Volume 11 Number 3 1983 Nucleic Acids Research

Consensus structure and evolution of 5S rRNA*

Hans Küntzel, Birgit Piechulla and Ulrich Hahn+

Max-Planck-Institut für experimentelle Medizin, Abteilung Chemie, Hermann-Rein-Str. 3, D-3400 Göttingen, FRG

Received 23 November 1982; Revised and Accepted 5 January 1983

ABSTRACT

A consensus structure model of 5S rRNA presenting all conserved nucleotides in fixed positions has been deduced from the primary and secondary structure of 71 eubacterial, archaebacterial, eukaryotic cytosolic and organellar molecules. Phylogenetically related groups of molecules are characterized by nucleotide deletions in helices III, IV and V, and by potential base pair interactions in helix IV. The group-specific deletions are correlated with the early branching pattern of a dendrogram calculated from nucleotide substitution data: the first major division separates the group of eubacterial and organellar molecules from a second group containing the common ancestors of archaebacterial and eukaryotic/cytosolic molecules. The earliest diverging branch of the eubacterial/organellar group includes molecules from Thermus thermophilus, T. aquaticus, Rhodospirillum rubrum, Paracoccus denitrificans and wheat mitochondria.

INTRODUCTION

The 5S rRNA component of the large ribosomal subunit (1) is a suitable marker molecule to study pre-cambrian phylogeny: a comparison of 5S rRNA sequences from eukaryotes, eubacteria, archaebacteria and organelles (2-14) has revealed a considerable conservation of primary (15,16) and secondary (17-20) structure. Here we present a general consensus structure for all types of 5S rRNAs. Our model is in agreement with previously published secondary structure models for eubacterial (19,20), eukaryotic cytosolic (18,19) and archaebacterial (7) molecules, and in addition presents all conserved nucleotides in fixed positions. The observed group-specific deletions and base pair interactions support the phylogenetic grouping of 5S rRNA molecules based on nucleotide substitution data.

METHODS

5S rRNA was extracted from the snail Arion rufus, purified by preparative gel electrophoresis and sequenced by chemical cleavage methods as described (21). Methods used for computer-aided construction of dendrograms have been reported in detail (16,25).

RESULTS AND DISCUSSION

Fig. 1 shows an alignment of selected 5S rRNA sequences representative for the groups of metazoa [sequence 1], protozoa [19,20], plants [32], archaebacteria [41-46], eubacteria [47,50,56,60,61,67-69], chloroplasts [63] and mitochondria [71].

The alignment is similar, but not identical, to previously published schemes (15,19) due to the inclusion of more sequence data.

The consensus structure of Fig. 2A is based on the alignment of 71 5S rRNA sequences (2-14) and contains 70 positions where either A,U,G,C or a

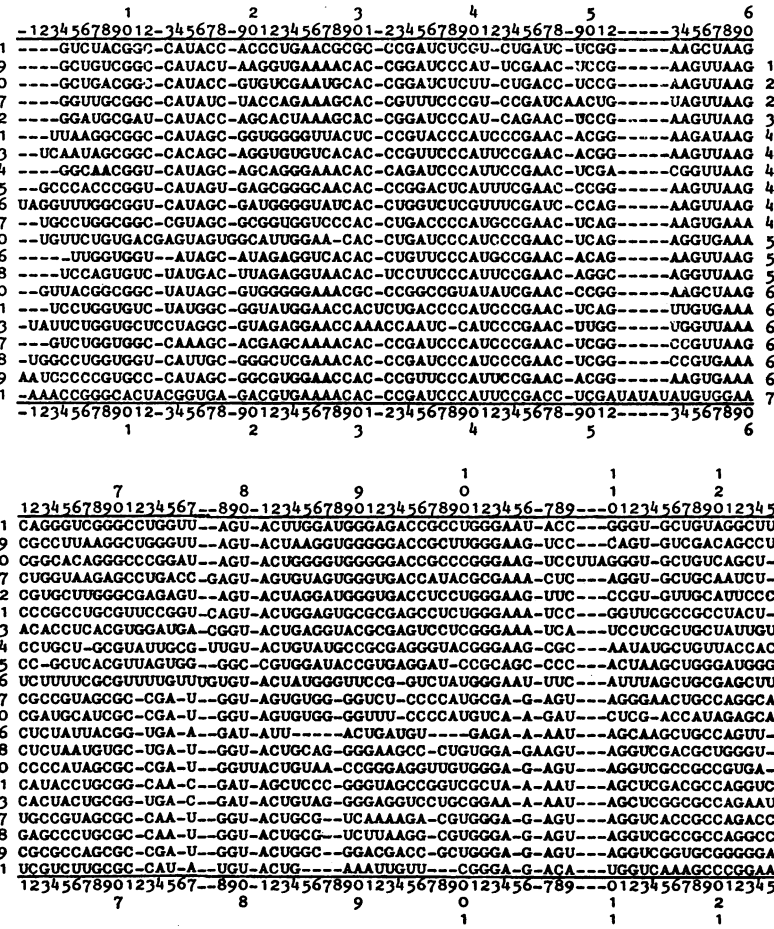

Figure 1. Alignment of 21 selected 5S rRNA sequences. The nucleotide numbers refer to the consensus structure of Fig. 2, the sequence numbers refer to the list of organisms shown in Fig. 4.

purine or pyrimidine nucleotide is conserved in more than 80 % of all sequences. The consensus sequence is folded according to the five helix model (18,19), and potential base pairs shown in straight lines are conserved in all molecules, with few exceptions. Some alternative or additional base pair interactions are possible, although in a smaller number of molecules: the base pairs 55/29 and 56/28 could be formed in 48 sequences (19), the base pairs 12/113 and 13/112 in six eubacterial molecules (20), and helix III could be extended by two base pairs (36/46 and 37/45) in 60 molecules (19). A tertiary interaction between nucleotides 42-45 and 80-77 observed in E.coli 5S rRNA (22) is possible in other eubacteria except cyanobacteria.

Fig. 2B shows a mollusc 5S rRNA (from the snail Arion rufus) written in

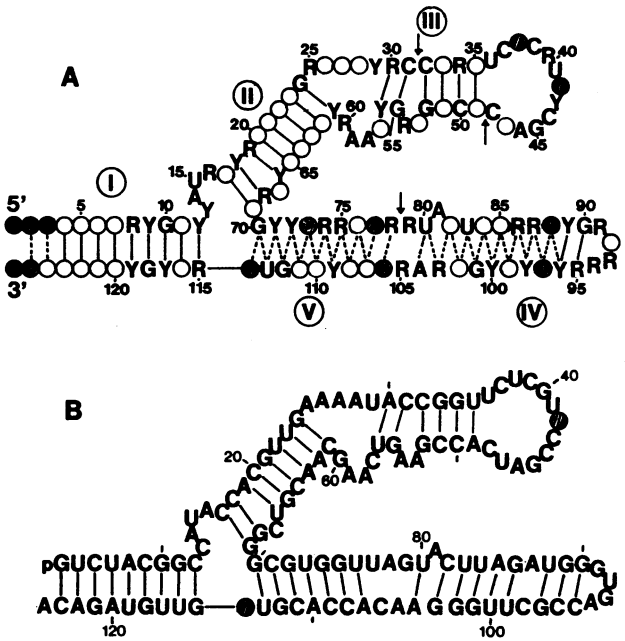

Figure 2 A: Consensus structure of 5S rRNA, based on 72 sequenced molecules (2-14). Nucleotides conserved in more than 80 % of all analyzed sequences are indicated (R, purine nucleotide, Y, pyrimidine nucleotide). Group-specific deletions are marked by shaded circles, group-specific insertions by arrows. Potential hydrogen bonds between the base pairs A/U, G/C and G/U conserved in at least 95 % of all molecules are shown in straight lines, alternative base pair interactions in helices IV and V in broken lines. B: Nucleotide sequence and potential secondary structure of 5S rRNA from the snail Arion rufus. The sequence was determined by chemical cleavage methods as described (21). Nucleotides 42 and 114 of the consensus sequence (shaded circles) are deleted in all eukaryotic (cytoplasmic) 5S rRNA molecules.

the consensus configuration. The nucleotide sequence was determined independently by us and by Fang et al. (14).

The group-specific patterns of nucleotide deletions and helical interactions are shown in Fig. 3. Nucleotides 42 and 114 of the consensus sequence are deleted in all eukaryotic cytosolic species, nucleotides 73, 77, 88 and 106 are deleted in all eubacterial and organellar sequences, and none of these nucleotides are deleted in the six archaebacterial sequences. Instead, some individual deletions (nucleotides 67, 94 and 99) are found in the archaebacterial sequences 44, 45 and 46 (sequence numbers refer to Fig. 4). An addi-

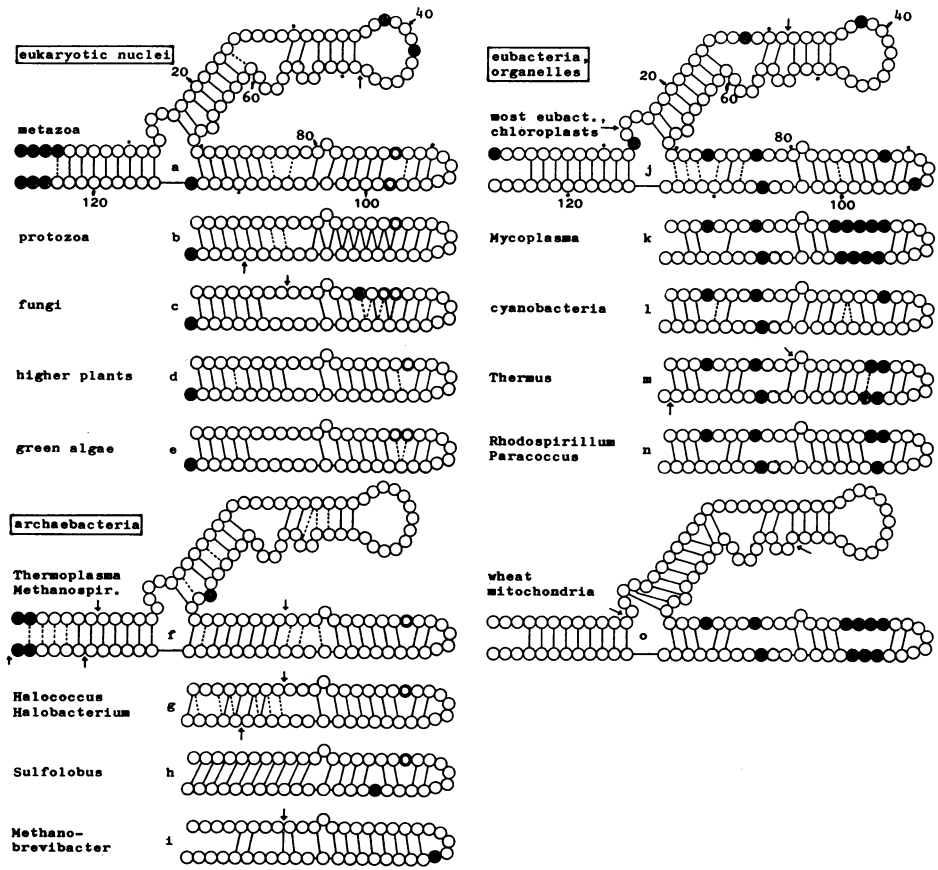

Figure 3. Deletions, insertions and potential base-pairing interactions in 5S rRNAs. Residues marked by black circles are deleted in all members of a given group, those marked by shaded circles are deleted in some organisms. Inserted nucleotides are indicated by arrows. See also the alignment of Fig. 1. The Halococcus molecule contains a long insert of 108 nucleotides at 109/110 (2).

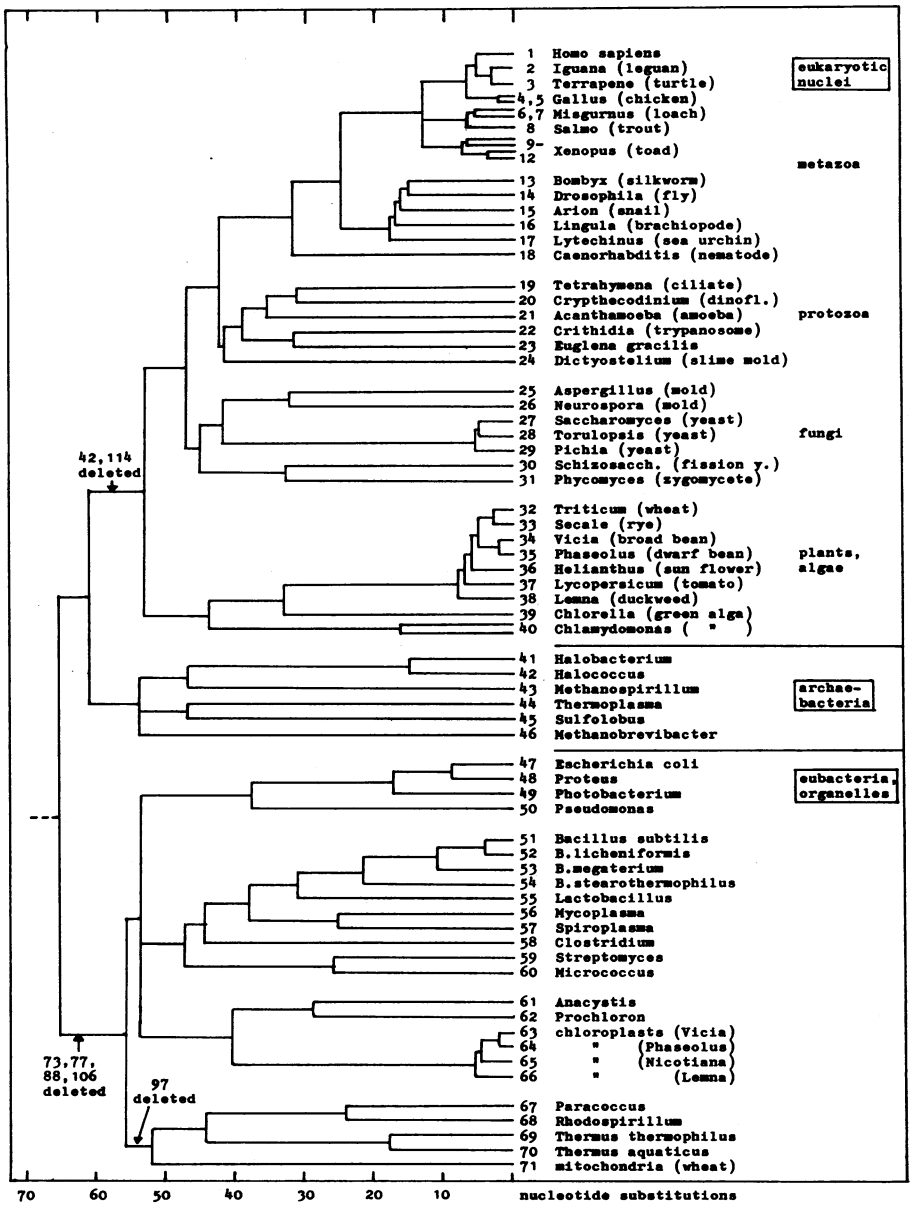

Figure 4. Phylogenetic tree based on the nucleotide substitution analysis of 5S rRNAs. Nodal points are expressed as the number of nucleotide substitutions between pairs of sequences or protosequences (16,24,25), and the early branching pattern is correlated with a phylogenetic interpretation of group-specific deletions (see Fig. 3). All sequences except 15 (14), 18 (3), 23 (4), 24 (5), 31 (6), 43 (7), 46 (7), 57 (8), 59 (9), 60 (10), 62 (11), 67 (11), 68 (12) and 69 (13) are compiled in ref. 2.

tional deletion of nucleotide 97 is a common feature of four eubacterial sequences from T. thermophilus (13), T. aquaticus (2), R. rubrum (12) and P. denitrificans (11). In two molecules (from wheat mitochondria (2) and Mycoplasma (2)) helix IV is further reduced by deletions of 7 to 9 nucleotides including positions 88 and 97. However, this could be a convergence phenomenon, because the two sequences belong to different affinity groups (see below).

The most conserved secondary interactions are in helices I, II and III: note the looped out base 67 of helix II (potentially base-paired only in the mitochondrial sequence) and the two looped out bases 53 and 54 of helix III (helix I contains a looped out base only in the two archaebacterial sequences 45 and 46).

Helices IV and V are more variable in their base pairing and deletion patterns: in helix IV nucleotides 80 and 82-86 can form base pairs either with 103-97 (configuration 1) or with 104-98 (configuration 2). The helix IV configurations shown in Fig. 3 are energetically (23) favourable over the alternative configurations, and a certain phylogenetic group specificity of these structural modulations cannot be overlooked: configuration 1 dominates in most eubacteria (j,n) and archaebacteria (f,g,i), in all plants and algae (d,e) and fungi (c, with the exception of Neurospora (2) due to the deletion of nucleotide 84). In all six protozoa both configurations are equivalent, and configuration 2 dominates in all 18 metazoan molecules. Helix IV is interrupted by a looped out base (87 or 88) in plants, algae, fungi and in all archaebacteria (one exception). In all protozoa configuration 1 produces a looped out base 87, and configuration 2 results in a mismatch at either 84/101, 85/100 or 86/99. In all metazoan molecules the bases 87/98 are mismatched (A/C).

A phylogenetic tree analysis of 46 sequences of 5S rRNAs has previously been described (16). The tree of Fig. 4 is constructed by the same method and is based on a larger number of sequences.

The nematode C. elegans (3) diverges from the metazoan branch prior to the separation between invertebrates (including the snail sequence of Fig. 2B) and vertebrates. Three base pairs in helix II and III (21/64, 30/56 and 35/49) are either C/G, G/C and A/U, respectively, in all 12 vertebrates, or G/C, A/U and U/A in the five invertebrate sequences 13-17. The nematode sequence is invertebrate-like in two positions and vertebrate-like in one position. The protozoan group includes Euglena gracilis (4) and the slime mold Dictyostelium discoideum (5) due to sequence affinities and the group-specific ambiguity of helix IV configurations. The lower fungus Phycomyces blakesleeanus (6) appears

to diverge from the fungal branch, together with the fission yeast S. pombe (2), prior to the separation between filamentous fungi and budding yeasts.

The subtree of Gram-positive bacteria includes two actinomycetes (Streptomyces griseus (9) and Micrococcus lysodeikticus (10)), and the two mycoplasmas M. capricolum (2) and Spiroplasma (8). The two related eubacteria P. denitrificans and R. rubrum appear to share a common ancestor with the two Thermus species, and the wheat mitochondrial sequence has a higher affinity to the protosequence of this early diverging group (53 substitutions) than to the protosequences of other eubacterial groups (between 61 and 66 substitutions). P. denitrificans and R. rubrum have earlier been recognized as possible relatives to proto-mitochondria on the basis of cytochrome c homologies (26).

The six archaebacterial sequences form an extremely heterogenous group (7) but appear to exhibit a common root, as suggested by substitution analysis and by the group-specific presence of the six nucleotides 42,73,77,88,106 and 114. The archaebacterial protosequence is significantly more related to the eukaryotic-cytosolic protosequence (52 substitutions) than to the eubacterial/organellar protosequence (65 substitutions), and a common ancestor of archaebacterial and eukaryotic molecules is also suggested by five eukaryotic features (presence of nucleotides 73,77,88,106, and looped out base 87) versus two prokaryotic features (presence of nucleotides 42 and 114).

ACKNOWLEDGEMENT

We thank Dr. R.T. Walker for communicating to us the Spiroplasma 5S rRNA sequence prior to publication and M. Heidrich for helpful advice.

*Dedicated to Friedrich Cramer on the occasion of his sixtieth birthday.

+Present address: Institut für Kristallographie, Freie Universität, 1000 Berlin, 33, FRG.

REFERENCES

1. Erdmann, V.A. (1976) Progr. Nucl. Acid Res. Mol. Biol. 18, 45-90.
2. Erdmann, V.A. (1982) Nucl. Acids Res. 10, r93-r115.
3. Butler, M.H., Wall, S.M., Luehrsen, K.R., Fox, G.E. and Hecht, R.M. (1981) J. Mol. Evol. 18, 18-23.
4. Hori, H., Osawa, S. and Iwabuchi, M. (1980) Nucl. Acids Res. 8, 5535-5539.
5. Delihas, N., Anderson, J., Andresini, W., Kaufman, L. and Lyman, H. (1981) Nucl. Acids Res. 9, 6627-6633.
6. Andersen, J., Andresini, W. and Delihas, N. (1982) J. Biol. Chem. 257, 9114-9118.
7. Fox, G.E., Luehrsen, K.R. and Woese, C.R. (1982) Zbl. Bakt. Hyg., I. Abt. Orig. C3, 330-345.
8. Walker, R.T., Chelton, E.T.J., Kilpatrick, M.W., Rogers, M.J. and Simmons, J. (1982) Nucl. Acids Res. 10, 6363-6367.
9. Simoncsits, A. (1980) Nucl. Acids Res. 8, 4111-4124.

10. Hori, H., Osawa, S., Murao, K. and Ishikura, H. (1980) Nucl. Acids Res. 8, 5423-5426.
11. MacKay, R.M., Salgado, D., Bonen, L., Stackebrandt, E. and Doolittle, W.F. (1982) Nucl. Acids Res. 10, 2963-2970.
12. Newhouse, N., Nicoghosian, K. and Cedergren, R.J. (1981) Can. J. Biochem. 59, 921-932.
13. Kumagai, I., Digweed, M., Erdmann, V., Watanabe, K. and Oshima, T. (1981) Nucl. Acids Res. 9, 5159-5162.
14. Fang, B., DeBaere, R., Vandenberghe, A. and DeWachter, R. (1982) Nucl. Acids Res. 10, 4679-4685.
15. Hori, H. and Osawa, S. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 381-385.
16. Küntzel, H., Heidrich, M. and Piechulla, B. (1981) Nucl. Acids Res. 9, 1451-1461.
17. Fox, G.E. and Woese, C.R. (1975) Nature 256, 505-507.
18. Luehrsen, K.R. and Fox, G.E. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 2150-2154.
19. DeWachter, R., Chen, M. and Vandenberghe, A. (1982) Biochimie 64, 311-329.
20. Pieler, T., Kumagai, I. and Erdmann, V.A. (1982) Zbl. Bakt. Hyg., I. Abt. Orig. C3, 69-78.
21. Piechulla, B., Hahn, U., McLaughlin, L. and Küntzel, H. (1981) Nucl. Acids Res. 9, 1445-1450.
22. Pieler, T. and Erdmann, V.A. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 4599-4603.
23. 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.
24. Küntzel, H. and Köchel, H.G. (1981) Nature 293, 751-755.
25. Küntzel, H. (1982) Zbl. Bakt. Hyg., I. Abt. Orig. C3, 31-39.
26. Schwartz, R.M. and Dayhoff, M.O. (1978) Science 199, 359-403.