Characterization of the *Caenorhabditis elegans* Tc1 transposase in vivo and in vitro

Jan C. Vos, Henri G.A.M. van Luenen, and Ronald H.A. Plasterk

The Netherlands Cancer Institute, Division of Molecular Biology, 1066 CX Amsterdam, The Netherlands

We have investigated the function of the *Tc1A* gene of the mobile element Tc1 of *Caenorhabditis elegans*. Tc1 is a member of a family of transposons found in several animal phyla, such as nematodes, insects, and vertebrates. Two lines of evidence show that *Tc1A* encodes the transposase of Tc1. First, forced expression of the Tc1A protein in transgenic nematodes results in an enhanced level of transposition of endogenous Tc1 elements. Second, DNase I footprinting and gel retardation assays show that Tc1A binds specifically to the inverted repeats at the ends of the element and that the Tc1A recognition site is located between base pairs 5 and 26 from the ends of Tc1. Functional dissection of the transposase shows the presence of two distinct DNA-binding domains. A site-specific DNA-binding domain is contained within the amino-terminal 63 residues of Tc1A; this region shows sequence similarity to the prokaryotic IS30 transposase. A second, general DNA-binding domain is located between amino acids 71 and 207. Our results suggest that Tc1 is more similar to prokaryotic insertion elements than to eukaryotic transposons such as P elements in *Drosophila* or Ac and En-I in plants.

[Key Words: C. elegans; transposase; Tc1; DNA binding]

Received February 11, 1993; revised version accepted April 21, 1993.

The Tc1 transposable element of *Caenorhabditis elegans* is the prototype of a family of transposons that are not only found in nematodes but also in arthropods and vertebrates (Harris et al. 1988; Henikoff and Plasterk 1988; Brezinsky et al. 1990; Heierhorst et al. 1992; Henikoff 1992; Prasad et al. 1991). The Tc1 transposon is 1611 bp long, has 54-bp terminal inverted repeats, and contains one major open reading frame (ORF) [Rosenzweig et al. 1983a]. As most members of the family, Tc1 is always flanked by a TA target site [Rosenzweig et al. 1983b; Mori et al. 1988a]. Although the terminal 4 nucleotides (CATG) are conserved among Tc1-like elements, the length and sequence of the inverted repeats are highly diverse [Henikoff 1992].

The Tc1 element is present in all *C. elegans* strains tested, but the number of copies (between 30 and 300) and the frequency of transposition are variable. Tc1 excision is detectable in somatic cells of the common laboratory strains N2 and Bergerac (Emmons and Yesner 1984). In contrast, germ-line Tc1 transposition is restricted to Bergerac animals. Genetic loci, called mutators, are essential for this germ-line activity (Collins et al. 1987; Mori et al. 1988b). The mobile nature of certain mutators themselves, together with the cosegregation of Tc1 elements with mutator activity, indicates that specific Tc1 elements may be the cause of germ-line transposition [Mori et al. 1988b; Mori 1989; R.H.A. Plasterk, unpubl.]. This could be the consequence of either differences in primary sequence or, more likely, the flanking sequences that influence the tissue-specific expression of the gene contained in Tc1.

A general characteristic of autonomous transposable elements is the presence of one or more genes encoding proteins involved in the transposition reaction. For instance, P elements in *Drosophila* contain a gene that can encode either a transposase or a repressor of transposition. Both proteins share the same DNA-binding domain, which mediates binding to a 10-bp internal sequence of P elements adjacent to the 31-bp terminal inverted repeats [Kaufman et al. 1989; Misra and Rio 1990]. Interestingly, the transposase-binding site is located 16 bp from the 5' inverted repeat and 4 bp from the 3' inverted repeat. The transposases encoded by the plant transposons En-1 and Ac also do not bind to the inverted repeat sequences but, rather, recognize subterminal repeat motifs that are asymmetrically located within the left and right arm of the transposon [Gierl et al. 1988; Kunze and Starlinger 1989]. In contrast, the transposases encoded by prokaryotic elements, such as IS30 and IS903, recognize sequences within the inverted repeat of their transposable element [Stalder et al. 1990; Derbyshire and Grindley 1992]. Mu contains a transposase-binding site at each end of the transposon, like IS30 and IS903, as well as additional binding sites that are asymmetrically spaced at the left and right arm of the transposon [Craigie et al. 1984].

The first complete sequence of a Tc1 element revealed the presence of an ORF capable of encoding 273 residues.
C. elegans Tcl transposase

(Rosenzweig et al. 1983a). The hypothesis that the encoded polypeptide is a transposase was supported by the observation that the 273-amino-acid protein, named TcA, when produced in Escherichia coli, showed a strong affinity for DNA. No sequence specificity, however, was detected in a South-Western analysis (Schukkink and Plasterk 1990). Sequence analysis of more Tcl elements and the sequence of a Tcl homolog in Caenorhabditis briggsae, indicated that the gene could actually contain an additional exon, 5' of the large exon identified previously (Schukkink and Plasterk 1990; Prasad et al. 1991). However, analysis of Tcl mRNA did not result in the identification of a spliced transcript; the analysis was hindered by readthrough transcripts resulting from the many copies of Tcl in the genome (R.H.A. Plasterk, unpubl.).

Here we report the identification of the intron–exon structure of the TclA gene by forced expression of this gene in transgenic animals and analysis of the induced mRNA and protein. Furthermore, we studied the effect of TclA expression on Tcl transposition and identified two distinct DNA-binding domains within TclA.

Results

Induced expression of the Tcl transposase

Previous efforts to detect either mRNAs or proteins encoded by the Tcl element of C. elegans failed to demonstrate directly the presence of the putative transposase (R.H.A. Plasterk, unpubl.). Therefore, we forced the expression of the genetic information of Tcl in transgenic nematodes. A heat shock promoter plus untranslated leader sequence (Stringham et al. 1992) was fused to the start codon of the first exon of the Tcl gene and cloned into the vector pPD26.50 (Fire et al. 1990), which contains an unc-54 polyadenylation signal (see Material and methods).

A transgenic N2 line, named NL224, containing the hsp–Tc1A fusion construct was established (see Materials and methods) and characterized. Northern analysis showed the appearance of an abundant RNA of ~1.7 kb in response to a heat shock, which is consistent with proper initiation at the heat shock promoter and the use of the unc-54 polyadenylation signal [Fig. 1A, cf. lanes 1 and 2]. The signal seen in both lanes near the origin of the gel resulted from residual genomic DNA. A polyclonal antiserum raised against the carboxyl terminus of TcA (Schukkink and Plasterk 1990) was used to analyze the induction of the putative transposase [Fig. 1B]. A polypeptide of ~40 kD, which we named TclA, was detectable after a heat shock of 2 hr at 33°C in transgenic NL224 animals [Fig. 1B, lanes 7–10]. Fractionation of protein extracts demonstrated that TclA is a nuclear protein (data not shown). The protein was not detected without a heat shock [lane 6], or in heat-shocked non-transgenic N2 animals [lanes 1–5].

The appearance of a polypeptide of 40 kD indicated that the coding information for TclA is contained within two exons. To determine the intron–exon borders precisely, an analysis of the mRNA induced with heat shock was performed by generating a cDNA that was subsequently amplified by PCR. Sequence analysis of the PCR product revealed that an intron of 41 nucleotides is removed from the pre-mRNA [Fig. 1C]. This demonstrates that the 3' splice site was predicted correctly but that the actual intron size is 41 nucleotides instead of the predicted 65 nucleotides (Schukkink and Plasterk 1990). Taken together, these results show that TclA encodes a protein of 343 amino acids, which is synthesized in the transgenic NL224 line upon induction.

Increase of Tcl transposition on expression of TclA

The effect of induction of TclA on the frequency of transposition of resident Tcl elements was tested by

Figure 1. Inducible expression of TclA in transgenic line NL224. (A) Two micrograms of total RNA from line NL224 extracted before (C) or after (HS) a 3-hr heat shock was analyzed on a Northern blot using Tcl DNA as probe. The position of the induced TclA mRNA is indicated with an arrow, as well as the positions of 18S and 28S RNAs. (B) Western blot analysis of N2 (lanes 1–5) and NL224 (lanes 6–10) using a TclA-specific polyclonal antiserum. From t = 0 hr, animals were heat shocked for 2 hr and subsequently returned to 18°C. Protein extracts were prepared at the times indicated (hr) and separated on a 10% SDS–polyacrylamide gel. The position of TclA is indicated with an arrow. Lane M contains prestained molecular mass markers of 106, 80, 49.5, 32.5, 27.5, and 18.5 kD. Lane 12 contains TclA produced in E. coli. (C) Sequence analysis of TclA cDNA. A cDNA was made of heat shock-induced TclA mRNA, amplified by PCR, and subsequently sequenced. The exon sequence is given in uppercase letters, and the intron sequence is shown in lowercase letters.
monitoring Tcl insertions into the gpa-2 gene (Fino Silva and Plasterk 1990) using the specificity and sensitivity of PCR. This method allowed us to detect somatic transposition events. We did not detect germ-line transposition, which was not unexpected, because the hsp-16 promoter is only active in somatic cells (Stringham et al. 1992). Two rounds of PCR were performed with gpa-2- and Tcl-specific primers. A PCR product was expected only when Tcl had integrated in the proper orientation in the vicinity of the gpa-2 gene. The frequency of Tcl transposition was measured as the number of PCR products per mass of DNA. Genomic DNA was prepared from both heat-shocked and non-heat-shocked N2 and transgenic NL224 worms. In 20 reactions, each containing 30 ng of DNA derived from the NL224 animals in which TclA was induced, we found 20 Tcl insertions in gpa-2 (Fig. 2B). The size distribution of the PCR products indicates that different, independent transposition events are detected. In contrast, 20 reactions with 100 ng of DNA derived from uninduced NL224 animals resulted in only three insertions (Fig. 2A). To exclude the possibility that the increase in transposition is the result of the heat shock itself, we performed a similar experiment with DNA prepared from N2 animals. Three insertion events were obtained in 20 PCRs with 100 ng of DNA each prepared from non-heat-shocked N2 animals (Fig. 2C), whereas the same amount of DNA from the heat-shocked N2 animals yielded five insertions (Fig. 2D). This demonstrates that a moderate level of somatic transposition occurs in wild-type N2 animals, a process not influenced significantly by a heat shock treatment. Therefore, the ~18-fold increase in transposition in heat-shocked NL224 animals is attributable to expression of TclA. The 18-fold increase is an underestimation of the effect of TclA expression, because the transgene is neither stably maintained within the nematode population nor within all cells of a transgenic animal, and because the heat shock promoter is not expressed in all somatic cells of C. elegans (Stringham et al. 1992). It is important to note that the expression of TclA did not result in detectable levels of transposition of a Tcl-related element, Tcl3 (Collins et al. 1989), indicating that the activity of TclA is transposon specific (data not shown).

Several Tcl insertions in gpa-2 were sequenced and found to be genuine transposition events as judged by the presence of the complete inverted repeat and the integration into a TA target sequence (data not shown). From the increase in Tcl insertions after induction of TclA, we conclude that TclA is a limiting factor in the transposition reaction in vivo.

Site-specific binding of TclA to the inverted repeat of Tcl

One of the activities expected of Tcl transposase is binding to a DNA sequence within the Tcl element. We prepared cytoplasmic and nuclear extracts of the transgenic NL224 line and used those extracts to probe for sequence-specific binding to Tcl fragments. The result of a gel-retardation assay is shown in Figure 3, in which the labeled DNA fragment contains the 26 terminal nucleotides of Tcl as well as the flanking TA target dinucleotide. A band shift was observed using the nuclear extract prepared from NL224 animals with induced TclA expression (lane 4) and not with the nuclear extract from uninduced animals (lane 2). Cytoplasmic extracts (lanes 1,3) yielded no specific complex. The band shift obtained was the result of sequence-specific binding, because it was competed with an excess of unlabeled Tcl oligonucleotide (lane 5) but not with an unrelated oligonucleotide (lane 6). These results suggest strongly that TclA has the ability to bind to the inverted repeat of Tcl but do not exclude formally the possibility that

**Figure 2.** Effect of TclA expression on the frequency of Tcl insertions in gpa-2. DNA was prepared from N2 (C,D) and NL224 (A,B) animals that had either received no heat shock (A,C) or a 2-hr heat shock followed by 18 hr of recovery at 18°C (B,D). Tcl insertions in gpa-2 were detected by PCR using Tcl- and gpa-2-specific primers. The products of 20 reactions using 30 ng of DNA (B) or 100 ng of DNA (A,C,D) were separated on a 1% agarose gel and stained with ethidium bromide. A 1-kb DNA ladder (BRL) was run in parallel as marker.
C. elegans Tc1 transposase

The polypeptide was not recognized by the antiserum against the carboxyl terminus of Tc1A [data not shown]. Therefore, the complex resulted from binding of a Tc1A derivative, which was possibly the consequence of proteolysis of full-length Tc1A or of a premature translational stop. Binding of the full-length Tc1A protein resulted in the retarded complex near the origin of the gel. This complex is far less abundant, although full-length Tc1A is expressed at a much higher level than the Tc1A derivative. Thus, full-length Tc1A appears to have a low specific DNA-binding activity.

Mapping of the site-specific DNA-binding domain of Tc1A

To map the domain of Tc1A that is responsible for binding to the inverted repeat of Tc1, we made carboxy-terminal deletion mutants and expressed them in E. coli (see Materials and methods). The expression of all of the deletion mutants was verified by SDS-PAGE [data not shown] and site-specific DNA binding was tested in a

Figure 3. Tc1 ends specific DNA-binding activity in C. elegans extracts. A 36-bp probe containing the 26 terminal bp of Tc1 flanked by the TA target sequence was mixed under standard conditions with 2 μl of either cytoplasmic (C) or nuclear (N) extract prepared from NL224 animals that had received either no heat shock (lanes 1, 2) or a 2.5-hr heat shock (lanes 3-6). In lane 7, no extract was added to the reaction. Ten picomoles of specific (lane 5) or nonspecific (lane 6) 28-mers were included in the binding assay. Products were separated on a native 5% polyacrylamide gel.

Tc1A expression induces or activates another protein that binds Tc1 ends.

To study the DNA-binding potential of Tc1A in more detail, we expressed the protein under the control of the T7 promoter in E. coli. The intron sequence was removed from the Tc1A gene (see Materials and methods), and the ORF was cloned into the vector pET3c (Rosenberg et al. 1987). Upon induction of T7 RNA polymerase, a prominent protein was synthesized, which comigrated with the 40-kD Tc1A protein produced in transgenic NL224 animals as shown in the Western blot analysis of Figure 1B (lane 12).

Bacterial lysates were made and analyzed in a bandshift assay. Two retarded complexes specific for Tc1 ends were obtained in the extract with the Tc1A expression vector [Fig. 4, lane 3] and were absent from the pET3c extract (lane 2). Surprisingly, neither of these complexes has the same mobility as the complex found with the nuclear NL224 extract (lane 1). Upon purification of the polypeptide responsible for the major shift, we obtained a fraction that did not contain the 40-kD Tc1A protein but which did contain a polypeptide of ~18 kD with reactivity against an antiserum that recognizes an epitope between amino acids 71 and 153 of Tc1A. The polypeptide was not recognized by the antiserum against the carboxyl terminus of Tc1A [data not shown]. Therefore, the complex resulted from binding of a Tc1A derivative, which was possibly the consequence of proteolysis of full-length Tc1A or of a premature translational stop. Binding of the full-length Tc1A protein resulted in the retarded complex near the origin of the gel. This complex is far less abundant, although full-length Tc1A is expressed at a much higher level than the Tc1A derivative. Thus, full-length Tc1A appears to have a low specific DNA-binding activity.

Figure 4. Specific DNA binding by recombinant Tc1A and carboxy-terminal deletion mutants. E. coli strain BL21 pLysS was transformed with pET3c-derived vectors for the expression of Tc1A as well as carboxy-terminal deletions thereof. Bacterial lysates were prepared from induced cultures and mixed under standard conditions with a 36-bp probe containing the ends of Tc1. The number of amino-terminal triplets of the Tc1A gene preserved in the expression vector are indicated at the bottom. Lane 2 represents the bacterial lysate obtained with the empty vector pET3c. In lane 1, 2 μl of a nuclear extract prepared from heat-shock treated NL224 animals was used. Reaction products were separated on a native 5% polyacrylamide gel.
gel-retardation assay with the terminal sequence of Tcl as probe (Fig. 4). Because the major shift observed with the TclA expression plasmid is presumably caused by a smaller polypeptide, it is possible that deletions of carboxy-terminal sequences do not abolish the appearance of the major retarded complex. Extracts containing deletion mutants N207 (lane 4) and N153 (lane 5) showed the same prominent complex as the extracts with full-length TclA (lane 3). Thus, the genetic information required for the generation of the major retardation complex (lanes 3–5) is contained within the first 153 triplets of the TclA gene. Complexes corresponding to either the full-length N207 or N153 mutant proteins were not detected. Mutant N78 (lane 6) resulted in a retarded complex with a higher mobility than the one generated by the TclA derivative, suggesting that the deletion was beyond the position of the premature translational stop or the proteolytic cleavage. Mutant N63 (lane 7) gave rise to a complex with the highest mobility. The deletion mutant N54 appeared to be unable to bind to the terminal Tcl sequence (lane 8). Apparently, the amino-terminal 63 amino acids of TclA are necessary and sufficient for sequence-specific binding to the inverted repeat of Tcl.

**DNase I footprinting analysis**

The site-specific binding of the TclA derivative was analyzed further in a DNase I footprint assay (Fig. 5A). The top strand was protected against DNase I attack by the TclA derivative from position 3 to 29 with respect to the end of the element. A slightly shifted protected region was obtained on the bottom strand—between positions 6 and 31. A schematic representation of the DNase I-protected region within the inverted repeat of Tcl is given in Figure 5B. The footprint data are consistent with the binding specificity seen in the gel-retardation assays.

The DNase I footprint results suggest that the conserved terminal TACAGT sequence at the end of the transposon is not part of the transposase recognition sequence. To investigate this, we tested DNA binding of the TclA derivative to transposon ends with a mutated TACAGT sequence in a gel-retardation assay (Fig. 6). The TclA derivative bound with equal affinity to two different mutated sequences (lanes 3,4) and to the fragment with the authentic Tcl sequence (lane 1), whereas an unrelated sequence was not recognized (lane 2). Therefore, the terminal TACAGT sequence is not essential for recognition of the ends of Tcl by the TclA derivative.

**Mapping of a nonspecific, second DNA-binding domain**

A general DNA-binding domain has been defined for the TclA amino acids 71–235 (Schukkink and Plasterk 1990). This domain does not coincide with the minimal DNA-binding domain required for site-specific binding to the Tcl inverted repeat defined above. We investigated the nonspecific DNA-binding domain further by analyzing the carboxy-terminal deletions in a South-
Figure 6. Gel retardation analysis of the effect of the outer base pairs on binding of Tcl transposase. Probes of 79 bp were made that contained either no similarity to Tcl ends (lane 2) or contained the Tcl inverted repeat sequences between 5 and 26, flanked by the conserved TACAGT sequence (lane 1) or the mutated sequences ATCCCC (lane 3) or AGAGTC (lane 4). Probes were mixed with 0.05 μl of the partially purified TclA derivative, and reaction products were analyzed on a 4% polyacrylamide gel.

Figure 7. South-Western analysis of the general DNA-binding capacity of TclA and carboxy-terminal deletion mutants thereof. Total bacterial lysates (10 μl) from induced BL21 pLysS carrying the expression vectors for TclA and various deletion derivatives were separated on a 12.5% SDS-polyacrylamide gel and transferred to nitrocellulose. The blot was probed with radiolabeled nonspecific DNA for 2 hr. The lanes are marked at the top with the number of amino-terminal codons of the TclA gene present in the expression vector. The positions and molecular masses of the marker proteins are shown at right.

Discussion

To test the hypothesis that the TclA gene of the transposable element Tcl of C. elegans encodes a transposase, we generated a transgenic line in which the TclA gene is under the control of a heat shock promoter. Analysis of the mRNA revealed the intron–exon structure of the TclA gene; the actual size of the intron is smaller than initially predicted from DNA sequence data (Schukkink and Plasterk 1990). Overexpression of the 40-kD TclA protein in C. elegans resulted in an enhanced level of transposition of resident Tcl elements. This demonstrates that TclA, like all autonomous mobile elements, contains a transposase gene and that its product is a limiting factor for transposition. The transposase is specific for Tcl transposition, as overexpression did not result in transposition of Tcl3, a Tcl-like transposable element in C. elegans. The transgenic NL224 line in which TclA expression is under the control of a somatic promoter can be characterized as an artificial somatic mutator. Likewise, it should be possible to establish an artificial germ-line mutator by fusion of the TclA gene to germ line-specific regulatory sequences. The natural germ-line mut-5 mutator activity is thought to be the result of enhanced expression of the TclA gene of a particular Tcl element that has inserted in the neighborhood of a germ-line enhancer (Mori et al. 1988b).

We show site-specific binding of TclA to the inverted repeat of the transposable element. The primary recognition sequence is within base pairs 5 and 26 of the inverted repeat; substitution of the conserved terminal sequence TACAGT has no effect on binding of TclA. It has been suggested for prokaryotic transposons, such as Tn10 and IS903, that the outer base pairs are not important for initial transposase binding, but are involved in a later step in the transposition reaction (Huisman et al. 1989; Derbyshire and Grindley 1992). Currently, we are

C. elegans Tcl transposase

protein has been ignored because the Tcl1A gene was considered to consist of only one exon. Once we mapped the site-specific DNA-binding domain to the amino terminus of TclA, we repeated the homology search using the BLAST program (Altschul et al. 1990) for the first 71 amino acids of TclA. The search revealed one striking homology (maximal segment pair score of 51) with IS30 transposase (Fig. 8) within an amino-terminal fragment of the transposase that is sufficient for site-specific binding to the terminal repeats of IS30 (Stalder et al. 1990). It requires further experimentation to verify the significance of the observed similarity.

Similarity of the amino terminus of Tcl1A with IS30 transposase

In previous computer searches for homology between TclA and other proteins, the amino-terminal part of the between amino acid residues 71 and 207, whereas the site-specific DNA-binding domain is within the first 63 residues.

Figure 8. Homology between the DNA-binding domain of Tcl1A and IS30 transposase. The similarity (+) and identity (uppercase letters) are indicated between amino acid residues 9 and 48 of Tcl1A (top) and residues 58 and 97 of IS30 transposase (bottom).
investigating the consequence of mutations within the terminal nucleotides of Tcl on transposition. The observation that TclA binds to the inverted repeat is more reminiscent of prokaryotic transposases than of eukaryotic transposases such as P transposase in Drosophila [Kaufman et al. 1989] or TnPA of En-1 (Gierl et al. 1988) and ORFAs of Ac [Kunze and Starlinger 1989] in Zea mays. These transposases do not bind to inverted repeat sequences but to sub-terminal control regions that are asymmetrically spaced with respect to the transposon ends. The similarity of the DNA-binding domain of TclA with IS30 transposase, which binds to the inverted repeat of IS30, is also suggestive of a closer relationship between the Tcl family of transposons and prokaryotic insertion elements than between Tcl and the above-mentioned transposons characterized in plants and in Drosophila. Furthermore, sequence similarities have been found between the central part of TclA (triplets 103–285) and prokaryotic insertion elements [Henikoff 1992].

The proposed mechanism of Tcl transposition is the excision of the element by double-strand cleavage and reintegration of the linear intermediate into a novel target site [Plasterk 1991]. TclA is the most likely candidate for the execution of the double-strand break, because it binds to the Tcl ends close to the cleavage sites. However, DNase I footprinting only shows protection by a TclA derivative of the DNA region from base pairs 3 to 31. It is possible that full-length TclA makes contacts with the terminal nucleotides, but, alternatively, contacts with the very ends of the transposon are only made after the formation of a synapse between the two Tcl ends. This has been shown for Mu, where the cleavage sites become protected from nuclease digestion after formation of the type 1 transpososome [Lavoie et al. 1991; Mizuuchi et al. 1991]. The nonspecific DNA-binding domain of TclA, which we find to be distinct from its site-specific recognition domain, might be responsible for interactions with the target DNA.

We observed that a recombinant amino-terminal derivative of TclA binds Tcl ends efficiently, whereas the full-length transposase has a very low affinity for Tcl ends. The results obtained with the nuclear extract of the transgenic NL224 line point to a similar conclusion. They suggest the presence in the transgenic nematode of a relatively rare polypeptide, comparable in size to the recombinant TclA derivative, with affinity for Tcl ends.

We have not yet identified the polypeptide responsible. Possibly, an internal region of the TclA protein influences the DNA-binding potential of the amino-terminal domain. The structural basis of this theory is not yet understood, but there are precedents for cryptic DNA-binding domains. Recently, it has been reported that amino-terminal amino acids of the prokaryotic σ70 transcription factor inhibit binding to promoter DNA [Dombroski et al. 1992]. Also, the carboxyl terminus of human p53 has been shown to regulate DNA binding [Hupp et al. 1992].

Synthesis of a truncated version of a transposase that plays a role in the regulation of transposition is known for the P element [Misra and Rio 1990], Tn5 [Johnson et al. 1982; Isberg et al. 1982], and IS1 [Escoubas et al. 1991]. These transposons regulate the level of transposition by controlling the relative abundance of specific sites on the transposon and act as inhibitors either by binding to the recognition sites within the element, or by protein–protein interaction with the transposase. We consider it possible that expression of TclA is similarly accompanied by the generation of a low level of a smaller version of the protein that regulates transposition.

Materials and methods

Cloning of various plasmids

Generation of the hsp–TclA fusion A PCR product containing the C. elegans heat shock promoter–untranslated RNA leader of hsp-16 (position 1018–1398, Stringham et al. 1992) was obtained from template pC31 using primers 2335 (5'-CCGGCATGCTGGAGGAAATAGTGGG) and 2123 (5'-CATTTCTTGAACTTTAGAGAATGCAAG). The TclA gene was obtained after PCR using primers 2334 (5'-ATGGAATTCTTAAATACCTTGCTGGCCTATCC) and 2122 (5'-CTCTAAAACCTCCAGATGTTAAATCTGTGTTGGG), together with template plM40, which contains a Tcl element that cosegregates with mutator activity (Mori 1988). As a result of an overlap extension [Horton et al. 1989] between the 5' end of the TclA PCR product and the 3' end of the hsp promoter PCR product, the two fragments could be fused and further amplified by PCR using primers 2334 and 2335 (25 cycles; 1 min at 95°C, 1 min at 55°C, 1 min at 72°C). Products were recovered and restricted with SplI and EcoRI and ligated into the Sphi–EcoRI-restricted plasmid backbone of pPD26.50 (a derivative of pP21.28, Fire et al. 1990), which resulted in plasmid pRf465.

Construction of pRP470 A similar PCR strategy was used for the removal of the 41-bp intron from the TclA gene from plM40 to express TclA in E. coli. The first exon was obtained using primers 2336 (5'-AGGCCATATGTTAAATCTGTGTTGGT) and 2407 (5'-CTCAATGGTATCTCAGTGTTTG) and the second exon was amplified using primers 2409 (5'-CAAGAAGATCCAAAATGAAATGTTTCATAG) and 2410 (5'-GAAGATTCTTACCTCAGGTTTG), and the second exon was amplified using primers 2409 (5'-CAAGAAGATCCAAAATGAAATGTTTCATAG) and 2410 (5'-GAAGATTCTTACCTCAGGTTTG), and the second exon was amplified using primers 2409 (5'-CAAGAAGATCCAAAATGAAATGTTTCATAG) and 2410 (5'-GAAGATTCTTACCTCAGGTTTG). The fused PCR product was amplified using primers 2336 and 964, recovered and subsequently restricted with Ndel and BglII, and ligated into the Ndel and BamHI sites of pET3c (Rosenberg et al. 1987) to yield pRP470.

Construction of deletion mutants Carboxy-terminal deletion pN207 resulted from linearization of pRP470 with ApaI, the site of which was made blunt by T4 DNA polymerase in the presence of DNTPs, and subsequent religation of the product, which results in an in-frame stop codon immediately downstream of the disrupted ApaI site. Bal31 deletions were generated from ApaI-linearized pRP470. Reaction products from different time points were digested with PsiI and run on an agarose gel. Carboxy-terminal deletion fragments were purified and cloned into pET3c digested with PsiI and BamHI [made blunt by Klenow and DNTPs]. Deletion endpoints were verified by sequencing the final constructs pN54, pN78, pN101, and pN153 in which the nomenclature corresponds with the number of authentic amino-terminal TclA codons. Finally, pN63 was generated by cloning the PsiI–Stul fragment with the 5' end of the TclA gene from pRP470 into pET3c digested with PsiI and BamHI. The use
of a translational stop codon in the vector-derived sequences resulted in the synthesis of a fusion protein with an additional 20 or 23 amino acids at the carboxyl terminus.

**Protein expression and purification**

Tc1A was expressed in *E. coli* strain BL21 pLysS grown in Luria broth at 30°C. When the cells reached an O.D. of 0.5 at 600 nm, IPTG was added to a concentration of 0.4 mM and incubation was continued for 2.5 hr. Bacteria were collected by centrifugation and resuspended in H buffer [25 mM HEPES (pH 7.5), 0.1 mM EDTA, 15% glycerol, 10 mM MgCl₂, 0.1% NP-40, 2 mM DTT] containing 1 mM PMSF and 1 mM NaCl. Cell lysis was achieved by sonication, and the insoluble material was removed by centrifugation. The Tc1A derivative was purified by the addition of one volume of saturated ammonium sulfate [pH 7.5] to the supernatant. Precipitated proteins were collected by centrifugation, re-suspended in 1 mM NaCl, and dialyzed against 200 mM NaCl in H buffer. This material was loaded onto a Sephrose FF column. The Tc1A-related polypeptide was eluted with a step from 350 to 550 mM NaCl in H buffer. This material was dialyzed against H buffer containing 100 mM NaCl and passed through a Sephrose FF (Pharmacia) column. The flowthrough fraction had a total protein concentration of ~1 mg/ml and contained ~25 µg/ml of a 18-kD polypeptide, which reacts with a polyclonal serum raised against Tc1A.

**Generation and analysis of transgenic animals**

A transgenic Bristol [N2] line (NL224) was obtained after microinjection [Mello et al. 1991] of 200 µg/ml of pRF4 [Kramer et al. 1990], together with 5 µg/ml of plasmid pRP465. Transgenic animals were grown at 18°C. DNA was isolated from *C. elegans* as described [Sulston and Hodgkin 1988]. Tc1 insertions in *gpa-2* were detected by two rounds of PCR [30 cycles; 1 min at 95°C, 1 min at 55°C, 1 min at 72°C]. The first PCR used as primer 964 and 2336. The cDNA was amplified by PCR using primers 964 and 2336 [5'-GTGCTAGTTTCAATCCAAGATC] and Tc1 primer L1 [5'-CGTGGGTATTCCTTGTTCGAAGCCAGCTACAATGGCTTTCT]. The second PCR used 1 µl of a 100-fold dilution of the first PCR as template in combination with primer 3173 [5'-GTGCTAGTTTCAATCCAAGATC] and Tc1 primer L2 [5'-TCAATAGTTTCAATCCAAGATC].

**Preparation of cytoplasmic and nuclear extracts of *C. elegans***

One volume of worms was mixed with two volumes of nuclear isolation buffer [NIB: 0.5 M sucrose, 25 mM HEPES (pH 7.5), 25 mM KCl, 0.1 mM EDTA, 0.15 M spermine, 0.15 M spermidine, 10 mM DTT, 10 mM MgCl₂, 0.1% PMSF]. This suspension was ground in liquid nitrogen until a fine white powder was obtained. Then, four volumes of NIB were added, and the cells were broken by 40 strokes in a Dounce homogenizer. Debris was removed by centrifugation at 200g for 2 min. Nuclei were collected by centrifugation of the supernatant for 5 min in an Eppendorf centrifuge. The supernatant contained the cytoplasmic fraction (~10 mg/ml of protein). Nuclei were extracted with two volumes of buffer H containing 500 mM NaCl. After homogenizing in a Dounce, the insoluble material was removed by centrifugation. The protein concentration of the nuclear extract was ~5 mg/ml.

**Probes and oligonucleotides**

Double-stranded oligonucleotides corresponding to regions of the ends of Tc1 [5'-AGTG-GATATCCTTTGCGGAGCACCCTGTA and its complement (bp -1 to +26) or 5'-ACCGCAAAGGTATCCTTTTGGC-CAGC and its complement (bp +5 to +32)] were subcloned into the polylinker site of pUC18. The nonspecific double-stranded oligonucleotide [5'-AAACCCGAGCCGAAAATCCTATCGACG and its complement] was also cloned into the polylinker of pUC18. Gel-retardation probes were obtained by end-labeling appropriate restriction fragments with [α³²P]dATP and Klenow. A footprint probe was obtained from pRSA, in which the terminal 225 bp of Tc1 was cloned as a BamH I-Rsal fragment from pRP4 into BamHI–SalI-digested pUC18. Footprint probes were obtained by restriction of pRSA with HindIII and end-labeling either with [α³²P]dATP and Klenow or by using T4 polynucleotide kinase and [γ³²P]ATP, followed by subsequent digestion with EcoRI.

**Western and South-Western blots**

Protein samples were subjected to electrophoresis in SDS–polyacrylamide gels, blotted to nitrocellulose, and either probed with a polyclonal antibody against Tc1A or with a polyclonal serum raised against Tc1A. The protein concentration of the nuclear extract was ~5 mg/ml.

**Acknowledgments**

We thank Drs. Stringham and Candido for pC31, Dr. Fire for pPD26.50, Dr. Mori for pM40, and Dr. Kramer for pRF4. We thank José Groenen for assistance with the microinjections. Stimulating discussions with our colleagues were greatly appreciated. We thank Piet Borst, Sean Colloms, Kenny Simmen, and Kees Vink for critically reading of the manuscript. This work was supported by a Pioneer grant of the Dutch science foundation NWO to R.H.A.P.

The publication costs of this article were defrayed in part by...
payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 USC section 1734 solely to indicate this fact.

References

Altschul, S.L., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215: 403–410.

Breznisky, L., G.V.L. Wang, T. Humphreys, and J. Hunt. 1990. The transposable element Uhu from Hawaiian Drosophila — Member of the widely dispersed class of Tcl-like elements. Nucleic Acids Res. 18: 2053–2059.

Collins, J., B. Saari, and P. Anderson. 1987. Activation of a transposable element in the germline but not the soma of Caenorhabditis elegans. Nature 328: 726–728.

Collins, J., E. Forbes, and P. Anderson. 1989. The Tcl family of transposable elements in Caenorhabditis elegans. Genetics 121: 47–55.

Craigie, R., M. Mizuuchi, and K. Mizuuchi. 1984. Site-specific recognition of the bacteriophage Mu ends by the Mu A protein. Cell 39: 387–394.

Derbyshire, K.M. and N.D.F. Grindley. 1992. Binding of the IS093 transposase to its inverted repeat in vitro. EMBO J. 11: 3449–3455.

Dombroski, A.J., W.A. Walter, M.T. Record Jr., D.A. Siegela, and C.A. Gross. 1992. Polypeptides containing highly conserved regions of transcription initiation factor σ54 exhibit specificity of binding to promoter DNA. Cell 70: 501–512.

Emmons, S.W. and L. Yesner. 1984. High-frequency excision of transposable element Tcl in the nematode Caenorhabditis elegans is limited to somatic cells. Cell 36: 599–605.

Escoubas, J.M., M.F. Prére, O. Fayet, I. Salvignol, D. Galas, D. Zerbib, and M. Chandler. 1991. Translational control of transposition activity of the bacterial insertion sequence IS1. EMBO J. 10: 705–712.

Feinberg, A.P. and B. Vogelstein. 1984. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 137: 266–267.

Fino Silva, I. and R.H.A. Plasterk. 1990. Characterization of a G-protein α-subunit gene from the nematode Caenorhabditis elegans. J. Mol. Biol. 215: 483–487.

Fire, A., S. White Harrison, and D. Dixon. 1990. A modular set of lacZ fusion vectors for studying gene expression in C. elegans. Gene 93: 189–198.

Gierl, A., S. Lutticke, and H. Saedler. 1988. TnpA product encodes the transposase of the bacterial insertion sequence IS1. EMBO J. 7: 4045–4053.

Harris, L.J., D.L. Baillie, and A.M. Rose. 1988. Sequence identity between an inverted repeat family of transposable elements in Drosophila and Caenorhabditis. Nucleic Acids Res. 16: 5991–5997.

Heierhorst, J., K. Lederis, and D. Richter. 1992. Presence of a member of the Tcl-like transposon family from nematodes and Drosophila within the vasotocin gene of a primitive vertebrate, the Pacific hagfish Eptatretus stouti. Proc. Natl. Acad. Sci. 89: 6798–6802.

Henikoff, S. 1992. Detection of Caenorhabditis transposon homologs in diverse organisms. New Biol. 4: 382–388.

Henikoff, S. and R.H.A. Plasterk. 1988. Related transposons in C. elegans and D. melanogaster. Nucleic Acids Res. 16: 6234.

Horton, R.M., H.D. Hunt, S.N. Ho, J.K. Pullen, and L.R. Pease. 1989. Engineering hybrid genes without the use of restriction enzymes: Gene splicing by overlap extension. Gene 77: 61–68.

Huisman, O., P.E. Errada, L. Signom, and N. Kleckner. 1989. Mutational analysis of IS10’s outside end. EMBO J. 8: 2101–2109.

Hupp, T.A., D.W. Meek, C.A. Midgley, and D.P. Lane. 1992. Regulation of the specific DNA binding function of p53. Cell 71: 875–886.

Isberg, R.R., A.L. Lazaar, and M. Syvanen. 1982. Regulation of Tn5 by the right-repeat proteins: Control at the level of the transposition reaction? Cell 30: 883–892.

Johnson, R.C., J.C.P. Yin, and W.S. Renzikoff. 1982. Control of Tn5 transposition in Escherichia coli is mediated by protein from the right repeat. Cell 30: 873–882.

Kaufman P.D., R.F. Doll, and D.C. Rio. 1989. Drosophila P element transposable recognizes internal P element DNA sequences. Cell 59: 359–371.

Kramer, J.M., R.P. French, E.-C. Park, and J.J. Johnson. 1990. The Caenorhabditis elegans rol-6 gene, which interacts with the sqt-1 collagen gene to determine organismal morphology, encodes a collagen. Mol. Cell. Biol. 10: 2081–2089.

Krosztrak, P.A. and J.J. Campos. 1990. Sequence-specific binding of DNA by the Moloney murine leukemia virus integrase protein. J. Virol. 64: 2796–2801.

Kunze, R. and P. Starlinger. 1989. The putative transposase of transposable element Ac from Zea mays L. interacts with subterminal sequences of Ac. EMBO J. 8: 3177–3185.

Lavoie, B.D., B.S. Chan, R.G. Allison, and G. Chaconas. 1991. Structural aspects of a higher order nucleoprotein complex: Induction of an altered DNA structure at the Mu-host junction of the Mu Type 1 transpososome. EMBO J. 10: 3051–3059.

Mello, C.C., J.M. Kramer, D. Stinchcomb, and V. Ambros. 1991. Efficient gene transfer in C. elegans: Extrachromosomal maintenance and integration of transforming sequences. EMBO J. 10: 3959–3970.

Misra, S. and D.C. Rio. 1990. Cytotype control of Drosophila P element transposition: The 66 kd protein is a repressor of transposase activity. Cell 62: 269–284.

Mizuuchi, M., T.A. Baker, and K. Mizuuchi. 1991. DNease protection analysis of the stable synaptic complexes involved in Mu transposition. Proc. Natl. Acad. Sci. 84: 1809–1813.

Mori, I. 1988. “Analysis of germline transposition and excision of Tcl transposable elements in Caenorhabditis elegans.” Ph.D. thesis. Washington University, St. Louis, MO.

Mori, I., G.M. Benian, D.G. Moerman, and R.H. Waterston. 1988a. Transposable element Tcl of Caenorhabditis elegans recognizes specific target sequences for integration. Proc. Natl. Acad. Sci. 85: 861–864.

Mori, I., D.G. Moerman, and R.H. Waterston. 1988b. Analysis of a mutator activity necessary for germline transposition and excision of Tcl transposable elements in Caenorhabditis elegans. Genetics 120: 397–407.

Plasterk, R.H.A. 1991. The origin of footprints of the Tcl transposon of Caenorhabditis elegans. EMBO J. 10: 1919–1925.

Prasad, S.S., L.J. Harris, D.L. Baillie, and A.M. Rose. 1991. Evolutionarily conserved regions in Caenorhabditis transposable elements deduced by sequence comparison. Genome 34: 6–12.

Rosenberg, A.H., B.N. Lade, D.-s. Chui, S.-W. Lin, J.J. Dunn, and F.W. Studier. 1987. Vectors for selective expression of cloned DNAs by T7 RNA polymerase. Gene 56: 125–135.

Rosenquist, T.A. and J. Kimble. 1988. Molecular cloning and transcription analysis of fem-3, a sex-determination gene in Caenorhabditis elegans. Genes & Dev. 2: 606–616.

Rosenzweig, B., L.W. Liao, and D. Hirsch. 1983a. Sequence of the C. elegans transposable element Tcl. Nucleic Acids Res.
C. elegans Tcl transposase

Rosenzweig, B., L.W. Liao, and D. Hirsch. 1983b. Target sequences for the C. elegans transposable element Tcl. *Nucleic Acids Res.* 11: 7137–7140.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Schukkink, R.F. and R.H.A. Plasterk. 1990. TcA, the putative transposase of the C. elegans Tcl transposon, has an N-terminal DNA binding domain. *Nucleic Acids Res.* 18: 895–900.

Stalder, R., P. Caspers, F. Olasz, and A. Werner. 1990. The N-terminal domain of the insertion sequence 30 transposase interacts specifically with the terminal inverted repeats of the element. *J. Biol. Chem.* 265: 3757–3762.

Stringham, E.G., D.K. Dixon, D. Jones, and E.P.M. Candido. 1992. Temporal and spatial expression patterns of the small heat shock (hsp16) genes in transgenic Caenorhabditis elegans. *Mol. Biol. Cell* 3: 221–233.

Sulston, J.E. and J. Hodgkin. 1988. Methods. In *The nematode Caenorhabditis elegans* (ed. W.B. Wood), pp. 587–606. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
Characterization of the Caenorhabditis elegans Tc1 transposase in vivo and in vitro.

J C Vos, H G van Luenen and R H Plasterk

Genes Dev. 1993, 7:
Access the most recent version at doi:10.1101/gad.7.7a.1244