Transposase Contacts with Mu DNA Ends*

(Received for publication, April 18, 1991)

Alihua Zou‡, Pak C. Leung§, and Rasika M. Harshey‡||

From the ‡Department of Microbiology, University of Texas, Austin, Texas 78712

We have investigated the interaction between phage Mu transposase (A protein) and the ends (att sites) of Mu by chemical and nuclease protection and interference studies. These studies define a 24-base pair contact region at five of the six att sites (L1, L3 at att L and R1, R2, R3 at att R). Hydroxyl radical footprints show that the transposase binds to one face of the DNA helix and covers two consecutive major grooves. Binding specificity is achieved primarily through the major groove. Strong contacts are found with 3 guanines which are conserved at five of the sites. Two of these guanines are missing in the weakest binding site (L2) where 13 base pairs are mainly contacted. A pair of DNAase I hypersensitive sites, one on each strand, appear at the back of only one of the two contacted major grooves at most sites except at L2, and can be correlated with the degree of A protein-induced bend (Kuo, C.-F., Zou, A., Jayaram, M., Getzoff, E. D., and Harshey, R. M. (1991) EMBO J. 10, 1585–1591) at these sites. No contacts are observed for 4–5 base pairs in the vicinity of L1 and R1, where the A protein nicks DNA during transposition.

The phage Mu transposase or A protein (663 amino acids; Harshey et al., 1985; Priess et al., 1987) brings about Mu DNA transposition (for reviews see Symonds et al., 1987; Pato, 1989). On the basis of proteolysis, the A protein has been assigned three structural domains (Nakayama et al., 1987). Distinct regions in the N-terminal domain of A protein interact separately with specific DNA sequences at the two Mu ends (att L at the left and att R at the right; Nakayama et al., 1987), as well as an internal enhancer region (Leung et al., 1985; Priess et al., 1987) brings about Mu DNA breakage. Strong contacts are found with 3 guanines which are conserved at five of the sites. Two of these guanines are missing in the weakest binding site (L2) where 13 base pairs are mainly contacted. A pair of DNAase I hypersensitive sites, one on each strand, appear at the back of only one of the two contacted major grooves at most sites except at L2, and can be correlated with the degree of A protein-induced bend (Kuo, C.-F., Zou, A., Jayaram, M., Getzoff, E. D., and Harshey, R. M. (1991) EMBO J. 10, 1585–1591) at these sites. No contacts are observed for 4–5 base pairs in the vicinity of L1 and R1, where the A protein nicks DNA during transposition.

* This work was supported by a Grant GM33247 from the National Institutes of Health. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Present address: Dept. of Immunology, Research Institute of Scripps Clinic, La Jolla California 92037.
§ Present address: Winnipeg Rh Inst., University of Mannitoba, Winnipeg, Canada R2T 2N2.
|| Present address: Winnie Rh Inst., University of Mannitoba, Winnipeg, Canada R2T 2N2.

1 Corresponding author Tel.: 512-471-6881; Fax: 512-471-5546 or -7088.

R2, R3) at att R (Craigie et al., 1984). The protein binds as a monomer to each att site and causes the DNA to bend through 80–90° at most sites (Kuo et al., 1991). All six att sites are occupied by the protein within the synaptic complex but three of these sites (L2, L3, R3) are only loosely held (Kuo et al., 1991). A protein interactions with the remaining three sites (L1, R1, R2) result in a core arrangement that is sufficient to confer stability to the synaptic complex and that is fully proficient in Mu DNA strand transfer. Sites L2 and L3 must be required only in the initial events that lead to formation of the synaptic complex, while site R3 is not necessary for complex formation (see Kuo et al., 1991).

This study reports a comprehensive analysis of the A protein-DNA contacts made at each of the six att sites. We present details of the protein-DNA contacts for site L1 and summarize those for the remaining five sites. Data for site L3 will be discussed in more detail elsewhere.1 Our results provide information on how the A protein approaches and interacts with DNA. We discuss the possible importance of the phasing of the A protein binding sites at att R for generating a stable nucleoprotein complex.

**EXPERIMENTAL PROCEDURES**

Reagents—Ethylendinitrosourea, (NH4)2Fe(S04)2.6H2O, L-ascorbic acid, and thioare was from Sigma, and dimethyl sulfate (DMS)§ from Aldrich Chemical Co. Methidiumpropyl-EDTA (MPE) was a gift from P. Dervan, Caltech, Pasadena, CA. Phage T4 polynucleotide kinase, Klenow polymerase, and restriction endonucleases were from Boehringer Mannheim Biochemicals. Mu A protein was purified as previously described (Craigie and Mizuuchi, 1985).

Plasmide—pRA-L29 contains approximately 200 bp of Mu att L DNA cloned into a multiple cloning site vector (Nakayama et al., 1987); pZA-R1 contains 120 bp of att R DNA, made by cloning a filled-in 200-bp DdeI-TaqI fragment from pCN201 (Teplow et al., 1988) into the Smal site of pUC8.

DNA Manipulations—5′ ends were labeled using T4 polynucleotide kinase and [γ-32P]ATP, while 3′ ends were labeled using Klenow polymerase and [α-32P]dCTP as described by Maniatis et al. (1982). HindIII or BamHI were used to cut pRA-L29 DNA close to the L1 or L3 sites, respectively. EcoRI or BamHI were used to cut pZA-R1 DNA close to the R1 or R3 sites, respectively. Att L or att R DNA fragments labeled at one end were released by cutting with a second enzyme and purified on acrylamide gels as described by Maniatis et al. (1982). DNA sequencing reactions were performed as described by Maxam and Gilbert (1977).

Footprinting—DNAase I footprinting was done essentially as described by Craigie et al. (1984). Typically, DNA-protein complexes were made by incubating 2–4 pmol of end-labeled DNA fragments with 3–30 pmol of A protein in 20 μl of 25 mM Hepes (pH 7.5), 150 mM KCl, 10 mM MgCl2, 1 mM DTT, and 5% glycerol for 20 min at 30 °C. Samples were electrophoresed on 6% denaturing polyacrylamide gels. MPE-(Fe)2 footprinting was done essentially as described by 1 C. F. Kuo, T. Macke, A. Zou, R. M. Harshey, and E. D. Getzoff, manuscript in preparation.

The abbreviations used are: DMS, dimethyl sulfate; MPE, methidiumpropyl-EDTA; bp, base pair(s); Hepes, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; DTT, dithiotreitol.
Hertzberg and Dervan (1984). DNA-protein complexes were formed as described above except that the reactions contained in addition 100 μg/ml bovine serum albumin and 100 μg/ml sheared salmon sperm DNA. MPE at 1 mM was premixed with an equal volume of 1 mM Fe(NH₄)₂SO₄ and immediately diluted 8-fold with water. 4 μl of this mixture was added to each tube and the reaction started by addition of 1 μl of 50 mM DTT. Reactions were left at room temperature for 30 min and stopped by freezing in dry ice/ethanol. Samples were thawed by adding 50 μl of 0.6 M sodium acetate (pH 7), 80 mM EDTA, precipitated twice with ethanol, washed with 70% ethanol, dried, and dissolved in formamide loading dye. After incubation for 1 min at 90 °C samples were electrophoresed on 6% polyacrylamide gels with 50% urea.

Hydroxy radical footprinting was performed essentially as described by Tullius and Dombojski (1986). DNA-protein complexes (100 μl) were made as described above for DNase I footprinting, except that DTT and glyceral were omitted and 2.5 μg/ml of sheared salmon sperm DNA was included. The DNA cleavage reaction was started by adding 10 μl of freshly prepared EDTA.Fe(II) solution (made by adding 5 μl of a solution containing equal volumes of 10 mM Fe(NH₄)₂SO₄ and 20 mM EDTA to 25 μl of 3% H₂O₂, 5 μl of 100 mM L-ascorbic acid, and 15 μl of distilled water). After 2 min at room temperature, the reaction was stopped by adding 10 μl of 100 mM thioglycolic acid EDTA, 10 μl of 3 mM sodium acetate, and 450 μl of ethanol. The DNA was prepared for electrophoresis as described above for MPE footprinting.

MPE and hydroxyl radical footprinting consistently required more A protein to achieve protection.

Methylation Protection—The protocol used in these experiments was essentially as that described by Siebenlist and Gilbert (1980). Sodium cacodylate buffer inhibited formation of DNA-protein complexes, hence these were formed in Hepes buffer as described above for MPE footprinting. DNA was methylated by DMS (1 μl of a 1.0 M solution) at room temperature for 3 min. Reactions were stopped by addition of 5 μl of a solution containing 15 μM magnesium acetate, 50 mM sodium acetate (pH 7) and 1.0 mM β-mercaptoethanol. DNA was precipitated with ethanol, washed, and dried. Pellets were resuspended in 100 μl of 1.0 M piperidine and treated as in chemical sequencing procedures (Maxam and Gilbert, 1980), or treated according to the “A greater than G” reaction of Maxam and Gilbert (1977). In the latter case, DNA was dissolved in 20 μl of water, chilled on ice, and mixed with 5 μl of 0.3 M HCl. After 2 h on ice, 200 μl of 0.3 M sodium acetate was added, DNA precipitated, washed, and dried. DNA was dissolved in 10 μl of 0.1 M NaOH, 1 mM EDTA, heated at 90 °C for 30 min, dried again, and dissolved in formamide dye for electrophoresis.

Ethylation Interference—Experiments were based on the method of Siebenlist and Gilbert (1980). DNA samples were methylated as described above except in the absence of protein and in 50 mM sodium cacodylate pH 8.0, 100 μg/ml salmon sperm DNA. After methylation and ethanol precipitation, the samples were resuspended in 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, and used for protein binding experiments. Protein complexes were formed with varying amounts of A protein as described for MPE footprinting. Samples were electrophoresed on 6% nondenaturing polyacrylamide gels. Free (not bound by protein) and complexed DNA was located by autoradiography and eluted by crushing the appropriate gel pieces in 0.5 M ammonium acetate, 0.01 M magnesium acetate, 0.1% sodium dodecyl sulfate, and 10 μg/ml RNA. DNA was precipitated, dried, and cleaved with 1.0 M piperidine, or dissolved in water and processed for the A greater than G reaction as described for methylation protection.

Ethylation Interference—DNA was ethylated as described by Siebenlist and Gilbert (1980) and used for protein binding experiments as described for methylation interference experiments. After electrophoresis and elution from gels, precipitated DNA was resuspended in 30 μl of 10 mM sodium phosphate, pH 7.0, 1 mM EDTA, and 5 μl of 1.0 M NaOH. DNA was cleaved by heating at 90 °C and precipitated as described by Siebenlist and Gilbert (1980). Precipitates were washed in 80% ethanol, dried, and resuspended in formamide dye solution for electrophoresis.

All autoradiograms were scanned with a Bio-Rad Model 620 video densitometer.

RESULTS

MPE-Fe(II) and DNase I Footprints—MPE-Fe(II) is a small molecule that intercalates into DNA via the minor groove and cleaves the sugar phosphate backbone by hydroxyl free radical attack at several positions 3' to the site of intercalation (Hertzberg and Dervan, 1984; Dervan, 1986). This molecule has a lower sequence specificity than DNase I (Van Dyke and Dervan, 1983). Its small size and low sequence specificity produce smaller and discrete footprints that allow a more accurate determination of the size and location of protein binding sites than those obtained with DNase I.

To determine the size of A protein binding sites on att L1 of Mu DNA, purified A protein was incubated with restriction fragments derived from plasmid pRA-L29 that contain all three A protein binding sites at att L. Fragments were labeled both at the 5' and 3' ends close to site L1. DNA-A protein complexes were cleaved with MPE.Fe(II) and analyzed on 6% denaturing polyacrylamide gels (Fig. 1). With increasing A protein concentrations three distinct protected regions at L1 (around 5–30 bp), L2 (120–135 bp), and L3 (150–175 bp) are visible. L3 is protected at the lowest A protein concentration shown followed closely by L1. L2 is the last site to be protected. On the 5'-labeled top (+) strand (running 5' to 3' from left to right on Mu DNA), a maximum of 24 nucleotides (G7 to A30) are fully protected at L1 (Fig. 1A; at att L or at R the bases are numbered starting at 1 from either end of the Mu sequence and progressing inwards; see Priess et al., 1987; Kahmann and Kamp, 1987). The protection pattern of the adenine residues (A26–30) is not very distinct because of lack of good cleavage in this region even in the control (no protein) sample (lane 7). An enhanced cleavage is observed at T5. Such an enhancement, indicative perhaps of an altered DNA structure upon protein binding, is not seen on the bottom strand of L1 (Fig. 1B) or at any of the other att sites. On the 3'-labeled bottom (−) strand (Fig. 1B), protection from MPE.Fe(II) cleavage is observed for at least 20 bp, from T4 to T23. Again, as in the top strand, the protection pattern of the T24–30 residues is not distinct due to lack of cleavage in this region in the control (lane 1). The protection is offset by two to three nucleotides toward the 3' end of the bottom strand in relation to the top strand. No cleavage protection is observed for 4–5 base pairs in the immediate vicinity of the A protein nicking site (3' side of A1 on the bottom strand; Craigie and Mizuuchi, 1987). Similarly for R1, the site for A protein nicking at att R, no MPE.Fe(II) cleavage protection is observed for 5 base pairs in the neighborhood of the nicking site (see Fig. 6, where the protection results for all six att sites are summarized).

DNase I footprints on both strands of att L are shown beside the MPE footprints in Fig. 1 (lanes 1 and 2 in A and 11 and 12 in B; these were previously reported for the top strand only by Craigie et al., 1984). At L1 (Fig. 1B, bottom strand), one can observe that these extend a few base pairs beyond either side of the MPE footprints and include the A protein nicking site. A DNase I hypersensitive site appears on each strand, the two sites being separated from each other by 4 base pairs. The exact location of this pair of sites on the L1 sequence is indicated (see Fig. 6). Patterns of enhanced cutting by nuclease are indicative of exposure of one face of the DNA helix to the solvent, the other face being masked as it lies along the surface of a protein (Liu and Wang, 1978; Rhodes and Klug, 1980; Klug and Lutter, 1981; Drew and Travers 1985). Consistent with this, the enhanced nuclease cleavage occurs in a region not protected from hydroxyl radical attack (see below).

Hydroxyl Radical Footprints—MPE.Fe(II) cleavage demarcates the DNA sequence that is bound by a protein but provides no information on how the protein interacts with the backbone of the helix, i.e. whether the protein lies along one face of the helix or surrounds it from all sides. The
hydroxyl radical method of Tullius and Dombroski (1986) provides a means to assess such an interaction. Hydroxyl radicals, generated by breakdown of $H_2O_2$, are presumed to attack the deoxyribose sugars by extracting a proton. Secondary reactions of the resulting deoxyribose-centered radicals eventually break the backbone by disintegration of the sugars. Unlike MPE Fe(II) cleavage, the hydroxyl radical produces an intercalation-independent attack. Protein contact with the DNA backbone or protein-induced changes in the DNA helix, cause alterations in the pattern of hydroxyl radical attack.

Fig. 2 shows the hydroxyl radical analysis of the L1 binding site for A protein. For both the top and bottom strands, three areas of protection are evident from examination of the reaction in the absence or presence of A protein (lanes 1 vs. 2, respectively). Each strand has a similar pattern of interspersed protection and cleavage. In agreement with the MPE Fe(II) cleavage, the region in which A protein affects DNA cleavage spans approximately 24 bp. We note that even in the absence of A protein, there is an uneven pattern of hydroxyl radical attack on the DNA backbone. Some of these regions correspond to tandem adenine residues (Burkhoff and Tullius, 1987). The data for A protein protection at L1 are summarized (see Fig. 6). Similar patterns of interspersed protection and cleavage are observed for five of the six sites.

Methylation Protection—DMS methylates the N3 of adenine in the minor groove and the N7 of guanine in the major groove (Gilbert et al., 1976). When a protein binds DNA, the presence of the protein affects the methylation of purines, either blocking or enhancing it.

To begin to define specific contacts between A protein and its DNA binding site, we first determined the protection from methylation by DMS, of guanine residues within site L1 (Fig. 3). End-labeled DNA fragments were incubated with A protein as in the experiments above, and DMS was added. Guanines and adenines would be methylated by DMS unless they were blocked by close contact with the A protein. Depending on the nature of the subsequent treatment, DNA sugar-phosphate backbone can be broken at positions of methylated guanines alone, or both guanines and adenines. On the 5' labeled top strand (Fig. 3A; cf. lanes 1 vs. 2), G7, G19, and

![Image](http://example.com/image.png)

**FIG. 1.** MPE Fe(II) and DNasel cleavage patterns on att L1-A protein complexes. **A.** Analysis of the top strand (+) of an att L DNA fragment, labeled at the 5' end. DNA-A protein complexes were formed as described under "Experimental Procedures" and electrophoresed on denaturing polyacrylamide gels. Lanes 1 and 2 show the DNasel cleavage protection pattern, where prior to nuclease digestion, 2 pmol of labeled fragment were incubated with either no protein (lane 1), or 8 pmol of A protein (lane 2). Lanes 3-6 contain Maxam-Gilbert chemical cleavage reactions G, A+G, T+C, and C, respectively, performed on the same fragments. Lanes 7-12 show the MPE Fe(II) cleavage protection pattern, where prior to cleavage, 2 pmol of labeled fragment were incubated with either no A protein (lane 7), or 3, 4, 8, 14, or 20 pmol (lanes 8–12) of protein, respectively. Nucleotides are numbered starting at 1 from the left end of Mu and progressing inward. The three blocks of cleavage protection observed between 1 to 175 bp correspond to binding to sites L1, L2, and L3. Progressively, the position of a DNasel hypersensitive site that appears upon A protein binding at att L1 (lane 2), is marked by an arrowhead. **B.** Analysis of the bottom strand (−) of the att L DNA fragment, 3' labeled at the same end. Lanes 1-6 as in lanes 7–12 of A. Lanes 7–10 are the Maxam-Gilbert sequencing reactions as in lanes 3–6 of A. Lanes 11 and 12 as in lanes 1 and 2 of A. All other symbols as in A.

![Image](http://example.com/image.png)

**FIG. 2.** Hydroxyl radical cleavage pattern on att L1-A protein complexes. DNA fragments (2 pmol) labeled at the 5' or 3' ends (top or bottom strands, respectively) near att L1, were incubated with hydrogen peroxide plus iron-EuTPA in the absence (lane 1) or presence (lane 2) of A protein (20 pmol) as described under "Experimental Procedures" and electrophoresed in denaturing polyacrylamide gels. All other conventions as in Fig. 1.
G23 are protected from methylation, with G23 showing the strongest protection followed by G7. Protection of G19 appears unconvincing to the eye but is consistently seen by the densitometer. Similarly, densitometer tracings indicate weak protection of A8, A18, A26, and A28 (lanes 3 vs. 4). On the 3′-labeled bottom strand (Fig. 3B), G11 shows good protection while G13 shows only a very slight protection (lanes 1 vs. 2). No adenines are protected on the bottom strand (lanes 3 vs. 4).

**Methylation Interference**—To determine if the purine contacts identified in the methylation protection experiments were functionally significant, i.e., were essential for rather than a consequence of A protein binding, a complimentary experiment was carried out, following the protocol of Siebenlist and Gilbert (1980). End-labeled DNA was first partially methylated with DMS and then incubated with A protein. The bound and unbound DNA was isolated by electrophoresis through polyacrylamide gels. If a base normally comes into close contact with A protein, methylation of that base should interfere with binding. When DNA is treated so as to break the phosphodiester backbone at the position of each methylated base and the resulting fragments separated in a denaturing polyacrylamide gel, comparison of bound versus unbound DNA fragments reveals those positions of the DNA sequence at which methylation interferes with formation of the protein-DNA complex, i.e., the positions that are underrepresented in DNA extracted from the complex.

The results of such an experiment are presented in Fig. 4. For the top strand (+), G7 and G23 show a strong underrepresentation in the DNA from the complex, whereas G19 again shows a weak requirement (as judged by densitometric scans). For the bottom strand (−), G11 shows a stronger signal than G13. These results are in complete agreement with the protection experiments in that the residues showing protection from methylation (Fig. 3), are also seen to be essential for A protein binding. However, none of the adenine contacts were found to be essential, when determined as described above for methylation protection (data not shown).

**Ethylation Interference**—These experiments are very similar to the methylation interference experiments described above except that the phosphates were modified instead of the purines. Ethylnitrosourea reacts with DNA and places an ethyl group on one of the phosphate oxygens. The ethyl group may inhibit binding by neutralizing a negative charge or by steric interference.

As in the methylation experiments, end-labeled DNA was partially ethylated with ethylnitrosourea, complexed with A protein, and isolated from polyacrylamide gels as described under "Experimental Procedures." DNA was cleaved with alkali and heat, which cleaves the DNA backbone next to the ethylated phosphates, creating a series of fragments which were separated on denaturing polyacrylamide gels. Results in Fig. 5 (confirmed by densitometric scans of the autoradiograph) show that for the top strand, ethylation of phosphates 3′ to positions 5–7, 10–12, 17–24, 27, and 29 interferes with binding at att L1. There is an enhanced cleavage at the phosphate 3′ to position 16. On the bottom strand, phosphates 5′ to positions 11–13, 16, 24–27, show interference. Thus all the phosphates around the essential purines identified in the methylation interference experiments, are seen to be important.

**Summary of Contacts at Other att Sites**—Fig. 6A displays A protein contacts (only hydroxyl radical footprints, DNase I hypersensitive sites, and strong guanine contacts) at the L1 site on a model of the DNA helix drawn with 10.5 bp per helical turn. The essential guanine contacts occur in two
Transposase Contacts with Mu DNA Ends

Fig. 6. Summary of A protein contacts at all att sites. A. contacts at site L1 (1-30 bp) are displayed on a DNA helix drawn with 10.5 bp per turn. Backbone residues protected from hydroxyl radical cleavage are black. Guanines that show strong interference are circled. Position of the pair of DNaseI hypersensitive sites is indicated by arrows, and the site where A protein nicks the left Mu end during strand transfer (Craigie and Mizuuchi, 1987) is indicated by a diamond. The sequence of the top strand is shown and has been assigned an arbitrary orientation designated by an arrow pointing to the left above the helical display. B, 30 bp at each att site, as originally aligned to accommodate DNase I footprinting data (Craigie et al., 1984) are shown. Nucleotides are numbered starting at 1 from either end of Mu and progressing inward. While comparing the A protein contacts at all the att sites, nucleotides are referred to in the text according to their alignment with site L1. Note that DNA strands are denoted top (+) or bottom (-) also with respect to the L1 site. On the Mu genome, this holds true for the att L sites and for site R1 (all of which are oriented in the same direction), but not for sites R1 and R2 which are in the opposite orientation. MPE.Fe(II) cleavage protection limits are marked by brackets. Letter Y with hash marks indicates that these limits could not be accurately determined due to lack of cleavage or a lack of control (no protein lane, but do not extend beyond the bracket). Enhanced MPE.Fe(II) cleavage is indicated by a diamond. Hydroxyl radical cleavage protection areas are boxed. The double-boxed T residue (at L3) shows enhanced hydroxyl radical cleavage. DNaseI hypersensitive sites are marked by filled triangles while A protein nicking sites (at L1 and R1) during strand transfer are indicated with diamonds. Purines that show strong and weak interference are enclosed with open circles or triangles, respectively. Features of interference shared by all sites (except L2) are represented by outlined letters while those shared by only some of the sites are represented by italicized letters. Enhanced methylation is indicated by † above the affected base. Enhanced methylation at +G14 at R2 and +G20 at R3 was observed only during indirect protection experiments (see text). Purines showing common protection (but not interference) are marked by *. For protected guanines not common to all sites, see text. Phosphates showing interference or enhanced ethylation are indicated by small dots or dots marked by † respectively. Ethylation interference data are available only for site L1. Note that sequences at R1 and R2 (23-30 bp) overlap in the drawing in order to show alignment over the larger DNase I-protected region.

The ends of Mu are complex regions containing an asymmetric distribution of transposase (A protein) binding sites. Initial nuclease footprinting studies (Craigie et al., 1984) had shown an approximately 30-bp stretch of DNA to be protected at each of the six binding sites at the two ends of Mu. The data presented in this paper define the specific contacts made by A protein at each of these sites. At most sites, the interaction occurs over 24 bp encompassing two major grooves, with the protein occupying only one face of the helix. L2 is the smallest and weakest binding site, with interactions observed mainly over one major groove (13 bp). Upon protein binding, a pair of DNase I hypersensitive sites appear at the back of the second contacted major groove at most sites except at L2. The A protein bends DNA through 90° at each of these sites except at L2, which is the least bent site (Kuo et al., 1991). It has been suggested that DNase I binds across the minor groove of the DNA helix and that its cutting efficiency might be affected by local variations in the width of this groove (Drew, 1984; Drew and Travers, 1984). An enhanced susceptibility to nuclease attack could therefore be

consecutive major grooves. The hydroxyl radical footprints are seen to map on one face of the helix, while the DNase I hypersensitive sites are found on the back of the second contacted major groove. Fig. 6B summarizes the protection and interference results for all att sites. MPE.Fe(II) and hydroxyl radical protection patterns (indicated by brackets and boxes, respectively) are similar for all sites except for L2. At L2, although the DNase I footprint shows a large protected region (see Fig. 1 and Craigie et al., 1984), the MPE.Fe(II) and hydroxyl radical footprints show only a 13-bp region of protection. At att R, the R1 and R2 sites show no demarcation of either MPE or hydroxyl radical cleavage patterns (note that in the figure, nucleotides 23-30 at the R1 site, are also included at the R2 site), while there is an enhanced MPE cleavage at four nucleotides between R2 and R3. The enhanced hydroxyl radical cleavage at T in site L3 (double-boxed on the bottom strand) is not always seen and not observed at equivalent positions in other sites.

A pair of DNase I hypersensitive sites (filled triangles) occur at most att sites except at the smaller L2 site. At R3, only a single DNase I hypersensitive site is found on the bottom strand as aligned in Fig. 6. Since all other A protein contacts at this site are similar to those found at the other att sites at which DNase I hypersensitive sites appear on both strands, it is likely that the phasing and orientation of the R3 site with respect to its neighbor R2 is preventing access of its top strand to DNase I.

The three essential guanine additions (+G7, -G11, and +G23; when comparing contacts, residue numbering will be discussed with reference to site L1) are shared by all sites with the exception of L2. -G13 shows weak methylation interference at all sites where it is present (L1, R1 and R2), as does +G21 (L2, R1-R3). L2 is unusual in that it shows strong interference at +G16, +G21, and -G20. At R1, a pronounced enhancement of methylation is seen at +G8. None of the other sites have a G residue at this position. At R2 and R3, indirect footprinting experiments (performed by Klenow polymerase extension of labeled primer on DNA cleaved after methylation; Kuo et al., 1991) show an enhancement of methylation at +G14 and -G20, respectively, not seen in direct protection experiments. On the bottom strand of all sites (except L1), -G22 showed protection but no interference. Not shown in the figure are -G5 and -G12 at L2, and +G5 and -G6 at R1-R2, which showed moderate to weak protection but no interference. Adenines at positions equivalent to 17, 18, and 26 at site L1, showed protection but not interference at all sites except at L2, where A17 and 18 showed weak interference (Fig. 6 shows adenine protection data only for residues that are common to most sites). At L3, A20 on the bottom strand showed weak interference.

DISCUSSION

The ends of Mu are complex regions containing an asymmetric distribution of transposase (A protein) binding sites. Initial nuclease footprinting studies (Craigie et al., 1984) had shown an approximately 30-bp stretch of DNA to be protected at each of the six binding sites at the two ends of Mu. The data presented in this paper define the specific contacts made by A protein at each of these sites. At most sites, the interaction occurs over 24 bp encompassing two major grooves, with the protein occupying only one face of the helix. L2 is the smallest and weakest binding site, with interactions observed mainly over one major groove (13 bp). Upon protein binding, a pair of DNase I hypersensitive sites appear at the back of the second contacted major groove at most sites except at L2. The A protein bends DNA through 90° at each of these sites except at L2, which is the least bent site (Kuo et al., 1991). It has been suggested that DNase I binds across the minor groove of the DNA helix and that its cutting efficiency might be affected by local variations in the width of this groove (Drew, 1984; Drew and Travers, 1984). An enhanced susceptibility to nuclease attack could therefore be
indicative of nearby alterations in the width of the minor groove (Fox and Waring, 1984). We note that no enhanced cleavage by MPE Fe(II) is seen in this region. An enhanced intercalation of this molecule (which intercalates via the minor groove) is correlated with a bend that γδ resolvase induces at its binding site, suggesting that a bend toward the major groove may open the minor groove for access of the intercalator (Hatfull et al., 1987). The Mu A protein may either induce a bend different from the γδ resolvase or make a similar bend inaccessible to the intercalator. Proteins can induce bends in DNA sequences by compression of either the major or minor grooves (Salvo and Grindley, 1985; Zinkel and Crothers, 1987). The composition of the DNA at the site of bending can affect the degree to which DNA can be bent: GC-rich sequences compress more easily toward the major groove while AT-rich sequences compress more easily toward the minor groove (Drew and Travers, 1985; Gartenberg and Crothers, 1988; Crothers et al., 1990). We note that all the att sites have 2-4 GC base pairs between the two DNase I sites (Fig. 6). Whether or not the above observations argue in favor of a compression of the major groove and an opening up of the minor groove to accommodate the bend, will need to be resolved by further experimentation.

Three strong guanine contacts are found at equivalent positions in all sites except at L2, which is missing 2 of these guanines from its sequence in the first major groove. A pronounced enhancement of methylation is seen at +G8 at the R1 site, which is the only site with a guanine at this position. The enhancement may reflect an increased DMS concentration in a hydrophobic protein pocket or local DNA helix unwinding (Johnsrud, 1978). At L1, there is an adenine at this position which shows no enhancement of methylation. This suggests that if similar changes occur at the L1 site, they are manifested only in the major groove. Interestingly, inspite of the many runs of As at all sites, no essential minor groove adenine contacts were observed (except for two weak contacts at L2 and one weak contact at L3). These results suggest that the A protein achieves binding specificity primarily through major groove contacts. The adenine runs may impart curvature and hence an appropriate structure for facilitating and stabilizing A protein binding. Also noteworthy is the observation that at L1 and R1, no contacts are found within 4-5 bp flanking the A protein nicking sites.

Mutation of specific nucleotides at the att sites and measurement of the magnitude of their effects on Mu transposition in vivo (Burlingame et al., 1986; Groenen and van de Putte 1986) are largely consistent with all the contacts delineated in our studies. However, mutation of some of the guanines showing weak interference in our studies (−13 and +19 at L1, −9 and −20 at R1, and +13 and +21 at R2) were seen to have no effect on transposition frequency in vivo (Groenen and van de Putte, 1986).

The N-terminal domain of Mu A protein encodes att DNA binding specificity (Nakayama et al., 1985). A sequence homologous to the helix-turn-helix motif, commonly employed in DNA binding by many proteins, occurs in this region between residues 130 and 150 (Harshey et al., 1985). Many protein binding DNA sites are symmetric, and the proteins that bind to these sites have two or four identical subunits (for reviews see Pabo and Sauer, 1984; Schleif, 1988; Knight et al., 1989). Although different kinds of protein structures are used for binding, for most of the proteins for which information is available, binding is achieved by two symmetrical protein units contacting adjacent major grooves of the DNA helix. Since the A protein binds DNA as a monomer (Kuo et al., 1991) to a sequence without dyad symmetry and protects two consecutive major grooves, it clearly contacts DNA differently from the examples known to date. Even if the helix-turn-helix motif were found to participate in DNA binding, some other structure must also be used to cover the entire observed contact region. Recent results with γδ transposase show that this protein also binds to one face of the DNA and interacts over three consecutive major and minor grooves (Wiater and Grindley, 1991). It may well be that this class of proteins uses a new type of structural motif for interaction with DNA.

During site-specific recombination the recombinating sites must synapse prior to strand exchange. A growing body of evidence suggests that the recombinating sites are melded into stable higher-order structures by an assembly of recombination proteins, which contort DNA appropriately by bending and looping (see Salvo and Grindley, 1988; Landy, 1989; Kanaar, 1988; Heichman and Johnson, 1990). A higher-order complex is also formed during Mu transposition (Surette et al., 1987; Craigie and Mizuuchi, 1987). The hallmark of this complex is its unusual stability (50% disruption in 4.6 M urea or at 56 °C; Surette et al., 1978). Indirect footprinting experiments have shown that although all six att binding sites are occupied in such a complex, three of these sites are loosely held (Kuo et al., 1991). A synaptic complex with only three sites occupied ("core" complex, comprised of sites L1, R1, and R2), is as stable (in 4 M urea, or 2 M NaCl or at 60 °C) and equally competent in the subsequent strand transfer process. It is rather remarkable that such extraordinary stability can result from interactions between only three att sites. The results reported in this paper allow us to speculate on how the arrangement of the core sites on Mu DNA ends might contribute to the stability of the synaptic complex. Inspection of the A-binding sequences and the hydroxyl radical footprints (see Fig. 6) at the core att R sites (R1−R2) reveals that if one were to assume a normal B DNA configuration, R1 would be out of phase with respect to R2. When a substantial bend is introduced at each of these sites by A protein, this end might be shaped into a hook-like frame. This special configuration of DNA at the R1−R2 sites could help wrap the att R end around a paired att L end and facilitate the kind of proximity required for fostering interactions between the core sites that would be necessary for stabilizing a higher-order structure adopted by the ends. Current experiments are aimed at assessing the importance of these features, i.e. DNA bends and helical phasing of the R1−R2 sites, for formation of the synaptic complex.

REFERENCES


Transposase Contacts with Mu DNA Ends


