A universal model for the secondary structure of 5.8S ribosomal RNA molecules, their contact sites with 28S ribosomal RNAs, and their prokaryotic equivalent

Jack C. Vaughn and Sally J. Sperbeck
Department of Zoology, Miami University, Oxford, OH 45056, and

W. Jay Ramsey and Charles B. Lawrence
Department of Cell Biology, Baylor College Medicine, Houston, TX 77030, USA

Received 14 August 1984; Accepted 4 September 1984

ABSTRACT

The phylogenetic approach (ref. 1) has been utilized in construction of a universal 5.8S rRNA secondary structure model, in which about 65% of the residues exist in paired structures. Conserved nucleotides primarily occupy unpaired regions. Multiple compensating base changes are demonstrated to be present in each of the five postulated helices, thereby forming a major basis for their proof. The results of chemical and enzymatic probing of 5.8S rRNAs (ref. 13, 32) are fully consistent with, and support, our model. This model differs in several ways from recently proposed 5.8S rRNA models (ref. 3, 4), which are discussed. Each of the helices in our model has been extended to the corresponding bacterial, chloroplast and mitochondrial sequences, which are demonstrated to be positionally conserved by alignment with their eukaryotic counterparts. This extension is also made for the base paired 5.8S/28S contact points, and their prokaryotic and organelle counterparts. The demonstrated identity of secondary structure in these diverse molecules strongly suggests that they perform equivalent functions in prokaryotic and eukaryotic ribosomes.

INTRODUCTION

A widely used approach in nucleic acid secondary structure determination was introduced by Fox and Woese (1), who used a phylogenetic comparison to formulate an early 5S ribosomal RNA (rRNA) structure. In this comparative phylogenetic approach, the fundamental premise is that functionally equivalent molecules will exhibit the same secondary structure. This method has been used to develop general models for 5S (2), 5.8S (3, 4), 18S (5) and 28S rRNAs (6). In any universal secondary structure model, each helix and unpaired tract must be present at the same position in homologous molecules from every species examined. Such models are best established by comparing sequences from species which are very distantly related. A good

example of the value of using widely separated species is evident in a study of prokaryotic 23S rRNA secondary structure (7), in which a model was derived based upon comparison of conserved helices in a gram-negative and a gram-positive bacterium. Although about 100 helices are present in the model, 10% of these had to be changed when the comparison was extended to an even more distantly related sequence: that of a chloroplast.

Prokaryotic ribosomes contain a 16S rRNA molecule within their small subunit whereas their large subunit contains a 5S and a 23S molecule. These RNAs are the evolutionary predecessors of the eukaryotic 18S, 5S and 25-28S rRNAs, respectively. One additional rRNA, the 5.8S molecule, is universally present in the large subunit of eukaryotic but not prokaryotic ribosomes. (For a recent review, see ref. 8). The 5.8S molecule, which functions in pulling together structural domains within the 25-28S rRNA, is too similar to the sequence at the 5' end of bacterial 23S rRNA for the similarity to have arisen by chance (9, 10). There is a growing consensus that the 5.8S rRNA molecule is base paired near both its 5' and 3' ends with complementary regions at the 5' end of eukaryotic 25-28S rRNA (11-15). It appears that homologous nucleotides may be involved in forming equivalent helices near the 5' end of the prokaryotic 23S molecule (7, 16).

Prokaryotes do not contain 5.8S rRNA, so that systematic application of the comparative phylogenetic approach to secondary structure determination for this molecule has thus far been largely limited to comparing the ever increasing number of available sequences between as evolutionarily diverse eukaryotic species as possible (3, 4, 17). Several proposals have been made for 5.8S rRNA secondary structure, as depicted in Fig. 1. None of these models are universal, especially when extended to the prokaryote level, and careful study of homologous prokaryotic sequences has led some workers to propose helices (7, 16, 18) which do not appear in any of the general 5.8S models, although detailed parallel comparative analysis of 5.8S sequences was not usually attempted. The 5.8S rRNAs exhibit considerable nucleotide sequence conservation, so

Figure 1. Diagrammatic representation of various general models which have been proposed for the secondary structure of 5.8S rRNAs and (D, E) their bacterial equivalent, using the *Homo sapiens* molecule as an example. In every model, the helix at the lower right is the "G-C-rich helix."

- (A): the "burp gun" model of Nazar et al. (20).
- (B): the "cloverleaf" model of Luoma and Marshall (21).
- (C): the modified burp gun model of Ursi et al. (3).
- (D): the minimal secondary structure model of Olsen and Sogin (4). The complete model fits higher eukaryotes. Dotted line component "a" is absent in lower eukaryotes and is not homologous to its eukaryotic counterpart in bacteria; dotted line component "b" is absent in bacteria.
- (E): our "universal" model.

that comparisons between only the eukaryotes have considerable potential both for including helices which are not proven by observation of compensatory base changes and for preventing discovery of the mutations which are a major criterion used in the "proving" of a suspected helix (1, 7), as has recently been pointed out in a review by Walker and Pace (19).

Since the cited 5.8S models were proposed, new evidence has appeared which led us to suspect that a more nearly universal 5.8S rRNA model could be developed if the phylogenetic comparison could be successfully extended to the prokaryotic and organelle level. In this study, we show that a universal secondary structure model which is energetically favorable is applicable to not only all eukaryotic rRNAs which have thus far been sequenced, but also to the corresponding 5' end of the

large subunit rRNAs of evolutionarily diverse bacteria as well as to chloroplast and mitochondrial sequences.

MATERIALS AND METHODS

Selection of representative 5.8S rRNAs and their equivalent

We utilized the annual 5.8S rRNA sequence compilation (22) as a source of most of the sequences in this study. This was supplemented with sequences not included in the compilation: a gnat (23), a thermophilic fungus (24), and the silkworm Bombyx (25). We also utilized sequences which have appeared subsequent to the compilation: Tetrahymena and an aphid (26), the silkworm Philosamia (27), the slime mold Physarum (28), yellow lupine (29), a snail and sponge (17) and a sea urchin (30). Finally, we took into account species for which sequence corrections have appeared: the frog Xenopus laevis (31), wheat (32), mouse (33) and brine shrimp (34).

The sequences of about thirty-two 5.8S rRNAs have thus far been reported, but not all were selected for this study because their inclusion would add little additional understanding. The Saccharomyces cerevisiae and S. carlsbergensis (15) sequences are identical, as are those for mouse and man (22), and those for Xenopus laevis and X. borealis differ by only one nucleotide; those of mouse (33) and rat (35) are identical. Eight vertebrate RNAs have been sequenced, but these are very similar to one another. We settled on frog and mouse because both their rRNA and corresponding gene have been sequenced. Given these considerations, the conclusions reached in this study are applicable to all of the thirty-two 5.8S rRNAs thus far sequenced.

The equivalent prokaryote and organelle large subunit rRNAs utilized were from two strains of Escherichia coli (16, 36), Bacillus stearothermophilus (7, 37), Anacystis nidulans (38), and chloroplast sequences from Euglena gracilis (39-41), corn (42), tobacco (43) and Chlamydomonas reinhardtii (44). Mitochondrial rRNAs from the yeast Saccharomyces cerevisiae (45) and protozoan Paramecium primaurelia (46) were selected because their 5' ends retain homology with the bacterial sequences, whereas the mammalian mitochondrial sequences are

reportedly both highly scrambled and exceedingly divergent relative to the corresponding bacterial sequences, with little homology in the first few hundred positions (7, 47).

In our search for universal 5.8S/28S rRNA complementary contact points, we utilized published partial and complete 25-28S rRNA sequences: slime mold *Physarum polycephalum* (48), yeasts *S. cerevisiae* (49) and *S. carlsbergensis* (15), sea urchin (30), mouse (50), rat (51) and frog (52), in addition to most of the cited prokaryotic and organelle sequences.

Alignment of sequences

The twenty-five 5.8S rRNA sequences selected for this study were individually aligned against the yeast, *S. cerevisiae*, for maximum homology. This work was simplified by first aligning conserved oligonucleotide sequences which occur at about the same position in most species: GGAU (at positions #15-18); GAAG (#39-42); GCAG (#75-78); GAAC (#103-106); and PyAPy at #137-139. A computer program which aligns sequences for maximum homology (M. Browner and C. Lawrence, manuscript in preparation) facilitated the initial recognition of these regions. Following this initial step, sequences were compared in pairs and notation was made of nucleotide differences with special emphasis on conservative substitutions. The alignment was completed by assuming as few deletions (gaps) and insertions as possible, while generating the maximum number of matching residues. This approach produced a satisfactory alignment for all 5.8S rRNA regions except for the region between positions #117-136, which was later recognized as the "G-C-rich helix" and eventually became "region F" in our model. The boundaries of this region are fixed by conserved nucleotide blocks.

Having aligned the 5.8S molecules, we then aligned this as a block to the already aligned 5' end block of the prokaryotic and organelle sequences. This alignment is basically as set forth in the original reports (9, 10) that the eukaryotic and prokaryotic molecules show extensive sequence homology, but should represent an improvement because this pairing of course involves more than comparison of only one prokaryote sequence to one eukaryote sequence.

Strategy for secondary structure determination

Our sequence alignment computer program was utilized to locate all positionally conserved helices having a stem length of four or more base pairs, with a toleration of up to 25% decay in homology. This program was initially utilized to compare only the *S. cerevisiae* and *X. laevis* sequences, but was then expanded to include *Drosophila* and *Artemia* in various combinations. This procedure generated models and comprised the first step in building a "phylogenetically proven" secondary structure. In order for a helix to be "proven" by this approach, it must contain a number of compensating base changes when different species are compared. In our study, we adopted the criteria suggested by Noller et al. (7) that "the existence of a helix is proven if there are two or more base pair replacements... if only a single base pair replacement is known, we consider the helix likely, but not proven."

Rules for base paired regions and computation of free energy of secondary structure formation

Although the set of rules devised by Tinoco et al. (53) have been widely used to predict secondary structures, Ninio (54) has recently pointed out that these rules in fact predict the tRNA cloverleaf in only about 52% of the molecules. Ninio derived an alternative set of rules for estimating secondary structures which reportedly results in up to 90% predictability of tRNA cloverleaves. His rules consist of a set of ΔG values for the stacking interactions of all possible slices of two base pairs upon which is superimposed a set of rules consisting of various energetic bonuses and penalties for certain topological features. In the calculation of the free energy of secondary structure formation for various structures we have utilized a combination of Ninio's topologically- and thermodynamically-based models, as has recently been done in developing a universal 5S rRNA secondary structure model (2).

A helix was assumed to conform to the following set of rules: 1) standard base pairs are G-C, A-U and G-U; 2) helices consist of at least three base pairs (in certain cases two base pairs); 3) stacking interactions for both standard and mismatched base pairs in a helix are calculated; 4) mismatched

base pairs in a helix may be adjacent to a standard base pair but not to one another (in the latter case, they are counted as an internal loop); 5) when a helix contains a mismatched base pair, it must (usually) be at least five base pairs long; 6) when a helix contains two or more mismatched base pairs, it must be at least seven base pairs long. A positionally conserved weak helix such as the GUUU/UAAG found in the DiHU helix of yeast aspartate tRNA is considered acceptable.

RESULTS AND DISCUSSION

Alignment of 5.8S rRNA sequences and equivalent prokaryote and organelle regions

In the derivation of a general secondary structure model, the first step is to determine the extent to which homologous nucleotides comprise positionally conserved helices or occupy unpaired positions. This requires alignment of the various sequences being compared, to assure that homologous structures (nucleotides) are indeed being compared. Fig. 2 shows our alignment for the twenty-five representative eukaryotic 5.8S rRNAs and also for the equivalent large subunit rRNA sequence from three quite diverse prokaryotic species and six organelle sequences.

Defining a conserved or semi-conserved nucleotide to be one which is present at a given location in at least 95% of all species examined, the 5.8S rRNAs show 28% sequence conservation (43 identical nucleotides out of the 156 positions which are present in virtually all species compared), while an additional 32% are semi-conserved. This result is a slight improvement over the 27% sequence conservation (41 identical nucleotides out of the 154 positions) and additional 25% semi-conserved nucleotides in a somewhat different alignment of virtually the same species (17). When a 5.8S rRNA nucleotide is conserved throughout the eukaryotes, it often is also conserved at the corresponding prokaryote 23S rRNA sequence position: 29 of the 43 conserved eukaryote positions, or 67%, are identical when comparison is made to the *Bacillus* sequence. These features were in part utilized in bringing the block of aligned prokaryote sequences into line with the block of aligned 5.8S sequences.

Olsen and Sogin (4, and Fig. 1) have recently proposed a 5.8S model in which relatively little secondary structure is present in lower eukaryotes and the equivalent bacterial 23S rRNA region. Their model was built on the basis of a proposed eukaryote vs prokaryote alignment of nine species which is quite different from ours and also from those originally proposed (9, 10). This resulted in their omission among prokaryotes of what in our model is "helix B/B'," and their rejection of "helix D/D'" (our model) among lower eukaryotes. Their alignment utilizes a very different region in prokaryotes for what in our model is strand "D'" within "helix D/D'," so

Figure 2. Alignment of selected eukaryotic 5.8S rRNA sequences with corresponding region in prokaryote and organelle large subunit rRNAs. Boxes show base paired tracts (B/B', C/C'...) in our universal secondary structure model, as illustrated in Fig. 3. Occasional use of superscript letters is for brevity. Locations of modified nucleotides, none of which are predicted to interfere with postulated base pairings, are not shown.

that they conclude that helix D/D' is not positionally conserved between prokaryotes and eukaryotes. Our proposed alignment has the attraction of bringing both the prokaryote strands B' and D' into line with what appears to us to be their

eukaryotic counterparts, which contributes to a universal 5.8S rRNA model which extends to the prokaryote and organelle level.

The universal 5.8S rRNA secondary structure model

When the steps described in Materials and Methods were applied to the numerous computer-generated potential helices predicted after comparison of *S. cerevisiae*, *X. laevis*, *Drosophila* and *Artemia*, the long "G-C-rich helix," which is predicted to be especially stable, was the first helix to be recognized. Its acceptance restricted the number of remaining potential helices. Repetition of this process, and extension to additional species, eventually resulted in a general 5.8S secondary structure model having four helices; a short fifth helix ("E/E" in our model) was later recognized by visual inspection and its existence has been postulated in some *E. coli* structure models (7, 16). Owing to the considerable base sequence conservation present in these RNAs, phylogenetic proof for some helices was obtained by extending the analysis to the prokaryotic and organelle level.

About 65% of the residues in our model exist in paired structures when the proposed terminal interactions with 25-28S rRNAs are included. Positionally conserved sequences which we find to contribute to a universal secondary structure model encompassing all 5.8S rRNAs and also the equivalent region at the 5' end of prokaryote and organelle large subunit rRNAs are boxed in Fig. 2. The terminology used in our model and also the structures of several representative molecules are illustrated in Fig. 3. The two strands within each proposed helix are alphabetically named as they appear, reading from the 5' end: B/B', C/C' and so forth. Corresponding hairpin loops are also named as they appear: H-1, H-2 ..., as are multi-branched loops: M-1 and M-2. The conserved G-C-rich helix F/F' is an evolutionarily volatile region, so that it is difficult to meaningfully align individual nucleotides here. We have therefore not attempted to box the provisionally aligned helical components "F" and "F'" and have simply designated this "region F." Its limits can however be closely estimated owing to conserved nucleotide segments at either end of this region.

A consistent generalization which emerges in comparing 5.8S

rRNAs is that where runs of three or more adjacent conserved nucleotides occur, they are virtually always located within unpaired regions of the molecule: GGAU within region A (see below); GAAG in multi-branched loop M-1; GCAG in hairpin loop H-2 and GAAC in multi-branched loop M-2. This is often observed in secondary structure models, and is believed to reflect the importance of primary structure at such positions in specific protein/nucleic acid interactions in the ribosome.

In a 5.8S model proposed by Ursi et al. (3, 17), the conserved GAAC at positions 103-106 is consistently present within a postulated helix which does not appear in other models nor in our own model. This proposal conflicts with observations that the two adenylate residues are unpaired, as evidenced by: 1). their reaction with diethyl pyrocarbonate and 2). degradation of the entire GAAC tract by S1 nuclease (13, 32). Acceptance of this helix has the additional feature of destroying the helix in region E (our model), which has already been shown (18, and the present study) to be phylogenetically proven among prokaryotes and organelles. We therefore do not accept "helix B" and "internal loop I" in the Ursi model, although except for the inclusion of our "helix E" the two models are otherwise in reasonably good agreement.

We have calculated the free energy of secondary structure formation for our model in comparison with that of several others (Table I), and conclude that our model compares favorably, although there are no other extensive attempts to extend such 5.8S rRNA models to the prokaryote and organelle level. The ΔG values for various individual 5.8S rRNAs differ widely in our model, but no more so than different 5S rRNAs in another universal model (2). It is likely, however, that the free energy of secondary structure formation is of limited value in arguing inter-species molecular structural homogeneity, since the stability of helices within the ribosome is surely affected by interactions with the constituent proteins. For this reason, we also do not believe that an occasional positive ΔG value is a problem.

Region F : This was the first conserved helix we recognized during our computer search due to its length, predicted

Table I. Free energy of secondary structure formation by 5.8S rRNAs and their equivalent for various models^a

Species	ΔG (kcal/mole)		
	Ursi-Vandenberghe- De Wachter model	Olsen-Sogin model	Our model ^b
S. CEREVISIAE	-28.6	-27.3	-28.4
S. POMBE	-18.4	-21.0	-22.1
N. CRASSA	-25.0	-23.7	-24.8
T. LANUGINOSUS	-25.3	-24.0	-25.1
V. FABA.....	- 1.9	0.0	- 1.6
T. AESTIVUM	-13.3	-14.9	-18.5
L. LUTEUS	- 7.0	- 8.6	- 9.1
A. MAGNOLIAE	-28.7	-28.1	-28.1
T. PYRIFORMIS	-10.4	- 6.0	- 6.0
S. COPROPHILA.....	-31.1	-35.8	-35.8
D. MELANOGASTER	-23.8	-22.5	-24.1
A. CASTELLANII	-12.0	-10.0	-10.5
B. MORI	-29.7	-28.3	-28.3
P. CYNTHIA	-29.7	-28.3	-28.3
ARTEMIA SP.....	-19.8	-11.3	-13.5
M. MUSCULUS	-47.4	-45.6	-49.0
X. LAEVIS	-50.1	-47.2	-47.2
L. VARIEGATUS	-36.4	-34.6	-34.3
A. RUFUS	-40.0	-40.6	-40.6
H. SANGUINEA.....	-14.7	-15.7	-15.9
C. FASCICULATA	- 7.8	-10.8	-10.1
C. COHNII	-12.0	-11.1	-12.7
C. REINHARDTII	+ 0.2	- 0.9	- 2.4
D. DISCOIDEUM	+ 3.8	+ 6.1	+ 6.7
P. POLYCEPHALUM.....	- 6.9	-12.7	-10.8
B. STEAROTHERMOPHILUS	N.F.	-12.9	-25.4
N. TABACUM (CHLOROPLAST)	N.F.	+ 5.1	+ 1.7
Z. MAYS (CHLOROPLAST)	N.F.	+ 7.3	+ 2.3
E. GRACILIS (CHLOROPLAST)	N.F.	+ 8.7	- 2.2
C. REINHARDTII (CHLOROPLAST)	N.F.	+ 1.0	-10.7
A. NIDULANS	N.F.	- 2.6	-15.0
S. CEREVISIAE (MITOCHONDRION)	N.F.	+13.5	+ 3.3
P. PRIMAURELIA (MITOCHONDRION)	N.F.	+18.8	+10.3
E. COLI	N.F.	- 3.9	-17.4

(a) Calculated with omission of 28S-paired 5' and 3' termini or the equivalent.

(b) Base pairing schemes defined by boxes in the sequence alignment figure.

(c) No satisfactory fit of sequence to model was found.

stability and positional consistency. This helix has been experimentally verified owing to its marked resistance to both chemical and enzymatic probes for unpaired nucleotides (13, 32). Comparison of more-or-less closely related prokaryotic and/or eukaryotic species, between which a suitable alignment is still possible, reveals extensive evidence of the compensating base changes which constitute the major criterion for a "phylogenetically proven" helix. There are seven compensating mutational changes in this helix between *Bacillus* and *Escherichia*, where the nucleotides are virtually super-imposable. A good example of compensatory base changes in eukaryotes is seen in the relatively closely related yeasts *S. cerevisiae* and *S. pombe*, where alignment can be done with confidence. The 5.8S rRNA sequences of these two species differ by only three nucleotides (2%) outside region F, whereas the sequences differ by ten nucleotides (50%) in the helix within this region. This helix is proven.

This region shows other evidences of being an evolutionary hot spot. One such evidence is the presence of a DNA insertion in the dipteran insects *Drosophila* (12) and *Sciara* (23). The insert is excised during rRNA processing in an event which does not include splicing, so that the mature 5.8S rRNA is in two pieces, which both remain weakly bonded to the 26S rRNA. The 5.8S rRNA of the protozoan *Crithidia* is longer than that of any other species, due to an insertion within this region which is not excised during processing (Fig. 2).

An interesting feature has become clear during this study in comparing chloroplasts with the blue-green algal sequence. As shown in Fig. 2 and 3, the region F helix is completely deleted in the chloroplast sequences and this region is also absent from the blue-green algal sequence. This is consistent with the common belief that chloroplasts arose, during evolution, from symbiotic blue-green algae. The deletion of the region F helix in the mitochondrial sequences is not so easily explained, unless these organelles also arose from a prokaryotic group in which this region is deleted. This situation illustrates the point that a helix need not necessarily be universally present in order for it to be

Figure 3. Secondary structure of representative 5.8S rRNA molecules from a wide diversity of phylogenetic positions, as illustrated by our universal model. The terminology for the various regions is shown for the *Xenopus laevis* molecule. The prokaryote and organelle diagrams show the corresponding region from the 5' end of their large subunit rRNAs. Modified nucleotides are indicated, where known.

acceptable as a constituent of a "universal" model, for region F, the "G-C-rich helix," is the only helix in the various 5.8S rRNA secondary structure models which appears to have been universally agreed upon (Fig. 1).

Region E : This short helix has not previously been proposed in

5.8S rRNA models derived from extensive phylogenetic comparisons, although it has recently been accepted as phylogenetically proven among prokaryote and organelle sequences (18). Our alignment places this prokaryotic helix directly in line with a short helix which is highly conserved and virtually universally present in the eukaryotes, the only clear exception being in the two slime molds. In the latter, region E is partly replaced by an extension of region F. (We question the apparent deletion of region E in *Tetrahymena*, owing to the suspicious reported absence of the otherwise universally conserved GAAPy in this species: see Fig. 2). Comparison of the prokaryote and eukaryote helices, which are superimposable, provides multiple examples of the compensating mutational changes required for the proof of this helix, as does comparison of the bacterial and organelle sequences. For example, four out of six possible mutational changes occur between *Xenopus* and *E. coli* in this helix. Extensive sequence conservation in this region among eukaryotes again underscores the value of extending the phylogenetic comparisons to the prokaryote level for evidence of the compensating base changes required for proof of postulated helices. Results of S1 nuclease digestion studies (32) are consistent with the existence of this helix. However, it should be pointed out that it is often necessary to borrow a base pair from region F in order to form a three base pair long helix E/E'. In addition, of the four mutational changes observed at position 116 in eukaryotes possessing helix E/E', two are compensated at position 137: extending the region F helix at the expense of region E. None of the changes at position 116 are compensated at position 106, the putative partner in helix E/E'. Accordingly, the existence of the region E helix in eukaryotes is problematic.

Region D : The nucleotides in this helix are virtually superimposable in comparing bacterial and organelle sequences, so that one can be sure that homologous structures are being compared in seeking compensating base changes. The helix contains two stems separated by an internal loop in the prokaryotes and organelles. There are multiple compensating

base changes in both stems of the helix (in comparing *E. coli* with yeast mitochondrion, two mutational changes in one stem and seven in the other), so that this region is considered proven among the prokaryotes. The position of the internal loop is invariant, and its primary structure is highly conserved. Its absence from the eukaryotes, owing in part to the deletion of a short purine-rich tract (Fig. 2), is marked by a single nucleotide bulge loop in many of the 5.8S rRNAs. This type of feature is known to serve as a recognition site for ribosomal protein in other rRNAs (7).

Region D is in approximate line with the corresponding tract in the eukaryotes, which however differ considerably from one another in the extent to which nucleotides have been borrowed from or donated to multi-branched loop M-1. This is the most variable and least convincing helix in our model, although where more-or-less closely related species are compared (e.g. yeasts and fungi, "plants" or cellular vs acellular slime molds), the helices are virtually superimposable. Where homologous nucleotides can be compared, the required examples of compensating base changes are present. This proposed helix is quite resistant to both chemical and enzymatic probes for unpaired nucleotides, while nucleotide positions corresponding to its hairpin loop H-2 are extensively degraded (13, 32), so that there is experimental evidence for its existence. This helix is tentatively accepted as proven.

Region C : This helix is present in most other 5.8S models, although its primary structure is so highly conserved among eukaryotes that its proof via observation of compensating base changes is tenuous. The value of extending this study to the prokaryote and organelle level is evident here. For example, virtually every base in this helix has been replaced in the yeast mitochondrial sequence, and several other compensating changes are observed in comparing the bacteria, thus providing the proof for existence of this helix. This helix is proven.

Region B : Maly and Brimacombe (18) compared available 21-23S rRNA sequences and concluded that a somewhat different region (16, 55) is responsible for formation of helix B/B' than we have proposed, but noted that this involves utilization of a

region whose position does not correspond with the one often proposed in 5.8S rRNA models. With the addition of the Bacillus sequence (Fig. 2), it is evident that strand B of their proposed helix utilizes almost totally different nucleotides (three of the four possible nucleotide positions differ) from those involved in the other bacterial pairings. Also, no helix can be formed using these nucleotides in the Paramecium mitochondrial rRNA (Fig. 2). Furthermore, their pairing scheme consistently incorporates the highly conserved GAAG region into a helix, which is contrary to our observation that such conserved regions tend to be present in unpaired regions of the molecule. This GAAG tract in 5.8S rRNAs is attacked by chemical and enzymatic probes (13, 32) and must be unpaired. These objections can all be eliminated by utilizing a slightly different pairing (Fig. 2). The compensating base changes required for proof of this proposed helix are present within the prokaryote and organelle sequences (three compensatory changes in comparing Bacillus to tobacco chloroplast), as well as among the eukaryotes. This helix is considered proven.

Universal 5.8S/28S rRNA sites of interaction

Contrary to early 5.8S rRNA secondary structure models (20, 21), the 5'- and 3'-termini are generally believed to be involved in intermolecular bonding with the 25-28S rRNA molecule near its 5'-terminus (11-15), resulting in the pinching off of a conformational domain. This is in part evidenced by the observations that the two separate 5.8S rRNA half-molecules in the dipteran insects Drosophila (12) and Sciara (23) remain bonded to 26S rRNA after RNA isolation, that nuclease generated mouse 5.8S rRNA 5'- and 3'-terminal "half-molecules" are capable of independently bonding to 28S rRNA (56), and that corresponding nucleotides in prokaryotes could be involved in intramolecular helices (7, 16, 18). We have carefully studied previous models and also applied our search techniques and criteria for proof of helices to search for possible universal 5.8S/28S rRNA sites of interaction. Since nucleotides at the proposed sites of interaction are often highly conserved, we have again extended the phylogenetic comparison to the corresponding homologous prokaryote and organelle sequences in order to maximize chances of observing

Figure 4. Simultaneous alignment of proposed 25-28S/5.8S rRNA contact sites with corresponding regions in prokaryote and organelle large subunit rRNAs. For each eukaryotic species, the relevant 5.8S rRNA sequence region is given in the lower line and the complementary 25-28S rRNA tracts in the upper line. The 5' end of 5.8S rRNAs (A) interacts with 25-28S rRNAs at three helical contact sites (A-1, A-2 and A-3); the 3' end (A') interacts along most of its length. The terminal seven nucleotides at the 3' end of 5.8S rRNAs and initial seven nucleotides at the 5' end of 25-28S rRNAs (X) do not have a counterpart in the prokaryotic sequences, which contain an end loop.

the compensating mutational changes required for proof of helices. To assure that homologous eukaryotic and prokaryotic 23-28S rRNA nucleotides are actually being compared, we have utilized a recent alignment (52). Our results, summarized in Fig. 4, are strongly supportive of the 5.8S/28S rRNA pairing scheme originally proposed for yeast by Veldman et al. (15), and considerably expand the number and diversity of species examined.

Region A : The 5'-terminus of 5.8S rRNAs contains three tracts, which we have designated A-1, A-2 and A-3, that could interact with complementary regions a few hundred nucleotides from the 5'-terminus of the 28S rRNA molecule. These three tracts are separated from one another by totally conserved nucleotides: GGAU at positions 15-18 and GG at positions 24-25. The latter is shown as comprising an additional short extension of our helix A-3, containing only two base pairs and a bulge loop, in several models (15, 16, 55). However, the total lack of any compensating base changes in every nucleotide of both strands in this proposed segment (Fig. 4) lends no support to this idea. The existence of helix A-3 is supported by numerous compensating mutational changes: for example, comparison of mouse and *E. coli* reveals twelve such changes.

Helix A-2 appears in most other models, but not in all (7), about as we have shown it. It is not always convincing that identical nucleotide positions are being utilized in the various models. During our study, we attempted to generalize an alternative helix A-2 pairing scheme proposed by Branlant et al. (16), but were unable to produce acceptable helices for several species. We therefore do not believe that their scheme can be either universal or comprise a "switch," as some have proposed (18). Observation of compensating base changes supports the existence of helix A-2: for example, comparison of *Zea mays* chloroplast and yeast shows seven such changes.

In a recent re-examination of published 23S rRNA secondary structure models, Maly and Brimacombe (18) tentatively concluded that helix A-1 (their "helix 2") is interrupted by an unpaired loop (positions 520-523 in *E. coli*), owing to the utilization of the second of two tandemly repeated GUAC

sequences in this and related molecules (55). Since the eukaryotic sequences must utilize the first GUAC sequence, and our model results in an acceptable alignment between all sequences examined which also contains numerous compensating base changes, we have represented this region as a single uninterrupted helix, except for a bulge loop in the Zea mays chloroplast sequence owing to the deletion of two nucleotides. Proof for existence of helix A-1 is shown by observation of numerous compensating base changes: for example, comparison of mouse and E. coli reveals eleven such changes.

Region A' : The 3'-terminus of 5.8S rRNAs could interact with complementary positions at the 5'-terminus of the 28S rRNA molecule to form a single long helix. As is the case in region A, the simultaneous alignment of the 5.8S and 28S rRNA paired nucleotides with those in the prokaryotes and organelles clearly shows that homologous nucleotides at identical positions could be utilized in the formation of a remarkably uniform secondary structure for all sequences studied. The existence of helix A' is proven by observation of many compensating mutational changes: for example, comparison of mouse and E. coli reveals thirteen such changes.

The short eukaryotic segment located distal to the terminal bulge loop, which does not align with prokaryotic or organelle sequences, is believed to have arisen during evolution as part of a much larger insertion. The bulk of this ancient insertion is believed to be ITS-2, the internal transcribed spacer separating the eukaryotic 5.8S and 28S rRNA genes (16). The remainder of the helix clearly has a counterpart in the prokaryotes and organelles, where the loop end within the single continuous 23S rRNA molecule aligns with positions near (but not at) both the 3'-terminus of 5.8S rRNA molecules and the 5'-terminus of 28S rRNAs.

Finally, there is experimental evidence in partial support of the region A and A' pairing schemes which we have proposed. Utilization of nucleases and chemical agents as conformational probes (13, 56) on mouse 5.8S rRNA alone and also on the 5.8S/28S rRNA complex shows that 5.8S rRNA positions 137-155 (our helix A') are susceptible to attack in the free molecule

but are protected when complexed with 28S rRNA. However, this experimental approach provides confusing results for the 5.8S rRNA 5'-terminal contact sites, even though 5'- and 3'-terminal "half-molecules" were utilized, owing to the tendency of this region to form intramolecular helices. Taken at face value, the results show that mouse 5.8S rRNA positions 4-14 (our helix A-1) and 18-23 (our helix A-2) are susceptible to attack in the free molecule but are protected in complex with 28S rRNA. Evidence in support of helix A-3 rests primarily upon the comparative phylogenetic results, which are both consistent and convincing.

CONCLUSIONS

Utilization of the phylogenetic approach has resulted in the construction of a universal 5.8S rRNA secondary structure model containing five helical regions. Each postulated helix has been extended to the corresponding prokaryote and organelle level, where it is positionally conserved in line with its eukaryotic counterpart. Multiple compensating base changes are demonstrated to be present in each of the five proposed helices, thereby forming the basis of proof. The results of chemical and enzymatic probing of naked 5.8S rRNAs, reconstituted 5.8S/28S rRNA complexes and ribosomal subunits (13, 32) are fully consistent with, and support, our proposed model. Since 5.8S rRNAs have now been sequenced from a great variety of phylogenetic positions, it appears unlikely to us that additional sequences will provide significant further insights into the secondary structure of these molecules. The demonstration that 5.8S rRNAs have a potentially identical secondary structure both with one another and with that at the 5' end of bacterial and organelle large subunit rRNAs strongly suggests that these molecules perform equivalent functions.

Our model differs in several ways from recently proposed 5.8S rRNA secondary structure models (3, 4, 17), and we have discussed these as they have arisen. Application of the comparative phylogenetic approach to elucidation of the sites of interaction between 5.8S and 28S rRNAs, and extension to the homologous prokaryote and organelle sequences, shows that

identical tracts could be utilized universally in intermolecular interaction. The proposed sites of interaction are shown to contain numerous compensating mutational changes.

Our results support the consensus recently arrived at by Maly and Brimacombe (18) as to the secondary structure at the 5' end of the prokaryotic 23S rRNA molecule, although a correction is proposed for their "helix 5" (helix B/B' in our model). Experimental proof for models such as we have proposed will likely come from the results of additional chemical probing and cross-linking experiments directed at the organelle itself (8, 32), for as yet unknown structural constraints are likely placed upon these molecules within the ribosome.

ACKNOWLEDGMENTS

We wish to thank Dr. H.F. Noller for generously sharing his unpublished Bacillus stearothermophilus sequence data. This work was supported, in part, by National Science Foundation Grants CDP-8002341 and PRM-8106947 to J.C.V. and by National Institutes of Health Grant AI-19578 to C.B.L.

REFERENCES

- 1 Fox, G.E. and Woese, C.R. (1975) *Nature* 256, 505-507
- 2 De Wachter, R., Chen, M.-W. and Vandenberghe, A. (1982) *Biochimie* 64, 311-329
- 3 Ursi, D., Vandenberghe, A. and De Wachter, R. (1982) *Nucl. Acids Res.* 10, 3517-3530
- 4 Olsen, G.J. and Sogin, M.L. (1982) *Biochemistry* 21, 2335-2343
- 5 Olsen, G.J., McCarroll, R. and Sogin, M.L. (1983) *Nucl. Acids Res.* 11, 8037-8049
- 6 Michot, B., Hassouna, N. and Bachellerie, J.-P. (1984) *Nucl. Acids Res.* 12, 4259-4279
- 7 Noller, H.F., Kop, J.A., Wheaton, V., Brosius, J., Gutell, R.R., Kopylov, A.M., Dohme, F., Herr, W., Stahl, D.A., Gupta, R., and Woese, C.R. (1981) *Nucl. Acids Res.* 9, 6167-6189
- 8 Nazar, R.N. (1982) in *The Cell Nucleus, rDNA*, Pt. B, Busch, H. and Rothblum, L., eds., Vol. 11, pp.1-28, Academic Press, New York
- 9 Nazar, R.N. (1980) *FEBS Lett.* 119, 212-214
- 10 Jacq, B. (1981) *Nucl. Acids Res.* 9, 2913-2932
- 11 Pace, N.R., Walker, T.A. and Schroeder, E. (1977) *Biochemistry* 16, 5321-5328
- 12 Pavlakis, G.N., Jordan, B.R., Wurst, R.M. and Vournakis, J.N. (1979) *Nucl. Acids Res.* 7, 2213-2238
- 13 Walker, T.A., Johnson, K.D., Olsen, G.J., Peters, M.A. and Pace, N.R. (1982) *Biochemistry* 21, 2320-2329
- 14 Michot, B., Bachellerie, J.-P., Raynal, F. and Renalier, M.-H. (1982) *FEBS Lett.* 140, 193-197

- 15 Veldman, G.M., Klootwijk, J., de Regt, V., Planta, R.J., Branlant, C., Krol, A. and Ebel, J.-P. (1981) Nucl. Acids Res. 9, 6935-6952
- 16 Branlant, C., Krol, A., Machatt, M.A., Pouyet, J., Ebel, J.-P., Edwards, K. and Kossel, H. (1981) Nucl. Acids Res. 9, 4303-4324
- 17 Ursi, D., Vandenberghe, A. and De Wachter, R. (1983) Nucl. Acids Res. 11, 8111-8120
- 18 Maly, P. and Brimacombe, R. (1983) Nucl. Acids Res. 11, 7263-7286
- 19 Walker, T.A. and Pace, N.R. (1983) Cell 33, 320-322
- 20 Nazar, R.N., Sitz, T.O. and Busch, H. (1975) J. Biol. Chem. 250, 8591-8597
- 21 Luoma, G.A. and Marshall, A.G. (1978) Proc. Natl. Acad. Sci. U.S.A. 75, 4901-4905
- 22 Erdmann, V.A., Huysmans, E., Vandenberghe, A. and De Wachter, R. (1983) Nucl. Acids Res. 11, r105-r133
- 23 Jordan, B.R., Latil-Damotte, M. and Jourdan, R. (1980) Nucl. Acids Res. 8, 3565-3573
- 24 Wildeman, A.G. and Nazar, R.N. (1981) J. Biol. Chem. 256, 5675-5682
- 25 Fujiwara, H., Kawata, Y. and Ishikawa, H. (1982) Nucl. Acids Res. 10, 2415-2418
- 26 Fujiwara, H. and Ishikawa, H. (1982) Nucl. Acids Res. 10, 5173-5182
- 27 Feng, Y.-X., Krupp, G. and Gross, H.J. (1982) Nucl. Acids Res. 6383-6387
- 28 Otsuka, T., Nomiyama, H., Sakaki, Y. and Takagi, Y. (1982) Nucl. Acids Res. 10, 2379-2385
- 29 Rafalski, J.A., Wiewiorowski, M. and Soll, D. (1983) FEBS Lett. 152, 241-246
- 30 Hindenach, B.R. and Stafford, D.W. (1984) Nucl. Acids Res. 12, 1737-1747
- 31 Hall, L.M.C. and Maden, B.E.H. (1980) Nucl. Acids Res. 8, 5993-6005
- 32 Wildeman, A.G. and Nazar, R.N. (1982) Eur. J. Biochem. 121, 357-363
- 33 Michot, B., Bachellerie, J.-P. and Raynal, F. (1982) Nucl. Acids Res. 10, 5273-5283
- 34 Vaughn, J.C., Sperbeck, S.J. and Hughes, M.J. (1984) Biochim. Biophys. Acta, in press
- 35 Subrahmanyam, C.S., Cassidy, B., Busch, H. and Rothblum, L.I. (1982) Nucl. Acids Res. 10, 3667-3680
- 36 Brosius, J., Dull, T.J. and Noller, H.F. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 201-204
- 37 Kop, J.A., Wheaton, V., Gupta, R., Woese, C.R. and Noller, H.F. (Unpublished data)
- 38 Kumano, M., Tomioka, N. and Sugiura, M. (1983) Gene 24, 219-225
- 39 Graf, L., Kossel, H. and Stutz, E. (1980) Nature 286, 908-910
- 40 Orozco, E.M., Rushlow, K.E., Dodd, J.R. and Hallick, R.B. (1980) J. Biol. Chem. 255, 10997-11003
- 41 Orozco, E.M. (1981) Ph.D. Thesis, Univ. of Colorado
- 42 Edwards, K. and Kossel, H. (1981) Nucl. Acids Res. 9, 2853-2869

- 43 Takaiwa, F. and Sugiura, M. (1982) *Eur. J. Biochem.* 124, 13-19
- 44 Rochaix, J.D. and Darlix, J.L. (1982) *J. Mol. Biol.* 159, 383-395
- 45 Sor, F. and Fukuhara, H. (1983) *Nucl. Acids Res.* 11, 339-348
- 46 Seilhamer, J.J. and Cummings, D.J. (1981) *Nucl. Acids Res.* 9, 6391-6406
- 47 Van Etten, R.A., Walberg, M.W. and Clayton, D.A. (1980) *Cell* 22, 157-170
- 48 Otsuka, T., Nomiyama, H., Yoshida, H., Kukita, T., Kuhara, S. and Sakaki, Y. (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80, 3163-3167
- 49 Georgiev, O.I., Nikolaev, N., Hadjiolov, A.A., Skryabin, K.G., Zakharyev, V.M. and Bayev, A.A. (1981) *Nucl. Acids Res.* 9, 6953-6958
- 50 Hassouna, N., Michot, B. and Bachellerie, J.-P. (1984) *Nucl. Acids Res.* 12, 3563-3583
- 51 Chan, Y.-L., Olvera, J. and Wool, I.G. (1983) *Nucl. Acids Res.* 11, 7819-7831
- 52 Ware, V.C., Tague, B.W., Clark, C.G., Gourse, R.L., Brand, R.C. and Gerbi, S.A. (1983) *Nucl. Acids Res.* 11, 7795-7817
- 53 Tinoco, I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) *Nature New Biol.* 246, 40-41
- 54 Ninio, J. (1979) *Biochimie* 61, 1133-1150
- 55 Glotz, C., Zwieb, C., Brimacombe, R., Edwards, K. and Kossel, H. (1981) *Nucl. Acids Res.* 9, 3287-3306
- 56 Peters, M.A., Walker, T.A. and Pace, N.R. (1982) *Biochemistry* 21, 2329-2335