Volume 15 Number 13 1987 Nucleic Acids Research

In vivo transcription from multiple spacer rRNA gene promoters during early development and evolution of the intergenic spacer in the brine shrimp *Artemia*

Hannelore T.Koller, Kathleen A.Frondorf[1], Patricia D.Maschner[2] and Jack C.Vaughn*

Department of Zoology, Miami University, Oxford, OH 45056, USA

Received February 2, 1987; Revised and Accepted June 4, 1987

ABSTRACT

The control of ribosomal RNA (rRNA) gene expression during development can be productively studied by examination of the relationship between promoter structure and function as well as the processing of primary transcripts. Toward this end, total cell RNA was extracted from embryos at various stages and probed with cloned rRNA genes using the "dot blot" method. This exercise showed that rRNA gene expression is a stage-specific process and is thus under developmental control. S1 nuclease protection experiments localized fourteen different upstream DNA sites encoding 5'-termini of pre-rRNAs during this synthetic phase of development. There is no indication of any spacer fail-safe terminator function. The S1 approach contributed to the sequencing of several of the sites. Comparative sequence alignments reveal short conserved regions in DNAs corresponding to these sites, which are shown to fall into two structural classes. Sites 3, 4, 6 and 9 are proposed to function in transcription initiation and are found to have the consensus sequence: 5' ...T-A-T-A-T-Pu-Pu-Pu-G-Pu-Pu-G-T-C-A... 3'. Sites 1, 2, 5 and 8, which are proposed to function in 5'-processing have the consensus sequence: 5' ...Pu-G-T-Pu-T-T-G... 3'. These short sequence conserved regions are hypothesized to serve as recognition signals for proteins within the rDNA transcription initiation complex and for 5'-processing enzymes, respectively. Sequencing of the intergenic spacer region, from which a model for spacer evolution is derived, shows that tandem ca. 600 bp subrepeats explain much of the multiplicity observed within control sites.

INTRODUCTION

Eukaryotic ribosomes contain four different individual ribosomal RNA (rRNA) molecules in stoichiometric amounts which are named on the basis of their differing sedimentation coefficients. These are the 5S, 5.8S and 25 - 28S rRNAs in the 60S ribosomal subunit and the 18S rRNA in the 40S subunit. The 5S genes are usually located at some distance from the others and are commonly organized into tandemly repeated units. In *Artemia*, however, virtually all the 5S rRNA genes are interspersed amongst histone gene repeat units (1). The other rRNA genes are typically located at the nucleolar organizer locus and are organized into hundreds of tandemly repeated units or rDNA units having the polarity 5'...18S/5.8S/26S... 3'. We have previously described the cloning of a complete rDNA repeat unit from the brine shrimp *Artemia* (2), and have shown that there are about 300 copies of this repeat unit per haploid genome (3). Within a repeat unit, the 5.8S rRNA coding region is flanked by internal transcribed spacers, and the immediate 18S upstream region comprises an external transcribed spacer (ETS). A transcription unit begins at the 5'-end of the ETS and apparently terminates in the vicinity of the 3'-end of the 26S rRNA coding region in *Artemia* (C.J. Lee and J.C. Vaughn unpublished data). Adjacent transcription units are separated from one another by long "nontranscribed" or intergenic spacer regions (NTS). Following transcription, the resulting

pre-rRNA molecules are processed to yield the mature 5.8S 18S and 26S rRNAs. The 5S (4) and 5.8S rRNA (5) molecules have been sequenced in *Artemia* as have the DNA regions encoding 5.8S (6) and 18S rRNAs (7). For recent pertinent reviews see ref. 8 and 9.

Existence of short conserved upstream promoter DNA tracts for bacterial (10) and eukaryotic RNA polymerase II genes (11), both within and between species, facilitated the early tentative identification of the critical "TATA" sequence element necessary for expression of these genes. Early progress in characterizing a model RNA polymerase I gene's expression was slowed in part by the general lack of conserved promoter DNA tracts between distantly related species (12) whose recognition could have signaled their potential significance and by the lack of reproducible *in vitro* transcription systems. Very closely related species (13) and moderately closely related species (14 - 16) do show considerable sequence conservation around the rRNA gene transcription initiation site, which in some cases may extend for one hundred or more nucleotides. This is especially the case for the regions surrounding the duplicate spacer promoters of some species (17, 18). In the well studied *Xenopus laevis* system, the existence of conserved tracts surrounding transcription initiation sites at the gene promoter and within duplicated spacer promoters correlates nicely with shorter regions identified from *in vitro* promoter mapping assays (19).

In this report, putative *Artemia* rRNA gene transcription initiation and pre-rRNA processing sites are precisely located by an S1 nuclease protection approach and their nucleotide sequences are determined. Potentially important nucleotide positions are identified using the comparative sequence approach. One of the most surprising findings is that during early development, at stages in which rapid synthesis of rRNA is occurring, multiple upstream promoters which are otherwise relatively inactive play a major role in transcription. This is effective because no functional fail-safe terminator appears to be present in this system.

MATERIALS AND METHODS

The brine shrimp *Artemia* inhabits saline lakes and pools, where its fertilized eggs are carried in the female's brood pouch until hatching occurs. Under conditions of low oxygen tension or high salinity, the developing embryo becomes encysted in a virtually impenetrable shell, dehydrates, and goes into a dormant stage. These dormant cysts, which have been analogized to plant seeds, remain developmentally frozen in time at approximately the gastrula stage of development until they are resuspended in salt water, at which time they resume synchronous normal development.

Animals

Dehydrated gastrula stage brine shrimp embryos collected from salterns in the vicinity of San Francisco Bay, Calif., were obtained from the Metaframe Corp. and stored at -20 C under desiccation. One gram aliquots were cultured at 30 C in 100 ml brine shrimp saline (20) in a gyrotory incubator operating at 100 rpm. Flask contents were harvested at periodic intervals and rinsed with distilled water on plankton netting. Swimming nauplius larvae from 24 h incubation and older cultures were separated from shells and unhatched cysts by phototactic response in large separatory funnels prior to filtration (21). In some runs, animals were fed yeast (22) starting at 24 h of incubation. Percent viability and numbers of larvae were determined by a modification of the method described in ref. 21.

Nucleic acid isolation and characterization from embryos

Glassware and solutions for RNA isolation were either autoclaved or baked for 6 h in a dry heat sterilizer to destroy RNAase activity prior to use wherever practicable. Animals were

blotted after filtration, transferred to a ground glass tissue grinder, and homogenized on ice in diethylpyrocarbonate-treated RNA isolation buffer: 50 mM sodium acetate, 5 mM $MgCl_2$, 0.5% (w/v) sodium dodecyl sulfate, pH 5.1, using a pulley-driven Thomas homogenizer. Examination of aliquots at all developmental stages by phase contrast microscopy verified that cyst, larval and cell disruption were complete. An aliquot was set aside for subsequent determination of DNA content. RNA was then isolated from the remainder of the original homogenate using a cold phenol method (modified from ref. 23), followed by proteinase K and RNAase-free DNAase digestions. The quantity of RNA obtained from each developmental stage was estimated by ultraviolet absorption, following dissolution in autoclaved distilled water. Although RNA recoveries were far from quantitative by this procedure, as expected, it was nevertheless apparent after repeated isolations that the relative quantities obtained per developmental stage within a given experimental series were reproducible to the extent that trends could be recognized and predicted from run to run.

DNA content was determined from an aliquot of original homogenate by the diphenylamine method (24) following adjustment to 10% trichloroacetic acid (TCA), hydrolysis for 15 min at 100 C, removal of particulate matter by Eppendorf centrifugation, and washing of the pellet with 10% TCA to recover all possible hydrolyzed DNA.

The relative quantity of (18S + 26S) rRNA per developmental stage was estimated from the recovered RNA samples using a dot blot assay procedure (25). RNAs were immobilized on nitrocellulose filters at several different serial dilutions. Ribosomal RNAs were determined following hybridization (26) with nick-translated (27) ^{32}P-labeled recombinant plasmids pXlr11 plus pXlr12, which together contain one complete *Xenopus* (18S + 28S) rDNA repeat unit, as previously described (2). Hybridization signals were quantitated by densitometry or direct counting in a scintillation counter, and it was determined that signal decreased in proportion to the extent of dilution of RNAs.

Preparation of rDNA subclones and restriction endonuclease mapping

The molecular cloning of a complete 13.9 kilobase (kb) *Artemia* rDNA repeat unit, designated λ Ch4A BSr1, has been described (2). The recombinant plasmid pBSr5 constructed in pBR322 from this cloned repeat unit after cutting with EcoRI and SalI contains about 2150 base pairs (bp) of spacer located immediately upstream of the 18S rRNA coding region. The plasmid which we designate pBSr11 was constructed in pBR322 from the cloned full repeat unit after cutting with EcoRI as previously described (6). The ca. 11 kb insert designated BSr11 contains an entire nontranscribed spacer (Fig. 1).

Restriction mapping of the 4.9 kb insert termed BSr5, the details of which have been reported (28) is summarized in Fig. 1. Restriction enzymes were utilized under conditions specified by the supplier (BRL).

Site mapping of DNA tracts encoding pre-rRNA 5'-termini

Approximate positions of DNA tracts encoding pre-rRNA 5'-termini were determined using an S1 nuclease protection method (29), adapted for use with similar systems (30, 13). Total cell RNAs obtained from various developmental stages were utilized as a source of pre-rRNA in these determinations. The method involved isolation (see below) and utilization of the coding strand of the 5'-end labeled 8.4 kb XbaI*/EcoRI DNA fragment obtained from pBSr11. Labeled DNA was combined with 200 ug of total cell RNA obtained from each developmental stage, denatured 10 min at 70 C and hybridized for 3 - 5 h at 45 C in 40 mM Tris-HCl, 0.3 M NaCl, 1 mM EDTA, 80% (v/v) deionized formamide, pH 8.0 in a total volume of 50 ul. The reaction mixture was then immediately diluted ten-fold into 37 C pre-warmed S1 nuclease buffer: 50 mM sodium

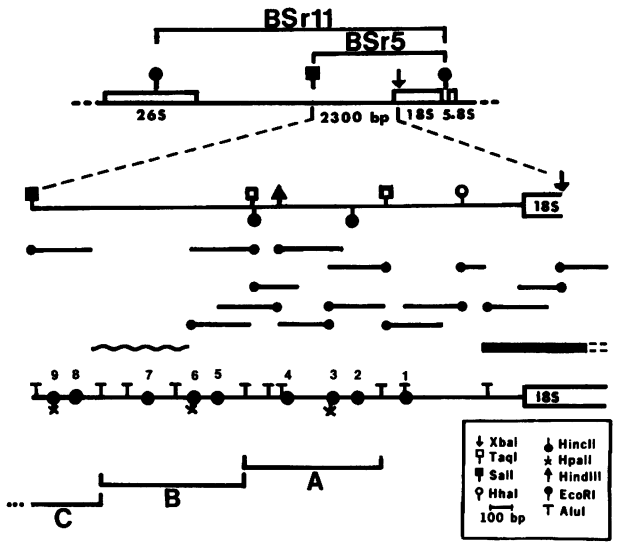

Figure 1. Restriction enzyme maps and experimental design for sequencing NTS within rDNA promoter region. Locations of 5.8S, 18S and 26S rRNA coding regions are in elevated boxes. Recombinant plasmids pBSr5 and pBSr11 contain the indicated inserts. The 2300 bp SalI/XbaI fragment cut from BSr5, is shown in expanded scale. Short overlapping ^{32}P end-labeled restriction fragments (—●) were isolated and sequenced. Positions to which the first nine pre-rRNA 5'-termini map are diagrammed along the lower map (solid black circles), as indicated by results shown in Fig. 3. The black bar represents the extent of spacer region sequence given in ref. 7. Wavy line region has not yet been sequenced. Subrepeat structure (A, B and C) is proven by results in Fig. 7.

acetate, 0.15 M NaCl, 0.5 mM $ZnSO_4$, pH 4.8 containing 10 - 50 U of S1 nuclease (BRL) in various experiments and incubated 1 h at 37 C. Approximate locations of DNA tracts encoding pre-rRNA 5'-termini were estimated relative to the end-labeled XbaI terminus of the DNA probe by measurement of the length of the protected S1-treated end-labeled DNA fragment(s), and equal the distance from the labeled terminus to the 3'-end of the probe located within the upstream spacer. This length was determined for denatured DNA fragments, following hybridization and S1 nuclease treatment, by alkaline agarose gel electrophoresis (31). End-labeled denatured size standards consisted of pBR322 DNA cut with AluI combined with pBR322 cut with TaqI and also lambda/EcoRI + HindIII DNA restriction fragments (32). Following electrophoresis, gels were neutralized, DNA fragments transferred to nitrocellulose filters (33), and autoradiographs were prepared.

DNA fragment isolation and DNA sequence analysis

Our experimental plan for sequencing much of the rDNA gene upstream region is summarized in Fig. 1. The 2.3 kb XbaI/SalI fragment was isolated by preparative low melting temperature agarose (IBI) gel electrophoresis, electroelution, DE-52 (Whatman) column chromatographic purification (34) and ethanol precipitation. Appropriate restriction fragments isolated in the same manner were then either 3'-end labeled (35) or 5'-end labeled following dephosphorylation (36) and singly end-labeled fragments were obtained following

secondary restriction enzyme digestion Determinations of putative transcription initiation and/or processing sites relative to the detailed restriction map made it possible to isolate short 5'-end labeled DNA coding strand fragments spanning these regions, for analysis of the precise nucleotide sequences of the DNA tracts encoding pre-rRNA 5'-termini. One half of each original end-labeled DNA fragment preparation was subjected to the procedure described above for site mapping, with the modification that after hybridization the sample was split into two aliquots which were digested with 25 U and 100 U of S1 nuclease, respectively. The other half of each original end-labeled DNA fragment preparation was subjected to chemical degradation steps for subsequent sequencing. The S1-treated protected DNA fragment preparation was then run in a polyacrylamide sequencing gel alongside the Maxam/Gilbert fragment preparation, where it represents the equivalent of one rung in the resulting sequencing ladder. In this manner, the precise sequences of the putative rRNA gene transcription initiation and/or processing sites were determined. The length of the short (usually less than 200 nucleotides) DNA fragment protected from S1 nuclease digestion by hybridization to RNA corresponds to the distance from the nucleotide encoding the 5'-terminus of the pre-rRNA molecule to the 5'-labeled end of the protected DNA coding strand (30). DNA sequencing was carried out using the Maxam and Gilbert (36) base specific chemistry protocols No. 10 and 12 - 14, followed by polyacrylamide gel electrophoresis in 20% gels to obtain the initial ca. 35 nucleotides and in double loaded 8% gels to obtain the next ca. 200 nucleotides. Sequence comparisons were carried out using computer-assisted analysis techniques (37). In calculations of percent sequence homology, gaps introduced to facilitate alignment and thus having no paired nucleotide were scored as half a mismatch (38).

RESULTS AND DISCUSSION

Our interest in the control of expression of developmentally regulated genes led us to study early developmental stages for evidence of the onset of rRNA synthesis following rehydration of gastrula stage *Artemia* embryos. It is not feasible in this system to determine at which developmental stage(s) rRNA synthesis resumes by the classical approach of scoring for uptake of labeled nucleic acid precursors into rRNA, for the cyst wall is virtually impenetrable by such molecules. Uptake of labeled precursors into rRNA of nauplius larvae has been demonstrated, and it has also been shown that total embryo RNA isolated from a high speed pellet (presumably largely rRNA) peaks at 30 - 36 h of development (39). However, the methods employed did not directly prove that rRNA was responsible for this peak. We elected to quantify rRNA levels during early development using a molecular hybridization approach, and chose the "dot blot" method for this purpose.

Onset of rRNA synthesis is under developmental regulation

Total embryo RNA recovered from cysts incubated for various periods of time remains approximately constant from the dehydrated gastrula stage until about 18 h of incubation under our conditions (Fig. 2A). This is the "hatching stage," at which time the developing embryo breaks out of its enveloping membranes and begins to actively swim. Total RNA then increases and peaks between 30 - 36 h, then gradually declines in unfed animals, although in some runs the rate of this decline is much less than in the illustrated example. Upwards of 75% of total cell RNA may be rRNA in various other systems, and we suspected that much of the RNA synthesis observed was due to rRNA. To confirm this, aliquots of RNAs isolated from each developmental stage (unfed series) were denatured in glyoxal, immobilized on nitrocellulose, and hybridized to cloned *Xenopus laevis* ^{32}P-labeled (18S + 28S) rRNA genes. It was found (Fig.

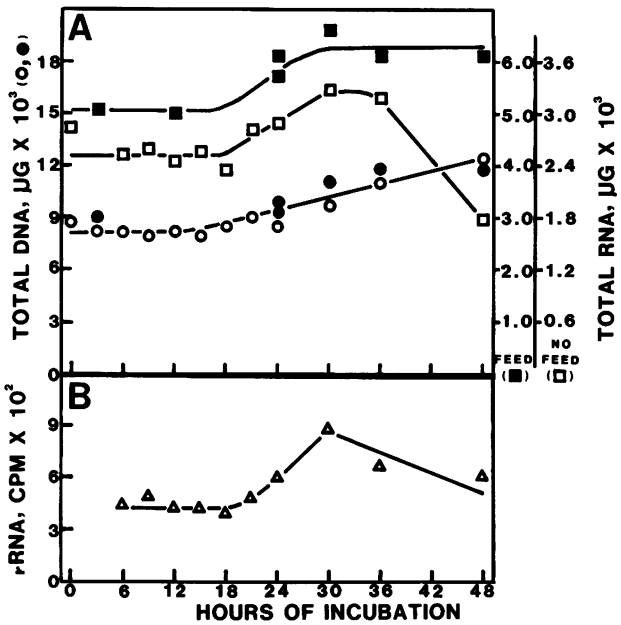

Figure 2. Changing nucleic acid contents during development. In **(A)**, total embryo RNA content is followed at various developmental stages in animals fed starting at 24 h (■) or unfed (□). The rise in RNA content starting at 18 h of incubation, which is the hatching stage, is a stage-specific process. Total embryo DNA content begins to also increase at 18 h of incubation and is consistent whether animals are fed (●) or unfed (○). In **(B)**, relative rRNA content per embryo is followed at various developmental stages in unfed animals, as indicated by dot blot assay. The point of onset and also the peak coincide with total cell RNA profiles shown in **(A)**. different RNA preparations were utilized in the two experiments.

2B) that the rRNA curve is largely superimposable with that of the total embryo RNA of the unfed series employed. Once again, accumulation begins following about 18 h of development and peaks at 30 h. We conclude that the onset of rRNA synthesis is a stage-specific process and is thus under developmental regulation in this system. It has been pointed out (40) that the decline in RNA content per embryo following 36 h of development is a characteristic observed in other systems upon starvation. We consistently observed such a decline in RNA content following 36 h in unfed animals, as have others (39). When animals were well fed, which is not the common practice in studying *Artemia*, total RNA leveled off following 36 h and did not decline (Fig. 2A). Starvation obviously represents an unnatural additional stress factor which could easily lead to altered expression of various physiological and biochemical parameters, so that results obtained in the same or different laboratories could easily differ in unexpected ways if this factor is not controlled.

We determined that each 1 g of dehydrated cysts contains about 387,500 organisms. Results of repeated countings showed that 84% of these cysts were viable after 24 h of incubation. It has been determined that a single encysted embryo contains about 4,000 cells and that this number is constant until hatching of the nauplius larvae at about 18 h, after which the cell

number begins to increase (41). We found that the total DNA content of each 100 ml culture is constant at about 8,100 ug up to about the 18 h hatching stage (Fig. 2A). It can then be calculated that the average individual nucleus should contain about 6.2 pg, a value which agrees exactly with that reported based on chemical analysis of counted, isolated nuclei obtained from 24 h stage nauplii (42). This correspondence indicates that our conditions of homogenization are breaking virtually all cysts at every stage, for we are able to extract all DNA expected. The changing RNA patterns we have observed during development should therefore be indicative of actual trends within embryos and not merely to differential breakage of increasingly more fragile organisms

Total DNA begins to gradually increase in a linear fashion following 18 h of incubation, reflecting its renewed synthesis. The period of renewed onset of rRNA synthesis coincides with that for DNA synthesis. The observed net increase in rRNA content per embryo following 18 h of incubation thus accompanies an increase in cell number during development. A plot of the ratio of total embryo RNA (fed embryos) to DNA on a weight-to-weight basis over the interval 0 - 48 h reveals a line with zero slope (not shown). Since total cell RNA is largely rRNA, this result indicates that the rRNA content per cell is not undergoing any appreciable net change during this interval, from which we conclude that the onset of rRNA synthesis during *Artemia* development is a direct response to the need for additional ribosomes as cell number, accompanied by cell growth, accelerates at hatching.

Site mapping of multiple DNA tracts encoding pre-rRNA 5'-termini

The 8.4 kb XbaI*/EcoRI DNA probe, which contains the entire NTS, was used under conditions of DNA excess for site mapping of DNA tracts encoding pre-rRNA 5'-termini during development. The results of site mapping are shown in Fig. 3, while the locations of the first nine sites relative to the restriction map are diagrammed in Fig. 1. The nucleotide sequences of these sites within that of the intergenic spacer are presented in Fig. 4. At least fourteen sites are recognizable in autoradiographs for RNA samples derived from embryos taken at hatching or somewhat later, i.e. between about 18 - 24 h of incubation. Transcripts mapping further out into the spacer than ca. site 4 are relatively rare prior to hatching and also in ca. 30 - 48 h stage embryos. There is a dramatic increase in utilization of sites located further upstream at about the time of hatching, in addition to a sharp increase in the hybridization signal attributable to sites 3 and 4. Since the hybridizations were carried out under conditions of DNA excess, every pre-rRNA molecule mapping to one of the sites should have had ample opportunity to hybridize with some labeled DNA strand under these conditions. Relative to the 5'-end labeled XbaI position, the sites are located at ca.: -650 (site 1), -850 (site 2), -950 (site 3), -1150 (site 4), -1500 (site 5), -1600 (site 6), -1750 (site 7), -2100 (site 8), -2200 (site 9), -2400 (site 10), -2700 (site 11), -2800 (site 12), -3000 (site 13) and -3500 ("site 14"). Following site 1, these are present in recurring groups showing identical spacing: 2, 3 and 4; 5, 6 and 7; 8, 9 and 10; 11, 12 and 13. This reflects an underlying subrepeat organization in this region within the NTS, as proven by DNA sequencing (discussed in a subsequent subsection). The region around "site 14" is too poorly resolved in our gels to be certain as to its detail. Insofar as the end label for the DNA probe was placed at the XbaI position, which has been precisely localized 160 bp into the 5'-end of the 18S rRNA coding sequence (see below), each of these fourteen pre-rRNA transcript size classes must actually pass directly through the 18S region. There is thus no evidence for the presence of a fail-safe terminator for spacer transcripts in *Artemia*, which differs in this regard from *Xenopus* (43). Early indications were that a fail-safe spacer terminator is also present in *Drosophila* (44). Recently however, detailed reinvestigation of

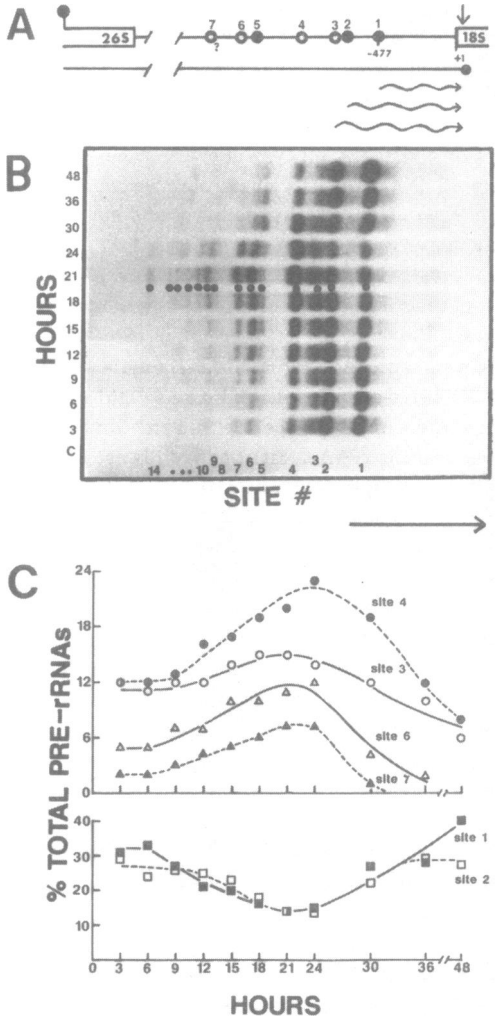

Figure 3. Locations of DNA tracts encoding pre-rRNA 5'-termini during development. In **(A)**, orientation diagram is provided. The 8.4 kb EcoRI/XbaI coding strand fragment, containing the entire NTS, was 5'-end labeled with ^{32}P (—●), denatured and hybridized with pre-rRNAs (wavy lines) isolated from various developmental stages. After S1 nuclease digestion of the resulting DNA*/RNA hybrids, DNA sites (1, 2, 3 . .) encoding pre-rRNA 5'-termini were located relative to the protected labeled XbaI terminus by electrophoretic fractionation. Solid circles represent processing sites and open circles represent promoter sites. The location of site 7 is known but this region has not yet been sequenced, so that its functional character is predicted based upon the recurring pattern and character of the other sites. In **(B)**, autoradiograph is shown for location of sites to which pre-rRNAs isolated from various developmental stages map. Upstream sites become activated around the 18 h hatching stage of development. Lane C is control, in which *E. coli* tRNA was substituted for *Artemia* RNA. The direction of electrophoretic migration is shown by the large arrow.

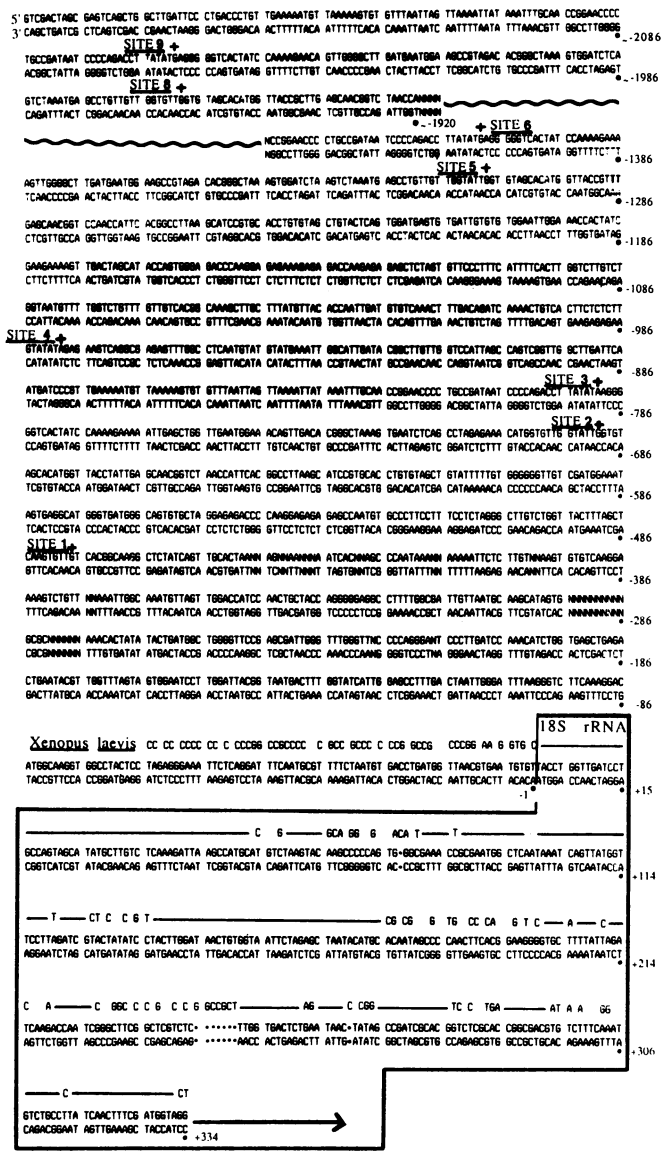

Figure 4. Non-coding strand DNA sequence of *Artemia* 18S rRNA gene/spacer boundary and adjacent upstream spacer control region. Precise experimentally derived locations of DNA nucleotides corresponding to 5'-termini of the first six gene proximal pre-rRNAs, and those inferred for sites 8 and 9, are marked by a +. Region shown by heavy wavy line has not yet been sequenced. Location of nucleotides corresponding to 5'-end of 18S rRNA (box) is in part by alignment with corresponding *Xenopus laevis* sequence (46). Only nucleotides differing between the two species are shown in this region. Corresponding nucleotides are indicated by thick black lines. Gaps introduced to facilitate alignment are shown by dots. Position numbers refer to the *Artemia* sequence relative to the start of the 18S rRNA region.

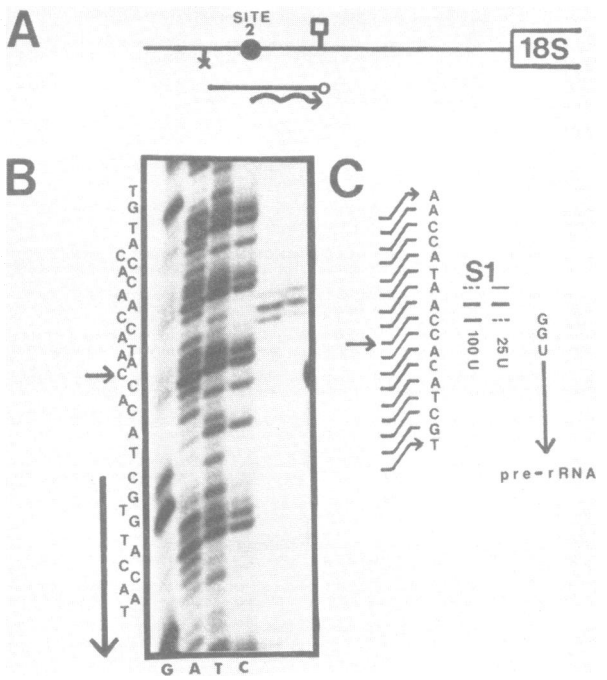

Figure 5. Sequence determination of DNA encoding pre-rRNA 5'-termini. In **(A)**, one aliquot of the single end-labeled (O) 230 bp TaqI*/HpaII coding strand fragment spanning site 2 (see Fig. 1) was subjected to the chemical degradation steps of the sequencing procedure. Another aliquot of this fragment was denatured, hybridized to RNA enriched for pre-rRNAs (wavy line) and aliquots digested with 25 U and 100 U of S1 nuclease. In **(B)**, the two S1 digested hybrids have been run alongside the complete corresponding sequencing ladder, whose autoradiograph is shown. In **(C)**, the DNA nucleotide encoding the first RNA nucleotide at the 5'-terminus of the pre-rRNA mapping to site 2 (short arrow) is read directly from the gel, following a $1^{1/2}$bp mobility correction necessitated (30) by the differing cutting modes of the two techniques employed.

this point (45) has revealed that no operative fail-safe terminator is present in *Drosophila*. Since the probe was labeled at a site internal to the 18S rRNA gene coding sequence, it is surprising that the great excess of 18S rRNA in the total cell RNA used in these hybridizations did not simply bind up all of the probe. Perhaps secondary structure within rRNA in this relatively short region prevented this from occurring.

Both the number of sites encoding pre-rRNA 5'-termini and the activity per site increase near the time of hatching, which is the developmental stage during which total embryo rRNA has been observed to dramatically increase (Fig. 2B). At the 21 h stage of incubation, for example, some 40% of all pre-rRNAs map upstream from site 4 in *Artemia*, as indicated by densitometric scanning of the autoradiographs. We conclude that net synthesis of rRNA is occurring at the hatching stage. It is later shown that activation of multiple upstream transcription initiation sites plays a major role in the observed increase in net synthesis within these somatic cells.

Two structural classes of DNA tracts encoding pre-rRNA 5'-termini

Except for site 7, all of the first nine sites have been sequenced. The precise nucleotides encoding pre-rRNA 5'-termini have been identified by combined S1 nuclease protection and DNA sequencing for sites 1 - 6.. Although site 7 and also sites 10 - 13 have not as yet been sequenced, nor the S1 protection approach applied to sites 8 and 9, their structures may be inferred from the regular pattern of all sites observed and also comparison to sites now characterized. The results from a typical analysis are shown in Fig. 5. Preliminary examination of the DNA sequences at the first four sites revealed interesting conserved features when alignment was made between sites 1, 2 and between sites 3, 4. These sequence combinations were therefore examined in detail, and revealed two different classes of DNA tracts encoding pre-rRNA 5'-termini. As discussed in a subsequent subsection, all sites beyond number 4 are recurrent and can be explained by the subrepeat structure of the intergenic spacer.

A. Pre-rRNA 5'-processing sites

A sequence of 156 positions surrounding and upstream of sites 1 and 2 was compared following alignment (Fig. 6A) to search for evidence of conserved nucleotide tracts which may represent important functional elements. Alignment was facilitated by precise identification of the DNA nucleotide encoding the 5'-end of each pre-rRNA size class. Except for an interesting short tract of seven nucleotides whose borders are clearly demarcated and which is located near the beginning of these two transcripts, the two sequences are quite different. The regions compared are only 26% identical, and randomly correlate. Each 156 nucleotide long sequence contains numerous palindromes and direct repeats, which are not positionally correlated. The short tract of seven conserved nucleotides, which also occurs at sites 5 and 8, has the non-coding strand consensus sequence: 5' ...Pu-G-T-Pu-T-T-G... 3', and terminates with what is the initial deoxyguanidyl residue at the 5'-end of the corresponding transcripts. Five out of seven positions within the conserved tract are identical and the remaining two differ by only conservative substitutions. A computer search revealed that this short conserved element does not occur anywhere else within the sequenced spacer, with the interesting exception of one tract having the very similar sequence: 5' ...G-T-G-T-T-A... 3' which directly spans the junction of the spacer/18S rRNA gene interface between positions -4 and +2 (Fig. 4).

It is commonly believed that evidence of nucleotide sequence conservation reflects the operation of strong evolutionary selective pressure and thereby implies the existence of a function for the conserved sequence. We therefore anticipate that this conserved element, located at DNA sites encoding the 5'-ends of the corresponding pre-rRNAs, may have an interesting function. This conserved seven nucleotide long sequence occurs at sites 1, 2, 5 and 8, while a very similar sequence appears at the 18S rRNA gene/spacer boundary, which is a 5'-processing site by definition. The similar structure of these nucleotide tracts strongly implies a common functional significance. We propose that these conserved elements represent 5'-processing site consensus sequence domains. If so, their significance presumably lies in their recognition by a processing enzyme(s). Sequence conserved pre-rRNA processing sites have been described in other species. The 5'- and 3'-cleavage sites for *Drosophila melanogaster* pre-rRNAs adjacent to the 18S rRNA region have been identified (47). A seven nucleotide long sequence surrounds both cleavage sites and is identical at both, although the processing cut site differs by one nucleotide at the two locations. Interestingly, this conserved processing site sequence is similar to the tract identified in *Artemia*, insofar as the sequence 5' ...T-A-T-T... 3' occurs within the processing sites of both arthropod species. In addition, this

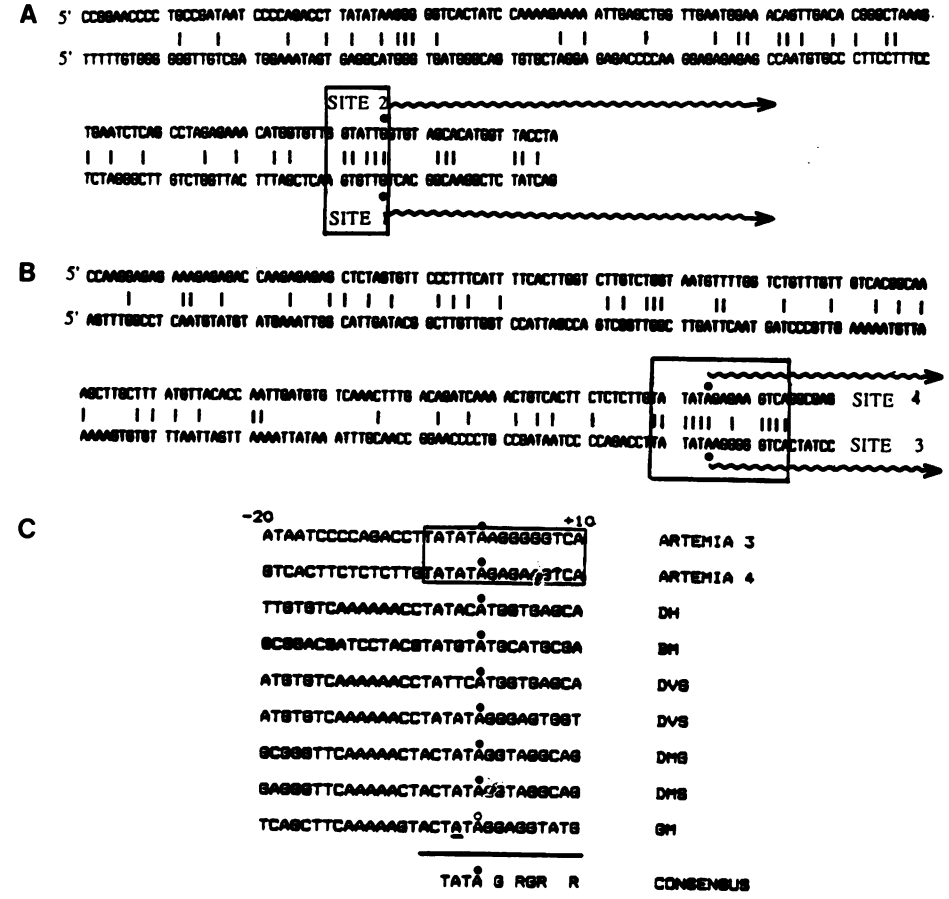

Figure 6. Alignment comparisons of non-coding DNA tracts corresponding to sequences surrounding arthropod rDNA transcription initiation sites and pre-rRNA 5'-termini. In **(A)**, *Artemia* DNAs mapping to sites 1 and 2 and their upstream regions are compared: box shows region of extensive similarity and vertical lines show corresponding nucleotides which are identical. In **(B)**, *Artemia* DNAs mapping to sites 3 and 4 and their upstream regions are compared in the same manner. In **(C)**, the *Artemia* site 3 and 4 consensus regions are compared with those of various other arthropods flanking their transcription initiation site: *Drosophila hydei*, DH (ref. 51); moth *Bombyx mori*, BM (ref. 48); *Drosophila virilis* gene promoter, DVG (ref. 44); *Drosophila virilis* spacer promoter, DVS (ref. 44); *Drosophila melanogaster* gene promoter, DMG (ref. 52); *Drosophila melanogaster* spacer promoter, DMS (ref. 53); and tsetse fly *Glossina morsitans*, GM (ref. 51). The initiation point of transcription as experimentally determined is indicated by a black dot, which is position +1 for all entries except tsetse fly (GM), which has been numbered relative to the alignment and whose experimentally determined point of initiation is underlined. Arthropod consensus sequence on bottom line for region corresponding to *Artemia* positions -5 to +10 only compares nine sequences and accepts a position as consensus if it is in common with seven or more. Purine = R; pyrimidine = Y.

identical sequence has been identified at a processing site within the ETS in the moth *Bombyx mori* (48) In the yeast *Saccharomyces carlsbergensis*, the comparative sequence analysis approach was successfully used to identify a consensus sequence thirteen nucleotides in length at the 5'-ends of the 26S. 17S and 5.8S rRNAs. A consensus sequence six nucleotides in length has also been identified in this species at the 3'-ends of 18S (precursor to 17S), 17S, 7S (precursor of 5.8S), 5.8S and 26S rRNAs (49, 50). These structures have been proposed to be involved in the precise recognition of pre-rRNA sites by processing enzymes. These authors present a model for processing pre-rRNA which involves base pairing between sequences at the 5'- and 3'-termini of various regions within the primary transcript to generate a similar configuration. We do not yet have sufficient 3'-terminal sequence data for the corresponding regions in *Artemia* to assess the possible applicability of this model to our system.

Our results provide hints that availability of processing enzymes may be under developmental regulation in the *Artemia* system and thus play a role in control of rDNA gene expression. During pre-hatching and also late larval stages, relatively little rRNA synthesis is occurring. Despite this, the hybridization signals attributable to sites 1 and 2, which we believe are 5'-processing sites, are surprisingly strong at these stages (Fig. 3). Comparison of these signal intensities to those at the corresponding sites during the 21 - 24 h stages, when rRNA synthesis is near its peak, shows a marked loss of signal at these two processing sites. This shows that the processing event at the spacer/18S rRNA junction is relatively rapid at the hatching stage, which is hardly surprising. Absence of processing enzyme activity prior to the hatching stage, followed by its activation (or synthesis) during hatching and subsequent diminution of activity during late larval stages could readily explain these results. These considerations suggest to us that some stable pre-rRNAs may actually be stored in the embryo until the hatching stage, at which time renewed activity of the processing enzyme(s) drives the mature rRNAs into the cellular pool along with those molecules arising due to renewed synthesis. These stable precursors are clearly visible in the autoradiographs during early developmental stages (Fig. 3) and are probably produced during early development prior to the desiccation step which occurs at about the gastrula stage.

B. Pre-rRNA transcription initiation sites

A sequence of 190 positions surrounding and upstream of sites 3 and 4 was compared following alignment (Fig. 6B) to search for evidence of conserved nucleotide tracts which may represent additional important functional domains worthy of further study. Alignment was again simplified by precise identification of the DNA nucleotide encoding the 5'-end of each pre-rRNA size class. A single short sequence conserved tract with clearly demarcated borders was identified surrounding the DNA nucleotide encoding the 5'-end of both transcripts. The compared sequences are otherwise quite different, and show only a 26% identity, which is a random correlation. Each of these two 190 nucleotide long sequences contains numerous palindromes and direct repeats, but the positions of these features do not correlate between the two sequences and their significance is therefore problematic. The conserved tract, which also occurs at sites 6 and 9, is fifteen nucleotides in length and its non-coding strand has the consensus sequence: 5' ...T-A-T-A-T-Pu-Pu-Pu-G-Pu-Pu-G-T-C-A... 3'. The sixth nucleotide in the sequence, a purine, occupies the initial position at the 5'-end of the corresponding transcripts. Ten out of fifteen positions within this conserved tract are identical in this comparison and the remaining five differ by only conservative substitutions. This conserved element of fifteen nucleotides occurs only at sites 3, 4, 6 and 9, and does not occur elsewhere in the sequenced spacer as shown by computer search (37). These conserved elements, located at

DNA sites encoding the 5'-ends of these several pre-rRNAs may share a common functional significance by virtue of their common structure

What kinds of important functions might conceivably be played by sequence conserved tracts located at DNA positions encoding the 5'-termini of pre-rRNA molecules synthesized *in vivo*? The two most obvious such functions would be for specific recognition by protein molecules associated with transcription initiation (19, 54) and subsequent processing (55, 49) of the primary pre-rRNAs. We have argued that positions 1 and 2 may serve in 5'-processing. Sites 3 and 4 are located relatively far out in the spacer and these two belong to the same structural class. Although it is not impossible but that they represent yet additional 5'-processing sites and the transcription initiation site is located quite far upstream this appears unlikely in that a) their consensus sequence differs from the site 1, 2 class and b) the recurring pattern of the two site classes thus far identified extends out into the spacer for about 3350 ntp, whereas sites for transcription initiation in arthropods are generally located around 800 - 900 ntp upstream of the 18S rRNA gene (44, 48, 51). Site 3 is located about 800 ntp upstream from the 18S rRNA coding region. Comparison of published arthropod rDNA transcription initiation site sequences with the *Artemia* site 3, 4 consensus element (Fig. 6C) reveals some shared features. The sequence T-A-T-A is held in common at this location. This identical sequence feature has also recently been noted at the rDNA transcription initiation site in a comparison of some twelve quite diverse eukaryotic species (48). Since this site class lies external to the ones which we have argued represent 5'-processing sites, and this site class shares common sequence features with the initiation site of many other animal species, we propose that it represents the *Artemia* rDNA transcription initiation site. If so, this consensus element shows an interesting similarity in position and size to the minimal proximal promoter element in the frog *Xenopus*. In three closely related *Xenopus* species, the sequence from -11 to +4 is identical (13), and this is very nearly the case for the region -1 to +18 in mouse, rat and human (16). In the closest approach to an *in vivo* assay yet attained, injection of truncated cloned rDNA fragments into *Xenopus laevis* oocytes identified the limits of the minimal proximal promoter in this species to a relatively short tract extending from ca. -7 to ca. +6: a region of only some thirteen nucleotide positions spanning the transcription initiation site (19). It is striking that this location identified by the assay of truncated sequence tracts in *Xenopus* by a pseudo *in vivo* method corresponds almost exactly, both in location and length, to the conserved fifteen nucleotides of sequence we have identified *in vivo* in *Artemia* by the comparative sequence analysis method: a tract with sharply demarcated borders which extends from exactly -5 to +10. It is surprising that no additional upstream conserved sequence elements are evident (Fig. 6B), in that the proximal promoter domain for Pol I is generally believed to span roughly the region from +10 to -45 or so (9). Perhaps features other than sequence are important here. Finally, it has been pointed out that two of the few seemingly universal sequence features present in rDNA promoters are the presence of a pyrimidine (usually a T) at position -1 with respect to the initiation nucleotide and (at least in mammals) the presence of a deoxyguanidyl residue at position -16. The proposed *Artemia* initiation site sequences obey the position -1 rule for both sites 3 and 4. There are however several exceptions to the -16 "rule" which are shown in the Fig. 6C alignment and which have been noted elsewhere for at least three additional species (57).

The functional significance of multiple spacer promoters in *Artemia* may lie in the increased efficiency of trapping components essential for construction of initiation complex(es). This function would confer some selective advantage, thereby providing for their

maintenance, as has been proposed in other systems (44, 58). Interaction between protein transcription factors and relatively short DNA sequence promoter elements constitutes a stable transcription complex which is subsequently recognized by RNA polymerase I (59, 60). The possible role of changing transcription factor protein concentration in controlling rRNA gene expression during *Artemia* development has yet to be explored. One positionally and sequence conserved short DNA element has now been identified within the putative *Artemia* rDNA promoter. Although the functional significance of this sequence conserved element within the *Artemia* rDNA promoter is yet to be proven, it would appear reasonable to suggest that absence of renewed rRNA synthesis in the presence of demonstrated RNA polymerase I activity prior to hatching (61) could arise because such factors have not as yet been synthesized. When such factors become relatively abundant, as during hatching, utilization of both gene and spacer promoters increases proportionately (Fig. 3). However, how does one explain the relative silence of identical upstream promoter sites homologous to sites 3 and 4 at stages when the latter sites are apparently active, for example at stages 36 - 48 h? Also, what explains the activation of additional upstream promoters during the 21 - 24 h peak activity period of early development? The subrepeat pattern of spacer organization proven by DNA sequencing (below) shows or implies a very high degree of sequence conservation between homologous initiation sites, so that we believe neither concentrations of critical protein factors nor sequence divergence alone can adequately explain these questions. Conformational changes within the rDNA chromatin at upstream sites may play a key role here. It has been shown, for example, that the degree of DNA supercoiling plays a critical role in the level of activity of cloned *Xenopus* rRNA genes injected into oocytes. Spacer promoters are scarcely transcribed at all within oocytes *in vivo*, even though there is a very high level of transcription factor protein present in these cells, but when spacer promoters on plasmids are injected into oocytes they are as active as injected gene promoters (62). Spacer promoters on injected plasmids can be selectively inactivated by treatment with intercalating agents such as ethidium bromide, raising the possibility that differential supercoiling could explain differences in spacer transcription activity in different situations (63). Upstream spacer promoters are transcribed at approximately the same level as the gene promoter in the early *Xenopus* embryo (S.C. Pruitt and R.H. Reeder, manuscript in preparation). This effect has also been reported several times in *Xenopus* tissue culture cells (e.g. ref. 64), which like early embryo cells are rapidly growing and dividing. Activation of upstream spacer rDNA promoters plays a major role during *Artemia* development at stages when there is a relatively sudden need for quantities of additional rRNA in rapidly growing and dividing somatic cells, so that this may be an important somatic counterpart to the amplification of rRNA genes utilized during oogenesis in germ line cells of many animal species.

Subrepeat pattern of organization within the rDNA multiple promoter region: a model for the evolutionary origin of the intergenic spacer in Artemia

The origins of the multiple putative transcription initiation and 5'-processing sites we have identified within the *Artemia* intergenic spacer are readily explained when one considers the sequence information now compiled for this region. The complete sequence we have obtained for 1778 nucleotides surrounding the boundary between the 18S rRNA coding region and the adjacent upstream spacer, in addition to 266 nucleotides surrounding sites 8 and 9, is shown in Fig. 4. Precise recognition of the first nucleotide within the 18S rRNA coding sequence was determined in part by alignment of our *Artemia* sequence with the corresponding conserved 18S region from *Xenopus laevis* (46) and has been confirmed by

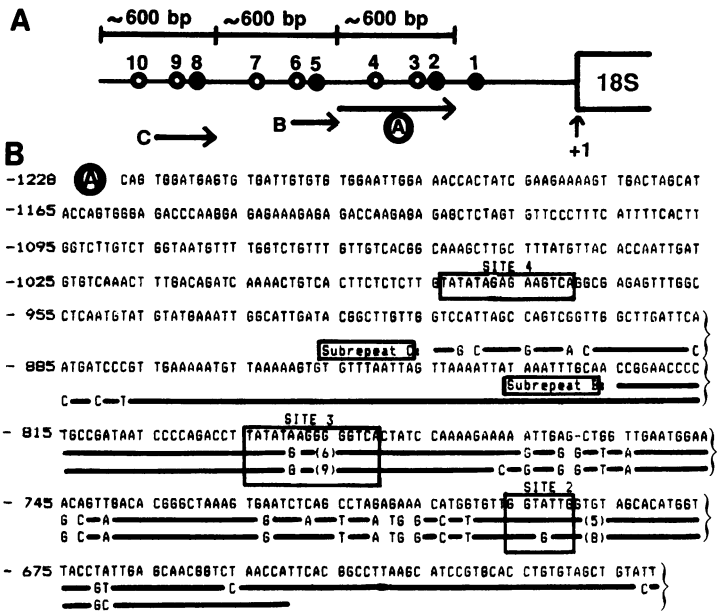

Figure 7. Alignment comparison of spacer DNA sequence tracts (heavy arrows) within ca. 600 bp subrepeats. In **(A)**, orientation diagram is provided, showing positions of sites to which the first ten gene proximal pre-rRNA 5'-termini map. Solid circles represent processing sites and open circles represent promoter sites. Location of sites 7 and 10 are known, but these have not yet been sequenced, so that their functional character is predicted based upon the recurring pattern and character of the other sites. In **(B)**, the complete non-coding strand DNA sequence for subrepeat A (upper continuous line) is aligned with partial sequences obtained for subrepeats B and C (lower two lines). Only nucleotides which differ from subrepeat A are shown in the two lower lines, and gaps introduced to facilitate alignment are indicated by dashes. Corresponding nucleotides are indicated by thick black lines. Conserved consensus sequence boxes corresponding to the various pre-rRNA map sites appear to have arisen due to ancestral duplication events which generated the ca. 600 bp subrepeats.

primer extension (7). The first 210 nucleotides within the ETS have been reported by others (7). Our results both confirm and extend this for an additional 1500 nucleotides. Combining the results of the independent sequencings by our two laboratories, the 1778 nucleotide positions illustrated in Fig. 4 out as far as site 6 have either been confirmed by reading both complementary strands or by independently sequencing different clones by our two different laboratories over about 93% of the total length. Except for the region surrounding site 7, which has not yet been attempted, regions of sequence uncertainty which remain are largely due to compression effects observed in sequencing gels owing to extensive secondary structure formation in the spacer.

Computer-assisted self-homology matrix analysis and related methods (37) were utilized as first steps in alignment of the sequences surrounding the several sites thus far characterized, followed by manual sequence comparisons. It is readily apparent (Fig. 7) that much of the NTS thus far sequenced consists of a series of tandem subrepeats, the region of repetition

commencing 134 nucleotides upstream of site 1, which is itself located some 477 nucleotides from the 18S rRNA gene coding region. Subrepeat A, which has been completely sequenced, is 617 nucleotides in length and contains a single HpaII restriction site. This explains the ca. 600 bp HpaII periodicity observed upon restriction mapping of this region of the NTS (Fig. 1), and the similar periodicity recently reported for other restriction enzymes in this region (65). Of the 215 nucleotides sequenced within subrepeat B, 90% homology to subrepeat A is observed. This is virtually identical to the degree of similarity noted between the sequenced region with subrepeats A and C, where 265 nucleotides in C show an 89% identity to A. Interestingly, comparison of the sequenced region common to subrepeats B and C shows that for the 23 positions differing from A, 18 are identical in both B and C. Subrepeats B and C are therefore more closely related to one another than either is to subrepeat A, and their sequenced common region is in fact 97% identical over a span of 175 positions compared. We do not know the significance of this feature, if any. It is clear from this comparison that the origins of the sites located upstream of site 4 can readily be explained as having arisen due to a series of duplications within the block containing sites 2, 3 and 4, presumably via saltatory replication events over a relatively short time span. Multiple spacer promoters have been described in other species. In *Xenopus*, the intergenic spacer is constructed from complex repetitive regions including subrepeats, which contain multiple copies of the 5'-end of the ETS, i.e. the promoter region (17, 30, 43). Occasional transcripts have been described as originating from these upstream initiation sites as short "prelude complexes" observable by electron microscopy (66, 67). Such transcripts however fail to proceed across the gene promoter, owing to a fail-safe terminator located upstream from the gene promoter (30, 43). In *Drosophila* species, upstream ca. 0.24 kb subrepeats comprising part of the NTS contain imperfect copies of the 5'-end of the ETS or promoter region (18). These spacer promoters have been shown to produce relatively infrequent transcripts *in vitro* (53) and also *in vivo* (44). In *Drosophila*, however, it has recently been reported (45) that a fail-safe terminator function is not present within the spacer, contrary to previous reports (44). As previously stated, there is also no evidence for a fail-safe terminator for *Artemia* spacer transcripts. If such were present, we would not have been able to observe transcription from the spacer promoters using the probe which was selected.

Each *Artemia* subrepeat presently contains two putative transcription initiation sites, in addition to one 5'-processing site, which raises the question as to the origin of the ancestral subrepeat itself. During the computer-assisted self-homology matrix analysis of the NTS sequence, additional sequence matches were observed which may bear on the origin of the ancestral ca. 600 bp subrepeat. A stretch of about 110 nucleotides surrounding site 1 ... a region located outside and just downstream of the first subrepeat... is 72% identical to a region within subrepeat A located just upstream of site 4 (Fig. 8). This alignment was improved by insertion of 10 gaps, and would scarcely have been noticed had these gaps been omitted. In the alignment, the site 1 processing consensus sequence box lies opposite a very similar sequence which may have once served a similar function, prior to the accumulation of base changes. We designate such sequence modified tracts M sites, and note that this one occurs about 100 bp upstream of site 4. Strict alignment of 190 nucleotides surrounding and upstream of sites 3 and 4, with no gaps inserted to facilitate alignment, showed only a 26% identity, as previously noted. However, it is possible to demonstrate a nearly 50% identity between nucleotides surrounding these sites and the upstream ca. 100 positions when 8 gaps are inserted (not shown). Sites 3 and 4 presumably serve similar functions, and the observed similarity in

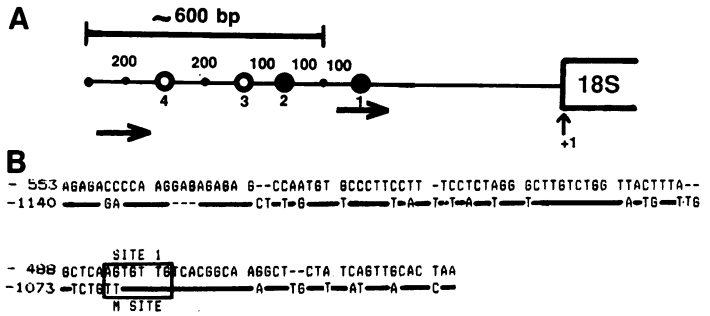

Figure 8. Alignment comparison of spacer DNA sequence tracts (heavy arrows) within ca. 600 bp subrepeats and region surrounding site 1. In **(A)**, orientation diagram is provided, showing promoter sites (open circles) and processing sites (solid circles). First or gene proximal subrepeat is diagrammed to emphasize the spacing between adjacent sites and spacer boundaries, which occurs in pattern (200, 200, 100, 100) bp within each subrepeat. In **(B)**, the non-coding strand DNA sequence surrounding site 1 (upper continuous line) is aligned with sequence downstream from site 4 within ca. 600 bp subrepeat (lower line). Only nucleotides which differ in the two sequences are shown in the lower line, and gaps introduced to facilitate alignment are indicated by dashes. Corresponding nucleotides are indicated by thick black lines. Conserved consensus sequence box corresponding to site 1 aligns with a (non-functional) proposed sequence-modified "M-site" processing site. Hypothetical original positions of such M-sites in ancestral ca. 600 bp subrepeat are shown in **(A)** by black dots.

surrounding nucleotide sequence implies ancestral homology. Site 4 is in fact a divergent form of site 3, even though both function in initiation of transcription, and we propose that site 4 is an example of an "M site." Finally, there is an interesting pattern of site positions within each subrepeat, illustrated in Fig. 8, which we believe provides an additional clue as to the origins of the ancestral ca. 600 bp subrepeat. The approximate spacing within each subrepeat appears to be (200-200-100-100) bp. Whenever adjacent sites are separated by the minimum distance, which is ca. 100 bp, one site initiates and the other processes. Whenever adjacent sites are separated by twice the minimum distance, or 200 bp, both sites have the same function.

Taking all of these observations into consideration, we suggest that the ancestral ca. 600 bp unit consisted of an alternating pattern of initiation sites and processing sites, separated from one another by ca. 100 bp intervals. Subsequent divergence led to the random loss of some site functions. The ancestral spacing has largely persisted. The original ca. 600 bp unit probably arose by duplications of an ancestral ca. 200 bp element containing only one start and one 5'-processing site. The proposed chronological sequence of events in the evolutionary origin of this part of the NTS is supported by the levels of homology actually observed: the extent of homology observed between elements within a subrepeat, ca. 50-70%, is less than that observed between different subrepeats, ca. 90%, as would be predicted in such an interrupted series of events. These considerations have been organized into a proposed model to account for some of the steps leading to the evolutionary origin of the region of the intergenic spacer containing the rDNA control elements in *Artemia* (Fig. 9). A somewhat similar model for the origin of the more complex *Xenopus* spacer, which is also organized into subrepeats in the region of the multiple rDNA promoters, has previously been proposed on the basis of sequence comparisons (17).

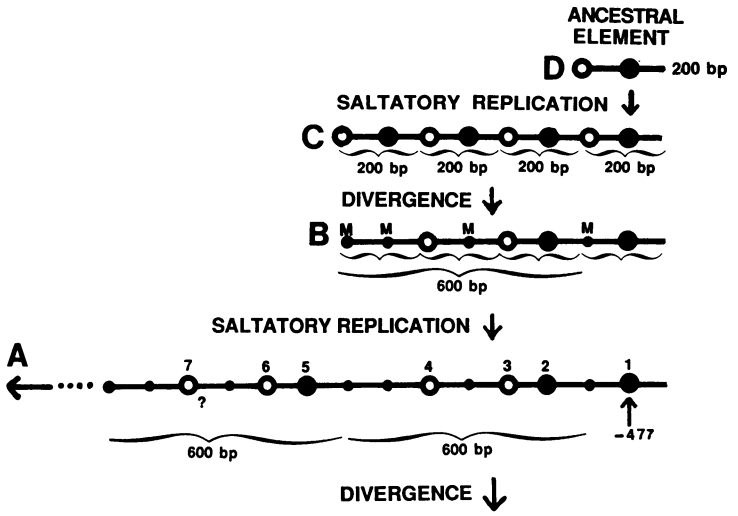

Figure 9. Model for evolutionary history of *Artemia* rDNA control region within the intergenic spacer. In **(A)**, the organization of the present-day spacer shows several regions corresponding to promoters (open circles) and 5'-processing sites (solid circles). Site 1 is located at position -477 with respect to the start of the 18S rRNA coding sequence. The ca. 600 bp subrepeats have very similar, but not identical sequences and arose by saltatory replication from an ancestral ca. 600 bp sequence **(B)**. This ancestral ca. 600 bp sequence arose by an earlier saltatory replication of an ca. 200 bp element, to produce an originally alternating array of start and processing sites **(C)**, some of which have since accumulated base substitutions and thereby lost their original function to become "M-sites." The original spacer ca. 200 bp element **(D)** contained one promoter and one 5'-processing site.

Concluding remarks

This study has identified a pattern of developmentally regulated gene expression correlated with what appears to be a developmentally regulated pattern of pre-rRNA processing events in *Artemia*. A major transcriptional role has been identified for multiple putative upstream "nontranscribed" spacer promoters during early embryogenesis *in vivo*. As is the case in *Drosophila*, there is no evidence for the presence of any fail-safe spacer terminator in *Artemia*. Further work will be required to establish the tentative assignments proposed for the two observed sequence classes of DNA tracts encoding pre-rRNA 5'-termini beyond any reasonable doubt. Verification that primary pre-rRNA transcripts initiate at sites 3 and 4 in *Artemia* should be attainable from projected *in vitro* 5'-capping experiments utilizing vaccinia virus capping enzyme. This enzyme is known to cap polyphosphate termini (68) and is predicted to cap transcripts mapping to sites 3 and 4 but not those mapping to sites 1 and 2. The probable evolutionary history of DNA sequences responsible for these structural characteristics has been traced. Several interesting questions have been raised for future investigation. It should now be possible to obtain a reasonably detailed molecular understanding of the factors underlying these observations, using biochemical, molecular cloning and *in vitro* genetic techniques. Availability of synchronously developing embryos

in virtually unlimited quantity at low cost should be a great benefit in the future exploration and elucidation of this promising new experimental system for the study of control of rRNA gene expression during development.

ACKNOWLEDGMENTS

This research was supported, in part, by Miami University Faculty Research grants to J.C.V. and by Graduate Research funds to H.T.K. and K.A.F. Components of this research were submitted in partial fulfillment of the requirements for the degrees Doctor of Philosophy (H.T.K.) and Master of Science (K.A.F.).

Present addresses: [1]Hipple Cancer Research Center, Dayton, OH 45439 and [2]School of Medicine, Wright State University, Dayton, OH 45435, USA
*To whom correspondence should be addressed

REFERENCES

1 Andrews, M.T., Vaughn, J.C., Perry, B.A. and Bagshaw, J.C. (1987) Gene 51, 61-67
2 Vaughn, J.C., Whitman, D.J., Bagshaw, J.C. and Helder, J.C. (1982) Biochim. Biophys. Acta 697, 156-161
3 Roberts, M.P. and Vaughn, J.C. (1982) Biochim. Biophys. Acta 697, 148-155
4 Diels, L., DeBaere, R., Vandenberghe, A. and DeWachter, R. (1981) Nucl. Acids Res. 9, 5141-5144
5 Ursi, D., Vandenberghe, A. and DeWachter, R. (1982) Nucl. Acids Res. 10, 3517-3530
6 Vaughn, J.C., Sperbeck, S.J. and Hughes, M.J. (1984) Biochim. Biophys. Acta 783, 144-151
7 Nelles, L., Fang, B.-L., Volckaert, G., Vandenberghe, A. and DeWachter, R. (1984) Nucl. Acids Res. 12, 8749-8768
8 Treco, D., Brownell, E. and Arnheim, N. (1982) in The Cell Nucleus, rDNA, Pt. C, Busch, H. and Rothblum, L., eds., Vol. 12, pp.101-126, Academic Press, New York
9 Sollner-Webb, B. and Tower, J. (1986) Annu. Rev. Biochem. 55, 801-830
10 Pribnow, D. (1975) Proc. Natl. Acad. Sci. U.S.A 72, 784-788
11 Goldberg, M. (1979) Ph.D. Thesis, Stanford Univ.
12 Grummt, I., Roth, E. and Paule, M. (1982) Nature 296, 173-174
13 Bach, R., Allet, B. and Crippa, M. (1981) Nucl. Acids Res. 9, 5311-5330
14 Financsek, I., Mizumoto, K. and Muramatsu, M. (1982) Gene 18, 115-122
15 Verbeet, M., Klootwijk, J., van Heerikhuizen, H., Fontijn, R., Vreugdenhil, E. and Planta, R. (1983) Gene 23, 53-63
16 Financsek, I., Mizumoto, K., Mishima, Y. and Muramatsu, M. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 3092-3096
17 Boseley, P., Moss, T., Machler, M., Portmann, R. and Birnstiel, M. (1979) Cell 17, 19-31
18 Coen, E.S. and Dover, G.A. (1982) Nucl. Acids Res. 10, 7017-7026
19 Sollner-Webb, B., Wilkinson, J., Roan, J. and Reeder, R. (1983) Cell 35, 199-206
20 Moens, L. and Kondo, M. (1976) Dev. Biol. 49, 457-469
21 Warner, A.H., MacRae, T.H. and Wahba, A.J. (1979) Methods Enzymol. 60, 298-311
22 Clark, L.S. and Bowen, S.T. (1976) J. Hered. 67, 385-388
23 Nakagawa, H. and Cohen, P.P. (1967) J. Biol. Chem. 242, 642-649
24 Dische, Z. (1955) in The Nucleic Acids, Chargaff, E. and Davidson, J.N., eds., Vol. 1, pp.285-305, Academic Press, New York
25 Thomas, P. (1983) Methods Enzymol. 100, 255-266
26 Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. and Efstratiadis, A. (1978) Cell 15, 687-701
27 Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237-251
28 Keating, K.A. (1985) Thesis, Miami Univ., Oxford, OH
29 Berk, A.J. and Sharp, P.A. (1977) Cell 12, 721-732

30 Sollner-Webb, B. and Reeder, R. (1979) Cell 18, 485-499
31 McDonell, M.W., Simon, M.N. and Studier, F.W. (1977) J. Mol. Biol. 110, 119-146
32 Southern, E.M (1979) Methods Enzymol. 68, 152-182
33 Southern, E.M. (1975) J. Mol. Biol. 98, 503-517
34 Haynes, J.R., Rosteck, P., Schon, E.A., Gallagher, P.M., Burks, D.J., Smith, K. and Lingrel, J.B. (1980) J. Biol. Chem. 255, 6355-6367
35 Donelson, J.E. and Wu, R. (1972) J. Biol. Chem. 247, 4654-4660
36 Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560
37 Gross, R.H. (1986) Nucl. Acids Res. 14, 591-596
38 Hori, H. and Osawa, S. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 381-385
39 McClean, D.K. and Warner, A.H. (1971) Dev. Biol. 24, 88-105
40 Hentschel, C.C. and Tata, J.R. (1976) Trends Biochem. Sci. 1, 97-100
41 Nakanishi, Y.H., Iwasaki, T., Okigaki, T. and Kato, H. (1962) Annot. Zool. Jap. 35, 223-228
42 Bagshaw, J.C., Bernstein, R.S. and Bond, B.H. (1978) Differentiation 10, 13-21
43 Moss, T. (1983) Nature 302, 223-228
44 Murtif, V.L. and Rae, P.M.M. (1985) Nucl. Acids Res. 13, 3221-3239
45 Tautz, D. and Dover, G.A. (1986) EMBO J. 5, 1267-1273
46 Maden, B.E.H., Moss, M. and Salim, M. (1982) Nucl. Acids Res. 10, 2387-2398
47 Simeone, A. and Boncinelli, E. (1984) FEBS Lett. 167, 249-253
48 Fujiwara, H. and Ishikawa, H. (1987) Nucl. Acids Res. 15, 1245-1258
49 Veldman, G., Brand, R., Klootwijk, J. and Planta, R. (1980) Nucl Acids Res 8, 2907-2920
50 Veldman, G., Klootwijk, J., van Heerikhuizen, H. and Planta, R. (1981) Nucl. Acids Res. 9, 4847-4862
51 Cross, N. and Dover, G.A. (1987) Nucl. Acids Res. 15, 15-30
52 Long, E., Rebbert, M. and Dawid, I. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 1513-1517
53 Kohorn, B.D. and Rae, P.M.M. (1982) Nucl. Acids Res. 10, 6879-6886
54 Clos, J., Normann, A., Ohrlein, A. and Grummt, I. (1986) Nucl. Acids Res. 14, 7581-7595
55 Miller, K. and Sollner-Webb, B. (1981) Cell 27, 165-174
56 Skinner, J., Ohrlein, A. and Grummt, I. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 2137-2141
57 Sommerville, J. (1984) Nature 310, 189-190
58 Reeder, R., Roan, J. and Dunaway, M. (1983) Cell 35, 449-456
59 Cizewski, V. and Sollner-Webb, B. (1983) Nucl. Acids Res. 11, 7043-7056
60 Wandelt, C. and Grummt, I. (1983) Nucl. Acids Res. 11, 3795-3809
61 Cano, A., Cruces, J., Estepa, I., Gallego, M.E., Sillero, M.A., Heredia, C.F., Llorente, P., Olalla, A., Osuna, C., Pestana, A., Renart, J., Ruiz, A., Sastre, L., Sebastian, J and Sillero, A. (1979) in Biochemistry of Artemia Development, Bagshaw, J.C. and Warner, A.H., eds., pp.150-166, University Microfilms Intl., Ann Arbor
62 Morgan, G.T., Roan, J.G., Bakken, A.H. and Reeder, R.H. (1984) Nucl. Acids Res. 12, 6043-6052
63 Pruitt, S.C. and Reeder, R.H. (1984) Mol. Cell. Biol. 4, 2851-2857
64 Rungger, D., Achermann, H and Crippa, M. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 3957-3961
65 Gallego, M.E., Diaz-Guerra, M., Cruces, J., Sebastian, J. and Renart, J. (1986) Gene 48, 175-182
66 Scheer, U., Trendelenburg, M.F., Krohne, G. and Franke, W.W. (1977) Chromosoma 60, 147-167
67 Trendelenburg, M.F. (1981) Biol. Cell 42, 1-12
68 Monroy, G., Spencer, E. and Hurwitz, J. (1978) J. Biol. Chem. 253, 4490-4498