Nucleic Acids Research

Conservation of sequences adjacent to the telomeric C_4A_2 repeats of ciliate macronuclear ribosomal RNA gene molecules

Peter B.Challoner[1] and Elizabeth H.Blackburn

Department of Molecular Biology, University of California, Berkeley, CA 94720, USA

ABSTRACT
    We sequenced and compared the telomeric regions of linear rDNAs from vegetative macronuclei of several ciliates in the suborder Tetrahymenina. All telomeres consisted of tandemly repeated C_4A_2 sequences, including the 5' telomere of the 11 kb rDNA from developing macronuclei of Tetrahymena thermophila. Our sequence of the 11 kb 5' telomeric region shows that each one of a previously described pair of inverted repeats flanking the micronuclear rDNA (Yao et al., Mol. Cell. Biol. 5: 1260-1267, 1985) is 29 bp away from the positions to which telomeric C_4A_2 repeats are joined to the ends of excised 11 kb rDNA. In general we found that the macronuclear rDNA sequences adjacent to C_4A_2 repeats are not highly conserved. However, in the non-palindromic rDNA of Glaucoma, we identified a single copy of a conserved sequence, repeated in inverted orientation in Tetrahymena spp., which all form palindromic rDNAs. We propose that this sequence is required for a step in rDNA excision common to both Tetrahymena and Glaucoma.

INTRODUCTION

    Telomeres are specialized structures required for the stabilization and replication of chromosome ends (reviewed in 1 and 2). The most distal portions of all known telomeric regions consist of tandem repeats of simple G + C-rich DNA sequences (2-9). For example, the termini of the subchromosomal linear macronuclear DNA molecules in holotrichous ciliates such as Tetrahymena thermophila and Glaucoma chattoni have ~30 to ~70 tandem repeats of the hexanucleotide C_4A_2 (4, 5). In addition to these simple sequence repeats, in Physarum and Dictyostelium rDNAs, and in Trypanosoma brucei and yeast chromosomes, the regions immediately proximal to these terminal telomeric repeats (designated telomere-associated regions (1)) include a variety of more complex tandemly repeated sequences (6-10). By in situ hybridization, repeated sequences have been localized to telomeric regions in a number of other species (reviewed in 1). However, a systematic study of telomere-associated sequences has not previously been done, in order to compare these regions in homologous molecules from a group of related organisms.

TABLE 1   Relationships between ciliates and rDNA telomeres analyzed

Suborder	Family	Genus	Species	Strain	rDNA Telomere*
Tetrahymenina	Glaucomidae	Glaucoma	chattoni	HZ-1	3'
"	"	"	"	"	5'
"	Tetrahymenidae	Tetrahymena	thermophila	B1868	3'
"	"	"	"	"	5'
"	"	"	"	C3-368	3'
"	"	"	pigmentosa	ALP15	3'
"	"	"	"	UM1060	3'

*5' and 3' refer to telomeres respectively upstream and downstream of the rRNA transcription unit.

In this paper, we analyze the sequences adjoining telomeric C_4A_2 repeats in the macronuclear ribosomal RNA gene (rDNA) molecules from several species of ciliates in the suborder Tetrahymenina. Table 1 lists the relationships between the various tetrahymenids whose rDNA telomere-associated sequences we examined in this study. The species compared were chosen to reveal sequence conservation resulting from functional constraints on telomere-associated sequences. Tetrahymena and Glaucoma are two evolutionarily distant genera; thus any conservation of sequences between them is likely to reflect strong functional constraints on those sequences. We expected that any such conserved sequences would be important for telomere function in the vegetative macronucleus (1, 11), for the specificity of rDNA excision during macronuclear development (12-14), and/or for de novo formation of macronuclear telomeres in the developing macronucleus (14, 15). Conversely, sequences specifically involved in the formation of the palindromic rDNA present in Tetrahymena should not be conserved in the rDNA of Glaucoma, which is non-palindromic (5). We also analyzed the rDNA telomere-associated sequences of different strains within the same species of Tetrahymena, to determine intraspecies rates of sequence divergence, and thus to reveal functional domains in telomere-associated sequences.

MATERIALS AND METHODS
Preparation of rDNAs from Tetrahymena thermophila, Tetrahymena pigmentosa and Glaucoma chattoni
    Tetrahymena thermophila strains B1868(IV) and C3-368(V) and Glaucoma chattoni strain HZ-1 were maintained and grown as described previously (5, 12,

16). Tetrahymena pigmentosa strains ALP15(II8) and UM1060(II6) were kindly provided by D. Nanney, University of Illinois, Urbana, and were maintained and grown in the same way as T. thermophila as described previously (12, 16). Macronuclear rDNA was prepared from all ciliates as described previously (12, 16, 17).

Recombinant DNA Techniques and DNA Sequence Analysis

All plasmids containing inserts of rDNA telomeric regions were constructed as described for pTre1, pGre2 and pGre7 (16), except plasmid prE-Y3, which was contructed as described previously (18). Plasmid preparations, DNA sequencing and computer analysis of DNA sequences were all done as described previously (15, 16).

RESULTS AND DISCUSSION

Cloning and sequence analysis of rDNA telomere-associated sequences

The macronuclear rDNA molecules of Tetrahymena are linear palindromic dimers. Early in development of a new macronucleus free, single RNA gene molecules, called 11 kb rDNA molecules, are generated by excision of the single micronuclear gene (12-14). The 11 kb rDNA is lost in subsequent vegetative growth. However, in Glaucoma chattoni, only single rRNA gene molecules, 9.1 kb in length, are found in the macronucleus (5). The telomeric regions of the rDNA molecules of Glaucoma chattoni, Tetrahymena thermophila and Tetrahymena pigmentosa were each cloned using a method which ensured that the junction of the telomeric C_4A_2 repeat block and the telomere-associated region was included in the insert (8). Briefly, the macronuclear rDNA molecules were isolated, digested with S1 nuclease to reduce the telomeres to blunt ends, and subsequently digested with a restriction endonuclease. The telomeric rDNA fragments were ligated into a pBR322 vector fragment produced by double digestion with PvuII and the restriction enzyme used on the rDNA. Recombinant clones were identified by hybridization to a repeated C_4A_2 probe. Restriction maps of the macronuclear rDNA molecules and the recombinant plasmids containing their telomeric regions are shown in Fig. 1. Plasmids pTre1, pGre2 and pGre7 have been described elsewhere (16). Restriction digestion mapping and hybridization to rDNA probes confirmed that in all plasmids the rDNA sequences were not rearranged by the cloning process. The DNA sequences of the telomere-associated regions, and of the C_4A_2 blocks themselves, were determined by Maxam-Gilbert sequencing.

The only sequence common to all tetrahymenid rDNA telomeres is repeated C_4A_2.

Hybridization with a repeated C_4A_2 sequence probe indicated that all the

cloned rDNA telomeric regions contained this sequence in the expected rDNA-
vector junction restriction fragment. Sequence analysis showed that all
cloned rDNA telomeric regions terminated in C_4A_2 repeats: pTre1 had 56 C_4A_2
repeats, pGre2 and pGre7 had 34 and 38 C_4A_2 repeats respectively, and pALP3
and pUM9 had 52 and 14 repeats respectively. In previous work (12), C_4A_2
repeats had not been detected by hybridization to the 5' telomere of the 11 kb
rDNA molecule, possibly because of some breakdown of the DNA sample analyzed,
but the cloned 11 kb rDNA 5' telomere terminated in 40 C_4A_2 repeats. The only
deviation from the canonical C_4A_2 hexanucleotide repeat in any of the clones
was a single C_5A_2 unit in pTre1 (M.L. Budarf and E.H. Blackburn, manuscript in
preparation).

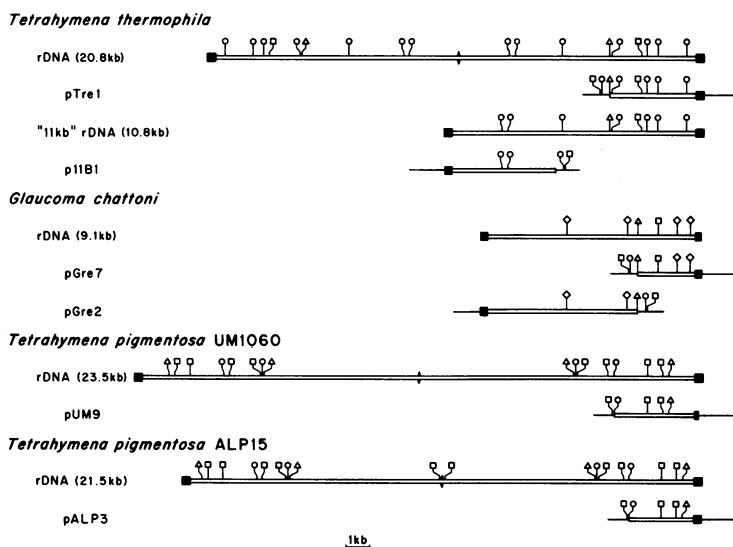To estimate the number of C_4A_2 repeats present in the original macro-

Fig. 1. Physical maps of plasmids with rDNA inserts.
The plasmids described in the text are shown aligned below the rDNA molecule
that was the source of the insert. All plasmid maps are presented as if
linearized at the PstI site at pBR322 position 3608, so that the single EcoRI
site at position 0/4362 in the pBR322 portion of the plasmids gives the rela-
tive orientation of the vector and insert. Thin line = pBR322; open bar =
rDNA; filled box = C_4A_2 repeats; arrowheads = dyad axis of symmetry in palin-
dromic rDNA molecules. All recognition sites for BamHI (Δ), EcoRI (□), and
HindIII (O) are shown. As G.chattoni rDNA has no HindIII sites, HpaI (◊)
recognition sites are given for this molecule and the plasmids pGre2 and pGre7
only. Cloned 5' and 3' telomeres are located respectively toward the left and
right sides of their plasmid maps. The direction of rRNA transcription on the
single rRNA gene molecules (G.chattoni 9.1 kb and T.thermophila 11 kb rDNAs),
and on the right halves of the palindromic rDNA molecules, is from left to
right.

nuclear rDNA molecules, the sizes of telomeric restriction fragments of rDNA molecules isolated from the respective ciliate nuclei were compared with the lengths of their cloned terminal restriction fragments. These size measurements showed that the Tetahymena macronuclear rDNA telomeres, which are variable in length and form diffuse bands in gel electrophoresis (4), contained 50 ± 15 C_4A_2 repeats. The Glaucoma macronuclear rDNA telomeres terminated in 30 ± 10 C_4A_2 repeats.

The telomere-associated sequences which adjoin the C_4A_2 repeats are shown in Fig. 2. Each telomere-associated sequence was compared with all others to locate regions of homology, using the computer programs Matrix and Align (see Materials and Methods). When corresponding 3' telomere-associated regions of the rDNAs were compared among the different tetrahymenid species, no long homologies apart from the C_4A_2 repeats were found. Even the telomeres at each end of rDNA molecules in the same species (the Glaucoma 9.3 kb macronuclear rDNA and the T. thermophila 11 kb rDNA) had no sequences in common besides the C_4A_2 repeats. Thus, the first general finding to emerge is that extensive co-

Fig. 2. DNA sequences of the terminal ~220–250 bp of the telomere-associated regions of tetrahymenid macronuclear rDNAs.
Numbering begins at the left at the first nucleotide of the innermost C_4A_2 repeats shown and proceeds in the 5'→3' direction toward the interior of the rDNA molecule. 5' and 3' telomeres are as defined in Table 1. C_4A_2 repeats are boxed. Sequences with homology to the yeast ARS consensus sequence are underlined. Dashes are inserted to maximize alignment between the sequences from different strains of the same species. 1 = T.pigmentosa ALP15 3' (from plasmid pALP3); 2 = T.pigmentosa UM1060 3' (plasmid pUM9); 3 = T.thermophila B1868 3' (plasmid pTre1); 4 = T.thermophila C3-368 3' (plasmid prE-Y3, ref. 18); 5 = G. chattoni HZ-1 3' (plasmid pGre7); 6 = T.thermophila B1868 5' (plasmid p11b1); 7 = G.chattoni HZ-1 5' (plasmid pGre2). Bases 9–33 and 62–103 in the T.thermophila B1868 5' telomere associated sequence are inverted M repeats (ref. 14; see text).

linear regions are found only between analogous telomeres of strains within
the same species.

We next searched for any shared sequence motifs or short homologies in
the rDNA telomere-associated sequences. No internally repeated sequences that
were statistically significant (see 15) were found in the telomere-associated
regions. In this way the rDNA telomeres of these ciliated protozoa differ
from those of the rDNA molecules in the slime molds Physarum or Dictyostelium,
which contain subtelomeric repeated sequences with repeat unit lengths of 140
and 29 bp respectively (6, 7). Likewise, no telomere-associated repeats have
been found in the macronuclear DNA molecules of the hypotrichous ciliate
Oxytricha (19, 20). However, macronuclear DNA molecules of the hypotrich
Euplotes share a common 5 bp sequence separated from the telomeric C_4A_2
repeats by 17 bp of non-conserved sequence (19). We could detect no such
sequence motif in the tetrahymenid rDNA telomere-associated sequences, al-
though a short sequence of this kind might not be apparent in the A + T-rich
telomere associated sequences of Tetrahymena and Glaucoma. The sequence
TTATT, located 0-3 bp away from the C_4A_2 repeats, has been noted in three
sequenced non-rDNA macronuclear telomeres of T.thermopila strain B (21).
However, out of four other non-rDNA macronuclear telomeres, the corresponding
sequence near the C_4A_2 repeats was found in only two (E. Spangler and E.
Blackburn, unpublished data). Thus, its statistical or functional signifi-
cance is unclear.

Each telomere-associated sequence, with the exception of the T. thermo-
phila 11 kb 5' sequence, included an 11 bp yeast ARS consensus sequence within
150 bp of the C_4A_2 repeats (underlined in Fig. 2). ARS function in yeast has
been demonstrated for the 5' and 3' telomere-associated sequences of the T.
thermophila rDNA (22, 23), although the corresponding regions of Glaucoma rDNA
do not have a similar activity in yeast (R. Pearlman, pers. comm.). Thus, a
function cannot be assigned to the ARS sequences in the ciliate rDNA mole-
cules, although it is interesting that the telomere-associated sequences of
yeast chromosomes contain ARS activity (10).

Since even within the same species there is no common telomere-associated
sequence adjacent to the $C_4A_2 \cdot G_4T_2$ repeats, these repeats alone must be both
necessary and sufficient for telomere function. The telomeric G_4T_2 repeat
sequence is recognized by a recently identified enzymatic activity, which adds
G_4T_2 repeats to the 3' ends of telomeric sequence DNA primers in vitro (11).
This activity most likely is required to stabilize and allow complete replica-
tion of telomeres (1, 11). Functional analysis of C_4A_2 repeats in vivo in

yeast has also shown that $C_4A_2 \cdot G_4T_2$ repeats alone are sufficient to specify telomere function in this heterologous system (24, J. Szostak, pers. comm.). Comparison of the 3' telomere-associated regions of the rDNA of two T. pigmentosa strains.

The occurrence of subtelomeric repeated sequences in a number of organisms (6-10) suggests that while telomere-associated sequences are not strictly required for telomere function, they may still be subject to selective pressure. Accordingly, we determined and compared more extensive lengths of telomere-associated sequences from two T. pigmentosa strains, to search for evidence of differential sequence conservation.

Restriction mapping of the rDNAs of T.pigmentosa strains UM1060 and ALP15 had indicated that the terminal restriction fragment in UM1060 rDNA is 600-700 bp longer than in ALP15 rDNA (17). The sequences of the terminal EcoRI restriction fragments of the rDNAs of these two T. pigmentosa strains were determined and compared (Fig. 3). The comparison shows that the length difference can be accounted for by a 625 bp "insertion" in UM1060 rDNA, relative to ALP15, that begins about 330 bp in from the C_4A_2 repeats. At each end of the inserted region there is a 21 bp homologous sequence, repeated in direct orientation, that is also present once in the ALP15 sequence. While the two 21 bp sequences in the UM1060 rDNA and the single 21 bp sequence in the ALP15 rDNA are not perfectly conserved, the 3' 10 bases, AAAAATGAAA, are identical in the two UM1060 examples, and differ by one base from the ALP15 sequence (Fig. 3b). This sequence arrangement is reminiscent of the target site duplication which commonly occurs upon the insertion of a mobile genetic element; thus, the "inserted" sequence in UM1060 could be a mobile genetic element. Although it does not have other structural features characteristic of some classes of mobile elements (15), several different types of apparently mobile but structurally diverse segments of DNA have been found in other lower eukaryotes (25, 26). Alternatively, we note that in the hypotrichous ciliate Oxytricha, sequences eliminated during macronuclear development from internal sites in macronuclear DNA molecules are also flanked by short direct repeats, and that the macronuclear DNA which is generated from such a region contains only one of the repeats (20). The "insertion" in the UM1060 strain could therefore be a sequence which is excised during macronuclear development in ALP15 but not in UM1060.

Regardless of the origin of the extra segment in the UM1060 rDNA, its presence shows that it is in a telomere-associated region of the rDNA spacer
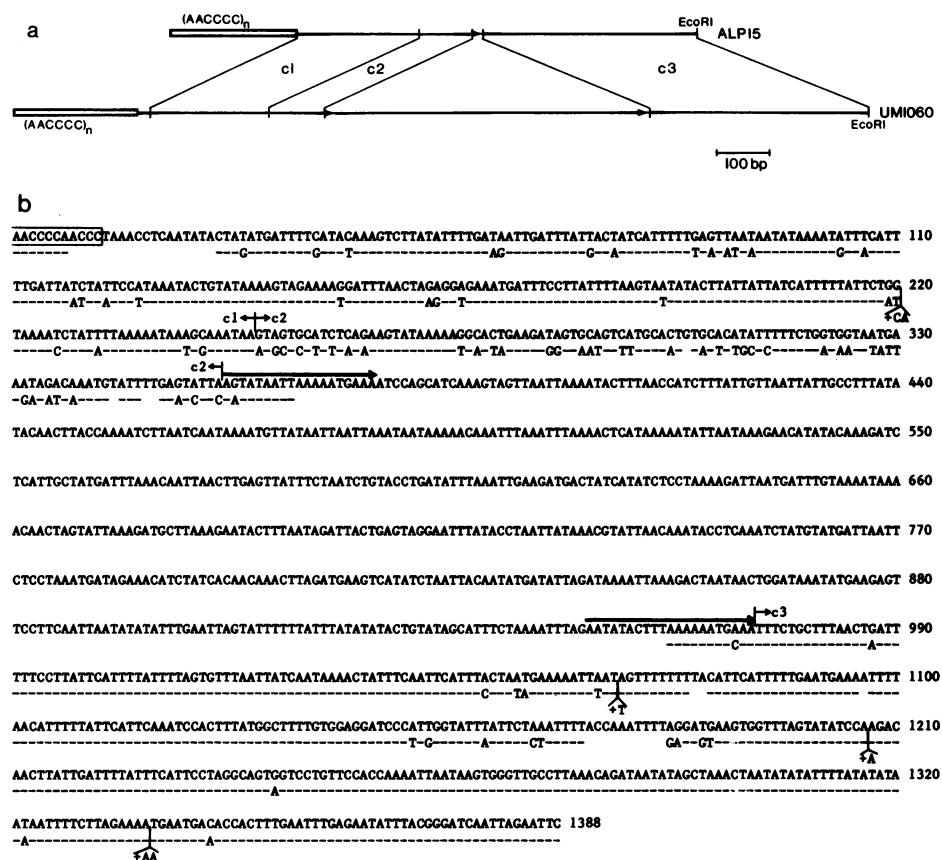
a

(AACCCC)n   EcoRI   ALP15

cl   c2   c3

(AACCCC)n   UMI060
                                                                         EcoRI

                                           ⊢——⊣
                                           100 bp

b

AACCCCAACCG̲TAAACCTCAATATACTATATGATTTTCATACAAAGTCTTATATTTTGATAATTGATTTATTACTATCATTTTTGAGTTAATAATATAAAATATTTCATT 110
------   ---C---------G---T----------------AG----------C--A---------T-A-AT-A-----------C--A----

TTGATTATCTATTCCATAAATACTGTATAAAAGTAGAAAAGGATTTAACTAGAGGAGAAATGATTTCCTTATTTTAAGTAATATACTTATTATTATCATTTTTATTCTGG 220
-------AT--A---T-----------------------T----------AG--T----------------------T----------------------AT
          cl⊣←→c2
TAAAATCTATTTTAAAAATAAAGCAAATAA̲GTAGTGCATCTCAGAAGTATAAAAAGGCACTGAAGATAGTGCAGTCATGCACTGTGCACATATTTTTCTGGTGGTAATGA 330
----C---A---------T-G------A-GC-C-T-T-A-A-----------T-A-TA-----GG--AAT--TT----A- -A-T-TGC-C-----A-AA--TATT
          c2⊣→
AATAGACAAATGTATTTTGAGTATTA̲AGTATAATTAAAAATGAAA̲ATCCAGCATCAAAGTAGTTAATTAAAATACTTTAACCATCTTTATTGTTAATTATTGCCTTTATA 440
-GA-AT-A--- --   --A-C--C-A------

TACAACTTACCAAAATCTTAATCAATAAAATGTTATAATTAATTAAATAATAAAAACAAATTTAAATTTAAAACTCATAAAAATATTAATAAAGAACATATACAAAGATC 550

TCATTGCTATGATTTAAACAATTAACTTGAGTTATTTCTAATCTGTACCTGATATTTAAATTGAAGATGACTATCATATCTCCTAAAAGATTAATGATTTGTAAAATAAA 660

ACAACTAGTATTAAAGATGCTTAAAGAATACTTTAATAGATTACTGAGTAGGAATTTATACCTAATTATAAACGTATTAACAAATACCTCAAATCTATGTATGATTAATT 770

CTCCTAAATGATAGAAACATCTATCACAACAAACTTAGATGAAGTCATATCTAATTACAATATGATATTAGATAAAATTAAAGACTAATAACTGGATAAATATGAAGAGT 880

                                                        ⊢→c3
TCCTTCAATTAATATATATTTGAATTAGTATTTTTTATTTATATATACTGTATAGCATTTCTAAAATTTAGAATATACTTTAAAAAATGAAA̲TTTCTGCTTTAACTGATT 990
--------C---------A---
TTTCCTTATTCATTTTATTTTAGTGTTTAATTATCAATAAAACTATTTCAATTCATTTACTAATGAAAAATTAATA̲GTTTTTTTTTACATTCATTTTTGAATGAAAATTTT 1100
--C---TA---------T-↓--------   ------------------ ----
                                                             ⬆⬆
AACATTTTTATTCATTCAAATCCACTTTATGGCTTTTGTGGAGGATCCCATTGGTATTTATTCTAAATTTTACCAAATTTTAGGATGAAGTGGTTTAGTATATCCA̲AGAC 1210
---T-G------A-----CT-----   GA--GT-- ---------------
                                                                                          ⬆A
AACTTATTGATTTTATTTCATTCCTAGGCAGTGGTCCTGTTCCACCCAAAATTAATAAGTGGGTTGCCTTAAACAGATAATATAGCTAAACTAATATATATTTTATATATA 1320
-----------------------------A---. -----------------

ATAATTTTCTTAGAAAA̲TGAATGACACCACTTTGAATTTGAGAATATTTACGGGATCAATTAGAATTC 1388
-A-------------↓-------A--
               ⬆AA

Fig. 3. Comparison of the rDNA 3' telomere-associated sequences of two
         strains of T.pigmentosa.

     a) Relationship between the T.pigmentosa ALP15 and UM1060 telomere-
associated regions. The ends of the rDNA molecules are on the left. Open
bars = telomeric C4A2 repeats. Arrowhead in the ALP15 sequence is a 21 bp
sequence, duplicated as shown by the two arrowheads in the UM1060 sequence.
Regions marked cl, c2 and c3 are described in the text.
     b) Optimal alignment of homologous sequences of the rDNA 3' telomere-
associated regions of UM1060 (top) and ALP15 (bottom). Identical bases are
shwon by dashes. Bases which differ between the two strains are shown by the
base in the ALP15 sequence. Gaps introduced to maximize the alignment are
shown as spaces. Telomeric C4A2 repeats are boxed. The 21 bp duplication of
the UM1060 sequence relative to the ALP15 sequence (see Fig. 3a) is shown by
thick arrows. The boundaries of the cl, c2 and c3 subregions described in the
text are indicated. Numbering of nucleotides is as in Fig. 2. Sequences were
obtained from plasmids pUM9 and pALP3 described in Figure 1.

where the colinearity with respect to ALP15 rDNA can be disrupted without im-

pairing rDNA function. To analyse further any constraints of this region of

the rDNA, as a function of position along the telomere-associated sequence,

the two sequences were aligned as shown in Fig. 3b. The colinear segments of the terminal EcoRl fragments are divisible into subregions, shown as c1, c2 and c3 in Fig. 3, based on differences in the rates of sequence divergence. Applying the formula of Kimura (27) to these sequences to derive K, the evolutionary distance per nucleotide site between the two strains, c3 (K=0.04) and c1 (K=0.14) are seen to be much more highly conserved in sequence than c2 (K=0.57), which is distal to the "insert" in UM1060 (Fig. 3b). These findings strongly suggest that c1 and c3 are under greater selective pressure than c2, and that despite the general lack of conservation of telomere-associated sequences of different species, these regions may play as yet unknown roles in macronuclear rDNA function or formation.

The role of telomere-associated sequences in de novo telomere formation during macronuclear development.

The lack of conserved sequences next to the terminal C_4A_2 repeats of the tetrahymenid rDNAs argues that there cannot be a strongly sequence-dependent target for the addition of C_4A_2 repeats during macronuclear development. Furthermore, we found that the exact points to which the C_4A_2 repeats are joined to the otherwise colinear 3' telomere-associated sequences differ in the rDNAs of different strains from the same species. These C_4A_2 repeat junction points differ by 3 bp in strain C_3-368 versus strain B-1868 of T. thermophila (Fig. 2), and by 18 bp in the corresponding junctions of the UM1060 and ALP15 strains of T. pigmentosa (Fig. 2).

Excision of rDNA from the micronuclear genome may be specified by sequence elements located outside the region which is retained as macronuclear rDNA. Sequences designated A-C occur in inverted orientation at each end of the chromosomally integrated micronuclear rDNA, in flanking micronuclear DNA which is eliminated from the macronucleus (14). The sequence of the 5' telomere of the 11 kb rDNA of T. thermophila strain B-1868 reported here allowed us to define precisely the positional relationship of the A-C sequence elements to both rDNA telomeres, as shown in Fig. 4. Each ~37 bp A-C element is 29 bp away from the position to which the C_4A_2 repeats become joined to form the 3' and 5' telomeres of the 11 kb rDNA.

These findings suggest that the A-C sequences specify the positions of breaks which excise the rDNA from the micronuclear locus (12-14), and hence specify the positions of $C_4A_2 \cdot G_4T_2$ addition to rDNA telomeres. We have previously proposed that a mechanism of telomere addition (11) acts during macronuclear development (15). Our results strongly suggest that the telomere-associated sequence to which the $C_4A_2 \cdot G_4T_2$ repeats are attached, to "heal"

```
                                          ...CCCCAACCCCAACCCCAA TTCAACAACATAGTGCTGA    5' MAC TELOMERE
-------------------------------------->                       *******************
...AAATTAAATTTATATTAAAGAAAGAGGTTGGTTTA-TTAATATTAAAAGCGATTTTCGAAGGTAAAA TTCAACAACATAGTGCTGA  5' MIC JUNCTION
  ****** ***** * **** *************** **  **  ** *         *    * **  *       *
...AAATTATATTTAAA-TAAACAAAGAGGTTGGTTTAATTTTTAAAAATAAGTTAAAACCCCAATTTAA TAATTTCTTTGACATTGAG  3' MIC JUNCTION
------------------------------------->                    ** *** *    **  *******************
                                          ...CCCCAACCCCAACCCCAA TAATTTCTTTGACATTGAG   3' MAC TELOMERE
```

Fig. 4. Alignment of the ends of the single integrated micronuclear rRNA gene
          copy with 11 kb rDNA telomeres.
Top and bottom sequences = the 11 kb rDNA 5' and 3' telomeric C_4A_2 repeats and
adjoining sequences. Only the innermost C_4A_2 repeats (boxed) are shown. 5'
mic junction and 3' mic junction = the micronuclear rDNA end regions and their
flanking micronuclear sequences, from (14). * = identical base in the aligned
sequences; arrows = A–C repeats (14); vertical line = position to which C_4A_2
repeats are joined to the macronuclear 11 kb rDNA. Note that for both
telomeres the exact point of joining may differ by up to two bases, because
two A residues in the micronuclear flanking sequence could be part of a C_4A_2
repeat unit. All sequences were determined from micronuclear or macronuclear
rDNAs of T.thermophila inbred strain B.

such breaks in macronuclear development, does not itself provide any speci-
ficity for that process.

A 5' telomere-associated sequence is conserved between Glaucoma and Tetra-
hymena rDNAs.

        The existence of a palindromic linear form of rDNA in the mature macro-
nucleus of all Tetrahymena species is enigmatic. First, this palindromic
rDNA must be accurately synthesized from the single, non-palindromic, chromo-
somally integrated rDNA copy at the micronuclear locus. Second, the non-
palindromic 11 kb rDNA generated during macronuclear development in Tetra-
hymena includes the entire rRNA transcription unit, together with 5' and 3'
non-transcribed spacers, and is able to self-replicate (12). That such a
non-palindromic form of the rDNA is also capable of providing all the func-
tions necessary for long term vegetative growth is evident from the fact that
the only rDNA in the macronucleus of vegetatively growing Glaucoma is non-
palindromic (5) . Similarly, the macronuclear rDNA in hypotrichous ciliates
also consists of non-palindromic linear single rRNA gene molecules (28).
From these observations, we reasoned that the ability to make palindromic
rDNA may not be intrinsically required for excision of rDNA from the Tetra-
hymena micronuclear genome.

        A comparison of the 5' telomere-associated sequences of Tetrahymena and
Glaucoma supports this view. First, T.thermophila rDNA has an inverted pair
of 42 bp conserved sequences (M sequences) on each side of the sequence which
becomes the non-palindromic central 29 bp of the palindromic rDNA (29, 30).
These two inverted 42 bp sequences are both present in the micronuclear gene

copy (14), and are highly conserved between different <u>Tetrahymena</u> species, al-
though the non-palindromic central sequence is variable (30). Based on these
sequence properties, a mechanism for generating palindromic rDNA from the
single micronuclear rDNA copy via intramolecular recombination between the M
repeats has been proposed (14). As shown diagramatically in Fig. 5, the 5'
telomere region of T.thermophila 11 kb rDNA contains all 42 bp of one M
sequence, the non-palindromic 29 bp sequence, and 25 bp of the other M
sequence adjacent to the C_4A_2 repeats (see Fig. 2). It is not known whether
the 11 kb rDNA is an intermediate in the formation of palindromic rDNA (12).
However, this sequence analysis shows that 11 kb rDNA does include the two
inverted M repeats that would be required to form palindromic rDNA through
intramolecular recombination.

Second, we found that the 5' telomere-associated sequence of the <u>Glaucoma</u>
rDNA has significant homology to part of the M sequence (Fig. 5). This homol-
ogy is striking because it is the only significant stretch of homology between
<u>Glaucoma</u> and <u>Tetrahymena</u> rDNAs in the 5' spacer until rRNA promoter elements
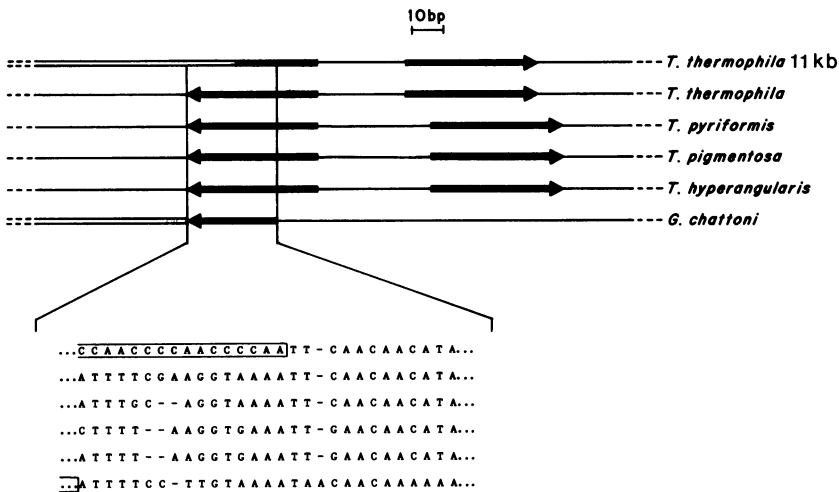

Fig. 5. A conserved sequence next to the 5' telomeres of <u>Glaucoma</u> and
<u>Tetrahymena</u>.
Schematic maps and sequence alignments of (top to bottom): <u>Tetrahymena ther-</u>
<u>mophila</u> 11 kb rDNA 5' telomere-associated region; centers of palindromic rDNAs
of <u>T.thermophila, T.pyriformis, T.pigmentosa</u> and <u>T.hyperangularis; G.chattoni</u>
9.1 kb rDNA 5' telomere-associated region. Thick arrows = inverted M repeats
(14, 30) or M-homologous sequence in <u>Glaucoma</u>; thin lines = non-conserved rDNA
5' spacer sequences (16, 30); open bar or boxed sequence = innermost bases of
telomeric C_4A_2 repeats. Dashes are included to maximize alignment between
sequences shown in the lower portion of the Figure. Sequence data from Fig. 2
and 29, 30.

are reached, over 700 bp away (16). However, in contrast to the pair of inverted M sequences in the Tetrahymena 11 kb and micronuclear rDNA, the M-homologous sequence is present only once in Glaucoma. This result provides the first evidence that the M sequence functions in a step of rDNA processing and/or amplification common to both species. We propose that during the evolution of Tetrahymena, a single M sequence underwent a duplication-inversion event, resulting in its conversion into a pair of inverted M repeats separated by the non-palindromic sequence. We suggest that this simple rearrangement produced a structure which allowed the generation of palindromic rDNA from the single micronuclear rRNA gene. The structure and formation of 11 kb rDNA in Tetrahymena, and the M sequence homology in Glaucoma, strongly suggest that the more "primitive" pathway of rDNA maturation, resulting in excision of a single gene copy, operates in both Tetrahymena and Glaucoma. In addition in Tetrahymena the pathway leading to palindrome formation operates, possibly but not necessarily via 11 kb rDNA.

Although during macronuclear development in Tetrahymena both 11 kb and palindromic rDNAs are formed and self replicate, after several vegetative generations the 11 kb molecules are lost (12). One explanation is that 11 kb rDNA is converted to palindromic rDNA. Alternatively, the palindromic form of the rDNA may have an advantage in replication over the 11 kb form in vegetative macronucleus. Competition between different allelic forms of palindromic rDNA molecules in the course of vegetative growth has been observed, and from sequence analysis of these alleles there is evidence that rDNA molecules compete for limiting replication factors (31). Hence, palindromic rDNA molecules, where every sequence is present twice, are expected to compete more effectively for such factors than single gene rDNA molecules. This model predicts that once the ability to form palindromic rDNA was evolutionarily established in Tetrahymena, selection for this molecular form would have occurred.

[1]Present address: Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, WA 98104, USA

REFERENCES

1. Blackburn, E.H. and Szostak, J.W. (1984) Ann. Rev. Biochem. 53, 163–194.
2. Blackburn, E.H. (1984) Cell 37, 7–8.
3. Ponzi, M., Pace, T., Dore, E. and Frontali, C. (1985) EMBO J. 4, 2991–2996.
4. Blackburn, E.H. and Gall, J.G. (1978) J. Mol. Biol. 120, 33–53.
5. Katzen, A.L., Cann, G.M. and Blackburn, E.H. (1981) Cell 24, 313–320.
6. Johnson, E.M. (1980) Cell 22, 875–886.
7. Emery, H.S. and Weiner, A.M. (1981) Cell 26, 411–419.
8. Blackburn, E.H. and Challoner, P.B. (1984) Cell 36, 447–457.
9. Van der Ploeg, L.H.T., Liu, A.Y.C. and Borst, P. (1984) Cell 36, 459–468.
10. Chan, C.S.M. and Tye, B.-K. (1983) Cell 33, 563–573.
11. Greider, C.W. and Blackburn, E.H. (1985) Cell 43, 405–413.
12. Pan, W.-C. and Blackburn, E.H. (1981) Cell 23, 459–466.
13. Pan, W.-C., Orias, E., Flacks, M. and Blackburn, E.H. (1982) Cell 28, 595–604.
14. Yao, M.-C., Zhu, S.-G. and Yao, C.-H. (1985) Mol. Cell. Biol. 5, 1260–1267.
15. Cherry, J.M. and Blackburn, E.H. (1985) Cell 43, 747–758.
16. Challoner, P.B., Amin, A.A., Pearlman, R.E. and Blackburn, E.H. (1985) Nucl. Acids Res. 13, 2661–2680.
17. Wild, M.A. and Gall, J.G. (1979) Cell 16, 565–573.
18. Shampay, J., Szostak, J. and Blackburn, E. (1984). Nature 310, 154–157.
19. Klobutcher, L.A., Swanton, M.T., Donini, P. and Prescott, D.M. (1981) Proc. Natl. Acad. Sci. USA 78, 3015–3019.
20. Klobutcher, L.A., Jahn, C.L. and Prescott, D.M. (1984) Cell 36, 1045–1055.
21. Yokoyama, R. and Yao, M.-C. (1986). Nucl. Acids Res. 14, 2109–2122.
22. Amin, A.A. and Pearlman, R.E. (1985) Nucl. Acids Res. 13, 2647–2659.
23. Amin, A.A. and Pearlman, R.E. (1986) Nucl. Acids Res. 14, 2749–2762.
24. Szostak, J.W. and Blackburn, E.H. (1982) Cell 29, 245–255.
25. Hasan, G., Turner, M.J. and Cordingley, J.S. (1984) Cell 37, 333–341.
26. Cappello, J., Cohen, S.M. and Lodish, H.F. (1984) Mol. Cell. Biol. 4, 2207–2213.
27. Kimura, M. (1980) J. Mol. Evol. 16, 111–120.
28. Swanton, M.A., Greslin, A.F. and Prescott, D.M. (1980) Chromosoma 77, 203–215.
29. Kiss, G.B. and Pearlman, R.E. (1981). Gene 13, 281–287.
30. Engberg, J. (1983) Nucl. Acids Res. 11, 4939–4946.
31. Larson, D.D., Blackburn, E.H., Yeager, P.C. and Orias, E. (1986), submitted for publication.