



Integration specificity of the *hobo* element of *Drosophila melanogaster* is dependent on sequences flanking the integration site

Kenneth J. Saville¹, William D. Warren², Peter W. Atkinson³ & David A. O'Brochta⁴

¹Department of Biology, Albion College, Albion, MI 49224, USA (Phone: 517-629-0388; Fax: 517-629-0509; E-mail: ksaville@albion.edu); ²Peter MacCallum Cancer Institute, St. Andrews Place, East Melbourne, Vic. 3002, Australia; ³Department of Entomology, University of California, Riverside, CA 92521, USA; ⁴Center for Agricultural Biotechnology, University of Maryland Biotechnology Institute, University of Maryland, College Park, MD 20742-4450, USA

Received 14 September 1998 Accepted 8 April 1999

Key words: *Drosophila*, *hobo*, hot spot, integration specificity, transposable elements

Abstract

We analyzed the integration specificity of the *hobo* transposable element of *Drosophila melanogaster*. Our results indicate that *hobo* is similar to other transposable elements in that it can integrate into a large number of sites, but that some sites are preferred over others, with a few sites acting as integration hot spots. A comparison of DNA sequences from 112 *hobo* integration sites identified a consensus sequence of NTNNNNAC, but this consensus was insufficient to account for the observed integration specificity. To begin to define the parameters affecting *hobo* integration preferences, we analyzed sequences flanking a donor *hobo* element, as well as sequences flanking a *hobo* integration hot spot for their relative influence on *hobo* integration specificity. We demonstrate experimentally that sequences flanking a *hobo* donor element do not influence subsequent integration site preference, whereas, sequences contained within 31 base pairs flanking an integration hot spot have a significant effect on the frequency of integration into that site. However, sequence analysis of the DNA flanking several hot spots failed to identify any common sequence motif shared by these sites. This lack of primary sequence information suggests that higher order DNA structural characteristics of the DNA and/or chromatin may influence integration site selection by the *hobo* element.

Introduction

Transposable elements are discrete genetic entities characterized by their mobility within genomes. Although the details of transposition mechanisms vary, all transposable elements have the ability to integrate into target DNA. An important feature of transposon integration is the specificity with which integration sites are selected. The degree of integration specificity varies considerably and is a characteristic feature of each element (for a recent review see Craig, 1997). Some elements exhibit a high degree of integration specificity, while others display relatively little preference. For example, the *R2* retrotransposon of *Bombyx mori* integrates into a defined sequence within ribosomal DNA (Xiong & Eickbush, 1988). Similarly, the bacterial transposon *Tn7* integrates at high

frequency into one chromosomal site (Craig, 1996). Unlike *R2* or *Tn7*, *Tn10* and bacteriophage *Mu* exhibit very limited integration specificity. *Tn10* inserts preferentially into sites conforming to the consensus sequence 5'-NGCTNGACN-3' (Halling & Kleckner, 1982), while *Mu* integration sites conform to the consensus NYG/CRN (Mizuuchi & Mizuuchi, 1993). However, both *Mu* and *Tn10* can insert into a large number of sites, including those that conform poorly to the respective consensus integration site.

Although our current understanding of transposon integration comes from the analysis of relatively few elements, a number of factors have been identified that influence integration site selection. Bender and Kleckner (1992) found that mutations in sequences immediately flanking a preferred *Tn10* integration site significantly decreased the frequency of integration

into that site. This effect was not attributable solely to the primary sequence of the flanking DNA, suggesting that higher order structural features of the DNA may play a role in integration site selection by this element (Bender & Kleckner, 1992). Similarly, although *Tc1* and *Tc3* elements of *Caenorhabditis elegans* always insert at a TA dinucleotide, some TA sequences are preferentially selected for integration over others, indicating that, like *Tn10*, additional characteristics of the target influence *Tc1* and *Tc3* integration (van Leunen & Plasterk, 1994). Interestingly, the distribution of integration sites within a defined target differed for *Tc1* and *Tc3*, indicating that these related elements exhibit different integration preferences (van Leunen & Plasterk, 1994). DNA structure can also influence integration specificity of retroviruses (reviewed by Craigie, 1992). Although retroviral integration occurs at many different sites, the distribution of sites is distinctly non-random (Pryciak, Sil & Varmus, 1992), and DNA bending induced by packaging of DNA into nucleosomes enhances retroviral integration relative to the same site in naked DNA (Pryciak & Varmus, 1992).

The functional status of DNA can also influence its targetability. For example, the *P*-element of *D. melanogaster* inserts preferentially into 5' regulatory sequences rather than coding sequences (reviewed by Engels, 1996). In yeast, the LTR element *Ty3* inserts specifically into sequences upstream of genes transcribed by RNA Pol III (Chalker & Sandemeyer, 1992), and *Ty1* exhibits a regional preference for Pol III transcription units (Devine & Boeke, 1996; Ji et al., 1993). Conversely, the *Ty5* element exhibits a preference for non-transcribed regions of the genome (Zou et al., 1996). Finally, at least one element, the *Bari-1* element of *D. melanogaster*, is found predominantly at one site within centromeric heterochromatin (Caizzi, Caggese & Pimpinelli, 1993).

We have analyzed the integration specificity of the *hobo* element of *D. melanogaster*. *Hobo* is a class II transposable element (Finnegan, 1989) with 12 base pair inverted DNA repeats flanking a transposase gene (Streck, MacGaffey & Beckendorf, 1986; Blackman et al., 1989). *Hobo* is a member of the *hAT* family of transposable elements (Atkinson, Warren & O'Brochta, 1993) based on sequence similarity between *hobo*, the *Ac* element of *Zea mays* and *Tam3* of *Antirrhinum majus* (Calvi et al., 1991). Additional *hAT* elements have been identified in diverse organisms including non-drosophilid insects (Warren, Atkinson & O'Brochta, 1994; Devault &

Narang, 1994; Warren, Atkinson & O'Brochta, 1995; Coates et al., 1996; Handler & Gomez, 1996), nematodes (Bigot, Auge-Gouillou & Periquet, 1996), fungi (Kempken & Klueck, 1996; Colot, Haedens & Rossignol, 1998), plants (Tsay et al., 1993; Grappin et al., 1996), fish (Koga et al., 1996), and humans (Smit & Riggs, 1996). Thus, *hAT* elements comprise a phylogenetically widespread family of transposable elements. *Hobo* and other *hAT* elements are mobile when introduced into a wide range of host organisms (Hehl & Baker, 1990; Atkinson, Warren & O'Brochta, 1993; Bancroft, Jones & Dean, 1993; Hehl, 1994; O'Brochta et al., 1994; Sarkar et al., 1997; O'Brochta et al., 1996; Lozovskaya, et al., 1996). This wide host range makes *hAT* elements particularly attractive as genetic control agents for organisms of medical and/or economic significance (O'Brochta & Atkinson, 1996). Thus, a thorough understanding of integration site selection by these elements will not only increase our basic understanding of transposable element movement, but may also enhance our ability to use these elements as tools for genome manipulation and analysis. For example, modifying the integration site specificity of an element may be desirable if that element is to be used as a gene tagging or enhancer trapping agent.

There is little direct evidence regarding the mechanism of *hobo* transposition. However, by analogy to *Ac* and other structurally similar elements, it is expected that *hobo* transposition occurs via a non-replicative mechanism whereby the element is excised from one chromosomal location and inserts into a new location. Consistent with such a 'cut and paste' mechanism are *hobo* excision footprints, which resemble those generated by *Ac* and *Tam3* (Atkinson, Warren & O'Brochta, 1993). These footprints suggest that *hobo* transposition proceeds via double-stranded DNA breaks between the inverted repeats and flanking sequences. Also, integrated *hobo* elements are flanked by eight base pair direct repeats, consistent with a staggered cut at the integration site during transposition.

Several lines of evidence suggest that *hobo* exhibits some integration site specificity. First, Streck, MacGaffey and Beckendorf (1986) observed that *hobo* integration sites recovered from the *D. melanogaster* genome tend to conform to a consensus sequence of NNNNNNAC. O'Brochta et al. (1994) reported a similar consensus sequence (NTNNNNAC) after analyzing 25 *hobo* integration sites recovered from a plasmid-based transposition assay. Moreover, of

38 independent *hobo* integration events recovered by O'Brochta et al. (1994), 10 occurred at one position. This integration hot spot, with the sequence ATCCTCAC, conformed to the general consensus for *hobo* integration sites, but it was unclear why this site was preferred over other sites that conformed equally well. Smith et al. (1993) also described a non-random distribution of *hobo* integration sites. In this case, a series of genomic *hobo* integrations were generated in an enhancer trapping experiment and the integration distribution was analyzed by *in situ* hybridization to polytene chromosomes. Not only was the distribution of *hobo* integrations non-random, the distribution was significantly different from that of *P*-element integrations. Similar results were reported by Ladevèze et al. (1994). These observations suggest that the use of *hobo* as an enhancer trapping and gene tagging agent in *D. melanogaster* may complement the use of the *P*-element and increase the number of genes amenable to such analyses (Smith et al., 1993; Ladevèze et al., 1994).

The parameters responsible for the integration site specificity exhibited by *hobo* are unclear. Characteristics of the target such as primary nucleotide sequence and higher order structural features of the DNA or chromatin, the nature of the donor element and sequences surrounding it, and specific properties of the *hobo* transposase may play a role in determining integration site choice. Here, we have used a series of plasmid-based transposition assays to expand the analysis of *hobo* integration site specificity with the aim of identifying the parameters that control integration site selection by the *hobo* element. We demonstrate that while the primary sequence of the integration site may play some role in site selection, it is insufficient to account for the level of specificity observed. We also demonstrate that one target site acts as an integration hot spot in experiments using donor elements flanked by different sequences, indicating that sequences flanking the donor element do not significantly affect targeting of *hobo* to this site. Finally, we provide experimental evidence that the sequences flanking an integration hot spot play an essential role in the selection of that site for integration.

Materials and methods

Plasmids

The plasmids used in this study are shown schematically in Figure 1. The transposition donor plas-

mid pBC:hobo:Kan (Donor A), the target plasmid pUC:SacRB (Target A), and the helper plasmid pK:hshobo are as described by O'Brochta et al. (1994). The donor plasmid pBC:hobo:Kan(Sac504) (Donor B) was constructed by subcloning into the cloning vector pBCKS+ (Stratagene), a *Kan^R*-marked *hobo* element inserted into position 504 of the *SacRB* gene (O'Brochta et al., 1994). This *hobo:Kan^R* element was released, along with flanking sequences from *SacRB*, by digestion with *Nde I* and *Msc I*. The resulting fragment was treated with Klenow DNA polymerase to create blunt ends, and was subcloned into *Sma I* digested pBCKS+ using standard protocols (Sambrook, Fritsch & Maniatis, 1989). To construct the transposition target plasmid pUC19: Sac1186–1224 (Target B), a synthetic oligonucleotide of the sequence: AGCTTCAACCATACGCTGAGAG ATCCTCACTACGTAGAAGATAAGAGCT was subcloned into *Hin DIII-Sac I* digested pUC19 plasmid DNA. Nucleotides 6–44 of this oligonucleotide correspond to sequences 1186–1224 of the *SacRB* gene. The sequences AGCTT and GAGCT were included at the 5' and 3' termini, respectively, to create *Hin DIII* and *Sac I* compatible ends for subcloning. To construct the target plasmid pUC19: Sac(1203–1210) (Target C), an oligonucleotide of the sequence: AGCTATCCTCACAGCT, which corresponds to *SacRB* sequences 1203–1210, flanked by AGCT at both the 5' and 3' ends, was subcloned into *Hin DIII-Sac I* digested pUC18. In both of the oligonucleotides described above, the eight base pair sequence defined as a *hobo* integration hot spot is underlined.

Transposition assays

Transposition assays were carried out as described by O'Brochta et al. (1994), and are illustrated schematically in Figure 2. For each assay a mixture of CsCl purified plasmid DNA (0.5 mg/ml donor DNA: 0.5 mg/ml helper DNA, and 1.0 mg/ml target DNA) suspended in embryo injection buffer (Spradling & Rubin, 1982), was injected into pre-blastoderm *Drosophila* embryos using standard micro-injection techniques (Spradling & Rubin, 1982). Embryos were collected from either wild-type (Canton-S) or *yw^{67c23}* stocks, both of which are devoid of functional *hobo* elements (Calvi & Gelbart, 1994). Following injection, embryos were allowed to develop for 16–24 h. Low molecular weight DNA was then isolated from viable embryos following

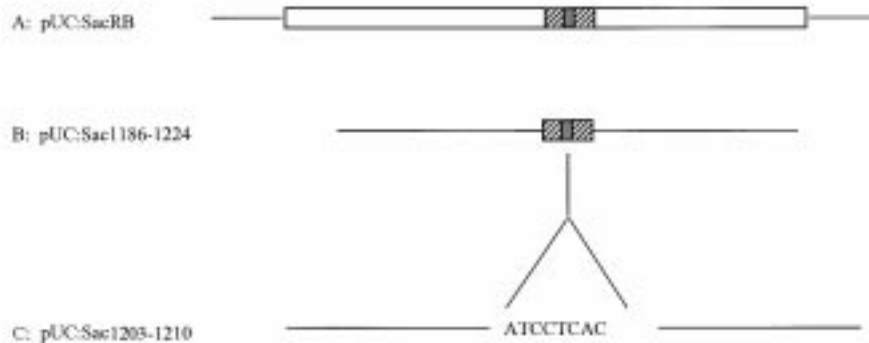
Transposition Donor Plasmids:**Transposition Targets:**

Figure 1. Donor and target plasmids used in *hobo* transposition assays. In all plasmids, thin lines represent vector sequences. *Donor plasmids:* Donor A and Donor B are identical except for sequences flanking the *hobo* element. The large arrowheads represent *hobo* sequences, including the terminal inverted repeats and internal sequences. The open vertical bars represent eight base pair duplications generated by *hobo* integration. Stippled regions represent sequences flanking the eight base pair repeats. *Cam^R* and *Kan^R* refer to chloramphenicol and kanamycin resistance genes, respectively. In Donor A, the eight base pair duplication and flanking sequences (designated Dm94E) are derived from region 94E of the *Drosophila* third chromosome. In Donor B, the eight base pair duplication and flanking sequences (designated *SacRB* 1–504 and *SacRB* 497–725) are derived from the region flanking a *hobo* integration at position 504 of the *SacRB* target gene. Nucleotides 497–504 are duplicated. *Target plasmids:* In Target A the open rectangle represents sequences 1–2000 of the *SacRB* gene subcloned into pUC19 (thin line). In targets A and B, the cross-hatched rectangles represent sequences 1186–1224 of the *SacRB* gene. The small filled boxes within the cross-hatched rectangles represents the eight base pair *hobo* integration hot spot, defined as sequences 1202–1210 of *SacRB*. In target C the sequence of nucleotides 1202–1210 of *SacRB* are shown.

the procedure of Hirt (1967). Recovered plasmid DNA was introduced into electrocompetent *E. coli* strain DH10B (Gibco-BRL) by electroporation as described by the supplier, and plated on LB agar containing the following: For transposition assays using Target A, plates contained 100 ug/ml ampicillin, 50 ug/ml kanamycin, and 10% sucrose. In these experiments only *hobo:Kan^R* integrations that disrupt the *SacRB* gene allow colony growth. For experiments using targets B and C, plates contained 100 ug/ml ampicillin, 50 ug/ml kanamycin, and 40 ug/ml X-gal. Integrations within the *lacZ* gene of these targets results in amp^R, kan^R, white colonies. Integrations outside the *lacZ* gene of these targets results in amp^R, kan^R, blue colonies. As co-transformation of the helper plasmid along with the target plasmid could also result in amp^R, kan^R, blue

colonies, plasmid mixtures isolated from injected embryos were digested with the restriction enzyme *Nco* I prior to electroporation of *E. coli*. This treatment linearizes both the donor and helper plasmids, but not the target plasmid or target plasmids containing a *hobo* integration, thereby minimizing false positives due to co-transformation of the helper and target plasmids. Plasmids with *hobo* integrations within the defined targets were initially identified by the expected combination of genetic markers. These plasmids were analyzed by restriction enzyme analysis and DNA sequencing using *hobo*-specific primers directed towards either the left or right junction between *hobo* and flanking sequences. Plasmid preparations were done using the alkaline lysis method as described (Sambrook, Fritsch & Maniatis, 1989) or using commercial plasmid puri-

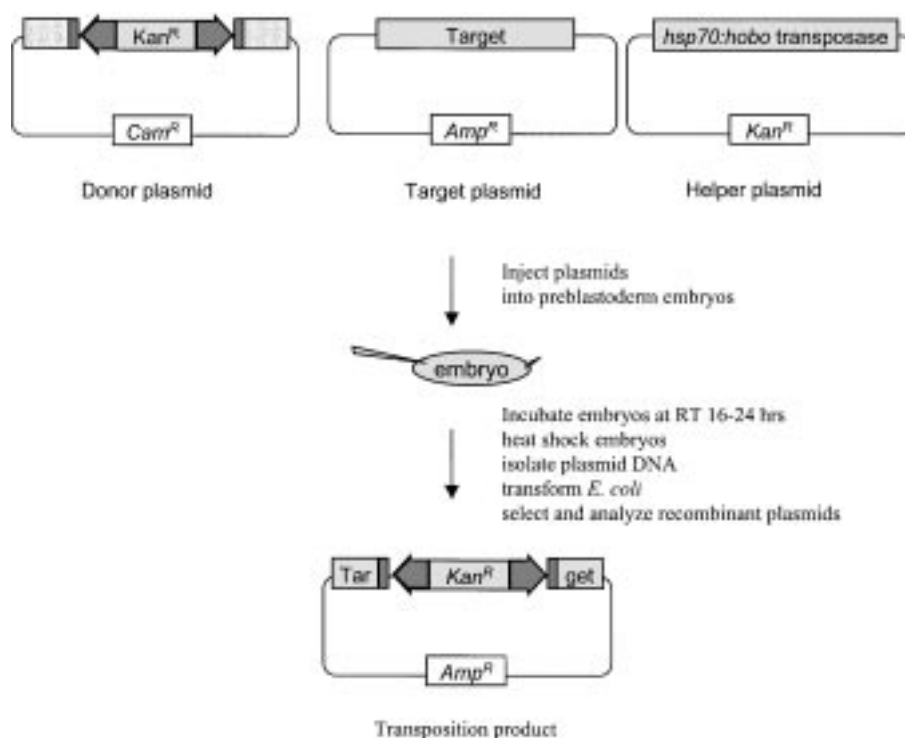


Figure 2. A generalized scheme depicting the transposition assays used in this study. The left and right ends of the *hobo* element are depicted as large arrowheads. Donor plasmids consist of a kanamycin resistance (*Kan^R*)-marked *hobo* element, along with flanking sequences (solid and stippled rectangles) subcloned into a plasmid carrying the gene for chloramphenicol resistance (*Cam^R*). Target plasmids consist of either the complete *SacRB* gene or portions thereof (see Figure 1) cloned into a plasmid carrying the gene for ampicillin resistance (*Amp^R*). The helper plasmid contains a *hobo* transposase gene, transcriptionally controlled by the *Drosophila* heat shock protein 70 (*hsp70*) promoter, cloned into a plasmid carrying the gene for kanamycin resistance (*Kan^R*). Recombinant plasmids resulting from transposition of the *Kan^R*-marked *hobo* element into the target are identified as *Amp^R/Kan^R* plasmids in *E. coli*. See Materials and methods for further details regarding transposition assays.

fication methods (Promega, Inc). DNA sequencing was carried out using standard dideoxy chain termination method (Sanger, Nicklen & Coulson, 1977) using the Sequenase DNA polymerase as described by the manufacturer (United States Biochemical), or by using fluorescent-dye labeled dideoxy terminator sequencing according to the manufacturer's description (Applied Biosystems, Inc). Integration sites within each target were determined by comparing the obtained DNA sequence information to the known sequence of the target using DNA Strider version 1.0.1 (C. Marck). The integration position is defined as the nucleotide of the target DNA sequence immediately preceding *hobo* sequences in transpositional recombinant plasmids, regardless of the orientation of the inserted element. Nucleotide positions for the *SacRB* target follow the numbering system of Gay et al. (1985), and the positions for *lacZ* are according to Norrander, Kempe and Messing (1983).

Results

Previous *hobo* transposition assays identified a weak consensus sequence shared by *hobo* integration sites and identified a *hobo* integration hot spot within the *Bacillus subtilis SacRB* gene (O'Brochta et al., 1994). Here we used similar assays to expand our analysis of the mechanism of *hobo* targeting. First, we generated a larger set of *hobo* integrations within the *SacRB* target in order to further analyze the influence of the target sequence on integration specificity. Second, we investigated the role of the donor element and its flanking sequences in directing integration to an integration hot spot. Third, we investigated the role of the DNA flanking an integration hot spot in conferring a preference for that site.

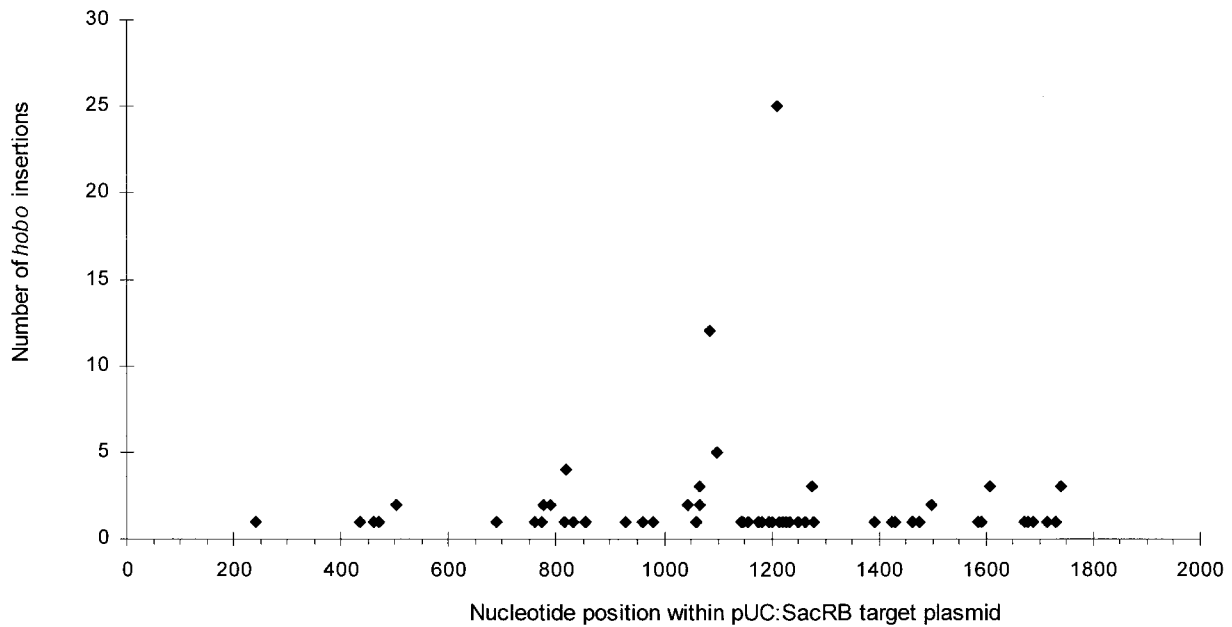


Figure 3. The distribution of *hobo* integrations within the *SacRB* target gene. The x-axis represents nucleotides 1–2000 of the pUC:SacRB gene as described by Gay et al. (1985). The y-axis represents the number of independent *hobo* integrations at various positions within the target gene.

Hobo exhibits a distinct integration specificity

Experiments using the pUC:SacRB target plasmid (Target A in Figure 1) resulted in 74 new independent integrations within the *SacRB* gene. The distribution of these integrations combined with those reported by O'Brochta et al. (1994) is shown in Figure 3. In total, 112 *hobo* integrations were recovered at 56 different sites within the 2 Kb *SacRB* target. Several characteristics of the integration distribution within this target are noteworthy. First, the integration of *hobo* is random relative to orientation of the target sequence. Of the 112 integrations, 52 were oriented with *hobo* left to right relative to *SacRB* and 60 were in the opposite orientation. This does not deviate significantly from an expected ratio of 1:1 if integration orientation were random ($P > 0.05$). This lack of an orientation preference suggests that target site recognition and transposon integration may occur as distinct steps during transposition. Second, considering the 2000 base pairs comprising the *SacRB* gene (Gay et al., 1985) as the effective target size, along with a total of 112 integrations, the distribution of integration sites differs significantly from the Poisson distribution (calculated as described by Wardlaw, 1985) expected if integration site selection were random ($P < 0.001$). Third, several integration sites, including the hot spot pre-

viously identified at position 1210 (O'Brochta et al., 1994), were clearly preferred sites for *hobo* integration. Based on the Poisson distribution, the number of sites expected to be selected two times is 2.85, and the number of sites expected to be selected three times is 0.06. Therefore, any site selected three times or more is clearly a preferred integration site. The observed number of sites selected twice was six, suggesting that some of these sites may also be preferred integration sites. However, we have considered only those sites selected three times or more as preferred sites of *hobo* integration. A fourth characteristic of the distribution of *hobo* integrations within *SacRB* is an apparent clustering of integration sites. The most prominent cluster consists of 17 integration sites within a region of 167 nucleotides surrounding the integration hot spot at position 1210. To determine if the apparent clustering of integration sites varies significantly from a distribution expected if integration were random, we divided the target into 200 base pair intervals centered on position 1210, and compared the observed proportion of integration sites within each interval to the expected proportion (0.1). Using the z-statistic as described by Devore and Peck (1993), only the cluster surrounding position 1210 varied significantly from the expected distribution ($P < 0.001$), suggesting that, in addition to containing a hot spot for integration, this interval

may represent a 'hot region' for integration of the *hobo* element.

The role of the target sequence in integration specificity

In an effort to determine the characteristics that determine whether a sequence is prone to *hobo* integration we analyzed the eight base pair duplications associated with the 56 integration sites represented in Figure 3 for any obvious sequence similarity. We first compared all 56 sites, then focused on those sites that were preferred sites for *hobo* integration. Comparison of all 56 integration sites yielded a consensus sequence of NTNNNNAC, with 27 of the 56 sites conforming to this sequence. This consensus is identical to that observed at *hobo* integration sites in *Drosophila* chromosomes (Streck, MacGaffey & Beckendorf, 1986), and in previous plasmid assays (O'Brochta et al., 1994). An inspection of the preferred sites yielded the same consensus, with six of the seven preferred sites conforming (Figure 4, panel A). However, no additional sequence information was apparent from the analysis of these preferred sites. Moreover, there are 47 additional sites within the *SacRB* gene which conform to the consensus sequence, yet were never recovered as integration sites in these experiments. Thus, sites that conformed equally well to the consensus were not equally preferred as *hobo* integration sites. Taken together, these observations indicate that the nucleotide sequence of the integration site may play a role in site selection, but that other factors must also be involved.

The role of sequences flanking the donor element

In addition to integration site analysis, the above experiments allowed us to analyze the role of the donor element and its flanking sequences in integration site selection. A potential role of sequences flanking the donor element was suggested by the observation that the eight base pair sequence flanking the original donor element used by O'Brochta et al. (1994) (Donor A in Figure 2) was identical in seven of eight positions to the integration hot spot at position 1210 of *SacRB*. This near identity may be due to chance, it may be attributable to the preference of *hobo* for sequences conforming to the consensus, or it may reflect a role of donor/target interaction in determining where the element will integrate. A precedent for donor/target interaction comes from studies of elements such as *Tn3* (Sherrat, 1989) and bacteriophage *Mu* (Mizuuchi, 1992) in which there is a requisite interaction between

the donor element and the target site before transposition can occur. There is also some evidence for donor/target interactions during *Tam3* transposition (Coen et al., 1989). Donor/target interaction would provide an opportunity for homologous sequence interaction between the donor and the target DNA. To test the role of sequences flanking the donor element we used two different donor plasmids in the experiments described above. The first donor (Donor A in Figure 1) was the same donor used by O'Brochta et al. (1994). The sequences flanking *hobo* in this donor originated from region 94E of the *Drosophila* genome. In the second donor (Donor B) the sequences flanking *hobo* were derived from a region of the *SacRB* gene that was selected for integration only once in previous experiments, and therefore is not a preferred site of *hobo* integration. If similarity between donor and target sequences were responsible for integration site choice we would have expected to see less use of site 1210, the hot spot described above, and more use of site 504, which corresponds to the sequences flanking *hobo* in the new donor. Of the total integrations represented in Figure 3, 55 were recovered using Donor B. Of these, 15 (27.2%) had integrated at position 1210, while only one inserted at 504. These results indicate that the sequences flanking this donor element did not alter the propensity for *hobo* to integrate into position 1210 of this target. Thus the preference observed in these experiments for *hobo* integration into this hot spot appears to be a property of the target DNA rather than the donor element and its flanking sequences.

The role of sequences flanking an integration hot spot

In order to determine where in the target the information resides, that determines the attractiveness of a site for *hobo* integration, we constructed two new target plasmids. Each of these targets consists of the integration hot spot, with or without *SacRB* flanking sequences, subcloned into the polylinker of a pUC plasmid (Figure 1, Targets B and C). In both plasmids, the target fragment was subcloned such that it did not disrupt the *lacZ* open reading frame and thus resulted in blue colonies in the presence of X-gal. Integration of *hobo* into the target sequence would result in white colonies due to the disruption of *lacZ*. The first such target plasmid, Target B, contained the eight base pair *hobo* hot spot along with a total of 31 nucleotides (17 from the left side of the hot spot and 14 from the right side) of flanking sequence from *SacRB*. The second target, Target C, contained only the *hobo* hot spot,

	<u>nt position</u>	<u>Sequence</u>	<u>n</u>						
A. <u>Target = pUCSacRB</u>									
	819	GATGACAC	4						
	1084	CTTTGAAC	12						
	1099	CTCTACAC	5						
	1210	ATCCTCAC	25						
	1276	ATGGCTAC	3						
	1608	TTAGAAAC	3						
	1740	ATCACGAC	3						
B. <u>Target = pUC: Sac 1186-1224</u>									
	134	TTAGGCAC	3						
	263	ATCCTCAC	25						
	324	GTCACGAC	4						
	663	GTGAAAAC	4						
C. <u>Target = pUC: Sac 1203-1210</u>									
	17	ACGCAAAC	4						
	134	TTAGGCAC	3						
	148	CTTTACAC	3						
	204	TTTCACAC	4						
	585	TTACAGAC	7						
	630	GTGAAAAC	3						
	670	CTCGTGAT	5						
D. <u>Consensus sequences</u>									
	<u>Position:</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>
	G:	4	0	3	7	3	4	0	0
	A:	5	1	4	2	8	4	18	0
	T:	5	16	5	4	5	1	0	1
	C:	4	1	6	5	2	9	0	17
	Consensus 1:	N	N	N	N	N	N	A	C (17/18)
	Consensus 2:	N	T	N	N	N	N	A	N (16/18)
	Consensus 3:	N	T	N	N	N	N	A	C (15/18)

Figure 4. Preferred *hobo* integration sites identified in three different targets. Eighteen preferred *hobo* integration sites recovered in transposition assays using three different targets are compared. *n* refers to the number of integrations recovered at each site. Sequences in A, B, and C correspond to the eight base pairs duplicated upon *hobo* integration. As the orientation of *hobo* integration is random relative to the target, both orientations of the eight base pair duplication were considered in each case, and the orientation which best conformed to the consensus NNNNNNAC was used. For those cases in which both orientations fit this consensus (integrations 317, 656, and 623), the orientation was chosen at random. Position 656 in target pUC: Sac 1186–1224 and position 623 in target pUC: Sac 1203–1210 correspond to the same site in the pUC plasmid. Position 1210 of pUC: SacRB is the previously identified *hobo* integration hot spot. Position 263 in target pUC: Sac 1186–1224 is this hot spot subcloned into pUC19. Panel D compares the nucleotides found at each position in all of the preferred integration sites, and shows three consensus sequences that can be derived from these data.

with no flanking sequences from *SacRB*. The results of transposition assays using Target B are summarized in Figure 5. Of 46 *hobo* integrations into this target, 25 (54.3%) were inserted precisely into the position defined previously as the *hobo* integration hot spot. When Target C was used, 79 independent *hobo* integrations were generated, and none of these were inserted into the hot spot (Figure 6). These results clearly indicate that information contained in the 31 nucleotides flanking the *hobo* hot spot, and included

in Target B, but not Target C, strongly influence the selection of this site for integration.

The transposition assays using Targets B and C resulted in a total of 125 new *hobo* integrations into 66 different sites. The distribution of these 66 integration sites revealed an integration pattern similar to that observed amongst the collection of integrations into *SacRB*. That is, a large number of sites were selected, but some sites were preferred over others, with the previously identified hot spot again being highly

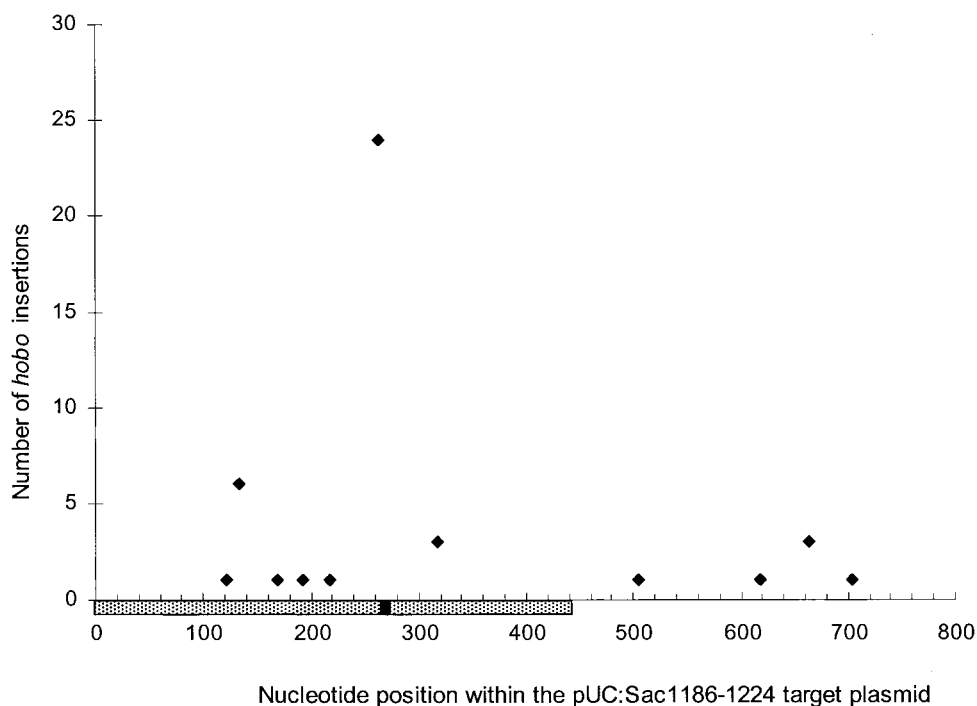


Figure 5. Distribution of *hobo* integrations within the pUC:SacRB 1186–1224 target plasmid. The numbering of the x-axis corresponds with the published pUC sequence (Norrander, Kempe and Messing, 1983). The stippled bar indicates the position of the lac operon sequences within this target. The vertical black bar represents sequences corresponding to nucleotides 1186–1224 of *SacRB*, which contains the *hobo* integration hot spot. Here the *SacRB* sequences are numbered 239–277, and nucleotide 263 corresponds to the integration hot spot at position 1210 of pUC:SacRB. Only integrations within the first 800 nucleotides of the target are shown. Additional integrations were obtained outside of this region and were used for sequence analyses of integration sites (see text).

preferred. Visual inspection of the eight base pairs duplicated at each new integration site revealed that, of the 66 new sites, 30 conformed to the sequence NNNNNNAC and nine to the sequence NTNNNNAC. Also, 11 new preferred sites were identified and of these, 10 conformed to the sequence NNNNNNAC, and eight to the sequence NTNNNNAC (Figure 4, Panels B and C). Again, considering all preferred sites identified, no additional consensus information could be derived by simple inspection of the sequences (Figure 4, Panel D).

The results of our experiments indicate that information contained within 39 base pairs, which include the eight base pair integration site along with 31 base pairs of flanking sequence, is essential for a high frequency of integration into a *hobo* integration hot spot. This is the smallest segment of DNA tested that could be placed in a new sequence context without altering the site's targetability by *hobo*. In an effort to identify primary nucleotide sequence information common to the flanking sequences of *hobo* integration sites, we analyzed these sequences according to

the method of Schneider and Stephens (1990) (Figure 7A). Comparing all of the *hobo* integration sites and flanking DNA did not reveal any evidence of nucleotide sequence information in the sequences flanking the eight base pair integration site. Similarly, when only the preferred integration sites reported in Figure 4 were compared, there was no evidence of nucleotide information in the flanking sequences. As expected from visual inspection of these integration sites, this analysis revealed that nucleotide positions two, seven, and eight of the integration site were important determinants of a site's targetability. For comparison, we conducted a similar analysis of the related *hAT* element *Hermes* based on published integration site data (Figure 7B). When all *Hermes* integration sites were compared, we found the nucleotide information content of the integration site to be similar in quantity and distribution to that found for *hobo*. Specifically, as was observed for *hobo*, positions two and seven of the integration site contributed most significantly to a site's targetability. However, even when only preferred *Hermes* integration sites were compared, po-

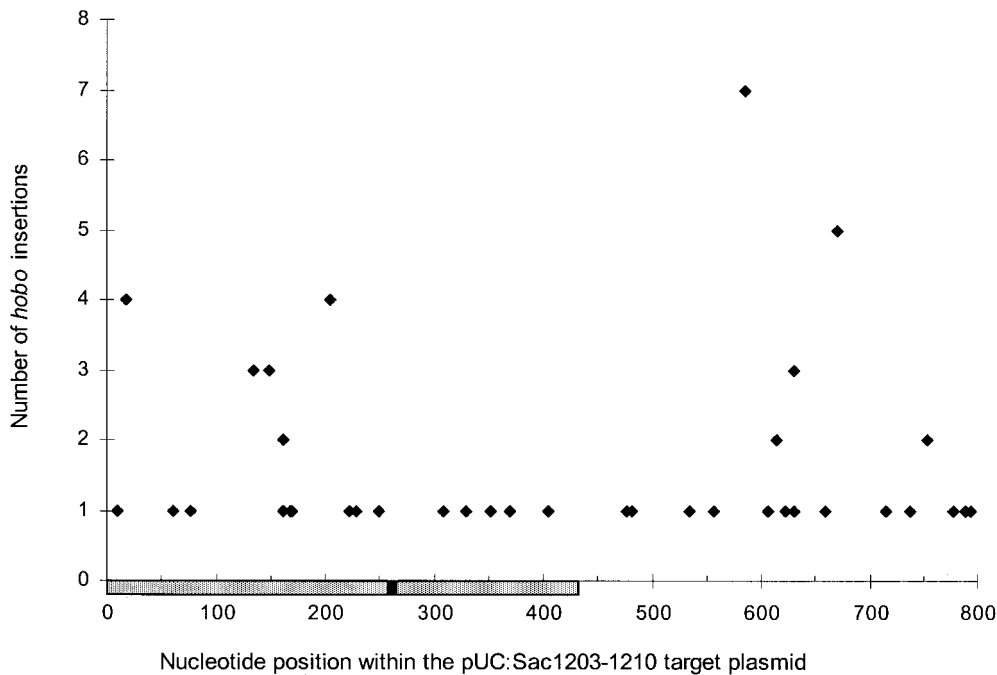


Figure 6. Distribution of *hobo* integration sites within the pUC:SacRB 1203–1210 target plasmid. The x-axis represents nucleotides 1–800 of target plasmid pUC:SacRB 1203–1210. The stippled bar indicates the position of lac operon sequences. The vertical black bar indicates the position of sequences 1203–1210 of *SacRB*, here re-numbered as nucleotides 244–251. Additional integrations were obtained outside of this region and were used for sequence analyses of integration sites (see text).

sition eight did not appear to contribute significantly to a site's targetability as it does in preferred *hobo* integration sites. When the sequences flanking all *Hermes* integration sites were compared, little primary nucleotide sequence information was detected. However, when flanking sequences from preferred *Hermes* sites were compared there was some evidence for significant information content within the flanking sequences. Finally, we performed a similar analysis for *Ac/Ds* integration sites (Figure 7C). As flanking sequence information was not available for all of the published *Ac/Ds* integrations, we analyzed only the eight base pairs of the integration site. Unlike *hobo* and *Hermes*, *Ac/Ds* integration sites appear to contain no primary nucleotide sequence information. That is, the sequence information contained within sites chosen by *Ac/Ds* did not differ significantly from that expected from random integration.

Discussion

Our experiments with the *hobo* element are similar in design to classical studies on targeting of the bacterial transposon *Tn10* (Halling & Kleckner, 1982; Bender & Kleckner, 1992), and have yielded similar results. *Tn10* inserts preferentially into sites conforming to the consensus sequence NGCTNGACN, probably reflecting sequence specific contact between the transposase and the bases at the integration site (reviewed by Craig, 1997). Similarly, the majority of *hobo* integration sites, and almost all preferred sites, conform to a consensus (NTNNNNAC), although this consensus contains less sequence information than the *Tn10* consensus. Also, DNA flanking the integration site plays a significant role in targeting of both *Tn10* and *hobo*. However, in both cases, analysis of flanking sequences does not reveal any clear primary sequence informa-

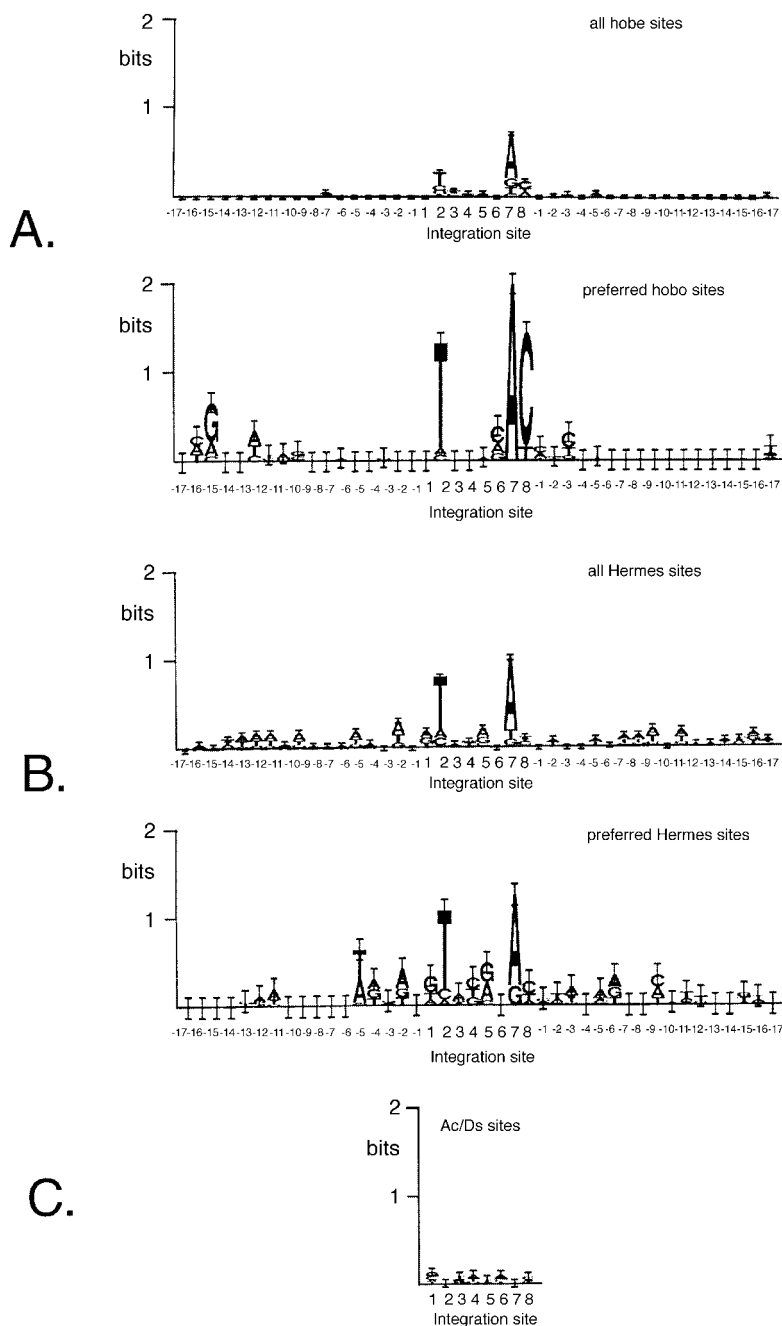


Figure 7. Sequence logo analysis of *hAT* element integration sites. The computer program Weblogo (<http://www.bio.cam.ac.uk/cgi-bin/seqlogo/logo.cgi/>) was used to analyze sequence information content within the integrations sites and corresponding flanking sequences for the *hAT* elements *hobo*, *Hermes*, and *Ac/Ds*. All sites included in the analysis were given equal weight regardless of the number of times they were selected for integration. The y-axis represents bits of information for each nucleotide position. The x-axis indicates the nucleotide position relative to the integration site. The height and order of each nucleotide reflects their relative contribution to the information content of that site. The error bars reflect the variance based on the number of sequences analyzed. (A) The analysis of *hobo* integration sites generated in our experiments using three plasmid targets. The upper panel includes all sites and the lower panel includes only preferred integration sites as reported in Figure 4. (B) The analysis of all *Hermes* integration sites and flanking DNA based on published data (Pinkerton, O'Brochta, & Atkinson, 1997; Sarkar et al., 1997), with all sites included in the upper panel and preferred sites in the lower panel. (C) The analysis of *Ac/Ds* sites based on published data (Dooner et al., 1988; Coen et al., 1989; Federoff, 1989; Hehl & Baker, 1990; Peterson, 1990; Osborne et al., 1991; Chen, Greenblatt & Dellaporta, 1992; Rommens et al., 1992; Bancroft, Jones & Dean, 1993; English, Harrison & Jones, 1993; Long et al., 1993; Marion-Poll et al., 1993; Weil & Wessler, 1993; Schmidt-Rogge et al., 1994; Rudenko, Nijkamp & Hille, 1992; Sugimoto et al., 1994).

tion to account for the influence of these sequences on integration site selection. Thus, as seems to be the case with *Tn10*, it is probably not the primary sequence, but rather higher order structural characteristics of the flanking DNA that affects *hobo* integration preferences. Integration site selection by other elements appears to be dependent on similar determinants. For example, the *Tc1* and *Tc3* elements of *C. elegans* always insert at TA dinucleotides with some TA sites preferred over others (van Luenen & Pasterk, 1994), and sequences within four base pairs of a preferred *Tc1* integration site are sufficient to account for this *Tc1* integration preference (Ketting, Fischer & Plasterk, 1997). Thus, the emerging picture for transposable element targeting is that both the primary nucleotide sequence of integration sites and DNA structure at or near the integration site are somehow 'sensed' by each transposable element. However, the relative influence of the integration site and the flanking DNA appears to be different for each element, and the mechanisms by which transposable elements distinguish these features of the target remain to be elucidated.

While recognition of the primary nucleotide sequence of an integration site is likely to involve base specific contacts between the transposase and the target DNA, the mechanisms by which structural characteristics of flanking DNA are recognized is less clear. Thus, an important step in understanding the mechanisms of integration site selection by transposable elements is to determine the structural characteristics of the DNA flanking integration sites that distinguish these sites from other regions of the DNA. There is some evidence that either intrinsic DNA bending, or DNA flexibility to facilitate protein induced bending, plays an important role in directing DNA integration. Hallet et al. (1994) suggest that sequences flanking preferred integration sites of the bacterial integration element *IS231A* consist of alternating AT and GC rich stretches that adopt an intrinsic S-curve conformation, and that this curved DNA structure may play a role in integration site selection by this class II element. Retroviral integration occurs preferentially into DNA which is bent as a consequence of its position within nucleosomes (Pryciak & Varmus, 1992; Muller & Varmus, 1994; Pruss, Bushman & Wolffe, 1994), illustrating a role for protein induced DNA bending in integration specificity. In another example, Lampe, Grant and Robertson (1998) found that a preferred region of integration for *Himar1* (a *mariner*-like element isolated from the horn fly) corresponded to a region of pUC18 previously shown to exhibit a high

degree of DNA bending based on increased sensitivity to digestion by DNase I (Brukner et al., 1995). This observation suggests that DNA bending may play a role in integration site selection by *mariner*-like elements.

To determine if DNA bending might play a role in *hobo* targeting we analyzed the sequences flanking *hobo* integration sites in an effort to identify nucleotide sequence motifs likely to adopt an intrinsic bent conformation or to confer flexibility to the DNA. First, by visual inspection of sequences flanking all *hobo* integration sites generated, we looked for periodic alternation of AT rich and GC rich sequences like those suggested to influence targeting of *IS231A* (Hallet et al., 1994). Next, we determined the frequency of all possible trinucleotides within each of the flanking sequences and looked for a prevalence of trinucleotides which confer increased DNase I sensitivity (Brukner et al., 1995). We were unable, by either of these criteria, to identify any consistent sequence motifs within the flanking sequences. We then compared our distribution of *hobo* integrations within a pUC18 based target (Target C in Figure 1) with the distribution of DNase I sensitive sites within pUC18 as reported by Brukner et al. (1995). As was observed for the *Himar1* element discussed above, a correlation between DNase I sensitivity and *hobo* integration would suggest that *hobo* integrates preferentially into regions of bent (or flexible) DNA. The 50 nucleotide region described in the most detail by Brukner et al. (1995) corresponds to nucleotides 1711–1761 of our Target C. As shown in Figure 8, there were three distinct peaks of DNase I sensitivity identified in this region. Intriguingly, we recovered three *hobo* integrations at three separate sites within this region. Two of the DNase sensitivity peaks were localized within an eight base pair *hobo* integration site, while the third peak was removed only five base pairs from the nearest integration site. Significantly, there were no other *hobo* integrations within 448 nucleotides on one side and 73 nucleotides on the other side of this small cluster of integration sites. These observations lead to the hypothesis that DNA bending plays a role in *hobo* integration specificity. We also observed that *hobo* integration sites were clustered within 167 nucleotides surrounding an integration hot spot, suggesting a regional influence on *hobo* integration specificity. If *hobo* integration were influenced by DNA bending due to the packaging of the target DNA into nucleosomes, similar cluster sizes might be expected. Further experiments need to be done to test these hypotheses.



Figure 8. The juxtaposition of DNase I sensitive sites and *hobo* integration sites suggests a role for DNA bending in *hobo* integration. Shown are pUC18 sequences corresponding to nucleotides 1703–1761 of Target C (see Figure 1). The underlined sequences identify *hobo* integration sites. The open arrows denote positions of high DNase I sensitivity as described by Brukner et al. (1995). The TA dinucleotide shown in bold corresponds to a preferred integration site of the *mariner*-like element *Himar1* (Lampe, Grant & Robertson, 1998).

In addition to understanding the structural features of the target itself, an understanding of the mechanism by which the transposase recognizes these features is critical to a complete understanding of transposon targeting. To address this question, a comparative analysis of integration site specificity of related elements may yield some insight not readily apparent by other types of analyses. The *hAT* family of transposable elements provides an excellent opportunity for such an investigation. For example, the *Hermes* element from the housefly *Musca domestica* (Warren, Atkinson & O'Brochta, 1994) is a *hAT* element, sharing 55% amino acid sequence identity with *hobo*. The integration pattern exhibited by *Hermes* is similar to that observed for *hobo* in that many sites are selected for integration, with some sites preferred over others (Sarkar et al., 1997). Moreover, the consensus sequence identified for *Hermes* integration sites was NTNNNNAC, identical to that described for *hobo*. However, an additional consensus of GTHYSVAC for highly preferred *Hermes* integration sites was also identified. The *hobo* integration hot spot ATCCTCAC does not fit this consensus. In contrast to *hobo* and *Hermes*, the *Ac* element exhibits very little, if any, integration site preference. Further research into the transposition mechanisms of these related, but distinctly different, elements may yield important insights into the structural aspects of the transposable elements, their respective transposase proteins, and/or characteristics of the target DNA which are responsible for similarities and differences in integration site selection.

Acknowledgements

This work was supported in part by the National Institutes of Health (GM8102) and the United States Department of Agriculture (NRI 9501880). Kenneth J. Saville would like to thank two anonymous reviewers for valuable comments, Cathy Saville for careful reading of the manuscript.

References

- Atkinson, P.W., W.D. Warren & D.A. O'Brochta, 1993. The *hobo* transposable element of *Drosophila* can be cross-mobilized in houseflies and excises like the *Ac* element of maize. Proc. Natl. Acad. Sci. USA 90: 9693–9697.
- Bancroft, I., J.D.G. Jones & C. Dean, 1993. Heterologous transposon tagging of the DRL1 locus in *Arabidopsis*. Plant Cell 5: 631–638.
- Bender, J. & N. Kleckner, 1992. *Tn10* integration specificity is strongly dependent upon sequences immediately adjacent to the target-site consensus sequence. Proc. Natl. Acad. Sci. USA 89: 7996–8000.
- Bigot, Y., C. Augé-Gouillou & G. Periquet, 1996. Computer analyses reveal a *hobo*-like element in the nematode *Caenorhabditis elegans*, which presents a conserved transposase domain common with the *Tc1-mariner* transposon family. Gene 174: 265–271.
- Blackman, R.K., M. Macy, D. Koehler, R. Grimaila & W.M. Gelbart, 1989. Identification of a fully-functional *hobo* transposable element and its use for germ-line transformation of *Drosophila*. EMBO J. 8: 211–217.
- Brukner, I., R. Sánchez, D. Suck & S. Pongor, 1995. Sequence-dependent bending propensity of DNA as revealed by DNase I: parameters for trinucleotides. EMBO J. 14: 1812–1818.
- Caizzi, R., C. Caggese & S. Pimpinelli, 1993. A new transposon-like family in *Drosophila melanogaster* with a unique heterochromatic organization. Genetics 133: 335–345.
- Calvi, B.R., T.J. Hong, D.D. Findley & W.M. Gelbart, 1991. Evidence for a common evolutionary origin of inverted repeat transposons in *Drosophila* and plants: *hobo*, *Activator* and *Tam3*. Cell 66: 465–471.
- Calvi, B.R. & W.M. Gelbart, 1994. The basis for germline specificity of the *hobo* transposable element in *Drosophila melanogaster*. EMBO J. 13: 1636–1644.
- Chalker, D.L. & S.B. Sandemeyer, 1992. *Ty3* integrates within the region of RNA polymerase II transcription initiation. Genes Dev. 6: 117–128.
- Chen, J., I.M. Greenblatt & S.L. Dellaporta, 1992. Molecular analysis of *Ac* transposition and DNA replication. Genetics 130: 665–676.
- Coates, C.J., K.N. Johnson, H.D. Perkins, A.J. Howells, D.A. O'Brochta & P.W. Atkinson, 1996. The hermit transposable element of the Australian sheep blowfly, *Lucilia cuprina*, belongs to the *hAT* family of transposable elements. Genetica 97: 23–31.
- Coen, E.S., T.P. Robbins, J. Almeida, A. Hudson & R. Carpenter, 1989. Consequences and mechanisms of transposition in *Antirrhinum majus*, pp. 523–529 in Mobile DNA, edited by D.E. Berg and M.M. Howe. American Society of Microbiology, Washington, DC.
- Colot, V., V. Haedens & J.-L. Rossignol, 1998. Extensive, non-random diversity of excision footprints generated by *Ds*-like

- transposon Ascot-1 suggests new parallels with V(D)J recombination. *Mol. Cell Biol.* 18: 4337–4346.
- Craig, N.L. 1996. Transposon *Tn7*. *Curr. Top. Microbiol. Immunol.* 204: 27–48.
- Craig, N.L., 1997. Target site selection in transposition. *Annu. Rev. Biochem.* 66: 437–474.
- Craigie, R., 1992. Hotspots and warm spots: integration specificity of retroelements. *Trends Genet.* 8: 187–190.
- DeVault, J.D. & S.K. Narang, 1994. Transposable elements in lepidoptera: *hobo*-like transposons in *Heliothis virescens* and *Helicoverpa zea*. *Biochem. Biophys. Res. Commun.* 203: 169–175.
- Devine, S.E. & J.D. Boeke, 1996. Integration of the yeast retrotransposon *Ty1* is targeted to regions upstream of genes transcribed by RNA polymerase III. *Genes Dev.* 10: 620–633.
- Devore, J. & R. Peck, 1993. pp. 464–465 in *Statistics: The exploration and analysis of data*. Wadsworth Publishing Co., Belmont CA.
- Dooner, H.K., J. English & E.J. Ralston, 1988. The frequency of transposition of the maize element *Activator* is not affected by an adjacent deletion. *Mol. Gen. Genet.* 211: 485–491.
- English, J., K. Harrison & J.D.G. Jones, 1993. A genetic analysis of DNA sequence requirements for Dissociation state 1 activity in tobacco. *Plant Cell* 5: 501–514.
- Engels, W.R., 1996. *P* elements in *Drosophila*, pp. 103–123 in *Transposable Elements*, edited by H. Saedler and A. Gierl. Springer-Verlag, Berlin.
- Finnegan, D.J., 1989. Eukaryotic transposable elements and genome evolution. *Trends Genet.* 5: 103.
- Fedoroff, N.V., 1989. Maize transposable elements, pp. 375–411 in *Mobile DNA*, edited by D.E. Berg & M.M. Howe. American Society of Microbiology, Washington, DC.
- Gay, P., D.L. Coq, M. Steinmetz, T. Berkelma & C.I. Kado, 1985. Positive selection procedure for entrapment of integration sequence elements in Gram-negative bacteria. *J. Bacteriol.* 164: 918–921.
- Grappin, P., C. Audeon, M.-C. Chupeau & M.-A. Grandastien, 1996. Molecular and functional characterization of Slide, an *Ac*-like autonomous transposable element from tobacco. *Mol. Gen. Genet.* 252: 386–397.
- Hallet, B., R. Rezsöházy, J. Mahillon & J. Delcour, 1994. IS231A integration specificity: consensus sequence and DNA bending at the target site. *Mol. Microbiol.* 14: 131–139.
- Halling, S. & N. Kleckner, 1982. A symmetrical six-base-pair target site sequence determines *Tn10* integration specificity. *Cell* 28: 155–163.
- Handler A.M. & S.P. Gomez, 1996. The *hobo* transposable element excises and has related elements in tephritid species. *Genetics* 143: 1339–1347.
- Hehl, R. & B. Baker, 1990. Properties of maize transposable element *Activator* in transgenic tobacco plants: a versatile interspecies genetic tool. *Plant Cell* 2: 709–721.
- Hehl, R., 1994. Transposon tagging in heterologous host plants. *Trends Genet.* 10: 385–386.
- Hirt, B., 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. *J. Mol. Biol.* 26: 365–369.
- Ji, H., D.P. Moore, M.A. Blomberg, L.T. Braiterman, D.F. Voytas, G. Natsoulis & J.D. Boeke, 1993. Hotspots for unselected *Ty1* transposition events on yeast chromosome III are near tRNA genes and LTR sequences. *Cell* 73: 1007–1018.
- Kempken, R. & U. Klueck, 1996. Restless, an active *Ac*-like transposon from the fungus *Tolyocladium inflatum*: structure, expression, and alternative RNA splicing. *Mol. Cell. Biol.* 16: 6563–6572.
- Ketting, R.F., S.E. J. Fischer & R.H. Plasterk, 1997. Target choice determinants of the *Tc1* transposon of *Caenorhabditis elegans*. *Nucleic Acids Res.* 25: 4041–4047.
- Koga, A., M. Suzuki, H. Inagaki, Y. Bessho & H. Hori, 1996. Transposable element in fish. *Nature* 383: 30.
- Lampe, D.J., T.E. Grant & H.M. Robertson, 1998. Factors affecting transposition of the *Himar1 mariner* transposon *in vitro*. *Genetics* 149: 179–187.
- Ladevèze, V., M.I. Galindo, L. Pasqual, G. Periquet & F. Lemeunier, 1994. Invasion of the *hobo* transposable element studied by *in situ* hybridization on polytene chromosomes of *Drosophila melanogaster*. *Genetica* 93: 91–100.
- Long, D., M.E. Sundberg, J. Swinburne, P. Paungomlee & G. Coupland, 1993. The maize transposable element system *Ac/Ds* as a mutagen in *Arabidopsis*: identification of an albino mutation induced by *Ds* integration. *Proc. Natl. Acad. Sci. USA* 90: 10370–10374.
- Lozovskaya, E.R., D.J. Nurminsky, D.L. Hartl & D.T. Sullivan, 1996. Germline transformation of *Drosophila virilis* mediated by the transposable element *hobo*. *Genetics* 142: 173–177.
- Marion-Poll, A., E. Marin, N. Bonnefoy & V. Pautot, 1993. Transposition of maize autonomous element *Activator* in transgenic *Nicotiana plumbaginifolia* plants. *Mol. Gen. Genet.* 238: 209–217.
- Mizuuchi, K., 1992. Transpositional recombination: mechanistic insights from studies of *Mu* and other elements. *Annu. Rev. Biochem.* 61: 1011–1051.
- Mizuuchi, M. & K. Mizuuchi, 1993. Target site selection in transposition of phage *Mu*. *Cold Spring Harb. Symp. Quant. Biol.* 58: 515–523.
- Muller, H.P. & H.E. Varmus, 1994. DNA bending creates favored sites for retroviral integration: an explanation for preferred integration sites in nucleosomes. *EMBO J.* 13: 4704–4714.
- Norlander, J.M., T. Kempe & J. Messing, 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* 26: 101–106.
- O'Brochta, D.A. & P.W. Atkinson, 1996. Transposable elements and gene transformation in non-drosophilid insects. *Insect Bioch. Molec. Biol.* 26: 739–753.
- O'Brochta, D.A., W.D. Warren, K.J. Saville & P.W. Atkinson, 1996. *Hermes*, a functional non-drosophilid insect gene vector from *Musca domestica*. *Genetics* 142: 907–914.
- O'Brochta, D.A., W.D. Warren, K.J. Saville & P.W. Atkinson, 1994. Interplasmid transposition of *Drosophila hobo* elements in non-drosophilid insects. *Mol. Gen. Genet.* 244: 9–14.
- Osborne, B.I., C.A. Corr, J.P. Prince, R. Hehl, S.D. Tanksley, S. McCormick & B. Baker, 1991. *Ac* transposition from a T-DNA can generate linked and unlinked clusters of integrations in the tomato genome. *Genetics* 129: 833–844.
- Peterson, T., 1990. Intragenic transposition of *Ac* generates a new allele of the maize *P* gene. *Genetics* 126: 469–476.
- Pinkerton, A.C., D.A. O'Brochta & P.W. Atkinson, 1996. Mobility of *hAT* transposable elements in the Old World American bollworm, *Helicoverpa armigera*. *Insect Mol. Biol.* 5: 223–227.
- Pruss, D., F.D. Bushman & A.P. Wolffe, 1994. Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. *Proc. Natl. Acad. Sci. USA* 91: 5913–5917.
- Pryciak, P.M., A. Sil & H.E. Varmus, 1992. Retroviral integration into minichromosomes *in vitro*. *EMBO J.* 11: 291–303.
- Pryciak, P.M. & H.E. Varmus, 1992. Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. *Cell* 69: 769–780.

- Rommens, C.M., G.N. Rudenko, P.P. Dijkwel, M.J. van Haaren, P.B. Ouwkerk, K.M. Blok, H.J. Nijkamp & J. Hille, 1992. Characterization of *Ac/Ds* behaviour in transgenic tomato plants using plasmid rescue. *Plant Mol. Biol.* 20: 61–70
- Rudenko, G.N., H.J. Nijkamp & J. Hille, 1992. *Ds* readout transcription in transgenic tomato plants. *Mol. Gen. Genet.* 243: 426–433.
- Sambrook, J., E.F. Fritsch & T. Maniatis, 1989. *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory Press, New York.
- Sanger, F., S. Nicklen & A.R. Coulson, 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74: 5463–5467.
- Sarkar, A., C.J. Coates, S. Whyhard, U. Willhoeft, P.W. Atkinson & D.A. O'Brochta, 1997. The *Hermes* element from *Musca domestica* can transpose in four families of cyclorrhaphan flies. *Genetica* 99: 15–29.
- Schmidt-Rogge, T., B. Weber, T. Boerner, E. Brandenburg, O. Scheider, & M. Meixner, 1994. Transposition and behavior of the maize transposable element *Ac* in transgenic haploid *Datura innoxia* Mill. *Plant Science* 99: 63–74.
- Schneider, T.D. & R.M. Stephens, 1990. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* 18: 6097–6100.
- Sherrat, D., 1989. *Tn3* and related transposable elements: site-specific recombination and transposition, pp. 163–184 in *Mobile DNA*, edited by D.E. Berg and M.M. Howe. American Society of Microbiology, Washington, DC.
- Smit, A.F.A. & A.D. Riggs, 1996. Tiggers and other DNA transposon fossils in the human genome. *Proc. Natl. Acad. Sci. USA* 93: 1443–1448.
- Smith D., J. Wohlgenuth, B.R. Calvi, I. Franklin & W.M. Gelbart, 1993. *hobo* enhancer trapping mutagenesis in *Drosophila* reveals an integration specificity different from *P* elements. *Genetics* 135: 1063–1076.
- Spradling, A.C. & G.M. Rubin, 1982. Transposition of cloned *P* elements into *Drosophila* germ line chromosomes. *Science* 218: 3410–347.
- Streck, R.D., J.E. MacGaffey & S.K. Beckendorf, 1986. The structure of *hobo* transposable elements and their integration sites. *EMBO J.* 5: 3615–3623.
- Sugimoto, K., Y. Otusuki, S. Saji, & H. Hirochika, 1994. Transposition of the maize *Ds* element from a viral vector to the rice genome. *Plant J.* 5: 863–871.
- Tsay, Y-F., M.J. Frank, T. Page, C. Dean & N.M. Crawford, 1993. Identification of a mobile endogenous transposon in *Arabidopsis thaliana*. *Science* 260: 342–344.
- van Luenen, H.G. A.M. & R.H.A. Plasterk, 1994. Target site choice of the related transposable elements *Tc1* and *Tc3* of *Caenorhabditis elegans*. *Nucleic Acids Res.* 22: 262–269.
- Wardlaw, A.C., 1985. pp. 124–126 in *Practical statistics for experimental biologists*. Wiley, New York.
- Warren, W.D., P.W. Atkinson & D.A. O'Brochta, 1994. The *Hermes* transposable element from the house fly, *Musca domestica*, is a short inverted repeat-type element of the *hobo*, *Ac*, and *Tam3* (*hAT*) element family. *Genet. Res. Camb.* 64: 87–97.
- Warren, W.D., P.W. Atkinson & D.A. O'Brochta, 1995. The Australian bushfly *Musca vetustissima* contains a sequence related to transposons of the *hobo*, *Ac*, and *Tam3* family. *Gene* 154: 133–134.
- Weil, C.F. & S.R. Wessler, 1993. Molecular evidence that chromosome breakage by *Ds* elements is caused by aberrant transposition. *Plant Cell* 5: 515–522.
- Xiong, Y. & T.H. Eickbush, 1988. The site-specific ribosomal DNA insertion element *RIBm* belongs to a class of non-long-terminal-repeat retrotransposons. *Mol. Cell. Biol.* 8(1): 114–123.
- Zou, S., N. Ke, J.M. Kim & D.F. Voytas, 1996. The Saccharomyces retroposon *Ty5* integrates preferentially into regions of silent chromatin at the telomeres and mating loci. *Genes Dev.* 10: 634–645.