Sequence and positional requirements for DNA sites in a Mu transpososome

Ilana Goldhaber-Gordon*, Michael H. Early*, Matthew K. Gray, and Tania A. Baker*\textsuperscript{*}

*Department of Biology and \textsuperscript{*}Howard Hughes Medical Institute
Massachusetts Institute of Technology
77 Massachusetts Ave, 68-523
Cambridge, MA 02139

Running title: DNA sites in a Mu transpososome
Summary

Transposition of bacteriophage Mu uses two DNA cleavage sites and six transposase recognition sites, with each recognition site divided into two half sites. The recognition sites can activate transposition of non-Mu DNA sequences, if a complete set of Mu sequences is not available. We have analyzed 18 sequences from a non-Mu DNA molecule, selected in a functional assay for the ability to be transposed by MuA transposase. These sequences are remarkably diverse. Nonetheless, when viewed as a group they resemble a Mu DNA end, with a cleavage site and a single recognition site. Analysis of these “pseudo-Mu ends” indicates that most positions in the cleavage and recognition sites contribute sequence-specific information that helps drive transposition, though only the strongest contributors are apparent from mutagenesis data. The sequence analysis also suggests variability in the alignment of recognition half-sites. Transposition assays of specifically designed DNA substrates support the conclusion that the transposition machinery is flexible enough to permit variability in half-site spacing, and also perhaps variability in the placement of the recognition site with respect to the cleavage site. This variability causes only local perturbations in the protein-DNA complex, as indicated by experiments in which altered and unaltered DNA substrates are paired.
Introduction

Some of the most fundamental of biological processes occur within nucleoprotein complexes. These processes include recombination, replication, transcription, and RNA splicing. The nucleoprotein complexes are often large and elaborate, containing features that permit sophisticated regulation. As a result, dissecting the chemistry of interactions within a complex is a challenging task, but one that is critical to understanding biological function.

Transposition of bacteriophage Mu, like that of other transposons, occurs within complexes called transpososomes. Transpososomes mediate at least two sequential chemical reactions, on pathway towards transferring the transposon DNA from one site to another. The two reactions are: (i) DNA cleavage, in which a nick is introduced precisely at the end of the transposon, on the 3’ strand, and (ii) DNA strand transfer, in which the nicked strand is joined to a separate DNA molecule called the target (1,2). The reaction sites on the transposon DNA, or “donor DNA”, are defined by specific DNA sequences, but Mu target sites are not very sequence-specific (3).

Transpososomes contain multiple subunits of a transposase protein, bound to DNA sequences from both of the transposon's ends (1). These protein-DNA complexes are also called "synaptic complexes", because they bring together the two ends of the transposon DNA. The phage Mu transposase, MuA, is monomeric in solution, but forms a tetramer upon binding to specific DNA recognition sites near the transposon ends (4-6).

Each end of the Mu transposon has three MuA recognition sites: R1, R2 and R3 on the right end and L1, L2 and L3 on the left, not all of which are essential for transposition (5,7-9). Five of the six sites are related by a degenerate 22 base-pair consensus sequence. The exception is L2, which is missing roughly half of the 22 base-pair sequence. In the transpososome tetramer, three of the MuA subunits are bound specifically to the R1, L1 and R2 sites. The fourth subunit probably binds the L2 site, but this contact is not as stable as that of the other three subunits and can be competed away by excess DNA (5,10). Additional subunits, not essential for catalysis, may be loosely associated with the R3 and L3 sites. In vitro, MuA can transpose "donor fragment" molecules—short DNA molecules carrying only the R2, R1 and cleavage sites (Fig. 1a) (11). These fragments support assembly of MuA tetramers, which synapse pairs of fragments.

MuA specifically binds its DNA recognition sites through its N-terminal domain (Fig. 1b) (12). This proteolytically defined domain contains three independently folding subdomains, Iα, Iβ, and Iγ, each of which has a helix-turn-helix (HTH) motif (13-15). Domain Iα binds to an internal enhancer-like site that is not essential in vitro. Domains Iβ and Iγ interact with the recognition sites and are essential. Each DNA recognition site can be thought of as two half-sites: the outer half-site is bound by domain Iγ and the
inner half by Iβ. This bipartite DNA binding has at least two consequences. It permits the recognition sequences to be unusually long, and therefore rare, and it allows for nanomolar binding affinities to the combined site (15). It is also possible that the bipartite structure achieves some additional, as yet undiscovered function.

The cleaved phosphodiester bond at the end of the transposon is separated from the outermost recognition site (R1 or L1) by five-base-pairs, the last two base-pairs being 5’CA (often described by the 5’TG complement). Thus, Mu has separate cleavage sites and recognition sites. Whereas the recognition sites are bound by domains Iβ and Iγ of MuA, the cleavage sites are contacted by the protein's active site which is in domain IIA (16).

Within the synaptic complex, individual MuA subunits bind a cleavage site and a recognition site on opposite DNA ends. Specifically, the MuA subunit bound to the R1 site engages the left-end cleavage site, and the subunit bound at L1 engages the right-end cleavage site (Fig. 1b) (17-20). This intertwined structure suggests that the various MuA domains and subunits function interdependently in a transpososome (20).

There are many unanswered questions about interactions between the protein and DNA components of the transpososome. For example, there is no direct structural information about the geometry of the protein-DNA interactions. On another level, although considerable mutagenesis has been performed on the MuA recognition sites (21-23), the sites are long and our understanding of sequence requirements is incomplete.

In the accompanying paper, we describe transposition of DNA molecules that have no defined MuA recognition sites (24). This unconventional activity depends on MuA recognition sites provided in trans on a donor fragment or a donor plasmid, so that "hybrid" synaptic complexes can assemble between the Mu DNA and a site on the non-Mu DNA (Fig. 1c). Here, we describe the sequences of 18 sites on φX174 RF1 that were transposed by MuA. These 18 sequences are extremely diverse. Nonetheless, looking at the most commonly represented nucleotide at each position, we find that as a group the 18 sequences resemble a Mu end with a “TG” cleavage site and a single MuA recognition site. We deduce from these 18 sequences the fundamental sequence requirements for transposition. Additionally, we draw conclusions about the arrangement of the individual DNA sites that contribute to the transpososome. Our conclusions, initially based on sequence analysis, are supported and extended through analysis of the transposition activity of defined DNA fragments. Our data point to localized regions of flexibility in the Mu transpososome.
Experimental Procedures

Proteins and DNA. MuA and MuB were prepared as described (25,26). φX174 RF1 was purchased from New England Biolabs. Oligonucleotides (fragment donors or PCR primers) were purchased either from MIT/HHMI biopolymers lab or from GeneLink, and fragment donors were purified by denaturing PAGE. Donor fragment sequences were as follows: **unjoinable Mu fragment:** gtttttcgattatcgtgaaacgctttcggcggttc/tcatttctggttcggcggttc/gattatcgtgaaacgctttcggcggtt

**wild-type:** gtttttcgattatcgtgaaacgctttcggcggttc/tcatttctggttcggcggttc/gattatcgtgaaacgctttcggcggttt

**fragment HSI:** gtttttcgattatcgtgaaacgctttcggcggttc/tcatttctggttcggcggttc/gattatcgtgaaacgctttcggcggttt

**fragment HSD:** gtttttcgattatcgtgaaacgctttcggcggttc/tcatttctggttcggcggttc/gattatcgtgaaacgctttcggcggttt

**fragment HSN:** gtttttcgattatcgtgaaacgctttcggcggttc/tcatttctggttcggcggttc/gattatcgtgaaacgctttcggcggttt

**fragment 17a:** cgttttcgattatcgtgaaacgctttcggcggttc/tcatttctggttcggcggttc/gattatcgtgaaacgctttcggcggttt

**fragment 17b:** cgttttcgattatcgtgaaacgctttcggcggttc/tcatttctggttcggcggttc/gattatcgtgaaacgctttcggcggttt

**fragment 17c:** cgttttcgattatcgtgaaacgctttcggcggttc/tcatttctggttcggcggttc/gattatcgtgaaacgctttcggcggttt

**fragment 17d:** cgttttcgattatcgtgaaacgctttcggcggttc/tcatttctggttcggcggttc/gattatcgtgaaacgctttcggcggttt

**fragment CSI:** gtttttcgattatcgtgaaacgctttcggcggttc/tcatttctggttcggcggttc/gattatcgtgaaacgctttcggcggttt

**fragment CSD:** gtttttcgattatcgtgaaacgctttcggcggttc/tcatttctggttcggcggttc/gattatcgtgaaacgctttcggcggttt

Isolation of pseudo-ends. The short target had the sequence: 5’ cagtgctggaaagataccggcaggccgcgcacccggccgccgattatatgagcagcagccagacgccccggggcggggg, annealed to a precise complement. This sequence was derived from an insertional hotspot in pBR322 (3), but was modified to remove any 3 base-pair sequence that resembled parts of a MuA recognition site. This short target [240 nM] was incubated with 15 ng/µl MuB and 0.2 mM ATP at 30°C for 20 minutes, under standard transposition buffer conditions (24). If MuB was not included in this step, we saw no transposition into the short target (data not shown). ATP-γ-S was then added to 2 mM, and the reaction was incubated for another 20 minutes at 30°C. If ATP-γ-S was not added, transposition into the short target was reduced (data not shown). We then added φX174 RF1 to 10 ng/µl, unjoinable fragment to 5 nM, and MuA to 50 nM, and continued incubating for two hours. Reactions were stopped with SDS/EDTA, phenol/chloroform extracted, and ethanol precipitated.

For the two sequences obtained without cloning, ~2.5 µg of DNA were used in a 50 µl thermal cycling reaction, using a "touchdown" protocol. Annealing temperature was reduced by 1° every 2 cycles for 26 cycles, starting at 60°C, and then amplification continued 12 more cycles with annealing
temperature 47°C. The primers were ccagtgcgtggaagatatc (anneals to the short target) and either
gtgaatgcacaagcc (to obtain sequence 17) or ggctgtggtgcattctac (to obtain sequence 18). The products
were run on a 1.5% MetaPhor (FMC Bioproducts) preparative agarose gel. In each case a ~700 bp band
was excised and purified with a Qiagen gel extraction kit. It was then amplified in a second round of PCR
(20 cycles, 49°C annealing temp), and sequenced with an Amersham Pharmacia cycle sequencing kit.

For the sequences obtained by cloning, the transposition products were purified away from
unused fragments (both the unjoinable Mu fragment and the target fragment) on a Superose 6 HR column
before being subjected to the "touchdown" PCR. This time, annealing temperature was decreased by
0.5°C every cycle for the first 20 cycles, followed by 12 cycles at annealing temperature of 51°C. Products
were purified with a Qiagen PCR Purification kit. The products were then simultaneously digested with
EcoRI and an appropriate second enzyme, cloned into a pUC19 vector, and transformed into DH5α cells.
Clones were sequenced at the MIT/HHMI biopolymers sequencing facility.

The target primer contained an EcoRI site: tagaaatgcataagcag tcgcag. The φX174 primer
was one of the following, containing one of the following restriction sites: AatII sites:
tgtgtgaccttgctctcc or cgtacctggaagactcataagtc; PstI sites: gattggtgtatccacagcg or
gaatgctgatacgctcgtgg; BamHI sites: tagatccatgtcttcctgacttgcgt or tagatccagatggaagagactcagc or
tagatccatgtcctgacttgcgtgc or tagatccagatggaagactcagc. Sequences 1, 2 and 3 (Fig. 2) were
obtained from faulty priming of the φX174 primer to an incorrect sequence on φX174 that happened to be
similar to the primer sequence and happened to be near the transposition joint.

Transposition reactions. Unless otherwise indicated, transposition reactions were done essentially
as described (24), except that protein and DNA concentrations were as follows: 200 nM donor fragment,
500 ng φX174, 200 ng MuB. MuA concentrations were as indicated in Figure legends. These conditions
were optimized for transposition of fragments carrying pseudo-MuA-recognition sites.

Results

Individual pseudo-ends have identifiable cleavage sites.

Using MuA transposition assays, we identified 18 non-Mu sequences that could function
analogously to the Mu DNA end sequences. These 18 "pseudo-Mu ends" were located on φX174 RFI
DNA, but the accompanying paper suggests that any large DNA molecule could have yielded a similar
group of pseudo-ends (24). In addition to the large (5386 bps) φX174 DNA, the transposition assays used
to isolate pseudo-Mu ends required two types of DNA fragments (Fig. 1c). An 80 base-pair fragment
served as a transposition target. The sequence of this fragment was based on a preferred target-site from
plasmid pBR322 (3), but the original plasmid sequence was modified slightly to remove any minor
sequence similarity to Mu DNA. We also needed a source of native MuA recognition sites, to permit
transposition of the non-Mu DNA. For this purpose we used a DNA fragment that contained two MuA recognition sites (R2 and R1) but lacked the pA3' from the cleavage site, and therefore could not participate in the covalent chemistry of transposition. We show in the accompanying paper that this fragment, called an "unjoinable fragment", efficiently activates non-Mu transposition (24).

The 80-mer target fragment was incubated with MuB, a Mu protein which controls selection of transposition targets (27). Subsequently, MuA transposase, the unjoinable Mu DNA fragment, and φX174 RFI DNA were added. DNA products in which the φX174 DNA became covalently joined to the target fragment were then amplified with a primer to the target and one of several primers designed to arbitrary locations on φX174 (Fig. 1c). This method yielded two PCR products that were sufficiently abundant to be directly sequenced. Sixteen additional sites were identified after cloning and sequencing.

The sequences obtained are listed in Fig. 2, beginning with the 5' complement of the nucleotide that was joined to target DNA. In sixteen cases, that complement was a "T", and in the remaining two cases it was an "A". Furthermore, in 8 cases (out of the 18) the following nucleotide was a “G”. Thus, these sequences revealed that functional pseudo-Mu DNA ends have cleavage sites that are close in sequence to the 5'TG of the natural cleavage sites.

Several additional lines of evidence confirm that these sequences are indeed pseudo-Mu ends; that is, that despite low similarity to the ends of the Mu genome, they served as transposition donor sites (not all data shown). (i) In control experiments, a 5' label on the target 80-mer revealed some target fragment joined to φX174 DNA. (ii) To be functional, the 80mer target needed to be pre-incubated with MuB. Since DNA molecules bound by MuB are known to be preferred by MuA as transposition targets, these data support the claim that φX174 DNA was transposing into the 80-mer, rather than vice versa. (iii) For the two sequences analyzed directly (without cloning), we could clearly read on the sequencing gel roughly 100 bp of φX174 sequence. Then the φX174 sequence ended abruptly, and we could read multiple sequences from the 80-mer target overlapping with each other on the sequencing gel. This sequence pattern suggests a family of transposition products, corresponding to a single site on φX174 recombining with multiple sites on the small target DNA. (iv) To control for faulty priming of the target primer on φX174 DNA, we rejected clones that did not include target sequence beyond the target primer. (v) We also cloned and sequenced DNA that had been treated as in our experiment, except that MuA and B were omitted from the transposition reactions. We were reassured that these controls did not yield sequences that we would have considered to be transposition products. (vi) Finally, as described below, further sequence analysis affirms that these 18 sequences, as a group, resemble a Mu DNA end.

**The pseudo-Mu ends have weak similarity to a MuA recognition site.**
A striking feature of the sequences in Fig. 2 is that most of them do not strongly resemble a MuA recognition site. All of the 16 pseudo-ends isolated by cloning (rather than direct sequencing) were located within 400 base-pairs of the PCR primer that isolated them. All but three were located within 250 base-pairs of their respective primers. This close proximity to an arbitrary set of primers suggested that use of a different set of PCR primers would have resulted in isolation of a different group of sequences, indicating minimal requirements for DNA to function as a transposition substrate in these experiments.

We developed a scoring system to address the resemblance of any sequence to the consensus sequence derived from native MuA recognition sites (right side of Fig. 2). Sites were scored 1 point for a match to the consensus at a position where a nucleotide is specified, and half a point if only purine or pyrimidine is specified. For analysis of a single 22 base-pair recognition site, without the cleavage site, the highest possible score is 17.5.

For the 18 pseudo-ends, the mean score was 6.50 points for the putative outer-site (i.e. nucleotide positions 6-27, equivalent to R1 or L1) (Table 1, row 1). This is indeed low, compared to, for example, the 13.5 points found for the natural R1 site. However, the mean score for all 10,772 possible sites on the two strands of φX174 is 5.41, and the standard deviation for randomly chosen groups of 18 such sites is 0.42. Thus, the pseudo-ends, as a group, are 2.6 standard deviations better than a typical group of 18 (Table 1, last column; \(\frac{6.5-5.41}{0.41} = 2.6\)). A random group of 18 sequences will score this well 0.5% of the time.

We next conducted the same analysis for the putative inner site (i.e. nucleotide positions 28-49, equivalent to R2 or L2). The results of this calculation gave a mean score of 4.61 (Table 1, row 10). For the group of 18, this is 1.9 standard deviations lower than the expected score of 5.41 for a randomly chosen group. The lower-than-mean score is puzzling, but perhaps not significant—a random group should score this poorly 3% of the time.

Analysis of pseudo-ends suggests alternative alignments.

The poor scores for the 18 pseudo-ends could be due to variability in the correct sequence alignment. For example, there might be flexibility in the placement of the recognition sites relative to the cleavage site. In addition, recall that the recognition sites consist of two half-sites, bound respectively by the Iγ and Iβ subdomains of MuA (Fig. 1b). Spacing between recognition half-sites could also vary. It would be interesting if the pseudo-end sequences showed variability in their alignments, because this variability might point to areas of flexibility within the Mu transpososome.

With this in mind, we reanalyzed the pseudo-end sequences, allowing variability in the spacing between pairs of sites. The computer program (a perlscript) used for this analysis can be made available on request. The point system used by the program to score sites was the same as described above.
Allowing variability between the two halves of the outer recognition site greatly improved the point score of the pseudo-end group. With flexibility of ±1 nucleotide between half sites, the group of 18 pseudo ends scored more than 4 standard deviations better than random groups of 18 (Table 1, row 6). On the order of $10^{-4}\%$ of random groups of 18 sites would score this well. Note that if MuA could not permit variability in the spacing of the two half-sites, an analysis that did permit variability would worsen the sites' mean score compared to a random group. If we accept the hypothesis that these 18 sequences were not randomly selected, then the fact that allowing flexibility improved the group’s score from 2.6 standard deviations (just significant) to 4.2 (highly significant) strongly suggests that MuA bound some of these sites in an unorthodox register.

In contrast, allowing variability in the spacing between cleavage site and outer recognition site did not improve the relative mean score for the pseudo-end population (Table 1, rows 2&3). On the other hand, this variability did not significantly decrease the score either. We will return to this point below.

Analyzing the φX174 sequences adjacent or nearly adjacent to those shown in Fig. 2 failed to reveal a set of inner recognition sites (equivalent to R2) (Table 1, rows 10-13, and data not shown). Note that we would not have detected a set of inner sites located at arbitrary distances from the outer sites. Most likely, in our experiments the transpososomes did contain four DNA-bound MuA subunits – two bound to φX174 and two bound to a Mu fragment – but the second φX174-bound subunit relied on cues from the other three subunits to position it in the transpososome.

**Transposition of diverse sequences**

It is worth noting that some individual pseudo-ends continued to score poorly, even allowing variable alignments (data not shown). For example, with variability of ±1 nucleotide between half-sites, sequences 2 and 3 continued to score only 3.5 and 4 points, respectively. Several others scored 5 or 6 points. Also note that not a single nucleotide was absolutely required among the 18 pseudo-ends; even at position one, two pseudo-ends contained an A rather than a T. We suggest that almost any sequence can transpose at some frequency, but that similarity to a Mu end determines the transposition frequency.

To test this assertion, we constructed precleaved donor fragments designed to not resemble Mu sequences except at the cleavage site. A radiolabel on the fragments showed that they did transpose, though at just barely detectable levels after an hour incubation. By comparison, a fragment with recognition sites taken from sequence 17 in Fig. 2 produced a similar amount of transposition product in only 7 minutes. A wild-type fragment, however, produced ten times as much product in two minutes or less. (Data not shown. Fragment sequences listed in experimental procedures. The "sequence 17" fragment is used again for Fig. 4 below, and also in the accompanying paper (24).)

**Phasing flexibility is permitted between half-sites.**
Table 1 suggested that, under our experimental conditions, MuA can recognize half-sites with non-native alignments. We performed two experiments to further test this hypothesis. First, we designed precleaved donor fragments with base-pair insertions or deletions between half sites. This experiment is the conceptually obvious one, but its results are only suggestive, as explained below. Second, we manipulated the sequence of a pseudo-recognition site whose half-sites seemed to be one base-pair out-of-phase with each other. This second experiment provided more conclusive evidence in favor of flexibility.

In both sets of experiments, each donor fragment was assayed at multiple MuA concentrations. This was done because high MuA concentrations were necessary to see activity for the less-active fragments, but the high MuA made quantitation difficult for the more-active fragments. Also in both sets of experiments, transposition assays were done as single time-points rather than time courses. As a result, we cannot say whether variability in transposition efficiencies reflected varying rate constants or varying end-points.

Fragment HSI ("half-site insertion"; Fig. 3a&b) contained an added "A/T" at roughly the boundary between the halves of each of R1 and R2. This fragment transposed at approximately 70% the efficiency of wild-type DNA (Fig. 3c, lanes 4-6). Fragment HSD (half-site deletion) had an A/T removed at each half-site boundary (Fig. 3a&b), and it transposed at ~50% the efficiency of a wild-type fragment (Fig. 3c). These results showed that fragments with altered half-site spacing can transpose, albeit with slightly reduced efficiencies. However, these results do not indicate whether MuA recognized fragments HSI and HSD with a rigid or a flexible alignment. Did the protein bind the optimal sequence, perhaps stretching or pinching to accommodate the altered spacing? Or did the protein recognize the rigid alignment, sacrificing sequence-specific contacts at the inner half of R1 and all of R2? As shown in detail in Fig. 3a, because the recognition sites contain stretches of repeating "A" a single base-pair insertion or deletion does not entirely remove specific contacts from the inner half-site.

We designed another fragment (fragment HSN -- "half-site null"), in which the inner half of the R1 site was exchanged for an "anti-consensus sequence", designed to contain the nucleotides most underrepresented at each position in the consensus (15) (Fig. 3a&b). This fragment transposed at only ~20% the efficiency of the wild type DNA – considerably worse than either fragment with altered spacing (Fig. 3c). The difference in efficiency between HSN and either HSI or HSD suggests that MuA did bind flexibly to fragments HSI and HSD. But we hasten to add that this experiment is not fully conclusive, since we cannot evaluate the stringency of the substitutions in fragment HSN.

Note that this experiment did conclusively show that, for the strand transfer reaction, the natural spacing is the best of the three spacings tested here, and is probably the best possible spacing altogether. This is true whether MuA bound fragments HSI and HSD with a rigid or a flexible alignment. If the latter,
the reduced transposition efficiencies of 70% for fragment HSI and 50% for HSD would be due to an energetic cost for components of the transpososome to assume a less-than ideal geometry.

We next used a pseudo recognition site to test the possibility of flexible half-site spacing. Sequence 17 in Fig. 2 scored relatively well as a MuA recognition site; 8.5 out of 17.5 possible points. However, a deletion of one nucleotide near the half-site boundary would dramatically improve the sequence match, producing a new score of 13, compared to 13.5 points for the native R1 site (Fig. 4a). Does MuA bind this particular pseudo-recognition site with its domain Iβ one nucleotide out of register, in order to maximize sequence-specific contacts? We could test this hypothesis with base-pair mutations at positions that contribute favorably to the flexible alignment but not to the rigid alignment. We constructed fragment 17a, containing tandem copies of the recognition sequence from pseudo-end 17. Fragments 17b, 17c and 17d were variants of 17a, designed to remove some of the advantage of flexible binding (Fig. 4b). Note that the cleavage site sequence on these fragments was taken from the natural Mu right end, rather than from sequence 17.

The changes made to generate fragment 17b were particularly conservative. Fragment 17b differs from the original sequence (fragment 17a) at just two positions: positions 21 and 24. Each of the two mutations improves the fragment's rigid alignment with the MuA consensus sequence, but weakens the flexible alignment. Consideration of published mutagenesis data and of the preferences seen among the sequences listed in Fig. 2 also indicates that the two changes in fragment 17b should weaken the flexible alignment but not the rigid alignment (see Table 2 and Fig. 6, below).

In five independent experiments, fragment 17a transposed at least twice as efficiently as any of the other three (Fig. 4c). These results show that the mutations in fragments 17b, c and d were at positions that, in the original sequence, had communicated sequence-specific information to the transpososome. Given that these positions already did not match the consensus sequence in a rigid alignment, the simplest explanation is that these nucleotides function in the original sequence via contacts involving flexible binding.

Although fragment 17a transposed more efficiently than its variants, it was substantially less efficient than a wild-type fragment substrate (data not shown.) This could be due in part to a penalty for changing the relative spacing of the two half-sites, as discussed above. In addition, although fragment 17a is a good match to the consensus sequence when aligned flexibly, it is still not a perfect match.

Spacing between recognition site and cleavage site is at least as stringent as between recognition half sites.

The statistical analysis in Table 1 did not indicate variability in spacing between the R1 and cleavage sites, but neither did it rule out this possibility. To directly investigate the effect of cleavage site
position on transposition, we constructed two fragments with altered R1-to-cleavage site spacing. The native Mu ends contain five nucleotides between outer recognition site and cleavage site: fragment CSI (cleavage-site insertion) contained six nucleotides and fragment CSD (cleavage-site deletion) contained four nucleotides. The fragments were precleaved, so these experiments again probed the strand transfer but not the cleavage reaction. In addition, to prevent an improper cleavage reaction this set of experiments were performed in calcium rather than magnesium. Calcium supports strand transfer but not cleavage (11).

Fragments CSI and CSD transposed at approximately 45% and 10% the efficiency of wild-type, respectively (Fig. 5a&b). Thus, our results indicate a tolerance for an improperly spaced cleavage site, though we cannot comment on the structural basis of that tolerance. Furthermore, alterations in cleavage site spacing were at least as deleterious as alterations in half-site spacing (Fig. 5a&b).

**Changes in cleavage site spacing alter the distribution of transposition products.**

Inspection of Fig. 5a reveals that fragments CSI and CSD produced an increase in single fragment transfer (SFT) products as compared to double fragment transfers (DFT). This change in product ratio is distinct from the decrease in total amount of product discussed above. We expect that both SFTs and DFTs require synapsis of two fragments, but that SFTs are produced when one of the two fragments fails to transpose. Even wild-type fragments produce SFTs, though it is not clear what causes the transposition failures that lead to wild-type SFTs. The additional SFTs caused by alterations in cleavage site spacing is likely due to architectural imperfections in the transpososome – for example, failure to engage the DNA cleavage site in the protein's active site.

Are the affects of altered spacing transmitted across the transpososome, such that if one fragment in a complex is imperfect its synapsis partner also suffers? This question interests us because it probes the degree of rigidity in the transpososome structure, by asking whether the transpososome accommodates these DNA imperfections through local versus global changes. To address the question, we performed transposition reactions containing pairwise mixtures of the cleavage-site-insertion, the half-site-insertion, and the wild-type fragments (Fig. 5c). In each case, one type of fragment was radiolabeled, and the other was unlabeled and present at ten-fold higher concentration. We found that the transposition frequencies of the labeled fragments were independent of the identity of the fragment in excess (Fig. 5c, white bars). Thus, under these conditions, the synaptic partners do not influence each other's transposition efficiencies.

Unlike the total transposition efficiency, the percentage of the total product that comprised SFTs depended only on the identity of the unlabeled, excess fragment (Fig. 5c, dark bars). For example, whenever CSI was the unlabeled partner, we saw the highest percentage SFT (~70%). Note that only radiolabeled products are visible in these experiments. Therefore, the dependence of SFTs on the
unlabeled fragment indicates that labeled and unlabeled fragments paired-up indiscriminately. Taken together, these data suggest that the alterations in sequence described here are accommodated through localized, rather than global, perturbations in transpososome structure. This experiment also shows that the structure of an active transpososome can accommodate some asymmetry between its two halves.

**Discussion**

**Functional pseudo-Mu DNA ends reveal sequence requirements for transposition.**

The Mu transpososome engages multiple DNA sites, including several 22 bp recognition sites and two separate cleavage sites (7). Defining the relevance of each nucleotide position in this array of sites is a tremendous task. Several labs have conducted mutagenesis and chemical protection assays of the Mu DNA ends, and these have gone a long way towards elucidating the sequence requirements for transposition (21-23,28,29). The present study asked the protein to select its favorite among disfavored sites, and thus could detect subtle sequence preferences that designed mutagenesis would have been unlikely to find.

We sequenced 18 sites on a non-Mu DNA molecule (φX174) that had served as pseudo-Mu end sequences. To distill the information contained in those 18 sequences into an easily comprehended picture, we used a Pictogram analysis developed by Chris Burge (30). Figure 6 includes three Pictograms, each of which represents the 18 pseudo-end sequences in a slightly different way (see figure legend for details). The results of the Pictograms generally agree with past mutagenesis and chemical protection assays, with some exceptions (Table 2). One study (22) found no effect on transposition from individual mutations at any of 11 positions. Yet in our study, MuA favored matches to its consensus at seven positions that did not appear important in the mutagenesis, as well as at four positions that were not tested by mutagenesis. To illustrate the point further, the top letters of the Pictogram are those that occur most frequently in the 18 pseudo-ends; these top letters correspond well to the consensus sequence derived from the natural sites (written above the Pictogram). This agreement suggests that most base-pairs in the recognition sequence contribute favorable energy for transposition, though in some cases the contribution may be too subtle to detect by point mutations.

The Pictogram represents only the sequences immediately adjacent to the cleavage sites; we found no further sequence similarity in the DNA adjacent to that. Nonetheless, we assume that the observed transposition activity is due to MuA tetramers (5,6). The absence of an identifiable second pseudo-recognition site suggests that the position and/or sequence requirements for an inner recognition site (R2/L2) are less rigid than for an outer site (R1/L1).
Cleavage site sequences are highly selected.

The pseudo-ends are highly selected for their cleavage sites (Fig. 6). Among transposons in general, cleavage sites tend to have an identical sequence on the two transposon ends (31). It is tempting to speculate that the cleavage site sequence is so constrained because it is intimately engaged in the protein's active site and perhaps directly involved in catalysis. However, the same would be true of the cleavage sites of modular restriction enzymes, for example the type II enzymes FokI and SfiI. Yet these enzymes have tightly constrained recognition sites rather than cleavage sites (32,33), arguing that there is no intrinsic mechanistic reason why the cleavage site should bear the tightest constraints. The cleavage site of a transposon defines which base-pairs will transpose and which will stay behind. Transposases may have evolved the ability to strongly discriminate their correct cleavage sites, so as to protect the integrity of the transposon ends from one transposition event to the next.

The particular sequence TG/CA marks the cleavage sites of many distantly related transposable elements, including retroviruses (34) and bacterial insertion sequences (31). It is not clear why this is so. The most favored cleavage site of the SfiI restriction enzyme is CA, but SfiI cleaves between the C and the A whereas transposases cleave after the A (33). Analysis of naked DNA by computational modeling and other methods reveals unique structural flexibility for the dinucleotide TG/CA (35). It is uncertain whether this intrinsic DNA structure is relevant to enzymatic cleavage reactions.

Flexibility and rigidity within a MuA transpososome

Within a transpososome, MuA subunits use three distinct domains to contact three DNA sites on the transposon ends (13-16). The cleavage site is contacted by the catalytic domain, and the two halves of the recognition site are contacted by the two subdomains of the bipartite domain Iβ-γ. How flexible are these various binding modules in their relative positioning? Statistical analyses of the 18 pseudo-ends, together with transposition assays of specifically designed Mu end fragments, provide a partial answer.

Why does cleavage site spacing matter? Within a transpososome, a cleavage site and its adjacent recognition site are bound by separate MuA subunits (17-20), whereas the two half-recognition sites are bound by tandem domains of a single subunit (5,10,15,28). One might therefore have expected higher tolerance for insertions near the cleavage site than within the recognition sites, but in fact the opposite is true (Table 1 and Fig. 5).

The constraints on cleavage-site position may indicate extensive intersubunit interactions. In addition, the crystal structure of the synaptic complex of a related transposase, from Tn5, shows that the catalytic domain and DNA-binding domain of opposite subunits lie immediately adjacent to each other on the DNA (36). The subunits' close proximity probably limits the ability to shorten the DNA of the Tn5 transposon, and similar constraints may apply to the Mu transpososome. Finally, it is possible that the...
catalytic domain of MuA makes some contacts in the outer-half recognition site, overlapping the contacts made by domain Iγ of the opposite subunit. If so, moving the cleavage site may disrupt the contacts to the catalytic domain, which could account for activity losses.

Altered DNA spacing is accommodated locally in the transpososome and affects a late step of transposition.Mixing heterologous dsDNA fragments revealed that the spacing of sites on one fragment does not effect its synaptic partner (Fig. 5c). This suggests that the transpososome accommodates altered spacing through localized changes, rather than through global changes in protein structure that may be transmitted between subunits.

Two lines of evidence suggest that, with the precleaved fragments used in this study, positional-shifts affect a step that occurs after initial transpososome assembly. The results are most striking for the fragments altered in cleavage-site spacing, and they contrast with a similar study of the restriction enzyme SfiI, for which altering the spacer between its two recognition sites does harm complex assembly (37). First, base-pair insertions or deletions in MuA fragments cause an increase in the relative number of single fragment transfers (SFTs) compared to double fragment transfers (DFTs) (Fig. 5a&c), and an assembly defect should not specifically increase complexes that perform SFT. Second, in the fragment-mixing experiments (Fig. 5c), altered and wild-type fragments seem to pair-up indiscriminately in synaptic complexes, as judged by the extent of SFTs in these reactions. This result implies that the altered fragments are defective at a post-synaptic step. Perhaps the altered fragments tend to fail in engaging the cleavage site in the protein's active site (38). Consistent with this idea, MuA can form competitor-stable complexes on DNA fragments that lack a cleavage site, indicating that the initial commitment to a pair of DNA molecules does not require engagement of the cleavage site ((24); and Personal communication, K. Yangihara and K. Mizuuchi.).

Flexibility between half-sites, at a cost. Analyses of 18 pseudo-Mu DNA ends suggest phasing flexibility of ±1 base-pair between recognition half-sites (Table 1). Transposition assays of various DNA fragments supported this assertion (Figs. 3&4), but also implied that the natural spacing is optimal. Others have found that adding or removing an A/T base-pair from an A/T stretch in the middle of R1 dramatically reduces in vivo transposition, consistent with the natural spacing being optimal (21).

Other protein families contain bipartite DNA binding domains, similar to domain Iβ-γ of MuA: for example, the POU, cut, myb and paired families of transcription factors (39-42). Pax6, a member of the paired family, has a rigid 15 residue tether connecting the two domains that comprise its bipartite DNA-binding domain. The tether interacts with DNA, directly contributing to binding specificity (42). By contrast, the structures of POU proteins Oct-1 and Pit-1 both suggest flexible tethers (43,44). Consistent with the structural data, a number of POU proteins can bind and regulate DNA sites with
variable arrangements of half-sites (39,43,45,46). Yet even flexible POU proteins usually have an optimal spacing, analogous to the situation described here for MuA. Optimal spacing can be determined in part by the length of the amino acid tether, but also by cooperative interactions between the protein's two DNA binding domains (47).

NMR structures have been solved for the isolated MuA domains Iβ (15) and Iγ (14). In the domain Iγ structure, the inter-domain tether is unstructured. The tether is only 13 amino-acids, whereas the tethers in POU proteins range from 15 to 56 amino-acids. The two MuA domains have been modeled docked on DNA (15), and the authors of this model suggest that upon binding DNA the tether becomes structured and interacts directly with the DNA. This suggestion was made to explain the pattern of nucleotides that are protected from chemical digestion when MuA is bound to its full recognition site (28). The flexibility in half-site spacing described in this study is perhaps most consistent with an unstructured tether. However, it is possible that the DNA-bound tether is structured, but that the transpososome overall retains enough flexibility to accommodate one nucleotide more or less – perhaps through changes in DNA structure rather than protein structure.

**Biological role for flexible binding.** Many transcription factors exploit their bipartite DNA binding properties for biological regulation (39,46,48). There is no evidence that bipartite binding plays as significant a role in Mu transposition, but flexible binding might play a role in interactions between MuA and the L2 site. L2 is considered to be a half-site, because it only contains the sequences recognized by domain Iβ of MuA (7) (Fig. 2), and because its chemical protection pattern is less extensive than the other recognition sites (28). However, positions 11, 12 and 13 of L2 read "TCA" (Fig. 2), which matches the consensus sequence through flexible alignment to positions 10, 11 and 12. Thus, it is possible that MuA obtains favorable binding energy through transient contacts with these three base-pairs.

**Multiple roles of the recognition sites in transposition.**

In the accompanying paper, we discussed two roles for the MuA recognition sites: positioning two MuA subunits close together, and allosterically activating transposase (9,24). Here we showed that the MuA recognition sites also help define the cleavage sites, since moving the cleavage site reduces transposition efficiency. Among transposons in general, each transposon has a defined spacing between cleavage site and outer recognition site: for example, 5 bps on Tn10 (49), 8 bps on Tn7 (50). In contrast, both the position and the number of recognition sites may differ on each of a transposon's two ends. Given the complexity in length and number of transposase recognition sites, it makes sense that they serve multiple functions during transposition.
Acknowledgements. We are indebted to Chris Burge for use of his Pictogram program, guidance in using the program, and helpful discussions, to Earnest Fraenkel for comments on the manuscript, and to Stephen Halford for helpful discussions. We are also deeply grateful to Stephen Bell, Briana Burton, Robert Gordon, Janet Lindow, Phoebe Rice and Frank Solomon for comments on the manuscript and/or helpful discussions, and we thank Katsuhiko Yanagihara and Kiyoshi Mizuuchi as well as Insuk Lee and Rasika Harshey for communicating results prior to publication. This work was supported by a U.S. Public Health Service grant (GM499224) from the National Institutes of Health. T.A.B. is an employee of the Howard Hughes Medical Institute.

References

DNA Sites in a Mu Transpososome

Figure Legends

**Fig. 1. Introductory schematics.** Not drawn to scale. **A. Fragment transposition assays.** Donor fragments can be paired in a synaptic complex and transposed by MuA (11). The fragments usually include two MuA recognition sites (e.g. R1 and R2) and a cleavage site. If both fragments are properly transferred to the target DNA, the target appears to become linear, producing a DFT (Double Fragment Transfer) product. If one of the two fragments fails transfer, a supercoiled target relaxes, producing an SFT (Single Fragment Transfer) product. Either way, if the fragments are radiolabeled the label is incorporated in the DNA product, providing an easy assay for transposition efficiency. In the schematic, DNA fragments are drawn parallel to each other, for simplicity of presentation. **B. DNA binding by MuA transposase** (14-16). MuA transposase contains a bipartite DNA binding domain. Its subdomains, Iβ and Iγ, recognize the inner and outer halves respectively of a recognition site. In addition, a separate catalytic domain engages the cleavage site on the synaptic partner. The cleavage site is shown in bold. Not shown are the reciprocal set of interactions which are mediated by a second subunit. The DNA molecules are drawn anti-parallel, as they appear in the crystal structure of the related Tn5 transposase (36). The placement of the protein domains on the DNA in this drawing is not intended to convey information about actual modes of binding, as the structures were all solved in the absence of DNA. **C. Strategy used to isolate non-Mu donor sequences from a "hybrid" transpososome.** Synaptic complexes were formed between a DNA fragment containing the R1 and R2 recognition sites and a larger non-Mu DNA molecule (φX174 RF1). A separate 80 bp DNA fragment served as transposition target. Transposition joints were specifically amplified with a pair of PCR primers, one to the target and one to φX174.

**Fig. 2. Sequences of 18 sites on a non-Mu DNA molecule (φX174 RF1) that were transposed by MuA.** The sequences are listed 5’ to 3’, beginning with the complement of the nucleotide that was joined to the target. Rows 19 – 26 list the natural MuA recognition sites. The top and bottom rows of the table list a consensus sequence derived from the 6 natural MuA recognition sites and the 6 natural D108 sites (7,22). In this and all figures, black squares indicate base-pairs that match this consensus. The numbers on the right side of the table (second-to-last column) are the point scores for the recognition sites (positions 6-27), as described in the text. The numbers in the last row identify sites that were isolated.
multiple times. We know that multiple isolates represent multiple transposition events, as opposed to sibling pairs of an identical clone, because each transposition was into a unique target site.

**Fig. 3. Evidence for flexibility between recognition half sites.**

A. Summary of fragment sequences. Shown are nucleotides 5-32 on the non-transferred strands of specially designed donor fragments. This includes the cleavage site and the first recognition site—not shown is the second recognition site and the 5' overhang. (The complete fragment sequences are described in experimental procedures.) Fragment names are: "HSI"—half-site insertion, "HSD"—half-site deletion, "HSN"—Half-site null. The capital "A" denotes an inserted base-pair. The point scores, on the right side of the figure, were calculated as for figure 2. B. Schematic of the same fragments. Light gray box – outer half-site sequence (R1 or R2). Dark gray box – inner half site sequence (R1 or R2) White box – spacer nucleotide. Hatched box – anti-consensus sequence. C. Autoradiogram showing transposition of radiolabeled fragments into $\phi X174$ target. In these experiments, $\phi X174$ served primarily as target, not donor. Smears at high MuA concentration result from reuse of individual target DNA molecules in multiple transposition reactions. [fragment] = 50nM. [MuA] = 50, 100, 150 nM.

**Fig. 4. Further evidence for flexibility between recognition half sites.**

A. Rigid and flexible alignments of the pseudo-recognition site from sequence 17. The arrow highlights the alignment shift.. B. Summary of fragment sequences. Donor fragments contained recognition sites based on sequence 17 in table 2 and cleavage sites from the natural Mu right end. Capital letters represent base-pairs that were mutated from the original sequence 17. The fragment sequences are shown aligned with the MuA consensus sequence in two ways: a rigid alignment, and a flexible alignment based on a one-base-pair shift after position 14. This figure does not show entire fragments, which include two copies of the same pseudo-recognition site. The complete fragment sequences are listed in experimental procedures. C. Autoradiogram showing transposition of fragments 17a,b,c and d. [MuA] = 400, 800, 1600 nM.

**Fig. 5. Analysis of effects of base-pair insertions or deletions.**

A. Comparison of cleavage site vs. half-site phasing. The autoradiogram shows the products of transposition of precleaved donor fragments into $\phi X174$ target. New fragments are CSI (cleavage site insertion) and CSD (cleavage site deletion), which contained a base-pair insertion and a deletion, respectively, between positions four and five. Fragments HSI (half-site insertion) and HSD (half-site deletion) are as described in Fig. 3. Transposition reactions were performed in 10mM CaCl$_2$, as opposed to 10mM MgCl$_2$, to prevent donor cleavage. (The previous experiments, shown in Figs. 3 and 4, were performed in MgCl$_2$.) [Fragment] = 125nM. [MuA] = 100nM.

B. The bar-graph summarizes results of four experiments, including the one shown in Fig. 5a. Error bars represent one standard deviation. C. Fragment mixing experiment reveals autonomy of the two transpososome halves. Transposition reactions were performed with 50 nM unlabeled fragment, 5 nM
labeled fragment, 150 nM MuA, and 10 mM CaCl2 in place of MgCl2. White bars represent the percentage of radiolabel that became incorporated in strand-transfer products. Dark bars represent the percentage of the total that comprised SFTs (single fragment transfers). Bars represent the mean of three independent experiments. Error bars represent one standard deviation.

**Fig. 6. Pictogram analysis of 18 pseudo-end sequences (30).** The sequences are listed individually in Fig. 2. The size and position of each letter in the Pictogram represent the frequency of appearance of that nucleotide at that position on the non-transferred strand of the pseudo-ends. The consensus sequence, listed above the pictogram, is derived from the 6 natural Mu recognition sites and the 6 natural D108 recognition sites (7,22). Underlined letters in the consensus are positions where the pictogram and the consensus match. A. Analysis of 18 pseudo-ends, without manipulation of the sequences. B. Analysis of 18 pseudo-ends, of which sequences 1 and 5 each had a nucleotide inserted at position 15, and sequence 17 had a nucleotide removed at position 15. These changes resulted in an improvement of • 5 nucleotides (or 3.5 points) per sequence. We chose to manipulate the alignment of only those three sequences because similar changes to other sequences gave improvements of no more than 3 nucleotides (or 2.5 points). C. For this analysis, sequences 1, 5 and 17 were manipulated as in Fig. 6b. In addition, sequences that were isolated multiple times were counted that many times in the analysis (see Fig. 2).
Paper II, Figure 3

A

<table>
<thead>
<tr>
<th>Rigid alignment</th>
<th>consensus</th>
<th>wild-type</th>
<th>HSI</th>
<th>HSD</th>
<th>HSN</th>
</tr>
</thead>
<tbody>
<tr>
<td>cleavage site</td>
<td>t g n n</td>
<td>t g a a g</td>
<td>t g a a g</td>
<td>t g a a g</td>
<td>t g a a g</td>
</tr>
<tr>
<td>outer half R1</td>
<td>y g t t c a y t n</td>
<td>a a a a c g e g a a a g</td>
<td>a a a a c g e g a a a g</td>
<td>a a a a c g e g a a a g</td>
<td>a a a a c g e g a a a g</td>
</tr>
<tr>
<td>inner half R1</td>
<td>r a a r e c g a a a r</td>
<td>13.5</td>
<td>9.5</td>
<td>8.5</td>
<td>7.0</td>
</tr>
<tr>
<td>point score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Flexible alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>consensus HSI</td>
</tr>
<tr>
<td>t g n n</td>
</tr>
<tr>
<td>HSD</td>
</tr>
<tr>
<td>point score</td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th>R1</th>
<th>R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>TC</td>
</tr>
<tr>
<td>HSI</td>
<td>TG</td>
</tr>
<tr>
<td>HSD</td>
<td>AC</td>
</tr>
<tr>
<td>HSN</td>
<td>TG</td>
</tr>
<tr>
<td></td>
<td>AC</td>
</tr>
</tbody>
</table>

C

Fragment: [MuA] WT HSI HSD HSN

SFT

DFT

fragment

<table>
<thead>
<tr>
<th>lane #</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>% product</td>
<td>34</td>
<td>23</td>
<td>11</td>
<td>22</td>
<td>17</td>
<td>7</td>
<td>17</td>
<td>13</td>
<td>5</td>
<td>7</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>% of wt</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>65</td>
<td>74</td>
<td>64</td>
<td>50</td>
<td>57</td>
<td>41</td>
<td>21</td>
<td>15</td>
<td>14</td>
</tr>
</tbody>
</table>
Paper II, Figure 4

A  

<table>
<thead>
<tr>
<th>Consensus</th>
<th>Rigid Alignment</th>
<th>Flexible Alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGGNNYGTTCAYTNTARAAYRCGAARR</td>
<td>TCGGGCGACTTCACGCGCGAAG</td>
<td>TCGGGCGACTTCACGCGCGAAG</td>
</tr>
</tbody>
</table>

B  

<table>
<thead>
<tr>
<th>Position</th>
<th>Rigid Alignment</th>
<th>Flexible Alignment</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>17a</td>
<td>TGGNNYGTTCAYTNTARAARYRCGAARR</td>
<td>TGGNNYGTTCAYTNTARAARYRCGAARR</td>
<td>8.5</td>
</tr>
<tr>
<td>17b</td>
<td>TGGNNYGTTCACGCGCGAAG</td>
<td>TGGNNYGTTCACGCGCGAAG</td>
<td>10</td>
</tr>
<tr>
<td>17c</td>
<td>TGGNNYGTTCACGCGCGAAG</td>
<td>TGGNNYGTTCACGCGCGAAG</td>
<td>8.5</td>
</tr>
<tr>
<td>17d</td>
<td>TGGNNYGTTCACGCGCGAAG</td>
<td>TGGNNYGTTCACGCGCGAAG</td>
<td>9.5</td>
</tr>
</tbody>
</table>

C  

<table>
<thead>
<tr>
<th>Fragment: [MuA]</th>
<th>17a</th>
<th>17b</th>
<th>17c</th>
<th>17d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wells —_________</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SFT —___________</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DFT —___________</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lane #</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Paper II, Figure 6

A Raw alignment

B Adjusted alignment

C Adjusted alignment, Multiples included
Table 1. Statistical analyses of 18 pseudo-ends indicate flexibility between recognition half-sites.

<table>
<thead>
<tr>
<th>1. cleavage to first site (nts)(^a)</th>
<th>2. outer to inner half of outer site (nts)(^b)</th>
<th>3. mean score for outer site (points)(^c)</th>
<th>4. mean score for random outer site (points)(^d)</th>
<th>5. std dev for random group of 18 (points)(^e)</th>
<th>6. # of std devs from random (points)(^f)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>1</td>
<td>6.50</td>
<td>5.41</td>
<td>0.42</td>
</tr>
<tr>
<td>2</td>
<td>4 to 5</td>
<td>1</td>
<td>7.14</td>
<td>6.28</td>
<td>0.38</td>
</tr>
<tr>
<td>3</td>
<td>5 to 6</td>
<td>1</td>
<td>7.00</td>
<td>6.28</td>
<td>0.38</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>1 to 2</td>
<td>7.56</td>
<td>6.00</td>
<td>0.41</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0 to 1</td>
<td>7.31</td>
<td>6.00</td>
<td>0.40</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>0 to 2</td>
<td>7.97</td>
<td>6.33</td>
<td>0.39</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>0 to 4</td>
<td>8.33</td>
<td>6.77</td>
<td>0.38</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>0 to 6</td>
<td>8.44</td>
<td>7.06</td>
<td>0.38</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>0 to 10</td>
<td>8.53</td>
<td>7.37</td>
<td>0.38</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>outer to inner site</th>
<th>outer to inner half of inner site</th>
<th>mean score for inner site</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>0 to 1</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>0 to 2</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0 to 2</td>
</tr>
</tbody>
</table>

\(^a\&\(b\) Describes spacing parameters. We defined the outer half site as the consensus YGTITTCAYT and the inner half site as RAARYRCGAAAC. The precise boundary between the two half-sites is not known; the two might even overlap slightly.

\(^c\)Mean point score for 18 sites listed in figure 2.

\(^d\)Mean point score for all 10,772 sites on both strands of φX174.

\(^e\)Standard deviation for the mean scores of randomly chosen groups of 18 sites, calculated as the standard deviation for the group of 10,772 individual sites, divided by √18.

\(^f\)The number of standard deviations between the mean for the 18 pseudo-ends and the mean for the random population. (column 6 = (column 3+column 4/column 5)).
<table>
<thead>
<tr>
<th>position</th>
<th>consensus(^a)</th>
<th>pseudo-ends(^b)</th>
<th>mutagenesis(^c)</th>
<th>protection/interference(^d)</th>
<th>other</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>T</td>
<td>strongly favors T, then A</td>
<td>Left end favors T &gt;&gt; C &gt; A &gt;&gt; G</td>
<td>Right end favors T &gt;&gt; A &gt;&gt; C &gt;&gt; G</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>G</td>
<td>favors G &gt; A &gt; T &gt; C</td>
<td>Favors G &gt; A &gt; C &gt; T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>N</td>
<td>favors T, then A</td>
<td>Left end favors T</td>
<td>Right end favors A</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>N</td>
<td>no bias</td>
<td>ND</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>N</td>
<td>no bias</td>
<td>ND</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Y</td>
<td>favors Y</td>
<td>R2 C to T no effect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>G</td>
<td>strongly favors G (then A)</td>
<td>L1 G to A down</td>
<td>strong purine interference</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>T</td>
<td>favors T</td>
<td>R1 G to A no effect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>T</td>
<td>no bias</td>
<td>R1 C to A no effect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>T</td>
<td>strongly favors T</td>
<td>ND</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>C</td>
<td>strongly favors C</td>
<td>L1 C to T down</td>
<td>strong purine interference</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>A</td>
<td>strongly favors A</td>
<td>ND</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Y</td>
<td>favors Y</td>
<td>R2 or L1 C to T no effect</td>
<td>weak purine interference</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>T</td>
<td>favors C or G</td>
<td>R1 or R2 G to A no effect</td>
<td>&quot;C&quot; is most favored here, yet no natural site has &quot;C&quot;</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>N</td>
<td>no bias</td>
<td>ND</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>R</td>
<td>favors T or A bias against C</td>
<td>L1 G to A no effect</td>
<td>&quot;R&quot; in consensus perhaps reflects steric clash with &quot;C&quot;</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>A</td>
<td>no bias</td>
<td>ND</td>
<td>protection but no interference</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>A</td>
<td>favors A</td>
<td>ND</td>
<td>protection but no interference</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>R</td>
<td>no bias</td>
<td>L1 G to A no effect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Y</td>
<td>favors Y</td>
<td>R1 C to T no effect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>R</td>
<td>strongly favors A, then G</td>
<td>R2 G to A no effect</td>
<td>weak purine interference</td>
<td>Of native sites, only L1 and L3 have A at this position.</td>
</tr>
<tr>
<td>22</td>
<td>C</td>
<td>favors C</td>
<td>R1, R2, L1, L2 C to T down</td>
<td>protection but no interference</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>G</td>
<td>favors G</td>
<td>R1, R2, L1 G to A down</td>
<td>strong purine interference</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>A</td>
<td>favors A</td>
<td>ND</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>A</td>
<td>no bias</td>
<td>ND</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>A</td>
<td>favors A</td>
<td>L1 A to G no effect</td>
<td>protection but no interference</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>R</td>
<td>no bias</td>
<td>L1 A to G no effect</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) See legend to Fig. 2.

\(^b\) Summarizes Fig. 6b&c, taking into account the consensus sequence in determining what constitutes “favoring”.

\(^c\) Positions 1-3, in vitro and in vivo data (23,29); position 5-27, in vivo data (22).

\(^d\) Effects of DMS methylation (28). Interference—methylation interferes with MuA binding. Protection—MuA protects the nucleotide from methylation.
Sequence and positional requirements for DNA sites in a Mu transpososome
Ilana Goldhaber-Gordon, Michael H. Early, Matthew K. Gray and Tania A. Baker

J. Biol. Chem. published online December 27, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M110342200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts