MOLECULAR AND CELLULAR BIOLOGY, Aug. 1983, p. 1488–1500
0270-7306/83/081488-13$02.00/0

Vol. 3, No. 8

Mouse rDNA: Sequences and Evolutionary Analysis of Spacer and Mature RNA Regions

WILLIAM E. GOLDMAN, GREGORY GOLDBERG,† LEWIS H. BOWMAN,‡ DENNIS STEINMETZ, AND DAVID SCHLESSINGER*

Department of Microbiology and Immunology, Division of Biology and Biomedical Sciences, Washington University School of Medicine, St. Louis, Missouri 63110

Received 8 March 1983/Accepted 26 May 1983

Two regions of mouse rDNA were sequenced. One contained the last 323 nucleotides of the external transcribed spacer and the first 595 nucleotides of 18S rRNA; the other spanned the entire internal transcribed spacer and included the 3′ end of 18S rRNA, 5.8S rRNA, and the 5′ end of 28S rRNA. The mature rRNA sequences are very highly conserved from yeast to mouse (unit evolutionary period, the time required for a 1% divergence of sequence, was 30×10^6 to 100×10^6 years). In 18S rRNA, at least some of the evolutionary expansion and increase in G+C content is due to a progressive accretion of discrete G+C-rich insertions. Spacer sequence comparisons between mouse and rat rRNA reveal much more extensive and frequent insertions and substitutions of G+C-rich segments. As a result, spacers conserve overall G+C richness but not sequence (UEP, 0.3×10^6 years) or specific base-paired stems. Although no stems analogous to those bracketing 16S and 23S rRNA in *Escherichia coli* pre-rRNA are evident, certain features of the spacer regions flanking eucaryotic mature rRNAs are conserved and could be involved in rRNA processing or ribosome formation. These conserved regions include some short homologous sequence patterns and closely spaced direct repeats.

rDNA is evolutionarily conserved as a multigene family. The long precursor rRNA transcripts always contain a spacer region at the 5′ end (external transcribed spacer) followed by the rRNA sequence for the smaller ribosomal subunit (16S to 18S). Another spacer (internal transcribed spacer [ITS]) is then transcribed before the rRNA for the larger subunit (23S to 28S). In addition, smaller RNAs usually interrupt the internal spacer: specific tRNAs between the bacterial 16S and 23S rRNAs, and 5.8S rRNA between the eucaryotic 18S and 28S rRNAs. The precursor RNA is cleaved at a number of sites to yield the mature RNAs, and spacer sequences are totally degraded during processing (16).

As one might expect for a structural RNA, the similarity in precursor organization and processing extends to features of the mature rRNA primary sequence. For example, particular bases are methylated, and some sequences are highly conserved in similar rRNAs from different species (8). Conservation of secondary structure features has also been suggested (9). The high degree of evolutionary relatedness among mature rRNAs is undoubtedly due to their important role in the synthesis of proteins.

Nevertheless, mature rRNA species tend to become longer and, among the eucaryotes, increasingly G+C rich during evolution, for unknown reasons. For example, the *Escherichia coli* 23S rRNA is 3,000 nucleotides long with 53.5% G+C; the human 28S rRNA is 4,200 nucleotides long with 68% G+C (8). Transcribed spacer sequences show an even greater increase in size and G+C content during evolution. The number of spacer nucleotides in the primary rRNA transcript increases from about 1,400 in *E. coli* to about 6,800 in humans, and the G+C content increases 1.5-fold (8).

Unlike mature rRNAs, however, transcribed spacer sequences diverge rapidly. Not only is there no detectable homology between yeast and *Xenopus laevis* spacers, but eucaryotic spacers have no known function. In contrast, *E. coli* spacer sequences flanking 16S and 23S rRNA base pair to form double-stranded stems which are the sites of initial cleavages in *E. coli* pre-rRNA (4). Such stems have not been detected in either yeast or *X. laevis* pre-rRNAs, but eucaryotic spacer sequences may have other conserved

† Present address: Division of Dermatology, Department of Medicine, Washington University School of Medicine, St. Louis, MO 63110.

‡ Present address: Department of Biology, University of South Carolina, Columbia, SC 29208.

features that might specify processing cleavage sites, regulate the structure and replication of rDNA, or control transport of nascent ribosomes to the cytoplasm.

To identify conserved features and the basis of the evolutionary trends in rDNA length and G+C content, we have sequenced 3,282 nucleotides of mouse rDNA; the regions include the entire internal transcribed spacer and the sequence surrounding the 5′ terminus of 18S rRNA (see Fig. 1). The primary and possible secondary structures of the mouse sequences were then extensively analyzed and compared with those from *E. coli*, yeast, *X. laevis*, and rat.

These analyses suggest that the primary sequence of the ITS diverges very rapidly. However, high G+C content and a capacity to form stable secondary structures are maintained in spacer sequences by the continued insertion and deletion of short G+C-rich regions. Other structural features are evident which, combined with the analysis of pre-rRNA processing sites, permit a search for possible cleavage signals (2, 3).

MATERIALS AND METHODS

Mouse rDNA sequencing. Regions near the termini of mouse (MOPC myeloma) 18S, 5.8S, and 28S rRNAs were sequenced from rDNA subclones 5A, 6, and 7, all containing inserts into plasmid pBR322 as described by Bowman et al. (2). Each subclone was mapped extensively with restriction enzymes; those sites used for end labeling of sequenced fragments are indicated in Fig. 1. Sequencing was essentially as described by Maxam and Gilbert (19), with restriction fragments 5′ end labeled with polynucleotide kinase. Labeled fragments were extracted from polyacrylamide gels by electroelution (B. K. Saha, S. Stretlow, and D. Schlessinger, J. Biochem. Biophys. Methods, in press) and were suitable for subsequent restriction enzyme cleavage, strand separation, or sequencing. Computer-predicted restriction sites calculated from the final confirmed sequence matched precisely with actual cleavage patterns for 31 restriction enzymes.

Computer analysis. Sequence editing and restriction enzyme analysis were facilitated by modifications by W. Barnes (Washington University School of Medicine) of the programs of Staden (26). Dot matrix analysis of sequence homologies (see Fig. 5) was generated by a program of W. Barnes and G. Zyda (unpublished, Washington University School of Medicine). More specific alignment of the rDNA sequences was accomplished with the aid of a gapping program (15; adapted for use here by M. Brandenberg) which selects regions of high homology and allows the user to introduce deletions for the best fit of those regions. Like the dot matrix program, this program was most useful for the rDNA sequence when we searched for regions with homology of at least 20 out of 30 bases. Prediction of possible pre-rRNA secondary structure was assisted by the SEQ program of the MOLGEN project (Stanford University) at the National Institutes of Health SUMEX-AIM facility. Dyad symmetries constructed by SEQ were plotted on a matrix (see Fig.

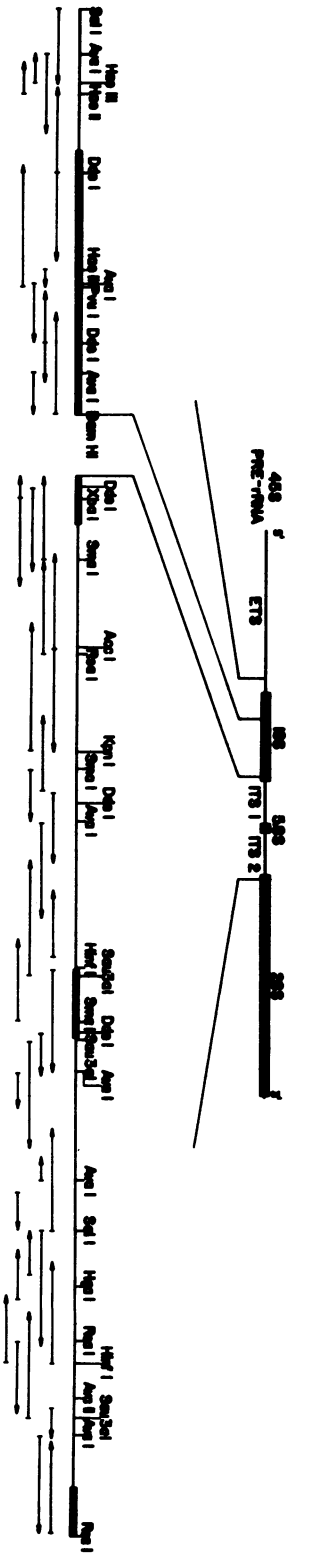


FIG. 1. Sequencing strategy for regions near pre-rRNA cleavage sites. The restriction enzyme sites indicated are those used for 5′ end labeling of fragments to be sequenced. Arrows represent length and direction (5′ to 3′) of sequence derived from these fragments.

7) with an adjunct program by M. Brandenberg. This program sorts the data into groups based on the predicted total free energy of each dyad symmetry and then plots the locations of base-paired regions with respect to each sequence. Additional programs by M. Brandenberg aided in displaying homologous sequence alignment and permitted searches for perfect and imperfect sequence repeats. The SEQ program was also used to determine nucleotide frequencies and to help identify homologous sequences between species.

Calculations of the rate of rDNA evolution. Sequence alignments of mature rDNA from different species and ITSs from mouse and rat have permitted us to estimate the rate of evolution for these regions. In calculating conservation and divergence of a given pair of sequences, a gap introduced in the alignment was scored as a single mismatch, and the percentage of divergence was corrected for multiple base change events (20). From fossil records of evolutionary divergence of species (10, 18), the corrected percentage of divergence can be used to estimate the unit evolutionary period (UEP), the time (in millions of years) required to accumulate a 1% divergence between two initially identical sequences (36).

RESULTS

Mouse rDNA sequences. Figure 1 shows the strategy used for sequencing the indicated rDNA segments. The Maxam-Gilbert method was adequate for these studies, although the high G+C content of the internal spacer tended to produce compressed bands on gels. Therefore, repeated analyses of both strands from multiple restriction sites was often necessary to yield an unequivocal sequence. The sequences derived are listed in Fig. 2. Termini of the mature rRNAs were located along the rDNA sequence from their previously determined terminal sequences (13, 14) or by S1 nuclease mapping techniques (2, 3). The G+C content of the various mature rRNA segments varied from 45.9 to 56.2%; the G+C content of spacer segments was much higher, ranging from 70.1 to 75.3%.

Analysis of mature rRNA sequences. The sequences determined for portions of mouse 18S, 28S, and 5.8S rRNAs were compared with the published sequences of *E. coli* (5), yeast (21, 22, 25, 33, 34), *X. laevis* (17, 23, 24), and rat (7, 29). As expected, alignment of these sequences (Fig. 3) indicates that the segments of mouse mature rRNAs are highly homologous to the mature rRNAs from other eucaryotic species (70 to 87% homologous to yeast, 89 to 100% homologous to *X. laevis*). This suggests that the majority of these sequences are critical for eucaryotic ribosomal assembly, structure, or function. In contrast, mouse and *E. coli* rRNAs were only 45.4% homologous in the best computer-assisted alignment of this region, although some regions of better homology were evident, such as nucleotides 367 through 384 and 767 through 820.

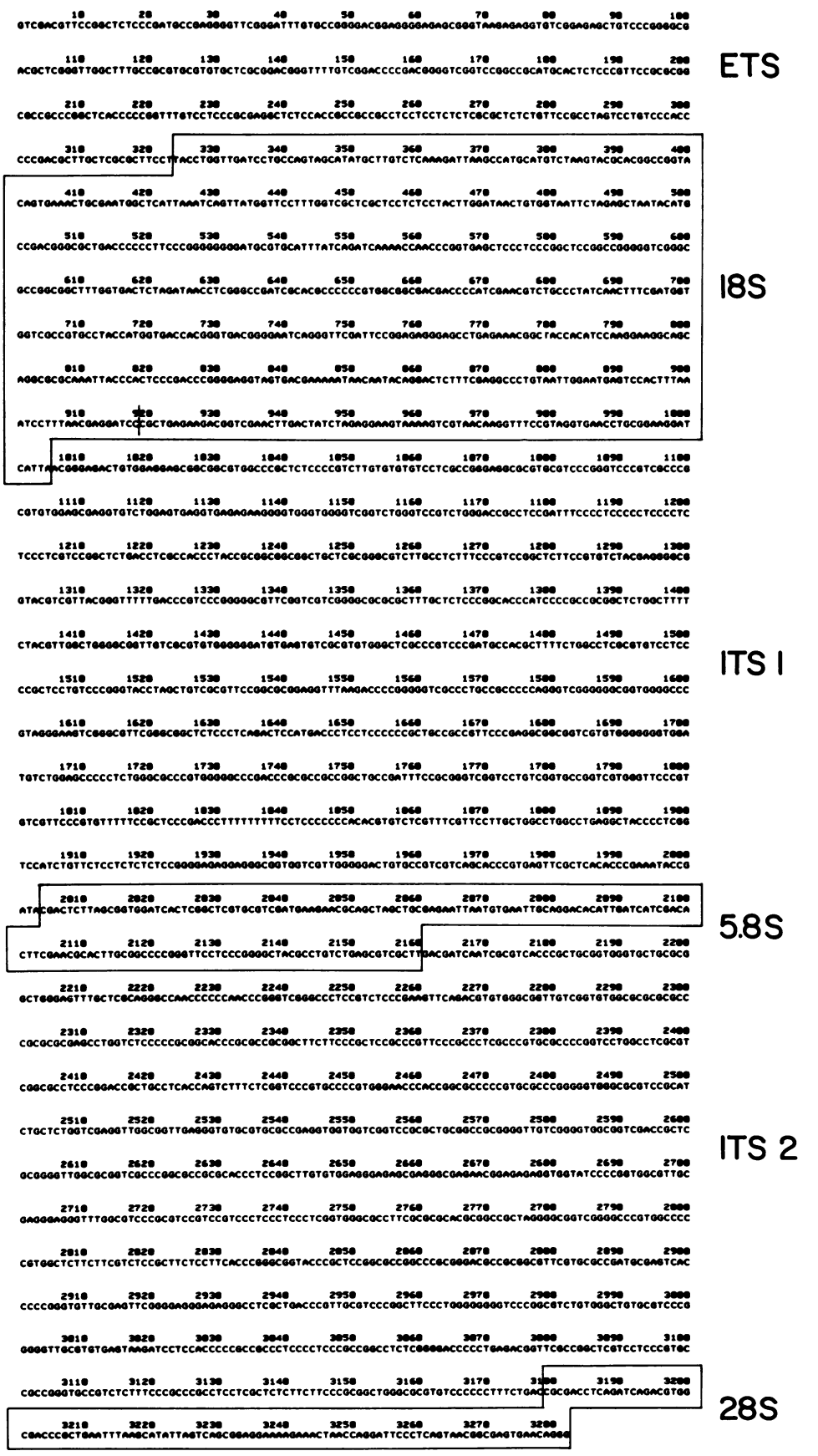
Of particular interest are the very G+C-rich insertions in mouse 18S rRNA as compared with *X. laevis* or yeast 18S rRNA. Figure 4 shows that the size and number of these insertions progressively increase from yeast to mouse sequences and that these insertions can form stable hairpin structures. The increasing size of mature rRNA during evolution may thus be due, at least in part, to the insertion of these G+C-rich sequences (see below).

Analysis of spacer rRNA sequences. The transcribed spacer sequences have diverged so rapidly that no alignment could be constructed between mouse and *X. laevis* or between mouse and yeast ITS sequences. Only when the mouse ITSs were compared with those of rat could an alignment be generated. Figure 5 displays the mouse-rat ITS comparison on a matrix, in which every line indicates a sequence homology of at least 20 out of 30 consecutive nucleotides. Because the ITS sequences are of comparable length in mouse and rat, homologies near the diagonal of the graph are regions similar in location as well as sequence.

An optimized sequence comparison of the mouse and rat ITS regions was derived from Fig. 5 and from searches with a computer program that introduces gaps to improve sequence alignment. The result (Fig. 6) indicates which stretches of sequence are more conserved between the two species; a histogram below the matrix plot in Fig. 5 shows the extent of divergence across segments of the spacer (calculated from the alignment of Fig. 6). The histogram also maps four regions (a, b, c, and d) of the rat ITS previously identified as highly homologous (~75%) to *X. laevis* ITS sequences (29). However, none of these regions corresponds to those ITS sequences that are the most conserved between rat and mouse. We have not found any long spacer sequences that are highly conserved in all three species.

The spacer sequences near the termini of mature mouse and rat rRNA sequences are no more homologous than other regions of their ITSs. However, a role for these particular sequences in rRNA processing or ribosome formation is more likely because of homologies with *X. laevis* flanking spacer sequences; the similarities extend for 10 to 20 nucleotides after the 3'

FIG. 2. Mouse rDNA sequence. Termini of mature RNA determined from known rRNA sequences (see text) and from S1 nuclease analysis (2). ETS, External transcribed spacer. Note that the sequence of 18S rRNA is interrupted at the vertical line (nucleotide 918); the remaining 18S sequence is from the 3' end of the rRNA.


ETS
18S
ITS 1
5.8S
ITS 2
28S

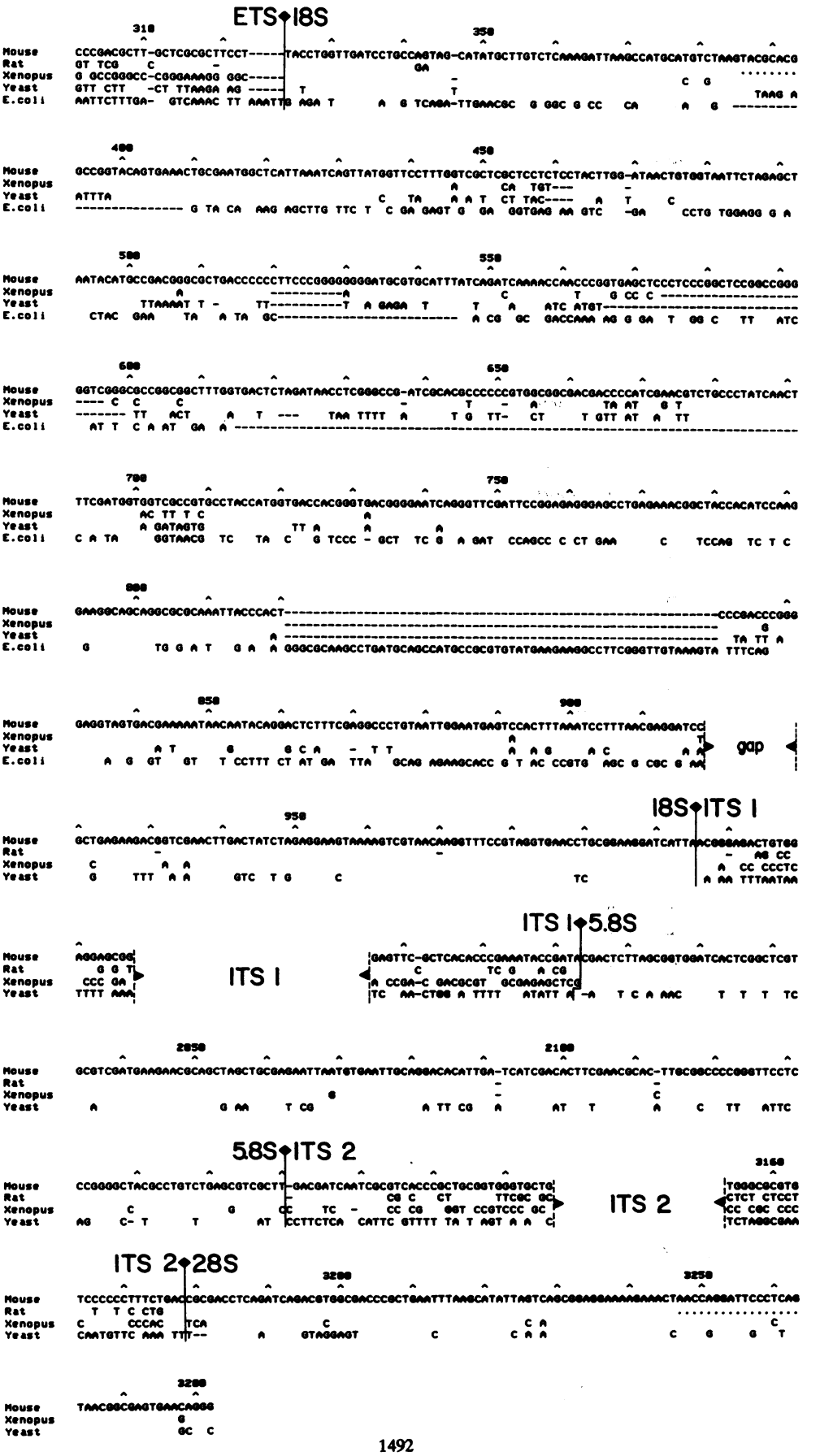

ETS
18S
gap
18S
ITS 1
ITS 1
5.8S
5.8S
ITS 2
ITS 2
ITS 2
28S
Mouse
Rat
Xenopus
Yeast
E.coli

FIG. 4. Location and possible structure of the insertions found in mouse and *X. laevis* 18S rDNA as compared with yeast 18S rDNA. The vertical arrows indicate the borders of homologous sequences, at which inserts occur in higher organisms. Asterisks denote the actual inserted nucleotides. Numbering corresponds to the sequence in Fig. 2.

end of 5.8S rRNA and possibly for 4 to 7 nucleotides after the 3′ end of 18S rRNA (Fig. 3). In addition, all three vertebrate species have a block of pyrimidines 5 to 15 nucleotides upstream from the 5′ terminus of 28S rRNA, and all the eucaryotes show at least five additional purines beyond the 3′ terminus of 18S rRNA.

Analyses of possible secondary structure. Possible dyad symmetries near the ends of mature rRNA termini and in the adjoining spacers were determined by computer-assisted analysis. The most stable structures ($\Delta G < -15$ kcal) are represented in a matrix (Fig. 7) for the available sequences of *X. laevis* and mice. Hairpin loops are located along the diagonal, indicating base pairing between nearby pre-rRNA sequences. Larger loops arising from possible long-range base pairing appear at appropriate distances from the diagonal. For each loop, the thickness of the enclosing oval corresponds to the ΔG^0 of the base-paired stem (see legend to Fig. 7).

Amid the complexity of the graphs, some features stand out. First, the potential for structure formation is very great within each spacer region, and many stable base pairings are possible, even between ITS 1 and ITS 2. Nearly every stretch of 20 nucleotides can potentially match with at least one other stretch, and some short regions show many possible matches with sequences scattered throughout the pre-rRNA.

Second, the capacity for the formation of stable secondary structure is far lower in mature rRNA sequences. These regional differences in the potential for secondary structure are most extreme in the *X. laevis* matrix, in which the complete 18S rRNA sequence can be compared with three spacer regions.

Third, the regions with a high density of potential secondary structure approach, but do not usually include, the termini of certain mature rRNA species. In fact, windows of sequence with far less capacity to form dyads extend up to 60 nucleotides from the termini into spacer regions. In some cases, the border of a window is very well defined, e.g., the spacer sequence proximal to the 5′ end of 5.8S rRNA (Fig. 7). The same windows appear in the matrix analyses of rat rDNA (data not shown).

Fourth, no dyad symmetries are observed with the stability, length, or relative position of

FIG. 3. Alignment of rDNA sequences corresponding to mature rRNA of different species. Mouse rDNA is compared with available sequences of rats, *X. laevis*, and *Saccharomyces cerevisiae* (ITS sequences from *Saccharomyces carlsbergenesis*); an alignment to *E. coli* 16S rRNA is also included for the 5′ portion of mouse 18S rRNA. Deletions (dashed lines) have been introduced in each sequence to maximize homology. The mouse rDNA sequence is presented in the top row, and numbers relate directly to the nucleotide sequence number in Fig. 2. Nucleotides that differ from the mouse rDNA sequence are indicated in the corresponding rows for each species, and spaces represent complete identity with the mouse sequence. Dots indicate absence of sequence data for a given region.

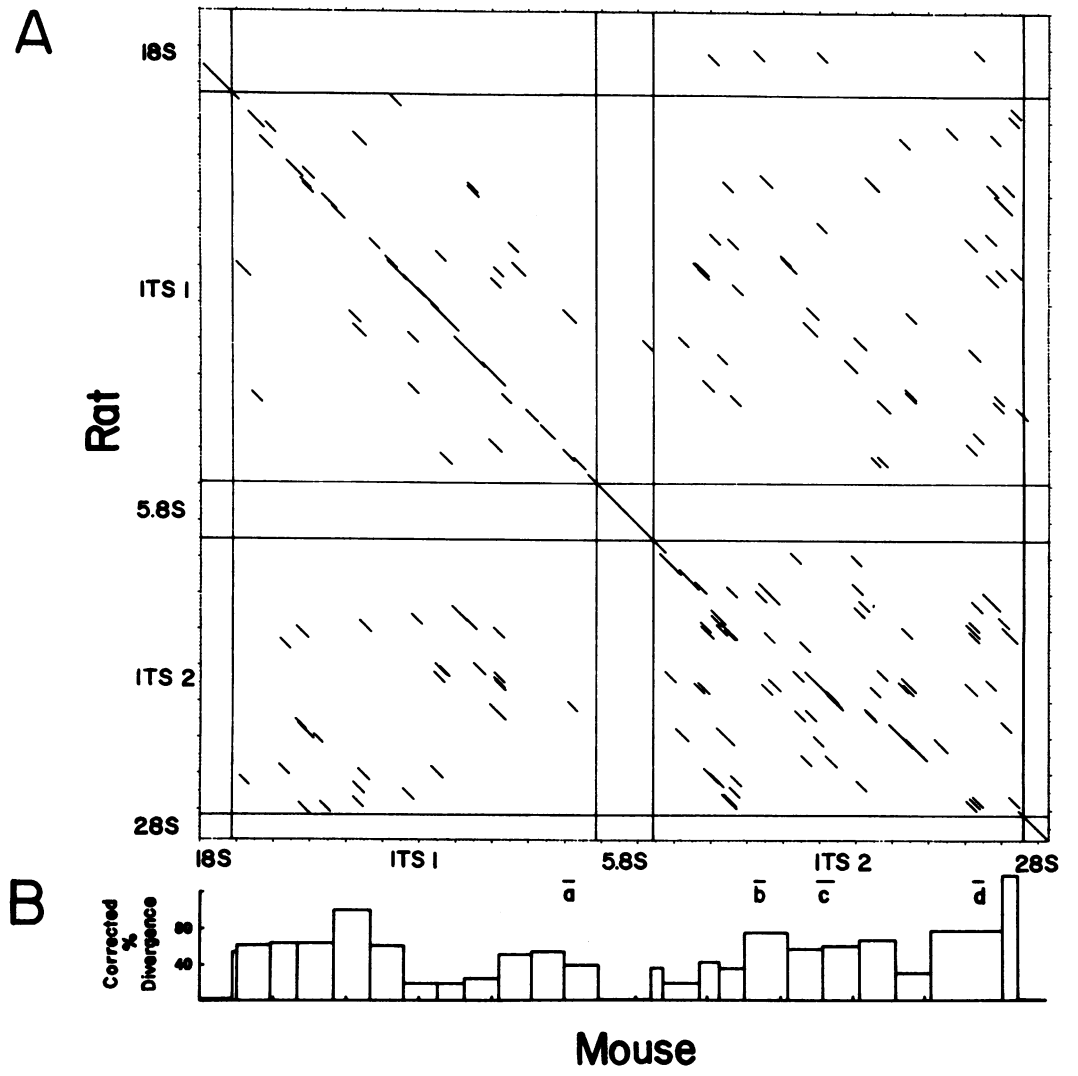


FIG. 5. Matrix comparison of mouse and rat rDNA sequences spanning the ITSs. (A) Each line indicates a region of homology of at least 20 out of 30 bases identical. Homologies appearing close to the diagonal are those that also correspond closely in relative location within the two sequences. (B) Bars represent corrected percentage of divergence (see text) between aligned regions of the mouse and rat rDNAs. The coordinates for sequence location match those in A, and each bar corresponds to a line of mouse rDNA sequence from Fig. 5 (mature rRNA sequences are represented by separate bars). Lines a through d show the location of mouse rDNA regions corresponding to those of rat spacer rDNA that have relatively high homology to *X. laevis* rDNA (29).

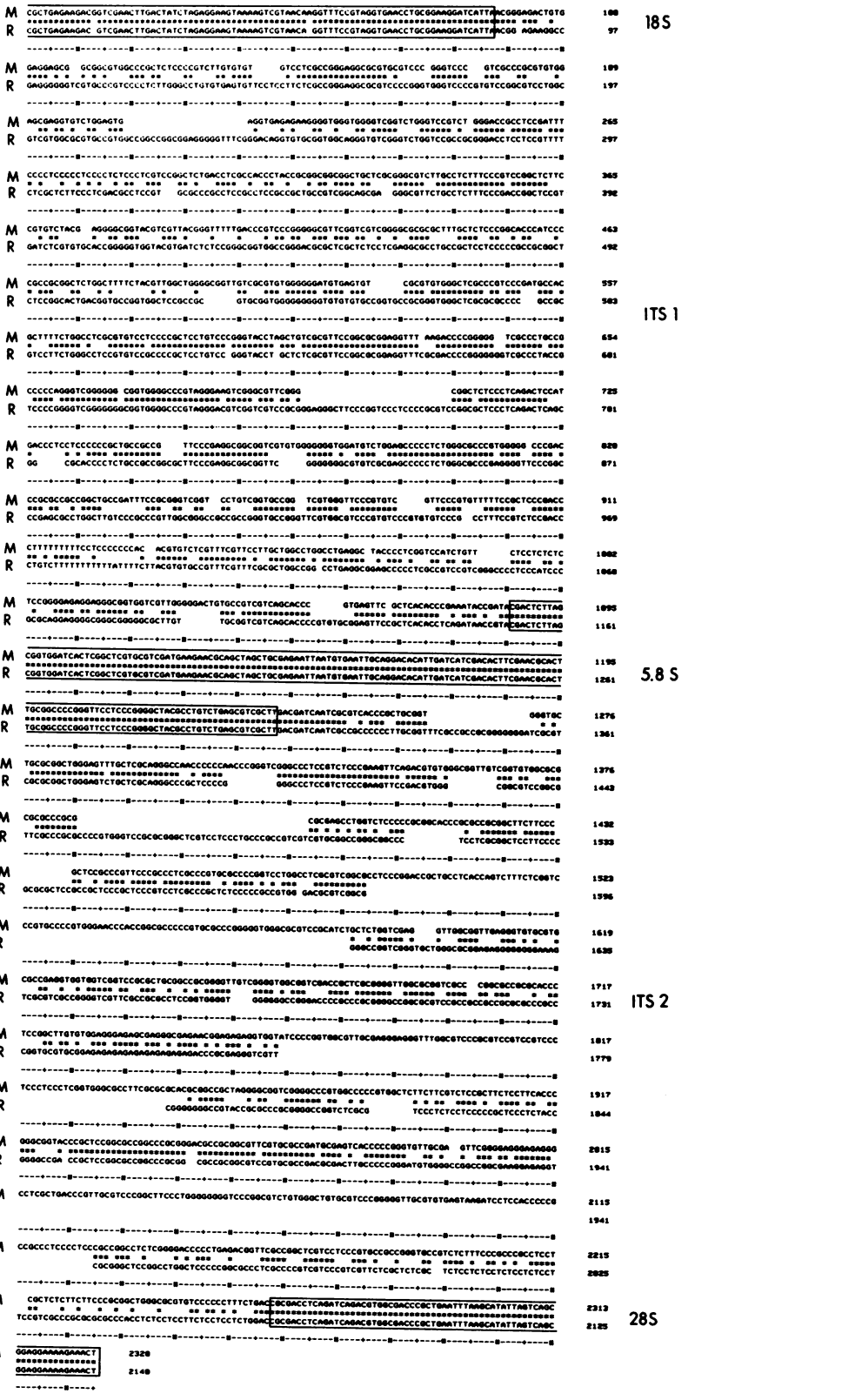
those that form stems to enclose 16S and 23S rRNA sequences in *E. coli* pre-rRNA (4).

DISCUSSION

A curious feature of phylogeny, namely the progressive increase in length and G+C content of mature and spacer rRNA, can now be examined by comparative sequence analysis. One can also assess whether the evolution of any spacer sequences is constrained in a way which implies functions in rRNA metabolism.

Evolution of mature rDNA sequences. Sequence comparisons of procaryotic rDNAs indicate that the primary sequences of rRNA are highly conserved. *E. coli* and *Bacillus brevis*

FIG. 6. Computer-assisted alignment of mouse (M) and rat (R) rDNA sequences spanning ITS 1, ITS 2, and adjacent mature rRNAs. Asterisks indicate homologous bases; gaps in both sequences have been introduced to maximize homology.


18S
ITS 1
5.8 S
ITS 2
28S

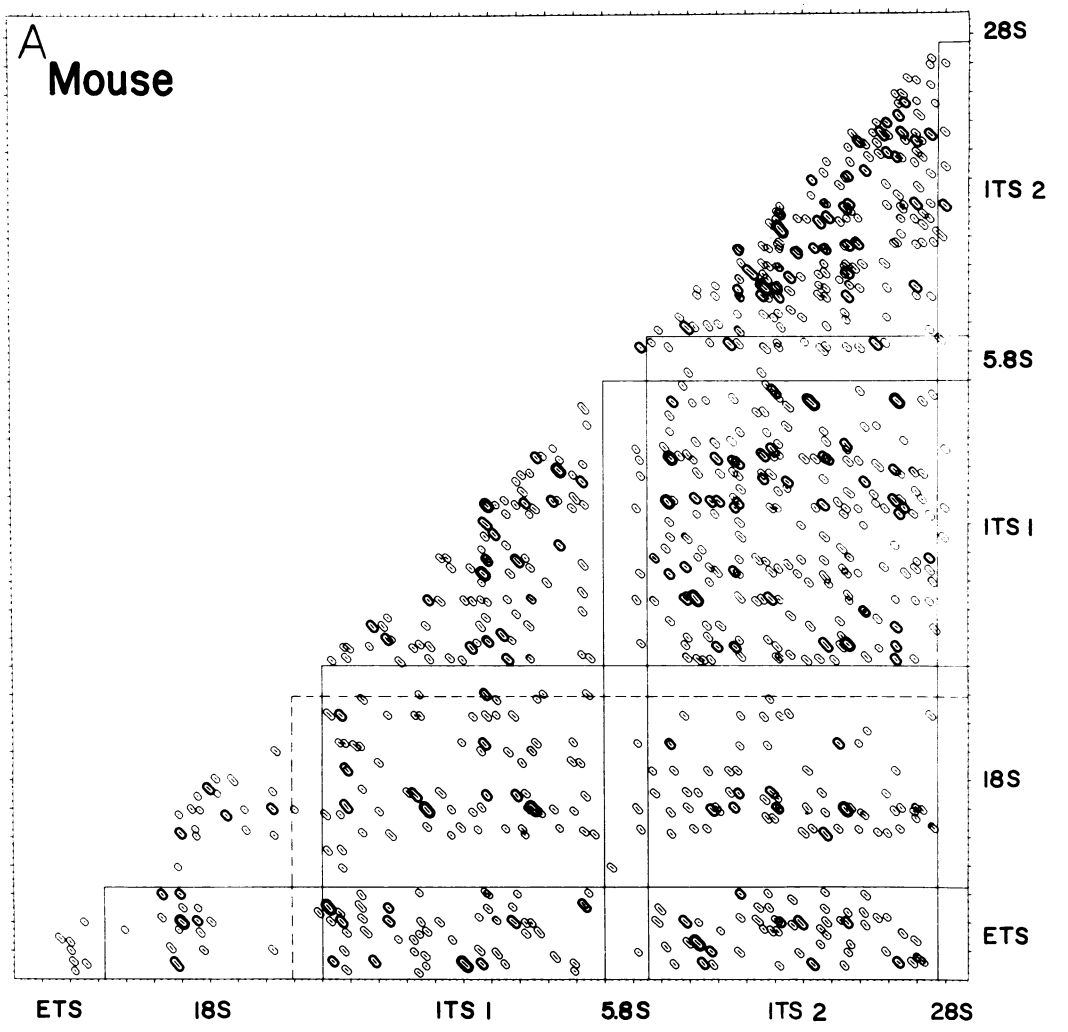


FIG. 7. Matrix representation of possible secondary structure in mouse (A) and *X. laevis* (B) pre-rRNA. Each matrix indicates the location of possible dyad symmetries based on a computer search for base pairing sequences. Finest gradations on each axis represent distances of 25 nucleotides. Lines within ovals correspond to the sequence coordinates and length of each strand in a base-paired stem; the surrounding oval indicates relative stability of the structure. The stability (ΔG^0) of each possible stem is calculated according to the rules of Tinoco et al. (30, 31) and the free energy calculations of Borer et al. (1). Thinnest oval = lowest stability range: -25 kcal $< \Delta G^0 \leq -15$ kcal. Middle range: -35 kcal $< \Delta G^0 \leq -25$ kcal. Highest range: $\Delta G^0 \leq -35$ kcal. Points near the diagonal indicate the smallest (hairpin) loops between closely spaced sequences. The mouse rDNA matrix has a dotted line denoting a gap in the available sequence data for 18S rRNA. Conditions for dyad symmetry search: homology of $\geq$70%; $\geq$7 bases paired; loopouts within the stem limited to single base bulges; any mismatches within the stem must be followed by $\geq$3 matched base pairs; G-U pairing is allowed.

rDNAs have long, nearly identical regions with an occasional sequence insertion or deletion (37). The extra or missing sequences in one or another species appear to be within largely self-contained regions dominated by short-range structure. Models of procaryotic rRNA secondary structure also show considerable overall conservation (27, 28, 37, 39). These observations are consistent with the similar function and common antibiotic sensitivity of ribosomes from many bacterial sources (32).

Analogous to the similarities among procaryotic rRNAs, antibiotic sensitivities and functional characteristics of yeast ribosomes are largely the same as those of mammalian ribosomes (32), and the mature rRNA sequences have corre-

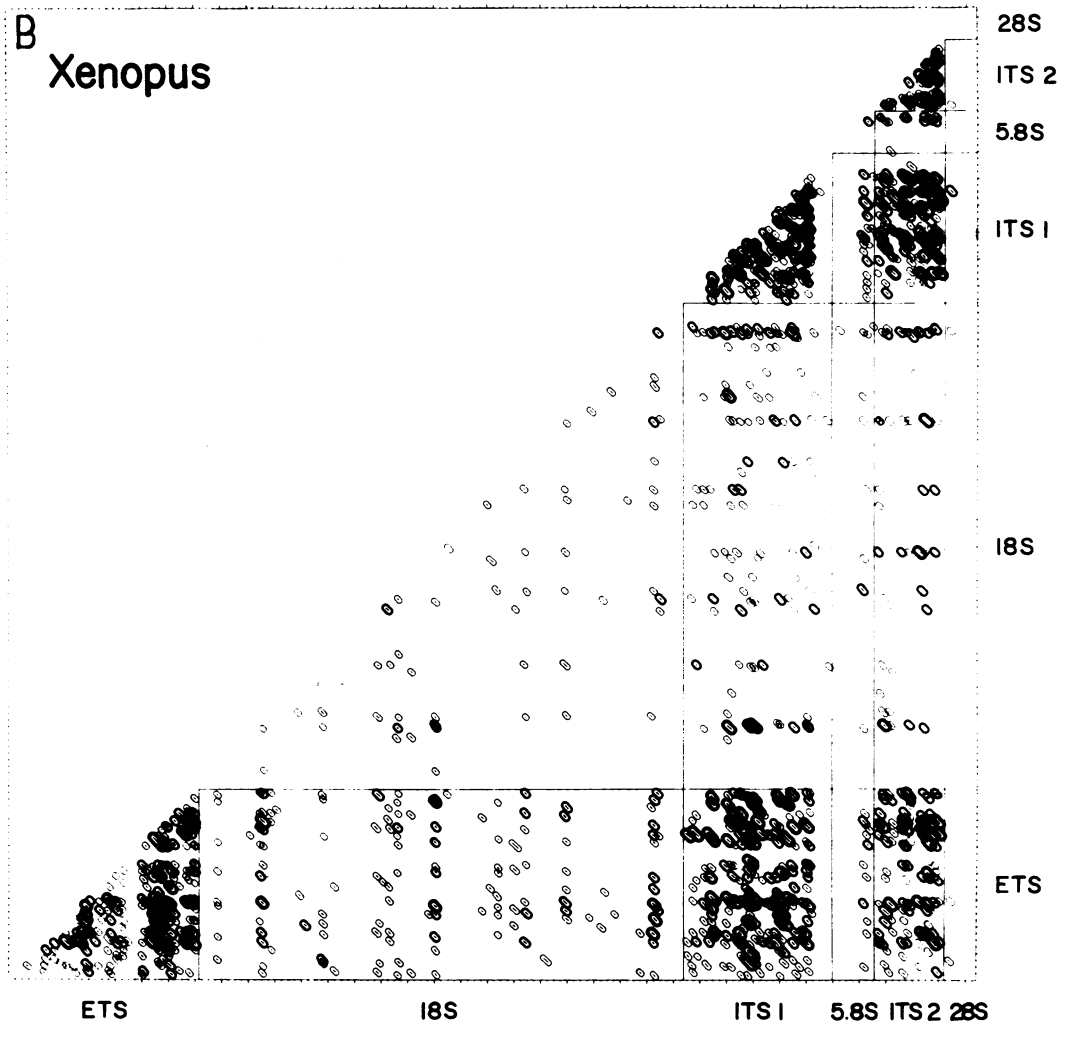


FIG. 7. *Continued.*

spondingly high homology. Some of the base substitutions from yeasts to *X. laevis* to mice (Fig. 3) may be compensatory base changes (see also reference 24). Even the G+C-rich inserts within the 18S rRNA sequences of mice and *X. laevis* probably form stable hairpin stems (Fig. 4) and may not disturb the structure of neighboring segments.

These G+C-rich inserts are quite unlike the introns observed in genes coding for proteins since they are not removed from the RNA by processing. The insertion of these elements can account for at least part of the progressive lengthening of mature rRNA during evolution as well as the trend toward higher G+C content. For example, the first 595 nucleotides of 18S rRNA from yeast, *X. laevis*, and mice show G+C contents of 40.9, 52.1, and 56.2%, respectively.

The rate of evolution of mature rRNA sequences can be estimated from the corrected percentage of divergence (calculated from pairwise comparison of the aligned mouse, *X. laevis*, and yeast sequences) and from the approximate dates of divergence of these organisms. A 1% change in the nucleotide sequence of the mature rRNAs (the UEP) takes 30×10^6 to 100×10^6 years. The divergence of mature rRNAs is much slower than that of protein-encoding genes like globin (from DNA sequence data: UEP; 10×10^6 years) (12) or preproinsulin (from DNA sequences: UEP, 5×10^6 to 25×10^6 years) (20). Although change occurs faster than in histone H4 (from protein sequences: UEP, 400×10^6 years) (36), the mature rRNA sequences are among the most highly conserved that have been studied.

Evolution and function of spacer sequences.

Two trends are apparent in the evolution of eucaryotic transcribed spacer regions. Analogous to the expansion of mature rRNAs, the length of spacers increases progressively from yeast to *X. laevis* to mouse (363 to 561 to 998 nucleotides for ITS 1; 235 to 263 to 1,018 nucleotides for ITS 2). Second, the proportion of G+C is much higher in vertebrate spacers than in yeast: for example, the yeast ITS 1 contains 35.2% G+C, compared with 83.9% in *X. laevis* and 70.1% in mouse.

One mechanism for the expansion of the spacers may involve repeated insertion of small G+C-rich sequences. The frequency of these events is apparent in the comparison of mouse and rat ITSs (Fig. 6), in which the optimized alignment contains 28 deletions and 25 insertions in the mouse spacer. Although rare deletions or insertions are as large as 120 nucleotides, nearly all are in the range of 4 to 40 nucleotides. Thus, small regions of the spacers may be continually deleted, and as in mature rRNA, small G+C-rich segments may be inserted. However, it is also possible that the rare insertion of large segments (see ITS 2, Fig. 6) may be the major mechanism for the expansion of spacer sequences during evolution.

The rate of insertion (greater than 3 nucleotides) in the mouse versus the rat ITSs is about 1.5 insertions per 100 nucleotides per 100×10^6 years. The corresponding rate for insertions into the first 500 nucleotides of mature 18S rRNA for mouse versus *X. laevis* or *X. laevis* versus yeast is about 10-fold lower (0.05 to 0.16 per 100 nucleotides per 100×10^6 years). The slower accumulation of insertions in the mature sequences is consistent with the slower increase in length of mature versus spacer sequences. It could reflect additional mechanisms for variation in the spacer sequence; alternatively, the rates of insertion might be comparable, but many insertions into mature rRNA could be lethal or disadvantageous.

The mechanism by which these G+C-rich insertions are generated can only be the object of speculation. *Zea mays* chloroplast 23S rRNA contains insertions compared with *E. coli* 23S rRNA (11). Many of these insertions are flanked by short direct repeats and contain inverted repeats. Similarly, the insertion into the mouse sequence at nucleotides 574 through 596 is flanked by a direct repeat (nucleotides 565 through 573 and 613 through 620) and contains an inverted repeat (see Fig. 3 and 4). Because flanking direct repeats and internal inverted repeats are characteristic of bacterial insertion sequences (6), it has been proposed that the *Z. mays* insertions are insertion sequence-like elements (11). Another possibility is that the insertions were generated by slippage of the two DNA strands during DNA replication (12). This mechanism has been suggested for the insertion or deletion of sequences near direct or inverted repeats. However, the sequences surrounding the mouse inserts and the inserts themselves are not entirely consistent with any present model for the generation of deletions or insertions. This may be due to divergence of these sequences after the insertion event.

The comparison of mouse and rat ITSs also demonstrates that base substitutions occur at a high rate within these sequences. The overall rate corresponds to a UEP of 0.3×10^6 years. This is comparable to the UEPs of 1×10^6 to 6×10^6 years calculated for base changes within noncoding regions, 5′ flanking regions, and intron sequences in human β-like globin genes (12). We conclude that the drift of internal spacer sequences is largely dispersive and at least as fast as the divergence rate of any other gene sequences analyzed. This suggests that the vast majority of the ITS has no functions strictly dependent upon its primary sequence.

The one feature of the spacer sequences that is highly conserved among vertebrates is their high G+C content. One speculation is that the structure or replication of rDNA involves a modified conformation (38) that is dependent upon high G+C DNA. The high G+C content could also be important for the formation of stems and loops within spacer sequences. Interestingly, transcribed spacers from yeast and other lower eucaryotes contain a high percentage of A+T instead of G+C. Perhaps the small size of these spacers (as compared with those in vertebrates) obviates the need for high G+C DNA, or perhaps the high A+T content favors a structure analogous to that favored by high G+C content.

Evolution of sequences near mature rRNA termini. Of any regions within the transcribed spacers, sequences bordering mature rRNA termini seem most likely to function during rRNA metabolism. Therefore, flanking spacer regions were searched for evolutionarily conserved features. Short stretches of sequence conservation near mature rRNAs have already been mentioned. Another type of evolutionary consistency is a conserved pattern of repeats near the mature rRNA termini (Fig. 8). Direct 4-nucleotide repeats are found at the 3′ ends of 18S and 5.8S rRNAs from mouse, *X. laevis*, yeast, and rat (all perfect repeats except near the rat 18S terminus). These repeats have different sequences in different organisms but are located at nearly identical positions. Less suggestive imperfect repeats of 5 to 7 nucleotides are seen at the 5′ ends of 18S, 5.8S, and 28S, rRNA in all four species, always with a single nucleotide stutter (e.g., ABCD . . ., ABXCD . . .). The

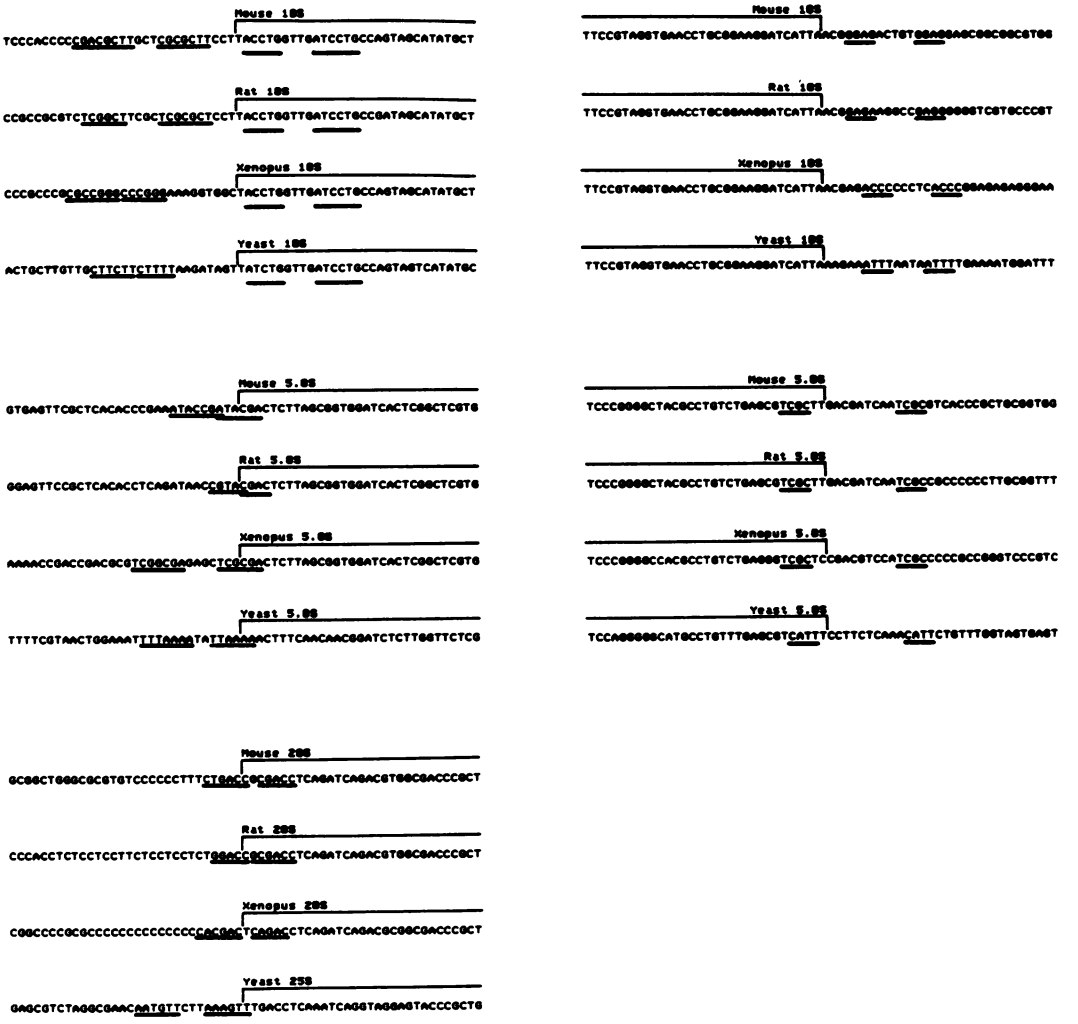


FIG. 8. Short repeated sequences near mature rRNA termini. Perfect repeats are associated with 3′ termini of each mature rRNA species; imperfect repeats are found near 5′ ends. Only those repeats that are the longest and show some evolutionary consistency in relative location are indicated here.

evolutionary conservation of these repeats suggests that they may play some role in ribosome formation or pre-rRNA processing (2).

Our analysis also suggests that pre-rRNA is composed of two types of evolutionarily conserved structural domains. The first is characterized by an enormous potential for secondary structure within and between transcribed spacers (Fig. 7). These sequences may be knotted into very stable stems, as suggested by visualization of pre-rRNA structure by electron microscopy (35). The mature rRNA sequences and the regions 10 to 50 nucleotides from their termini seem to represent another type of structural domain without much possibility for the formation of stable stems. The boundaries of these two domains are fairly sharp and are found in all vertebrate pre-rRNA sequences analyzed (rat data not shown). Of the thousands of possible dyad symmetries distributed throughout both domains, only a few involve base pairing with sequence at a mature rRNA terminus. These stems are not conserved in sequence or location and therefore are not analogous. Thus, the mechanism of rRNA processing in mammalian cells may be very different from the bacterial paradigm and is considered further in the accompanying paper (3).

ACKNOWLEDGMENTS

We thank Laverne Wodraska for her assistance in preparation of this manuscript.

This work was supported by Public Health Service grant GM 10447 from the National Institutes of Health and by Postdoctoral Fellowship 1F32AI06621 from the National Institutes of Health to W.E.G.

LITERATURE CITED

1. **Borer, P. N., B. Dengler, I. Tinoco, Jr., and O. C. Uhlenbeck.** 1974. Stability of ribonucleic double-stranded helices. J. Mol. Biol. **86**:843–853.
2. **Bowman, L. H., B. Rabin, and D. Schlessinger.** 1981. Multiple ribosomal RNA cleavage pathways in mammalian cells. Nucleic Acids Res. **9**:4951–4966.
3. **Bowman, L. H., W. E. Goldman, G. I. Goldberg, M. B. Hebert, and D. Schlessinger.** 1983. Location of the initial cleavage sites in mouse pre-rRNA. Mol. Cell. Biol. **3**:1501–1510.
4. **Bram, R. J., R. A. Young, and J. A. Steitz.** 1980. The ribonuclease III site flanking 23S sequences in the 30S ribosomal precursor RNA of *E. coli*. Cell **19**:393–401.
5. **Brosius, J., M. L. Palmer, P. J. Kennedy, and H. F. Noller.** 1978. Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **75**:4801–4805.
6. **Calos, M. P., and J. H. Miller.** 1980. Transposable elements. Cell **20**:579–595.
7. **Cassidy, B. G., C. S. Subrahmanyam, and L. I. Rothblum.** 1982. The nucleotide sequence of the 5′ region of rat 18S rDNA and adjoining spacer. Biochem. Biophys. Res. Commun. **107**:1571–1576.
8. **Choi, Y. C., and T.-S. Ro-Choi.** 1980. Basic characteristics of cellular RNA's: a directory, p. 609–667. *In* L. Goldstein and D. Prescott (ed.), Cell biology: a comprehensive treatise, vol. 3. Academic Press, Inc., New York.
9. **Cox, R. A.** 1977. Structure and function of prokaryotic and eukaryotic ribosomes. Prog. Biophys. Mol. Biol. **32**:193–231.
10. **Dickerson, R. E.** 1971. The structure of cytochrome c and the rates of molecular evolution. J. Mol. Evol. **1**:26–45.
11. **Edwards, K., and H. Kossel.** 1981. The rRNA operon from Zea mays chloroplasts: nucleotide sequence of 23S rDNA and its homology with E. coli 23S rDNA. Nucleic Acids Res. **9**:2853–2869.
12. **Efstratiadis, A., J. W. Posakony, T. Maniatis, R. M. Lawn, C. O'Connell, R. A. Sprintz, J. K. DeRiel, B. G. Forget, S. M. Weissman, J. L. Slightom, A. E. Blechl, O. Smithies, F. E. Baralle, C. C. Shoulders, and N. J. Proudfoot.** 1980. The structure and evolution of the human β-globin gene family. Cell **21**:653–668.
13. **Eladari, M. E., and F. Galibert.** 1975. Sequence determination of 5′-terminal and 3′-terminal T_1 oligonucleotides of 18 S ribosomal RNA of a mouse cell line (L 5178 Y). Eur. J. Biochem. **55**:247–255.
14. **Erdmann, V. A.** 1979. Collection of published 5S and 5.8S RNA sequences and their precursors. Nucleic Acids Res. **6**:r29–r44.
15. **Goldberg, G. I., E. G. Vanin, A. M. Zrolka, and F. R. Blattner.** 1981. Sequence of the gene for the constant region of the μ chain of Balb/c mouse immunoglobulin. Gene **15**:33–42.
16. **Hadjiolov, A. A., and N. Nikolaev.** 1976. Maturation of ribosomal ribonucleic acids and the biogenesis of ribosomes. Prog. Biophys. Mol. Biol. **31**:95–144.
17. **Hall, L. M. C., and B. E. H. Maden.** 1980. Nucleotide sequence through the 18S–28S intergene region of a vertebrate ribosomal transcription unit. Nucleic Acids Res. **8**:5993–6005.
18. **Jacobs, L. L., and D. Pilbeam.** 1980. Of mice and men: fossil-based divergence dates and molecular "clocks." J. Hum. Evol. **9**:551–555.
19. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
20. **Perler, F., A. Efstratiadis, P. Lomedico, W. Gilbert, R. Kolodner, and J. Dodgson.** 1980. The evolution of genes: the chicken preproinsulin gene. Cell **20**:555–566.
21. **Rubin, G. M.** 1973. The nucleotide sequence of *Saccharomyces cerevisiae* 5.8S ribosomal ribonucleic acid. J. Biol. Chem. **248**:3860–3875.
22. **Rubtsov, P. M., M. M. Musakhanov, V. M. Zakharyev, A. S. Krayev, K. G. Skryabin, and A. A. Bayev.** 1980. The structure of the yeast ribosomal RNA genes. I. The complete nucleotide sequence of the 18S ribosomal RNA gene from *Saccharomyces cerevisiae*. Nucleic Acids Res. **8**:5779–5794.
23. **Salim, M., and B. E. H. Maden.** 1980. Nucleotide sequence encoding the 5′ end of *Xenopus laevis* 18S rRNA. Nucleic Acids Res. **8**:2871–2884.
24. **Salim, M., and B. E. H. Maden.** 1981. Nucleotide sequence of *Xenopus laevis* 18S ribosomal RNA inferred from gene sequence. Nature (London) **291**:205–208.
25. **Skryabin, K. G., V. M. Zakharyev, P. M. Rubtsov, and A. A. Bayev.** 1979. Full sequence of the nucleotides of the spacer region located between the genes for yeast 18S and 5.8S RNA. Dokl. Akad. Nauk SSSR **247**:1278–1280.
26. **Staden, R.** 1977. Sequence data handling by computer. Nucleic Acids Res. **4**:4037–4051.
27. **Stiegler, P., P. Carbon, J.-P. Ebel, and C. Ehresmann.** 1981. A general secondary structure model for procaryotic and eucaryotic RNAs of the small ribosomal subunits. Eur. J. Biochem. **120**:487–495.
28. **Stiegler, P., P. Carbon, M. Zuker, J.-P. Ebel, and C. Ehresmann.** 1981. Structural organization of the 16S ribosomal RNA from *E. coli*. Topography and secondary structure. Nucleic Acids Res. **9**:2153–2172.
29. **Subrahmanyam, C. S., B. Cassidy, H. Busch, and L. I. Rothblum.** 1982. Nucleotide sequence of the region between the 18S rRNA sequence and the 28S rRNA sequence of rat ribosomal DNA. Nucleic Acids Res. **10**: 3667–3680.
30. **Tinoco, I., O. C. Uhlenbeck, and M. D. Levine.** 1971. Estimation of secondary structure in ribonucleic acids. Nature (London) **230**:362–367.
31. **Tinoco, I., Jr., P. N. Borer, B. Dengler, M. Levine, O. Uhlenbeck, D. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nature (London) New Biol. **246**:40–41.
32. **Vazquez, D.** 1979. Inhibitors of protein biosynthesis. Springer-Verlag KG, Berlin.
33. **Veldman, G. M., R. C. Brand, J. Klootwijk, and R. J. Planta.** 1980. Some characteristics of processing sites in ribosomal precursor RNA of yeast. Nucleic Acids Res. **8**:2907–2920.
34. **Veldman, G. M., J. Klootwijk, H. Van Heerikhuizen, and R. J. Planta.** 1981. The nucleotide sequence of the intergenic region between the 5.8S and 26S rRNA genes of the yeast ribosomal RNA operon. Possible implications for the interaction between 5.8S and 26S rRNA and the processing of the primary transcript. Nucleic Acids Res. **9**:4847–4862.
35. **Wellauer, P. K., I. B. Dawid, D. E. Kelley, and R. P. Perry.** 1974. Secondary structure maps of ribosomal RNA. II. Processing of mouse L-cell ribosomal RNA and variations in the processing pathway. J. Mol. Biol. **89**:397–407.
36. **Wilson, A. C., S. S. Carlson, and T. J. White.** 1977. Biochemical evolution. Annu. Rev. Biochem. **46**:573–639.
37. **Woese, C. R., L. J. Magrum, R. Gupta, R. B. Siegel, and D. A. Stahl.** 1980. Secondary structure model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. Nucleic Acids Res. **8**:2275–2293.
38. **Zimmerman, S. B.** 1982. The three-dimensional structure of DNA. Annu. Rev. Biochem. **51**:395–427.
39. **Zwieb, C., C. Glotz, and R. Brimacombe.** 1981. Secondary structure comparisons between small subunit ribosomal RNA molecules from six different species. Nucleic Acids Res. **9**:3621–3640.