Gene Amplification in *Tetrahymena thermophila*: Formation of Extragloboidal Palindromic Genes Coding for rRNA

MENG-CHAO YAO,* SHENG-GENG ZHU,† AND CHING-HO YAO

Department of Biology, Washington University, St. Louis, Missouri 63130

Received 26 November 1984/Accepted 25 February 1985

*Tetrahymena thermophila* contains in the macronucleus multiple copies of extrachromosomal palindromic genes coding for rRNA (rDNA) which are generated from a single chromosomal copy during development. In this study we isolated the chromosomal copy of rDNA and determined the structure and developmental fate of the sequence surrounding its 5' junction. The result indicates that specific chromosomal breakage occurs at or near the 5' junction of rDNA during development. The breakage event is associated with DNA elimination and telomeric sequence addition. Similar results were also found previously for the 3' junction of this gene. These results could explain how the extrachromosomal rDNA is first generated. Near both junctions of the chromosomal rDNA, a pair of 20-nucleotide repeats was found. These sequences might serve as signals for site-specific breakage. In addition, we found a pair of perfect inverted repeats at the 5' junction of this gene. The repeats are 42 nucleotides long and are separated by 28 nucleotides. The existence of this structure provides a simple explanation for the formation of the palindromic rDNA.

Selective gene amplification is a prominent feature of many eucaryotes, in which the rRNA genes (rDNAs) are selectively amplified several hundredfold during development (5, 12). Amplification of rDNA is now known to occur in a wide variety of organisms (reviewed in references 31 and 36). Amplification of other genes, such as the genes coding for chorion proteins, has also been observed in some insects (14, 30). The process clearly plays an essential role in the normal development of many organisms. Recently gene amplification has also been found in cultured mammalian cells in the presence of specific metabolic inhibitors (reviewed in references 7 and 27) and in conjunction with the formation of some tumor cells (1, 28). Thus, this process apparently occurs widely and participates in both normal and abnormal cellular growth.

The amplification of rDNA in *Tetrahymena thermophila* is among the best studied cases of gene amplification (reviewed in reference 36). *T. thermophila*, a ciliated protozoan, contains a macronucleus which produces most of the cellular RNA and a micronucleus which is inactive in transcription during vegetative growth. The micronucleus is the germinal nucleus. It gives rise to both the micronucleus and the macronucleus of the next sexual generation, after the sexual event of conjugation. The diploid micronucleus contains a single copy of rDNA per haploid genome (37). During development of the macronucleus, this chromosomal copy of rDNA is replaced by approximately 9,000 copies of linear extrachromosomal molecules (11, 13, 38). Each of the extrachromosomal molecules is 21 kilobases (kb) in size and contains two copies of the rDNA arranged as reverse repeats, with the 5' end of the genes located near the center of the molecule (10, 20). These molecules are maintained as autonomous replicating units during the subsequent vegetative life of the cell. Thus, the amplification of rDNA in *T. thermophila* involves the formation of the extrachromosomal dimeric genes from the single integrated copy and the selective replication of this molecule during development.

To understand how the extrachromosomal palindromic rDNA is generated in *T. thermophila*, we have been analyzing the structures and the developmental fates of the DNA surrounding the two ends of the single integrated rDNA. Nucleotide sequencing of the extrachromosomal rDNA has revealed that there are 29 nucleotides of nonpalindromic sequence at the very center of this otherwise palindromic molecule (18, 23). Moreover, the free ends contain from 20 to 70 tandem repeats of the hexanucleotide CCCCCA (4). This sequence, which is also found at the termini of other macronuclear DNA molecules in *T. thermophila* (39) and several other ciliates (21, 39), probably functions as a telomere to stabilize the free ends. The structure of the integrated copy of rDNA was first determined by hybridization of the genomic DNA and was found to be a single copy sequence and not a palindrome (37). Cloning and direct analysis of the DNA sequences at the 3' end of the integrated micronuclear gene have revealed some interesting features of the amplification process. First, the 3' end of the integrated gene must become disconnected from the adjacent flanking sequence by a break, since the macronucleus contains both of these sequences located near free ends (35). Second, approximately 3 kb of the adjacent flanking DNA is missing from the macronucleus (35). Third, the CCCCCA repeats found at the termini of the extrachromosomal rDNA are not present at the 3' end of the integrated copy. Instead, a single CCCCCA hexanucleotide is found near this site. Clearly, the telomeric CCCCCA repeats are not inherited and are added on during development (22). These results suggest that the formation of the extrachromosomal rDNA is by way of site-specific DNA breakage and that the breakage event is associated with the elimination of DNA from and the addition of terminal repeats to the free ends.

To further understand how the extrachromosomal palindromic rDNA is generated, we have now cloned the entire integrated copy of rDNA and determined the structure and developmental fate of the DNA surrounding its 5' end. We found evidence for DNA breakage, DNA elimination, and telomere addition at this site also. Moreover, DNA sequence

* Corresponding author.
† Present address: Department of Biology, Peking University, Beijing, People's Republic of China.
analysis revealed several interesting structures which may have crucial roles in chromosome breakage and palindrome formation.

MATERIALS AND METHODS

Cell cultures and DNA isolation. The inbreeding line B1868 mating type IV of T. thermophila was obtained from P. Bruns of Cornell University and maintained in axenic culture as described previously (15). The isolation of nuclei and nuclear DNA was by the methods reported earlier (15, 38).

DNA labeling, mapping, and sequencing. Nick translation of DNA was carried out by the method of Rigby et al. (26). 32P-labeled dCTP and dATP (3,000 Ci/mmol; Amersham Corp.) are used to label the DNA. Restriction enzymes and Bal 31 were purchased from New England BioLabs, Inc., or Bethesda Research Laboratories. Digestion with Bal 31 was carried out as previously described (35). Agarose gel electrophoresis of DNA, blotting of DNA onto nitrocellulose filters, and hybridization of the filters were carried out as previously reported (37). The sizes of DNA fragments were determined by gel electrophoresis, with HindIII-digested phage λ DNA as references. Nucleotide sequences of DNA fragments were determined by the chemical method of Maxam and Gilbert (25). The DNA fragments were labeled at the 3' end by using the large fragment of Escherichia coli DNA polymerase. The sequencing gel electrophoresis was carried out with 1- by 3-ft (ca. 30- by 91-cm) gels.

DNA cloning. To construct the genomic library, micronuclear DNA was digested to various degrees with the restriction enzyme BglII. The digested DNAs were combined, and fragments ranging from 10 to 20 kb were isolated by sedimentation through a sodium chloride gradient. These fragments were ligated to BamHI-digested DNA of λ1059 (19) and packaged by a modified version of the method of Hohn (16). The recombinant DNA library was screened by the plaque hybridization method of Benton and Davis (3). Subcloning of cTt 506 was carried out with the vector pUC13 (32).

RESULTS

Cloning the integrated rDNA. A major problem in cloning the integrated copy of rDNA lies in the fact that there are approximately 9,000 copies of the extrachromosomal rDNA and only two copies of the integrated rDNA in a cell (37, 38). Although micronuclei can be isolated in large quantities, the preparation usually contains 10 to 20% of the macronuclear DNA as a contaminant (15), resulting in 20 to 40 times more extrachromosomal rDNA than integrated rDNA in the preparation. To overcome this problem, we designed an approach to discriminate against the macronuclear rDNA during cloning. This approach relies primarily on the fact that the free ends of the extrachromosomal rDNA are different from ends produced by restriction enzyme digestion and thus cannot be directly ligated to vector DNA by conventional methods (35). A phage λ1059 DNA library was constructed which contained micronuclear DNAs partially digested by BglII as inserts. Under this condition, the integrated rDNA can be cloned in its entirety as part of a 14-kb fragment (Fig. 1). This DNA library was screened by hybridization with a HindIII fragment of the extrachromosomal rDNA as a probe. This probe, HindIII-E, is located near the 3' end of the rDNA and should hybridize only with the terminal BglII fragment of the extrachromosomal rDNA or the corresponding fragment of the integrated rDNA. Since the terminal extrachromosomal rDNA fragment could not be incorporated into our library due to the presence of its telomeric end, the only clones hybridized should have contained the integrated rDNA. Of 9 × 10^4 clones screened, 9 were positive in hybridization and were thus expected to contain the 3' end of the integrated rDNA. To determine which of the clones also contained the 5' end of the rDNA, they were hybridized with HindIII-B, a fragment which is located near the center of the extrachromosomal rDNA. Five clones, i.e., cTt 504, cTt 506, cTt 507, cTt 509, and cTt 510, hybridized strongly and were analyzed further.

cTt 504, cTt 507, and cTt 510 were indistinguishable from one another when analyzed by EcoRI, BglII, HindIII, and HaeIII digestion and by hybridization with the rDNA. cTt 506 contains the same Tetrahymena DNA insert in the opposite orientation. A restriction map of cTt 506 is shown in Fig. 2. This clone contains a 14-kb insert that hybridizes strongly with the extrachromosomal rDNA and has a restric-
tion map indistinguishable from that of the integrated rDNA determined previously by genomic hybridization (37; M.-C. Yao, unpublished data). We thus conclude that these clones indeed contain the desired integrated rDNA sequence. The other clone, cTt 509, is basically the same as cTt 504, except that there is a 0.2-kb deletion near the 5' end of the rDNA. It should be noted that the Tetrahymena strain we have used in this study is a product of 18 generations of inbreeding and contains a homozygous micronucleus. Thus, although the deletion observed in clone cTt 509 could represent a rare, mutated allele in the population of cells used as the source of DNA for the library construction, it is more likely to be the result of cloning.

Site-specific chromosomal breakage. We showed previously that chromosomal breakage occurs at or near the 3' end of the integrated rDNA during macronuclear development. To understand how the integrated rDNA was removed from the chromosome, it was important to determine whether breakage also occurred at the 5' end of the rDNA. An 800-base-pair (bp) TaqI fragment of cTt 506 (cTt 506-T4), which contains the chromosomal DNA immediately outside the 5' end of the rDNA, was subcloned into pUC13 and used as a hybridization probe to determine the structure of the rDNA flanking sequence in the macronucleus. Figure 3 shows the result of the hybridization of this sequence to restriction enzyme-digested macro- and micronuclear DNA, and Fig. 4 shows the restriction maps derived from this study. In either nuclear DNA, a single band was detected. Thus, this sequence is probably present only once in the micronuclear genome, and at least some part of it is retained in the macronucleus. However, in most cases, the macronuclear fragment detected was different in size from the micronuclear fragment. Detailed analysis of the restriction maps suggested that the differences were due to the presence of a free end in the macronuclear DNA (Fig. 4). This argument was best supported by the fact that regardless of which of the 11 restriction enzymes (BglII, HaeIII, XbaI, HindIII, EcoRI, MnlI, Sau3AI, Hinfl, AluI, TaqI, and HhaI) was used, one aligned the left cutting site of the macronuclear DNA with that of the micronuclear DNA, the right end of the macronuclear DNA fragment always fell at the same point near the rDNA junction (Fig. 4).

The presence of a free DNA end in this region was further supported by the study with the exonuclease Bal 31. We showed previously that Bal 31 can progressively shorten free ends of macronuclear DNA (35, 39). Figure 5 shows how the flanking sequence of the rDNA is also sensitive to Bal 31 treatment. When high-molecular-weight whole-cell DNA, which contains mostly macronuclear DNA, was treated with Bal 31 and analyzed by restriction enzyme digestion and Southern hybridization, the 5' flanking sequence was progressively shortened. From this result and the restriction enzyme mapping data, we conclude that the 5' end of the rDNA, like the 3' end, is disconnected from the rest of the chromosomal DNA by a specific double-strand DNA break during the formation of the macronucleus.

At the 3' end of the rDNA, chromosomal breakage is accompanied by the elimination of approximately 4 kb of the flanking sequence from the macronucleus (35). Here we find that DNA elimination also occurs at the 5' end, although to a much lesser extent. Judging from the hybridization result, the macronucleus apparently contains a significant portion of the 800-bp sequence of cTt 506-T4, which is immediately adjacent to rDNA. The fact that some flanking sequence is indeed eliminated is suggested by the study with MnlI digestion. From the 3' end sequence it is known that there are three MnlI cutting sites within the 102 bp of DNA immediately flanking the rDNA (see below). None of these sites is found in the flanking sequence of the macronucleus (Fig. 3 and 4). Furthermore, using a synthetic probe of a 20-nucleotide sequence from this region (A' in Fig. 6), we failed to detect any homologous sequence in the macronucleus (data not shown). Thus it is likely that this 102-bp sequence is eliminated from the macronucleus. From these results it is concluded that more than 102 and much less than 800 bp of DNA is probably eliminated from the flanking region after chromosomal breakage at the 5' end of the integrated rDNA.

The free end of the flanking DNA sequence bears some interesting features. First, it appears to be heterogeneous in size. This conclusion is based on the observation that the restriction fragments containing this end migrate as broad bands in gels (Fig. 3). This kind of heterogeneity is common among Tetrahymena macronuclear DNA ends studied (4, 35) and has also been found in other eucaryotes (8, 17). Second,
approximately 1 kb of DNA near the free end is different in sequence from the corresponding region of the micronuclear genome (indicated by the hatched box in Fig. 4). This 1-kb area contains none of the cutting sites of TaqI and MnlI found in the micronuclear DNA and fails to hybridize with the cloned micronuclear sequence (data not shown). Since many, if not all, of the free DNA ends in the Tetrahymena macronucleus contain the repeated hexanucleotide (CCCCAA)₄ (4, 39), this terminal sequence is most likely also a repeated CCCCCA. Regardless of its nature, the sequence is not directly derived from the micronuclear sequence and is probably added to the free end after breakage occurs.

**Sequence analysis of the 5' end of the rDNA.** To further understand the mechanism of chromosomal breakage and palindrome formation, we determined the nucleotide sequence near the 5' end of the integrated copy of the rDNA. The nucleotide sequences at the center of the extrachromosomal rDNA were reported previously (23). A 29-nucleotide nonpalindromic sequence which separates the palindromic sequences on both sides was detected at the very center of the extrachromosomal rDNA molecule. Comparison of this sequence with the sequence at the 5' end of the integrated rDNA should allow us to localize the exact 5' end of the integrated copy. Restriction fragments of cTt 506 containing the 5' end of the integrated rDNA were cloned into the vector pUC13 and sequenced by the method of Maxam and Gilbert (25). The sequencing strategy is shown in Fig. 2. To improve accuracy, all sequences were determined at least three times, and in some cases both DNA strands were analyzed. The 1,935-nucleotide sequence determined from this region (Fig. 6) reveals several interesting features which might be important in the formation of the extrachromosomal palindromic genes.

First, the very 5' end of the integrated rDNA contains a pair of perfect inverted repeats (M in Fig. 6). The repeats are 42 bp long and are separated from each other by 28 nucleotides. This structure is essentially identical to that of the 112 nucleotides at the very center of the extrachromosomal rDNA. This structure could help explain how the palindromic rDNA is generated during development (see below for details). Except for a difference of a single nucleotide in the central nonpalindromic sequence, our sequence of the integrated rDNA agrees well with the published sequence of...
the extrachromosomal rDNA. Thus, the amplified rDNA has faithfully retained the sequence of the inherited rDNA in this region.

Second, there are repetitive sequences flanking this end of the integrated rDNA. Within 116 bp of the 5' end of the rDNA, we found two copies each of an 8-nucleotide sequence (B in Fig. 6) and a 21-nucleotide sequence (C in Fig. 6) and three copies of a 14-nucleotide sequence (A in Fig. 6). The sequences are arranged in the order ABCABCA, with B and C separated from each other by 2 nucleotides. Most significantly, a 20-nucleotide sequence (A' in Fig. 6), which makes up part of sequences C and A, is also present once in reverse orientation in the flank region at the 3' end of the integrated rDNA (Fig. 7; 22). The sequence outside the 3'
end differs from the two copies outside the 5' end by 1 nucleotide and is located 21 nucleotides from the rDNA. Due to their locations and close homology, it is likely that the A' sequences could play a significant role in site-specific chromosomal breakage.

Finally, we did not find any CCCCCA sequence within the 1,935-nucleotide region sequenced. This region surrounding the 5' end of the integrated rDNA should include sequences for at least one breakage site and two free ends, one of which is maintained as such in the macronucleus, whereas the other eventually becomes the center of the palindromic rDNA. Although we do not know the exact location of the breakage site or the exact sequence maintained at the free end, it is clear that, at least in this case, the presence of the hexanucleotide CCAACC in the micronucleus is not required for chromosomal breakage or for the addition of terminal sequences during development.

DISCUSSION

T. thermophila is the only eucaryote known to contain a single integrated copy of rDNA in the germline genome (37). During development, selective amplification of this gene occurs such that the integrated copy is replaced by a large number of extrachromosomal palindromic copies in the somatic nucleus. This unique feature makes T. thermophila a good system in which to study the early events of gene amplification. In this study we isolated the single integrated copy of rDNA by molecular cloning. Analysis of the structure and the developmental fate of the DNA surrounding this gene reveals many interesting features and for the first time provides us with a coherent picture of the formation of the extrachromosomal palindromic rDNA. Figure 8 summarizes these results and our speculations on this process.

Chromosomal breakage. There is little doubt that the first extrachromosomal copy of rDNA is generated from the integrated copy through chromosomal breakage at or near the two ends of the gene. Breakage near the 3' end was detected in an earlier study from this laboratory (39). In the current study, we documented the occurrence of breakage near the 5' end by a similar method involving the use of restriction enzymes and Bal 31 exonuclease. Thus, the first event in rDNA amplification in T. thermophila appears to involve site-specific chromosomal fragmentation.

Since the breakage occurs selectively at or near the two ends of the integrated rDNA, it would be interesting to know how the specificity is achieved. A simple mechanism might involve recognition of specific nucleotide sequences near or at the breakage sites by an endonuclease produced during this time of development. In this regard, it is highly signifi
cant that a pair of nearly identical sequences are present near the two ends of the integrated rDNA. If one compares the sequence at the 5’ end shown in Fig. 6 with the sequence at the 3’ end which we determined previously (22), a region of homology is found (Fig. 7). These two sequences are identical in 19 of the 20 nucleotides and are located 21 nucleotides from the 3’ end and 14 nucleotides from the 5’ end of the rDNA. It is tempting to suggest that they are the recognition sequences for breakage or that they are at least somehow involved in the breakage process. In this sense, it is interesting that the hexanucleotide CCAACC is part of the homologous sequence. Tandem repeats of CCAACC are indistinguishable in internal structure from repeats of CCCCAA, which is found at most macronuclear DNA ends. Since chromosomal breakage is usually accompanied by the addition of CCCCAA repeats at the broken ends (22, 39), it would be interesting if similar hexanucleotides are involved in both processes.

**DNA elimination.** The breakage events are accompanied by DNA elimination. Approximately 3 kb of the DNA flanking the 3’ end and 0.1 to 0.8 kb of DNA flanking the 5’ end of rDNA are eliminated from the macronucleus. It is possible that these sequences are eliminated as the result of exonuclease digestion at the broken ends, although other possibilities also exist and cannot be excluded at present. Although the precise extent of the elimination is not known, the entire eliminated sequence is included in the region that has been sequenced. Since this region is the first of its kind to be sequenced, it is worthwhile to note that no apparent structure, such as a long open reading frame, is present. Thus, a defined region of germ line-specific DNA does not necessarily code for a germ line-specific protein.

**Telomeric DNA addition.** After chromosomal breakage and DNA elimination, the free ends generated are modified by the addition of telomeric DNA such as CCCCAA repeats. The best-known case involves the 3’ end of the rDNA, to which 20 to 70 tandemly repeated copies of CCCCAA are added after breakage occurs (22). A similar situation also exists for the 5’ flanking sequence. From restriction mapping and hybridization studies, it is clear that approximately 1 kb of DNA next to this free end in the macronucleus is not derived from the same region of the micronucleus and must therefore be added after breakage occurs. Although this particular macronuclear DNA end has not been isolated, it probably also contains the CCCCAA repeat typical of most free macronuclear DNA ends (39). A similar analysis was also made on the 3’ flanking sequence, and the results suggest that approximately 1 kb of DNA is also added to this free end after breakage occurs (M.-C. Yao, unpublished data). Thus, the addition of telomeric sequences appears to be a common phenomenon associated with chromosomal breakage in this ciliate. It is conceivable that this process is necessary for stabilizing and propagating the free end generated. How CCCCAA repeats are added to the free ends is not known. Analysis of the 3’ end of rDNA suggests that the presence of a single copy of CCCCAA near this site may be important in the formation of the tandem repeats. This feature is probably the case for this site, but is certainly not a general feature for CCCCAA addition. In the entire region surrounding the 5’ end sequenced here, not a single CCCCAA sequence is found. Thus, this hexanucleotide cannot be a general requirement for chromosomal breakage or telomere formation in *Tetrahymena*.

Besides being found at the two ends of rDNA, chromosomal breakage is found in many other sites of the *Tetrahymena* genome during development (39). It is also found in other ciliates (24), as well as in *Ascaris* sp. (reviewed in reference 34) and *Cyclops* sp. (2), as a phenomenon associated with DNA elimination from somatic cells or nuclei. In all these cases, the free ends generated are propagated for many generations during the somatic life. It is conceivable that chromosomal breakage, DNA elimination, and telomere addition are tightly linked events in these cases, as they probably are in *Tetrahymena* rDNA. Thus, the molecular details we have learned here may lead to a better understanding of these developmental processes.

**Palindrome formation.** A distinctive feature associated with rDNA amplification in *T. thermophila* is the formation of the palindromic molecule (10, 20). The first extrachromosomal copy of rDNA produced through chromosomal breakage would be a 10.5-kb monomer. Somehow this molecule becomes the 21-kb palindromic molecule found in the mature macronucleus. Our study here reveals an interesting structure of the rDNA which may explain how the palindromic molecule is formed. In a previous study we suggested that a giant palindromic molecule could be formed from a single copy sequence if this sequence contains a pair of short reverse repeats at one end (37). When the 5’ end of the rDNA is analyzed, we indeed find such a structure. At the 5’ end are a pair of reverse repeats which are 42 nucleotides long and are separated from each other by 28 nucleotides. This 112-nucleotide sequence is essentially the same structure which makes up the very center of the palindromic rDNA. We suggest that this structure is crucial in the formation of the giant palindromic rDNA. Figure 8 shows how the short reverse repeats might participate in the formation of the 21-kb palindrome. Essentially, the short repeats could facilitate intramolecular recombination between the repeats. If only one of the two DNA strands is recombined, a long hairpin molecule is produced, which contains the entire rDNA sequence, including its replication origin, and which would become the 21-kb palindromic molecule after one round of replication.

The idea that the 42-nucleotide repeats are important in palindrome formation is also supported by the fact that this sequence is highly conserved during evolution. Sequences near the center of the extrachromosomal rDNA have been determined in four *Tetrahymena* species (9, 18). The first 38 nucleotides of the 42-nucleotide repeats described here are nearly identical among these species, whereas the rest of the sequences show little homology. The conservation of this sequence may reflect the needs of the organism to maintain a perfect reverse repeat at the 5’ end of the integrated rDNA.

Long stretches of inverted duplications have been observed in rho+ mitochondrial DNA of yeast cells. Recent analysis indicates that the duplications are formed at sites which in the wild-type genome contain short reverse repeat sequences (29). This situation closely resembles that observed here for *Tetrahymena* rDNA. Thus, the formation of a long reverse repeat from a short one is not a unique feature of *T. thermophila*. It is possible that this process occurs widely in eukaryotes. In this regard, it would be interesting to know whether the extrachromosomal palindromic rDNAs found in *Dictyostelium* sp. (6) and *Physarum* sp. (33) are formed by a similar mechanism.

**ACKNOWLEDGMENTS**

We thank Marsha Altschuler for critical reading of the manuscript and Tom Cech for valuable comments on the model of palindrome formation.

This work was supported by research grant GM26210 from the National Institutes of Health. M.-C.Y. is the recipient of research
COLORADO

LITERATURE CITED

Downloaded from http://mcb.asm.org on July 6, 2017 by guest