Nucleic Acids Research, 1998, Vol. 26, No. 15 **3521–3527**

Altering the DNA-binding specificity of Mu transposase *in vitro*

Soon-Young Namgoong, Senthil Sankaralingam and Rasika M. Harshey*

Department of Microbiology and Institute of Cellular and Molecular Biology, University of Texas at Austin, Austin, TX 78712, USA

Received April 20, 1998; Revised May 20, 1998; Accepted June 16, 1998

ABSTRACT

We describe the isolation of a variant of Mu transposase (MuA protein) which can recognize altered *att* sites at the ends of Mu DNA. No prior knowledge of the structure of the DNA binding domain or its mode of interaction with *att* DNA was necessary to obtain this variant. Protein secondary structure programs initially helped target mutations to predicted helical regions within a subdomain of MuA demonstrated to harbor *att* DNA binding activity. Of the 54 mutant positions examined, only two showed decreased affinity for *att* DNA, while eight others affected assembly of the Mu transpososome. A variant impaired in DNA binding [MuA(R146V)], and predicted to be in the recognition helix of an HTH motif, was challenged with altered *att* sites created from degenerate oligonucleotides to select for novel DNA binding specificity. DNA sequences bound to MuA(R146V) were detected by gel-retardation, and following several steps of PCR amplification/enrichment, were identified by cloning and sequencing. The strategy allowed recovery of an altered *att* site for which MuA(R146V) showed higher affinity than for the wild-type site, although this site was bound by wild-type MuA as well. The altered association between MuA(R146V) and an altered *att* site target was competent in transposition. We discuss the strengths and limitations of this methodology, which has applications in dissecting the functional role of specific protein–DNA associations.

INTRODUCTION

DNA transactions such as transposition, site-specific recombination, replication and transcription are performed by large protein–DNA assemblies (1, and references therein). Not unlike protein chaperones that guide the final form of partially folded proteins (2), multiple binding sites for one or more proteins provide scaffolds for guiding specific functional interactions within these assemblies.

The Mu transpososome is an example of a large nucleoprotein complex within which a tetrameric form of the Mu transposase (MuA protein) executes the chemistry of Mu DNA transposition (3). Several cofactors are essential for formation of the transpososome. These include a negatively supercoiled DNA substrate, *Escherichia coli* HU protein, and metal ions. Two families (*att* and enhancer) of multiple sites interact with individual monomers of MuA, promoting assembly of a tetramic complex which footprints on only three of the six binding sites within the left (*att*L) and right (*att*R) ends (L1, R1 and R2, but not L2, L3 and R3; Fig. 1) (the individual sites will be referred to here with an *att* prefix). MuA interactions with the enhancer sites have been shown to help overcome an intrinsic barrier in the protein structure that prevents its tetramerization when free in solution (4). Interactions with two of the three accessory *att* binding sites (L2, L3, R3) are also important for tetramer assembly (5). What role do the multiple DNA sites and protein factors play in unfolding the dormant activity of MuA, but only within the context of the tetramer? Is there an ordered set of interactions between individual DNA–protein associations? Does each subunit perform a distinct role in the finished complex? How does position dictate function?

Complementation experiments between two sets of catalytically inactive variants have suggested that monomeric MuA may carry only a partial active site, with full active sites for DNA cleavage and strand transfer being assembled by sharing polypeptide domains between MuA monomers (6). Experiments designed to determine the functional contribution of each MuA subunit to the different steps of transposition have thus far relied on assembly of the tetramer under artificial conditions [addition of dimethyl sulfoxide (DMSO)], where a functional MuA–DNA complex can be assembled efficiently on linear right end substrates, bypassing the requirement for HU protein, supercoiled DNA, the Mu enhancer and the multiple *att* sites (7–10). While these experiments have provided insights into the arrangements of the individual subunits within a tetramer assembled on linear *att*R substrates (11–13), they have left open the possibility that the linear topology of the *att* substrates might permit pairing possibilities not available to sites arranged to interact within the constraints of a normally supercoiled DNA substrate (14,15). The linear substrates also do not provide the opportunity to address how the complex arrangement of *att* and enhancer sites found on a natural Mu substrate influences subunit association and function within the tetramer. A powerful tool to investigate such questions would be an altered *att* specificity variant of MuA, that could be directed to the different MuA binding sites to report on their role in the

*To whom correspondence should be addressed at: Department of Microbiology, University of Texas at Austin, Austin, TX 78712, USA. Tel: +1 512 471 6881; Fax: +1 512 471 7088; Email: rasika@uts.cc.utexas.edu

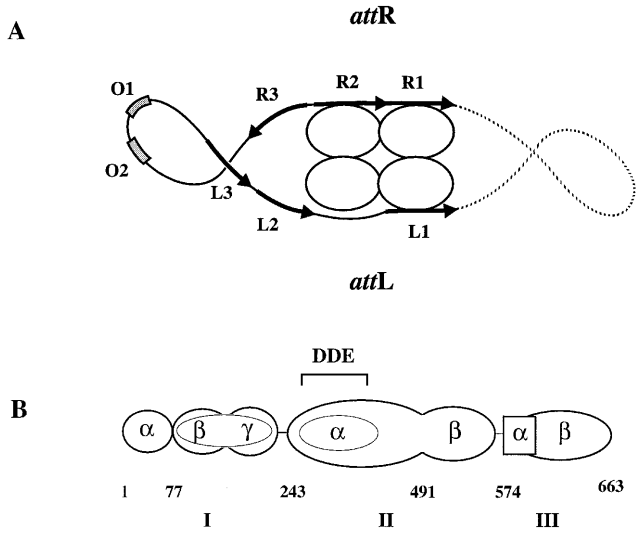

Figure 1. (**A**) Disposition of *att*L (L1–L3), *att*R (R1–R3) and enhancer (O1–O2) sites on supercoiled Mu DNA. The MuA tetramer footprints on only three *att* sites (L1, R1, R2). Non-Mu DNA is indicated by broken lines. (**B**) Domainal organization of MuA. On the basis of limited proteolysis, three domains (I–III) were assigned to MuA protein. Amino acid numbers corresponding to the N-terminus of each major subdomain (designated α, β or γ) are shown beneath the structure. Domain Iβγ is responsible for *att* DNA binding. A catalytic triad of catalytic DDE residues is found in domain IIα. See text for details.

assembly and function of the transpososome. The problem was particularly challenging given that the *att* site spans two turns of the DNA helix (~25 bp) and requires two protein subdomains (comprising together ~200 amino acids) for recognition. We describe here the isolation of such a variant. The properties of this variant agree well with the structural features of the MuA DNA-binding domain, which became available after this work was completed (16,17). As such, the experimental approach that was successfully applied to MuA should be suitable for other DNA-binding systems that lack structural information.

MATERIALS AND METHODS

DNA and protein reagents

Plasmids pMK21 (mini-Mu donor substrate) and pET158 (vector for high-level protein expression) have been described (18). pUC19 was used for cloning individual wild-type and mutant *att* sites; *att* substrates (~100 bp in length) for DNA binding assays were generated by PCR amplification, such that a 30 bp *att* site was flanked by ~35 bp of pUC19 DNA on either side. Double-stranded R1–R2 and *att*L substrates have been described previously (11 and 19, respectively).

MuA and HU proteins were purified as described (18). DNA and purified protein concentrations were determined as previously described (20). DNA-binding and transposition activities of MuA variants were assayed either in crude lysates or after purification.

Procedures for DNA manipulation were as described (21). PCR reactions employed Vent Polymerase from New England Biolabs.

Site-directed mutagenesis

Residues in domain Iβγ were altered by PCR mutagenesis using the megaprimer method (22), and cloned directly into pET158. Mutagenic primers contained a centrally placed degenerate codon of the targeted residue, so as to recover a spectrum of amino acid changes at that residue. All mutations were verified by DNA sequencing.

Design of oligonucleotides for altering MuA specificity

Five sets (S1–S5) of single-stranded *att*R2 oligonucleotides (100mers) were synthesized, each set differing from the next in the position of five contiguous degenerate nucleotides within the centrally placed 30 nt R2 sequence, flanked on either side by sequences encompassing the *Eco*RI–*Hin*dIII polylinker region of pUC19 plasmid. Double-stranded forms of these substrates were generated by PCR amplification using appropriate primers.

PCR amplification/selection protocol for MuA(R146V)-altered DNA association

Aliquots of 0.02 pmol of 5′ ^{32}P-labeled *att*R2 substrates (S1–S5) were incubated with 1.5 pmol MuA or 10 pmol MuA(R146V) in 20 µl 25 mM Tris–HCl (pH 7.5), 10 mM $MgCl_2$, 1 µg salmon sperm DNA and 150 mM NaCl for 15 min at 30°C. Bovine serum albumin (50 µg/ml) and 5% glycerol (final) were added to the reactions prior to electrophoresis on 6% native polyacrylamide gels. MuA(R146V)-bound S2 DNA was detected by auto-radiography, recovered, eluted (21) and amplified by PCR using appropriate primers. Enrichment of the cognate DNA sequences for MuA(R146V) was repeated through three more cycles of binding, isolation of bound DNA and PCR amplification. During each enrichment step, a fixed amount of the amplified substrate (0.02 pmol) was titrated with a range of MuA(R146V) concentrations. Only the binding reaction that yielded ~10–20% of the input DNA in the bound complex was subjected to the next enrichment cycle. Reactions that yielded >20% conversion of the substrate into bound form were discarded so as to maintain high-stringency selection against non-specific sequences. Compared to the initial binding reaction, the first and second enrichment steps required 8 and 5 pmol of MuA(R146V), respectively, to yield 10% of the substrate in the bound form. However, no reduction in the protein amount was observed between the second and third cycles to obtain 10% level of substrate binding. The protein-associated DNA from the third enrichment cycle was PCR amplified and cloned into pUC19 digested with *Eco*RI and *Hin*dIII. Plasmid DNA was isolated from several independent clones, and the inserts were sequenced.

In vitro assays for DNA transposition

Type 0 (18), type I (23) and cleavage assays for linear substrates (11) have been described. Crude extracts proved unreliable for type 0 assays, which were performed only on purified proteins.

Gel retardation assays

Relative dissociation constants (K_d values) for MuA and MuA(R146V) were estimated by titrating a constant amount of *att* or *att*^{S2} DNA with increasing amounts of the proteins (24). K_d values for binding were estimated by determination of the protein concentration needed to complex half of the DNA under conditions where the protein was in excess over the DNA [0.02 pmol 5′ ^{32}P-labeled DNA and 0.1–10 pmol of purified MuA or MuA(R146V)]. Protein–DNA complexes were formed and

analyzed as described above, except salmon sperm DNA concentration was reduced to 0.2 µg.

DNA-binding activity of domain Iβγ mutants was assayed using end-labeled *att*L DNA.

Quantitation

Gels were dried, visualized by autoradiography or Bio-Rad Phosphor Imager and analysed by Molecular Analyst 2.0 video densitometer.

Protein secondary structure prediction

The SOPMA (Self Optimized Prediction Method for Alignment; 25,26) protein sequence analysis package was used for secondary structure prediction (http://www.ibcp.fr/serv_pred.html).

RESULTS AND DISCUSSION

Rationale

Most strategies to alter the DNA-binding specificity of proteins thus far have relied on knowledge of DNA–protein contacts gained either from DNA footprinting experiments, or from crystal structures of DNA–protein complexes. These strategies have generally included: (i) construction of chimeric DNA-binding proteins from different DNA-binding domains; (ii) site-directed alterations of known amino acid–base pair contacts; and (iii) random mutagenesis of the gene encoding the DNA binding protein, followed by genetic selection strategies aimed at decreased affinity for the wild-type DNA site and for increased affinity for altered DNA (reviewed in 27,28). We describe here a strategy by which we have altered the DNA-binding specificity of MuA protein without prior knowledge of specific protein–DNA contacts. The strategy involves: (i) site-directed mutagenesis of predicted secondary structure motifs within a known DNA-binding domain; (ii) choice of a DNA-binding mutant based on the failure of most mutations at that position to bind DNA; (iii) restoration of DNA binding to altered oligonucleotides *in vitro*; and (iv) enrichment of bound DNA by several cycles of PCR amplification.

The outcome of the above strategy was an altered DNA site for which a MuA variant shows a significantly higher relative affinity than for the wild-type site. However, this affinity was lower than that of wild-type MuA for the wild-type site. The altered site was bound by MuA and the variant with roughly equal affinities. In the context of the results discussed below, we define 'altered specificity' as the ability of a protein (the MuA variant) to discriminate between two DNA targets (the native or the altered binding site).

Secondary structure prediction and mutagenesis of the *att* DNA-binding domain Iβγ

The MuA *att* site is large (29,30), and residues in two separate subdomains of MuA (Iβγ; Figs 1B and 2; 19,31,32) bind two consecutive major grooves and an intervening minor groove

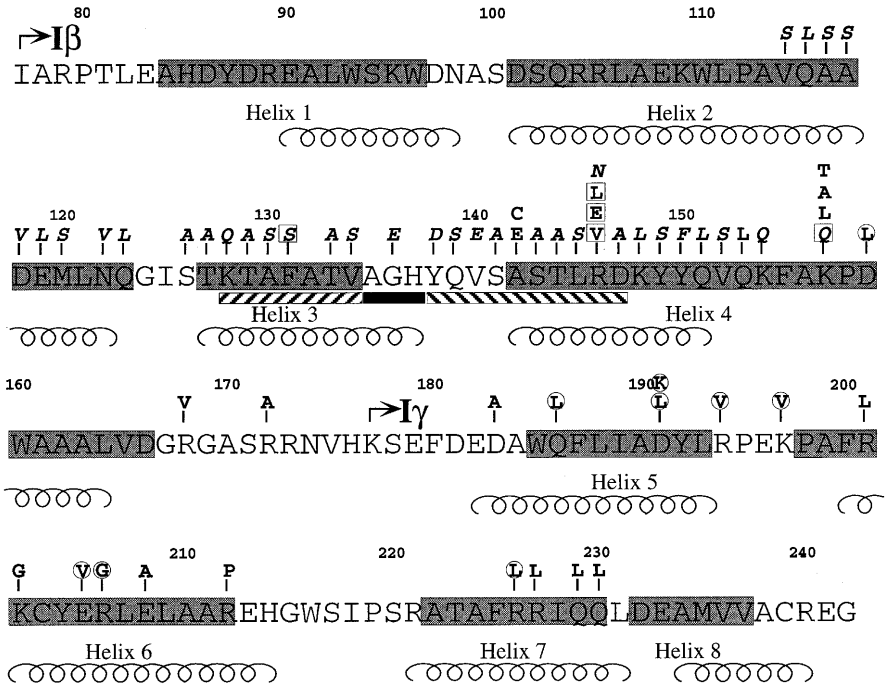

Figure 2. Summary of targeted mutagenesis of domain Iβγ. The domain Iβ and Iγ subdivision (arrows) is based on partial proteolysis of MuA (19). Eight predicted helical regions (Helix 1–8, shaded grey) within domain Iβγ were identified using the SOPMA package (see Materials and Methods). The consensus secondary structure predicted by this method for amino acid residues 77–242 of MuA was as follows (H, helix; E, sheet; C, coil): CHCCHHCCHHHHHHHHHHHHHCCCCHHHHHHH-HHHHHHHHHHHHHHHHHCCCHHHHHHHHCCCEEEEHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCHHCCCHHHCCHHCHHHHHHHHHCCCCHH HHHHHHHHHHHHHCCCCCCCCCHHHHHHHHHCHHHHHHEHHCC. A previously identified HTH motif is indicated by hatched boxes (34). Residues in the predicted helical regions were primarily targeted for mutagenesis (vertical bars with indicated changes). Mutants isolated in a previous study (32) are indicated by italicized font. Mutations affecting DNA-binding are boxed, while those affecting assembly are circled (see text for details). Recently determined helical regions using NMR (16,17) are indicated as helices below the peptide sequence.

(30,33). In an earlier study, targeted mutagenesis of residues within a putative HTH motif in domain Iβ had identified three mutants (F131S, R146N, K157Q) that were compromised in DNA binding (Fig. 2; 32). Based on the predicted secondary structure of domain Iβγ (Materials and Methods), we targeted residues primarily within helical regions of Iγ for mutagenesis. Of particular interest were several acidic and basic residues that might potentially be involved in base-specific or phosphate contacts with DNA. A summary of the mutagenesis results is presented in Figure 2. [While this manuscript was in preparation, Clubb *et al*. (16) and Schumacher *et al*. (17) published solution structures for isolated domains Iβ and Iγ. The congruence between the predicted and determined helical regions was excellent (Fig. 2).] The mutants were first tested in crude cell extracts for cleavage of Mu ends (type I assay) (Materials and Methods). Mutants defective in these assays were further tested for DNA binding, as described in Materials and Methods. Assembly of the Mu transpososome (type 0 assay) was monitored using purified proteins.

Consistent with our previous inference that Arg146 was important in DNA recognition (32), the present analysis also revealed several alterations at this position that resulted in loss of binding (Fig. 2). However, unlike the K157Q mutation that abolished DNA binding (32), the mutations introduced at the Lys157 position in this study showed little effect on DNA binding (Fig. 2). Therefore, Arg146 variants were chosen for further analysis. This residue is located within a helix which was predicted earlier to be the recognition helix of an HTH motif (34). This prediction has now been confirmed by the recently published NMR studies (17). Essentially all of the mutations at the 146 position failed to bind DNA, as would be expected for a residue that makes a critical contact with DNA (Fig. 2). The only exception was an Asn substitution for Arg146. Although the R146N mutant was able to bind DNA, the migration of the resultant DNA–protein complexes in gel-retardation experiments was different from that of wild-type MuA complexes (32). Selection of altered target specificity was carried out using MuA(R146V). In a precedent provided by the CAP protein, substitution of Glu181 by Val, Leu or Lys resulted in a protein that had specificity for A-T base pairs at positions 7 and 16 of the DNA recognition site, rather than G-C base pairs favored by wild-type CAP (35,36).

Eight other mutants (D159L, Q186L, D191LK, R194V, K197V, E205V, R206G, R226L) were found to be defective in type I formation (cleavage of Mu ends), but not in DNA binding (Fig. 2; data not shown). No type 0 (assembled but uncleaved complex) was detected with these mutants in the presence of either Mg^{2+} or Ca^{2+} ions. These results suggest that domain I of MuA, in addition to its role in DNA binding, may also contribute to the later steps in transposition.

Association of MuA(R146V) with altered *att* DNA sites

In order to determine if any alteration in the *att* DNA sequence would allow MuA(R146V) to re-establish binding, we targeted five contiguous regions separately within *att*R2 (the first site to be occupied on supercoiled Mu DNA upon titration with increasing MuA; 20) for alteration (Fig. 3A). The rationale for this strategy was that if Arg146 was involved in a direct DNA contact, specific alterations within or in the vicinity of the contacted nucleotide might restore interaction with the

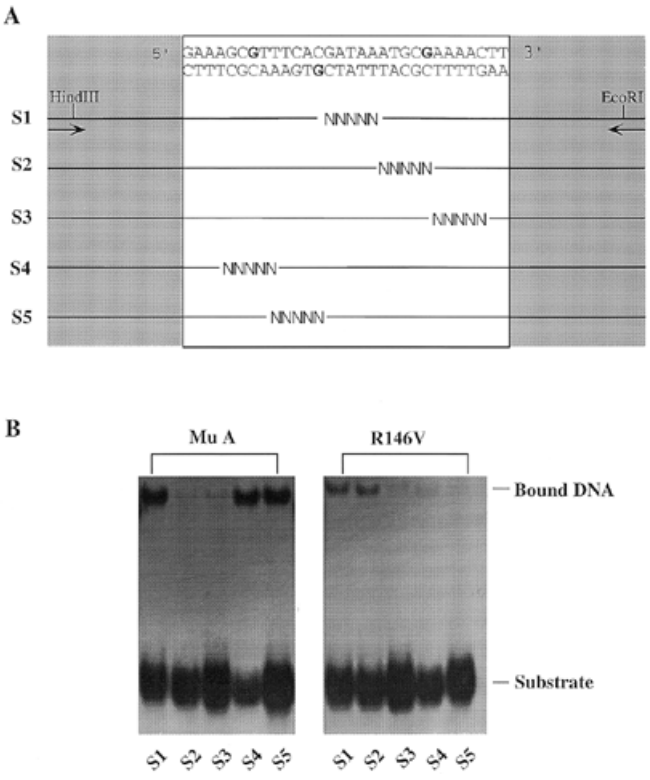

Figure 3. Strategy for altering the specificity of MuA. (**A**) Five sets of degenerate *att*R2 oligonucleotides (S1–S5) spanning a 25 nt region (essential wild-type MuA contacts with G residues at R2 indicated by boldface; 30) were synthesized, each set differing from the next in the location of five contiguous degenerate (N is any one of the four bases) nucleotides. Extension of appropriately placed primers (arrows) in flanking residues (pUC19 sequences; grey areas) converted these single-stranded substrates into double-stranded DNA. (**B**) Aliquots (0.02 pmol) of end-labeled *att*R2 substrates (S1–S5) were incubated with either 1.5 pmol of MuA or 10 pmol of MuA(R146V), and electrophoresed in 6% polyacrylamide gels. DNA was visualized by autoradiography. See Materials and Methods for details.

Arg146Val substituent. In principle, the restoration of binding could result via a novel DNA contact or via the removal of binding interference caused by an unfavorable interaction. A gel-retardation assay was carried out with complexes of the altered (S1–S5) double-stranded oligonucleotides and either wild-type MuA or MuA(R146V) (Fig. 3B; Materials and Methods). The results of this initial experiment clearly indicated that this strategy was likely to succeed. Substitutions within the S2 and S3 regions were not well tolerated by wild-type MuA when compared to those at S1, S4 and S5, as indicated by poor binding to the former substrates. (The extent of binding obtained with S1, S4 or S5 was ~2–3-fold lower than that with wild-type *att*R2 under identical reaction conditions; data not shown.) This is consistent with the observed higher affinity of the right half of the *att* consensus sequence for domain Iβγ (17). On the other hand, MuA(R146V) appeared to associate with the S1 and S2 substrates preferentially. Because wild-type MuA showed good binding to S1, the S2 substrate was chosen for further selection, in the hope that it would provide a better discrimination between wild-type MuA and MuA(R146V). [Note, however, that the binding reactions contained ~6–7-fold higher MuA(R146V) concentration relative to MuA. The choice of the protein amount

	A	T	G	C	G
	T	T	G	C	C
	T	G	A	G	C
	G	G	C	C	C
	G	C	C	G	C
	G	C	A	C	C
	C	T	G	C	G
	C	G	C	G	A
	C	G	C	A	C
	C	C	T	C	C
	C	C	C	C	C
	C	C	C	C	C
	C	C	C	C	A
	C	C	A	C	T
	C	C	A	C	C
	A	G	C	A	G
	A	G	C	A	C
	A	C	C	C	G
	A	C	C	C	C
	A	C	A	C	A
A	5	0	5	3	3
C	9	11	11	13	12
G	3	6	2	3	3
T	2	2	1	0	1

Figure 4. Sequence of 19 independent *att*R2(S2) clones competent in binding MuA(R146V). The 'S2' region on one strand of the R2 site is indicated at the top. Frequency of the four bases (A, C, G, T) recovered at each of the five altered nucleotides within 'S2' is shown at the bottom.

(yielding ~10–20% of the DNA in complexed form) was based upon prior titration assays.]

MuA(R146V)–S2 DNA complexes were isolated, and the bound DNA amplified by PCR (Materials and Methods). Complex formation, DNA isolation and PCR amplification of MuA(R146V)-bound DNA was repeated for three more cycles, until there was no further enrichment in binding (Materials and Methods). Bound DNA from the final selection cycle was cloned into pUC19. The sequence of 19 independent clones is shown in Figure 4. The recovered DNA sequences that restored binding to MuA(R146V) showed an increase in the number of C residues in the top strand (in the S2 region), rather than changes at a specific position. Two of the 19 sequences recovered had all C nucleotides in the top strand. All of these DNA substrates showed comparable affinities (within a factor of 2) for MuA(R146V) (data not shown). An R2 site whose S2 region consisted of all C nucleotides in the top strand, was chosen for functional characterization. We shall henceforth indicate this altered site with an 'S2' superscript.

Specificity of MuA(R146V) for *att*^{S2}

The specificity and affinity of wild-type MuA and MuA(R146V) for *att*R2^{S2} were determined from band-shift assays (Table 1; see Materials and Methods). Since the six *att* sites (L1–L3 and R1–R3) are not all identical (29,30), and since wild-type MuA is known to have slightly different affinities for these sites (20,29), the affinity of MuA(R146V) for all other *att*^{S2} sites (Table 1A; Materials and Methods) was also determined. The binding conditions employed were similar to those used for standard *in vitro* transposition assays, except that the reactions contained 10 µg/ml salmon sperm DNA (Materials and Methods). MuA(R146V) showed an ~60-fold lower affinity for *att*R2 compared to wild-type MuA (Table 1B). Similar affinities

Table 1. DNA binding affinities of wild-type and Mu(R146V)

A

```
                1         10        20        30
L1    5'TGTATTGATTCACTTGAAGTACGAAAAAAA 3'
L1S2  TGTATTGATTCACTTGAACCCCCAAAAAAA

L2    TAGTCGTTAATCAATGAAACGCGAAAGATA
L2S2  TAGTCGTTAATCAATGAACCCCCAAAGATA

L3    GCTTTTGTTTCATTGAAAATACGAAAAACA
L3S2  GCTTTTGTTTCATTGAAACCCCCAAAAACA

R1    TGAAGCGGCGCACGAAAAACGCGAAAGCGT
R1S2  TGAAGCGGCGCACGAAAACCCCCAAAGCGT

R2    GAAAGCGTTTCACGATAAATGCGAAAACTT
R2S2  GAAAGCGTTTCACGATAACCCCCAAAACTT

R3    ACATCTGTTTCATTTGAAGCGCGAAAGCTA
R3S2  ACATCTGTTTCATTTGAACCCCCAAAGCTA
```

B

Att DNA	MuA (‡K_d, M)	Mu A(R146V)(K_d, M)
R2-(wt)	3.3 x 10^{-8}	1.9 x 10^{-6}
R2^{S2}	2.4 x 10^{-7}	2.4 x 10^{-7}
R1^{S2}	NT	2.7 x 10^{-7}
R3^{S2}	NT	3.1 x 10^{-7}
L1^{S2}	2.6 x 10^{-7}	2.2 x 10^{-7}
L2^{S2}	NT	3.2 x 10^{-7}
L3^{S2}	NT	2.2 x 10^{-7}

(**A**) Sequences of wild-type and *att*^{S2} sites. Nucleotide numbers refer to L1, 1 being the first nucleotide at the left end of the Mu genome. Sites are aligned based on consensus sequences (29,30). Only one DNA strand is shown. Essential G residues (or their complementary bases) are indicated in bold face. The 'S2' region is boxed. (**B**) Dissociation constants for MuA and MuA(R146V) binding to wild-type and *att*^{S2} sites. ‡ Dissociation constants (K_d) were determined by gel retardation as decribed in Materials and Methods. The values reported are from single experiments. The affinities of MuA(R146V) for wild-type R1, L1 and L3 were similar to that shown for R2 (within a factor of ~2). wt, wild-type; NT, not tested.

(~1–2 × 10^{-6} M) were obtained for the binding of MuA(R146V) to other individual wild-type *att* sites as well (R1, L1 and L3; data not shown). By contrast, the affinity of MuA(R146V) for *att*R2^{S2} relative to *att*R2 was enhanced ~8-fold. However, it should be noted that MuA(R146V) was a weaker DNA-binding protein than MuA, as revealed by the difference in the affinities (~7-fold) of the two proteins for their respective cognate sites. Although initial experiments had suggested possible reverse discrimination by MuA against *att*R2^{S2}, i.e. lower affinity for *att*R2^{S2} compared to MuA(R146V) (Fig. 3B), the binding analysis with the pure target sites (Table 1) revealed that this was not the case [K_d values for *att*R2^{S2} were nearly the same for MuA and MuA(R146V)]. Not withstanding this outcome, the large decrease in the binding of MuA(R146V) to *att*R2 combined with its significant preference for *att*R2^{S2} over *att*R2 permitted us to target MuA subunits to specific *att* sites on Mu DNA.

Functionality of *att*R2^{S2} in DNA transposition

To test whether the MuA(R146V)–R2^{S2} DNA partnership was proficient in transposition, we monitored Mu end cleavage on an R1–R2^{S2} substrate under DMSO assay conditions (Fig. 5; 11). We had shown earlier that under these conditions a cleavage proficient configuration is one where the active [with respect to a triad (DDE) of catalytic residues] MuA subunit is positioned at R2 (Fig. 5A; 11). Positioning of an active (DDE$^+$) subunit at R2 and an inactive (DDE$^-$) subunit at R1 on the R1–R2^{S2} substrate (S), was achieved as follows. The substrate was first incubated with MuA(R146V) at a molar ratio of ~1:10 of substrate to protein, in order to saturate the R2^{S2} sites, and thus block them from occupancy by a partner lacking the R146V mutation. The DDE$^-$ mutant MuA(E392A) was added subsequently at a similar DNA to protein ratio (Fig. 5B). The 11 nt cleavage product (CL) could be readily detected on a denaturing polyacrylamide gel. Control reactions showed that wild-type MuA yielded strand cleavage at a DNA to protein molar ratio of ~1:20 (lane 2), while MuA(R146V) (lane 5) was inactive at the same ratio [MuA(R146V) was also inactive on a wild-type R1–R2 substrate; data not shown]. This is to be expected, since wild-type MuA can bind both R1 and R2^{S2} sites, while MuA(R146V) shows specificity for R2^{S2} (Table 1). Also as expected, the catalytically inactive variant MuA(E392A) (lane 4) did not yield a cleavage product. When MuA(R146V) was paired with MuA(E392A), however, strand cleavage occurred (lane 3). The level of cleavage was ~80% of that seen in the MuA reaction (lane 2). The lack of cleavage with MuA(R146V) alone (lane 5) certifies that the cleavage obtained with the protein mixture could not have resulted from fortuitous association of MuA(R146V) with R1. This result is therefore consistent with the arrangement of the MuA subunits deduced from subsite complementation assays using wild-type DNA and protein substrates (11). Thus, MuA(R146V) can not only discriminate between wild-type and *att*^{S2} sites, but also participate in the chemical step of strand cleavage.

We have now extended the above analysis to supercoiled DNA substrates by substituting each of the six wild-type *att* sites with *att*^{S2} sites. Using a mixture of MuA and MuA(R146V) we have analyzed DNA strand cleavage and strand transfer executed by the oriented tetramers (37). The results have established the utility of MuA(R146V) as an altered specificity variant in elucidating the functional arrangement of MuA subunits within its normal tetrameric configuration. Furthermore, they have led to the unexpected revelation that the catalytic contributions of the individual MuA subunits can be different in native versus artificial substrates (37).

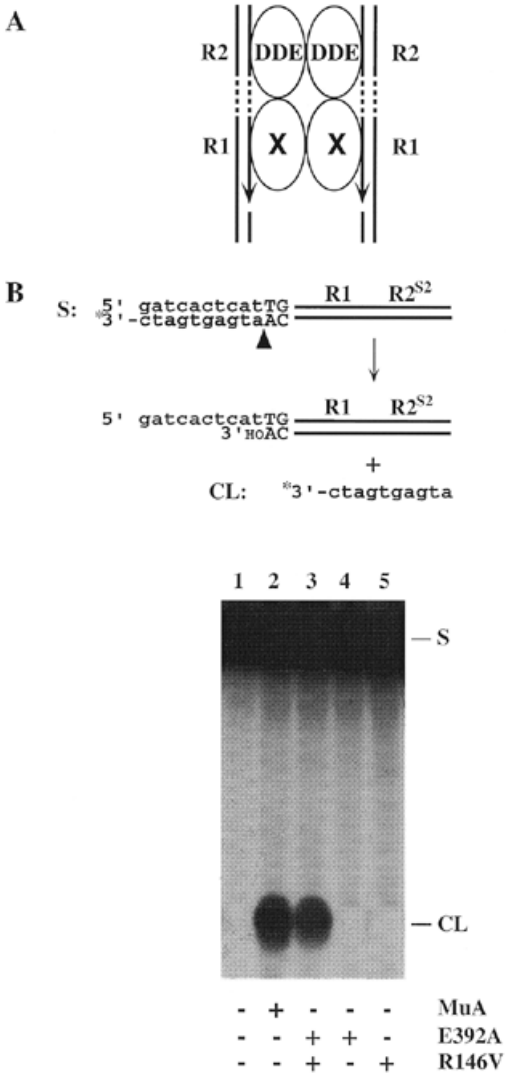

Figure 5. MuA(R146V)–*att*R2^{S2} association is functional in transposition *in vitro*. (**A**) Position of 'DDE' domains during DNA cleavage and strand transfer derived from complementation experiments performed using linear R1–R2 substrates (11). DDE$^-$ subunits are indicated by X, and DDE$^+$ subunits by DDE. The subunits were loaded separately on R1 or R2 'subsites' that were brought together by complementary base-pairing (dotted lines). The R1–R2/R1–R2 synapse is presumed to mimic a normal *att*L/*att*R synapse. The arrowheads indicate the 3′ OH groups generated upon cleavage. (**B**) An aliquot of 0.2 pmol of the substrate (S) (obtained by hybridizing two deoxyoligonucleotides that span the R1 and R2^{S2} sequences, and labeling the strand that undergoes cleavage with [α-^{32}P]cordycepin phosphate at the 3′ end; indicated by *) was incubated with 4 pmol each of wild-type MuA (lane 2), MuA(E392A) (lane 4), MuA(R146V) (lane 5), or a complementing mixture of 2 pmol of MuA(R146V) added first, followed by 2 pmol of MuA(E392A) (lane 3) under DMSO reaction conditions (see Materials and Methods; reaction volume was 20 µl). Reaction products were electrophoresed in a 12% denaturing acrylamide gel, and detected by autoradiography. Strand breakage (arrowhead) produces a labeled product (CL) that is 11 nt long.

CONCLUSIONS

We have used an *in vitro* selection scheme to isolate a transpositionally functional variant of MuA with an altered specificity for *att* DNA. The method we have used should have general application for the isolation of altered specificity variants of DNA-binding proteins, even in the absence of structural information on the nature of the DNA–protein contacts. A potential limitation of the method is that if the assay is done with an inappropriate substitution of a critical residue involved in DNA contact, the *in vitro* selection may not yield an altered binding site from the pool of quasi-randomized cognate sequences. This problem could, in principle, be circumvented by performing the selection with multiple substitutions, and at more than one amino acid position. Furthermore, as has been demonstrated here for MuA, it is important to establish the functional correspondence between the mutant protein and the altered DNA site before probing the productive configuration of a DNA–protein assembly.

The recently obtained NMR structure of the Iβ domain of MuA (17) accommodates our results quite nicely, even though the structure was not available to us until after our studies had been completed. Our rationale for the choice of MuA(R146V) for

selection of new DNA binding specificity has been vindicated by the ability of MuA(R146V) to discriminate between wild-type and *att*^{S2} sites on linear substrates, as well as to participate in the chemical step of strand cleavage (Fig. 5B). We have extended these findings successfully to supercoiled substrates as well, allowing important new insights into the mechanism of transposition (37).

Each Mu *att* site is relatively large (~25 bp), and two separate HTH motifs in the Iβ and Iγ subdomains have now been shown to bind the distal and proximal halves of the *att* site, respectively (16,17). The low recovery of non-*att* binding point mutants in domain Iβγ (Fig. 2) is consistent with the notion that the abrogation of a single DNA contact may be insufficient to significantly diminish DNA-binding, since the overall contribution to binding is provided by two DNA-binding regions contacting DNA over two turns of the DNA helix. That the two positions at which DNA-binding mutants were recovered in our studies (Phe131 and Arg146) mapped to the HTH domain in Iβ is consistent with the higher affinity of this subdomain relative to Iγ (17). Indeed, changes in *att* DNA that restored binding of MuA(R146V) were recovered in the distal half of the *att* site (Fig. 3), consistent with the possibility that Arg146 may make a specific contact within this region (17). However, the recovered changes were not located at one or two specific base pairs, but were spread over several base pairs (Fig. 4). It is possible that the Arg146Val substitution does not just abolish a specific contact, but rather also diminishes the strength of other contacts. It is quite likely that the specific recognition of the *att* site by MuA is achieved by the additivity of multiple protein–DNA contacts. The observed properties of MuA(R146V) can be accounted for by a loss of these co-operative interactions as a result of mutating Arg146. Restoration of binding may then require multiple changes in the DNA in order to re-establish co-operativity. A general increase in GC content of the DNA in the S2 region (Fig. 4) might lead to suppression of the binding defect of MuA(R146V) by perturbing the local geometry (note that S2 is flanked on both sides by AT-tracts; Table 1A) (38).

An interesting finding from this study is the occurrence in the DNA-binding region (domain I) of mutations that affect assembly of the MuA tetramer. Such assembly-defective mutants have been mapped earlier to domain II as well as III (18,39,40). Assuming that these defects are due to direct effects on inter-subunit interactions, all three domains of MuA must participate in inter-subunit protein contacts within the MuA tetramer. Interestingly, footprinting studies on linear *att* DNA have determined that MuA contacts only one face of the DNA (30,33). If protein–protein interactions occur through the DNA binding domain, the simplest picture of the transposition complex containing the MuA tetramer consistent with the footprinting results would be one in which the protein core occupies the inside, and the DNA is on the outside.

In applying the strategy described here to other DNA–protein interaction systems, modifications can be incorporated to improve its efficiency. Inclusion of a subtractive selection procedure (using the wild-type protein to bind and exclude the normal DNA target) between successive enrichment cycles can, in principle, enhance the discriminatory capacity of a variant protein for an altered DNA site. Furthermore, for a protein that interacts with a relatively large DNA segment using more than one peptide domain, the binding affinity of a specificity variant of the protein may be increased by step-wise application of the selection/enrichment process separately to each peptide domain and its cognate sequence. Given the intrinsic merits of an altered specificity variant in dissecting reaction mechanisms within multiprotein–nucleic acid assemblies (27,28), the MuA analysis reported here is likely to provide another useful paradigm for application in related systems.

ACKNOWLEDGEMENTS

We thank Jin Zhong for isolation of a subset of domain Iγ mutants, Adam Toguchi for help with using the SOPMA package, Susana Mariconda for technical assistance, and M.Jayaram and Ian Grainge for critically reading the manuscript. This work was supported by a grant from the National Institutes of Health (GM33247).

REFERENCES

1 Alberts,B. (1998) *Cell*, **92**, 291–294.
2 Bukau,B. and Horwich,A.L. (1998) *Cell*, **92**, 351–366.
3 Chaconas,G. Lavoie,B.D. and Watson,M.A. (1996) *Curr. Biol.*, **6**, 817–820.
4 Yang,J.-Y., Jayaram,M. and Harshey,R.M. (1995) *Genes Dev.*, **9**, 2545–2555.
5 Lavoie,B.D., Chan,B.S., Allison,R.G. and Chaconas,G. (1991) *EMBO J.*, **10**, 3051–3059.
6 Yang,J.-Y., Kim,K., Jayaram,M. and Harshey,R.M. (1995) *EMBO J.*, **14**, 2374–2384.
7 Mizuuchi,M. and Mizuuchi,K. (1989) *Cell*, **58**, 399–408.
8 Baker,T.A. and Mizuuchi,K. (1992) *Genes Dev.*, **6**, 2221–2232.
9 Namgoong,S.-Y., Jayaram,M., Kim,K. and Harshey,R.M. (1994) *J. Mol. Biol.*, **238**, 514–527.
10 Savilahti,H., Rice,P. and Mizuuchi,K. (1995) *EMBO J.*, **14**, 4893–4903.
11 Yang,J.-Y., Jayaram,M. and Harshey,R.M. (1996) *Cell*, **85**, 447–455.
12 Aldaz,H., Schuster,E. and Baker,T.A. (1996) *Cell*, **85**, 257–269.
13 Savilahti,H. and Mizuuchi,K. (1996) *Cell*, **85**, 271–280.
14 Lee,J., Whang,I. and Jayaram,M. (1996) *J. Mol. Biol.*, **257**, 532–549.
15 Lee,J. and Jayaram,M. (1997) *Genes Dev.*, **11**, 2438–2447.
16 Clubb,R.T., Schumacher,S., Mizuuchi,K., Gronenborn,A.M. and Clore,G.M. (1997) *J. Mol. Biol.*, **273**, 19–25.
17 Schumaker,S., Clubb,R.T., Cai,M., Mizuuchi,K., Gronenborn,A.M. and Clore,G.M. (1997) *EMBO J.*, **16**, 7532–7541.
18 Kim,K., Namgoong,S.-Y., Jayaram,M. and Harshey,R.M. (1995) *J. Biol. Chem.*, **270**, 1472–1479.
19 Nakayama,C., Teplow,D.B. and Harshey,R.M. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 1809–1813.
20 Kuo,C.-F., Zou,A., Jayaram,M., Getzoff,E. and Harshey,R.M. (1991) *EMBO J.*, **10**, 1585–1591.
21 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
22 Sarkar,G. and Sommer,S.S. (1990) *Biotechniques*, **8**, 404–407.
23 Surette,M.G., Buch,S.J. and Chaconas,G. (1987) *Cell*, **49**, 253–262.
24 Fried,M.G. (1989) *Electrophoresis*, **10**, 366–376.
25 Geourjon,C. and Deleage,G. (1994) *Protein Engng.*, **7**, 157–164.
26 Geourjon,C. and Deleage,G. (1995) *Comput. Applic. Biosci.*, **11**, 681–684.
27 Ebright,R.H. (1991) *Methods Enzymol.*, **208**, 620–640.
28 Rebar,E.J., Greisman,H.A. and Pabo,C.O. (1996) *Methods Enzymol.*, **267**, 129–149.
29 Craigie,R., Mizuuchi,M. and Mizuuchi,K. (1984) *Cell*, **39**, 387–394.
30 Zou,A., Leung,P.C. and Harshey,R.M. (1991) *J. Biol. Chem.*, **266**, 20476–20482.
31 Leung,P.C., Teplow,D.B. and Harshey,R.M. (1989) *Nature*, **338**, 656–658.
32 Kim,K. and Harshey,R.M. (1995) *Nucleic Acids Res.*, **23**, 3937–3943.
33 Ding,Z.-M., Harshey,R.M. and Hurley,L.H. (1993) *Nucleic Acids Res.*, **21**, 4281–4287.
34 Harshey,R.M., Getzoff,E.D., Baldwin,D.L., Miller,J.L. and Chaconas,G. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 7676–7680.
35 Ebright,R.H., Cossart,P., Gicquel-Sanzey,B. and Beckwith,J. (1984) *Proc. Natl Acad. Sci. USA*, **81**, 7274–7278.
36 Ebright,R.H., Kolb,A., Buc,H.B., Kunkel,T.A., Krakow,J.S. and Beckwith,J. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 6083–6087.
37 Namgoong,S.-Y. and Harshey,R.M. (1998) *EMBO J.*, **17**, 3775–3785.
38 Hagerman,P.J. (1990) *Annu. Rev. Biochem.*, **59**, 755–781.
39 Baker,T.A., Mizuuchi,M., Savilahti,H. and Mizuuchi,K. (1993) *Cell*, **74**, 723–733.
40 Namgoong,S.-Y., Kim,K., Saxena,P., Yang,J.-Y., Jayaram,M., Giedroc,D.P. and Harshey,R.M. (1998) *J. Mol. Biol.*, **275**, 221–232.