4230–4240 *Nucleic Acids Research, 1998, Vol. 26, No. 18* *© 1998 Oxford University Press*

Conserved DNA sequences adjacent to chromosome fragmentation and telomere addition sites in *Euplotes crassus*

Lawrence A. Klobutcher*, Scott E. Gygax, Jed D. Podoloff, Joris R. Vermeesch[1,+], Carolyn M. Price[1], Christopher M. Tebeau[2] and Carolyn L. Jahn[2]

Department of Biochemistry, University of Connecticut Health Center, Farmington, CT 06030, USA, [1]Department of Chemistry and Department of Biochemistry, University of Nebraska, Lincoln, NE 68588, USA and [2]Department of Cell and Molecular Biology, Northwestern University Medical School, Chicago, IL 60611, USA

Received May 13, 1998; Revised and Accepted July 23, 1998

DDBJ/EMBL/GenBank accession nos[§]

ABSTRACT

During the formation of a new macronucleus in the ciliate *Euplotes crassus*, micronuclear chromosomes are reproducibly broken at ~10 000 sites. This chromosome fragmentation process is tightly coupled with *de novo* telomere synthesis by the telomerase ribonucleoprotein complex, generating short linear macronuclear DNA molecules. In this study, the sequences of 58 macronuclear DNA termini and eight regions of the micronuclear genome containing chromosome fragmentation/telomere addition sites were determined. Through a statistically based analysis of these data, along with previously published sequences, we have defined a 10 bp conserved sequence element (E-Cbs, 5′-HATTGAAaHH-3′, H = A, C or T) near chromosome fragmentation sites. The E-Cbs typically resides within the DNA destined to form a macronuclear DNA molecule, but can also reside within flanking micronuclear DNA that is eliminated during macronuclear development. The location of the E-Cbs in macronuclear-destined versus flanking micronuclear DNA leads us to propose a model of chromosome fragmentation that involves a 6 bp staggered cut in the chromosome. The identification of adjacent macronuclear-destined sequences that overlap by 6 bp provides support for the model. Finally, our data provide evidence that telomerase is able to differentiate between newly generated ends that contain partial telomeric repeats and those that do not *in vivo*.

INTRODUCTION

All ciliated protozoa that have been examined at the molecular level have been found to undergo developmentally controlled chromosome fragmentation. This is possible, in part, because of the nuclear dimorphism displayed by this group of organisms (reviewed in 1–3), in which cells contain both micronuclei and macronuclei. The micronucleus contains conventional chromosomes, but is transcriptionally inert during asexual reproduction. The second nucleus, the macronucleus, is responsible for transcription during vegetative growth. It possesses a polyploid genome of subchromosomal DNA molecules that represent a subset of the micronuclear genome. In the hypotrich group (e.g. *Euplotes*, *Oxytricha* and *Stylonychia*) the macronuclear DNA molecules are linear, very short (~2 kb on average) and generally contain single coding regions. The ends of the macronuclear DNA molecules are capped by telomeres, which consist of 5′-CCCCAAAA-3′/3′-GGGGTTTT-5′ (C_4A_4 or G_4T_4) repeats in the hypotrichs (4).

The macronucleus is destroyed during sexual reproduction and regenerated from a copy of the micronuclear genome. The process begins with cell pairing (conjugation), followed by micronuclear meiosis. Haploid products are exchanged between the paired cells and fuse with a resident haploid nucleus to form the zygotic nucleus. The zygotic nucleus then undergoes one or more mitotic divisions. Some division products are retained as micronuclei, while others undergo a DNA rearrangement process that culminates in the formation of a new macronucleus. In hypotrichs, macronuclear development begins with multiple rounds of DNA replication that result in the formation of polytene chromosomes. During this period, thousands of interstitial DNA segments, termed IESs (internal eliminated sequences), are excised from the chromosomes with rejoining of flanking DNA. IESs consist of both short (<600 bp) unique DNA sequences and large families of transposable elements (Tec elements) in *Euplotes crassus* (2,5). Once the polytene chromosomes have fully formed, they are fragmented to generate the individual macronuclear DNA molecules. Telomeric repeat sequences are added to the ends of the liberated DNA molecules by telomerase (6). In *E.crassus*, the telomeres are initially synthesized in an extended form (7–9) and subsequently trimmed later in development (10). The final stages

*To whom correspondence should be addressed. Tel: +1 860 679 2816; Fax: +1 860 679 3408; Email: butcher@panda.uchc.edu

+Present address: Center of Human Genetics, VIB, Flanders Interuniversity Institute of Biotechnology, Campus Gasthuisberg O7N6, Herestraat 49, B-3000 Leuven, Belgium

§AF061108–AF061112, AF061334, AF063084, AF068575–AF068614, AF069914–AF069917, AF069919–AF069931, AF077371, AF072706–AF072708

of macronuclear development involve additional rounds of DNA replication that result in a mature macronucleus.

The chromosome fragmentation/telomere addition process has been studied at the molecular level in only a few ciliate species. It is generally a very reproducible process, but most ciliates that have been examined display some heterogeneity with regard to the exact position at which telomeric repeats are added (reviewed in 1). The sole known exception is *E.crassus*, where chromosome fragmentation/telomere addition reproducibly occurs at the same nucleotide position (11). Sequence elements controlling chromosome fragmentation/telomere addition have been best characterized in *Tetrahymena thermophila*. A 15 bp sequence termed the Cbs (chromosome breakage sequence) resides in the developmentally eliminated DNA flanking macronuclear-destined sequences (12) and has been shown to be necessary and sufficient to direct chromosome fragmentation/telomere addition (13). However, the Cbs element is found only in *Tetrahymena* and closely related species (14). Candidate *cis*-acting sequences have not been identified for most other ciliates, perhaps because of the observed heterogeneity in the process.

For *E.crassus*, previous studies have suggested a candidate sequence element that may direct chromosome fragmentation/telomere addition. First, bulk sequencing of total macronuclear DNA termini using the chemical sequencing method has provided evidence that many macronuclear DNA molecules share the sequence element 5′-TTGAA-3′ beginning at base 18 following the telomeric repeats (4,15). Sequencing of individual macronuclear DNA molecules has confirmed this, but has also shown that a significant number of termini lack the sequence element (11,16,17). However, sequencing of a limited number of micronuclear chromosome fragmentation sites (i.e. the junctions between macronuclear-destined and micronuclear-limited DNA) suggests that if the 5′-TTGAA-3′ element is absent in the macronuclear-destined DNA, it is then found in the flanking micronuclear-limited DNA (11). In these instances, the sequence element is in an inverted orientation and located 12 bp from the position at which telomeric sequences will ultimately be added. A model was proposed to reconcile the different spacing of the sequence element when in micronuclear-limited as opposed to macronuclear-destined DNA (11; Fig. 6A). Briefly, the model suggests that the 5′-TTGAA-3′ sequence serves to direct chromosome fragmentation/telomere addition to a defined position 5′ of itself and results in a 6 bp staggered cut in the DNA. The 6 bp staggered cut is then filled-in prior to or during the process of telomere addition. In support of this model, an adjacent pair of macronuclear-destined sequences has been characterized that overlap by 6 bp (18).

In the current study we more thoroughly examine sequences near chromosome fragmentation sites. A large sample of macronuclear DNA termini has been analyzed and a more extensive conserved sequence element containing the 5′-TTGAA-3′ core has been identified. Additional micronuclear fragmentation sites have also been examined to test the hypothesis that the sequence element may be positioned in either macronuclear-destined or flanking micronuclear-limited DNA. The results of these analyses provide new insights into the chromosome fragmentation/telomere addition process in *E.crassus* and lead us to suggest a modified model of the process.

MATERIALS AND METHODS

Cell culture and DNA isolation

Euplotes crassus strains X1, X2, ST9 and ST11 were grown in artificial seawater prepared using Reef Crystals (Aquarium Systems, Mentor, OH) and supplemented as described in Roth *et al.* (19), except that vitamin B12 was omitted. Either the alga *Dunaliella salina* or the bacterium *Escherichia coli* were used as the food source. Matings were carried out at 24°C using strains X1 and X2 as previously described (8,19).

Euplotes crassus vegetative total cellular DNA was isolated using established procedures (4). Micronuclear DNA used in PCR reactions was obtained by subjecting total vegetative cellular DNA to electrophoresis in a low melting point agarose gel (Gibco BRL Life Technologies, Gaithersburg, MD). Limit mobility DNA was excised from the gel and used directly for PCR.

Developing macronuclei were isolated and DNA purified as previously described (8,19), except that 3 mM $CaCl_2$ and 50 mM NaCl were used in place of spermidine in nuclear isolation buffers.

Southern hybridizations and other molecular biological methods

Southern blots of agarose gels were prepared using GeneScreen Plus membrane filters (NEN Research Products, Boston, MA) and the alkaline blotting method (20). Hybridizations and washes were carried out as previously described (21,22) using restriction fragments or PCR products labeled with ^{32}P by the random hexamer priming method (23) as probes.

Random clones of macronuclear DNA molecules were isolated from previously described plasmid recombinant libraries (10). Isolation of the micronuclear EC1 clone and identification of the Mac C and Mac J macronuclear-destined sequences has been described (24). Restriction fragments derived from EC1 were used to screen a recombinant macronuclear DNA library (24) in order to isolate clones of the Mac C and Mac J macronuclear DNA molecules.

Other molecular biological methods were performed using established procedures (20).

DNA sequencing

For sequencing PCR products, the reactions were first run on low melting point agarose gels. The specific PCR products were excised from the gels and the DNA purified as described (25). The DNAs were then directly sequenced using the same primers employed in PCR. Sequencing was carried out by the University of Connecticut Health Center Molecular Core facility using the Taq Dideoxy Termination Cycle Sequencing Kit (Perkin Elmer Cetus, Norwalk, CT).

Sequencing of the cloned DNA segments was carried out using Sequenase kits (US Biochemicals, Cleveland, OH). For sequencing of the EC1 micronuclear clone and the Mac J and Mac C macronuclear clones, selected restriction fragments were subcloned into pBS or pKS plasmids (Stratagene, La Jolla, CA). Synthetic oligonucleotide primers (Gibco BRL Life Technologies Inc.) were then used in some cases to extend sequences determined from the vector priming sites.

Analysis of DNA sequences

Conserved sequence elements in macronuclear DNA termini and micronuclear junctions were defined using a previously described statistical method (26). Briefly, sequences in the sample were aligned relative to a reference point, which in this case was either the first subtelomeric base or the chromosome fragmentation/telomere addition site. The overall frequencies of each base (f_A, f_T, f_C and f_G), as well as the overall frequencies of each combination of three bases (f_{ATC}, f_{TCG}, etc.), were then calculated for the entire sample. The average expected occurrence (expected mean = M) for each base or combination of three bases at each position was next calculated as nf_B, where n is the number of sequences in the sample. The expected standard deviation was calculated as the square root of $nf_B(1 - f_B)$. Positions where the observed occurrence of a base or a combination of three bases exceeded M + 2 SD were considered to be significant.

The sample of macronuclear DNA termini analyzed included the newly sequenced macronuclear DNA ends noted above. In addition, the sample included sequence data from the following previously sequenced termini of macronuclear DNA molecules, most of which contain genes of known function, that were available in GenBank (accession nos are given in parentheses): telomerase RNA (M33461); *TLEE* (M73025), phosphoglycerate kinase (U97355); V3 (U47679), *conZA8* (AF061334); actin (J04533); *MACET3* (M63336); γ-tubulin 1 (X85234); γ-tubulin 2 (X85235); ribosomal protein L29 (U13207); histone H3 (U75429); histone H4 (U75430); telomere binding protein (M96818); telomere binding protein homolog (M96819); V2 (M28500); β-tubulin (J04534); V1 (M28500); D7 (M28498); D8 (M28498 and M28499); *conF5* (AF063084); *conZB1* (U65646). A compilation of the sequence data used in the analysis is available on request.

Using the above method, some single positions displayed non-random base composition for both a single base and a combination of three bases. For the consensus sequences, we have chosen to display significant single bases in the consensus when their frequency is >0.60 in the sample and the three base combination when the frequency of the single base is <0.60. Also included in the consensus are positions within or adjacent to conserved clusters, where a single base was present at a frequency of >0.5. These positions are indicated in lower case.

Polymerase chain reaction

PCR was carried out using either a mixture of Taq polymerase (Gibco BRL Life Technologies Inc.) and Vent polymerase (New England Biolabs Inc., Beverly, MA) or KlenTaq polymerase (Sigma, St Louis, MO), using buffers and reaction conditions supplied by the manufacturers. All reactions were carried out in a MJ Research PTC-150 MiniCycler (Cambridge, MA).

For inverse PCR (27) reactions to isolate sequences flanking the micronuclear *TelRNA* and *RPL29* genes, DNA isolated from developing macronuclei 40 h after mating was used as a substrate. The 40 h DNA was digested with *Hin*dIII and then ligated at a concentration of 2.5 ng/µl to foster intramolecular circle formation. Aliquots of 12.5 ng ligated DNA were then used as the substrate in 50 µl PCR reactions.

For inverse PCR of the *TelRNA* gene, the first step PCR reaction used primers ETEL-1 (5′-CATTTTGAATCCCTGACCTC-3′) and ETEL-2 (5′-CTAAACGCACCTATTGATCG-3′). Five microliters of this PCR reaction were then used as the substrate for a second round of PCR using the ETEL-3 (5′-CCTGACCTCCATTTTTAAGG-3′) and ETEL-4 (5′-TTGTTTGACAGATTTGACGG-3′) oligonucleotides as primers. For the *RPL29* gene the first round of PCR employed oligonucleotides r29-1 (5′-TGTCTCATACCGTGCTTTCC-3′) and r29-2 (5′-AGTTCCTCTCGTTGTCAAGG-3′) as primers and the second round used oligonucleotides r29-3 (5′-GTGACCTCTGAGTTTCCTCG-3′) and r29-4 (5′-AGGTGCCTGTGTATTGAGAG-3′). In all cases, 25 cycles of amplification were carried out, with a cycle consisting of 95°C for 1 min, 45 or 46°C for 75 s and 72°C for 135 s.

For PCR amplification of the macronuclear LTEL terminus, 1.5 µg *E.crassus* strain X1 total cellular DNA were run on a 1.2% low melting point agarose gel and the region of the gel containing DNA molecules ~1.4 kb in size was excised for use as the substrate. The first round of PCR was carried out using oligonucleotides LTEL1 (5′-GGCTTTAAGGTGAATCTGAG-3′) and C_4A_4 28mer (5′-CCCCAAAACCCCAAAACCCCAAAACCCC-3′) as primers. The second round of PCR employed LTEL2 (5′-CATCATTCTCTCAAGGTAAG-3′) and C_4A_4 28mer as the primers. Each round of PCR consisted of 25 cycles of 95°C for 1 min, 50°C for 75 s and 72°C for 1 min.

Some of the above two-step PCR strategies resulted in a weak specific product band with additional background PCR products. To enhance specificity and product yield, once candidate PCR products were identified, first round PCR reactions were subjected to electrophoresis on low melting point agarose gels. The region of the gel containing DNA fragments of the size predicted for first round PCR products (usually weak or not visible) was excised and this material was used as the substrate DNA for the second round PCR reactions.

RESULTS

Analysis of subtelomeric regions of macronuclear DNA

The analysis of macronuclear subtelomeric sequences examined a total sample of 100 macronuclear DNA termini. Forty two of the terminal sequences were from previously characterized macronuclear DNA molecules that are available in GenBank (Materials and Methods). To increase the sample size, one or both termini of an additional 33 randomly selected clones of *E.crassus* macronuclear DNA molecules were sequenced. These new terminal sequences have been deposited in GenBank under accession nos AF068575–AF068614, AF069914–AF069917, AF069919–AF069931 and AF077371.

A previously described statistical method (26) was employed to identify conserved sequence elements near the ends of macronuclear DNA molecules. This involved aligning the first 50 subtelomeric bases of the 100 macronuclear DNA ends and then calculating the overall frequency of each base for the entire sample. The first subtelomeric base is operationally defined as the one following the last block of four C residues in the 5′-C_4A_4-3′ telomeric repeats. Based on the overall base composition, an expected mean occurrence and standard deviation for each of the bases at any given position, as well as for all possible combinations of three different bases (ATG, TGC, GCA and CAT), was calculated (analyses of all possible combinations of two bases were also performed, but provided little additional information and will not be discussed). Subtelomeric positions where the actual (observed) frequency of a particular base or a combination

of three bases exceeded M + 2 SD were considered to have significant deviations from random base composition. The analysis of three base combinations, in effect, provides a means of assessing whether single bases are under-represented at a particular position. It should also be noted that this statistical approach compensates for the overall AT-richness of the subtelomeric sequences (the sample of 100 termini had the following base composition: A = 0.40, T = 0.35, C = 0.10 and G = 0.15).

Figure 1A graphically displays the results of the analysis for the entire sample. Black bars are shown at all subtelomeric positions where single bases displayed significant deviations from the average expected occurrence; open bars are shown for positions where a combination of three bases were significant *and* where the frequency of a single significant base did not exceed 0.60 (note that a single position can display a significant deviation from random base composition for more than one single base and/or combination of three bases). Three clusters of positions with significant deviations from the overall base composition are evident. First, within positions 46–50 there are over-abundances of ACG, G and C. Both the 5′ and 3′ subtelomeric non-coding regions of macronuclear DNA molecules in *E.crassus* are extremely short, frequently <50 bp in length (26). Thus, this small cluster of significant positions is likely the result of simply impinging on coding regions, which are more GC-rich than the subtelomeric non-coding DNA.

A second cluster of positions that has significant deviations from random base composition is positioned adjacent to the telomere. The first subtelomeric base is rarely C (f_C at position 1 = 0.02) and positions 2–4 all show a moderate bias for T.

The third cluster of significant positions extends from subtelomeric position 16 to 24. It centers on the 5′-TTGAA-3′ motif previously identified as a conserved element at the ends of macronuclear DNA molecules (4,15). The overall consensus sequence that can be deduced is 5′-$a_{0.53}V_{0.79}T_{0.85}T_{0.65}G_{0.72}$-$A_{0.73}A_{0.63}NH_{0.93}$-3′ (V = ACG, H = ACT and N = ACGT; subscripts denote the observed frequency of the base or combination of bases at the given position in the sample).

Inspection of individual sequences for strong matches to the well-conserved core 5′-TTGAA-3′ sequence (i.e. matches at four of the five positions) produced two intriguing observations. First, 62 of the 100 macronuclear termini contained strong matches to the core beginning at subtelomeric position 18. A significant number of the remaining macronuclear DNA termini (16 of 38) that lack a strong match to the E-Cbs core beginning at position 18 displayed a strong match to the core sequence positioned 1–3 bp closer to the chromosome fragmentation/telomere addition site. Only two termini had a strong match to the core sequence located 1–3 bp more distal of the usual placement. This suggests that the sequence element can sometimes be positioned closer to the telomere of macronuclear DNA molecules. Second, for the macronuclear DNA molecules where the subtelomeric sequences of both ends were known, one could find instances where the core was present at both ends, neither end, only the end 5′ of the coding region or only the end 3′ of the coding region. This is significant, as the conserved sequence element could potentially play a role in transcription, translation, DNA replication or chromosome fragmentation/telomere addition. The fact that the sequence element exists at the 5′- and/or 3′-ends of coding regions lessens the likelihood that it is involved in transcriptional or translational control. While it remains formally possible that the sequence element plays a role in DNA replication, we consider it more likely to play a role in directing chromosome fragmentation/telomere addition and will refer to the element as the *Euplotes* chromosome breakage sequence (E-Cbs).

The above considerations indicated that a mixed sample was being considered in the original analysis of the 100 macronuclear DNA termini, i.e. the sample contained ends with the E-Cbs core beginning at position 18, as well as ends lacking the E-Cbs or having it in a slightly different position. As a result, we carried out the same type of analysis on the 62 macronuclear termini possessing a strong match to the E-Cbs core beginning at subtelomeric position 18 to determine if a more robust consensus could be derived from this selected sample (Fig. 1B). The overall results were quite similar, but the actual frequencies of conserved bases (or combinations of three bases) increased in the E-Cbs region and additional positions with significant deviations from random base composition were observed. The increase in the absolute frequencies of the 5′-TTGAA-3′ core bases themselves are not unexpected, as the sample was selected on this basis. However, the increases in frequencies at the flanking positions provides an indication that these positions co-vary with the core and are likely significant for the function of the E-Cbs. The overall E-Cbs consensus sequence derived for subtelomeric positions 16–25 is $H_{0.95}A_{0.69}T_{0.95}T_{0.90}G_{0.98}A_{0.95}A_{0.76}a_{0.50}H_{0.97}H_{0.95}$ (note that the a at position 23 is not statistically significant by our criteria, but is included in the consensus because it is within the cluster of conserved residues and is present in 50% of the sample). This sequence is shorter than, but generally similar to, the previously suggested consensus (11) of 5′-ATAUUWUTTGAWWW-3′ (W = AT, U = AG and Y = CT). This previous consensus was based on the analysis of only seven sequences. The much larger sample size in the present study and the statistically based analysis of the sequences indicate to us that the current E-Cbs consensus is a much better representation of the well-conserved positions at the ends of macronuclear DNA molecules.

Micronuclear chromosome fragmentation sites

Even allowing for some variation in the position of the E-Cbs, there are numerous macronuclear DNA termini that lack reasonable matches to the consensus. It was previously suggested (11) that for these cases, the same sequence element exists in the flanking DNA in the micronucleus in inverted orientation, but is positioned six bases closer to the site of telomere addition. To investigate this further, we sequenced eight new micronuclear regions containing chromosome fragmentation/telomere addition sites (junction sequences), as well as compiled previously determined junction sequences available in GenBank.

The first approach to obtaining new junctions involved sequencing portions of a previously characterized cloned segment of micronuclear DNA (clone EC1; 24), as well as the termini of two macronuclear DNA molecules (Mac J and Mac C) that are derived from this region of the micronuclear genome. The sequencing analysis is outlined in Figure 2 and the obtained junction sequences are shown in Figure 4. As is typical for other macronuclear-destined sequences, Mac J and Mac C are separated by a 417 bp spacer DNA segment that is eliminated during development (Fig. 2). The micronuclear copy of Mac J was also found to contain one 62 bp IES. Micronuclear Mac C completely lacked IESs, which is highly unusual, as only one other macronuclear-destined sequence has been found to lack IESs in all hypotrich species studied at the molecular level (18,28). In

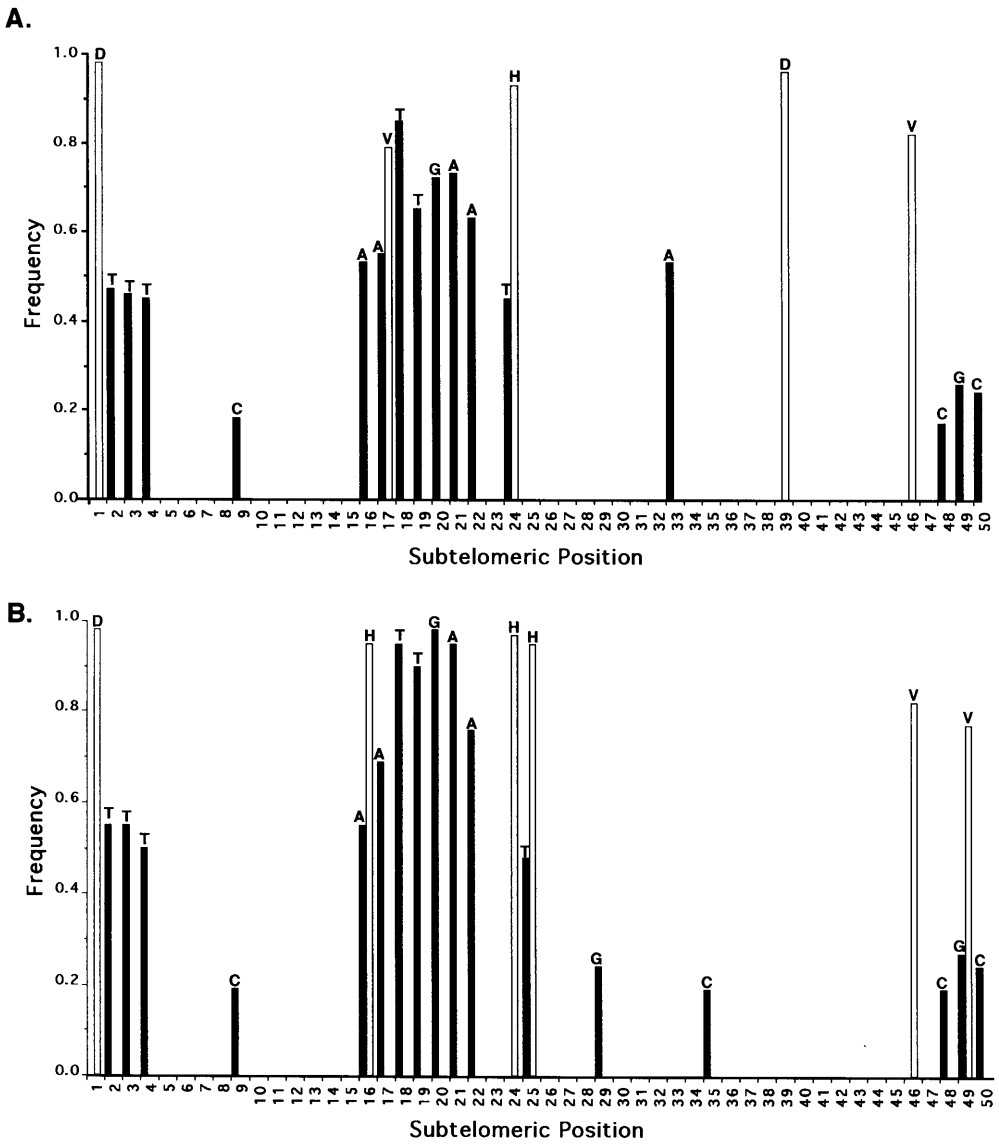

Figure 1. Analysis of the subtelomeric sequences of macronuclear DNA molecules. Histograms display the results of statistical analyses of base composition at each subtelomeric position for the entire data set of macronuclear termini (**A**) (n = 100; Materials and Methods) and for the selected macronuclear DNA termini (**B**) (n = 62). Black bars are shown only at those positions that have statistically significant deviations from random base composition for single bases (A, C, G or T). Open bars are shown at positions where a combination of three bases (ATC = H, TCG = B, CGA = V or GAT = D) displayed a statistically significant deviation from random base composition, except for those positions where a statistically significant single base whose frequency was >0.60 was also present. The overall base composition for the 100 termini analyzed in (A) was: A = 0.40, T = 0.35, C = 0.10, G = 0.15. Bases or combinations of three bases had to be present at a particular position in excess of the following values to be considered significant: A > 49, T > 44, C > 16, G > 21, H > 92, B > 69, V > 74 and D > 95. For the sequences analyzed in (B), the overall base composition was A = 0.40, T = 0.36, C = 0.10, G = 0.14. In this case, bases or combinations of three bases had to be present at a particular position in excess of the following values to be considered significant: A > 32, T > 29, C > 11, G > 14, H > 58, B > 45, V > 47 and D > 60.

regard to the E-Cbs, all four of the newly sequenced junctions possess good matches to the E-Cbs consensus within the regions that will form the ends of the Mac J and Mac C macronuclear DNA molecules (Fig. 4).

In the second approach, we specifically sought to characterize micronuclear junctions where the corresponding macronuclear DNA molecule was known to lack the E-Cbs at one end. This was accomplished using inverse PCR (27) to determine the micronuclear junctions for the two previously characterized macronuclear chromosomes that carry the telomerase RNA (*TelRNA*; 29) and ribosomal protein L29 (*RPL29*; 17) genes. The inverse PCR strategy for the *TelRNA* gene is outlined in Figure 3A. DNA isolated from developing macronuclei at the late polytene chromosome stage (40 h after the start of mating) was used as the substrate for inverse PCR, rather than micronuclear DNA. The rationale for this is that the inverse PCR procedure requires that the substrate DNA be digested with a restriction enzyme that does not cut within the region of known sequence and then ligated at a low DNA concentration to foster circle formation (Fig. 3A). While the sequences of the macronuclear *TelRNA* and *RPL29*

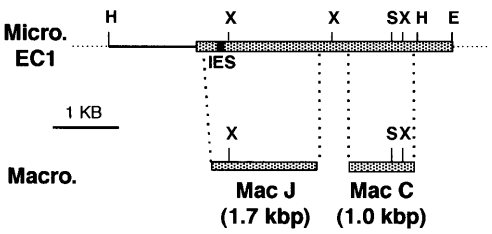

Figure 2. Restriction maps of a portion of micronuclear clone EC1 and two macronuclear DNA molecules (Mac J and Mac C) derived from this region of the genome. Sequenced regions are indicated by stippled rectangles and the position of an IES is indicated by a black box. S, *Sst*I; E, *Eco*RI; H, *Hin*dIII; X, *Xba*I.

genes are known, the micronuclear copies of the genes may contain IESs that introduce unanticipated restriction sites. This potential problem is reduced by using the 40 h DNA from the developing macronucleus, as most IESs have been removed by this time but chromosome fragmentation has not yet occurred (8,30–32).

We obtained a 1.6 kb PCR product for the *TelRNA* gene from the two-step inverse PCR procedure (Fig. 3B and C) and a 0.9 kb PCR product for the *RPL29* gene (data not shown). Sequencing of the junction regions from the PCR products revealed that for the two ends where the macronuclear-destined sequence lacked the E-Cbs, a good match to the element was present in the flanking micronuclear DNA (Fig. 4). In both cases, the orientation of the E-Cbs is inverted and the core 5′-TTGAA-3′ is positioned beginning at base 12 from the fragmentation/telomere addition site.

Figure 4 presents a compilation of all the known micronuclear junction regions. Seventeen of the 21 junction sequences have strong matches to the E-Cbs core within the macronuclear-destined sequence, beginning at either 17 or 18 bases from the chromosome fragmentation/telomere addition site. The four junctions that do not have a good E-Cbs core match in the macronuclear-destined DNA (*TelRNA*-5′, *RPL29*-5′, V2-l and V3-l), all have a strong match to the E-Cbs core in the flanking micronuclear DNA (Fig. 4). In each of these instances, the element is inverted and positioned with the core beginning at position 12 from the telomere addition site. Two of the junctions (*EFA1*-3′ and D8-r) have good matches to the E-Cbs in both the macronuclear-destined and flanking micronuclear DNA.

A statistical analysis of the micronuclear junction sequences using the same procedure employed for the macronuclear DNA termini was also carried out (data not shown). In this case we aligned all the junction sequences relative to the 5′-TTGAA-3′ core (i.e. junctions with the E-Cbs in flanking micronuclear DNA were inverted for the analysis and junctions where the core began at position 17 were shifted by 1 bp). The results again indicated that there is a cluster of positions with significant deviations from random base composition with the consensus $V_{0.96}T_{0.87}T_{0.87}G_{1.00}A_{0.91}A_{0.83}NA_{0.61}$. This consensus sequence is quite similar to that obtained for the samples of total and selected macronuclear DNA ends (Fig. 1), but some of the significant positions observed for the macronuclear DNA ends are not observed. This is likely the result of the smaller sample size for the micronuclear junctions. Overall, the results indicate that the E-Cbs is most frequently present in the macronuclear-destined DNA. However, in cases where it is absent in this position, it can

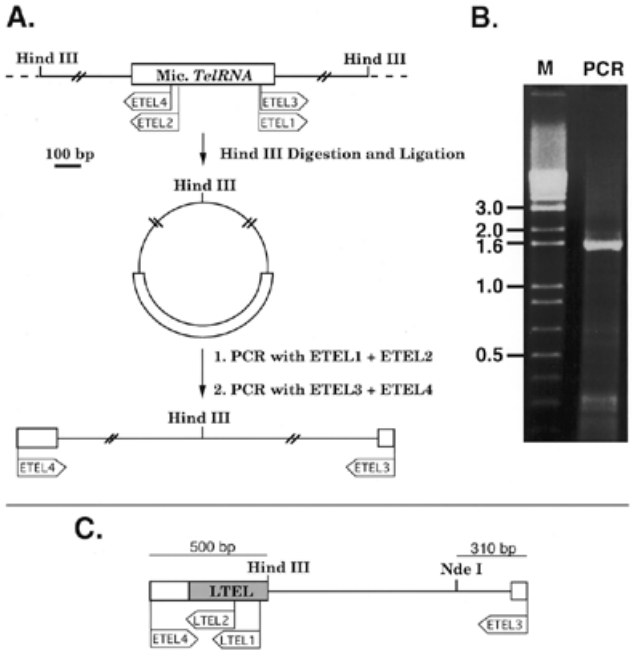

Figure 3. Inverse PCR of the chromosomal copy of the *TelRNA* gene. (**A**) The inverse PCR strategy is diagrammed, beginning with a representation of the micronuclear *TelRNA* gene (open rectangle; any IESs present are not shown) and flanking micronuclear DNA (lines). Positions of oligonucleotides used as PCR primers are indicated below the maps, with the arrows denoting the direction in which they prime DNA synthesis. (**B**) Agarose gel displaying the PCR product obtained from the *TelRNA* inverse PCR procedure (lane PCR), along with size markers (lane M). Sizes of selected marker DNAs are indicated in kb. (**C**) Restriction map of the 1.6 kb PCR product obtained using the inverse PCR strategy for the *TelRNA* gene. *TelRNA* macronuclear-destined sequences are indicated as open boxes. The location of the LTEL macronuclear-destined sequences are also indicated (shaded box), along with locations of PCR primers (LTEL1 and LTEL2) used in later analyses.

be found in the adjacent micronuclear DNA in inverted orientation and positioned with the core beginning 12 bp from the site of telomere addition.

Finally, we would like to note two additional aspects of the known micronuclear junctions. First, the V2-r junction contains the poorest match to the E-Cbs (4/10 positions overall and only 3/5 positions in the core; Fig. 4). Interestingly, one of us (L.A.Klobutcher, unpublished results) has recently obtained evidence that this particular fragmentation site is sometimes not utilized during macronuclear development, i.e. in some cell lines generated by independent episodes of macronuclear development (exconjugant clones), a subset of the V2 macronuclear DNA molecules retain the flanking spacer micronuclear DNA. The poor match to the E-Cbs may explain the inefficient use of this fragmentation site. Second, there are six junctions where the micronuclear-limited base, or bases, directly adjacent to the chromosome fragmentation/telomere addition sites (position –1, Fig. 4) are C residues. These junctions, in fact, represent cases where we cannot unambiguously define the chromosome fragmentation/telomere addition site, because the telomeric repeat begins with four C residues on the 5′-ends of macronuclear DNA molecules. Intriguingly, three of the four cases (*RPL29*-3′, C1-r and Ec2-5′) where the core 5′-TTGAA-3′ of the E-Cbs is positioned beginning at position 17 from what is shown as the ‘telomere

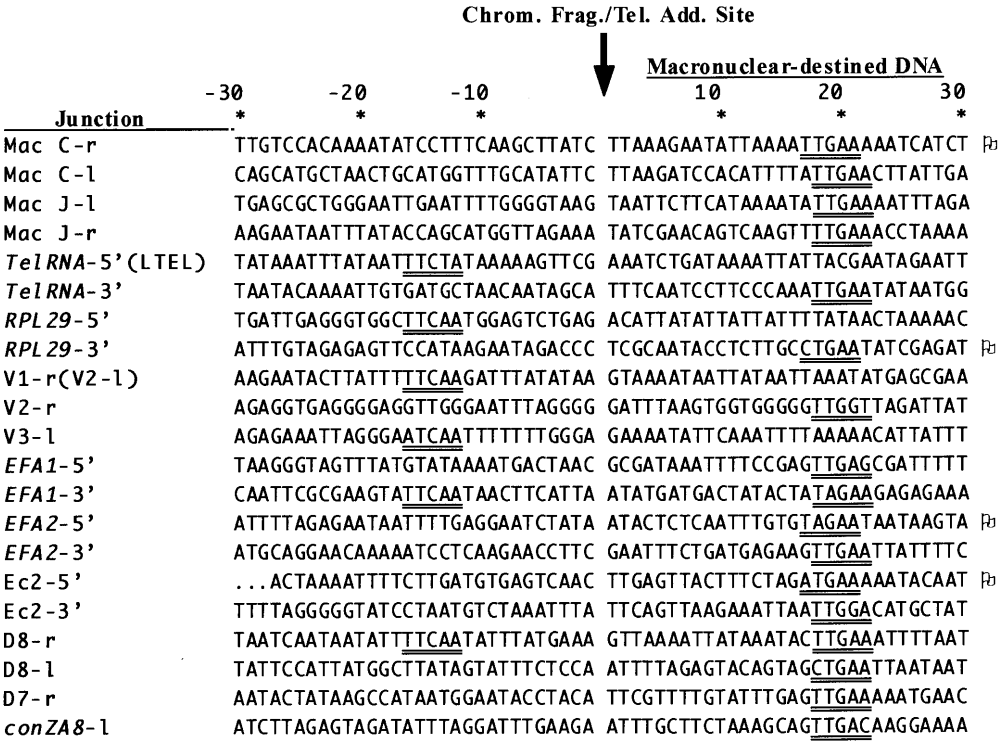

Figure 4. Micronuclear junction sequences. The first eight junction sequences shown are those determined in this study, while the remainder were published previously. The positions of the chromosome fragmentation/telomere sites are indicated and macronuclear-destined sequences are generally shown to the right. Matches to the 5′-TTGAA-3′ E-Cbs core are double underlined and flags denote cases where the E-Cbs core begins at position 17, as opposed to the more typical position 18, when positioned in macronuclear-destined DNA. In the sequence names, -5′ and -3′ denote the junctions upstream and downstream of coding regions in macronuclear-destined sequences, where known. In cases where coding regions have not been identified, -l and -r denote the left and right junctions of macronuclear destined sequences relative to previously published maps. For the two junctions where overlapping macronuclear-destined sequences are present, the name of the second macronuclear DNA molecule is given in parentheses. The full names of the genes present in the macronuclear-destined sequences of the junctions and their GenBank accession numbers or appropriate reference sources are as follows: Mac C (AF072706 and AF072708); Mac J (AF072707 and AF072708); *TelRNA* (telomerase RNA, AF061108 and AF061109); *RPL29* (ribosomal protein L29, AF061111 and AF061112); V2 and V1 (M28500); V3 (protein kinase, M28500); *EFA1* (elongation factor-1α-1; 41); *EFA2* (elongation factor-1α-2; 41); Ec2 (M73022 and M73023); D8 (M28498 and M28499); D7 (M28498); *conZA8* (AF063083).

addition site' in Figure 4 fall into this ambiguous category. This raises the possibility that chromosome fragmentation/telomere addition is actually occurring to the left of the –1 position for these three junctions, so that the distance relationship to the E-Cbs is actually constant. The potential significance of this latter observation with regard to the telomere addition process will be addressed further in the Discussion.

A second macronuclear-destined sequence overlaps the micronuclear *TelRNA* gene

In the micronuclear genome, adjacent macronuclear-destined sequences are typically separated by a developmentally eliminated spacer DNA segment (18; this study). However, in *E.crassus*, one pair of adjacent macronuclear-destined sequences has been found to overlap by 6 bp (18). To determine if other macronuclear-destined DNA sequences were located adjacent to the *TelRNA* and *RPL29* genes in the micronucleus, Southern hybridizations to macronuclear DNA were carried out using restriction fragments from the inverse PCR products as probes. A 310 bp *Nde*I restriction fragment from the *TelRNA* inverse PCR product (Fig. 3C), containing 64 bp of the 3′-end of the *TelRNA* macronuclear-destined sequence and ~250 bp of flanking micronuclear DNA, showed no hybridization to macronuclear DNA (Fig. 5A). This indicates that the sequences adjacent to the chromosome fragmentation/telomere addition site at the 3′-end of the *TelRNA* gene represent spacer DNA that is eliminated during macronuclear development. Similar hybridizations with probes from the *RPL29* micronuclear PCR product indicated that both ends of the *RPL29* macronuclear-destined sequence are flanked by eliminated spacer DNA (data not shown). In contrast, a 500 bp *Hin*dIII fragment, containing only 164 bp of the 5′-end of the *TelRNA* macronuclear-destined sequence, hybridized strongly to 1.4 kb macronuclear DNA molecules (Fig. 5A; note that hybridization to the 0.67 kb *TelRNA* macronuclear DNA was also seen on longer exposure). This indicates that a second macronuclear-destined sequence exists in close proximity to the fragmentation/telomere addition site at the 5′-end of the *TelRNA* macronuclear-destined sequence. We refer to this as the LTEL (left of Tel*RNA*) macronuclear-destined sequence (Fig. 3C).

To determine the position of the chromosome fragmentation/telomere addition site that generates LTEL, we employed a procedure termed MAC-end PCR (33) to isolate a terminus of the

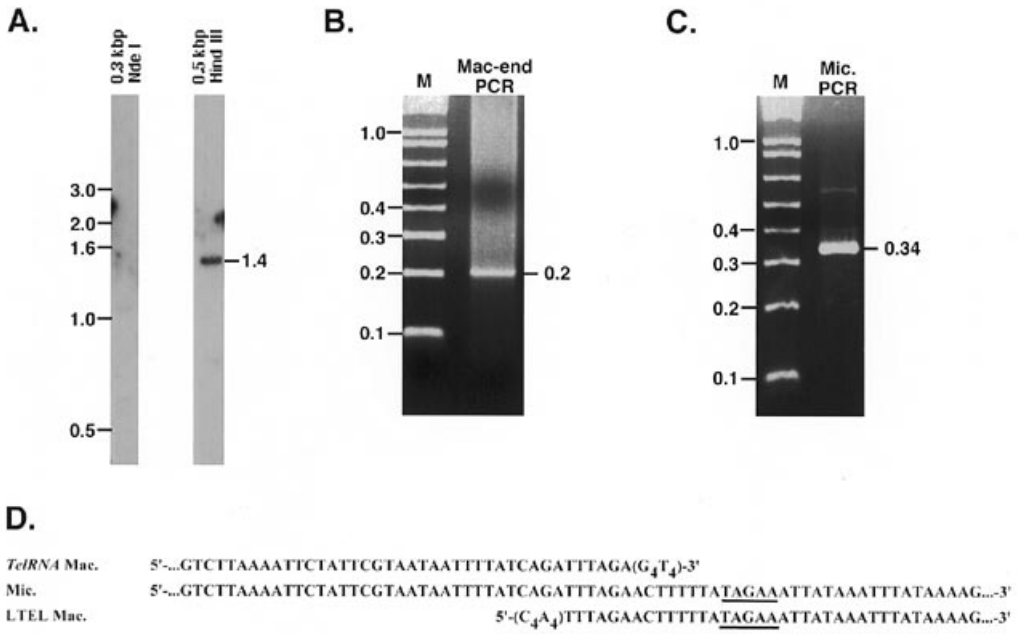

Figure 5. Analyses of the LTEL macronuclear-destined sequence. (**A**) Autoradiographs of Southern hybridizations of the 0.3 kb *Nde*I and 0.5 kb *Hin*dIII fragments of the 1.6 kb inverse PCR product from the micronuclear *TelRNA* locus to *E.crassus* whole cell DNA. (**B**) Agarose gel displaying the product obtained from ~1.4 kb *E.crassus* macronuclear DNA after two cycles of PCR using primers LTEL1 + C_4A_4 28mer and LTEL2 + C_4A_4 28mer (lane Mac-end PCR), along with DNA size markers (lane M). (**C**) Agarose gel displaying the product obtained from *E.crassus* micronuclear DNA after two cycles of PCR using primers ETEL2 + LTEL1 and ETEL4 + LTEL2 (lane Mic PCR), along with DNA size markers (lane M). Sizes of selected marker DNA fragments, hybridizing DNA fragments and PCR products are shown in kb. (**D**) Comparison of the micronuclear sequence of the *TelRNA* gene and its left-flanking region (Mic.) with the termini of the *TelRNA* and LTEL macronuclear DNA molecules (*TelRNA* Mac. and LTEL Mac., respectively). The telomeric repeats at the 5′- and 3′-ends of the macronuclear DNA molecules are abbreviated as C_4A_4 and G_4T_4, respectively, and the region matching the E-Cbs core is underlined. The sequence of the macronuclear LTEL terminus is deposited in GenBank under accession no. AF061110.

LTEL macronuclear DNA molecule. MAC-end PCR takes advantage of the fact that all macronuclear DNA molecules have telomeric repeats at their ends, so that a C_4A_4 oligonucleotide can serve as a universal PCR primer. A PCR reaction using gel-purified 1.4 kb macronuclear DNA molecules from *E.crassus* strain X1 as substrate was first carried out using LTEL1 (Fig. 3C) and a 28 bp oligonucleotide consisting of C_4A_4 telomeric repeats (C_4A_4 28mer) as primers. A portion of this PCR reaction was then re-amplified using LTEL2 (Fig. 3C) and C_4A_4 28mer as the primers. This procedure resulted in an ~200 bp PCR product (Fig. 5B) that was then sequenced using LTEL2 as the primer. The obtained sequence exactly matches the micronuclear DNA flanking *TelRNA* in the micronucleus, except for the presence of telomeric repeats at the LTEL macronuclear terminus (Fig. 5D). More important, the data reveal that, in the micronucleus, the LTEL macronuclear-destined sequence overlaps that of *TelRNA* by 6 bp (Fig. 5D).

Because the ‘micronuclear’ sequence of the *TelRNA*/LTEL region was obtained from a PCR product generated from 40 h developing macronuclear DNA, it remained formally possible that an IES separated the *TelRNA* and LTEL macronuclear-destined sequences. To test for this possibility, a step-in PCR procedure was used to PCR amplify across the *TelRNA*/LTEL junction directly from *E.crassus* strain X1 micronuclear DNA. The first round of PCR employed oligonucleotides ETEL2 and LTEL1 as the primers, while the second round used oligonucleotides ETEL4 and LTEL2 (Fig. 3A). A PCR product of 0.34 kb was observed (Fig. 5C), which is the size predicted from the sequence obtained from the 40 h DNA. This indicates that there are no IESs in the *TelRNA*/LTEL junction region in the micronucleus and that the two adjacent macronuclear-destined sequences do indeed overlap by 6 bp. This represents the second example of macronuclear-destined sequences with a 6 bp overlap (11) and suggests that this situation may be relatively common in the micronucleus. Thus, any model of the chromosome fragmentation/telomere addition process must provide an explanation for how such adjacent macronuclear-destined sequences can be resolved.

DISCUSSION

In this study we have examined a large number of macronuclear DNA termini, as well as junction sequences in micronuclear DNA. Our results provide a number of new insights into the *E.crassus* chromosome fragmentation/telomere addition process, as well as substantiate and clarify previous observations. First, through analyzing 100 macronuclear DNA termini, we have presented strong evidence for a conserved DNA element which we term the E-Cbs. The E-Cbs is found at one or both ends of most macronuclear DNA molecules, most typically with the highly conserved 5′-TTGAA-3′ core beginning at subtelomeric position 18. By selecting and analyzing the subset of macronuclear DNA termini that show a strong match to the core, we have deduced an E-Cbs consensus sequence of $H_{0.95}A_{0.69}T_{0.95}T_{0.90}G_{0.98}A_{0.95}$-$A_{0.76}a_{0.50}H_{0.97}H_{0.95}$ for subtelomeric positions 16–25. This sequence is generally similar to a previously derived consensus (11), but the current statistically based analysis of a much larger

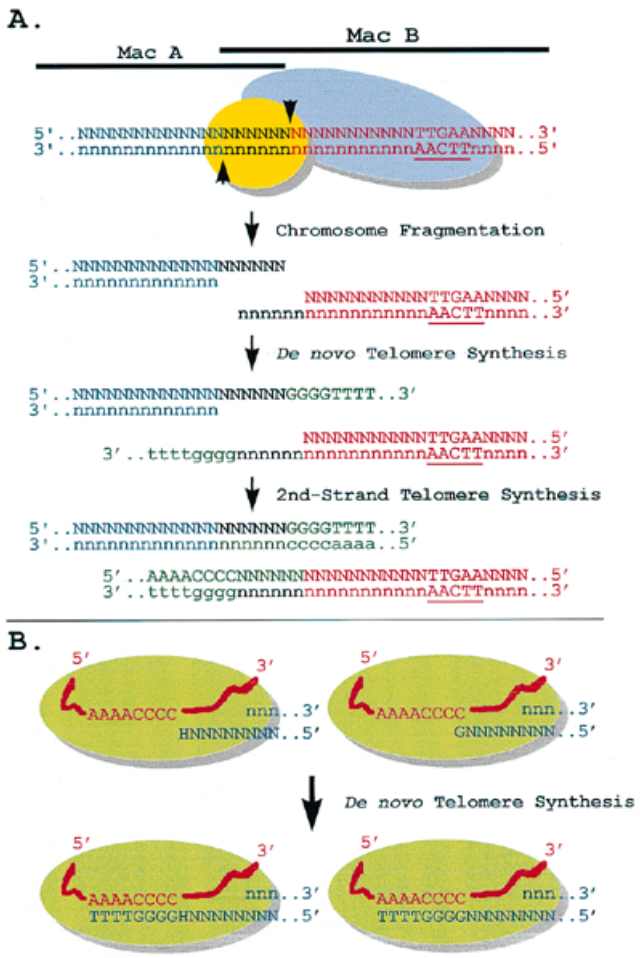

Figure 6. (**A**) Model of chromosome fragmentation and *de novo* telomere formation in *E.crassus*. A stretch of micronuclear DNA containing overlapping macronuclear-destined sequences (Mac A and Mac B) is shown at the top in association with proteins that interact with the E-Cbs (blue oval) and carry out chromosome fragmentation (yellow circle). Mac A and Mac B macronuclear-destined sequences are shown in blue and red, respectively, except for the 6 bp shared between them (black) and the core of the E-Cbs is underlined. Bases synthesized during the chromosome fragmentation/telomere addition process are shown in green. See text for additional details of the model. (**B**) *De novo* telomere addition by telomerase. Telomerase is represented as a green oval with a portion of telomerase RNA shown in red with the telomeric repeat template region (5′-CCCCAAAA-3′) highlighted. At the left, telomerase is shown interacting with a newly generated terminus of a macronuclear-destined sequence (blue lettering) where the 3′-terminus does not have a G residue (H). *De novo* telomere synthesis initiates by adding four G residues followed by four T residues. At the right, the 3′-end of the macronuclear-destined DNA has a G residue that aligns with part of the template of telomerase RNA. In this situation, telomere synthesis begins by adding three G residues followed by four T residues.

sample indicates to us that the newly derived consensus provides the better representation of the E-Cbs.

Second, although most macronuclear DNA termini possess the E-Cbs, there is a significant minority that lacks it. By analyzing micronuclear junction sequences, we have found that strong matches to the E-Cbs can be found in the flanking micronuclear DNA in all cases where an E-Cbs is absent from the terminus of the macronuclear-destined sequence (Fig. 4). When in micronuclear DNA, the E-Cbs is in inverted orientation and the 5′-TTGAA-3′ core is positioned beginning at 12 bp from the telomere addition site. The observation that the E-Cbs sometimes resides in flanking DNA suggests to us that the most likely role of the E-Cbs is in specifying chromosome fragmentation, rather than some function specific to macronuclear DNA molecules, such as replication or transcriptional control. In addition, the E-Cbs is unlikely to be directly responsible for telomerase recognition during *de novo* telomere synthesis, as there are numerous cases where the E-Cbs is in flanking micronuclear DNA yet the macronuclear-destined end still receives a telomere.

The E-Cbs sequence is intriguing in that a very similar conserved sequence element resides within the terminal inverted repeat of Tec transposon IESs and at one end of the short IESs of *E.crassus* (5 and references therein). This has led to the suggestion that there may be a link between chromosome fragmentation and IES excision (e.g. a shared interacting protein; 5 includes a discussion of this issue). With regard to protein recognition, studies of both macronuclear chromatin and naked macronuclear DNA indicate that the 5′-TTGAA-3′ core of the E-Cbs contains DNase I-hypersensitive sites and is particularly sensitive to DMS methylation (15). This suggests that the element has an alternative DNA structure, which might aid in protein recognition.

It should be emphasized, however, that even the extended E-Cbs consensus does not appear to contain sufficient information to restrict chromosome fragmentation to all appropriate positions in the genome during macronuclear development. We have carried out computer searches for the E-Cbs and have found strong matches at internal regions of mature macronuclear DNA molecules. Since these sites are not subject to fragmentation, this implies that other features must be involved in specifying chromosome fragmentation. This could include a second sequence element that was not identified in our analyses. Alternatively, the sequence context of the E-Cbs may be important, and the overall AT-richness of the subtelomeric regions is noteworthy in this regard. Also, recent studies suggest that alternative chromatin structures are involved in the DNA rearrangement processes of macronuclear development in both *E.crassus* (34) and *Tetrahymena* (35). It is possible that some E-Cbs-like sequences in the genome are masked from the fragmentation machinery by virtue of their chromatin structures.

Finally, the E-Cbs shares little sequence similarity with the 15 bp Cbs element (5′-AAAGAGGTTGGTTTA-3′) in *Tetrahymena* that has been shown to be required for chromosome fragmentation/telomere addition (reviewed in 1). Moreover, in contrast to the E-Cbs, the *Tetrahymena* Cbs always resides within flanking, eliminated micronuclear DNA and appears to work in an orientation-independent manner. This suggests that there may be significant differences in the chromosome fragmentation processes of these two distantly related ciliates. With regard to other hypotrichs, our results are probably relevant to other *Euplotes* species, as many macronuclear DNA termini in *E.aediculatus*, *E.octocarinatus* and *E.eurystomus* possess at least the core of the E-Cbs at the same position as *E.crassus* (4; GenBank entries). For more distantly related hypotrichous ciliates, such as *Stylonychia* and *Oxytricha* spp., one can often find reasonable matches to the E-Cbs core near fragmentation sites and some of these are positioned as in *E.crassus*. Moreover, there is one case of overlapping macronuclear-destined sequences in *Oxytricha nova*, where the overlap can be viewed as being 6 bp (36). However, the

observed heterogeneity in the fragmentation/telomere addition process in these more distantly related hypotrichs (reviewed in 1) greatly complicates the interpretation of these data and makes it unclear if the E-Cbs is a relevant feature for hypotrichs other than *Euplotes*.

A refined model of chromosome fragmentation/telomere addition

Our results generally support a previously proposed model of chromosome fragmentation/telomere addition in *E.crassus* (11). A refined version of the model is presented in Figure 6, considering a case of overlapping macronuclear-destined sequences. This arrangement was chosen because it, in effect, mimics the two general situations we have observed; the E-Cbs resides within what will become the subtelomeric region for one of the macronuclear-destined sequences (Fig. 6, Mac B) and in what is in effect flanking micronuclear DNA for the other macronuclear-destined sequence (Fig. 6, Mac A). In our model the E-Cbs serves as a positioning element that interacts with the chromosome fragmentation machinery. This interaction may be with the protein that actually carries out strand cleavage or, alternatively, with a second protein that interacts with the proposed endonuclease. The positioning serves to direct the machinery to produce a double-stranded break in the DNA 5′ of the E-Cbs. The top strand is cut 11 bp from the 5′-TTGAA-3′ core of the E-Cbs, while the bottom strand is cut 17 bp upstream of E-Cbs, producing a 6 bp 3′-overhang. The 3′-overhang serves as a substrate for *de novo* telomere addition by telomerase, which adds repeats of the sequence G_4T_4 in a 5′→3′ direction. Finally, the C_4A_4 strand of the telomere appears to be synthesized by a conventional DNA polymerase (37) and we postulate that the gap formed by the initial cleavage event is also filled-in by this process.

A key virtue of the model is that the proposed 6 bp staggered cut explains the two observed positions of the E-Cbs, relative to the telomere addition site, in macronuclear-destined as opposed to flanking micronuclear DNA. As can be seen in Figure 6, the E-Cbs core is located beginning at base 18 from the telomere for Mac B, while for the Mac A macronuclear DNA molecule, it is located 12 bp away from the ultimate site of telomere addition. The model also provides a straightforward scheme for how overlapping macronuclear-destined sequences can be processed into two separate macronuclear DNA molecules that share 6 bp of subtelomeric sequence (11 contains a discussion of other means of resolving overlapping sequences). It should be noted that while the data suggest an end with a 6 bp overhang as the product of fragmentation, they do not predict whether it will be a 3′- or 5′-overhang. The previous model of chromosome fragmentation suggested a 5′-overhang (11), which was filled-in by a DNA polymerase prior to the telomere addition process. Based on what is now known about telomerase activity, we favor a 3′-overhang. *In vitro*, telomerase prefers to add telomeric repeats to oligonucleotides with single-stranded 3′-ends, as compared with blunt-ended double-stranded oligonucleotides (38). Ultimately, studies aimed at detecting the predicted fragmentation intermediates with either 5′- or 3′-overhangs will be needed to resolve this issue.

Finally, while the model shown in Figure 6 suggests that the chromosome fragmentation proteins and telomerase are separate entities, this may not be the situation *in vivo*. Chromosome fragmentation and telomere addition appear to be closely coupled *in vivo*, as Southern hybridization studies have consistently failed to detect newly fragmented forms that lack telomeres (8,9). This suggests that a protein complex may be carrying out the two processes. While there is as yet no direct evidence for the association of fragmentation activity with telomerase, there are other indications that telomerase interacts with other proteins. During *de novo* telomere addition, synthesis of the G-rich and C-rich strands are coordinated, suggesting that telomerase and DNA polymerase are part of an interacting complex (37). Moreover, Bednenko *et al*. (39) have presented evidence that telomerase interacts with a novel factor during macronuclear development that appears to be important for *de novo* telomere repeat addition.

Implications for telomere addition

The analysis of macronuclear DNA termini also revealed a non-random base composition at the first four subtelomeric positions: the first position is almost never a C residue, while positions 2–4 have a significant bias for T residues (Fig. 1). Since the absolute frequencies of the bases at these positions do not increase substantially in the data set selected for matches to the E-Cbs core (Fig. 1B), it seems unlikely that the sequences at these positions serve to direct chromosome fragmentation. We suggest that they are important for the telomere addition process. Specifically, the T residues, which would be A residues on the opposite DNA strand to which telomerase adds telomeric repeats, may be involved in interacting with either the protein or RNA components of telomerase. Such interactions could serve to position the 3′-end for telomeric repeat addition.

The strong bias against having a C at the first subtelomeric position, which corresponds to an absence of G on the opposite strand, is likely a reflection of the mode of action of the *E.crassus* telomerase. *In vitro*, telomerase from both vegetative cells and cells undergoing macronuclear development efficiently adds telomeric repeats to the 3′-ends of oligonucleotides consisting of the telomeric repeat sequence (39,40). The developmental, but not the vegetative, telomerase will also add telomeric repeats to single-stranded oligonucleotides that have a single telomeric repeat at the 5′-end but non-telomeric DNA at the 3′-end (39,40). The latter situation somewhat mimics the type of *de novo* telomeric repeat addition that must occur on the ends generated by chromosome fragmentation. Melek *et al*. (40) have shown that the developmental telomerase usually begins telomere addition onto non-telomeric 3′-ends by adding four G residues, followed by four T residues (Fig. 6B). However, when the non-telomeric 3′-end of the oligonucleotide contains a G residue, only three additional G residues are added, followed by four T residues (Fig. 6B), i.e. telomerase recognizes the terminal G (or G residues), presumably through Watson–Crick base pairing with the template region of telomerase RNA (Fig. 6B) and adds only enough additional G residues to form a complete repeat. The almost complete lack of C residues (G on the strand subject to telomere addition) at the first subtelomeric position provides an indication that the same mechanism is operating *in vivo*.

This type of model also helps explain the subset of macronuclear DNA termini containing the E-Cbs sequence located 1–3 bp closer to the telomere. In our analysis of micronuclear junctions, four junctions had the E-Cbs shifted 1 bp closer to the telomere (Fig. 4). For three of these junctions, the first micronuclear base flanking the ‘chromosome fragmentation/telomere addition’ site

is a C (Fig. 4), corresponding to a G on the opposite strand. As we have noted, these represent cases where the telomere addition site is ambiguous, i.e. fragmentation/telomere addition could be occurring at the position indicated in Figure 4. Alternatively, chromosome fragmentation may have occurred 1 bp to the left of the position indicated (Fig. 4, position –1), if telomerase initiated telomere synthesis by adding only three G residues in forming the first repeat. We favor the latter situation, as it implies that the E-Cbs consistently directs cleavage to a defined upstream position. Further evidence for this hypothesis may come from analyzing the micronuclear copies of some of the macronuclear DNA ends where the E-Cbs appears to be shifted 2 or 3 bp closer to the telomere. We would predict that the immediately flanking micronuclear DNA in these instances should consist of two and three C residues, respectively.

ACKNOWLEDGEMENTS

We thank Tom Doak and Dorothy Shippen for their insightful comments on models of chromosome fragmentation/telomere addition and Michelle Kelley for technical assistance. This work was supported by grant 4737 from The Council for Tobacco Research to L.A.K., National Science Foundation grant MCB-9319009 to C.L.J. and National Institutes of Health grant GM41803 to C.M.P.

REFERENCES

1 Coyne,R.S., Chalker,D.L. and Yao,M.-C. (1996) *Annu. Rev. Genet.*, **30**, 557–578.
2 Klobutcher,L.A. and Herrick,G. (1997) *Prog. Nucleic Acid Res. Mol. Biol.*, **56**, 1–62.
3 Prescott,D.M. (1994) *Microbiol. Rev.*, **58**, 233–267.
4 Klobutcher,L.A., Swanton,M.T., Donini,P. and Prescott,D.M. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 3015–3019.
5 Jacobs,M.E. and Klobutcher,L.A. (1996) *J. Eukaryot. Microbiol.*, **43**, 442–452.
6 Yu,G.L., Bradley,J.D., Attardi,L.D. and Blackburn,E.H. (1990) *Nature*, **344**, 126–132.
7 Vermeesch,J.R. and Price,C.M. (1994) *Nucleic Acids Res.*, **21**, 5366–5371.
8 Tausta,S.L. and Klobutcher,L.A. (1990) *Nucleic Acids Res.*, **18**, 845–853.
9 Roth,M. and Prescott,D.M. (1985) *Cell*, **41**, 411–417.
10 Vermeesch,J.R. and Price,C.M. (1993) *Mol. Cell. Biol.*, **14**, 554–566.
11 Baird,S.E. and Klobutcher,L.A. (1989) *Genes Dev.*, **3**, 585–597.
12 Yao,M.-C., Zheng,K. and Yao,C.H. (1987) *Cell*, **48**, 779–788.
13 Yao,M.-C., Yao,C.H. and Monks,B. (1990) *Cell*, **63**, 763–772.
14 Coyne,R.S. and Yao,M.-C. (1996) *Genetics*, **144**, 1479–1487.
15 Price,C.M. (1990) *Mol. Cell. Biol.*, **10**, 3421–3431.
16 Shippen-Lentz,D. and Blackburn,E.H. (1990) *Science*, **247**, 546–552.
17 Jahn,C.L., Erbeznik,M., Jaraczewski,J.W., Melek,M. and Shippen,D.E. (1994) *Gene*, **151**, 231–235.
18 Baird,S.E., Fino,G.M., Tausta,S.L. and Klobutcher,L.A. (1989) *Mol. Cell. Biol.*, **9**, 3793–3807.
19 Roth,M., Lin,M. and Prescott,D.M. (1985) *J. Cell. Biol.*, **101**, 79–84.
20 Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1998) *Current Protocols in Molecular Biology.* John Wiley and Sons Inc., New York, NY.
21 Knecht,K. and Klobutcher,L.A. (1995) *Eur. J. Protistol.*, **31**, 201–207.
22 Boswell,R.E., Klobutcher,L.A. and Prescott,D.M. (1982) *Proc. Natl Acad. Sci. USA*, **79**, 3255–3259.
23 Feinberg,A.P. and Vogelstein,B. (1983) *Anal. Biochem.*, **132**, 6–13.
24 Jahn,C.L., Nilles,L.A. and Krikau,M.F. (1988) *J. Protozool.*, **35**, 590–601.
25 Qian,L. and Wilkinson,M. (1991) *BioTechniques*, **10**, 736–737.
26 Ghosh,S., Jaraczewski,J.W., Klobutcher,L.A. and Jahn,C.L. (1994) *Nucleic Acids Res.*, **22**, 214–221.
27 Ochman,H., Gerber,A.S. and Hartl,D.L. (1988) *Genetics*, **120**, 621–625.
28 Prescott,D.M. and DuBois,M.L. (1996) *J. Eukaryot. Microbiol.*, **43**, 432–441.
29 Shippen,L.D. and Blackburn,E.H. (1990) *Science*, **247**, 546–52.
30 Klobutcher,L.A. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 1979–1983.
31 Frels,J.S. and Jahn,C.L. (1995) *Mol. Cell. Biol.*, **15**, 6488–6495.
32 Frels,J.S., Tebeau,C.M., Doktor,S.Z. and Jahn,C.L. (1996) *Mol. Biol. Cell*, **7**, 755–768.
33 Seegmiller,A., Williams,K.R., Hammersmith,R.L., Doak,T.G., Messick,T., Witherspoon,D., Storjohann,L.L. and Herrick,G. (1996) *Mol. Biol. Evol.*, **13**, 1351–1362.
34 Jahn,C.L., Ling,Z., Tebeau,C.M. and Klobutcher,L.A. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 1332–1337.
35 Madireddi,M.T., Coyne,R.S., Smothers,J.F., Mickey,K.M., Yao,M.-C. and Allis,C.D. (1996) *Cell*, **87**, 75–84.
36 Klobutcher,L.A., Huff,M.E. and Gonye,G.E. (1988) *Nucleic Acids Res.*, **16**, 251–263.
37 Fan,X. and Price,C.M. (1997) *Mol. Biol. Cell*, **8**, 2145–2155.
38 Lingner,J. and Cech,T.R. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 10712–10717.
39 Bednenko,J., Melek,M., Greene,E.C. and Shippen,D.E. (1997) *EMBO J.*, **16**, 2507–2518.
40 Melek,M., Greene,E.C. and Shippen,D.E. (1996) *Mol. Cell. Biol.*, **16**, 3437–3445.
41 Florian,V. and Klein,A. (1996) *Nucleic Acids Res.*, **24**, 3195–3200.