Volume 15 Number 10 1987 Nucleic Acids Research

An unusually compact ribosomal DNA repeat in the protozoan *Giardia lamblia*

John C.Boothroyd[1], Alice Wang[2], David A.Campbell[1,3] and Ching C.Wang[2]

[1]Department of Medical Microbiology, Stanford University School of Medicine, Stanford, CA 94306, [2]Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, CA 94143 and [3]Department of Microbiology and Immunology, School of Medicine, University of California, Los Angeles, CA 90024, USA

Received January 12, 1987; Revised and Accepted April 29, 1987 Accession no. Y00331

ABSTRACT

The ribosomal RNA (rRNA) genes of the protozoan parasite Giardia lamblia have been analyzed with respect to size, composition and copy number. They are found to be remarkable in several respects. First, the rRNAs themselves are the smallest yet reported for any eukaryotic organism. Second, the genes encoding them are found as an exceptionally small tandemly repeated unit of only 5.4 kilobase-pairs. Third, the genes are extraordinarily G:C rich, even in regions which are highly conserved between all other eukaryotic rRNA genes. Finally, by analogy to other organisms, the 5.8S RNA appears to lack about 15 nucleotides from its 3'-end, a region previously thought to be essential for 5.8S RNA function. We also provide the first estimates of the genomic complexity and total G:C content of this important protozoan pathogen.

INTRODUCTION

The ribosomal RNA (rRNA) genes are amongst the most conserved of all genes so far studied from both prokaryotes and eukaryotes (for review, see refs. 1,2,3). Within eukaryotes, they are typically transcribed by RNA polymerase I to produce a large primary transcript (45S or pre-rRNA) which is processed by endonucleolytic cleavage within the nucleolus. This processing results in the removal of the internal and external transcribed spacers (ITS and ETS, respectively) which are rapidly degraded. For vertebrates, the stable transcripts constitute the 18S, 5.8S and 28S rRNA components of the ribosome. Because of variability in their size, particularly in lower eukaryotes, we will use the terminology SSRNA and LSRNA to describe the RNAs found in the small and large subunits of the ribosome, repsectively. Typically, the coding regions are arranged 5'-ETS/ SSRNA /ITS1/ 5.8S RNA /ITS2/ LSRNA /ETS-3', frequently as tandem, head-to-tail

repeats separated by a non-transcribed spacer region. The gene encoding rRNA is also termed ribosomal DNA or rDNA.

One of the most interesting applications of the wealth of data on rRNA has been the construction of probable secondary structures based on evolutionarily conserved base pairing. The sequence data have also proven useful in deducing phylogenetic relationships between widely divergent organisms. This approach is likely to prove extremely fruitful in making sense of the remarkable variety of single-celled organisms within the protozoa.

The rRNA genes described in this paper are from the protozoan parasite Giardia lamblia which is of interest to us for its unusual biological and pathogenic properties. G. lamblia belongs to the class Zoomastigophora and order Diplomonadorida and is the causative agent of a severe form of diarrhoea. It lacks mitochondria, depending primarily on anaerobic glycolysis for its energy production and electron transport. Here, we present the characterization of the rDNA repeat unit from this protozoan and show that it is unusual in being exceptionally G:C rich and compact with the smallest stable transcripts and ITS regions yet reported for any eukaryote.

METHODS

Growth of Parasites and Preparation of RNA.

The Portland 1 strain of Giardia lamblia was obtained from Dr. Donald Lindmark of the Cleveland State University. A single organism was used to establish a culture of trophozoites which was expanded in axenic medium as described (4). Total cellular RNA was prepared from about 4 x 10^8 organisms by SDS lysis and extraction with hot phenol as described (5). This yielded about 5 mg of RNA. Total RNA was prepared from Toxoplasma gondii and Trypanosoma brucei by the same procedure. For Trichomonas vaginalis, total RNA was prepared as described (6) to eliminate degradation.

Sizing of RNA.

High molecular weight RNA was resolved by agarose gel electrophoresis under native and denaturing conditions. The native conditions were: 1.5% agarose in TBE (90 mM Tris-borate (pH

8.3), 2.5 mM EDTA) with a glycerol loading buffer (0.1 x TBE, 10% glycerol, 0.04% bromophenol blue (BPB), 0.04% xylene cyanol FF (XCFF)). The denaturing gel was 1.2% agarose in 2.2 M formaldehyde, 40 mM MOPS (pH 7.0), 10 mM sodium acetate and 1 mM EDTA. Samples were heated in denaturing loading buffer (50 % deionized formamide, 1.5 M formaldehyde, 20 mM MOPS (pH 7.0), 5 mM sodium acetate, 1.5 mM EDTA, 0.04 % BPB, 0.04% XCFF and 5% glycerol) at 85°C for one min. and then chilled on ice before loading.

Low molecular weight RNA was resolved by electrophoresis in 7 M urea/5% polyacrylamide gels, using TBE buffer. The loading buffer was 80% formamide, 0.5 x TBE, 0.04% BPB, 0.04% XCFF. Samples were heated to 90°C for one minute before loading. These gels were run at 400V so that the gel was about 60°C during electrophoresis.

RNA was visualized by staining with ethidium bromide. Sizes were estimated by comparison with RNA markers.

Preparation of DNA and Cosmid Libraries.

High molecular weight DNA was isolated by the method of Blin and Strafford (7). A cosmid library containing G. lamblia DNA was constructed in the vector pc2xB according to the method of Bates and Swift (8). Briefly, genomic DNA was partially digested with MboI to give fragments of about 30-35 kilobase-pairs (kb). After dephosphorylation with calf alkaline phosphatase, the DNA was ligated to BamHI and SmaI double-digested c2xB vector. The ligated DNA was packaged in vitro and transfected into E. coli.

Screening and Mapping Recombinants.

RNA, enriched for ribosomal RNA by agarose gel electrophoresis, was electroeluted and 5'-^{32}P labelled by incubation with polynucleotide kinase and [γ^{32}P]-ATP following mild alkaline hydrolysis, as described (9). This was used to screen a portion of the cosmid library described above by colony hybridization (10). Colonies giving a positive signal were picked and rescreened in an identical manner. The coding capcity of the recombinants was confirmed as being for the ribosomal RNA by labelling the insert (through nick translation (11)) and using this as a probe in Northern blot analysis of total RNA (10).

Following limited restriction mapping of several of the positive cosmid recombinants, it became clear that they contained

several copies of a tandemly repeated unit of about 5.4 kb. One copy of this repeat was subcloned (as a PvuII fragment) into the SmaI site of the plasmid vector pUC8 (12). This recombinant (termed pGRP1) was used for all subsequent analyses.

C_ot Analysis of Genomic Complexity.

DNA from either E. coli B (Sigma) or G. lamblia was sheared by sonication (Heat System W-375) to an average length of 400 bp. The fragments were precipitated with ethanol and redissolved in 200µl of pH 7.0 sodium phosphate buffer of either 0.15 M or 0.41 M (13). E. coli DNA was dissolved in the 0.15M phosphate buffer only, to a final concentration of about 5 µg/ml; G. lamblia DNA was dissolved in 2 cuvettes, one at about 5 µg/ml in 0.15 M phosphate and the other at about 120 µg/ml in 0.41 M phosphate buffer. The solutions were topped with mineral oil, heated to 100°C in a spectrophotometer equipped with thermal control and automatic plotter (Gilford 2600 and thermoprogrammer 250) and then held at 65°C for 72 hours for the fragments to reanneal. The difference in A_{260} at 100°C and at the end of 72 hours was taken as the increment in A_{260} when the DNA is 100% denatured, and the fractions of A_{260} increment at each time point taken as the fraction of dissociated DNA.

Quantitation of the Number of rDNA Repeat Units in the Genome.

Genomic DNA (2 µg) from G. lamblia was digested with PvuII and co-electrophoresed with various amounts of pGRP1 cut with EcoRI and HindIII to release the insert. To ensure comparable transfer efficiencies and mobilities, the digested pGRP1 was mixed with 2 µg of similarly digested Trypanosoma brucei DNA prior to electrophoresis. The resulting gel was blotted to nitrocellulose paper and probed with nick-translated pGRP1 insert. Hybridization was in 50% formamide, 0.9 M NaCl, 50 mM $NaPO_4$ (7.4), 5mM EDTA, 0.1% SDS, 2 mg/ml ficoll, 2 mg/ml PVP, 2 mg/ml BSA, 100 µg/ml sonicated salmon sperm DNA. Final washing was in 0.1% SDS, 0.1 X SSC (1 X SSC is 0.15 M NaCl, 0.015 M Na-citrate, pH 7.0) at 65°C. Given the genomic complexity of G. lamblia and the size of the plasmid, the molar amounts of the repeat unit could be calculated.

Determination of G:C Content of G. lamblia DNA.

The G:C content of G. lamblia DNA was determined by thermal denaturation using the method of Mandel and Marmur (14).

Briefly, 200 ng of G. lamblia DNA was dissolved in 200 µl of 1 X SSC. The solution was sealed with mineral oil and heated from 50°C to 100°C in a spectrophotometer (Gilford 2600 with thermoprogrammer 250) at a rate of 0.25°C per minute. Absorbance at 260 nm during this period was plotted and this information used to calculate the G:C content as described (14).

The G:C content of the G. lamblia rDNA as cloned in pGRP1 was similarly calculated except that 0.1 X SSC was used as the buffer to enable an accurate estimation of the G:C content. This was necessary because preliminary experiments in 1 X SSC gave a melting point near 100°C, preventing an accurate determination of the G:C content.

DNA Sequencing.

Restriction fragments of pGRP1 were 5'- or 3'-^{32}P labelled by kinasing or end-fill-in, respectively (10). These were then sequenced by the chemical method of Maxam and Gilbert (15) with gel electrophoresis and autoradiography as described (16) except that 6% polyacrylamide gels 80 cm long were also used and autoradiography was also without intensifying screens at room temperature to improve the resolution of the bands. All restriction sites were sequenced through, except those at the end of the insert.

S1-nuclease Mapping of the rDNA Repeat.

The recombinant pGRP1 was digested with one restriction enzyme and the site 5'- or 3'-labelled with ^{32}P by kinasing or end fill-in as described (10). This material was further cut with one or more other enzymes and the resulting uniquely end-labelled fragment purified by gel electrophoresis and electroelution. These fragments were individually mixed with G. lamblia RNA, denatured, annealed and then subjected to treatment with S1 nuclease as described (10). The products were analyzed by polyacrylamide gel electrophoresis in the presence of 7 M urea followed by autoradiography of the dried gel. Sizes of protected species were determined by comparison to either the homologous chemical sequence ladder or end-labelled, heterologous markers. Note that the sequence ladder is shifted 1.5 nucleotides relative to the S1-protected species because a band in a given chemical sequencing lane corresponds to a 3'-phosphorylated fragment (vs.

3'-OH for the S1 species) which is lacking the targeted nucleotide.

RESULTS

Sizing the ribosomal RNAs of Giardia lamblia.

Fig. 1a shows the relative electrophoretic mobility of the ribosomal RNAs of Giardia lamblia in comparison with the rRNAs of two other protozoan parasites, Trichomonas vaginalis and

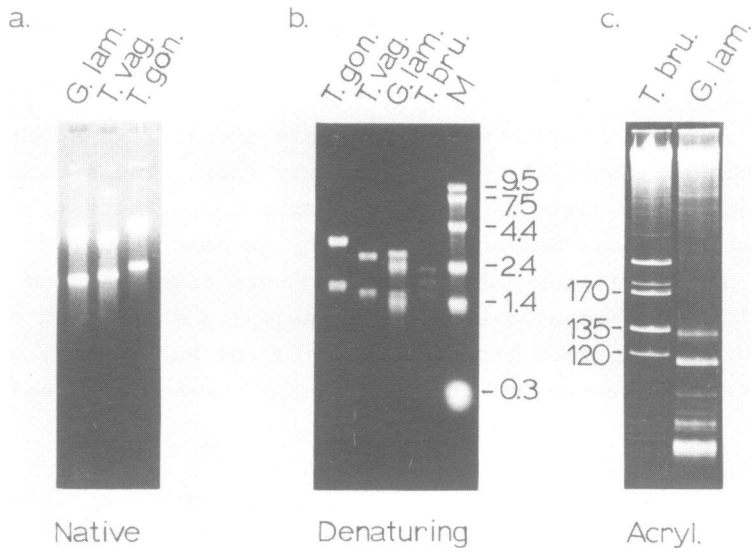


Figure 1. Electrophoretic sizing of G. lamblia rRNA. a. Agarose electrophoresis under native conditions. Approximately 0.5 µg of total RNA from each organism were analyzed in parallel as described in Methods. The sources of the RNA were Giardia lamblia (G.lam.), Toxoplasma gondii (T.gon.) and Trichomonas vaginalis (T.vag.). The two major bands in each sample correspond to the rRNAs. The minor band near the top of the gel is due to a small amount of contaminating DNA in the RNA preparations. b. Denaturing agarose gel electrophoresis. The source of the RNAs is as in part a., except that a sample of RNA from Trypanosoma brucei (T.bru.) was also analyzed. Approximately 1.4 µg of each RNA (0.2 µg for T. brucei) was treated as described in Methods followed by electrophoresis in the presence of formamide. RNA markers (M) are from Bethesda Research Laboratories. c. Polyacrylamide gel electrophoresis in the presence of 7 M urea. RNA (7 µg) from G. lamblia and T. brucei were treated as described in Methods and electrophoresed in 8% polyacrylamide. The sizes of the trypanosome RNAs are based on previously published estimates (24,25) and are ±3 nucleotides.

Toxoplasma gondii. Two discrete bands in the approximate size range expected for rRNA were seen for each sample. To determine their sizes more precisely, electrophoresis was also performed under strongly denaturing conditions (as described in Methods) in parallel with markers of known size (Fig. 1b). In comparison to the other RNA samples analyzed, four bands were apparent by ethidium bromide staining in the lane containing G. lamblia RNA. Two of these had a similar relative mobility to that seen under native conditions (Fig. 1a) whereas the other two bands were present as diffuse smears of faster mobility. This we attribute to a possibly incomplete denaturation of the rRNAs, so that the slower, sharper bands represent completely denatured molecules while the faster, more diffuse bands are partially renatured forms of the rRNA. These latter forms are likely only present in the samples containing rRNA of G. lamblia because of its extraordinarily high G:C content as discussed below. Based on comparison with the standards and assuming the slower migrating species are the denatured forms of the LSRNA and SSRNA, these data predict their respective sizes to be 2.9 and 1.5 kilobases. More accurate estimations of their sizes were made by S1-nuclease protection experiments as described below.

To determine the approximate sizes of the 5S and 5.8S (or their equivalent) RNAs in G. lamblia, RNA was analyzed by electrophoresis in polyacrylamide/7 M urea gels. After ethidium bromide staining, several bands were detected (Fig. 1c). These include two major bands with sizes larger than 100 nucleotides (nt) based on comparison with the size markers. One corresponds to about 134+3 nt, whereas the other is an apparent doublet at about 118±3 nt. The intensity of this doublet varies inversely with the intensity of the band at 134 nt, depending on the gel conditions. For example, when the gels were made up at a lower percentage acylamide (5%) and run at lower voltages so that the temperature of the gel was also lower (about 40°C), no band at 134 nt was detectable, and instead a more intense doublet at 118 nt was seen (data not shown). This suggests that, as with SSRNA and LSRNA described above, the band at 134 nt migrates as two forms depending on the extent to which it is denatured. Given their sizes, the constant band at 118 nt is most likely the 5S

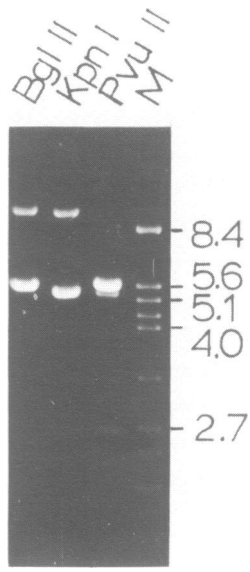


Figure 2. Restriction analysis of the rDNA repeat. A cosmid clone, identified as containing rDNA by colony hybridization (see Methods) was digested with the enzymes shown and analyzed by agarose gel electrophoresis. The sizes were estimated by ethidium bromide staining and comparison with DNA standards of known sizes (previously calibrated against lambda DNA cut with PstI), as indicated.

RNA of G. lamblia while the band at 134 nt (and, thus, the variable component of the doublet) is likely the 5.8S RNA equivalent of this protozoan. These designations have been confirmed by limited sequence analysis on end-labelled RNA excised from the gels corresponding to the two bands (J.C.B., unpublished results). The migration of the 5.8S RNA in either of two manners (depending on the extent of denaturation) is again likely due to its unusually high G:C content (see below).

Identification of recombinants containing the rDNA gene.

Using ^{32}P-labelled RNA enriched for rRNA as a probe, cosmids containing the rDNA gene from G. lamblia were identified as described in Methods. Digestions of such a cosmid with three restriction enzymes are shown in Fig. 2. From this, it can be seen that there is a strong band at about 5.4 kb with PvuII and BglII while KpnI gave a slightly smaller band at about 5.1 kb.

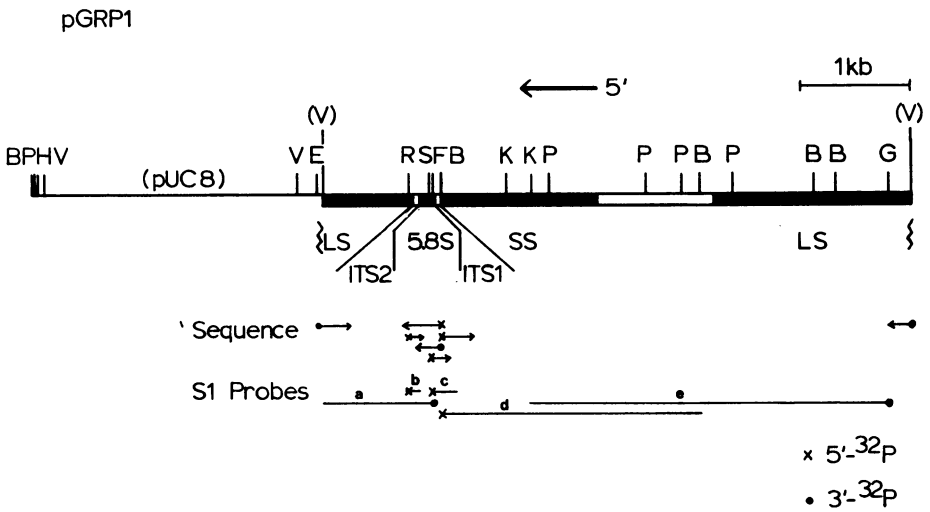


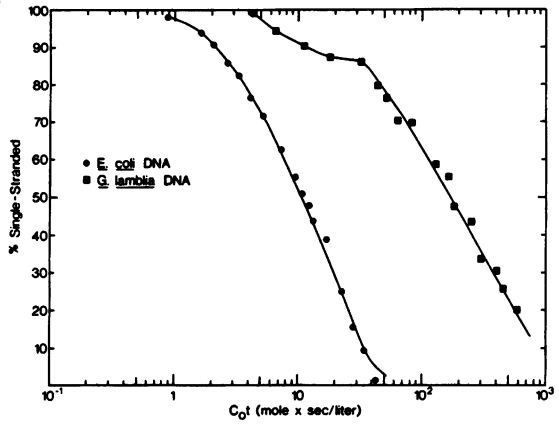
Figure 3. Map of a cloned rDNA repeat unit. The plasmid pGRP1 was constructed by inserting a PvuII fragment containing a complete repeat unit of the rDNA gene into the SmaI site of the polylinker in pUC8. The regions represented in mature rRNA are indicated by the solid black boxes (abbreviations are described in the text). The direction of transcription (indicated by the arrow) is from right to left from an unknown promoter. The sequencing strategy is given below the map. B, BamHI; E, EcoRI; F, FokI; G, BglII; H, HindIII; K, KpnI; P, PstI; R, RsaI; S, SmaI; V, PvuII. All sites for each enzyme are shown except for FokI, RsaI and SmaI. Fragments used in S1-nuclease protection studies are shown at the bottom of the figure. These are (a) 3'-5.8S; (b) 5'-LS; (c) 5'-5.8S; (d) 5'-SS; (e) 3'-LS.

One or two lighter bands were also seen in each digest. This basic pattern was seen in several independent cosmid clones suggesting that the rDNA is a tandemly repeated unit of about 5.4 kb, containing two closely spaced KpnI sites and one site each for BglII and PvuII. The lighter fragments presumably represent junction fragments comprising one repeat and all or part of the cosmid vector. One repeat unit (as generated by PvuII) was subcloned into the BamHI site of pUC8. A map of the resulting recombinant, pGRP1, is shown in Fig. 3. The coding regions and portions sequenced are also shown.

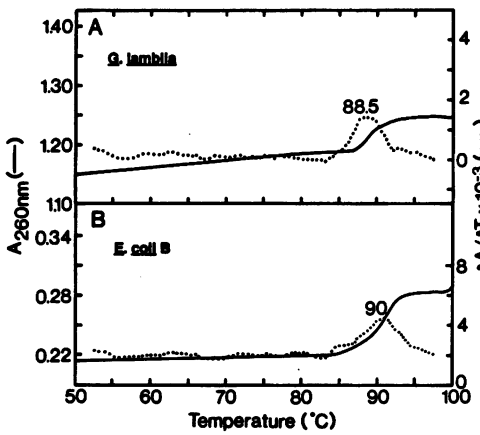
$C_o t$ Analysis of Genomic Complexity.

Using the procedure described in Materials and Methods, a renaturation profile for G. lamblia DNA compared with that of E. coli was determined (Fig. 4a). These results indicate that G.

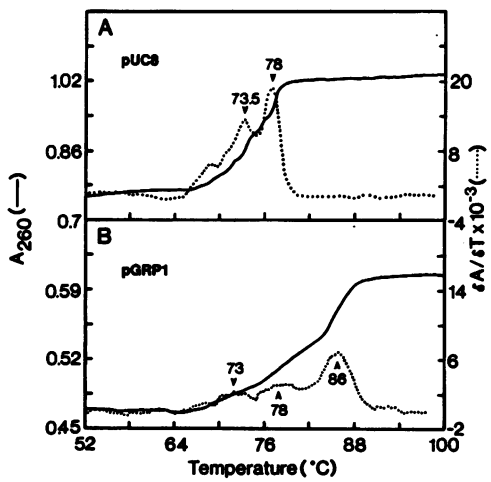
a.


% Single-Stranded
E. coli DNA
G. lamblia DNA
C_0t (mole x sec/liter)

b.


A
G. lamblia
88.5
B
E. coli B
90
A_{260nm} (—)
$\delta A/\delta T \times 10^{-3}$ (.....)
Temperature (°C)

c.


A
pUC8
78
73.5
B
pGRP1
73
86
78
A_{260} (—)
$\delta A/\delta T \times 10^{-3}$ (.....)
Temperature (°C)

lamblia DNA has two components: 14% having a $C_ot_{1/2}$ of 12 and 86% having a $C_ot_{1/2}$ of 220. Assuming that E. coli B has a genome of $4x10^6$ bp, the estimated complexity of the G. lamblia genome is $8x10^7$ bp.

Estimation of G:C Content of G. lamblia Total DNA and rDNA.

Thermal denaturation of total G. lamblia DNA, as described in Methods, showed an increase in A_{260} beginning at 85°C and reaching a plateau at 92°C with a midpoint of 88.5°C (Fig. 4b). This corresponds to an overall G:C content of 46.8%. The control sample, comprised of E. coli DNA, melted from 88°C to 92°C with a midpoint of 90°C indicating a G:C content of 50.5%, in good agreement with the previously published estimates of 49.8-52.2% (17).

Using lower salt concentrations, purified pGRP1 gave three melting domains as against two for the vector pUC8 alone. The additional peak present in the pGRP1 profile represented about 60% of the total profile consistent with the size of the insert (5.4 kb) relative to the entire recombinant (8.2 kb). The T_m of this domain was 86°C which corresponds to 78.3% G:C. This, then, is the approximate G:C content of one complete rDNA repeat unit.

Quantitation of the rDNA repeat unit.

An estimation of the number of rDNA genes in the G. lamblia genome was made by co-electrophoresing known amounts of pGRP1 and genomic DNA of G. lamblia. Given a genomic complexity of about 8 x 10^7bp and the size of the cloned fragment (5400 bp), an estimate of the number of copies of the rDNA genes can be made by comparing the amount of plasmid DNA necessary to give a similar signal to a defined amount of genomic DNA. In Fig. 5, it can be seen that 2 µg of G. lamblia genomic DNA gives a signal corresponding to between 5 and 10 ng of the cloned fragment. (The

Figure 4. Analyses of G. lamblia DNA. a. Estimation of genomic complexity. Total genomic DNA from G. lamblia and E. coli were analyzed by renaturation kinetics (% single stranded vs. C_ot) as described in Methods. b. Estimation of G:C content. Total genomic DNA was analyzed by thermal denaturation in 1 X SSC and measurement of the hypochromism at A_{260}. The G:C content was estimated from the T_m, as described in methods. c. Estimation of G:C content of rDNA as cloned in pGRP1. Details as in (b) except that the control was the vector pUC8 and the buffer was 0.1 X SSC (see Methods).

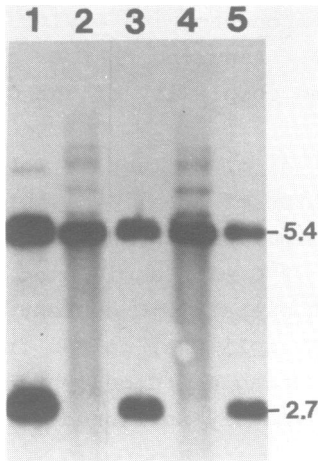


Figure 5. Copy Number of the rDNA Repeat. G. lamblia DNA (2 µg) was digested with PvuII (lanes 2 and 4) and electrophoresed in parallel with different amounts of the plasmid pGRP1 (which contains one complete rDNA repeat) cut with EcoRI and HindIII to release the insert. Lanes 1, 3 and 5 contained 20, 5 and 1 ng of pGRP1 DNA, respectively. Following electrophoresis, the DNA was transferred to nitrocellulose, hybridized to ^{32}P-labelled pGRP1 and autoradiographed.

band at 2.7 kb in the lanes containing plasmid DNA is the released vector portion of the recombinant.) Interpolation of the signals determined by densitometric scanning gave the more precise figure of 8.2 ng for the equivalent amount of cloned fragment (not shown). This predicts that there are about 63 copies of the rDNA gene in the genome of G. lamblia. The comigration of the genomic and cloned DNA confirms that the repeat cloned is a representative member of the rDNA genes of G. lamblia. The minor bands visible in the genomic lanes could be due to terminal fragments of repeat clusters, repeat units containing inserted sequences and/or single repeats in isolation of the major cluster(s) ("orphons"). They appear to be present at about one copy each.

Nucleotide sequencing of the rDNA gene.

Given the high conservation of rRNA throughout evolution, the quickest way to localize the particular coding regions in the rDNA repeat is to sequence random portions of the insert in

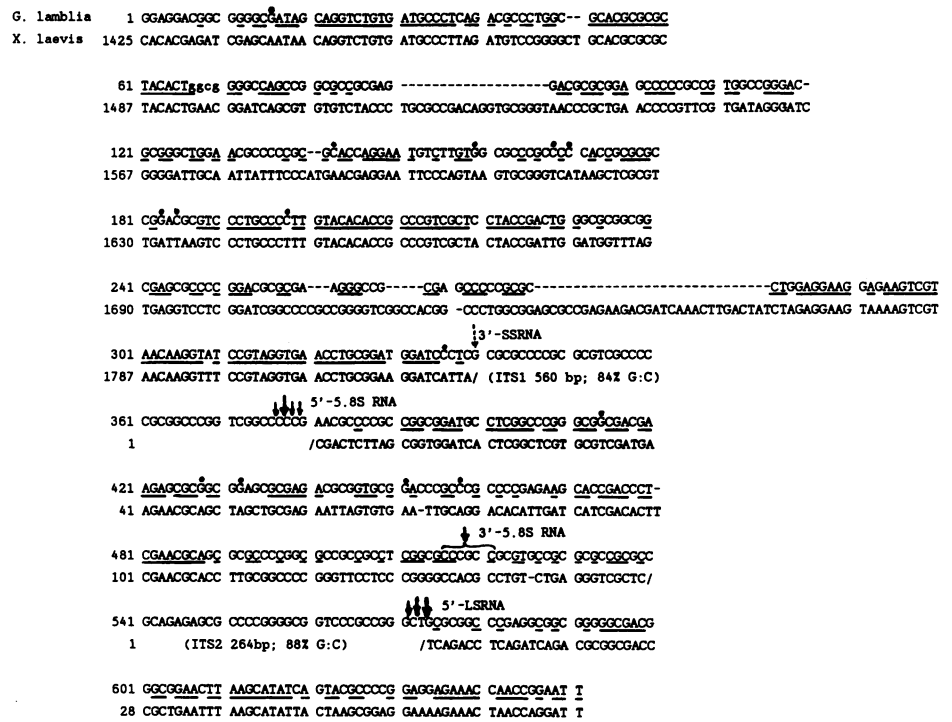


Figure 6. Nucleotide Sequence of G. lamblia rDNA. The plasmid pGRP1 was sequenced in the regions containing the 3'-end of SSRNA, 5.8S RNA and 5'-end of LSRNA using the strategy described in Fig. 3. This sequence is given on the top line of each pair and is numbered from an arbitrary site in the plasmid. Below this is given the corresponding sequence from Xenopus laevis (26-29), numbered from the 5'-end of each molecule. The sequences are aligned for maximal homology (underlined). Gaps (dashes) were introduced in some regions to optimize this homology. The known ends of the RNAs from Xenopus are indicated by a slash. The size and composition of the two ITS regions from Xenopus are also given. Lower case letters indicate ambiguity in the sequence due to extreme "band-compression" which could not be resolved even at extremely high temperatures. This is due to the very high G:C content of the G. lamblia rDNA gene. Ambiguities in the Xenopus sequence are as reported by Ware et al. (25). Dots above some positions indicate sites where a G or C is substituted for evolutionarily conserved A or T residues. Arrows mark predicted 5'-termini based on the results in Figs. 8 and 9.

pGRP1. This was done using the sites indicated in Fig. 3. By this approach, we quickly determined the approximate locations of the LSRNA, SSRNA and 5.8S RNA coding regions. The sequences ob-

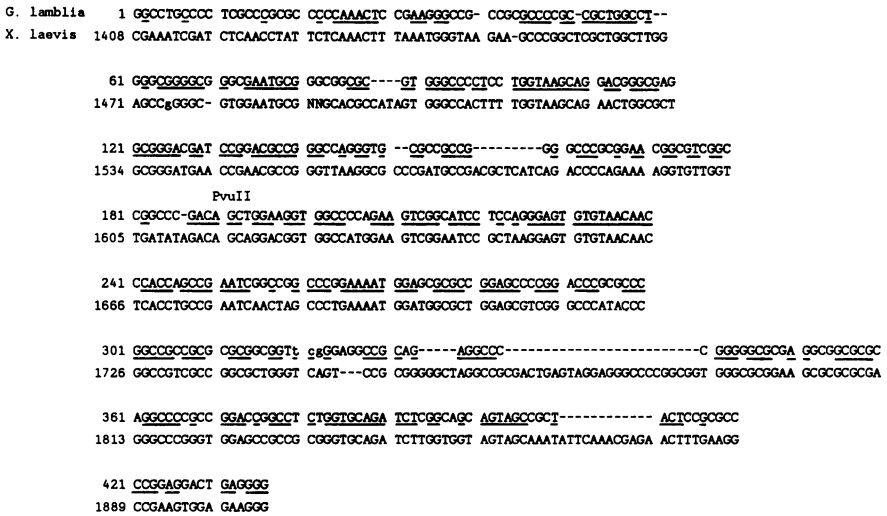


Figure 7. Demonstration of a Complete rDNA Repeat. The sequence of pGRP1 at the ends of the insert were determined and compared with the corresponding sequence from Xenopus. Details are as for Fig. 6. The PvuII site used to clone the rDNA repeat is boxed. Numbering of the G. lamblia sequence is arbitrary whereas that of Xenopus is relative to the 5'-end of LSRNA.

tained are presented in Fig. 6 and 7 along with the corresponding sequences from Xenopus laevis.

From these data, it first can be seen that pGRP1 does indeed contain a complete copy of the rDNA gene of G. lamblia (note the continuity of the LSRNA sequence through the PvuII site used to clone the repeat, Fig. 7). Second, the G. lamblia rRNA sequences presented are strikingly G:C rich (76% and 85% for the sequences in Fig. 6 and 7, respectively, compared with 53% and 62% for Xenopus in the same regions). These figures are consistent with the G:C content (78%) calculated from the T_m of the rDNA repeat, above.

S1 nuclease mapping of RNA termini.

Given the approximate sizes of the rRNAs (Fig. 1) and the sequence data presented in Fig. 6, the approximate termini of each RNA could be localized on the map. To confirm these, S1-nuclease protection studies were employed. The probes used are shown in Fig. 3 and the results presented in Fig. 8a and b (5'-

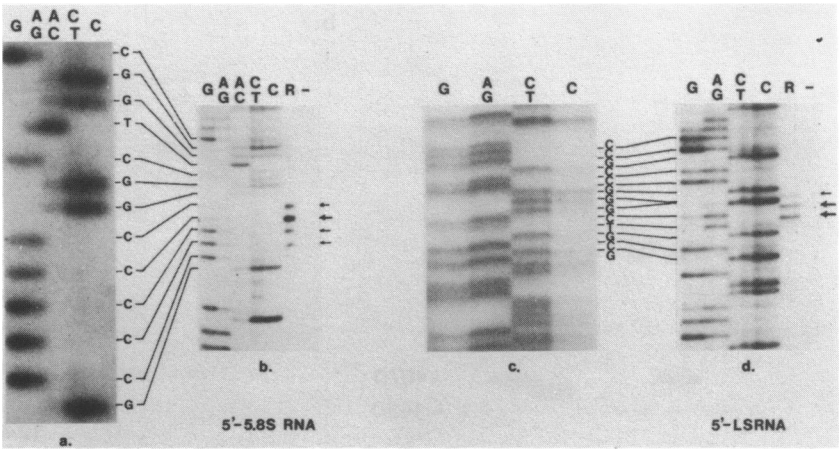


Figure 8. Precise determination of the 5'-end of 5.8S RNA (a. and b.) and LSRNA (c. and d.). The 5'-ends of these two RNA molecules were determined by S1 nuclease digestion using 2 µg of G. lamblia RNA as described in the Methods. The sequence ladders run in parallel were of the end-labelled probe (Fig. 3) used in the protection experiments and thus they can be used to directly read the position of the end of the RNAs. Sequence beside gels gives opposite strand to ease interpretation. a. Blow-up of sequence ladder where too faint to reproduce in part b. b. S1-nuclease protection showing 5'-end of 5.8S RNA. Beside the five sequencing reactions is the S1-nuclease treated hybrid (R). A control reaction where RNA was omitted from the reaction is also shown (-). c. Blow-up of sequence ladder from different gel showing three G residues where only two are evident in d. due to band compression. d. S1-nuclease protection showing 5'-end of LSRNA; details as in a.

end of 5.8S), 8c and d (5'-end of LS), 9a (5'-end of SS and 3'-end of LS) and 9b (3'-end of 5.8S). The precise 3'-end of SSRNA was not determined, but the very high homology of this region with other organisms strongly suggests that it will be at about the position indicated in Fig. 6.

Taken together, the S1 results indicate sizes of about 140, 1450 and 2650 nt for 5.8S, SS and LSRNA, respectively. These compare with about 134, 1500 and 2900 nt, respectively by gel mobility. Given the greater accuracy of the gel system and DNA standards used in the S1-nuclease protection experiments, we have used the results from these experiments in assigning the termini shown in Fig. 3. The termini indicated for the 5.8S region are shown in the sequence of Fig. 6.

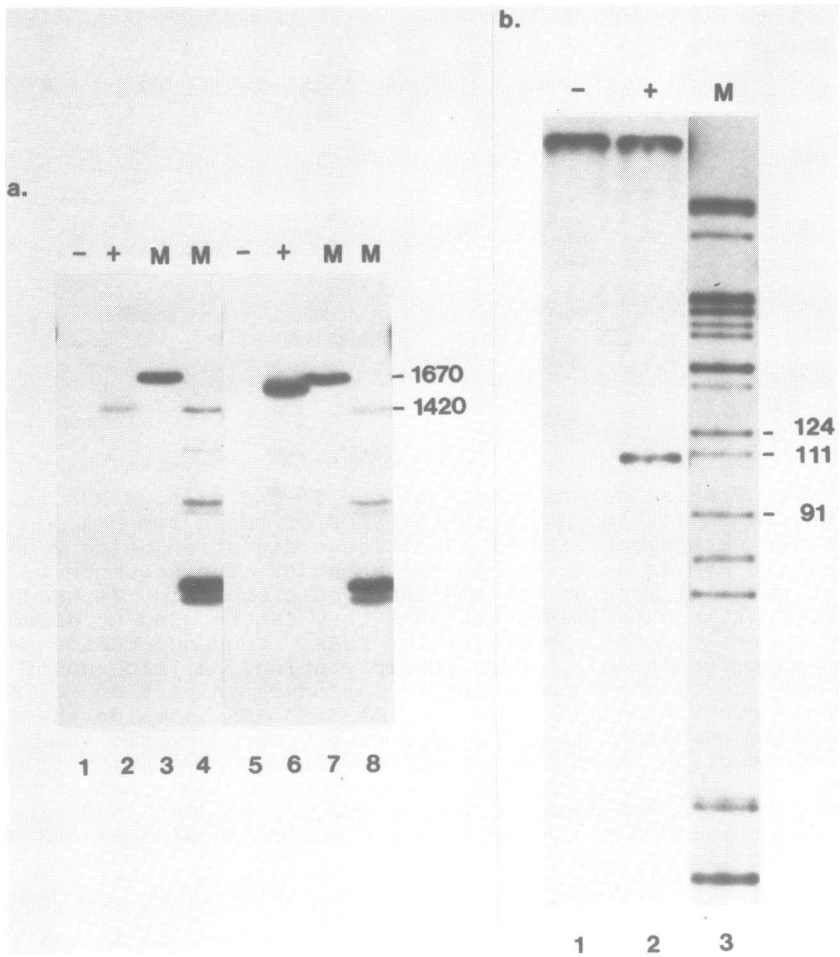


Figure 9. Approximate determination of the 5'-end of SSRNA, 3'-end of LSRNA and 3'-end of 5.8S RNA. These ends were determined by S1-nuclease mapping as described in the Methods. a. 5'-end of SSRNA and 3'-end of LSRNA. Lanes 1 and 5; control reactions lacking G. lamblia RNA. Lanes 2 and 6, S1-nuclease protections in the presence of 2 µg of G. lamblia RNA localizing the 5'-end of SSRNA and 3'-end of LSRNA, respectively. Markers are in lanes 3 and 7 (pUC8/HinfI fragment of 1670 nucleotides) and lanes 4 and 8 (pGB117.1/HinfI (30) fragment of 1420 nucleotides as calibrated against lambda DNA digests). b. 3'-end of 5.8S RNA. Lane 1, control reaction lacking G. lamblia RNA. Lane 2, S1-nuclease protection in the presence of 0.2 ug of G. lamblia RNA. Lane 3, marker lane containing ^{32}P-labelled pAT153 cut with MspI.

With the exception of the 3'-end of 5.8S RNA, these termini are as predicted by homology to the Xenopus rRNA sequences. Within the coding regions, however, there are several regions where "deletions" occur relative to the Xenopus sequence. Such "deletions" presumably account for the exceptionally small size of the G. lamblia rRNAs. The predicted small size of ITS1 (41nt) and ITS2 (33nt) further accounts for the small size of the rDNA repeat in this organism.

The 3'-end of the G. lamblia 5.8S RNA maps to position 518 of Fig. 6. This predicts that this RNA is only about 140 nucleotides as compared with at least 155 nucleotides in all other eukaryotes (18) and that the missing nucleotides are at the 3'-end of this RNA. As discussed below, this has been previously thought to be a critical portion of the 5.8S RNA molecule for its proper functioning.

DISCUSSION

We have shown here that the rRNA genes of G. lamblia are exceptional in four respects. These are discussed individually below.

First, the basic repeat unit of 5.4 kb is the smallest yet reported for any eukaryote. Although this might suggest some pressure on the overall genome size of G. lamblia, the two are not obviously connected as the latter figure is $8x10^7$ bp which is typical for protozoa (17).

Second, the sizes of SSRNA, LSRNA and 5.8S RNA are all substantially smaller than for any other eukaryote which has been studied. This is a surprising result and extends the trend for the protozoa to have highly varied sizes compared to the remainder of eukaryotes (trypanosomes and Euglena have much larger SSRNA than any other organism yet studied (19)). The full implications of this finding on the function of rRNA in the ribosome must await the determination of the complete sequence for both LSRNA and SSRNA.

Third, the rDNA repeat is extremely G:C rich. The pressure in this direction seems to be even stronger than the pressures which have conserved certain nucleotides in these genes

throughout eukaryotic evolution. Others have shown that such conservation applies to even the trypanosomes and Euglena which are the most ancient eukaryotes yet studied in this detail (19). It is noteworthy that the overall G:C content of the G. lamblia genome is not similarly skewed.

Fourth, about 15 nucleotides are missing from the 3'-end of 5.8S RNA relative to the structure of this molecule in other organisms. This region is believed to be one of the interacting sites 5.8S RNA has with the 5'-end of LSRNA (typically there are 15-20 base-pairs between nucleotides 137-155 of 5.8S RNA and 2-20 of LSRNA (18)).

Given that the 5.8S RNA is apparently missing this region, one might expect the 5'-end of LSRNA to be released from the sequence constraint it is otherwise under. This appears to be indeed the case as examination of Fig. 6 shows that the first 30 nucleotides of LSRNA are not conserved with the Xenopus sequence. This region is highly conserved in all other eukaryotes, so far studied, including trypanosomes (D.A.C., K. Kubo and J.C.B., unpublished results). Further work is necessary to confirm that the missing nucleotides in 5.8S are, in fact, at the 3'-end and whether this does, indeed, abolish a critical interaction between these two RNAs. Along these lines, it should be noted that in another protozoan, Vairimorpha necatrix, it has been recently reported that the "5.8S RNA" is fused to the LSRNA in a manner similar to that observed in prokaryotes (21). Trypanosomes, on the other hand, are exceptional in having the longest 5.8S RNA yet reported (170 nt; (22)). It is difficult to speculate on the significance of these exceptional structures but one recent finding may have some bearing. Wong and Clayton (23) have recently shown that 5.8S RNA plays some role in the replication of mammalian mitochondrial DNA. The fact that neither Giardia nor Vairimorpha has mitochondria may explain the apparent "relaxation" of selective pressure maintaining the 5.8S RNA structure in other eukaryotes.

It is clear from all of the above that the rRNA of G. lamblia is exceptional in many respects. Because the evolutionary position of Giardia has not yet been examined through the detailed analysis of any well conserved gene, it is difficult to

put the above results into context. However, it is clear that these findings will force us to rethink some of the conclusions regarding rRNA function which have been based on the conservation of structure previously observed for a limited number of samples.

ACKNOWLEDGEMENTS

We thank Dr. David Clayton for providing a preprint of the work on 5.8S RNA and mitochondrial DNA replication, Dr. Ted White for a preprint of the trypanosome small rRNA sequences and our other colleagues for many useful discussions. This work was supported by a grant from the MacArthur Foundation to J.C.B. and NIH grants to C.C.W. (AI19391) and D.A.C. (RR5354). Both J.C.B. and C.C.W. are Burroughs Wellcome Scholars in Molecular Parasitology.

REFERENCES

1. Mandal,R.K. (1984) Prog. Nuc. Acid Res. and Mol. Biol. 31, 115-160.
2. Brimacombe,R. and Stiege,W. (1985) Biochem. J. 229, 1-17.
3. Gerbi,S.A. (1985) in Molecular and Evolutionary Genetics., R. J. MacIntyre Ed., pp. 419-517. Plenum, New York.
4. Keister,D. (1983) Trans. R. Soc. Trop. Med. Hyg. 72, 431-432.
5. Hoeijmakers,J.H.J., Borst,P., Van den Burg,J., Weissman,C. and Cross,G.A.M. (1980) Gene 8, 391-417.
6. Wang,A.L. and Wang,C.C. (1985) J. Biol. Chem. 260, 3697-3702.
7. Blin,N. and Strafford,D.W. (1976) Nucl. Acids Res. 3, 2303-2308.
8. Bates,P.F. and Swift,R.A. (1983) Gene 26, 137-146.
9. Richardson,C.C. (1971) Prog. Nuc. Acids Res. 2, 815-828.
10. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) Molecular Cloning: A Laboratory Manual., Cold Spring Harbor Laboratory, Cold Spring Harbor.
11. Rigby,P.W.J., Dickman,M., Rhodes,C. and Berg,P. (1977) J. Mol. Biol. 113, 237-251.
12. Vieira,J. and Messing,J. (1982) Gene 19, 259-268.
13. Britten,R.J., Graham,D.E. and Neufeld,B.R. (1974) Meth. Enzymol. 29E, 363-418.
14. Mandel,M. and Marmur,J. (1968) Meth. Enzymol. 12B, 195-205.
15. Maxam,A.M. and Gilbert,W. (1980) Meth. Enzymol. 65, 499-560.
16. Sanger,F. and Coulson,A.R. (1978) FEBS letts. 87, 107-110.
17. Laskin,A.I. and Lechevalier,H.A. (1973) Handbook of Microbiology. pp. 625-626, CRC Press, Cleveland.
18. Vaughn,J.C. and Sperbeck,S.J. (1984) Nucl. Acids Res. 12, 7479-7502.
19. Sogin,M.L., Elwood,H.J. and Gunderson,J.H. (1986) Proc. Natl. Acad. Sci. U.S.A. 83, 1383-1387.
20. Subrahmanyam,C.S., Cassidy,B., Busch,H. and Rothblum,L.I. (1982) Nucl. Acids Res. 10, 3667-3680.
21. Vossbrinck,C.R. and Woese,C.R. (1986) Nature 320, 287-288.
22. Dorfman,D.M., Lenardo,M.J., Reddy,L.V., Van der Ploeg,L.H.T and Donelson,J.E. (1985) Nucl. Acids Res. 13, 3533-3549.

23. Wong,T.W. and Clayton,D.A. (1986) Cell 45, 817-825.
24. Schnare,M.N., Spencer,D.F. and Gray,M.W. (1983) Can. J. Biochem. 61, 38-45.
25. White,T.C,, Rudenko,G. and Borst,P. (1986) Nucl. Acids Res. 14, 9471-9489.
26. Ware,V.C., Tague,B.W., Clark,C.G., Gourse,R.L., Brand,R.C. and Gerbi,S.A. (1983) Nucl. Acids Res. 11, 7795-7817.
27. Salim,M. and Maden,B.E.H. (1981) Nature 291, 205-208.
28. Boseley,P.G., Tuyns, A. and Birnstiel, M.L. (1978) Nucl. Acids Res. 5, 1121-1137.
29. Hall,L.M.C. and Maden,B.E.H. (1980) Nucl. Acids Res. 8, 5993-6005.
30. Bernards,A., Van der Ploeg,L.H.T., Frasch,A.C.C., Borst,P., Boothroyd,J.C., Coleman,S.L. and Cross,G.A.M. (1981) Cell 27, 497-505.