Volume 10 Number 12 1982 Nucleic Acids Research

Nucleotide sequence of the region between the 18S rRNA sequence and the 28S rRNA sequence of rat ribosomal DNA

Chirala S.Subrahmanyam, Brandt Cassidy, Harris Busch and Lawrence I.Rothblum

Department of Pharmacology, Baylor College of Medicine, Houston, TX 77030, USA

Received 2 March 1982; Revised and Accepted 25 May 1982

ABSTRACT

The DNA sequence of the intragenic region of the rat 45S ribosomal RNA precursor was determined. This sequence contains 2282 nucleotides and extends from the conserved EcoR I site near the 3' terminus of 18S rRNA to 69 nucleotides downstream of the 5' terminus of 28S rRNA. The sequences corresponding to 18S and 5.8S rRNA were identified by comparison with previously published data. The 5' terminus of rat 28S rRNA was identified by S1 nuclease protection and reverse transcriptase elongation assays. The internal transcribed spacers were found to be 1066 and 765 nucleotides long and had little homology with those of Xenopus and yeast. Regions of sequence homology between rat and Xenopus were found at the junctions of the internal transcribed spacers with 18S, 5.8S and 28S rRNA. These homologies suggest that these sequences may function as recognition sites for the processing of the ribosomal precursor RNA.

INTRODUCTION

Ribosomal RNA is synthesized as a 45S preribosomal precursor and processed to yield the 18S, 5.8S and 28S ribosomal RNA molecules (1,2). The fidelity of formation of these products implies that there is a stringent requirement for the recognition of the termini of the preserved portions of the initial transcript. The nucleotide sequences of these processing points may provide the information required by these reactions in a manner analogous to that which has been hypothesized for the excision and ligation events involved in the processing of mRNA precursors (3). To evaluate processing sites, a portion of a cloned rat ribosomal DNA repeat was sequenced from the 3' terminus of 18S RNA to the 5' terminus of 28S RNA. This region contains four processing sites. We have used the terms of Hall and Maden (4) to define the spacer between the 18S and 5.8S RNA sequences and the spacer between the 5.8 and 28S RNA sequences as ITS1 and ITS2, the first

0305-1048/82/1012-3667$2.00/0

and second internal transcribed spacers, respectively.

The portion of the rat rDNA sequenced was examined for regions of interest and was compared with the available analogous portions of the Xenopus (4), yeast (5,6,7), Bombyx (8) and Lytechninus (9, and Hindenach and Stafford, personal communication).

MATERIALS AND METHODS

p59 is a fragment of ChR-B4 (10) bordered by EcoR I and BamHI sites, cloned in the vector pBR322. This subclone has been characterized by both restriction endonuclease digestion and Southern hybridization analysis and was shown to contain the 3' end of 18S RNA, 5.8S RNA and the 5' end of 28S rRNA (10). Starting at the EcoR I site near the 3' end of 18S rRNA the nucleotide sequence was determined by the method of Maxam and Gilbert (11). The 5' end of 28S rRNA was defined both by S1 endonuclease mapping (12) using a BstN1-Hinf I fragment labeled at the Hinf I site and by reverse transcriptase elongation (13) using a Sau 3A-Hinf I fragment labeled at the Hinf I site (Fig. 1). The sequence was analyzed by using the computer program of Korn and Queen (14).

RESULTS

Sequence determination. The DNA sequenced contained the 3' end of 18S, ITSI, 5.8S, ITSII and the 5' end of 28 rRNA coding regions and the sequencing strategy is shown in Fig. 1. The 2282 nucleotides of the sequence are shown in Fig. 2. The regions of mature ribosomal RNAs are indicated as boxed regions. Approximately 93% of the sequence was confirmed by sequencing both strands. The sequence between nucleotides 1604 and 1750 was confirmed by sequencing twice in the same direction and had no ambiguities. All of the restriction sites used in obtaining the sequence were confirmed by sequencing through these sites. There were some ambiguities when fragments with runs of "G" and "GA" were sequenced. These are indicated with wavy lines in Fig. 2. In these regions only the number of repeated nucleotides varied in different sequencing runs. Thus the number of "G"s starting at nucleotide 675 was found to be 8 or 9; the sequence at 1811 could be G_4TG_5 or G_4TG_6; and the run of 11 "GA" residues ending

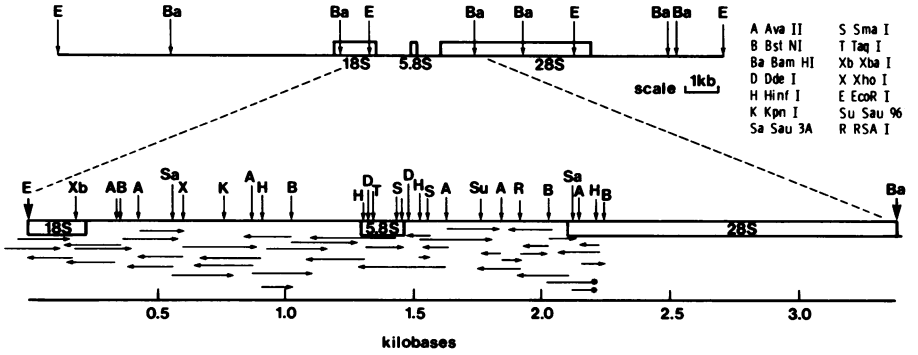

Figure 1 The structural organization of the rat ribosomal DNA repeat. The expanded portion of the map describes the subcloned EcoR I- Bam HI fragment that contains the intragenic region. The restriction sites that are relevant for the sequence determination only are indicated. Arrows indicate the direction and extent of the sequence obtained starting from a restriction site. * represents the two fragments used in the S1 mapping and reverse transcriptase elongation assay.

at nucleotide 1906 could be 10. The exact numbers of these nucleotides could not be confirmed from the complimentary strand sequence.

Location of the mature ribosomal RNA regions in the sequence. The terminal fifty nucleotides of the 3' end of 18S rRNA of rat were determined by Azad and Deacon (15). This sequence agreed with nucleotides 177-227 of the sequence presented in Fig. 2. It was concluded that nucleotides 1 through 227 constitute the 3' end of 18S rRNA. Comparison with the published sequence of rat 5.8S rRNA (16) showed nucleotides 1294 through 1449 to be the corresponding rDNA sequence. The location of 28S rRNA was first determined by comparison of the sequence of the rat rDNA with the Xenopus rDNA sequence (4).

The 5' terminal nucleotides of 28S rRNA were determined by S_1 nuclease mapping (12) and reverse transcriptase extension (13). The fragments used for this purpose were BstNI-Hinf I fragment or Sau 3A-Hinf I fragment both labeled at the Hinf I site (shown in Fig. 1). The sequence of Bst NI-Hinf I fragment which is complimentary to the 28S rRNA is shown in Fig. 3, as well as the results of S_1 mapping and reverse transcriptase extension. From the se-

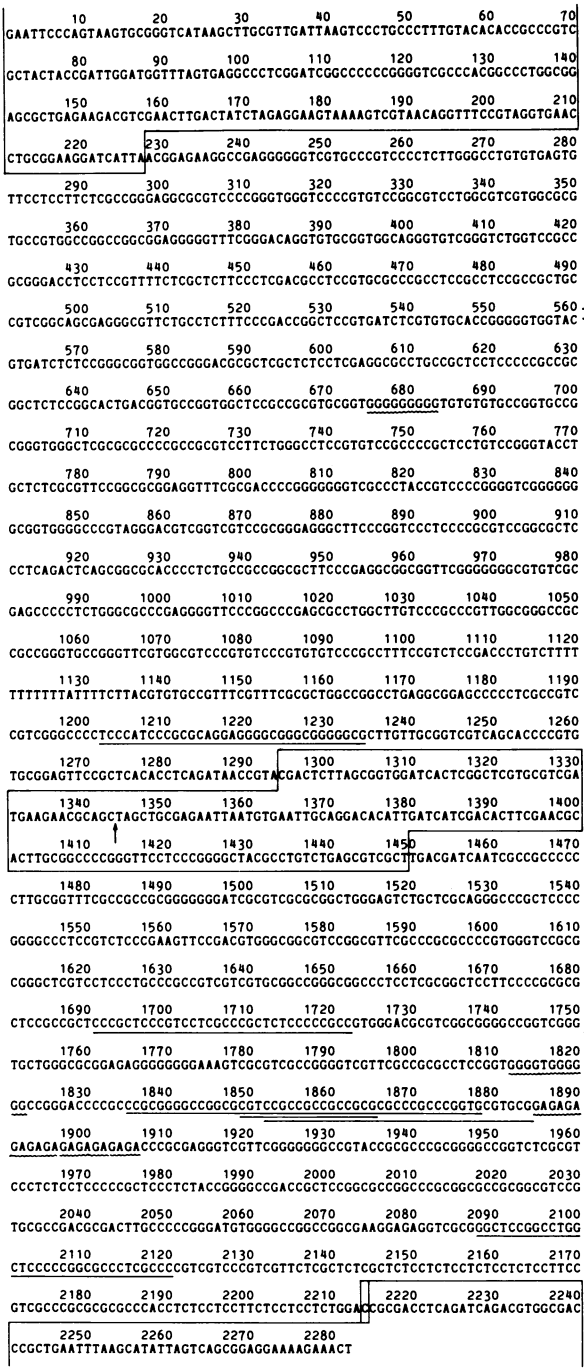

Figure 2

Sequence of the intragenic region of rat ribosomal DNA repeat. Boxed regions designate the 18S, 5.8S and 28S coding regions. The sequence shown is that of the non-coding strand of the DNA. The wavy lines indicate portions of the sequence where the number of nucleotides varied slightly from one sequence determination to the next. Maximum number of nucleotides determined are indicated in such instances. The arrow in 5.8S rDNA sequence indicates the location of the GC dinucleotide reported in the RNA sequence (16). The horizontal lines indicate the regions in the ITS that were greater than 75% homologous with regions of the ITS of Xenopus.

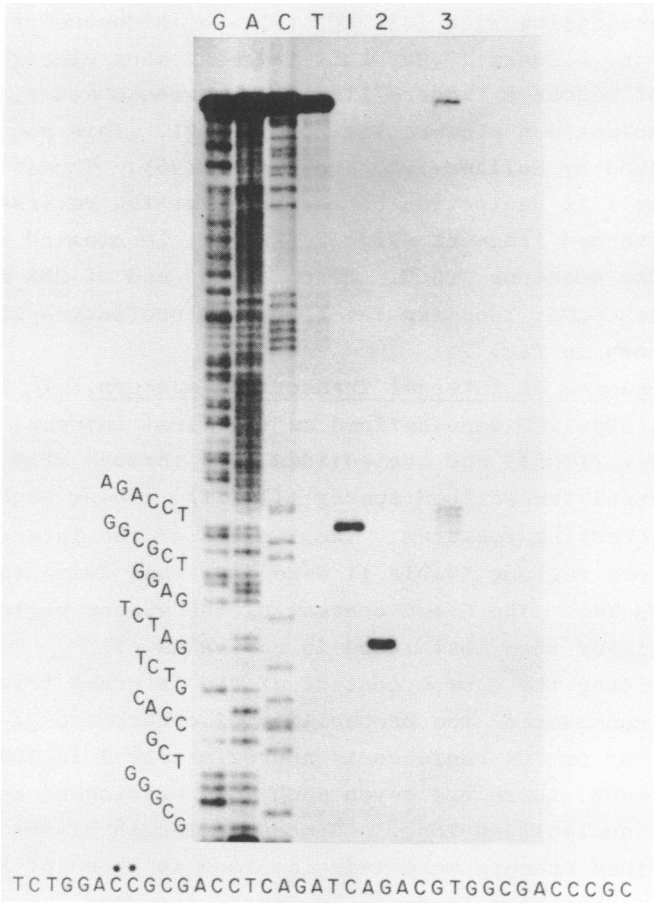

Figure 3

Determination of the 5' terminus of rat 28S rRNA. A Bst NI - Hinf I fragment (nucleotides 2103-2295) or Sau 3A - Hinf I fragment (nucleotides 2230-2295), both labeled at the Hinf I site using polynucleotide kinase were used in these experiments (see Fig. 1 for location of these fragments). Lane 1. Sau 3A - Hinf I fragment extended with reverse transcriptase after hybridization to 28S rRNA. Lane 2. Sau 3A - Hinf I fragment. Lane 3. Bst NI - Hinf I fragment hibridized to 28S and digested with S1 nuclease. Maxam and Gilbert sequence ladder of Bst NI - Hinf I fragment (GACT). This sequence corresponds to the coding strand of DNA. The non-coding strand sequence is written below the autoradiograph. *, indicates the 5' terminus of 28S rRNA.

quence specificity of the restriction enzyme Sau 3A (GATC); the Hinf-Sau 3A fragment labeled at its Hinf site has the 3' end sequence terminating with CCACGTCT (Fig. 3; sequence written alongside the ladders). But this fragment when electrophoresed alongside of sequence ladders lines up between G and A, hence it moved 1½ nucleotides slower (Fig. 3, Lane 2). This phenomenon has been described by Sollner-Webb and Reeder (25). Thus it was concluded after a 1½ nucleotide correction that the reverse transcriptase extended fragment (Fig. 3, Lane 1) terminated at the last G in the sequence TCGCG. Hence the 5' end of 28S rRNA has the sequence CGCGA; locating the 5' end at nucleotide 2215 in the sequence shown in Fig. 2.

The sequence of internal transcribed spacers. In Fig. 2, nucleotides 228-1293 were defined as the first internal transcribed spacer (ITS I) and nucleotides 1450 through 2214 as the second internal transcribed spacer (ITS II). These sequences had several interesting features. The lengths of the internal transcribed spacers regions (Table I) were more than twice those of Xenopus rDNA (4). The G + C content of the spacer regions is about 20% higher than that found in rat rRNAs (1).

When either the G or C content of the internal transcribed spacers is considered, the probability of occurrence of seven contiguous "G" or "C" residues is approximately 1 in 1000 nucleotides. However, there are seven such "G" "stretches" and no "C" stretches 7 nucleotides long. Since the length of the rat internal transcribed spacers were twice as long as those of Xenopus, it was possible that this increase in length occurred through the endoreduplication of a portion of the sequence. Hence we analyzed the sequence for direct repeats. There are about 270 direct repeats between 7 and 14 nucleotides long. This number is an overestimate because overlapping sequences were not excluded from the analysis. Except for the "G" stretches, each direct repeat occurs only once. These sequences are scattered in the ITS sequences without any apparent order. Thus it appears that the rat ITS sequence did not increase in size due to the endoreduplication of a smaller region.

There are two sequences of unique nucleotide composition. One consists of 11 "T's" starting at nucleotide 1117 and the

TABLE I

NUCLEOTIDE COMPOSITION RAT AND XENOPUS INTERNAL TRANSCRIBED SPACERS

	RAT		XENOPUS	
	ITS I	ITS II	ITS I	ITS II
A	5.3	5.5	12.5	4.9
C	37.4	43.2	42.8	53.6
G	37.1	36.8	41.1	34.3
T	20.1	14.6	3.6	7.2
Length	1067bp	765bp	559	256

Base composition is expressed as percent of total ITS I or ITS II nucleotides.

other consists of 10 or 11 alternating "GA" residues starting at nucleotide 1885. Comparison of the rat sequence with the sequence of the spacer region of Xenopus (4) resulted in the finding that there are several 10 nucleotide long sequences and thirty 20 nucleotide long sequences with greater than 75% homology between the two sequences. There are six 25 nucleotide long sequences with greater than 75% homology. These six sequences are underlined in Fig. 2. Homologous sequences between rat and Xenopus are scattered with no apparent order and do not suggest that these sequences have been conserved during evolution.

DISCUSSION

Sequence homologies at the 3' end of 18S rRNA. The last 20 nucleotides of the small subunit ribosomal RNA, 18S rRNA, are highly conserved both in eukaryotes and prokaryotes, except that the "CCTCC" sequence implicated in prokaryotic translation is missing in the eukaryotic ribosomal RNA molecules (18). Among the last 230 nucleotides of the 3' end of 18S rRNA compared in Fig. 4, there is greater than 90% homology among eukaryotes except the sequence from nucleotide 113 to nucleotide 158. A similar result was found when yeast 18S rRNA (5,6) and Xenopus

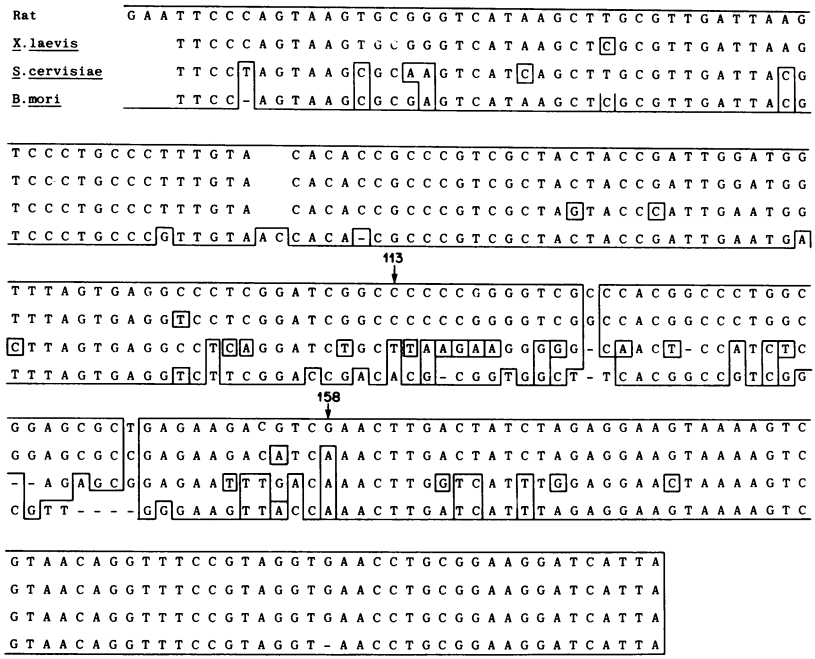

Figure 4 Comparison of the 3' terminal sequences of 18S rDNA from rat, X.laevis (4), S.cerevisiae (5), and B.mori (8). Homologous sequences are boxed. The variable region lies between nucleotides 113 and 158.

18S rRNA were compared (4). From nucleotide 113 through nucleotide 158, the sequence of eukaryotic 18S rRNAs diverge from each other. In this variable region, rat and yeast have only 50% homology. The rat and Xenopus sequences are 92% homologous in this variable region, demonstrating that this region reflects the relative evolutionary distances between these organisms (26).

5.8S rDNA. The 5.8S rDNA sequence agreed well with the published 5.8S rRNA sequence (16) except that a GC dinucleotide reported in the RNA sequence was not found in the rDNA sequence. The location of this GC dinucleotide in Figures 2 and 5 is indicated by an arrow. Fig. 5 is an autoradiograph of a sequencing gel of this region of 5.8S rDNA. This same discrepancy was observed when Xenopus rDNA was sequenced (4,17). Rat 5.8S rRNA separates into two isomeric bands on 7 M urea gels (19). It is

possible that there are two different 5.8S rRNAs and that one sequence contains this GC dinucleotide, or there is an error in the 5.8S rRNA sequence.

28S rDNA. The 5' end of rat 28S rRNA has been reported to be heterogeneous (20). We were able to define the 5' end as one of two nucleotides in the rDNA sequence, nucleotide 2214 or 2215 (Fig. 3). The sequence of the first 70 nucleotides at the 5' end of rat 28S rRNA is highly homologous (95%) with Xenopus. Yeast and Xenopus are only 80% homologous in this region (7). The overall sequence homology between the mature rRNAs (the 3' end of 18S, 5.8S and 5' end of 28S rRNAs) of rat and Xenopus are more extensive (95-98%) than that of rat and yeast (75-80%) or yeast and

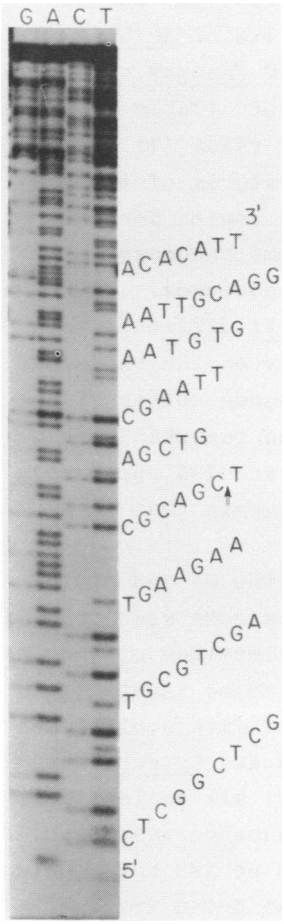

Figure 5 Nucleotide sequence of a portion of the 5.8S rDNA. A Hinf I fragment labeled at the 5' ends was cut with Dde I and the fragment corresponding to the non-coding strand of 5.8S rDNA was sequenced. The sequence written along side of the sequence ladders corresponds to the sequence between nucleotides 1314 to 1379 as shown in Fig. 2. The arrow indicates the position where the 'GC' dinucleotide reported to be present in 5.8S rRNA (16) would be found.

Xenopus (75-80%), which is consistant with the taxonomic distances of these species (26).

Internal transcribed spacer sequences. The rat ITS sequences are not homologous to those of Xenopus or yeast. This was not surprising, as in 1972 Brown et al (21) showed that there was little homology between the spacer sequences of two closely related Xenopus species. There are some characteristics of the ITS sequences that are specific for the ITS of each species and which distinguish them from the remainder of the gene. For example, in rat there are very few "A's" in the spacer sequence (5%, Table I), and there are "unique" sequences such as runs of "T" residues or "GA" residues, as were described earlier. A large number of "G strings" (7 consecutive G's) are found in the rat ITS, whereas there was a preference for "C strings" in the ITS of Xenopus (4). The yeast ITS is "U" rich (7). In both rat and Xenopus, the internal transcribed spacer regions are 75-80% "GC" (Table I) which is 15-20% more "GC" than the 60% of the mature rRNAs (1). It is not known whether any of these or any other features of the sequence make the spacer sequence a recognizable region for processing. It is also possible that there is some secondary structure conserved among the various spacer sequences. The high G + C content and the length of the ITS makes it difficult to select a significant secondary structure solely on the basis of computer simulation. As there was little sequence conservation within the internal transcribed spacers, it was remarkable that the sequences flanking the 3' termini of 5.8S and 18S apparently were conserved (Fig. 6). This observation suggests that these sequences are involved in processing.

When the DNA sequence immediately beyond the 3' end of 18S rRNA of rat was compared with the corresponding rDNA sequences of various species (Fig. 6A), the first three nucleotides of the internal transcribed spacer (ITS 1), were found to be identical in rat, Xenopus and Bombyx, i.e. A-C-G. The first five nucleotides were found to be A-C-G-Pu-Pu. In rat, Xenopus and Lytechninus, there is an A-G-A or G-A-G-A within these first six nucleotides. Beyond these five nucleotides, no consensus sequence was found.

The sequence similarity beyond the 3' end of 18S rRNA could have arisen by chance. The probability that an A-C-G triplet

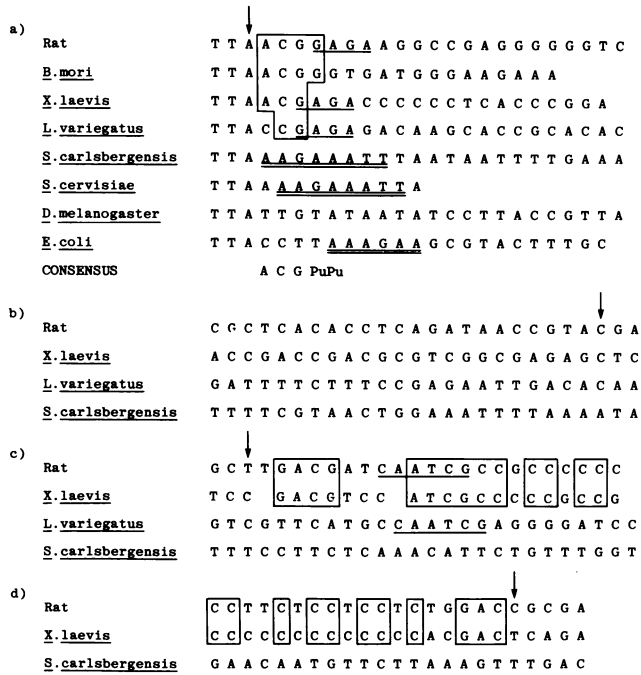

Figure 6 Comparison of the sequences of the internal transcribed spacers immediately adjacent to the 18S, 5.8S and 28S rRNA sequences of rat, B.mori, (8), X.laevis (4), L.-variegatus (9, and Hindenach and Stafford personal communication), S.carlsbergensis (6), S.cerevisiae (5), D.melanogaster (24) and E.coli (22). The arrow indicates in a) the 3' end of 18S rRNA, b) the 5' end of 5.8S rRNA, c) the 3' end of 5.8S rRNA and in d) the 5' end of 28S rRNA. Homologous sequences are indicated either by enclosing those sequences in boxes or by single or double lines under the sequences.

would be present immediately adjacent to the 3' end of 18S rRNA is 1/64, and that the same triplet would occur by random chance in three species of rDNA is $(1/64)^3$. In the consensus sequences shown in Fig. 2, the triplet ACG occurs at a frequency of 0.45. Thus, the consensus sequence appears to be a preferred or conserved sequence rather than one of random probability. Therefore, sequences similar to the consensus sequence A-C-G-Pu-Pu could be recognition sites for processing at the 3' end of 18S rRNA in higher eukaryotes. The sequences A-A-G-A-A or A-A-A found in E. coli and yeast could be involved in a similar way in prokaryotes.

The sequence of the terminal 20 nucleotides of 18S rRNA, which is more than 95% conserved in eukaryotes, could constitute a part of this recognition site for processing.

At the 5' end flanking regions of 5.8S ribosomal RNA (Fig.6B), there is very little sequence conservation.

The 3' end flanking region of rat 5.8S was compared with the sequence from Xenopus and sea urchin for possible homologies. In rat and Xenopus, G-A-C-G, A-T-C-G-C-C, and C-C were common sequences (boxed sequences in Fig. 6C). In rat and sea urchin, C-A-A-T-C-G is common in this region. Assuming that this homology is not a chance occurrence, at least in Xenopus and rat, the sequence at the 3' end of 5.8S and the corresponding flanking region could constitute a recognition sequence for processing.

The 5' end flanking regions of 28S rDNA of rat and Xenopus are similar (Fig. 6D). There is a pyrimidine stretch followed by G-A-C immediately 5' to the 28S coding sequences.

In yeast the 5' termini of precursor and mature rRNA, the 5.8S, 17S and 26S, share a consensus sequence (7). Similarly the 3' termini of 5.8S, 7S, 17S and 18S rRNA of yeast have a consensus sequence (6,7). Except for the conserved 3' end of 18S rRNA, the sequences at the termini of mature ribosomal RNAs of rat and yeast are different. Other consensus sequences were not found among the sequences of rat rRNA termini shown in Fig. 2.

Veldman et al (7) have proposed that in yeast the 3' end of 5.8S rRNA and 5' end of 26S rRNA could interact to form a secondary structure. A similar structure can be constructed for the 5' end of rat 28S rRNA and the 3' end of 5.8S rRNA:

$$\text{5.8S} \quad {}^{3'}\text{OH}\text{TCGCTGCGAGTCT-GTCCGCATCG--GGGCCC}^{5'}$$

$$\text{28S} \quad {}^{5'}\text{p}\text{CGCGAC-CTCAGATCAGACGTGGCGACCCGCT}_{3'}$$

In the absence of further information on the processing sites that generate the secondary precursors and any evidence of definitive secondary structure interactions, our findings are consistant with the possibility that whether or not the secondary structure of the ribosomal RNA precursors are conserved, the nucleotide sequence of the processing sites are involved in generating the proper termini of the mature ribosomal RNA. The

homologies in the regions which flank 18S, 5.8S and 28S rRNA are more pronounced in the comparisons of the rat and Xenopus sequences than when the other eukaryotic sequences were compared. This finding suggests that processing mechanisms may vary with phylogeny and that the higher vertebrates might share a common process.

ACKNOWLEDGEMENTS

These studies were supported by the Cancer Research Grant CA-10893, P9; the Bristol-Meyers Fund; the Pauline Sterne Wolff Memorial Foundation and the Michael E. DeBakey Medical Foundation.

REFERENCES

1. Busch, H. and Smetana, K. (1970) The Nucleolus. Academic Press, New York,
2. Perry, R. P. (1976) Ann. Rev. Biochem., 45, 605-629.
3. Breathnach, R. and Chambon, P. (1981) Ann. Rev. Biochem., 30, 349-383.
4. Hall, L. M. C. and Maden, B. E. (1980) Nucleic Acids Res., 8, 5993-6005.
5. Skryabin, K. G., Krayev, A. S., Rubtsov, P. M. and Bayev, A. A. (1979) Dokl. Akad. Nank. USSR. 247, 761-765.
6. Veldman, G. M., Brand, R. C., Klootwijk, J. and Planta, R. (1980) Nucleic Acids Res., 8, 2907-2920.
7. Veldman, G. M., Klootwijk, J., Heerikhuizen, H. V. and Planta, R. J. (1981) Nucleic Acids Res., 9, 4847-4862.
8. Samols, D. R., Hagenbuchle, O. and Gage, L. P. (1979) Nucleic Acids Res., 7, 1109-1119.
9. Hindenach, B. R. and Stafford, D. (1981) Fed. Proc., 40, 1752.
10. Rothblum, L. I., Parker, D. L. and Cassidy, B. (1982) Gene, 17, 74-77.
11. Maxam, A. and Gilbert, W. (1980) Methods in Enzymology, 65, 499-580.
12. Berk, A. J. and Sharp, P. A. (1977) Cell, 12, 721-732.
13. Bina-Stein, M., Thoren, M., Salzman, N. and Thompson, J. A. (1979) Proc. Natl. Acad. Sci. USA, 76, 731-735.
14. Queen, C. L. and Korn, L. J. (1980) In, Methods in Enzymology 65, 595-609.
15. Azad, A. A. and Deacon, N. J. (1980) Nucleic Acids Res., 8, 4365-4376.
16. Nazar, R. N., Sitz, T. O. and Busch, H. (1975) J. Biol. Chem., 250, 8591-8597.
17. Boseley, P. G., Tuyns, A. T. and Birnstiel, M. L. (1978) Nucleic Acids Res., 5, 1121-1137.
18. Hagenbuchle, O., Santer, M., Steitz, J. A. and Mans, R. J. (1978) Cell, 13, 551-563.
19. Reddy, R., Li, W. Y., Henning, D., Choi, Y. C., Nohga, K. and Busch, H. (1981) J. Biol. Chem. 256, 8452-8457.
20. Nazar, R. N. (1977) Biochemistry, 16, 3215-3219.
21. Brown, D. D., Wensink, P. C. and Jordan, E. (1972) J. Mol. Biol., 63, 57-73.
22. Brosius, J., Dull, T. J., Sleeter, D. D. and Noller, H. F.

(1981) J. Molec. Biol. 148, 107-127.
23. Bowman, L. H., Rabin, B. and Schlessinger, D. (1981) Nucleic Acids Res., 9, 4951-4966.
24. Jordan, B. R., Latil-Damotte, M. and Jourdan, R. (1980) FEBS Lett., 117, 227-231.
25. Sollner-Webb, B. and Reeder, R. H. (1979) Cell, 18, 485-499.
26. Hori, H. and Osawa, S. (1979) Proc. Natl. Acad. Sci., USA, 76, 381-385.