Volume 11 Number 21 1983 Nucleic Acids Research

Refined secondary structure models for the 16S and 23S ribosomal RNA of *Escherichia coli*

Peter Maly and Richard Brimacombe

Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, D-1000 Berlin-Dahlem, FRG

Received 13 September 1983; Accepted 11 October 1983

SUMMARY

The complete range of published sequences for ribosomal RNA (or rDNA), totalling well over 50,000 bases, has been used to derive refined models for the secondary structures of both 16S and 23S RNA from E. coli. Particular attention has been paid to resolving the differences between the various published secondary structures for these molecules. The structures are described in terms of 133 helical regions (45 for 16S RNA and 88 for 23S RNA). Of these, approximately 20 are still tentative or unconfirmed. A further 20 represent helical regions which definitely exist, but where the detailed base-pairing is still open to discussion. Over 90 of the helical regions are however now precisely established, at least to within one or two base pairs.

INTRODUCTION

The derivation of accurate secondary structure models for ribosomal RNA is a vital step towards reaching an understanding of the three-dimensional organisation and function of these molecules. In the case of the Escherichia coli ribosome, three research groups have independently proposed secondary structures for both the 16S (1-3) and the 23S (4-6) RNA, the models being based on various types of experimental evidence, combined with comparative studies (cf. 7) using sequence data from other organisms. Although the latest versions of the models (1-6) agree with one another to a very substantial extent, there are nevertheless a number of differences between them, some trivial and some important, and these differences were summarized in a recent review article from our laboratory (8). Since that time, several new rDNA sequences have appeared in the literature (see Table 1 below), and the number of residues in completed rRNA or rDNA sequences now totals well over 50,000 bases. In this paper we have made use of all the available sequence data to refine our models for the secondary structure of both 16S and 23S RNA from E. coli, with particular emphasis on resolving as far as possible the differences between the various models cited above.

RESULTS AND DISCUSSION

Table 1 lists the published sequences of ribosomal RNA or DNA which have been considered in this study. In addition, the sequence of 28S rDNA from X. laevis has recently been completed (S. Gerbi, personal communication). Two other sequences, that of Bacillus brevis 16S rDNA (1) and that of Bacillus stearothermophilus 23S rDNA (6), have also been used to derive secondary structure models (1,6), but the sequences themselves have not to our knowledge been published and therefore could not be taken into account.

When any one of the sequences listed in Table 1 is compared with the 16S or 23S RNA of E. coli, as appropriate, four categories of sequence region can be distinguished. Firstly, there are short regions of highly conserved primary structure, which serve to locate the "comparing" sequence (cf. Table 1) precisely with respect to the E. coli sequence. Secondly, there are double-helical regions in which the comparing sequence shows compensating base changes with respect to the E. coli sequence (e.g. an A-U base pair in E. coli becomes a G-C pair in the comparing sequence). The majority of the secondary structural elements fall into this category, and it is important to note that in the refined models described below the helical elements of the comparing sequence can in such cases be superimposed base-for-base upon the corresponding elements of the E. coli structure. This contrasts with several of the published secondary structure models for the various sequences (Table 1) or our own earlier models (e.g. 3), where similar but not necessarily superimposable structural elements were often drawn for the comparing sequences.

The third type of sequence region comprises a minority of helical elements where the evidence is conflicting. That is to say some comparing sequences support the proposed helical element with compensating base changes, whereas others are contradictory with mis-matching base changes. In such cases the correct structure is still open to discussion. The fourth and final category consists of those sequence regions where major deletions or insertions occur between the RNA molecules of the different size classes (12S-18S in the small subunit, 16S-28S in the large subunit, Table 1, and cf. refs. 3,4,30). Since our interest is focussed on the E. coli RNA, we have only considered these highly variable regions in this study insofar as they provide information relevant to the 16S and 23S structures.

In the following sections the status of each helical element in the 16S and 23S RNA secondary structures is given in the form of an extended "Table" for each of the two molecules, and we use the following terminology. The

Table 1: Sequences of rDNA or rRNA molecules. Known sequences are listed, showing the S-value of the corresponding rRNA molecule and (in parentheses) the appropriate literature reference. A dash indicates that the sequence is not available.

Organism	Small subunit	Large subunit
Human mitochondrion	12S (9)	16S (9)
Mouse mitochondrion	12S (10)	16S (10)
Rat mitochondrion	12S (11)	16S (12)
Saccharomyces cerevisiae mitochondrion	15S (13)	21S (14)
Aspergillus nidulans mitochondrion	15S (15)	-
Paramecium primaurelia mitochondrion	-	20S (16)
Escherichia coli	16S (17)	23S (18)
Proteus vulgaris	16S (19)	-
Halobacterium volcanii	16S (20)	-
Anacystis nidulans	16S (21)	-
Zea mays chloroplast	16S (22)	23S (23)
Euglena gracilis chloroplast	16S (24)	-
Chlamydomonas reinhardii chloroplast	16S (25)	-
Nicotiana tabacum chloroplast	16S (26)	23S (27)
Saccharomyces cerevisiae cytoplasm	18S (28)	26S (29)*
Saccharomyces carlsbergensis cytoplasm	-	26S (30)*
Xenopus laevis cytoplasm	18S (31)	-
Physarum polycephalum cytoplasm	-	26S (32)

*These two sequences are almost identical

models of Noller and Woese and their co-workers (1,6) are for convenience referred to as the "american", those of Ebel and his group (2,5) as the "french", and ours (3,4) as the "german" versions of the structures. "Undisputed and reconfirmed" means that the structural element concerned was in all three models (1-3, or 4-6), and that the new sequence data provide further confirmatory evidence in the form of compensating base changes. In cases where a helical element or part of one was disputed between the models, the refined version of the element is designated "confirmed" or "preferred", according as to whether the data are now unequivocal or still to some extent ambiguous (cf. the foregoing discussion). Sets of base pairs in hairpin loops are referred to as "distal" or "proximal" to the closed end of the hairpin, whereas in more complex helical regions they are either described in terms of their orientation in the various Figures (e.g. "right-hand, central, left-hand", or "upper, lower"), or are denoted by the numerical positions of the bases in the E. coli sequence. The term "compensating base change" is abbreviated to "CBC". "One-base slippage" means that the equivalent base-

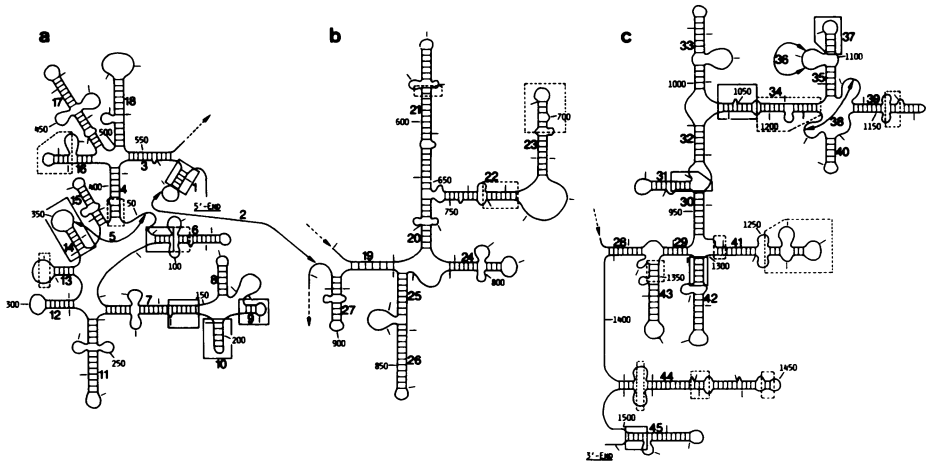


Figure 1: The previous secondary structure model of *E. coli* 16S RNA (3). The sequence is divided into three domains (a to c), and is numbered from the 5'-end (every 50 bases, with a stroke at every 10th base). The bars denote base pairs (A-U, G-C and G-U not being distinguished). The helices are numbered, and are discussed individually in the text. Differences in base-pairing between this model and the american (1) and french (2) models are indicated by the boxed-in regions, dotted lines indicating minor differences and solid lines major differences (see text). Arrows (e.g. interactions 2, 38) denote helices proposed in the other models (1,2).

pairing in the comparing sequence is achieved by displacing one of the strands by one base.

THE MODEL OF 16S RNA

Figure 1 shows our previous model of 16S RNA (3) in skeleton form, and indicates the major and minor discrepancies in relation to the american (1) or french (2) versions (cf. ref. 8). Minor discrepancies we define as extra or modified base-pairings at the proximal ends of hairpin loops, or other small differences which would not seriously influence the overall three-dimensional topography of the molecule. Major discrepancies on the other hand are those which would affect the overall topography. Figure 2 shows the refined version of the structure, which has for graphical convenience been condensed into two rather than three domains; the numbering of the helical elements is however identical to that in Figure 1. In Figure 3, some examples are shown of the actual sequence comparisons on which our structural assignments are based. (It should be noted that the position numbers given for the comparing sequence regions in Figure 3 correspond to those in the respective publications (Table 1)).

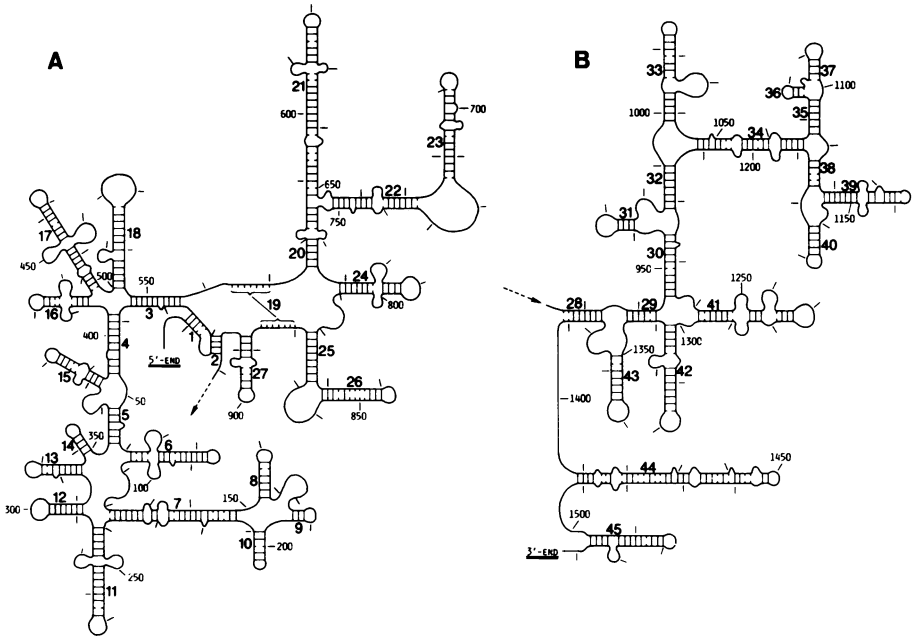


Figure 2: The revised secondary structure model of 16S RNA. The sequence is divided into two domains (A and B), and the bases are numbered as in Fig. 1. Solid bars denote G-C or A-U base pairs, broken bars G-U pairs. The helix numbers correspond to those of Fig. 1; Helix 19 is a special case (see text).

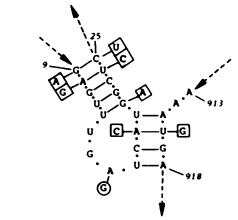
HELIX 1: The german interaction (absent in the other models) is confirmed. CBC's are found in all 12S and 18S RNA's, and also in H. volcanii and S. cerevisiae mitochondrion (see Fig. 3 for example).

HELIX 2: The american interaction (absent in the other models) is confirmed. CBC's are found in many organisms (see Fig. 3 for example). Helices 1 and 2 are both shown in Fig. 2A, but model-building studies will be required to establish to what extent they can co-exist. Opening up of the loop-end proximal base pair in helix 1 (usually a G-U pair) would probably be sufficient to allow helix 2 to form.

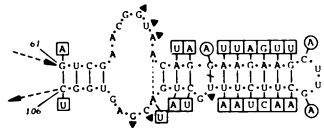
HELIX 3 is undisputed and re-confirmed.

HELIX 4: The upper 6 base pairs are undisputed and re-confirmed. The lower 3 base pairs (in the german and american versions) are also confirmed by CBC's in many organisms. It is noteworthy however that the status of this interaction in 18S RNA's is unclear (cf. helix 19, below).

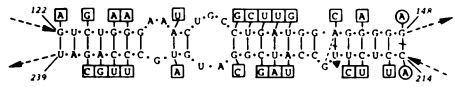
HELIX 5: The newly-described american interaction (33) is included in Fig. 2A, but is not fully confirmed. The upper 3 base pairs (Fig. 2A) are confirmed by



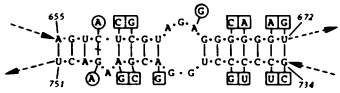
1,2 (HUMAN MITOCHONDRION)



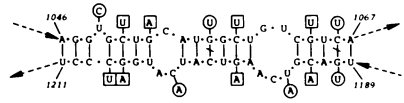
6 (A. NIDULANS MITOCHONDRION)



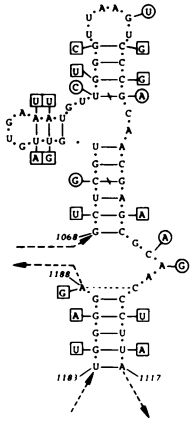
7 (E. GRACILIS CHLOROPLAST)



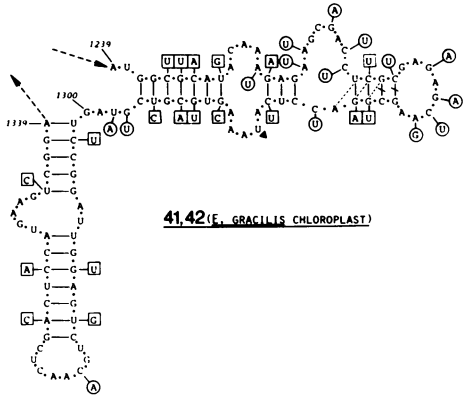
22 (C. REINHARDII CHLOROPLAST)



34 (S. CEREVISIAE MITOCHONDRION)



35-38 (S. CEREVISIAE CYTOPLASM)



41,42 (E. GRACILIS CHLOROPLAST)

Figure 3: Examples of structure comparisons in 16S RNA. The diagrams show the *E. coli* helices (numbered as in Fig. 2), with changes in the comparing sequence (indicated in each case by name, cf. Table 1) being denoted by the bases in boxes. Base changes in square boxes are compensating, those in round boxes are mis-matching or in single-stranded regions. Solid triangles denote deletions, and dotted lines or "crossed-out" base pairs show modified base-pairing in the comparing sequence. Arrows denote the 5'-3' direction of the RNA chains and the numbers are those of the *E. coli* sequence. The corresponding regions of the comparing sequences are: HELICES 1,2, human mitochondrion bases 4-20, 519-524; HELIX 6, *Asp. nidulans* bases 61-100; HELIX 7, *E. gracilis* bases 114-139, 210-234; HELIX 22, *C. reinhardii* bases 608-625, 687-704; HELIX 34, *S. cerevisiae* mitochondrion bases 1541-1562, 1655-1677;HELICES 35-38, *S. cerevisiae* cytoplasm bases 1276-1320; HELICES 41,42, *E. gracilis* bases 1193-1290.

CBC's in H. volcanii and 18S RNA's, but the lower 3 pairs are not yet established, as the primary sequence is universally conserved. This interaction involves the same bases as the german helix 14 (Fig. 1a), but the latter is not included in the revised model (see below).

HELIX 6: The german version is confirmed. The proximal 10 base-pairs are confirmed in many organisms (although the loop end is partially deleted in some chloroplasts and in Ana. nidulans), and the distal 4 base pairs are confirmed by CBS's in Asp. nidulans and S. cerevisiae mitochondria (see Fig. 3 for example).

HELIX 7: A modified version of the french/german model is given in Fig. 2A. In this preferred version the left-hand 16 base pairs (including the two additional pairs 131-132/230-231) are definitively confirmed by CBC's in all organisms, whereas the right-hand pairs are not always present. Several organisms have lost up to three pairs from the extreme right-hand side, and H. volcanii has lost all six. The example in Fig. 3 however is one of several which support the version shown in Fig. 2A (cf. helix 34, below).

HELIX 8 is undisputed and re-confirmed.

HELIX 9: The german interaction (absent in the other models) is preferred. This short helix cannot universally be formed, but is nevertheless present with CBC's in Ana. nidulans, C. reinhardii and N. tabacum. In addition, the helix can be formed with one- or two-base slippage in some other organisms (e.g. Z. mays chloroplast (3)).

HELIX 10: The french/german version is included in Fig. 2A, but is by no means certain. The evidence for the interaction is "negative", namely that the loop is clearly deleted in Z. mays and N. tabacum chloroplasts and in Ana. nidulans (cf. helix 17 below). On the other hand, no universal loop can be drawn here for other sequences, and this appears to be a very variable region.

HELIX 11 is undisputed and re-confirmed.

HELIX 12 is undisputed and re-confirmed.

HELIX 13: The distal 5 base pairs are undisputed and reconfirmed. The extra 3 proximal pairs (Fig. 2A) in the french version are included, as they can be formed with an occasional CBC in most (but not all) organisms.

HELIX 14: The french version of this helix is confirmed (Fig. 2A). It can be drawn with CBC's in all organisms, although in several cases a mis-matched pair is also introduced.

HELIX 15: The distal 4 base pairs are undisputed and re-confirmed. The proximal 5 base pairs are also re-confirmed by CBC's, but it is noteworthy that

the bases corresponding to 386 and 387 in E. coli are deleted in Ana. nidulans and two of the chloroplast sequences.

HELIX 16: The distal 4 base pairs are undisputed and re-confirmed. The french version is preferred for the proximal 4 pairs (Fig. 2A), primarily on the basis of CBC's in Ana. nidulans.

HELIX 17: The distal 9 base pairs are undisputed and reconfirmed. The proximal 8 pairs are in all the models (1-3), but are not confirmed by CBC's. The evidence for their existence is negative (cf. helix 10 above), namely that these base pairs are clearly deleted in all chloroplasts, and in Ana. nidulans and H. volcanii.

HELIX 18 is undisputed and re-confirmed.

HELIX 19 is in all the models (1-3), but we have already pointed out the uncertainty of this interaction (3,8). In fact the sequences concerned are either conserved, or else contain base substitutions which destabilize the interaction (e.g. in the latest russian model for S. cerevisiae 18S RNA (34), the base-pairing is reduced to 6 G-U pairs, 1 A-U, and 1 C-C mismatch). Only Asp. nidulans shows convincing CBC's (15), but even here one base must be looped out. This topographically very important interaction has therefore been "opened up" in Fig. 2A, to emphasize the lack of confirmatory evidence.

HELIX 20 is undisputed and re-confirmed, although an occasional mis-match occurs in addition to many CBC's.

HELIX 21 is essentially undisputed and re-confirmed, the extra 2 base pairs in the german version (Fig. 1b) being confirmed by CBC's in Ana. nidulans and the four chloroplast RNA's.

HELIX 22: The slightly different base-pairing scheme in the french/german models is confirmed by many CBC's in all organisms (see Fig. 3 for example).

HELIX 23: The distal 8 base pairs are undisputed and re-confirmed. The 5 proximal pairs in the german version are also preferred (Fig. 2A), being supported by CBC's in Asp. nidulans, H. volcanii and 18S RNA's, with a one- or two-base slippage in the latter two cases. An occasional mis-match is observed in some other organisms however.

HELIX 24 is undisputed and re-confirmed.

HELIX 25 is undisputed and re-confirmed, although it is noteworthy that an alternative interaction for the bases in X. laevis corresponding to 872-879 has been experimentally observed (35).

HELIX 26 is undisputed and re-confirmed.

HELIX 27: The proximal 4 base pairs are undisputed and re-confirmed. The 4 distal pairs (present in all models (1-3) are also included in Fig. 2A,

but are in fact almost universally conserved in primary sequence and therefore unconfirmed. One pair of CBC's occurs in H. volcanii, but in contrast there are 2 mis-matches in the 12S RNA's (3).

HELIX 28 is undisputed and re-confirmed, although sometimes a one-base slippage occurs, and an occasional mis-match.

HELIX 29 is undisputed and re-confirmed.

HELIX 30: The lower 8 base pairs (Fig. 1c) are undisputed and re-confirmed, and, in addition, the two extra base pairs in the french/american versions (954-955/1225-1226, Fig. 2B) are confirmed by CBC's in S. cerevisiae mitochondrion, Asp. nidulans, C. reinhardii, and in 18S RNA's.

HELIX 31: The extra 2 base pairs in helix 30 automatically result in the loss of the distal 4 base pairs of helix 31 in the german model (Fig. 1c). The proximal 4 pairs are in all the models (1-3) and in Fig. 2B, and (although the primary sequence is almost universally conserved) are confirmed by CBC's in H. volcanii and mammalian mitochondrial RNA.

HELIX 32 is undisputed and re-confirmed. N. tabacum has one base looped out.

HELIX 33 is undisputed and re-confirmed, although this is by no means a universal loop in 16S RNA's. Some species (most notably N. tabacum, C. reinhardii and Ana. nidulans) show many CBC's, whereas others show a progressive deletion of the loop, and in Asp. nidulans it has disappeared entirely.

HELIX 34: Various base-pairings can be made here, and the best fit is the version shown in Fig. 2B (see Fig. 3 for example). In this, the left-hand 7 base pairs of the german version (Fig. 1c) are confirmed by CBC's in several organisms, and the german/american base-pairing is confirmed for the central group of 6 base pairs by CBC's in 18S RNA, although an occasional mis-match occurs here. On the other hand the french version gives the best CBC's for the right-hand 5 base pairs. Here again mis-matches sometimes occur, resulting in a loss of the right-hand 2 pairs (cf. helix 7, above), but these are usually recovered elsewhere (e.g. in the example in Fig. 3, the base pairs corresponding to 1066-1067/1189-1190 are lost in the yeast mitochondrion, but can be replaced by pairing the bases corresponding to 1061-1062/1194-1195).

HELIX 35 is undisputed and re-confirmed.

HELIX 36 : This interaction (only in the american model) is confirmed by CBC's in almost all organisms (see Fig. 3 for example of helices 35-38).

HELIX 37: This interaction (only in the german model, Fig. 1c) is confirmed by CBC's in several organisms, although an occasional mis-match or one-base slippage is observed.

HELIX 38: This interaction, (only in the french model, but slipped along by one base (cf. 2)), is confirmed by CBC's in almost all organisms.
HELIX 39: The 7 distal pairs are undisputed and re-confirmed. The proximal 6 base pairs are not confirmed by CBC's, as the end of the loop varies in size from one species to another and often cannot be directly superimposed on the E. coli structure. These 6 pairs are however preferred, as almost all sequences are able to form similar base-pairings right up to the end of the individual loops.
HELIX 40 is undisputed and re-confirmed.
HELIX 41: The distal 7 base pairs in Fig. 2B are undisputed and re-confirmed, the 2 extra distal pairs in the german version (Fig. 1c) being removed on the basis of mis-matches in most organisms (see Fig. 3 for example). The central 3 base pairs of the german model (1253-1255/1282-1284) are confirmed by several CBC's in the four chloroplast sequences, and the proximal 4 pairs of the american/german models are preferred, as they can be formed (sometimes with one-base slippage (cf. Fig. 3)) in most organisms.
HELIX 42: The proximal 7 base-pairs are undisputed and reconfirmed, whereas the distal 5 pairs in the german version (Fig. 1c) are anomalous. These pairs are retained in Fig. 2B, on the basis of CBC's in several organisms, including H. volcanii and the chloroplast RNA's (see Fig. 3 for example), but other organisms show mis-matches.
HELIX 43: The french-american version of this loop is confirmed. The extra base pairs in the german model (Fig. 1c) are not supported; the sequences are either fully conserved, or else mis-matches occur.
HELIX 44: This long and rather irregular helix can be drawn for all organisms although the helix cannot always be superimposed upon that of E. coli. The 3 extra base pairs near the distal end of the helix in the french version (cf. Figs. 1c and 2B) are confirmed by occasional CBC's, but the disputed base-pairing in the centre of the helix (Fig. 1c) cannot yet in our opinion be resolved, as a number of mis-matches occur in this region in several organisms. The extra 2 proximal base pairs in the german model are however preferred, on the basis of a pair of CBC's in E. gracilis, and similar loop ends in almost all organisms.
HELIX 45: The proximal 9 base pairs of this loop are undisputed and re-confirmed. The distal 4 pairs are also retained in Fig. 2B, on the grounds that they can be formed in many 16S RNA species (cf. also ref. 36).

THE MODEL OF 23S RNA

The original secondary structure model of 23S RNA (4) is given in skele-

ton form in Fig. 4, with the discrepancies between this model and the french (5) and american (6) versions being indicated as in Fig. 1. The refined structure is shown in Fig. 5, and this has been condensed from six into four domains, for graphical convenience as with the 16S structure in Fig. 2. Some examples of the actual structure comparisons are shown in Fig. 6 (cf. Fig. 3), and Fig. 7 illustrates some anomalous features of the model, discussed below. In the eukaryotic 25-28S RNA molecules, it should be remembered that the 5'--region of the structure, corresponding to the first 160 bases of E. coli 23S RNA, is formed by the 5.8S RNA (37, 4, 30). Similarly, the 3'-region of the structure in the chloroplast RNA's is formed by the 4.5S RNA (38, 4).

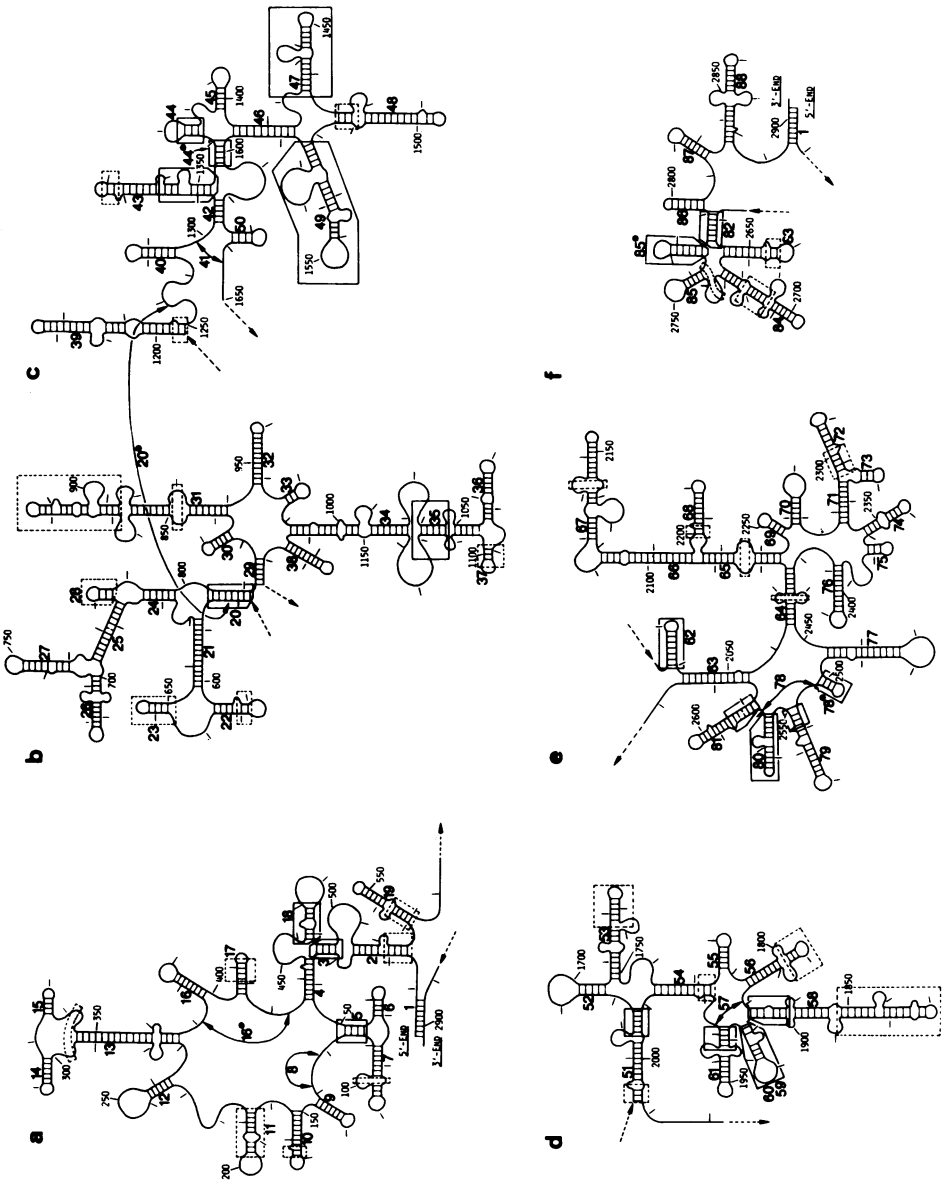
HELIX 1: This is in all the models (4-6), but in fact the only CBC's so far observed are in Z. mays or N. Tabacum chloroplasts. The interaction cannot exist in eukaryotes, since the 5'-end of the 5.8S RNA corresponds to position 12 of the E. coli 23S RNA (cf. 30). Additional evidence for the existence of the helix in E. coli is that it forms part of a long proposed helical element in the precursor RNA molecule (39).

HELIX 2 is anomalous, since in E. coli the sequence GUAC occurs twice in tandem at positions 520-523 and 524-527. In Z. mays CBC's are only observed if the second of these GUAC sequences is paired, whereas in eukaryotes the first must be used (see Fig. 7). In N. tabacum, the homology to Z. mays shows clearly that the 3 looped-out bases corresponding to positions 521-524 are deleted, leading to a continuous helix. It is therefore not clear which GUAC sequence is involved in the pairing in E. coli, and our version is retained in Fig. 5A in order to draw attention to this anomaly; it may well be, however, that the Z. mays sequence is the exceptional one here.

HELIX 3 is also anomalous, and may represent a genuine "switch" situation. CBC's are observed in most organisms both in the german version and in the french/american loop involving the same sequence (bases 484-496, see Fig. 7 for example). Again, we retain our version in Fig. 5A. (It should be noted that we were unable to correlate P. primaurelia and S. cerevisiae mitochondrial RNA satisfactorily with E. coli in this part of the structure, as was the case with the mammalian 16S mitochondrial RNA's (4)).

HELIX 4 is undisputed and re-confirmed.

HELIX 5: The german/french version is preferred, as the helix can be formed with one-base slippage and CBC's in the chloroplasts and in S. cerevisiae mitochondrion. (The corresponding helix in 5.8S RNA (cf. 37) is however not directly superimposable on that of E. coli, and therefore this helix, and also helix 7 below, are given "preferred" rather than "confirmed" status).



HELIX 6 is undisputed and re-confirmed.

HELIX 7: The german version is preferred. This differs from the other versions only in the absence of base pairs 83-84/99-100 (cf. Fig. 4a), which are mismatched in Z. mays chloroplast, whereas CBC's are observed in both the distal and proximal regions of the helix. The corresponding helix in 5.8S RNA's is however not superimposable (cf. helix 5 above).

HELIX 8: The french/american version (see Fig. 5A) is confirmed, being supported by CBC's in 5.8S RNA's.

HELIX 9 is undisputed (except that the terminal distal base pair in the german version (Fig. 4a) is sacrificed by the formation of helix 8). Direct evidence from CBC's is however not available, as the helix is very variable in length, and is deleted in the chloroplast RNA's (cf. helices 10 and 17 in 16S RNA).

HElIX 10: The loop end corresponds to the 3'-end of 5.8S RNA and the 5'-end of 25-28S RNA, and the helix is also very variable in length. In 23S RNA, the identity of the helix is confirmed by many CBC's in Z. mays, and the slightly different german version for the two proximal base pairs (Fig. 4a, 5A) is preferred for the same reason.

HELIX 11: The german/american version for the 5 distal base pairs is confirmed by CBC's in all organisms. The 3 proximal pairs can also be drawn in all organisms except for S. cerevisiae cytoplasmic RNA, where a mis-match occurs.

HELIX 12 is undisputed and re-confirmed.

HELIX 13: The french/american version, with 3 extra base pairs at the top end of the helix (cf. Figs. 4a, 5A), is preferred, as similar structures can always be drawn. The helix is variable however, and cannot be described in terms of CBC's. (S. cerevisiae 25S RNA can be arranged in a structure very like that of Z. mays 23S (4), in contrast to the model proposed in ref. 30).

HELIX 14 is undisputed and re-confirmed.

HELIX 15 is essentially undisputed, although the american/french interaction is slipped along by one base. Both versions are supported by CBC's, and cannot be distinguished.

HELIX 16* is proposed in the french and american models (see Fig. 4a), but

Figure 4: The previous secondary structure model of *E. coli* 23S RNA (4). The sequence is divided into six domains (a to f), and is numbered as in Fig. 1. Helix numbers, and differences between this model and the french (5) or american (6) models are indicated as in Fig. 1. Helices marked with an asterisk (e.g. 44*) are those which have no counterpart in the revised structure.

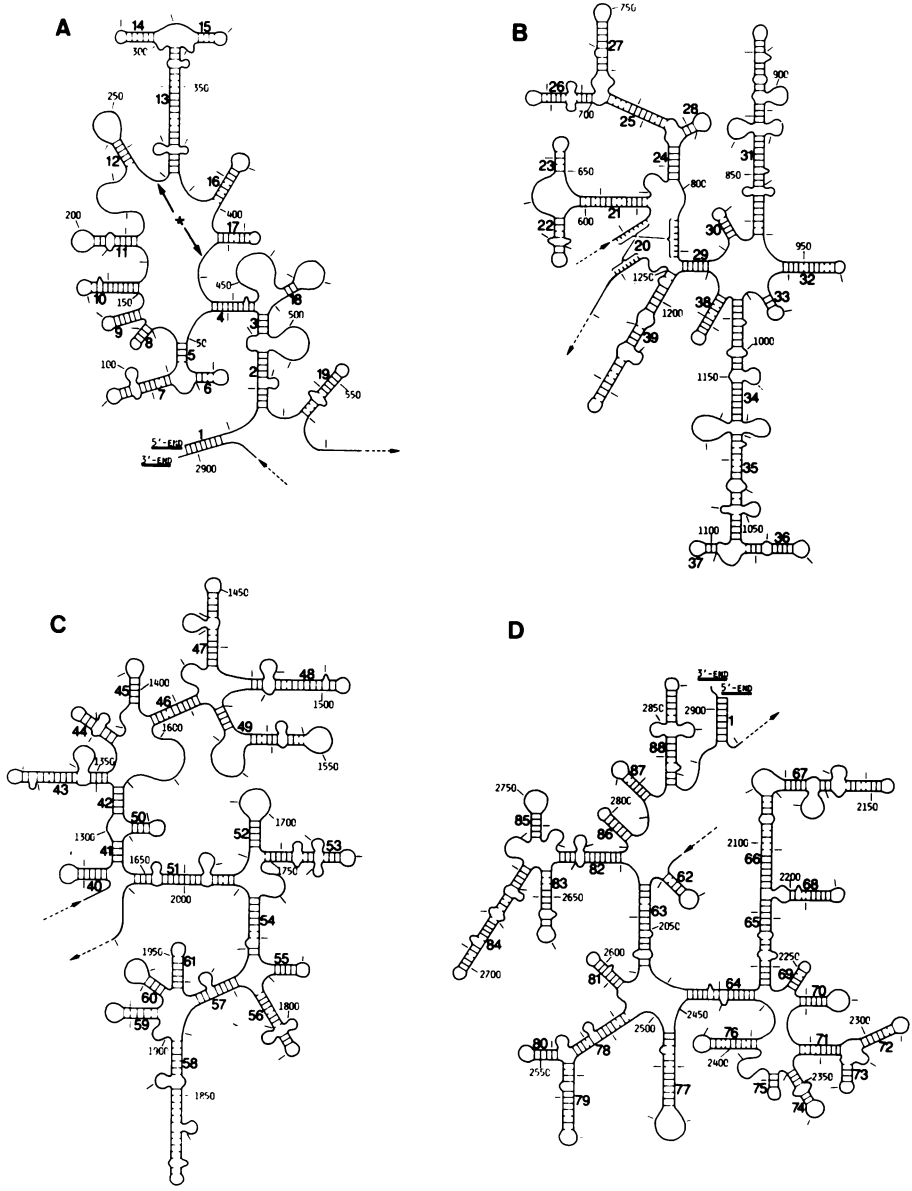


Figure 5: The revised secondary structure model of 23S RNA. The sequence is divided into four domains (A to D), and the helix numbers correspond to those of Fig. 4. Base pairs are denoted as in Fig. 2. The arrows and asterisk joining positions 267 and 426 (domain A) indicate an interaction which we observed while this manuscript was in the final stage of preparation (cf. Helix 16* in Fig. 4a, and see text). Helix 20 is a special case (see text).

the primary sequence is either conserved in other organisms, or else leads to mismatched pairs with no CBC's. The helix is not included in Fig. 5A. Instead an interaction can be drawn between bases 265-268 and 424-427 (see Fig.5A), which is confirmed by CBC's in S. cerevisiae cytoplasmic and mitochondrial RNA's.

HELIX 16 is undisputed and re-confirmed.

HELIX 17: The german/french version is confirmed by many CBC's in all organisms.

HELIX 18 is only in the german model (Fig. 4a). Although some mis-matches occur, a short helix can always be formed (sometimes with one- or two-base slippage) here. Fig. 5A shows a preferred minimal version of this helix. Note however that the "switched" version of helix 3 (Fig. 7, and see above) involves the same bases.

HELIX 19: The proximal 6 base pairs are undisputed and re-confirmed. The french version for the distal 4 pairs (Fig. 5A) is confirmed by CBC's in S. cerevisiae and P. primaurelia mitochondria.

HELIX 20 is anomalous. In the german/french version bases 578-584 are paired with 805-811 (see Fig. 4b). This structure shows CBC's in other organisms, including mammalian mitochondrial RNA (4), but only in one strand (see Fig. 7 for example). The american model on the other hand (HELIX 20*, Fig. 4b) pairs bases 579-585 with 1255-1261 (Fig. 7), and other organisms have CBC's in both strands. This latter pairing can however only be formed in mammalian mitochondrial RNA at the expense of the highly conserved helix 40 (see below). The cytoplasmic 25-28S RNA's contain a large insertion (200 bases or more) in the region corresponding to bases 560-585 in E. coli, and therefore do not offer an unambiguous clarification of this problem. Both possible versions of this topographically very important interaction are indicated "opened up" in Fig. 5B (cf. helix 19 in 16S RNA).

HELIX 21 is undisputed and re-confirmed.

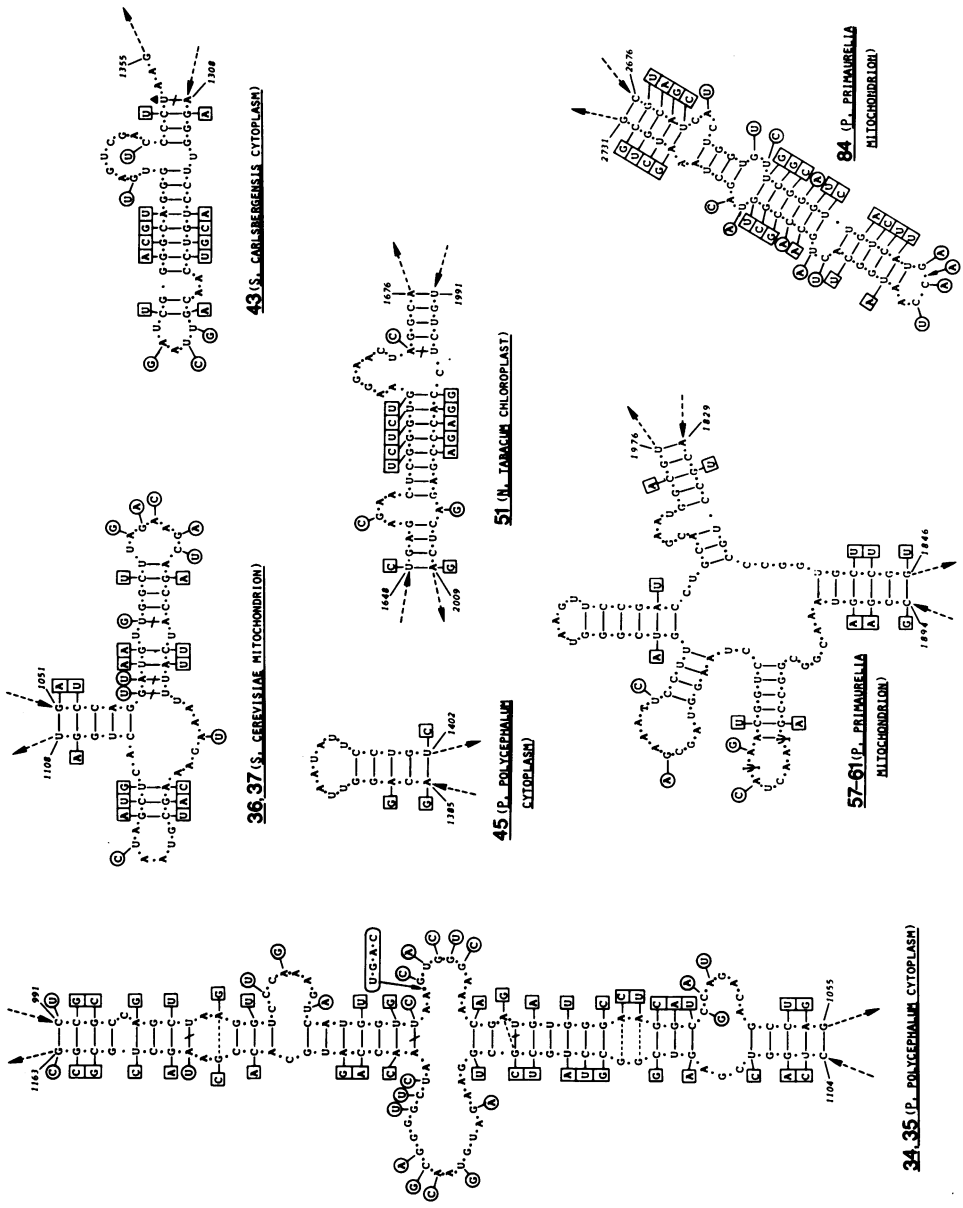
HELIX 22 is variable in length, the slightly different french/american version (cf. Figs. 4b, 5B) being preferred for 23S RNA, as it shows CBC's between E. coli and the chloroplast RNA's.

HELIX 23: The german/french version is preferred on the basis of several CBC's in S. cerevisiae mitochondrion. The longer RNA molecules however contain insertions in this region which obscure the issue.

HELIX 24 is undisputed and re-confirmed.

HELIX 25 is undisputed and re-confirmed.

HELIX 26 is undisputed and re-confirmed. The proximal 6 base pairs are deleted in all mitochondria.


43 (S. CARLSBERGENSIS CYTOPLASM)
36,37 (S. CEREVISIAE MITOCHONDRION)
45 (P. POLYCEPHALUM CYTOPLASM)
51 (N. TABACUM CHLOROPLAST)
84 (P. PRIMAURELIA MITOCHONDRION)
57-61 (P. PRIMAURELIA MITOCHONDRION)
34,35 (P. POLYCEPHALUM CYTOPLASM)

HELIX 27 is undisputed and re-confirmed.

HELIX 28: The french version is preferred (Fig. 5B), on the basis of CBC's in S. cerevisiae and P. polycephalum cytoplasmic RNA's. S. cerevisiae mitochondrion on the other hand has a mis-matched pair in this version.

HELIX 29 is undisputed and re-confirmed.

HELIX 30 is undisputed and re-confirmed.

HELIX 31: The distal 7 base pairs (838-844/934-940) are undisputed and re--confirmed. The proximal region of this long loop is however rather variable, and difficult to describe universally in terms of CBC's. The preferred version shown in Fig. 5B is essentially the german version (Fig. 4b) with extra base pairs (848-850/928-930) from the american/french versions, but is by no means confirmed.

HELIX 32 is undisputed and re-confirmed, although in addition to many CBC's an occasional mis-match is observed.

HELIX 33 was in all three models (4-6), but without comparative evidence, as the primary sequence is generally highly conserved. The helix is now confirmed by CBC's in S. cerevisiae and P. polycephalum cytoplasmic RNA's.

HELIX 34, extending as far as the large looped-out regions (positions 1020 and 1140), is undisputed and re-confirmed.

HELIX 35: The american version is confirmed by many CBC's in all organisms, and two additional base pairs (1030-1031/1123-1124) are also confirmed (cf. Figs. 4b, 5B). Fig. 6 shows a typical example of helices 34 and 35, in which a total of no less than 40 CBC's can be observed.

HELIX 36 is undisputed, and many CBC's are observed here, although mis-matches are also quite common (see Fig. 6 for example).

HELIX 37 is only in the german model, but is clearly confirmed by CBC's in S. cerevisiae mitochondrion (see Fig. 6). In other species the sequence is in general conserved.

HELIX 38 is undisputed and re-confirmed, although the loop is of variable length in some species.

HELIX 39 is undisputed and re-confirmed, with the exception of the two extreme distal base pairs in the german/french versions (Fig. 4c). These two

Figure 6: Examples of structure comparisons in 23S RNA. Symbols are explained in the legend to Fig. 3; in addition, boxed-in bases with an arrow denote insertions in the comparing sequence. The regions of the comparing sequences concerned are: HELICES 34,35,*P. polycephalum* bases 1216-1284, 1333-1392; HELICES 36,37, *S. cerevisiae* mitochondrion bases 978-1035; HELIX 43, *S. carlsbergensis* cytoplasm bases 1488-1534; HELIX 45, *P. polycephalum* bases 1725-1742; HELIX 51, *N. tabacum* bases 1680-1708, 2000-2018; HELICES 57-61, *P. primaurelia* bases 1337-1354, 1395-1477; HELIX 84, *P. primaurelia* bases 2147-2203.

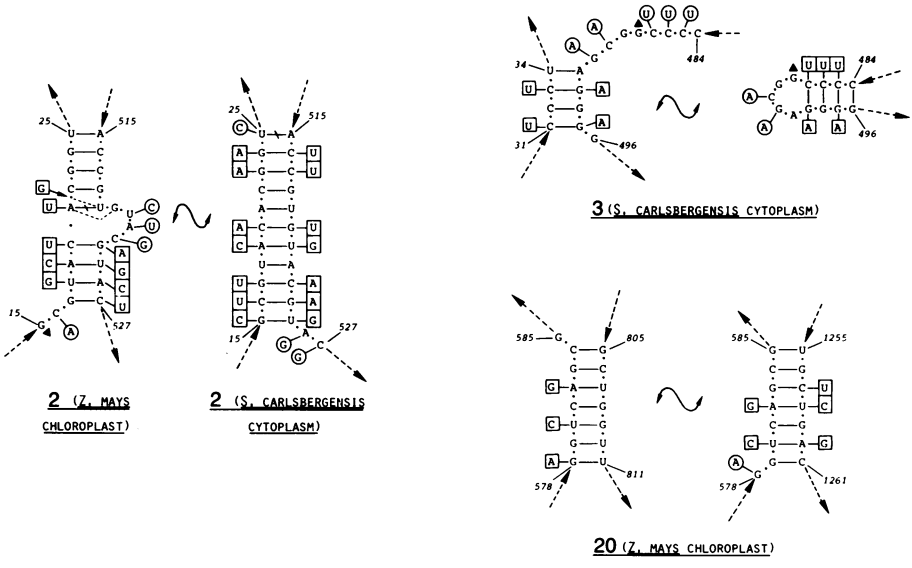


Figure 7: Some unresolved anomalies in the 23S RNA structure. In each example a pair of alternative structures is shown. See Fig. 3 for explanation of symbols. The regions of the comparing sequences concerned are: HELIX 2, *Z. mays* bases 16-26, 529-541, and *S. carlsbergensis* 5.8S RNA bases 4-14 and 25S RNA bases 407-419 (30); HELIX 3, *S. carlsbergensis* 5.8S RNA bases 20-23, and 25S RNA bases 377-388; HELIX 20, *Z. mays* bases 595-602, 828-834, 1286-1292. See text for further information.

pairs are included in Fig. 5B on the basis of a CBC in Z. mays (4), but are not further confirmed. Note also that the relatively large number of mismatches in the proximal region of the helix in Z. mays (4) is not typical; most organisms show perfect CBC's here.

HELIX 40 is undisputed and re-confirmed.

HELIX 41: The french/american interaction is confirmed by CBC's in most organisms, with occasional one-base slippage (cf. Figs. 4c, 5C).

HELIX 42 is undisputed and re-confirmed.

HELIX 43: A slightly modified version of the german model (cf. Figs. 4c, 5C) is confirmed by CBC's in several organisms. This is particularly clear in S. carlsbergensis (Fig. 6), (although it should be noted that in ref. 30 a different structure is proposed for this region of the yeast RNA, which is not superimposable on the E. coli structure).

HELIX 44* was one of two alternative proposals for this region in our original model (4). No support for this interaction has been found, and it is not included in Fig. 5C.

HELIX 44: The modified version of helix 44 (Fig. 4c), which we suggested as an alternative loop (4), is tentatively included in Fig. 5C. All organisms can form a similar loop here, but the region is very variable, and the loops cannot be described in terms of CBC's.

HELIX 45 is in all the models (4-6), but previously without confirmatory evidence as the primary sequence is highly conserved. This loop is now confirmed by CBC's in P. polycephalum (Fig. 6) and in S. cerevisiae (with one-base slippage, again differing from the proposal of ref. 30 (cf. helix 43)).

HELIX 46 is undisputed and re-confirmed.

HELICES 47-49 comprise the most uncertain region of the 23S RNA. Helix 48 is in all the models, but cannot be regarded as confirmed, as the corresponding loops in other organisms cannot be satisfactorily described in terms of CBC's in a common structure. The other two models differ in the lengths and positions of helices 47 and 49 (see ref. 8 for comparison). In our opinion the evidence marginally favours our structure (in particular in Z. mays and S. cerevisiae), and for this reason our version is retained in Fig. 5C.

HELIX 50, is undisputed and re-confirmed. (P. primaurelia however shows a one-base slippage.)

HELIX 51: The american version for the left-hand 4 base pairs of the interaction (Fig. 5C) is confirmed by CBC's in most organisms, and the central 8 base pairs are undisputed and re-confirmed (see Fig. 6 for example). The german/french version of the right-hand 5 pairs is also included in Fig. 5C, although the sequence is too highly conserved here for CBC's to be observed. The alternative american proposal (in which bases 1755-1759 are paired to 1991-1995) is rather clearly discounted by mis-matching base changes in the region corresponding to positions 1755-1759 in several organisms.

HELIX 52 is undisputed and re-confirmed.

HELIX 53 is a very variable loop. Eukaryotic cytoplasmic RNA's have a large insertion here (over 100 bases), whereas in other organisms (e.g. Z. mays (4)) the loop is partially deleted. The helix cannot be described by CBC's, and the french/american version of the loop (cf. Figs. 4d, 5C) is preferred, on the basis of a reproducible cobra venom nuclease cut in the proximal part of the helix (ref. 5, and C. Zwieb, personal communication).

HELIX 54: The upper 7 base pairs are undisputed and re-confirmed, and the lower 2 pairs in the german/american versions are confirmed by CBC's in several organisms.

HELIX 55 is in all the models (4-6) but is not confirmed, as the primary sequence is totally conserved except for the residue corresponding to position

1778, which occasionally shows a mis-matching base change.

HELIX 56: The distal 8 base pairs are undisputed and re-confirmed, and the german/american version of the proximal 3 base pairs is confirmed by CBC's in all organisms.

HELIX 57 is in the french and american models, and is formed at the expense of the distal pairs of helices 58 and 61 in our model (Fig. 4d). The french/american version is confirmed by CBC's in many organisms, although an occasional mis-match or looped-out base is also observed, and we suggest that additional stabilization of the helix could be achieved by pairing bases 1834-1836 with 1964-1966 (see Figs. 5C and 6). These bases are however conserved in all organisms, except for rat mitochondrion (cf. helix 5 in 16S RNA).

HELIX 58: The distal 6 pairs (Fig. 4d) are removed to form helix 57. The next 6 pairs (bases 1841-1846/1894-1899) are undisputed and re-confirmed (see Fig. 6 for example). The distal region of the loop is however variable and cannot be described in terms of CBC's. Our version is retained in Fig. 5C on the grounds that all sequences can form similar base-pairings, right up to the end of the loop (cf. helix 39 in 16S RNA).

HELICES 59, 60: The french/american two-helix version gives better CBC's than the german one-helix version (cf. Figs. 4d, 5C, and see Fig. 6 for example). The primary sequence is however very highly conserved in this region, and the CBC's are only observed in helix 59 (Fig. 5C), but not in helix 60. The latter helix therefore remains unconfirmed.

HELIX 61: The distal 4 base pairs in the german version (Fig. 4d) are removed to form helix 57 (see above), and the remaining 6 proximal pairs are undisputed and re-confirmed (see Fig. 6 for example).

HELIX 62: We have already pointed out (40) that we overlooked this helix in our original model (4). The french/american version is shown in Fig. 4e, and is reconfirmed with the exception that the 2 proximal base pairs (both G-U in E. coli) are mis-matched in several organisms, and are therefore omitted in Fig. 5D.

HELIX 63: The inclusion of helix 62 means that 63 is now undisputed and also reconfirmed. In addition the extra 2 base pairs at the lower end of the american version (2057-2058/2611-2612 (Fig. 5D)) are confirmed by CBC's in eukaryotic cytoplasmic and mammalian mitochondrial RNA's.

HELIX 64: The extra 2 base pairs in the centre of the helix (2069-2070/2441-2442) in the american version are confirmed by CBC's in several organisms, the remainder of the helix being undisputed and re-confirmed.

HELIX 65: Here again the extra base pairs (2083-2085/2234-2236 (Fig. 5D)) in

the american version are confirmed by CBC's in several organisms, sometimes with one-base slippage. Otherwise the helix is undisputed and re-confirmed.
HELIX 66 is undisputed and re-confirmed.
HELIX 67 is undisputed with the exception of 2 extra base pairs (2130-2131/2157-2158 (Fig. 4e)) in the american version. The 8 proximal base pairs of the loop are however deleted in the more recently determined sequences, so that the pairing cannot be followed beyond positions 2128/2160. In Z. mays the extra american pairs are both mis-matched (4), and are therefore not included in Fig. 5D. The remaining base pairs in this helix were already confirmed by the previous data.
HELIX 68: The 7 proximal base pairs are undisputed and re-confirmed. The german version, with extra base-pairs and one looped-out base, is confirmed for the distal portion of the helix by CBC's in several organisms. Eukaryotic cytoplasmic RNA's have a large insertion within or adjacent to this helix.
HELIX 69 is undisputed and re-confirmed.
HELIX 70 is undisputed and re-confirmed, although the helix is slightly longer in S. cerevisiae, and P. polycephalum shows a one-base slippage.
HELIX 71 is undisputed and re-confirmed.
HELIX 72: The slightly different base-pairing scheme for the 7 proximal pairs in the french/american version is confirmed by CBC's in several organisms. In P. primaurelia the loop is extended by 4 base pairs.
HELIX 73: The loop is in all the models (4-6), but the data are contradictory. P. primaurelia shows CBC's, whereas P. polycephalum and S. cerevisiae have mis-matches. The loop is tentatively retained in Fig. 5D.
HELIX 74 is undisputed and re-confirmed, sometimes with one-base slippage, with the exception that the proximal G-U pair (Fig. 4e) has been removed from Fig. 5D (cf. helix 62).
HELIX 75 is undisputed and re-confirmed, sometimes with one-base slippage.
HELIX 76 is undisputed, and can be drawn with many CBC's and an occasional mis-match in most organisms. The base-pairing appears to be somewhat different (with a four-base slippage) in eukaryotic cytoplasmic RNA's, but nevertheless can be regarded as confirmed for the 23S RNA.
HELIX 77 is undisputed and re-confirmed, although as an exception the 3 distal base-pairs (Fig. 5D) are mis-matched in eukaryotic cytoplasmic RNA's.
HELICES 78-81: The structure in our model (4) was proposed as an alternative to the original Baer and Dubin version for this region (41). The new sequence data do not support our model, and the french/american version (which is essentially that of ref. 41) is confirmed by CBC's in all organisms, with an

occasional mis-match. As a result, helices 78* and 80, the distal region of helix 81, and the extreme distal base pair of helix 79 (Fig. 4e) disappear, and are replaced by helix 78 and a new version of helix 80 (Fig. 5D). The remainder of helix 79 is undisputed and re-confirmed (with the exception that the looped-out residue at position 2546 (Fig. 4e) is removed by a one-base slippage), and the proximal 7 base pairs of helix 81 are also undisputed and re-confirmed.

HELIX 82: The version in our original model (4) is not supported by the new sequence data. The french/american version (Fig. 5D) is confirmed by CBC's in Z. mays and S. cerevisiae. As a result of introducing this interaction, helix 85* (Fig. 4f) disappears.

HELIX 83 is undisputed with the exception of the extra 2 proximal base pairs in the german model (Fig. 4f). These pairs are preferred (Fig. 5D) as they can be drawn in all organisms, sometimes with a one-base slippage. The remaining 7 base pairs of the helix are re-confirmed by many CBC's in all organisms.

HELIX 84: The modified base-pairing of the french version is confirmed by CBC's in several organisms, most strikingly in P. primaurelia (see Fig. 6).

HELIX 85: The german/american version is confirmed by CBC's in S. cerevisiae and P. polycephalum, the extra distal base pairs in the french version (cf. Fig. 4f) not being supported by CBC's.

HELIX 85* is not in Fig. 5D (see helix 82 above).

HELIX 86 is in all three models (4-6), but is not confirmed. It contains the junction between 23S RNA and 4.5S RNA in chloroplasts (4), and therefore cannot be described in terms of CBC's with respect to Z. mays or N. tabacum. On the other hand we have been unable to correlate the other available sequences unambiguously with that of E. coli in this region (or further towards the 3'-end), as a result of major insertions or deletions. The helix is nonetheless reasonable (cf. helix 10, Fig. 5A), and is included in Fig. 5D.

HELICES 87, 88 are undisputed, and satisfactorily confirmed by the older data (4-6).

CONCLUSIONS

The secondary structure models presented here should serve two purposes. Firstly, we hope that they will have largely resolved the confusion which has inevitably surrounded the publication of three similar but not identical models for these large molecules (1-3, 4-6), all presented in different formats (cf. 8). Secondly, by objectively considering each model helix by helix in the light of the latest sequence data, the refined versions which we have

developed are without doubt a much closer approximation to the actual structures. Of the 133 helices described in the 16S and 23S molecules (Figs. 2 and 4), approximately 15% are still tentative, unconfirmed, or apparently species specific. A further 15% are helices which definitely exist, but in which possible prolongations of the helices or the detailed base-pairing at the proximal ends of hairpin loops are still open to discussion. Most important is that the remaining 70% of the helices are now definitively established, at least to within one or two base pairs.

The occurence of mis-matched pairs still needs to be clarified. A glance at Figs. 3 and 6 shows that virtually all possible mis-matched pairs do occasionally occur in the comparing sequences, and it follows that mis-matching bases which are drawn as loop-outs in the E. coli structure may in fact be incorporated in the helices (cf.42,43,8). Occasional sequencing errors will also contribute to this problem, since it is inconceivable that in over 50,000 sequenced bases no errors have been incorporated.

The further application of the comparative sequencing approach to the solution of the 16S and 23S structures will now take on a role of diminishing importance, since most of the uncertainties in the models are in regions where the sequences are either too highly variable or too highly conserved for the approach to be useful. The next obvious step in our opinion will be to try to develop a first crude approximation to the three-dimensional topography of the molecules, by combining the secondary structures with experimental data obtained directly from the ribosomal subunits, such as intra-RNA cross-linking (40) or exposure to chemical modification (1,6). Such model-building studies are currently in progress in our laboratory.

ACKNOWLEDGEMENTS

The authors are grateful to Dr. H.G. Wittmann for his continued support.

REFERENCES

1. Noller, H.F. and Woese, C.R. (1981) Science 212, 403-411
2. Stiegler, P., Carbon, P., Zuker, M., Ebel, J.P. and Ehresmann, C. (1981) Nucleic Acids Res. 9, 2153-2172
3. Zwieb, C., Glotz, C. and Brimacombe, R. (1981) Nucleic Acids Res. 9, 3621-3640
4. Glotz, C., Zwieb, C., Brimacombe, R., Edwards, K. and Kössel, H. (1981) Nucleic Acids Res. 9, 3287-3306
5. Branlant, C., Krol, A., Machatt, M.A., Pouyet, J., Ebel, J.P., Edwards, K. and Kössel, H. (1981) Nucleic Acids Res. 9, 4303-4324
6. Noller, H.F., Kop, J., Wheaton, V., Brosius, J., Gutell, R.R., Kopylov, A.M., Dohme, F., Herr, W., Stahl, D.A., Gupta, R. and Woese, C.R. (1981) Nucleic Acids Res. 9, 6167-6189

7. Fox, G. and Woese, C.R. (1975) Nature, 256, 505-507
8. Brimacombe, R., Maly, P. and Zwieb, C. (1983) Progr. in Nucleic Acids Res. 28, 1-48
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. Kobayashi, M., Seki, T., Yaginuma, K. and Koike, K. (1981) Gene, 16, 297-307
12. Saccone, C., Cantatore, P., Gadaleta, G., Gallerani, R., Lanave, C., Pepe, G. and Kroon, A.M. (1981) Nucleic Acids Res. 9, 4139-4148
13. Sor, F. and Fukuhara, H. (1980) C.R. Acad. Sci. Paris, D. 291, 933-936
14. Sor, F. and Fukuhara, H. (1983) Nucleic Acids Res. 11, 339-348
15. Köchel, H.G. and Küntzel, H. (1981) Nucleic Acids Res. 9, 5689-5696
16. Seilhammer, J. and Cummings, D.J. (1981) Nucleic Acids Res. 9, 6391-6406
17. Brosius, J., Palmer, M.L., Kennedy, P.J. and Noller, H.F. (1978) Proc. Nat. Acad. Sci. USA, 75, 4801-4805
18. Brosius, J., Dull, T.J. and Noller, H.F. (1980) Proc. Nat. Acad. Sci. USA, 77, 201-204
19. Carbon, P., Ebel, J.P. and Ehresmann, C. (1981) Nucleic Acids Res. 9, 2325-2333
20. Gupta, R., Lanter, J.M. and Woese, C.R. (1983) Science, 221, 656-659
21. Tomioka, N. and Sugiura, M. (1983) Mol. Gen. Genet. 191, 46-50
22. Schwarz, Z. and Kössel, H. (1980) Nature, 283, 739-742
23. Edwards, K. and Kössel, H. (1981) Nucleic Acids Res. 9, 2853-2869
24. Graf, L., Roux, E., Stutz, E. and Kössel, H. (1982) Nucleic Acids Res. 10, 6369-6381
25. Dron, M., Rahire, M. and Rochaix, J.D. (1982) Nucleic Acids Res. 10, 7609-7620
26. Tohdoh, N. and Sugiura, M. (1982) Gene, 17, 213-218
27. Takaiwa, F. and Sugiura, M. (1982) Eur. J. Biochem. 124, 13-19
28. Rubtsov, P.M., Musakhanov, M.M., Zakharyev, V.M., Krayev, A.S., Skryabin, K.G. and Bayev, A.A. (1980) Nucleic Acids Res. 8, 5779-5794
29. Georgiev, O.I., Nikolaev, N., Hadjiolov, A.A., Skryabin, K.G., Zakharyev, V.M. and Bayev, A.A. (1981) Nucleic Acids Res. 9, 6953-6958
30. Veldman, G.M., Klootwijk, J., de Regt, V.C.H.F., Planta, R.J., Branlant, C., Krol, A. and Ebel, J.P. (1981) Nucleic Acids Res. 9, 6935-6952
31. Salim, M. and Maden, B.E.H. (1981) Nature 291, 205-208
32. Otsuka, T., Nomiyama, H., Yoshida, H., Kukita, T., Kuhara, S. and Sakaki, Y. (1983) Proc. Nat. Acad. Sci. USA, 80, 3163-3167
33. Brow, D.A. and Noller, H.F. (1983) J. Mol. Biol. 163, 27-46
34. Mankin, A.S., Kopylov, A.M. and Bogdanov, A.A. (1981) FEBS Lett. 134, 11-14
35. Atmadja, J., Brimacombe, R. and Maden, B.E.H., manuscript in preparation
36. Wickstrom, E. (1983) Nucleic Acids Res. 11, 2035-2052
37. Nazar, R.N., (1980) FEBS Lett. 119, 212-214
38. Mackay, R.M. (1981) FEBS Lett. 123, 17-18
39. Bram, R.J., Young, R.A. and Steitz, J.A. (1980) Cell, 19, 393-401
40. Stiege, W., Glotz, C. and Brimacombe, R. (1983) Nucleic Acids Res. 11, 1687-1706
41. Baer, R.J. and Dubin, D.T. (1981) Nucleic Acids Res. 9, 323-337
42. De Bruijn, M.H.L. and Klug, A. (1983) EMBO J. 2, 1309-1321
43. Stahl, D.A., Luehrsen, K.R., Woese, C.R. and Pace, N.R. (1981) Nucleic Acids Res. 9, 6129-6137

Note added in proof.

Correction to HELIX 17 in 16S RNA:

The proximal 8 base pairs are confirmed, by CBC's in P. vulgaris (19).