Volume 13 Number 7 1985 Nucleic Acids Research

Conserved arrangements of repeated DNA sequences in nontranscribed spacers of ciliate ribosomal RNA genes: evidence for molecular coevolution

Peter B.Challoner[1]*, Anthony A.Amin[2], Ronald E.Pearlman[2] and Elizabeth H.Blackburn[1]+

[1]Department of Molecular Biology, University of California, Berkeley, CA 94720, USA, and
[2]Department of Biology, York University, 4700 Keele Street, Downsview, Ontario M3J 1P3, Canada

Received 19 September 1984; Revised and Accepted 26 September 1984

ABSTRACT

We have analyzed the nucleotide sequences of the nontranscribed spacer (NTS) and transcription initiation and termination regions of the extrachromosomal rDNAs of the ciliated protozoans Tetrahymena thermophila and Glaucoma chattoni. The sequences surrounding the sites of transcription initiation and termination are highly conserved. The only extensive homologies of the NTS regions occur in five sets of dispersed repetitive sequences. Type I, II and III repeats in the 5' NTS are strongly conserved in sequence between Tetrahymena and Glaucoma in the case of the type I and III repeats, and in location relative to the transcription initiation site in the case of type I and II repeats. We identify two new repeat types, designated IV and V, in the 3' NTS. The sequence of type IV repeats, and the location relative to the transcription termination site of type IV and V repeats, are conserved. All five types of repeats are interspersed with nonconserved DNA sequences. These results suggest that the five repeat types in the 5' and 3' NTSs are important in rRNA gene function; the sequence organization, and the differing rates of divergence between species of the repeat types, provide strong evidence for their functional selection through the process of molecular coevolution.

INTRODUCTION

The ribosomal RNA genes (rDNA) in metazoan eukaryotes are organized as hundreds of tandemly repeated units at one or more chromosomal loci (reviewed in 1). Each unit consists of a nontranscribed spacer (NTS) and pre-rRNA coding region. Nontranscribed spacers typically contain several families of repeated sequence elements, and some types of these elements are likely to function as promoters and enhancers of transcription in Xenopus laevis (2,3) and Drosophila (4,5). The NTS sequence organization is poorly conserved, even between closely related species (6-8). Experiments utilizing in vitro transcription assays have shown that only promoter sequences and cellular extracts from the same species will correctly initiate transcription (9-11), suggesting that the rapid evolution of spacer sequences occurs even in functionally important regions of the NTS. Through a process termed molecular coevolution (12), rapid changes in spacer sequences that are

necessary for rDNA function must in turn be accompanied by compensatory changes in the proteins that interact with the NTS.

The divergence of NTS sequences seen when rDNA units from different species are compared contrasts sharply with the homogeneity of NTSs in rDNA units within and between individuals of the same species. Molecular mechanisms such as unequal crossing over and gene conversion operating in a multigene family could result in the spread and eventual fixation of a variant rDNA unit in a population of repeated rDNA units (reviewed in 7). Thus, these mechanisms can account for both the rapid evolution of NTS sequences and the intraspecies NTS homogeneity. The presence of repeated elements in the NTS of a variant rDNA, by effectively increasing the frequency of recombination at that unit, would favor the spread of that variant throughout the rDNA population. This has led to the suggestion that at least some repeated elements in the NTS may be maintained simply as a consequence of unequal crossing-over in the multigene rDNA family (6).

To define those structural elements of the NTS that are conserved in the absence of selective mechanisms which function because of the tandem repetition of rDNA units, we have determined and compared complete NTS sequences of two ciliate species: Tetrahymena thermophila and Glaucoma chattoni. In T. thermophila, the germline (micronuclear) haploid genomic complement of rDNA consists of a single rDNA unit (13). This micronuclear copy is the source of several thousand identical amplified extrachromosomal rDNA molecules in the transcriptionally active macronucleus. In many other inbred ciliates, including G. chattoni (14), the amplified macronuclear rDNA consists of a homogeneous population of molecules, arguing that they are also generated from one or very few germline rDNA copies. Macronuclear rDNA in T. thermophila is organized as linear 21 kb palindromic molecules containing two head-to-head copies of divergently transcribed rRNA genes flanked by 5' and 3' nontranscribed spacers (5' and 3' NTSs). G. chattoni macronuclear rDNA consists of linear 9.3 kb molecules, each with a single rRNA transcription unit and flanking 5' and 3' NTSs. The nontranscribed spacers of these macronuclear rDNAs must include sequences involved not only in the promotion regulation and termination of rRNA transcription (15-19), but also in replication initiation (20,21), copy number control, packaging in nucleoli, and excision and subsequent amplification of rDNA molecules from the single micronuclear chromosomal locus during macronuclear development.

Our analysis of the Tetrahymena and Glaucoma 5' and 3' NTSs shows that the only conserved structures in these spacers are five types of repeated

elements interspersed with unconserved sequences. Since the structure and arrangement of these repeated elements cannot be accounted for by conventional mechanisms of unequal crossing-over or gene conversion, we propose that they are maintained by functional selection through molecular coevolution.

MATERIALS AND METHODS

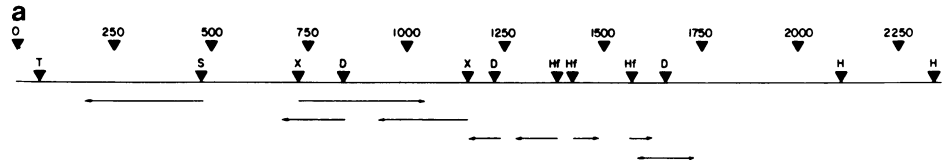
Preparation of rDNAs from Glaucoma chattoni and Tetrahymena thermophila

G. chattoni, strain GH-1, was maintained in stock cultures and grown as described previously (14). Macronuclear rDNA was prepared by fractionation of total undigested cellular DNA by agarose gel electrophoresis, and purification of the 9.3 kb size class which consists of the macronuclear rDNA molecules (14). T. thermophila inbred strain B was grown and macronuclear rDNA was purified as described by Wild and Gall (22), or by a modification of the method of Din and Engberg (23).

Recombinant DNA Techniques and DNA Sequencing

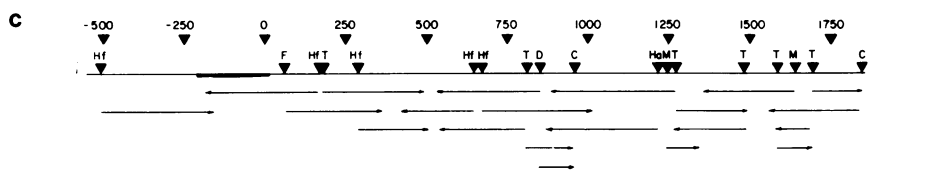
For the determination of rDNA sequences, three plasmids containing inserts of rDNA telomeric fragments were constructed from T. thermophila or G. chattoni rDNA by treatment with S1 nuclease followed by complete digestion with BamHI and ligation into PvuII-BamHI digested pBR322 DNA as described previously for cloning other telomeric regions (24). pTre1 and pGre7 contain the telomeric BamH1 fragments that include the 3' NTS from the T. thermophila and G. chattoni rDNAs respectively. pGre2 contains the other telomeric BamH1 fragment from G. chattoni rDNA including the 5'NTS. Each of these cloned telomeric rDNA segments carries a terminal stretch of C_4A_2 repeats (14,25,26). The plasmid pTtr1, which contains the central HindIII fragment of T. thermophila rDNA inserted into the HindIII site of pBR322, was also used for DNA sequence analysis. Techniques used in the construction and analysis of recombinant DNA were essentially as described in Maniatis et al. (27). Plasmid DNA for sequencing was isolated using the alkaline extraction procedure of Birnboim and Doly (28) followed by either CsCl-ethidium bromide density gradient centrifugation or phenol extraction and ethanol precipitation. DNA sequence analysis was according to Maxam and Gilbert (29).

Computer analysis of DNA sequences was carried out using a program obtained from A. Delaney, University of British Columbia (30) extensively modified by Yau-Hing Pak, York University. A program to find the optimal alignment based on homology of a pair of similar sequences was written by W. Gish, University of California, Berkeley and used with the asistance of Mike Cherry, University of California, Berkeley.



b

```
GTTTTTTTTG CTTTTTGTTG TTAGTTTTAT AGCCTTCAGC ACTATGTTGT TGAATTTTAC CTTCGAAAAT CACTTAAAAT TGAGTAATAA TTGGGTTTAA 100
AATTTAAATT TGAGTAGATA AGAATTAGAT GTTTATATTC TGCTAATTTC ACTGGTGAAA ATGTAGCAAA TAGAAATTAT TTTAATCTAA TAAACTAGCA 200
AATAGTATTT AAAACAAAAA TATTTGTTTT TTATGTTGTA AAATGTTTTA AATTAGATAA AATTTACAAA TTTACAAATT TTCAAGCAAA ATAGGTTCTA 300
AAAAATGAGA AAATATTACA TATTTTAGCT ATTTGACTAC TTTAATGCTA GTAAATTAAA ATGAATTTAA TTCATTTTCA CTTTAAAACA CTTATTTTAA 400
TAAAATATAT GATTTTAAAA TGATAAAATA TTTTTTAAGA GGTAAATTTA AGAAATTAGT TAAATTTTAA AGAAAAAGCA TCTAAAAATG GACAAAAATG 500
AAGTATTTCC TTTTTTTATA CATTTAAATG CTAGAAAATT TAAGTAAAAC ATTTATAAAT AAAAGTAAAA TAGTTTTAGG AATATGAGTA AATAGTTTTT 600
TTTATGTAAA AAACATTTTA TCAATTTCAT TTATTCATTT TAGTTAAATT TTTCATTCAC AAAAAACTTT TTTTTGGTAA AATAAAGACT TTATAAAGAT 700
AACTTAAAGA AAAAGTTTAT CTAGATTAAA AATATTGATT TTGAAAATTG CTCATTAGAT ATTTTTTTGG CAAAAAAAAA AACAAAAATA GTAAAAAATC 800
ACTTTTTTTG AGAGTTGAAA AAAAGACTTA GAAAAAATTT TAAAAGTGTA AAAAAAGACT TAGAGAAAAA ATCAAAAAGA GATAAAAAGA CTTAGAGAAA 900
ATTTATAAAT TAAAATGTAG AAAGTAAATT TATTTTATAT TTTTTAATCA TTTAAATGCT AGTAAATTTA AATAAAACAT ATATAAAAAA ACATAAAACA 1000
ATTTTAACAG CATGCGTATA TCATTTTTTA TATGTAAAAA ACATTTTATC AATTTCATTT ATTCATTTTA GTTAAATTTT ACATTCAAAA TAAATTTTTT 1100
TTGATTAAAT AAAGAGTTCA TAAAGAGAAC TTAAAGAAAA AGTTTATCTA GATTAAAAAT ATTGATTTTG AAAATTGCTC ATTAGAAATT TTTTTGGCAA 1200
AAAAAAAAAC AAAAATAGTA AACTTAGAAA AATTTTTGAA AAATGAAAAA AAAAGACTTA GAGAAAAAAA TCAAAAAGTC AAAAAAGACT TAGAAAAATT 1300
TTTAAAAATG AAAAAATGAT TTAAGGAGAA AATTTTGAGT TGCGCTTAGA TTTTGTGTGA AGCCACTTAC AAAAAATGAC GGGACTCGCT CAAATATTTA 1400
AGTGGACTCG CATAAAATGA GTGAGTCACT AAAAAAATTA AGTGAACTCA CTTAAAAATG AGTGGAGCCA CTCAAAAAAT TAAGTGACTC ACTTAATATT 1500
CGGCGGAGTT AAACAAAAAT AAGTGGACTC ACAGAAAATT AAGCGGATTG CGCTAAAAAA TGAGTGGACT CGCTTAAAAA TGAGTGGAGC CACTCAAAAG 1600
TTCAGCAGAG CCACTTAAAA ATTTAGCTTA AAATCAGCTC TAAATTAAAT TAGACTTAGT GAAAATAGC GAAAATGAAA AAAATGAAAA AATGAATGAA 1700
AACTGAAAAA TTTACAAGGG ATTGAAAATT TTGGCAGAGT CTTTTTTTTG GCAAAAAAAA AAACAAAAAT AGTAAACCTT CCGAACTTTT TTGACTTTGA 1800
GAAAAATTCT TTGGCAAAAA AAATAAAAAT AATATCAGGG GGGTAAAAAT GCATATTTAA GAAGGGGAAA CATCTCCGGA TCAAAAATAA AATATCAGCT 1900
```



d

```
AACCCCAAAT TTTCCTTGTA AAATAACAAC AAAAAATACA AGGGATTTTC TAAATTATAA ATGGATGATT ATGCTACTTT CGCTTAATCT AGTTCTCGGA 100
ATTGTTGATT GAGAGCATAG AGCAGACTGT CATTAATCGA ATTAGGTATG GTAGTGCTTT AGTATGATGA CTCGATCGCG TCGAAAACAG GGATTTACTG 200
GTTTTTCAGA TTTTCATTCA AAGTGAATGA AGGATAAAAT GGTATAAAAT AAAACGTTTA AATAAGTTGA TGTTCTGGAG AAAGCGATGA ATCTTTTTTT 300
CGTTTGTGCG ATAGAGCATA CCTATTACTT CAAGAACCGC ATGTTCTTGG ATTATCCAAT CGGCGGCAAC TACTTCCATT TTTGCGATTA CTCTCAAAAC 400
AGGACTAATA AATTAAAAAA AATCACCTTT TTACTCACTT GAGATATGTT ATTTTTTAAA CGGCTTATCA TACCTAGCAT TTTTTTCGTC TGCTATGTTG 500
ACAACAATCT CCTACACACT TTCAATTACT CCTTACATAC CTTATTGTCA GCGTAAGTGC CTCAAATTGT TGAAATACTG CGTTTTTTTG CCATTTTTAA 600
AATGAATGAG AAAAAAATCG TAAAATTTAC AGTCAAAAAA GATTGGATTC AAGACTACCC TTATATTTAG ATTCTTCAGT TTTTTTACCA CTAGTATTGT 700
AGAACAGAGT TCAACACACA TTTCTATCTA GGTCATTTAA CGTAATTTGT AGATTTAATG TTGAAAAATA TGAGAATACA TCAAAAAATT GACATTTTTT 800
TGAATGGAGC TCGATTTTAA TGATTTCGGT TTTTTAACTA TTATAATAAT CTTAGGTAGC ATTTTATTCG TTGGATAAAC GTTTTAAACT ATGGCAAATA 900
AGAATAACAT ACTTTATTTG TTGAAATAAA GTTGGTTATA GGAGATTTAA AAGATCGATA AGTTTAGAAA TGGGAGTAAA AAAATTTTCA TTCAATTTTG 1000
AATTGAGATT ATTAGGTTCT ATCTCCTAAA TTAAACATAG ATTGCAAAAA CTTGATTTTG AGATTGCCCC TTGTAAAAAT TTTTTGGCAA AAAAAATCTC 1100
AGGTATACCA GGTTTTTCAA AAAAAAGACT TAGAAAAATT TTCAAAAATT TTTTCAAAAA AAAAAAAAAT TTCAGCGGGA TTTTTTGGAA TTTGGCTGGG 1200
ATTTTTGTGA TTTGGCCGTG TTTTTTTCTG ATTTTTCTCA TCTCCGGCCA AATTTTCAGC GGGGAATCAC TCGATTTTTC AGCGGAGGAA AATATAATTA 1300
ATTTCAGCGG GGATTCGCTT AATTTCAGCG GAGAAATTAA AATCCCAGCC AAATTTCTGC GGGGAGAAGG GTTGTTCCCG CTGAATTAAT CAGGTGGGGA 1400
GAAAATTTCA GAAATTTTAT CAGAGTCTTT TTTTTGGCAA AAAAAATCG GATATATCAG GGGTGGGGGT AAATTGGCAA AATCGACTTC GATTTCTTGG 1500
CAAAAAAAAA ATCTGAGATA TCAGGGGGGT AAAAATGCAT ATTTAAGAAG GGGAAACATC TCCAGATTAA AAATAAAACG TGAGTTCGAT TTGAGCTTCT 1600
GGTGGTTCTT TTGGGATCTT TCGGGAACCC AAGGATAAAA CCGGTTAAGT TCTAAGTGCA CTGGCGTTTT TCTCTATGGC TTCGGCGTCG AGCTGATAGC 1700
CGCTGAGGGT CTGGCGAAGG TTTTTCTGGA TTACGGCTCG TATTAGAGCA AATGGCCCTA CTGGAATTTC CACGTGGGCT GTGCAATTGT CGCTGCAAAA 1800
GATTCGCAAG AATCTTTTCC CAGTGAGAGT TGTTGTATCG AT 1842
```

RESULTS

The 5'NTS Sequence of Tetrahymena and Glaucoma rDNA Molecules

We determined the complete nucleotide sequence of the 5'NTS and part of the external transcribbed spacer (ETS) of the extrachromosomal rDNA molecules of the ciliated protozoa T. thermophila and G. chattoni. The sequencing strategies are given in Fig. 1a and c. For T. thermophila the sequence presented (Fig. 1b) includes sequences determined in this study, sequences near the center of the molecule we have previously published (31) and ETS sequences recently reported by Engberg et al. (17) and confirmed by ourselves. The sequence numbering starts at the 5' end of the central 29 bp nonpalindromic region (31), and the orientation of this region with respect to the palindromic region is the same as in the linear 11 kb form of the rDNA found in developing but not mature vegetatively growing cells (32; Challoner and Blackburn, manuscript in preparation) and in the single integrated rDNA copy in the micronucleus (M.-C. Yao, personal communication). There are a few minor modifications to previously published sequences (31). The telomere associated sequence of the G. chattoni rDNA (bp 1-200) will be discussed in detail elsewhere (Challoner and Blackburn, manuscript in preparation), but the sequencing strategy and sequence of the entire NTS and flanking regions is presented here for completeness and ease of reference.

We compared the T. thermophila and G. chattoni sequences with each other and with the sequence of the analogous region from T. pyriformis (15,33) using computer generated two-dimensional matrix plots (data not shown). Pairwise

Fig. 1. Sequence of T. thermophila and G. chattoni rDNA 5'NTS and adjacent regions.

a. T. thermophila 5'NTS sequencing strategy and partial restriction map. Sequencing of the regions shown (→) was according to Maxam and Gilbert (29). Fragments were end labelled and sequences read towards the arrowheads. All regions not sequenced have been previously published (17,31). C = ClaI; D = DdeI; E2 = EcoRII; F = FokI; H = HindIII; Ha = HaeIII; Hf = HinfI; Hh = HhaI; M = MspI; R = RsaI; S = SfaNI; T = TaqI; X = XbaI.

b. T. thermophila rDNA 5'NTS sequence reading in the 5'→3' direction beginning at the non-palindromic region (bp1→29, double underline) at the center of the molecule. Transcribed region begins at bp 1859 (17) and is underlined. Repeated sequences are boxed and named below each box.

c. G. chattoni 5'NTS sequencing strategy and partial restriction map. Sequences were determined from the plasmid pGre2. The thicker segment of the map from +6 to -200 is telomeric C_4A_2 repeats (14) with pBR322 vector to the left and rDNA 5'NTS to the right of the C_4A_2 repeats. Sequencing strategy and restriction sites are as described for Fig. 1a.

d. G. chattoni rDNA 5'NTS sequence reading in the 5'→3' direction beginning at the most proximal telomeric C_4A_2 repeat (bp 1→6, double underline). Transcribed region and repeats are indicated as described for Fig. 1b.

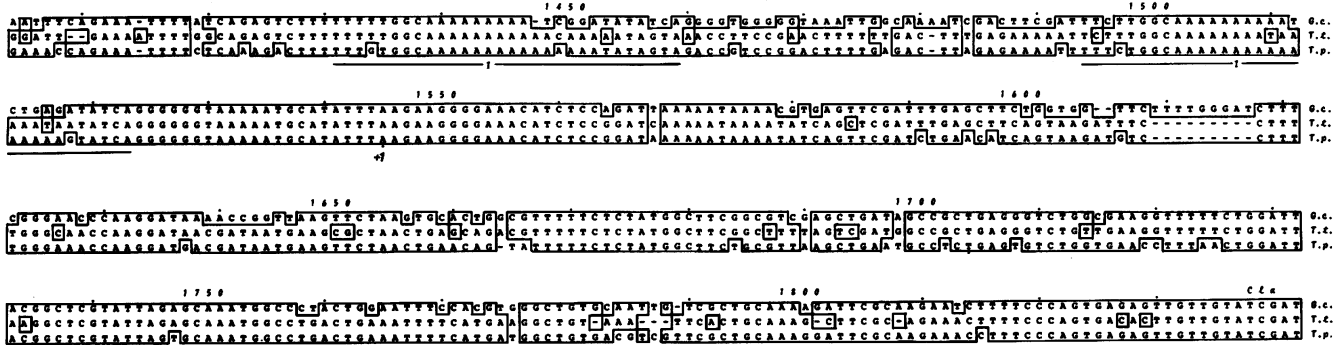


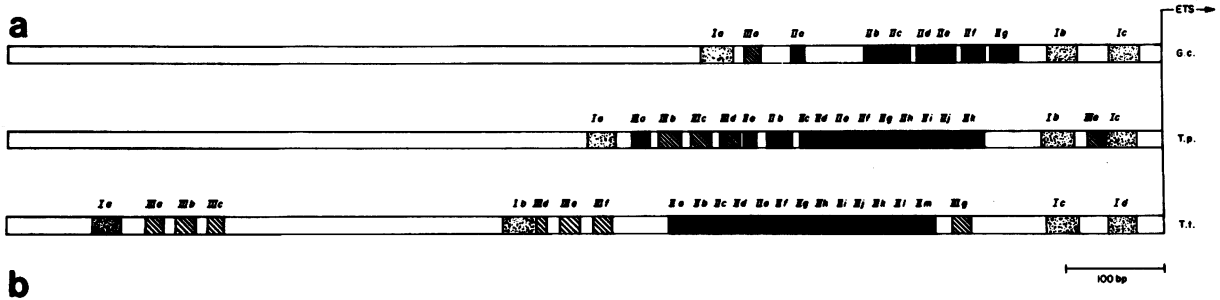
Fig. 2. Comparison of the pre-rRNA transcription initiation site and flanking regions. The optimal alignment of the region surrounding the pre-rRNA transcription initiation site that is conserved among G. chattoni (G.c.), T. thermophila (T.t.) and T. pyriformis (T.p.) is shown (see 15,17) - indicates 1 bp gap introduced to maximize alignment. Sequences are 5'→3' left to right so that transcription occurs rightward from the transcription initiation site (+1) (16,17). Boxes enclose positions where identical bases are found in at least two of the species. The numbering refers to the G. chattoni sequence and is consistent with Fig. 1d. Underlinings show the positions of Type I repeats.

comparisons of the Tetrahymena and Glaucoma sequences showed that the pre-rRNA ETS and about 140 bp of 5'NTS adjacent to the site of transcription initiation are the only sequences well conserved among the three species. The high degree of homology in this region is illustrated in Fig. 2, where sequences flanking the transcription initiation region from the three species are arranged to give optimal alignment.

Niles et al. (15) have reported that an array of three types of repeated sequences, referred to as type I, II and III repeats, are present in the NTS region of Tetrahymena pyriformis. We found similar repeat arrays in the analogous regions of the T. thermophila and G. chattoni NTS, as shown by the boxed sequences in Fig. 1b and d. A detailed comparison of the organization of the repeat arrays in these species is presented in Fig. 3a. The The sequences of the individual members of the three repeat types are aligned in Fig. 3b to produce the maximum homology of repeat members for both intra- and interspecies comparisons. These conserved repeated sequences are embedded in a completely nonconserved matrix. From the data in Fig. 3 it is seen that both the sequences and positions of the three repeat types are remarkably well conserved among the three ciliate species. Considering only the DNA sequence, the conservation of the repeat elements is: type III >type I >>type II, while the constraints on the number and location of the repeats are: type I >typeII >type III. It is in fact because of the extremely high conservation of both the position of the type I repeats, and the sequence around the transcription start site, that by analogy with the two Tetrahymena species we infer transcription in Glaucoma to initiate with the A residue at position 1545. Finally, it is notable that the sub-domain of a type I and three type III repeats proximal to the transcription initiation site in T. thermophila is duplicated approximately 400 bp upstream in this species. This subdomain is part of a larger ~420 bp tandem duplication from bp 500 to bp 1340 that is not found in T. pyriformis or G. chattoni.

The 3'NTS Regions of Tetrahymena and Glaucoma rDNA

We determined the complete nucleotide sequence of the G. chattoni rDNA 3' NTS and 260 bp of the adjacent 26S rRNA coding sequence (Fig. 4c,d). We also completed the sequence of the corresponding region in T. thermophila rDNA (Fig. 4a,b). The sequences in Fig. 4 both read from 35S pre-rRNA towards the telomeres and are numbered so that transcription termination is at bp 260. Again, the telomere associated sequences of the G. chattoni 3'NTS (bp 1151-1351) will be discussed in detail elsewhere (Challoner and Blackburn, manuscript in preparation), but are included here for completeness.

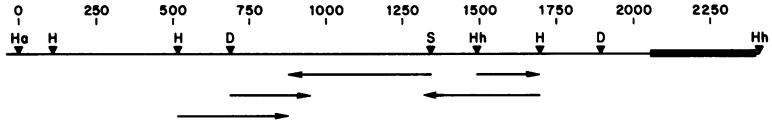

a
ETS
Gc.
T.p.
T.t.
100 bp
b
Glaucoma chattoni
Repeat I
a TTTTTTGGCAAAAAAAATCTCAGGTATACCAG
b TTTTTTGGCAAAAAAAAATCGGATATATCAG
c TTTCTTGGCAAAAAAAAAATCTGAGATCTCAG
Repeat II
a AAATTTCAGCGGGAT
b TCCGGCCAAATTTTCAGCGGGGAA
c TCACTCGATTTTTCAGCGGAGGAA
d TTAATTTCAGCGGGGAT
e TCGCTTAATTTCAGCGGAGAA
f TCCCAGCCAAATTTCTGCGGGGAG
g TCCCGCTGAATTAATCAGGTGGGGAG
Repeat III
a AAAAAAGACTTAGAAAAA
Tetrahymena pyriformis
Repeat I
a TTTCCTTGGCAAAAAAAAAAAAAAAAAAGAGTGG
b TTTTTGTGGCAAAAAAAAAAAAAAAATATAGTAG
c TTTTTCTGGCAAAAAAAAAAAAAAAAAGTATCAG
Repeat II
a AAAAGTCAGA
b GTCCAGTCAAAAAAGTGAGCCAG
c AAAAGTGAGCGAG
d TCCACTCAAAAATGTGAGCCAG
e ACCAGTCAAAAAAAGTGAGTGAG
f TCCACTCAAAAAAGTGACCCAG
g CCACTCAAAAAAAGTGAGTGAG
h CCCACTCAAAAAAATGAGTGAG
i TCCACTCAAAAAAAGTCATAAGT
j TCCAGTCAAAAAAGTGAGCCAG
k TTCCGTCAAAAAAAATGAGTGGAG
Repeat III
a AGACTTAGAGAAAAAAA
b AAAAAGACTTAGAAAAAAAAT
c AAAAAGACTTAGAGAAAAAAA
d AAAAAGACTTAGAGAAAA
e AGACTTAGAGAAAA
Tetrahymena thermophila
Repeat I
a TTTTTTTGGCAAAAAAAAAAACAAAAATAGTAA
b TTTTTTTGGCAAAAAAAAAAACAAAAATAGTAA
c TTTTTTTGGCAAAAAAAAAAACAAAAATAGTAA
d ATTCTTTGGCAAAAAAAATAAAAATAATATCAG
Repeat II
a CACTTACAAAAAATGACGGGA
b CTCGCTCAAATATTTAAGTGG
c ACTCGCATAAAATGAGTGAG
d TCACTAAAAAAATTAAGTGAA
e CTCACTTAAAAATGAGTGGAG
f CCACTCAAAAAATTAAGTGA
g CTCACTTAATATTCGGCGGAG
h TTAAACAAAAATAAGTGGA
i CTCACAGAAAATTAAGCGGAT
j TGCGCTAAAAAATGAGTGGA
k CTCGCTTAAAAATGAGTGGAG
l CCACTCAAAAGTTCAGCAGAG
m CCACTTAAAAATTTAGCTTA
Repeat III
a AAAAAAAGACTTAGAAAAAA
b AAAAAAAGACTTAGAGAAAA
c ATAAAAAGACTTAGAGAAAA
d AATAGTAAACTTAGAAAAAT
e AAAAAAAGACTTAGAGAAAA
f GAAAAAAGACTTAGAAAAAT
g TAAATTAGACTTAGTGAAAA

Computer analysis of these Glaucoma and Tetrahymena sequences revealed that the most extensive homology is found in the 3' end of the 35S pre-rRNA gene and the adjoining 40 bp of the NTS. The alignment of this region for maximum homology, including the published sequence from T. pyriformis rDNA (19), is shown in Figure 5. Although the 35S pre-rRNA transcription termination site of G. chattoni has not been directly mapped by S1 nuclease protection experiments, it has been mapped for both T. thermophila and T. pyriformis rDNA (18,19). The strong conservation of this sequence in G. chattoni argues that this sequence is also the transcription termination region of G. chattoni. This conclusion is reinforced by restriction mapping and sequencing data obtained for G. chattoni sequences upstream in the 26S rRNA coding region (data not shown), where the conservation of restriction sites and DNA sequences between the T. thermophila and G. chattoni pre-35S rRNA genes allows the unambiguous alignment of the upstream 26S rRNA coding sequences.

The analyses of the 3'NTS regions of Tetrahymena and Glaucoma rDNA revealed two new types of spacer repeats. First, as shown in Figures 4b and 4d, the 3'NTS's of G. chattoni and T. thermophila each contain a characteristic array of repeated DNA sequences, which we denote type IV repeats. The alignment of sequences surrounding the transcription termination regions of G. chattoni, T. thermophila and T. pyriformis also shows that the two type IV repeats closest to the transcription unit are common to all three species (Fig. 5). The position relative to the transcription termination site of the first type IV repeat, IVa, is highly conserved, being 10 bp downstream from the termination site in all three species; the second repeat, IVb, is separated from IVa by a distance which varies between 79 and 89 bp in these ciliates. In G. chattoni seven other type IV repeats (IVc-IVi) are distributed downstream from repeats IVa and IVb as shown by the boxed sequences in Fig. 4d. The positions and sequences of the type IV repeats from

Fig. 3. Organization and sequences of three repeat types in the rDNA 5'NTS.
a. Organization of the three repeat types, using momenclature of Niles et al. (15), in the 5'NTS of G. chattoni (G.c.), T. pyriformis (T.p.) and T. thermophila (T.t.), aligned from transcription initiation sites at the extreme right of the figure. Members of each repeat type are named above the filled boxed areas representing the repeat. Open boxed areas between repeats represent non-repetitive sequence.
b. Comparison of the primary sequence of all Type I, II and III repeats of G. chattoni, T. pyriformis and T. thermophila. Nomenclature is consistent with part a of this figure. Sequences for T. pyriformis are from refs. 15 and 33.

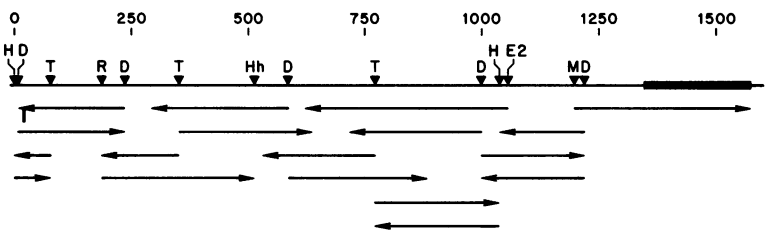
a



b

```
GAATTTCTAC GATCTGCTGA GATTCAGCCC GTCTCCTTAG ATTTATCTCA TCTCCCTTTA TTTTTTACTT CTGCTGGGGT TGTTAACTCT TAAGAAATTT 300
                                                                    IVa
TTTATGTTTT GATTTGTTTA ATTTAATTTT GTTTAACTTT AGTAAATTTT TTTCCTTTTT TCACTCACTG GGTTATTAAA TACTTAGAGA TTTTACATTT 400
                                                          IVb
TATCAAATAA TTTATGAACT CATTAAAACA AAACAAAACA AAGTTAAAAA AAACTCAATA GATTTGCTAA TAAGATGCAA AGCAGCTGGA AGGCAATTTT 500
TCTCATTTGG AAAGCTTCAA TCAACTGCTT ATAAATAAAT TATAAATCAA TATTAAAAAT GTTAAAGTTT TATGTTATTT GTTAGTAAAA AAATTGAATA 600
GTTGTGTTTA AGCTGATATA AGTCTTTATG CATGATATGT TAAAAGTTAC GCTTAAAATT ATGCTTTTTA CGCAGAATGA GCTTAGCTAA ATTTTTTCTC 700
AAAGTAAATT TTTTTTAATC CAAAATGAAT GAAAAAATTT TAGTATTTTA TAAAAATTCT TCATTCAAAT TTTACCCACT TATCAATTTA TTTTTTTTTG 800
                                                                                    Va
TGACTAAAGC AGTCGCAGAG CCTTTCTCTA AAAGTTGAAT TTTATTAACA ATGCCACTTT TATAGAAAAT TTTGCATGGA TTTCCTGGGG CTCCAATGGA 900
AAAATTGCGA AAGTGGATTC GAATGAAAAA GTGAATGTAA AAATTAAAGT AAAATTTTGC TTTATAAAAT GAATGAAAAT TTAAAACAAC CAAAATGAAT 1000
                Vb3          Vb2                                   Vb1
GAAAAAATCT TTGCATTTTA ACAAAATTTT TCATTCAAAA TTTCACCCAC TTATCAATAC TTTTTTTTCT CCAGCCGTCA AAGACCTATA TTGTTTTCCT 1100
                                                          Vb
AAAAGTTGAC TTTTATTAAG AACACAAATG AATGAAAAAA TCTTTGTATT TTAACAAAAT TTTTCATTCA AAATTTCACC CACTTATCAA TATATTTTTT 1200
                                                                                    Vc
TTAAGTCCTG CCTGCAGGCT AATATTCTTT CCTAAAAGTT GACTTTTATT AAGAACCAAA ATGAATGAAA AAATCTTTGT ATTTTAACCA AATTTTTCAT 1300
TCAAAATTTC ACCCACTTAT CAATATTTTT TTTGGGACCA AACCAAAAGA CCTAAAGATT TGATTTAAAA AGTTGACTTT TTTCAAGAAA ACCACTTTAT 1400
                Vd
TAGATAAATC TCTTTTTTAC CACGGCTTGT CCAATGAATA ATTTGCTAAA GTGGATTTGA ATAAAAATTT TTTTGCGTGT AAAAATGCGC TAAACTACGC 1500
TTAGATTTTA ACTTTATCCC ACTTTAATTT CAAGCGTAAA AATAAAAATC CCACACAAAA ATTAAGTGGA AATTGATGCA AAAATTTCAC TAAAATTTAA 1600
TTCAATAAAT ATGTAAAAAT GGTTGATCTC TATAATTTAT GAGATTTGCA TTATTTAAGG CTTATAAGAA ATTTTAAATT TAACGCGGAA GCTT 1694
```

c



d

```
TCTAAGTCAG AATCCATGCT GGAAAGCGAT GTCTTGGTGA TGATAAAACG AAAAAAACAT AAGAAATTAA GTTCGAAAGG TAGAGCGGAG AAGAGCGGAA 100
GAGCTTGATC TTAACTGCTA ATCGTAATTC CGAATTATCA TCAGCATAAA TCTTTTGTAG ACGACTTAAA ACGGAACGGG TATTGTACAC GTGAGAGTAG 200
AATCTTCTAC GATCCGCTGA GATTCAGCCC GTCTCCTCAG ATTTATCTCA TCTCCCTTTC GTTTCTTTCT CTGCTGGGGT TATTAACTAC TTATACAAAA 300
                                                                    IVa
AAAAAAAAGA AAATTAATTT TTGATTTTTT AATCTTGTGA TTCATCTGTC GATTTATTGT TTTGTCTCTC CTCCCCTGCT GGGGTTATTA AAATTAGTAA 400
                                                                              IVb
AATGAAATCT ATCTAAAAGT TTGCTGGGGT TATTAAATAA TGATTGGTCT TTCGTGGATT TTAAGTCAAA TTTAGTAATT TCTGCTGGGG TTTTTTAACT 500
                IVc                                                                 IVd
TTGATACTAA AGCGCTAAAT TCTGCTGGGG TTACTAAATA AATTGATGAC ATCTCTCTCT ATTATCCTGCT GGGGTTGTTA ACTACTTAGT TGAAATTAAA 600
                IVe                                                       IVf
AAATATCAAT CAATACTCTA CTGGGGTTAC TAAATCTTAT TTAGAATAAT TGTTATACTC GCAGGTCTGC TGGGGTTAAT AATTAGAATT TTTTAGAATA 700
                IVg                                                       IVh
ATCTCTCCTG GGGTTATTAA ACCCTTCTGA ATATGATAAA ATATTTGTCC ACTTTTTGGT TGGGTCTCGA GAGGATCGTT CACTGCGAGG TGCAATCGCC 800
IVi
CAATGAAACT TTAATTTCGC TTGTTTTTTT GTTTCATGTT AAAATGAATA AATATTGAAT GAAAAAATTA CATCCACTTT TTGCAGATTT TGAGATCCCT 900
Va
ATATATCATT TATTAAAAAC AAAAAATTTA TAAATAATTT TTCCTTTATA TTTTGTTTCA TTCAATTTTA TATGAGAACT AGTAAAAGTC ATTGTTAGCT 1000
                                                Vb
TGAAAATGTC CACTTTTTGG AAAAAGAATG TTTGGCCCAA GGGGTGGAGA CGAGGTGGGA CAGGCAAATT TGAAATTTTC CTTTATTTTT GGTTCCATTC 1100
                                                                    Vc
AAAAACCTTC AAAAATAAAA AATGAAGTTA ACAAGAGAAA ATATTATCGT TCATTTTCAA TGAAGTTAGT TCATAAAATG AAAATGATTT TTTCCGGCAG 1200
CGAGATAATT GAACTAAGAA CTATTTTATA TTCTTATATT TGTAGTATTC GACTTCTACA CTTAGATATA GCTCAAACAG CATTAATTAT ATCCATTTTC 1300
CATTCATCTT GCATGAGAAT CTATGCTGTT AGATAGAGTA TAGTAGGGGT T 1351
```

G. chattoni, T. thermophila and T. pyriformis are compared in Figures 6a and 6b respectively, showing that the 17 bp sequence is highly conserved both within and between these three ciliates. Repeats IVb to IVi are quite regularly spaced in G. chattoni, with the length of nonconserved sequence between the repeats varying from 21 to 44 bp.

The second type of repeated sequences in the 3'NTS regions of G. chattoni and T. thermophila, occuring distal to the type IV repeats, are termed type V repeats. The length of each type V repeat unit is ~130 bp. They are most obvious in the 3'NTS of T. thermophila, which has four type V repeats, Va-Vd, as shown in Figs. 4b and 6a. Repeats Vb, Vc and Vd are tandemly repeated. Repeats Va and Vb are separated by 138 bp that are notable not only for a length very close to the type V repeat length, but also for including three 11 to 16 bp sequences (termed Vb1, Vb2, and Vb3) that are duplications of one end of the type V repeats (Fig. 6b). In the lower portion of Fig. 6b the type V repeats of T. thermophila are aligned to show that while the majority of the repeat sequence is very well conserved, a 24 bp section beginning 31 bp from one end is different in each type V repeat unit. These "variable" regions have an average G+C content of 45%, or more than twice the 20% G+C found in the rest of the repeat sequences. The presence of these G+C rich variable sequences suggests that different sections of the type V repeats are subject to different selective pressure.

Fig. 4. Sequence of T. thermophila and G. chattoni rDNA 3'NTS and adjacent regions.

a. T. thermophila 3'NTS sequencing strategy and partial restriction map. Sequences were determined from the plasmid pTre1. The thicker segment of the map from 2053 to 2388 represents telomeric C_4A_2 repeats with rDNA 3'NTS to the left and pBR322 vector to the right of the C_4A_2 repeats. All regions not sequenced have been previously published (18,26). Sequencing strategy and restriction sites are as described for Fig. 1a, except S = Sau96I.

b. T. thermophila rDNA 3'NTS sequence reading in the 5'→3' direction beginnning with the last 60 bp of the pre-26S rRNA coding region (underlined, 18). Repeated sequences are boxed and named below each box. Thick underline in Type V repeat boxes indicates region of variable sequence. The listed sequence ends at the most distal HindIII site of the rDNA molecule (bp1689-1694). The remaining 358 bp of the 3'NTS is published in ref. 26.

c. G. chattoni 3'NTS sequencing strategy and partial restriction map. Sequences were determined from the plasmid pGre7. The thicker segment of the map from 1346 to 1574 represents telomeric C_4A_2 repeats with rDNA 3'NTS to the left and pBR322 vector to the right of the C_4A_2 repeats. Sequencing strategy and restriction sites are as described for Fig. 1a.

d. G. chattoni rDNA 3'NTS sequence reading in the 5'→3' direction beginning with the last 260 bp of the pre-26S rRNA coding region. Transcribed region and repeats are indicated as described for Fig. 4b. The listed sequence ends at the most proximal telomeric C_4A_2 repeat (bp 1346→1351, double underline).

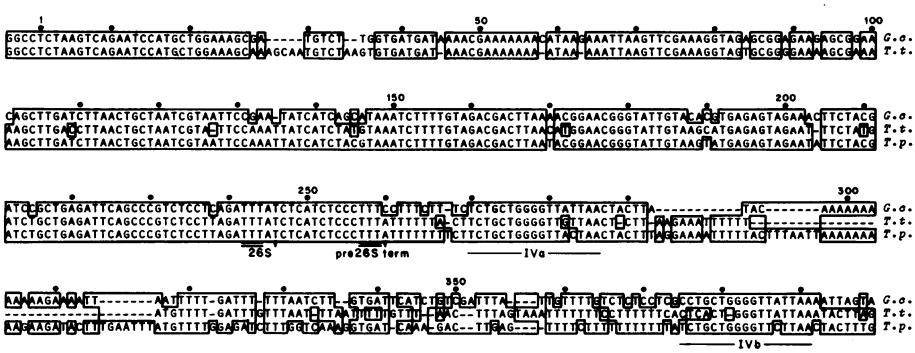


Fig. 5. Comparison of the pre-rRNA transcription termination site and 3' flanking regions. The optimal alignment of the region surrounding the pre-26S rRNA transcription termination site that is conserved among G. chattoni (G.c.), T. thermophila (G.t.) and T. pyriformis (T.p.) is shown (see 18,19). - indicates 1 bp gap introduced to maximize alignment. Sequences are 5' to 3' left to right so that transcription is rightward toward the terminal nucleotides of 26S and pre-26S rRNA as determined for T. pyriformis (=; 19) and T. thermophila (▼; 18). Boxes enclose positions where identical bases are found in at least two of the species. The numbering refers to the G. chattoni sequence and is consistent with Fig. 4d. Type IV repeats are underlined.

The structure and location of the analogous G. chattoni type V repeats are given in Figures 6a and 6b. Like the T. thermophila type V repeats they contain regions of variable sequence, but in G. chattoni each type V unit has two regions of variable sequence, and the ratio of variable to conserved sequence is much greater. The Glaucoma and Tetrahymena type V repeats have short sequences in common, such as CCACTT and TTTCATTCAA, but in general the sequence of the type V repeats is poorly conserved between the two organisms. However, three features of the Glaucoma and Tetrahymena type V repeats are very similar. As mentioned above, their length and the presence of variable sequences within the repeats are conserved between Glaucoma and Tetrahymena. Also, the distances from the most proximal repeat, Va, to the transcription unit are 485 bp and 459 bp in Glaucoma and Tetrahymena respectively. Therefore, the most important characteristics of the type V repeats based on conservation among Glaucoma and Tetrahymena seem to be size, location, and morphology (i.e. the interspersion of "variable" and conserved sequences within the type V repeat unit).

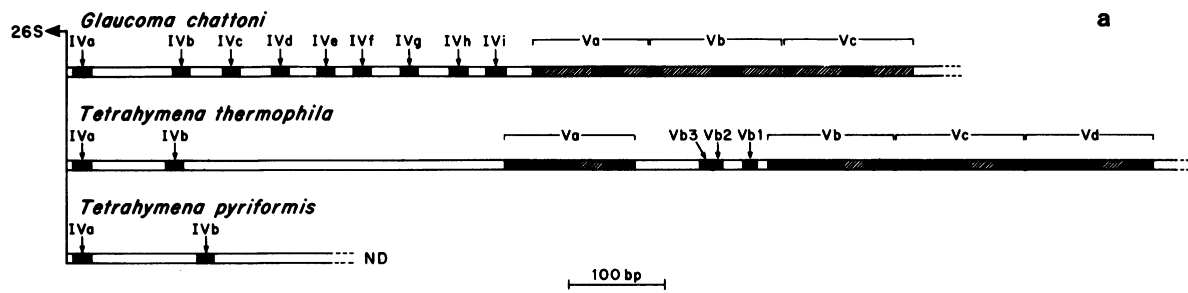
DISCUSSION

Structure of the Transcription Initiation Region

As in other eukaryotes, 5'NTS sequences and possibly sections of the 5' ETS in *Tetrahymena* and *Glaucoma* probably function as promoters and enhancers of rRNA transcription. Niles et al. (15,33) showed that the 590 bp of 5'NTS adjacent to the initiation site in *T. pyriformis* rDNA contain three types of dispersed repeated sequences. *T. thermophila* also has three repeated sequence types, organised in a similar array, in the 670 bp immediately adjoining the rRNA initiation site. We find a very similar pattern in the corresponding region of the *Glaucoma* 5'NTS, indicating that the array of promoter associated repeats is highly conserved (see Fig. 3). However, while the sequences of the type I and type III repeats in *Glaucoma* and *Tetrahymena* are conserved, the type II repeats at positions 1170 to 1400 in *Glaucoma* and 1360 to 1630 in *T. thermophila* are not homologous. A similar pattern of homology is observed when *Glaucoma* sequences are compared with those of *T. pyriformis*. There are no other long homologies between the *Glaucoma* and *Tetrahymena* 5'NTS sequences distal to the repeat array.

The most constant feature of the repeat domain in the 5'NTS adjacent to the 35S pre-rRNA coding region is the position of the two type I repeats within 120 bp of the transcription initiation site. A type I repeat is also found in all three organisms at the distal end of the repeat region, although the distance from this repeat to the initiation site varies from 432 to 637 bp. Next to this distal type I repeat, all three rDNA molecules have one to four type III repeats with a highly conserved primary sequence. The position of the type III repeats relative to the transcription initiation site and the position and number of the less conserved type II repeats are both quite variable. Nevertheless, the strong homologies between the *Glaucoma* and *Tetrahymena* upstream repeats, and the overall similarities of the arrays in each species suggest that these repeats have a function in the initiation of pre-rRNA transcription.

Several features of the ciliate rDNA promoter and upstream region are similar to the rRNA promoters of higher eukaryotes. Data from other organisms suggests that some sequences 3' to the NTS/ETS boundary may be essential for transcription initation (34-36). The extremely strong sequence conservation in this region among ciliates (see Fig. 2) is consistent with this hypothesis. In *Xenopus laevis* the segment from -142 to +6 bp (relative to the transcription start site) is required for maximal rRNA transcription (37 38); the corresponding region of the *D. melanogaster* rRNA promoter is located



b

```
TCTGCTGGGGTTATTAA IVa
CCTGCTGGGGTTATTAA IVb
TTTGCTGGGGTTATTAA IVc
TCTGCTGGGGTTTTTTA IVd
TCTGCTGGGGTTACTAA IVe   G. chattoni
TCTGCTGGGGTTGTTAA IVf
TCTACTGGGGTTACTAA IVg
TCTGCTGGGGTTAATAA IVh
TCTGCTGGGGTTATTAA IVi

TCTGCTGGGGTTGTTAA IVa   T. thermophila
CTCACT-GGGTTATTAA IVb

TCTGCTGGGGTTCTTAA IVa   T. pyriformis
TCTGCTGGGGTTATTAA IVb
```

Glaucoma chattoni

```
                   variable                                                      variable
TGTCCACTTTTTGGTtgggtctcgagaggatcgttcactgcgacgtgcaatcccccaatgaaagtttAATTT-CGCTTGTTTTTTTGTTTCATGTTAaaatg------aataaatattgaatg-------aaaaaatta Va
CATCCACTTTTTGGAgatttgagatccctatatat--catttattaaaaacaaaaaatttataaatAATTTTTCCTTTATATTTTGTTTCATTCAAttttatatgagaactactaaaagtcatt-cttagcttgaaaa Vb
TGTCCACTTTTTGGAaaaacaatgtttggcccaagggctggagaccaggtcggacacgcaaatttgaAATTT-TCCTTTATTTTTGGTTCCATTCAAaaaccttgaaaaataaaaaatgaagttaacaagagaaaatat Vc
```

Tetrahymena thermophila

```
                                                                                 variable
GCAAAATGAATGAAAAAATTTTAGTATTTTATAAAAATTCTTCATTCAAA-TTTTACCCACTTATCAATTTATTTTTTTtgtgactaaagcagtcccagag-CCTTTCTCTAAAAGTTGAATTTTATTAACAA Va
GCAAAATGAATGAAAAAATCTTTGCATTTTAACAAAATTTTTCATTCAAAATTTCACCCACTTATCAATACTTTTTTTTctccagccgtcaaagacctatattGTTTTC-CTAAAAGTTGACTTTTATTAAGAA Vb
ACAAAATGAATGAAAAAATCTTTGTATTTTAACAAAATTTTTCATTCAAAATTTCACCCACTTATCAATATATTTTTTTtaagtcctgcctgcaggctaatatTCTTTC-CTAAAAGTTGACTTTTATTAAGAA Vc
GCAAAATGAATGAAAAAATCTTTGTATTTTAACCAAATTTTTCATTCAAAATTTCACCCACTTATCAATATTTTTTTTTGggaccaaaccaaaagacctaaagaTTTGATTTAAAAAGTTGACTTTTTTCAAGAA Vd
ATAAAATGAATGAAAA    Vb1
      TGAATGTAAAAAT Vb2
      TGAATGAAAAA   Vb3
```

between -43 to -27 bp and -13 to +4 bp (39,40). A section of these functionally defined Xenopus and Drosophila promoters are duplicated a variable number of times in the rDNA NTS. In Xenopus the duplicated region is from -145 to about +1 (41,42); in Drosophila the duplicated region is from -18 to +24 bp (5). Clearly, the organization of the type I repeats in ciliates is qualitatively similar to these duplicated rRNA promoters in higher eukaryotes. Moss (3) has shown that in Xenopus oocytes injected duplicated promoters can bind RNA polymerase I and initiate transcription in the NTS. Transcription has also been observed to initiate in vitro in the analogous region of the Droosphila NTS (4). By analogy to these promoters and NTS promoter-like duplications in Xenopus and Drosophila, the type I repeats also probably interact with RNA polymerase I.

A second similarity between the Xenopus and ciliate NTS is that the Xenopus promoter duplications are separated by a series of tandem "60/80" repeats (41). The type II repeats in the Glaucoma and Tetrahymena 5'NTS are found in a location equivalent to the 60/80 repeats. Niles et al. (15) have postulated that if the type I repeats are in fact RNA polymerase I binding sites, then the type II and type III repeats may have a role in regulating or enhancing transcription initiation.

Comparison of the 3'NTS Regions of Glaucoma and Tetrahymena

The major conserved structure in the 3'NTS are the type IV and V repeats, embedded in otherwise diverged sequences. The alignment of Glaucoma and Tetrahymena sequences in Fig. 5 shows that the only highly conserved sequences in the 3'NTS immediately adjoining the termination site are the IVa and IVb repeats, although by fixing the position of these repeats, some scattered homology in the sequences between IVa and IVb suggests a common origin for this region. In addition, a T-rich sequence, characteristic of RNA polymerase I termination sites, precedes both IVa and IVb. Thus, it is possible that IVb and the T-rich sequence may be an additional, or "failsafe", termination site for the 35S pre-rRNA transcript. While the strong conservation of the

Fig. 6. Organization and sequences of two repeat types in the rDNA 3'NTS.
a. Organization of the Type IV and Type V repeats in rDNA 3'NTSs aligned from transcription termination sites at the extreme left of the figure. Members of each repeat type are named above the filled boxed areas representing the repeat. Variable regions of Type V repeats are shown by hatched bars. Open boxed areas between repeats represents non-repetitive sequence. ND = not determined.
b. Comparison of the primary sequence of all known type IV and V repeats of G. chattoni, T. pyriformis and T. thermophila. Nomenclature is consistent with part a of this figure. Sequences for T. pyriformis are from ref. 19.

sequences and positions of IVa and IVb relative to the transcription termination site makes it very probable that they play a role in transcription termination, this does not explain the function of repeats IVc to IVi in the Glaucoma 3'NTS, which lack adjacent T-rich sequences.

A peculiarity of the highly conserved type IV repeats is that the central 7 bp, AACCCCA, of the 17 bp sequence is also found in the telomeres of Glaucoma and Tetrahymena macronuclear DNAs. The telomeric sequence in these organisms consists of tandem repeats of the hexanucleotide CCCCAA (14,25). Furthermore, the AACCCCA sequence in the type IV repeats has the same strand orientation as the distal telomeric sequence. In the hemoflagellate Trypanosoma brucei a similarly non-random occurrence of telomere-like sequences occurs just downstream from some copies of variable surface glycoprotein genes (24,43). Like the Tetrahymena rRNA gene, these expressed or expressable genes are located near the ends of nuclear DNA molecules, and the strand orientation of the telomere-related sequences also matches that of the telomeric repeats (43). The significance of the non-random occurrence of telomere-related sequences downstream from telomeric genes is not currently understood.

The type V repeats which we have identified in G. chattoni and T. thermophila are conserved in position relative to the transcription unit, but not relative to the end of the macronuclear rDNA molecule. In data presented elsewhere (Challoner and Blackburn, manuscript in preparation), we have shown that in two strains of Tetrahymena pigmentosa, the sequences extending 700 and 1300 bp from the rDNA telomeres do not contain type V repeats. Thus, a role for the type V repeats in telomere formation or function seems unlikely. An analysis of T. thermophilia 3'NTS chromatin structure (Budarf and Blackburn, manuscript in preparation) revealed that the type V repeats define a region of unique non-nucleosomal chromatin that is bounded on the distal side by phased nucleosomes. Taken together, these results suggest that type V repeats function either in the packaging of rRNA genes in the nucleolus or in transcriptional control.

There are several striking similarities between ciliate type V repeats and a set of tandemly repeated sequences recently identified in the 3'NTS of Physarum polycephalum (44). Specifically, the Physarum repeats are also 130 bp long, and are located in the same position relative to the rRNA coding region (i.e. they begin approximately 470 bp from the termination site). The ciliate type V repeats and the Physarum 130 bp repeats do not share any significant sequence homologies: this is not surprising, since even at the

evolutionary distance between Glaucoma and Tetrahymena the type V repeat primary sequence has diverged.

Evolution of Spacer Repeats in Eukaryotic rRNA Genes

Our analysis of both the 5' and 3' NTS of extrachromosomal rDNA from two ciliated protozoan species reveals conservation of both sequence and position of five types of repeated elements. All of these conserved sequence elements are embedded in a matrix of nonconserved sequences. Repeat types I-III were identified previously in the T. pyriformis 5'NTS (15); our finding that they are highly conserved in two other ciliate species provides evidence for their functional importance. Repeat types IV and V have not been identified before.

The domain of repeats adjacent to the transcription initiation site in ciliates (types I-III) is the only example in which classes of repeat elements upstream from the rRNA transcription initiation site have been distinguished by different rates of sequence divergence. Similarly, of the 3'NTS repeats, we find that the type IV repeats are analogous to type I repeats in the strong conservation of both sequence and position relative to the transcription unit. In contrast, the type V repeats, like the type II repeats, consist of sequences that have diverged between Glaucoma and Tetrahymena, but are highly conserved in position relative to the transcription unit.

The rapid rate of type II repeat sequence divergence may be relevant to studies showing strong species specificity of the rRNA promoter in cell-free transcription systems from other eukaryotes (11,45,46). In these experiments, cellular extracts for RNA polymerase I transcription initiated transcription from the homologous, but not the heterologous, rDNA promoter, even when promoter-homologous RNA polymerase I was included in the assay. Thus, the species specificity of rRNA transcription initiation is not solely a function of the polymerase and promoter interaction, but must be mediated by other trans-acting factors. The molecular coevolution of such factors and corresponding rapidly evolving NTS recognition sites provides the simplest explanation for this specificity. In the ciliate 5'NTS sequences described here, the divergent type II repeats are a probable source of species specificity at the ciliate rDNA promoter.

The type III repeats, although variable in number and position, have the most highly conserved sequence of the 5'NTS repeat types. These repeats have been shown in Tetrahymena to be sites for cleavage (O. Westergaard, personal communication) by a strand specific, endogenous SDS dependent nuclease (47). Although the type III repeats in T. pyriformis and the upstream copy of the repeats in T. thermophila are near the in vivo origin of bidirectional

replication of these rDNAs (20,21), they are not necessarily involved with replication function; only one of the two ARS elements in the 5'NTS of T. thermophila contains type III repeats (26,48). Further discussion of the replication of the protozoan rDNAs is presented in the accompanying paper (48).

The structural organization of spacer repeats that we find in ciliate nontranscribed spacers have obvious implications not only for their functional significance, but also for the evolutionary relationship of spacer repeats in general. Clearly, as the sequence of a set of repeat types changes, there is some mechanism which ensures that all elements of the set remain homogeneous. It has been suggested that tandemly repeated sequences will be homogenized by unequal crossing-over between the repeats in sister chromatids (49,50). However, in each ciliate species all the arrays of NTS repeats are characterized by the presence of highly conserved regions interspersed with divergent sequences. This interspersion of conserved and divergent sequences in an array precludes simple mechanisms of unequal crossing over or gene conversion to maintain homogeneity within a species, since neither unequal crossing-over nor gene conversion alone would result in the observed dispersal of conserved repeats in a matrix of nonconserved DNA.

Furthermore, because there is only one rDNA copy in the T. thermophila germline genome, and probably one or only a few in G. chattoni, no opportunity exists for the spread of changes in these rRNA gene spacers by unequal crossing over between tandem repeated rDNA units. These considerations argue strongly that there must be functional selection to maintain the sequence and relative locations of each type of repeat, and that the selective pressure must be different at each set of repeats because of the observed difference in the rate of sequence divergence. Thus, the concept of molecular coevolution (12) agrees well with our data on the rDNA sequence similarities and differences between ciliate species. Type I and type IV repeats are highly conserved in sequence and position relative to the transcription unit, suggesting that they interact with transcriptional enzymes or factors which do not differ greatly in DNA recognition specificity between Glaucoma and Tetrahymena. In contrast, type II and type V repeat sequences, while well conserved within a species, are not well conserved between Glaucoma and Tetrahymena; this suggests an interaction with transcriptional or nucleolar protein factors that have also changed in the two ciliate species.

In summary, the unusual germline genetic content of rRNA genes in ciliates (one to at most a few gene copies), and the unusual structure of

their transcribed genes (one or two genes per linear macronuclear molecule) have provided a useful system for relating spacer features both to the constraints imposed by rDNA function, and to the mechanisms operating on these spacers in evolution. Molecular coevolution of spacer sequences with transcriptional and other factors can account for the differences we observe in NTS repeat types between ciliate species; functional cosntraints appear to act selectively on the repeats within a species to keep them homogeneous without a contribution from passive unequal crossing over mechanisms.

ACKNOWLEDGEMENTS

This work was supported by grants from NIH (GM 32565 and GM 26259) to E.H.B. and from NCI and NSERC to R.E.P. We thank Yau-Hing Pak and Mike Cherry for expert assistance with the computer analysis, Gordon Temple for the photography and Drena Larson and Marietta Dunaway for critical reading of the manuscript.

+To whom correspondence should be sent

*Present address: Fred Hutchinson Cancer Center, 1124 Columbia Street, Seattle, WA 98104, USA

REFERENCES

1. Long, E. O. and Dawid, I. B. (1980) Ann. Rev. Biochem. 49, 727-764.
2. Reeder, R. H. (1984) Cell 38, 349-351.
3. Moss, T. (1983) Nature 302, 223-228.
4. Kohorn, B. D. and Rae, P. M. M. (1982) Nucl. Acids Res. 10, 6879-6886.
5. Coen, E. S. and Dover, G. A. (1982) Nucl. Acids Res. 10, 7017-7026.
6. Federoff, N. V. (1979) Cell 16, 697-710.
7. Dover, G. (1982) Nature 299, 111-117.
8. Coen, E., Strachan, T. and Dover, G. (1982) J. Mol. Biol. 158, 17-35.
9. Grummt, I., Roth, E. and Paule, M. R. (1982) Nature 296, 173-174.
10. Learned, R. M. and Tjian, R. (1982) J. Mol. Appl. Gen. 1, 575-584.
11. Miesfield, R. and Arnheim, N. (1984) Mol. Cell. Biol. 4, 221-227.
12. Dover, G. A. and Flavell, R. B. (1984) Cell 38, 622-623.
13. Yao, M.-C. and Gall, J. G. (1977) Cell 12, 121-132.
14. Katzen, A. L., Cann, G. M. and Blackburn, E. H. (1981) Cell 24, 313-320.
15. Niles, E. G., Sutiphong, J. and Haque, S. (1981) J. Biol. Chem. 256, 12849-12856.
16. Saiga, M., Mizumoto, K., Matsui, T. and Higashinakagawa, T. (1982) Nucl. Acids Res. 10, 4223-4235.
17. Engberg, J., Din, N., Saiga, M. and Higashinakagawa, T. (1984) Nucl. Acids Res. 12, 959-972.
18. Din, N., Engberg, J. and Gall, J. (1982) Nucl. Acids Res. 10, 1503-1513.
19. Niles, E. G., Cunningham, K. and Jain, R. (1981) J. Biol. Chem. 256, 12875-12860.
20. Truett, M. A. and Gall, J. G. (1977) Chromosoma 64, 295-303.
21. Cech, T. R. and Brehm, S. L. (1981) Nucl. Acids Res. 9, 3531-3543.

22. Wild, M. A. and Gall, J. G. (1979) Cell 16, 565-573.
23. Din, N. and Engberg, J. (1979) J. Mol. Biol. 134, 555-574.
24. Blackburn, E. H. and Challoner, P. B. (1984) Cell 36, 447-457.
25. Blackburn, E. H. and Gall, J. G. (1978) J. Mol. Biol. 120, 33-53.
26. Kiss, G. B., Amin, A. A. and Pearlman, R. E. (1981) Mol. Cell. Biol. 1, 535-543.
27. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) in Molecular Cloning. Cold Spring Harbor Laboratory, Box 100, Cold Spring Harbor, New York.
28. Birnboim, H. C. and Doly, J. (1979) Nucl. Acids Res. 7, 1513-1522.
29. Maxam, A. M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
30. Delaney, A. D. (1982) Nucl. Acids Res. 10, 61-67.
31. Kiss, G. B. and Pearlman, R. E. (1981) Gene 13, 281-287.
32. Pan, W. C. and Blackburn, E. H. (1981) Cell 23, 459-466.
33. Niles, E. G. (1985) J. Biol. Chem. 260, 672-678.
34. Urano, Y., Kominami, R., Mishima, Y. and Muramatsu, M. (1980) Nucl. Acids Res. 8, 6043-6058.
35. Rothblum, L. R., Reddy, R. and Cassidy, B. (1983) Nucl. Acids Res. 10, 7345-7362.
36. Miesfield, R. and Arnheim, N. (1982) Nucl. Acids Res. 10, 3933-3949.
37. Moss, T. (1982) Cell 30, 835-842.
38. Sollner-Webb, B., Wilkinson, J. K., Roan, J. and Reeder, R. M. (1983) Cell 35, 199-206.
39. Kohorn, B. D. and Rae, P. M. M. (1983) Proc. Natl. Acad. Sci. USA 80, 3265-3268.
40. Kohorn, B. D. and Rae, P. M. M. (1983) Nature 304, 179-181.
41. Boseley, P. G., Moss, T., Machler, M., Portmann, R. and Birnstiel, M. L. (1979) Cell 17, 19-31.
42. Moss, T., Boseley, P. G. and Birnsteil, M. L. (1980) Nucl. Acids Res. 8, 467-485.
43. Van der Ploeg, L. H. T., Liu, A. Y. C. and Borst, P. (1984) Cell 36, 459-468.
44. Hattori, M., Ljljana, A. and Sakaki, Y. (1984) Nucl. Acids Res. 12, 2047-2053.
45. Cizewski, V. and Sollner-Webb, B. (1983) Nucl. Acids Res. 11, 7043-7056.
46. Wandelt, C. and Grummt, I. (1983) Nucl. Acids Res. 11, 3795-3809.
47. Gocke, E., Bonven, B. and Westergaard, O. (1983) Nucl. Acids Res. 11, 7661-7678.
48. Amin, A. A. and Pearlman, R. E. (1984) Nucleic Acids Res., 13, 2647-2659
49. Smith, G. P. (1976) Science 191, 528-535.
50. Keil, R. L. and Roeder, G. S. (1984) Cell 39, 377-386.