

Target choice determinants of the Tc1 transposon of *Caenorhabditis elegans*

René F. Ketting, Sylvia E. J. Fischer and Ronald H. A. Plasterk*

Division of Molecular Biology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066CX Amsterdam, The Netherlands

Received July 7, 1997; Revised and Accepted September 4, 1997

ABSTRACT

The Tc1 transposon of *Caenorhabditis elegans* always integrates into the sequence TA, but some TA sites are preferred to others. We investigated a TA target site from the *gpa-2* gene of *C.elegans* that was previously found to be preferred (hot) for Tc1 integration *in vivo*. This site with its immediate flanks was cloned into a plasmid, and remained hot *in vitro*, showing that sequences immediately adjacent to the TA dinucleotide determine this target choice. Further deletion mapping and mutagenesis showed that a 4 bp sequence on one side of the TA is sufficient to make a site hot; this sequence nicely fits the previously identified Tc1 consensus sequence for integration. In addition, we found a second type of hot site: this site is only preferred for integration when the target DNA is supercoiled, not when it is relaxed. Excision frequencies were relatively independent of the flanking sequences. The distribution of Tc1 insertions into a plasmid was similar when we used nuclear extracts or purified Tc1 transposase *in vitro*, showing that the Tc1 transposase is the protein responsible for the target choice.

INTRODUCTION

We studied integration of the Tc1 transposon, a member of the most widespread family of DNA transposons: the Tc1/*mariner* family (1–5). Members of this family are found in a wide variety of organisms, ranging from fungi to vertebrates. Also in the human genome elements of this family have been identified (6–9). Transposase protein alone is sufficient to catalyse the complete transposition reaction, at least for the Tc1 and *mariner* elements (10,11). This feature makes these elements attractive vectors for transgenesis, and thus it is important to know what factors determine the integration specificity of these transposons.

It has become apparent that most transposons do not integrate their DNA randomly into their host genome. Appropriate target sites may be selected by various means and for various reasons. The target choice can be dependent on primary sequence. The 297 and 17.6 elements of *Drosophila melanogaster* always integrate into the sequences ATAT and TATATA, respectively (12,13); the *gypsy* transposon of *Drosophila melanogaster* has TA(C/T)ATA as target site (14,15), whereas *pogo* uses a TA dinucleotide (16). P elements have a less defined preference, but a consensus sequence

has been determined: GGCCAGAC (17). Also prokaryotic elements can be attracted by specific sequences, ranging from a preference for G/C base pairs in the case of Tn5 (18), to more specific insertion consensus sequences for bacteriophage Mu [see review by Mizuuchi and Craigie (19)], IS630 (20), Tn3 (21) and Tn10 (22). The most dramatic example is the *Escherichia coli* Tn7 transposon, which, when using the *TnsD* protein, integrates very efficiently into only one specific site in the *E.coli* genome: *attTn7* [reviewed by Craig (23)].

Besides the primary DNA sequence, the DNA structure can also have a significant effect on target site choice. It has been shown for IS231A that, in addition to the consensus sequence, the flanking DNA should be curved in opposite directions (24). The target site choice of Tn3 and Tn10 is probably also influenced by local DNA structures (21,25). The Tn5 transposon has a requirement for supercoiled target DNA (26), and retroviruses integrate more efficiently into nucleosomal DNA than into naked DNA, probably due to bending of the DNA induced by the nucleosomes (27,28). Integration of the Ty elements of *Saccharomyces cerevisiae* is affected by a variety of features ranging from DNA structure to protein–protein interactions [reviewed by Curcio and Morse (29)].

Transcription and replication may further affect target choice. P elements preferentially integrate in the vicinity of transcription start sites (30,31), and retroviruses integrate most often into DNase I hypersensitive sites, which are associated with transcriptional activity [see review by Craigie (32)], and also Tn5 seems to prefer transcriptionally active regions of the DNA (33). Tn10 and Mu, on the other hand, preferentially integrate into non-transcribed DNA (34,35). A very interesting case is the Tn7 transposon. The *TnsE* protein of Tn7 targets the element to the replication fork of conjugating plasmids, thus enabling the Tn7 transposon to spread among bacteria (36).

The Tc1 transposon of *Caenorhabditis elegans*, as all elements of the Tc1/*mariner* family, always integrates into a TA dinucleotide, which is duplicated upon insertion (37). However, not all TA dinucleotides are equally used. When independent *in vivo* Tc1 insertions into a 1 kb region of the *C.elegans* genome are analysed, a distinct insertion pattern emerges (38). Some TA dinucleotides are used frequently, whereas others, which may be only base pairs away, are not used at all. This insertion pattern within this 1 kb region is not influenced by chromatin structure or transcriptional activity, as *in vitro* a similar insertion pattern is obtained (11).

*To whom correspondence should be addressed. Tel: +31 20 512 2081; Fax: +31 20 669 1383; Email: rplas@nki.nl

Insertion consensus sequences have been identified for Tc1, by analysing both new insertions into known genes (39,40), and flanks of endogenous Tc1 elements present in the genome of a high Tc1 copy number strain of *C.elegans* (41). The consensus sequences obtained are very similar, and define a stretch of 10 bp centred around the TA dinucleotide: CAYATATRTG. Here we take advantage of the Tc1 *in vitro* transposition system described by Vos *et al.* (11), to define the minimal sequences that make a site hot for Tc1 integration, using deletion mapping and site directed mutagenesis. We also show that the Tc1 transposase (Tc1A) is the sole protein responsible for the observed insertion pattern.

MATERIALS AND METHODS

Plasmids

All plasmids used in this study as target plasmid were made by cloning double-stranded oligonucleotides into the polylinker of pUC19. Target plasmid pRP1209 contains the oligos TAT1 (5'-TGTATCTGGTGTATGTCTATTGAC-3') and TAB1 (5'-GTCAATAGACATACACCAGATAACA-3'), cloned into the *Hind*II site. pRP1210 and pRP1211 were obtained by cloning the following oligonucleotides into the *Sph*I-*Xba*I site; pRP1210: TS8T (5'-CTAGACATACACCATG-3') and TS8B (5'-GTGTATGT-3'); pRP1211: TSCST (5'-CTAGGTCAATAGACATACACCACATACACACATG-3') and TSCSB (5'-TGTGTATGTGGTGTATGTCTATTGAC-3'). The target plasmids pRP1218-pRP1223 were made by cloning the following oligonucleotide pairs into the *Bam*HI-*Pst*I site; pRP1218: 8TS4T (5'-GATCAGACATACACCTGCA-3') and 8TS4B (5'-GGTG-TATGTCT-3'); pRP1219: 8TS5T (5'-GATCAAACATACACATGCA-3') and 8TS5B (5'-TGTGTATGTTT-3'); pRP1220: 8TS6T (5'-GATCAATCATAACAGATGCA-3') and 8TS6B (5'-TCTGTATGATT-3'); pRP1221: 8TS7T (5'-GATCAGTCATACAGCTGCA-3') and 8TS7B (5'-CTGTATGACT-3'); pRP1222: 8TS8T (5'-GATCAGGTGTACACCTGCA-3') and 8TS8B (5'-GGTGTA-CACCT-3'); pRP1223: 8TS9T (5'-GATCAGACATATGTCTGCA-3') and 8TS9B (5'-GACATATGTCT-3').

Donor plasmids pRP1214 and pRP1215 were made by cloning the Tc1-containing *Pvu*II fragment of plasmid pRP1209, with a Tc1 insert into sites I or IV respectively, into the filled up *Eco*RI-*Hind*III site of pRP490 (11).

Relaxed pRP1218 plasmid was made by ligating 2 µg of *Eco*RI digested DNA in a volume of 200 µl with 5 U of T4 ligase overnight at 16°C. The DNA was phenolised, precipitated and dissolved in 12 µl water, of which 9 µl was used in a transposition reaction. As judged from an agarose gel, 50% of the DNA was religated.

In vitro transposition

In vitro transposition reactions were performed as described by Vos *et al.* (11). In a reaction of 50 µl, typically 2 µg of target plasmid was incubated together with 500 ng of donor plasmid (pRP490) and 20 ng Tc1A from nuclear worm extract or 2.5 ng Tc1A purified from inclusion bodies. The DNA from the reaction was transformed to *E.coli* strain DS941, and the colonies were stained with IPTG/X-gal. Only white colonies were analysed further.

Localisation of Tc1 insertions within the target plasmids

The precise integration sites of Tc1 were determined by sequencing or length determination of DNA fragments obtained using PCR.

Sequencing of new inserts was done as follows: the DNA to be sequenced was amplified using the primers VIP8 (5'-CTGGTG-AGTACTCAACCAAG-3') and L2 (42), in a 25 µl reaction. DNA was isolated from a 1% agarose, and sequenced using the Dye Terminator Cycle Sequencing Ready Reaction kit of Perkin Elmer. Sequencing primer was L2.

Alternatively, the site of insertion was determined by running PCR products obtained with the primers 8TS0 (5'-TTCGCCAT-TCAGGCTGCGC-3') and ³²P-labelled vip65 (5'-GGATATCT-TTTTGGCCAG-3') on a 6% denaturing polyacrylamide gel, and comparing the length of the PCR products to markers. The marker fragments were obtained by performing the same PCR on plasmids containing previously sequenced inserts. Two inserts obtained with purified Tc1A protein, that were not comigrating with any of the marker bands, were sequenced. These proved to be odd integration events into non-TA sequences and were discarded.

Transposase preparation

A nuclear extract containing Tc1A was isolated from *C.elegans* as described by Vos *et al.* (11). Recombinant Tc1A was prepared as described by Lampe *et al.* (10). The final protein concentration was ~5 µg/ml. The protein was estimated to be >90% pure.

Statistics

Statistical analysis was performed using the χ^2 method. Differences with a *P* < 0.05 were considered statistically significant.

RESULTS

A hot site for Tc1 integration from the *gpa-2* gene remains a hot site when cloned into the *lacZ* polylinker

To be able to analyse a frequently used TA dinucleotide from the *gpa-2* gene, we cloned a fragment of 24 bp, containing the hot TA together with two flanking TA dinucleotides, into the *Hind*II site of the pUC19 polylinker. The *lacZ* gene remains functional after insertion of this fragment, and thus Tc1 insertions into this region, containing the TA of interest, are easily scored using blue/white selection. We used this plasmid as an integration target in the *in vitro* Tc1 transposition system described by Vos *et al.* (11), using a nuclear extract from worms expressing Tc1 transposase. Transposition events were detected with a frequency between 10⁻⁵ and 10⁻⁶ (Table 1, pRP1209).

Sequence analysis of 65 independent white colonies revealed the insertion pattern depicted in Figure 1A. The distribution of Tc1 elements over the *lacZ* gene was not random, as could be expected from previous studies (38-40). Of the 27 TA dinucleotides, only four are actually used: two sites within the promoter region (sites I and II), one overlapping the start codon (site III), and the hot TA contained within the cloned oligonucleotide (site IV). All these sites have at least a thymidine 3 bp downstream of the TA in common with the consensus (Table 2). Also between these four sites a clear preference is observed. Sites I and III together

contain 6% of the insertions, whereas sites II and IV contain 43% and 51%, respectively. In other words 50% of all insertions choose the subcloned *gpa-2* site out of 27 available target sites in the *lacZ* gene. The two other TA dinucleotides within the 24 bp oligo (sites IVa and IVb in Fig. 1A) are not used at all. This is consistent with previous *in vivo* and *in vitro* experiments (11,38), in which the frequency of usage of these sites was found to be very low. We can conclude that the sequences immediately flanking the TA dinucleotide determine the usage of that dinucleotide for Tc1 integration.

Tc1 flanks do not influence excision frequency

To address whether the flanks of a Tc1 transposon have an effect on the excision of the element, we subcloned a Tc1 element containing either the hot flanks of site IV, or the cold flanks of site I, into the

donor plasmid used in the *in vitro* transposition assay, and determined the transposition frequency out of these plasmids pRP1214 and pRP1215 (Table 3). We found that both donor plasmids support transposition to the same extent, suggesting that the excision frequency out of these plasmids is the same. We also looked directly at the amount of excised Tc1 element from plasmids containing the Tc1 transposon in either a cold or a hot insertion site. This was done by incubating the respective Tc1 containing plasmids with Tc1 transposase and analysing the reaction products on an agarose gel. Excision is detected as a band corresponding to the free linear transposon. Again we could detect no significant difference between the amounts of excised Tc1 elements from either plasmid (data not shown), indicating that the flanking DNA sequences have no effect on excision efficiency. This could be expected as even the TA dinucleotide, although essential for integration, is not essential for Tc3 excision *in vivo* (37).

Table 1. Transposition frequencies obtained with the various target plasmids

Target plasmid	Frequency of transposition (10^{-5})	No. of inserts analysed
pRP1209	0.2	65
pRP1210	0.6	32
pRP1211	1.0	32
pRP1218	2.8	48
pRP1218 ^a	1.7	47
pRP1218 ^a (relaxed)	0.5	21
pRP1219	5.2	47
pRP1220	3.7	35
pRP1221	0.4	43
pRP1222	0.7	47
pRP1223	0.7	49

^aUsing purified Tc1A from *E.coli*, all other data were obtained using nuclear worm extract as Tc1A source.

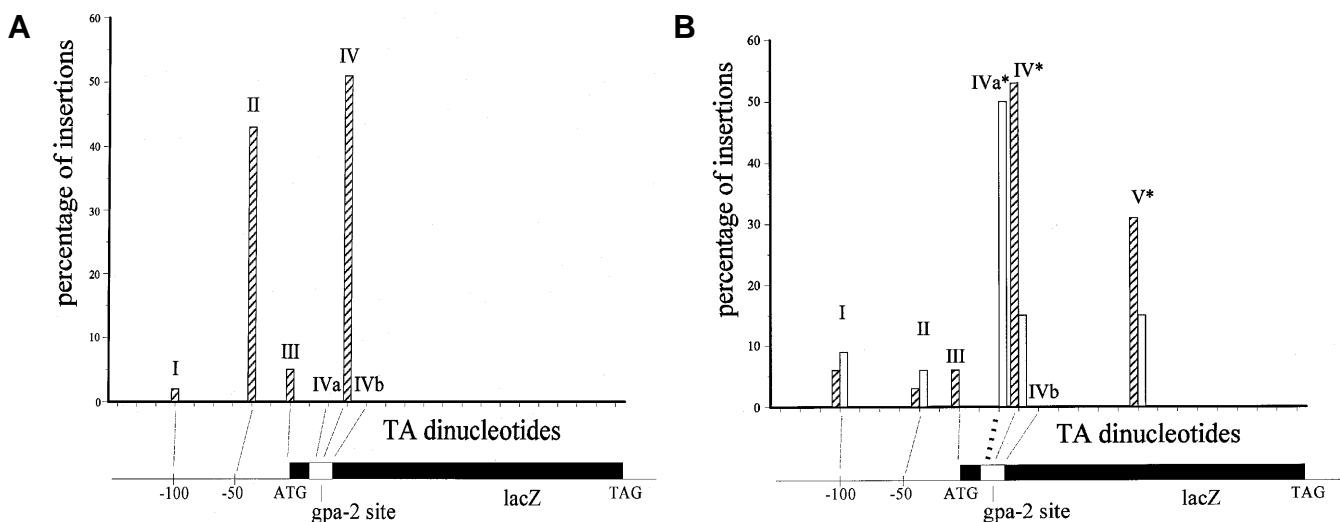


Figure 1. A hot site for Tc1 insertion from the *gpa-2* gene of *C.elegans* is still a hot site when subcloned into a plasmid. (A) Distribution of Tc1 insertions into a *lacZ* gene, represented by the black bar, containing a small piece of *gpa-2* sequence (24 bp), represented by the white bar, cloned into the *Hind*II site of the pUC polylinker (pRP1209). The insertion sites that were found to be used are numbered in roman numerals. Every tick mark on the x-axis represents a TA dinucleotide (B) Mutations in the flank of site IVa can make this site hot (white bars). This oligonucleotide was cloned into the *Sph*I-*Xba*I site (pRP1211). The hatched bars represent the insertion distribution when the *gpa-2* sequence is shortened to 13 bp (also cloned into *Sph*I-*Xba*I, pRP1210). Sites showing statistically significant differences between the two distributions are indicated with an asterisk.

Table 2. Sequences of the various integration sites

Potential integration site	Flanking sequences
Consensus ^a	CAYA TA TRTG
Site I	GAGT TA GCTC
Site II	CTCG TA TGTT
Site III	GTCA TA GCTG
Site IVa	TCTG TA TCTG
Site IVa mutated (pRP1211)	TGTG TA TGTG
Site IV (= <i>gpa-2</i> hot site)	GGTG TA TGTC
Site IV +/- 4 mutated (pRP1219/1220)	TGTG TA TGTT
Site IV +/- 3 mutated (pRP1220/1221)	GCTG TA TGAC
Site IV symmetric (pRP1222)	GGTG TA CACC
Site IV symmetric (pRP1223)	GACA TA TGTC
Site IVb (from <i>gpa-2</i> gene)	TGTC TA TTGA
Site V	CGAT TA AGTT

^aFrom Korswagen *et al.* (41); Y = T or C, R = A or G.

Specific mutations can turn a cold site hot

The sequence around site IVa very much resembles the flanking sequence of site IV (Table 2). However, a dramatic difference in usage as an insertion site is observed. Comparison of the flanks shows that the two flanks differ at the -4, -3, +2 and +4 positions. To investigate whether this explains the difference we mutated the base pairs at -3 and +2 in the flanks of site IVa (Tables 1 and 2, pRP1211). As a result site IVa becomes a hot site (Fig. 1B, white bars). This confirms that these sequence differences can fully explain the differential usage of these two sites.

One additional point should be noted. Compared to Figure 1A (pRP1209), usage of sites II and IV decreases ($\chi^2 = 19.1$, $P < 0.0005$; $\chi^2 = 15.1$, $P < 0.0005$) and another TA dinucleotide, located 120 bp downstream of the oligo is used: site V. For practical reasons in pRP1211 the oligo was cloned into a different site of the pUC19 polylinker compared with pRP1209. Maybe this can account for the usage of site V in pRP1211, as also another clone (pRP1210), containing a shorter oligo cloned into this same site, showed frequent usage of site V instead of site II (Fig. 1B, hatched bars), whereas no other clone analysed showed this phenomenon. It appears that the usage of some TA dinucleotides can be influenced by sequences up to 120 bp away, indicating that not only the base pairs directly flanking the TA dinucleotide are important in determining whether a given TA will be hot (also see the experiment below addressing the role of supercoiling).

Table 3. Tc1 donor flanks do not influence transposition frequency

Site of insertion	Frequency of insertion	Site of excision	Frequency of transposition
Site I (pRP1209)	2%	Site I (pRP1215)	1.0×10^{-5}
Site IV (pRP1209)	51%	Site IV (pRP1214)	1.7×10^{-5}

Four base pairs flanking a TA dinucleotide are sufficient to define a potential hot site

To further fine map the sequences around the TA dinucleotide necessary for efficient Tc1 integration, we made a clone containing site IV, with only 4 bp of *gpa-2* flanking sequence on both sides (pRP1218), and used it as target DNA. The results are shown in Table 1 and Figure 2A (hatched bars). Usage of site IV does not change upon shortening the *gpa-2* flanks from 11 to 4 bp: 51% versus 56%. Only when the base pairs at + and -4 are mutated (Tables 1 and 2, pRP1219) the usage of site IV drops significantly ($\chi^2 = 32.7$, $P < 0.0005$) to 15% (Fig. 2A, white bars). This shows that a stretch of 4 bp on each side of a TA dinucleotide determines the potential usage of that TA. When the base pairs at + and -3 are mutated (Tables 1 and 2, pRP1221) the usage of the *gpa-2* site drops to zero (Fig. 2B, hatched bars), showing that at least one of the base pairs at these positions are essential for recognition by the Tc1 integration complex. When these same mutations are introduced together with the mutations at the + and -4 positions (Tables 1 and 2, pRP1220), the insertion distribution does not change significantly (Fig. 2B, white bars). In summary, mutations at three or four base pairs distance from the TA dinucleotide can seriously affect Tc1 integration efficiency, while changes further removed have no apparent effect.

The right flank of the *gpa-2* site is the determining factor

When the flanks of the *gpa-2* hot site are compared to the Tc1 insertion consensus sequence (Table 2), it appears that the right flank of this site fulfils the consensus nicely, the only mismatch is at the weakest position of the consensus (+4), whereas the left flank does not: it has only one match at the -2 position. To determine which of these two flanks is responsible for the observed integration preference, we made constructs containing symmetrical *gpa-2* sites with either two left (pRP1222) or two right (pRP1223) flanks and analysed the Tc1 insertion pattern. The clone containing the two left flanks had no insertions in this palindromic *gpa-2* site (Fig. 3, hatched bars). When the site is made symmetrical for the right flank 67% of the insertions were found in this palindromic site (Fig. 3, white bars); a slightly higher percentage than the wild type flank, but we could not show this to be significant ($\chi^2 = 2.56$, $P = 0.110$). These results show that it is indeed the right flank, resembling the consensus sequence, that is responsible for the frequent use of the *gpa-2* site. Furthermore, we can state that the mutations introduced into the left flank of various clones have been of little importance to the usage of the site, as these mutations never introduce base pairs present in the consensus (Table 2). Consequently, the observed differences will be due to the mutations introduced into the right flank. Also, it appears that symmetry at the site of insertion does not significantly increase the usage of the site; when one flank of a TA dinucleotide is found to be suitable by the transpososome, the integration reaction initiates, apparently independent of the sequence at the other flank.

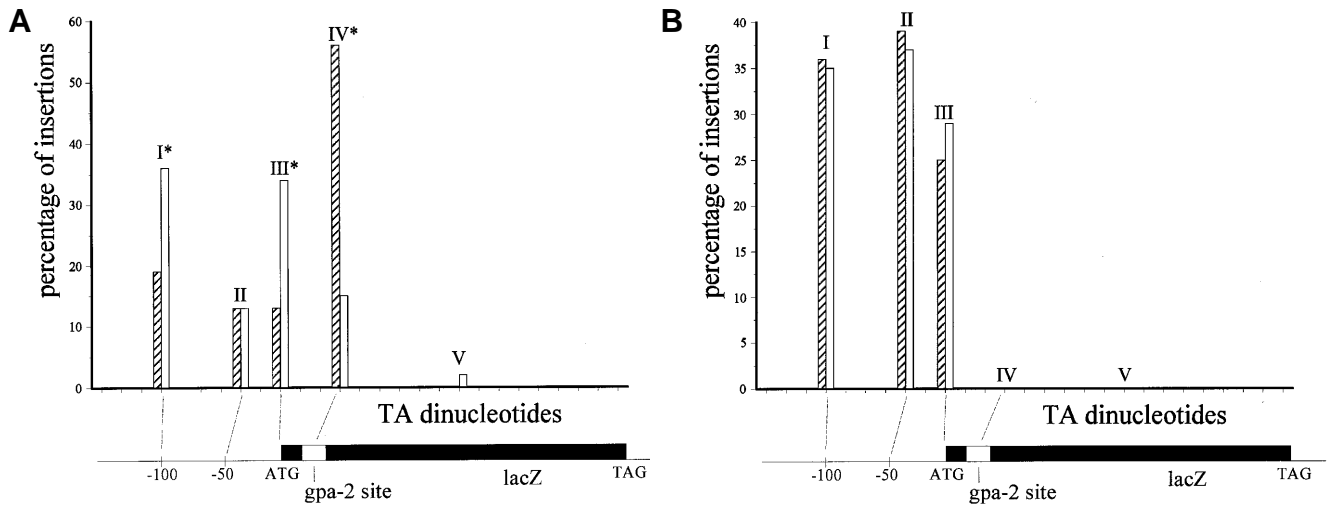


Figure 2. The 4 bp flanking a TA dinucleotide are important to make it a hot site. (A) The hatched bars represent the insertion distribution of a construct containing the *gpa-2* hot site with only 4 bp of *gpa-2* sequence flanking the TA dinucleotide on both sides (pRP1218). When the fourth base pair is mutated on both sides (pRP1219), the site becomes much less preferred (white bars). Significant differences are indicated with an asterisk. (B) When the base pairs at the + and -3 positions are also mutated, we did not detect any insertions into this site anymore (white bars, pRP1220). The same insertion distribution is found when only the base pairs at the third position are mutated (hatched bars, pRP1221).

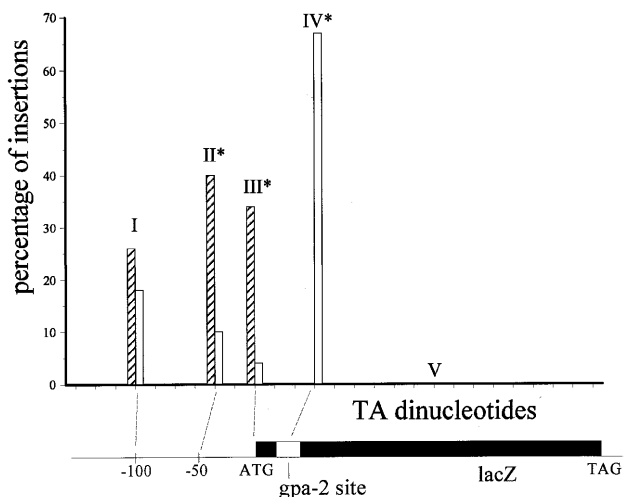


Figure 3. The right flank of the *gpa-2* site makes it hot. In hatched bars the insertion distribution is shown of a construct containing the *gpa-2* site made palindromic for the left flank (pRP1222). The white bars represent a *gpa-2* site that is palindromic for the right flank (pRP1223). Significant differences are indicated with an asterisk.

Tc1A is the only protein responsible for the observed preferences

To address the question whether it is the Tc1A protein alone, or the Tc1A protein together with other nuclear factors, that is determining the observed insertion distribution, we repeated the experiment with the 4 bp *gpa-2* flank (pRP1218), but now using Tc1A protein purified from *E.coli*. The insertion distribution is shown in Figure 4. The observed pattern is similar to that obtained

with nuclear worm extract. Only site III shows a significant difference, 26% versus 13% ($\chi^2 = 9.875$, $P = 0.002$). One notable difference between reactions with nuclear worm extract and purified recombinant protein is that nuclear extract, containing several nucleases, induces significant nicking of DNA, leading to relaxation of the target DNA. After 10 min reaction time >90% of the supercoiled target DNA has become relaxed, and after 15 min no supercoiled target DNA is detected any more on an ethidium bromide stained gel. Incubation of supercoiled target DNA with purified recombinant protein for 1 h results in only marginal relaxation (Fig. 5A). It is conceivable that this structural difference of target plasmid DNA can account for the differential usage of site III. To investigate this hypothesis, we analysed the insertion pattern of Tc1 into either supercoiled or relaxed target DNA, using the Tc1A protein purified from *E.coli*.

The transposition frequency obtained when using relaxed target plasmid DNA [Table 1, pRP1218 (relaxed)] suggests that supercoiling in general is not essential for the integration reaction. Secondly, the insertion distribution pattern, depicted in Figure 5B (white bars), is indeed different from the pattern obtained with supercoiled target plasmid (hatched bars): site III again shows a significantly lower usage ($\chi^2 = 4.923$, $P = 0.026$) when the target DNA is not supercoiled. The distribution of insertions over other sites shows no significant difference.

DISCUSSION

In the study presented here, we demonstrate that a TA dinucleotide together with 4 bp of flanking sequence are sufficient to define a hot site for Tc1 integration. When an insertional hot site from the *gpa-2* gene of *C.elegans* is taken out of its chromosomal context and cloned into a pUC vector, with only 4 bp of original *gpa-2* sequence on each side, this site is still hot. It is not very likely that this is caused by the presence of a particular *lacZ* sequence, as various oligonucleotides cloned into three different

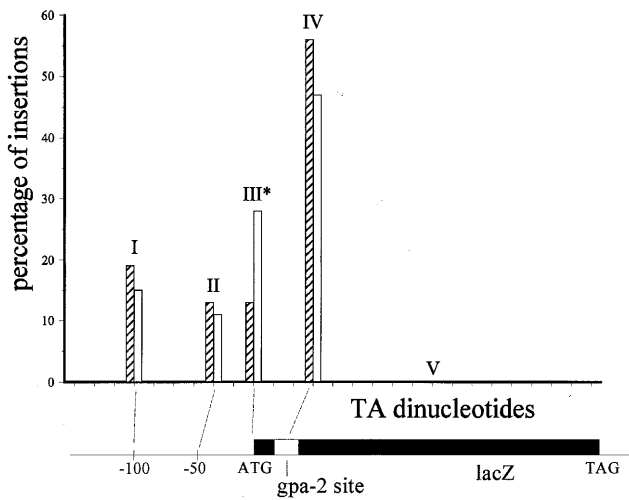


Figure 4. Insertion distribution patterns obtained with different Tc1A protein preparations using pRP1218 as a target vector. Hatched bars indicate the distribution obtained using nuclear worm extract; white bars indicate the distribution obtained using purified Tc1A from *E.coli*. Significant differences are indicated with an asterisk.

sites in *lacZ* all give similar results for the *gpa-2* hot site. In addition, we showed that this high frequency of use is conferred by only one of the two flanks, namely the flank that resembles the previously identified consensus sequence (41). Making the site symmetric did not increase integration efficiency. These results indicate that the transposon complex, next to recognising the essential TA dinucleotide, also has to recognise the DNA on either side of that TA dinucleotide in order to integrate the transposon DNA. The symmetry found in the consensus sequence is most likely due to the fact that the Tc1 transposon integrates orientation independently (38), and that, therefore, the symmetrical

consensus sequence is the result of addition of randomly orientated one-sided targets into one consensus. This was already suggested by Korswagen *et al.* (41), although they did find that there is a preference for symmetry at the ± 3 position.

We found that the flanks of the TA dinucleotide are not of importance for the excision reaction. For the Tc3 transposon even the TA dinucleotide itself is not essential in this step of the reaction (37). A similar independence of the flank of a donor element has been found for the bacterial transposon Tn10 (25). Tc1 excision *in vitro* however, does seem to depend on the flanking TA dinucleotide (11). This difference between Tc1 and Tc3 could be caused by the fact that both transposons have different transposase proteins or by the fact that the TA dinucleotide has been mutated to different residues in both studies. Alternatively, it may be that the method of detection in the assay used for Tc3 is more sensitive than that of the Tc1 *in vitro* system.

Tc1A protein is sufficient for proper target site recognition

We showed that virtually identical distribution patterns are obtained when either a crude nuclear worm extract or purified Tc1A from *E.coli* is used as transposase source (except for site III, see below). This result indicates that the only protein required for proper target site recognition is the Tc1A protein itself. This does not rule out the possibility that various regions in the genome may behave differently with respect to Tc1 integration.

Also, for Tn10 it has been shown that the transposase protein itself recognises the target site: mutants of Tn10 transposase have been identified that display an altered target specificity (43), showing that it is the Tn10 transposase protein that discriminates between the various potential insertion sites. In case of retroviral integration, it has been shown for HIV-1 that the central core region of the integrase protein, harbouring the catalytic domain, contributes to target site selection (44).

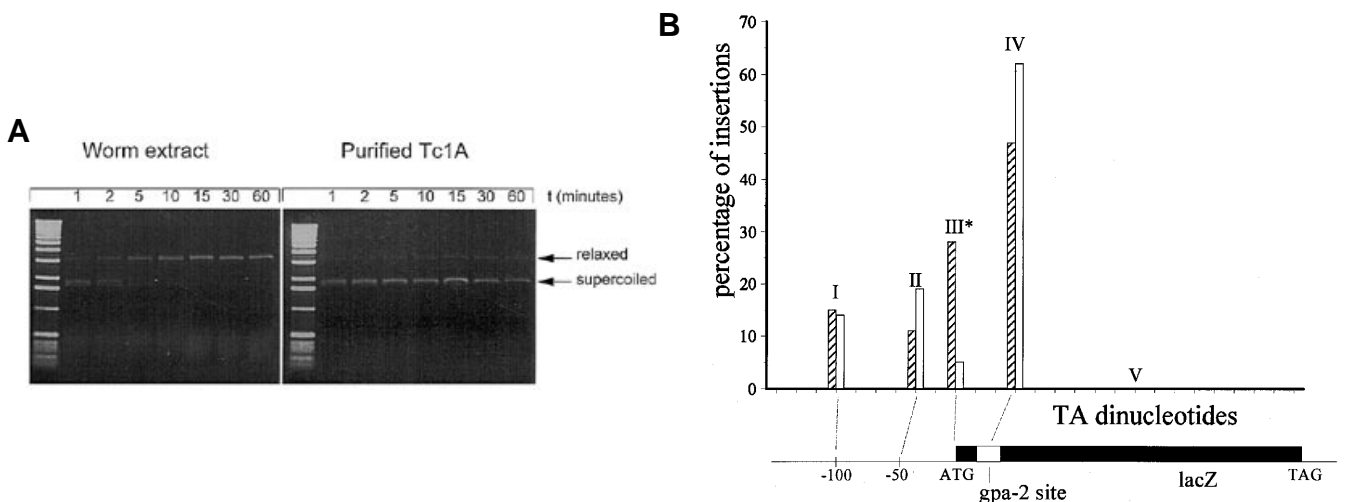


Figure 5. Supercoiling affects Tc1 integration. (A) Nuclease activity in Tc1A preparations. The left panel shows a time series of supercoiled plasmid (pRP1218) incubated with nuclear worm extract and the right panel a time series with Tc1A protein purified from *E.coli*. (B) Tc1 insertion distribution into supercoiled (hatched bars) and relaxed (white bars) target plasmid. Significant differences are indicated with an asterisk.

Tc1 integration and target DNA structure

One site, site III, present in the promoter region overlapping the start codon of *lacZ*, was found to be differentially used when the transposase preparation was either a nuclear worm extract or Tc1A purified from *E.coli*. A possible explanation could be a difference in nuclease activity between the two protein sources, leading to relaxation of the target plasmid within minutes when using the crude nuclear worm extract. When the target DNA is incubated with Tc1A purified from *E.coli* no relaxation was observed. We showed that this difference in DNA superstructure indeed affects Tc1 integration, but only into one of the potential integration sites. Integration efficiency into the other sites of *lacZ* is not affected. The efficiency of transposition (Table 1) indicates that supercoiling does not greatly stimulate the integration reaction, similar to what was found for the bacterial transposon Tn7 and bacteriophage Mu, where the strand transfer reaction can be carried out *in vitro* using oligonucleotides (45,46). For Tn10 it has been speculated that the helical structure in the vicinity of the integration consensus sequence is of importance for integration efficiency (25), and DNA supercoiling has been shown to be essential for efficient Tn5 integration (26). A differential effect of supercoiling on different potential integration sites has also been reported for Tn5 (47). This study showed that negative supercoiling was necessary for efficient integration into one site, but not for integration into another site located only 41 bp downstream of the first. This situation very much resembles the results found here. Apparently, supercoiling affects the sequences around site III making it a good substrate for Tc1 integration.

Possibly this effect can also explain the effect of changes in nucleotide sequence on the usage of certain TA dinucleotides up to 120 bp away (for example Fig. 1, compare A with B). These changes in nucleotide sequence could affect the supercoiling structure of the plasmid (48,49), and thus affect the usage of TA dinucleotides all over the plasmid DNA. These results also suggest that transcription may have an effect on Tc1 integration, as transcription transiently changes the supercoiling structure of DNA (50).

We conclude that the observed insertion distributions are a consequence of interactions between target DNA and the Tc1A protein alone, and that the effects of mutations introduced into the flanks, reflect an altered interaction between target DNA and Tc1A protein. All critical interactions are within a sequence of 5 bp: the TA plus the three 3' flanking nucleotides. Probably, the base pair 4 nt downstream of the TA has weak, but important interactions with the transposase complex. In addition, the interaction between target DNA and the Tc1 transpososome can be influenced by the structural status of the DNA.

ACKNOWLEDGEMENTS

We thank G.Verlaan for Tc1A protein purified from *E.coli*. We thank H.G.A.M.van Luenen and P.Borst for critical reading of the manuscript. R.F.K. is supported by grant 700-35-210 from NWO/SON.

REFERENCES

- Heierhorst, J., Lederis, K. and Richter, D. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 6798–6802.
- Henikoff, S. (1992) *New Biol.*, **4**, 382–388.

- Robertson, H.M. (1993) *Nature*, **362**, 241–245.
- Robertson, H.M. (1995) *J. Insect Phys.*, **41**, 99–105.
- Plasterk, R.H.A. (1996) In Saedler, H. and Gierl, A. (eds) *Transposable Elements*. Springer-Verlag, Berlin Heidelberg, pp. 125–143.
- Auge-Gouillou, C., Bigot, Y., Pollet, N., Hamelin, M., Meunier-Rotival, M. and Periquet, G. (1995) *FEBS Lett.*, **368**, 541–546.
- Morgan, G.T. (1995) *J. Mol. Biol.*, **254**, 1–5.
- Oosumi, T., Belknap, W.R. and Garlick, B. (1995) *Science*, **378**, 672.
- Reiter, L.T., Murakami, T., Koeth, T., Pentao, L., Munzy, D.M., Gibbs, R.A. and Lupski, J.R. (1996) *Nature Genet.*, **12**, 288–297.
- Lampe, D.J., Churchill, M.E.A. and Robertson, H.M. (1996) *EMBO J.*, **15**, 5470–5479.
- Vos, J.C., de Baere, I. and Plasterk, R.H.A. (1996) *Genes Dev.*, **10**, 755–761.
- Ikenaga, H. and Saigo, K. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 4143–4147.
- Inouye, S., Yuki, S. and Saigo, K. (1984) *Nature*, **310**, 332–333.
- Freund, R. and Meselson, M. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4462–4464.
- Marlor, R.L., Parkhurst, S.M., Corces, V.G. (1986) *Mol. Cell. Biol.*, **6**, 1129–1134.
- Tudor, M., Lobočka, M., Goodell, M., Pettitt, J. and O'Hare, K. (1992) *Mol. Gen. Genet.*, **232**, 126–134.
- O'Hare, K. and Rubin, G.M. (1983) *Cell*, **34**, 25–35.
- Lodge, J.K., Weston-Hafer, K. and Berg, D.E. (1988) *Genetics*, **120**, 645–650.
- Mizuuchi, K. and Craigie, R. (1986) *Annu. Rev. Genet.*, **20**, 385–429.
- Tenzen, T., Matsutani, S. and Ohtsubo, E. (1990) *J. Bacteriol.*, **172**, 3830–3836.
- Davies, C.J. and Hutchinson, C.A., III (1995) *Nucleic Acids Res.*, **23**, 507–514.
- Halling, S.M. and Kleckner, N. (1982) *Cell*, **28**, 155–163.
- Craig, N.L. (1991) *Mol. Microbiol.*, **5**, 2569–2573.
- Hallet, B., Rezsóhazy, R., Mahillon, J. and Delcour, J. (1994) *Mol. Microbiol.*, **14**, 131–139.
- Bender, J. and Kleckner, K. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 7996–8000.
- Isberg, R.R. and Syvanen, M. (1982) *Cell*, **30**, 9–18.
- Pryciak, P.M. and Varmus, H.E. (1992) *Cell*, **69**, 769–780.
- Pryciak, P.M., Sil, A. and Varmus, H.E. (1992) *EMBO J.*, **11**, 291–303.
- Curcio, M.J. and Morse, R.H. (1996) *Trends Genet.*, **12**, 436–438.
- Tsubota, S., Ashburner, M. and Schedl, P. (1985) *Mol. Cell. Biol.*, **5**, 2567–2574.
- Kelley, M.R., Kidd, S., Berg, R.L. and Young, M.W. (1987) *Mol. Cell. Biol.*, **7**, 1545–1548.
- Craigie, R. (1992) *Trends Genet.*, **8**, 187–190.
- McKinnon, R.D., Waye, J.S., Bautista, D.S. and Graham, F.L. (1985) *Gene*, **40**, 31–38.
- Casadesus, J. and Roth, J.R. (1989) *Mol. Gen. Genet.*, **216**, 210–216.
- Pato, M.L. (1989) In Berg, D.E. and Howe, M.M. (eds) *Mobile DNA*. American Society for Microbiology, Washington, DC pp. 23–52.
- Wolkow, C.A., DeBoy, R.T. and Craig, N.L. (1996) *Genes Dev.*, **10**, 2145–2157.
- Van Luenen, H.G.A.M., Colloms, S.D. and Plasterk, R.H.A. (1994) *Cell*, **79**, 293–301.
- Van Luenen, H.G.A.M. and Plasterk, R.H.A. (1994) *Nucleic Acids Res.*, **22**, 262–269.
- Eide, D. and Anderson, P. (1988) *Mol. Cell. Biol.*, **8**, 737–746.
- Mori, I., Benian, G.M., Moerman, D.G. and Waterston, R.H. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 861–864.
- Korswagen, H.C., Smits, M.T., Durbin, R.M. and Plasterk, R.H.A. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 14680–14685.
- Zwaal, R.R., Broeks, A., Van Meurs, J., Groenen, J.T.M. and Plasterk, R.H.A. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 7431–7435.
- Bender, J. and Kleckner, N. (1992) *EMBO J.*, **11**, 741–750.
- Shibagaki, Y. and Chow, S.A. (1997) *J. Biol. Chem.*, **272**, 8361–8369.
- Bainton, R.J., Kubo, K.M., Feng, J. and Craig, N.L. (1993) *Cell*, **72**, 931–943.
- Savilahti, H., Rice, P.A. and Mizuuchi, K. (1995) *EMBO J.*, **14**, 4893–4903.
- Lodge, J.K. and Berg, D.E. (1990) *J. Bacteriol.*, **172**, 5956–5960.
- Ellison, M.J., Fenton, M.J., Ho, P.S. and Rich, A. (1987) *EMBO J.*, **6**, 1513–1522.
- Sproul, D. and Harvey, S.C. (1996) *Biophys. J.*, **70**, 1893–1908.
- Liu, L.F. and Wang, J.C. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 7024–7027.