Volume 11 Number 20 1983 Nucleic Acids Research

Target sequences for the *C. elegans* transposable element Tc1

Bradley Rosenzweig, Louise W.Liao and David Hirsh

Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO 80309, USA

Received 23 August 1983; Accepted 16 September 1983

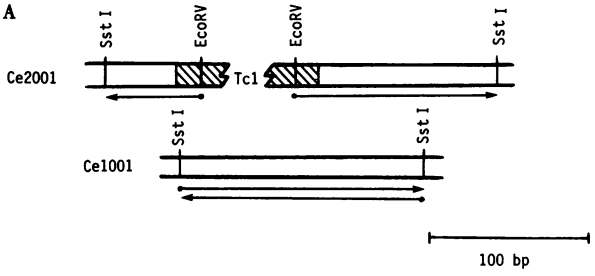
ABSTRACT

The target sequences for two independent insertions of the transposable element Tc1 from Caenorhabditis elegans show homology. Because both insertions are at palindromic TA/AT sequences, the exact boundaries of Tc1 cannot be distinguished; Tc1 could be 1610 bp and flanked by a 2-bp duplication of the target site or it could be 1612 bp and without target site duplication. The latter possibility implies a novel manner for insertion of a transposable element.

INTRODUCTION

Tc1, a transposable element in the nematode Caenorhabditis elegans, was originally recognized and isolated because it is the basis for many of the DNA polymorphisms between the Bristol and Bergerac strains (1,2). A polymorphism due to a Tc1 insertion aided in the mapping of actin genes to chromosome V (2). The Bristol strain contains about 30 copies of Tc1 and the Bergerac strain contains about 300 copies. The Tc1 family exhibits a highly conserved restriction map but it is hypervariable in copy number and genomic locations among several C. elegans strains, indicating mobility in evolutionary time (3,4). Tc1 excision can occur within a few generations in a laboratory stock, but no direct evidence exists for induced transposition (3).

A constant feature of prokaryotic and eukaryotic transposable elements is that they duplicate a sequence at the site of insertion. A Tc1-filled site in the Bergerac strain and its corresponding empty site in the Bristol strain were sequenced earlier in our laboratory and it appeared that a 2 base-pair (bp) duplication occurred upon insertion (5). The study presented here compares the target sequences for a second pair of Tc1-filled and empty sites with the original sequences. We have found that the insertion points for both Tc1 integrations are at palindromic $\frac{TA}{AT}$ sequences, creating a dilemma in assigning the junction points between the empty sites and the Tc1 insertions. The sequences can be interpreted either as creating a duplication or not.



B

5' (G)AGCTCTGGGTGTTGGGATATTACTGAAAACTATAGTACATATTTATGTACTTACAGTGG
Sst I

TATAACAAATCGGAAAATTGCGTATACCGTGTTGGAAAAATCTTGAAAATTTATAAGTGAAT

TGCTCGCAAATTATAACGAAGTTTCAGCAGAGAGCT(C) 3'
Sst I

Figure 1. Analyses of the target sequences in the fragments Ce2001 and Ce1001. (A) The fragments shown represent enlargements of the insertion region mapped by Emmons et al. (3). The hatched region designates the Tc1 insertion. Arrows indicate the direction and extent of the sequence analyses. (B) The target sequence from 45 bp 5' to 110 bp 3' to the insertion point (boxed) is shown.

MATERIALS AND METHODS

The Bergerac recombinant plasmid pCe2001 and the Bristol recombinant plasmid pCe1001 were isolated and provided by Emmons et al. (3). DNA restriction fragments from pCe2001 were labeled with [γ-^{32}P]ATP at the 5' termini by T4 polynucleotide kinase and sequenced as described by Maxam and Gilbert (6). The SstI restriction fragment from pCe1001 that contains the target site was subcloned into M13mp10 in both orientations and sequenced by the Sanger dideoxy method (7,8).

RESULTS AND DISCUSSION

Sequence determination of Tc1 target sites

Pairs of subcloned DNA fragments were isolated from the Bergerac and Bristol strains in order to compare the target sequences with and without an inserted Tc1 element (3,4). The members of a pair correspond to the same genomic location, but differ in that the Bergerac-derived fragment contains a Tc1 element. The sequencing strategies for Ce2001, a Tc1-filled site, and Ce1001, its corresponding empty site, are outlined in Fig. 1A. The region in Ce2001 that was sequenced was isolated from a restriction endonuclease digest using EcoRV, which cleaves within the inverted terminal repeats of Tc1, and

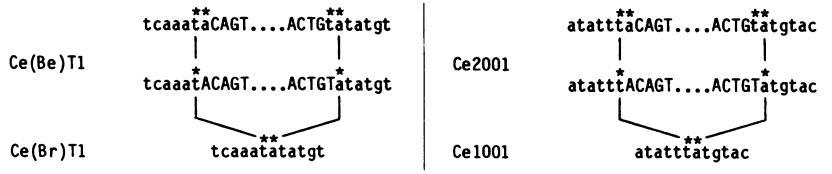


Figure 2. Comparisons of two Tc1-filled sites with their empty sites. Possible boundaries for Tc1 are in upper case, and the target sites are in lower case. The TA insertion point is marked (*). The top sequences for the Tc1-filled sites illustrate a duplication of the insertion point, while the alternative sequences for the Tc1-filled sites illustrate insertion without target site duplication.

SstI, which cleaves in the flanking regions. The target sequence in Ce1001 was determined by analysis of the corresponding SstI restriction fragment. Except at the Tc1 insertion point, the sequences in Ce2001 and Ce1001 are identical for the flanking 155 bp analyzed in both strains (Fig. 1B). The target DNA sequences for another pair of Tc1-filled and empty sites, designated Ce(Be)T1 and Ce(Br)T1 respectively, have been described earlier (5).

Does Tc1 insertion generate a duplication of the target sequence?

The nucleotide sequences at the sites of these two Tc1 insertions are shown in Fig. 2. Examination of these sites shows that the exact limits of Tc1 cannot be defined. In both cases, the Tc1 element inserted at a $\frac{TA}{AT}$ sequence, which allows two possible interpretations for the Tc1 boundaries. One interpretation, presented previously, is that the boundaries of Tc1 are 5' CAGT...ACTG 3' and that the two nucleotides (TA) of the target sequence are duplicated immediately flanking the Tc1 sequence (5). This interpretation would make the Tc1 element 1610 bp long with 54-bp inverted terminal repeats. However, the juxtaposition of a palindromic insertion point with the palindromic inverted terminal repeats of Tc1 allows another interpretation. The boundaries of Tc1 could be 5' ACAGT...ACTGT 3' with no target site duplication accompanying insertion. Tc1 would be 1612 bp long with 55-bp inverted terminal repeats. Duplication has been hypothesized to result from a staggered cleavage of the target sequence (9). Therefore, if Tc1 does not generate a duplication, insertion must occur by a novel, blunt cleavage of the target site.

Comparison of the two different insertion sites

Immediately surrounding the Tc1 insertion sites, 12 of 16 bp are identical between the two empty sites (Fig. 3A), although direct alignment of the two insertion points shows little homology (Fig. 3B). It is unclear whether this homology bears any association with site-selective transposition.

```
A.   Ce(Br)T1  ...TTCAAATATATGTGCT...
                    * ** ** ***** **
     Ce1001    ...TACATATTTATGTACT...

B.   Ce(Br)T1  ...TTTCAAATATATGTGC...
                      *  *** *
     Ce1001    ...ACATATTTATGTACTT...
```

Figure 3. Comparison between the two empty sites. (A) Alignment of the empty sites to give maximum homology, where 12 of 16 nucleotides are identical, requires shifting the [TA] insertion points relative to each other. (B) Direct alignment of the [TA] insertion points results in only 5 of 16 nucleotides being identical.

Unlike the Tc1 insertion site analyzed previously, the site sequenced in this study does not share features with preferred integration sites of particular prokaryotic transposons (5). There is no homology between the target sequence and sequences within the terminal repeats of Tc1. The 155-bp target region is 66% A+T, the other target region is 72% A+T, and C. elegans DNA is 64% A+T overall, so it is uncertain if AT-rich regions are preferred (5,10). The sequence surrounding the insertion point is not palindromic, except for the $\frac{TA}{AT}$ base pairs.

ACKNOWLEDGEMENTS

We thank Dr. S. Emmons for his gift of the recombinant plasmids pCe2001 and pCe1001. We also thank Richard Jefferson for provocative and helpful discussions. This research was supported by the National Institutes of Health Grant GM26515 to D.H. and a postdoctoral fellowship from the American Cancer Society (PF-1720) to L.W.L.

REFERENCES

1. Emmons, S.W., Klass, M.R. and Hirsh, D. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 1333-1337.
2. Files, J.G., Carr, S. and Hirsh, D. (1983) J. Mol. Biol. 164, 355-375.
3. Emmons, S.W., Yesner, L., Ruan, K. and Katzenberg, D. (1983) Cell 32, 55-65.
4. Liao, L.W., Rosenzweig, B. and Hirsh, D. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 3585-3589.
5. Rosenzweig, B., Liao, L.W. and Hirsh, D. (1983) Nucleic Acids Res. 11, 4201-4209.
6. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
7. Messing, J. and Viera, J. (1982) Gene 19, 269-276.
8. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467.
9. Kleckner, N. (1981) Annu. Rev. Genet. 15, 341-404.
10. Sulston, J.E. and Brenner, S. (1974) Genetics 77, 95-104.