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 TATTERSALL1,3*

Departments of Laboratory Medicine1 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, Denmark2

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-ethylatation 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...
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 poly(dI-dC)-poly(dI-dC) (4).

In this report, we describe the purification of an identical protein from the 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. (27). Briefly, the cell culture was collected by centrifugation and resuspended in a buffer containing 0.32 M sucrose, 10 mM HEPES-KOH (pH 7.6), 0.1 mM EDTA, 1 mM dithiothreitol, 10 mM MgCl2, and 0.5 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 cell culture), which was then washed with buffer A containing 50 mM NaCl, and bound proteins were eluted in buffer A adjusted to 150 mM NaCl and 0.5 mM imidazole. Eluted fractions were dialyzed against buffer A containing 50 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 N2 before storage at -80°C.

Probes for gel mobility shift and UV cross-linking assays were prepared as described previously (4) by 39-labeled probes. Plasmid pL1-2TC (10), containing the minimal active left-end origin which 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 HindIII or XhoI and 39 end labeled on one strand with [32P]dATP or [32P]dCTP, respectively. The substrate was then digested with either XhoI or HindIII, respectively, and DNA fragments were purified from agarose gels.

For methylation interference analyses, DNA fragments were methylated by incubating 108 cpm of probe in 200 ml of 1 mM EDTA-50 mM sodium cacodylate buffer (pH 8.0) with 1 ml of dimethyl sulfate (DMS) for 5 min at room temperature. Reactions were terminated by adding 50 ml of 1 M sodium acetate containing 1 M β-mercaptoethanol, 60 μg of oxygen glycerol, and 750 μg of cold ethanol. For carboxy-ethylation interference assays, probes were prepared essentially as described by Sturm et al. (25). Briefly, 39-labeled DNA fragments were boiled, chilled, and diluted in 50 mM cacodylate (pH 8.0)-1 mM EDTA. Dithiothreitol (DTT) 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 MgCl2-50 mM NaCl-1 mM DTT, and reannealed.

Probes for nicking analyses were prepared as described previously (4) by 39 end labeling a 95-bp EcoRI fragment from plL1-2TC which contains the minimal left-end origin. The DNA for gel mobility shift assay and UV cross-linking shift was prepared by labeling a double-stranded oligonucleotide (P1Fw; 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 59-[GATC]TGACTGAAAGTAGCTGATCGA-39, where the [GATC] sequence is a four-base 59-overhang present on each strand to facilitate labeling. The duplex oligonucleotide SCRAM, 59-[GATC]TJAGAGGTAAGTAGCTGATCGA-39, used as a nonspecific competitor in various assays, is the same length and has the same base composition as P1Fw but has the MVM nucleotide numbers in the published sequence (1).

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 solid line indicates the position of ATF binding site. The position of DNA sequence indicates the position protected by NS1 from DNase I digestion, and the broken line indicates the position of the consensus ATF binding site. MVM 9t denotes the MVM nucleotide numbers in the published sequence (1).

**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
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 32P 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-
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 theirs 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-ethylyating 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 theirs 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-ethylyating 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
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 32P-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 32P-labeled probe. However, insertions one or four nucleotides beyond this first ACGT motif (mutant IN3) 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 32P-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 compete 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 dideoxynucleotides carrying insertions at selected positions as indicated at the top and shown in Fig. 6A. Reactions in lanes 1 and 2 lacked PIF, whereas 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.

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.

PIF interacts with the binding site as a multimer. Mobility shift assay gels were UV irradiated to covalently cross-link the 32P-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-
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-
semblles 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 nuclelease 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 tandem duplication within its binding site would tend 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) 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. Mutation 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


