Binding of Simian Virus 40 A Protein to DNA with Deletions at the Origin of Replication

BETSY A. LEWTON, ANGELO L. DeLUCIA, AND PETER TEGTMAYER*

Department of Microbiology, State University of New York, Stony Brook, New York 11794

Received 18 May 1983/Accepted 23 September 1983

Previous studies with wild-type simian virus 40 DNA have shown that the sequence 5'-GAGGC-3' directs the binding of A protein (T antigen). The functional origin of replication contains four recognition pentanucleotides each of which is separated by a single base pair and arranged as two pairs of direct repetitions that are inverted relative to each other. Analysis of A protein binding to a series of nonviable mutants progressively deleting these contact sites leads to the following conclusions: (i) stable binding of subunits of A protein to three origin pentanucleotides is not sufficient for the initiation of DNA replication, (ii) the stability of DNA binding depends on interactions between bound protein subunits, and (iii) a single pentanucleotide is sufficient to bind and orient a subunit of A protein.

The origin of replication in simian virus 40 (SV40) DNA has been mapped to a 60- to 65-base-pair (bp) region centered at the unique *Bgl*I restriction site (1, 3). Binding of the A protein (T antigen) to the origin regulates the initiation of each round of DNA replication (10, 12, 14, 16, 18). According to Hay and DePamphilis, RNA oligonucleotides at various positions on the early strand within the origin prime the synthesis of the first forward arm of DNA replication (7).

Studies with wild-type SV40 DNA show that a consensus family of recognition sequences directs the binding of A protein (2, 17). Different arrangements of the pentanucleotide determine different patterns of protein binding in three adjacent regions at the origin of replication. In region I on the early side of the origin, three contact sites arranged as direct repetitions lead to high-affinity binding but are not essential for DNA replication (2, 3, 10, 17). Region III on the late side of the origin contains six recognition sequences arranged as direct repetitions. Binding to these sites in vitro is stable only at low salt concentrations (2, 15).

Region II corresponds well to the functional origin of replication defined by the mapping of viable deletion mutants (3). It contains four ideal recognition pentanucleotides that are each separated by a bp and arranged as two pairs of direct repetitions inverted relative to each other. Base substitutions between pentanucleotides are associated with wild-type or even increased levels of DNA synthesis. In contrast, substitutions within pentanucleotides reduce, but do not eliminate, replication (11). Deletions of a part of a pentanucleotide appear to eliminate replication altogether (6, 10).

In the present study we analyze the binding of A protein to nonviable deletion mutants of region II to gain insight into the minimal binding requirements for DNA replication. Our findings also confirm the major features of a model for DNA binding established with wild-type DNA (15).

MATERIALS AND METHODS

Protein purification. The SV40 A protein was purified from productively infected CV-1 cells as previously described (13).

DNA. Wild-type as well as mutant DNAs 6-1 (*dl6*) and 6-17 (*dl9*), cloned in plasmid pMK16, were a gift from Y. Gluzman (5). The deletion mutant 295 (*dl18*), cloned in pBR322,

was given to us by R. Martin. The plasmid was grown in *Escherichia coli* HB101, and the DNA was prepared as described by Kahn et al. (8). For simplicity we refer to these mutants as *dl6*, *dl9*, and *dl18*, the numbers of which indicate the number of bp deleted from the origin of replication.

DMS footprinting. SV40 DNA was digested with *Nco*I and 3' end labeled with [³²P]deoxyribosylthymine triphosphate by use of avian myeloblastosis virus reverse transcriptase as previously described (18). The transcriptase was inactivated by heating at 70°C for 10 min, and the labeled DNA was then cut with *Hinf*I. The origin-containing fragment, radiolabeled on the 3' late end of the early strand, was purified by gel electrophoresis and was ethanol precipitated. After SV40 A protein was bound to the DNA for 1 h on ice, the guanine residues were methylated by the addition of 1 µl of 10 M dimethyl sulfate (DMS) for 10 min. The DMS reaction was stopped, and the DNA was rapidly processed by the Maxam and Gilbert DNA sequencing method specific for cleavage at guanine residues (9). The resulting fragments were run on an 8.3 M urea, 8% polyacrylamide gel electrophoresis system at a constant temperature of 50°C.

Footprinting of the late strand was accomplished by radiolabeling the 3' early end of the late strand of the origin-containing restriction fragment *Hinf*I-*Bst*NI as previously described (2).

DNase footprinting. The *Bst*NI-*Hinf*I origin-containing fragment of each DNA was radiolabeled on the 3' end of the early or the late strand by previously described methods (2). SV40 A protein was bound to the fragment, and then 0.0005 units of DNase was added. The reaction was stopped, and the sample was processed by the method of DeLucia et al. (2). Protected DNA was analyzed by polyacrylamide gel electrophoresis as described above.

RESULTS

DMS footprinting. Nucleotides in close proximity to bound protein were identified with the DMS protection assay. For convenience we will refer to these as contact sites. In these experiments, an origin-containing SV40 DNA fragment was radiolabeled on the 3' end of either the early or the late strand. The guanine residues were then treated with DMS in the presence or absence of A protein at two ionic strengths to determine the binding locations and affinities of the mutant DNAs. Because these experiments were de-

* Corresponding author.

FIG. 1. DMS footprints of the SV40 A protein bound to both strands of the origin regions of wild-type (WT) and mutant DNAs. Restriction fragments were exposed to DMS in the presence or absence of A protein in 0.005 M NaCl or 0.125 M NaCl and processed as described in the text. All guanine (G) residues of region II are placed on the same line where guanines occur at adjacent positions in the sequence, and small brackets numbered in an increasing order identify recognition pentanucleotides. A circled G represents guanines protected at low ionic strength, and a dotted G represents hypersensitive guanines. (A) *NcoI-HinI* origin fragment was labeled on the 3' late end of the early strand. (B) The *BstNI-HinI* origin fragment was labeled on the 3' early end of the late strand. ▲6, ▲9, and ▲18 are *dl6*, *dl9*, and *dl18*, respectively.

signed to compare the strength of binding with various origins. A protein concentrations were adjusted to protect 50 to 70% of wild-type origins from DMS. Excess concentrations of protein would have decreased the sensitivity of the comparative assay.

Protein contact sites on the early and late strands are shown in Fig. 1A and B, respectively. The large brackets to the left of the figure outline DNase protection regions I, II, and III to be described below. All of the mutant DNAs maintain the wild-type guanine sequence and DMS protection pattern in regions I and III. The results are described in detail and correlated with the DNase protection analyses below. However, it is evident that deletions of pentanucleotides in region II have little effect on binding to the remaining recognition sites at 0.005 M NaCl and have variable effects on binding at 0.125 M NaCl, depending on the extent and location of the deletion.

DNase footprinting. The limits of SV40 A protein bound to DNA were defined by the use of DNase footprinting. In this assay each molecule of DNA is cut once producing a ladder of fragments displayed by gel electrophoresis. When the DNA is first bound with protein and then cut, gaps appear in the ladder representing the region of DNA protected by T

antigen. Double-stranded restriction fragments containing the origin of replication were radiolabeled at the 3' end of the early or late strands (Fig. 2A and B). The figure shows DNase protection patterns obtained under low salt conditions. A protein protects wild-type DNA in the regions indicated by the brackets to the left of the gels (Fig. 2). Each region is flanked by DNase hypersensitive sites and each has a characteristic stability at different salt concentrations (15). Nevertheless, the boundary between regions I and II is somewhat arbitrary because A protein bound to region I sometimes protects DNA in the early end of region II from DNase (15). As shown in Fig. 2, deletions in region II have little if any effect on binding of A protein to regions I and III. Protein protects all of the undeleted region II of mutants *dl6* and *dl9* but fails to protect the early end of region II of *dl18*. The hypersensitive sites flanking each end of region II are gradually lost with increasing deletion size. Thus, these sites must be determined, in part, by binding to region II.

The DNase protection appears to be more complete than DMS protection even though the assays were carried out under the same conditions. This difference is typical and can be explained in several ways. First, protection from the larger DNase molecule is indeed more efficient because of

FIG. 2. DNase footprints of the SV40 A protein bound to both strands of the origin regions in wild-type (WT) and mutant DNAs. Restriction fragments were cut with DNase in the presence or absence of A protein in 0.005 M NaCl and processed as described in the text. (A) The *Bst*NI-*Hinf*I origin fragment was labeled on the 3' late end of the early strand. (B) The same fragment was labeled on the 3' early end of the late strand. $\Delta 6$, $\Delta 9$, and $\Delta 18$ are *dl6*, *dl9*, and *dl18*, respectively.

the size of the probe. Second, protection by subunits of A protein bound to adjacent pentanucleotides provides domains of overlapping DNase protection.

DISCUSSION

To understand the process of initiation of DNA replication, we chose to focus on the binding of A protein to SV40 DNA deleted at the *Bgl*I restriction site in the origin. All the mutants are nonviable in permissive host cells. We used DMS footprinting to define close contacts between origin DNA and A protein and DNase footprinting to determine the broader limits of the DNA-protein interaction. The results from both studies are summarized and interpreted in Fig. 3.

In the case of wild-type DNA, A protein protects all circled guanines (Fig. 3) within the four pentanucleotides of region II and two of three guanines between the pentanucleotides. Weak protection of a nearby guanine is indicated by parentheses and hypersensitive guanines are dotted (Fig. 3). Protein also protects the entire region from DNase. The arrow below each pentanucleotide shows our interpretation of the DMS and DNase protection results (15). In our model, each pentanucleotide directs the binding of a monomer of A protein in an orientation that protects 20 bp of DNA at the 5'

end of the 5'-GAGGC-3' sequence and 10 bp at the 3' end (15).

Mutant *dl6* deletes site 2 completely without altering the single-bp spacing between the remaining sites. Protein binding protects all the remaining pentanucleotides from DMS and induces hypersensitivity of the outlying guanines of the mutant DNA in the same pattern as in the case of wild-type DNA. Protein also protects the entire undeleted portion of region II from DNase. Both DMS (Fig. 1) and DNase assays (data not shown) indicate that binding is only slightly sensitive to 0.125 M NaCl. These results are consistent with the predictions of our binding model (15). The loss of site 2 would decrease the number of protein subunits bound to region II but would not affect other interactions with pentanucleotides.

Mutant *dl9* deletes all of site 2 but only a part of site 3 so that the remaining pentanucleotides are separated by four rather than one bp. At low salt concentrations, protein protects sites 1 and 4 from DMS and the entire span of undeleted region II from DNase. It is interesting that this deletion also creates a new 5'-GAGGC-3' sequence that overlaps site 1 and is separated from site 4 by a single bp as shown by the brackets in Fig. 3. The DMS protection pattern

FIG. 3. Methylation and DNase protection results. The region II sequences of wild-type (WT) and mutant DNAs are shown and numbered by the method of Fiers et al. (4). The upper and lower strands are the early and late strands, respectively; the curved arrow above each sequence specifies the location of the deletion. Boxes identify recognition pentanucleotides numbered in an early to late direction. Dotted boxes indicate sites that are incompletely protected at 0.125 M NaCl. Circled guanines (G) indicate DMS protection, and the parentheses signify weak protection. Large and small dots above guanines indicate different degrees of hypersensitivity. The elongated arrows below the recognition sequences represent our interpretation of the DNase protection results. Each arrow corresponds to the DNase protection domain of a monomer of protein bound to a single pentanucleotide (15). The bracketed pentanucleotide site in mutant *dl9* represents a newly formed contact site that is apparently unused. $\Delta 6$, $\Delta 9$, and $\Delta 18$ are *dl6*, *dl9*, and *dl18*, respectively.

indicates that A protein prefers binding to site 1 rather than to the new site. Furthermore, binding to site 1 is more stable than binding to site 4 at 0.125 M NaCl. These results are consistent with our previous finding that binding to region II occurs in a sequential process in an early to late direction (15). We conclude that the higher affinity of binding to site 1 does not depend on adjacent recognition pentanucleotides in region II.

Mutant *d/l18* deletes all of sites 1, 2, and 3 so that site 4 is the only site remaining in region II; it occupies the original position of site 1 relative to region I but in the opposite orientation. A protein protects this single recognition sequence at low, but not at high, ionic strength. Thus, high-affinity binding to site 1 may depend not only on sequences or protein-protein interactions on the early side of region II, but also on the orientation of the pentanucleotide. Binding to site 4 in *d/l18* provides the first direct evidence that confirms our model for the mechanism of A protein binding. A single recognition sequence directs binding of a unit protein and results in asymmetric DNase protection. The protection of a guanine adjacent to site 4 raises a question as to whether the recognition sequence is a pentanucleotide or a hexanucleotide. We prefer the former alternative because a guanine in the sixth position is not consistently present and more importantly is not always used. For example, the guanine at the center of the region II palindrome is not protected by A protein (Fig. 1A and 3).

The deletions in region II change not only the number of contact sites but also the distance and rotational angles between remaining sites. Yet, the A protein can accommodate these changes and engage all available sites under low salt conditions. These results can be easily explained if the primary binding form is a monomer. Binding of preformed dimers or tetramers to altered DNA would require considerable flexibility of protein subunits. Studies using quantitative transmission electron microscopy (I. A. Mastrangelo, unpublished data) demonstrate that A protein monomers are frequent binding forms in region II, although dimer and tetramer forms are also readily seen at higher protein concentrations. Even though monomers may be the primary binding form, protein-protein contacts would account for the higher affinity binding evident when recognition sequences correctly position monomers relative to one another.

The 27-bp palindrome in the origin of replication could have a number of additional functions as a consequence of protein binding. For example, binding to the palindrome could induce a cruciform structure, initiate melting in the adjacent AT-rich sequences, or alter conformation in another manner. Interestingly, one of the major initiation sites for RNA priming of DNA replication appears to occur within the third pentanucleotide in the protein-binding palindrome (7). The production of a primer could require either de novo synthesis or possibly processing of preexisting early transcripts by an RNase H-like activity at this site. Thus, deletions in this region could have multiple effects on the initiation of replication. Mutant *d/l6* deletes site 2 but maintains a single-bp space between the remaining sites so that the spacing and rotation of bound protein subunits would not be altered. This positioning effect is likely to account for the stable binding observed. Because this mutant is not viable, we conclude that binding of four monomers of

A protein in specific positions and orientations is a requirement for successful DNA replication. Additional possible functions of the palindrome in DNA replication, if any, remain to be demonstrated.

ACKNOWLEDGMENTS

This work was supported by Public Health Service Grants CA-18808 and CA-28146 from the National Cancer Institute.

We thank Y. Gluzman and R. Martin for their deletion mutants.

LITERATURE CITED

1. Danna, K. J., and D. Nathans. 1972. Bidirectional replication of simian virus 40 DNA. *Proc. Natl. Acad. Sci. U.S.A.* **69**:3097-3100.
2. DeLucia, A. L., B. A. Lewton, R. Tjian, and P. Tegtmeyer. 1983. Topography of simian virus 40 A protein-DNA complexes: arrangement of pentanucleotide interaction sites at the origin of replication. *J. Virol.* **46**:143-150.
3. DiMaio, D., and D. Nathans. 1980. Cold-sensitive regulatory mutants of simian virus 40. *J. Mol. Biol.* **140**:129-142.
4. Fiers, W., R. Contreras, G. Haegeman, R. Rogiers, A. Van de Voorde, H. Van Heuverswyn, J. Van Herreweghe, G. Volckaert, and M. Ysebaert. 1978. Complete nucleotide sequence of SV40 DNA. *Nature (London)* **273**:113-120.
5. Gluzman, Y., R. J. Frisque, and J. F. Sambrook. 1980. Origin-defective mutants of SV40. *Cold Spring Harbor Symp. Quant. Biol.* **44**:293-299.
6. Gluzman, Y., J. F. Sambrook, and R. J. Frisque. 1980. Expression of early genes of origin-defective mutants of Simian Virus 40. *Proc. Natl. Acad. Sci. U.S.A.* **77**:3898-3902.
7. Hay, R. T., and M. L. DePamphilis. 1982. Initiation of SV40 DNA replication in vivo: location and structure of 5' ends of DNA synthesized in the ori region. *Cell* **28**:767-779.
8. Kahn, M., R. Kolter, C. Thomas, D. Figurski, R. Meyer, E. Remaut, and D. R. Helinski. Plasmid cloning vehicles derived from plasmids ColE1, F, R6K and RK2. *Methods Enzymol.* **68**:268-280.
9. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**:499-560.
10. Myers, R. M., and R. Tjian. 1980. Construction and analysis of simian virus 40 origins defective in tumor antigen binding and DNA replication. *Proc. Natl. Acad. Sci. U.S.A.* **77**:6491-6495.
11. Shortle, D., and D. Nathans. 1979. Regulatory mutants of simian virus 40: constructed mutants with base substitutions at the origin of replication. *J. Mol. Biol.* **131**:801-817.
12. Tegtmeyer, P. 1972. Simian virus 40 deoxyribonucleic acid synthesis: the viral replicon. *J. Virol.* **10**:591-598.
13. Tegtmeyer, P., and B. Andersen. 1981. Partial purification of SV40 A protein and a related cellular protein from permissive cells. *Virology* **115**:67-74.
14. Tegtmeyer, P., B. Andersen, S. B. Shaw, and V. G. Wilson. 1981. Alternative interactions of the SV40 A protein with DNA. *Virology* **115**:75-87.
15. Tegtmeyer, P., B. A. Lewton, A. L. DeLucia, V. G. Wilson, and K. Ryder. 1983. Topography of simian virus 40 A protein-DNA complexes: arrangement of protein bound to the origin of replication. *J. Virol.* **46**:151-161.
16. Tenen, D. G., L. L. Haines, and D. M. Livingston. 1982. Binding of an analog of the simian virus 40 T antigen to wild-type and mutant viral replication origins. *J. Mol. Biol.* **157**:473-492.
17. Tjian, R. 1978. Protein-DNA interactions at the origin of simian virus 40 DNA replication. *Cold Spring Harbor Symp. Quant. Biol.* **43**:655-662.
18. Wilson, V. G., M. J. Tevethia, B. A. Lewton, and P. Tegtmeyer. 1982. DNA binding properties of simian virus 40 temperature-sensitive A proteins. *J. Virol.* **44**:458-466.