Nucleic Acids Research, 2001, Vol. 29, No. 10 **2191–2198**

Analysis of intergenic spacer transcripts suggests 'read-around' transcription of the extrachromosomal circular rDNA in *Euglena gracilis*

Spencer J. Greenwood, Murray N. Schnare, James R. Cook[1] and Michael W. Gray*

Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada and [1]Department of Zoology, University of Maine, Orono, ME 04469, USA

Received as resubmission March 15, 2001; Accepted March 21, 2001

DDBJ/EMBL/GenBank accession no. X53361

ABSTRACT

We report here the sequence of the 1743 bp intergenic spacer (IGS) that separates the 3′-end of the large subunit ribosomal RNA (rRNA) gene from the 5′-end of the small subunit (SSU) rRNA gene in the circular, extrachromosomal ribosomal DNA (rDNA) of *Euglena gracilis*. The IGS contains a 277 nt stretch of sequence that is related to a sequence found in ITS 1, an internal transcribed spacer between the SSU and 5.8S rRNA genes. Primer extension analysis of IGS transcripts identified three abundant reverse transcriptase stops that may be analogous to the transcription initiation site (TIS) and two processing sites (A′ and A0) that are found in this region in other eukaryotes. Features that could influence processing at these sites include an imperfect palindrome near site A0 and a sequence near site A′ that could potentially base pair with U3 small nucleolar RNA. Our identification of the TIS (verified by mung bean nuclease analysis) is considered tentative because we also detected low-abundance transcripts upstream of this site throughout the entire IGS. This result suggests the possibility of 'read-around' transcription, i.e. transcription that proceeds multiple times around the rDNA circle without termination.

INTRODUCTION

In typical multicellular eukaryotes, ribosomal RNA (rRNA) genes are present in many copies organized as tandemly arrayed head-to-tail repeating units that are integrated into the chromosome (1,2). In the context of this paper, we define the intergenic spacer (IGS) as the sequence that separates the large subunit (LSU = 5.8S plus 25–28S) rRNA gene at the 3′-end of one repeat from the adjacent downstream small subunit (SSU = 18–20S) rRNA gene. Each repeat is transcribed by RNA polymerase I (RNAP-I) to yield a long pre-rRNA that contains internal transcribed spacer (ITS) as well as external transcribed spacer (ETS) sequences (the latter derived from the ETS). Genes encoding 5S rRNA are transcribed by rRNA polymerase III and in most cases are not physically linked to genes for the other rRNAs.

Protozoa (unicellular eukaryotes) provide examples of 'typical' chromosomally integrated rRNA genes organized as tandem arrays. However, many protozoa exhibit rRNA gene organizations that deviate from this pattern. *Plasmodium* species have only a few copies of the rDNA transcriptional unit and these are dispersed in the genome (3). In *Paramecium tetraurelia*, rDNA units are tandemly repeated but are found in linear and circular extrachromosomal molecules (4). In hypotrichous ciliates, the macronuclear DNA is present in discrete, gene-size fragments; thus, the rDNA units in these organisms are found as linear monomers (5,6). Extrachromosomal rDNA dimers appear in the form of linear palindromes in *Dictyostelium discoideum* (7), *Physarum polycephalum* (8,9) and *Tetrahymena pyriformis* (10,11) and as a circular palindrome in *Entamoeba histolytica* (12,13). Finally, rDNA transcriptional units are found as circular monomers in *Naegleria gruberi* and related schizopyrenid amoebae (14,15) and in *Euglena gracilis* (16,17).

Among the organisms that contain small circular rDNA molecules, it has been demonstrated for *Naegleria* (14) and *Entamoeba* (18) that there are no integrated chromosomal copies and that there are few, if any, integrated copies in *Euglena* (19). The extrachromosomal nature of rRNA genes in these organisms suggests that the rDNA circles must be able to replicate autonomously in order to maintain copy number within the cell. Intermediates in the replication of rDNA circles have been detected in *Entamoeba* (12,20) and *Euglena* (19).

Because of the diversity in rRNA gene organization evident among protozoa, the processes that regulate rRNA gene expression probably exhibit novel features in these organisms. We are interested in the expression of the extrachromosomal circular rDNA in *Euglena*. Estimates of the number of rDNA circles in *Euglena* range from 800 to 4000 copies per cell

*To whom correspondence should be addressed. Tel: +1 902 494 2521; Fax: +1 902 494 1355; Email: m.w.gray@dal.ca
Present address:
Spencer J. Greenwood, Department of Biomedical Sciences, Ontario Veterinary College, University of Guelph, Guelph, Ontario N1G 2W1, Canada.
James R. Cook, deceased.

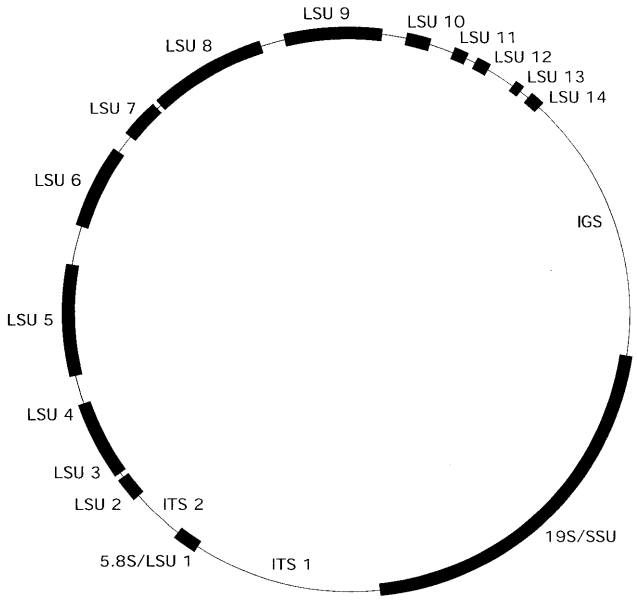

Figure 1. The 11 056 bp circular extrachromosomal rDNA of *Euglena*. Positions of genes for mature rRNA components are indicated by solid arcs. The IGS that separates the LSU 14 gene from the SSU rRNA gene is represented by a thin line. All other thin lines represent ITS sequences. The LSU 2–14 genes encode the fragmented 28S rRNA. The 10–11 nt ITS that separates the genes for LSU 3 (350–351 nt) and LSU 4 (116 nt) is not shown.

(16,17,21–23), with the actual amount depending upon growth phase (16,17) and culture conditions (24). The RNAP-I responsible for transcription of the *Euglena* rDNA circle has been purified to homogeneity (25). A ~10.2 kb pre-rRNA has been shown to undergo multiple processing reactions (26,27) to yield a 2.3 kb SSU rRNA (28) and a highly fragmented LSU rRNA that consists of 5.8S rRNA and 13 additional small rRNAs (29,30) (Fig. 1). Some of these small rRNAs have been shown to interact by base pairing to generate secondary structural elements that resemble those of their covalently continuous homologs in other eukaryotes (31). 5S rRNA genes are not found on the rDNA circle but instead are present in a tandemly repeated stretch of DNA that also encodes spliced-leader RNA (32).

Pre-rRNA processing in eukaryotes is complex and is facilitated by many protein factors and small nucleolar RNAs (snoRNAs) (33–35). A homolog of U3 snoRNA, the most abundant snoRNA known to be involved in rRNA processing in other systems, has been identified in *Euglena* (36). In this paper, we report an investigation of the structure of the IGS region of *Euglena* rDNA and an analysis of RNA transcripts that map to this region. We conclude that the entire rDNA circle is transcribed, with no indication of efficient termination of transcription within the IGS.

MATERIALS AND METHODS

Sequence analysis

Total RNA was isolated from mid-log phase cultures of the UCLA variety of *E.gracilis* strain Z as previously described (29). SSU rRNA was gel purified from total RNA (29) and its 5′-terminal sequence determined by the enzymatic method (37,38). Clones and subclones of the circular rDNA (16,27,30) were sequenced as described (30).

Sequence comparisons were performed using the Genetic Data Environment (39), the Basic Local Alignment Search Tool (40), the MicroGenie Sequence Analysis Program (41), DNASIS v.2.5 (MiraiBio Inc., Alameda, CA) and the Nip4 program of the Staden package (42).

Transcript mapping

Prior to mung bean nuclease mapping and primer extension analysis, aliquots of RNA were treated with DNase I (43). Reverse transcriptase (RT) sequencing and primer extension analysis of RNA were performed using 5′-end-labeled IGS-specific oligonucleotides according to the protocol of Geliebter (44) but without actinomycin D. Alternatively, 5′-end-labeled RT products were gel purified and subjected to chemical sequence analysis (45).

Mung bean nuclease protection assays followed a protocol for S1 nuclease mapping with reaction conditions adjusted for mung bean nuclease (43). The probe was an *Xho*I–*Sst*I restriction fragment derived from subclone pPvSs-467 (27). Hybridization of the double-stranded DNA probe (5′-labeled at its *Sst*I end) to 20 µg RNA was performed in the presence of 80% formamide (46). Hybrid-protected mung bean nuclease digestion products were analyzed using 6% polyacrylamide, 7.0 M urea sequencing gels with a sequencing ladder generated from subclone pPvSs-467 using the Sequenase Version 2 Kit (United States Biochemical). The standard reaction products from the dideoxy sequencing run (labeled by incorporation of [α-^{32}P]dATP) were digested with *Sst*I (the same restriction endonuclease used to produce the ^{32}P-labeled end of the probe) (47). This allowed direct alignment of the mung bean nuclease digestion products with the sequencing ladder.

RESULTS

DNA sequence of the IGS region of *Euglena* rDNA

The IGS sequence (Fig. 2) was determined by the chemical method from both strands with overlapping data for each restriction site. Portions of the sequence were verified by the dideoxy-mediated chain termination method. During the course of this study and our previous determination of the LSU rRNA gene sequence (30) we also isolated and sequenced the ends of many restriction fragments that covered large portions of the SSU rRNA gene. Our data are in agreement with a previously published sequence (48), with one exception: we find that there are two G residues where the published sequence has a single G at SSU rRNA position 1126. Data obtained in the present study also overlap the ends of previously determined sequences (30,48) and complete the sequence of the entire circle. The actual length of the rDNA circle (11 056 bp) is in close agreement with estimates (16) based upon restriction mapping (11.15 kb) and contour length (11.3 kb, with a standard deviation of 0.6 kb). Similarly, the base composition (57.75% G + C) agrees well with estimates (16) based upon buoyant density (58%) and melting temperature (59%). Similar estimates have been published by other groups (17,23).

The *Euglena* IGS begins immediately after the gene for LSU rRNA species 14 (Fig. 1), which has been precisely localized

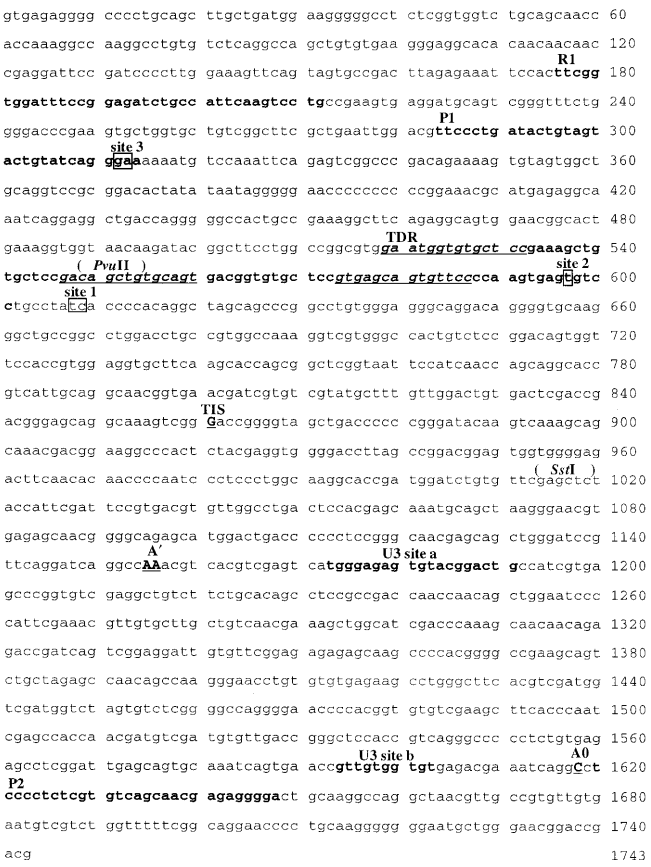

Figure 2. Primary sequence of the IGS region of *Euglena* rDNA. Positions of restriction enzyme recognition sequences that delineate subclone pPvSs-467 are indicated above the sequence (*Pvu*II and *Sst*I). A repeated sequence (R1), two imperfect palindromes (P1 and P2) and two stretches of sequence that are complementary to U3 snoRNA (U3 sites a and b) are highlighted in bold. Six copies of an imperfect TDR are presented in bold with every second copy in italics and underlined. Three abundant RT stop sites (TIS, A′ and A0) are uppercase, bold and underlined, while three minor RT stop sites (sites 1–3) are boxed.

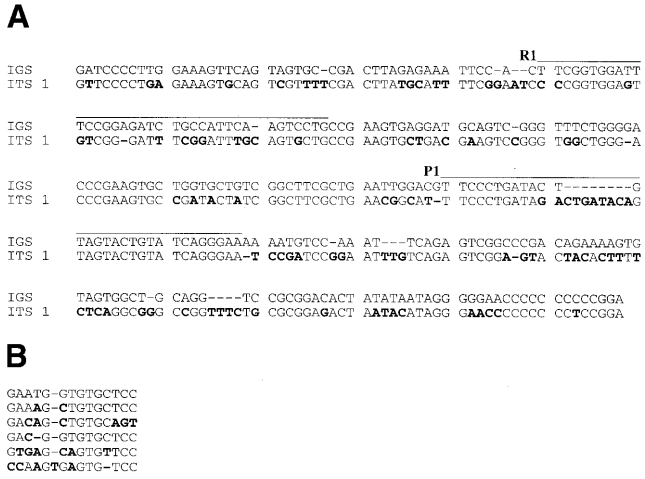

Figure 3. Alignment of repeated sequences. Residues that differ from the top sequence are in bold. (**A**) IGS positions 131–405 aligned against ITS 1 positions 468–760. The R1 and P1 sequences are overlined (Fig. 2). (**B**) Alignment of the TDRs (IGS positions 519–601).

by direct sequencing of the RNA transcript (29). The other end of the IGS is adjacent to the 5′-end of the SSU rRNA gene, which had not been mapped precisely. We therefore isolated the SSU rRNA and determined its 5′-terminal sequence (data not shown), which confirmed the 5′-terminus assigned by Sogin *et al.* (48) through comparison of the *Euglena* sequence with known 5′-terminal sequences from other eukaryotes. The IGS sequence is 1743 bp long and contains six copies of a 14 nt imperfect tandem direct repeat (TDR) located at positions 519–601 (Figs 2 and 3). We also identified two imperfect palindromes located at positions 284–314 (P1) and at positions 1620–1648 (P2). Surprisingly, we found that the P1 palindrome is specifically related to a sequence that we had previously identified in ITS 1, the spacer that separates the SSU and 5.8S rRNA genes [sequence P in (30)]. Further analysis revealed that a 277 nt stretch of the IGS (positions 131–405, containing the P1 sequence) is homologous to a stretch of ITS 1 (468–760 nt after the 3′-end of the SSU rRNA gene) that also includes one of two copies of a 37 nt repeat (R1 in Figs 2 and 3) (30).

Three abundant RT stop sites map to the 3′-half of the IGS

Using RT with total RNA and an IGS-specific primer that maps just upstream of the SSU rRNA gene, primer extension analysis identified three strong RT stop sites (Fig. 4A). This result is reminiscent of what we (45) and others (49) have found in trypanosomatid protozoa. Considering the close evolutionary relationship shared by trypanosomatids and *Euglena* (50) we inferred that these RT stops are most likely to represent the transcription initiation site (TIS), which marks the 5′-end of the 5′ ETS, and two processing sites (A′ and A0) as defined in the trypanosomatid studies (45,49). These three major RT stop sites were mapped to IGS positions 861 (TIS), 1155/1156 (A′) and 1618 (A0) (Figs 2 and 4A–C). Note that the A0 site is located in close proximity to the 5′-end of the P2 palindrome (Fig. 2). In the course of mapping these major RT stop sites, we also detected several less abundant RT stops within the 5′ ETS; the most prominent of these minor stops mapped to IGS position 1134 (data not shown).

To verify that the RT stop labeled TIS in Figure 4 represented a true 5′ terminus, we performed a mung bean nuclease protection experiment. This experiment suggested the presence of a heterogeneous TIS that is centered around IGS position 861 (Fig. 5A). However, these nuclease protection results should be interpreted with caution because apparent TIS heterogeneity was not as prominent in the RT mapping experiments (Fig. 4C and data not shown).

Transcription proceeds through the entire IGS

Mung bean nuclease protection experiments suggested the presence of additional low-abundance transcripts whose 5′-termini mapped upstream of the TIS (Figs 2 and 5B, sites 1 and 2). In the analysis depicted in Figure 5B, a small amount of full-length probe was protected from nuclease digestion because of re-annealing with the unlabeled DNA strand. However, in experiments where total RNA was present, the major protected band was a few nucleotides shorter than the full-length probe, mapping to the insert/vector boundary. This

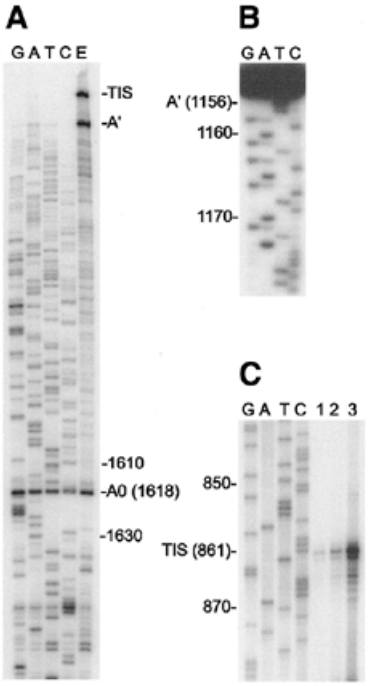

Figure 4. Mapping of three abundant RT stop sites in IGS transcripts present in total RNA. Numbers indicate positions in the IGS sequence. (**A**) Localization of site A0 by RT sequencing using a 5′-end-labeled primer complementary to IGS positions 1691–1717. Lane E represents primer extension in the absence of dideoxy-NTPs. Two additional RT stop sites were also detected (TIS and A′). (**B**) A 5′-end-labeled primer (complementary to IGS positions 1235–1259) was used to generate an RT product that ended at site A′. The sequence of this product (two bands differing in length by 1 nt) was determined by the chemical method. The band in lane T labeled A′ (1156) was generated by chemical removal of a T residue from the end of the RT product. The band that is generated by removal of a residue from the end of the '1155' RT product is obscured by the full-length '1156' RT product. (**C**) The location of the RT stop site labeled TIS in (A) was determined by electrophoresing the products of RT reactions adjacent to a DNA sequencing ladder (subclone pPvSs-476) that was generated using the same primer (complementary to IGS positions 910–940). Different amounts of RNA template were used; 5, 10 or 20 µg (lanes 1–3, respectively).

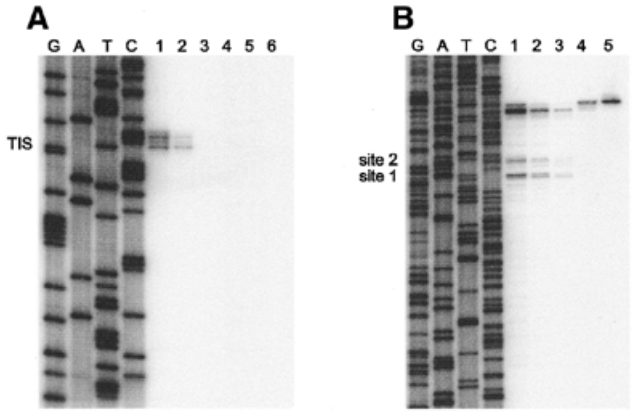

Figure 5. Mung bean nuclease protection experiments using a 5′-end-labeled probe. (**A**) The probe was hybridized with total *Euglena* RNA (lanes 1–3), *Escherichia coli* tRNA (lane 4) or no RNA (lane 5). Hybrids were digested with mung bean nuclease (200 U, lanes 1, 4 and 5; 300 U, lane 2; 500 U, lane 3). Lane 6 contained undigested probe that migrated above the region of the gel shown. (**B**) Lanes 1–4 contained hybrids as described above. These hybrids were digested with mung bean nuclease (100 U, lanes 1 and 4; 200 U, lane 2; 300 U, lane 3). Lane 5 contained undigested probe.

result indicated that transcripts were present that spanned the entire length of the cloned insert, which contained the putative TIS (Fig. 2). As in Figure 5A, each of the protected fragments displayed length heterogeneity.

RT sequencing experiments were performed to verify the presence of IGS-derived transcripts that extended upstream of the putative TIS. The results of these experiments (Fig. 6) confirmed the location of the TIS as well as sites 1 and 2; however, no heterogeneity was observed for site 2. These experiments also yielded RNA sequence data that extended from the 5′ ETS into the 3′-end of the LSU 14 coding region. In order to ensure that the observed primer extension products were produced from the RNA template rather than from contaminating rDNA, samples of total RNA were treated with RNase A prior to RT sequencing. In these control experiments no primer extension products were observed (data not shown); in contrast, when RNA samples were pre-treated with DNase I, primer extension still occurred (Fig. 6). This result confirmed that the low-abundance primer extension products were being generated from RNA templates while the sequencing ladders show that these templates were IGS-specific. In each of these experiments, primer extension products were detected (Fig. 6A–D, lanes E) that migrated very slowly near the top of the gel (data not shown), suggesting that transcripts are present that include 5′ ETS sequence and extend into the upstream LSU rRNA sequence. The data presented here, combined with northern hybridization data (27), establish that the entire *Euglena* rDNA circle is transcribed, including the IGS.

The RT sequencing experiments also detected putative processing sites at the 3′-end of the LSU 14 sequence (Fig. 6A) and in the 3′ ETS at IGS positions 312/313 (Fig. 6B, site 3). Because site 3 is located at the 3′-end of the P1 palindrome (Fig. 2), it could be argued that this RT stop was caused by stable secondary structure blocking the progress of the enzyme. However, this interpretation seems unlikely in view of the observation that RT was not impeded to the same extent at the 3′-end of the P2 palindrome (Figs 2 and 4A), which has the potential to form a hairpin similar to the putative P1 hairpin. Processing at site 3 may be mediated by RNase III, which is known to cleave at a hairpin downstream of the LSU rRNA sequence in yeast pre-rRNA (51,52).

DISCUSSION

Sequence features of the IGS

The 1743 bp sequence reported here for the IGS region of *Euglena* rDNA completes the sequence of the 11 056 bp extrachromosomal rDNA circle. The IGS contains two imperfect palindromes (~30 bp), which are expected to form hairpin structures (putative processing signals as judged by RT mapping) in pre-rRNA transcripts. A 277 bp sequence that contains one of these palindromes is related to a stretch of sequence in ITS 1. This finding could partially explain the unusually large size of the *Euglena* ITS 1 (30) and raises the possibility of pre-rRNA processing at the ITS 1 palindrome.

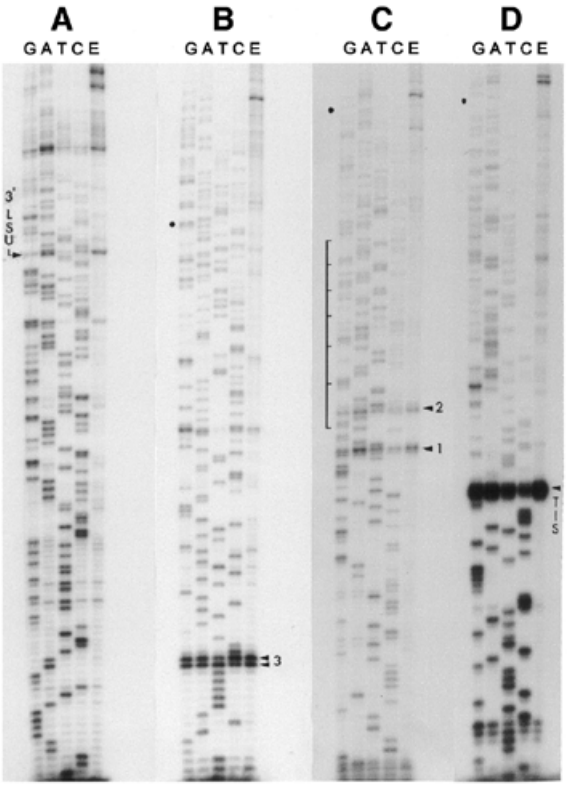

Figure 6. Primer extension and RT sequencing within the *Euglena* IGS. Primer extension analysis employed 5′-end-labeled IGS-specific primers and total RNA that had been pre-treated with DNase I. Lanes labeled G, A, T and C indicate the dideoxynucleotide incorporated during RT sequencing. Lanes E represent primer extension in the absence of dideoxy-NTPs. The dot in (B), (C) and (D) indicates the approximate location of the sequence complementary to the primer used in the previous panel. Note that lanes E in (B), (C) and (D) also contain the RT stops that are precisely mapped in the preceding panel. (**A**) The primer was complementary to IGS positions 145–175. The RT stop corresponding to the 3′-end of the LSU rRNA sequence is indicated. (**B**) The primer was complementary to IGS positions 329–358. The two adjacent RT stops at site 3 are indicated (Fig. 2). (**C**) The primer was complementary to IGS positions 669–698. The two arrowheads indicate the positions of putative rRNA processing events (Figs 2 and 5B, sites 1 and 2). The locations of the six imperfect repeats are marked (Figs 2 and 3). (**D**) The primer was complementary to IGS positions 910–940. The arrowhead marks the putative TIS at the 5′-terminus of an abundant transcript (Figs 2, 4C and 5A).

The IGS also contains six copies of a 14 bp imperfect tandem direct repeat. Length heterogeneity is often associated with repeated sequences in the rDNA of other eukaryotes (13,45); however, length heterogeneity of *Euglena* rDNA has not been detected by restriction enzyme analysis (16,17,23,53). Furthermore, RT sequencing of *Euglena* IGS transcripts clearly demonstrates that transcribed copies of the rDNA do not display length heterogeneity in the region that contains the tandem 14mers (Fig. 6). These repeats could potentially be involved in (i) enhancement of transcription (54–56), (ii) pre-rRNA processing at sites 1 and 2 (located in the same region of the IGS), or (iii) rDNA replication [(note that a replication origin has been identified but not localized on the circle (19)].

TIS or processing site?

Primer extension analysis indicated the presence of three abundant RT stop sites upstream of the SSU rRNA sequence when total *Euglena* RNA was used as a template. By analogy with similar results in trypanosomatid species (45,49), we infer that the 5′-most of these RT stops most likely represents the TIS. In northern hybridization experiments a probe that spans this site (clone pPvSs-467, Fig. 2) detected pre-rRNA transcripts that extended 0.9 kb upstream of the SSU rRNA sequence (27). Probes located further upstream in the IGS did not detect transcripts in northern analysis (27). Our localization of the TIS at position 861 was verified by nuclease protection experiments and is in agreement with the northern results yielding a 5′ ETS length of 883 nt. However, our identification of the TIS must be considered tentative because we did detect low amounts of 5′-termini that map further upstream in the IGS (Figs 2 and 6, sites 1–3) and we cannot rule out the possibility of transcription initiation at one or more of those sites followed by rapid processing between IGS positions 860 and 861.

Our initial attempts to verify the position of the TIS by guanylyltransferase capping experiments have been unsuccessful. Although such negative results should be interpreted with caution, lack of a cappable transcript could indicate that (i) position 861 is the TIS but the transcript is not a substrate for guanylyltransferase (possibly due to pyrophosphatase activity), (ii) the 5′-terminus of the primary transcript is removed by rapid processing/degradation, or (iii) many copies of the rRNA transcript may be produced from a single initiation event ('read-around' transcription), resulting in very low levels of cappable transcript. Thus, a final definition of the TIS may have to await the development of an *in vitro* transcription system for *Euglena* rDNA.

Conservation of 5′ ETS processing sites

Northern hybridization analysis indicated that *Euglena* SSU rRNA is generated from the pre-rRNA through several alternative pathways, resulting in a mixture of ~3.2 kb intermediates that contain different 5′- and 3′-ends (27). The 5′-termini that we have mapped in this study (the TIS, sites A′ and A0, and the 5′-end of mature SSU rRNA) most likely correspond to the 5′-ends of the various ~3.2 kb processing intermediates.

It has been proposed that a pre-rRNA processing site located near the beginning of the 5′ ETS may be universally conserved among eukaryotes (57–59). This site is thought to be functionally equivalent to the mammalian primary processing site (A′) although cleavage at this site is not necessarily the first processing event in non-mammalian systems. In *Euglena*, the first major RT stop site found downstream of the putative TIS is a strong candidate for this conserved processing site and has been designated A′ in this study.

A second major 5′ ETS processing site (A0) has been identified in fungi (57,60,61) and trypanosomatids (45,49). In each of these cases the cleavage site is located within a few hundred nucleotides of the SSU rRNA sequence on the 5′ side of a structural element that places the A0 site in close proximity to the beginning of the SSU rRNA sequence. We have detected a major RT stop site in the *Euglena* 5′ ETS that corresponds to a possible A0 processing site. In the *Euglena* case the A0 site is also found at the 5′-side of a structural element (the P2 palindrome/hairpin, Fig. 2) but in this case the A0 site and the 5′-end of the SSU rRNA sequence are not juxtaposed. It now seems likely that an A0 processing site is also present in animals, considering that a recently discovered SSU rRNA precursor in *Xenopus laevis* has a 5′-terminal extension of

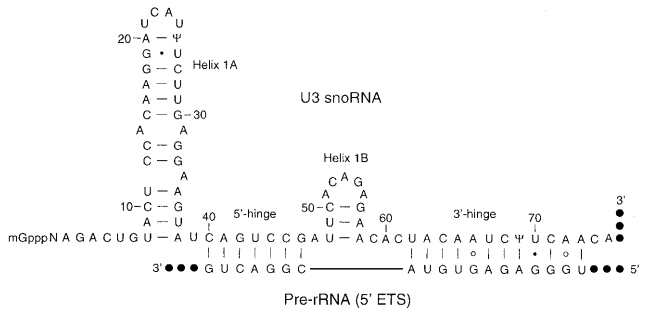

Figure 7. Potential interactions between U3 snoRNA and the pre-rRNA 5′ ETS in *Euglena*. U3 helix 1A is separated by the hinge region (positions 39–75) from the 3′-domain. The 5′- and 3′-hinge are drawn paired to adjacent regions of the 5′ ETS (IGS positions 1173–1191; Fig. 2, U3 site a). The remaining, central portion of the hinge region contains a possible helix 1B. Note that an extended helix 1A and two alternative helix 1B structures have been proposed for the free U3 snoRNA (36,72); however, those structures are incompatible with the newly proposed U3 snoRNA:5′ ETS interactions. A secondary structure model for the 3′-domain (not shown) is available (36).

~200 nt (62). This would place the *Xenopus* cleavage site on the 5′ side of an extended hairpin structure (63,64), as expected for a homolog of site A0.

As summarized above, a pattern is emerging that suggests conservation of two major 5′ ETS processing sites (A′ and A0) among a broad range of eukaryotes. This does not necessarily mean that these are the only 5′ ETS processing sites present in a particular system. Indeed, as discussed above, we cannot completely rule out the possibility that the putative *Euglena* TIS is actually a processing site (another A′ candidate). Furthermore, a processing site has been identified 105 nt upstream of the SSU rRNA sequence in the mouse 5′ ETS (65,66); however, this site is not a convincing A0 candidate because it does not occur in the expected structural context (45). Finally, additional RT stops have been mapped to the 5′ ETS in *Schizosaccharomyces pombe* (67) and *Euglena* (see Results). At the moment, it is difficult to evaluate whether these additional RT stops represent (i) artifacts, (ii) real processing sites necessary for rRNA maturation, or (iii) intermediates in degradation of the spacer after its release from the long pre-rRNA [the 5′ ETS is rapidly removed from the *S.pombe* pre-rRNA (67) and the free 0.9 kb 5′ ETS has been detected in *Euglena* total RNA (27)].

Potential interactions between U3 snoRNA and the 5′ ETS in *Euglena*

It is well documented that U3 snoRNA plays an essential role in pre-rRNA 5′ ETS processing in other eukaryotes and that this role is mediated by base pairing between pre-rRNA and U3 snoRNA (34,35). By analogy, it is reasonable to propose that *Euglena* U3 snoRNA is likely to be involved in processing at the A′ and A0 sites described above. We previously noted that U3 snoRNA from *Euglena* contains a sequence (positions 58–67, Fig. 7) that has the potential to base pair with a sequence located in the 5′ ETS just upstream of site A0 (Fig. 2, U3 site b) (36). More recently, we (45) and others (62) noted that two sequences in the hinge region of U3 snoRNA have the potential to base pair with the 5′ ETS in several eukaryotes. As a test of this proposal we searched for sequences within the *Euglena* 5′ ETS that could pair with the appropriate regions of *Euglena* U3 snoRNA. The results of this analysis (Fig. 7) reinforce the idea that 5′ ETS sequences are recognized by the two ends of the U3 snoRNA hinge region (45,62) and that these interactions may be conserved across a broad phylogenetic range (62). Interestingly, the two adjacent ETS sequences involved in the proposed interactions (Fig. 7) are located just downstream of site A′ (Fig. 2, U3 site a).

'Read-around' transcription of *Euglena* rDNA

In eukaryotes that have tandemly repeated rDNA units, transcription begins within one IGS and ends at a specific termination signal within the next downstream IGS. Thus, although the tandemly repeated rDNA units are physically linked, they are transcribed as independent units; this is facilitated, at least in part, by the presence of termination signals downstream of the LSU rRNA gene (upstream of the next promoter). These termination signals and associated proteins play several important roles during rRNA biosynthesis and rDNA replication (54–56). Of particular interest to this discussion, terminators protect against 'promoter occlusion', which occurs when RNAP-I that has initiated transcription in an upstream repeat moves through a downstream promoter and disrupts the semistable pre-initiation complex (68–70). Because rDNA transcriptional units are already physically separate in *Euglena*, in the form of circular monomers, terminators may not be necessary in this system.

In other systems, as a consequence of RNAP-I termination, there is a non-transcribed region of IGS that maps immediately upstream of the TIS. For this reason, significant levels of transcript that span the TIS are not produced. Accordingly, it is reasonable to conclude that termination of RNAP-I transcription does not occur at a specific site in *Euglena* rDNA. Although we have not definitively identified the *Euglena* TIS, our data demonstrate that there is not a single non-transcribed nucleotide in the rDNA circle, every position being overlapped by readily detectable transcripts. Furthermore, because probes derived from the 5′ half of the IGS did not detect bands in northern hybridization experiments (27), we know that transcripts encompassing that portion of the IGS are not present in stable, discrete-sized pre-rRNAs. Therefore, these IGS transcripts must be heterogeneous in length, having 3′-termini produced by processing at multiple sites or through non-specific termination. It should be noted that the techniques employed in this study were designed to detect RNA 5′-termini; thus, we cannot rule out the possibility that a portion of the transcripts do terminate within the IGS, generating 3′-termini that would have gone undetected. However, if such termination does occur, our data clearly show that it cannot be very efficient.

Our results suggest that 'read-around' transcription occurs in *Euglena*, i.e. that RNAP-I transcription may continue multiple times around the rDNA circle without termination and re-initation. In this scenario, sequence-specific transcription factors would bind to the promoter and form a stable initiation complex with RNAP-I. Multiple initiation events could occur before the first RNAP-I completed transcription of the entire circular template. Instead of termination and release, the RNAP-I would then move through the promoter region, displacing any assembled transcription initiation factors, and continue to transcribe multiple times around the rDNA circle. In this case,

promoter occlusion would be beneficial because dislodged initiation factors would then be available to form initiation complexes on other copies of the extrachromosomal circle. Rapid and extensive pre-rRNA processing would produce the mature SSU rRNA and the 14 pieces of LSU rRNA before any significant amount of multimeric precursor could accumulate.

Evidence for 'read-around' transcription is available from the mouse (*in vitro*) and *Xenopus* (*in vivo*) systems, in studies employing circular plasmid constructs that contained RNAP-I promoters but lacked terminators (69–71). The *Euglena* system provides the first example where this mode of rDNA transcription is likely to be operating on natural templates.

ACKNOWLEDGEMENTS

We thank Dr. D.F.Spencer for assistance with Figure 1. This work was supported by an operating grant (MT11212) from the Medical Research Council of Canada (M.W.G.), an operating grant (DCB-8408588) from the National Science Foundation USA (J.R.C.), a Walter C. Sumner Memorial Fellowship (S.J.G.), and a Fellowship (M.W.G.) from the Canadian Institute for Advanced Research (Program in Evolutionary Biology).

REFERENCES

1. Long,E.O. and Dawid,I.B. (1980) Repeated genes in eukaryotes. *Annu. Rev. Biochem.*, **49**, 727–764.
2. Mandal,R.K. (1984) The organization and transcription of eukaryotic ribosomal RNA genes. *Prog. Nucleic Acid Res. Mol. Biol.*, **31**, 115–160.
3. McCutchan,T.F. (1986) The ribosomal genes of *Plasmodium*. *Int. Rev. Cytol.*, **99**, 295–309.
4. Findly,R.C. and Gall,J.G. (1978) Free ribosomal RNA genes in *Paramecium* are tandemly repeated. *Proc. Natl Acad. Sci. USA*, **75**, 3312–3316.
5. Lipps,H.J. and Steinbrück,G. (1978) Free genes for rRNAs in the macronuclear genome of the ciliate *Stylonychia mytilis*. *Chromosoma*, **69**, 21–26.
6. Swanton,M.T., McCarroll,R.M. and Spear,B.B. (1982) The organization of macronuclear rDNA molecules of four hypotrichous ciliated protozoans. *Chromosoma*, **85**, 1–9.
7. Cockburn,A.F., Taylor,W.C. and Firtel,R.A. (1978) *Dictyostelium* rDNA consists of non-chromosomal palindromic dimers containing 5S and 36S coding regions. *Chromosoma*, **70**, 19–29.
8. Vogt,V.M. and Braun,R. (1976) Structure of ribosomal DNA in *Physarum polycephalum*. *J. Mol. Biol.*, **106**, 567–587.
9. Campbell,G.R., Littau,V.C., Melera,P.W., Allfrey,V.G. and Johnson,E.M. (1979) Unique sequence arrangement of ribosomal genes in the palindromic rDNA molecule of *Physarum polycephalum*. *Nucleic Acids Res.*, **6**, 1433–1447.
10. Engberg,J., Andersson,P., Leick,V. and Collins,J. (1976) Free ribosomal DNA molecules from *Tetrahymena pyriformis* GL are giant palindromes. *J. Mol. Biol.*, **104**, 455–470.
11. Karrer,K.M. and Gall,J.G. (1976) The macronuclear ribosomal DNA of *Tetrahymena pyriformis* is a palindrome. *J. Mol. Biol.*, **104**, 421–453.
12. Bhattacharya,S., Bhattacharya,A., Diamond,L.S. and Soldo,A.T. (1989) Circular DNA of *Entamoeba histolytica* encodes ribosomal RNA. *J. Protozool.*, **36**, 455–458.
13. Huber,M., Koller,B., Gitler,C., Mirelman,D., Revel,M., Rozenblatt,S. and Garfinkel,L. (1989) *Entamoeba histolytica* ribosomal RNA genes are carried on palindromic circular DNA molecules. *Mol. Biochem. Parasitol.*, **32**, 285–296.
14. Clark,C.G. and Cross,G.A.M. (1987) rRNA genes of *Naegleria gruberi* are carried exclusively on a 14-kilobase-pair plasmid. *Mol. Cell. Biol.*, **7**, 3027–3031.
15. Clark,C.G. and Cross,G.A.M. (1988) Circular ribosomal RNA genes are a general feature of schizopyrenid amoebae. *J. Protozool.*, **35**, 326–329.
16. Cook,J.R. and Roxby,R. (1985) Physical properties of a plasmid-like DNA from *Euglena gracilis*. *Biochim. Biophys. Acta*, **824**, 80–83.
17. Ravel-Chapuis,P., Nicolas,P., Nigon,V., Neyret,O. and Freyssinet,G. (1985) Extrachromosomal circular nuclear rDNA in *Euglena gracilis*. *Nucleic Acids Res.*, **13**, 7529–7537.
18. Bagchi,A., Bhattacharya,A. and Bhattacharya,S. (1999) Lack of a chromosomal copy of the circular rDNA plasmid of *Entamoeba histolytica*. *Int. J. Parasitol.*, **29**, 1775–1783.
19. Ravel-Chapuis,P. (1988) Nuclear rDNA in *Euglena gracilis*: paucity of chromosomal units and replication of extrachromosomal units. *Nucleic Acids Res.*, **16**, 4801–4810.
20. Dhar,S.K., Choudhury,N.R., Mittal,V., Bhattacharya,A. and Bhattacharya,S. (1996) Replication intermediates at multiple dispersed sites in the ribosomal DNA plasmid of the protozoan parasite *Entamoeba histolytica*. *Mol. Cell. Biol.*, **16**, 2314–2324.
21. Scott,N.S. (1973) Ribosomal RNA cistrons in *Euglena gracilis*. *J. Mol. Biol.*, **81**, 327–336.
22. Gruol,D.J. and Haselkorn,R. (1976) Counting the genes for stable RNA in the nucleus and chloroplasts of *Euglena*. *Biochim. Biophys. Acta*, **447**, 82–95.
23. Curtis,S.E. and Rawson,J.R.Y. (1981) Characterization of the nuclear ribosomal DNA of *Euglena gracilis*. *Gene*, **15**, 237–247.
24. Bertaux,O., Mederic,M. and Valencia,R. (1991) Amplification of ribosomal DNA in the nucleolus of vitamin B_{12}-deficient *Euglena* cells. *Exp. Cell Res.*, **195**, 119–128.
25. Falchuk,K.H., Ulpino,L., Mazus,B. and Vallee,B.L. (1977) *E. gracilis* RNA polymerase I: a zinc metalloenzyme. *Biochem. Biophys. Res. Commun.*, **74**, 1206–1212.
26. Brown,R.D. and Haselkorn,R. (1971) Synthesis and maturation of cytoplasmic ribosomal RNA in *Euglena gracilis*. *J. Mol. Biol.*, **59**, 491–503.
27. Greenwood,S.J. and Gray,M.W. (1998) Processing of precursor rRNA in *Euglena gracilis*: identification of intermediates in the pathway to a highly fragmented large subunit rRNA. *Biochim. Biophys. Acta*, **1443**, 128–138.
28. Gunderson,J.H. and Sogin,M.L. (1986) Length variation in eukaryotic rRNAs: small subunit rRNAs from the protists *Acanthamoeba castellanii* and *Euglena gracilis*. *Gene*, **44**, 63–70.
29. Schnare,M.N. and Gray,M.W. (1990) Sixteen discrete RNA components in the cytoplasmic ribosome of *Euglena gracilis*. *J. Mol. Biol.*, **215**, 73–83.
30. Schnare,M.N., Cook,J.R. and Gray,M.W. (1990) Fourteen internal transcribed spacers in the circular ribosomal DNA of *Euglena gracilis*. *J. Mol. Biol.*, **215**, 85–91.
31. Smallman,D.S., Schnare,M.N. and Gray,M.W. (1996) RNA:RNA interactions in the large subunit ribosomal RNA of *Euglena gracilis*. *Biochim. Biophys. Acta*, **1305**, 1–6.
32. Keller,M., Tessier,L.H., Chan,R.L., Weil,J.H. and Imbault,P. (1992) In *Euglena*, spliced-leader RNA (SL-RNA) and 5S rRNA genes are tandemly repeated. *Nucleic Acids Res.*, **20**, 1711–1715.
33. Eichler,D.C. and Craig,N. (1994) Processing of eukaryotic ribosomal RNA. *Prog. Nucleic Acid Res. Mol. Biol.*, **49**, 197–239.
34. Maxwell,E.S. and Fournier,M.J. (1995) The small nucleolar RNAs. *Annu. Rev. Biochem.*, **64**, 897–934.
35. Gerbi,S.A. (1995) Small nucleolar RNA. *Biochem. Cell Biol.*, **73**, 845–858.
36. Greenwood,S.J., Schnare,M.N. and Gray,M.W. (1996) Molecular characterization of U3 small nucleolar RNA from the early diverging protist, *Euglena gracilis*. *Curr. Genet.*, **30**, 338–346.
37. Donis-Keller,H., Maxam,A.M. and Gilbert,W. (1977) Mapping adenines, guanines and pyrimidines in RNA. *Nucleic Acids Res.*, **4**, 2527–2538.
38. Donis-Keller,H. (1980) PhyM: an RNase activity specific for U and A residues useful in RNA sequence analysis. *Nucleic Acids Res.*, **8**, 3133–3142.
39. Smith,S.W., Overbeek,R., Woese,C.R., Gilbert,W. and Gillevet,P.M. (1994) The genetic data environment an expandable GUI for multiple sequence analysis. *Comput. Appl. Biosci.*, **10**, 671–675.
40. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
41. Queen,C. and Korn,L.J. (1984) A comprehensive sequence analysis program for the IBM personal computer. *Nucleic Acids Res.*, **12**, 581–599.
42. Staden,R., Beal,K.F. and Bonfield,J.K. (1999) The Staden package, 1998. *Methods Mol. Biol.*, **132**, 115–130.
43. Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1987). *Current Protocols in Molecular Biology*. Greene Publishing Associates, New York and John Wiley and Sons, PA.
44. Geliebter,J. (1987) Dideoxynucleotide sequencing of RNA and uncloned cDNA. *Focus*, **9**(1), 5–8.
45. Schnare,M.N., Collings,J.C., Spencer,D.F. and Gray,M.W. (2000) The 28S-18S rDNA intergenic spacer from *Crithidia fasciculata*: repeated sequences, length heterogeneity, putative processing sites and potential interactions between U3 small nucleolar RNA and the ribosomal RNA precursor. *Nucleic Acids Res.*, **28**, 3452–3461.
46. Dean,M. (1987) Determining the hybridization temperature for S1 nuclease mapping. *Nucleic Acids Res.*, **15**, 6754.

47. Kudo,T., Yoshitake,J., Kato,C., Usami,R. and Horikoshi,K. (1985) Cloning of a developmentally regulated element from alkalophilic *Bacillus subtilis* DNA. *J. Bacteriol.*, **161**, 158–163.
48. Sogin,M.L., Elwood,H.J. and Gunderson,J.H. (1986) Evolutionary diversity of eukaryotic small-subunit rRNA genes. *Proc. Natl Acad. Sci. USA,* **83**, 1383–1387.
49. Hartshorne,T. and Toyofuku,W. (1999) Two 5′-ETS regions implicated in interactions with U3 snoRNA are required for small subunit rRNA maturation in *Trypanosoma brucei*. *Nucleic Acids Res.*, **27**, 3300–3309.
50. Dooijes,D., Chaves,I., Kieft,R., Dirks-Mulder,A., Martin,W. and Borst,P. (2000) Base J originally found in Kinetoplastida is also a minor constituent of nuclear DNA of *Euglena gracilis*. *Nucleic Acids Res.*, **28**, 3017–3021.
51. Abou Elala,S., Igel,H. and Ares,M.,Jr (1996) RNase III cleaves eukaryotic preribosomal RNA at a U3 snoRNP-dependent site. *Cell*, **85**, 115–124.
52. Kufel,J., Dichtl,B. and Tollervey,D. (1999) Yeast Rnt1p is required for cleavage of the pre-ribosomal RNA in the 3′ ETS but not the 5′ ETS. *RNA*, **5**, 909–917.
53. Neyret-Djossou,O., Freyssinet,G., Ravel-Chapuis,P. and Heizmann,P. (1986) Comparison between the organization of nuclear ribosomal DNA unit of *Euglena gracilis* Z and var. *Bacillaris*. *Plant Mol. Biol.*, **6**, 111–117.
54. Moss,T. and Stefanovsky,V.Y. (1995) Promotion and regulation of ribosomal transcription in eukaryotes by RNA polymerase I. *Prog. Nucleic Acid Res. Mol. Biol.*, **50**, 25–66.
55. Grummt,I. (1999) Regulation of mammalian ribosomal gene transcription by RNA polymerase I. *Prog. Nucleic Acid Res. Mol. Biol.*, **62**, 109–154.
56. Reeder,R.H. (1999) Regulation of RNA polymerase I transcription in yeast and vertebrates. *Prog. Nucleic Acid Res. Mol. Biol.*, **62**, 293–327.
57. Beltrame,M. and Tollervey,D. (1992) Identification and functional analysis of two U3 binding sites on yeast pre-ribosomal RNA. *EMBO J.*, **11**, 1531–1542.
58. Mougey,E.B., Pape,L.K. and Sollner-Webb,B. (1993) A U3 small nuclear ribonucleoprotein-requiring processing event in the 5′ external transcribed spacer of *Xenopus* precursor rRNA. *Mol. Cell. Biol.*, **13**, 5990–5998.
59. Enright,C.A., Maxwell,E.S., Eliceiri,G.L. and Sollner-Webb,B. (1996) 5′ ETS rRNA processing facilitated by four small RNAs: U14, E3, U17 and U3 [published erratum appears in RNA (1996), **2**, 1318]. *RNA*, **2**, 1094–1099.
60. Hughes,J.M.X. and Ares,M.,Jr (1991) Depletion of U3 small nucleolar RNA inhibits cleavage in the 5′ external transcribed spacer of yeast pre-ribosomal RNA and impairs formation of 18S ribosomal RNA. *EMBO J.*, **10**, 4231–4239.
61. Intine,R.V.A., Good,L. and Nazar,R.N. (1999) Essential structural features in the *Schizosaccharomyces pombe* pre-rRNA 5′ external transcribed spacer. *J. Mol. Biol.*, **286**, 695–708.
62. Borovjagin,A.V. and Gerbi,S.A. (2000) The spacing between functional *cis*-elements of U3 snoRNA are critical for rRNA processing. *J. Mol. Biol.*, **300**, 57–74.
63. Maden,B.E.H., Moss,M. and Salim,M. (1982) Nucleotide sequence of an external transcribed spacer in *Xenopus laevis* rDNA: sequences flanking the 5′ and 3′ ends of 18S rRNA are non-complementary. *Nucleic Acids Res.*, **10**, 2387–2398.
64. Michot,B. and Bachellerie,J.-P. (1991) Secondary structure of the 5′ external transcribed spacer of vertebrate pre-rRNA. Presence of phylogenetically conserved features. *Eur. J. Biochem.*, **195**, 601–609.
65. Mishima,Y., Mitsuma,T. and Ogata,K. (1985) Coupled transcription and processing of mouse ribosomal RNA in a cell-free system. *EMBO J.*, **4**, 3879–3886.
66. Mishima,Y., Katayama,M. and Ogata,K. (1988) Identification of a protein factor and the nucleotide sequence required for processing of mouse precursor rRNA. *J. Biochem.*, **104**, 515–520.
67. Good,L., Intine,R.V.A. and Nazar,R.N. (1997) The ribosomal-RNA-processing pathway in *Schizosaccharomyces pombe*. *Eur. J. Biochem.*, **247**, 314–321.
68. Bateman,E. and Paule,M.R. (1988) Promoter occlusion during ribosomal RNA transcription. *Cell*, **54**, 985–992.
69. Henderson,S.L., Ryan,K. and Sollner-Webb,B. (1989) The promoter-proximal rDNA terminator augments initiation by preventing disruption of the stable transcription complex by polymerase read-in. *Genes Dev.*, **3**, 212–223.
70. McStay,B. and Reeder,R.H. (1990) An RNA polymerase I termination site can stimulate the adjacent ribosomal gene promoter by two distinct mechanisms in *Xenopus laevis*. *Genes Dev.*, **4**, 1240–1252.
71. Henderson,S. and Sollner-Webb,B. (1986) A transcriptional terminator is a novel element of the promoter of the mouse ribosomal RNA gene. *Cell*, **47**, 891–900.
72. Antal,M., Mougin,A., Kis,M., Boros,E., Steger,G., Jakab,G., Solymosy,F. and Branlant,C. (2000) Molecular characterization at the RNA and gene levels of U3 snoRNA from a unicellular green alga, *Chlamydomonas reinhardtii*. *Nucleic Acids Res.*, **28**, 2959–2968.