Biochem. J. (1984) **220**, 105–116
Printed in Great Britain

Localization and structure of endonuclease cleavage sites involved in the processing of the rat 32S precursor to ribosomal RNA

Krassimira V. HADJIOLOVA,* Oleg I. GEORGIEV,* Valery V. NOSIKOV† and Asen A. HADJIOLOV*

**Department of Molecular Genetics, Institute of Molecular Biology, Bulgarian Academy of Sciences, Sofia, Bulgaria, and †Institute of Molecular Biology, Academy of Sciences of the U.S.S.R., Moscow, U.S.S.R.*

(*Received 21 November 1983/Accepted 2 February 1984*)

The initial endonuclease cleavage site in 32S pre-rRNA (precursor to rRNA) is located within the rat rDNA sequence by S_1-nuclease protection mapping of purified nucleolar 28S rRNA and 12S pre-rRNA. The heterogeneous 5'- and 3'-termini of these rRNA abut and map within two CTC motifs in tS_i2 (internal transcribed spacer 2) located at 50–65 and 4–20 base-pairs upstream from the homogeneous 5'-end of the 28S rRNA gene. These results show that multiple endonuclease cleavages occur at CUC sites in tS_i2 to generate 28S rRNA and 12S pre-rRNA with heterogeneous 5'- and 3'-termini, respectively. These molecules have to be processed further to yield mature 28S and 5.8S rRNA. Thermal-denaturation studies revealed that the base-pairing association in the 12S pre-rRNA : 28S rRNA complex is markedly stronger than that in the 5.8S : 28S rRNA complex. The sequence of about one-quarter (1322 base-pairs) of the 5'-part of the rat 28S rDNA was determined. A computer search reveals the possibility that the cleavage sites in the CUC motifs are single-stranded, flanked by strongly base-paired GC tracts, involving tS_i2 and 28S rRNA sequences. The subsequent nuclease cleavages, generating the termini of mature rRNA, seem to be directed by secondary-structure interactions between 5.8S and 28S rRNA segments in pre-rRNA. An analysis for base-pairing among evolutionarily conserved sequences in 32S pre-rRNA suggests that the cleavages yielding mature 5.8S and 28S rRNA are directed by base-pairing between (i) the 3'-terminus of 5.8S rRNA and the 5'-terminus of 28S rRNA and (ii) the 5'-terminus of 5.8S rRNA and internal sequences in domain I of 28S rRNA. A general model for primary- and secondary-structure interactions in pre-rRNA processing is proposed, and its implications for ribosome biogenesis in eukaryotes are briefly discussed.

Processing of preribosomes and their constituent pre-rRNA is an important intranucleolar stage of ribosome biogenesis (Hadjiolov, 1980). Generally, the primary transcript, identified as 45S pre-rRNA in mammalian cells, undergoes a series of nucleolytic cleavages, producing 18S, 5.8S and 28S rRNA (Perry, 1976; Hadjiolov & Nikolaev, 1976). It is now known that in several organisms the early endonuclease attacks may take place in a stochastic manner, thus resulting in the simultaneous occurrence of multiple pre-rRNA processing pathways (Weinberg & Penman, 1970; Winicov, 1976; Dabeva *et al.*, 1976; Dudov *et al.*, 1978; Bowman *et al.*, 1981). However, in most mammalian species endonucleolytic cleavages follow a predominant sequential pattern, thus generating a major pre-rRNA processing pathway (Maden *et al.*, 1972; Hadjiolov & Nikolaev, 1976; Perry, 1976).

The main pre-rRNA species in rat liver have been identified and a major processing pathway, 45S→41S→32S+21S→28S (5.8S)+18S rRNA, has been deduced (Dabeva *et al.*, 1976). Tracer-kinetics studies with normal and regenerating rat liver revealed that an early and rapid phase in

Abbreviations used: pre-rRNA, precursor to rRNA; preribosome, ribonucleoprotein particle containing pre-rRNA; tS_e, external transcribed spacer in rDNA or pre-rRNA; tS_i, internal transcribed spacer in rDNA or pre-rRNA; L-rRNA, rRNA of the large ribosomal particle; rDNA, DNA of the repeating unit of rRNA genes; snRNA, small nuclear RNA; bp, base-pairs in DNA.

processing of 45S pre-rRNA generates 18S rRNA (with or without the intermediate formation of 21S pre-rRNA) and 32S pre-rRNA (Dudov *et al.*, 1978; Dudov & Dabeva, 1983). The processing of 32S pre-rRNA to 28S and 5.8S rRNA in the nucleolus is markedly slower and appears to present more rigid structural requirements (Hadjiolov, 1980). Basically, the process results in the ultimate formation of specifically base-paired 28S and 5.8S rRNA (Pace *et al.*, 1977; Peters *et al.*, 1982; Walker *et al.*, 1982, 1983; see Walker & Pace, 1983). Several subsequent steps appear to be involved. A 12S precursor to 5.8S rRNA was detected in mouse liver (Hadjiolova *et al.*, 1973) and shown to correspond to the whole 5'-end segment of 32S pre-rRNA, including the 5.8S rRNA and tS_i2 sequences (Khan & Maden, 1976; Bowman *et al.*, 1981, 1983; Dudov *et al.*, 1983). Also, a shorter 8S precursor to 5.8S rRNA was identified, encompassing part of the tS_i2 sequence (Reddy *et al.*, 1983). On the other hand, 5'-terminal trimming of nucleolar 28S rRNA is also a compulsory step in the formation of mature 28S rRNA in the mouse (Kominami *et al.*, 1978). The finding that some snRNA molecules (U3 snRNA in particular) in the nucleolus are associated with 32S pre-rRNA or nucleolar 28S rRNA suggests their participation in processing of 32S pre-rRNA (Prestayko *et al.*, 1971; Bachellerie *et al.*, 1983; Reddy & Busch, 1983). All the above results illustrate the complexity of processing of 32S pre-rRNA and correlate with its role as a critical regulatory step in ribosome biogenesis (Hadjiolov *et al.*, 1978; Hadjiolov, 1980).

In the present work we located the first 32S-pre-rRNA endonuclease-processing site by S_1-nuclease protection mapping of the nucleolar 28S rRNA and 12S pre-rRNA termini within the tS_i2 sequence of the rat rDNA. It is shown that multiple cleavages occur at CUC sites in tS_i2, generating 28S rRNA and 12S pre-rRNA with heterogeneous 5'- and 3'-termini, respectively. These molecules have to be processed further to yield mature 28S and 5.8S rRNA. Thermal-denaturation studies revealed a markedly higher stability for the nucleolar 12S:28S rRNA complex than for the cytoplasmic 5.8S:28S rRNA complex. To search for possible base-pairing interactions, we determined the sequence of 1322 bp from the 5'-terminal domain of the rat 28S rRNA gene. These data were correlated with the available sequence information for the 5'-part of 32S pre-rRNA (encompassing tS_i1, 5.8S rRNA and tS_i2 segments) in the rat and other eukaryotes. Possible base-pairing of sequences in the 5'-terminal region of 32S pre-rRNA was deduced, and a general model for the structural features in pre-rRNA processing is proposed.

Materials and methods

Preparation and sequencing of rDNA

The pRr 19 rDNA segment was a *Eco*RI–*Bam*HI restriction fragment derived from a λCharon 4A phage, containing fragment-IV rDNA (Braga *et al.*, 1982), subcloned in the *Escherichia coli* plasmid pBR 322 by standard methods (Maniatis *et al.*, 1982). This fragment contains the 3'-terminal part of 18S rDNA, tS_i and the 5'-terminal part of 28S rDNA (Braga *et al.*, 1982). The sequence of the pRr 19 rDNA was determined by the method of Maxam & Gilbert (1977, 1980) by using appropriate subfragments obtained with different restriction nucleases.

Isolation, purification and analysis of pre-rRNA and rRNA

Total rat liver nucleolar rRNA was isolated as described previously (Dabeva *et al.*, 1976, 1978). The precipitate of nucleolar rRNA was dissolved in a small volume of 0.1M-Tris/acetate (pH5.2), containing 1mM-EDTA and 0.5% sodium dodecyl sulphate, and deproteinized at 25°C with an equal volume of phenol, saturated with the same buffer. The rRNA was precipitated from the water phase with 3vol. of 96% ethanol/1% sodium acetate at −20°C overnight. This deproteinization step is critical in minimizing RNA degradation during further fractionation and storage (Venkov & Hadjiolov, 1969).

Nucleolar rRNA was fractionated by centrifugation in linear 15–30% (w/v) sucrose gradients in 0.1M-Tris/acetate (pH7.0) plus 0.5% sodium dodecyl sulphate, in a vertical rotor to shorten the separation time and minimize degradation. Centrifugation in a Beckman VTi 50 rotor was for 105min at 46000 rev./min, at 4°C in a Beckman L5-65 ultracentrifuge. Gradient fractions, monitored at 254nm, were pooled and the RNA was precipitated as above. The purified nucleolar 28S rRNA was obtained after two consecutive sucrose-density-gradient runs. To isolate 12S pre-rRNA, we used 32S pre-rRNA isolated by five consecutive runs. The 32S pre-rRNA fraction was denatured by heating for 5min at 70°C in the presence of 50% formamide (Dudov *et al.*, 1983). The released 12S pre-rRNA was isolated by sucrose-density-gradient centrifugation as above, its position being monitored by a parallel run of denatured 32S pre-rRNA, labelled with [^{14}C]-orotate *in vivo*.

The rRNA fractions were analysed by gel electrophoresis in 1.5% agar (or agarose)/5M-urea (Dudov *et al.* 1976). The 'melting' of 12S:28S or 5.8S:28S rRNA complexes was performed essentially as described by Pace *et al.* (1977), but heating at the indicated temperatures was for 5min, and

dissociated rRNA components were analysed by the agar/urea-gel-electrophoresis method.

S_1-nuclease protection mapping

The precise location of the 5'- and 3'-termini of pre-rRNA and rRNA was performed by S_1-nuclease protection mapping (Berk & Sharp, 1977) of hybrids between purified pre-rRNA or rRNA and appropriate ^{32}P-labelled restriction-nuclease subfragments from our rat rDNA clone. The RNA:DNA hybrids were digested with 30, 50 or 100 units of S_1 nuclease for 1h at 37°C in the standard medium [30mM-sodium acetate (pH4.5)/0.1M-NaCl/0.1mM-$ZnSO_4$]. It is shown that hydrolysis of RNA:DNA hybrids under these conditions (37°C) avoids possible incomplete protection at long oligo(U) tracts in RNA and yields results identical with those obtained by independent techniques, i.e. reverse-transcriptase extension (Rothblum *et al.*, 1982*b*). For mapping of 5'-termini the rDNA restriction fragments were labelled at 5'-ends with T_4-phage polynucleotide kinase and [γ-^{32}P]ATP. For 3'-termini the restriction fragments were labelled at their 3'-ends by filling in the protruding 5'-ends with the use of DNA polymerase I (large fragment) and labelled [α-^{32}P]dATP. The size of the rDNA fragments protected against S_1 nuclease was determined by electrophoresis on to acrylamide/urea sequencing gels in parallel with an appropriate rDNA sequence 'ladder' (Maxam & Gilbert, 1980).

Materials

Analytical-grade reagents were used throughout. Restriction nucleases *Ava*II, *Hinf*I, *Eco*RII, *Taq*I, *Sma*I and *Bam*HI, as well as DNA polymerase I (large fragment), were products of BioLabs Inc., Beverly, MA, U.S.A. Restriction nucleases *Eco*RI, *Sau*3A, *Xho*I, *Kpn*I and *Mva*I, and T_4-phage polynucleotide kinase, were kindly given by Dr. A. Yanulaitis (Institute of Applied Enzymology, Vilnus). The preparation of S_1 nuclease (type III) was a product of Sigma Chemical Co., St. Louis, MO, U.S.A. The [γ-^{32}P]ATP and [α-^{32}P]dATP were prepared by the methods developed by Walseth & Johnson (1979), by using [^{32}P]orthophosphate from The Radiochemical Centre, Amersham, Bucks., U.K.

Results

Mapping of the endonuclease-cleavage site in 32S pre-rRNA generating 28S rRNA and 12S pre-rRNA

The first step in rat liver 32S pre-rRNA processing appears to be its cleavage into 12S pre-rRNA and 28S rRNA (Dudov *et al.*, 1983). The 12S pre-rRNA remains hydrogen-bonded to 28S rRNA and co-sediments with 32S pre-rRNA. Therefore we decided to map the 5'-termini of nucleolar and cytoplasmic 28S rRNA and the 3'-terminus of 12S pre-rRNA. The native state and purity of the RNA fractions is of critical importance in such studies. The purified nucleolar and cytoplasmic 28S rRNA yielded (after three subsequent sucrose-density-gradient runs) single symmetric peaks, shown to be homogeneous by gel electrophoresis. Also, the nucleolar 28S rRNA has been shown to be virtually free of cytoplasmic contaminants (Dudov *et al.*, 1978). The 12S pre-rRNA fraction is released from 32S pre-rRNA (see the Materials and methods section) and is likely to contain mainly the initial cleavage products. This fraction does not contain heavier RNA components, but is contaminated by 5.8S rRNA (results not shown). Since the endonuclease cleavage is expected to be near the 5'-terminus of 28S rRNA (Dudov *et al.*, 1983), the presence of 5.8S rRNA should not interfere in mapping the 3'-terminus of 12S pre-rRNA.

To map pre-rRNA and rRNA termini, we used appropriate rDNA restriction fragments covering parts of tS_i2 and 28S rDNA sequences. The complete sequence of these rDNA fragments is now known (see below and Subrahmanyam *et al.*, 1982). To map the 5'-termini of 28S rRNA molecules we used a *Sma*I (2070)–*Hinf*I (2310) rDNA subfragment, obtained from the 5'-end-labelled *Hinf*I (1526)–*Hinf*I (2310) fragment. [Here and elsewhere the numbers in parentheses indicate the position of a nucleotide (or restriction site) in the sequence presented in Fig. 4.] To map the 3'-terminus of 12S pre-rRNA, we used a 3'-end-labelled *Eco*RII (2115)–*Eco*RII (2306) rDNA fragment. After annealing with the respective rRNA and S_1-nuclease digestion, the protected fragments were run in parallel with a *Sma*I–*Hinf*I sequence ladder (Fig. 1*a*). The results show that the 5'-terminus of cytoplasmic 28S rRNA is homogeneous and maps at nucleotides 2231 and 2232, in complete agreement with previous studies (Subrahmanyam *et al.*, 1982). Unexpectedly, the 5'-termini of nucleolar 28S rRNA were heterogeneous and mapped in two zones (*A* and *B*) within the tS_i2 rDNA segment, with most of the material in zone *A*. Both zones of protected rDNA represent a CTC motif extending between positions 2171 and 2184 (zone *A*) and 2218–2230 (zone *B*), i.e. the most distant protected nucleotide in *A* and *B* being at 60 and 13 bp upstream from the 5'-end of 28S rDNA. The 3'-termini of 12S pre-rRNA are also heterogeneous and map within the same CTC motif zones (*A'* and *B'*). Here the protected zones encompass nucleotides 2166–2182 (*A'*) and 2211–2218 (*B'*), the first nucleotides being at 65 and 20 bp upstream from the 5'-end of 28S rDNA. By using

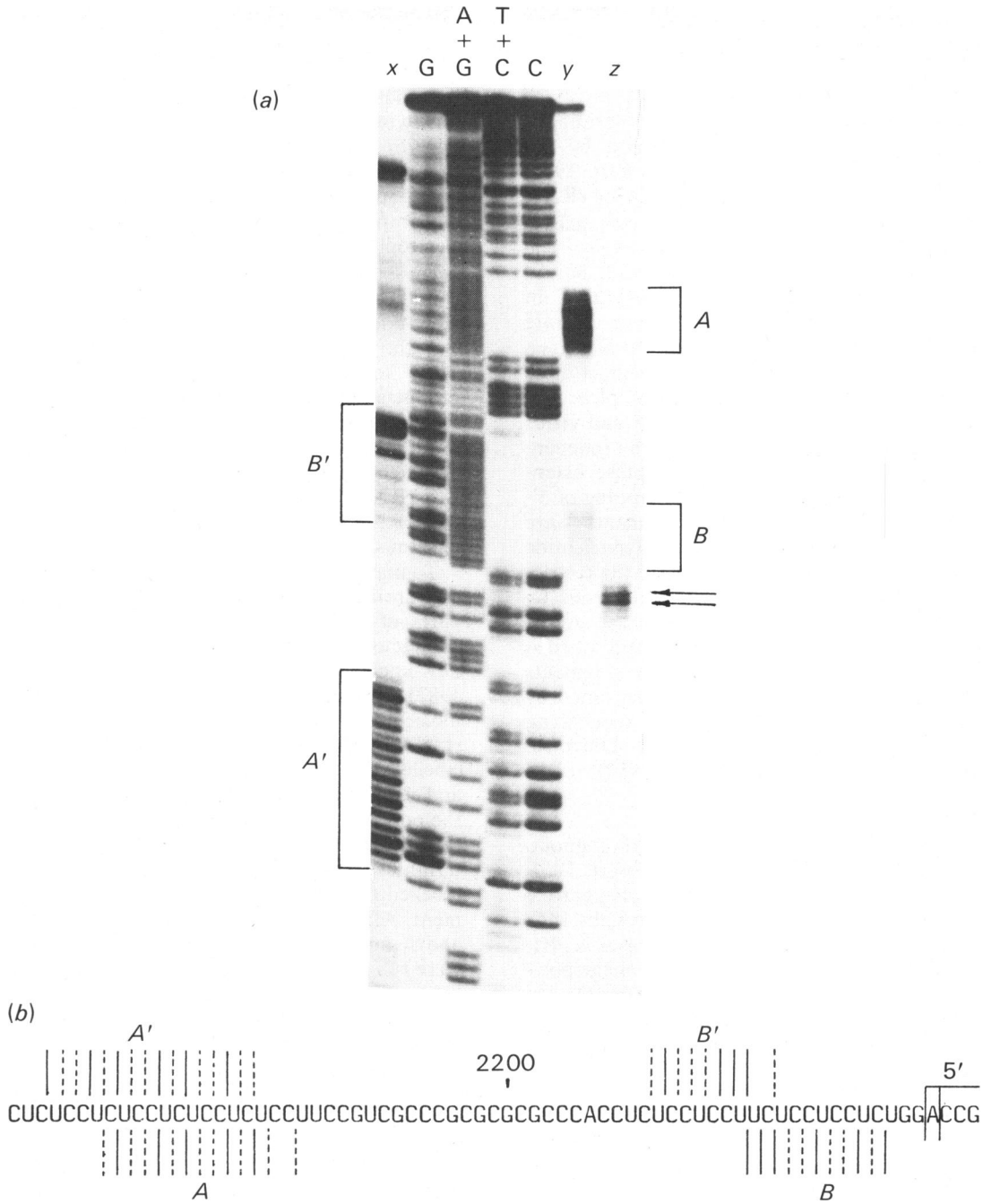

Fig. 1. *Nuclease-S_1 protection mapping of the 5′-termini of nucleolar and cytoplasmic 28 S rRNA and the 3′-terminus of 12 S pre-rRNA*

(*a*) Lane *x*, hybridization of 12 S pre-rRNA with a 3′-end-labelled *Eco*RII (2115)–*Eco*RII (2306) rDNA subfragment; lanes *y* and *z*, hybridization with a 5′-end-labelled *Sma*I (2070)–*Hin*fI (2310) rDNA subfragment of nucleolar (*y*) and cytoplasmic (*z*) 28 S rRNA. The parallel sequence ladder is from the same *Sma*I–*Hin*fI rDNA subfragment. The horizontal arrows indicate the position of the rDNA protected from S_1-nuclease digestion by hybridization with cytoplasmic 28 S rRNA. The boxes encompass the heterogeneous rDNA fragments protected by nucleolar 28 S rRNA (*A* and *B*) and by 12 S pre-rRNA (*A′* and *B′*). (*b*) Precise mapping of the rDNA fragments protected by nucleolar 28 S rRNA (*A* and *B*) and 12 S pre-rRNA (*A′* and *B′*) within the respective segment of the pRr 19 rDNA sequence. Continuous and broken lines indicate the position of strong and weak blackening of the autoradiogram respectively. 5′ denotes the terminus of cytoplasmic 28 S rRNA.

overexposed films we attempted to locate more precisely the termini of protected rDNA fragments and to estimate the position of the more highly labelled bands, best illustrated in zone *A'* (Fig. 1*b*).

The location of nucleolar 28S rRNA 5'-termini and of 12S pre-rRNA 3'-termini at identical positions within rDNA identifies unambiguously (from both sides) the first endonuclease-cleavage site in 32S pre-rRNA. These results show also that the endonuclease attack is site-heterogeneous and located within two CUC motifs in 32S pre-rRNA, upstream from the 5'-end of the mature cytoplasmic 28S rRNA. It is noteworthy that there are no endonuclease cleavages in the GC-rich zone located between the two CUC motifs. Closer observation of the pattern of the fragments resistant to S_1-nuclease digestion (*A* and *A'* in particular) suggests that the endonuclease cleaves preferentially CpU and UpU rather than UpC and CpC bonds. Location of the first endonuclease-cleavage site within the tS_i2 sequence reveals that both nucleolar 28S rRNA and 12S pre-rRNA have to be processed further in order to generate mature 28S and 5.8S rRNA, respectively.

Association of 12S pre-rRNA with 28S rRNA

We have shown previously that 12S pre-rRNA is tightly bound to nucleolar 28S rRNA and the complex co-migrates with 32S pre-rRNA (Dudov *et al.*, 1983). To gain further insight into the interactions involved, we studied the temperature-dependence of the denaturation of the 12S:28S rRNA complex, as compared with that of the 5.8S:28S rRNA complex. The results (Fig. 2) reveal that the thermal transitions of the two complexes are markedly distinct. The 'melting' temperature (T_m) for the 5.8S:28S rRNA complex is 50°C, in agreement with previous studies with the same complex from mice (Pace *et al.*, 1977; Walker *et al.*, 1982). In contrast, the nucleolar 12S:28S rRNA complex displays a T_m in the range 80–85°C. These results show that, in addition to the 5.8S:28S rRNA interactions, the association of 12S pre-rRNA involves also strong interactions of its non-conserved segment (corresponding to tS_i2) with complementary segments in 28S rRNA. Since the tS_i2 segment in 12S pre-rRNA is extremely GC-rich (see below), the rather high T_m for the nucleolar 12S:28 SrRNA complex strongly suggests the presence of base-pairing interactions between rather long GC-rich tracts in 12S pre-rRNA and 28S rRNA.

Sequence of the 5'-part of the rat 32S pre-rRNA

The rat *Eco*RI–*Bam*HI rDNA segment studied here contains the whole 18S–28S rRNA intergene region and about one-quarter of the 28S rRNA gene (Rothblum *et al.*, 1982*a*; Braga *et al.*, 1982).

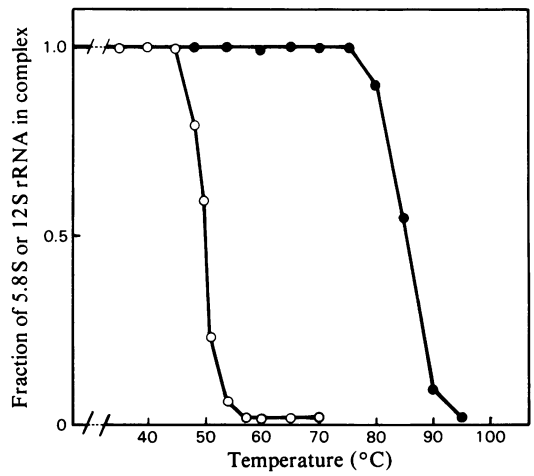

Fig. 2. *Thermal denaturation of cytoplasmic 5.8S:28S rRNA* (○) *and nucleolar 12S:28S rRNA* (●) *complexes*
The RNA samples were dissolved in 0.1M-NaCl/3mM-EDTA and heated for 5min at the indicated temperatures. RNA was fractionated by electrophoresis in agar/urea gels (see the Materials and methods section).

The sequence of the 18S–28S rRNA intergene region in one rDNA clone was determined (Subrahmanyam *et al.*, 1982). The sequence of the same region in the homologous pRr 19 rDNA segment (see the Materials and methods section) was also determined (O. I. Georgiev, V. V. Nosikov, E. A. Braga & A. A. Hadjiolov, unpublished work).

In the present work we have established the sequence of the 5'-terminal part of the rat 28S rRNA gene. It is likely that the corresponding part in the 32S pre-rRNA molecule participates in important interactions shaping the structure of endonuclease sites (see above). The location of the pRr 19 rDNA segment within the rat rDNA repeating unit and the sequencing strategy for its 28S rDNA segment are depicted in Fig. 3. The whole pRr 19 *Eco*RI–*Bam*HI rDNA segment contains 3552 bp, including the 3'-part of 18S rDNA (231 bp), tS_i1 (1072 bp), 5.8S rDNA (156 bp), tS_i2 (771 bp) and the 5'-part of 28S rDNA (1322 bp). We have previously mapped the 5'-terminus of 32S pre-rRNA at 161–163 bp upstream from the 5'-end of 5.8S rRNA (Hadjiolova *et al.*, 1984). These data permit us to derive the sequence for the 5'-part of 32S pre-rRNA, encompassing tS_i1, 5.8S rRNA, tS_i2 and 28S rRNA segments (Fig. 4).

It is known that only limited, if any, 3'-terminal trimming is involved in pre-rRNA processing (see, e.g., Bayev *et al.*, 1981; Kominami *et al.*, 1982; Din

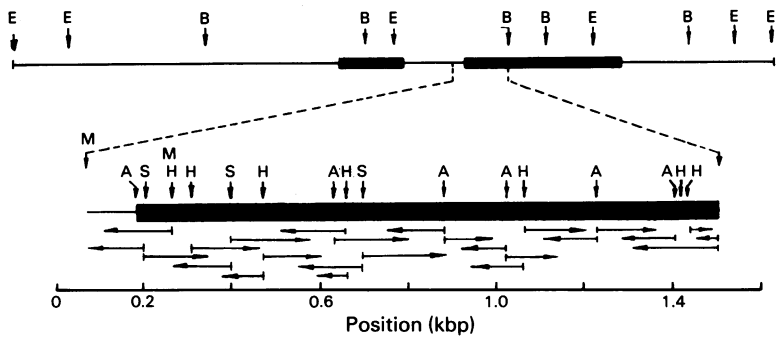

Fig. 3. *Structure of the rat rRNA repeating unit (above) and expanded map of the sequenced segment of the subcloned pRr 19 rDNA fragment containing the 5'-terminal part of the 28S rRNA gene*

The restriction-nuclease sites used in sequence determination are designated by vertical arrows: E, *Eco*RI; B, *Bam*HI; S, *Sau*3A; A, *Ava*II; H, *Hin*fI; M, *Mva*I. The horizontal arrows indicate the direction and the size of sequenced rDNA restriction subfragments.

et al., 1982). Therefore, all processing-endonuclease cleavage sites in 32S pre-rRNA are evidently located within its 5'-end sequence shown in Fig. 4. In this respect the following features of the rat 32S pre-rRNA sequence are noteworthy. (i) The tS_i2 segment is extremely GC-rich (79.8%); however, although very long GC tracts dominate the whole sequence, single-nucleotide (G or C) tracts longer than 9 residues are not present. (ii) The 28S rRNA segment encompassing nucleotides 2232–3552 is also extremely GC-rich (79.1%); in this case too, single-nucleotide tracts (G or C) longer than 8 residues are not found. (iii) The GC-rich segment in the rat 28S rRNA does not display any apparent homology with the corresponding region in the previously sequenced 25S rRNA of *Saccharomyces cerevisiae* (Georgiev *et al.*, 1981) or *Saccharomyces carlsbergensis* (Veldman *et al.*, 1981*b*).

Possible base-pairing interactions in 32S pre-rRNA processing

There is now general agreement that base-pairing interactions play a crucial role in pre-rRNA processing. As shown in this work, the initial endonuclease cleavage is located in tS_i2, within the CUC motifs upstream from the 5'-end of the 28S rRNA segment. Since the termini of mature 5.8S and L-rRNA are homogeneous, the spacer/rRNA junctions are obvious nuclease-cleavage sites in the terminal stages of pre-rRNA processing. Accordingly, we searched for possible base-pairing around these processing sites.

In a comparative analysis we used presently available sequence data on spacer/rRNA junctions in *S. cerevisiae* (Skryabin *et al.*, 1979; Bayev *et al.*, 1981), *Xenopus laevis* (Hall & Maden, 1980), *Mus musculus* (Michot *et al.*, 1983; Goldman *et al.*, 1983) and *Rattus norvegicus* (Subrahmanyam *et al.*, 1982; the present work). The results summarized in Fig. 5 show: (i) considerable conservation of terminal sequences in mature 5.8S and L-rRNA; (ii) an abrupt and rapid divergence of spacer sequences adjacent to the 5'-terminus of mature L-rRNA, even between closely related species; and (iii) limited conservation of spacer sequences at 5.8S rRNA/spacer junctions. The homologous sequences at the 5.8S rRNA/tS_i2 junction appear evolutionarily more stable than the tS_i1/5.8S rRNA junction. These observations indicate that base-pairing involving 5'- and 3'-termini of mature rRNA molecules may be similar in most eukaryotes, whereas that involving spacer sequences is expected to be species- or genus-specific.

Considering the information conveyed by the comparative studies, we carried out a computer search for complementarity between sequences in 32S pre-rRNA which might be involved in its processing. In agreement with a similar study for the mouse tS_i2 segment (Goldman *et al.*, 1983), we found a high base-pairing potential among tS_i2 sequences. In addition, numerous base-paired tracts can be formed between tS_i2 and the GC-rich tract in the 28S rRNA segment. Looking specifically for possible base-pairing interactions in the neighbourhood of established or putative pre-rRNA processing sites, some reasonably stable base-paired structures may be discerned, as listed in Fig. 6, including the following. (i) Strong base-pairing is possible at the tS_i2 sequences adjacent to the two processing-site CUC motifs, one of them involving a GC-rich tract in 28S rRNA, whereas the other is in tS_i2 at about 150 nucleotides downstream from the 3'-end of 5.8S rRNA (see Figs. 6*a* and 6*b*). It is also noteworthy that GAG tracts with reasonable complementarity to the

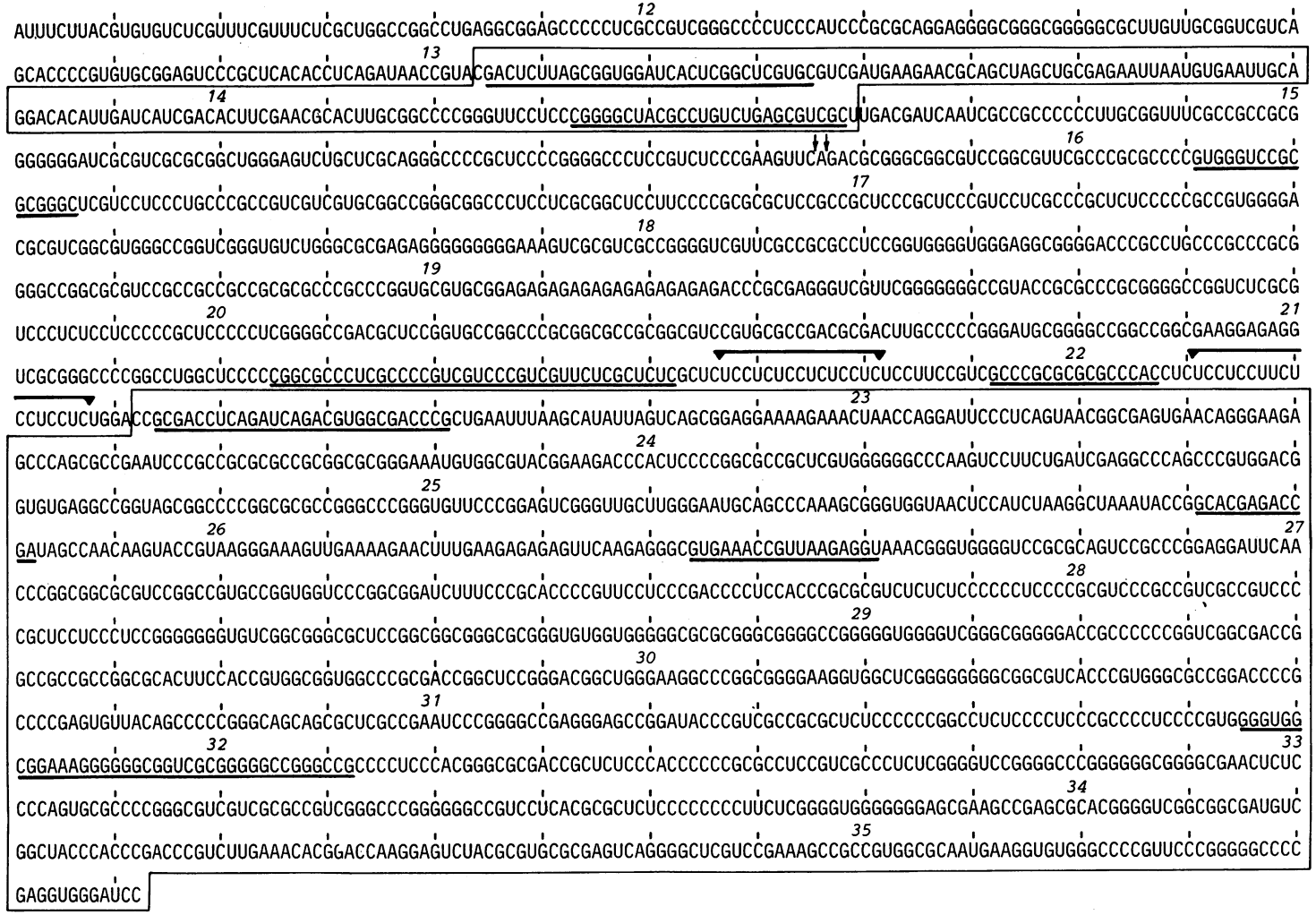

Fig. 4. *Primary structure of the 5′-terminal part of the rat 32S pre-rRNA*

The structure of 32S pre-rRNA is derived from the sequence of the respective rat rDNA segment (clone pRr 19). The sequence of tS_i1, 5.8S rRNA and tS_i2 segments corresponds to that established by Subrahmanyam *et al.* (1982), corrected according to the sequence found for our rDNA clone (O. I. Georgiev, V. V. Nosikov, E. A. Braga & A. A. Hadjiolov, unpublished work) and data from Reddy *et al.* (1983). The sequence of the 28S rRNA segment is determined in the present work. Location of the 5′-terminus of 32S pre-rRNA is based on data by Hadjiolova *et al.* (1984). The numbering (in hundreds) starts from the *Eco*RI site at the 5′-end (non-coding strand) of the pRr 19 rDNA fragment (see Fig. 3) and is used throughout the text. Positions 1200 and 2200 correspond respectively to positions 1186 and 2182 in the sequence published by Subrahmanyam *et al.* (1982). The 5.8S and 28S rRNA segments are boxed. Vertical arrows and arrowheads at ends of horizontal lines indicate the position of identified endonuclease-cleavage sites. The underlined sequences are possibly involved in base-pairing interactions related to the processing of 32S pre-rRNA.

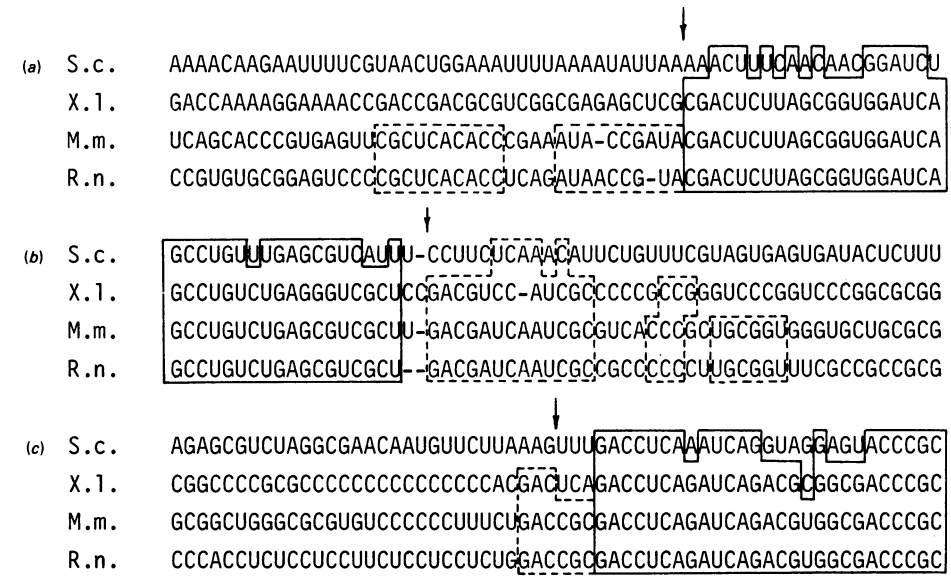

Fig. 5. *Comparison of eukaryotic sequences at rRNA/spacer junctions*
(*a*) tS_i1/5.8S rRNA; (*b*) 5.8S rRNA/tS_i2; (*c*) tS_i2/28S rRNA. For sources of sequence data see the text. The vertical arrows indicate the positions of the 5′- or 3′-termini in rRNA. Homologous sequences are boxed with a continuous line. Spacer sequences of limited homology are boxed with a broken line. Abbreviations: S.c., *Saccharomyces cerevisiae*; X.l., *Xenopus laevis*, M.m., *Mus musculus*; R.n., *Rattus norvegicus*.

CUC motifs are not present in the whole tS_i2 or 28S rRNA sequence. (ii) There is rather strong base-pairing between the 5′-terminus of the 28S rRNA segment and 26 nucleotides from the 3′-terminus of 5.8S rRNA. As noted by others (Olsen & Sogin, 1982), this interaction is remarkably conserved in evolution. (iii) The 5′-terminal sequence of 5.8S rRNA can be stably base-paired with two sequences (60 nucleotides apart) in the 5′-end domain of 28S rRNA (see Fig. 6*c*). These base-paired tracts involving the 3′-end of 5.8S rRNA are similar to structures proposed for yeasts (Veldman *et al.*, 1981*b*) and the mouse (Michot *et al.*, 1982).

Discussion

The results obtained in the present work provide information on the mechanisms involved in the processing of 32S pre-rRNA in mammalian cells. It is shown that the initial endonuclease cleavages (generating nucleolar 28S rRNA and 12S pre-rRNA) are site-heterogeneous and occur in tS_i2 within two CUC motifs upstream from the 5′-end of mature 28S rRNA. Cleavage of CpU, UpU, UpC and CpC bonds takes place, but the observed protection pattern (Fig. 1*b*) suggests the preferential hydrolysis of CpU and possibly UpU bonds. Owing to the extensive divergence of tS_i2 sequences, it is likely that the location of this cleavage site will be species-specific. A CUC motif in tS_i2 may be discerned in the mouse (Michot *et al.*, 1983; Goldman *et al.*, 1983), but not in yeast (Skryabin *et al.*, 1979; Veldman *et al.*, 1981*a*; Bayev *et al.*, 1981) or *Xenopus laevis* (Hall & Maden, 1980). While our manuscript was in preparation, a S_1-nuclease-mapping study on the initial cleavage in the mouse 32S pre-rRNA was published (Bowman *et al.*, 1983). Those authors mapped 5′-termini of nuclear 28S rRNA at the 5′-end of 28S rDNA or within tS_i2 at 4–6 bp upstream. However, the 3′-terminus of 12S pre-rRNA mapped at 295 bp upstream from the 5′-end of 28S rDNA, thus leaving uncertain the location of the initial cleavage site in 32S pre-rRNA. Since those authors used denatured total nuclear RNA (instead of purified nucleolar 32S pre-rRNA), it seems likely that the presence of processing products (12S pre-rRNA with processed 3′-termini and 28S rRNA with processed 5′-termini) could account for the gap between the mapping sites of these two RNA molecules. Therefore further studies are needed to clarify whether the CUC motifs present in the mouse tS_i2 have, like those of the rat, some signal role in 32S pre-rRNA processing.

The location of the initial processing site in the rat 32S pre-rRNA provides information on its

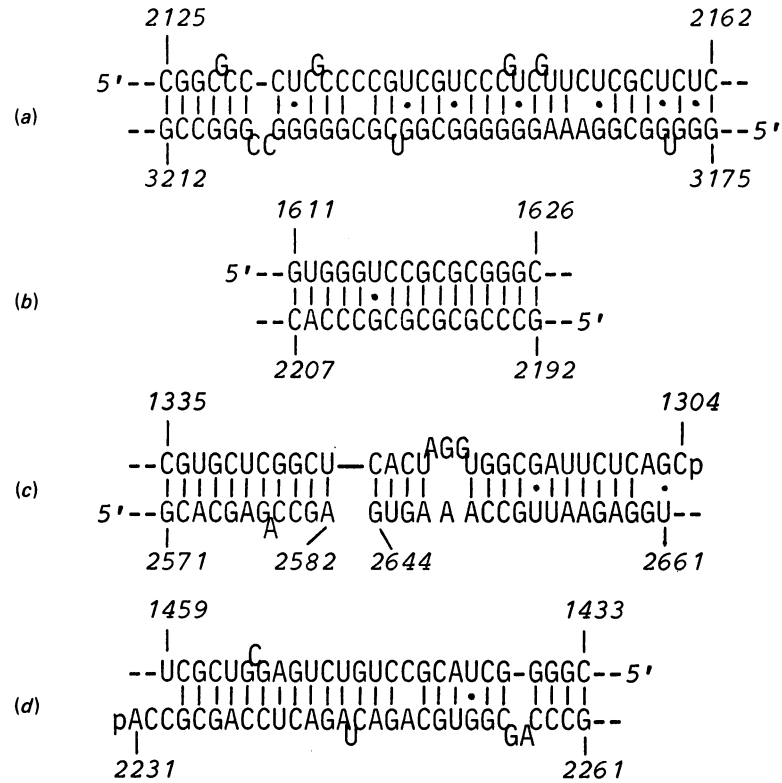

Fig. 6. *Possible base-pairing within 32S pre-rRNA related to its processing*
(*a*) tS_i2 sequence (part of 12S pre-rRNA) complementary to sequence in domain I of 28S rRNA; (*b*) self-complementary sequences in tS_i2; (*c*) 5′-terminal part of 5.8S rRNA and sequences in domain I of 28S rRNA; (*d*) 3′-terminus of 5.8S rRNA and 5′-terminal part of 28S rRNA. The numbering of the respective nucleotide positions is the same as in Fig. 4.

structure. It seems that a single-strand-specific endonuclease is responsible for the cleavage within the two CUC motifs in tS_i2. This possibility is supported by the following facts. (i) The two CUC motifs are surrounded by strongly base-paired segments involving tS_i2 and 28S rRNA sequences (Figs. 6*a* and 6*b*); (ii) no complementary GAG motifs are present in tS_i2 or 28S rRNA. The single GA tract (positions 1905–1928) in tS_i2 does not show satisfactory complementarity to either CUC motifs. It is noteworthy that the CUC motifs in the mouse tS_i2 appear also to be single-stranded (Goldman *et al.*, 1983; Michot *et al.*, 1983). Of course, the above evidence is only suggestive, and direct enzymological studies should decide whether a single-strand-specific endonuclease participates in the first step of 32S pre-rRNA processing. Several authors described the presence in mammalian cell nucleoli of double-strand-specific nucleases (see Hadjiolov, 1980), but a single-strand-specific nuclease was also reported (Eichler & Eales, 1982). The likely involvement of single-strand-specific endonucleases in pre-rRNA processing, if directly proved, will be basically distinct from the pre-rRNA processing mechanisms in prokaryotes, where the double-strand-specific ribonuclease III plays a leading role (see, e.g., Nikolaev *et al.*, 1975; Bram *et al.*, 1980).

An unexpected finding in the present work is that the multiple site cleavages in tS_i2 generate nucleolar 28S rRNA and 12S pre-rRNA with heterogeneous 5′- and 3′-termini, respectively. These results pose the question about the mechanisms of further processing to yield the mature 28S and 5.8S rRNA.

Considering the processing of 12S pre-rRNA, its strong association with 28S rRNA should be accounted for. A 29.5S precursor to 28S rRNA was observed previously in rat liver (Fujisawa *et al.*, 1973*a*). The 29.5S pre-rRNA is converted into 28S rRNA on thermal denaturation, releasing an RNA fragment of about 500 nucleotides (Fujisawa *et al.*, 1973*b*). It is likely that this fragment is a 12S pre-rRNA processing product, but its existence

remains to be ascertained. Another 12S pre-rRNA processing product is the 8S pre-rRNA, associated with nucleolar 28S rRNA (Prestayko *et al.*, 1971; Fujisawa *et al.*, 1973*b*). More recently, the complete sequence of 8S pre-rRNA was determined, showing the same 5′-end as mature 5.8S rRNA, and its homogeneous 3′-end is at 117–118 nucleotides within the tS_i2 sequence (Reddy *et al.*, 1983; see Fig. 4). Thus the 8S rRNA is a stable processing product of 12S pre-rRNA, lacking the extra 160 nucleotides at the 5′-end and most of the tS_i2 sequence. Accumulation of a stable 8S pre-rRNA suggests that: (i) 12S pre-rRNA processing is catalysed by endonucleases and (ii) a cleavage at the 5′-end of 5.8S rRNA precedes that generating 8S pre-rRNA. Further processing is obviously needed to shape the 3′-end of 5.8S rRNA. The presence of a homogeneous 5′-end in 8S pre-rRNA (coinciding with that of 5.8S rRNA) supports the proposal that the shaping of 5′- and 3′-ends in 5.8S rRNA is directed by its base-pairing with 28S rRNA sequences (see Fig. 6). The proposal that interactions with U3 snRNA also participate in shaping processing structures in some eukaryotes (Bachellerie *et al.*, 1983) requires direct experimental evidence.

The presence of excess 5′-terminal nucleotides in nucleolar 28S rRNA provides direct proof that 5′-terminal processing is needed to generate mature 28S rRNA. The fact that we do not observe nucleolar 28S rRNA molecules with mature 5′-termini indicates that the 5′-terminal processing takes place immediately before the release of nucleolar 28S rRNA or, alternatively, in the nucleoplasm or the cytoplasm. These results agree with previous indirect evidence for the cytoplasmic 5′-terminal processing of mouse 28S rRNA (Kominami & Muramatsu, 1977; Kominami *et al.*, 1978). However, the finding of mature 5′-ends in mouse nuclear 28S rRNA (Bowman *et al.*, 1983) and the differences in mouse and rat tS_i2 sequences (Goldman *et al.*, 1983; Michot *et al.*, 1983) make uncertain the conclusion about the identity of nucleolar 28S rRNA 5′-processing in the two species.

Our results, correlated with data on the processing of pre-rRNA in eukaryotes (see Hadjiolov & Nikolaev, 1976; Hadjiolov, 1980), permit us to outline a general model for the processing of pre-rRNA in mammalian cells. The model emphasizes the leading role of secondary-structure interactions in defining endonuclease-cleavage sites and the shaping of 5′- and 3′-termini in mature rRNA. The following aspects deserve further comment (Fig. 7).

(i) Primary pre-rRNA processing can be delineated into two distinct phases designated as the rapid (*r*) and slow (*s*) phases. The *r*-phase generates 32S pre-rRNA and 18S rRNA and the respective ribonucleoprotein particles. The *s*-phase generates 28S and 5.8S rRNA and the large ribosomal particle. Endonuclease cleavages within spacer segments in pre-rRNA are at species- or genus-specific sites. In contrast, the final shaping of

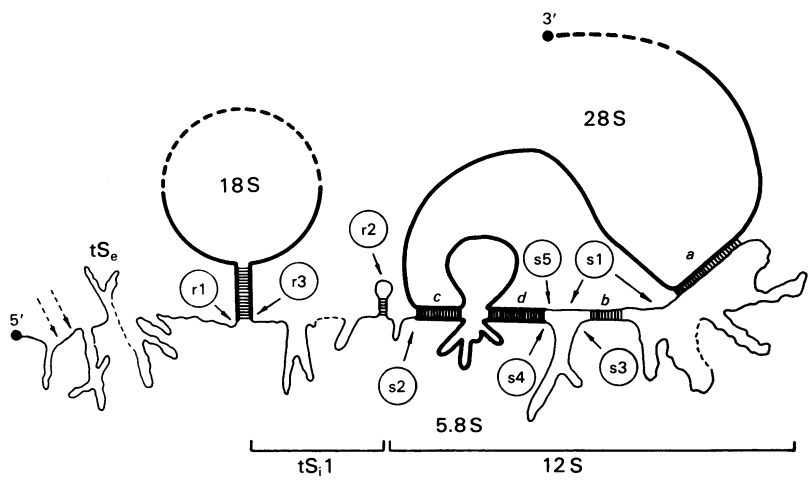

Fig. 7. *A general model for the possible secondary-structure interactions in the processing of pre-rRNA in mammalian cells*
The arrows indicate endonuclease-cleavage sites during the rapid (*r1–r3*) and slow (*s1–s5*) phases of pre-rRNA processing. The numbering reflects the preferred order of endonuclease attacks. The segments corresponding to mature rRNA species are indicated by thicker lines. 12S pre-rRNA encompasses tS_i1, 5.8S rRNA and tS_i2 segments. The sequences involved in base-pairing are expanded, and *a–d* denote the interactions depicted in Fig. 6. The wavy lines indicate unspecified secondary-structure interactions within spacer sequences. For details and discussion see the text.

homogeneous 5′- and 3′-termini in mature rRNA is defined by their involvement in evolutionarily conserved base-pairing interactions.

(ii) Considerable flexibility in the sequence of endonuclease cleavages may occur in *r*-phase, resulting in the formation of different intermediate pre-rRNA species and processing along alternative pathways (Dudov *et al.*, 1978; Bowman *et al.*, 1981, 1983; Dudov & Dabeva, 1983). Endonucleases showing a preference for UpU or NpU bonds seem to catalyse *r*-phase cleavages. The location of site *r2* (see Fig. 7) is species-specific, whereas sites *r1* and *r3* are at the 5′- and 3′-termini of mature 18S rRNA (Bowman *et al.*, 1983). We propose that cleavages at sites *r1* and *r3* are directed by possible base-pairing of 5′- and 3′-termini in 18S rRNA (not shown), rather than by interactions involving the highly divergent sequences in adjacent spacers (Maden *et al.*, 1982). Cleavages within oligo(U) tracts, present in the external spacer of some species, may cause further heterogeneity of *r*-phase pre-rRNA (Dabeva *et al.*, 1976; Mishima *et al.*, 1981).

(iii) The *s*-phase seems to involve more rigid structural requirements in defining the site and sequence of endonuclease cleavages. Here too, the final shaping of 5.8S termini and the 5′-terminus of 28S rRNA is directed by their involvement in base-pairing interactions, which are evolutionarily conserved. In contrast, the initial cleavages at site *s1* (and possibly *s3*) seem to be species- or genus-specific. The *s1* cleavage within tS_i2 is close to the 5′-end of mature 28S rRNA. A single-strand-specific endonuclease, with a preference for UpU and CpU bonds, appears to be involved. In mammalian cells, the cleavage at site *s1* is further specified by strong base-pairing interactions between GC-rich sequences in tS_i2 (12S pre-rRNA) and the 5′-terminal domain of 28S rRNA. The participation of GC-rich sequences in pre-rRNA processing is a novel acquisition in the evolution of eukaryotes (see Hadjiolov & Nikolaev, 1976).

The complex conformation requirements in pre-rRNA processing in eukaryotes suggest the participation of selective and phase-specific interactions in the assembly of structural ribosomal proteins within the two ribosomal particles, in particular the large one. These interactions and the 5′-terminal processing of nucleolar 28S rRNA are likely to play an important regulatory role in ribosome biogenesis. We hope that the proposed model may be helpful in the elucidation of such interactions.

We are indebted to Dr. K. P. Dudov for preparing the computer program and to Mrs. Z. Stoymenova for her technical assistance in some experiments.

References

Bachellerie, J. P., Michot, B. & Raynal, F. (1983) *Mol. Biol. Rep.* **9**, 79–87

Bayev, A. A., Georgiev, O. I., Hadjiolov, A. A., Nikolaev, N., Skryabin, K. G. & Zakharyev, V. (1981) *Nucleic Acids Res.* **9**, 789–799

Berk, A. J. & Sharp, P. A. (1977) *Cell* **12**, 721–732

Bowman, L. H., Rabin, B. & Schlessinger, D. (1981) *Nucleic Acids Res.* **9**, 4951–4966

Bowman, L. H., Goldman, W. E., Goldberg, G. I., Hebert, M. B. & Schlessinger, D. (1983) *Mol. Cell. Biol.* **3**, 1501–1510

Braga, E. A., Yussifov, T. N. & Nosikov, V. V. (1982) *Gene* **20**, 145–156

Bram, R. J. R., Young, R. A. & Steitz, J. A. (1980) *Cell* **19**, 393–401

Dabeva, M. D., Dudov, K. P., Hadjiolov, A. A., Emanuilov, I. & Todorov, B. N. (1976) *Biochem. J.* **160**, 495–503

Dabeva, M. D., Dudov, K. P., Hadjiolov, A. A. & Stoykova, A. S. (1978) *Biochem. J.* **171**, 367–374

Din, N., Engberg, J. & Gall, J. G. (1982) *Nucleic Acids Res.* **10**, 1503–1513

Dudov, K. P. & Dabeva, M. D. (1983) *Biochem. J.* **210**, 183–192

Dudov, K. P., Dabeva, M. D. & Hadjiolov, A. A. (1976) *Anal. Biochem.* **76**, 250–258

Dudov, K. P., Dabeva, M. D., Hadjiolov, A. A. & Todorov, B. N. (1978) *Biochem. J.* **171**, 375–383

Dudov, K. P., Hadjiolova, K. V., Kermekchiev, M. B., Stanchev, B. S. & Hadjiolov, A. A. (1983) *Biochim. Biophys. Acta* **739**, 79–84

Eichler, D. C. & Eales, S. J. (1982) *J. Biol. Chem.* **257**, 14384–14389

Fujisawa, T., Abe, S., Kawada, T., Satake, M. & Ogata, K. (1973*a*) *Biochim. Biophys. Acta* **324**, 226–240

Fujisawa, T., Abe, S., Satake, M. & Ogata, K. (1973*b*) *Biochim. Biophys. Acta* **324**, 241–256

Georgiev, O. I., Nikolaev, N., Hadjiolov, A. A., Skryabin, K. G., Zakharyev, V. M. & Bayev, A. A. (1981) *Nucleic Acids Res.* **9**, 6953–6959

Goldman, W. E., Goldberg, G., Bowman, L. H., Steinmetz, D. & Schlessinger, D. (1983) *Mol. Cell. Biol.* **3**, 1488–1500

Hadjiolov, A. A. (1980) *Subcell. Biochem.* **7**, 1–80

Hadjiolov, A. A. & Nikolaev, N. (1976) *Prog. Biophys. Mol. Biol.* **31**, 95–144

Hadjiolov, A. A., Dabeva, M. D., Dudov, K. P., Gajdardjieva, K. C., Georgiev, O. I., Nikolaev, N. & Stoyanova, B. B. (1978) *FEBS Symp.* **51**, 319–328

Hadjiolova, K. V., Golovinsky, E. V. & Hadjiolov, A. A. (1973) *Biochim. Biophys. Acta* **319**, 373–382

Hadjiolova, K. V., Georgiev, O. I., Nosikov, V. A. & Hadjiolov, A. A. (1984) *Biochim. Biophys. Acta* in the press

Hall, L. M. C. & Maden, B. E. H. (1980) *Nucleic Acids Res.* **8**, 5993–6005

Khan, M. S. N. & Maden, B. E. H. (1976) *FEBS Lett.* **61**, 10–13

Kominami, R. & Muramatsu, M. (1977) *Nucleic Acids Res.* **4**, 229–240

Kominami, R., Hamada, H., Fujii-Kuriyama, Y. & Muramatsu, M. (1978) *Biochemistry* **17**, 3965–3970

Kominami, R., Mishima, Y., Urano, Y., Sakai, M. & Muramatsu, M. (1982) *Nucleic Acids Res.* **10**, 1963–1979

Maden, B. E. H., Salim, M. & Summers, D. F. (1972) *Nature (London) New Biol.* **237**, 5–9

Maden, B. E. H., Moss, M. & Salim, M. (1982) *Nucleic Acids Res.* **10**, 2387–2398

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning*, pp. 1–520, Cold Spring Harbor Laboratory, New York

Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560–564

Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–580

Michot, B., Bachellerie, J. P. & Raynal, F. (1982) *Nucleic Acids Res.* **10**, 5273–5283

Michot, B., Bachellerie, J. P. & Raynal, F. (1983) *Nucleic Acids Res.* **11**, 3375–3391

Mishima, Y., Yamamoto, O., Kominami, R. & Muramatsu, M. (1981) *Nucleic Acids Res.* **9**, 6773–6785

Nikolaev, N., Birge, C. H., Gotoh, S., Glazier, K. & Schlessinger, D. (1975) *Brookhaven Symp. Biol.* **26**, 175–193

Olsen, G. J. & Sogin, M. L. (1982) *Biochemistry* **21**, 2335–2343

Pace, N. R., Walker, T. A. & Schroeder, E. (1977) *Biochemistry* **16**, 5321–5328

Perry, R. P. (1976) *Annu. Rev. Biochem.* **45**, 605–629

Peters, M. A., Walker, T. A. & Pace, N. R. (1982) *Biochemistry* **21**, 2329–2335

Prestayko, A. W., Tonato, M. & Busch, H. (1971) *J. Mol. Biol.* **47**, 505–515

Reddy, R. & Busch, H. (1983) *Prog. Nucleic Acid Res. Mol. Biol.* **30**, 127–162

Reddy, R., Rothblum, L. I., Subrahmanyam, C. S., Liu, M. H., Henning, D., Cassidy, B. & Busch, H. (1983) *J. Biol. Chem.* **258**, 584–589

Rothblum, L. I., Parker, D. L. & Cassidy, B. (1982*a*) *Gene* **17**, 75–77

Rothblum, L. I., Reddy, R. & Cassidy, B. (1982*b*) *Nucleic Acids Res.* **10**, 7345–7362

Skryabin, K. G., Krayev, A. S., Rubtsov, P. M. & Bayev, A. A. (1979) *Dokl. Akad. Nauk SSSR* **247**, 761–765

Subrahmanyam, C. S., Cassidy, B., Busch, H. & Rothblum, L. I. (1982) *Nucleic Acids Res.* **10**, 3667–3680

Veldman, G. M., Klootwijk, J., Heerikhuizen, H. & Planta, R. J. (1981*a*) *Nucleic Acids Res.* **9**, 4847–4862

Veldman, G. M., Klootwijk, J., Regt, V., Planta, R. J., Branlant, C., Krol, A. & Ebel, J. P. (1981*b*) *Nucleic Acids Res.* **9**, 6935–6952

Venkov, P. V. & Hadjiolov, A. A. (1969) *Biochem. J.* **115**, 91–94

Walker, T. A. & Pace, N. R. (1983) *Cell* **33**, 320–322

Walker, T. A., Johnson, K. D., Olsen, G. J., Peters, M. A. & Pace, N. R. (1982) *Biochemistry* **21**, 2320–2329

Walker, T. A., Endo, Y., Wheat, W. H., Wool, I. G. & Pace, N. R. (1983) *J. Biol. Chem.* **258**, 333–338

Walseth, T. F. & Johnson, R. A. (1979) *Biochim. Biophys. Acta* **562**, 11–31

Weinberg, R. A. & Penman, S. (1970) *J. Mol. Biol.* **47**, 169–178

Winicov, I. (1976) *J. Mol. Biol.* **100**, 141–155