Target site choice of the related transposable elements Tc1 and Tc3 of Caenorhabditis elegans

Henri G.A.M. van Luenen and Ronald H.A. Plasterk*
The Netherlands Cancer Institute, Division of Molecular Biology, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

Received November 16, 1993; Revised and Accepted January 4, 1993

ABSTRACT
We have investigated the target choice of the related transposable elements Tc1 and Tc3 of the nematode C. elegans. The exact locations of 204 independent Tc1 insertions and 166 Tc3 insertions in an 1 kbp region of the genome were determined. There was no phenotypic selection for the insertions. All insertions were into the sequence TA. Both elements have a strong preference for certain positions in the 1 kbp region. Hot sites for integration are not clustered or regularly spaced. The orientation of the integrated transposon has no effect on the distribution pattern. We tested several explanations for the target site preference. If simple structural features of the DNA (e.g. bends) would mark hot sites, we would expect the patterns of the two related transposons Tc1 and Tc3 to be similar; however we found them to be completely different. Furthermore we found that the sequence at the donor site has no effect on the choice of the new insertion site, because the insertion pattern of a transposon that jumps from a transgenic donor site is identical to the insertion pattern of transposons jumping from endogenous genomic donor sites. The most likely explanation for the target choice is therefore that the primary sequence of the target site is recognized by the transposase. However, alignment of the Tc1 and Tc3 integration sites does not reveal a strong consensus sequence for either transposon.

INTRODUCTION
Transposable elements insert at many different positions in the genome, but the integrations are not necessarily randomly distributed over the genome. A detailed analysis of the insertion specificity of a transposon may be revealing with respect to the transposition mechanism of the element, since the integration preference may reflect how the integration complex interacts with the target sequence.

A limited number of transposable elements insert into a specific sequence. The closely related transposable elements Tc1 and Tc3 of Caenorhabditis elegans always integrate into the dinucleotide TA (1–5). Other Tc elements (Tc2 and Tc6) from C. elegans and members of the class of Tc1-like elements found in arthropods and vertebrates also integrate only into TA dinucleotides (6–11). The copia-like elements 17.6 and 297 always insert into the sequence ATAT and TATATA respectively (12, 13). Gypsy integrates into TACA (14) and pogo into TA (15). The Tc1 transposable element found in the ciliate Euplotes crassus most likely integrates into the sequence TA although only excision of the element has been detected and no integration (16). The IS630 associated transposons are thus far the only prokaryotic transposable elements which integrate at many different positions but always into TA (17).

Elements with a specific target sequence do not integrate into each possible site with equal frequency. This suggests that the element looks at the sequences flanking the integration site. A consensus sequence for the flanking sequence of Tc1 integration sites (G A G/T A G T A T/C G/C T) has been proposed based on 16 germ line insertion events (2, 3).

Hot spots for integration have also been observed for elements which do not integrate into a specific sequence. The consensus insertion sequence for P elements is GGCCAGAC (18). However, some hot spots for P integration have no homology to this consensus sequence whereas insertions into some potential integration sites, which perfectly fit the consensus sequence, have never been isolated (19). The 6–9 bp flanking the Tn10 consensus insertion sequence (20) can confer variations of at least 1000-fold in insertion frequency (21). However, no discernible consensus for the Tn10 flanking sequence has been found.

The primary sequence of the integration site but also the functional state of the DNA (e.g. transcription, replication) affects the distribution of integration sites. P element insertions are preferentially found at or near transcription start sites (22, 23). The Tn5 transposon integrates more frequently into regions with increased negative supercoiling (24, 25), and in the vicinity of promoter regions of transcribed genes (26, 27). Insertions of the yeast elements Ty1 and Ty3 are preferentially into regions containing tRNA genes, long terminal repeats or pre-existing transposable elements (28, 29; see also review by F. Bushman, 30). Tn10 (31) as well as Mu (32) preferentially integrate into non-transcribed regions. Retroviral integrations are mostly into DNAseI hypersensitive sites which are associated with transcribed regions of the genome (33–38; see also review by R. Craigie,)* To whom correspondence should be addressed
39). The retroviral target site selection is furthermore affected by the positioning of nucleosomes on the DNA (40, 41).

We have analyzed the distribution of Tc1 and Tc3 integrations in an 1 kbp region of the genome. For this purpose we used a PCR based detection system which does not involve phenotypic selection for insertions and is therefore unbiased for the distribution of insertions. We sequenced 370 independent, somatic insertion events. All transposons were found to be integrated into the sequence TA. The integrations of Tc1 and Tc3 were not randomly distributed among the 82 TA dinucleotides within the 1 kbp region. We considered several explanations for the observed strong target site preference. 1. Does the local structure of the DNA make certain sites preferred targets for transposon insertion? 2. Does the donor site influence the choice of the target site? 3. Does the sequence of the target site make some sites hot spots for integration?

**MATERIALS AND METHODS**

Generation of transgenic animals with Tc3Δkan′ transposon

A Tc3 element has been cloned after insertion into the unc-22 gene (TR #10; 4). An EcoRI fragment from pTR #10 containing 239 bp flanking unc-22 sequence and 694 bp Tc3 sequence (containing the complete inverted repeat) and a BamHI fragment from pTR #10 containing 391 bp flanking unc-22 sequence and 859 bp Tc3 sequence (containing the other inverted repeat) were cloned into pUC18. A HincII fragment from pUC4K (Pharmacia LKB Biotechnology) containing the aminoglycoside 3′-phosphotransferase gene was cloned in the Smal site in between the two pTR #10 fragments. The resulting plasmid (pRP749) was injected into the gonads of Bristol N2 nuc-1 animals (150 µg/ml) (42) together with 50 µg/ml rol-6(p) (pRP74) (43) and 5 µg/ml m pRP716 (5). The resulting semi-stable transgenic line was irradiated with a γ source (137Cs, 40 Gray) in order to induce integration of the extrachromosomal array (which contains the plasmids) into the genome. After several generations a stable transgenic line was obtained (NL266).

DNA isolation

DNA from the transgenic lines NL224 [expressing Tc3 transposase after heat shock induction (44)] NL261 [expressing Tc3 transposase after heat shock induction (5)] and NL266 was isolated as described by Sulston and Hodgkin (45). DNA from a collection of frozen pools of animals (strain MT3126) was isolated as described by Zwaal et al. (46). Transposase expression in the transgenic lines was induced by incubating the worms for 2 hours at 33°C and a subsequent recovery at room temperature for 4 to 12 hours.

PCR detection and sequence analysis of insertions

Transposon insertions in an 1 kbp region of the gpa-2 gene (47) were detected by two rounds of PCR. In the first reaction, genomic DNA was used in combination with a gpa-2 specific primer (AB3550 5′gaacctcataaccaactcc) and a transposon specific primer [Tc1: L1 5′gaggttatcctggcgaagacgctcttgaag; Tc2: AB2730 5′eggaaatctccaaactccc]. A 25 µl PCR reaction contained 1 µl genomic DNA (50–300 ng, depending on the number of insertions found per PCR), 10 pmoles of each primer, 0.05 M KCl, 0.01 M Tris–Hcl pH9, 1.5 mM MgCl2, 0.2 mM dNTPs and 0.5 units Taq DNA polymerase (Gibco BRL). 30 cycles were performed; each cycle 1° 94°C, 1° 55°C and 1° 72°C. The PCR products were analyzed on an 1% agarose gel and the bands were excised using pasteur pipettes. DNA was eluted from the agarose plugs in 10 µl water for several hours. The DNA (2 µl per reaction) was used in a PCR based sequencing method with 5′ 32P labelled L2, R2 or 2730 primers as described by Craxton (48). Sequence reactions were run on 6% denaturing polyacrylamide gels. Tc3Δkan′ insertions in gpa-2 were detected using a primer in the aminoglycoside 3′-phosphotransferase gene (5′ aaagagctcataacccggtctg) instead of the Tc3 inverted repeat primer AB2745 in the first PCR.

**RESULTS**

Determination of a large number of independent transposon insertion sites

Transposon insertions in C.elegans can be detected using PCR. Insertions were detected in genomic DNA, using a transposon specific primer and a primer with homology to a chosen genomic sequence (see Material and Methods and Fig. 1). The region between the primers is exponentially amplified if a transposon has integrated close to the location of the genomic primer. A second PCR with nested primers is performed to increase the sensitivity and the specificity of this assay. This method is sensitive enough to detect single molecules as has been shown with limiting dilution series.

The genomic primers and PCR conditions were chosen such that insertions were detected in a region of approximately 1 kbp of the genome containing the gpa-2 gene (47). This gene is expressed in only a few neurons (pers. comm. J.Mendel and P.Sternberg). For the detection of Tc1 insertions we used primers specific for either the left or the right end of Tc1. In this way we could independently determine the distribution of Tc1

![Figure 1](https://www.nature.com/nrx/journal/v22/n3/images/nrc21263f1.jpg)
insertions in both orientations into the gpa-2 gene (Fig. 1A). For Tc3 we used PCR primers located in the terminal inverted repeats; as a consequence, both orientations of the Tc3 insertions are amplified in the same reaction (Fig. 1B). The amount of DNA used in the first PCR was adjusted so that we got on average less than one PCR product per reaction. This is to prevent selection of smaller products which are better amplified.

DNA for the PCR detection of Tc1 insertions was obtained from two different sources. The major source of Tc1 insertions was a large collection of independent pools of frozen animals.

![Figure 2](link-to-figure)

**Figure 2.** Comparison of the distribution of Tc1 insertions detected using either the R primer set or the L primer set. Every mark on the X-axis represent a TA dinucleotide: a potential target site for Tc1 insertion. The first TA dinucleotide indicated on the X-axis is located at position 2058 and the last at position 3098. The exact positions and flanking sequences of the TA dinucleotides containing insertions are listed in table I. The intron/exon structure as well as the actual position of the TA dinucleotides within the sequence are indicated underneath the X-axis. The two sites most closely resembling the consensus sequence of Eide and Anderson (2) and Mori et al. (3) are marked I and II.

![Figure 3](link-to-figure)

**Figure 3.** Comparison of the distribution of Tc1 insertions (detected using the L primer set) and Tc3 insertions (see also legend figure 2).
animals. Aliquots of these DNA samples were used in several large independent populations of heat shock induced NL224 transposition in this strain (49). Genomic DNA was isolated from transposition (44). There is no detectable germ line Tel promoter. The Tel transposase is produced in somatic cells after induction, and results in an increased level of somatic Tel expression, and happens 4 hours before the induction of transposition, which took place 4 hours before DNA was isolated within hours of the heat shock induction and DNA was isolated from these cultures and again the insertions could have occurred. That the insertions from one DNA sample are thus almost certainly independent.

This collection has been used to isolate germ line Tc1 insertions in a variety of genes (46). DNA originating from a different pool of animals was used for each PCR; thus all the insertions are independent events. The Tc1 insertions can either be of somatic or germ line origin. The other source of Tc1 insertions was a transgenic Bristol N2 line (NL224; 44). This line contains the Tel transposase under the control of a heat shock inducible promoter. The Tc1 transposase is produced in somatic cells after induction, and results in an increased level of somatic Tc1 transposition (44). There is no detectable germ line Tc1 transposition in this strain (49). Genomic DNA was isolated from 6 large independent populations of heat shock induced NL224 animals. Aliquots of these DNA samples were used in several PCRs. The insertions obtained were in principle independent, because the large majority of detected insertions arose as a result of the heat shock induction and DNA was isolated within hours after the induction of transposition, so no extensive replication could have occurred. That the insertions from one DNA sample represent independent events is illustrated by the fact that the distribution patterns obtained from the different DNA samples is similar. Tc3 insertions were obtained from several independent cultures of transgenic Bristol N2 animals containing a transgene which expresses Tc3 transposase upon induction (5). There is no endogenous Tc3 transposition (germ line or somatic) in Bristol N2 (4, 5). The insertions that we analyzed are therefore caused by the transposase induction, which took place 4 hours before DNA was isolated from these cultures and again the insertions are thus almost certainly independent.

Table I. Tc1 and Tc3 insertions in the gpa-2 gene

<table>
<thead>
<tr>
<th>Tc1 primer set L</th>
<th>Tc1 primer set R</th>
<th>Tc3</th>
</tr>
</thead>
<tbody>
<tr>
<td>position</td>
<td>sequence</td>
<td>insertions</td>
</tr>
<tr>
<td>2174</td>
<td>ATGAACTACCTAT</td>
<td>1</td>
</tr>
<tr>
<td>2178</td>
<td>ATACCTATTGTG</td>
<td>3 (2)</td>
</tr>
<tr>
<td>2187</td>
<td>TGGCTATTGTG</td>
<td>1</td>
</tr>
<tr>
<td>2220</td>
<td>TGGCTATTGTG</td>
<td>0 (1)</td>
</tr>
<tr>
<td>2222</td>
<td>TGATATATTTG</td>
<td>1</td>
</tr>
<tr>
<td>2251</td>
<td>GAACTTTAAA</td>
<td>0 (3)</td>
</tr>
<tr>
<td>2331</td>
<td>CATGATCTATTC</td>
<td>1</td>
</tr>
<tr>
<td>2542</td>
<td>CATGATCTATTC</td>
<td>1 (1)</td>
</tr>
<tr>
<td>2546</td>
<td>CATGATCTATTC</td>
<td>2 (3)</td>
</tr>
<tr>
<td>2548</td>
<td>CATGATCTATTC</td>
<td>1</td>
</tr>
<tr>
<td>2644</td>
<td>CATGATCTATTC</td>
<td>2 (2)</td>
</tr>
<tr>
<td>2838</td>
<td>CATGATCTATTC</td>
<td>8 (2)</td>
</tr>
<tr>
<td>2855</td>
<td>CATGATCTATTC</td>
<td>3</td>
</tr>
<tr>
<td>2862</td>
<td>CATGATCTATTC</td>
<td>26 (12)</td>
</tr>
<tr>
<td>2868</td>
<td>CATGATCTATTC</td>
<td>4 (2)</td>
</tr>
<tr>
<td>2870</td>
<td>ATGAACTACCTAT</td>
<td>1</td>
</tr>
<tr>
<td>2870</td>
<td>ATGAACTACCTAT</td>
<td>1</td>
</tr>
<tr>
<td>2884</td>
<td>ATGAACTACCTAT</td>
<td>1</td>
</tr>
<tr>
<td>2892</td>
<td>ATGAACTACCTAT</td>
<td>1</td>
</tr>
<tr>
<td>2900</td>
<td>ATGAACTACCTAT</td>
<td>2</td>
</tr>
<tr>
<td>2906</td>
<td>ATGAACTACCTAT</td>
<td>1</td>
</tr>
<tr>
<td>2912</td>
<td>ATGAACTACCTAT</td>
<td>1</td>
</tr>
<tr>
<td>2918</td>
<td>ATGAACTACCTAT</td>
<td>2</td>
</tr>
<tr>
<td>2940</td>
<td>ATGAACTACCTAT</td>
<td>1</td>
</tr>
<tr>
<td>2948</td>
<td>ATGAACTACCTAT</td>
<td>1</td>
</tr>
<tr>
<td>2952</td>
<td>ATGAACTACCTAT</td>
<td>1</td>
</tr>
<tr>
<td>2956</td>
<td>ATGAACTACCTAT</td>
<td>1</td>
</tr>
<tr>
<td>2964</td>
<td>ATGAACTACCTAT</td>
<td>1</td>
</tr>
<tr>
<td>2970</td>
<td>ATGAACTACCTAT</td>
<td>1</td>
</tr>
<tr>
<td>2970</td>
<td>ATGAACTACCTAT</td>
<td>1</td>
</tr>
<tr>
<td>2970</td>
<td>ATGAACTACCTAT</td>
<td>1</td>
</tr>
<tr>
<td>3068</td>
<td>ATGAACTACCTAT</td>
<td>1</td>
</tr>
<tr>
<td>3088</td>
<td>ATGAACTACCTAT</td>
<td>1</td>
</tr>
<tr>
<td>3088</td>
<td>ATGAACTACCTAT</td>
<td>1</td>
</tr>
<tr>
<td>3088</td>
<td>ATGAACTACCTAT</td>
<td>1</td>
</tr>
<tr>
<td>3088</td>
<td>ATGAACTACCTAT</td>
<td>1</td>
</tr>
</tbody>
</table>

1The positions correspond to the sequence of gpa-2 published by Fino Silva and Plasterk (47).
2The nucleotides in bold indicate the site where Tc1 or Tc3 inserted into the gpa-2 gene.
3The number in brackets are the insertions found after heat shock induction of the Tc1 transposase in the transgenic line.
4The numbers in brackets are the insertions found for the Tc3Δkan4 transposon.
Tel insertions are not randomly distributed

The exact positions of 95 independent Tc3 insertions and 166 independent Tel insertions were determined (Table I). All the insertions took place at a TA dinucleotide. This further illustrates that Tel and Tc3 only integrate into TA.

The distribution of the 166 Tel insertions among the 82 TA dinucleotides present within the 1 kbp region of the gpa-2 gene is depicted in Figure 2. The Tel insertions have a non-random distribution. Some sites are frequently used whereas other sites are seldom or never used. Frequently used target sites are not clustered or regularly spaced. A frequently used TA dinucleotide for Tel integration at position 2564 (containing 13 insertions) is directly flanked by another TA dinucleotide (position 2562) having no insertions. This shows that the 'hotness' is not a regional property of the DNA but rather determined by the primary sequence flanking the TA dinucleotide.

Tel integration site preference is not influenced by the orientation of the transposon

We used two sets of primers for the detection of Tel insertions (Fig. 1A). Each set specifically detects Tel insertions in one orientation with respect to the gpa-2 gene. Figure 2 shows the distribution of both orientations separately. The distribution pattern of insertions is the same for both orientations of the transposon. This suggests that the integration complex is symmetrical with respect to Tel.

Tel and Tc3 have different integration site preferences

We determined whether the Tel and Tc3 integration patterns differ. 95 Tc3 insertion sites were determined and compared to the distribution of the Tel insertions (Fig. 3). The distribution of Tel and Tc3 insertions is different. The difference in insertion patterns could reflect a different target choice of the two elements, or it might be caused by the fact that the Tc3 insertions were obtained from worms in which transposase was expressed from a transgene after heat shock induction, whereas the Tel insertions were mainly obtained from a strain with endogenous transposition activity. The heat shock treatment could have affected the structure of the DNA and thus result in a different insertion pattern, or the high expression level of transposase after heat shock induction could possibly have affected the integration site preference. Furthermore, the heat shock promoter is not activated in all cells with the same efficiency (50). This could also be the case for the endogenous transposase expression, and we would then be comparing insertion patterns from different cell populations. To study whether any of these factors was responsible for the observed difference we compared the integration sites of 38 Tel insertions obtained from a transgenic line in which transposition was induced by heat shock to the insertion pattern of the 'natural' Tel mutator strain (76 insertions) (Table I and Fig. 4). The distribution of heat shock induced Tel insertions is similar to that of Tel insertions in the mutator strain, and not to that of heat shock induced Tc3 insertions (Fig. 3). This confirms that the difference in integration patterns between Tel and Tc3 is transposon specific, since the difference is also observed when both transposases are expressed in an identical fashion.

Distribution of Tc3 insertions originating from one donor site

Thus far we have shown that Tel and Tc3 have different integration site preferences but they always integrate into TA. The non-random distribution of the Tel and Tc3 insertions is probably not caused by the local structure of the DNA; we would expect a similar distribution for Tel and Tc3 if, e.g., bent DNA was the sole determinant for transposon insertion.

What is then the cause of the non-random and transposon specific distribution of insertions? The difference in target choice could be explained if the donor site (the site where a transposon is integrated before it excises and integrates somewhere else) determines the choice of insertion site. Mechanistically one could for example imagine that the flanking sequences at the donor site interact with the target site prior to excision and integration. If
that were the case then the difference in target choice between Tcl and Tc3 could be caused by the different sequences flanking the transposons at the donor sites of the 50–100 endogenous Tcl elements and the approximately 15 endogenous Tc3 elements. To test whether the target choice of Tc3 is influenced by the sequence of the flanking DNA at the donor site we made a transgenic C. elegans line containing a marked Tc3 transposon (Tc3Δkan^r). If the target choice of endogenous Tc3 elements reflects the sequences at the approximately 15 donor sites in the genome, then one would expect a different distribution of insertion sites when a marked Tc3 transposon located at one specific donor site is used. Using a PCR primer in the gene coding for kanamycin resistance, which replaces an internal Tc3 sequence in Tc3Δkan^r, we specifically amplified Tc3Δkan^r insertions into gpa-2. The distribution of insertions of the marked Tc3 transposon coming from a single donor site, was compared to the distribution of Tc3 insertions originating from genomic Tc3 donor sites (Fig. 5). No difference between the two distribution patterns was found. This shows that the donor site does not determine the choice of insertion site.

Sequence alignment of the Tcl and Tc3 integration sites
A third explanation for the specific target choice could be that the primary sequence at the insertion site is recognized by the incoming transposon. One can imagine that the Tcl and Tc3 transposons prefer to interact with different nucleotides at the target site. We aligned the Tcl and Tc3 insertion sites to determine whether a consensus sequence for Tcl and Tc3 integration could be derived (Table II). The Tcl insertions obtained in the different experiments (left/right orientation and with/without heat shock) were combined because the distribution of insertions was identical; the same has been done for the Tc3 insertions. The insertion sites were aligned according to the 5′–3′ orientation of the gpa-2 gene and not to the orientation of the transposon [since the insertion preference of Tcl was found to be independent of the transposon orientation (Fig. 2)]. Table IIa shows the occurrence of each nucleotide in the region from 5 nucleotides upstream to 5 nucleotides downstream of the TA dinucleotide for either Tcl (all 24 different integration sites) or Tc3 (all 29 different integration sites). In this comparison the preference for one site over the other is ignored; a site for which one insertion is found contributes as much as a frequently used integration site. To correct for this, in Table IIb, the nucleotide usage is given for all the 204 Tcl insertions and all the 166 Tc3 insertions. Although the sequences of the hot sites for integration contribute more to the nucleotide distribution they do not exclusively determine them, e.g. a G at position +2 is found in 134 out of the 204 Tcl insertion events and the hot spot for Tcl integration contributes only 62 times to this score. So even without taking into account the hot site, the G at position +2 is still often found (72 out of 142 insertions). Whether a consensus sequence for Tcl and Tc3 integration can be deduced from our data will be discussed in the next section.

DISCUSSION
We have determined the exact locations of a large series of independent, unselected Tcl and Tc3 transposon insertions in an 1 kbp region of the genome. The integrations were not randomly distributed over this area; a few integration sites were frequently used, whereas others were not used. We have considered several explanations for the strong target site preference.
All Tcl and Tc3 insertions took place at a TA dinucleotide. This absolute requirement for the sequence TA could be the result of specific sequence recognition by the transposase or result from the way in which the elements integrate into the genome (e.g. integration via base pairing between short single stranded transposon ends and the target site). The latter explanation seems to be excluded because the ends of excised linear Tc3 elements do not contain the TA sequence (our unpublished observations).

The orientation of the Tcl transposon does not affect the choice of integration site. This has also been noted for Ty3 and Tn5 (27, 51, 52) as well as for in vitro retroviral integration (41),
although analysis of \textit{in vivo} retroviral integration sites indicates that there might be an orientation dependence (36). The orientation independent distribution of Tc1 insertions suggests that the integration complex or the integration site is symmetrical. We favor the first option since there is hardly any sequence symmetry in the insertion sites we have obtained (Table I). The transposon sequence is therefore most likely incorporated in a transposition complex independent of its orientation.

What determines the frequency at which a TA dinucleotide is used for Tc1 or Tc3 integration: the tertiary structure of the DNA, the donor site, the primary sequence flanking the TA dinucleotide, or a combination of these factors? Since hot spots for integration are not clustered or regularly spaced and since Tc1 and Tc3 have different insertion preferences we presume that the regional structure of the DNA does not determine the target site specificity unless Tc1 and Tc3 recognize completely different structural features. We have not studied whether chromatin structure has an effect on the accessibility of target sites. This regional effect of the chromatin structure has been noted for several integration systems, e.g. viral integrations are frequently in transcribed regions of the genome compared to non-transcribed regions (33—38). Our study does not focus on the regional distribution of Tc insertions but on the local distribution within a small region of 1 kbp. This region is only expressed in a few neurons, the majority of cells does not express this sequence. We can not exclude that the integration pattern is affected by the transcription in these few cells.

The generation of a transgenic line containing a marked Tc3 transposon with unique flanking sequences allowed us to investigate the effect of the donor site on the target selection. No effect was found. Based on this result there is no need to assume a direct sequence interaction between the donor and the target site during transposition.

The most likely explanation for the local target site preference is the recognition of the sequence flanking the TA dinucleotide by the incoming transposon. We investigated whether we could recognize a pattern of preferred sequences in the close vicinity of the insertion site. In other words: is there a consensus sequence for Tc1 and Tc3 integration? To answer this question we have aligned all the Tc1 and Tc3 insertion sites according to the 5′—3′ orientation of the \textit{gpa}-2 gene (Table II). This alignment is only meaningful if the integrating transposon knows the orientation of the gene, which is very unlikely. Therefore one has to consider the insertion sites in Table II randomly ordered. We have no criteria by which we could order the insertion sites more meaningfully. If there is a strong consensus sequence for integration we should still be able to extract it from Table II because a consensus sequence should be palindromic (e.g. if an A at position +1 is part of a consensus sequence we expect a similarly preferred T at position —1 due to the random ordering of the sequences). With this in mind we can not extract an obvious consensus sequence from Table II. We have no indications that specific nucleotides are preferentially absent in the flanking sequence of the integration sites and we have also found no consensus sequence for the cold TA sites (without insertions) (data not shown). It is possible that a combination of 2 or 3 nucleotides define a consensus sequence, or that sequences located further away (e.g. one turn of the helix) are important but that is very difficult to determine. A and T nucleotides are more frequently found than G or C nucleotides, possibly, because Tc1 and Tc3 are more frequently found than Tc1 and Tc3, possibly, because Tc1 and Tc3 integrate into the dinucleotide TA which is often found in A/T rich regions, such as introns.

Eide and Anderson (2) and Mori et al. (3) have proposed a consensus sequence for Tc1 integration (G A G/T A/G T A/T C G/ C T). We compared the sequences of the integration sites obtained in our study with the consensus sequence. Although there was only a very weak correlation between the number of insertions at a certain TA dinucleotide and the match of the flanking DNA with the consensus sequence, it is striking that the hot site for Tc1 integration in this region (62 of the 204 insertions; marked I in figure 2) has a match of 8 out of 9 with the Tc1 consensus sequence and that the other site (marked II in figure 2) with a 8 out of 9 match contains 18 of 204 insertions.

This is the first eukaryotic system in which the integration patterns of two transposons within a short region of the genome have been determined at the sequence level. Before the days of PCR only a few insertion sites were determined at the nucleotide level and these sites were mostly selected because they led to a phenotype. Our study shows that two closely related transposons have strong but different integration preferences even within a very small region of the genome. The mechanistic basis of target preference is not clear. The difference between Tc1 and Tc3 target site preferences argues against structural motifs (bends, melting of DNA) as determinants for transposon insertions (unless Tc1 and Tc3 have different liking for these). A simple explanation, that the target site choice reflects the sequence flanking the donor site, could be eliminated by demonstration that a transgenic element has an unaltered target preference. The

\begin{dcolumn}
\begin{tabular}{|c|c|c|c|c|c|c|c|c|}
\hline
 & -5 & -4 & -3 & -2 & -1 & TA & +1 & +2 \\
\hline
\textit{gpa}-2::Tc3 insertion sites (n = 29) & & & & & & & & \\
G & 4 & 7 & 6 & 2 & 6 & 1 & 5 & 3 \\
A & 8 & 9 & 10 & 11 & 8 & 3 & 4 & 15 \\
T & 11 & 7 & 11 & 14 & 4 & 20 & 16 & 15 \\
C & 6 & 6 & 2 & 2 & 15 & 5 & 4 & 3 \\
\hline
\end{tabular}
\end{dcolumn}

\begin{dcolumn}
\begin{tabular}{|c|c|c|c|c|c|c|c|c|}
\hline
 & -5 & -4 & -3 & -2 & -1 & TA & +1 & +2 \\
\hline
\textit{gpa}-2::Tc3 insertion events (n = 166) & & & & & & & & \\
G & 11 & 54 & 13 & 5 & 15 & 1 & 5 & 3 \\
A & 53 & 42 & 62 & 7 & 4 & 2 & 1 \\
T & 66 & 23 & 82 & 58 & 7 & 132 & 111 & 86 \\
C & 36 & 47 & 22 & 14 & 74 & 29 & 7 & 0 \\
\hline
\end{tabular}
\end{dcolumn}
most likely explanation for the target site preference is the recognition of the sequence flanking the TA dinucleotide by the transposition complex. We could however not detect a consensus sequence. The conclusion at this stage can only be that there is recognition of the sequence flanking the TA dinucleotide by the

ACKNOWLEDGEMENTS

We thank P. Anderson for the sequence of Tc3 and pTR #10, J. Kramer for pRF4 and S. Colloms for cloning pRP749. We thank J. Mendel and P. Sternberg for information about the expression of gpa-2. We also thank our colleagues for their valuable advice and P. Borst, S. Colloms and J. C. Vos for critical reading of the manuscript. This work was supported by a Pioneer grant of the Netherlands Organization for Scientific Research N.W.O.

REFERENCES