' arm, as demonstrated here. Resolution must therefore occur by a more complex mechanism than originally proposed. We are currently exploring an alternative model in which the template DNA adopts a cruciform configuration resembling a Holliday structure, prior to the initiation of replication, and in which the products of synthesis are ultimately separated from one another by recombination.

The experiments described in Figure 2 establish that the replication origin at the 3′ end of the viral genome resides exclusively in one arm of the palindromic sequence which bridges adjacent genomes in dimer RF. Since it is contained within a duplex copy of sequences from the stem region of the outboard arm of the viral telomere, recognition cannot depend upon any of the potential secondary structure elements normally associated with this terminus.

Mapping the in vitro nick site

Although the *in vivo* nicking and NS1-attachment site at the 3′ end of the genome has not been mapped precisely, its approximate location can be deduced from the sizes of palindromes in cloned forms of MVM and the closely related LuIII virus (Astell *et al.*, 1985; Diffoot *et al.*, 1993). Moreover, in MVM there are two quite disparate replication origins, one associated with each terminus of the negative-sense viral genome (Cotmore and Tattersall, 1992; Tam and Astell, 1993). Although these two termini are remarkably different in size, primary sequence, predicted secondary structure and sequence heterogeneity (Astell *et al.*, 1985), they are both activated by a site- and strand-specific nick introduced by the replication-initiator protein NS1 (Cotmore *et al.*, 1992, 1993; Nüesch *et al.*, 1992). NS1 is known to be associated *in vivo* with the 5′ end of the MVM genome, predominantly through the adenine residue at nucleotide 5170 (Cotmore and Tattersall, 1989). Significantly, this adenine is the third nucleotide in the sequence CTWWTCA, which is also found at the 3′ end of the genome in the vicinity of the predicted nick site.

To map the *in vitro* nicking and NS1 attachment site at the 3′ end of the genome, a 169 bp DNA fragment containing the viral insert from pL1-2TC was isolated and labeled with [^{32}P]dGTP at the 3′ end of the DNA strand which becomes nicked during the initiation reaction, as shown in Figure 3. This fragment was used as an *in vitro* replication substrate, and the NS1-associated DNA then

purified by immunoprecipitation with anti-NS1 serum. The immunoprecipitated product was a single-stranded fragment of ~100 bases, covalently linked at one end to NS1. This was digested with proteinase K in 0.5% SDS at 60°C for several hours before being fractionated on a sequencing gel (Figure 3, lanes 2 and 3) alongside an A + G tract and a total nucleotide tract derived by chemical cleavage of the same DNA fragment (Figure 3, lanes 4 and 1, respectively). The immunoprecipitated DNA co-migrated with three of the bands from the nucleotide track, corresponding to positions 2, 3 and 4 in the nick site consensus CTWWTCA, although that corresponding to the third position (MVM nucleotide − 4, according to Astell *et al.*, 1985) was most intense. This heterogeneity was highly reproducible and might mean that *in vitro* the site at which NS1 is able to nick and attach itself to the DNA is somewhat flexible. Alternatively, this microheterogeneity may be a function of the limit digest products of the covalent NS1–DNA substrate with proteinase K. The oligonucleotide product seen in Figure 3 would be expected to retain at least a terminal tyrosyl residue, and perhaps an additional amino acid or two, from the degraded NS1 molecule, which could influence its migration on the gel. Nevertheless, the major band observed in this analysis co-migrates with the third nucleotide in the consensus sequence, as is observed at the 5′ end of the MVM genome (Cotmore and Tattersall, 1989).

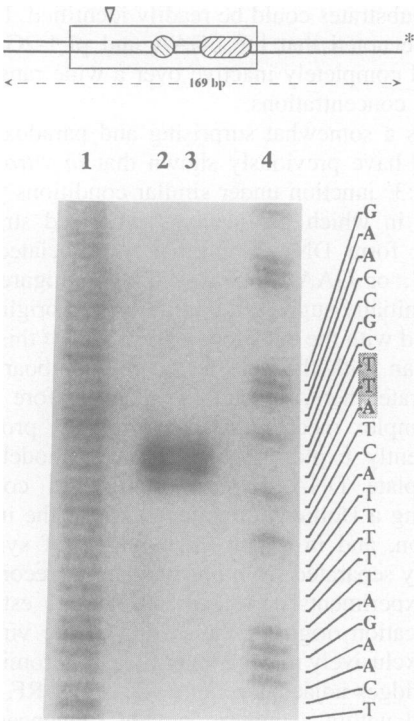

Fig. 3. Denaturing PAGE of *in vitro* cleavage products, immunoprecipitated with anti-NS1 antibody, from a reaction containing a 169 bp, 3′-end-labeled fragment derived from pL1-2TC (lanes 2 and 3). Lane 1 contains a hydrazine cleavage reaction of the same labeled 169 bp fragment showing all possible end-labeled products, and lane 4 contains the G + A reaction products. For lanes 1 and 4, chemical cleavage was achieved by the method of Maxam and Gilbert (1980).

One border of the replication origin is located near the nick site

To ask whether a specific DNA sequence is required at the initiation site, we prepared plasmid clones containing pL1-2TC sequences deleted at or beyond the position of the major nick site identified above (Figure 4). When compared *in vitro* for their ability to serve as templates for NS1-dependent DNA replication, sequences deleted up to the nick site were found to be completely inactive

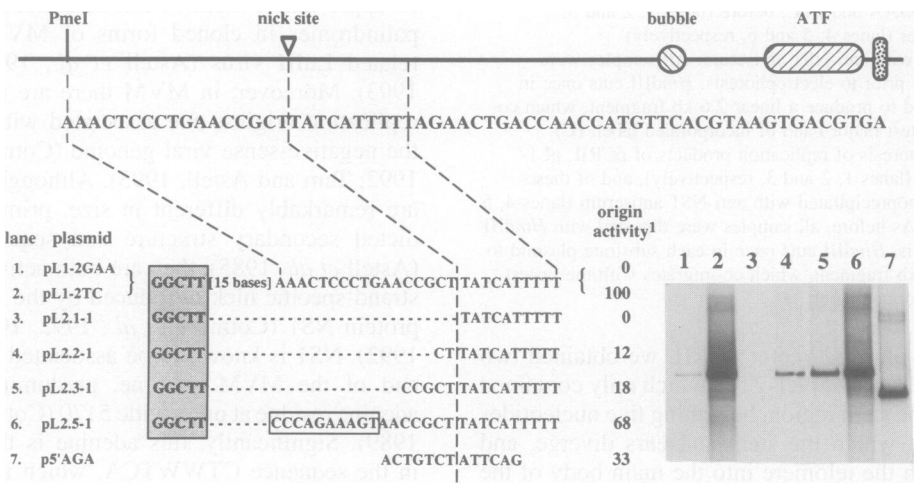

Fig. 4. A schematic representation of the origin and the DNA sequence of the nicked strand are shown above the structures of the vector–virus junctions near the nick sites in pL1-2GAA, pL1-2TC and four deletion mutants derived from pL1-2TC. In this and subsequent diagrams, the positions of the 'bubble' sequence and ATF-binding site are indicated by cross-hatched boxes. The stippled box next to the ATF site shows the position of the A residue which is complementary to the asymmetric T at nucleotide 41 in the viral strand, as depicted in Figure 1A. The boxed, shaded sequence GGCTT represents the vector sequence at the plasmid–insert junctions where the deletions were made. The open-boxed sequence was mutated away from viral sequence as described in the text. Also shown is the equivalent sequence from p5′AGA, a clone derived from an origin at the right-hand end of the MVM genome. The panel on the lower right shows an agarose gel electrophoretic analysis of the replication products of each of these constructs. As before, the replicated DNA was digested with *Hin*dIII to yield linear molecules. The fastest major band of labeled material in each case co-migrates with the linearized input plasmid. [1]Origin activity, shown for each construct, was determined by adsorption of labeled product DNA on to DE81 paper, as described in Materials and methods.

(Figure 4, lane 3) and those which retained the full consensus sequence, but extended only two and five nucleotides beyond the nick site, exhibited 12 and 18% of wild-type activity, respectively, under the limiting conditions of this assay (Figure 4, lanes 4 and 5). A fourth clone was constructed which contained the wild-type pL1-2TC sequence for seven nucleotides beyond the nick site, but then substituted a 10 bp sequence in which the MVM nucleotides were specifically mutated, replacing purines with non-complementary pyrimidines and vice versa. This was done in order to make sure that adjacent vector sequences would not contribute to the viral origin. This clone served as a rather effective replication origin, exhibiting ~68% of wild-type activity (Figure 4, lane 6).

In contrast, p5′AGA, a single-arm clone derived from the 5′ end of the viral genome which contains the full origin region, the nick site and hundreds of base pairs beyond, could only elicit ~33% of the DNA synthesis associated with pL1-2TC. This demonstrates that the MVM 3′ origin is relatively potent *in vitro* provided that it retains the correct sequence at and beyond the nick site.

Defining the origin border at the top of the hairpin stem

Although pL1-2TC and pL1-2GAA only differ from each other at the bubble sequence, these clones both contain 13 bp of stem sequence beyond this site. If these 13 nucleotides are deleted, the pL1-2TC origin is completely inactivated in this assay (Figure 5, lane 3). In designing the primers for creating pL1-2TC, we had originally sought to maintain a seven nucleotide sequence near the end of the stem which appeared to conform to the binding-site consensus for members of the ATF family of basic leucine zipper DNA-binding proteins (Hai *et al.*, 1989). To evaluate the need for this site, we generated forms of pL1-2TC which lacked one or two of the correct nucleotides from the 3′ end of this consensus sequence. Substitution of the terminal nucleotide still gave a highly active origin (Figure 5, lane 5), whereas substitution of the two most 3′ nucleotides reduced origin activity to 40% of wild-type levels (Figure 5, lane 4).

Although the sequence of cognate ATF binding sites can vary substantially, certain residues are known to be more critical than others. Mutation of the third guanine in the sequence GTGACGT to a thymine is known to be particularly damaging (Sakai *et al.*, 1991), such that when this mutation is present in the promoter region of the human *Rb* gene the interaction between ATF and its cognate site is so impaired that a retinoblastoma phenotype results. To explore the effect of this mutation on MVM replication, we restored the pL1-2TC clone to its original length, but incorporated the aforementioned mutation into the penultimate base. This single mutation reduced origin-dependent DNA synthesis >4-fold, compared with the wild-type level (Figure 5, lane 6). Thus, activation of the viral origin requires specific recognition of sequences within the consensus ATF site, and a mutation which is known to be severely debilitating for ATF binding is also highly deleterious for the initiation of MVM DNA replication at this origin.

The stem of the genomic 3′ telomere has an additional mismatched thymine residue which is located three nucleotides away from the base of the ear structure (Figure 1A), and thus was not included in the pL1-2TC construct (Figure 2A). This residue is represented by its complement, adenine, in the sequence of the DNA strand which is nicked during initiation (Figure 4). In order to see whether inclusion of this additional nucleotide would augment the *in vitro* origin, we created plasmid clones (designated pL6–2, GA, GT, GC and GG) which extended the sequence from pL1-2TC by one base in the direction of the ears, and introduced each of the four possible nucleotides in the position of this genomic asymmetry (Figure 5, lanes 7–10). All of these plasmids contained potent replication origins, but none was more efficient than pL1-2TC, and there was no preference for the wild-type sequence.

The bubble regulates origin activity

The deletions and mutations described above show that the MVM 3′ origin is made up of a minimal core region of ~50 nucleotides, in which sequences at both extremities

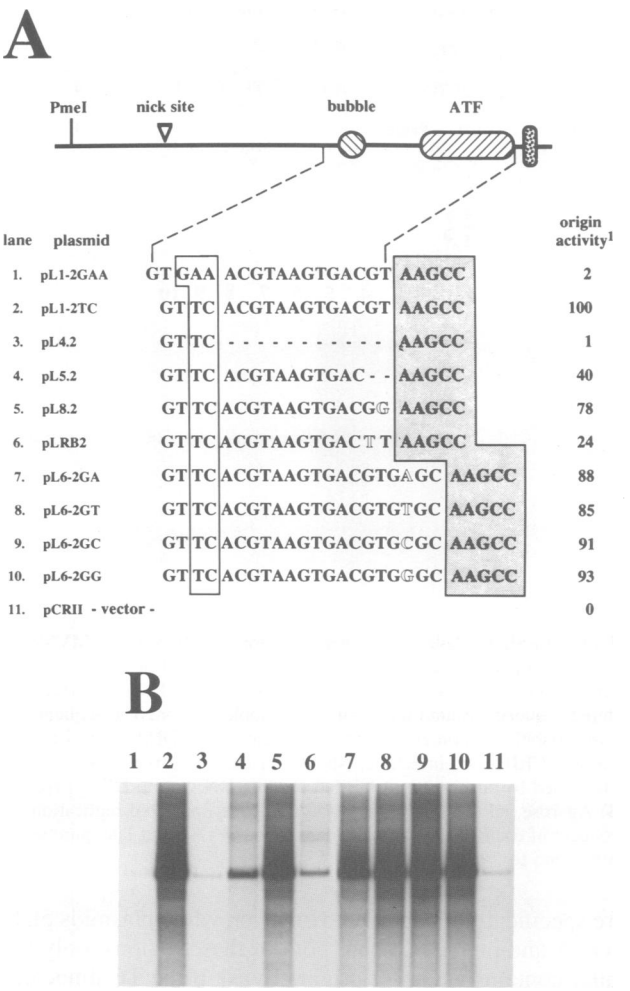

Fig. 5. **(A)** Shown below the schematic representation of the MVM origin are the structures of the relevant vector–virus junctions in pL1-2GAA, pL1-2TC and eight deletion or substitution mutants derived from pL1-2TC, each with altered sequence around the consensus ATF binding site. The sequences of the 'bubble' region are boxed, and vector sequences (AAGCC) are boxed and shaded. The vector-alone DNA control is marked pCRII. [1]Origin activity, shown for each construct, was determined by adsorption of labeled product DNA on to DE81 paper. **(B)** Agarose gel electrophoretic analysis of the linearized replication products of each of the constructs described in (A), and lane numbers correspond to those shown in (A).

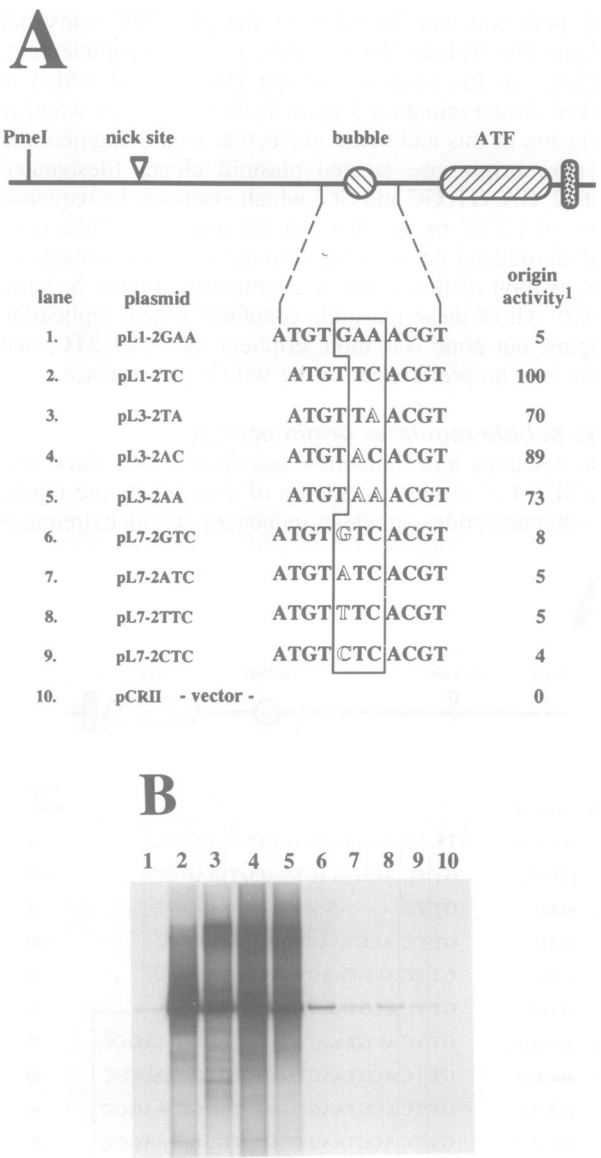

Fig. 6. **(A)** Shown below the schematic representation of the MVM origin are the structures of pL1-2GAA, pL1-2TC and of seven substitution or insertion mutants derived from pL1-2TC, each with altered sequence around the mismatch 'bubble' region. The sequences of the 'bubble' region are boxed. The vector-alone DNA control is marked pCRII. [1]Origin activity, shown for each construct, was determined by adsorption of labeled product DNA on to DE81 paper. **(B)** Agarose gel electrophoretic analysis of the linearized replication products of each of the constructs described in (A), and lane numbers correspond to those shown in (A).

are specifically recognized. However, while plasmids pL1-2GAA and pL1-2TC share both of these regions, only the latter contains an active origin. To ask if the TC dinucleotide in the bubble sequence was itself specifically recognized, we created plasmids in which either or both of these bases were mutated (Figure 6, lanes 3–5). Although the resulting plasmids evoked slightly less DNA synthesis in the *in vitro* assay than their progenitor, they remained highly efficient templates, suggesting that there is no specific recognition of this sequence. In contrast, if any additional nucleotide was introduced into the site, DNA synthesis was drastically curtailed (Figure 6, lanes 6–9).

The G of the GAA triplet in the replication-minus arm of the viral palindrome is the final nucleotide of a consensus 'E box' binding site, CANNTG, which could potentially be recognized by transcription factors such as USF and MLTF (Z.Gu, S.Plaza, M.Perros, C.Cziepluch, J.Rommelaere and J.J.Cornelis, submitted). Thus, it could be argued that in the GAA arm this sequence is recognized by a transcription factor which blocks replication initiation at this site, and suggesting that inhibition of the GAA arm, rather than positive selection of the TC arm, could govern expression of the replication origin. Significantly, one of the insertion mutants shown in Figure 6, pL7-2GTC, actually recreates this 'E box' motif in the context of the TC arm. However, while this mutation does indeed block initiation (Figure 6, lane 6), so also does the introduction of any other base (lanes 7–10) in this same position, suggesting that inactivation of the origin does not involve the creation of a specific inhibitory sequence. Instead, the TC bubble sequence appears to be acting as a critical spacer element within the origin, presumably facilitating the interaction of essential sequence recognition factors. Thus, by the simple insertion or omission of one nucleotide, the virus is able to modulate the establishment of replication complexes at the 3′ end of its genome, and thus to select the viral forms which will be committed to progeny DNA synthesis.

Discussion

In this paper, we show that the replication origin at the 3′ end of the MVM genome is a duplex DNA sequence of ~50 bp which is derived from the outboard arm of the palindromic viral telomere. This duplex sequence is not present in virion DNA, where the terminal palindrome is somewhat imperfect and contains a critical sequence mismatch, the 'bubble', where the triplet GAA on the inboard arm is juxtaposed to the dinucleotide sequence GA on the outboard arm. During replication, this terminal palindrome is extended and copied to form the junction sequence which bridges adjacent duplex genomes in dimer RF. In the bridge structure, the bubble nucleotides are paired with complementary residues, so that the sequence derived from the inboard arm of the telomere contains 3 bp, while that from the outboard arm contains 2 bp. An active origin is only created in the latter configuration, not because there is specific recognition of the dinucleotide, but apparently because steric considerations in the origin absolutely require that there be only two bases in this position. Segregation of an additional asymmetric nucleotide in the 'bubble' thus provides the virus with a highly effective mechanism for deriving both active and inactive origins from the same basic sequence.

Since the replication fork initiated at this site is unidirectional and supports the synthesis of a single, continuous DNA strand (S.F.Cotmore and P.Tattersall, in preparation), selection of a single origin from the dimer palindrome results in the conservation of a single DNA sequence (called the 'flip' sequence) in the progeny viral telomere, as is observed *in vivo* (Astell *et al.*, 1985). In contrast, if both arms of the junction palindrome contained active origins, two complementary but inverted sequences ('flip' and 'flop') would be generated in equimolar concentrations. This latter arrangement is encountered in the palindromic tetramer-junction sequences derived from the 5′

end of the MVM genome (Cotmore and Tattersall, 1992), and would be expected to provide a more efficient replication template. However, telomeric sequences serve a variety of functions in addition to initiating DNA replication, and the 3′ palindrome is likely to be particularly constrained because of the close proximity of the P_4 transcriptional promoter. Although most of the autonomous, or helper-independent, parvoviruses have mismatched bubble nucleotides and show conservation of a single DNA sequence in their 3′ telomeres (Astell *et al.*, 1979, 1985; Sahli *et al.*, 1985; Diffoot *et al.*, 1989; Martyn *et al.*, 1990; Vasudevacharya *et al.*, 1990), it was not obvious how such features could be beneficial. Our data suggest that these two features are linked, and that this asymmetry provides the virus with a mechanism for restricting replication initiation to certain telomere configurations and so, by default, liberating others.

Initial studies suggested that parvoviral replication-initiator proteins might specifically recognize some aspect of the complex secondary structure associated with the ear region of viral telomeres (Im and Muzyczka, 1990; Snyder *et al.*, 1993). However, such recognition is clearly not required at the MVM 3′ origin, since the ear regions are absent from most of the potent origin constructs described in this paper. Nevertheless, activation of this origin does require specific recognition of two separate elements: the nick site and an upstream ATF-binding consensus. In the helper-dependent parvoviruses, such as adeno-associated virus-2 (AAV-2), replication-initiation also requires the recognition of two elements: the nick site (Snyder *et al.*, 1993) and a sequence located at the top of the telomeric stem in approximately the same position as the MVM ATF binding site consensus. However, in this case the viral initiator protein *Rep* binds directly to the upstream site (Chiorini *et al.*, 1994; Weitzman *et al.*, 1994). Although it is possible that NS1 also binds directly to the MVM upstream recognition element, several lines of evidence argue against this. First, it has proven difficult to demonstrate any direct binding between NS1 and its cognate origin, although a variety of cellular proteins can be shown to interact specifically with this sequence (S.F.Cotmore *et al.*, unpublished; Faisst *et al.*, 1994; Z.Gu, S.Plaza, M.Perros, C.Cziepluch, J.Rommelaere and J.J.Cornelis, submitted). Second, in the MVM 3′ origin the upstream sequence is an excellent fit to the consensus ATF binding site (Hai *et al.*, 1989), and oligonucleotides containing the MVM origin sequence bind *in vitro* translated ATF-1 and ATF-2 molecules efficiently (S.F.Cotmore and P.Tattersall, unpublished observations). Third, a point mutation, which is known to severely impair the binding of ATF to the identical sequence, GTGACGT, in the promoter of the retinoblastoma gene (Sakai *et al.*, 1991), is similarly debilitating for replication initiation at the MVM 3′ origin, and, finally, the MVM 5′ origin, which is also nicked by NS1, has no ATF-like site but does have a consensus binding site, of unknown function, for the transcription factor Sp1 located ~50 nucleotides from the nick site. These observations suggest that replication complexes initiated by NS1 may utilize cellular site-specific DNA binding proteins to mediate critical interactions with distant sequences in the origin, a possibility we are currently investigating.

Like the AAV *Rep* molecules, many of the proteins which are known to initiate rolling-circle replication in prokaryotic systems are multivalent DNA-binding molecules, containing one protein domain which recognizes and binds the initiator to the origin, and another separate domain which contains the active-site tyrosine(s) which mediate the nick (Baas and Jansz, 1988; Reygers *et al.*, 1991; Matson *et al.*, 1993). However, there are other examples, such as the conjugal transfer plasmid RP4, where these two functions are specified on separate gene products which must interact with each other in order to effect initiation (Pansegrau *et al.*, 1990). Possibly, MVM has evolved a third strategy, using a cellular DNA-binding protein to mediate site specificity between the initiator and the origin. Since NS1 also upregulates transcription from both the viral P_4 and P_{38} promoters (Rhode, 1985; Doerig *et al.*, 1988, 1990; Hanson and Rhode, 1991), protein–protein interactions of this type with cellular transcription factors could serve multiple roles in the viral life cycle.

Materials and methods

Plasmid constructs

The plasmid pLEB 407, containing the MVM (3′) dimer bridge region, was obtained by excising the MVM 3′ bridge fragment from pLEB 711 (Cotmore and Tattersall, 1992) with *Nco*I (MVM nucleotide 259), and re-cloning it into a pUC 19 vector which had been cut at the *Sma*I site and modified by the addition of an *Nco*I linker (d[CCCATGGG]). The single-arm clones pGAA and pTC were obtained by digesting pLEB 407 with *Bss*HII, which cuts four times in the 'ears' region of the MVM bridge insert (Figure 2A). Digested plasmid was re-cut once, in the polylinker of the vector, with either *Kpn*I (to obtain the plasmid pTC) or *Bam*HI (to obtain the plasmid pGAA). Clones were selected which retained one *Bss*HII site, derived either from the MVM sequence (pGAA) or from the fusion of blunted *Bss*HII and *Kpn*I sites (pTC), and their DNA sequences confirmed using Sequenase® (USB, Cleveland, OH).

PCR amplification of selected MVM sequences from pTC and pGAA was carried out using GeneAmp™ (Perkin Elmer Cetus, Norwalk, CT) as described by the manufacturers. Products were ligated directly into the vector pCRII as described by the supplier (Invitrogen, San Diego, CA). Recombinant clones were checked by sequencing the viral insert.

A plasmid, pREB 1412, containing palindromic MVM 5′ sequences derived from the junction region of tetrameric MVM RF DNA, has been described previously (Cotmore and Tattersall, 1992; Cotmore *et al.*, 1992). This insert resides in the *Xba*I site of pUC 19, and was excised by cutting in the vector polylinker with *Bam*HI and *Pst*I. After gel purification, the insert was cut near the palindromic axis with *Msp*I, and the mixture ligated into a pUC 19 vector cut with *Acc*I and *Pst*I. The resulting plasmid, p5′AGA, contains the three additional bases, 5′AGA, which differentiate one arm of the 5′ telomere from the other (Astell *et al.*, 1985).

Preparation of replication templates

Unlabeled supercoiled plasmid DNA was used as the template for most replication assays, although unlabeled, linearized substrates also work well (data not shown). To determine the NS1 nick site (Figure 4), the insert from pTC was excised with *Hin*dIII and *Xba*I, 3′ end-labeled with [^{32}P]dGTP and Klenow polymerase, and re-digested with *Bam*HI which cuts next to *Hin*dIII in the vector sequence. This gave a DNA fragment of 169 bp, labeled with [^{32}P]dGMP at the 3′ of the TC strand containing the NS1 nick site.

Cell extracts and replication assays

Replication extracts were prepared from uninfected HeLa cells as described by Stillman and Gluzman (1985). Nuclear extracts containing wild-type copies of NS1 were prepared from HeLa cells 15 h after infection with 15 p.f.u./cell of each of the two recombinant vaccinia viruses vTF7-3, which expresses T7 polymerase, and vv-$NS1_{wt}$, which expresses full-length copies of NS1 from the T7 promoter. Replication assays contained extracts from uninfected cells supplemented with nuclear extracts containing NS1, deoxynucleotides, ATP and an ATP-regenerating system, template DNA (2 μg/ml) and a ^{32}P-labeled dNTP

as previously described (Cotmore *et al.*, 1992). Incorporation of [^{32}P]dNMP into DNA was monitored by spotting a sample onto DEAE paper (DE81, Whatman), followed by three successive washes with 0.5 M Na_2HPO_4, one wash with H_2O, one wash with ethanol and liquid scintillation counting.

In some assays, [^{32}P]dNTP was omitted and a ^{32}P-labeled template employed. Reactions were terminated and the products analyzed as previously described (Cotmore *et al.*, 1992). Briefly, samples to be immunoprecipitated were incubated in buffer containing 2% SDS for 30 min at 60–70°C, diluted to 0.2% SDS and precipitated overnight with rabbit antiserum directed against the NH_2-terminal domain of NS1 (Cotmore and Tattersall, 1986). In most cases, the reaction products were linearized with restriction endonucleases, as specified in the text, prior to electrophoresis in order to restrict the complexity of the products. All samples were digested with proteinase K in 0.5% SDS at 60°C prior to electrophoresis. Immunoprecipitates to be analyzed on sequencing gels were extracted with phenol and ethanol precipitated prior to analysis.

Acknowledgements

We are indebted to Jürg Nüesch for help in the preparation of vaccinia-infected cell extracts and for critical discussion of the work as it progressed. This work was supported by Public Health Service grants AI26109 and CA29303 from the National Institutes of Health.

References

Astell,C.R., Smith,M., Chow,M.B. and Ward,D.C. (1979) *Cell*, **17**, 691–703.
Astell,C.R., Chow,M.B. and Ward,D.C. (1985) *J. Virol.*, **54**, 171–177.
Baas,P.D. and Jansz,H.S. (1988) *Curr. Top. Microbiol. Immunol.*, **136**, 31–70.
Chiorini,J.A., Weitzman,M.D., Owens,R.A., Urcelay,E., Safer,B.B. and Kotin,R.M. (1994) *J. Virol.*, **68**, 797–804.
Cotmore,S.F. and Tattersall,P. (1986) *J. Virol.*, **58**, 921–926.
Cotmore,S.F. and Tattersall,P. (1987) *Adv. Virus Res.*, **33**, 91–174.
Cotmore,S.F. and Tattersall,P. (1989) *J. Virol.*, **63**, 3902–3911.
Cotmore,S.F. and Tattersall,P. (1992) *J. Virol.*, **66**, 420–431.
Cotmore,S.F., Nüesch,J.P.F. and Tattersall,P. (1992) *Virology*, **190**, 365–377.
Cotmore,S.F., Nüesch,J.P.F and Tattersall,P. (1993) *J. Virol.*, **67**, 1579–1589.
Diffoot,N., Schull,K.C., Chen,K.C., Stout,E.R., Lederman,M. and Bates,R.C. (1989) *J. Virol.*, **63**, 3180–3184.
Diffoot,N., Chen,K.C., Bates,R.C. and Lederman,M. (1993) *Virology*, **192**, 339–345.
Doerig,C., Hirt,B., Beard,P. and Antonietti,J.P. (1988) *J. Gen. Virol.*, **69**, 2563–2573.
Doerig,C., Hirt,B., Antonietti,J.P. and Beard,P. (1990) *J. Virol.*, **64**, 387–396.
Faisst,S., Perros,M., Deleu,L., Spruyt,N. and Rommelaere,J. (1994) *Virology*, **202**, 466–470.
Faust,E.A. and Ward,D.C. (1979) *J. Virol.*, **32**, 276–292.
Hai,T., Liu,F., Coukos,W.J. and Green,M.R. (1989) *Genes Dev.*, **3**, 2083–2090.
Hanson,N.D. and Rhode,S.L. (1991) *J. Virol.*, **65**, 4325–4333.
Ilyina,T.V. and Koonin,E.V. (1992) *Nucleic Acids Res.*, **20**, 3279–3285.
Im,D.-S. and Muzyczka,N. (1990) *Cell*, **61**, 447–457.
Inamoto,S., Yoshioka,Y. and Ohtsubo,E. (1991) *J. Biol. Chem.*, **266**, 10086–10092.
Martyn,J.C., Davidson,B.E. and Studdert,M.J. (1990) *J. Gen. Virol.*, **71**, 2747–2753.
Matson,S.W., Nelson,W.C. and Morton,B.S. (1993) *J. Bacteriol.*, **175**, 2599–2606.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 490–560.
Nüesch,J.P.F., Cotmore,S.F. and Tattersall,P. (1992) *Virology*, **191**, 406–416.
Pansegrau,W., Balzer,D., Kruft,V., Lurz,R. and Lanka,E. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 6555–6559.
Reygers,U., Wessel,R., Muller,H. and Hoffmann-Berling,H. (1991) *EMBO J.*, **10**, 2689–2694.
Rhode,S.L. (1985) *J. Virol.*, **55**, 886–889.
Sahli,R., McMaster,G.K. and Hirt,B. (1985) *Nucleic Acids Res.*, **13**, 3617–3633.
Sakai,T., Ohtani,N., McGee,T.L., Robbins,P.D. and Dryja,T.P. (1991) *Nature*, **352**, 83–86.
Snyder,R.O., Im,D.S., Ni,T., Xiao,X., Samulski,R.J. and Muzyczka,N. (1993) *J. Virol.*, **67**, 6096–6104.
Stillman,B.W. and Gluzman,Y. (1985) *Mol. Cell. Biol.*, **5**, 2051–2060.
Tam,P. and Astell,C.R. (1993) *Virology*, **193**, 812–824.
Tattersall,P. and Ward,D.C. (1976) *Nature*, **263**, 106–109.
Vasudevacharya,J., Basak,S., Srinivas,R.V. and Compans,R.W. (1990) *Virology*, **178**, 611–616.
Wang,P.-Z., Projan,S.J., Henriquez,V. and Novick,R.P. (1993) *EMBO J.*, **12**, 45–52.
Weitzman,M.D., Kyostio,S.R.M., Kotin,R.M. and Owens,R.A. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 5808–5812.

Received on April 29, 1994; revised on June 28, 1994