Nucleic Acids Research, 1994, Vol. 22, No. 21 4535–4542

Developmentally regulated, low abundance Tec element transcripts in *Euplotes crassus*—implications for DNA elimination and transposition

John W.Jaraczewski, John S.Frels and Carolyn L.Jahn*
Department of Cell and Molecular Biology, Northwestern University Medical School, 303 E. Chicago Avenue, Chicago, IL 60611, USA

Received March 25, 1994; Revised and Accepted August 30, 1994

ABSTRACT

During macronuclear development in the ciliated protozoan, *Euplotes crassus*, >105 Tec elements are precisely eliminated from the genome in a 2–4 h time interval, generating extrachromosomal circular forms of the elements. Various models have proposed a transposition-based mechanism for this excision. We have tested this hypothesis by determining the abundance of transcripts of Tec element open reading frames (ORFs) and the timing of their appearance. Transcripts are very low in abundance and are only detected by PCR amplification techniques. Thus, the low levels of transcripts argue against the participation of element-encoded functions in the Tec element elimination process. The element transcripts are only detected in RNA samples from mated cells, indicating that the micronucleus and/or developing macronucleus are transcriptionally active during the sexual phase of the life cycle. The transcription detected could allow a low level of germline-specific transposition for these elements.

INTRODUCTION

In ciliated protozoa, genome-wide, developmentally coordinated DNA rearrangements occur during the formation of a transcriptionally active macronucleus from a micronucleus (1,2). The DNA sequence differences between micronuclei and macronuclei are the most dramatic in the 'hypotrichous' ciliates such as *Oxytricha*, *Stylonychia* and *Euplotes* (reviewed in 3,4). The macronucleus contains 5–10% of the sequence information present in the micronucleus organized as small, gene-sized, linear DNA molecules and amplified 1000- to 10,000-fold. Some of the micronuclear sequences that are eliminated during formation of the macronuclear molecules reside internal to the macronuclear-destined sequences, hence they are referred to as Internal Eliminated Sequences (IESs) (reviewed in 1,2). Because many of these IESs interrupt coding regions, their precise elimination is critical for the construction of active genes. Two classes of IESs have been defined. One class is comprised of short, unique sequences. The other is comprised of transposon-like repetitive elements.

In *Euplotes crassus* two different, but related, tranposon-like repetitive element families, Tec1 and Tec2, are eliminated during macronuclear development (5–8). As many as one-third of the 30,000 elements in each family interrupt macronuclear-destined sequences as IESs. Elements inserted within elements form very large interruptions of macronuclear-destined sequences. We have demonstrated that early in macronuclear development, while micronuclear chromosomes are undergoing polytenization, the Tec elements are excised *en masse* to form extrachromosomal circular forms (7,8). Due to the copy number increase during polytenization, 10^5-10^6 Tec elements per nucleus are excised in a discrete 2–4 h time period of macronuclear development. A similar excision and circle formation during macronuclear development has been described for the TBE transposon-like elements of another hypotrichous ciliate *Oxytricha trifallax* (9).

Although we do not have any direct means to demonstrate that Tec elements are transposons, several lines of evidence suggest that they are. First is their structure. The Tec1 and Tec2 elements are 5.3 kb in size with large inverted repeats (700 bp) at their ends (7,8,10). They are surrounded by a TA direct repeat, one copy of which is present in the macronuclear gene that is the product of excision (5,8). This suggests that the TA is a target site duplication created when the element inserted, with excision precisely reconstituting the target. The open reading frames (ORFs) identified in the sequenced members of the Tec1 and Tec2 elements are very similar in organization for the two element families (10). Comparison of ORFs between members of each family indicates that each ORF within a family has diverged by ~10%, whereas the ORF amino acid sequences from the Tec1 family differ from the corresponding ORFs of the Tec2 family by 30–80%. Sequence comparisons between Tec element ORFs and *O.fallax* TBE element ORFs demonstrate sequence similarities between an 1140 bp ORF from Tec1 (ORF1, Fig. 1), a homologous 1149 bp ORF from Tec2 and a 1062 bp ORF from TBE (11). Furthermore, these related ORFs have sequence

*To whom correspondence should be addressed

similarity to transposases from a wide range of transposable elements (11). The structure of the Tec elements and the presence of an ORF with similarity to known transposases taken together with the high copy number and extensive dispersion throughout the *E.crassus* genome as well as other Euplotid genomes indicates that they are *bona fide* transposons.

Our initial hypothesis upon finding the concerted excision of Tec elements was that excision might involve element-encoded transposition functions (7,8). The possibility that excision was related to transposition was based on the occurrence of extrachromosomal circular forms of transposable elements in other organisms and their association with transposition (12–16). Transposons that utilize a conservative, cut and paste type mechanism of transposition have been shown to produce double-strand breaks at the ends of the element to release a free transposon intermediate (17). It seemed possible that the extrachromosomal Tec elements represent excised transposition intermediates that are prevented from reinsertion (7). This possibility is supported by *in vitro* assays with Tn10 and IS911 transposases (15,16,18,19). High amounts of these transposases result in circular excision products that appear to be generated by transposition of the element into itself. Thus, mechanisms involving high amounts of transposase have been invoked to explain transposon excision during macronuclear development (9). Suggestions of an altered transposase or an interaction of transposase with sequences other than transposon ends (i.e. unique sequence IESs) via interactions with host-encoded proteins have been hypothesized to explain the similarities between Tec element and unique sequence IES excision in *E.crassus* (7,20–23).

Several aspects of Tec element excision differ from the transposition of other elements shown to produce circular extrachromosomal elements in other organisms (12–16). First is the precision of excision (22). This contrasts with the predominately imprecise excision of other transposons (24). Second is the concerted nature of the process, whereby the majority of the $>10^5$ copies of each element present in the polytenized chromosomes excise in a 2–4 h period (7,8). Because excision of Tec elements and formation of circular products occurs at an amount per cell that is $>10^5$ higher than that observed for other transposons, we expected that transcripts of Tec element ORFs should be readily detected at a developmentally appropriate time if the element-encoded functions were involved in the elimination process. We summarize below our results on the detection of Tec element transcripts.

MATERIALS AND METHODS

Cell culture and mating

E.crassus lines 8, 24, B6 and G12 were fed algae and bacteria and mated and harvested as described previously (8,10,25). Developmental timepoints are described in hours post-mixing of starved cultures of two different mating types. For the RNA preparations spanning 6 h time intervals (1–6, 7–12, 13–18 and 19–24) cultures of two different mating types (lines 8 and 24) were mixed at 1 h intervals over a 6 h time period. At 6 h intervals corresponding to 7, 13, 19 and 25 h after the first mating, equal aliquots of each of the 6 matings were combined and harvested to represent each of the intervals.

RNA preparation and Northern blotting

Total RNA was prepared by guanidine–isothiocyanate lysis and sedimentation through CsCl (26) with DNase I treatment as described by Price *et al.* (27). RNA preparations were further purified through a Qiagen tip-100 column (Qiagen Inc., Chatsworth, CA) prior to cDNA synthesis. RNA was denatured with glyoxal–DMSO and electrophoresed in agarose gels in 0.01 M NaH_2PO_4 (pH 7.0) (26). Northern blots using non-charged Qiabrane nylon membranes (Qiagen Inc.) were prepared according to the manufacturer's suggested protocols using capillary transfer and 20× SSC (1× SSC is 0.15 M NaCl and 0.015 M Na_3citrate). Hybridizations were carried out using a heat-treated milk–formamide mixture (28). Post-hybridization washes utilized 0.1× SSC, 0.5% SDS at 65°C. Radioactive hybridization probes for Northerns were synthesized using [^{32}P]CTP and T3 or T7 RNA polymerase with templates cloned in pBS or pKS (Stratagene, La Jolla, CA). Quantitation of hybridization utilized a Fujix BAS 2000 Bio-Imaging Analyzer (Fuji Film Co. Ltd, Tokyo, Japan) and associated data analysis computer programs.

RACE and RT–PCR reactions

RACE (rapid amplification of cDNA ends) reactions were carried out essentially as described by Frohman *et al.* (29). The cDNAs were synthesized using the Superscript Kit (BRL Inc., Gaithersburg, MD) essentially as described in the kit protocol for first strand synthesis utilizing 1 μg of total RNA and 10 pmol of the poly-T adapter primer. The cDNA synthesis included an initial DNase I treatment prior to reverse transcription. All developmental timepoint cDNA preparations compared in Fig. 2 were made simultaneously with all reaction components other than the RNA aliquoted from one pre-mix. For the PCR reactions, cDNA equivalent to 25 ng of total RNA was amplified using 25 pmol of both the upstream adapter primer (UAP) and a 5′ gene-specific primer (GSP) for RACE reactions or with 10 pmol of the 3′ end GSP for RT–PCRs (reverse transcriptase–polymerase chain reactions). PCR was carried out with the hot-start technique with Ampliwax Beads (Perkin-Elmer Cetus, Norwalk, CT) as described previously (30). The reactions were analyzed by electrophoresis in 1.5% agarose gels, which were Southern blotted to Zeta-bind (BioRad, Richmond, CA) using 0.4 N NaOH, after which the blot was UV crosslinked in a Stratagene Stratalinker and hybridized with ORF-specific probes labeled by nick-translation or random-hexamer labeling as described previously (6–8).

Oligonucleotide primers

The sequences of the primers used in the RT–PCR and RACE reactions are derived from the published Tec1-1 and Tec2-1 sequences (10) (GenBank numbers L03359 and L03360, respectively), but were chosen to match other versions of the elements that have been sequenced. The listing in Table 1 gives the primer name, sequence and the nucleotide position of the first (5′ end) and last (3′ end) base relative to the Tec1-1 or Tec2-1 sequences. Note that when the position of the first base is a higher number than the last, the primer corresponds to the opposite strand than that listed in GenBank. The sequence of the poly-T adapter primer used for all cDNA synthesis was 5′ GACTCG-AGTCGACATCGATTTTTTTTTTTTTTT 3′ and the sequence of the adapter primer used in the RACE reactions was

5′ GACTCGAGTCGACATCG 3′ (29). The primer for actin RACE reactions was 5′ CGAACGCCAACAAAAGA 3′. Histone H4 primers were 5′ GCCAAGAGACACGCCAAGAAG 3′ and 5′ CTTGTCTCTTGAGAGCGTAGAC 3′, the former being the primer used in histone H4 RACE reactions.

RESULTS

Detection of Tec element transcripts

In *E.crassus*, the differentiation of a macronucleus from a mitotic copy of the micronucleus begins 18–20 h after mating and is initiated by mixing of starved cultures of two different mating types (25,27; Fig. 5). Thus, time of mixing is taken as time zero and all other developmental times are referred to as hours post-mixing. Since excision of Tec elements occurs between 26 and 30 h post-mixing (7,8), we expected to see active transcription prior to or concurrent with this time interval if Tec element gene products are involved in excision. Thus we analysed total RNA and PolyA$^+$ RNA prepared from the following cell populations or developmental stages: vegetative cells, either fed or starved; and mated cells at time intervals of 1–6, 7–12, 13–18, 19–24 h post-mixing and at 24, 26, 28, and 30 h post-mixing. The pooled samples (i.e. 1–6, 7–12 h, etc.) were made by combining 6 samples of cells from matings done at 1 h intervals. We have used several methods to look for Tec1 and Tec2 transcripts, including hybridization of DNA or RNA probes to Northern blots, RNase protection, primer extension and cDNA PCR amplification. The most sensitive Northern analyses we performed, utilizing high specific activity antisense RNA probes and Northern blots of polyA$^+$ RNA showed only faint smears of hybridization with no discrete bands after 7–10 day exposures of the autoradiograms. Our radioactive probes detected low copy numbers of Tec elements in dilutions of total DNA. In addition, transcripts of other known genes were readily detectable in our Northern blots with no evidence of RNA degradation. Thus, we are confident that if discrete transcripts of Tec ORFs were present, they were rare in abundance and below the sensitivity of detection afforded by Northern blots. Although this lack of detection by Northerns already indicated that transcript abundance was not high enough to account for a transposition-based mechanism, we pursued further analyses of Tec transcripts using PCR methods. Whereas hybridization of radioactive probes to Northern (or Southern) blots has a lower limit of detection of 10^5 molecules, PCR methods are capable of detecting single molecules in a huge excess of a complex mixture of DNA sequences (31,32). Using PCR methods, we have defined developmental time periods when Tec transcripts accumulate and have demonstrated that the transcripts are specific to the Tec element ORFs and not the entire element. The faint smears that we detect in Northern blots may represent partial transcripts of the ORFs.

Two methods of PCR amplification of cDNAs were used to detect Tec element transcripts: (i) a RACE (rapid amplification of cDNA ends) protocol (29) where cDNA synthesis with reverse transcriptase uses an oligo-dT adapter primer and the subsequent PCR utilizes one gene-specific primer (GSP) and a universal adapter primer that corresponds to the oligo-dT adapter primer sequence; (ii) a reverse transcriptase PCR protocol where the oligo-dT adapter primer from the RACE protocol was used in the reverse transcriptase reaction and the subsequent PCR utilized two GSP (31,33). Because of the sensitivity of these methods, special precaution was taken to remove DNA from the RNA preparations (see Materials and Methods). Because of the high copy number of Tec elements (30,000 of each family per micronuclear genome equivalent) even small amounts of DNA contamination could interfere with the detection of transcripts. As described below, several controls were performed to determine the amount of DNA contamination in each sample.

Although numerous copies of Tec elements exist in the *E.crassus* micronuclear genome, it is possible that only a small subset of each element family is capable of expression. Moreover, the element ORFs that we have sequenced may differ from those that are transcribed. To address this concern, we have previously compared the sequences of the ORFs from several members of each family (10). Overall the ORFs differ at 10% of the nucleotide positions. However alignment of the sequences indicates that certain regions are identical among all copies of the elements that have been sequenced. To optimize the match to all possible Tec elements, all of the GSP were chosen to correspond to these highly conserved sequences. In some cases, sequences of Tec1 and Tec2 were similar enough that the same primer could be used for either family (primers 41 and 63 in Fig. 1). All combinations of GSP used (Fig. 1) gave equivalent yields of products in reactions with 1 pg of total *E.crassus* genomic DNA (corresponding to ~0.02 pg of micronuclear DNA, or ~500 copies of each element) indicating that the combinations of primers we are using have similar annealing properties under the PCR conditions used. Sequence analysis of PCR products amplified from genomic DNA with these primers indicates that the variation in sequence is equivalent to that

Table 1. Tec element oligonucleotide primers

Oligo	Tec1/2	ORF	Start[a]	End[a]
2	1	IR	316	293[b]
7	1	IR	1	28
12	2	2B	3562	3541
13	2	3	4604	4580
36	2	2B	2953	2977
38	2	2B	3384	3405
39	1	2	3490	3513
40	2	3	4337	4363
41	1	1	1106	1086
	2	1	1120	1100
43	1	1	743	743
47	1	2	2552	2535
60	2	1	824	842
61	2	2A	2773	2751
62	2	2B	3208	3188
63	1	3	4328	4348
	2	3	4343	4363
76	1	2	2198	2217
102	2	2A	2446	2466
129	1	1–2[c]	1894	1910
130	1	1–2	2164	2143
131	1	2–3	4073	4050

[a]Start (5′ end) and end (3′) nucleotide positions are given relative to the complete sequences of Tec1-1 and Tec2-1 (GenBank numbers, L03359 and L03360, respectively) (10).

[b]Nucleotide positions that are numerically higher for the start than for the end are complementary to the strand reported in GenBank for which the numbers read 1 = 5′.

[c]Primers from regions that lie outside an ORF are designated according to the ORFs that they lie between.

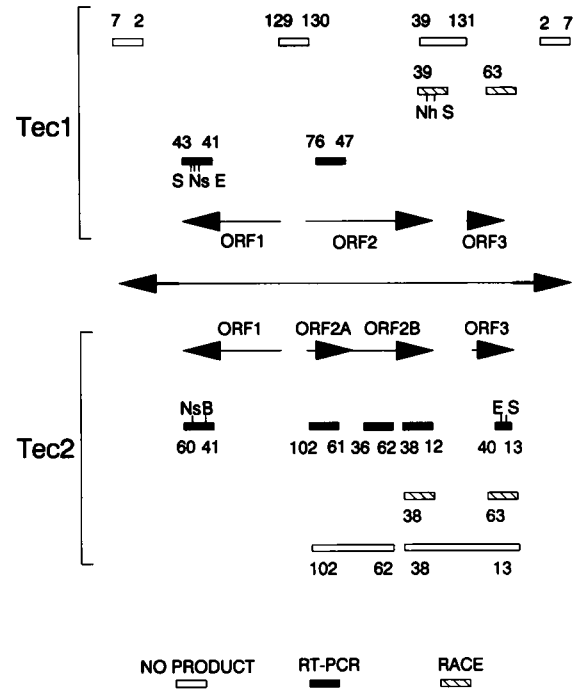

Figure 1. Identification of transcripts using reverse transcriptase PCR (RT−PCR) and rapid amplification of cDNA ends (RACE). The numbered open reading frames of both Tec1 and Tec2 are shown as arrows in their relative locations either above or below (respectively) the structure of Tec elements, where large arrows represent the inverted repeats located at the ends of the element. ORFs bearing the same numbers, and in the same relative position in Tec1 and Tec2 are related in sequence. RT−PCR and RACE products identified by electrophoresis on agarose gels, Southern blotting and hybridization with Tec1 or Tec2 specific probes (see Fig. 2) are shown as black boxes and cross-hatched boxes, respectively. Open boxes represent products that are not detected in RNA. Numbers above or below the boxes signify the primers used for each reaction. Lines extending upwards or downwards from the boxes with letters associated identify restriction sites shown to exist within the products which correspond to conserved sites in Tec elements. Restriction enzyme abbreviations are as follows: S, *Sau*3A; B, *Bgl*II; Ns, *Nsi*I, E, *Eco*RI; Nh, *Nhe*I.

obtained when randomly selected genomic clones are sequenced (10% mismatch). Thus, the primers anneal to a heterogeneous mixture of elements. The GSP have also been used in comparisons of Tec elements from other species and, again, appear to amplify a heterogeneous mixture of Tec sequences. Thus, we believe they are capable of detecting transcripts that represent the diversity of sequences present in Tec elements in *E.crassus*.

In order to detect Tec element transcripts, the Taq polymerase amplified products from the reverse transcriptase reactions were electrophoresed in agarose gels, Southern blotted and hybridized with radioactive probes specific for each of the ORFs. Two types of control amplification reactions were performed to determine whether DNA contamination was present. A mock cDNA reaction without the addition of reverse transcriptase to the cDNA synthesis was amplified in parallel with the reverse transcriptase cDNA reactions in order to determine the extent of DNA contamination of the RNA samples. In addition, all PCR analyses included controls utilizing identical reaction components without added cDNA to check for other sources of DNA contamination. All of the results that we present are based on reactions where this control reaction showed no detectable products. To facilitate comparison of PCR reactions on cDNAs made from different RNA samples, the cDNAs were synthesized in parallel with identical reaction components on equal amounts of RNA. PCR reactions with 1 pg of total *E.crassus* DNA were also performed in parallel as a positive control. As seen in Figs 1 and 2, products that are reverse transcriptase dependent are observed for all of the Tec1 and Tec2 ORFs. In the RACE reactions no hybridizing products were ever detected in the RT-minus reactions with any of the developmental time course RNA samples. In addition, no hybridizing products were obtained from genomic DNA or plasmid DNAs with combinations of one GSP and either the RACE oligo-dT or adapter primers, indicating that DNA contamination would not be detected (or interfere) in the RACE reactions. Thus, in subsequent RACE reactions with the RNA time course samples (shown in Fig. 2A) amplifications of the RT-minus reactions were omitted. However, with RT−PCR, where two GSP are used, some of the samples gave detectable products in the RT-minus reactions (Fig. 2B), indicating detectable amounts of DNA. Thus in order to use the RT−PCR technique to determine whether Tec element transcripts were present in these RNA samples it was essential to compare the plus and minus RT reactions for each RNA sample.

As seen in Fig. 2, Tec transcripts are only detected in RNA samples from mated cells. They are most abundant in the 1−6, 19−24 and 26 h samples. Transcripts for all of the Tec1 and Tec2 ORFs are detected in these three samples and not in samples from starved cells (0 h, Fig. 2) or vegetative, fed cells (not shown). Tec-specific transcripts are variably detected in the 7-12 and 13−18 h samples. ORF1-specific RT−PCR reactions for Tec1 show considerably higher amounts of PCR products in the RT-plus reaction than the RT-minus reactions for the 7−12 and 13−18 h samples (Fig. 2B); although the amounts are lower than that seen for the 1−6, 19−24 and 26 h samples. Tec2 ORF1 and ORF2A RT−PCRs are positive for the 13−18 h sample, but again, the amounts of PCR products are lower than that seen for the 1−6, 19−24 and 26 h samples. In general, no Tec-specific products were detected in RACE reactions with the 7−12 and 13−18 h samples (Fig. 2A). To verify that no Tec transcripts are detectable in vegetative cells we performed the Tec1 ORF2, ORF3 and Tec2 ORF3 reactions on 2 other samples from fed cells and 4 samples from starved cells and have never detected an RT-dependent product that hybridizes with the ORF-specific probes.

We have established the specificity of these transcripts as follows. First, Tec-specific RT−PCR products are the sizes expected from the DNA sequences and all RACE products were ~50−100 bp longer than the distance from the primer to the ORF termination. In addition, these products hybridized only with probes specific to the ORF from which the primers were derived and hybridization was always greater than that observed with primers alone. Second, we have digested several of the products with restriction enzymes (designated in Fig. 1) that are known to cut a majority of the copies of an element family and all or most of the products yield the expected fragment sizes. For instance, ORF1 of both Tec1 and Tec2 possess a conserved *Nsi*I site that was present in their respective RT−PCR products. Third, we have tried combinations of primers from within the inverted repeats or spanning between two open reading frames (Fig. 1 primer pairs 7,2; 129,130; 39,131; 102,62; and 38,13). None of these primer sets gave detectable products from RNA. This indicates that DNA contamination is not the source of the products we detect since these other regions would be equally

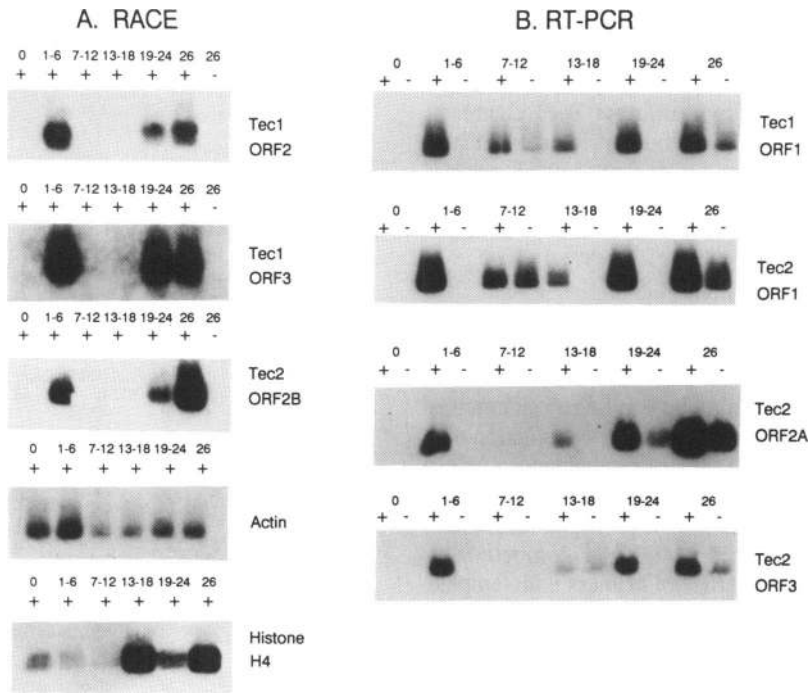

Figure 2. PCR products from Tec ORFs produced from RNA samples derived from mated cells. The results of hybridization of Tec ORF or gene-specific probes to Southern blots of agarose gels used to analyze the PCR reactions with Tec ORF-specific primers and actin or histone gene-specific primers are shown for RNA samples from starved cells and from the early stages of mating and macronuclear development. The RACE reactions (**A**) were electrophoresed and blotted without corresponding RT-minus reactions because all such control reactions were negative. In the RT−PCR reactions (**B**), RT-minus and -plus reactions were analysed in parallel so that the contribution of product from DNA contamination could be determined by comparison of the plus and minus RT lanes.

represented in DNA. In addition, 'read through' transcription from promoters flanking Tec elements is not responsible for the products we detect, as this would lead to inverted repeat transcripts. Fourth, we can readily detect PCR products for transcripts from several known genes (actin and histone H4). RACE products are obtained for both the actin and histone transcripts in the 7−12 and 13−18 h samples, indicating that the lower amounts of Tec RT−PCR products and lack of Tec-specific RACE products from these samples is not due to differences in the cDNA preparations. The high amounts of histone H4 RACE products obtained from the 13−18 and 26 h samples corresponds to high amounts of transcripts detected in Northern blots of these samples (see Fig. 4). The somewhat higher amounts of actin RACE products in the 0 and 1−6 h samples may also correspond to higher amounts of transcripts detected early in development by Northern analyses (27), although we have not examined our starved (0 h) or 1−6 h samples by Northern analysis with actin probes. The range of transcript abundance detectable in Northerns for histone H4 is less apparent in the PCR reactions (see below). We believe this is due to saturation of PCR for samples with high amounts of transcripts (34). The RACE reactions for histone H4 required fewer cycles than are used for analysis of Tec transcripts. In addition, with several of the cDNA samples, dilution of the cDNA by 1000-fold relative to the amount used in the reactions with Tec-specific primers resulted in detectable histone PCR products with fewer than 30 amplification cycles. When the amount of cDNA used in the histone H4 reaction is the same as that used for the Tec-specific reactions, even the reactions that show much lower amounts of products in Fig. 2 yield products that can be visualized by ethidium bromide staining of a gel rather than hybridization to a Southern blot of the gel.

Tec1 ORF2 cDNA 20 *TAA*CTCTTAATTCATACATA*TAGATTATGTTAAAAAAAAAAAAAAAAAAAA

Tec1 ORF2 cDNA 25 *TAA*CTC*ATAATTCATACATAAAGATTATGTTAAAAAAAAAAAAAAAAAAAAAAAA

Tec1-1 ORF2 *TAA* CTCTTAATTCATACATAAAGATTATGTTAAATTTTATTTTCATTAATTTTATTTT

Figure 3. Sequence of 3′ cDNA clones from the OP39 primer RACE reaction. Shown are the 3′ untranslated sequences and polyA tails of two cloned cDNAs obtained from RACE reactions using the OP39 primer (39 in Fig. 1) compared to the corresponding Tec1-1 sequence. The TAA stop at the end of ORF2 is italicized and the base mismatches in the cDNAs relative to the Tec1-1 sequence are indicated by asterisks. Clone 25, which contains a 200 bp insert, contains 9 additional mismatches in the 160 bases of translated sequence contained in the cDNA. The exact location of polyadenylation cannot be determined because of the three adenines in the Tec1-1 sequence at the position of the ends of the transcripts.

Taken together, the above data indicate that our RNA and cDNA preparations are intact and representative of the overall RNA population with minimal DNA contamination. Because of the low abundance of the Tec transcripts that we are detecting we cannot easily determine whether they are discrete in length and representative of a translatable mRNA. However, we have cloned PCR products of a RACE reaction for the 3′ end of ORF2 from Tec1 and sequenced two of the cloned products (Fig. 3). Comparison of the sequences with that of the Tec1-1 element shows that the polyA sequences begin 29 base pairs after the TAA stop codon that terminates ORF2. This is similar to what has been observed for transcripts from known *E.crassus* genes, where

no polyA addition consensus sequence could be defined in this interval (30). Each ORF2 RACE product contains mismatched bases relative to the Tec1-1 ORF2 sequence, thus both products represent unique transcripts from two different Tec1 elements. Since no combination of the RACE oligo-dT primer or adapter primer with a GSP yielded PCR products from total *E.crassus* DNA or clones of the appropriate ORF under the PCR conditions used, we are confident that these represent the 3′ ends of transcripts.

Estimating transcript abundance

How abundant are the transcripts we are detecting? We have examined the sensitivity of our RT−PCR reactions by comparing the amplification of cDNA to the amplification of known amounts of total genomic DNA. Although this method of quantitation may not be as accurate as comparison to an internal standard or use of a competitor (35), the comparison to genomic DNA allows us to amplify a heterogeous mixture of Tec sequences, which we would not be able to mimic by adding a competitor or standard. Since the annealing properties of primers with template are major determinants of the PCR efficiency (32−34), we reasoned that for the purposes of quantitation, the primers should show similar annealing with a mixture of total cDNA sequences as they do to the mixture of sequences present in total DNA. To determine whether this method of quantiation was realistic, we used the same procedure to compare the amounts of H4 histone transcripts that we detect in the cDNA samples used for quantitation of the Tec transcripts. This allowed us to quantitate histone H4 transcripts by both Northern blot analysis and PCR comparison to total DNA. The RNA samples used to make the cDNAs were Northern blotted and hybridized with an H4 histone antisense RNA probe (Fig. 4A) and the extent of hybridization was quantitated using a phosphoimager. The range of transcript abundance seen in Fig. 4A is similar to that seen with all of the Northern blots analyzed. The H4 histone transcripts detected in total RNA samples from starved cells are 30−35 times less abundant than those detected in 26 or 28 h total RNA samples, which showed the greatest hybridization. At the exposure times used for quantitation of the starved and 26 or 28 h samples, histone H4 transcripts are not detectable above background in the 1−6 h total RNA sample (Fig. 4A). Visualization of a band in the 1−6 h total RNA sample requires a 10-fold increase in exposure time for the autoradiogram. Thus, we reasoned that comparison of the 1−6 h and 26 h cDNA samples by RT−PCR with histone H4 primers should show differences of 10^2-10^3 in amounts of histone H4 transcripts.

To quantitate the amounts of PCR products, we analyzed the reactions by Southern blotting and hybridization and quantitated the hybridization using the phosphoimager. This indicated that reactions with total DNA at or below 10 pg of total DNA showed a linear increase in the hybridization as a function of the logarithmic increase in DNA quantity. Thus, below 10 pg, it was possible to graph the hybridization results for the reactions with total DNA and directly determine a relative concentration for the amount of hybridization seen for the cDNA reactions. For histone H4 (Fig. 4B) the amount of PCR product obtained from the 1−6 h cDNA sample (25 ng total RNA equivalent) is similar to that obtained from ~5 pg of total DNA. A 10^{-3} dilution of the 26 h cDNA sample (25 pg total RNA equivalent) yields a similar amount of product to the 25 ng equivalent of 1−6 h cDNA. Thus, this comparison appeared to detect the range of abundance of histone H4 transcripts that we detect in our Northern blots. For Tec1 ORF1 (Fig. 4C), the amount of product obtained from cDNA equivalent to 25 ng of total RNA is similar to that obtained from 0.1 pg of total DNA. The amounts of PCR products quantitated with reactions for Tec2 ORF1, Tec1 ORF2 and Tec2 ORF3 (not shown) were nearly identical to the amounts for Tec1 ORF1 in the 1−6 h and 26 h cDNA samples. Because all of our RACE and RT−PCR results included a positive control with 1 pg of total DNA, we are certain that all of the reactions that yielded products (Fig. 1) were similar in the quantities of products made and thus are representative of similar amounts of transcripts.

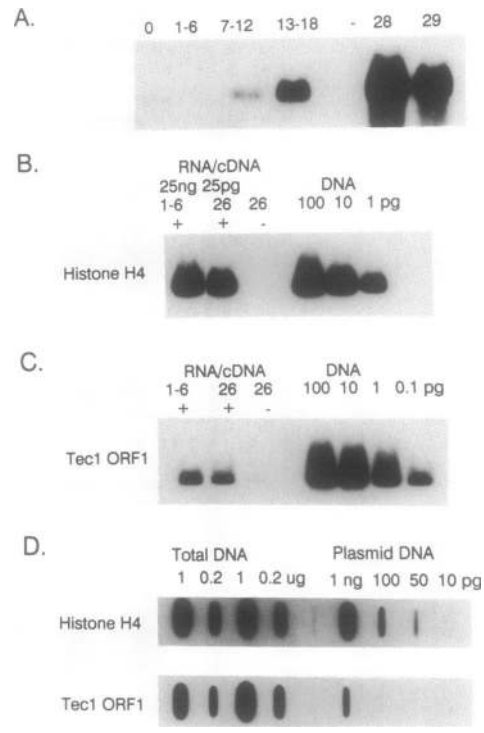

Figure 4. Northern analysis of histone H4 transcripts and quantitation of PCR products. (**A**) Total RNA from *E.crassus* at various time points post-mixing of cells of opposite mating type were Northern blotted and hybridized with a labeled antisense RNA probe for histone H4. The samples labelled '0, 1−6, 7−12, and 13−18' correspond to the samples used to make cDNA for the PCR analyses in Fig. 2. (**B**) PCR with histone H4-specific primers (sequences are given in Materials and Methods) of cDNAs from 1−6 and 26 h RNA samples were compared to PCR of total genomic DNA at different concentrations. For the 26 h cDNA sample, PCR was carried out on a 10^{-3} dilution of the cDNA relative to that used for the 1−6 h sample (cDNA equivalent to 25 ng of total RNA). (**C**) Equal amounts (equivalent to 25 ng of total RNA) of the 1−6 h cDNA sample and 26 h cDNA sample were amplified with primers from Tec1 ORF1 (41 and 43 in Fig. 1) alongside reactions with total genomic DNA at different concentrations. (**D**) The hybridization of histone H4 and Tec1 ORF1 probes to total genomic DNA was compared to known concentrations of plasmids carrying each sequence in order to determine their relative concentration in genomic DNA. The plasmids were used as the hybridization probes.

In order to equate the above PCR comparisons to known quantities of total DNA with a known number of molecules, we determined the copy number of Tec elements and histone H4 sequences in total DNA relative to a dilution series of plasmid clones by slot-blot hybridization (Fig. 4D). Although these quantitations have been made numerous times with different Tec element probes (5−8), the copy number of histone H4 molecules was unknown. The slot-blots (Fig. 4D) indicate that the copy number per macronucleus of H4 histone molecules is 20,000

molecules. These copy number determinations indicate that in cDNA from 25 ng of total RNA we are detecting Tec ORF1 sequences equivalent to 70 molecules. Histone H4 mRNA levels in cDNA from 25 ng of total RNA are equivalent to 2000 molecules in the 1–6 h sample and 2×10^6 molecules in the 26 h sample. These calculations assume a 1:1 conversion of the RNA to cDNA. Thus, if the efficiency of conversion were low, our cDNA samples might represent fewer molecules than are present in 25 ng of RNA. We can estimate the lowest reverse transcription efficiency possible by taking into consideration the amount of histone H4 detected in the 26 h sample. Quantitation of total RNA and polyA$^+$ RNA yields from known numbers of *E.crassus* indicates 100 pg of total RNA and 1 pg (2×10^6 molecules 10^3 nucleotides in size) of polyA$^+$ RNA per cell (30 and this work). Since the histone H4 transcripts are unlikely to comprise more than 10% of the polyA$^+$ RNA (2×10^5 molecules), the 2×10^6 molecules we detect are representative of a maximum of 10 pg of polyA$^+$ RNA or 1 ng of total RNA. This is 25-fold higher than we predict from a 1:1 conversion of RNA to cDNA. Based on this assessment of efficiency of conversion of RNA to cDNA, the 70 molecules of Tec1 ORF1 mRNA we detected could represent a maximum of 7 molecules per cell. It should be noted that the $\sim3\times10^4$-fold difference in amounts of H4 histone transcript and Tec transcripts in the 26 h sample is similar to what we would predict from our Northern blots. The H4 transcripts are detected in total RNA samples at 1 h exposures of autoradiograms while the Tec transcripts are detected in polyA$^+$ RNA samples (as a faint smear) in 7–10 day exposures. The 10^2-fold difference in concentrations between total and polyA$^+$ RNA and the 10^2-fold difference in exposure time amounts to an $\sim10^4$ difference in transcript abundance.

DISCUSSION

The identification of Tec element transcripts described above addresses several issues. First, is the question of whether Tec elements are active as transposable elements. Second, is the general problem of how transposable elements that are restricted to an inactive micronucleus can be expressed. Third, is the possible role of element-encoded functions in elimination of Tecs and IESs during macronuclear development.

Our results indicate that Tec element transcripts are present in very low amounts during the sexual phase of the *E.crassus* life cycle. These low abundance Tec element transcripts may be sufficient for a low frequency of transposition, although we do not as yet have evidence that discrete, translatable Tec element mRNAs are made. Further analysis of cloned 5′ and 3′ RACE products may allow us to identify discrete transcripts that include intact ORFs. We have made antibodies to fusion proteins for regions of ORF2 and ORF3 of Tec1 and have not detected any Tec-specific proteins in Western blots of *E.crassus* total proteins from various developmental timepoints. However, given the low transcript abundance, detection of the Tec element-encoded proteins may be difficult.

The developmental timing of the appearance of detectable Tec element transcripts is suggestive of transcriptional activity in the pre-meiotic micronucleus and in the anlagen. The timing of appearance of Tec transcripts corresponds well with transcriptional activation of micronuclei and developing macronuclei in *Tetrahymena*, where micronuclei in prophase of meiosis and early anlagen have been shown to incorporate [^{3}H]uridine (37,38). Studies of TATA-binding protein and micronuclear- versus macronuclear-specific histone localizations in *Tetrahymena* (39–43) indicate that transcriptional repression of micronuclei is reversed prior to meiosis, thus, the premeiotic micronucleus and early anlagen are capable of some transcription.

An alternative explanation for the origin of Tec element transcripts is that they arise in the macronucleus. This possibility is difficult to rule out. Tec elements have not been detected in macronuclear DNA. However, the high copy number in the micronucleus makes it difficult to be certain that a very low number of Tecs does not exist in the macronuclei. The very low abundance of transcripts that we detect could arise from a copy number of Tecs in the macronucleus of less than one per cell. Nevertheless, if this were true, Tec element transcript abundance must be regulated such that trancripts do not accumulate in vegetative cells or during meiosis in conjugating cells.

The developmental timing of the appearance of transcripts is not restricted to the time interval when the Tec elements are excised. Although our studies only address the amounts of Tec element transcripts that accumulate during development and not the gene products, it seems unlikely that the low abundance of transcripts detected would result in the accumulation of sufficient amounts of the proteins to carry out the massive amounts of Tec excision seen at 26–30 h post-mixing. We estimate a maximum of 7 molecules per cell or 0.0004% of the polyA$^+$ RNA. The transcript abundance observed is similar to that seen for transposable elements where less than one transposition event occurs per cell (44–47). For instance, transcripts of Ac elements comprise 0.002–0.013% of the polyA$^+$ RNA (1.5–10 mRNA molecules/cell) and allow a transposition rate of 1 per generation per 0.2–1% of all cells (47). Similarly, the P element transcripts in the Pc[ry]2 strain, which carries a single Pc[ry] element, comprise 0.001% of the polyA$^+$ RNA and result in transposition in 2.4% of the germ cells exposed per generation (45,46). The amount of Tec excision per cell is 10^6–10^7 higher.

Our data lead us to a view of Tec excision that does not involve element-encoded proteins and suggests the possibility of a general mechanism for DNA elimination that is similar in all ciliates, rather than the accumulation of numerous different site-specific transposition-like mechanisms that would be required if transposon-encoded functions are utilized. A 'host-encoded' system fits with the high similarity of circular excision products for Tecs and unique sequence IESs in *E.crassus*, which do not share any sequences that could provide site-specificity, but appear to be excised by related machinery (7,20,21,23).

Studies of reversion of Tn*5*- and Tn*10*-induced insertion mutations have demonstrated the involvement of 'host-encoded' functions without element-encoded functions in transposon excision in *Escherichia coli* and *Saccharomyces cerevisiae* (48–56). Precise and 'nearly precise' excision of Tn*10* is mechanistically unrelated to the excision of the element from the donor locus during transposition (49), which appears to result in either loss of the donor locus or reconstitution of the parental locus via double-strand break repair (54). Excision of Tn*5* and Tn*10* is dependent on the presence of the large inverted repeats (48–50) and the frequency is greatly increased when the elements are present as single-stranded DNA (51). Thus, one model of excision involves replication slippage due to interaction of the inverted repeats (48,50,53). Mutations affecting replication, recombination and repair (including *recB*, *recC*, *mutH*, *mutL*, *mutS*, and *ssb* in *E.coli* and *pol2* and *pol3* in *S.cerevisiae*) can increase the frequency of these excision events (50,52–56). These processes may be involved in *E.crassus* Tec element and

unique sequence IES excision, but, at present, it is unclear how they might act in such a concerted fashion. Again, the differences in frequencies are mind-boggling. Despite the above studies, the major determinants of large inverted repeat stability or instability are largely unknown. Ultimately, identification of the host functions that govern Tec element excision in *E.crassus* is likely to uncover processes important to genome stability in all organisms.

ACKNOWLEDGEMENTS

The authors thank Dr Dorothy Shippen for critical reading of the manuscript. This work was supported by Public Health Service grant GM37661 to C.L.J.

REFERENCES

1. Yao, M.-C. (1989) In Berg, D.E. and Howe, M.M. (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 715–734.
2. Klobutcher, L.A. and Jahn, C.L. (1991) *Curr. Opin. Gen. Devel.*, **1**, 397–403.
3. Klobutcher, L.A. and Prescott, D.M. (1987) In Gall, J. (ed.), *The Molecular Biology of Ciliated Protozoa*. Academic Press Inc., New York, NY, pp. 111–154.
4. Jahn, C.L. (1991) *J. Protozool.*, **38**, 252–258.
5. Baird, S.E., Fino, G.M., Tausta, S.L. and Klobutcher, L.A. (1989) *Mol. Cell. Biol.*, **9**, 3793–3807.
6. Jahn, C.L., Nilles, L.A. and Krikau, M.F. (1988) *J. Protozool.*, **35**, 590–601.
7. Jahn, C.L., Krikau, M.F. and Shyman, S. (1989) *Cell*, **59**, 1009–1018.
8. Krikau, M.F. and Jahn, C.L. (1991) *Mol. Cell. Biol.*, **11**, 4751–4759.
9. Williams, K., Doak, T.G. and Herrick, G. (1993) *EMBO J.*, **12**, 4593–4601.
10. Jahn, C.L., Doktor, S.Z., Frels, J.S., Jaraczewski, J.W. and Krikau, M.F. (1993) *Gene*, **133**, 71–78.
11. Doak, T.G., Doerder, F.P., Jahn, C.L. and Herrick, G. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 942–946.
12. Sundaresan, V. and Freeling, M. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 4924–4928.
13. Ruan, K.-S. and Emmons, S.W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4018–4022.
14. Rose, A.M. and Snutch, T.P. (1984) *Nature*, **311**, 485–486.
15. Morisato, D. and Kleckner, N. (1984) *Cell*, **39**, 181–190.
16. Morisato, D. and Kleckner, N. (1987) *Cell*, **51**, 101–111.
17. Mizuuchi, K. (1992) *Ann. Rev. Biochem.*, **61**, 1011–1051.
18. Benjamin, H.W. and Kleckner, N. (1989) *Cell*, **59**, 373–383.
19. Polard, P., Prere, M.F., Fayet, O. and Chandler, M. (1992) *EMBO J.*, **11**, 5079–5090.
20. Jaraczewski, J.W. and Jahn, C.L. (1993) *Genes Dev.*, **7**, 95–105.
21. Tausta, S.L. and Klobutcher, L.A. (1989) *Cell*, **59**, 1019–1026.
22. Tausta, S.L., Turner, L.R., Buckley, L.K. and Klobutcher, L.A. (1991) *Nucleic Acids Res.*, **19**, 3229–3236.
23. Klobutcher, L.A., Turner, L.R. and LaPlante, J. (1993) *Genes Dev.*, **7**, 84–94.
24. Berg, D.E. and Howe, M.M., (eds), (1989) *Mobile DNA*. American Society for Microbiology, Washington, DC.
25. Roth, M.R., Lin, M.-Y. and Prescott, D.M. (1985) *J. Cell Biol.*, **101**, 79–84.
26. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
27. Price, C.M., Adams, A.K. and Vermeesch., J.R. (1994) *J. Eukaryot. Microbiol.*, **41**, 267–275.
28. Monstein, H.-J., Geiger, T. and Bakalkin, G.Y. (1992) *BioTechniques*, **13**, 842–844.
29. Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 8998–9003.
30. Ghosh, S., Jaraczewski, J.W., Klobutcher, L.A. and Jahn, C.L. (1994) *Nucleic Acids Res.*, **22**, 214–221.
31. Kawasaki, E.S. (1990) In Innis, M.A., Gelfand, M.A., Sninsky, J.J. and White, T.J. (eds), *PCR Protocols: A Guide to Methods and Applications*. Academic Press Inc., San Diego, CA, pp. 21–27.
32. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) *Science*, **239**, 487–491.
33. Rappolee, D.A., Wang, A., Mark, D., and Werb, Z. (1989) *J. Cell. Biochem.*, **39**, 1–11.
34. Innis, M.A. and Gelfand, D.H. (1990) In Innis, M.A., Gelfand, M.A., Sninsky, J.J. and White, T.J. (eds), *PCR Protocols: A Guide to Methods and Applications*. Academic Press Inc., San Diego, CA, pp. 1–12.
35. Foley, K.P., Leonard, M.W. and Engel, J.D. (1993) *Trends Genet.*, **9**, 380–385.
36. Luporini, P. and Dini, F. (1975) *J. Protozool.*, **22**, 541–544.
37. Martindale, D.W., Allis, C.D. and Bruns, P.J. (1985) *J. Protozool.*, **32**, 644–649.
38. Sugai, T. and Hiwatashi, K. (1974) *J. Protozool.*, **21**, 542–548.
39. Stargell, L.A. and Gorovsky, M.A. (1994) *Mol. Cell. Biol.*, **14**, 723–734.
40. Stargell, L.A., Bowen, J., Dadd, C.A., Dedon, P.C., Davis, M., Cook, R.G., Allis, C.D. and Gorovsky, M.A. (1993) *Genes Dev.*, **7**, 2641–2651.
41. Chicoine, L.G., Wenkert, D.W., Richman, R., Wiggins, J.C. and Allis, C.D. (1984) *Dev. Biol.*, **109**, 1–8.
42. Wenkert, D. and Allis, C.D. (1984) *J. Cell Biol.*, **98**, 2107–2117.
43. Allis, C.D. and Wiggins, J.C. (1984) *Dev. Biol.*, **101**, 282–294.
44. Raleigh, E.A. and Kleckner, N. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1787–1791.
45. Karess, R.E. and Rubin, G.M. (1984) *Cell*, **38**, 135–146.
46. Laski, F.A., Rio, D.C. and Rubin, G.M. (1986) *Cell*, **44**, 7–19.
47. Fusswinkel, H., Schein, S., Courage, U., Starlinger, P. and Kunze, R. (1991) *Mol. Gen. Genet.*, **225**, 186–192.
48. Egner, C. and Berg, D.E. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 459–463.
49. Foster, T.J., Lundblad, V., Hanley-Way, S., Halling, S.M. and Kleckner, N. (1981) *Cell*, **23**, 215–227.
50. Collins, J., Volckaert, G. and Nevers, P. (1982) *Gene*, **19**, 139–146.
51. Berg, D.E., Egner, C. and Lowe, J.B. (1983) *Gene*, **22**, 1–7.
52. Lundblad, V. and Kleckner, N. (1984) *Genetics*, **109**, 3–19.
53. Lundblad, V., Taylor, A.F., Smith, G.R. and Kleckner, N. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 824–828.
54. Gordenin, D.A., Lobachev, K.S., Degtyareva, N.P., Malkova, A.L., Perkins, E. and Resnick, M.A. (1993) *Mol. Cell. Biol.*, **13**, 5315–5322.
55. Gordenin, D.A., Malkova, A.L., Peterzen, A., Kulikov, V.N., Pavlov, Y.I., Perkins, E. and Resnick, M.A. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 3785–3789.