Volume 13 Number 10 1985 Nucleic Acids Research

rRNA processing: removal of only nineteen bases at the gap between 28Sα and 28Sβ rRNAs in *Sciara coprophila*

Vassie C.Ware, Rainer Renkawitz* and Susan A.Gerbi

Division of Biology and Medicine, Brown University, Providence, RI 02912, USA

Received 21 November 1984; Revised and Accepted 18 April 1985

ABSTRACT

We have determined the sequence of the rDNA region between the 28Sα and 28Sβ rRNA coding segments (termed a "gap") in the insect Sciara coprophila, and have used S1 nuclease mapping and cDNA primer extension to define the 5' and 3' boundaries of the gap. Only 19 bases found in rDNA at the gap region are absent from mature 28S rRNA. Eukaryotic rRNAs contain stretches of nucleotides ("expansion segments") which are absent in E. coli rRNA. The gap region in Sciara is located within expansion segment V. Therefore, the excision of 19 bases in the Sciara gap suggests that a large portion of expansion segment V plays no function in mature ribosomes. Specific sequences conserved in Sciara and Drosophila are considered as candidates for recognition signals for the excision of the gap transcript.

INTRODUCTION

The fungus fly, Sciara coprophila, is among a group of organisms that exhibits a specific break in the 26S-28S rRNA at a position approximately halfway down the molecule. Upon hybridizing 26S-28S rRNA to rDNA in R-loop analysis, the break can be visualized in the electron microscope as a "gap" (e.g., [1-3]), thereby dividing 26S-28S rRNA into α and β moieties. Under nondenaturing conditions the α and β halves remain hydrogen bonded together. The rDNA gap is not to be mistaken for an intervening sequence which is found in some systems approximately three quarters from the 5' end of 28S rRNA (within 28Sβ rRNA).

Historically it was unclear if the fragmentation of 26S-28S rRNA was due to nicking during RNA extraction; subsequent analyses, however, determined that the fragmentation pattern is highly reproducible (e.g., [4]). This specific 26S-28S dissociation pattern has since been documented in a wide range of organisms within the phyla Protozoa, Mollusca, Annelida, and Arthropoda. Essentially this phenomenon has been demonstrated in protostomes, protozoa, and some coelenterates, but not in deuterostomes, sponges, platyhelminths, or prokaryotes (5).

Whether or not the break in rRNA is accompanied by a loss of

nucleotides or simply represents a clip in the rRNA has been a subject of controversy. Pellegrini et al. (2) measured a gap of 170 bp in Drosophila, but since its occurrence was limited to only 10% of their hybrid molecules, they concluded that the majority of Drosophila mature 26S rRNA must have a single cleavage in the polynucleotide chain with no concomitant loss of nucleotides. Others have measured gap sizes in Drosophila of 120-340 bases by electron microscopy (e.g., [1, 6, 7]).

This additional cleavage or removal of nucleotides is generally believed to occur within the cytoplasm (8-10), although one group has reported nuclear processing (11). Lava-Sanchez and Puppo (8) could even mimic in vitro the exact 28S fragmentation phenomenon by using a mild pancreatic RNase treatment on newly synthesized ribosomes, suggesting that a nuclease which is active in the cytoplasm cleaves exposed 26S-28S rRNA on the ribosome surface.

Is gap processing of 26S-28S rRNA the result of random nuclease attack due to the apparent accessibility of the gap region on the ribosome surface or is specific endonuclease action involved? Earlier fingerprinting studies in a number of organisms hinted that a specific evolutionarily conserved dinucleotide around the gap region may be recognized by the gap processing machinery (12, 13). There is a dearth of sequence information for the gap region. Direct sequence analysis of the gap region would possibly allow us to identify enzyme recognition signals for gap processing and to determine any differences in sequence between organisms with or without gap processing. In this paper we report the 28S rDNA sequence of the gap region in the fungus fly, Sciara coprophila and the boundaries of the gap (equivalent to the region between the 3' end of 28Sα rRNA and the 5' end of 28Sβ rRNA).

MATERIALS AND METHODS

DNA

Sciara coprophila 28S rDNA was prepared as described by Brand and Gerbi (14) from a plasmid (pBC2) containing a complete rDNA repeat unit. The construction of this pBR322-derived clone has been described previously (15). NIH Guidelines were followed for recombinant DNA work.

RNA

Total RNA was extracted from Sciara larvae using a modification of the guanidinium hydrochloride extraction method of Ernest and Feigelson (16). Three to six hundred larvae were homogenized in 8 mls of guanidinium HCl solution and layered over 4 mls of the cesium chloride solution in a

SW41 polyallomer tube (all buffers and solutions as per [16]). Following centrifugation overnight for 16-17 hours at 25°C at 25K rpm in a Beckman SW41 rotor, the RNA pellet was dissolved in sterile H_2O, adjusted to 2% potassium acetate pH 5.2, and ethanol precipitated. After precipitation at -20°C for at least 4 hours, the RNA was pelleted, dried under vacuum, resuspended in sterile water and reprecipitated with ethanol as above. Following 2-3 washes with 80% ethanol and vacuum drying, the RNA pellet was either stored under 95% ethanol or dissolved in a minimum volume of 0.1 M NaCl/0.01 M Na acetate pH 5.1, buffer. One RNA preparation usually yielded 0.8 to 1.6 mg of RNA. Total RNA was fractionated on 5-20% sucrose gradients as described by Renkawitz et al. (15). Total RNA was extracted from Drosophila melanogaster flies (a generous gift from Marie Levesque) as above.

Enzymes and Nucleotides

Restriction endonucleases were purchased from New England Biolabs or Bethesda Research Labs. DNA polymerase (Klenow fragment), terminal transferase, and polynucleotide kinase were from New England Nuclear (NEN). S1 nuclease was from Boehringer Mannheim. AMV reverse transcriptase was a generous gift from Chris Petropoulos received from Dr. James W. Beard. Deoxynucleotide triphosphates and dideoxynucleotide triphosphates were purchased in the Amersham M13 sequencing kit. The following labeled nucleotides were used: $^{32}P\gamma$-ATP (ICN crude preparation) for 5' end kinase labeling, 3'-dCTP (α-^{32}P) from NEN for Klenow labeling, and 3'-dATP (α-^{32}P) from NEN for 3' end labeling with terminal transferase.

DNA Sequencing

Labeled ends were separated by secondary restriction digestion. DNA sequencing reactions were performed following the method of Maxam and Gilbert (17). All sequencing gels were run in 1X Peacock's Buffer (18).

S1 Nuclease Protection

Hybridization and S1 nuclease digestion were performed as described by Favaloro et al. (19) with the following modifications: 5' or 3' end labeled DNA was denatured for 15 minutes at 85°C; hybridization in 80% formamide was carried out at 62°C for 2 hours. S1 nuclease treatment of the RNA-DNA hybrids was performed at 37°C for 30 minutes in a 300 µl mixture with 300 U of S1 nuclease. Following phenol extraction, reaction mixtures were ethanol precipitated. The following control experiments were done: 1) DNA was denatured and then S1 nuclease digested with no added RNA to assure that the denaturation step was adequate and 2) native DNA was

incubated with S1 nuclease to assure that double stranded nucleic acids were protected from S1 nuclease digestion. DNA resistant to S1 nuclease was displayed on a 10% acrylamide sequencing gel alongside a sequencing ladder of the original 5' or 3' end labeled DNA fragment.

cDNA Primer Extension

The 75 bp primer, 5' end labeled at the Taq I end, was derived from a 250 bp Taq I fragment which was subsequently cut with Acc I; this primer was hybridized to native total rRNA or 28S rRNA. In the presence of deoxynucleotides and reverse transcriptase, cDNAs were transcribed off the RNA templates (20). Sequence analysis on adjacent gel lanes was carried out using intact RNA templates, deoxynucleotides, dideoxynucleotides, and reverse transcriptase.

Heterologous primer extension experiments were performed using intact total rRNA from *Drosophila* and a 30 bp *Sciara* rDNA primer from a secondary BstN1 digestion of the 250 bp Taq I fragment. All primer extension experiments were analyzed on 8% acrylamide sequencing gels.

RESULTS

Determination of Gap Boundaries

The S1 nuclease data and restriction map of *Sciara* rDNA clone pBC2, previously reported by Renkawitz *et al.* (15), were used to delimit the gap region to a 1.0 kb Hae II/Hinc II restriction fragment (Fig. 1). Within this rDNA segment additional restriction sites relative to the Hind III site were mapped by double digestion. Using the strategy shown in Fig. 1, 384 bases of sequence spanning the gap region were determined (Fig. 2).

It was estimated previously by S1 nuclease analysis using uniformly labeled DNA of the coding regions for 28Sα and β rRNAs that the *Sciara* gap is 100 bases in length, and its approximate map position had been located (15). As a first experiment to locate the gap boundaries more precisely, *Sciara* 28S rRNA was hybridized to the 1.0 kb Hae II/Hinc II fragment which was 3' end labeled at the Hae II site. Sizing the protected hybrid, which was S1 nuclease resistant, on a 5% acrylamide sequencing gel suggested that the 3' end of 28Sα rRNA (equivalent to the 5' end of the gap) was approximately 300 nucleotides from the labeled Hae II site.

To further estimate the location of the gap boundaries, the *Sciara* rDNA sequence was aligned with the rDNA sequences of *Xenopus laevis* 28S rDNA (21) and *E. coli* 23S rDNA (22); the latter two sequences lack a gap. By sequence homology one could predict where the gap boundaries might be located. However, just by comparison of the *Xenopus* and *Sciara* rDNA

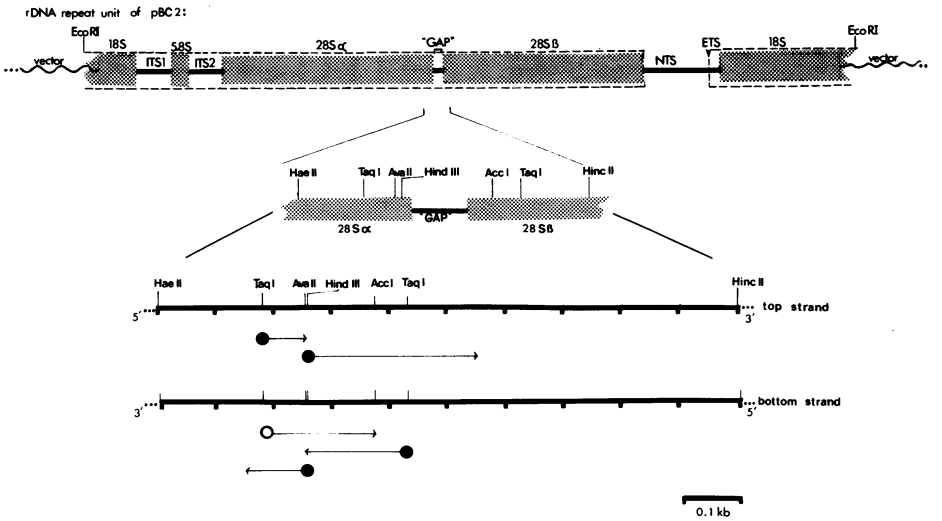

Figure 1. Sequencing strategy used for the gap region in Sciara 28S rDNA. The upper line shows a complete rDNA unit carried as an Eco RI insert in clone pBC2 containing the 3' end of 18S rDNA, internal transcribed spacer (ITS 1), 5.8S rDNA ("5.8S"/2S rDNA - see discussion), ITS 2, 28Sα and 28Sβ rDNA with gap DNA between these two coding regions, nontranscribed spacer (NTS), external transcribed spacer (ETS), and the 5' end of 18S rDNA attached to the vector pBR322 (15). The dotted line shows the transcribed regions of the rDNA unit. The 1.0 kb Hae II/Hinc II restriction fragment spans the gap region as shown. In the lower portion of the figure restriction sites used for 5' (●) or 3' (○) end labeling are indicated; arrows mark the amount of sequence information from the labeled end.

sequences, it appeared that the gap size would be smaller than the original estimate of 100 bases (15). It was clear that the gap region in Sciara is within eukaryotic-specific expansion segment V in Domain IV of 28S rRNA (21, 23). We have previously defined "expansion segments" as regions which are found within the mature rRNAs of eukaryotes but not prokaryotes (21, 23); the sequence and length of expansion segments varies between eukaryotic species, but their positions within rRNA are conserved in all eukaryotes.

In order to refine the map position of the 3' end of 28Sα rRNA and the 5' end of 28Sβ rRNA at the sequence level, a smaller restriction fragment (250 bp) bounded by Taq I sites was used in S1 nuclease protection experiments (Fig. 3). For the 5' labeled fragment, the longest major protected band was 125 bp from the 5' end label after S1 nuclease digestion thereby defining the position of the 3' border of the gap. Similarly, when the same Taq I fragment was 3' end labeled, S1 nuclease digestion located the 5' boundary of the gap 110 bp from the 3' end label.

```
0         10          20          30          40          50          60
GATCTTGGTG  GTAGTAGCAA  ATAATCGAAT  GAGATCTTGG  AGGACTGAAG  TGGAGAAGGG

          70          80          90         100         110         120
TTTCGTGTGA  ACAGTGGTTG  ATCACGAGTT  AGTCGGTCCT  AAGCTTTATG  CGAAAGCTGT

         130         140         150         160         170         180
TTAATTTATG  TTACACACAC  TTAATTTTGT  TGTACATAAT  TCAATAAAAG  CGAAAGGGAA
                   →     →←  →  ———→   →
                             →     →

         190         200         210         220         230         240
TACGGTTCCA  ATTCCGTAAC  CTGTTGAGTA  TACGTTTGTT  TATTAAAAAT  GGGTCTTATT

         250         260         270         280         290         300
ACACTCATCC  TGGCAACAGG  AACGACCATA  GAGAAGCCTT  CGAGAGATAC  TGGAAGAGTT

         310         320         330         340         350         360
TTCTTTTCTG  CTAAACAAAC  GTACCACCAT  GGAAGTCTTT  TATAGAGAGA  TATGGTGGAT

         370         380
GGCTTGGAAG  AGCATGACAT  TTAC
```

Figure 2. DNA sequence of the RNA-like strand of the gap region in *Sciara coprophila*. The boxed-in region represents the bases of the gap which are removed from the mature 28S rRNA. Arrows within box indicate direct and inverted repeats. The sequence presented is from the extreme left arrowhead to the extreme right arrowhead shown in Fig. 1.

It is well-known that S1 nuclease may degrade a hybrid artifactually giving rise to a population of shortened RNA molecules (24). For this reason we considered the largest major band after S1 nuclease digestion to be the true 3' end of 28Sα or the 5' end of 28Sβ rRNA in the experiments described above. We confirmed these conclusions by primer extension. In this case a cDNA copy is extended in a 5' to 3' direction by reverse transcriptase along the rRNA template until the site where the rRNA ends due to the gap. The restriction fragment used as a primer must be downstream on the rRNA relative to the direction of reverse transcription. Therefore it was possible to use this approach to confirm the 5' end of 28Sβ rRNA. Because of the polarity of reverse transcription this tactic cannot be used to ascertain the 3' end of any RNA. Figure 4 shows the results of a primer extension experiment to confirm the 5' end of 28Sβ rRNA. A 5' end labeled 75 bp DNA primer (250 bp Taq I fragment subsequently cut with Acc I) located 3' to the beginning of 28Sβ rRNA was hybridized to total rRNA or 28S rRNA. cDNAs were transcribed up to a break

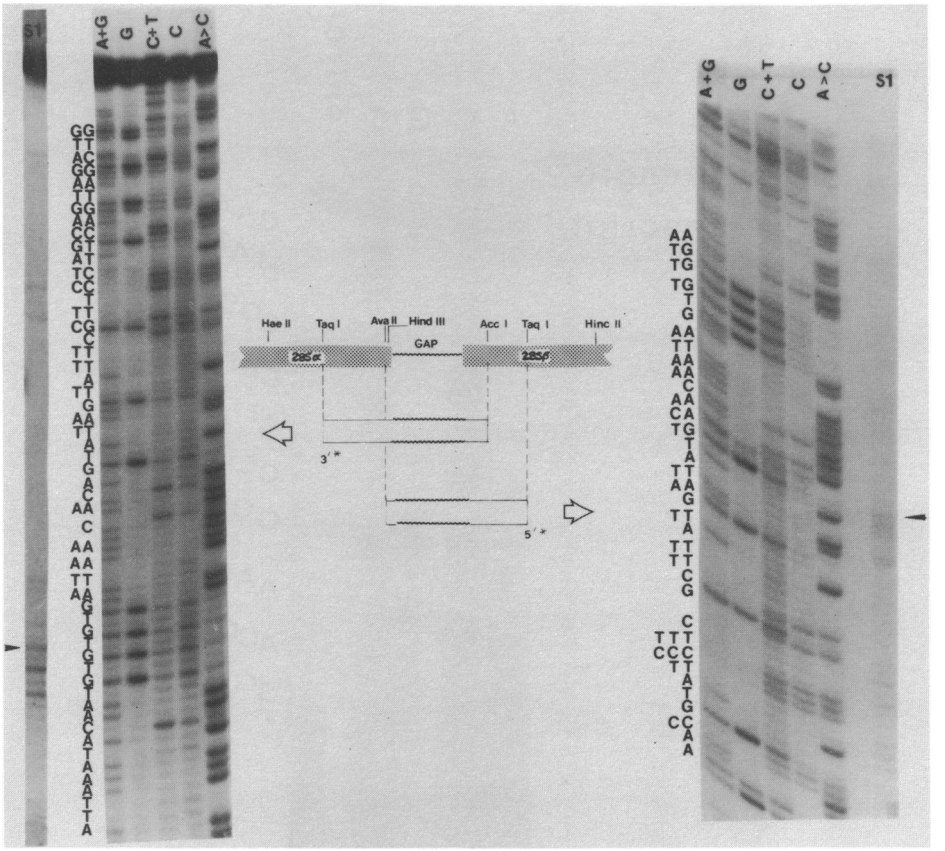

Figure 3. Determination of gap boundaries. A Taq I (250 bp) restriction fragment, 3' or 5' end labeled and subsequently digested to remove one labeled end, was hybridized to Sciara 28S rRNA as described in Materials and Methods. RNA-DNA hybrids were subjected to S1 nuclease treatment. S1 nuclease resistant hybrids were displayed on a 10% acrylamide sequencing gel adjacent to a sequencing ladder of the original full length fragment. The position of the largest major S1 nuclease resistant fragment is indicated by the arrow. Sequences shown are for the template strand (not RNA-like strand). Letters for sequences are read from right to left. The schematic diagram (center of figure) indicates which labeled restriction fragment was used to define the 5' or 3' end of the gap. LEFT PANEL: 3' end labeling experiment to define the 3' end of 28Sα rRNA (equivalent to the 5' boundary of the gap). RIGHT PANEL: 5' end labeling experiment to define the 5' end of 28Sβ rRNA (equivalent to the 3' boundary of the gap).

in the RNA template caused by the naturally occurring 5' boundary of 28Sβ rRNA (= 3' boundary of gap region). These experiments corroborated the S1 nuclease mapping results for the 3' boundary of the gap.

A primer extension experiment was also carried out for Sciara using a 5' end labeled 175 bp rDNA primer (250 bp Taq I fragment cut with Hind III)

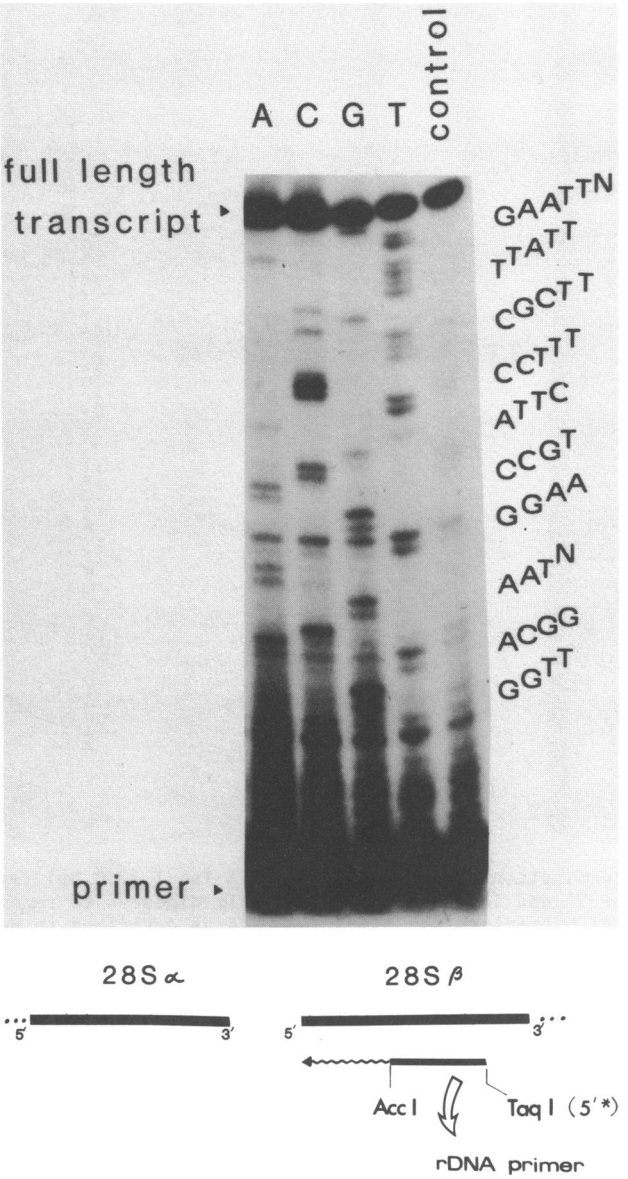

Figure 4. Primer extension analysis. A 5' end labeled (^{32}P) 75 bp rDNA fragment (derived from an Acc I digest of Taq I, 250 bp) was hybridized to Sciara 28S rRNA. The RNAs served as templates for cDNA synthesis (see schematic) (as per [20]). cDNA products were displayed on a 8% acrylamide sequencing gel. No bands larger than the full length transcript shown appeared from that point to the top of the gel.

which traversed the gap sequence. One would predict that cDNA extension would result only if the Hind III site is a part of the 28Sα coding region; if the Hind III site were among the bases processed out of mature rRNA, then there could be no extension beyond the DNA primer. The results agreed with the former prediction and confirmed that the Hind III site must constitute a part of the 28Sα rRNA coding segment and not be within the gap (data not shown).

Since there was such good agreement between the S1 nuclease mapping and primer extension results for Sciara, we felt that it would be useful to compare the recent preliminary results of Delanversin and Jacq (25) for the Drosophila gap determined by S1 nuclease mapping with primer extension results for Drosophila rRNA (current study). Therefore a heterologous primer extension experiment was done using a 30 bp Sciara rDNA primer (250 bp Taq I fragment cut with BstN1) and Drosophila total rRNA (Fig. 5). Using this approach we found the 5' end of Drosophila 26Sβ rRNA to extend twelve bases further upstream (beginning with UAAUU) from the largest S1 nuclease resistant fragment reported by Delanversin and Jacq (25). This sequence is identical to the sequence at the beginning of the Sciara 28Sβ rRNA. Note that the mature 28Sβ transcript in Sciara is longer than that for Drosophila by eight bases in the region between the 3' boundary of the gap and the 3' end of the DNA primer (origin of cDNA synthesis) (shown in part in Fig. 6).

DISCUSSION

A. Structural features of the Gap Region

The amount of material excised by gap processing is extremely small; we have shown here that only nineteen bases are removed in Sciara. The discrepancy in the actual length of the gap from this study compared to the previous estimate of 100 bases (15) can be accounted for in the latter case due to the inability to resolve small changes in restriction fragment sizes with the gel conditions used. The gap size also differs from electron microscopic measurements of R-loops formed between Sciara rDNA clones and 28Sα and 28Sβ rRNAs (3); however, this probably reflects a limitation of the electron microscopy technique as similar overestimates have been noted for the Drosophila 26S gap region as well (25).

The Sciara gap sequence is located within the eukaryotic specific expansion segment V. Expansion segments are found in the same positions in different eukaryotic species, but their length and base composition varies between species (21). Indeed, the base compositional bias of expansion

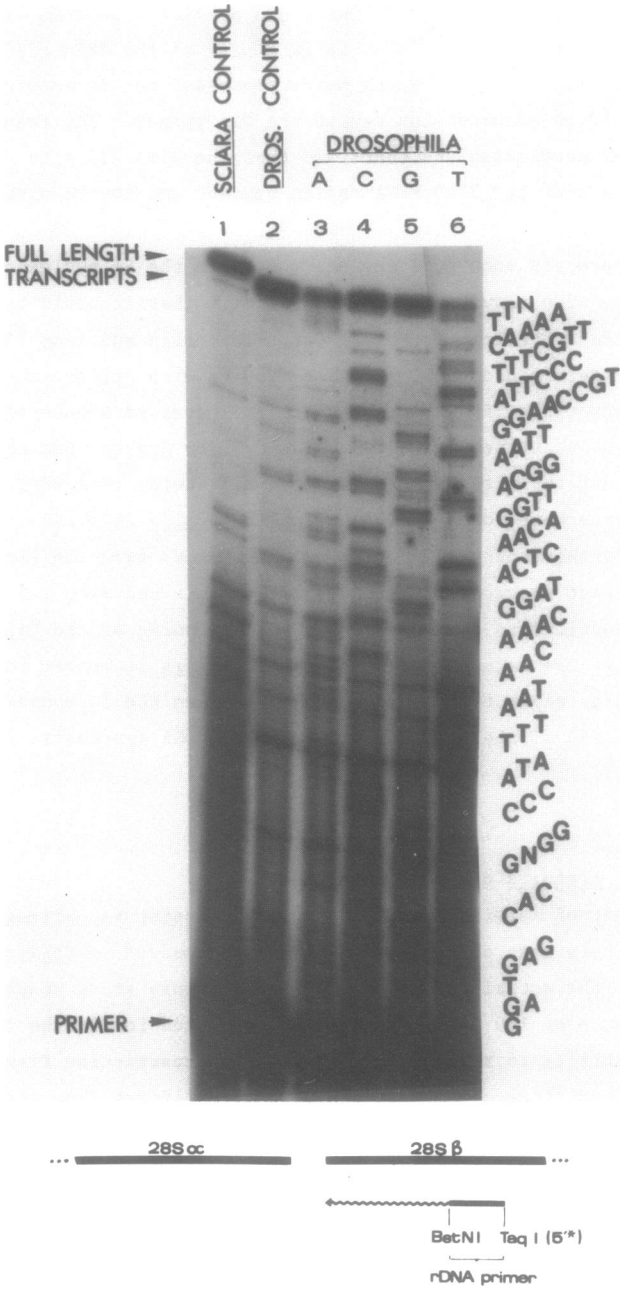

Figure 5. Heterologous primer extension analysis. A 5' end labeled (^{32}P) 30 bp Sciara DNA fragment (derived from a Bst NI digest of Taq I, 250 bp) was hybridized to Sciara total rRNA (lane 1) or Drosophila total rRNA (lanes

2-6). The RNA served as a template for cDNA synthesis in the presence of deoxynucleotides or deoxynucleotides plus dideoxynucleotides and reverse transcriptase. The elongated cDNA products (lanes 1 and 2) were electrophoresed on a 8% acrylamide sequencing gel adjacent to a dideoxy-sequencing ladder of Drosophila cDNA (lanes 3-6). cDNA synthesis continues up to the 5' end of 26S-28Sβ rRNA. Note that the full length transcript for Sciara is longer than the Drosophila transcript (see text for explanation). The Drosophila sequence shown is for the cDNA (not RNA-like strand). Our sequence data are in agreement with the sequence data of Delanversin and Jacq (25). No bands larger than the full length transcript appeared from that point to the top of the gel.

segments contributes to the overall difference in rRNA base compositions between species. This base compositional bias is seen in the Sciara gap sequence which is very A+T rich (74%), as is the remainder of expansion segment V.

Using the recent sequence data for Drosophila 26S rDNA (25), we have aligned segments of the Sciara and Drosophila 26S-28S rDNA sequences in the gap region (Fig. 6). This alignment supports our data which show that the 5' end of 26Sβ extends further upstream than previously reported by Delanversin and Jacq (25). Areas which are adjacent to the gap region (outside the boundaries of expansion segment V) have been strongly conserved between all eukaryotes (see sequence alignment in [21]) and especially between Sciara and Drosophila. It was the strong sequence homology, particularly on the 3' side of the expansion segment, that made the heterologous primer extension experiment possible (see Fig. 5); Drosophila and Sciara rDNAs are 86% homologous in this region.

Using the alignment between the Sciara and Drosophila gap regions, we can ask if these two insects share any features in common which may be recognized as signals for gap processing and are absent in other eukaryotes such as yeast and Xenopus which lack gap processing. Comparison of the primary rDNA sequence around the processing site for 26S-28Sβ reveals features common to Sciara and Drosophila, most notably the hexanucleotide ATAATT (Fig. 6); this sequence is absent from expansion segment V in yeast (26, 27) and Xenopus (21). Conservation of sequences within the gap region, especially at the beginning of 26S-28Sβ rRNA, implies that there is some specificity associated with the processing reaction and that there may be selective pressure to maintain certain sequences in the gap region.

We have previously noted that base duplications occur near the borders of rRNA expansion segments (23). For Sciara and Drosophila, however, the pattern of duplication differs from that of the other eukaryotes: an identical duplication CGAAAG exists on each side of the

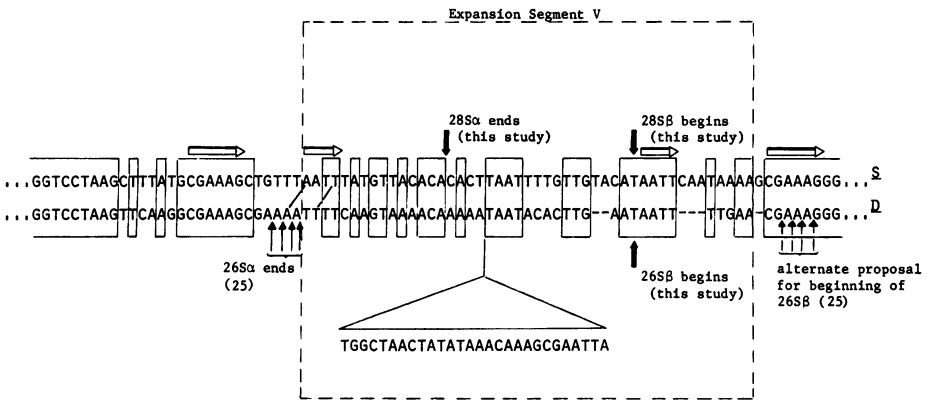

Figure 6. Alignment of Sciara and Drosophila 26S-28S rDNA sequences in the gap region. "S" represents Sciara coprophila 28S (this paper); "D" represents Drosophila 26S (25). The sequence is from the RNA-like strand in both cases. The alignment starts at Sciara base 95 (from Figure 5) and extends to Sciara base 178. The solid boxes are regions of sequence homology; sequence homologies extend beyond the region shown, as this alignment includes only the area in the immediate vicinity of expansion segment V (dotted boxed-in region). The large open arrows indicate direct repeats.

expansion segment boundary within the conserved 28S rRNA core sequence and an additional duplication containing the consensus sequence AATT is situated generally within the expansion segment boundaries (Fig. 6). What role, if any, the base duplications play in gap processing remains to be determined.

Secondary structure as well as primary sequence may also play a role for specifying gap processing. However, a completely proven model for secondary structure of the gap cannot be obtained unless sufficient amounts of precursor (with gap bases still present) are available. It was only by using in vitro transcription and cell-free splicing that the secondary structure of the purified intron from Tetrahymena rRNA could be experimentally proven (28, 29). Using the method of Qu et al. (20) we have experimentally determined the secondary structure of the 5' end of mature 28Sβ rRNA from Sciara (data not shown). These results support the notion that expansion segment V forms a stem structure, and that the gap processing boundaries are partway up this stem. The stem of expansion segment V in organisms such as Xenopus (23) and yeast (26), which lack gap processing, is shorter than in Sciara and Drosophila. Moreover, in these insects a larger apical loop can be drawn, and this may be a recognition

signal for gap processing to occur. However, this awaits experimental proof.

Once the gap bases are excised, a relatively short expansion segment stem of 5 bp would remain in our preliminary secondary structure model for Sciara (not shown). An additional nearby long range interaction of 10 bp has been proposed for Domain IV of Xenopus (23), and this may also serve to hold 28Sα and 28Sβ rRNA moieties together after gap processing in Sciara. Evidence from studies on other gap processing systems suggests that the amount of hydrogen bonding between the 26S-28Sα and 28Sβ halves is probably limited to a relatively small number of adjacent base pairs: 1) Electron microscopy of psoralen crosslinked mature 26S rRNA from Drosophila revealed a central hairpin in which the two parts of 26S rRNA are crosslinked (probably near the 26Sα and 26Sβ ends) in the hairpin at its base (30); 2) Thermal denaturation studies on the 26S rRNA from the silkmoth species, Hyalophora cecropia, showed that the dissociation temperature for this rRNA is quite low at 45-50°C (4). Similarly, Sciara 28S rRNA also melts at 45°C in 0.1 M NaCl/0.01 M Na acetate pH 5.1, buffer (our unpublished observations).

B. Comparison to other rRNA processing systems.

It could be hypothesized that no enzymatic machinery is involved in gap processing, but that the processing activity is intrinsic to the gap region precursor rRNA, in a similar fashion to the self-splicing capability of the Tetrahymena rRNA intron (31). We examined the Sciara gap sequence for similarities to the Tetrahymena intron boundaries and found no consensus sequences; yet, we cannot rule this out as a formal possibility. However, since intron self-excision is a rapid first event in Tetrahymena rRNA processing (32, 33), and gap excision is a later event in processing (10), we feel that self-excision of the gap is unlikely.

Other rRNA processing systems were examined for structural similarities to gap processing in Sciara and Drosophila. The Sciara gap is comparable to other internal transcribed spacers in rDNA in which the spacer is removed during processing leaving intact two separate moieties that generally are hydrogen bonded together (e.g., 5.8S/28S in eukaryotes: [34]; "5.8S"/2S in insects: [35, 36]; 4.5S/23S in higher plant chloroplasts: [37]).

(1) Comparison to 4.5S RNA processing. It is noteworthy that the oligonucleotide ATAA has been mapped to the 5' end of 4.5S rDNA in higher plant chloroplasts (37); the same sequence encompasses the 5' end of insect 28Sβ rRNA (Fig. 6 this paper). The 4.5S rRNA is located at the 3' end of

23S rRNA, and thus shares a similar position in the rRNA precursor with respect to the spacer DNA as does the 28Sβ rRNA of Sciara relative to the gap. It is likely that the chloroplast 23S/4.5S spacer includes eukaryotic-specific expansion segment IX, again being similar to the situation of the insect gap which constitutes a part of expansion segment V. If a similar enzyme is involved in 23S/4.5S and gap processing, then one has to explain its localization in two different cellular compartments: in the chloroplasts of higher plants and the cytoplasm of organisms with gap processing.

(2) Comparison to "5.8S"/2S RNA processing. In the case of "5.8S"/2S processing described in Drosophila and Sciara (35, 36), 2S rRNA corresponds to the 3' part of 5.8S rRNA from other species. Like gap processing, "5.8S"/2S processing is a cytoplasmic event (11). The 5' end of 2S is hydrogen bonded to the 3' end of the insect "5.8S" rRNA. Although there are no direct sequence equivalents at the 5' end of 2S rRNA, there are certain features that have been conserved between "5.8S"/2S and gap processing:

a) Base duplications surround the rDNA sequence that is removed in "5.8S"/2S processing. The base duplications in the "5.8S"/2S spacer region in Sciara and Drosophila are not conserved in sequence with those present in the gap region in Sciara or Drosophila. However, the relative positions of the spacer region duplications are similar to those in the gap region with respect to the spacer boundaries: in "5.8S"/2S one set of duplications (C A Py A U) clearly falls outside the boundaries of the spacer; the other set (containing the consensus sequence CUG) falls outside the spacer boundary on the "5.8S" side and within the cleavage site on the 2S side. The presence of two sets of base duplications as described in "5.8S"/2S and gap processing may be a general feature recognized by the spacer and gap excision machinery.

b) A secondary structure model has been described in Drosophila (35) and Sciara (36) that allows for "5.8S"/2S association and spacer processing to occur simultaneously without invoking changes in rRNA conformation to achieve a base pairing relationship. Our preliminary Sciara secondary structure model accomplishes the same end: once the gap bases are removed, the 28Sα and β moieties remain hydrogen bonded together in this region. These conditions cannot be fulfilled for Drosophila if one accepts the gap boundaries as described by Delanversin and Jacq (25); however, our 26Sβ data for Drosophila would support a hydrogen bonded stem in the gap region.

c) The spacer between "5.8S" and 2S loops out as an A-U rich region in the secondary structure model possibly with some internal base pairing; adjacent inverted repeats (Sciara: GUUUCUUUUAUUA; Drosophila: CUUUAAUUAAUUU) reside within this A-U rich "5.8S"/2S loop. Similarly, a short inverted repeat UUAAUU is present within the Sciara expansion segment V hairpin loop, and a different inverted repeat GUAAAACAAAAAUG occurs on the left side of the larger loop in Drosophila. Unlike the loop in Sciara and Drosophila, the smaller expansion segment loop in yeast and Xenopus lacks any tandem or inverted repetition.

Therefore, to summarize, our clues for gap processing signals in Sciara and Drosophila include 1) the sequence AUAAUU which is present at the start of 26S-28Sβ rRNA; AUAA is also at the 5' end of 4.5S rDNA in higher plant chloroplasts. 2) CGAAAG and an additional sequence containing the consensus AAUU which are duplicated within the gap region; base duplications are also observed in the spacer region in "5.8S"/2S processing in insects. 3) The gap region probably exists as a hairpin structure in shich the A-U rich hairpin loop is larger and contains inverted repeats unlike the counterpart hairpin structure in yeast or Xenopus. "5.8S"/2S, like gap processing, also has inverted repeats in a loop which is excised.

C. Functional and Evolutionary Considerations

Most likely the recognition signals necessary for gap processing are encoded within expansion segment V since the surrounding core sequence is generally conserved among organisms with and without a central break in the 26S-28S rRNA. If expansion segments represent insertion events relative to E. coli (as they do have features reminiscent of mobile elements), then one must assume that gap processing was "introduced" after expansion segment V was inserted into this domain. Why, then, do some organisms such as Xenopus, yeast, rat, and mouse lack processing at this site? Have these organisms merely lost the capability for gap processing as a consequence of changes in recognition elements or processing machinery? Hints to the evolutionary directionality of the gap processing phenomenon might possibly arise from the development of heterologous systems to assay the competency of organisms such as Xenopus or yeast to process rRNAs at the gap site.

We previously speculated that expansion segments represent nonfunctional regions within rRNA among eukaryotes (with the exception of the 3' end of expansion segment II where 25 bases are completely conserved between eukaryotes) and that these segments have been tolerated in rRNA because they do not disrupt function (21, 23). The fact that the bases

within the gap are removed in some species suggests that they play no role in the mature ribosomes of those organisms, although whether these bases leave the ribosome after gap processing and are degraded remains to be shown.

Localized primary and secondary structure conservation within the Sciara and Drosophila gap regions suggests that there is positive selection to retain those features that specify gap processing. Although it is clear that organisms that lack processing at this site can conduct ribosome functions in the absence of this extra excision, there may be some functional advantage for the species that have gap processing.

ACKNOWLEDGEMENTS

We gratefully acknowledge Dr. Judith Furlong for helpful discussions during the course of this work. We thank Dr. J.-P. Bachellerie for preprints that led us to the discovery of the Delanversin and Jacq (25) paper at the conclusion of our experiments. Susan DiBartolomeis and Heidi Smith helped in the collection of Sciara larvae. Mrs. Carol King is to be commended for excellence in typing this manuscript. This work was supported by grant PHS-GM 20261 to S.A.G. The work presented in this paper was given the Young Investigator Award at the Society for Developmental Biology 43rd Annual Symposium, 1984.

* Present Address: Gene Technology Groups, Max-Planck-Institut, D-8033 Martinsried, West Germany

REFERENCES

1. Wellauer, P.K. and Dawid, I.B. (1977). Cell 10, 193-212.
2. Pellegrini, M., Manning, J., and Davidson, N. (1977). Cell 10, 213-224.
3. Renkawitz-Pohl, R., Matsumoto, L., and Gerbi, S.A. (1981). Nucl. Acids. Res. 9, 3747-3764.
4. Applebaum, S.W., Ebstein, R.P., and Wyatt, G.R. (1966). J. Mol. Biol. 21, 29-41.
5. Rubin, N.A. and Eckhardt, R.A. (1977). J. Cell Biol. 75, 355a.
6. White, R.L. and Hogness, D.S. (1977). Cell 10, 177-192.
7. Dawid, I.B., Wellauer, P.K., and Long, E.O. (1978). J. Mol. Biol. 126, 749-768.
8. Lava-Sanchez, P.A. and Puppo, S. (1975). J. Mol. Biol. 95, 9-20.
9. Serfling, E. (1976). Biol. Zbl. 95, 713-723.
10. Eckert, W.A., Kaffenberger, W., Krohne, G., and Franke, W.W. (1978). Eur. J. Biochem. 87, 607-616.
11. Jordan, B.R., Jourdan, R., and Jacq, B. (1976). J. Mol. Biol. 101, 85-105.
12. Shine, J., Hunt, J.A., and Dalgarno, L. (1974). Biochem. J. 141, 617-625.
13. Ishikawa, H. (1976). Biochim. Biophys. Acta 425, 185-195.
14. Brand R.C. and Gerbi, S.A. (1979). Nucl. Acids Res. 7, 1497-1511.
15. Renkawitz, R., Gerbi, S.A., and Glätzer, K.H. (1979). Molec. Gen. Genet. 173, 1-13.
16. Ernest, M.J. and Feigelson, P. (1978). J. Biol. Chem. 253, 319-322.

17. Maxam, A.M. and Gilbert, W. (1980). Meth. Enzymol. 65, 449-560.
18. Peacock, A.C. and Dingman, C.W. (1968). Biochem. 7, 668-674.
19. Favaloro, J., Treisman, R., and Kamen, R. (1980). Meth. Enzymol. 65, 718-749.
20. Qu, L.H., Michot, B., and Bachellerie, J.-P. (1983). Nucl. Acids Res. 11, 5903-5920.
21. 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.
22. Brosius, J., Dull, T.J., and Noller, H.F. (1980). Proc. Nat. Acad. Sci. USA 77, 201-204.
23. Clark, C.G., Tague, B.W., Ware, V.C., and Gerbi, S.A. (1984). Nucl. Acids Res. 12, 6197-6220.
24. Miller, K.G. and Sollner-Webb, B. (1981). Cell 27, 165-174.
25. Delanversin, G. and Jacq, B. (1983). C.R. Acad. Sc. Paris, 296, Série III, 1041-1044.
26. Veldman, G.M., Klootwijk, J., de Regt, V.C.H.F., Planta, R.J., Branlant, C., Krol, A., and Ebel, J.-P. (1981). Nucl. Acids Res. 9, 6935-6952.
27. Georgiev, O.I., Nikolaev, N., Hadjiolov, A.A., Skryabin, K.G., Zakharyev, V.M., and Bayev, A.A. (1981). Nucl. Acids Res. 9, 6953-6958.
28. Cech, T.R., Tanner, N.K., Tinoco, I., Weir, B.R., Zuker, M. and Perlman, P.S. (1983). Proc. Nat. Acad. Sci. 80, 3903-3907.
29. Inoue, T. and Cech, T.R. (1985). Proc. Nat. Acad. Sci. 82, 648-652.
30. Wollenzien, P.L., Youvan, D.C., and Hearst, J.E. (1978). Proc. Nat. Acad. Sci. USA 75, 1642-1646.
31. Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R. (1982). Cell 31, 147-157.
32. Cech, T.R. and Rio, D.C. (1979). Proc. Nat. Acad. Sci. USA 76, 5051-5055.
33. Din, N., Engberg, J., Kaffenberger, W., and Eckert, W.A. (1979). Cell 18, 525-532.
34. Pene, J.J., Knight, E., JR., and Darnell, J.E., JR. (1968). J. Mol. Biol. 33, 609-623.
35. Pavlakis, G.N., Jordan, B.R., Wurst, R.M., and Vournakis, J.N. (1979). Nucl. Acids Res. 7, 2213-2238.
36. Jordan, B.R., Latil-Damotte, M., and Jourdan, R. (1980). Nucl. Acids Res. 8, 3565-3573.
37. Edwards, K. and Kössel, H. (1981). Nucl. Acids Res. 9, 2853-2869.