The EMBO Journal vol.11 no.1 pp.287-290, 1992

Targeted alterations of the *Caenorhabditis elegans* genome by transgene instructed DNA double strand break repair following Tc1 excision

Ronald H.A.Plasterk[1] and José T.M.Groenen

Netherlands Cancer Institute, Division of Molecular Biology, Amsterdam, The Netherlands

[1]Corresponding author

Communicated by P.Borst

Excision of a Tc1 transposon of *Caenorhabditis elegans* is thought to leave a DNA double strand break. We report here that sequence polymorphisms in a transgenic DNA template are copied into the corresponding chromosomal gene upon excision of Tc1 from the chromosome. This shows that the double strand DNA break resulting from Tc1 excision is repaired with the extrachromosomal DNA as template and that sequences flanking the break can be replaced by sequences from the transgene. Transgene instructed break repair provides a method for the targeted introduction of precise alterations into the *Caenorhabditis elegans* genome.

Key words: *Caenorhabditis elegans*/reverse genetics/targeting/Tc1/transposition

Introduction

Detailed analysis of the *Caenorhabditis elegans* genome will aid in the understanding of the genetic control of development. The physical map of the nematode genome is approaching completion (Coulson *et al.*, 1986) and current efforts are aimed at the determination of the sequence of the genome (Roberts, 1990). However, strategies are lacking for the precise directed alteration of the sequence. Gene interruption by homologous recombination with transgenic DNA has not been reported until now. We recently found evidence that excision of the Tc1 transposon probably leaves a double strand break that is repaired from a homologous template and suggested that this could be exploited in a strategy for directed alteration of the nematode genome (Plasterk, 1991).

The evidence for double strand break (DSB) repair following Tc1 excision is the following: the frequency of loss of Tc1 elements (from the *unc-22* gene) depends on the presence of a wild type sequence on the homologous chromosome: heterozygous Tc1 mutations revert at high frequency, and by precise loss of the element, whereas homozygous Tc1 mutants revert 100× less frequently and always with a 'footprint' (i.e. in an imprecise fashion) (Plasterk, 1991; Mori *et al.*, 1990; Kiff *et al.*, 1988; Moerman and Waterston, 1988). Similar observations have been made for the P element of the fruit fly *Drosophila melanogaster* (Engels *et al.*, 1990) and have led to the suggestion that transposon excision was followed by repair of the remaining chromosome break in a homologue-dependent fashion. In homozygotes the transposon is usually copied back into the excision site and therefore loss of the transposon is not seen; only rare cases of interrupted repair lead to (imprecise) loss of the element.

By providing a marked transgenic template we wanted to investigate whether these marked sequences can be introduced into the chromosome. This experiment would serve two purposes: (i) to prove the DSB repair model for Tc1 and (ii) to demonstrate that specific alterations can thus be shuttled into the nematode genome.

Results

Generation of a transgenic strain containing a marked template for DSB repair

We generated a nematode that has a Tc1 insertion at a known position of a chromosomal gene (*unc-22*) and contains transgenic DNA that corresponds to the region of the Tc1 insertion but contains some (silent) polymorphisms (at −2, +26, +29, +167 and +191 bp from the position of the Tc1 in the chromosome, see Figure 1). The homology between the template and the chromosome extends in both directions 1500 bp from this position. Transgenic DNA in *Caenorhabditis elegans* is present in extrachromosomal tandem arrays; in our experiment the copy number of the transgenes is at least 50 per genome (Figure 2). The animal is mutant for Unc-22, since the transgene contains only a 3 kb region out of the 50 kb of the *unc-22* gene (Benian *et al.*, 1989).

Selection and analysis of animals with targeted alterations

The reversion frequency for Unc-22 was determined: 11 revertants were found among 450 500 progeny of a transgenic line (2×10^{-5}). This is similar to the reversion frequency previously reported for homozygous *unc-22 (st192::Tc1)* mutants and ~500× lower than the frequency in heterozygotes (Mori *et al.*, 1990; Plasterk, 1991). This shows that the transgene is not used as a template for repair synthesis as easily as the homologous chromosome, perhaps due to the absence of meiotic pairing between transgene and endogenous gene. Apparently this effect is not counterbalanced by the high copy number of the transgenic template. Revertants were allowed to propagate by self-fertilization and progeny were picked that had lost the transgene as witnessed by loss of the Rol phenotype. These were allowed to grow for 2−6 generations, and single animals were picked and allowed to propagate further. Their DNA was isolated and analysed by Southern blotting (Figure 2), showing that all transgenic DNA was absent from this line. The region surrounding the reversion site was analysed by direct sequencing of PCR products: as shown in Figure 3 two classes of revertants were found: 32% of the revertants contained 'footprints' as have been described before (Kiff *et al.*, 1988; Plasterk, 1991); these do not show the polymorphisms that are present in the transgene. Since

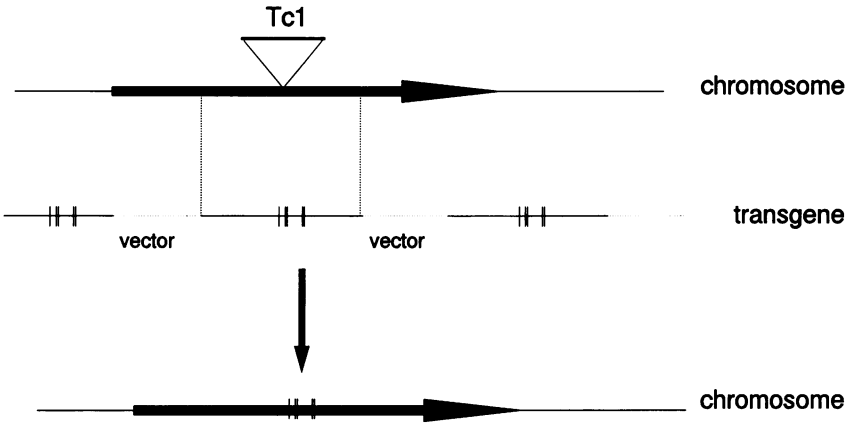

Fig. 1. Polymorphism marked transgenic DNA. Shown (not to scale) is the configuration of the *unc-22* gene: the complete gene is in the chromosome and is interrupted by Tc1, the transgene contains (apart from the *rol-6*^{D} marker, not shown here) several copies of a 3 kb fragment of the gene that is marked by some polymorphisms (shown by small tags). These polymorphisms are here numbered with respect to the Tc1 insertion site *st192* (see Moerman and Waterston, 1989, and the *unc-22* sequence kindly supplied by Dr Guy Benian, Emory University, Atlanta): −2 T to A, +26 A to T, +29 T to A, +167 A to T (which removes the *Sau*IIIA site GATC), +191 T to A (which creates an *Nde*I site from CTTATG).

the reversion frequency of the transgenic line was found to be in the same order as that of the non-transgenic isogenic strain, one would expect that a considerable number of revertants would be of the footprint type found in the non-transgenic strain; this is what we found. These revertants presumably result from incomplete repair that has used the sister chromatid or the homologous chromosome as template. The other 68% of revertants show precise loss of the transposon: 56% contain polymorphisms derived from the transgene. This demonstrates that the transgene has been used as a template for double strand gap repair that follows Tc1 excision and is definite proof for double strand break repair following Tc1 excision.

These 68% revertants without footprints fall into three classes. (i) Three revertants were precise, but had not picked up any of the polymorphism. These are here interpreted as resulting from transgene instructed repair with very short repair tracks, but we cannot rule out the possibility that they result from either direct ligation of excision products, or interrupted repair that uses the allelic gene as template. (ii) One revertant was precise and had picked up only the −2 polymorphism. (iii) 13 revertants were precise and had all the polymorphisms of the transgene. This means that 13 out of 17 (or, if the precise revertants are not included, 13 out of 14) revertants that used the transgene as template have repair tracks that extend >191 bp at one side of the break. In conclusion, 76% of the transgene-directed repair tracks extend at least 191 bp to one end of the excision site. Since the polymorphisms tested here were mostly on one side of the Tc1 insertion site, we do not know whether the track at the other side is similar, but we consider that likely. Recent quantitative analysis of repair track length in *Drosophila* showed an average length of ~1400 bp, and an 80% score for the presence of a polymorphism 200 bp from the break site (Gloor *et al.*, 1991). Our work does not allow a quantitative description of the length distribution of repair tracks, but the numbers we find could be in agreement with a similar distribution of the length of the repair track in *C.elegans*. In none of those 15 revertants were any other changes of *unc-22* sequence observed in the ~300 nucleotides analysed, suggesting that the repair replication is fairly accurate.

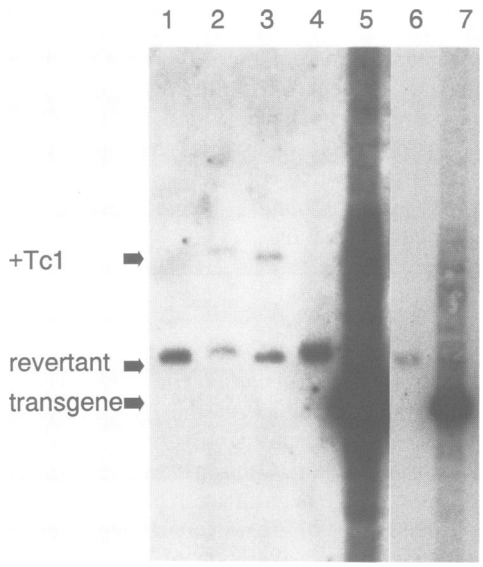

Fig. 2. Southern blot of DNA of transgenic animal and of revertants. Lanes 1−4, revertants; lane 5, transgenic line. Lanes 6 and 7 are shorter exposures of lanes 4 and 5. As shown in lanes 5 and 7 the transgene is at a higher copy number than the endogenous gene (e.g. lanes 4 and 6). The figure shows that the revertants have lost the Tc1 insertion at the position *st192* (e.g. lane 3 shows a heterozygous animal where one allele is reverted and the other is mutant). It can also be concluded from this blot that after segregation of the Rol phenotype no transgenic DNA is present in the DNA from revertants. DNA was digested with *Hin*dIII and the blot was probed with a PCR fragment that covers the Tc1 insertion site (primers AB1594 and AB1595, on genomic *C.elegans* DNA). The expected size of the transgenic fragment is 1820 bp, the genomic sequence after reversion is 2280 bp, and the corresponding fragment before reversion is 2280+1612=3892 bp. Lanes 2 and 3 show animals that are heterozygous for the reverted *unc-22* allele, and therefore they show the larger fragment as well. The estimated copy number of the transgene is >50 (also based on shorter exposures, not shown here); this has to be a lowest estimate, since any population of transgenic animals contains WT segregants that have lost the transgene.

Discussion

Tc1 excision and double strand break repair

We here provide conclusive evidence for the recently proposed model for DSB repair following Tc1 excision.

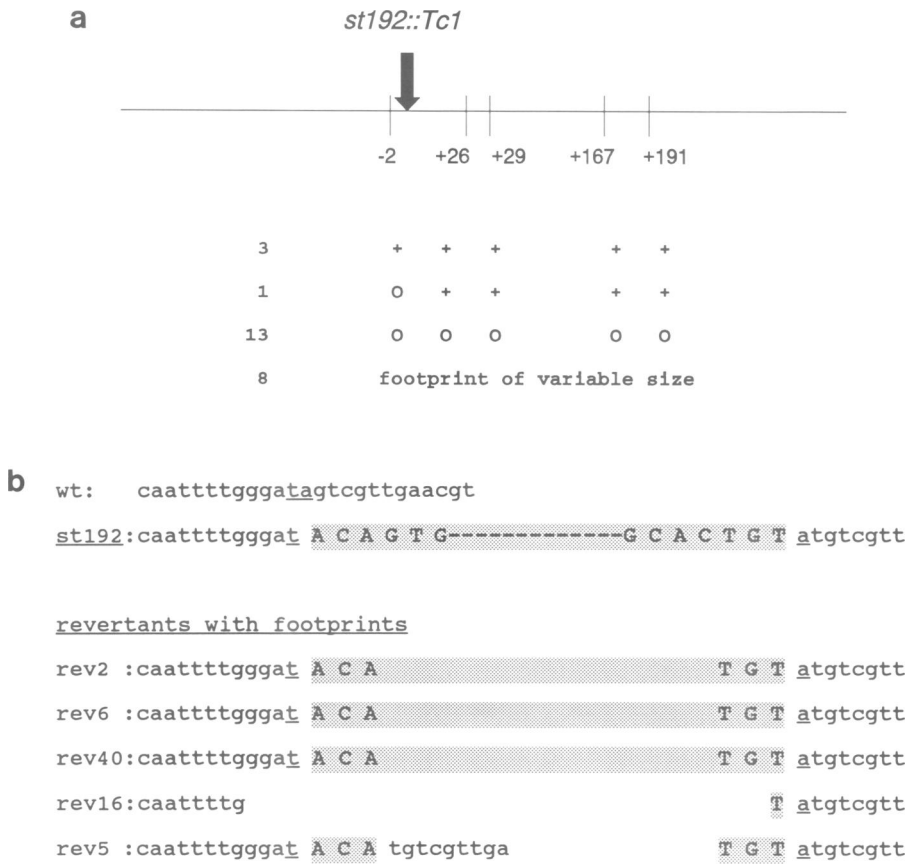

Fig. 3. Sequence of revertants of a transgenic *unc-22* (*st192::Tc1*) mutant. **a**. A + indicates the wild type sequence, an O indicates a polymorphism (also see Figure 1). No other sequence alterations were observed in these revertants. The sequence around the site of *st192::Tc1* was determined for 25 revertants: eight showed a 'footprint', i.e. imprecise restoration of the sequence. Five footprints are shown in **b**: three of these have a 6 bp insertion (3 bp from each end of the Tc1 element), one of them has a small deletion and one has a repetition of nine nucleotides flanking the insertion site. These types of footprints have been described before (Moerman and Waterston, 1989; Plasterk 1991); none of these revertants had picked up any of the polymorphisms of the transgene.

Probably the first step after Tc1 excision is degradation of the ends of the broken DNA (at least a few hundred base pairs), followed by a search for a homologous template and extension of the 3′ ends of the break (compare Szostak *et al.*, 1983; Schiestl *et al.*, 1988; Rudin *et al.*, 1989). The role of Tc1 in this process is probably only the generation of a double strand break; the subsequent repair process is part of the cellular strategy for dealing with broken chromosomes in a template directed manner to avoid errors. Usually the repair template will be the sister chromatid or the homologous chromosome, containing a transposon at the position of the excision site, and in that case the only way such repair will lead to loss of the element is when it is incomplete. Such incomplete repair will result in a transposon footprint, and therefore the length and sequence of transposon footprints is not a result of transposase action, but of properties of the repair system, and also of the constraints of the selection for revertants. If we had chosen for our analysis a Tc1 allele at a position in a gene where the presence of footprints was incompatible with restoration of gene function, we would have expected all revertants to be derived from transgene instructed repair. The allele studied here seems to be very tolerant for footprints (Kiff *et al.*, 1988); as a consequence the fraction of revertants that have not used the transgene as template (32%) is probably higher in this case than it will be for most other Tc1 alleles.

In the fruit fly *Drosophila*, breaks resulting from the excision of the P transposon are repaired by double strand break repair (Engels *et al.*, 1990) and internally deleted versions of the element are attributed to interrupted DSB repair. It may be useful to check whether the transposon footprints in plants, which are currently thought to result from imprecise excision of the element (Schwartz-Sommer *et al.*, 1985; Haring *et al.* 1991), also result from interrupted gap repair. A recent discovery of significant homology between transposons in flies and plants (Calvi *et al.*, 1991) makes it all the more likely that similar mechanisms account for the footprints left by transposons in plants and animals.

Targeted alteration of the Caenorhabditis elegans genome

The introduction of specific alterations into the *C.elegans* genome by Tc1 excision induced gap repair could be a powerful technique for the analysis of the *C.elegans* genome. A similar approach has recently been found to work for specific alteration of the *Drosophila* genome, where a single ectopically integrated gene can be used as repair template (Gloor, *et al.* 1991). In *C.elegans*, gene targeting by homologous recombination has not yet been described, and transposon induced DSB repair is thus far the only strategy for the precise targeted alteration of the genome. We have shown sequence replacement to happen at a frequency of

$\sim 10^{-5}$, and strains have been described in which the frequency of Tc1 excision, and thus probably of repair, is at least 10× higher (Collins *et al.*, 1987). This is sufficient for most purposes. We have shown that a mutation 200 bp away from the position of the Tc1 insertion is copied into the chromosome in 76% of the cases where the transgene was used as template, which suggests that the repair track can be quite a bit longer than 200 bp, so that it will probably be possible to introduce mutations that are further removed. Many *C.elegans* genes have been cloned by Tc1 tagging, and therefore Tc1 alleles of these genes are available. New Tc1 insertions may be obtained by a PCR based detection strategy (Ballinger *et al.*, 1990; Kaiser and Goodwin, 1990; O'Hare, 1990), that has recently been effectively applied to Tc1 (A.Rushforth and P.Anderson, personal communication). Once a Tc1 insertion mutant has been obtained, the strategy is as follows (also see Figure 1): the animal is made transgenic for an altered version of the gene, and Tc1 excision provides frequent DNA breaks at a defined chromosomal position, resulting in transgene instructed repair. Detection of this can be by a PCR/sib-detection protocol (similar to protocols that have successfully been used to detect Tc1 insertions), or in some cases by genetic selection. In this paper we describe the introduction of single base polymorphisms; in principle, double strand break repair is expected to copy any sequence that is flanked by a sufficient stretch of homology to the ends of the broken DNA, thus it should also be possible to use Tc1 initiated break repair to create gene fusions, insertions and replacements *in situ*.

Materials and methods

Generation of marked repair template

To obtain a suitable template for gap repair we amplified the region surrounding Tc1 insertion *st192* in *unc-22* by PCR: three adjacent regions were first amplified separately and then fused in two steps by PCR. The primers contained alterations in a few positions that correspond to third base positions in the *unc-22* reading frame and care was taken that in all cases these alterations left the coding capacity intact and did not result in triplets that are rare in *C.elegans*. All are T to A or A to T transversions. Primers were as follows:

AB1939, 5′-TAATCTCGAGTCGATTTCATTGGAGCTCCC;
AB1938, 5′- TTTTCTCGAGTTTCAACGGCTTCTGGTTCT;
AB1937, 5′-CTGAGCCAGTCACCTGCAGACCATATGTCG;
AB1936, 5′-GTCTGCAGGTGACTGGCTCAGAACCGTTAC;
AB1935, 5′-CGTTGAACGTTTTGAGAAGAGAGGAGGCG;
AB1934, 5′-CTCTTCTCAAAACGTTCAACGACATAACCC.

Genomic *C.elegans* DNA was used as template, and the initial three PCRs were with primers AB1939 + AB1934, AB1938 + AB1937 and AB1936 + AB1935. The products of the latter two reactions were then gel purified and mixed and a PCR was done: in the first three cycles the annealing time was 3′ and the temperature 37°C; no primers were added. This allowed priming of the two fragments on each other (primers AB1936 and AB1937 are largely complementary); subsequently the outer primers (AB1935 and AB1938) were added, and PCR was continued for 25 cycles with an annealing temperature of 55°C. The product was again gel purified and fused to the third fragment in a similar PCR, now using AB1939 and AB1938 as outer primers. The final product was gel purified, digested with *Xho*I (sites for this enzyme had been included in the primers AB1938 and AB1939) and cloned into the *Sal*I site of the Bluescript M13(+) plasmid. Sequence analysis of the resulting plasmid (pRP411) revealed one extra base pair substitution, in the area of overlap of the two primers AB1936 and AB1934 (A to T transversion 26 nucleotides to the right of the insertion site *st192*). Subsequent experiments showed that this alteration has no effect on Unc-22 phenotype, and the polymorphism was included in the further analysis (polymorphism '+26').

Genetic experiments and DNA sequence analysis

Strain NL203 [*vab-9(e1744) II, mut-6(st702) unc-22(st192::Tc1)IV*] was made transgenic using a 1:3 mixture of plasmids PRF4 [containing *rol-6*^{D} (Kramer *et al.*, 1990)] and pRP411 (described above). Rol progeny was picked and cloned, and two lines were selected that had a high fraction transgenic progeny ($\sim 50\%$). These were grown and Unc-22^{+} revertants were picked. These were often Rol; in these cases WT segregants were picked from the next generation, from the progeny of these animals individual animals were brought in culture, and after a few generations a clonal culture was again started. DNA was isolated (Sulston and Hodgkin, 1988). The first series of DNA preparations was checked by Southern blot analysis to confirm the absence of transgenic DNA (see Figure 2). To determine the sequence around the reversion site, a PCR was done using primers AB1594 (5′-AGAGAATGCAGTCGGACTTTC) and AB1595 (5′-CTTGCAAGTTGGCTTGGATGGTTCACCTTC). The product was analysed for the polymorphisms as follows: asymmetric PCR was done with excess primer AB1595 and the product was sequenced using primer AB1594. Apart from the polymorphisms no other changes were ever observed, showing the accuracy of the break repair synthesis; in most cases the sequencing gel could be read until the area of polymorphisms +167 and +191. In a few cases we took advantage of the fact that polymorphism +167 removes a *Sau*IIIA site and +191 creates an *Nde*I site, and did restriction analysis of double stranded PCR products to investigate their presence.

Acknowledgements

We thank Hanneke van der Gulden and Ellen Wientjens for the synthesis of oligonucleotides, and Drs W.R.Engels, A.Rushforth and P.Anderson for unpublished information. Dr Guy Benian is thanked for making available the sequence of *unc-22*. Drs Piet Borst and Anton Berns are thanked for critical reading of the manuscript.

References

Ballinger,D.G. and Benzer,S. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 1986–1690.
Benian,G.M., Kiff,J.E., Neckelmann,N., Moerman,D.G. and Waterston,R.H. (1989) *Nature*, **342**, 45–50.
Calvi,B.R., Hong,T.J., Findley,S.D. and Gelbart,W.M. (1991) *Cell*, **66**, 465–471.
Collins,J., Saari,B. and Anderson,P. (1987) *Nature*, **328**, 726–728.
Coulson,A., Sulston,J., Brenner,S. and Karn,J. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 7821–7825.
Engels,W.R., Johnson-Schlitz,D.M., Eggelston,W.B. and Sved,J. (1990) *Cell*, **62**, 515–525.
Gloor, G.B., Nassif, N.A., Johnson-Schlitz, D.M., Preston, C.R. and Engels, W.R. (1991) *Science*, **253**, 1110–1117.
Haring,M.A., Teeuwen-de Vroomen,M.J., Nijkamp,J.J. and Hille,J. (1991) *Plant Mol. Biol.*, **16**, 39–48.
Kaiser,K. and Goodwin,S.G. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 1686–1690.
Kiff,J.E., Moerman,D.G., Schriefer,L.A. and Waterston,R.H. (1988) *Nature*, **331**, 631–633.
Kramer,J.M. French,R.P. Park,E.C. and Johnson,J.J. (1990) *Mol. Cell. Biol.*, **10**, 2081–2089.
Moerman,D.G. and Waterston,R.H. (1989) In Berg,D.E. and Howe,M.M. (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 537–556.
Mori,I., Moerman,D.G. and Waterston,R.H. (1990) *Mol. Gen. Genet.*, **220**, 251–255.
O'Hare,K. (1990) *Trends Genet.*, **6**, 202–203.
Plasterk,R.H.A. (1991) *EMBO J.*, **10**, 1919–1925.
Roberts,L. (1990) *Science*, **248**, 1310–1313.
Rudin,N. Sugarman,E. and Haber,J.E. (1989) *Genetics*, **122**, 519–534.
Schiestl,R.H., Igarashi,S. and Hastings,P.J. (1988) *Genetics*, **119**, 237–247.
Schwartz-Sommer,Z., Gierl,A., Cuypers,P.A. and Saedler,H. (1985) *EMBO J.*, **4**, 591–597.
Szostak,J.W., Orr-Weaver,T.L., Rothstein,R.J. and Stahl,F.W. (1983) *Cell*, **62**, 25–35.
Sulston,J. and Hodgkin,J. (1988) In Wood,W.B. (ed.), *The Nematode Caenorhabditis elegans*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 587–606.

Received on September 6, 1991; revised on October 10, 1991