Volume 9 Number 14 1981 Nucleic Acids Research

Secondary structure of the large subunit ribosomal RNA from Escherichia coli, Zea mays chloroplast, and human and mouse mitochondrial ribosomes

Carola Glotz, Christian Zwieb and Richard Brimacombe
Max-Planck Institut für Molekulare Genetik (Abteilung Wittman), Ihnestrasse 63-73, Berlin-Dahlem, GFR, and

Kaylene Edwards and Hans Kössel
Institut für Biologie III, Universität Freiburg, Schänzlestrasse 1, 7800 Freiburg, GFR

Received 20 May 1981

SUMMARY

Short base-paired RNA fragments, and fragments containing intra-RNA cross-links, were isolated from E. coli 23S rRNA or 50S ribosomal subunits by two-dimensional gel electrophoresis. The interactions thus found were used as a first basis for constructing a secondary structure model of the 23S rRNA. Sequence comparison with the 23S rDNA from Z. mays chloroplasts, as well as with the 16S (large subunit) rDNA from human and mouse mitochondria, enabled the experimental model to be improved and extrapolated to give complete secondary structures of all four species. The structures are organized in well-defined domains, with over 450 compensating base changes between the two 23S species. Some ribosomal structural "switches" were found, one involving 5S rRNA.

INTRODUCTION

Three groups have published models for the secondary structure of the 16S ribosomal RNA from E. coli (1-4). The experimental approach used by two of these groups (1,4) involved detailed analyses of the sites of chemical modification by single-strand specific reagents, or the sites of hydrolysis by nucleases. Our model (3) on the other hand was based on the direct identification of short base-paired fragments of the RNA (2), and the analysis of sites of intra-RNA cross-linking (5). In all three models a comparison with other ribosomal RNA sequences was used to refine and extend the experimental data. This comparative approach had already been applied successfully to models of tRNA and 5S rRNA (e.g. 6), and is the factor primarily responsible for the large measure of agreement which now exists between the three 16S models. The sequence of the 16S rDNA from Z. mays chloroplasts (7) was particularly useful for these comparisons, since it shows 74% homology to the E. coli 16S rRNA. This degree of homology is

large enough for the two sequences to be compared directly, while at the same time the degree of divergence is sufficient to allow a considerable number of "compensating base changes" to be found in the double-helical regions of the structure.

More recently, sequences for the large ribosomal subunit RNA genes from several species have become available. These include the 23S rDNA from E. coli (8), the 16S rDNA from human (9) and mouse (10) mitochondria, and also the 23S rDNA from Z. mays chloroplasts (11). In this paper we present secondary structure models for the corresponding four rRNA species, based on the same experimental strategy as that which we employed in the case of the E. coli 16S rRNA (2,3,5,12). Base-paired fragment and cross-link site analyses yielded a network of secondary structural interactions (including several long-range interactions) covering a large proportion of the E. coli 23S rRNA, and clear evidence was found for some structural "switches" (cf. 3,13), one involving 5S rRNA. The optimal base-pairings for these interactions were determined by searching for compensating base changes between the E. coli and Z. mays chloroplast 23S sequences in the regions concerned, and this sequence comparison was used to extend the structures into intervening regions where there was no direct experimental evidence. When the two 23S molecules were compared with the two mitochondrial 16S sequences, a number of areas of primary structural homology could be observed (more extensive than those already published (9)), and this enabled secondary structures for the mitochondrial sequences to be built up simultaneously, all four sequences being compared together for the final refinement of the structures.

MATERIALS AND METHODS

^{32}P-labelled 50S ribosomal subunits were prepared from E. coli strain MRE 600 as described (14), and 23S rRNA was isolated by centrifugation of the subunits through sucrose gradients containing 0.1% dodecyl sulphate and 2 mM EDTA (15). The isolated RNA was partially digested with ribonuclease T_1 for 30 min at 25^{o} or 37^{o} under a variety of conditions (10-100 units enzyme per A_{260} unit of RNA, in buffers containing 10 mM Tris-HCl pH 7.8, 0.1% dodecyl sulphate, and either 0.3 mM EDTA or 0.3-10 mM mag-

nesium acetate). After incubation, the digests were treated with proteinase K (cf. 2) and applied to two-dimensional gels, exactly as described (2,16). In other experiments, 50S subunits were dialysed against 10 mM Tris-HCl pH 7.8, 0.03 mM magnesium acetate, treated with ribonuclease T_1 (12-50 units enzyme per A_{260} unit 50S) as described previously (17), and applied to a 5-20% polyacrylamide gradient gel containing 5 mM magnesium at pH 6 (cf. 18): This treatment releases a large number of free RNA fragments from the subunit, and regions of the gel containing these fragments (cf. 17) were applied directly to the second dimension of the two-dimensional gel system above (cf. 2). In all cases, pairs or families of fragments were extracted from the two-dimensional gels and identified exactly as described (2,16).

Intra-RNA cross-links were generated by ultra-violet irradiation of ^{32}P-labelled 50S subunits, and the cross-link sites identified as described in the case of the 30S subunit (5).

RESULTS AND DISCUSSION

The results of several hundred base-paired and cross-linked fragment analyses are summarized in Fig. 1, which is divided into sections corresponding to the various domains of the secondary structure (Figs. 2a-7a, see below). As in our studies on 16S rRNA (2,16), the same fragments were often cut at slightly different points by the ribonuclease in the different experiments. Consequently the 3'- and 5'-ends of the fragments listed should be taken as approximate to the extent of 3 or 4 bases. No significant pattern of difference was observed between those fragments obtained from isolated 23S rRNA and those from 50S subunits (see Materials and Methods). The fragment complexes involving 5S rRNA (18a-d, 27a-b) were of course only found in digests of the 50S subunits (see below and Fig. 9).

In the case of the cross-linked complexes (Fig. 1), both oligonucleotide components of the cross-link site could sometimes be identified, and these complexes are illustrated in Fig. 8 (see later). In other cases, however, only one of the oligonucleotides involved in the cross-link could be determined, most probably as a result of the other component of the cross-link being at a position which appears as a short uncharacteristic oligonucleotide

in our ribonuclease T_1 analysis (cf. 5). In such cases, the cross-linked complexes served either to indicate or confirm long-range interactions (e.g. complexes X4, X5, X10), or to give an indication as to where short-range secondary structure could be expected (e.g. complexes X1, X2). As was previously found (5), all the cross-links so far determined are compatible with the secondary structure.

As explained in the Introduction, the fragment and cross-link data were used as a first basis for building the secondary structure models. The complete structures derived for E. coli and Z. mays 23S rRNA are given in Figs. 2a-7a, and those for human and mouse mitochondrial 16S rRNA in Figs. 2b-7b. The structure diagrams show the base changes between the two pairs of sequences, and also indicate sequence homology between E. coli 23S and human mitochondrial 16S rRNA (see General Legend to Figs.2-7). In Figs. 2a-7a, the sites of kethoxal modification of E. coli 50S subunits (8) are included, and the 5'- and 3'-ends of the fragment pairs (but not of the cross-linked complexes, Fig. 1) are also shown. It should be noted that since some of the nuclease digestion conditions used were quite vigorous (see Materials and Methods), the ends of the fragments sometimes occur in double-stranded regions of the structure.

The various elements of the secondary structure vary from those which can be regarded as very well established (elements present in all four sequences, with several pairs of compensating base changes, and supported by fragment and/or cross-link data) to others which must obviously be regarded as tentative. There is no space here to describe the structures in detail, and the following brief notes are intended only to help the reader to find the salient features, and to give a few illustrative examples of various points of interest.

FIGURE 2: The interaction predicted (19,20) between the 5'- and 3'-ends of the E. coli 23S rRNA is supported by comparison with the Z. mays chloroplast 23S and 4.5S rRNA species (Fig. 2a), the 4.5S rRNA being equivalent to the 3'-terminus of E. coli 23S rRNA (20-22, and see Fig. 7a).Nazar (23) has pointed out the equivalence of eukaryotic 5.8S rRNA to the 5'-region of the 23S rRNA. If the stem region of his proposed structure (24) for trout

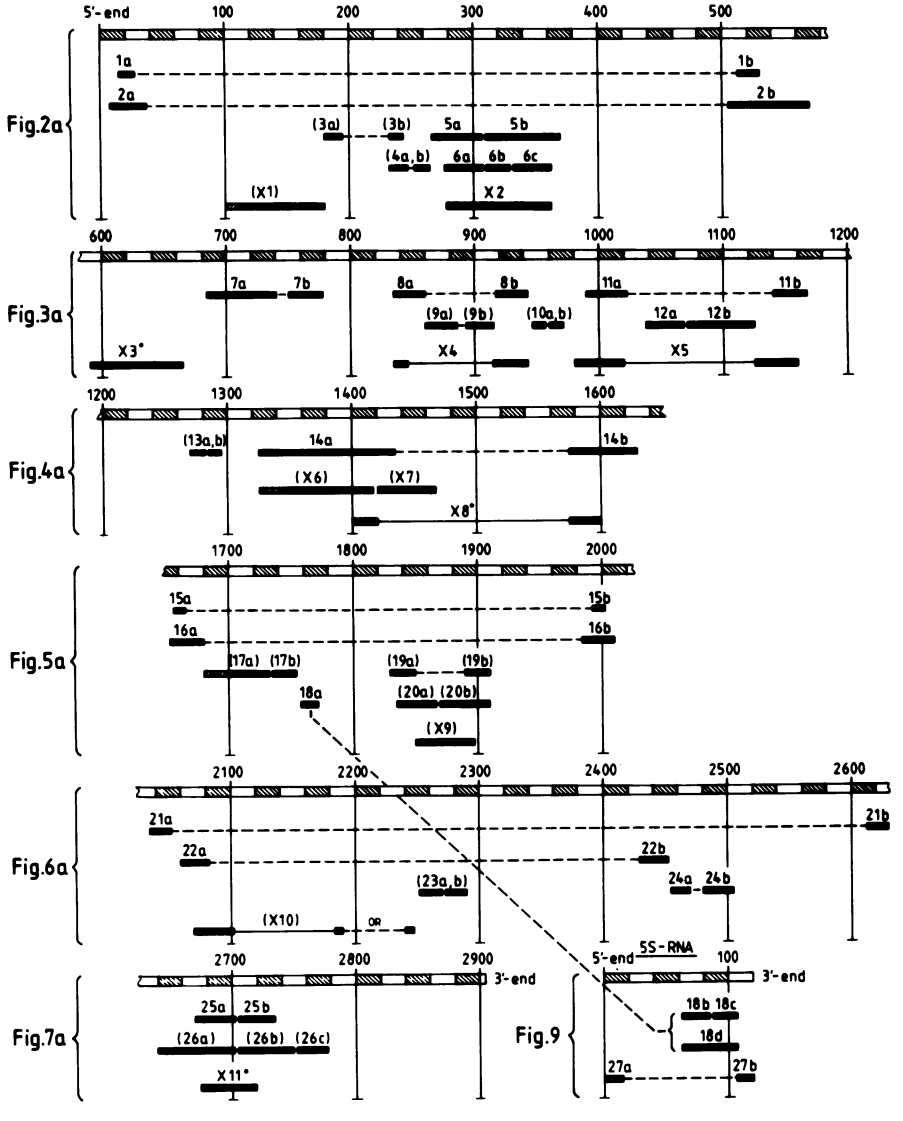

Figure 1: Location of base-paired and cross-linked fragments in E. coli 23S (and 5S) rRNA. Base-paired fragments are denoted by bars numbered 1a, 1b etc., and cross-linked complexes by bars numbered X1, etc. Numbers in brackets indicate cases where the fragment pair was only observed once or twice, and cross-hatched regions of the fragment bars indicate sequences which were sometimes present, sometimes not. The sequence is divided into sections corresponding to the domains of the secondary structure (Figs. 2a-7a, 9). Cross-linked complexes with an asterisk are shown in detail in Fig. 8.

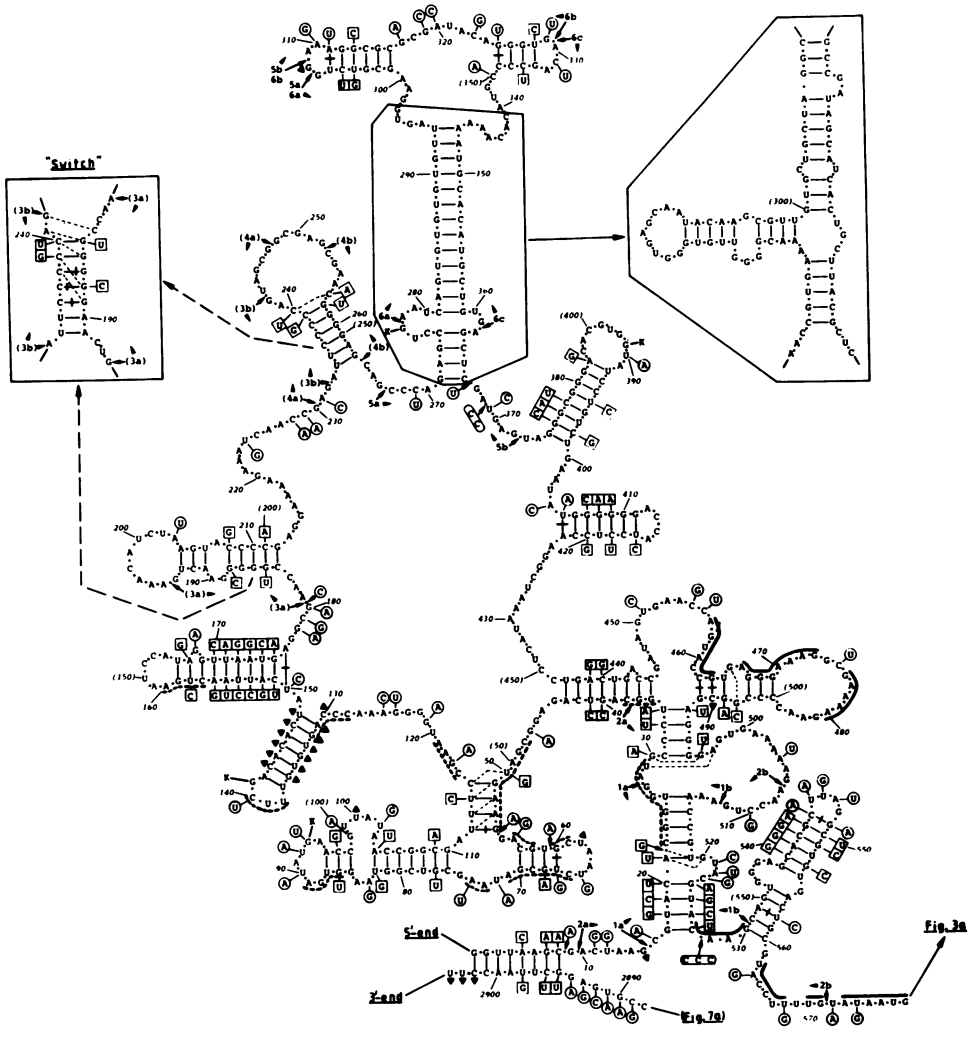

Figure 2a: E. coli/Z. mays 23S rRNA, bases 1 to 577.

GENERAL LEGEND TO FIGS. 2 - 7: *The principal sequence joined by dots is that of E. coli 23S rRNA in Figs. 2a-7a, and of human mitochondrial 16S rRNA in Figs. 2b-7b. These sequences are numbered every 10 bases from the 5'-end. Base changes in the sequence used for comparison (Z. mays chloroplast 23S rRNA in Figs. 2a-7a, and mouse mitochondrial 16S rRNA in Figs. 2b-7b) are indicated by nucleotides in square or round boxes; these latter sequences are numbered (in brackets) every 50 bases. Base changes in square boxes are those which are compensating or which enhance the secondary structure, whereas those in round boxes are in single-stranded regions or are non-compensating. A-U, G-C and G-U base-pairs in the principal sequence are denoted by bars; thin dotted lines indicate modified base-pairing in the second (chloroplast or*

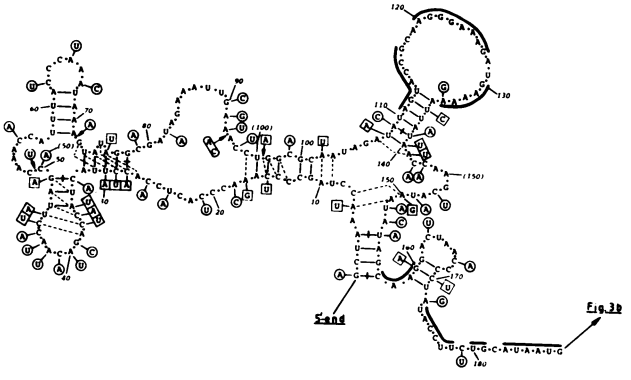

Figure 2b: Human/mouse mitochondrial 16S rRNA, a tentative structure for bases 1 to 188 (see text).

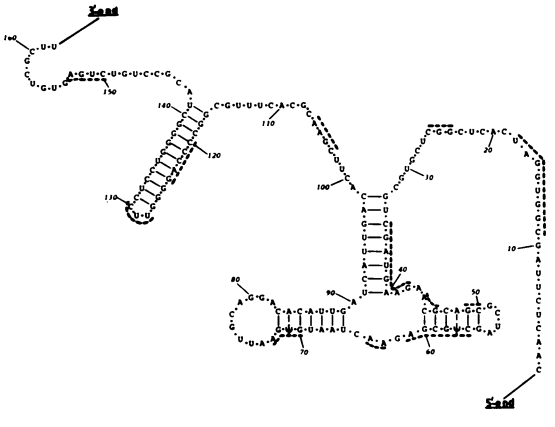

Figure 2c: Trout 5.8S rRNA (cf. refs. 23, 24, and see text).

mouse mitochondrion) sequence, and a bar "crossing out" a base-pair indicates that this pair is not present in the second sequence. A base change with an arrow pointing between two bases of the principal sequence is an insertion, and a solid triangle is a deletion. Major differences between E. coli and Z. mays or between human and mouse mitochondrion are shown by the boxed inset diagrams. Alternative structures are also indicated by inset diagrams (see text). Thick lines along the sequence denote tracts of homology between E. coli 23S rRNA and human mitochondrion 16S rRNA, at the identical or nearly identical position in the secondary structure. (Similarly, thick dashed lines (Fig. 2a) denote homology between E. coli 23S rRNA and trout 5.8S rRNA (23, Fig. 2c); see text). The approximate ends of the base-paired fragments (numbered as in Fig. 1) are indicated in Figs. 2a-7a, and the sites of reaction of kethoxal with 50S subunits (8) are denoted by "K". It should be noted that in variable regions of the sequences, the homology has been arranged to give the best fit to the secondary structure; the alignment is therefore sometimes slightly different to that cited in the primary sequence (11).

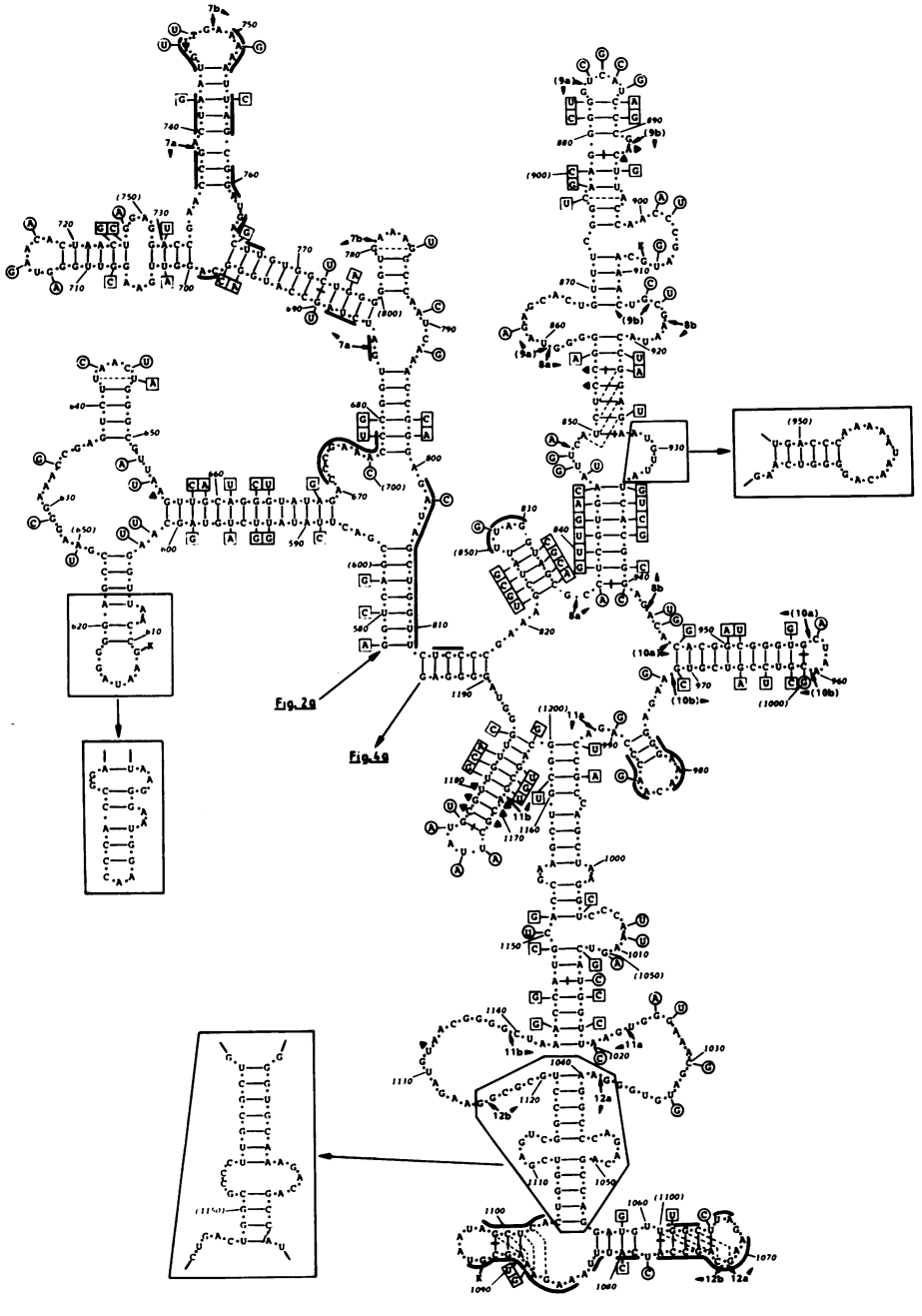

Figure 3a: E. coli/Z. mays 23S rRNA, bases 578 to 1195.

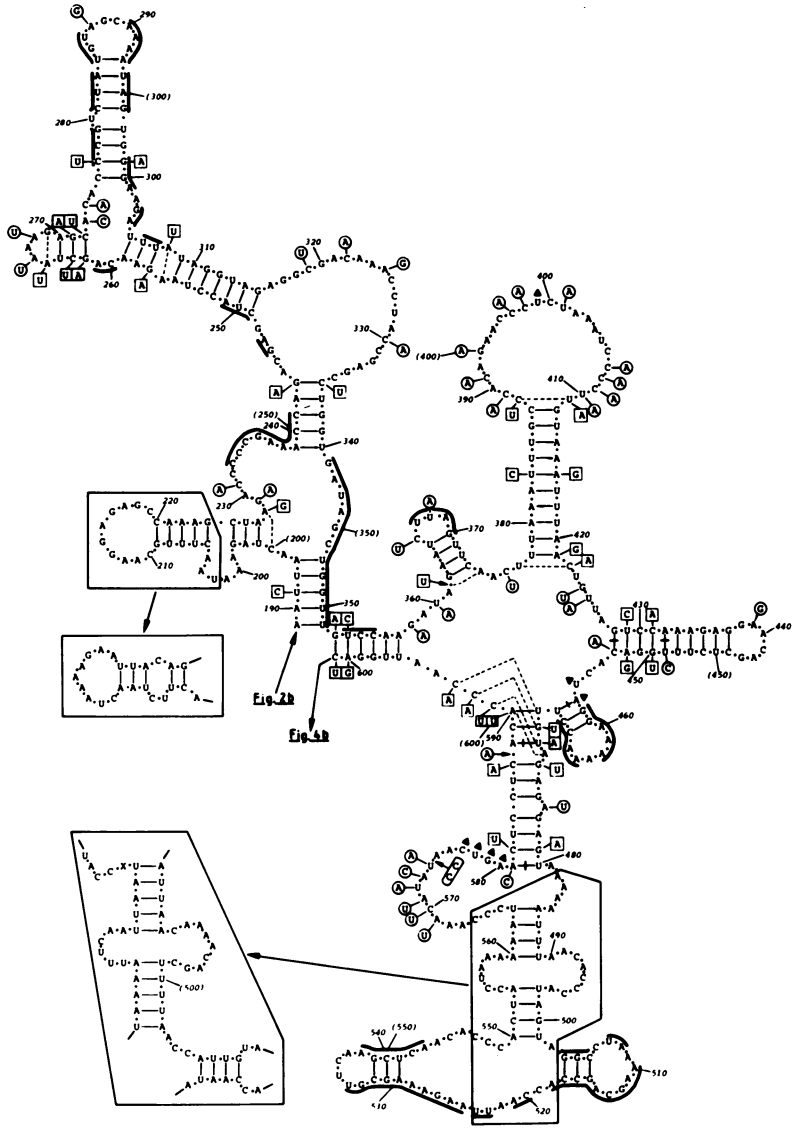

Figure 3b: Human/mouse mitochondrial 16S rRNA, bases 189 to 601.

5.8S rRNA is opened up in a manner similar to that suggested by Pace et al (25), then the secondary structure becomes very similar to that shown for the corresponding region in Fig. 2a. (G. Clark and S. Gerbi (personal communication) have also proposed

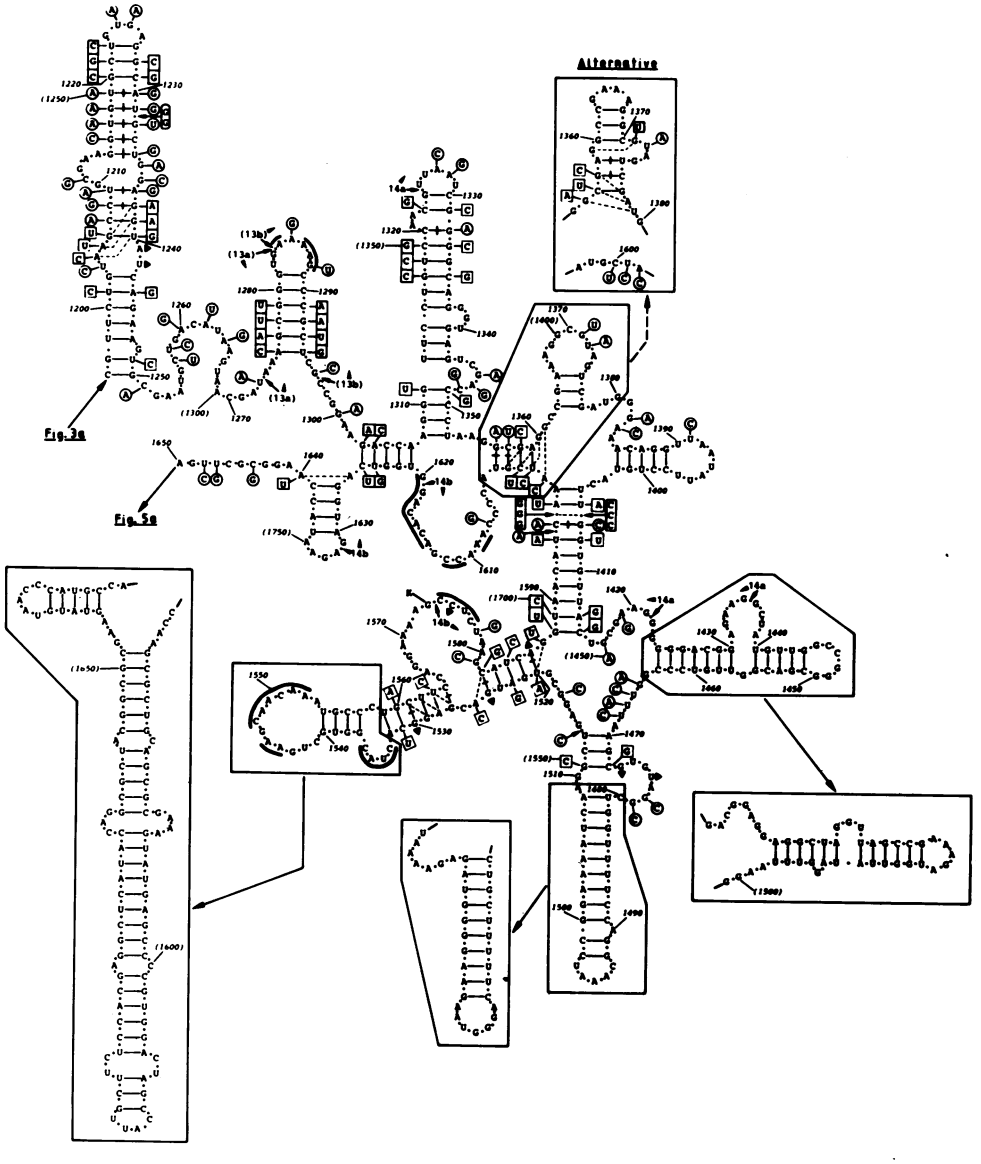

Figure 4a: E. coli/Z. mays 23S rRNA, bases 1196 to 1650.

some of the elements of this structure). The "opened-up" structure of the trout 5.8S rRNA is given in Fig. 2c, and the primary sequence homology to E. coli (23) is also indicated. In E. coli however, the region from bases 17-25 is involved in a very well-

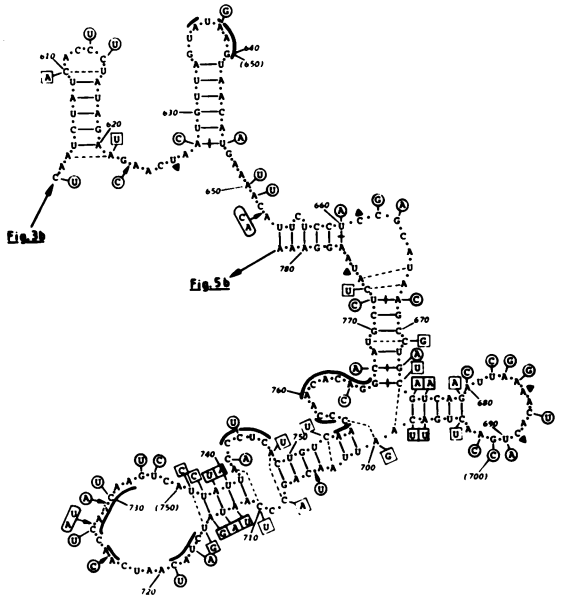

Figure 4b: Human/mouse mitochondrial 16S rRNA, bases 602 to 782.

established interaction with bases 515-527 (fragment pairs 1 and 2, Fig. 1). These structures would predict that 5.8S rRNA is base-paired to both the 5'- and 3'-ends of the 28S rRNA, and evidence for the latter interaction has recently been published (26).

The remainder of the domain shown in Fig. 2a contains a number of loops strongly supported by the phylogenetic evidence, one loop (bases 130-150) being deleted in the chloroplast. There is also a "switch" (fragment pair 3) which is acceptably compatible with the chloroplast sequence. The latter shows an extra loop in the 270-370 region (inset diagram), the "T"-shaped structure for this region in E. coli being clear from the fragment data. Bases 450-490 show strong homology with the mitochondrial RNA (Fig. 2b), but the two corresponding regions cannot be arranged in an identical structure, and may therefore be involved in other interactions. The whole 5'-region of the mitochondrial RNA (Fig. 2b) is drastically shorter, shows many mutations between the two species, and the possibilities for secondary structure are very few. It may be that non-Watson-Crick base-pairs occur in this region (as have been found in mitochondrial tRNA (e.g. 9,10)), and it is

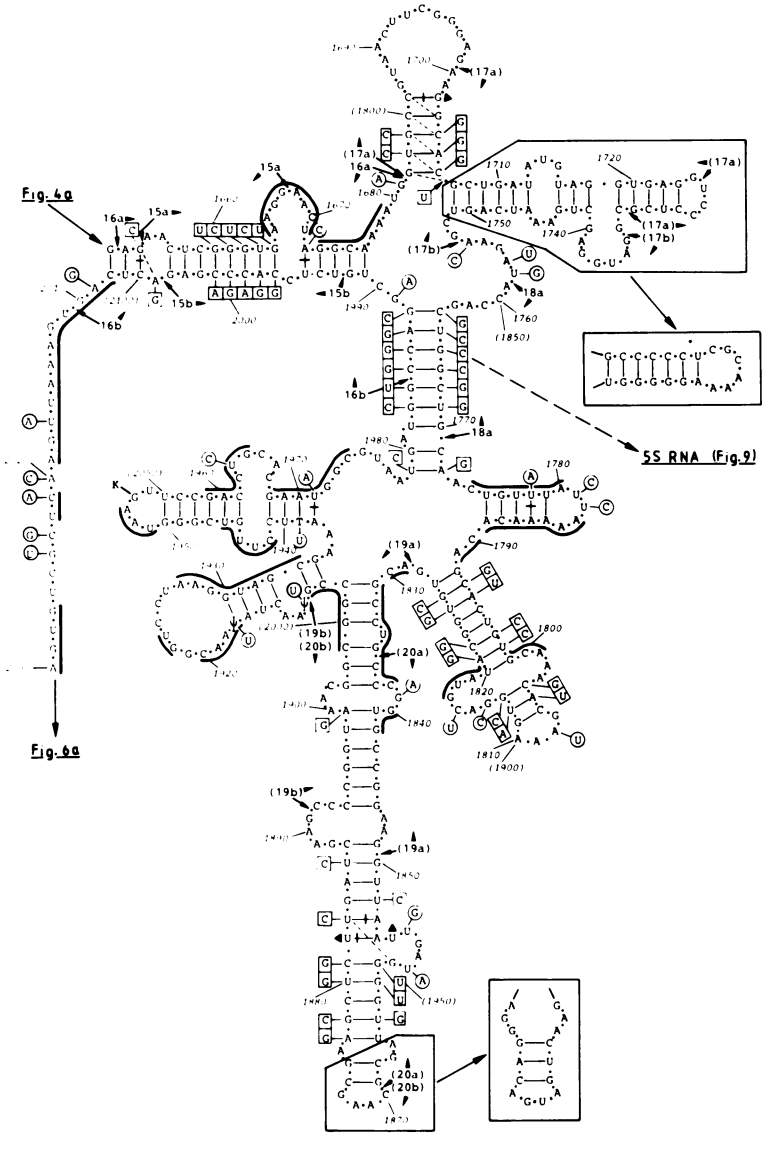

Figure 5a: E. coli/Z. mays 23S rRNA, bases 1651 to 2030.

noteworthy in this context that a number of A-C "mis-matches" occur in all four structures. Several possible interactions can be drawn with the 3'-end of the rRNA in both mitochondrial se-

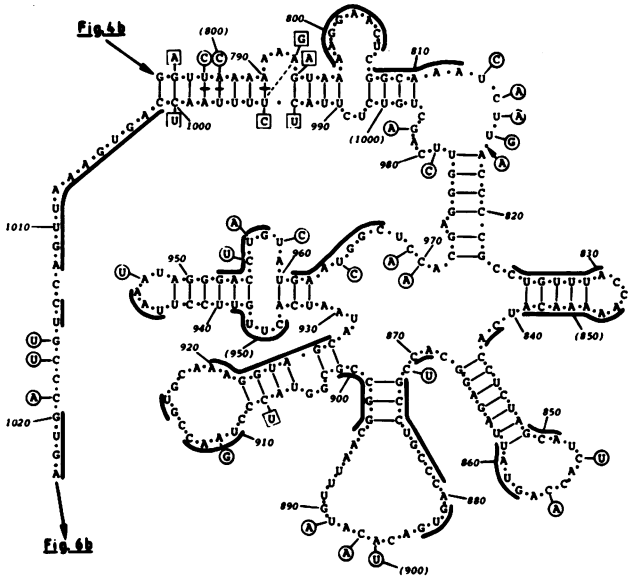

Figure 5b: Human/mouse mitochondrial 16S rRNA, bases 783 to 1023.

quences, but in contrast to E. coli/Z. mays none is very convincing. The whole structure shown in Fig. 2b must be regarded as tentative, but the 3'-end of the domain is however clearly defined by the strong homology (bases 172-188) with the E. coli sequence (bases 562-577).

FIGURE 3: This is a region of highly-conserved secondary structure which is largely self-explanatory. Both structures (Fig. 3a and 3b) are well supported by compensating base changes, the mitochondrial structures showing some clean "amputations". The 590-670 region (Fig. 3a) is supported by a reproducible cross-link (Fig. 8). The interaction between bases 812-817 and 1190-1195 is a good example of the interplay between all four sequences which was exploited in building the structures; the corresponding region in Fig. 3b shows four compensating base changes. There is an extra loop in the chloroplast sequence at position 930, and it is noteworthy that the whole lower domain of the structure (bases 990-1165, Fig. 3a) is shortened in the mitochondrion by approximately one helix-turn.

FIGURE 4: This region contains some clearly-defined elements,

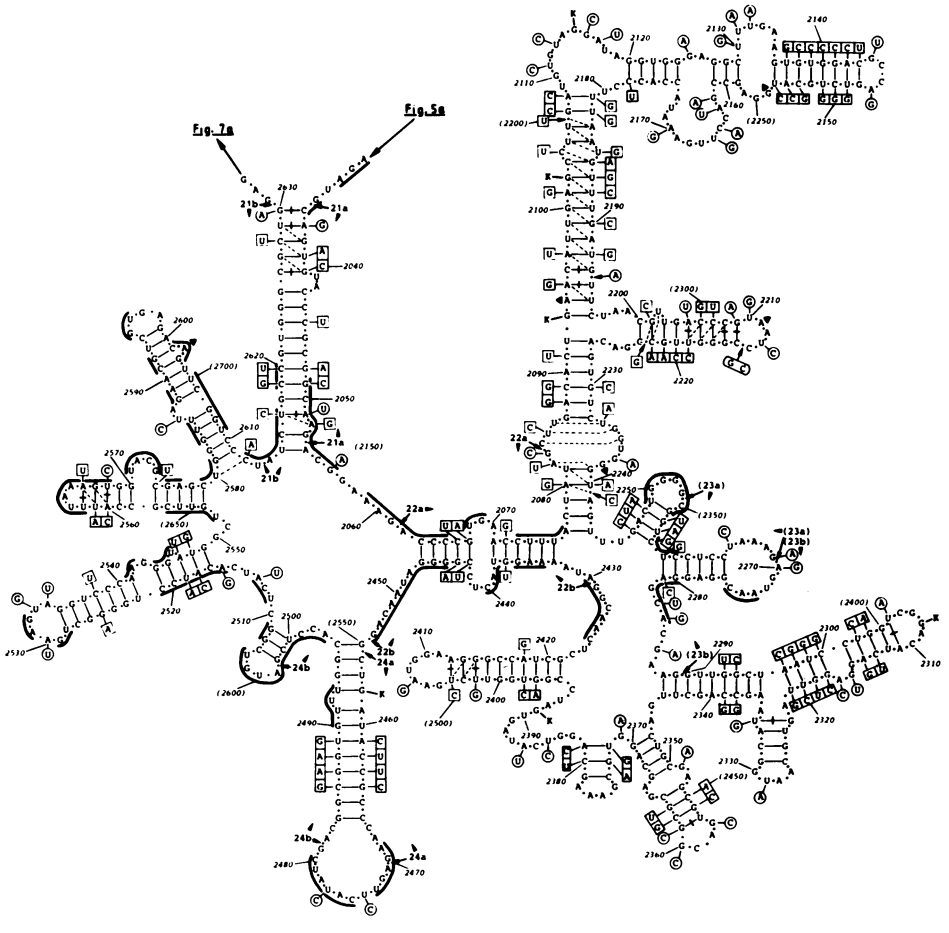

Figure 6a: E. coli/Z. mays 23S rRNA, bases 2031 to 2633.

and some regions where the data are not so satisfactory. Two possible structures are suggested for the 1350-1380 area (Fig. 4a). The main feature in Fig. 4a is the well-supported interaction between bases 1405-1414 and 1588-1597, which was also cross-linked (Fig. 8). The domain enclosed by this interaction contains the largest insert in the chloroplast RNA (11), and is altogether rather variable. The corresponding mitochondrial structure (Fig. 4b) is drastically reduced, but, in contrast to Fig. 2b, can be arranged in a reasonable structure with several compensating base changes. The sequence homologies between the structures in Figs.

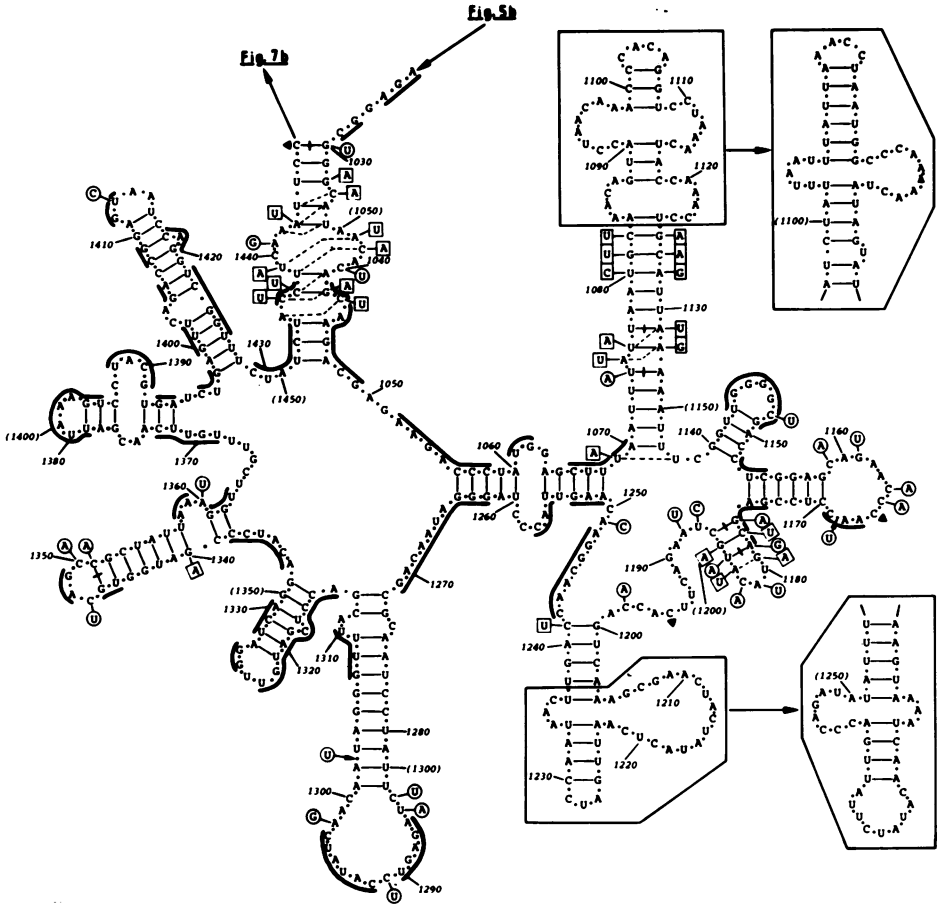

Figure 6b: Human/mouse mitochondrial 16S rRNA, bases 1024 to 1447.

4a and 4b are also noteworthy.

FIGURE 5: The structure here, both primary and secondary, is highly conserved. The domain is enclosed by the well-defined interaction between bases 1651-1676 and 1991-2008 (Fig. 5a). The mitochondrial structures show several amputations, and the chloroplast sequence also shows an amputated loop (corresponding to the 1710-1750 region in E. coli). The structure for the most highly-conserved region in E. coli and Z. mays (bases 1910-1980) was deduced by comparison with the mitochondrial sequences. (Pseudo-uridine and ribothymidine residues in the E. coli sequ-

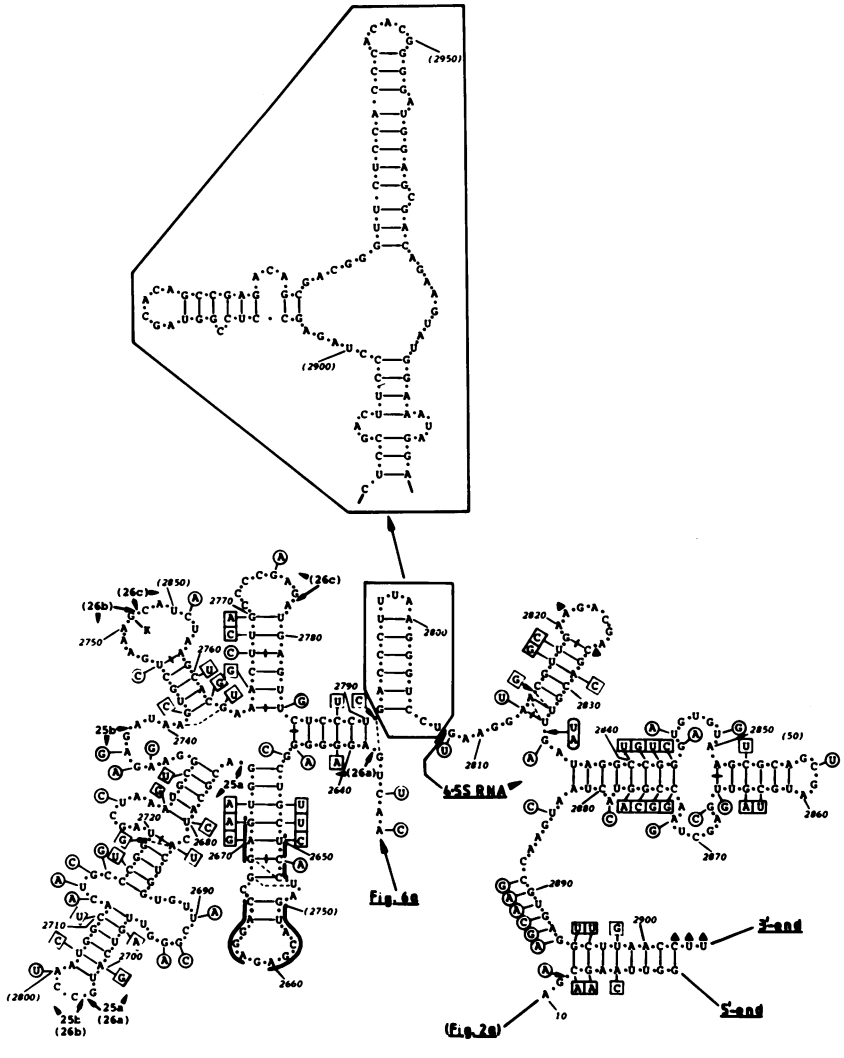

Figure 7a: E. coli/Z. mays 23S rRNA, bases 2634 to 2904. The region corresponding to chloroplast 4.5S rRNA is indicated (see text).

ence in this region are denoted as uridines in the Z. mays sequence, there being as yet no evidence (11) for base modifications in the latter). Fig. 5a also contains the "switch" structure (bases 1759-1770) involving 5S rRNA, which was found in the fragment analyses (fragment family 18, Fig. 1), and this is illustrated in Fig. 9. A corresponding interaction can also be drawn

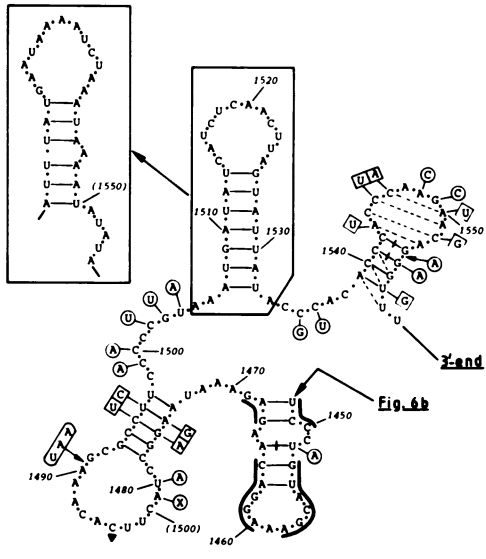

Figure 7b: Human/mouse mitochondrial 16S rRNA, bases 1448 to 1559.

between Z. mays 5S and 23S rRNA (Fig. 9). The fragment data (pair 27) served in addition to confirm the stem structure of 5S rRNA in situ in the 50S subunit.

FIGURE 6: Here again the structure is highly conserved, the domain being enclosed by the interaction between bases 2036-2054 and 2615-2630 (Fig. 6a), although the corresponding interaction in Fig. 6b is not very satisfactory. The domain includes the binding site for ribosomal protein L1 (bases ca. 2090-2200), for which very similar secondary structures have already been proposed (27, 28). This region is shortened in the mitochondria, the latter also showing some amputations in the lower right-hand corner of the diagram. Baer and Dubin (29) have proposed a structure for the region between bases 2450 and 2610, which differs from the structure shown in Fig. 6a in that they pair bases 2506-2513 with bases 2575-2582. Both possibilities are supported by compensating base changes (between all four sequences) so it is likely that both versions are correct; the situation closely parallels the "switch" in Fig. 2a. It is noteworthy that Fig. 6a contains three kethoxal sites (8) which are not compatible with the secondary structure; these are at positions 2093, 2102 and 2458. All other kethoxal

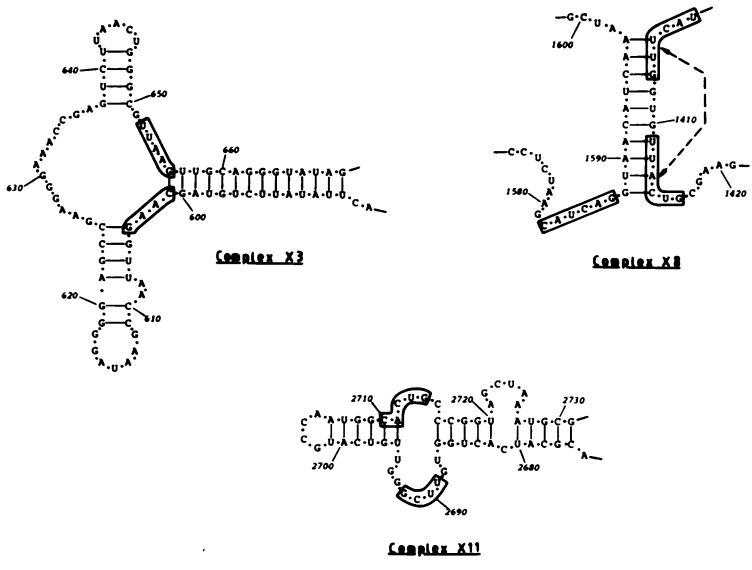

Figure 8: Structures of cross-linked complexes (cf. Fig. 1). Boxed-in oligonucleotides were absent (i.e. contained the cross-link site, cf. ref. 5). The two arrowed oligonucleotides in X8 could not be distinguished in our analysis; only one was absent.

sites throughout the 23S rRNA are in single-stranded regions (Figs. 2a-7a).

FIGURE 7: The region up to base 2790 (Fig. 7a) is characterized by good fragment data (Fig. 1), and a clear cross-link site (Fig. 8), as well as by compensating base changes. As already mentioned above, the 3'-end of the E. coli 23S rRNA corresponds to the 4.5S rRNA in chloroplast (20-22), the 5'-terminus of the latter being indicated in the Figure. The structure for this region in Fig. 7a is almost identical to that proposed by Branlant et al (20). The precise 3'-end of the 23S rRNA in Z. mays is not yet certain (11), and the inserted bases corresponding to the rDNA can be arranged in the secondary structure shown in the inset diagram. The structures suggest that the 4.5S rRNA is not hydrogen-bonded to 23S rRNA, in agreement with published observations (30) and in contrast to 5.8S rRNA (see above). The corresponding region in mitochondrial rRNA (Fig. 7b) is drastically shortened, but shows one loop which is obviously homologous to the 23S rRNA structure.

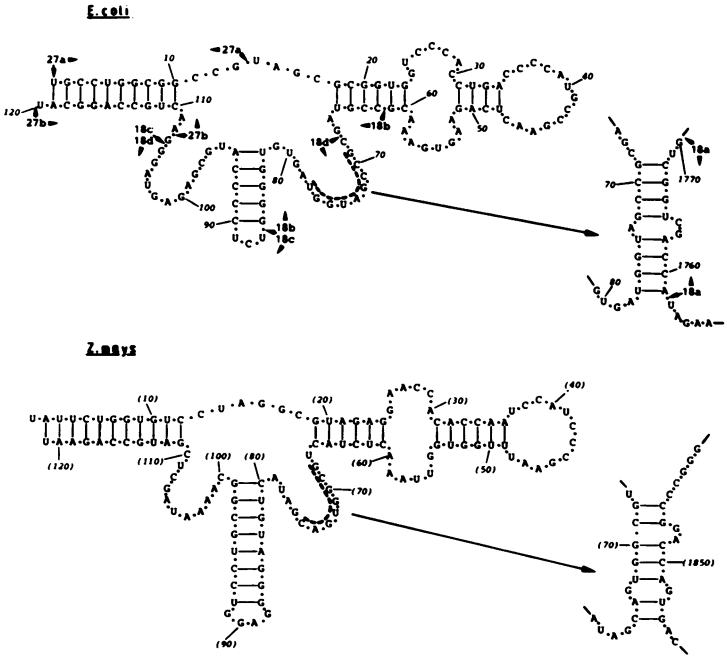

Figure 9: E. coli 5S rRNA (in the structure of ref. 6), showing the fragment families 18 and 27 (Fig. 1). The arrowed structure is the interaction found with 23S rRNA (Fig. 5a). (Fragment 18a was reproducibly formed by a contaminating ribonuclease A activity). The corresponding structure for Z. mays 5S rRNA (31) and its possible interaction with 23S rRNA are shown below.

In all there are over 450 compensating base changes between the E. coli and Z. mays chloroplast 23S rRNA species, and over 100 such changes between the two mitochondrial 16S sequences. The structures presented here provide a firm basis for further work on ribosomal RNA, and demonstrate that base-paired fragment and cross-link site analysis, coupled with sequence comparison, offers the most powerful combination currently available for the determination of secondary structure in large RNA molecules.

ACKNOWLEDGEMENTS

The authors are grateful to Drs. T. Dyer, S. Gerbi and R.A. Zimmermann for making their data available to us prior to publication. We also thank Dr. H.G. Wittmann for his continued support. The work was aided by grants from the Deutsche Forschungsgemeinschaft (SFB 9 to C.Z. and SFB 46 to H.K.).

REFERENCES

1. 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. and Noller, H.F. (1980) Nucleic Acids Res. 8, 2275-2293
2. Glotz, C. and Brimacombe, R. (1980) Nucleic Acids Res. 8, 2377-2395
3. Brimacombe, R. (1980) Biochemistry Int. 1, 162-171
4. Stiegler, P., Carbon, P., Zuker, M., Ebel, J.P. and Ehresmann, C. (1980) C.R. Acad. Sci. Paris, Ser. D. 291, 937-940
5. Zwieb, C. and Brimacombe, R. (1980) Nucleic Acids Res. 8, 2397-2411
6. Fox, G. and Woese, C.R. (1975) Nature, 256, 505-507
7. Schwarz, Zs. and Kössel, H. (1980) Nature, 283, 739-742
8. Brosius, J., Dull, T.J. and Noller, H.F. (1980) Proc. Nat. Acad. Sci. USA, 77, 201-204
9. Eperon, I.C., Anderson, S. and Nierlich, D.P. (1980) Nature, 286, 460-467
10. Van Etten, R.A., Walberg, M.W. and Clayton, D.A. (1980) Cell, 22, 157-170
11. Edwards, K. and Kössel, H. (1981) Nucleic Acids Res. submitted
12. Brimacombe, R. (1980) in "Biological Implications of Protein-Nucleic Acid Interactions", Augustyniak, J. Ed., pp. 44-62, Elsevier/North Holland Press
13. Noller, H.F. (1980) in "Ribosomes", Chambliss, G.R., Davies, J., Davis, K., Kahan, L. and Nomura, M. Eds., pp. 3-22, University Park Press, Baltimore
14. Yuki, A. and Brimacombe, R. (1975) Eur. J. Biochem. 56, 23-34
15. Möller, K. and Brimacombe, R. (1975) Mol. Gen. Genet. 141, 343-355
16. Ross, A. and Brimacombe, R. (1979) Nature, 281, 271-276
17. Newton, I. and Brimacombe, R. (1974) Eur. J. Biochem. 48, 513-518
18. Rinke, J., Yuki, A. and Brimacombe, R. (1976) Eur. J. Biochem. 64, 77-89
19. Branlant, C., Sriwidada, J., Krol, A. and Ebel, J.P. (1976) Nucleic Acids Res. 3, 1671-1687
20. Machatt, M.A., Ebel, J.P. and Branlant, C. (1981) Nucleic Acids Res. 9, 1533-1549
21. MacKay, R.M. (1981) FEBS Lett. 123, 17-18
22. Edwards, K., Bedbrook, J., Dyer, T. and Kössel, H. (1981) Biochemistry Int. 2, 533-538
23. Nazar, R.N. (1980) FEBS Lett. 119, 212-214
24. Nazar, R.N. and Roy, K.L. (1978) J. Biol. Chem. 253, 395-399
25. Pace, N.R., Walker, T.A. and Schroeder, E. (1977) Biochemistry, 16, 5321-5328
26. Kelly, J.M. and Cox, R.A. (1981) Nucleic Acids Res. 9, 1111-1121
27. Branlant, C., Krol, A., Machatt, A. and Ebel, J.P. (1981) Nucleic Acids Res. 9, 293-307
28. Gourse, R.L., Thurlow, D.L., Gerbi, S.A. and Zimmermann, R.A. (1981) Proc. Nat. Acad. Sci. USA, in the press
29. Baer, R.J. and Dubin, D.T. (1981) Nucleic Acids Res. 9,323-337
30. Whitfield, P.R., Leaver, C.J., Bottomley, W. and Atchison, B.A. (1978) Biochem. J. 175, 1103-1112
31. Bedbrook, J. and Dyer, T. (1981) Nucleic Acids Res. submitted