JOURNAL OF VIROLOGY, Aug. 1997, p. 5733–5741 Vol. 71, No. 8
0022-538X/97/$04.00+0

Parvovirus Initiation Factor PIF: a Novel Human DNA-Binding Factor Which Coordinately Recognizes Two ACGT Motifs

JESPER CHRISTENSEN,[1,2] SUSAN F. COTMORE,[1] AND PETER TATTERSALL[1,3]*

Departments of Laboratory Medicine[1] and Genetics,[3] Yale University School of Medicine, New Haven, Connecticut 06510, and Department for Virology and Immunology, The Royal Veterinary and Agricultural University of Copenhagen, 1870 Frederiksberg C, Denmark[2]

Received 5 February 1997/Accepted 23 April 1997

A novel human site-specific DNA-binding factor has been partially purified from extracts of HeLa S3 cells. This factor, designated PIF, for parvovirus initiation factor, binds to the minimal origin of DNA replication at the 3′ end of the minute virus of mice (MVM) genome and functions as an essential cofactor in the replication initiation process. Here we show that PIF is required for the viral replicator protein NS1 to nick and become covalently attached to a specific site in the origin sequence in a reaction which requires ATP hydrolysis. DNase I and copper *ortho*-phenanthroline degradation of the PIF-DNA complexes showed that PIF protects a stretch of some 20 nucleotides, covering the entire region in the minimal left-end origin not already known to be occupied by NS1. Methylation and carboxy-ethylation interference analysis identified two ACGT motifs, spaced by five nucleotides, as the sequences responsible for this binding. A series of mutant oligonucleotides was then used as competitive inhibitors in gel mobility shift assays to confirm that PIF recognizes both of these ACGT sequences and to demonstrate that the two motifs comprise a single binding site rather than two separate sites. Competitive inhibition of the origin nicking assay, using the same group of oligonucleotides, confirmed that the same cellular factor is responsible for both mobility shift and nicking activities. UV cross-linking and relative mobility assays suggest that PIF binds DNA as a heterodimer or higher-order multimer with subunits in the 80- to 100-kDa range.

Minute virus of mice (MVM) is a member of the autonomous parvoviruses, a group of small (260-Å) nonenveloped viruses with linear, single-stranded DNA genomes approximately 5 kb in length. Parvoviruses replicate their DNA through a series of duplex concatemeric intermediates by a unidirectional rolling hairpin mechanism (11), which closely resembles the rolling circle replication mechanisms previously characterized in certain prokaryotic systems (15). For MVM, a single multifunctional viral phosphoprotein, NS1, serves as a site-specific DNA-binding protein which recognizes sequences within specialized viral origins and initiates replication by introducing a single-stranded nick within a sequence located close to its core recognition site. This cleavage reaction leaves the NS1 polypeptide covalently attached to the 5′ end at the nick site, via a phosphotyrosine bond, and generates a base-paired 3′ nucleotide which serves as a primer for DNA synthesis. Since NS1 is the only viral nonstructural gene product absolutely required for rolling hairpin synthesis (6, 20, 21), all of the remaining replication factors must be provided by the host cell.

The negative-sense, single-stranded viral genome is bracketed by short palindromic sequences which fold back on themselves to form imperfect terminal hairpin duplexes (1). Although the sequences and structures of these two duplex telomeres are very different, they both contain sequence elements which can serve as replication origins in one or other of the various replicative intermediate DNA forms (2, 7–9, 17). In this report, we focus on the duplex sequences which create the origin at the left-hand (3′) end of the genome. In the viral single strand, the 3′ telomere assumes a Y-shaped structure in which there are two internal hairpins, forming the ears, and a third palindrome forming a stem. Within the stem there is an important mismatched "bubble" sequence where a triplet, 5′-GAA-3′, on one strand opposes a dinucleotide, 5′-GA-3′, on the other, so that when replication through this hairpin unfolds and copies the palindrome, a double-stranded intermediate is generated in which these tri- and dinucleotide sequences are located on either side of the axis of symmetry. Although these arms have nearly identical sequences, only the arm containing the GA dinucleotide functions as a replication origin (10). Genetic mapping studies revealed that the minimal active origin, shown in Fig. 1, is a 50-bp sequence which contains three distinct recognition elements: an NS1 nick site (CTWWTCA); an $(ACCA)_2$ sequence which is a high-affinity NS1 binding site; and a third block of sequence which contains a consensus ATF (activated transcription factor) binding site. The NS1 binding site and the consensus ATF site are separated by an 11-bp sequence which includes the bubble dinucleotide. This dinucleotide is not itself a specific recognition element, since its actual sequence is unimportant, but instead it serves as a precise spacer domain, such that insertion of any third nucleotide here abolishes nicking and hence initiation (10). By itself, purified NS1 binds the minimal origin in an ATP-dependent manner but is unable to initiate replication by nicking the DNA (4). DNase I footprints show that NS1 protects a region of some 43 nucleotides, including most of the minimal origin and extending over flanking sequences beyond the nick site. However, the NS1 footprint ends abruptly in the middle of the bubble dinucleotide, leaving the 14-bp region harboring the ATF site unprotected (5).

We have partially purified a cellular factor from leached cytoplasmic extracts of human 293 cells which we have designated PIF (for parvovirus initiation factor, [4]). PIF functions as an essential cofactor which allows NS1 to nick and become covalently attached to the left-end origin. This factor appears

* Corresponding author. Mailing address: Departments of Laboratory Medicine and Genetics, Yale University School of Medicine, 333 Cedar St., New Haven, CT 06510. Phone: (203) 785-4586. Fax: (203) 785-7340. E-mail: Peter_Tattersall@QM.Yale.edu.

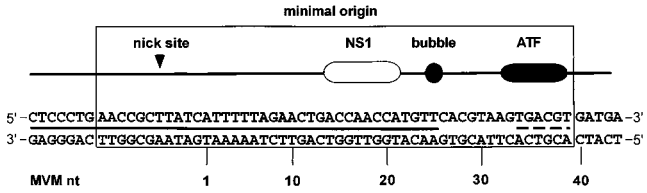

FIG. 1. Major elements in the minimal left-end origin of MVM. The nick site indicates the position at which NS1 nicks the DNA to initiate replication. The NS1 box indicates the specific NS1 recognition sequence $(ACCA)_2$, the bubble sequence (5′-TC-3′) is an important spacer element in the origin, and ATF indicates a consensus ATF binding site. The solid line between the two strands of DNA sequence indicates the region protected by NS1 from DNase I digestion, and the broken line indicates the position of the consensus ATF binding site. MVM nt denotes the MVM nucleotide numbers in the published sequence (1).

to bind to the region of the origin harboring the ATF site, but binding analyses show that PIF is not a member of the ATF family and that its ability to bind to its recognition sequence is inhibited by duplexes formed from the synthetic alternating polymer poly(dI-dC)-poly(dI-dC) (4).

In this report, we describe the purification of an identical activity from nuclear extracts of HeLa S3 cells and map the specific nucleotides in the left-end origin which are responsible for its binding.

MATERIALS AND METHODS

Cell extracts and fractionation. HeLa S3 cells were grown in Joklik-modified minimal essential medium supplemented with 5% horse serum at 37°C in spinner bottles, and nuclear extracts were prepared essentially as described by Dignam et al. (12) except that the dialysis step was omitted. To isolate PIF, extracts were diluted threefold in buffer A (25 mM Tris-HCl [pH 7.5], 1 mM EDTA, 0.1 mM phenylmethylsulfonic acid, 0.01% Nonidet P-40, 1 mM dithiothreitol [DTT], 10% glycerol) and loaded on to Q-Sepharose (3 ml of resin for each liter of original cell culture) equilibrated in buffer A adjusted to 175 mM NaCl. After washing, bound proteins were eluted in buffer A containing 300 mM NaCl. This eluate was applied to Zn^{2+}-metal chelate Sepharose (1.5 ml of resin for each liter of original cell culture), which was then washed with buffer A containing 300 mM NaCl, and bound proteins were eluted in buffer A adjusted to 150 mM NaCl and 50 mM imidazole. Eluted fractions were dialyzed against buffer A containing 50 mM NaCl and 20% sucrose and flash frozen in liquid N_2 before storage at −80°C. For nicking assays, eluates from Zn^{2+}-chelate Sepharose were further purified by using fast protein liquid chromatography on a MonoQ column (HR5/5; Pharmacia, Uppsala, Sweden) as described previously (4).

Recombinant MVM NS1. Recombinant histidine-tagged MVM NS1 was expressed in *Spodoptera frugiperda* Sf9 cells by using a baculovirus vector and purified as previously described (3).

Plasmid and probes. Plasmid pL1-2TC (10), containing the minimal active MVM left-end origin, was used as template for the analysis of most protein-DNA interactions. For DNase I protection, copper *ortho*-phenanthroline (Cu-OP) degradation, and methylation and carboxy-ethylation interference analyses, pL1-2TC was digested with either *Hin*dIII or *Xba*I and 3′ end labeled on one strand with [^{32}P]dATP or [^{32}P]dCTP, respectively. The substrate was then redigested with either *Xba*I or *Hin*dIII, respectively, and DNA fragments were purified from agarose gels.

For methylation interference assays, DNA fragments were methylated by incubating 10^6 cpm of probe in 200 μl of 1 mM EDTA–50 mM cacodylate buffer (pH 8.0) with 1 μl of dimethyl sulfate (DMS) for 5 min at room temperature. Reactions were terminated by adding 50 μl of 1.5 M sodium acetate containing 1 M β-mercaptoethanol, 60 μg of oyster glycogen, and 750 μl of cold ethanol. For carboxy-ethylation interference assays, probes were prepared essentially as described by Sturm et al. (25). Briefly, 3′-end-labeled DNA fragments were boiled, chilled, and diluted in 50 mM cacodylate (pH 8.0)–1 mM EDTA. Diethyl pyrocarbonate (DEPC) was added, and samples were vortexed and incubated at 37°C for 20 min with frequent mixing. DNA was recovered by ethanol precipitation, resuspended in 10 mM Tris-HCl (pH 7.9)–10 mM $MgCl_2$–50 mM NaCl–1 mM DTT, and reannealed.

Probes for nicking assays were prepared as described previously (4) by 3′ end labeling a 95-bp *Eco*RI fragment from pL1-2TC which contains the minimal left-end origin. The probe for gel mobility shift and UV cross-linking assays was prepared by end labeling a double-stranded oligonucleotide (PIFwt; referred to as ATF oligo in reference 4) covering the region of the minimal replication origin which is not protected by NS1 from DNase digestion. This oligonucleotide has the sequence 5′-[GATC]TTCACGTAAGTGACGTGATGA-3′, where the [GATC] sequence is a four-base 5′ overhang present on each strand to facilitate labeling. The duplex oligonucleotide SCRAM, 5′-[GATC]TAGAGAGTCGAT GTATCTGCA-3′, used as a nonspecific competitor in various assays, is the same length and has the same base composition as PIFwt but has the MVM sequence scrambled.

DNase I protection analysis. PIF fractions were preincubated for 10 min at room temperature in buffer B (20 mM HEPES-KOH [pH 7.8], 100 mM NaCl, 0.5 mM EDTA, 2 mM $MgCl_2$, 1 mM DTT, 1 μg of double-stranded SCRAM oligonucleotide, 2 μg of sonicated salmon sperm DNA, 10% glycerol) in a total volume of 30 μl. DNA fragments (approximately 4×10^4 cpm) were added, and reaction mixtures were incubated for 45 min at room temperature before addition of various amounts of DNase I (0.5 to 1 U per reaction) and further incubation for 1 min at room temperature. Digestion was terminated with 0.3 ml of 10 mM Tris-HCl (pH 8.0)–10 mM EDTA–0.5% sodium dodecyl sulfate (SDS) and 300 μg of proteinase K, followed by incubation at 50°C for 45 min. Samples were extracted with phenol-chloroform, ethanol precipitated, and analyzed by electrophoresis through 6% denaturing acrylamide gels. Control samples of all DNA probes used for protection assays were also chemically cleaved at G residues by the procedure of Maxam and Gilbert (18) and electrophoresed as markers in adjacent lanes to allow precise sequence alignment.

Cu-OP protection analysis. Cu-OP protection analyses were performed essentially as described by Kuwabara and Sigman (16). Briefly, 3′-end-labeled probe (10^5 cpm) was incubated for 30 min at room temperature in 25 μl of buffer B in the presence of sufficient PIF to bind approximately 50% of the probe. Bound probe and free probe were separated by electrophoresis as described below, and the gel was incubated in a mixture of 0.045 mM cuprous sulfate, 0.2 mM 1,10-phenanthroline, and 5.8 mM mercaptopropionic acid for 10 min. The reaction was quenched by addition of 2.8 mM 2,9-dimethyl-1,10-phenanthroline, and the wet gel was exposed for autoradiography. Free probe and shifted complexes were localized, excised, and eluted overnight in 0.4 ml of 10 mM Tris-HCl–1 mM EDTA–0.5 M NaCl–0.1% SDS. Eluted DNAs were ethanol precipitated, repurified by using GeneClean (Bio 101, Vista, Calif.), and analyzed by electrophoresis through 6% denaturing acrylamide gels.

Methylation and carboxy-ethylation interference assays. End-labeled probe (10^5 cpm), modified by either DMS or DEPC treatment as described earlier, was incubated with PIF and subjected to electrophoresis as described for Cu-OP protection analyses. Wet gels were exposed for autoradiography, and free and bound probes were localized, excised, eluted, and ethanol precipitated as described above. To remove contaminating acrylamide, the precipitated probes were further purified by resuspension in 200 μl of 0.3 M sodium acetate–1 mM EDTA and reprecipitated by adding 30 μl of 1% cetyltrimethyl ammonium bromide. Samples were resuspended in 100 μl of 1 M piperidine and incubated for 30 min at 90°C, and the piperidine was removed by multiple rounds of lyophilization. Samples were analyzed by electrophoresis through 6% denaturing acrylamide gels.

Gel mobility shift and UV cross-linking assays. DNA binding reactions and UV cross-linking experiments were carried out as described previously (4). The sequences of the double-stranded oligonucleotides used as competitors are listed in the relevant figures.

Nicking assays. Samples of purified PIF were added to a reaction mixture containing 100 ng of purified recombinant NS1 in 20 mM HEPES-KOH–100 mM NaCl–0.5 mM $MgCl_2$–10% glycerol–0.5 mM DTT–0.05% Nonidet P-40–1 μg of SCRAM oligonucleotide–50,000 cpm of ^{32}P-labeled minimal origin DNA–2 mM ATP in a total volume of 25 μl. Reaction mixtures were incubated at 37°C for 45 min; 25 μl of 10 mM Tris-HCl–1 mM EDTA–100 mM NaCl–20% glycerol containing 1% SDS was added, and the incubation continued at 60°C for 30 min. Samples were analyzed by electrophoresis through 6% native acrylamide gels.

Size estimation of protein-DNA complexes by native gel electrophoresis. The molecular weight of the PIF-DNA complex was estimated indirectly by electrophoresis through native acrylamide gels essentially as described by Mueller et al. (19). Briefly, the migration of protein-DNA complexes generated in gel mobility shift assays was measured relative to the tracking dye after electrophoresis through 5, 6, 7, and 8% acrylamide gels. After electrophoresis, gels were fixed and stained in 10% acetic acid–25% methanol–0.0125% Coomassie brilliant blue to visualize the molecular weight marker standards before being dried and exposed for autoradiography. The relative migrations were plotted against the acrylamide concentration of the gels, and the slopes of the plots were determined by linear regression. The slope of the curve is independent of the charge and dependent only on the size and shape of the protein-DNA complex. A standard curve relating size to relative migration was obtained by plotting, against their molecular masses, the slopes for nondenatured standard proteins (α-lactalbumin [14.2 kDa], carbonic anhydrase [29 kDa], chicken albumin [45 kDa], bovine serum albumin [66 kDa], bovine serum albumin dimer [132 kDa], urease trimer [272 kDa], and urease hexamer [545 kDa]) from a MW-ND-500 kit (Sigma, St. Louis, Mo.) run on the same gels.

RESULTS

A factor purified from HeLa cell nuclear extracts can activate NS1 to nick the minimal left-end replication origin. We used a gel mobility shift assay to monitor the isolation of an

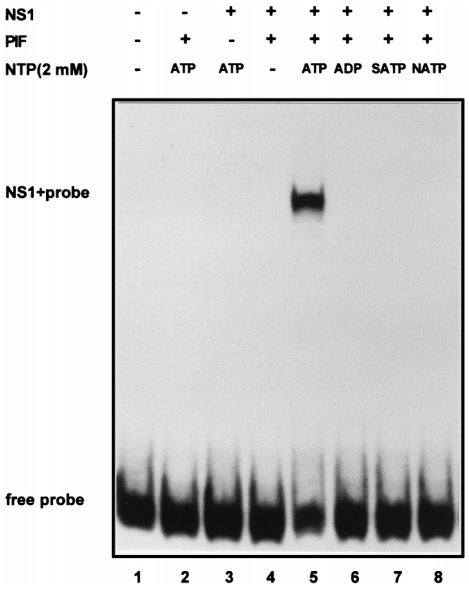

FIG. 2. NS1 requires PIF and ATP hydrolysis to nick the minimal left-end origin. The minimal left-end origin sequence was incubated in the presence or absence of NS1, HeLa PIF, and a variety of different nucleotide energy sources, as indicated at the top. Reactions were terminated by heating to 60°C in 0.5% SDS, and the products were electrophoresed through 6% native acrylamide gels. NS1+probe indicates the position of the probe with NS1 covalently attached at the nick site.

origin-binding factor from HeLa cell nuclei, using a purification protocol which was similar to, but substantially more stringent than, that used previously to isolate PIF fractions from leached cytoplasmic extracts of 293 cells (4). To determine whether this HeLa fraction was also able to function as a cofactor for NS1-mediated nicking and covalently attachment to the left-end origin, we incubated it in the presence of purified NS1, ATP, and a ^{32}P 3′-end-labeled DNA fragment containing the minimal active origin and analyzed the reaction products on native polyacrylamide gels. When HeLa PIF and purified NS1 were added on their own, they were not able to modify the substrate (Fig. 2, lanes 2 and 3, respectively), but when PIF and NS1 were added together, the origin was nicked, and a low-mobility DNA complex was formed by the covalent attachment of NS1 to the new 5′ end generated at the nick site (Fig. 2, lane 5). Boiling the samples in 0.5% SDS to melt all double-stranded DNA prior to electrophoresis resulted in complexes with increased mobility, indicating that the initial product was double stranded (data not shown). Thus, in the left-end nicking assay, PIF fractions isolated from HeLa nuclei behave in the same way as the previously described 293 cell factor.

NS1 binds DNA in a site-specific manner, but this reaction is entirely dependent on its ability to bind, but not necessarily to hydrolyze, ATP (3, 5). To determine whether ATP was similarly required in the nicking assay just to allow NS1 and/or PIF to bind the DNA or whether nicking required energy from nucleotide hydrolysis, a panel of nucleotides was substituted for ATP in the reaction mixture. As expected, in the absence of ATP or in the presence of ADP, no nicking occurred (Fig. 2, lanes 4 and 6). However, γS-ATP and N-ATP, two nonhydrolyzable analogs of ATP which allow NS1 to bind DNA, failed to support nicking (Fig. 2, lanes 7 and 8), indicating that strand scission requires both ATP binding and hydrolysis.

The PIF footprint adjoins that of NS1 in the left-end origin. We first mapped the PIF DNA binding sequence in the left-end origin by DNase I protection analysis (Fig. 3). This revealed a protected region of some 21 bp extending over both strands of the minimal origin, starting just beyond the bubble sequence and extending through the consensus ATF site to a position three bases into the pCRII vector sequence. To minimize any effects on the footprint caused by minor contami-

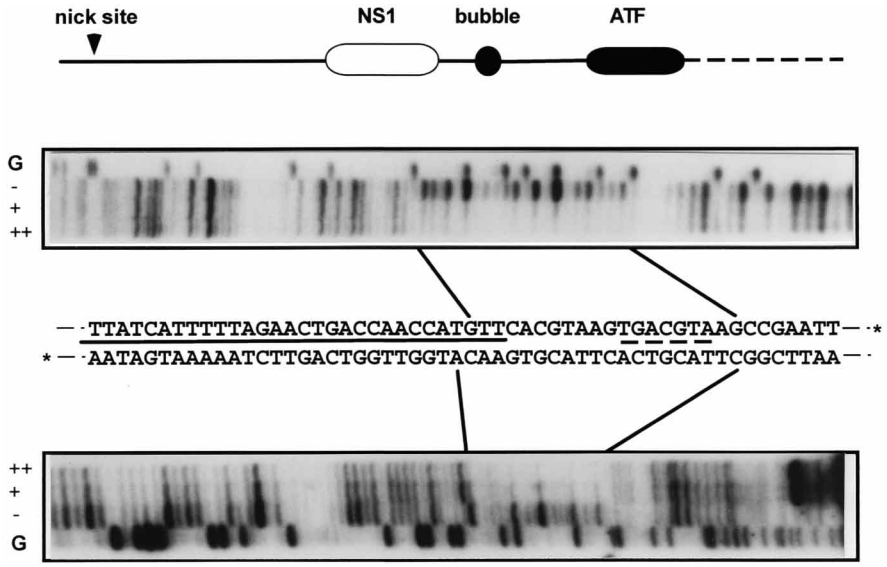

FIG. 3. DNase I protection analysis of the left-end origin in the presence of PIF. A diagram of the major elements in the cloned minimal left-end origin is shown; solid lines indicate MVM-derived sequence, and the dashed line indicates sequence from the pCRII vector. Abbreviations are detailed in the legend to Fig. 1. Oriented below the diagram are autoradiographs of sequencing gels showing the DNase I digest products obtained from the ^{32}P-labeled upper or lower strands of the minimal origin sequence shown in the middle. Asterisks at the 3′ ends of the DNA sequence denote the alternate positions of the ^{32}P labels used to identify the two strands. The solid line between the two strands of sequence indicates the region protected from DNase I digestion by NS1, and the broken line indicates the consensus ATF binding site. Lanes marked G contain the products of a G-specific chemical cleavage reaction run on each substrate. DNase I products obtained in the absence of PIF (−) or in the presence of 4 μl (+) or 8 μl (++) of the PIF fraction are indicated. Boundaries of the sequences protected by PIF are marked by solid lines.

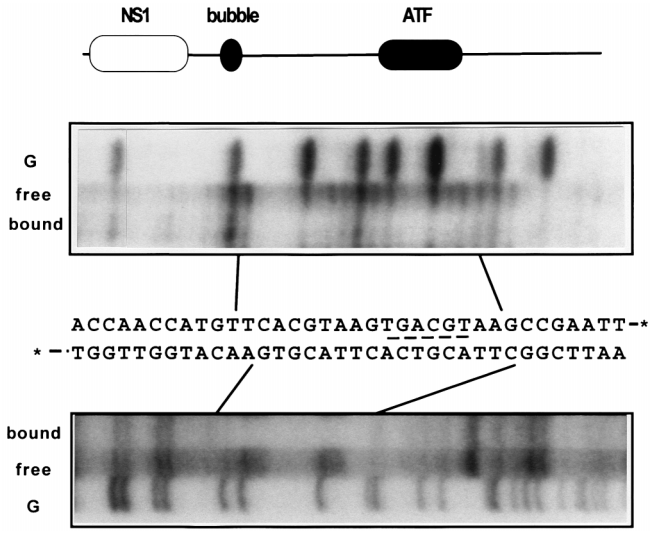

FIG. 4. Cu-OP protection analysis of the PIF–left-end origin complex. A diagram of major elements in the left-end origin is positioned above autoradiographs of sequencing gels showing the Cu-OP cleavage products obtained from ^{32}P-labeled upper or lower strands of the minimal origin sequence detailed in the middle. Asterisks at the 3′ ends of the DNA sequence denote the alternate positions of the ^{32}P labels used to identify the two strands, and the broken line identifies the consensus ATF binding site. Lanes marked G contain the products of a G-specific cleavage reaction run on each substrate, and lanes marked free contain the Cu-OP cleavage products of free probe; lanes marked bound contain the cleavage products the gel purified PIF-origin complex. Boundaries of the sequences protected by PIF are indicated by solid lines.

nants in the PIF fraction, we next separated PIF-DNA complexes by gel electrophoresis and analyzed them directly by limited copper Cu-OP degradation. Cu-OP cleaves the ribose backbone of unprotected DNA chemically, and because it is small, it can provide a more precise map of the borders of the complex than DNase I. As seen in Fig. 4, PIF protected a region of 19 bp from Cu-OP, starting immediately within the bubble sequence and extending again through the ATF site. This protected region was interrupted by a single Cu-OP-sensitive site (G) in the middle of the upper strand, indicating that at this position the minor groove was not in continuous close contact with protein.

Thus, PIF protects a region of the left-end origin which includes the consensus ATF site and projects into the bubble sequence. This observation suggests that, while PIF and NS1 can likely bind the origin simultaneously, they would be very closely juxtaposed in such a ternary complex.

ACGT motifs are essential elements in the PIF recognition sequence. To identify the nucleotides directly involved in the PIF-DNA interaction, we analyzed the effects of methylating or carboxy-ethylating the N7 position of purines in the DNA, using DMS or DEPC, respectively. DMS modifies guanines predominantly, while DEPC favors adenines but will react with both purines. In general, any modified bases that normally make contact with a bound protein will be underrepresented in the complexed probe relative to the free probe. PIF-DNA complexes were formed with the modified probes and separated from free probe by gel electrophoresis. Free and bound probes were then separately eluted, hydrolyzed at the sites of modification by using piperidine, and analyzed by denaturing polyacrylamide gel electrophoresis (PAGE). Figure 5A shows the result of a methylation interference analysis. In the upper strand, three guanines, at positions 29, 35, and 38 in the minimal origin, were clearly underrepresented in the bound probe, while G33 was very slightly reduced. In the lower strand, modified guanines at positions 28 and 37 were clearly underrepresented in the complex.

Carboxy-ethylation interference analysis (Fig. 5B) showed that modification of A27, A36, and G38 in the upper strand affected their inclusion in PIF-DNA complexes, while modification of A31 and A32 did not. Although very weak interference can be observed, residues G29 and G35, previously identified in the methylation interference assay, are not strongly implicated by this interference procedure, presumably because

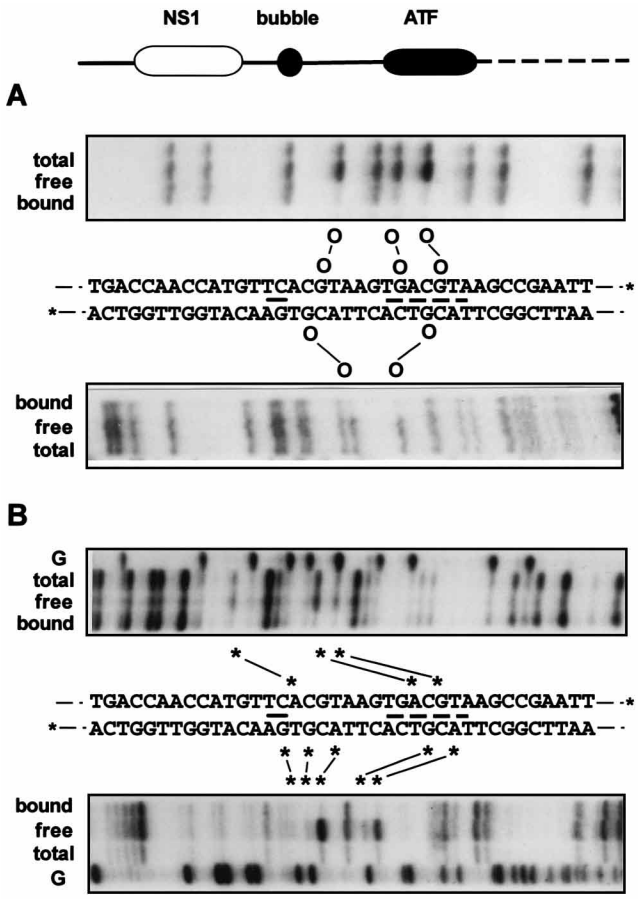

FIG. 5. Chemical modification interference analysis of PIF binding to the left-end origin. (A) Methylation interference analysis. A diagram of the major elements in the cloned minimal left-end origin is shown; solid lines indicate MVM sequence, and the dashed line indicates sequence from the pCRII vector. Autoradiographs of sequencing gels show methylation interference patterns obtained from ^{32}P-labeled upper or lower strands of the DNA sequence detailed in the middle. Asterisks at the 3′ ends of this sequence denote the alternate positions of the ^{32}P labels used to identify the two strands, a solid line indicates the position of the bubble sequence, and a broken line identifies the consensus ATF binding site. Lanes marked total contain piperidine cleavage products of the total methylated DNA sample used in the binding assay, lanes marked free contain cleaved probe which remained free after the binding assay, and lanes marked bound contain the cleavage products of methylated DNA from the PIF-DNA complex. Methylated residues which impair PIF binding are indicated by circles connected by solid lines. (B) Carboxy-ethylation interference analysis. Autoradiographs of sequencing gels show carboxy-ethylation interference patterns obtained from ^{32}P-labeled upper or lower strands of the DNA sequence detailed in the middle. Lanes marked total contain piperidine cleavage products of the total carboxy-ethylated DNA sample used in the binding assay, lanes marked free contain cleaved probe which remained free after the binding assay, and lanes marked bound contain the cleavage products of carboxy-ethylated DNA from the PIF-DNA complex. Carboxy-ethylated residues which impair PIF binding are indicated by asterisks connected by solid lines. The lanes marked G indicate the products of G-specific cleavage reactions run on each substrate.

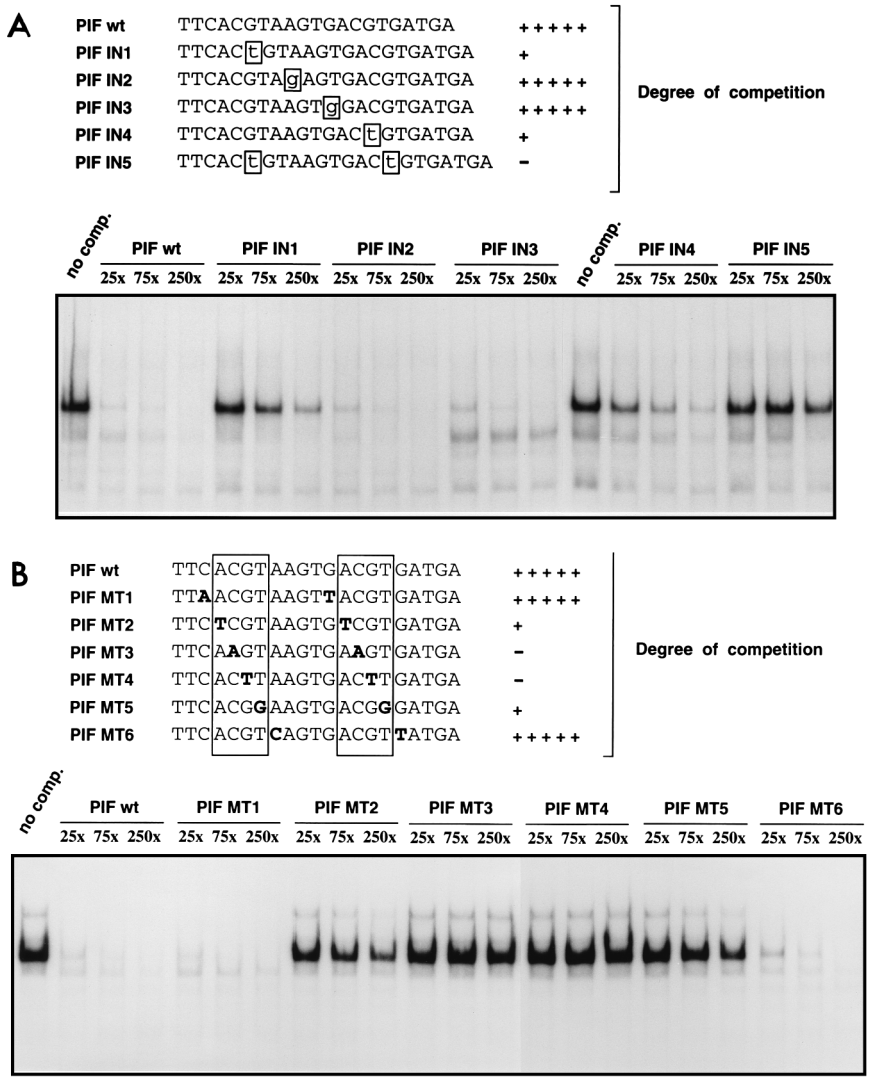

FIG. 6. Mutational analysis of the PIF-DNA interaction using competitive gel shift assays. (A) Autoradiographs of gel mobility shift assays showing the PIF-probe complex positioned in the center of each lane. The ability of PIF to bind to the ^{32}P-labeled PIFwt duplex oligonucleotide was analyzed alone (no competitor [comp.]) or in the presence of a 25-, 75-, or 250-fold molar excess of a panel of unlabeled, duplex oligonucleotides carrying insertions at selected positions, as indicated above the lanes. Sequences of the unlabeled competitor oligonucleotides are shown at the top, with the positions of insertions boxed. ACGT core sequences identified by interference assays are underlined. (B) The ability of PIF to bind to the ^{32}P-labeled PIFwt duplex oligonucleotide was analyzed alone (no competitor [comp.]) or in the presence of a 25-, 75-, or 250-fold molar excess of a panel of duplex oligonucleotides carrying transversions at selected positions, as indicated above the lanes. Sequences of unlabeled competitor oligonucleotides are shown at the top, with ACGT core sequences boxed and the transversions printed in boldface.

guanines are not modified so efficiently by DEPC. In the lower strand, modification of purines A30, A39, G26, G28, and G37 clearly interfered with PIF binding. Of these, guanines at positions 28 and 37 were also implicated by methylation interference analysis, while G26 was sensitive to carboxy-ethylation but not methylation. Since DEPC modification introduces a larger adduct than DMS, the differential sensitivity of G26 suggests that this residue does not itself make direct contact with PIF but lies immediately adjacent to a residue which does.

In summary, DNase I and Cu-OP protection analysis suggest that the DNA sequence which interacts with PIF starts at adenine A27 in the minimal origin and terminates at T39, while methylation and carboxy-ethylation interference analyses implicate two spaced ACGT motifs which appear to be in close contact with PIF in the major groove. Since these ACGT motifs are also protected from degradation with Cu-OP, PIF presumably also makes contacts in the minor groove at these positions.

The two ACGT motifs create a single PIF binding site rather than two separate sites. To determine whether the ACGT sequences function as two separate PIF binding sites or are part of a single site, we designed a series of double-stranded oligonucleotides carrying single-base insertions both within and between the two ACGT motifs and assessed their ability to compete with a duplex ^{32}P-labeled wild-type sequence (PIFwt) for PIF in a gel mobility shift assay. As seen in Fig. 6A, a single-base insertion between the C and G of the first ACGT motif (mutant IN1) rendered this oligonucleotide approximately 1,000-fold less efficient than the unlabeled PIFwt sequence at competing with the ^{32}P-labeled probe. However, insertions one or four nucleotides beyond this first ACGT motif, at positions 31 and 34, did not affect the inhibitory

activity of these oligonucleotides (IN2 and IN3) compared to the wild type. The inhibition pattern observed for oligonucleotide IN4, which has an insertion between the C and G of the second ACGT motif, resembled that of IN1 in being some 1,000-fold decreased compared to the wild-type level, while insertions at both of these positions in the same oligonucleotide (IN5) totally abolished inhibition of PIF binding to the ^{32}P-labeled PIFwt oligonucleotide. These results show that there is cooperation between the two ACGT motifs and that both must be present to allow PIF to bind. Thus, we conclude that the two ACGT sequences are part of a single PIF binding site and not two separate sites. However, single-base insertions between these two motifs gave oligonucleotides which could competed for PIF as efficiently as the wild-type sequence, indicating that the spacing between these core motifs must be somewhat flexible.

To evaluate the relative importance of nucleotides at different positions in the ACGT core motifs, we performed similar competitive mobility shift assays using a series of unlabeled double-stranded competitor oligonucleotides which had transversions at the same nucleotide position with respect to both motifs. As shown in Fig. 6B, oligonucleotides MT1 and MT6, which have transversions at nucleotides C26 plus G35 and A31 plus G40, respectively, flanking the ACGT core sequences, competed as efficiently as the unlabeled PIFwt sequence. In contrast, transversion of nucleotides at the first and fourth positions in the ACGT core (at A27 plus A36 in MT2 and T30 plus T39 in MT5) resulted in a drastic decrease in the ability of these oligonucleotides to inhibit PIF binding to the ^{32}P-labeled PIFwt probe, and transversions at the second and third positions in the core sequence (at C28 plus C37 in MT3 and G29 plus G38 in MT4) eliminated all competition. This result indicates that all four positions in the ACGT motif are important but that the central positions contribute relatively more to the PIF-DNA interaction than the two flanking positions.

Oligonucleotides which bind PIF can competitively inhibit the NS1-mediated nicking reaction. Although the PIF fractions used in this study were relatively highly purified, they still contained multiple protein species. Thus, the DNA binding activity evaluated in the preceding mobility shift assays might, theoretically, not be mediated by the same factor that activates NS1 to nick the left-end origin. To evaluate this possibility, we assayed the inhibitory potential of mutant double-stranded oligonucleotides, selected from those shown in Fig. 6A, in the PIF- and NS1-dependent, origin-based nicking assay represented in Fig. 2. In the presence of a constant amount of NS1, addition of increasing amounts of PIF had little effect on the observed nicking activity, suggesting that PIF was present in saturating amounts in these assays (Fig. 7, lanes 3 to 5). Nonetheless, under these conditions, addition of 50 ng of the PIFwt oligonucleotide strongly inhibited the formation of the slowly migrating NS1-origin complex (Fig. 7; compare lanes 5 and 7), but oligonucleotides IN1, IN4, and IN5, which do not compete efficiently in the mobility shift assay (as seen in Fig. 6A), failed to inhibit the nicking reaction. In contrast, oligonucleotides IN2 and IN3, which compete as well as the wild-type sequence in the mobility shift assay, inhibited the nicking reaction as efficiently as unlabeled PIFwt. Thus, the inhibitory potentials of the mutated oligonucleotides were identical in the specific DNA binding assay and the functional nicking assay, indicating that the same factor was responsible for both. Moreover, this analysis ruled out the possibility that the contaminating DNA binding activity which was responsible for a minor band seen migrating slightly faster than PIF in the gel mobility shift assays (Fig. 6A) might contribute to the nicking reaction, since this binding activity was not inhibited by, for instance, oligonucleotide IN3, which was a potent inhibitor of the nicking assay.

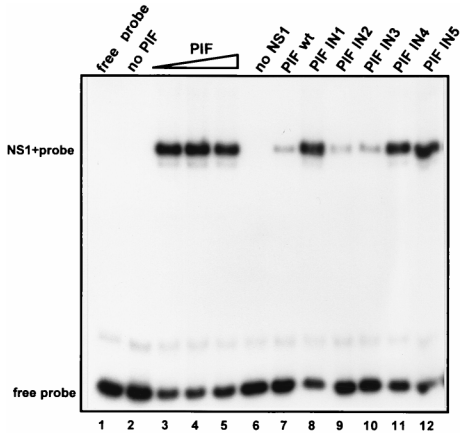

FIG. 7. Mutational analysis of the PIF-DNA interaction, using competitive NS1-mediated origin nicking assays. The autoradiograph of a nicking assay shows unreacted (free) probe and covalent NS1-origin complexes (NS1+probe) generated when the left-end origin was incubated in the presence or absence of NS1, HeLa PIF, ATP, and 50 ng of a variety of double-stranded, competitor oligonucleotides carrying insertions at selected positions as indicated at the top and shown in Fig. 6A. Reactions in lanes 1 and 2 lacked PIF, those in lanes 3 and 4 contained 1 and 2 μl of PIF, respectively, and all others contained 4 μl of PIF. Reactions in lanes 1 and 6 lacked NS1, and those in lanes 1 to 6 lacked competitor oligonucleotides.

PIF interacts with the binding site as a multimer. Mobility shift assay gels were UV irradiated to covalently cross-link the ^{32}P-labeled probe to the associated PIF proteins. Cross-linked complexes were then excised from the gel and analyzed by SDS-PAGE (Fig. 8A). PIF fractions purified from both HeLa cells and the murine L-cell derivative A9 gave a major cross-linked band of 105 to 110 kDa and a fainter, more diffuse band of around 91 kDa. Since the intact, duplex oligonucleotide had a molecular mass of 17.6 kDa, these two bands could represent the same polypeptide (of 85 to 100 kDa) bound either to a full-length duplex form of the probe, to give the 110-kDa band, or to a truncated or single-stranded form of the probe, to give the 91-kDa species. More likely, the two bands represent two different proteins, of 85 to 100 kDa and 70 to 85 kDa, which have been cross-linked to the intact probe with different efficiencies. Such a pattern might be expected if the two proteins formed a heterodimer in which only one of the partners made direct contact with the DNA. Interestingly, a diffuse band migrating at 220 kDa, a species which could represent two such PIF subunits cross-linked to the same probe, was observed at the top of the gel in both samples. These results suggest that PIF binds to the DNA as a dimer or higher-order multimer.

To estimate the native molecular weight of the PIF-probe complex, we analyzed its migration relative to the tracking dye on 5, 6, 7, and 8% native acrylamide gels (Fig. 8B). By plotting the relative migrations of the PIF complex and a series of marker proteins against the acrylamide concentration, we were able to obtain a relative measure for the size and shape of the protein-DNA complex, independent of charge, by linear regression (Fig. 8C). This analysis indicated that the mass of the PIF-DNA complex was about 385 kDa. When the molecular weight of the probe (17.6 kDa) was subtracted and the resulting mass was divided by the molecular mass of a single PIF polypeptide (~90 kDa as estimated by UV cross-linking), the result suggested that PIF interacts with the binding site as a tetramer.

However, we have shown previously that Ku antigen copu-

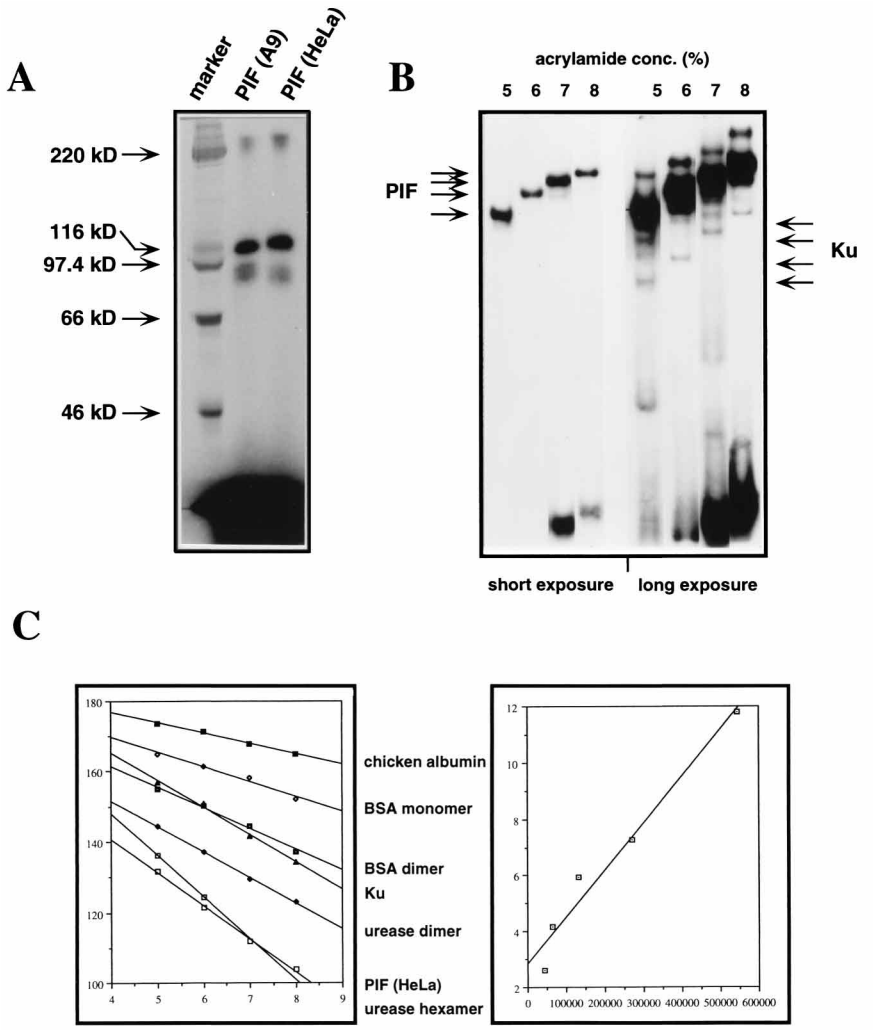

FIG. 8. Analysis of the molecular mass of denatured and native PIF-DNA complexes. (A) UV cross-linking analysis of proteins binding to the MVM PIFwt oligonucleotide. The PIF-DNA gel mobility shift complexes obtained by using Zn^{2+}-Sepharose fractions from murine fibroblasts or HeLa cells were UV irradiated to induce covalent cross-linking of the ^{32}P-labeled probe to the associated protein, then excised from the gel, and analyzed by discontinuous SDS-PAGE. ^{14}C-labeled molecular weight markers are shown in the first lane. (B) Relative migration of PIF-DNA complexes on gels of different acrylamide concentrations. The relative migration of PIF–^{32}P-labeled PIFwt oligonucleotide complexes was determined on 5, 6, 7, and 8% native acrylamide gels. The left half of the figure shows the migration of PIF complexes on such gels, while the right side shows these same strips after longer autoradiographic exposure, allowing the Ku antigen-DNA complexes to be visualized. (C) Size estimation of the PIF-DNA complex. The left panel shows the relative migration (R_f) of a number of standard molecular weight marker proteins and the PIF- and Ku-DNA complexes relative to the tracking dye in gels of different acrylamide concentrations, plotted with relative migration [100 log ($R_f \times 100$)] on the *x* axis and acrylamide concentration on the *y* axis. The right panel shows the standard curve obtained when the negative slopes of the standard molecular weight marker proteins (obtained by linear regression analysis) are plotted against their molecular weights. The molecular masses of the PIF-DNA complex (385 kDa) and the Ku-DNA complex (295 kDa) were then determined from their slopes, using this standard curve. BSA, bovine serum albumin.

rifies with PIF and gives a minor band in the gel shift analysis which migrates faster than PIF (4). Ku is known to be a heterodimer composed of 70- and 86-kDa subunits. Determining the molecular mass of the Ku-DNA complex in this way gave an estimate of 295 kDa, which, after deduction of the molecular mass of the probe, overestimates the size of the Ku heterodimer by 121 kDa. Since the standard curve for this assay is generated by using globular monomeric, dimeric, and trimeric standard proteins, this disparity indicates that the Ku-probe complex assumes a nonglobular configuration which reduces its relative mobility. Whether the size of the PIF-probe complex is similarly overestimated is uncertain, but Ku-DNA and PIF-DNA complexes are likely to differ in shape because PIF binds in the center of the probe while Ku binds preferentially to the ends. However, if we base our estimate on the mobility of the Ku complex, rather than the mobility of the standard proteins, the result would suggest that PIF binds DNA as a dimer.

DISCUSSION

PIF is a novel human site-specific DNA-binding protein which we have shown is absolutely required to allow the viral NS1 polypeptide to initiate DNA replication by introducing a nick into the viral left-end origin (4). PIF appears to activate the required endonuclease function by binding to a DNA sequence in the origin, detailed in this report, which is located immediately next to the NS1 complex, although the mechanism of activation remains to be determined. Such dependence on PIF is remarkable because parvovirus replication closely re-

sembles prokaryotic rolling circle mechanisms which do not require exogenous host factors to activate their replicon-encoded initiators (15). Likewise, the initiator protein of the closely related parvovirus, adeno-associated virus, acts independently of any known cellular factors (14).

MVM has a highly compact, 5-kb genome, so that much of its DNA serves multiple functions. In this report, we show that the sequence in the left-end origin which is recognized by PIF overlaps a consensus ATF binding site implicated in transcriptional control of the viral P4 promoter (22). However, while ATFs bind to a single contiguous DNA sequence, PIF recognizes an ACGT motif within the ATF consensus in conjunction with a second ACGT motif located nearer the NS1 binding site. Modification of the PIF binding site by DMS and DEPC allowed us to identify contact points within the major groove of both ACGT motifs, and transversions introduced at each position within these motifs showed that the central CpG doublet contributed significantly more to the affinity with which PIF bound DNA than did the flanking A's and T's. This strict requirement for the CpG doublet probably explains why the alternating polymer poly(dI-dC)-poly(dI-dC) is able to inhibit PIF binding at high molar excess (4).

NS1 protects most of the minimal left-end origin from nuclease digestion, as shown in Fig. 1, but this protection stops abruptly over the bubble dinucleotide sequence, while the region protected by PIF starts at the bubble sequence and extends through the rest of the minimal origin. Thus, DNA sequences covered by each of the two complexes are tightly juxtaposed across the bubble dinucleotide but do not substantially overlap. This finding suggests that PIF and NS1 can occupy the origin simultaneously, but whether they also interact in solution remains to be determined. The bubble sequence is known to be a critical spacer element, since the insertion of a single extra base at this position, as in the GAA arm of the dimer bridge, totally inactivates origin function (10), suggesting that NS1 and PIF might have to be in exact register in order to interact appropriately. Surprisingly, however, we show here that insertion of an extra base between the two ACGT motifs in the PIF binding site has no effect on its ability to bind. We have subsequently found that such an insertion also fails to disrupt PIF's ability to cooperate with NS1 to activate the origin (2a), and we are currently exploring this aspect of the binding site in greater detail, because the allowable spacing between the ACGT motifs does appear to be exceptionally flexible. Thus, the critical spacing determined by the bubble dinucleotide is a 5-bp sequence between the $(ACCA)_2$ motif of the NS1 binding site and the position of the proximal PIF-DNA contact point, but the way in which the PIF complex is then bound along the DNA helix seems of secondary importance.

Although spaced five nucleotides apart in the left-end origin, the two ACGT motifs create a single PIF binding site. Mutational analysis of replication initiation from this origin had previously shown that the guanine in the distal ACGT sequence (i.e., in the ATF binding site) was required for efficient replication initiation, even though the NS1-proximal ACGT motif in this mutated origin was intact (10). Moreover, an oligonucleotide carrying a single ACGT sequence, as part of a consensus ATF binding site, was shown to be less efficient than poly(dI-dC)-poly(dI-dC) in competing for PIF (4). Since single ACGT motifs form the core of a number of transcription factor binding sites commonly found in the control regions of mammalian genes, this absolute requirement for two adjacent sites presumably prevents PIF from competing with many of the basic leucine zipper transcription factors for these loci. In the MVM origin, the distance between equivalent nucleotides in the two motifs is 10 nucleotides, or one helical turn, which might suggest that the two sequences have to face the same side of the DNA helix, but as stated above, the spacing requirement does appear exceptionally flexible. Searching available databases, such as TransFac (26), for DNA-binding proteins which recognize repeated ACGT sequences, even with variable numbers of spacing nucleotides and even allowing mismatches in the A or T positions, has so far failed to reveal any previously described candidates for PIF.

The internal arrangement of the PIF binding site, with its flexibly spaced, reiterated motif, suggests that PIF probably interacts with the site as a dimer (or multimer) rather than as a monomer. Although the PIF fractions used in this study were more highly enriched than those described previously (4), they were still fairly complex. Thus, to determine the denatured mass of the human and murine proteins responsible for this DNA-binding activity, we resorted to cross-linking the complex by using UV irradiation. In both species, this analysis identified a pair of proteins of 90 to 100 kDa and 70 to 85 kDa, respectively, which is in close agreement with our current purification data (2a). We have recently purified the active factor from HeLa nuclei close to homogeneity by using a protocol which involves repeated DNA affinity chromatography. This procedure ultimately results in the isolation of two polypeptides, present in equimolar amounts, which give distinct tryptic peptide maps without any obviously coinciding peaks, strongly suggesting that they are two distinct gene products. Comparison of the native and denatured sizes of the PIF-DNA complex thus suggests that PIF binds the site as a heterodimer or tetramer, but if PIF does contain two distinct polypeptide chains, the tandem duplication within its binding site would tend to suggest that it forms a tetramer made of two heterodimers. Previous attempts to determine the native size of PIF by gel filtration indicated that PIF is a dimer when not complexed to DNA (4). This leaves open the possibility that it may form a tetramer cooperatively only when bound to DNA, as is observed for basic leucine zipper proteins when examined at normal cellular concentrations (13).

At present we do not know if PIF can actively regulate viral or cellular transcription, but clearly PIF will interact with a variety of cellular promoters which contain multiple cyclic AMP-responsive elements, since these binding sites usually contain the ACGT tetranucleotide. We know that PIF does bind to a region upstream of the human transferrin receptor (data not shown) which appears to be involved in the positive regulation of transcription (23, 24), and it could be that PIF plays a significant role in cellular transcription but has been overlooked previously due to the small size of the oligonucleotides used to detect such activities and the widespread use of poly(dI-dC)-poly(dI-dC) as a nonspecific competitor. Characterization of the PIF binding site presented in this report should allow us to design promoter constructs which can discriminate between the binding of PIF and ATF and thus allow us to reevaluate the influence of PIF and ATF in the transcriptional regulation of viral and cellular promoters.

ACKNOWLEDGMENTS

We thank Keith Derbyshire, John Mueller, and Sherman Weissman for helpful discussions and Ulla Toftegaard for excellent technical assistance.

This work was supported by Public Health Service grant AI26109 (to P.T.) from the National Institutes of Health. J.C. is supported by grants from the Danish Agricultural and Veterinary Research Council and Danish Center for Biotechnology.

REFERENCES

1. **Astell, C. R., M. B. Chow, and D. C. Ward.** 1985. Sequence analysis of the termini of virion and replicative forms of minute virus of mice DNA suggests a modified rolling hairpin model for autonomous parvovirus DNA replication. J. Virol. **54:**171–177.
2. **Baldauf, A. Q., K. Willwand, E. Mumtsidu, J. F. K. Nuesch, and J. Rommelaere.** 1997. Specific initiation of replication at the right-end telomere of the closed species of minute virus of mice replicative-form DNA. J. Virol. **71:**971–980.

2a. **Christensen, J., et al.** Unpublished data.

3. **Christensen, J., S. F. Cotmore, and P. Tattersall.** 1995. The minute virus of mice transcriptional activator protein, NS1, binds directly to the trans-activating region (*tar*) of the viral P38 promoter in a strictly ATP-dependent manner. J. Virol. **69:**5422–5430.
4. **Christensen, J., S. F. Cotmore, and P. Tattersall.** 1997. A novel cellular site-specific DNA-binding protein co-operates with the viral NS1 polypeptide to initiate parvoviral DNA replication. J. Virol. **71:**1405–1416.
5. **Cotmore, S. F., J. Christensen, J. P. F. Nuesch, and P. Tattersall.** 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.
6. **Cotmore, S. F., A. M. D'Abramo, L. F. Carbonell, J. Bratton, and P. Tattersall.** 1997. The NS2 polypeptide of parvovirus MVM is required for capsid assembly in murine cells. Virology **231:**267–280.
7. **Cotmore, S. F., J. P. Nuesch, and P. Tattersall.** 1992. In vitro excision and replication of 5′ telomeres of minute virus of mice DNA from cloned palindromic concatemer junctions. Virology **190:**365–377.
8. **Cotmore, S. F., J. P. Nuesch, and P. Tattersall.** 1993. Asymmetric resolution of a parvovirus palindrome in vitro. J. Virol. **67:**1579–1589.
9. **Cotmore, S. F., and P. Tattersall.** 1992. In vivo resolution of circular plasmids containing concatemer junction fragments from minute virus of mice DNA and their subsequent replication as linear molecules. J. Virol. **66:**420–431.
10. **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.
11. **Cotmore, S. F., and P. Tattersall.** 1996. Parvovirus DNA replication, p. 799–813. *In* M. DePamphilis (ed.), DNA replication in eukaryotic cells. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
12. **Dignam, J. D., P. L. Martin, B. S. Shastry, and R. G. Roeder.** 1983. Eukaryotic gene transcription with purified components. Methods Enzymol. **101:**582–598.
13. **Ellenberger, T. E., C. J. Brandl, K. Struhl, and S. C. Harrison.** 1992. The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex. Cell **71:**1223–1237.
14. **Im, D. S., and N. Muzyczka.** 1990. The AAV origin binding protein Rep68 is an ATP-dependent site-specific endonuclease with DNA helicase activity. Cell **61:**447–457.
15. **Kornberg, A., and T. A. Baker.** 1991. DNA replication, 2nd ed. W. H. Freeman, New York, N.Y.
16. **Kuwabara, M. D., and D. S. Sigman.** 1987. Footprinting DNA-protein complexes in situ following gel retardation assays using 1,10-phenanthroline-copper ion: Escherichia coli RNA polymerase-lac promoter complexes. Biochemistry **26:**7234–7238.
17. **Liu, Q., C. B. Yong, and C. R. Astell.** 1994. In vitro resolution of the dimer bridge of the minute virus of mice (MVM) genome supports the modified rolling hairpin model for MVM replication. Virology **201:**251–262.
18. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65:**499–560.
19. **Mueller, J. E., D. Smith, M. Bryk, and M. Belfort.** 1995. Intron-encoded endonuclease I-TevI binds as a monomer to effect sequential cleavage via conformational changes in the td homing site. EMBO J. **14:**5724–5735.
20. **Naeger, L. K., J. Cater, and D. J. Pintel.** 1990. The small nonstructural protein (NS2) of the parvovirus minute virus of mice is required for efficient DNA replication and infectious virus production in a cell-type-specific manner. J. Virol. **64:**6166–6175.
21. **Naeger, L. K., N. Salome, and D. J. Pintel.** 1993. NS2 is required for efficient translation of viral mRNA in minute virus of mice-infected murine cells. J. Virol. **67:**1034–1043.
22. **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.
23. **Roberts, M. R., Y. Han, A. Fienberg, L. Hunihan, and F. H. Ruddle.** 1994. A DNA-binding activity, TRAC, specific for the TRA element of the transferrin receptor gene copurifies with the Ku autoantigen. Proc. Natl. Acad. Sci. USA **91:**6354–6358.
24. **Roberts, M. R., W. K. Miskimins, and F. H. Ruddle.** 1989. Nuclear proteins TREF1 and TREF2 bind to the transcriptional control element of the transferrin receptor gene and appear to be associated as a heterodimer. Cell Regul. **1:**151–164.
25. **Sturm, R., T. Baumruker, R. J. Franza, and W. Herr.** 1987. A 100-kD HeLa cell octamer binding protein (OBP100) interacts differently with two separate octamer-related sequences within the SV40 enhancer. Genes Dev. **1:**1147–1160.
26. **TransFac.** 1997. TransFac database: release 3. http://transfac.gbf-braunschweig.de/TRANSFAC/index.html.