Testing the palindromic target site model for DNA transposon insertion using the Drosophila melanogaster P-element

Raquel S. Linheiro and Casey M. Bergman (Michael Smith Building, Faculty of Life Sciences, University Of Manchester)

Abstract
Understanding the molecular mechanisms that influence transposable element target site preferences is a fundamental challenge in functional and evolutionary genomics. Growing evidence from a wide variety of prokaryotes and eukaryotes indicates that DNA transposons recognize palindromic target site motifs. Here we use over 10,000 accurately mapped P-element insertions in Drosophila melanogaster genome to test predictions of the palindromic target site model for DNA transposon insertion. We provide evidence that the P-element targets a 14 bp palindromic motif that can be identified at sequence level, which predicts the local spacing, hotspots and strand orientation of P-element insertions. Intriguingly, we find that although P-element destroys the complete 14 bp target site upon insertion, the terminal three nucleotides of the P-element inverted repeats complement and restore the original target site motif, suggesting a link between transposon target sites and their terminal inverted repeats.

Introduction
The P-element is a naturally occurring mutagenic agent that has been engineered to facilitate a large number of genetic and genomic manipulations in Drosophila, including gene disruption, chromosomal aberrations, reporter gene analysis, gene and enhancer trapping and misexpression of endogenous genes. Understanding the mechanisms of target site selection for new insertions is critical for functional genomics applications using the Drosophila P-element.

O’Hare and Rubin (1) first demonstrated that P-elements prefer to insert into an 8-bp GC-rich consensus sequence (GGCCAGAC), Liao et al. (2) analyzed a much larger set of 1469 P-element insertions and concluded “although there are base preferences at each position, these are not strong enough to generate a clear consensus sequence.” Instead, these authors argued that the P-element recognizes a 14-bp palindromic structural motif based on a pattern of hydrogen bonding at the target site. More recently, Julian (3) analyzed a sample of 795 P-elements and reported a 14-bp nonpalindromic pattern of hydrogen bonding at the target site. More recently, Julian (3) analyzed a sample of 795 P-elements and reported a 14-bp nonpalindromic pattern of hydrogen bonding at the target site. These conflicting results have lead us to clarify whether the P-element targets a specific sequence motif and, if so, whether this motif is a palindrome in order to better understand the target site selection of the P-element and other DNA transposons. Results presented here are also reported in (4).

Results

<table>
<thead>
<tr>
<th>P-element family name</th>
<th>Number of Insertions</th>
<th>Number Mapped to 1 bp</th>
<th>Number Mapped to +/- Strand</th>
<th>Number on + Strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>GT1</td>
<td>556</td>
<td>531 (95.50%)</td>
<td>496 (93.4%)</td>
<td>260 (52.42%)</td>
</tr>
<tr>
<td>SUPor-P</td>
<td>2297</td>
<td>2288 (99.61%)</td>
<td>2134 (92.27%)</td>
<td>1065 (49.91%)</td>
</tr>
<tr>
<td>EPgy2</td>
<td>3496</td>
<td>3473 (99.34%)</td>
<td>3258 (93.81%)</td>
<td>1630 (48.03%)</td>
</tr>
<tr>
<td>XP</td>
<td>5311</td>
<td>4974 (93.65%)</td>
<td>4972 (99.96%)</td>
<td>2479 (49.89%)</td>
</tr>
<tr>
<td>Total</td>
<td>11690</td>
<td>11266 (96.62%)</td>
<td>10860 (96.40%)</td>
<td>5434 (50.94%)</td>
</tr>
</tbody>
</table>

Table 1. Summary of reliably mapped P-element insertions in the Release 5.6 Flybase genome annotation. Numbers reported include redundant insertions in the same insertion site.

Figure 1. The P-element targets a 14 bp palindromic target site motif. (A) Sequence logo depicting the relative base usage for a 51 bp window centered around 10,221 P-element insertion sites. The insertion site on the positive strand is just before position zero, and the insertion site on the negative strand is just after position seven. The Y-axis represents the usage of bases in the motif relative to the random expectation of equal frequency. (B) The terminal three nucleotides of the P-element inverted repeats restore and complement the optimal target sequences flanking the target site duplication (TSD) on both ends of the P-element insertion. Specifically, the terminal three bp flanking the TSD at the 5' (ATR...) and 3' (…WAY) end of the target site motif are complementary to the terminal three bp of the 3’ (…ATG) and 5’ (CAT…) ends, respectively, of the P-element terminal inverted repeats. Note that this occurs on both ends of the P-element regardless of whether the 5’ or 3’ insertion site (and the corresponding orientation of the P-element insertion) is used in the palindromic staggered-cut target site.

Figure 2. The 14 bp target site motif predicts non-random local spacing between P-element insertion sites and reveals two types of hotspots. (A) Distances, in base pairs (bp), between all consecutive P-element insertions in the genome. (B) Distances between consecutive P-element insertions on the same strand (+/+) or (-/ -), showing same-strand hotspots at a distance of zero bp. Distances between consecutive P-element insertions on opposite strands show opposite-strand hotspots at a distance of eight bp only in the expected +/- configuration (C) but not in the-/+- configuration (D). Note that the X-axis has been truncated at 50 bp in all four panels for clarity.

Conclusions
• The P-element targets a 14 bp staggered-cut palindromic target motif (Figure 1A).
• The target site motif is destroyed on P element insertion but restored and complemented by the P element terminal inverted repeats (Figure 1B).
• The palindromic nature of the target motif predicts random strand integration (Table 1).
• The staggered cut nature of the target motif predicts the local spacing of P element insertions (Figure 2).
• The 14 bp target site motif predicts hotspots for P element insertion (Figure 3).
• The palindromic target site model can confirm 90.4% (9,243/10,221) of the annotated insertion sites, hotspots and background DNA.

Acknowledgements
We thank Sam Griffiths-Jones, Stefan Roberts, Don Rio, Roger Hoskins and Hugo Bellen and members of the Bergman Lab for helpful comments during the project.

References