

Segregation of a Single Outboard Left-End Origin Is Essential for the Viability of Parvovirus Minute Virus of Mice[▽]

Erik Burnett,^{2†} Susan F. Cotmore,¹ and Peter Tattersall^{1,2*}

Departments of Laboratory Medicine¹ and Genetics,² Yale University Medical School,
333 Cedar Street, New Haven, Connecticut 067510

Received 13 July 2006/Accepted 11 August 2006

During DNA replication, the hairpin telomeres of Minute Virus of Mice (MVM) are extended and copied to create imperfectly palindromic duplex junction sequences that bridge adjacent genomes in concatameric replicative-form DNA. These are resolved by the viral initiator protein, NS1, but mechanisms employed at the two telomeres differ. Left-end:left-end junctions are resolved asymmetrically at a single site, OriL_{TC}, by NS1 acting in concert with a host factor, parvovirus initiation factor (PIF). Replication segregates doublet and triplet sequences, initially present as unpaired nucleotides in the bubble region of the left-end hairpin stem, to either side of the junction. These act as spacers between the NS1 and PIF binding sites, and their asymmetric distribution sets up active (OriL_{TC}) and inactive (OriL_{GAA}) forms of OriL. We used a reverse genetic approach to disrupt this asymmetry and found that neither opposing doublets nor triplets in the hairpin bubble were tolerated. Viable mutants were isolated at low frequency and found to contain second-site mutations that either restored the asymmetry or crippled one PIF binding site. These mutations either inactivated the inboard or activated the outboard form of OriL, a polarity that strongly suggests that, in the genus *Parvovirus*, an active inboard OriL is lethal.

Parvoviral genomes are single-stranded, linear DNA molecules, of about 5 kb, with small palindromic telomeres that fold into self-priming hairpins to mediate the viruses' rolling hairpin replication strategy (11, 12). These hairpins serve as "hinges" that repeatedly unfold and refold during replication, reversing the direction of synthesis at each end of the linear molecule and creating long palindromic replicative-form (RF) DNA intermediates in which inverted copies of the viral genome are separated by duplex copies of the terminal hairpins. The latter also serve as replication origins, where the viral initiator protein, NS1, is able to nick one strand of the DNA in a sequence- and strand-specific fashion, generating a base-paired 3' nucleotide that can prime a cellular leading-strand DNA polymerase. This initiates unidirectional strand displacement synthesis, which resolves the junction while also regenerating the terminal sequences on each of the daughter molecules. In many parvoviral genera the terminal palindromes form part of inverted terminal repeats, so that the same DNA origin sequence occurs at the two ends of the genome and each terminus is resolved by the same mechanism with equal efficiency. However, termini of species belonging to the *Parvovirus*, *Amdovirus*, *Bocavirus*, and *Brevidensovirus* genera differ from each other in size, sequence, and predicted secondary structure (9). This renders their replication conspicuously asymmetric and ultimately drives the encapsidation of predominantly negative-sense progeny single strands (8). Although the replication of Minute Virus of Mice (MVM), one such virus

from the genus *Parvovirus*, has been studied in some detail, it is not clear how it may benefit from this asymmetric strategy.

Most parvoviral telomeres are nicked and copied in a process called "hairpin transfer," which inverts the palindrome with each round of synthesis. However, the left end of MVM is resolved by an alternate mechanism, termed "junction resolution," that preserves it in a single-sequence orientation in progeny genomes. As shown in Fig. 1, the unique 121-nucleotide left-end sequence of MVM folds into a Y-shaped hairpin containing small internal palindromes, forming the "ears" of the Y, and a 43-bp duplex stem region that is interrupted by a mismatched "bubble" sequence, where the triplet 5'-GAA-3' on the inboard arm is opposed by the doublet 5'-GA-3' on the outboard strand. Viral DNA synthesis is initially primed from the 3' nucleotide of this hairpin, generating a duplex molecule in which the two strands are effectively cross-linked by the hairpin. However, potential origin sequences 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 (Fig. 1). 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 bp in length, extending from two 5'-ACGT-3' motifs spaced 5 nucleotides apart at one end to a position some 7 bp beyond the nick site (7). The two 5'-ACGT-3' sequences serve as half sites that cooperatively bind a heterodimeric cellular factor called PIF, for parvovirus initiation factor (4, 12), also known as glucocorticoid modulating element-binding protein, or GMEB (19), which is abundantly expressed in many cell types. The binding

* Corresponding author. Mailing address: Departments of Laboratory Medicine and Genetics, Yale University Medical School, 333 Cedar St., CB408, New Haven, CT 067510. Phone: (203) 785-4586. Fax: (203) 688-7340. E-mail: peter.tattersall@yale.edu.

† Present address: Lawrence Livermore National Laboratory, Livermore, CA 94551.

[▽] Published ahead of print on 23 August 2006.

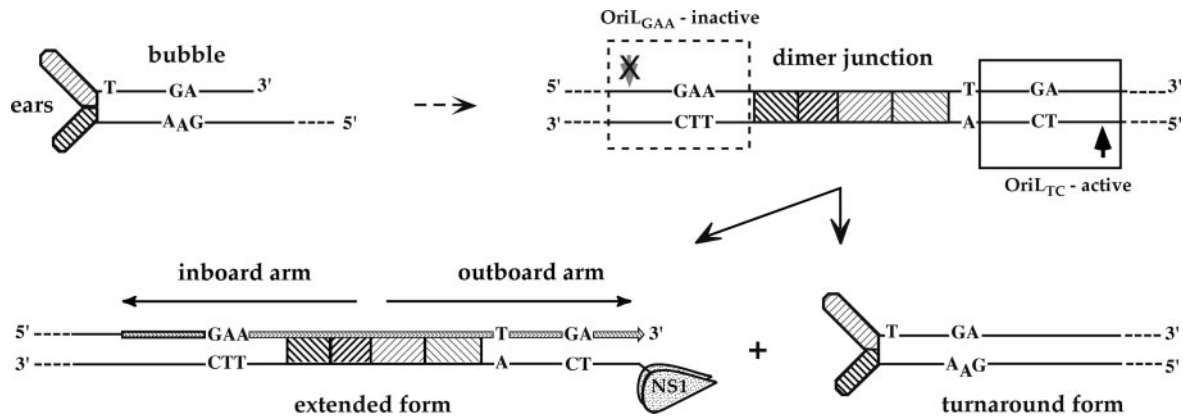


FIG. 1. Asymmetric resolution is controlled by the bubble. Shown is a diagram of the left-end hairpin of MVM, showing sequences of the bubble nucleotides that direct the segregation of active and inactive origins on either side of left-end:left-end junctions. The lower panel depicts the major products resulting from the ensuing replication/resolution reaction.

specificity of this complex is unusual in that while it requires both half sites, these sites can be flexibly spaced from 1 to 9 nucleotides apart, with an optimal spacing of 6 nucleotides (1). In OriL_{TC} , PIF makes contact with NS1 across the bubble dinucleotide sequence, stabilizing its binding to its cognate DNA binding site, activating NS1 to melt and subsequently nick its duplex nick site (3, 5). 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 OriL_{GAA} . We have previously shown that, while the actual sequence of the OriL_{TC} doublet is relatively unimportant, 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 (7, 12). The bubble mismatch therefore restricts origin activity to duplex forms of

the outboard arm of the hairpin and causes them to be resolved asymmetrically, as shown in Fig. 1.

In the current study, we explore the effects of three mutant hairpins, detailed in Fig. 2A, which destroy the dimer bridge asymmetry by introducing sequences that modify or eliminate the mismatch. We use a reverse genetic system described previously (2), in which synthetic oligonucleotides corresponding to the wild-type and mutant OriL sequences are ligated to a truncated genome and the products transfected into A9 cells. Our initial expectation was that the GAA-GAA mutant would not be viable, since it should be unable to form a functional replication origin on either side of the dimer junction. In contrast, we predicted that the other two mutants might be viable but impaired, the extent of their disability reflecting the importance of maintaining asymmetry in the replication process. Unexpectedly, each of the three mutant constructs was defective, giving rise to between 30- and 90-fold fewer plaques than

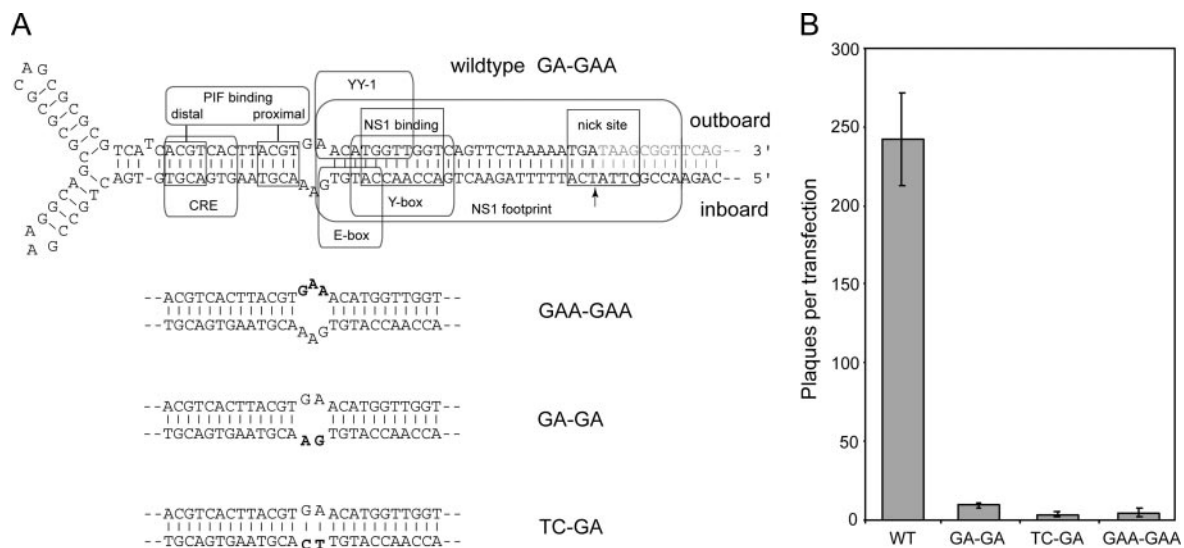


FIG. 2. Reverse genetic analysis of the bubble. Panel A details the distribution of transcription and replication factor binding sites within the left-end hairpin of MVM. Below this are shown the three mutant bubble sequences introduced into the infectious clone by oligonucleotide ligation (2). Panel B shows results from multiple transfections of ligation products generated with wild-type (WT) and mutant oligonucleotides.

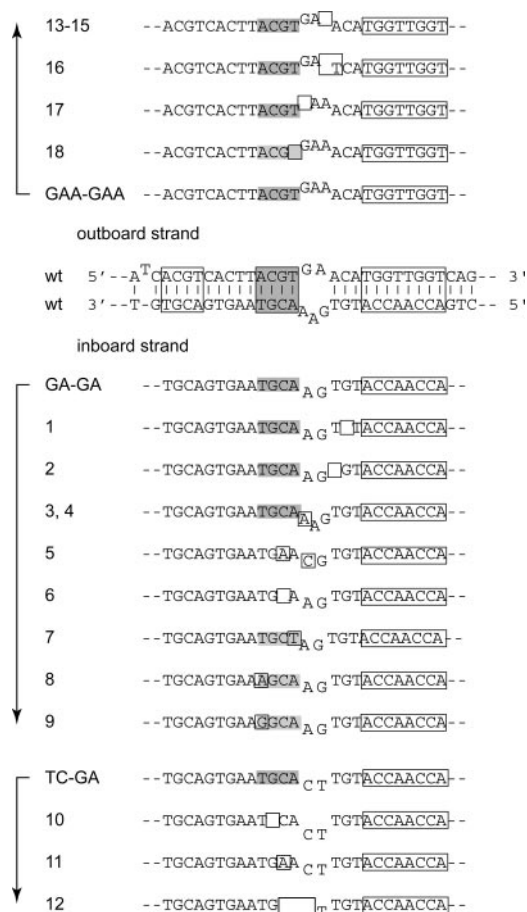


FIG. 3. Second-site revertants of the lethal phenotype. DNA sequences were determined for revertants recovered as individual plaques, as shown in Fig. 2B. For each revertant, only the strand containing nucleotide changes from the input is shown (boxed), with deleted nucleotides indicated by an empty box. Proximal PIF half sites are shaded dark gray. Those with mutations predicted to be debilitating are shaded light gray, whereas those predicted to be inactive are unshaded. wt, wild type.

the wild type (Fig. 2B). Since oligonucleotide synthesis typically generates a low level of erroneous product, we considered that these plaques might represent second-site suppressor mutations that recreate a viable genome from otherwise lethal mutant forms. Therefore, we picked several plaques from each transfection, established infections with these stocks, and cloned and sequenced extended forms of the viral left end from RF DNA.

Left-end termini of viruses created using an oligonucleotide identical to the wild-type sequence always generated virus with unaltered input sequence. In contrast, nine viruses recovered from three separate transfections using the GA-GA construct had sequences on the inboard arm that differed from the input oligonucleotide (Fig. 3) while retaining the wild-type sequence, containing the GA dinucleotide, on the outboard arm. Thus, there appears to be various viable solutions to the apparently lethal condition of having doublet spacers on both sides of the dimer junction. For example, GA-GA-derived clones 1 and 2 contain single base deletions in this critical spacer region that

separates the NS1 and PIF binding sites, while clones 3 and 4 contain a single base insertion in the bubble, thereby regenerating the inactivating trinucleotide in the inboard arm. Deriving an alternate solution, clones 5 to 9 retain the size of the input bubble region but contain mutations predicted to eliminate the proximal PIF binding site. Accordingly, in clones 5 and 6 the CpG core of the proximal PIF half site was destroyed, while in clones 7 to 9 transversions in positions 1 and 4 of this proximal PIF site would be expected to severely impair, but perhaps not totally eliminate, PIF binding to the inboard arm. Thus, failure to isolate virus with the GA-GA input sequence suggests that having two apposed dinucleotides in the bubble, and hence generating two active origin sequences, creates a nonviable genome. The universal inactivation of the inboard, rather than the outboard, OriL further indicates that the lethal condition is due to creating an active origin in the prospective inboard arm of the telomere.

To explore this possibility further, we retrieved and sequenced viruses from transfections using the TC-GA dinucleotide hairpin. Like the GA-GA oligonucleotide, this construct contains two potentially active origins but lacks the helical disturbance of noncomplementary bases created by the bubble. Thus, by eliminating the bubble mismatch from the hairpin configuration of the telomere and replacing it with a base-paired dinucleotide, this mutation would generate a left-end hairpin that was theoretically capable of supporting hairpin transfer. While only three plaques were obtained following two separate transfections with this TG-GA construct, these three isolates (clones 10 to 12 in Fig. 3) contained mutations in the proximal PIF half site derived from the inboard arm while retaining wild-type sequence in the outboard arm. Each of these mutations destroyed the invariant CpG at the core of the PIF site, thus precluding PIF binding and inactivating the potential origin sequence associated with this inboard arm, either when presented in duplex RF DNA or in its hairpin form in progeny virus.

Similarly, the six viruses isolated from genomes reconstituted with the GAA-GAA mutation each failed to conform to the input sequence. However, in this instance they all retained wild-type sequence in the inboard arm but contained a single base deletion in the outer arm between the NS1 and PIF binding sites that restored the correct spacing for an active origin (Fig. 3). Three of the six viruses sequenced (clones 13 to 15) had regenerated the wild-type sequence in the outer arm. The frequency of this particular solution might be expected because the input sequence contains three consecutive adenine bases, and a deletion of any one of them would be indistinguishable. In clones 16 and 17, an alternate single base has been lost from this spacer element, while in clone 18 the final T from the proximal PIF half site has been lost, allowing its place to be taken by the supernumerary G residue from the bubble. This transversion would be predicted to create a sub-optimal PIF binding site; however, the correct spacing of the NS1 and PIF sites is regenerated, suggesting that in an active origin, the need for correct spacing is of greater importance than perfect PIF affinity. However, it should be noted that a mutation in a similar position in clone 7, in this case a transition, is predicted to restore enough origin asymmetry to render the resulting virus viable. We are currently exploring these two mutant viruses in greater detail, assessing their ability to am-

plify against wild-type virus at low input multiplicities and exploring their life cycle in the hope that this will reveal the nature of the defect arising from having just one, rather inefficient, left-end origin in the outboard arm, as in clone 18, versus having an active outboard origin sequence and a less effective inboard sequence.

Unlike the inverted terminal repeats of many parvoviruses, the left-end hairpin of MVM is not simply dedicated to replication but also contains a series of important upstream control elements for the nearby P4 promoter, interspersed and overlapping with its replication motifs (13, 17). Thus, the asymmetric sequence conservation observed might reflect a need to segregate elements involved in these two processes, such as transcription factor binding sites, to different arms, or forms, of the hairpin. As indicated in Fig. 2, NF-Y would bind equally well to Y-boxes present in the inboard and outboard palindromic arms and in the hairpin, while USF would bind to an E-box created exclusively within the inboard arm. However, mutagenesis of these motifs in reporter constructs indicate that NF-Y binding influences P4 transcription while USF does not (15), suggesting that neither site is likely to be a physiologically significant asymmetric binding module. Indeed, although the nucleotide changes in variants 2 and 3 likely inactivate NF-Y binding, these mutations occur on the arm that is presumed to be transcriptionally active. On the other hand, as indicated in Fig. 2, the outboard palindromic arm contains a CCAT-core YY1 consensus binding sequence (18), while the inboard arm does not. Since YY1 can act as both a repressor and an activator of transcription, depending on the particular promoter (14), it will be important to determine whether or not the selected mutant sequences bind YY1 and whether a requirement for specific repression of transcription exercised through the TC bubble sequence might be responsible for the selected asymmetry.

An alternative explanation for the observed polarity would be that, while it is essential to have one active origin, an active inboard origin is lethal in some way. One consequence of such an arrangement might be that replication initiating on an inboard OriL would set up a fork that proceeded in the outboard direction. If this happened on the extended form of the left end, depicted in Fig. 1, it would lead to the continuous displacement of isolated hairpins, as DNA synthesis repeatedly copied just the terminal palindrome. Possibly, isolated 3' hairpins might interfere with the establishment of viral packaging complexes, or, alternatively, this nonproductive amplification might not be lethal per se but simply impede the amplification process. Significantly, each of the displaced hairpins would have an NS1 molecule covalently attached to its 5' end, so that active replication at an inboard left-end origin might create a "sink" for those NS1 molecules that are activated for nicking a species that may be limiting for infection (16). If this is the explanation, then an alternative mechanism must exist for suppression of the inboard origin at the right end of the genome, where NS1 nicking is activated by HMG1/2 rather than by PIF (10) and where both outboard and inboard origins are active in vitro (6).

Parvoviruses are unable to transcribe their genes until the cell enters S phase, at least in part because complementary strand synthesis is required to create a duplex transcription template. Thus, during the first hours of S phase, a balance

must be achieved between the use of duplex strands as transcription versus replication substrates. During this critical time, it may prove equally or more efficient to restrict the number of replication forks that can be established in the vicinity of the P4 promoter. Thus, given the virus' precarious life style, mutations that suppress the efficiency of the inboard left-end origin might initially simply have provided kinetic advantages during the first hours of S phase, when the virus must usurp control to ensure the maximum virus burst from each infected cell. If such a strategy did emerge, it would provide an opportunity for the inboard sequences to evolve, allowing them to serve additional functions or to dispense with constraints required to limit repeated initiation at this site in extended-form RF, as discussed above. Thus, an early arbitrary or minimally advantageous selection could lead eventually to modifications that would render asymmetric resolution essential. This appears to have been an easily adopted and successful strategy, as it occurs in four of the nine currently recognized parvovirus genera. Presumably this has occurred as the result of convergent evolution, since some of these genera are phylogenetically remote from one another.

DNA oligonucleotides used in this study were synthesized at the HHMI Biopolymer Laboratory and W. M. Keck Foundation Biotechnology Resource Laboratory at Yale University.

This work was supported by U.S. Public Health Service grant number AI26109. E.B. was supported, in part, by T32 GM07499.

REFERENCES

1. Burnett, E., J. Christensen, and P. Tattersall. 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.
2. Burnett, E., and P. Tattersall. 2003. Reverse genetic system for the analysis of parvovirus telomeres reveals interactions between transcription factor binding sites in the hairpin stem. *J. Virol.* **77**:8650–8660.
3. Christensen, J., S. F. Cotmore, and P. Tattersall. 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.
4. Christensen, J., S. F. Cotmore, and P. Tattersall. 1999. Two new members of the emerging KDWK family of combinatorial transcription modulators bind as a heterodimer to flexibly spaced PuCGPy half-sites. *Mol. Cell. Biol.* **19**:7741–7750.
5. Christensen, J., and P. Tattersall. 2002. Parvovirus initiator protein NS1 and RPA coordinate replication fork progression in a reconstituted DNA replication system. *J. Virol.* **76**:6518–6531.
6. Cotmore, S. F., J. Christensen, and P. Tattersall. 2000. Two widely distributed initiator binding sites create an HMG1-dependent parvovirus rolling-hairpin replication origin. *J. Virol.* **74**:1332–1341.
7. Cotmore, S. F., and P. Tattersall. 1994. An asymmetric nucleotide in the parvoviral 3' hairpin directs segregation of a single active origin of DNA replication. *EMBO J.* **13**:4145–4152.
8. Cotmore, S. F., and P. Tattersall. 2005. Genome packaging sense is controlled by the efficiency of the nick site in the right-end replication origin of parvoviruses minute virus of mice and LuIII. *J. Virol.* **79**:2287–2300.
9. Cotmore, S. F., and P. Tattersall. 2006. Genome structure and organisation. Chapter 7, p. 5–16. *In* J. Kerr, S. F. Cotmore, M. E. Bloom, R. M. Linden, and C. R. Parrish (ed.), *The parvoviruses*. Hodder Arnold, London, United Kingdom.
10. Cotmore, S. F., and P. Tattersall. 1998. High-mobility group 1/2 proteins are essential for initiating rolling-circle-type DNA replication at a parvovirus hairpin origin. *J. Virol.* **72**:8477–8484.
11. Cotmore, S. F., and P. Tattersall. 2006. Parvoviruses, p. 593–608. *In* M. L. DePamphilis (ed.), *DNA replication and human disease*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
12. Cotmore, S. F., and P. Tattersall. 2006. A rolling hairpin strategy: basic mechanisms of DNA replication in the parvoviruses. Chapter 14, p. 5–16. *In* J. Kerr, S. F. Cotmore, M. E. Bloom, R. M. Linden, and C. R. Parrish (ed.), *The parvoviruses*. Hodder Arnold, London, United Kingdom.
13. Faisst, S., M. Perros, L. Deleu, N. Spruyt, and J. Rommelaere. 1994. Mapping of upstream regulatory elements in the P4 promoter of parvovirus minute virus of mice. *Virology* **202**:466–470.

14. **Gordon, S., G. Akopyan, H. Garban, and B. Bonavida.** 2006. Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. *Oncogene* **25**:1125–1142.
15. **Gu, Z., S. Plaza, M. Perros, C. Cziepluch, J. Rommelaere, and J. J. Cornelis.** 1995. NF-Y controls transcription of the minute virus of mice P4 promoter through interaction with an unusual binding site. *J. Virol.* **69**:239–246.
16. **Nuesch, J. P., R. Corbau, P. Tattersall, and J. Rommelaere.** 1998. Biochemical activities of minute virus of mice nonstructural protein NS1 are modulated in vitro by the phosphorylation state of the polypeptide. *J. Virol.* **72**:8002–8012.
17. **Perros, M., L. Deleu, J. M. Vanacker, Z. Kherrouche, N. Spruyt, S. Faisst, and J. Rommelaere.** 1995. Upstream CREs participate in the basal activity of minute virus of mice promoter P4 and in its stimulation in ras-transformed cells. *J. Virol.* **69**:5506–5515.
18. **Yant, S. R., W. Zhu, D. Millinoff, J. L. Slightom, M. Goodman, and D. L. Gumucio.** 1995. High affinity YY1 binding motifs: identification of two core types (ACAT and CCAT) and distribution of potential binding sites within the human beta globin cluster. *Nucleic Acids Res.* **23**:4353–4362.
19. **Zeng, H., D. A. Jackson, H. Oshima, and S. S. Simons, Jr.** 1998. Cloning and characterization of a novel binding factor (GMEB-2) of the glucocorticoid modulatory element. *J. Biol. Chem.* **273**:17756–17762.