Nucleic Acids Research, Vol. 19, No. 20 **5669–5672**

Germline excision of the transposable element Tc1 in *C.elegans*

Donald G.Moerman, Jane E.Kiff[1] and Robert H.Waterston[1]
Department of Zoology, University of British Columbia, Vancouver, V6T 2A9, Canada and
[1]Department of Genetics, Washington University School of Medicine, St Louis, MO 63110, USA

Received July 3, 1991; Revised and Accepted September 23, 1991

ABSTRACT

We have examined eight germline revertants generated by the excision of Tc1 from a site within the *unc-22* gene of *Caenorhabditis elegans*. A rich variety of rearrangements accompanied Tc1 excision at this site, including transposon 'footprints', deletions of sequences flanking the insertion site and direct non-tandem duplications of flanking DNA. With only modest modification the double-strand gap repair model for transposition, recently proposed by Engles and co-workers (Cell 62: 515–525 1990), can explain even the most complex of these rearrangements. In light of this model rearrangements of the target site accompanying transposition/excision may not be the end result of imprecise excision of the element. Instead, these rearrangements may be the result of imprecise repair of the double-strand gap by the host replication and repair machinery. Sequences surrounding an insertion site influence the fidelity of gap repair by this machinery. This may lead to a number of possible resolutions of a double-strand gap as documented here for a Tc1 site in *unc-22*.

INTRODUCTION

There is still controversy over whether transposition occurs by a conservative 'cut and paste' mechanism similar to lambda excision (1) or a copying (replicative) mechanism (2,3). In the former model a necessary step in transposition is the excision of the element from its original genomic site. Germline excision of transposons has been documented in a number of systems (for examples see reviews 4–7). If germline excision is part of a conservative (nonreplicative) mechanism of transposition then the sequence of events associated with excision might reflect this mechanism of transposition. For this reason we have examined a variety of genetic revertants generated by the excision of the transposon Tc1 in *C. elegans*.

The 1.6 kb transposon Tc1 of *C. elegans* is similar in structure to insertion sequence (IS) elements of prokaryotes and P and Ac elements in eukaryotes: Tc1 has 54 bp perfect terminal inverted repeats and potentially encodes a protein of 273 aa (8). Upon insertion the element is flanked by a TA dinucleotide duplication. Whether this TA is generated by a staggered cut upon insertion or whether the element has 55 bp terminal inverted repeats and makes a blunt end cut at TA is unknown (9–12,6; also see Discussion). Germline activity of Tc1, including excision and transposition, is regulated by chromosomal genetic components termed mutators (13–15). By sequencing somatic and germline excision sites precise and imprecise excision of Tc1 in somatic and germline cells has been documented (16,11).

Our interest in mutator induced excision was stimulated by mutants of *unc-22*, a gene that encodes a protein located within muscle sarcomere myosin containing A bands (17,18). Revertants of an *unc-22::Tc1* allele were found to include both phenotypically full and partial revertants suggesting a variety of excision events. We have sequenced the region encompassing the former Tc1 site of several partial and full revertants. Most excision events examined, including 3 of 4 apparent full revertants were accompanied by rearrangements (DNA sequence changes). The nature of these rearrangements sheds light on the excision process and also suggests a major role for DNA repair after or accompanying the excision event.

MATERIALS AND METHODS

Genetic strains and isolation of revertants

All reversion experiments were done using the *C. elegans* strain RW7096. Isolation and detailed characterization of this strain have previously been described (15,19). Briefly, this strain has essentially a Bristol (N2) genetic background but contains *mut-6(st702)*, an activator of Tc1 transposition, on chromosome IV along with *unc-22(st192::Tc1)*. Germline reversion at this site in RW7096 occurs at a frequency of 10^{-3} (15,19). Individual RW7096 animals were placed on separate 100 mm petri plates and screened after about two generations for well moving non-twitching animals and only single revertants were picked per plate. Revertants were propagated to generate homozygous wild type revertant lines.

Molecular cloning and sequencing of revertants

For DNA isolation, Southern analysis, cloning and sequencing, standard techniques were utilized (20; also see 21). The original insertion mutant, *st192::Tc1*, and the 8 revertants were cloned into either lambda 2001 or pUC18/19. The region of interest for each was subcloned into M13mp18/19, and the sequence of both strands was determined by the chain termination method (22; 23). Analysis of the sequences, i.e. searches for direct and inverted repeats, was done by eye and using DNA sequence analysis programs (24).

RESULTS

Identification of imprecise Tc1 excision

For our studies on Tc1 excision we chose the *unc-22* allele, *unc-22(st192::Tc1)*. This allele is stable in non-mutator backgrounds: the reversion frequency is less than 10^{-6}. However, reversion frequency is 3 orders of magnitude higher (i.e., approximately 10^{-3}) in the *mut-6* strain RW7096 (15,19). The *st192::Tc1* site resides within the coding region of the *unc-22* gene (18,21). The 700kd *unc-22* gene product has similarities to titin and smooth muscle light chain kinase and is a component of the thick filaments in nematode muscle (25). Animals homozygous for mutations in the gene exhibit a characteristic 'twitching' phenotype under normal culture conditions. Heterozygotes can be induced to twitch violently for extended periods (eg overnight) by placing them in a 1% nicotine solution, but are indistinguishable from wild type animals under normal culture conditions (26). Revertants of the *unc-22(st192::Tc1)* allele are easily recognized under normal culture conditions by their lack of twitching, and by their greater size relative to Unc animals.

In order to assess the range of potential excision events at this site we collected 51 revertants in the *mut-6* background. All revertants were phenotypically wild type under normal culture conditions. We further tested each revertant by immersion in 1% nicotine, and even under these conditions 47 of the revertants could not be induced to twitch. The remaining 4 exhibited a weak and transient twitch, lasting several minutes. Extragenic suppressors of *unc-22* mutants are rare and all those identified to date are unusual alleles of the *unc-54* myosin heavy chain locus (17,27). These *unc-54* suppressors exhibit their own phenotype, namely varying degrees of paralysis, and none of our revertants were slow moving or paralyzed. Since all of the revertants we isolated were phenotypically wild type, and since revertants were found at a high frequency and only in strains possessing mutator activity, we assumed that they were intragenic revertants, presumably involving excision of the element.

To confirm this and to examine the molecular events underlying the phenotypic reversion we first performed Southern blots using probes spanning the entire gene on DNA from 8 of the revertants, including all 4 of the partial phenotypic revertants (*st530*, *st531*, *st529*, and *st528*) and 4 randomly selected, full phenotypic revertants (*st534*, *st532*, *st527*, and *st533*). Within the level of sensitivity of Southern analysis, 5 of the revertants were apparently wild type, while 3 others, all from the partial phenotypic revertant class, had deletions of 1, 1.2, and 2 kb, extending from the empty site in either direction into the *unc-22* gene.

We next cloned and sequenced the regions of the *unc-22* gene that the Southern analysis indicated were altered in the 3 partial revertants and the region surrounding the *st192::Tc1* target site in the other revertants. The sequenced revertants were compared to the sequence of the wild type gene (18) and are shown in Figure 1.

Nature of the small rearrangements

Only one of four apparently full revertants (*st534*) is actually a precise excision event. The other full revertants involve insertion/deletion (*st532*), or, insertions (*st527*, *st533*), all of which maintain the reading frame of the *unc-22* gene (see 21). In *st532*, *st527*, and, *st533* some of the inserted bases appear to derive from the ends of Tc1. We say appear because there is an ambiguity inherent at the ends of Tc1: either Tc1 duplicates the TA dinucleotide into which it inserts, or else there is an extra A at the 5′ end of the element and an extra T at the 3′ end (for details see review 6 and references therein). Regardless of the origin of these bases, we consider them here as inserted bases, relative to the wild type sequence. The simplest case is *st532* in which there is a net GA to TG substitution (this results in a single conservative residue alteration). This could have resulted if during the process of excision and repair of the empty site, 3 bases (GAT) were removed from the upstream *unc-22* sequence (relative to the former insertion site) and 3 bases (TGT) from the 3′ end of Tc1 were added to the downstream *unc-22* sequence (relative to the former insertion sequence). Similarly, in *st527* and *st533* the last three bases (TGT) from the 3′ end of Tc1 are still present at the insertion site, but, in addition 1 or 2 bases (A or AC), respectively, from the 5′ end of Tc1 are also still at the site. These latter two events are complicated by duplicated *unc-22* sequence as well as sequence from the ends of Tc1 (see below). To summarize, the TGT sequence from the 3′ end of Tc1 is present in all 3 revertants, and none (*st532*), one (*st527*), or two (*st533*), bases may be present from the 5′ end of the element.

Of the partial revertants, one (*st530*) also fell into the insertion class, with a net six base insertion. Like the two full revertants *st527* and *st533*, the inserted sequence duplicates flanking *unc-22* sequence. The end points of all three duplications, and their lengths, are ambiguous; some bases may represent residue of the Tc1 ends or could represent duplications. Figure 1 presents one possible combination of duplication with ends of the Tc1 element that accounts for origins of all the inserted and/or substituted bases.

For example, for *st530* the six base insertion could represent a 6 base duplication between the TA dinucleotide where the element resided as shown. Alternatively the event could result from an 8 base duplication after a two base deletion at the Tc1 insertion site. Examination of the sequence shows this can happen in either of 2 ways: ATTTTGGG repeated twice in tandem with deletion of the AT immediately 5′ of the insertion site, or TTTGGGAT repeated twice in tandem with deletion of the TA

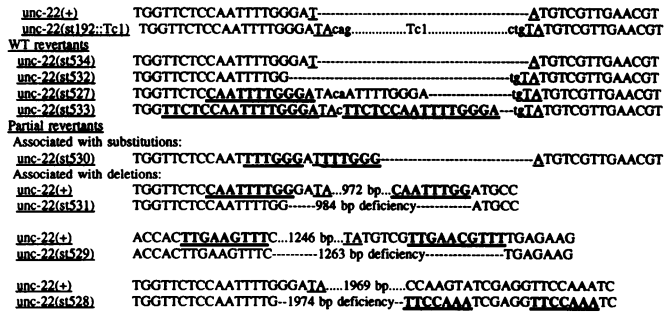

Figure 1. Germline revertants of *unc-22(st192:Tc1)*. The duplicated TA dinucleotide target sequence is singly underlined. Nucleotides of the Tc1 element are in lower case. Extra nucleotides accompanying the excision event are aligned where possible with the excision mutant. In those cased where the extra nucleotides appear to represent duplication of flanking sequences, both the extra nucleotides and the sequences they may be derived from are doubly underlined and are bold. Sequence identities possibly involved in deficiency formation are also doubly underlines and in bold. (...) Sequences omitted from the figure; (---) gaps introduced to permit alignment of flanking sequences.

dinucleotide target site. With the possible exception of *st530*, none of the duplications are tandem.

Extensive deletions may accompany excision

Large target sequence deletions accompany excision of the element in three of the partial revertants (*st531*, *st528*, and *st529*). The deletions are 984, 1974 and 1246 bases in length, respectively (the effects of these deletions on the *unc-22* gene product are described in 21). Two of these deletions, *st531* and *st528*, extend downstream from the empty site, while another, *st529*, extends in the upstream direction. All three deletions have their proximal breakpoints within a few bases of the TA insertion site while deletion breakpoints distal to the empty site are different in all three cases. An interesting feature of two of the deletion revertants, *st531* and *st529*, is that we find short sequences (8 to 10 bases) at the distal end of the deletion that bear strong homology to sequences proximal to the break in the wild type sequence (see figure 1). This suggests involvement of a recombinational mechanism in the repair of the strands following excision/deletion. The third deletion revertant, *st528*, is more complicated in that it has both a large deletion (1974 bases) and a direct non-tandem duplication of part of the downstream *unc-22* sequence at which it is joined. It does not seem that there is any strong homology between the proximal and distal sequences in this case.

DISCUSSION

Our data points out the diversity of rearrangements that can result from Tc1 excision. Tc1 excision can be precise (16), and *st534* is a further example of this type of event. Rearrangements accompanying Tc1 excision have been observed previously (11,16), but in all of these cases the rearrangements consist of a small portion of the ends of the element remaining at the target site (a footprint). *St532* falls into this category, and several others include this feature in conjunction with the additional feature of duplications of flanking sequences. We have also observed large deletions accompanying excision of the element, and we have observed one case in which both deletion and duplication of genomic sequences occurred.

A model of P transposition in *Drosophila melanogaster* has been proposed which suggests that transposition is initated by a 'cut and paste' mechanism but that the P element is then restored at the donor site by double-strand gap repair (28). An implication of the model relevant to this study is that rearrangements at target sites, which had previously been attributed to 'imprecise excision' of the transposon, are actually the result of 'imprecise repair' of the double-strand break caused by the excision of the transposon. A variety of excision products occur because of copying errors made during double-strand gap repair.

Recently Plasterk (29) has shown that this double strand gap repair model for transposition is compatible with many observations on the behaviour of Tc1. In particular, the model explains a set of puzzling observations concerning enhanced Tc1 excision at *unc-22* in certain heteroallelic strains (19). In this new model enhanced reversion rates in heteroallelic combinations can be explained as double strand gap repair from the homologous chromosome (29).

In the context of this study an attraction of the model is that it allows us to explain even the most complicated rearrangements at the *unc-22(Tc1:st192)* site. Where both homologs have Tc1 inserts, as in the current study, a precise event, like *st543*, can occur only if the ends are religated without gap repair after Tc1 excision. Restoration of a wild type sequence without a duplicated TA at the target site suggests the double-strand break generates blunt, not staggered ends. This conclusion is further supported by earlier observations which showed that Tc1 upon forming a circle includes a T and an A at its ends (30). Taken together these results suggest that Tc1 has a 55 base pair terminal inverted repeat, not a 54 base pair repeat.

In the simplest rearrangement, *st532*, TG substitutes for GA, which results in a single residue alteration at the target site. If excision is always precise then two events must have occurred: gap repair, although initiated at the 3′ end, was aborted early, and at the same time exonuclease activity trimmed back the 5′ end but gap repair of this strand did not occur. The net result after religation is a small deletion 5′ of the target site and a 3′ Tc1 'footprint' at the site.

The deletion revertants represent a new class of Tc1 excision event. In two cases we found homologous sequences on either side of the deletion breakpoint which suggests illegitimate recombination may have played a part in generating the deletions (reviewed in 31). The internal deletions may have arisen via a process similar to that proposed to explain P element internal deletions (28). The presence of homologies near the ends of these transposon mediated deletion events contrast with the spontaneous deletions previously reported in *C. elegans* where short homologies are not found at deletion end points (32).

The direct non-tandem duplications, *st527*, *st533*, and possibly *st530*, are separated by Tc1 end sequences. The non-tandem nature of these duplications is most curious, and an explanation of how they arose is a challenge for any model of transposon excision/transposition. With only modest modification the model proposed by Engels et al., (28) can readily explain the origin of such rearrangements (see figure 2). We propose that in these three revertants Tc1 excised precisely and double-strand gap repair was initiated, but processed only a few nucleotides before aborting the task. If the ends had been ligated and filled in we

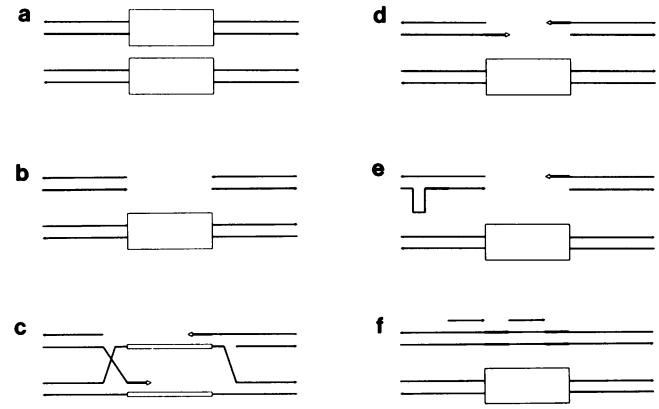

Figure 2. Double-strand gap repair model of Tcl excision and illegitimate repair. a. 2 homologs with Tc1 inserts. b. Tc1 precise excision from one homolog. c. Gap repair initiated from homolog (or sister chromatid). d. Gap repair aborted after extension a few bases into Tc1 inverted repeat. Effectively, the strand and repair enzymes switch templates. e. During reannealling with normal pairing partner a hairpin or loop has formed causing a shift in point of extension by DNA polymerase. f. Ends have been extended and ligation and filling in has occurred. End result is a direct duplication with the 3′ duplication flanked by Tc1 bases (Figure modified from Engels et al., 1990).

would simply have observed a 'footprint' (see revertant *st532*; and references 11,16) as explained in Plasterk (29). However, if rather than aborting repair, the replication machinery and the strand to be duplicated actually switch templates, and are displaced 5′ in this new pairing, then one would observe a direct non-tandem duplication. It would be non-tandem because it initiated copying from the Tc1 bearing homolog before switching templates.

If the double strand gap repair model for transposition is correct and general then repair of an empty site includes the host replication and repair machinery. In the repair of gaps this machinery will be affected by sequences surrounding a target site. Therefore, there may be a number of ways to resolve a double-strand gap at any particular site, some of which we have documented here for a Tc1 site in *unc-22*.

ACKNOWLEDGEMENTS

We thank Ron Plasterk for communicating results prior to publication. We also thank Teresa Rogalski, Mary Gilbert and Erin Gilchrist for comments on the manuscript. This work was supported by U.S. Public Health Service grant GM23883 to R.H.W. D.G.M. acknowledges support by the National Science and Engineering Research Council of Canada.

REFERENCES

1. Berg, D.E. 1983 Structural requirements for IS50-meditated gene transposition. Proc. Natl. Acad. Sci. USA 80: 792–796.
2. Shapiro, J. 1979 Molecular model for the transposition and replication of bacteriophage Mu and other elements. Proc. Natl. Acad. Sci. USA 76: 1933–37.
3. Craig, N.L. 1990 P element transposition. Cell 62: 399–402.
4. Engels, W.R. 1989 P elements is Drosophila. In Mobile DNA. D.E. Berg and M.M. Howe, eds. (Washington, D.C.: American Society of Microbiology), pp. 437–484.
5. Fedoroff, N.V. 1989 Maize transposable elements. In Mobile DNA. D.E. Berg and M.M. Howe, eds. (Washington, D.C.: American Society of Microbiology), pp. 375–411.
6. Moerman, D.G. and R.H. Waterston 1989 Mobile elements in *Caenorhabditis elegans* and other nematodes. In Mobile DNA. D.E. Berg and M.M. Howe, eds. (Washington, D.C.: American Society of Microbiology), pp. 537–556.
7. Berg, D.E. and M.M. Howe, eds. 1989 Mobile DNA (Washington, D.C.: American Society for Microbiology).
8. Rosenzweig, B., L. Liao and D. Hirsh 1983a Sequence of the *C. elegans* transposable element Tc1. Nucleic Acids Res. 11:4201–4209.
9. Rosenzweig, B., L. Liao and D. Hirsh 1983b Target sequences for the *C. elegans* transposable element Tc1. Nucleic Acids Res. 11: 7137–7140.
10. Mori, I., G.M. Benian, D.G. Moerman and R.H. Waterston 1988a The transposon Tc1 of *Caenorhabditis elegans* recognizes specific target sequences for integration. Proc. Natl. Acad. Sci. USA 85: 861–864.
11. Eide, D. and P. Anderson 1988 Insertion and excision of the *Caenorhabditis elegans* transposable element Tc1. Mol. Cell. Biol. 8: 737–746.
12. Herman, R.H. and J. Shaw 1987 The transposable genetic element Tc1 in the nematode *Caenorhabditis elegans*. Trends in Genetics 3: 222–225.
13. Moerman, D.G. and R.H. Waterston 1984 Spontaneous unstable *unc-22* IV mutations in *Caenorhabditis elegans* var. Bergerac. Genetics 108: 859–877.
14. Collins, J., B. Saari and P. Anderson 1987 Activation of a transposable element in the germ line but not the soma of *Caenorhabditis elegans*. Nature (London) 328: 726–728.
15. Mori, I., D.G. Moerman and R.H. Waterston 1988b Analysis of a mutator activity necessary for germline transposition and excision of Tc1 transposable elements in *Caenorhabditis elegans*. Genetics 120: 397–407.
16. Ruan, K.S. and S.W. Emmons 1987 Precise and imprecise somatic excision of the transposon Tc1 in the nematode *C. elegans*. Nucleic Acids Res. 15: 6875–6881.
17. Moerman, D.G., S. Plurad, R. H. Waterston and D.L. Baillie 1982 Mutations in the *unc-54* myosin heavy chain gene of *Caenorhabditis elegans* that alter contractility but not muscle structure. Cell 29: 773–781.
18. Benian, G.M., J.E Kiff, N. Neckelmann, D.G. Moerman and R.H. Waterston 1989 Sequence of an unusually large protein implicated in regulation of myosin activity in *C. elegans*. Nature 342: 45–50.
19. Mori, I., D.G. Moerman and R.H. Waterston 1990 Interstrain crosses enhance excision of Tc1 transposable elements in *Caenorhabditis elegans*. Mol. Gen. Genet. 220: 251–255.
20. Sambrook, J., E.F. Fritsch and T. Maniatis 1989 Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
21. Kiff, J.E., D.G. Moerman, L.A. Schriefer and R.H. Waterston 1988 Transposon-induced deletions in *unc-22* of *Caenorhabditis elegans* associated with almost normal gene activity. Nature (London) 310: 332–333.
22. Sanger, F., S. Nicklen and A.R. Coulson 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74: 5463–5467.
23. Messing, J. 1983 New M13 vectors for cloning. Meth. Enzymol. 101:20–78.
24. Staden, R. 1986 The current status and portability of our sequence handling software. Nucleic Acids Res. 14: 217–231.
25. Moerman, D.G., G.M. Benian, R.J. Barstead, L. Schriefer and R.H. Waterston 1988 Identification and intracellular localization of the *unc-22* gene product of *Caenorhabditis elegans*. Genes and Dev. 2: 93–105.
26. Moerman, D.G. and D.L. Baillie 1979 Genetic organization in *Caenorhabditis elegans*: fine structure analysis of the *unc-22* gene. Genetics 91: 95–103.
27. Dibb, N.J., D.M. Brown, J. Karn, D.G. Moerman, S.L. Bolten and R.H. Waterston 1985 Sequence analysis of mutations that affect the synthesis, assembly, and enzymatic activity of the *unc-54* myosin heavy chain on *Caenorhabditis elegans*. J. Mol. Biol. 183: 543–551.
28. Engels, W.R., D.M. Johnson-Schlitz, W.B. Eggleston and J. Sved, 1990 High-frequency P element loss in Drosophila is homolog dependent. Cell 62: 515–525.
29. Plasterk, R.H.A. 1991 The origin of footprints of the Tc1 transposen of *Caenorhabditis elegans*. EMBO J. (in press).
30. Ruan, K.S. and S.W. Emmons 1984 Extrachromosomal copies of transposon Tc1 in the nematode *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA 81: 4018–4022.
31. Ehrlich, S.D. 1989 Illegitimate recombination in Bacteria. In Mobile DNA. D.E. Berg and M.M. Howe, eds. (Washington, D.C.: American Society of Microbiology), pp. 799–832.
32. Pulak, R.A. and P. Anderson 1988 Structure of spontaneous deletions in *Caenorhabditis elegans*. Mol. Cell. Biol. 8: 3748–3754.