a conformational switch has any role in ribosome function. In this regard, however, it would be interesting to inspect the dissociation properties of the 5.8S fragment/28S rRNA complexes from other organisms, with diverse sequences.

Acknowledgments

We thank Dr. Harry Noller for providing us with the secondary structure model for *E. coli* 23S rRNA prior to publication and Dr. Dona Chikaraishi, Chris Harrington, David Lane, and Gary Olsen for their critical readings of the manuscript.

References

Donis-Keller, H., Maxam, A. M., & Gilbert, W. (1977) *Nucleic Acids Res. 4*, 2527–2538.

Ford, P. J., & Mathieson, T. (1978) *Eur. J. Biochem. 87*, 199–214.

Hassur, S. M., & Whitlock, H. M. (1974) *Anal. Biochem. 59*, 162–164.

Khan, M. S. N., & Maden, B. E. H. (1976) *FEBS Lett. 72*, 105–110.

Luoma, G. A., & Marshall, A. G. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 4901–4905.

Maxam, A. M., & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 560–564.

Nazar, R. N., Sitz, T. O., & Busch, H. (1975) *J. Biol. Chem. 250*, 8591–8597.

Noller, H. F., Kop, J., Wheaton, V., Brosius, J., Guteu, R. R., Kopylov, A. M., Dohme, F., Herr, W., Stahl, D. A., Gupta, R., & Woese, C. R. (1981) *Nucleic Acids Res. 9*, 6167–6198.

Pace, N. R., Walker T. A., & Schroeder, E. (1977) *Biochemistry 16*, 5321–5328.

Salser, W. (1977) *Cold Spring Harbor Symp. Quant. Biol. 42*, 985–1002.

Sanger, F., Brownlee, G. G., & Barrell, B. G. (1965) *J. Mol. Biol. 13*, 373–398.

Sitz, T. O., Szu-Chuan, K., & Nazar, R. N. (1978) *Biochemistry 17*, 5811–5815.

Stahl, D. A., Walker, T. A., Meyhack, B., & Pace, N. R. (1979) *Cell (Cambridge, Mass.) 18*, 1133–1143.

Stern, R., & Littauer, U. Z. (1971) in *Procedures in Nucleic Acid Research* (Cantoni, G. L., & Davies, D. R., Eds.) Vol. 2, pp 567–587, Harper and Row, New York.

Walker, T. A., & Pace, N. R. (1977) *Nucleic Acids Res. 4*, 595–601.

Walker, T. A., Johnson, K. D., Olsen, G. J., Peters, M. A., & Pace, N. R. (1982) *Biochemistry* (preceding paper in this issue).

Nucleotide Sequence of *Dictyostelium discoideum* 5.8S Ribosomal Ribonucleic Acid: Evolutionary and Secondary Structural Implications†

Gary J. Olsen and Mitchell L. Sogin*

ABSTRACT: We have determined the nucleotide sequence of the *Dictyostelium discoideum* 5.8S ribosomal RNA (rRNA). The sequence has relatively low homology with other 5.8S rRNAs and is further distinguished by its lack of modified nucleotides. The homology data indicate that *D. discoideum* diverged from the mainstream of eukaryotic descent at the earliest branch yet characterized by molecular phylogeny. Taking advantage of the considerable divergence of the *D. discoideum* 5.8S rRNA sequence from those of other eukaryotes, we have concluded that there is very little phylogenetically conserved, intramolecular secondary structure. This conclusion is discussed in the light of the variety of evidence for a highly ordered structure of 5.8S rRNA in vitro. We also offer comparative evidence in support of a specific model for the base pairing between the 3′ end of 5.8S rRNA and the 5′ end of eukaryotic 28S rRNA.

The large subunit of most eukaryotic ribosomes contains the 28S, 5S, and 5.8S rRNAs. When isolated under nondenaturing conditions, the 5.8S rRNA is found associated with the 28S rRNA (Pene et al., 1968). Chemical modification and enzymatic susceptibility measurements have been used with limited success to probe the secondary structure of the 5.8S rRNA (Walker et al., 1982; Nazar et al., 1975; Pace et al., 1977; Kelly & Maden, 1980; Wildeman & Nazar, 1981) and its association with the 28S rRNA (Walker et al., 1982; Pace et al., 1977). A complementary method of secondary structure analysis involves comparison of the potential base pairings of homologous RNAs from a variety of organisms. This approach has been successful in establishing universal secondary structures for tRNAs (e.g., Rich & RajBhandary, 1976), 5S rRNAs (Fox & Woese, 1975), and eubacterial 16S (Woese et al., 1980) and 23S rRNAs (Glotz et al., 1981; Noller et al., 1981). However, a lack of sequence diversity has limited the effectiveness of a similar comparative analysis of 5.8S rRNAs.

Molecular phylogeny (Zuckerkandl & Pauling, 1965) indicates that the cellular slime mold *Dictyostelium discoideum* is the deepest known branch within the eukaryotic kingdom. This is based upon the sequences of the 5S rRNA (Hori et

† From the Department of Molecular and Cellular Biology, National Jewish Hospital and Research Center, Denver, Colorado 80206, and the Department of Biochemistry, Biophysics and Genetics, University of Colorado Health Sciences Center, Denver, Colorado 80262. *Received November 4, 1981.* This investigation was supported by National Institutes of Health Research Grant GM23464 to M.L.S.

* Address correspondence to this author at the Department of Molecular and Cellular Biology, National Jewish Hospital and Research Center.

0006-2960/82/0421-2335$01.25/0

al., 1981) and the 3′-terminal region of the 17S rRNA,[1] as well as an analysis of the complete set of oligonucleotides generated by RNase T_1 digestion of the 17S rRNA.[2] These data are corroborated by comparison of the primary structures of putative tryptophan tRNA (Peffley & Sogin, 1981) and valine tRNA[3] genes from *D. discoideum* with homologous tRNAs from other eukaryotes. The observed variation of RNA sequences from *D. discoideum* suggested that knowledge of its 5.8S rRNA primary structure would add significantly to the diversity of known sequences.

We have determined the nucleotide sequence of *D. discoideum* 5.8S rRNA by a combination of enzymatic partial digestion techniques and RNase T_1 oligonucleotide cataloging. As with other RNAs which are integral components of the cytoplasmic translation apparatus of eukaryotes, the most divergent of the characterized 5.8S rRNA sequences is that of *D. discoideum*. Our examination of possible intramolecular secondary structures indicates that only two of the previously proposed duplex regions are universal. We have also utilized the available 28S rRNA sequence information to propose a specific model for the base pairing of the 3′ end of the 5.8S rRNA with the 5′ end of the 28S rRNA. A preliminary determination of the 5′-terminal sequence of the *D. discoideum* 26S rRNA supports the proposed pairing.

Experimental Procedures

Materials. The sources of most materials are given elsewhere (Walker et al., 1982). Cellulose acetate strips were from Schleicher & Schuell. Cellulose thin-layer plates were from Macherey-Nagel. RNase PhyM was from a generous supply provided by Dr. H. Donis-Keller to Dr. N. R. Pace.

Growth and Labeling of Cells. Unlabeled RNA was isolated from whole cells of *D. discoideum* strain Ax3 grown in HL5 + 1% glucose (Cocucii & Sussman, 1970). RNA labeled in vivo with ^{32}P was isolated from cells grown in phosphate-depleted MES-HL5 + 1% glucose containing 0.1 mCi/mL $H_3{}^{32}PO_4$ (ICN). For preparation of the modified MES-HL5, solid $MgCl_2$ was added to a solution of 5% (w/v) yeast extract and 10% (w/v) proteose peptone to a concentration of 0.1 M. Following titration with concentrated NH_4OH to pH 9.0, the solution was allowed to sit at 4 °C overnight. The precipitate was removed by centrifugation at 4000*g* for 5 min, and the supernatant was titrated back to pH 7.0 with concentrated HCl. The resulting proteose peptone, yeast extract solution was used as a 10× stock for preparing the phosphate-depleted MES-HL5 (Jacobson, 1976). Use of the modified media yielded *D. discoideum* RNA with specific radioactivities greater that 10^6 cpm/μg.

Extraction and Purification of RNA. RNA was isolated by extraction of $NaDodSO_4$[4] lysed cells with phenol/chloroform/isoamyl alcohol (50:48:2 v/v/v) as previously described (Jacobson, 1976). The 5.8S/26S rRNA complex was purified from total cell RNA by two rounds of sedimentation through 5–30% (w/v) linear sucrose gradients in STE buffer (0.1 M NaCl, 50 mM Tris-HCl,[4] and 1 mM EDTA,[5] pH 7.2) containing 0.1% $NaDodSO_4$. Following heat denaturation of the complex in 7 M urea and 10 mM EDTA, pH 7.2, the 5.8S rRNA was electrophoretically purified on an 8% polyacrylamide gel containing 7 M urea, 90 mM Tris–borate, and 1.25 mM EDTA, pH 8.3. The RNA was located by "UV shadowing" (Hassur & Whitlock, 1974) and eluted from the gel according to Donis-Keller et al. (1977).

Sequence Analysis by Partial Enzymatic Hydrolysis. Sequence determinations were performed on both 5′- (Donis-Keller et al., 1977) and 3′- (England & Uhlenbeck, 1978) end-labeled 5.8S rRNA molecules by partial degradation with alkali, RNase T_1, RNase U_2, and PhyM activity as previously described (Donis-Keller et al., 1977; Donis-Keller, 1980; Krupp & Gross, 1979). Terminal nucleotides were determined by digestion of the appropriate end-labeled RNA with nuclease P_1 (5′ end) or RNase T_2 (3′ end) and identification of labeled nucleotides by pH 3.5 paper electrophoresis.

RNase T_1 Oligonucleotide Cataloging. The sequence was confirmed by characterization of the oligonucleotides generated by total RNase T_1 digestion of in vivo labeled 5.8S rRNA. The products resulting from digestion of 1.0 μg of RNA with 0.1 μg of RNase T_1 in 10 μL of H_2O were separated by two-dimensional fingerprinting (Sanger et al., 1965). The base composition of each oligomer was determined by RNase T_2 digestion followed by paper electrophoresis at pH 3.5. The RNase A secondary digestion products of most oligomers were identified by electrophoresis on diethylaminoethyl paper (Whatman DE81) in 0.4 M pyridinium acetate, pH 3.5, buffer (Pace et al., 1977).

Modified Nucleotide Analysis. The possible presence of pseudouridine was assayed by complete digestion of in vivo labeled 5.8S rRNA with RNases T_2 and A and analysis of the products by two-dimensional chromatography on a cellulose thin-layer plate (Nishimura, 1972).

Results

Sequencing by Partial Enzymatic Digestion. The partial hydrolysis of in vitro labeled *D. discoideum* 5.8S rRNA with alkali, RNase T_1, or RNase U_2 was performed as described by Donis-Keller et al. (1977) except that the enzymatic digestions were performed at pH 3.5 to improve the uniformity of RNase U_2 cutting (Krupp & Gross, 1979). The specificity of RNase T_1 is not altered by the reduced pH.[5] PhyM activity was used to discriminate between the pyrimidine nucleotides (Donis-Keller, 1980). The sequencing gels allowed unambiguous analysis of over 90 nucleotides from each end. An additional 40 nucleotides were checked for consistency with the presumed sequence. The primary structure deduced from molecules labeled at opposite ends was in complete agreement within the region of overlap. There was no observable terminal heterogeneity as evidenced by a lack of doublet bands in the ladder gels. The identities of the 5′ and 3′ termini were found to be U and C, respectively.

RNase T_1 Oligonucleotide Analysis. Trace quantities of a nuclease activity nicked the 5′-end-labeled RNA preparations in the sequence 5′-Y-A-3′,[4] rendering the identity of several pyrimidine residues in the 5′ half of the molecule ambiguous. Characterization of RNase T_1 oligonucleotides of uniformly labeled 5.8S rRNA permitted the identification of these pyrimidines and also provided confirmation of the sequence deduced from the partial enzymatic hydrolysis data. [The data from the oligonucleotide analyses are included in the supplementary material to this paper (see paragraph at end of paper regarding supplementary material).] The 5′-terminal residue, uridine 3′,5′-bisphosphate, was found in a single T_1 oligomer, pU-U-A-A-Gp. The 3′-terminal oligomer, U-C, was not retained during the two-dimensional separation of the oligonucleotides. The nucleotide sequence deduced from the partial

[1] Y. D. Stahl, R. M. McCarroll, M. L. Sogin, and G. J. Olsen, unpublished data.

[2] C. R. Woese and M. L. Sogin, unpublished data.

[3] M. L. Sogin, D. M. Peffley, and G. J. Olsen, unpublished data.

[4] Abbreviations: $NaDodSO_4$, sodium dodecyl sulfate; Tris, tris(hydroxymethyl)aminomethane; EDTA, ethylenediaminetetraacetic acid; Y, pyrimidine nucleoside.

[5] T. A. Walker, personal communication.

```
                                    10        20        30        40
Dictyostelium discoideum      UUAAGCAUAAACGGUGAAUACCUCGAC-UCCCAAAUUGAUGAAGAC
Xenopus laevis                 CGACUCUUAGCGGUGGAUCACUCGGC-UCGUGCGUCGAUGAAGAA
Sciara coprophilia              AACCCUAAGCGGGGGAUCACUUGGU-UUGUGGGUCGAUGAAGAA
Drosophila melanogaster         AACUCUAAGCGGUGGAUCACUCGGC-UCAUGGGUCGAUGAAGAA
Triticum vulgare          CACACGACUCUCGGCAACGGAUAUCUCGGC-UCUCGCAUCGAUGAAGAA
Acanthamoeba castellanii        AACUCCUAACAACGGAUAUCUUGGU-UCUCGCGAGGAUGAAGAA
Neurospora crassa               AAACUUUCAACAACGGAUCUCUUGGU-UCUGGCAUCGAUGAAGAA
Saccharomyces cerevisiae        AAACUUUCAACAACGGAUCUCUUGGU-UCUCGCAUCGAUGAAGAA
Escherichia coli           ...ACUAAGCGUACACGGUGGAUGCCCUGGCAGUCAGAGGCGAUGAAGGA

                          50        60        70        80        90        100
D. di.  CGUAGCAAACUGCGAUAAU-UCACUUGAAUUGCAGCCUACUGGGA-U-AGUUGAAAUGUUGA
X. la.  CGCAGCUAGCUGCGAGAAU-UAGUGUGAAUUGCAGGACACAUUGA-U-CAUCGACACUUCGA
S. co.  CGCAGCAAACUGCGUGUUG-ACAUGUGAACUGCAGGACACA-UGA-A-CAUUGACAUUUUGA
D. me.  CGCAGCAAACUGUGCGUCA-UCGUGUGAACUGCAGGACACA-UGA-A-CAUCGACAUUUUGA
T. vu.  CGUAGCGAAAUGCGAUACC-UGGUGUGAAUUGCAGAAUCCCGCGA-ACCAUCGAGUCUUUGA
A. ca.  CGCAGCGAAAUGCGAUACG-UAGUGUGAAUCGCAGGGAUCAGUGA-AUCAUCGAAUCUUUGA
N. cr.  CGCAGCGAAAUGCGAUAGG-UAAUGUGAAUUGCAGAAUUCAGUGA-AUCAUCGAAUCUUUGA
S. ce.  CGCAGCGAAAUGCGAUACG-UAAUGUGAAUUGCAGAAUUCCGUGA-AUCAUCGAAUCUUUGA
E. co.  CGUGCUAAUCUGCGAUAAGCGUCGGUAAGGUGAUAUGAACCGUUAUAACCGGCGAUUUCCGA

                          110       120       130       140       150       160
D. di.  ACGCACAUGAUGACAUCGG-UCCUUU-CGGAUUAGGUGUUAUACUUGGGUGAGAGUGGUC
X. la.  ACGCACCU--UGCGGCCCC-GGGUUCCUCCCGG-GGC--CACGCCUGUCUGAGGGUCGCUCC
S. co.  ACGCAUAU--UGCGGUCCA-UACUG   UGUAUG-GAC--CACACAUGGUUGAGGGUCGUU
D. me.  ACGCAUAU--CGCAGUCCA-UGCUG   UGCUUG-GAC--UACAUAUGGUUGAGGGUUGUA
T. vu.  ACGCAAGU--UGCGCCCGA--GCCAC-UC-GCC-GAGGGCACGCCUGCCUGGGCGUCACGC
A. ca.  ACGCAAGU--UGCGCUCUCGUGGUUUAACCCCCCGGGAGCACGUUCGCUUGAGUGCCGCUU
N. cr.  ACGCACAU--UGCGCUCGC-CAGUAU-UCUGGC-GAG--CAUGCCUGUUCGAGCGUCAUUU
S. ce.  ACGCACAU--UGCGCCCCU-UGGUAU-UCCAGG-GGG--CAUGCCUGUUUGAGCGUCAUUU
E. co.  AUGGGGAA-ACCCAGUGUG----UUU-CG----ACACACUAUCAUUAACUG...
```

FIGURE 1: rRNA (5.8S) sequence from *D. discoideum* and its alignment with other 5.8S sequences. The sequence of the *D. discoideum* 5.8S rRNA (see results) is shown aligned with those from a representative collection of other eukaryotes as well as the homologous region of the *E. coli* 23S rRNA (Nazar, 1980; Jacq, 1981). Numbering is for the *D. discoideum* sequence. Nucleotides 1–116 and 143–162 are aligned by primary structural homology. Positions 117–142 were aligned principally according to secondary structure, i.e., the "G-C rich hairpin". There are no modified residues in the *D. discoideum* 5.8S rRNA, and for convenience those in the other sequences are not shown, although all sequences except the *E. coli* 23S rRNA segment contain at least one modified nucleotide. Sequences are according to Erdmann (1981) except the following: *N. crassa* (Selker & Yanofsky, 1978); *X. laevis* is amended by Boseley et al. (1978); *T. vulgare* (MacKay et al., 1980); *A. castellanii* (MacKay & Doolittle, 1981); *E. coli* (Brosius et al., 1980). Contrary to the implicit assumption of Nazar (1980) that the homology with 23S rRNA extends to the end of the 5.8S rRNA, the secondary structure of 23S rRNA (see helix V in Figure 4) suggests that the functional homology ends about 10 nucleotides earlier. Walker (1981) points out that the homology of the eubacterial 23S rRNA with the eukaryotic 28S rRNA begins near the end of the 5.8S homologous region, but chooses to overlap both eukaryotic rRNAs with about 20 nucleotides of the 23S, rather than coming to our conclusion that the termini of the eukaryotic RNAs have no eubacterial counterpart.

enzymatic hydrolysis data and the RNase T_1 oligonucleotide catalog is presented in Figure 1 along with a representative collection of other 5.8S rRNA sequences.

Absence of Modified Nucleotides. Two types of posttranscriptionally modified nucleotides have been found in 5.8S rRNAs, 2′-*O*-methyl nucleotides and pseudouridylic acid residues [see Erdmann (1981)]. The absence of gaps in the alkaline hydrolysis lane of the sequencing gels argues against the presence of 2′-*O*-methyl groups in the *D. discoideum* 5.8S rRNA. This conclusion is supported by the absence of dinucleotides in the RNase T_2 digests of the RNase T_1 oligonucleotides.

The presence of pseudouridine was assayed by two-dimensional chromatographic resolution (Nishimura, 1972) of the nucleotides released by digestion of uniformly labeled 5.8S rRNA to completion with RNase T_2 and RNase A. The autoradiogram presented in Figure 2 reveals only the expected nucleoside 3′-phosphates and the terminal uridine 3′,5′-bisphosphate. Liquid scintillation counting of the region of the chromatogram in which pseudouridine 3′-phosphate would be expected (Nishimura, 1972) places an upper bound of 0.2 mol of pseudouridine per mol of 5.8S rRNA. This limit probably represents an overestimate because of streaking in the second dimension ahead of the Ap and Gp spots, possibly due to partial depurination in the acidic chromatography solvent. This analysis also failed to reveal the presence of any dinucleotides resulting from 2′-*O*-methylated residues, even at submolar levels.

Discussion

Phylogenetic Position of Dictyostelium discoideum. The homologies of a representative set of 5.8S rRNA sequences, including that of *D. discoideum*, are presented in Table I. As in the case of the rRNA and tRNA sequence comparisons mentioned above, it can be seen that *D. discoideum* 5.8S rRNA consistently displays the lowest homologies with other sequences. The *D. discoideum* sequence is also unique in its lack of posttranscriptionally modified nucleotides. Two alternative explanations of the low homology of the *D. discoideum* 5.8S rRNA sequence to those of other eukaryotes are (1) *D. discoideum* has an unusually high mutational acceptance rate or (2) it diverged from the mainstream of eukaryotic descent on the deepest known branch. Hori et al. (1980) have implicitly chosen the former explanation by interpreting the 5S rRNA homology data in terms of *D. discoideum* branching off of the animal line of descent after the separation of plants

Table I: Homologies between 5.8S rRNA Sequences[a]

organism	fractional homology with							
	D. di.	*X. la.*	*S. co.*	*D. me.*	*T. vu.*	*A. ca.*	*N. cr.*	*S. ce.*
Dictyostelium discoideum		0.56	0.61	0.60	0.54	0.52	0.57	0.59
Xenopus laevis	0.64		0.74	0.75	0.69	0.68	0.69	0.72
Sciara coprophilia	0.63	0.78		0.86	0.60	0.62	0.68	0.68
Drosophila melanogaster	0.63	0.83	0.88		0.62	0.62	0.62	0.65
Triticum vulgare	0.62	0.73	0.65	0.70		0.72	0.77	0.78
Acanthamoeba castellanii	0.59	0.72	0.68	0.69	0.81		0.76	0.78
Neurospora crassa	0.64	0.74	0.73	0.71	0.83	0.86		0.92
Saccharomyces cerevisiae	0.65	0.73	0.72	0.71	0.85	0.87	0.97	
Escherichia coli	0.49	0.46	0.50	0.48	0.43	0.46	0.47	0.47

[a] The upper right half of the table gives the fraction of homologous positions containing identical nucleotides for all pairs of the 5.8S rRNA sequences in Figure 1. The sequence to sequence variation of terminal length has been eliminated by confining the analysis to the region of each sequence corresponding to nucleotides 3–162 of the *D. discoideum* sequence. The internal length variation has been treated by scoring locations with a deletion in one sequence opposite a nucleotide in the second sequence as half a mismatch. When both sequences in the pair have deletions at the same location, the position was not counted. The lower left half of the table contains the fractional homologies between pairs of sequences when the analysis is limited to the region corresponding to nucleotides 3–117 of the *D. discoideum* sequence, i.e., up the beginning of the "G-C rich hairpin", arm IV in Figure 4. Elimination of the hairpin is advantageous when constructing phylogenies since the great variation of sequence length and primary structure makes alignments between sequences from distant organisms quite uncertain in this region. In addition, exclusion of the 3′ tail of the 5.8S sequences allows comparison with the *E. coli* 23S rRNA sequence which lacks a homologue to the last ca. 10 nucleotides of 5.8S rRNA (see Figures 1 and 4).

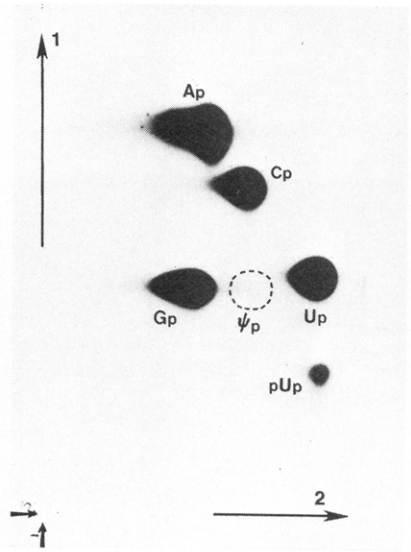

FIGURE 2: Modified nucleotide analysis. The nucleotides released by complete digestion of in vivo labeled *D. discoideum* 5.8S rRNA with RNases T_2 and A were resolved by the two-dimensional chromatographic system of Nishimura (1972). The radioactive products are shown in this autoradiogram. Ap, Cp, Gp, and Up were identified by their mobilities (Nishimura, 1972). The identity of pUp was inferred from its mobility relative to Up, its absence in nuclease P_1 digests of the same RNA (data not shown), and its molar yield (1.6 mol of phosphate/mol of 5.8S rRNA). The region of the chromatogram corresponding to pseudouridylic acid, ψp, was also analyzed for radioactivity (see Results).

and animals. For two reasons we prefer the alternate interpretation of the homology data. First, *D. discoideum* does not show a convincing affiliation with any specific organism or group of organisms. The tendency of its 5S sequence to be more similar to those of animals than to those of plants (Hori et al., 1980; Küntzel et al., 1981) is only about one standard deviation from equidistant (Hori & Osawa, 1979) and is not reflected in the 5.8S homology data. Second, the sequence data are not consistent with *D. discoideum* 5.8S rRNA having an above average mutation rate. Specifically, since the 5′ end of *Escherichia coli* 23S rRNA is homologous to the 5.8S rRNA of eukaryotes (Nazar, 1980; Jacq, 1981), the average rate of mutation within a eukaryotic line of descent will be reflected in the organism's divergence from *E. coli*, the higher the mutation rate, the greater the divergence. Table I also gives the homologies of approximately the first 110 nucleotides (a region of fairly unambiguous sequence alignment due to its strong sequence conservation) of the 5.8S rRNAs as well as their homologies with the 23S rRNA of *E. coli*. It can be seen that the *D. discoideum* sequence shows no evidence of an above average mutation rate, i.e., divergence from that of *E. coli*. A similar comparison of the 3′-terminal regions of the 18S rRNA sequences of yeast (Rubtsov et al., 1980), *Bombyx mori* (Samols et al., 1979), and *D. discoideum*[1] with the 16S rRNAs of *E. coli* (Brosius et al., 1978) and *Zea mays* chloroplast (Schwarz & Kössel, 1980) leads to the same conclusion. It follows that *D. discoideum* lies on the earliest characterized branch within the eukaryotic phylogeny.

A phylogenetic tree of the organisms in Table I, constructed by using the distance metric of Hori & Osawa (1979) and minimized with respect to the standard deviation of the percent difference between the data and the tree distances (Fitch & Margoliash, 1967), is presented in Figure 3. Beyond the position of *D. discoideum*, the grouping of the yeast *Saccharomyces cerevisiae* and the mold *Neurospora crassa* with the plant *Triticum vulgare* (wheat) is of special note. Phylogenies based upon 5S rRNA sequences and 18S rRNA RNase T_1 oligonucleotides tend to place the fungi either on the animal branch or as a branch which precedes the plant/animal division (Woese & Fox, 1977; Hori & Osawa, 1979; Küntzel et al., 1981). The statistical uncertainty in the 5S rRNA homologies is great enough to allow any of the above possibilites (Hori & Osawa, 1979). Final resolution of the branching order must await additional data; of particular value would be a plant 18S rRNA sequence.

Secondary Structure of 5.8S rRNA. We have undertaken a comparative study of the possible secondary structures of the various 5.8S rRNA sequences available (Erdmann, 1981; Boseley et al., 1978; Selker & Yanofsky, 1979; MacKay et al., 1980; Jordan et al., 1980; Darlix & Rochaix, 1981; MacKay & Doolittle, 1981; Wildeman & Nazar, 1981), taking advantage of the considerable divergence of *D. discoideum* 5.8S sequence from those of other eukaryotes. Only those helices which are functionally important are expected to be conserved in all sequences, these being primarily, if not exclusively, those which are present in the 5.8S/28S complex. This differs from the 5.8S rRNA structure mapping work, most of which has emphasized analysis of isolated 5.8S rRNA. There is no a priori reason to expect that there is a universal

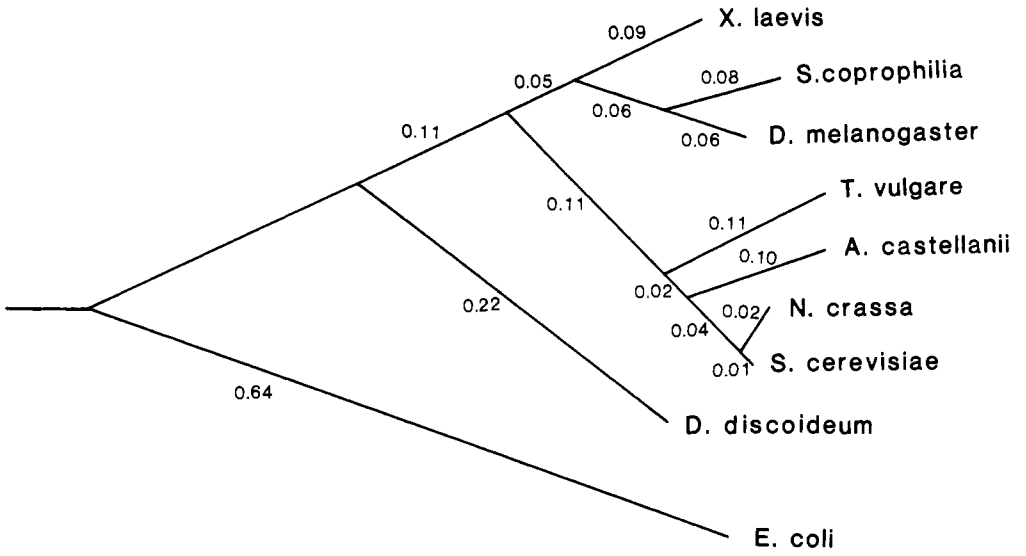

FIGURE 3: Phylogeny of 5.8S rRNA sequences. The homology data in Table I for the first ca. 110 nucleotides of the 5.8S rRNA sequences and the corresponding region of the *E. coli* 23S rRNA sequence were used to construct this phylogenetic tree. The homologies were converted into distances (average number of base changes per sequence position) according to Hori & Osawa (1979). The segment lengths in the tree were chosen so as to minimize the standard deviation of the fractional differences between the tree path lengths and the measured distances (Fitch & Margoliash, 1967). The evolutionary distance between nodes of the tree is given alongside the segment connecting them and is represented in the horizontal component of their separation in the figure. Alternative weightings of the path-length errors do not significantly alter the internodal distances or the tree geometry.

structure for the *free* 5.8S molecule or that such a structure is biologically meaningful.

Figure 4 presents the 5.8S rRNA sequences of *D. discoideum*, *S. cerevisiae*, *T. vulgare*, *Xenopus laevis*, and *Drosophila melanogaster* folded into homologous (with the exception of one arm) secondary structures. Also shown is a specific association of the 3′ end of three of the 5.8S rRNAs with the 5′ end of their respective 28S rRNAs (see below). Additionally, the corresponding region of the *E. coli* 23S rRNA is shown in a partially homologous folding. The individual elements of the structures are discussed below.

Arm II in Figure 4 [helix c and loop III of Nazar et al. (1975)] can be drawn in all eukaryotic 5.8S rRNA sequences and in the *E. coli* 23S rRNA. The structure mapping data are consistent with the existence of arm II in free 5.8S (Nazar et al., 1975; Kelly & Maden, 1980; Wildeman & Nazar, 1981) and in the 5.8S/28S complex (Walker et al., 1982). It is, however, composed of nearly invariant primary structure, and where sequence variation is observed, it tends to disrupt the terminal base pair of the helix [if one includes prokaryotic sequences, then the *Z. mays* and *Euglena gracillis* chloroplast 23S rRNAs (Edwards & Kössel, 1981; Orozco et al., 1980) can only pair three bases in this stem]. This suggests that it is primary structure, not secondary structure, which is of paramount importance in this region.

Arm III [helix d and loop IV of Nazar et al. (1975)] is not a phylogenetically conserved structure. Although nuclease and chemical modification susceptibility data support its existence in the free mammalian 5.8S rRNA (Walker et al., 1982; Nazar et al., 1975; Kelly & Maden, 1980) and in the 5.8S/28S rRNA complex (Walker et al., 1982), two vastly different pairings are required to accommodate the metazoan sequences (represented here by *X. laevis*) vs. the plant and fungal sequences (as represented by yeast). Further, the *D. discoideum* sequence does not fit either pairing. This region of the eubacterial 23S rRNA is involved in an entirely distinct pairing (Glotz et al., 1981; Noller et al., 1981).

The stalk shown at the base of arms II and III [helix b of Nazar et al. (1975)] is similar to arm II in that it pairs regions of relatively conserved sequence. Out of at least six known base transitions and five base transversions within the region of complementarity defined by the *X. laevis* sequence, there are no sequences in which the presumptive pairing partner also varies so as to maintain Watson–Crick complementarity. Because of this uncompensated variation, the apparent length of this stalk varies from five to eight base pairs. The *D. discoideum* sequence can form five base pairs, three of which are G-U pairs, a very unsatisfying helix. Although a five-base-pair helix can also be drawn with the *Acanthamoeba castellanii* sequence (MacKay & Doolittle, 1981), only three of these pairs overlap those of the *D. discoideum* sequence. This helix cannot be formed in the eubacterial 23S rRNA sequences (Brosius et al., 1980; Orozco et al., 1980; Edwards & Kössel, 1981). Finally, a portion of the apparent complementary region is inconsistent with the structure mapping data for isolated human (Kelly & Maden, 1980), mouse (Walker et al., 1982), rat (Nazar et al., 1975), and *Thermomyces lanuginosus* 5.8S rRNAs (Wildeman & Nazar, 1981). These data suggest that a Watson–Crick duplex is not present in vivo.

Arm IV, the "G-C rich hairpin", is the only phylogenetically proven, intramolecular helix in 5.8S rRNA. Although the length and fidelity of pairing vary considerably, there are numerous examples of pairwise base changes in the two halves of the duplex so as to maintain complementarity. Also, the positions of the termini of the hairpin within the 5.8S primary structure are nearly universal, *D. discoideum* being unusual by virtue of extending the apparent pairing into regions which are unpaired in all other sequences. This might be related to an unusual pairing between the 5.8S and 26S rRNAs of *D. discoideum* (see below). The great variability in duplex length and pairing fidelity suggests that the details of the structure are unimportant relative to the existence and position of the arm within the primary structure. This conclusion is reinforced by the observation that the eubacterial 23S rRNA contains a homologous hairpin (Glotz et al., 1981; Jacq, 1981; Noller et al., 1981) which is preceded by a C residue and followed by an A residue as are all eukaryotic examples (except *D. discoideum*). The removal of the apex of this hairpin during the maturation of insect 5.8S rRNA (Pavlakis et al., 1979) is a dramatic example of the tolerable variation in this

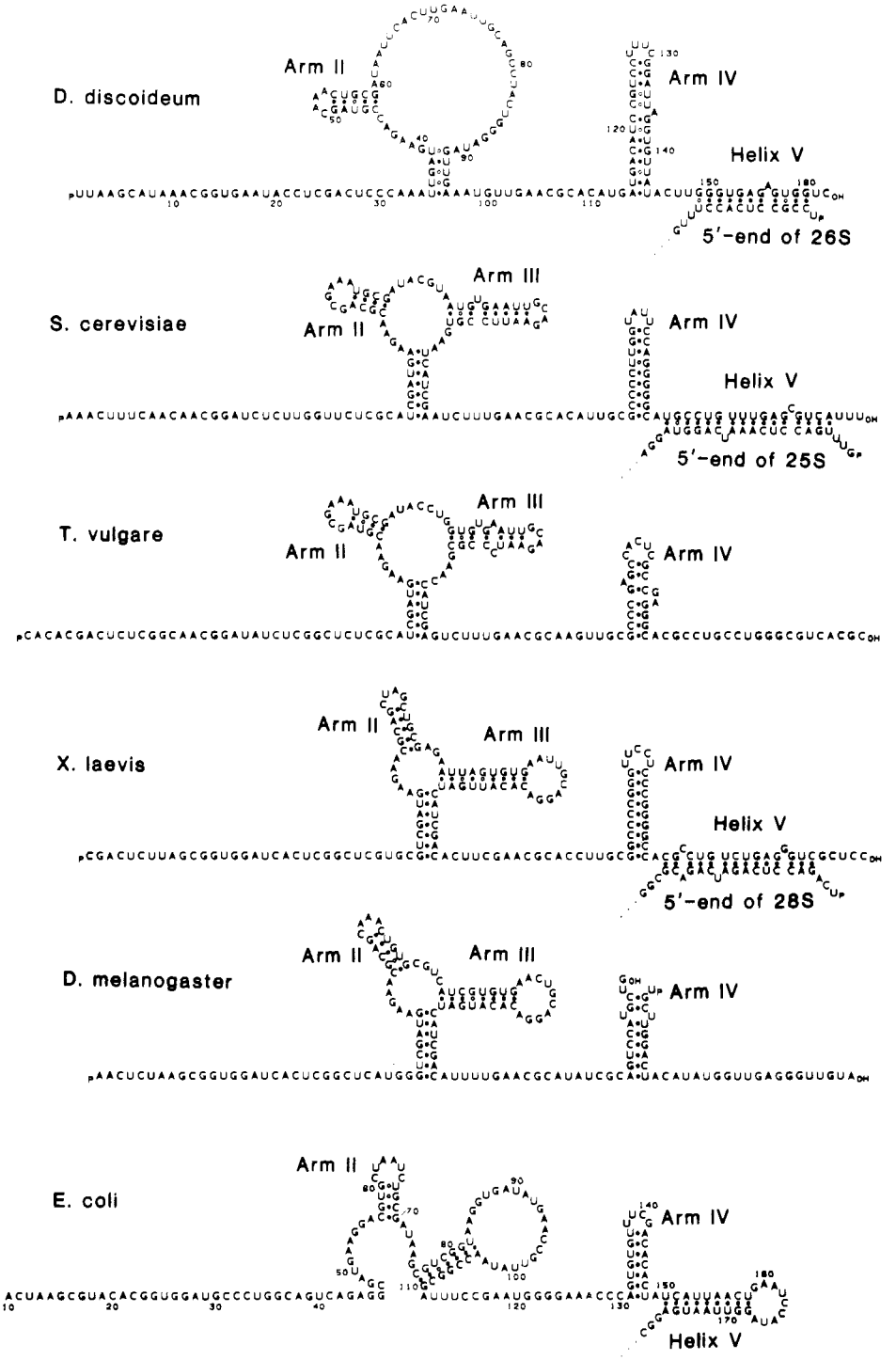

FIGURE 4: Secondary structures of 5.8S rRNAs and the 5′ end of the *E. coli* 23S rRNA. The sequences of the 5.8S rRNAs from *D. discoideum*, the yeast *S. cerevisiae* [see Erdmann (1981)], *T. vulgare* (MacKay et al., 1980), *X. laevis* (Boseley et al., 1978), and *D. melanogaster* (Pavlakis et al., 1979) are shown folded into partially homologous (see text) secondary structures. Structural features are identified according to Walker et al. (1982). The 5′-terminal sequences of three of the corresponding 28S rRNAs[7] (Hall & Maden, 1980; Bayev et al., 1981) are shown in a specific pairing with the 3′ ends of their respective 5.8S rRNAs. Also shown is the homologous region of the *E. coli* 23S rRNA sequence (Brosius et al., 1980) in a folding which has hairpins homologous to three of the eukaryotic pairings (Noller et al., 1981).

structure. Pace et al. (1977) have suggested that arm IV provides a stabilizing influence, by coaxial stacking, on the pairing of the 3′ end of 5.8S rRNA with the 28S rRNA (see below).

The metazoan 5.8S rRNA sequences tend to have a very G-C rich arm IV helix with perfect Watson–Crick complementarity. In contrast, the *D. discoideum*, plant, and fungal sequences contain considerable numbers of A-U and G-U pairs, as well as internal loops or bulges. In spite of their imperfections, all arm IV pairings are predicted to be stable under physiological conditions.

Walker et al. (1982) have noted that a G residue in the 3′ half of the arm IV helix and, to a lesser extent, a C residue in the 5′ half became accessible to nucleases upon association of mouse 5.8S and 28S rRNAs. It might be significant that the plant and fungal arm IVs all have G-U base pairs, bulges, or internal loops in the general region of the labilized G and C residues, suggesting the existence of a structural landmark. An observation which might relate to the perturbation of the arm IV helix upon 28S rRNA binding is that there are two

energetically distinct, mutually exclusive states in which the 3′ half of mouse 5.8S rRNA can bind to 28S rRNA (Peters et al., 1982). Since it is not known whether the disruption of the arm IV helix is present in all bound 5.8S rRNA molecules or only in a subpopulation, it creates the question of whether the two bound states are distinguished by the presence or absence of this helical perturbation.

A recurring theme in models of 5.8S secondary structure is a pairing between the 5′ and 3′ ends [see, for example, Nazar et al. (1975)]. We have been unable to find any phylogenetically conserved pairing. There are multiple counterexamples to every duplex of four consecutive base pairs. Further, no pair of positions undergoes coordinated base changes so as to maintain Watson–Crick or wobble pairs in all sequences. Walker et al. (1982) have pointed out that the only two processing events which we are certain the 32S precursor rRNA must undergo are the maturation of the 5.8S rRNA 3′ terminus and the 28S rRNA 5′ terminus, i.e., removal of internal transcribed spacer 2. This suggests that one should seek a mechanism to bring these two termini into proximity (see below), not the 5.8S rRNA 5′ and 3′ ends.

The phylogenetic analysis of secondary structure presented above creates somewhat of a dilemma as it portrays 5.8S rRNA as devoid of secondary structure with one, or at most two, exceptions, arms IV and II. This is in striking contrast to tRNA and 5S rRNA, which are rich in conserved secondary structure (Rich & RajBhandary, 1976; Fox & Woese, 1975; Studnicka et al., 1981). Because 5.8S rRNA is unusually rich in conserved primary structure, it is possible that it is qualitatively different from these other RNAs. However, it is still necessary to reconcile the structure with the nuclease and chemical modification susceptibility data (Walker et al., 1982; Nazar et al., 1975; Kelly & Maden, 1980; Wildeman & Nazar, 1981) and the physical chemical measurements (e.g., Luoma & Marshall, 1978) which indicate the existence of a largely ordered molecule. The resolution probably lies in a combination of several factors. First, most of the experimental data on the 5.8S rRNA structure have been obtained in the absence of 28S rRNA and the ribosomal proteins. Under these nonphysiological conditions, some aspects of 5.8S rRNA structure are apt to be changed, this being a certainty with respect to the portions of the molecule involved in intermolecular interactions. These portions of the 5.8S rRNA will appear to be unpaired in the preceding analysis of intramolecular pairing, yet to in vitro analyses of the isolated molecule they might offer some sort of orderly, albeit biologically irrelevant, structure. The 5′ and 3′ ends of 5.8S rRNA are obvious candidates for this phenomenon. A related artifact would be expected to plague studies involving high concentrations of 5.8S rRNA, e.g., those of Luoma & Marshall (1978), conditions under which 5.8S rRNA forms intermolecular complexes, i.e., dimers and higher order aggregates (Pace et al., 1977), creating an unresolved superposition of intramolecular and irrelevant intermolecular structures. A third source of disparity between our analysis and the experimental analyses might reside in the existence of a highly ordered tertiary structure. This would not appear in our secondary structure analysis, yet will often be indistinguishable from secondary structure to probes of nucleotide accessibility. This seems to be a likely explanation for the failure of the structure probes to accurately define the complementary nucleotides at the base of arms II and III, i.e., there might be little or no double helical structure, as we argued above based on the sequence comparisons, yet substantial amounts of tertiary structure might exist in the region, yielding a generally ordered structure. A final, less pleasing, possibility is that different organisms within the eukaryotic kingdom use radically different secondary structures to achieve the same function. There is no precedent for such a dramatically diverse set of secondary structures as would be required to accommodate the 5.8S rRNA sequences. The region of 5.8S rRNA most suggestive of differing structures for different groups of organisms is arm III, the metazoan version of this arm being well supported by the structure mapping data (Walker et al., 1982; Nazar et al., 1975) and having the virtue of placing a region of conserved primary structure within the single-stranded loop of the arm, yet this structure cannot accommodate the *D. discoideum*, plant, or fungal sequences.

Association of 5.8S and 28S rRNAs. It has been shown that there are at least two components to the 5.8S/28S rRNA interaction, one involving the 5′-terminal region of 5.8S rRNA and the second involving the 3′ terminus of the molecule (Walker et al., 1982; Peters et al., 1982; Pace et al., 1977; Nazar & Sitz, 1980). At this point little can be said about the interaction of the 5.8S 5′ end with 28S rRNA except to draw attention to the observation that the homologous region of the eubacterial 23S rRNA pairs with sequences approximately 440 and 520 nucleotides into the 23S sequence (Glotz et al., 1981; Jacq, 1981; Noller et al., 1981). Due to the great length difference between the eukaryotic and eubacterial molecules, the position of any eukaryotic counterpart to these pairings is difficult to predict. It has been proposed that the 5′ end of 5.8S rRNA pairs with the 3′-terminal region of the 28S rRNA (Kelly & Cox, 1981; Cox & Kelly, 1981). However, this pairing is not phylogenetically conserved, all base variation within the available sequences leading to a loss of complementarity or requiring substitution of alternative pairing partners.[6]

In Figure 4 we propose a specific model for the pairing of the 3′ end of 5.8S rRNA and the 5′ end of 28S rRNA, helix V. In addition to the published sequences for the 5′ termini of yeast 25S (Bayev et al., 1981) and *X. laevis* 28S rRNAs (Hall & Maden, 1980), we include a preliminary sequence of the 5′ end of the *D. discoideum* 26S rRNA.[7] The proposed pairing is homologous in all three cases with the exception of a segment proximal to arm IV which cannot pair in the *D. discoideum* sequences, a pecularity which might relate to the extension of arm IV pairing in *D. discoideum* beyond that observed in other organisms which are noted above. Also shown is the homologous region of the *E. coli* 23S rRNA sequence folded into a hairpin. The existence of a single unpaired A residue between arm IV and helix V in the yeast, *X. laevis*, and *E. coli* suggests a possible single base bulge and coaxial stacking of the two helices as was predicted by Pace et al. (1977). This configuration is expected to be energetically favorable (Salser, 1977) and so would contribute to the stability of the intermolecular pairing.

This proposed pairing is supported by the existence of six pairs of compensatory base changes among the three pairs of sequences presented. Beyond the base changes which shorten helix V in *D. discoideum* there are three other uncompensated base changes: one creates a wobble pair, one changes the end point of the *X. laevis* pairing, eliminating one base pair, and one which opposes an A residue with a C, creating an internal loop in the *X. laevis* pairing. Since the internal loop is surrounded by G-C base pairs, it would not be expected to disrupt

[6] Unpublished observation based on the sequence data of Kelly & Cox (1981), Mandal & Dawid (1981), Bayev et al. (1981), and Sollner-Webb & Reeder (1979).

[7] G. J. Olsen and M. L. Sogin, unpublished data.

the stacking (Salser, 1977). Overall, the organism to organism variations of the pairing are minor relative to the large number of compensated changes (contrast helix V with arm III, where there are no compensated changes in spite of much more sequence data). An appealing feature of helix V is the resulting proximity of the 5.8S 3′ end and the 28S 5′ end. This makes the removal of internal transcribed spacer 2 formally equivalent to removing the apex from a hairpin. Stahl (1978) has demonstrated that the apex of the homologous hairpin is indeed accessible to RNase A in the case of *E. coli* 50S ribosomal subunits.

Peters et al. (1982) have measured the equilibrium binding constant between the 3′ half of 5.8S rRNA and 28S rRNA from mouse. Their value of 1.3×10^{-7} M in 0.4 M salt at 50 °C corresponds to a binding energy, ΔG, of −10.1 kcal/mol. We estimate that helix V of *X. laevis*, the closest relative to mouse for which the sequence data are available, has a binding energy of −16.3 kcal/mol at 50 °C in 1.0 M salt (Gralla & Crothers, 1973). Under the assumption that the change to a G-U base-pair closing arm IV (Gralla & Crothers, 1973; Salser, 1977) is the only energetically significant sequence difference between the mouse and *X. laevis* pairings, we expect ΔG for formation of helix V in mouse to be −15.2 kcal/mol in 1.0 M salt at 50 °C.

The numerous assumptions used to arrive at the above binding energy estimate make serious comparisons with the measured value difficult. Of particular significance are (1) the assumption of identical 5′-terminal sequences of mouse and *X. laevis* 28S rRNAs, (2) the uncertainties in the ΔG and ΔS estimates, (3) the implicit assumption that there are no competing intramolecular structures [structures with a few kilocalories per mole of binding energy cannot be ruled out by the considerations of Peters et al. (1982)], and (4) our disregarding the observed biphasic melting curve of the mouse 5.8S 3′ end/28S complex and its implication of the existence of two distinct bound states (Peters et al., 1982). Given these reservations, we feel that the agreement between the measured and estimated binding energies is reasonable but cannot be used as support for the specific pairing proposed.

In summary, the 5.8S rRNA sequence of *D. discoideum* is the most divergent of those reported to date and is also unusual in lacking modified nucleotides. Comparison of 5.8S sequence to the homologous portion of *E. coli* 23S rRNA indicates that the mutation acceptance rate of *D. discoideum* is no higher than that of other eukaryotes; therefore its sequence divergence is a reflection of it being the sole representative of the earliest characterized branching within the 5.8S rRNA phylogeny.

Comparisons of the available 5.8S rRNA sequence data predict the existence of very little phylogenetically conserved *intramolecular* secondary structure. Only one hairpin, arm IV in Figure 4, can be phylogenetically proven. A second hairpin, arm II, is consistent with all of the sequence data but is likely to be a secondary consequence of primary structural constraints. No other universal duplexes of four or more consecutive base pairs appear to be possible in the available sequences, the other intramolecular pairings in Figure 4 being shown only for clarity of discussion.

We can find no phylogenetic evidence for a pairing between the 5′ and 3′ ends of 5.8S rRNA. Instead, we propose a specific model for the pairing between the 3′ end of 5.8S rRNA and the 5′ end of 28S rRNA, helix V in Figure 4. Among the yeast, *X. laevis*, and *D. discoideum* ribosomal RNAs, this pairing is supported by six pairs of compensatory base changes in the two halves of the complex, and it is homologous to a very stable hairpin in the eubacterial 23S secondary structure (Glotz et al., 1981; Jacq, 1981; Noller et al., 1981). The observed primary and secondary structural homologies between the 5′ end of the eubacterial 23S rRNAs and the eukaryotic 5.8S/28S rRNA complex lead us to assume that the 5′ end of 5.8S rRNA will pair with sequences several hundred nucleotides into the 28S rRNA, creating a structural domain analogous to that observed in the *E. coli* 23S rRNA (Glotz et al., 1981; Jacq, 1981; Noller et al., 1981).

Acknowledgments

We thank Drs. N. R. Pace and T. A. Walker for helpful discussions and unpublished data, Drs. R. M. MacKay and W. F. Doolittle for providing a preprint of the *Acanthamoeba castellanii* 5.8S rRNA sequence, and Drs. H. F. Noller and C. R. Woese for information on the eubacterial 23S rRNA secondary structure prior to its publication. We are also grateful to C. I. Reich, M. L. Pato, T. A. Walker, and N. R. Pace for useful comments on the manuscript.

Supplementary Material Available

Additional information on the determination of the nucleotide sequence presented in Figure 1 and a sample sequencing gel and a characterization of the RNase T_1 oligonucleotides (3 pages). Ordering information is given on any current masthead page.

References

Bayev, A. A., Georgiev, O. I., Hadjiolov, A. A., Nikolaev, N., Skryabin, K. G., & Zakharyev, V. M. (1981) *Nucleic Acids Res. 9*, 789–799.

Boseley, P. G., Tuyns, A., & Birnstiel, M. L. (1978) *Nucleic Acids Res. 5*, 1121–1137.

Brosius, J., Palmer, M. L., Kennedy, P. J., & Noller, H. F. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 4801–4805.

Brosius, J., Dull, T. J., & Noller, H. F. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 201–204.

Cocucci, S. M., & Sussman, M. (1970) *J. Cell Biol. 45*, 399–407.

Cox, R. A., & Kelly, J. M. (1981) *FEBS Lett. 130*, 1–6.

Darlix, J.-L., & Rochaix, J.-D. (1981) *Nucleic Acids Res. 9*, 1291–1299.

Donis-Keller, H. (1980) *Nucleic Acids Res. 8*, 3133–3142.

Donis-Keller, H., Maxam, A. M., & Gilbert, W. (1977) *Nucleic Acids Res. 4*, 2527–2538.

Edwards, K., & Kössel, H. (1981) *Nucleic Acids Res. 9*, 2853–2869.

England, T. E., & Uhlenbeck, O. C. (1978) *Biochemistry 17*, 2069–2076.

Erdmann, V. A. (1981) *Nucleic Acids Res. 9*, r25–r42.

Fitch, W. M., & Margoliash, E. (1967) *Science (Washington, D.C.) 155*, 279–284.

Fox, G. E., & Woese, C. R. (1975) *Nature (London) 256*, 505–507.

Glotz, C., Zwieb, C., Brimacombe, R., Edwards, K., & Kössel, H. (1981) *Nucleic Acids Res. 9*, 3287–3306.

Gralla, J., & Crothers, D. M. (1973) *J. Mol. Biol. 73*, 497–511.

Hall, L. M. C., & Maden, B. E. H. (1980) *Nucleic Acids Res. 8*, 5993–6005.

Hassur, S. M., & Whitlock, H. W., Jr. (1974) *Anal. Biochem. 59*, 162–164.

Hori, H., & Osawa, S. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 381–385.

Hori, H., Osawa, S., & Iwabuchi, M. (1980) *Nucleic Acids Res. 8*, 5535–5539.

Jacobson, A. (1976) *Methods Mol. Biol. 8*, 161–209.

Jacq, B. (1981) *Nucleic Acids Res. 9*, 2913–2932.

Jordan, B. R., Latil-Damotte, M., & Jourdan, R. (1980) *Nucleic Acids Res. 8*, 3565–3573.
Kelly, J. M., & Maden, B. E. H. (1980) *Nucleic Acids Res. 8*, 4521–4534.
Kelly, J. M., & Cox, R. A. (1981) *Nucleic Acids Res. 9*, 1111–1121.
Krupp, G., & Gross, H. J. (1979) *Nucleic Acids Res. 6*, 3481–3490.
Küntzel, H., Heidrich, M., & Peichulla, B. (1981) *Nucleic Acids Res. 9*, 1451–1461.
Luoma, G. A., & Marshall, A. G. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 4901–4905.
MacKay, R. M., & Doolittle, W. F. (1981) *Nucleic Acids Res. 9*, 3321–3334.
MacKay, R. M., Spencer, D. F., Doolittle, W. F., & Gray, M. W. (1980) *Eur. J. Biochem. 112*, 561–576.
Mandal, R. K., & Dawid, I. B. (1981) *Nucleic Acids Res. 9*, 1801–1811.
Nazar, R. N. (1980) *FEBS Lett. 119*, 212–214.
Nazar, R. N., & Sitz, T. O. (1980) *FEBS Lett. 115*, 71–76.
Nazar, R. N., Sitz, T. O., & Busch, H. (1975) *J. Biol. Chem. 250*, 8591–8597.
Nishimura, S. (1972) *Prog. Nucleic Acid Res. Mol. Biol. 12*, 49–85.
Noller, H. F., Kop, J., Wheaton, V., Brosius, J., Gutell, R. R., Kopylov, A. M., Dohme, F., Herr, W., Stahl, D. A., Gupta, R., & Woese, C. R. (1981) *Nucleic Acids Res. 9*, 6167–6198.
Orozco, E. M., Jr., Rushlow, K. E., Dodd, J. R., & Hallick, R. B. (1980) *J. Biol. Chem. 255*, 10997–11003.
Pace, N. R., Walker, T. A., & Schroeder, E. (1977) *Biochemistry 16*, 5321–5328.
Pavlakis, G. N., Jordan, B. R., Wurst, R. M., & Vournakis, J. N. (1979) *Nucleic Acids Res. 7*, 2213–2238.
Peffley, D. M., & Sogin, M. L. (1981) *Biochemistry 20*, 4015–4021.
Pene, J. J., Knight, E., Jr., & Darnell, J. E., Jr. (1968) *J. Mol. Biol. 33*, 609–623.
Peters, M. A., Walker, T. A., & Pace, N. R. (1982) *Biochemistry* (second paper of three in this issue).
Rich, A., & RajBhandary, U. L. (1976) *Annu. Rev. Biochem. 45*, 805–860.
Rubtsov, P. M., Musakhanov, M. M., Zakharyev, V. M., Krayev, A. S., Skryabin, K. G., & Bayev, A. A. (1980) *Nucleic Acids Res. 8*, 5779–5794.
Salser, W. (1977) *Cold Spring Harbor Symp. Quant. Biol. 42*, 985–1002.
Samols, D. R., Hagenbuchle, O., & Gage, L. P. (1979) *Nucleic Acids Res. 7*, 1109–1119.
Sanger, F., Brownlee, G. G., & Barrell, B. G. (1965) *J. Mol. Biol. 13*, 373–398.
Schwarz, Zs., & Kössel, H. (1980) *Nature (London) 283*, 739–742.
Selker, E., & Yanofsky, C. (1979) *Nucleic Acids Res. 6*, 2561–2567.
Sollner-Webb, B., & Reeder, R. H. (1979) *Cell (Cambridge, Mass.) 18*, 485–499.
Stahl, D. A. (1978) Ph.D. Thesis, University of Illinois, Urbana, IL.
Studnicka, G. M., Eiserling, F. A., & Lake, J. A. (1981) *Nucleic Acids Res. 9*, 1885–1904.
Walker, T. A., Johnson, K. D., Olsen, G. J., Peters, M. A., & Pace, N. R. (1982) *Biochemistry* (first paper of three in this issue).
Walker, W. F. (1981) *FEBS Lett. 126*, 150–151.
Wildeman, A. G., & Nazar, R. N. (1981) *J. Biol. Chem. 256*, 5675–5681.
Woese, C. R., & Fox, G. E. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5088–5090.
Woese, C. R., Magrum, L. J., Gupta, R., Siegel, R. B., Stahl, D. A., Kop, J., Crawford, N., Brosius, J., Gutell, R., Hogan, J. J., & Noller, H. F. (1980) *Nucleic Acids Res. 8*, 2275–2293.
Zuckerkandl, E., & Pauling, L. (1965) *J. Theor. Biol. 8*, 357–366.