Volume 9 Number 8 1981 Nucleic Acids Research

A unique secondary folding pattern for 5S RNA corresponds to the lowest energy homologous secondary structure in 17 different prokaryotes

Gary M.Studnicka*†, Frederick A.Eiserling*+ and James A.Lake*†

*Molecular Biology Institute, †Department of Biology and +Department of Microbiology, University of California at Los Angeles, Los Angeles, CA 90024, USA

Received 5 January 1981

ABSTRACT

A general secondary structure is proposed for the 5S RNA of prokaryotic ribosomes, based on helical energy filtering calculations. We have considered all secondary structures that are common to 17 different prokaryotic 5S RNAs and for each 5S sequence calculated the (global) minimum energy secondary structure (300,000 common structures are possible for each sequence). The 17 different minimum energy secondary structures all correspond, with minor differences, to a single, secondary structure model. This is strong evidence that this general 5S folding pattern corresponds to the secondary structure of the functional 5S rRNA.

The general 5S secondary structure is forked and in analogy with the cloverleaf of tRNA is named the 'wishbone' model. It contains 8 double helical regions; one in the stem, four in the first, or constant arm, and three in the second arm. Four of these double helical regions are present in a model earlier proposed (1) and four additional regions not proposed by them are presented here. In the minimum energy general structure, the four helices in the constant arm are exactly 15 nucleotide pairs long. These helices are stacked in the sequences from gram-positive bacteria and probably stacked in gram-negative sequences as well. In sequences from gram-positive bacteria the length of the constant arm is maintained at 15 stacked pairs by an unusual minimum energy interaction involving a C_{26}-G_{57} base pair intercalated between two adjacent helical regions.

INTRODUCTION

The small 5S ribosomal RNA is necessary for the structural integrity of the large ribosomal subunit and for activities associated with protein synthesis (2; 3). Many prokaryotic 5S rRNAs including *E. coli*, are competent to replace a gram positive 5S RNA in the *B. steorothermophilus* 50S subunit reconstitution system (4; 5); although eukaryotic 5S rRNAs, thus far, have not been reconstituted in the *B. stearothermophilus* system (5; 6).

Numerous investigations have been made of the single- and double-stranded regions of 5S RNA using chemical modification (7; 8; 9; 10; 11; 12) and enzymatic digestion (13; 14; 15; 16; 17; 18; 19). These experiments have been performed both on 5S RNA free in aqueous solution and also 5S RNA asso-

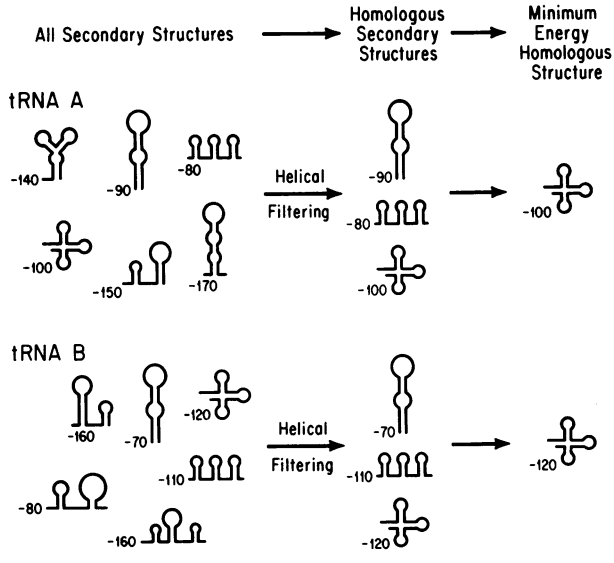
ciated with ribosomal proteins or *in situ* (i.e. in ribosomal subunits). Biophysical experiments, including low angle X-ray diffraction (21; 20), Raman spectroscopy (22), infared, ultraviolet, circular dichroism, optical rotary dispersion and nuclear magnetic resonance spectra (23; 24; 25; 26; 27), oligonucleotide binding (28; 29), and thermal denaturation (30; 31; 32), have also provided information about possible conformational states of 5S RNA. The *E. coli* 5S rRNA, in particular, can exist in two conformations as well as a denatured form (12; 24; 33; 34; 35). Although these studies have told us much about the structure of 5S RNA still no completely satisfactory secondary structure model has emerged.

The early (36; 37; 38) and rapid acceptance of the "cloverleaf" secondary structure for tRNA was primarily determined by its ability to fit known tRNA sequences. This was accepted as strong evidence for the correctness of the cloverleaf because it was anticipated that molecules having the same function (in different species) would have similar secondary and tertiary structures. Another observation, made for tRNA, was that helical regions could be conserved of an altered nucleotide sequence on one side of a double helical region was compensated by a base change on the other that permitted the two new nucleotides to form a base pair. This reasoning applied to prokaryotic 5S RNA molecules, predicted that they too would have common, secondary folding patterns, regardless of species. Fox and Woese developed and extended this technique (1; 35) and used it to show that prokaryotic 5S rRNAs could be folded into a forked secondary structure defined by four helical regions of nearly equal length. These four helices are all present in the 'wishbone' model presented here. They also showed that eukaryotic 5S rRNAs share three of these four common helices. The comparative approach, however, is 'local,' so that one has no quantitative assurance that the 'best' of many alternative sequences has been chosen. For example, the 5S RNA structure consisting of the four helices previously described (35) is only one of approximately 300,000 alternative common structures.

A 'global' approach to determining secondary structure has been to calculate the structure with the minimum energy. Computer programs (39; 40; 41; 42; 43) to predict secondary structures from primary sequences have been written using unique and novel algorithms. Recently, an algorithm has been developed (44) that is sufficiently efficient to perform complete searches for tRNA's, and is feasible for 5S rRNA. This algorithm is impractical to apply in an exact form to larger RNA's such as the major small and large subunit ribosomal RNA's. In addition, for tRNA's, at least, minimum energy

secondary structures do not always correspond to the true secondary structure (presumably because tertiary and other interactions form an important contribution to the total energy). Pipas and McMahan (41), for example, have shown that although the cloverleaf is consistently among the 1-2% of lowest energy structures, i.e. within the first 10,000 lowest energy structures, other secondary structures frequently have lower energies.

We have devised a comparative global method that is feasible to apply to even large ribosomal RNA's. Our helical filtering technique, outlined in Figure 1, involves calculating the minimum energy homologous structures. It is illustrated using two different, although related, hypothetical tRNA sequences. Both sequences can form a variety of secondary structures (approximately 1,000,000 for an average tRNA). Since any correct folding must be consistent with <u>both</u> sequences, only the secondary structures that can be made by both molecules need to be considered. Hence from these two groups, the subgroups corresponding to homologous secondary structures are



<u>Figure 1: Generalized Scheme for Helical Filtering.</u> A diagrammatic representation of the process of calculating the minimum energy homologous secondary structure is shown using two hypothetical tRNA sequences. In practice, the selection of homologous structures (helical filtering) is done by first accepting homologous helical regions and then using them to construct the sets of common structures. To compare directly entire structures would be prohibitively time consuming.

selected (helical filtering). In the final step, the lowest energy structure from each homologous subgroup is calculated and the minimum energy structures are compared. One can show that the method converges, as a sufficiently large number of sequences are included. We estimate the degree of convergence by comparing the solutions obtained from each sequence, and consider that convergence occurs when all of the minimum energy secondary structures are the same or nearly the same.

In this paper we demonstrate that a single bifurcated secondary structure pattern, which in analogy with the cloverleaf of tRNA has been named the wishbone model, corresponds to the minimum energy homologous solution for 17 different prokaryotic 5S RNA's. Details of the structures are described in this paper.

METHODS

(a) Computer Analysis

Homologous secondary structures were selected as individual helical regions using a computer to ensure that important structural possibilities were not overlooked. Once common helical regions were obtained (helical filtering) then homologous structures were formed and ordered according to the energy of their secondary structure using the algorithms and the program previously developed and described by Studnicka et al. (44).

(i) Sequence Alignment

The nucleotide sequences (reviewed in 45; 46) (see Figure 2 for a detailed listing) were aligned. Principally conserved bases (A, C, G, U) were aligned but occassionally purines and pyrimidines were aligned when it seemed appropriate. We did not use a formal set of alignment rules but fit the overall pattern using the principles mentioned above. Similar alignments (35; 45; 47; 48; 49) have been published by others, although they differ from ours in minor details.

(ii) Helical Filtering

Region A in one species and region B in another species are defined to be homologous helices if they satisfy the requirements described in the legend to Figure 3. This is less stringent than strict isomorphism (identical locations and lengths), and allows for the possibility that a helix may have been shortened, lengthened, or moved in either direction during its evolution. In addition, the alignment procedure is expected to compensate for insertion and deletion mutations. If one or more non-pairing bases occur within a helix, that helical region is treated in our program as two separate

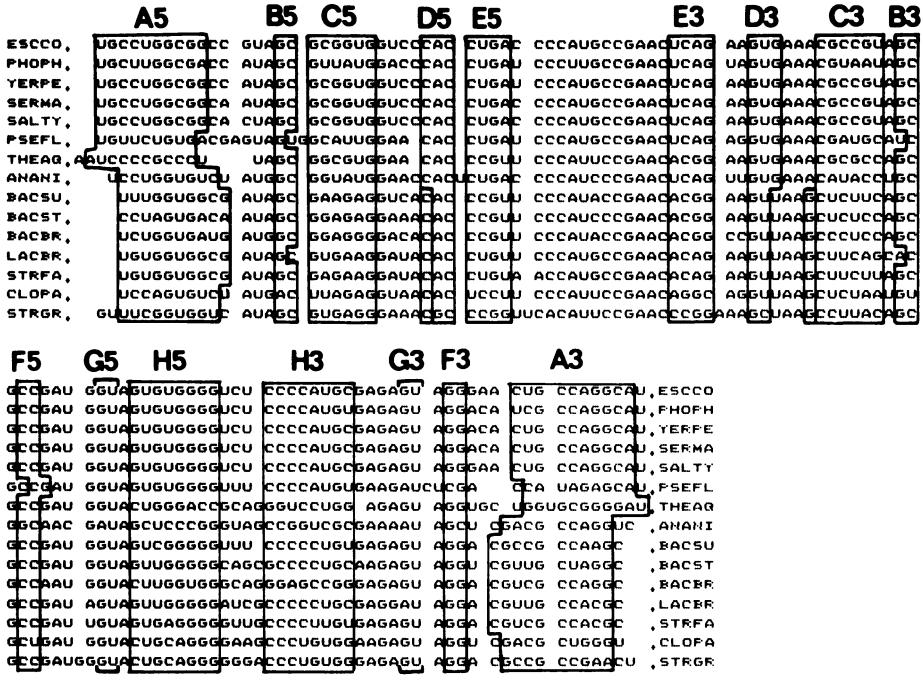


Figure 2: Alignment of Procaryotic 5S rRNA Sequences. Fifteen representative procaryotic 5S rRNAs have been aligned with spaces inserted to maximize the sequence homology. The large numbers denote base numbers of each set of aligned sequences. Spaces between the nucleotides are counted as nonpairing bases. The small numbers are the unspaced nucleotide positions of E. coli 5S rRNA. Since much of the discussion focuses on the E. coli sequence, its numbering system can be used to compare features with those of other species. Each "common" helix appearing in the best filtered structures has also been highlighted by lines in the spaced sequences, to make the degree of universality readily apparent. In Figure 3A each helical region is identified by a letter (A through H), and each halfregion (44) is specified. Thus, B5 is the 5' halfregion of helix B, and B3 is the 3' complementary half-region. Species names have been abbreviated to five letters: ESCCO = Escherichia coli, PHOPH = Photobacterium phosphoreum, YERPE = Yersinia pestis, PSEFL = Pseudomonas fluorescens, THEAQ = Thermus aquaticus, ANANI = Anacystis nidulans, BACSU = Bacillus subtilis, BACST = Bacillus stearothermophilus, BACBR = Bacillus brevis, LACBR = Lactobacillus brevis, STRFA = Streptococcus faecalis, CLOPA = Clostridium pasteurianum, and STRGR = Streptomyces griseus. Original sources of sequences are listed in references 45 and 46 (see also 66), and two minor changes are mentioned in 67.

helices, but if blank spaces have been inserted by us for alignment purposes, helices are temporarily broken into subregions, which are subsequently rejoined.

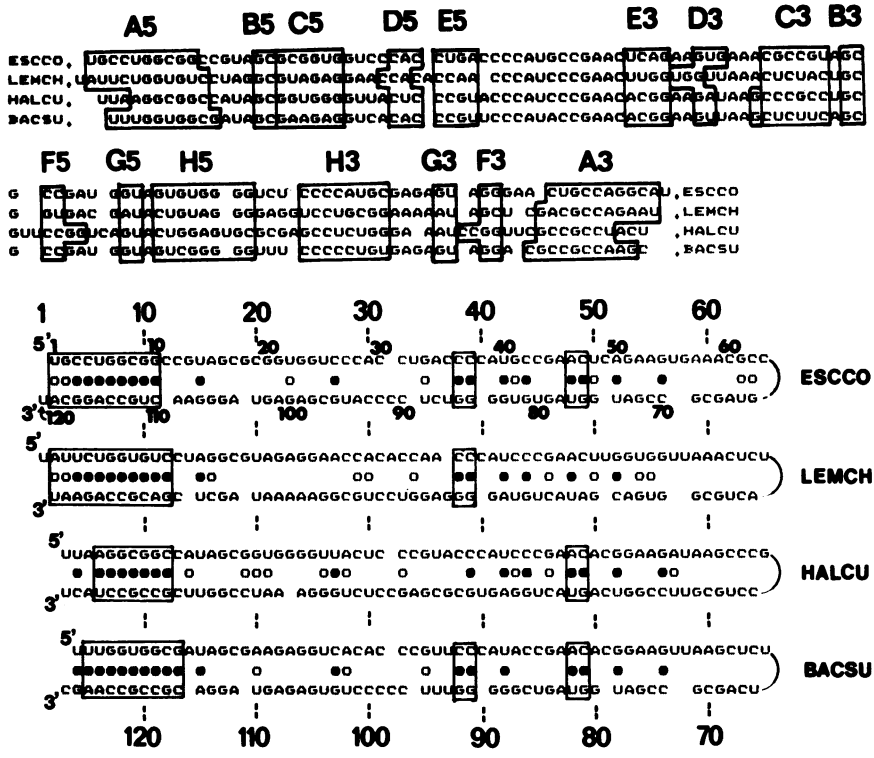


Figure 3: Helical Filtering. TOP: A chloroplast and a halophilic bacterium 5S rRNA sequences are aligned with representative Gram-negative and Gram-positive sequences. The species are: ESCCO = Escherichia coli, LEMCH = Lemna minor chloroplast, HALCU = Halobacterium cutirubrum, BACSU = Bacillus subtilis. BOTTOM: The four species are shown as they would appear during one step (group number 130) of the filtering process. The large numbers are the "spaced basenumbers": those at the top are for the 5' halves of the aligned sequences, and those at the bottom are for the 3' halves. Each "spaced" sequence has been folded back on itself so that the sum of the 5' and 3' basenumbers is 130 in every column. Thus every basepair and helical region formed in group number 130 is repeated in this diagram. Complementary nucleotides able to form basepairs are indicated by a dot. For this filtering the strigency was set at three, meaning that a helix (of at least two adjacent basepairs) must be common to at least three out of the four sequences to be saved. Black dots denote basepairs satisfying the stringency, and white dots those which do not. Isolated black dots (those having no other black dots immediately adjacent to them) are first eliminated. Then, if a helix contains at least one black dot, the entire helix is saved as a preregion. After GU's are trimmed from the ends, those helices still having two or more basepairs are saved, symbolized here by being enclosed in boxes. Group 130 contains helix A (Figure 4) and two more short helices.

As illustrated in Figure 3, patterns of basepair formation (permitted pairs are AU, GC, and GU) are examined for all group numbers. Helices are filtered according to a stringency criterion set by the user (typically one missing helix would be permitted in a group of 6-8 sequences since most 5S sequences were not determined by rapid methods and could have errors.) Helices which have too few homologs in other species are discarded, and those which survive filtering are added to a list of preregions. After temporary breaks due to spacing have been repaired, preregions are assigned energies according to published data (44). No biochemical data exist for energies of loops bounded by GU pairs (17; 44), so that although such groups exist in tRNA, all GU pairs occurring at ends of a helix have been removed. Since helices must have at least two basepairs to be saved, occasionally a helix of two standard base pairs will be kept in one sequence while a helix of one AU pair and one GU pair will be discarded in another species even though both helices will have survived the filtering step.

The RNA structure program was modified to either generate a set of unfiltered structures from a given nucleotide sequence, or to accept a region table produced by the filtering program to yield a set of filtered structures. In either case, if all helical regions are used to generate a structure the result is called a complete solution, but if some regions are eliminated by the user prior to structure building the result is a partial solution. In addition, regions may be subjected to greater or lesser degrees of branch migration (44), in order to resolve conflicts in basepairing between different helices that share common nucleotides.

The computer system available at the time of this study (APL*PLUS implemented on UCLA's IBM 370/3033) provided only 64 kilobytes of in-memory workspace, sufficient space to solve structures involving no more than about 225 possible regions. The unfiltered region table of a typical 5S rRNA molecule contains about 350 regions (of two or more basepairs) and that of 5.8S rRNA contains about 650 regions, not including subregions that are added by branch migration. Hence only partial unfiltered solutions of 5S and 5.8S RNA were obtained on our computer. We had no difficulty obtaining complete filtered solutions in the cases discussed here since filtering greatly reduced the number of regions to be considered.

The helical filtering process is analogous to other types of noise filtering in the sense that only those data that are consistent with some higher level organizational principle are kept. For example, in optical diffraction of electron micrographs only data that fit a predetermined sym-

metry are kept, and in RNA filtering only "homologous" helices are kept.

RESULTS

Generalized secondary structures for the 5S RNA's from gram negative and gram positive bacteria are shown in Figure 4. These structures illustrate the helical regions found in the minimum energy filtered structures common to the sequences listed in Figure 2. The constant nucleotides are indicated as described in the legend. Both general models contain eight helical regions, however, three differences in secondary structure, in addition to the many minor differences in sequence details, are also present. These differences involve helix H, helix D and the total number of nucleotides (120

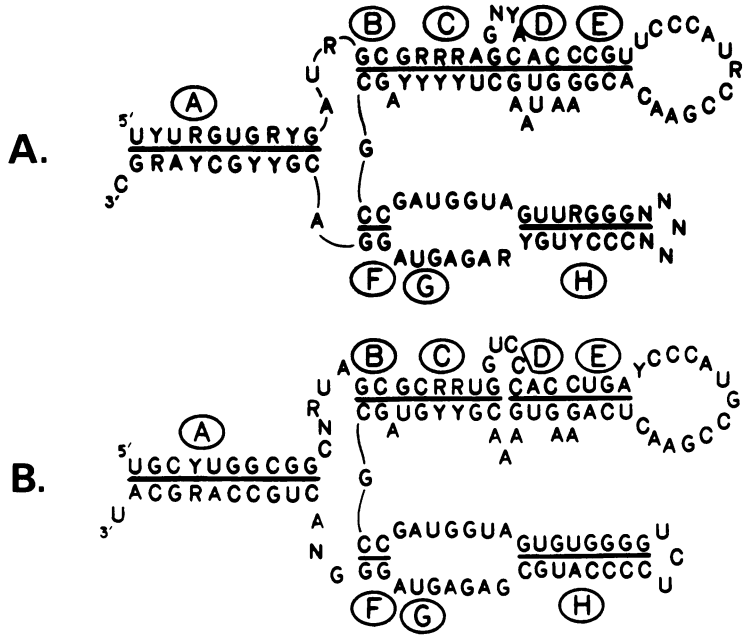


Figure 4: "Wishbone" Models for Prokaryotic 5S rRNA. General models of 5S rRNA from gram positive bacteria and from gram negative bacteria are shown in (A) and (B) respectively. Constant specific nucleotides are indicated by A, C, G, U, with constant purines indicated by R and pyrimidines by Y. Non-conserved bases are labelled N. The data were compiled from Hori and Osawa (45). Constant, here, means present in at least 12 of 15 of gram-negative bacteria (lines 25-PHO to 39-ECKD) or in at least 11 of 14 gram-positive bacteria (lines 41-BME1 to 54-CPA). A. nidulans, P. flourescens and T. aquaticus were not counted. Since their list contains many variants of E. coli and Bacillus, it is probably not representative sampling of prokaryotes.

in gram positive and 116 in gram negative bacteria). The E. coli sequence (Figure 5A) is typical of gram negative bacteria and the B. subtilis sequence (Figure 5C) is typical of the gram positive bacteria. Detailed solutions for specific sequences are described in the following paragraphs.

The lowest energy filtered secondary structures for E. coli 5S rRNA and

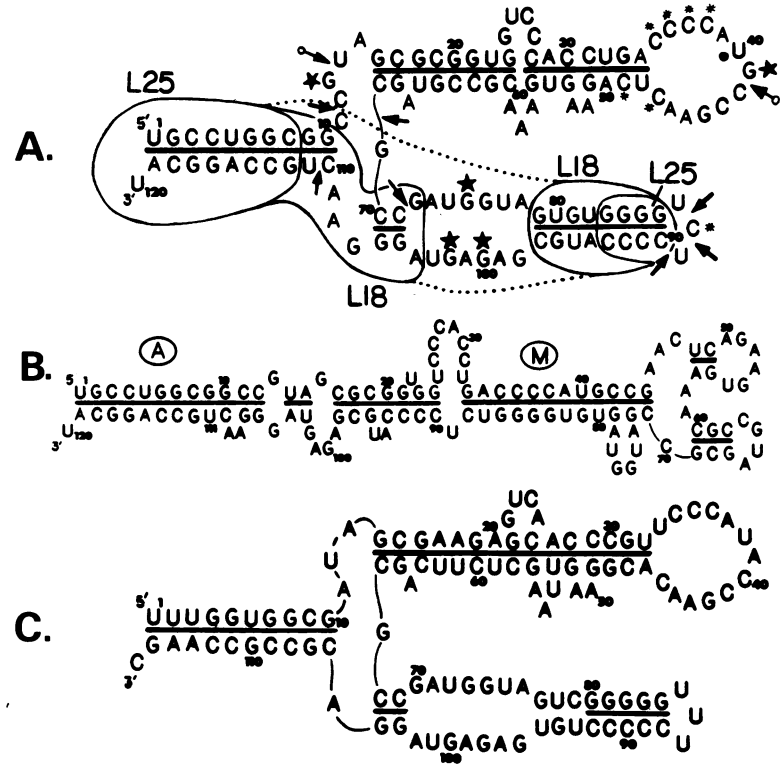


Figure 5: E. coli and B. subtilis 5S rRNAs. The best structure for E. coli (A) and for B. subtilis (C) 5S RNA generated by helical filtering using generated helices common to at least 12 out of 14 of the spaced sequences (in Figure 2). This structure is thought to correspond to the A or native form of 5S RNA. (B) The best computer generated structure of an unfiltered partial solution carried out in four steps, as described in the text. This structure is likely to be related to the B form of E. coli 5S RNA. In (A), (B), and (C), helices are highlighted by lines between the basepairs and are identified by capital letters. Domains of E. coli thought to be protected by ribosomal proteins are enclosed, and sites of chemical modification (✱ = kethoxal, ● = carbodiimide, ✳ = methoxamine) and partial enzymatic cleavage (↙ = pancreatic RNase, ↙ = T1 RNase) are indicated in (A) and (B). Base-numbers of some of the nucleotides are shown.

for B. subtilis 5S rRNA (when filtered in a group of 14 gram positive and gram negative sequences) are shown in Figures 5A and 5C. These structures have energies of -229.3 kilojoules/mole (kJ) and of -217.2 kJ, respectively, and both fit the general gram negative and gram positive models. The E. coli model corresponds exactly to the computer produced structure except that base pair G_{18}-U_{85} has been added since the structure building program does not accept G-U pairs at the end of helices. The B. subtilis intercalated base pair C_{26}-G_{57} was added manually as were the G_{47}-U_{95} and U_{48}-G_{94} pairs.

The upper three basepairs in helix H (distal to the hairpin loop) of E. coli are weak (two GU pairs and one GC pair) but still energetically favorable, according to our rules. In B. subtilis however, two stacked GU pairs are separated from helix H by a CU "pair," and are shown as weakly basepaired in Figure 5C. Nearly all prokaryotes have five strong (usually GC) pairs proximal to the hairpin loop and three weaker (GU) pairs distal to the loop. This suggests that extremely weak but complementary partners may be required. In this region, a U_{80}-G_{96} (gram negative) pair is strictly conserved in helix H in all procaryotic sequences (see Figure 4), and the pair below it is usually UU in gram positive sequences and always GU in gram negative sequences. Pairs of pyrimidines occasionally occur in tRNA helices (see for a discussion 50), but there are no rules available for calculating their energies in our models. Thus we do not show them as strictly paired although by analogy with the gram negative structure this probably occurs.

Although these are the minimum energy homologous secondary structures, they are not the lowest energy secondary structures. In general, much lower energy structures are obtained if non-homologous solutions are allowed. As an example, we have calculated a partial unfiltered solution for the E. coli 5S sequence. This structure, unlike the lowest energy homologous structure, is linear. Indeed, a virtual continuum of linear structures have lower energies than the homologous solution. Although we calculated only a partial solution (due to lack of sufficient in-memory computer storage) a reasonable approximation to the minimum energy unfiltered solution was obtained by the following approach. First, partial solutions always contained the two long helices (helix A and helix M) that are labeled in Figure 5B. Using these two helical regions to partition the molecules into two independent domains (44), we could solve each domain completely and separately. If a better unknotted secondary structure exists, it must lack at least one of the long helices, A or M, which seems very unlikely. The energy of this linear model -292.4 kJ

(-69.9 kcal/mole), is significantly greater than that of the lowest energy E. coli homologous structure. We assume, however, that the homologous solution is favored in situ by tertiary interactions as well as by interactions with other ribosomal components.

The sequences of two thermophilic bacteria, B. stearothermophilus and T. aquaticus have all the "common" helices. Their secondary structures have two principal features that distinguish them from other 5S sequences (Figure 6). Their total energies are particularily large (-243.8 kJ for B. stearothermophilus and -278.0 kJ for T. aquaticus) and an atypical arrangement is found in the A helix of both. The A helix of T. aquaticus contains one unpaired nucleotide (U_{112}) and the A helix of B. sterothermophilus can form 11 nucleotide pairs (rather than the normal 10). In addition, helix C is more stable in these thermophiles than in B. subtilus. T. aquaticus, however, derives most of its added energy from the many stacked GC pairs (even though U_{112} is unpaired) of helix A. This is in agreement with the evidence that 5S rRNA from these two thermophiles has a significantly increased T_m for denaturation (32). T. aquaticus has two clearly resolved

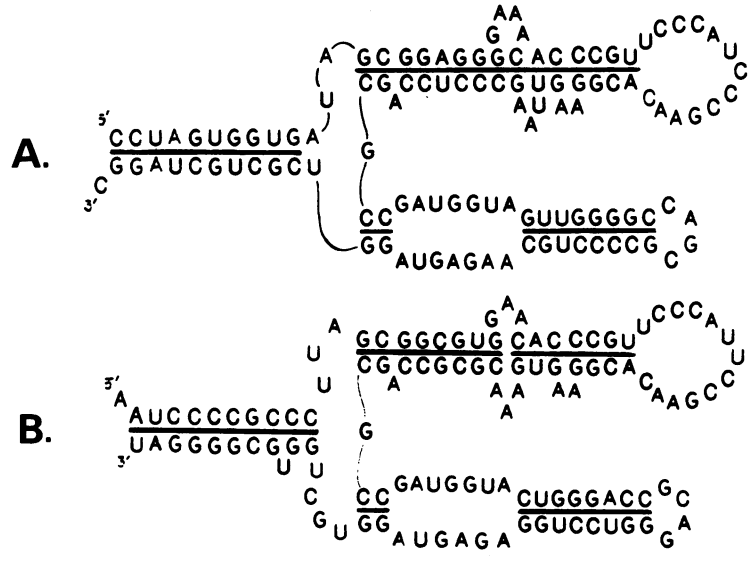


Figure 6: 5S rRNA's from Themophilic Bacteria. The best complete filtered (12 of 14) solutions for Bacillus stearothermophilus (A) and for Thermus aquaticus (B). No chemical modification or enzymatic cleavage data are available.

T_m derivative peaks (at 77°C and 86°C) that could reflect the melting of the very stable A helix, in contrast with E. coli and B. stearothermophilus where the higher peak is evident only as a shoulder (32).

The minimum energy homologous secondary structure calculated for Halobacterium cutiribrum 5S rRNA (shown in Figure 7A) contains all of the helices of the common 5S structure. Its most unusual features are that helix F is four base pairs long rather than two, and that the connection between

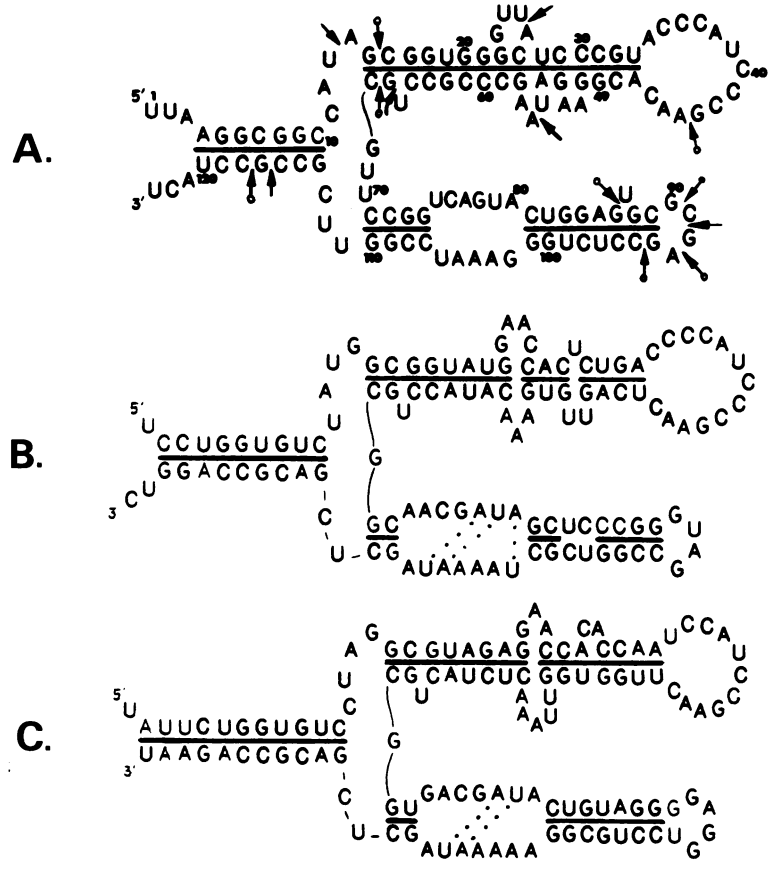


Figure 7: 5S rRNA from a Halophilic Bacteria, a Blue-Green Bacteria and a Chloroplast. (A) The best filtered solution (3 of 4) for H. cutiribrum. (B) The best filtered (12 of 14) solution for A. nidulans 5S rRNA and for (C) L. minor chloroplast (3 of 4). No chemical modification or enzymatic cleavage data are available for A. nidulans and L. minor. This lowest energy Lemna minor structure corresponds to a slightly different alignment in the vicinity of helix D than in Figure 3.

helices C and D is of the gram positive type rather than the gram negative type (see the discussion of the intercalated base pair). Enzymatic degradation studies (51) show that the major site of protein binding in H. cutiribrum corresponds to that found for E. coli (Figure 5A), suggesting that the common model exists in situ. H. cutiribrum 5S rRNA has been successfully reconstituted into B. stearothermophilus ribosomes under low-salt conditions, further suggesting H. cutiribrum 5S rRNA is homologous to other 5S rRNA species.

The minimum energy homologous secondary structures for Duckweed (Lemna minor) chloroplast 5S rRNA and for the blue green bacterium Anacystis nidulans 5S structure are shown in Figure 7B and 7C, respectively. A structure nearly identical to the A. nidulans folding has been listed among partial computer solutions (44). Both the blue green bacteria and the chloroplast sequences contain all eight helical regions. The arrangement of base pairs in helix E in chloroplast 5S depends upon details of sequence spacing. Other spacings result in higher, but similar energy structures. The arrangement of the C, D and E helices in the constant arm, although differing from both the gram positive and gram negative pattern, most closely resembles that of a gram negative bacterium, since it lacks the intercalated base pair. The D and E helices and the arrangement of the constant arm of the blue green bacterium resembles the chloroplast and the gram negative pattern, while the A helix of A. nidulans is closest to the gram positive type.

DISCUSSION

"Correlation with other Data"

The 5S rRNA from E. coli is the most extensively studied molecule of its class, and for this reason it is emphasized in this discussion. E. coli 5S exists with two secondary foldings (24; 31; 33; 34), an "A-form" and a "B-form." The A-form binds ribosomal proteins specifically under reconstitution conditions (52) suggesting that it is the functional form and that a homologous secondary structure solution should correspond to and be compared with the A-form. No functional role has been found for the B-form.

The enzymatic cleavage and chemical modification data have been indicated on the "wishbone" model for E. coli 5S RNA in Figure 5A. In general these data support the details of the model. When the A-form is freed from ribosomal proteins, nucleotide G_{41} is the nucleotide most sensitive to kethoxal (7; 33), and U_{40} is the nucleotide most sensitive to carbodiimide (53). The first cleavage of T1 RNase occurs at G_{41} (54; 55), and

both RNase IV and sheep kidney nuclease reveal a reactive region (7). Although five of the 7 C's contained in this loop are sensitive to methoxamine (7), the sequence CCG_{44} is not reactive with most enzymes and reagents [G_{44} can be slightly kethoxalated (10)]. G_{13} is readily kethoxalated [nearly as reactive as G_{41} (7; 56)], and is also very sensitive to T1 RNase (55). All four of the sites reported for nitrous acid deamination [A_{15}, A_{66}, A_{78}, and A_{104} (7)] are unpaired in the wishbone model, a methoxamine site C_{88} (7), and three additional kethoxal sites [G_{75}, G_{100}, and G_{102} (7)] are also unpaired in the model. Others (57) however did not find these last 3 kethoxal sites reactive.

The A-form, but not the B-form (17; 34), can bind specifically and stably to ribosomal proteins L18 and/or L25, and that complex has been reported to further bind a complex of L6 and 23S rRNA (58). Under those conditions, pancreatic RNase cleaves after C_{11} and C_{68}, detaching an entire stretch of 57 nucleotides from the complex and totally degrading it to oligonucleotides. The remainder of the molecule remains bound to the protein complex and is unreactive with pancreatic RNase except for UCU_{89} which is cleaved (58). The lack of T1 RNase cleavage data for bound complexes and also the large number of purines that are present in single-stranded loops in the wishbone model make more precise mapping and verification difficult. Nevertheless, the patterns of protection for L18 and L25, both individually and together, suggest that helices F and H could be close to helix A in the three dimensional 5S structure.

An interaction has been proposed between the sequence T CG in tRNA and the sequence CGAAC in 5S rRNA (30; 59; 60; 61; 62). In the wishbone model, as well as in most models proposed, this sequence is single stranded.

Although we can relate the wishbone model to the functional, A-form of 5S RNA, we cannot easily relate the B-form to a single secondary structure. Indeed, the B-form is thought to correspond to a spectrum of related secondary structures. A feature characteristic of the E. coli sequence is that the lowest energy secondary solutions, unlike the homologous solutions, are linear. These structures are candidate "B" structures. A clue to this interpretation comes from experiments (14) demonstrating that a complex of fragments 25-41 and 80-96 could be found after T1 digestion of the B-form, but not of the A-form. This sequence contains the M helix characteristic of linear 5S models (see Figure 5B). In this model G_{41}, (which is easily kethoxylated in the A-form is protected in the B-form) is within a double stranded region, paired to U_{80}. We have no direct evidence, however, that

the lowest energy non-homologous structure corresponds to the B-form.

Lecanidou et al. (30) have shown that 5S rRNA in the absence of protein and Mg^{++} approaches equilibrium as a mixture of both the A-form and the B-form, and that the addition of Mg^{++} drives the conformation equilibrium completely to the A-form at all temperatures. If the total free energy of the A-form is lower than the B-form, then tertiary energy contributions compensate for the difference in secondary structure energy between the B- (-292.4 kJ) and the A-forms (-229.3 kJ). This suggests that tertiary interactions of at least -63.1 kJ contribute to the stability of the A-form and implies that they make a significant contribution toward the total 5S energy. This is consistent with our rationale of using only the homologous component of the secondary structure energy to carry out minimization searches.

If the A-form were only slightly more favorable than the B-form at equilibrium, kinetic barriers could exist to prevent transformation from the A-form to the B-form. The A-form and the B-form would thus be local minima separated by a high activation energy. This hypothesis is given some support by a measurement of the activation energy required for "renaturation" of E. coli 5S rRNA, which was +259.8 kJ (31; 63). Since transformation between our models for the A- and B-forms requires breaking all of the basepairs except those in helix A, this measured value agrees fairly well with our predictions.

Our calculations suggest that certain prefered folding pathways are likely. The unfiltered partial solution for Model B was the best secondary structure from more than 100 million possible structures (not including substructures) and the complete solution (had it been calculated) would have involved more than 10^{12} possible structures. For 5S rRNA to assume a different secondary structure every 10^{-8} second, it would need nearly three hours to search each structure. Experiments on the folding of tRNA (64) argue that folding begins as soon as the 5' end of the tRNA molecule is synthesized. Viewed in the light of this folding pathway, the entire constant length arm of the "A" form containing helices B, C, D and E could be folded, before helix M (of the "B" form) would be available for base pairing, since the M3 region is on the 3' side of the B3 region.

Gram negative 5S rRNA sequences have 120 nucleotides, gram-positive sequences have only 116 and the length of the blue green bacteria are intermediate. The alignment in Figure 2, taking into account conserved purines, pyrimidines, and nucleotides shows that the four missing bases were added to the 5' and 3' ends of the 5S molecule, so that the loop connecting helices A, B, and F is four bases longer in gram-negative sequences. The apparent

evolutionary requirement for a constant number of basepairs in helix A seems to have forced coordinated base changes adjacent to helix A that resulted in the movement of helix A by two basepairs. The blue green bacteria, although resembling the gram positive bacteria in their A helix, resemble the gram negative bacteria in that they have lost the C_{26}-G_{57} intercalated base pair described in the following section.

"An Intercalated Base Pair"

An unusual stacked arrangement of base pairs may occur in 5S RNA that has not been previously reported. We have named this arrangement an "intercalated base pair" since it is inserted between two helical regions. Since our program searches for base paired regions two pairs and longer it did not appear in our computer calculated minimum energy homologous structure. It was discovered when we compared the lengths of the arm containing helices B, C, D and E in gram positive and in gram negative bacteria.

Helices B, C and E are two, six and four base pairs long, respectively, in both gram negative and positive 5S sequences (see Figure 8). Helix D is three and two base pairs long in gram negative and positive bacteria, respectively. If a C_{26}-G_{57} base pair is intercalated between helices C and D in the gram positive structure, then the constant arm is increased to 15 nucleotide pairs, as in the gram negative sequence. The intercalated base pair is strongly supported by all the criteria used in this paper to identify helical regions. It is energetically favored (by -2.0 kJ) and evolutionarily conserved, being exclusively present in gram positive sequences and absent in gram negative sequences. In addition, it satisfies the isomorphism criterion by preserving the length of the entire intercalated base arm at 15 base pairs in all bacterial sequences. In a second paper, (Studnicka, Eiserling and Lake, in preparation) we show that this arm is also 15 nucleotides long in 5S molecules from eukaryotic organisms.

"The Bulge Loops: A Coordinated Base Pair"

Another unusual feature of the constant length arm is the role of the bulge loops between helices. The sequences of three of these loops are constant, or nearly so, within the positive and negative classes. An A occurs at the 3' connection between helices B and C; an AA is present at the 3' connection between helices D and E; and an AAA is found at the 3' connection between helices C and D in gram negative sequences while a UAA is present at the 3' connection between the intercalated base pair and helix D in gram positive sequences. In this respect, H. cutirubrum fits the gram positive pattern, the L. minor chloroplast and A. nidulans fit the gram negative pat-

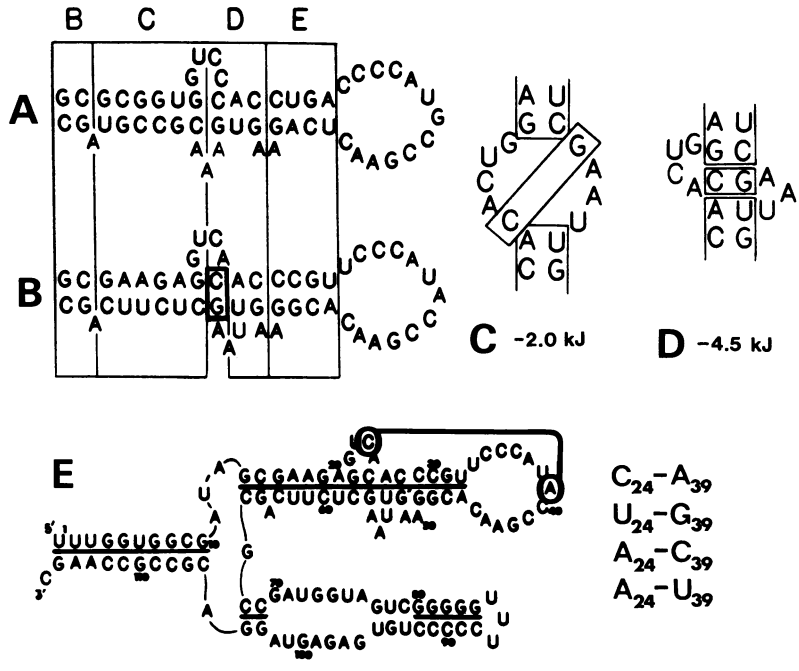


Figure 8: An Intercalated Base Pair. The sequences and secondary structures of the constant length arm are shown for E. coli and B. subtilis, in (A) and (B), respectively. The intercalated C_{26}-G_{57} pair appears in the helix D region in the B. subtilis sequence. Energy calculations show the energy of the unstacked sequence (C) and intercalated pair (D) and indicte the increased stability of the intercalated pair. The coordinated basepair (E), is shown by a line connecting C_{24} and A_{39} on the B. subtilis sequence.

tern. Finally, the bulge loop connecting the 5' side of helices C and D, is variable and may offer a clue to the function of the loops. It is either three or four bases long. It's sequence is G (U/A) (C/A) (C/) in gram negative bacteria and G (U/A) (U/C/A) A in gram positive bacteria. In gram positive bacteria we have evidence that this loop specifically interacts with the E loop at the end of the constant arm. Comparison of gram positive sequences shows that nucleotides 24 and 39 form the coordinated base pairs: C_{24}-A_{39}; A_{24}-$(C/U)_{39}$ and U_{24}-G_{39}. A purine in position 24 is always associated with a pyrimidine in position 39, and vice versa. The relationship is indicated in Figure 8E and emphasized by circled bases. In the tRNA tertiary structure, direct interaction invariably occured between bases having coordinated patterns of change. This is a good indication that

a specific interaction occurs between the C-D bulge loop and the loop at the end of the constant arm.

"A Pivotal Role for G_{69}?"

Nucleotide G_{69} (G_{67} in gram positive sequences) is strictly conserved in all 5S sequences and is positioned immediately adjacent to both helix B and helix F. It occupies a keystone position separating the two arms of the wishbone and could play an important role in the overall tertiary conformation of 5S RNA. It is a candidate component of a conformational switching mechanism (such as the tertiary switch observed by Kao and Crothers, 65) since it can potentially stack on either helix F or on helix B. In particular, it could function as an intercalated base pair to connect helix A with helix F. This would involve breaking the G_{10}-C_{110} (G_{10}-C_{107}) base pair in helix A, when it is present, and substituting the pair G_{69}-C_{110} (G_{67}-C_{107}). Energetically this is slightly favored in gram positive sequences (-0.6 kJ) and slightly disfavored in gram negative sequences (+0.1 kJ). It has the following points to recommend it: 1) It is universal since it can be made in all prokaryotic 5S sequences. 2) It converts the A helix into a region 9 base pairs long in gram positive and negative sequences (except T. aquaticus). 3) It makes the A helix in prokaryotic 5S-sequences the same length as in eukaryotic sequences (discussed in Studnicka, Eiserling and Lake, in preparation). 4) And it is consistent with the resistance of E. coli 5S helices A, F and H (in the presence of L18 and L25) to pancreatic RNase (58). Although the G_{69}-C_{110} intercalated pair satisfies the evolutionary requirements for an interaction, its slight energetic change and the presence of a competing pairing scheme makes us hesitant to classify it as an interaction.

In summary, we have presented our evidence for the "wishbone" model of 5S rRNA, based on a combined comparative and energetic analysis using helical filtering. We believe the evidence strongly supports some novel features, such as the intercalated base pair C_{26}-G_{57} and the constant length of the combined B, C, D and E helices. [Other features such as a possible G helix in prokaryotes and a possible role of the invariant G_{67} (G_{69} in gram positive sequences) are speculative and will require more information to be evaluated.] In light of the general applicability and universal minimum energy properties of the "wishbone" model, we regard its principal features as being proven. We hope that this structural model will aid in the ultimate understanding of the structure and function of 5S rRNA.

ACKNOWLEDGEMENTS

We thank J. Beyer and D. Williams for photographic assistance and M. Kowalczyk for drawing Fig. 1. This work was supported by the National Science Foundation (PCM 76-15868 to FAE and PCM 76-14718 to JAL) and the National Institutes of Health (AI-14092 to FAE and GM-22150 to JAL). Computing was provided by intramural funds from the Office of Academic Computing, UCLA, to FAE and JAL. GMS is a University Fellow in Biology.

REFERENCES

1. Fox G.E., Woese C.R. (1975) Nature 256, 505-507.
2. Erdmann V.A., Fahenstock S., Higo K., Nomura M. (1971) Proc. Nat. Acad. Sci., USA 68, 2932-2936.
3. Dohme F., Nierhaus K.H. (1976) Proc. Nat. Acad. Sci., USA 73, 2221-2225.
4. Horne J.R., Erdmann V.A. (1972) Mol. Gen. Genet. 119, 337-344.
5. Wrede P., Erdmann V.A. (1973) FEBS Lett. 33, 315-319.
6. Wrede P., Erdmann V.A. (1977) Proc. Nat. Acad. Sci., USA 74, 2706-2709.
7. Bellemare G., Jordan B.R., Rocca-Serra J., Monier R. (1972) Biochimie 54, 1453-1466.
8. Cramer F., Erdmann V.A. (1968) Nature 92-93.
9. Delihas N., Dunn J.J., Erdmann V.A. (1975) FEBS Lett. 58, 76-80.
10. Larrinua I., Delihas N. (1979) FEBS Lett. 108, 181-184.
11. Nishikawa K., Takemura S. (1978) Biochem. 84, 259-266.
12. Noller H.F., Garrett R.A. (1979) J. Mol. Biol. 132, 621-636.
13. Hatlen L.E., Amaldi F., Attardi G. (1969) Biochem. 8, 4989-5005.
14. Jordan B.R. (1971) J. Mol. Biol. 55, 423-439.
15. Vigne R., Jordan R.B. (1977) J. Mol. Evol. 10, 77-86.
16. Ross A., Brimacombe R. (1979) Nature 281, 271-276.
17. Douthwaite S., Garrett R.A., Wagner R., Feunteun J. (1979) Nucleic Acids Res. 6, 2453-2470.
18. Nishikawa K., Takemura S. (1974) Biochem. 76, 935-947.
19. Nicols J.L., Welder L. (1979) Biochim. Biophys. Acta 561, 445-451.
20. Connors P.G., Beeman W.W. (1972) J. Mol. Biol. 71, 31-37.
21. Osterberg R., Sjoberg B., Garrett R.A. (1976) Eur. J. Biochem. 68, 481-487.
22. Luoma G.A., Marshall A.G. (1978) J. Mol. Biol. 125, 95-105.
23. Bellemare G., Vigne R., Jordan B.R. (1973) Biochemie 55, 29-35.
24. Scott J.F., Monier R., Aubert M., Reynier M. (1968) Biophys. and Biochem. Res. Comm. 33, 794-800.
25. Cantor C.R. (1968) Proc. Nat. Acad. Sci., USA 59, 478-483.
26. Kearns D.R., Wong Y.P. (1974) J. Mol. Biol. 87, 755-774.
27. Appel B., Erdmann V.A., Stulz J., Ackerman T. (1979) Nucleic Acids Res. 7, 1043-1057.
28. Lewis J.B., Doty P. (1970) Nature 225, 510-512.
29. Wrede P., Pongs O., Erdmann V.A. (1978) J. Mol. Biol. 120, 83-96.
30. Lecanidou R., Richards E.G. (1975) Eur. J. Biochem. 57, 127-133.
31. Richards E.G., Lecanidou R., Geroch M.E. (1973) Eur. J. Biochem. 34, 262-267.
32. Nazar R.N., Sprott G.D., Matheson A.T., Van N.T. (1978) Biochim. Biophys. Acta 521, 288-294.
33. Aubert M., Bellemare G., Monier R. (1973) Biochemie 55, 135-142.
34. Aubert M, Scott J.F., Reynier M., Monier R. (1968) Proc. Nat. Acad. Sci., USA 61, 292-299.
35. Fox G.E., Woese C.R. (1975) J. Mol. Evol. 6, 61-76.

36. Holley R.W., Apgar J., Everett G.A., Madison J.T., Marquisee M., Merrill S.H., Penswick J.R., Zamir A. (1965) Science 147, 1462-1465.
37. Zachau H.G., Dutting D., Feldmann H. (1966) Angew. Chem. 78, 392-402.
38. Madison J.T. (1968) Ann. Rev. Biochem. 37, 131-148.
39. Jordan B.R. (1971) J. Theor. Biol. 34, 363-378.
40. Nussinov R., Pieczenik G., Griggs J.R., Kleitman D.J. (1978) SIAM J. Appl. Math.35, 68-82.
41. Pipas J.M., McMahon J.E. (1975) Proc. Nat. Acad. Sci., USA 72, 2017-2021.
42. Waterman M.S., Smith T.F. (1978) Math. Biosci. 42, 257-266.
43. DeLisi C., Crothers D.M. (1971) Proc. Nat. Acad. Sci., USA 68, 2682-2685.
44. Studnicka G.M., Rahn G.M., Cummings I.W., Salser W.A. (1978) Nucleic Acids Res. 5, 3365-3387.
45. Hori H., Osawa S. (1979) Proc. Nat. Acad. Sci., USA 76, 381-385.
46. Erdmann V.A. (1978) Nucleic Acids Res. 41-47.
47. Denis H., Wegnez M. (1978) J. Mol. Evol. 12, 11-15.
48. Hori H. (1975) J. Mol. Evol. 7, 75-86.
49. Hori H. (1976) Mol. Gen. Genet. 145, 119-123.
50. Clark B.F.C. (1980) in Ribosomes: Structure, Function, and Genetics, Chambliss et al., Eds., pp. 413-444 University Press, Baltimore.
51. Nazar R.N., Willick G.E., Matheson A.T. (1979) J. Biol. Chem. 254, 1506-1512.
52. Bellemare G., Jordan B.R., Monier R. (1972) J. Mol. Biol. 71, 307-315.
53. Lee J.C., Ingram V.M. (1969) J. Mol. Biol. 41, 431-441.
54. Vigne R., Jordan B.R., Monier R. (1973) J. Mol. Biol. 76, 303-311.
55. Luoma G.A., Marshall A.G. (1978) Proc. Nat. Acad. Sci., USA 75, 4901-4905.
56. Noller H.F., Herr W. (1974) J. Mol. Biol. 90, 181-184.
57. Noller H.F., Garrett R.A. (1979) J. Mol. Biol. 132, 621-636.
58. Gray P.N., Bellemare G., Monier B. (1972) FEBS Lett. 24, 156-160.
59. Erdmann V.A., Sprinzl M., Pongs O. (1973) Biophys. and Biochem. Res. Comm. 54, 942-948.
60. Sprinzl M., Wagner T., Lorenz S., Erdmann V.A. (1976) Biochem. 15, 3031-3039.
61. Richter D., Erdmann V.A., Sprinzl M. (1973) Nature New Biology 246, 132-135.
62. Schwarz U., Menzel H.M., Gassen H.G. (1976) Biochem. 15, 2484-2490.
63. Erdmann V.A. (1977) Progress in Nucleic Acids Research and Molecular Biology Prog. Nucleic Acids Res. Mol. Biol. 18, 45-90.
64. Boyle J., Robillard G.T., Kim S.-H. (1980) J. Mol. Biol. 139, 601-625.
65. Kao T.H., Crothers D.M. (1980) Proc. Nat. Acad. Sci., USA 77, 3360-3367.
66. Simoncsits A. (1980) Nucleic Acids Res. 4111-4124.
67. Matheson A.T., Nazar R.N., Willick G.E., Yaguchi M. (1980) in Genetics and Evolution of Transcriptional and Translational Apparatus, Osawa et al. Eds., Kedansha Scientific Press, Tokyo.