Evolutionary Conservation of Sequences Directing Chromosome Breakage and rDNA Palindrome Formation in Tetrahymenine Ciliates

Robert S. Coyne and Meng-Chao Yao

Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98104

Manuscript received June 5, 1996
Accepted for publication August 26, 1996

ABSTRACT

Extensive, programmed chromosome breakage occurs during formation of the somatic macronucleus of ciliated protozoa. The cis-acting signal directing breakage has been most rigorously defined in *Tetrahymena thermophila*, where it consists of a 15-bp DNA sequence known as Cbs, for chromosome breakage sequence. We have identified sequences identical or nearly identical to the *T. thermophila* Cbs at sites of breakage flanking the germline micronuclear rDNA locus of six additional species of *Tetrahymena* as well as members of two related genera. Other general features of the breakage site are also conserved, but surprisingly, the orientation and number of copies of Cbs are not always conserved, suggesting the occurrence of germline rearrangement events over evolutionary time. At one end of the *T. thermophila* micronuclear rDNA locus, a pair of short inverted repeats adjacent to Cbs directs the formation of a giant palindromic molecule. We have examined the corresponding sequences from two other *Tetrahymena* species. We find the sequence to be partially conserved, as previously implied from analysis of macronuclear rDNA, but of variable length and organization.

In some organisms, chromosome breakage is a normal and highly regulated event in the development of somatic nuclei. This phenomenon was first observed over 100 years ago by BOVERI in the ascidian nematodes (BOVERI 1887). It has been most well studied in the ciliated protozoa, in which breakage occurs at hundreds to thousands of sites during the process of sexual conjugation (reviewed in YAO 1989; PRESCOTT 1994). Although unicellular, ciliates contain two types of nucleus of very divergent character. One is a transcriptionally active, polyploid, "somatic" macronucleus and the other a transcriptionally silent, diploid, "germline" micronucleus. Despite these differences, the two nuclei are derived from mitotic division of the same zygotic nucleus generated during conjugation (reviewed in ORIAS 1986). The mitotic product destined to form the new macronucleus undergoes extensive genome remodelling, including DNA elimination and chromosome fragmentation.

The chromosome fragmentation phenomenon has been most well characterized in *Tetrahymena thermophila*. Breakage occurs at between 50 and 200 reproducible sites (ALTSCHULER and YAO 1985), each apparently defined by the presence of a 15-bp sequence known as Cbs, for chromosome breakage sequence, which has been shown to be the necessary and sufficient signal for breakage (YAO et al. 1987, 1990). With one known exception (described below), breakage is coupled to the healing of chromosome ends by addition of telomeric repeats by telomerase (YAO and YAO 1981; KING and YAO 1982; YAO et al. 1987; YU and BLACKBURN 1991). Telomere addition occurs at a somewhat variable distance (usually 5–20 bp) from both ends of Cbs, with the result that Cbs itself and some flanking DNA are eliminated from the macronucleus (KING and YAO 1982; YAO et al. 1985; FAN and YAO 1996).

Although chromosome breakage is common to all ciliates yet examined molecularly, other phylogenetic subgroups differ considerably from *T. thermophila* in its extent, variability and cis-acting sequence requirements (BAROIN et al. 1987; FORNEY and BLACKBURN 1988; KLOBUTCHER et al. 1988; BAIRO and KLOBUTCHER 1989; CARON 1992). These differences, as well as other fundamental differences in the development, division and DNA replication of ciliate macronuclei, have led ORIAS (1991) and HERRICK (1994) to propose that fragmentation arose independently in at least three separate ciliate lineages. Examination of breakage sites in a wider variety of contemporary ciliate species may help to evaluate this hypothesis. In addition, the overall pattern of sequence conservation at breakage sites may provide insight into breakage mechanisms as well as the organization of micronuclear chromosomes.

In this study, we explore the evolutionary conservation of breakage sites in several *Tetrahymena* species, as well as members of two closely related genera. A phylogenetic tree depicting the species' relatedness is shown in Figure 1 (SOGIN et al. 1986; PREPARATA et al. 1989; BRUNK et al. 1990; GREENWOOD et al. 1991). We present evidence that Cbs and other general features of the breakage mechanism have been extremely well...
conserved throughout this group of organisms. Unexpectedly, however, the orientation and number of copies of Cbs at particular sites have not been conserved.

As mentioned above, chromosome breakage at one site in *T. thermophila* is not followed by telomere addition. At the 5′ end of the ribosomal RNA locus (rDNA), a pair of 42-bp inverted repeats (also known as M repeats), separated by a 28-bp spacer, is located adjacent to a pair of 42-bp inverted repeats (also known as M repeats) of *T. thermophila*. A conserved 3′ region near the 5′ end of the 17S rRNA-encoding region (5′-AGACAACCATATGTAACTGGCAGGA-3′) or, for *G. chattoni* (from a well conserved region near the 3′ end of the 26S rRNA-coding region and the other, Tel (5′-CCCCAACCCCAACCCC3′), which is specific for the telomeric repeats of all the species used (see Figure 2A). The primers were annealed at 53°C in 1.5 mM MgCl₂. Amplification products were cloned and sequenced as described above, except that the Taq DNA polymerase was purchased from Gibco/BRL. PCR was performed according to standard protocols (SAKII et al. 1988). PCR products were isolated from agarose gel slices by a glass bead spin column method (Qiagen, Inc.) and cloned using the TA vector pCR II (Invitrogen Corp.). DNA sequencing was performed using an automated sequencer, Model 373, from Perkin Elmer, Applied Biosystems Division, and Dye Terminator reagents, except as noted.

**PCR amplification of the rDNA telomere-adjacent sequences:** Approximately 100 ng of genomic DNA from each species was amplified for 30 cycles using as one primer, r3′ (5′-CACAGGGATAACTGGCTTGT-3′), from a well conserved region near the 3′ end of the 26S rRNA-coding region and as the other, Tel (5′-CCCCAACCCCAACCCC3′). Amplification generated a primary product ~2 kb in length, which was cloned and partially sequenced as described above. For *G. chattoni*, the 5′ telomere-adjacent sequences were also amplified using the Tel primer and r5′ (see below).

**PCR amplification of the 5′ rDNA micronuclear region flanking the chromosome breakage site:** Approximately 100 ng of genomic DNA was amplified for 30 cycles using as one primer the 5′-AAAGAGGTGGTTTAATTT-3′ and as the other either Tel p, derived from the telomeric region of *T. piscicola* (from PETERS BRUNS), or, for *T. thermophila*, derived from the nonpalindromic center of the *T. piscicola* rDNA (5′-GCTACCTTACACCATAC-3′) or, for *T. thermophila*, r5′, from a well conserved region near the 5′ end of the 17S rRNA-encoding region (5′-AGACAACCATATGTAACTGGCAGGA-3′). Amplification products were cloned and sequenced as described above, except that the *T. thermophila* inverted repeats could only be sequenced through by using Dye Primer reagents. The upstream *T. piscicola* inverted repeat and flanking DNA was amplified from primers Tel p (5′-GCTTATAGGCGTTTTT-3′) and r5′ and sequenced directly.
Ciliate DNA Processing Signals

A. 

B. 

C. 

D. 

**Figure 2.**—PCR strategies. (A) Amplification of the 3\textsuperscript{r} rDNA telomere-adjacent region. (B) Amplification of the 5\textsuperscript{r} rDNA breakage site and M repeats. (C) Vectorette amplification of the 3\textsuperscript{r} rDNA flanking region, including Cbs. (D) Same as C for the 5\textsuperscript{r} rDNA end. Symbols are as follows: rRNA-transcribed region, ■; telomeres, □; Cbs, ■; flanking micronuclear DNA, wavy line; vectorette oligonucleotide pair, oval; M repeats, open triangles; nonpalindromic center, ▽; PCR and sequencing primers, arrows. Identity of most primers is explained in MATERIALS AND METHODS. V3'-1 and 2 and V5'-1 and 2 are the species-specific nested vectorette primers for the 3\textsuperscript{r} and 5\textsuperscript{r} rDNA ends, respectively. Distances are not to scale.

without cloning from primer Tp3 (5'CTGAATTACATCTTATACTCCTCG-3') (see Figure 2B).

**Vectorette PCR:** This method, like inverse PCR, is designed to amplify a DNA region when sequence information is only available from one end of the region (Riley et al. 1990; Arnold and Hodgson 1991). Four pairs of partially complementary “vectorette” oligonucleotides were designed such that, when annealed, the “top” strand will have a 5-base 5' overhang and there will be a 26-bp unpaired internal “bubble” region. The sequence of the “bottom” strand oligo (VKET3) is the same for each pair: 5'TCCTCTCTTGACGAACTGTCCTCTTCTCTC CC-3'. The four top strand oligos (VKET7-BamHI, -Nol, -XbaI and -EcoRI) differ only in their four-base 5' overhangs, each corresponding to that generated by its namesake restriction enzyme. For example, the sequence of VKET7-Nol is 5'CATGGAAGAAGAGACGGACCAGTG AAATGGTAATACGACAGGAGG-3'. Pairs of oligos were mixed in equimolar amounts, heated to 65°C and annealed by slow cooled. Whole cell genomic DNA from each ciliate species was digested with one or more restriction enzyme that generates ends compatible with one of the four vectorette oligo pairs. Approximately one microgram of digested DNA was ligated to 10-20 pmol annealed oligos at 15°C overnight. One-tenth of this ligation mix was then used for the first of two nested rounds of PCR amplification. Primers for the first round were a gene-specific primer, its 3' end facing the micronuclear region to be amplified, and B3 (5'CTCTGTGACGACGGCCAAGCTCGAAA-3'), whose sequence corresponds to a portion of the unpaired bubble region of the bottom vectorette oligo, but whose complementary target strand can only be made by extension from another primer (see Figure 2, C and D). Thus, the only region amplified by this pair of primers should be adjacent to the gene-specific primer. Amplification was for 30 cycles at 2 and 4 mM MgCl\textsubscript{2} with a 58°C annealing temperature in a Stratagene Robocycler. One-fiftieth of the first round PCR reaction was then used as template for a second round with a nested gene-specific primer and T3 (5'CGCGAAAGCTCGAAATACCGCT-3') under the same conditions as for the first round. Aliquots of second round reactions were run on 3% NuSieve agarose gels (FMC Bioproducts). Specific products were isolated from agarose gel slices, cloned and sequenced as described above.

**RESULTS**

**General strategy:** To compare breakage sites from several tetrahymenine ciliates, we focused our attention on the rDNA locus. The *T. thermophila* rDNA locus is present as a single copy gene, flanked by one copy of Cbs (dark arrowhead) to its 3' end and three copies to its 5' end. A pair of 42-bp inverted repeats (open arrowheads), also known as M repeats, flanking a 28-bp nonpalindromic center (C) are found adjacent to the 5' breakage site. After processing during macronuclear development, the mature palindromic macronuclear rDNA is amplified to ~200-fold the copy number of other macronuclear chromosomes. Distances are not to scale.

**Figure 3.**—Processing of the *T. thermophila* rDNA locus. The micronuclear rDNA locus is present as a simple copy gene, flanked by one copy of Cbs (dark arrowhead) to its 3' end and three copies to its 5' end. A pair of 42-bp inverted repeats (open arrowheads), also known as M repeats, flanking a 28-bp nonpalindromic center (C) are found adjacent to the 5' breakage site. After processing during macronuclear development, the mature palindromic macronuclear rDNA is amplified to ~200-fold the copy number of other macronuclear chromosomes. Distances are not to scale.

other tetrahymenine ciliates either share this palindromic rDNA form (Engberg et al. 1976; Karrer and Gall 1976; Din and Engberg 1979) or, at least in the case of *G. chattoni* (Katz et al. 1981), have a nonpalindromic, monomeric rDNA. A likely place, then, to look for breakage signals in these species would be adjacent to the micronuclear copy of their rDNA locus. Since tetrahymenine micronuclear DNA is present in substantially lower amounts than macronuclear DNA, we used a two-step PCR strategy to amplify the desired micronuclear regions. The first step was to amplify, clone and sequence the macronuclear DNA's adjacent to the 3' rDNA telomeres and the 5' palindromic centers of
several species. The second step was amplification of adjacent micronuclear DNA by a nested vectorette PCR method (Riley et al., 1990; Arnold and Hodgson, 1991), as described below and in Materials and Methods.

Organization of micronuclear DNA at 3′ rDNA breakage sites: The 3′ subtelomeric micronuclear rDNA was amplified from six species of Tetrahymena, G. chattoni, and C. campylum using one primer from a highly conserved portion of the 26S rRNA-encoding region close to the 3′ end of the transcribed portion of the rDNA (r3′) and one primer specific for the telomeric G4T2 repeats (Tel) (see Materials and Methods and Figure 2A). The ~2-kb product from each species was cloned and partially sequenced from the telomeric end (data not shown). This sequence information was used to design nested PCR primers facing out toward the telomere or, in the case of micronuclear DNA, toward the presumed breakage signal (see Figure 2C). These primers were then used in the second step of the strategy, a nested vectorette PCR (see Materials and Methods), to amplify the adjacent micronuclear DNA. This strategy was not successful with every species attempted, perhaps due to the absence of restriction sites at a suitable distance from the macronuclear-retained sequence for ligation to the vectorette oligonucleotides (all of the products successfully amplified were <1 kb in length), or for unknown technical reasons.

The sequences obtained from T. eliotii, T. borealis, T. caudata, T. hegewischi and G. chattoni are shown in Figure 4A and compared with T. thermophila. As indicated, a sequence identical to or having only a single base pair difference from the T. thermophila 15-bp Cbs is found in every species, at a similar distance from the site of macronuclear telomere addition (at least in the particular cloned telomeric PCR product chosen for sequencing). As also indicated, the orientation of Cbs in G. chattoni is reversed relative to the five Tetrahymena species, as discussed further below.

We were unable to amplify by vectorette PCR the 3′ micronuclear regions from two other Tetrahymena species, T. malacensis and T. pigmentosa, or from C. campylum. However, we were able to amplify the micronuclear region flanking the site of macronuclear telomere addition from each of these species by conventional PCR using either of the nested telomere-adacent macronuclear primers and a 15-mer oligonucleotide matching the T. thermophila Cbs sequence. The PCR products were cloned and sequenced, as shown in Figure 4B. The PCR reactions were fully reproducible and, as mentioned, performed independently with two species-specific primers. In addition, the sequences were in full agreement with the macronuclear rDNA sequence determined previously (see above) up to the site of telomere addition. We are therefore confident that these sequences represent the true organization 3′ of the micronuclear rDNA locus in each of these species.

**Figure 4.**—DNA sequence organization of the 3′ rDNA chromosome breakage sites of several tetrahymenellinae ciliates. The top line for each species shows micronuclear DNA sequence. The bottom line presents the corresponding macronuclear sequence, up to the point, in the individual cloned representative, where the micronuclear sequence gives way to the telomeric G4T2 repeats, indicated in parentheses. For most of the species, it is unknown whether the first two T's of the telomeric repeats were added by telomerase or were derived from the chromosomal sequence. The Cbs sequence of each species is underlined, with an arrow above indicating its orientation. Base pair differences from the T. thermophila Cbs are in lowercase and boldface. Also underlined is a short, G/T-rich sequence, conserved to varying degrees, at or near the telomere addition site in several species. (A) Regions amplified by vectorette PCR. (B) Regions amplified by direct PCR using one primer oligonucleotides directed against the T. thermophila 15-bp sequence. Therefore, the underlined Cbs sequence is not necessarily an exact match to the actual micronuclear DNA sequence of these three species.

Although we cannot state that Cbs in each of these species is identical to the T. thermophila sequence, it must be at least remarkably similar to be able to be amplified by a 15-mer primer at the annealing temperatures used (48–50°C). In the case of C. campylum, the base closest to the telomere addition site appears to be an A, as in T. thermophila, rather than a C, as in G. chattoni, since a 15-mer oligonucleotide matching the G. chattoni Cbs will not amplify this region from C. campylum (data not shown). As with G. chattoni, the Cbs...
in *C. campylum* is in a reversed orientation to that of all the Tetrahymena species examined.

Other conserved features of the breakage site region: Besides the Cbs sequence itself and the distance between it and the site of telomere addition, two other conserved features of the breakage site are noteworthy. The first, which is shared among all species examined, is the highly A/T-rich nature of the DNA sequence surrounding Cbs. This feature has been noted at other Cbs sites in the *T. thermophila* genome (Yao et al. 1987; Yu and Blackburn 1991).

The second is the presence in all the Tetrahymena species, except *T. caudata*, of a short G/T-rich sequence (underlined in Figure 4, A and B) at or near the site of telomere addition, conserved in a manner correlated roughly with the species’ evolutionary relatedness (see Figure 1). This sequence feature has not been observed at other sites of breakage in *T. thermophila*. The similarity of the central, most highly conserved portion of this sequence to the G4T2 telomeric repeat suggests that it may serve as a preferred site for telomere addition. *In vitro*, telomerase shows a requirement for oligonucleotide primers resembling telomeric repeats (Greider and Blackburn 1987; Harrington and Greider 1991). *In vivo*, at sites of chromosome breakage, no such requirement is evident, although some preference is apparent for telomere addition to initiate following a T or G nucleotide (FAN and Yao 1996). Perhaps, for some reason, the 3′ rDNA breakage site is under more stringent selection for efficient telomere addition. This putative selection could not be absolute since the G/T-rich sequence is not conserved in *T. caudata*, *G. chattoni* or *C. campylum*.

Organization of micronuclear DNA at 5′ rDNA breakage sites: Short inverted repeats near the site of chromosome breakage 5′ to the Tetrahymena rDNA locus directly the formation of the palindromic micronuclear chromosome through a mechanism that involves intramolecular recombination (Yasuda and Yao 1991; Butler et al. 1995). Upstream of these inverted repeats in *T. thermophila* are found three copies of Cbs, one of which has a single base pair alteration from the standard sequence (see Figure 5A) (Yao et al. 1985). We amplified the region between Cbs and micronuclear-retained sequences from two Tetrahymena species, *T. pigmentosa* and *T. hegewischi*, using in both cases one primer matching the *T. thermophila* 15-bp Cbs sequence. The second primer for *T. pigmentosa* (Tp1) was from the previously sequenced nonpalindromic center of the rDNA (Engberg 1988), and for *T. hegewischi* from a well conserved 5′-proximal region of the 17S rRNA-encoding DNA (r5′) (see materials and methods and Figure 2B). This sequence information was used to design two nested PCR primers that are partially or wholly micronuclear-specific and abut the Cbs sequence (see Figure 2D). These primers were used in vectorette PCR (see materials and methods) to amplify Cbs and the sequences further upstream. A portion of the resulting DNA sequence determined is shown in Figure 5A. As expected, the sequence of the *T. hegewischi* 5′ Cbs matches exactly that of its previously determined 3′ counterpart that, in turn, matches the standard *T. thermophila* Cbs. The sequence of the *T. pigmentosa* 3′ Cbs had not been unambiguously defined, but, as shown in Figure 5A, the 5′ Cbs is identical to that of *T. thermophila*.

The 5′ Cbs of *T. pigmentosa* and *T. hegewischi* are located 14 and 11 bp, respectively, further from the junction between micronuclear-limited and micronuclear-retained sequences than is the first of the three *T. thermophila* Cbs’s. Also, the sequence immediately upstream of this junction is strikingly more G/C-rich in the two newly determined sequences. This feature is unexpected from our knowledge of *T. thermophila* breakage sites; all that have been sequenced are A/T-rich.
rich (YAO et al. 1987; YU and BLACKBURN 1991). The most striking difference, however, between T. thermophila and the other two species examined is the presence of only a single copy of Cbs in these two, as compared with the three copies of Cbs in T. thermophila. In over 300 bp of additional upstream sequence (data not shown), no other copies of Cbs were present than those indicated in Figure 5A. The 5' rDNA breakage site of T. thermophila is the only location yet identified where multiple copies of Cbs are found. We will return to this issue below.

As noted, the T. hegewischi 5' and 3' Cbs's are identical in sequence. An additional copy of Cbs from a randomly selected breakage site in T. elliotti was cloned and sequenced (data not shown) and found to be identical to the 3' Cbs of this species. In neither species was there any conservation of sequence beyond the 15-bp consensus. This parallels the situation in T. thermophila, in which Cbs is the only conserved sequence at sites of breakage (YAO et al. 1987).

Structure of the micronuclear inverted repeats that direct palindrome formation: ENGBERG (1983) and KAN and GALL (1981) have previously cloned and sequenced the centers of the macronuclear rDNAs from four Tetrahymena species and described the conservation of 38-bp inverted repeats surrounding their nonpalindromic centers (see Figure 5B). In T. thermophila, these regions have since been shown to correspond to the major portions of the 42-bp micronuclear inverted repeats, also known as M repeats, that direct the formation of the macronuclear palindrome. The significance of this conservation is unclear since unrelated inverted repeat sequences have been shown to substitute for the M repeats in directing palindrome formation in a T. thermophila transformation-based assay (YASUDA and YAO 1991). A portion of this sequence is also conserved in G. chattoni, which does not carry out palindrome formation (CHALLONER and BLACKBURN 1986).

In this study, we have sequenced the corresponding micronuclear region from two additional species of Tetrahymena, T. pigmentosa and T. hegewischi. Figure 5B shows an alignment of micronuclear DNA sequence from the center of the rDNA, with the sequences derived from the micronuclear inverted M repeats underlined. These repeats vary in length from 42 bp in T. thermophila to 41 in T. pigmentosa and 35 in T. hegewischi. The ends of the repeats are also shifted with respect to the nonpalindromic center between T. thermophila and the other two species. The 38-bp conserved repeats identified by ENGBERG are indicated by brackets. The conserved region extends slightly beyond the center-distal end of the M repeats of T. pigmentosa and T. hegewischi. The nonpalindromic center is of a similar length in each species: 28 bp in T. thermophila, 26 in the two others. This central sequence is quite similar between T. pigmentosa and T. hegewischi, and of limited similarity between either of these and the more distantly related (see Figure 1) T. thermophila.

Glaucoma chattoni does not convert its rDNA into a palindrome; instead, the macronuclear rDNA is a monomer, with telomeric repeats at both ends (KATZEN et al. 1981). We amplified, cloned and sequenced the 5' telomere-adjacent sequence from this species. As described previously for another strain (CHALLONER and BLACKBURN 1986), there is similarity between this sequence and part of the M repeats of Tetrahymena species, although no second, inverted copy is found in G. chattoni. Figure 5C shows an alignment of sequences from the T. thermophila M repeat and the G. chattoni 5' rDNA telomere-adjacent region. Unexpectedly, PCR primers from this region of G. chattoni DNA failed to amplify a product in conjunction with Cbs primers of either strand, matching either the T. thermophila or G. chattoni 3' rDNA sequence or the 14 bp they have in common.

**DISCUSSION**

Sequences directing rDNA palindrome formation: Our study is the first to explore the micronuclear organization of the sequences directing rDNA palindrome formation in Tetrahymena species other than T. thermophila. Previous studies have shown that palindrome formation in T. thermophila proceeds by a mechanism involving intramolecular recombination, guided by intrastrand pairing of the 42 bp inverted M repeats, and presumably passing through a hairpin loop intermediate (YASUDA and YAO 1991; BUTLER et al. 1995). Our study has shown that, in T. pigmentosa and T. hegewischi, the M repeat lengths are somewhat shorter (41 and 35 bp, respectively). Such lengths are not unreasonable mechanistically since it has been shown that inverted repeat sequences as short as 29 bp will promote palindrome formation in a T. thermophila transformation assay (YASUDA and YAO 1991).

Evolutionary conservation of the micronuclear sequences derived from these repeats has previously been described (KAN and GALL 1981; ENGBERG 1983). The reason for such conservation is unclear, since unrelated inverted DNA sequences can substitute for the M repeats in promoting palindrome formation in transformation assays (YASUDA and YAO 1991). One possibility is that the particular sequence is indeed unimportant, but mutational drift is slowed by the need to maintain intrastrand pairing ability. It would be instructive to compare the number of nucleotide substitutions in the M repeats over evolutionary time with the number in another region where base pairing is important but primary sequence is apparently not. The D8 variable region of the rRNA seems to fulfill these criteria (SWEENEY et al. 1994). In a region of D8 predicted to contain about 31 Watson-Crick base pairs, five compensatory pairs of nucleotide substitutions have occurred.
between *T. thermophila* and *T. pigmentosa/hegewischi* (both species were identical in D8 sequence) (R. Sweeney and M.-C. Yao, unpublished results). This is quite comparable to the four pairs of substitutions in the 30 bp of the M repeats shared between all three of these species (see Figure 5B). Although comparing constriction of mutational drift due to RNA and DNA secondary structure may not be entirely valid, it seems not unreasonable that all or most of the M repeat conservation observed may be due solely to the requirement for intrastand pairing in palindrome formation. This conclusion is also supported by the fact that recombinant rDNAs containing inverted repeats unrelated to M transform *T. thermophila* efficiently and appear normal in copy number (Yasuda and Yao 1991).

However, two observations are inconsistent with this conclusion. One is the conservation of a portion of the M repeat between Tetrahymena species and *G. chattoni*, which does not form rDNA palindromes (see Figure 5C) (this study and Challoner and Blackburn 1986). The other is that sequence conservation between the Tetrahymena species in this study extends slightly outward from the edges of the M repeats (especially notable in *T. hegewischi*, see Figure 5B) and conservation of the nonpalindromic center is somewhat greater than of the surrounding macronuclear destined sequences. It is thus possible that this region plays an additional subtle role in rDNA replication, amplification or function. It would be interesting to compare the sequence of this region from other tetrahymenine ciliates that do or do not form palindromes. It is notable that the rDNA of *C. campylum* is ~20 kb in size (data not shown) and is therefore likely to be palindromic rather than monomeric. The ability of rDNA to form palindromes may have arisen independently in the Colpidium genus or might have existed in a common ancestor and been lost in *G. chattoni*.

**Implications of breakage site conservation:** As described in the Introduction, chromosome breakage mechanisms in the few ciliate phylogenetic classes that have been studied in detail are substantially dissimilar with regard to the conservation and organization of cis-acting sequences as well as the precision and consequences of the breakage event itself (reviewed in Yao 1989; Prescott 1994). Among the species of our study, however, the most remarkable finding was the lack of diversity in breakage site sequence organization. As in *T. thermophila*, Cbs itself is eliminated from the macronuclear genome, and the site of telomere addition at the 3′ rDNA end is at an approximately equivalent distance from Cbs. The DNA sequence surrounding Cbs is not conserved and its highly A/T-rich nature is probably the result of random mutational drift in a generally A/T-rich genome.

No more than a single base pair difference from the *T. thermophila* Cbs sequence was observed in any of the species examined. All three of these differences were in the terminal two base pairs, corresponding to the 3′ end of the G-rich strand. Interestingly, the only known case of a solo Cbs in *T. thermophila* deviating from the standard 15-bp sequence has a T to A substitution at the second to last base in this same strand (Yu and Blackburn 1991). The furthest upstream of the three Cbs′s at the 5′ rDNA breakage site has the same substitution (Yao et al. 1985), although in this case, due to the triplication of Cbs, it is unknown whether the variant copy is active. Taken together, these results suggest that the trans-acting factor(s) that recognize Cbs must have a high degree of specificity for this particular sequence, since changes in it have only been tolerated in a few positions. Whatever drift may be allowed seems to be confined to one end, at least primarily, where the specificity of interaction with recognition factors may not be as great.

The high level of Cbs conservation may be compared to other DNA sequence signals, such as those promoting transcription, or directing other DNA rearrangement events. Brunke and Sadler (1990) have made an extensive comparison among 22 species of Tetrahymena of the intergenic region between the histone genes H3II and H4II, which are divergently transcribed. Two copies, with opposite polarities, of a 12-bp sequence (11 highly conserved) are found in each species. This sequence contains a canonical “CCAAT” box and is the only well-conserved putative promoter element yet recognized in Tetrahymena. Its conservation is quite high, most species agreeing with the consensus in at least 10 of the 11 highly conserved positions. Similarly well conserved copies are found at comparable upstream positions in four other *T. thermophila* histone genes as well as one unrelated gene each from *T. pigmentosa* and the hypotrich *Stylonychia lemmae*. Thus, a transcriptional promoter element of comparable size to Cbs appears to be conserved to a similar degree. The core of this sequence, apparently being the ubiquitous (among eukaryotes) “CCAAT” box, is presumed to have emerged very early in the eukaryotic lineage.

Another example of short DNA rearrangement signals present in multiple copies in the genome is given by the sequences directing V-D-J joining in vertebrate immunoglobulin gene assembly. Conserved heptamer and nonamer sequences separated by either 12 ± 1 or 23 ± 1 bp of unconserved spacer DNA are found flanking the coding regions to be joined (reviewed in Tonegawa 1983). Although these sequences are not nearly as invariant as Cbs within a given organism, the conservation of the overall consensus during vertebrate evolution has been quite high (reviewed in Reth and Leclercq 1987; Litman et al. 1989). Signal sequences in the primitive horned shark have been found to match the mammalian consensus to a high degree (Litman et al. 1985; Hinds and Litman 1986).

Thus, the degree of conservation of the Cbs sequence and its overall mechanism of action may be comparable.
to that of other important DNA sequence elements. Although this cis-acting signal and the general features of the breakage site are very well conserved in the tetrahymenine ciliates, there is little indication of similarity between this group and other more distantly related ciliates. Thus it may be that the mechanisms of chromosome breakage in separate ciliate lineages are of independent evolutionary origin. Whether the occurrence of breakage itself had several independent origins (ORTAS 1991; HERRICK 1994) is a matter of speculation. It may be, for example, that chromosome breakage by a Cbs-based mechanism took over from, or is a refinement of, a more primitive, less precise mechanism found in a common ancestor of, for example, Paramaecium and the tetrahymenine ciliates.

Differences in Cbs copy number and orientation: The inversion of the 3′ Cbs in G. chattoni and C. campylum as compared to the Tetrahymena species and the triplication of the 5′ Cbs′s of T. thermophila as compared with T. pigmentosa and T. hegewichi suggest that two separate micronuclear rearrangements have occurred and become fixed during the evolutionary time period that has separated the relevant lineages.

Another line of evidence that supports the hypothesis of a recent evolutionary origin for the 5′ Cbs triplication in T. thermophila is that the 30-bp sequences on either side of the middle Cbs, separating it from the outer copies, are nearly identical to each other (28/30 bp identity. See Figure 5A) (YAO et al. 1985). We would not expect this to be the result of functional conservation for two reasons. First, the DNA sequences surrounding Cbs have not been shown to have a strong influence on the efficiency of chromosome breakage or end-healing by telomerase in transformation assays (YAO et al. 1990; FAN and YAO 1996). Second, no interspecies conservation of sequences surrounding Cbs has been observed at the 5′ or 3′ rDNA breakage sites, nor intraspecies conservation at the various sites in T. thermophila (YAO et al. 1987; YU and BLACKBURN 1991).

We suggest a more likely explanation to be an initial species conservation of sequences surrounding Cbs has been observed at the 5′ or 3′ rDNA breakage sites, followed by a recent expansion to three copies by an undefined mechanism, followed by the development of a more precise, less primitive mechanism found in a common ancestor of, for example, Paramaecium and the tetrahymenine ciliates.

LITERATURE CITED


ENGBERG, J., 1983 Strong sequence conservation of a 36 bp region near the center of the extrachromosomal rDNA palindrome in different Tetrahymena species. Nucleic Acids Res. 11: 4939–4946.


KING, B. O., and M. C. YAO, 1982 Tandemly repeated hexanucleotide at *Tetrahymena* rDNA free end is generated from a single copy during development. *Cell* 51: 177–182.


Communicating editor: S. L. ALLEN