Nucleic Acids Research, 1993, Vol. 21, No. 1 59–67

A variant Tc4 transposable element in the nematode *C.elegans* could encode a novel protein

Wei Li and Jocelyn E.Shaw*
Department of Genetics and Cell Biology, University of Minnesota, St Paul, MN 55108, USA

Received October 1, 1992; Revised and Accepted November 24, 1992 GenBank accession no. L00665

ABSTRACT

A variant *C.elegans* Tc4 transposable element, Tc4-*rh1030*, has been sequenced and is 3483 bp long. The Tc4 element that had been analyzed previously is 1605 bp long, consists of two 774-bp nearly perfect inverted terminal repeats connected by a 57-bp loop, and lacks significant open reading frames. In Tc4-*rh1030*, by comparison, a 2343-bp novel sequence is present in place of a 477-bp segment in one of the inverted repeats. The novel sequence of Tc4-*rh1030* is present about five times per haploid genome and is invariably associated with Tc4 elements; we have used the designation Tc4v to denote this variant subfamily of Tc4 elements. Sequence analysis of three cDNA clones suggests that a Tc4v element contains at least five exons that could encode a novel basic protein of 537 amino acid residues. On northern blots, a 1.6-kb Tc4v-specific transcript was detected in the mutator strain TR679 but not in the wild-type strain N2; Tc4 elements are known to transpose in TR679 but appear to be quiescent in N2. We have analyzed transcripts produced by an *unc-33* gene that has the Tc4-*rh1030* insertional mutation in its transcribed region; all or almost all of the Tc4v sequence is frequently spliced out of the mutant *unc-33* transcripts, sometimes by means of non-consensus splice acceptor sites.

INTRODUCTION

Tc4 is a distinctive family of transposable elements in the nematode *Caenorhabditis elegans* (1). The first member of the family to be sequenced, Tc4-*n1416*, was shown to be composed of two almost perfect 774-bp terminal inverted repeats joined by a 57-bp unique internal sequence. Tc4-*n1416* is thus a fold-back element, although its long inverted terminal repeats, unlike the terminal repeats of fold-back elements in Drosophila (2, 3) and the sea urchin (4), do not contain extensive regions of short direct repeats. No long open reading frame was found within Tc4-*n1416*. Tc6, a probable transposable element of *C.elegans* (5), has a fold-back structure similar to that described for Tc4-*n1416*, containing near-perfect inverted repeats of 765 bp, an internal loop of 73 bp, and no long open reading frames; however, this element bears no obvious sequence similarity to Tc4. Other reported families of transposable elements in *C.elegans*, Tc1 (6, 7), Tc2 (8, 9), and Tc3 (10), contain inverted terminal repeats, but their repeats are much shorter—constituting a minor fraction of each element. The extreme termini of Tc1, Tc3 and Tc6 share some sequence identity, however, this sequence similarity is not shared by Tc4.

Tc4 elements have been shown to transpose in the *C.elegans* mutator strain TR679, which contains a *mut-2* mutation (11): Tc4-*n1416* transposed into the *ced-4* gene (1), Tc4-*n1351* was identified as an insertional mutation in *unc-86* (12), and Tc4-*mn260* and Tc4-*rh1030* were found in *unc-33* (13). In addition, Southern blots probed with Tc4-*n1416* sequence identified extra copies of Tc4 elements in TR679 as compared to its progenitor strains, Bristol (N2) and Bergerac, which were estimated to contain about 20 copies of Tc4 each per haploid genome (1). The elements whose transposition is enhanced in TR679—Tc1, Tc3, Tc4, and Tc5 (J.Collins and P.Anderson, pers. commun.; 14)—do not have any obvious features in common. The Tc1 and Tc3 elements have identical sequence in 8 of 9 terminal base pairs; however, this sequence is not shared by Tc4 or Tc5, which themselves have similar terminal sequences (J.Collins, pers. commun.). It is unclear how the different classes of transposons are activated in TR679.

In general the mechanism of transposition of fold-back elements is not known. In the Drosophila P element family of transposons, which contain short inverted repeats, a specific transposase acts on the inverted terminal repeats of a P element to promote the element's transposition (for review see 15). The gene for transposase is found in certain P elements, but defective P elements lacking a functional transposase gene may be activated as long as they have intact inverted terminal repeats. By analogy, it has been proposed that those FB fold-back elements of Drosophila that seem to lack protein-coding capacity may be activated by the protein products of other FB fold-back members that have an extensive central loop region (16) containing three long open reading frames (17). Similar considerations may apply to the Tc4 family. Yuan *et al.* (1) suggested from their analysis of Southern blots that the Tc4 family is heterogeneous. The first three Tc4 elements identified, Tc4-*n1416*, Tc4-*n1351*, and Tc4-*mn260*, all proved to be 1.6-kbp elements. The element Tc4-*rh1030* appeared to be 3.5 kbp in size; therefore, we have analyzed it in detail.

* To whom correspondence should be addressed

MATERIALS AND METHODS

C.elegans strains and culture

The *C.elegans* strains used in this work were the Bristol wild-type strain N2, the mutator strain TR679 (11), and NJ294, a transposon-induced *unc-33(rh1030)* mutant isolated from TR679 by E. Hedgecock. Nematodes were grown at 20°C on NG plates seeded with *E.coli* strain OP50 (18).

Cloning and sequencing Tc4-*rh1030* and Tc4v cDNAs

General methods for manipulating and analyzing nucleic acids were by standard procedures (19, 20). Previous Southern blot analysis (13) showed that the *unc-33(rh1030)* mutant had a Tc4 element inserted in the *unc-33* gene such that the size of a *Sac*I restriction fragment increased from 0.4 kbp to 3.9 kbp. Genomic DNA from *unc-33(rh1030)* was digested with *Sac*I and the fragments were separated by electrophoresis through a 0.7% agarose gel. The 3.9-kbp size fraction of *Sac*I-digested DNA was isolated and cloned into pUC119 (21). Clones containing the *Sac*I fragment of interest were identified by their hybridization to ^{32}P-labeled (22) Tc4-*n1416* (1).

A ^{32}P-labeled (22) 2.1-kbp *Eco*RV fragment from Tc4-*rh1030* was used as probe to isolate cDNA clones specific to the Tc4v subset of Tc4 elements; this probe does not hybridize to Tc4-*n1416*. Two clones, EH#J1 and EH#J9, were recovered from a λgt10 library made from early embryonic N2 RNA (provided by J. Ahringer and J. Kimble), and one clone, EH#B5, was recovered from a λ-ZAP library prepared from mixed-stage N2 RNA (provided by R. Barstead and R. Waterston). The cDNA insert in EH#B5 contains a poly(A) tail and approximately 1.2 kbp of Tc4v-specific sequence; the 1-kbp insert in EH#J1, representing the 5′ portion of the message, overlaps with EH#B5 and extends approximately 0.7 kbp further 5′; the 0.9-kbp insert in EH#J9 overlaps entirely with EH#B5 and represents the 3′ portion of the message. Together, the three cDNA clones represent 1.9 kb of continuous sequence of a transcript from Tc4v. Each of the cDNA inserts was subcloned into pUC119 for sequencing.

The sequences of Tc4-*rh1030* and of the cDNA inserts were determined by dideoxy sequencing (23) of double-stranded templates from sets of nested deletions generated by exonuclease III and S1 nuclease digestion (24). Both DNA strands of Tc4-*rh1030* were sequenced; one strand of each of the cDNA inserts was sequenced. Six percent polyacrylamide-8 M urea sequencing gels were run as described by Sheen and Seed (25).

Southern blot analysis

Genomic DNA was isolated from worms cultured on NG agarose plates by methods described previously (13, 26). DNA samples digested with either *Hin*dIII or *Bfa*I were electrophoresed through 0.8% agarose in 1×TBE (19) buffer and transferred to Zeta-Probe nylon membranes (Bio-Rad) by capillary blotting with 0.4 M NaOH. After neutralization with 2×SSPE (19), the DNA blots were prehybridized in 50% formamide, 0.25 M $NaHPO_4$, pH 7.2, 0.25 M NaCl, 7% SDS, 1 mM EDTA, and 100 μg/ml denatured sonicated herring sperm DNA at 42°C for more than 6 hr. ^{32}P-labeled Tc4-*n1416* probe (1), generated by random priming (22), was added to the same solution, and the blots were hybridized overnight with agitation. The membranes were washed at 50°C four times, 15 min each, in 0.2×SSPE and 0.1% SDS. Hybridizing probe was removed from the membranes by incubation in 0.1×SSPE and 0.5% SDS at 92°C; blots were then re-hybridized with a ^{32}P-labeled 2.1-kbp *Eco*RV fragment from Tc4-*rh1030*.

Northern blot analysis

Mixed-stage populations of *C.elegans* were collected from enriched NG agar plates (NG plates with 20g/liter peptone) and washed several times with water to remove contaminating *E.coli*. Two ml of pelleted worms were homogenized at 4°C in 15 ml of 4 M guanidine HCl, 0.5% sodium lauryl sarcosinate, 1 mM EDTA, 0.2 M sodium acetate, pH 5.2, and 0.1 M β-mercaptoethanol with a French pressure cell at 11,000 psi. The homogenates were immediately extracted 3−4 times with an equal volume of ice-cold phenol:chloroform:isoamyl alcohol (25:24:1) and once with chloroform:isoamyl alcohol (24:1). RNA was precipitated with 0.5 volume of 95% ethanol at −20°C overnight. Enrichment for poly(A) RNA was achieved by one round of selection by oligo(dT) chromatography (27). Poly(A)-enriched RNA samples (5 μg) were electrophoresed through a 1% agarose-6.6% formaldehyde gel, transferred to a Zeta-Probe membrane using 20×SSPE, and fixed to the membrane by baking for 2 hr at 80°C under vacuum. The prehybridization, hybridization and washing solutions were the same as those used for DNA blots, but the temperatures were 50°C for prehybridization and hybridization and 55°C for washing. The blot was probed with a ^{32}P-labeled 2.1-kbp *Eco*RV fragment from Tc4-*rh1030* and then re-hybridized to ^{32}P-labeled pT7/T3-18-103 (provided by M.Krause), which hybridizes to the 3′ end of the *act-1* transcript (28).

PCR analysis of *unc-33(rh1030)* transcripts

The sequence of a 20mer (RT) primer (5′-CAGCCAGTCGAGAGATCATC-3′) designed for reverse transcription of *unc-33(rh1030)* transcripts is complementary to a sequence located 254 bp downstream of the Tc4-*rh1030* insertion site, in the *unc-33* mRNA sequence (13). To synthesize the first strand of cDNA, one μg of poly(A)-enriched RNA from an asynchronous *unc-33(rh1030)* population was incubated with 10 units of AMV reverse transcriptase in a 25 μl mixture containing 1×PCR buffer (Promega Corporation), 1 mM of each deoxyribonucleoside triphosphate (dNTP), 10 units of RNasin (Promega), 6.5 mM $MgCl_2$, and 10 pmol of RT primer at 42°C for 30 min and then at 52°C for 30 min. The reaction mixture was further incubated at 65°C for 15 min and then diluted with 500 μl of 0.1×TE. Ten μl of the cDNA was amplified in 100 μl of 1×PCR buffer, 0.2 mM of each dNTP, 50 pmol of 5′ amplification primer (5′-AATGGAGAGACGCCGACAGA-3′, a sequence located in *unc-33* mRNA 63 bp upstream of the Tc4-*rh1030* insertion site, see Fig. 4A), 50 pmol of 3′ amplification primer (5′-TCGGTGACCTGAGTGTAGAC-3′, a sequence complementary to *unc-33* mRNA 223 bp downstream of the Tc4-*rh1030* insertion site, see Fig. 4A) and 5 units of Taq polymerase (Promega Corporation) in a DNA Thermal Cycler (Perkin-Elmer Cetus). The 5′ ends of the *unc-33(rh1030)* messages were also amplified with 50 pmol of *C.elegans* SL1 trans-spliced leader (29) primer (5′-GGTTTAATTACCCAAGTTCGAG-3′) in place of the 5′ amplification primer. The first PCR cycle consisted of denaturation at 94°C for 5 min, annealing at 50°C for 3 min, and extension for 15 min at 72°C. The remaining 35 cycles of amplification were carried out by running a step-cycle file of 93°C, 1 min; 50°C, 1 min; 72°C, 2 min,

followed by a 15 min final extension at 72°C. For cloning, the PCR products were extracted once with phenol, extracted once with choloroform:isoamyl alcohol (24:1), and precipitated with 95% ethanol at −20°C overnight. The products were blunt-ended using Klenow fragment (24) and digested with *Cla*I, which cuts once within the amplified cDNA immediately 5′ of the 3′ amplification primer. The restriction fragments were then cloned into *Sma*I- and *Cla*I-cleaved plasmid pUC119. Transformants that hybridized to a probe specific to the cDNA fragment between the 5′ and 3′ amplification primers were isolated for sequencing. Single strands of the inserts were sequenced as described above.

Sequence analysis

Nucleotide and putative amino acid sequences were analyzed using IntelliGenetics programs. The IntelliGenetics FastDB program was used to search the PIR, Swiss-Prot, and Prosite databanks for polypeptides with amino acid sequences similar to sequences within the predicted Tc4v polypeptide and for amino acid sequence similarities with polypeptides predicted from other ORFs within Tc4v. The PIR and SwissProt databanks were also searched using the BLAST program (30). The Tc4-*rh1030* DNA sequence was compared to sequences deposited in the GenBank and EMBL databanks using the IntelliGenetics FastDB program.

RESULTS

Identification of a variant Tc4 transposable element, Tc4-*rh1030*

In previous work related to an analysis of mutations in *unc-33*, which lead to severely uncoordinated movement and abnormalities in the guidance and outgrowth of many neurons in *C.elegans* (13), we identified two Tc4 elements: Tc4-*mn260* and Tc4-*rh1030*. Each element was associated with a spontaneous *unc-33* mutation occurring in the *C.elegans* mutator strain TR679. Both elements were classified as Tc4 elements by virtue of their hybridization to probes made from Tc4-*n1416* (1); however, Tc4-*rh1030* appeared to be an insert of 3.5 kbp (13), whereas Tc4-*mn260* (13) and the previously-studied Tc4 elements, Tc4-*n1416* and Tc4-*n1351* (1), were all 1.6 kbp in size. We also found that the distribution of restriction endonuclease cleavage sites differed between Tc4-*rh1030* and the three 1.6-kbp Tc4 elements. For example, Tc4-*rh1030* contains an *Eco*RI site that is absent from the 1.6-kbp elements and the 1.6-kbp elements have a *Sac*I site that is not present in Tc4-*rh1030*. On the basis of these differences it appeared that Tc4-*rh1030* was a variant of the Tc4 elements that had been analyzed previously; therefore, we cloned Tc4-*rh1030* from the *unc-33(rh1030)* mutant for further analysis.

Tc4-*rh1030* has one of its inverted repeats disrupted by a novel 2343-bp sequence

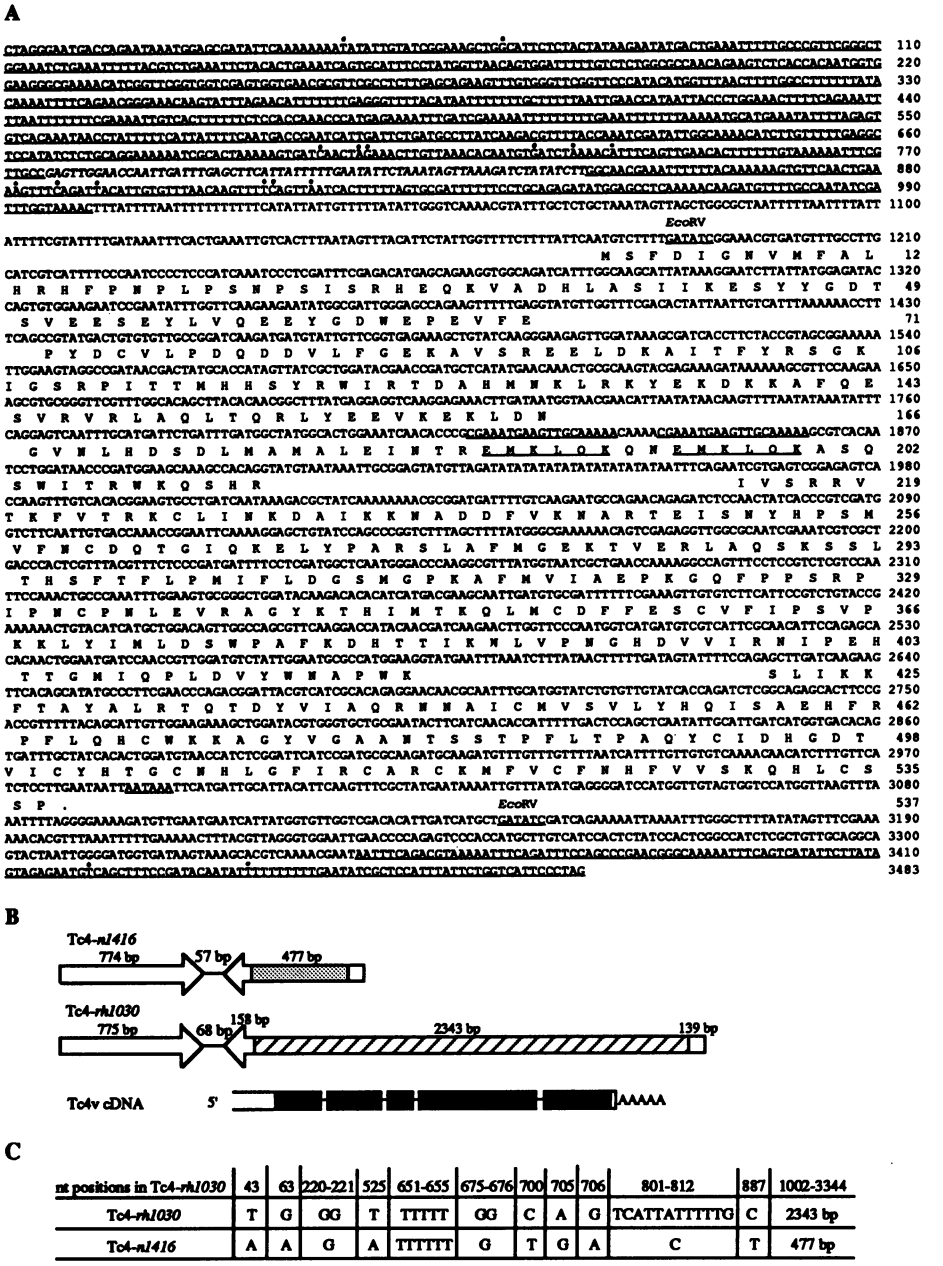
The sequence of the entire 3483-bp Tc4-*rh1030* element is presented in Fig. 1, along with a listing of the positions at which its sequence differs from that of the 1605-bp Tc4-*n1416* element (Fig. 1C). The latter element consists of almost perfect inverted terminal repeats of 774 base pairs (fold-back region) joined by a 57-bp unique internal sequence (loop region). The major difference between the two Tc4 elements is that a novel 2343-bp sequence in Tc4-*rh1030* replaces a 477-bp segment near the middle of the right inverted repeat of Tc4-*n1416* (Fig. 1B), whereas the left inverted repeats of the two elements are nearly identical. Thus the fold-back structure of the Tc4-*rh1030* element is drastically different from that of Tc4-*n1416*. The internal loop sequence of Tc4-*rh1030* is identical to that of Tc4-*n1416* except for the substitution of 12 bp (at positions 801−812) for a single C. This substitution results in a 68-bp loop sequence for Tc4-*rh1030*, 11 bp longer than the loop sequence of Tc4-*n1416*. [Note that alignment of the loop sequences of the two Tc4 elements is the basis upon which we orient the elements and make comparisons between the left (or right) inverted repeats of the two elements. In our description of Tc4 elements we have inverted the orientation, left vs. right, from the orientation presented previously by Yuan *et al.* (1)]. There are also a few relatively minor differences between the Tc4-*rh1030* and the Tc4-*n1416* elements. The Tc4-*rh1030* element contains a left inverted repeat of 775 bp, one base pair longer than the Tc4-*n1416* repeat. It has nine sequence differences from the left repeat of Tc4-*n1416*: six base pair substitutions, two single base pair additions, and one base pair deletion. The sequence of the remaining 297 bp of the disrupted right repeat of Tc4-*rh1030*, which includes a 158-bp segment proximal to the internal loop region and a 139-bp segment distal to the loop, differs from the right inverted repeat of Tc4-*n1416* at only one position, a single base substitution in the proximal segment (Fig. 1C).

We have compared the sequences of the 158-bp and 139-bp portions of the disrupted right inverted repeat of Tc4-*rh1030* with the corresponding segments of its left inverted repeat. There are eight positions throughout the 297 bp at which the left and right repeats differ from perfect inverted repeats; all depart from perfect symmetry by single base pair substitutions (Fig. 1A). Two of the left-right differences, at positions 727/892 and 737/882, were also present in Tc4-*n1416* (1). Yuan et al. (1) found only two other differences between the left and right inverted repeats of Tc4-*n1416*, a single base pair addition and a single base pair deletion; both of the these left-right differences are absent from Tc4-*rh1030* and account for two of the differences between Tc4-*rh1030* and Tc4-*n1416*, at positions 651−655 and 675−676 (Fig. 1C). At these positions the left and right inverted repeats of Tc4-*rh1030* have sequences identical to the corresponding positions of the right inverted repeat of Tc4-*n1416*.

The 2343-bp novel sequence in Tc4-*rh1030* is 60% AT overall (79% AT throughout the first 175 bp in Fig. 1A) and has a few rather long open reading frames (ORFs) and several conserved intron splice sites on both strands. Databank searches did not reveal any significant similarities of this segment with other known DNA sequences. The novel sequence also has no apparent homology to any portion of Tc4-*n1416*. We considered the possibility that the 2343-bp sequence may represent a transposable element that transposed into a Tc4 inverted repeat; however, the ends of the novel segment do not have inverted or direct repeats, nor do the regions immediately flanking each end of the 2343-bp sequence. We are thus left with no evidence that the novel segment within Tc4-*rh1030* represents a transposable element.

Copies of the 2.3-kbp novel sequence in Tc4-*rh1030* are associated exclusively with members of a Tc4 subfamily, Tc4v

If the 2343-bp sequence is an integral component of a subset of the Tc4 family, we would expect that it would be present only in additional Tc4 copies, and that it would not be found isolated from Tc4. To test the relative locations of Tc4 elements and the 2343-bp novel sequence, a 2.1-kbp *Eco*RV fragment generated wholly from the novel sequence (Fig. 1A) was used to re-probe



C

nt positions in Tc4-*rh1030*	43	63	220-221	525	651-655	675-676	700	705	706	801-812	887	1002-3344
Tc4-*rh1030*	T	G	GG	T	TTTTT	GG	C	A	G	TCATTATTTTTG	C	2343 bp
Tc4-*n1416*	A	A	G	A	TTTTTT	G	T	G	A	C	T	477 bp

Figure 1. The structure of Tc4-*rh1030*. (**A**) Nucleotide sequence and predicted amino acid sequence of Tc4-*rh1030*. The sequence data are available under GenBank accession No. L00665. The inverted terminal repeat sequences are underlined, and the internal or loop sequence is in italics. The 2343-bp novel sequence, which disrupts the right inverted repeat, is displayed in plain text. The eight single base pair differences between the left and right inverted repeats are indicated by dots in both repeats. The *Eco*RV sites near the ends of the novel sequence are indicated. The determined cDNA sequence differs from the genomic sequence of Tc4-*rh1030*, given here, in four respects: three single base substitutions, from A in the genomic sequence to G in the cDNA, were found at positions 930, 936, and 948, and a 22-bp segment at positions 1837–1858 of the genomic sequence was replaced by CTTC in the cDNA. This net 18-bp deletion occurred within a region that contains two 19-bp direct repeats—underlined in the figure—separated by 5 bp. The proposed polyadenylation signal beginning at position 2986 is underlined. (**B**) Diagrammatic representations of Tc4-*n1416* and Tc4-*rh1030*. The stippled box represents the 477-bp segment in Tc4-*n1416* that is replaced by a 2343-bp novel sequence in Tc4-*rh1030*, represented by a striped box. At the bottom is represented the structure of the Tc4v cDNA derived from three cDNA clones. Filled boxes are ORFs. Open boxes represent the untranslated regions. The 5' end of the Tc4v message has not been determined. (**C**) Listing of nucleotide sequence differences between Tc4-*rh1030* and Tc4-*n1416*, using the Tc4-*rh1030* position numbers, as given in Fig. 1A. The published Tc4-*n1416* sequence has been inverted to facilitate presentation of the amino acid sequence encoded by Tc4-*rh1030*.

Southern blots of N2 and TR679 genomic DNA that had been probed previously with the 1.6-kbp Tc4-*n1416* element. After digestion with *Hin*dIII, which does not cut within any of the known copies of Tc4, the Tc4-*n1416* probe hybridized to about 25 fragments of DNA from N2 and about 30 fragments from TR679, in which Tc4 is known to transpose (1, 12, 13) (Fig. 2). When the Southern blot was re-probed with the 2.1-kbp *Eco*RV sequence, we identified five fragments from both N2 and TR679, four of which were common to the two strains and all of which were larger than 3.5 kbp. For each strain, all five

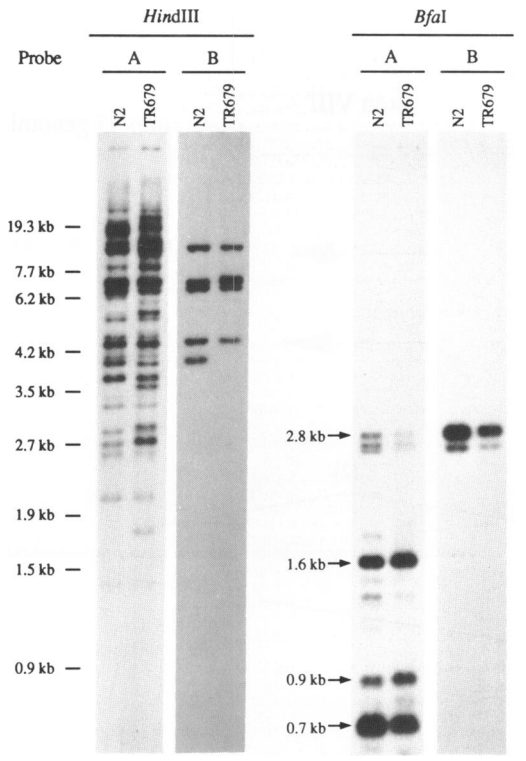


Figure 2. Southern blot analysis of Tc4 elements in *C.elegans* strains N2 and TR679. Genomic DNA was digested with *Hin*dIII or *Bfa*I, electrophoresed through 0.7% agarose gels, and transferred to nylon membranes. The membranes were probed with Tc4-*n1416* (probe A), stripped and then re-probed with a 2.1-kbp *Eco*RV sequence (probe B) from the novel sequence of Tc4-*rh1030*. Fragments that hybridized to probe B contain members of the Tc4v subfamily of Tc4 elements.

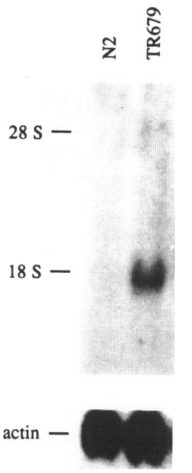


Figure 3. Northern blot analysis of Tc4v transcripts. Five μg of poly(A)-enriched RNA isolated from strains N2 or TR679 were run through 1% agarose-6.6% formaldehyde gels. The Northern blot was hybridized with a probe generated from a 2.1-kbp *Eco*RV sequence from the novel sequence of Tc4-*rh1030*. The *C.elegans* 18s and 28s rRNAs are 1,750 and 3500 bases long, respectively (35). A 1.6-kilobase transcript was detected in strain TR679 but not in N2. A re-probing of the blot with an *act-1* probe showed that the two lanes contained equivalent amounts of RNA.

fragments appeared to electrophorese to the same positions as five Tc4-*n1416*-hybridizing fragments. This co-migration of fragments containing the novel sequence with Tc4-*n1416*-hybridizing fragments was also detected on Southern blots of N2 genomic DNA cleaved with *Bfa*I (Fig. 2) and with six other restriction enzymes (data not shown). We conclude that the 2343-bp sequence is invariably associated with sequences that are homologous to Tc4-*n1416*. We refer to the Tc4 subfamily containing sequence homology to the 2.1-kbp *Eco*RV sequence as Tc4v. The intensities of hybridization with the 2.1-kbp *Eco*RV probe were similar for all five hybridizing fragments of N2 and TR679 (Fig. 2); therefore, it appears that the N2 and TR679 strains each contain five Tc4v elements per haploid genome.

As reported by Yuan et al. (1) and evident from *Hin*dIII-digested DNA shown in Fig. 2, Tc4-containing fragments exhibited varying intensities of hybridization when probed with Tc4-*n1416*, which could be explained by size and/or sequence heterogeneity among Tc4 elements (1). The Tc4v elements represent one type of variant within the Tc4 family. To explore further the question of heterogeneity of Tc4 elements, Southern analysis was performed on DNA digested with *Bfa*I, which recognizes CTAG. The two occurrences of CTAG in Tc4-*n1416* are at the very ends of the Tc4 inverted terminal repeats. Tc4-*rh1030* has these sites as well as one other, at positions 703–706 (as noted in Fig. 1C, Tc4-*rh1030* has AG at positions 705–706 rather than GA). On a Southern blot of *Bfa*I-digested DNA probed with Tc4-*n1416*, elements like Tc4-*n1416* should give a prominent 1.6-kbp DNA fragment, and elements like Tc4-*rh1030* should give 0.7- and 2.8-kbp fragments. Strongly hybridizing fragments of these sizes were indeed detected (Fig. 2), although the 0.7-kbp fragment hybridized much more intensely than would be expected if it arose only from cleavage of Tc4v elements, and at least eight additional bands, of varying intensity, were observed. The hybridization patterns were identical for N2 and TR679, suggesting that the CTAG termini are present in elements that have transposed to different chromosomal locations in TR679. Although it is possible that there is a large class of Tc4 elements only 0.7 kbp in length, we speculate that the intense hybridization to *Bfa*I fragments of 0.7 kbp may be due to the presence of 1.6-kbp Tc4 elements containing one or two *Bfa*I restriction sites at positions comparable to the one found in the left terminal repeat of Tc4-*rh1030*. An element with a *Bfa*I site in one inverted repeat would give one 0.7-kbp fragment and one 0.9-kbp fragment, a size which is also prominent; an element with symmetrical *Bfa*I sites would give two 0.7-kbp fragments and a 0.2-kbp fragment (not present on our gels). These results indicate that although there is heterogeneity among the Tc4 family, the majority of elements may be approximately the same size.

When *Hin*dIII-digested N2 or TR679 DNA was probed with the 2.1-kbp *Eco*RV fragment unique to Tc4v elements, the intensity of hybridization was approximately the same for all fragments (Fig. 2), suggesting that members of the Tc4v subfamily may be fairly homogenous. The question of heterogeneity within the Tc4v subfamily was addressed by probing *Bfa*I-digested DNA with the 2.1-kbp *Eco*RV probe. Fig. 2 shows that both N2 and TR679 DNA yielded a 2.8-kbp hybridizing fragment, as expected, but they also gave a 2.6-kbp hybridizing fragment. We suggest that perhaps one member of the Tc4v family has a *Bfa*I restriction site in its disrupted right repeat at the position comparable to the one found in the left repeat of Tc4-*rh1030*. This would reduce the size of the 2.8-kbp *Bfa*I fragment to 2.6 kbp. An alternative possibility is that a Tc4v

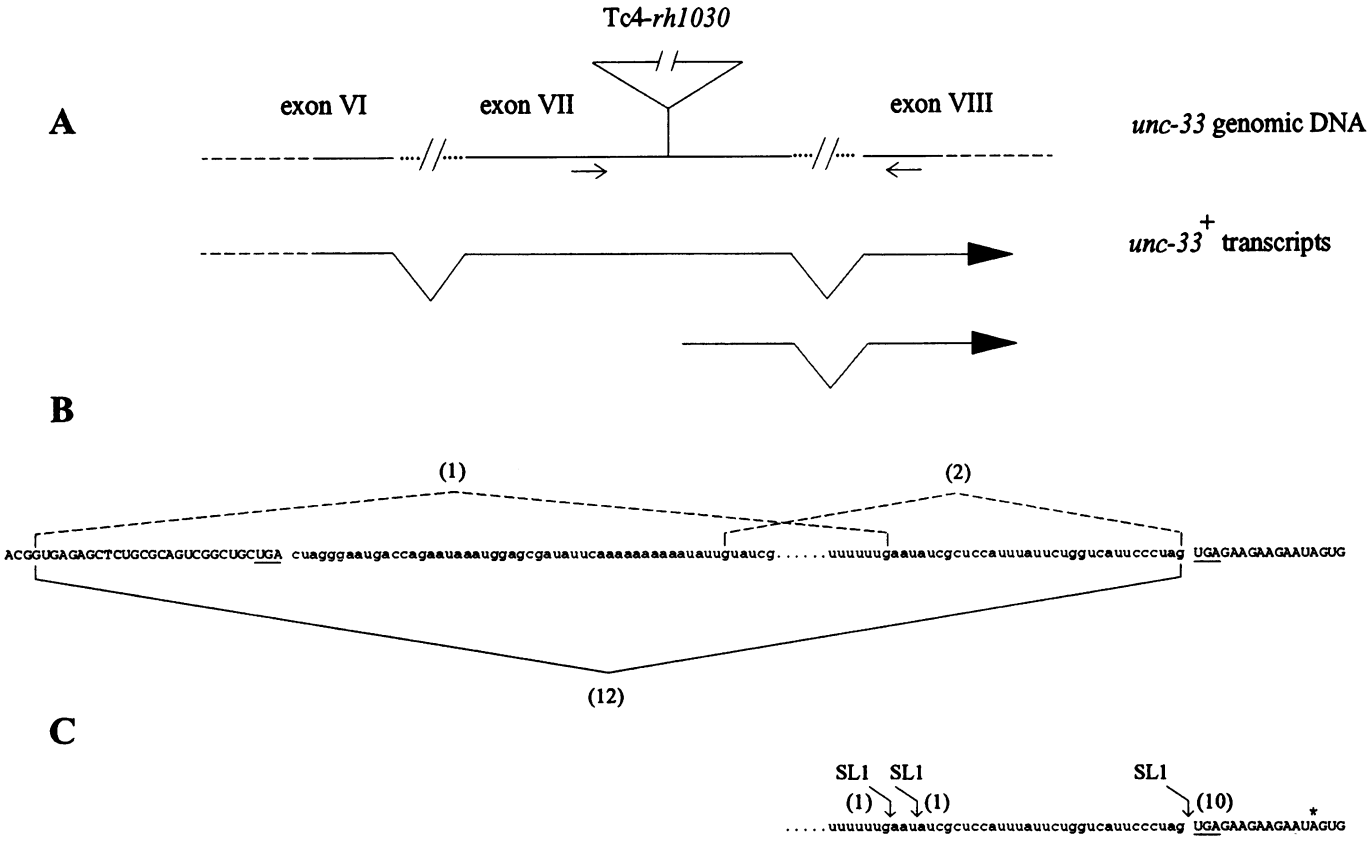


Figure 4. Structures of spliced transcripts from *unc-33(rh1030)* determined by RT-PCR. (**A**) A portion of the *unc-33* gene is depicted with the insertion site of Tc4-*rh1030* indicated. Transcripts that extend through this region or are initiated within this region are illustrated. Small arrows beneath the *unc-33* genomic region indicate the position of primers used for RT-PCR amplification. (**B**) Segments that were spliced out of transcripts from *unc-33(rh1030::Tc4v)* that extended through the region are depicted. The sequences in lower-case letters are the sequences of the termini of the 3.5-kbp Tc4-*rh1030* sequence, which is shown here in the opposite orientation as in Fig. 1. The duplicated UGAs that flank the Tc4-*rh1030* sequence are underlined. Transcription of the *unc-33* gene is from left to right. Twelve clones were completely missing Tc4-*rh1030* sequence along with 28 nucleotides of upstream *unc-33* sequence. Two clones were missing all but the 48 left-most nucleotides from Tc4-*rh1030*, and one clone was missing all but the 33 right-most nucleotides from Tc4-*rh1030* as well as the 28 nucleotides of upstream *unc-33* sequence. (**C**) The splicing patterns of transcripts that were trans-spliced to SL1. The site of transcriptional initiation of the smallest wild-type *unc-33* transcipt (which is not trans-spliced) is marked by an asterisk. Ten of the 12 sequenced clones had SL1 spliced to *unc-33* sequence at the point of insertion of Tc4-*rh1030*. The other two clones were trans-spliced to SL1 at points slightly upstream, within the terminus of Tc4-*rh1030*.

element has suffered a deletion; however, if this is the case, the deletion does not appear to be within its *Eco*RV fragment. When *Eco*RV-digested N2 genomic DNA was probed with the 2.1-kbp *Eco*RV fragment from Tc4-*rh1030*, only a fragment of approximately 2.1 kbp was detected (data not shown). These results are consistent with the view that members of the Tc4v subfamily are fairly homogenous and that the novel sequence which distinguishes the Tc4v subclass is close to the same size in all Tc4v elements.

Tc4v has protein-coding capacity

We used the 2.1-kbp *Eco*RV fragment as probe to screen cDNA libraries prepared from N2 RNA. Three cDNA clones were isolated and sequenced; together they represent 1.9 kb of a transcript from Tc4v. The 5′ end of the cDNA sequence lies within the right Tc4-*rh1030* inverted repeat at position 922 (see Fig. 1B); however, we do not know whether the cDNA represents a full-length transcript since we have not mapped the 5′ end of the Tc4v message. The cDNA extends through most of the novel 2343-bp sequence and terminates within this region at position 3002–3003. The transcribed region of Tc4v contains five exons and four very short introns, ranging in length from 43 to 58 nucleotides. The 5′ and 3′ intron splicing sequences are all consistent with *C.elegans* consensus splicing sequences (31). The poly(A) tract in the cDNA clone representing the 3′ end of the message begins 11–12 nucleotides downstream from a potential polyadenylation signal AATAAA at positions 2986–2991. There is another AATAAA sequence 33 bp downstream.

The longest putative ORF of the message is 1.6 kb long, beginning from the first ATG near the 5′ end of the cDNA sequence at position 1172 and terminating at the stop codon TGA starting at position 2976 (Fig. 1A). The 5′ untranslated region (UTR) represented in the cDNA (260 nucleotides long) has stop codons in all three frames, and the 3′ UTR is only 23–24 nucleotides long. The ORF could encode a polypeptide of 537 amino acid residues, with a mass of 62 kD. The predicted protein is highly hydrophilic and overall is positively-charged with a pI of 8.83. Twelve of the 46 amino acid residues at the carboxyl terminus are cysteine or histidine, suggesting that this segment of the polypeptide could form a metal-binding domain; however, the sequence and arrangement of amino acids in this region is somewhat different from that of zinc-finger proteins (32, 33).

The polypeptide does not have significant amino acid sequence similarity to any other protein described in the databanks; however, there may be some similarity between the predicted Tc4v polypeptide and a polypeptide hypothetically encoded by the Tc5 transposable element (J.Collins, pers. commun.).

The Tc4v cDNA sequence differs from the Tc4-*rh1030* genomic sequence at four positions. The cDNA has three single base base pair substitutions of A for G corresponding to positions 930, 936 and 948 in the genomic sequence (Fig. 1A). These changes are all located in the 5′ UTR of the message. An additional difference occurred in exon 3 at positions 1837−1858 of the genomic sequence. A segment of 22 nucleotides in the Tc4-*rh1030* DNA (ACAAAACGAAATGAAGTTGCAA) was replaced by CTTC in the cDNA insert, leading to a net change of 18 bp, which does not alter the reading frame. Two cDNA clones isolated from two independent N2 cDNA libraries extended through this region, and they both contained the same nucleotide variation. Extending through this region in the Tc4-*rh1030* sequence, starting at position 1819, are two 19-bp perfect direct repeats separated by 5 bp (underlined in Fig. 1A). The cDNAs are missing parts of both repeats and the intervening five nucleotides. It seems likely that the cDNAs were derived from an N2 message that was copied from a Tc4v element different from Tc4-*rh1030*. Indeed, Southern blots of *Eco*RV-digested DNA probed with the 2.1-kbp *Eco*RV suggested the presence of a doublet band that would be explained by the existence of two classes of Tc4v elements differing by 18 bp (data not shown).

Tc4v transcripts are more abundant in the mutator strain TR679 than in N2

Northern blots of poly(A)-enriched RNA prepared from mixed-stage populations of N2 and TR679 were probed with the 2.1-kbp *Eco*RV fragment unique to Tc4v elements. A 1.6-kb Tc4v RNA was detected in the mutator strain TR679 in which Tc4 elements are mobile; however, no transcript from N2 was apparent (Fig. 3). A probe that hybridizes to *act-1* demonstrated that approximately equal amounts of N2 and TR679 RNA were present on the blot. The Tc4-*n1416* element was also used to probe Northern blots. A few faint Tc4-hybridizing bands in N2 as well as TR679 were detected (data not shown); none of these transcripts was the size of the 1.6-kb Tc4v transcript detected in TR679.

Tc4-*rh1030* insertion in the *unc-33* gene results in the splicing out of Tc4 sequences from mutant *unc-33* transcripts

We previously showed that the wild-type *unc-33* gene produces three messages, which have different 5′ ends, common 3′ ends, and identical reading frames (13). The 3.5-kbp Tc4-*rh1030* and the 1.6-kbp Tc4-*mn260* insertions are at identical sites within an exon that is common to the two larger transcripts but at a position about 13 nucleotides upstream of the transcriptional start site of the smallest wild-type *unc-33* message (Fig. 4A). In Northern blot analysis with RNA prepared from *unc-33(rh1030)* and from *unc-33(mn260)*, we observed at least two very large *unc-33* transcripts from each mutant, corresponding to the inclusion of the 3.5-kbp Tc4-*rh1030* or the 1.6-kbp Tc4-*mn260* sequences. However, the predominant transcripts, almost as abundant as those of N2, were close to the three wild-type sizes (13). Southern blots indicated that somatic excision of Tc4-*rh1030* or Tc4-*mn260* from the genome was undetectable and hence could not account for the wild-type-size transcripts. Tc4-*rh1030* was inserted into the *unc-33* gene such that transcription through the Tc4v element was on the opposite strand from the known Tc4v message; therefore transcriptional termination and polyadenylation at the known polyadenylation site within Tc4v could not account for the wild-type-size messages in *unc-33(rh1030)*.

The wild-type-size messages in *unc-33(mn260)* were shown to arise from splicing out of the Tc4-*mn260* sequences from the transcripts by either cis- or trans-splicing (13). Wild-type-size transcripts that arose from cis-splicing were missing the entire Tc4-*mn260* sequence plus 28 nucleotides immediately 5′ of the site of Tc4-*mn260* insertion. Splicing of these transcripts appeared to use a cryptic 5′ splice donor site within an *unc-33* exon and a 3′ splice acceptor site at the downstream terminus of Tc4-*mn260*. Wild-type-size transcripts that arose from trans-splicing contained the SL1 sequence, a 22-nucleotide trans-spliced leader on many *C.elegans* messages (29), spliced to sequences immediately 3′ of the Tc4-*mn260* insertion. In this case the downstream terminus of Tc4 appeared to act as a 3′ trans-splice acceptor site.

To elucidate the nature of the wild-type-size transcripts in *unc-33(rh1030)*, we have now analyzed poly(A)-enriched RNA isolated from the mutants by PCR amplification of reverse transcription products (RT-PCR). The results revealed that one class of transcripts, presumably corresponding to the two larger wild-type sizes, was generated by the cis-splicing out of all or nearly all of the Tc4-*rh1030* sequence. Of the 15 RT-PCR clones we sequenced that extend through the region in which Tc4-*rh1030* is situated, 12 had the Tc4-*rh1030* insert completely removed along with 28 nucleotides of *unc-33* exon sequence immediately upstream of the Tc4-*rh1030* insertion site (Fig. 4B), as was observed for *unc-33(mn260::Tc4)* (13). Inspection of the sequence at the 5′ splice site indicates that ACG|GUGAGAG, located upstream of the Tc4v element, normally within an exon of *unc-33*, was a cryptic 5′ splice donor and that the downstream terminus of the Tc4, CCCTAG, provided the 3′ splice acceptor site (CCCUAG|U). In two of the other three clones analyzed, all of the Tc4-*rh1030* sequence except for 48 nucleotides from the upstream terminus was spliced out: a cryptic 5′ splice donor, AUU|GUAUAG, was used, and the 3′ terminus of the element again provided the 3′ splice acceptor site (Fig. 4B). The final RT-PCR clone extending through the region made use of the same 5′ splice donor as the first 12 but used a non-consensus 3′ splice acceptor, UUUUUG|A, to leave 33 nucleotides of downstream Tc4 sequence in the spliced message (Fig. 4B).

A second class of mutant transcripts was identified by RT-PCR when the SL1 sequence was used as 5′ primer. The 12 RT-PCR clones that were sequenced had their 5′ terminal SL1 sequences spliced to sequences slightly upstream of the 5′ beginning of the smallest wild-type message (the smallest wild-type *unc-33* message does not carry SL1). In 10 of the 12 clones analyzed, an SL1 sequence was followed by 13 extra nucleotides in the *unc-33* exon immediately downstream of Tc4-*rh1030* (Fig. 4C), as was observed for *unc-33(mn260::Tc4)* transcripts (13). In the other two clones, 30 or 33 nucleotides from the 3′ end of the Tc4-*rh1030* sequence were present in addition to the 13 extra *unc-33* nucleotides. We presume that transcription of these mutant transcripts began within the transposon, which has apparently separated the promoter of the smallest wild-type transcript from the normal site of transcriptional initiation. In ten clones, the 3′ terminus of Tc4-*rh1030* was used as a 3′ trans-splice acceptor site. The other two splicing events used cryptic non-consensus trans-splice acceptors, UUGAAU|A or UUUUUG|A, upstream of the 3′ end of the Tc4-*rh1030* insertion (Fig. 4C).

DISCUSSION

The distinguishing property of the Tc4v element described here is the presence of a novel 2343-bp sequence that replaces a 447-bp sequence in the middle of a long inverted terminal repeat of the previously-studied Tc4 elements. This alteration results in a variant element that is 3483-bp long and distinct from the 1605-bp Tc4 elements. [Because of the sequences at the insertion sites of all analyzed Tc4 elements there is some ambiguity in the sizes of these elements. All elements have inserted so that the trinucleotide TNA of their target site is duplicated and flanks Tc4. It is possible that the element itself begins with A and ends with T and that only the N of the target site is duplicated; therefore, Tc4 elements could each be 2 bp longer.] There appear to be five copies of Tc4v per haploid genome of N2 and TR679. The members of the Tc4v subfamily appear to be fairly homogeneous in structure but we have detected some differences among these elements. We estimate that the Tc4v subfamily comprises approximately one-fifth and one-sixth of the Tc4 elements found in strains N2 and TR679, respectively.

The evolutionary relationship between Tc4 and Tc4v elements is not at all clear; however, there is no evidence that the novel 2343-bp sequence of Tc4v is itself a transposable element that inserted into an inverted repeat of a Tc4 element. The disrupted right inverted terminal repeat of Tc4-*rh1030* has two segments of similarity to the left repeat, a 158-bp proximal segment and a 139-bp distal segment. Among a total of eight left-right differences over these 297 bp, two were also present in Tc4-*n1416*. This finding suggests that these two asymmetries predate the divergence of the two Tc4 elements. The other six asymmetries were not present in Tc4-*n1416*, which had two left-right differences that were not present in Tc4-*rh1030*. The element-specific left-right differences could have arisen after the divergence of the two elements or, alternatively, a pre-existing asymmetry might have been removed by gene conversion in one of the elements. It may be relevant that both of the left-right differences that are common to the two elements are near the internal loop region, where they might be less subject to gene conversion.

Using the novel Tc4v sequence as probe, we recovered three overlapping cDNAs that correspond to a 1.9-kb mRNA with substantial protein-coding capacity: a long open reading frame capable of encoding a polypeptide of 537 amino acid residues was found, as well as an appropriate polyadenylation signal and poly(A) tail. The coding region was located wholly within the novel Tc4v sequence. Previously-identified Tc4 elements appear to contain no protein-coding capacity, although they are capable of transposition. It seems likely that their transposition depends upon transposase produced by other elements. Thus, it seems possible that Tc4v elements encode a Tc4-specific transposase. The predicted novel polypeptide produced by Tc4v shows no similarity to any other protein present in sequence databanks; however, it may have similarity to a potential Tc5 polypeptide (J.Collins, pers. commun.). The predicted Tc4v polypeptide contains a high percentage of basic amino acid residues, having a predicted pI of 8.83. An overall positive charge for the Tc4v polypeptide is consistent with the idea that it may interact with Tc4 DNA to promote transposition.

Tc4 transposition appears to be activated by the *mut-2* mutation carried by strain TR679. The mechanism of this activation is not clear. However, it may be relevant to Tc4 transposition that Tc4v-specific transcripts were readily detected in TR679 but not in N2. Thus we find a correlation of activation of Tc4 transposition with enhanced Tc4v transcript levels. Although we detected no Tc4v transcripts on Northern blots of N2 RNA, the cDNAs that we used to define a Tc4v-specific message were recovered from N2 libraries, suggesting some level of Tc4v-specific transcription in N2.

Further investigation of Tc4v transcripts from TR679 may prove interesting. The size of the observed transcript(s) from TR679, about 1.6 kb, was at least 300 nucleotides shorter than that expected from the cDNAs recovered from N2 libraries. It is thus possible that the cDNA sequence from N2 misrepresents the structures of the transcripts produced in TR679. The 5′ end of the cDNA sequence lay within the right repeat (Fig. 1). We do not know where this transcript was initiated; no conserved promoter sequences are apparent in the inverted repeats of Tc4v. It is even possible that the Tc4v mRNA in N2 might have been initiated by an extragenic promoter located upstream of a particular Tc4v member. Inspection of the Tc4v sequence reveals a consensus promoter site, including a CAT box and TATA box, within the Tc4v novel sequence; the TATA box, starting at position 1294, is 0.37 kbp downstream of the 5′ end of the cDNA sequence. It is possible that this site may serve as a good promoter in TR679 but not N2.

For many transposons, inverted terminal repeats are necessary for transposition to occur, although it is not clear what role the lengthy inverted repeats of foldback elements play in transposition. The Tc4-*rh1030* element was able to transpose into the *unc-33* gene and is also able to excise from *unc-33* (13) despite having inverted terminal repeats that are only 158 bp long. Yuan et al. (1) noted the presence of four copies of a 7-bp repeat, C-TGAAAT, in each of the sub-terminal regions of Tc4-*n1416*. The left terminal inverted repeat of Tc4-*rh1030* also has four copies of the 7-bp repeat; the right inverted repeat has three copies of the repeat, with one of the three abutting the Tc4v novel sequence. Whether this repeat is involved in transposition remains to be seen.

We previously showed that *unc-33(rh1030)* mutants have predominantly wild-type-size *unc-33* mRNAs despite having a 3.5-kbp Tc4 within an *unc-33* exon (13). We have shown here that we can account for the wild-type size messages by the splicing out of all or nearly all of the Tc4-*rh1030* sequence. The splice sites that are used in removing Tc4 sequences do not match perfectly the *C.elegans* consensus splice donor (A/GAG|GUAAGUU) or splice acceptor (UUUCAG|G/A) sequences; however, all 5′ splice donor sites that were identified do share the obligatory GU of the consensus sequence. This was not the case for some of the 3′ splice acceptor sites that were observed. Although 24 of the 27 RT-PCR clones that were analyzed share the normally obligatory AG consensus of 3′ splice sites, three did not (the exceptions being UG, in two independent clones, and AU). We have observed splicing at non-AG 3′ splice sites for both cis- and trans-splicing. This use of splice acceptor sites that lack the supposedly obligatory AG of eukaryotic introns is very unusual but has been observed before in *C.elegans* mutants whose splice acceptor AG sequence had been mutated (Aroian, R.V., Levy, A., Koga, M., Ohshima, Y., Kramer, J., and Sternberg, P., personal communication). In these cases introns were also spliced at non-AG splice acceptor sites located at or very near the position of the normal splice acceptor. This finding suggests that the internal sequence or structure of *C.elegans* introns strongly influences splicing events. We further suggest, then, that the sequence or structure of Tc4 and Tc4v elements is recognized

as an intron by the *C.elegans* splicing machinery. In maize, Ds transposable elements inserted in exons of the waxy gene can be spliced out of transcripts. The splice sites that are used can be near the ends of the Ds element, either within the Ds element or within the waxy gene (34). This has led to the idea that transposons able to behave as introns would have a strong selective advantage because they would be able to insert into genes with little deleterious effect on gene expression. The splicing we have seen does not precisely excise Tc4 sequence nor does it usually restore wild-type function to a mutant message; however, we have observed the behavior of Tc4 inserted at only a single genomic location chosen because it did confer a mutant phenotype. It seems likely that Tc4's ability to behave as an intron would reduce the deleterious effects of its insertion into some genes. It may be noteworthy that the mutant phenotypes of *unc-33(mn260::Tc4)* and *unc-33(rh1030::Tc4v)* are less severe than that of a null mutation in *unc-33* (13).

ACKNOWLEDGMENTS

We thank E.Hedgecock for *unc-33(rh1030)*, J.Yuan and R.Horvitz for Tc4-*n1416*, J.Ahringer, J.Kimble, R.Barstead, and R.Waterston for cDNA libraries, and M.Krause for pT7/T3-18-103. We thank J.Collins for sharing unpublished data with us and we thank S.Gantt for help with the manuscript. W.L. was supported by NIH grant GM22387 to R.K.Herman. This research was supported by NIH grant HD22163 to J.E.S.

REFERENCES

1. Yuan, J., Finney, M., Tsung, N. and Horvitz, H. R. (1991) *Proc. Natl. Acad. Sci USA*, **88**, 3343–3338.
2. Truett, M.A., Jones, R.S. and Potter, S.S. (1981) *Cell*, **24**, 753–763.
3. Potter, S.S. (1982) *Nature*, **297**, 201–204.
4. Liebermann, D., Liebermann, B.H., Weinthal, J., Childs, G., Maxson, R., Mauron, A., Cohen, S.N. and Kedes, L. (1983) *Nature*, **306**, 342–347.
5. Dreyfus, D.H. and Emmons, S.W. (1991) *Nucleic Acids Res.*, **19**, 1871–1877.
6. Emmons, S.W. Yesner, L., Ruan, K.-S. and Katzenberg, D. (1983) *Cell*, **32**, 55–65.
7. Rosenzweig, B., Liao, L.W. and Hirsh, D. (1983) *Nucleic Acids Res.*, **11**, 4201–4209.
8. Levitt, A. and Emmons, S. W. (1989) *Proc. Natl. Acad. Sci USA*, **86**, 3232–3236.
9. Ruvolo, V., Hill, J.E. and Levitt, A. (1992) *DNA Cell Biol.*, **11**, 111–122.
10. Collins, J., Forbes, E. and Anderson, P. (1989) *Genetics*, **121**, 47–55.
11. Collins, J., Saari, B. and Anderson, P. (1987) *Nature*, **328**, 727–728.
12. Finney, M., Ruvkun, G. and Horvitz, H.R. (1988) *Cell*, **55**, 757–769.
13. Li, W., Herman, R.K. and Shaw, J.E. (1992) *Genetics*, **132**, 675–689.
14. Moerman, D.G. and Waterston, R.W (1989) In Berg, D.E. and Howe, M.M. (eds.) Mobile DNA, American Society for Microbiology, Washington, D.C., pp. 537–556.
15. Engels, W.R. (1989) In Berg, D.E. and Howe, M.M. (eds.) Mobile DNA, American Society for Microbiology, Washington, D.C., pp. 437–484.
16. Brierly, H.L. and Potter, S.S. (1985) *Nucleic Acids Res*, **13**, 485–500.
17. Templeton, N.S. and Potter, S.S. (1989) *EMBO J.*, **8**, 1887–1894.
18. Brenner, S. (1974) *Genetics*, **77**, 71–94.
19. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
20. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning, Ed. 2., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
21. Vieira, J. and Messing, J. (1987) *Methods Enzymol.*, **153**, 3–11.
22. Feinberg, A.P. and Vogelstein, B. (1983) *Anal. Biochem.*, **132**, 6–13.
23. Sanger, R., Nicklen, S. and Coulsen, A.R. (1977) *Proc. Natl. Acad. Sci. USA*. **74**, 5463–5467.
24. Henikoff, S.(1987) *Methods Enzymol.*, **155**, 156–157.
25. Sheen, J.-Y. and Seed, B. (1988) *BioTechniques*, **6**, 942–944.
26. Emmons, S.W., Klass, M.R. and Hirsh, D. (1979) *Proc. Natl. Acad. Sci USA*, **76**, 1333–1337.
27. Jacobson, A. (1987) *Methods Enzymol.*, **152**, 254–261.
28. Krause, M., Wild, M., Rosenzweig, B. and Hirsh, D. (1989) *J. Mol. Biol.*, **208**, 381–392.
29. Krause, M. and Hirsh, D. (1987) *Cell*, **49**, 735–761.
30. Atschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) *J. Mol. Biol.*, **215**, 403–410.
31. Emmons, S. W. (1988) In Wood, W.B. (ed.), The Nematode *Caenorhabditis elegans*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp. 47–79.
32. Klug, A. and Rhodes, D. (1987) *Trends Biochem. Sci.* **12**, 464–469.
33. Berg, J.M. (1990) *Ann. Rev. Biophys. Biophys. Chem.*, **19**, 405–421.
34. Wessler, S. R. (1991) *Mol. Cell. Biol.*, **11**, 6192–6196.
35. Files, J.G. and Hirsh, D. (1981) *J. Mol. Biol.*, **149**, 223–240.