*Proc. Int. Symp. Biomol. Struct. Interactions, Suppl. J. Biosci.*,
Vol. 8, Nos 3 & 4, August 1985, pp. 791–798. © Printed in India.

# Prediction of secondary structures of 16S and 23S rRNA from *Escherichia coli*

A. S. KOLASKAR*, T. A. THANARAJ and M. W. PANDIT
Centre for Cellular and Molecular Biology, Hyderabad 500 007, India

**Abstract.** Small and large subunits of *Escherichia coli* ribosome have three different rRNAs, the sequences of which are known. However, attempts by three groups to predict secondary structures of 16S and 23S rRNAs have certain common limitations namely, these structures are predicted assuming no interactions among various domains of the molecule and only 40% residues are involved in base pairing as against the experimental observation of 60% residues in base paired state. Recent experimental studies have shown that there is a specific interaction between naked 16S and 23S rRNA molecules. This is significant because we have observed that the regions (oligonucleotides of length 9–10 residues), in 16S rRNA which are complementary to those in 23S rRNA do not have internal complementary sequences. Therefore, we have developed a simple graph theoretical approach to predict secondary structures of 16S and 23S rRNAs. Our method for model building not only uses complete sequence of 16S or 23S rRNA molecule along with other experimental observations but also takes into account the observation that specific recognition is possible through the complementary sequences between 16S and 23S rRNA molecules and, therefore, these parts of the molecules are not used for internal base pairing. The method used to predict secondary structures is discussed. A typical secondary structure of the complex between 16S and 23S rRNA molecules, obtained using our method, is presented and compared briefly with earlier model building studies.

**Keywords.** RNA secondary structure; 16S–23S RNA complex; nucleic acid–nucleic acid interaction.

## Introduction

The knowledge of the three dimensional structure of the ribosomal proteins and the rRNA molecules will help in understanding the mechanism of interactions between these constituent molecules and, in turn, will throw light on the function of ribosomes. A first step towards this goal would be to study the secondary structure of small as well as large naked ribonucleic acid molecules which would help in predicting three dimensional structures of these molecules and the sites of interactions between proteins and nucleic acids.

A good model system, on which already a large number of experimental studies (Fox and Woese, 1975) and some theoretical studies have been carried out towards the prediction of secondary structure, is *Escherichia coli* ribosome. The primary structure of both 16S and 23S rRNA molecules are known (Brosius *et al.*, 1978, 1980). Attempts have been made by three different groups to predict the secondary structures of rRNAs (Noller and Woese, 1981; Stiegler *et al.*, 1981; Zwieb *et al.*, 1981; Glotz *et al.*, 1981; Branlant *et al.*, 1981; Noller *et al.*, 1981) and these secondary structures have been

* To whom correspondence should be addressed.

refined recently (Maly and Brimacombe, 1983). Inspite of the fact that all the three groups have used almost similar strategy to predict the secondary structure, their results are not in complete agreement with each other. When compared, several parts of the secondary structures of 16S and 23S rRNA are quite different. One of the possible reasons for these differences seems to be the over-emphasis on the sequence homology and the identification of helices based on such sequence homology. It is known from the studies on protein structures that very different structures are possible for the same primary sequence; on the other hand entirely different sequence of amino acids can give fairly similar three dimensional structure (Novotny *et al.*, 1984). Therefore, we feel that invarience in certain regions of the sequence cannot be taken as a proof for the existence of double helical regions in the structure. Further, recent experimental studies have revealed specific interactions between 16S and 23S rRNA molecules suggesting that in the assembly of ribosomal particle, nucleic acid – nucleic acid interactions are primary and nucleic acid–protein interactions are secondary (Burma *et al.*, 1983). The effect of such interactions on the secondary structure of individual rRNAs has not been considered in any of the earlier studies on the prediction of structures. Apart from this, a major drawback of all previous studies is the assumption that there are no interactions through complementary base pairing between one domain of the nucleic acid structure and the other domain. Such assumption is required to make the computations on secondary structure relatively simpler, as one has to handle only about 600 nucleotides at any given time as against total of 1542 nucleotides of 16S rRNA and 2904 nucleotides of 23S rRNA of *E. coli*. To overcome these lacunae and to take into account the experimental observations, we have adopted a different approach by using modified algorithm originally suggested by Nussinov and Pieczenik (1984). Our basic approach and the algorithm is briefly discussed below. Secondary structure thus obtained from the complex of 16S and 23S rRNAs of *E. coli* is also discussed.

## Basic approach used to predict secondary structures of 16S and 23S rRNAs

It has been shown experimentally that 16S and 23S rRNA molecules interact with each other specifically and form energetically stable complex. Such specific interactions are possible if and only if there are sufficiently long stretches of oligonucleotides in 16S rRNA which are complementary to the stretches in 23S rRNA, and these *inter*molecular complementary regions are not involved in *intra*molecular complementation. We have, therefore, carried out a search for *intra*- as well as *inter*molecular complementary subsequences in 16S and 23S rRNA of *E. coli*. These search studies revealed that, even at the nanonucleotide level, there are a good number of *intra*- as well as *inter*molecular complementary sequences of 16S and 23S rRNAs as given in table 1. These long *inter*molecular complementary regions can interact and form stable double helical conformations. We, therefore, suggest that in all probability complementary stretches given in table 1 are the regions which interact specifically during complex formation of 16S and 23S rRNA molecule. Further our approach treats each of the 16S and 23S rRNA molecules as total entity without taking into consideration various domains in these molecules as reported by earlier workers. However, we hope that if the approach described here is reasonably correct, the calculated secondary structures will consequentially lead to the experimentally observed domains also.

**Table 1.** *Intra-* and *inter*molecular complementation in rRNA molecules of *E.coli* at deca- and nanonucleotide level.

| Length | Complementary sequences | | |
|---|---|---|---|
| | 16S–16S | 23S–23S | 16S–23S |
| 10 | — | (16–25) — (524–515)<br>(71–80) — (1176–1167)<br>(260–269) — (2652–2643) | — |
| 9 | (392–400) — (1061–1053)<br>(413–421) — (422–414)<br>(430–438) — (439–431)<br>(863–871) — (872–864) | (88–96) — (2480–2472)<br>(95–103) — (1923–1915)<br>(246–254) — (1299–1291)<br>(519–527) — (528–520)<br>(561–569) — (2440–2432)<br>(728–736) — (2611–2603)<br>(792–800) — (1468–1460)<br>(1052–1060) — (2524–2516)<br>(1074–1082) — (1985–1977)<br>(1505–1513) — (2109–2101)<br>(1590–1598) — (2195–2187) | (307–315) — (775–767)<br>(630–638) — (2039–2031)<br>(653–661) — (2009–2001)<br>(783–791) — (2303–2295)<br>(891–899)* — (1912–1904)*<br>(1156–1164) — (41–33)<br>(1195–1203)* — (768–760)*<br>(1204–1212) — (2419–2411)<br>(1404–1412) — (1831–1823)<br>(1444–1452) — (1854–1846)<br>(1464–1472) — (2572–2564) |

* These stretches of oligonucleotides are shown as an example of intermolecular contacts envisaged from our search for *inter*molecular complementary sequences in 16S and 23S rRNA molecules.

Recently, an algorithm based on graph theory has been suggested by Nussinov and Pieczenik (1984) to study the folding of a large nucleic acid molecule. We have modified this approach so that one can use micro-computer of PDP 11/23 type, which has a limited memory and speed. Experimental studies using bisulphite, kethoxal, glyoxal, and *m*-chloroperbenzoate as reagents have given a very valuable information regarding those C, G and A residues which lie in single-stranded regions. Constraints arising out of this information were placed during the computer model building studies.

## Folding algorithm

**Step 1:** *Intra*molecular complementary nucleotide subsequences of length $\geqslant 5$ are searched in 16S and 23S rRNAs of *E. coli*. Such complementary subsequences are stored along with their location $I$, from 5′-end of the parent sequence.

**Step 2:** Complementary sequences are arranged in two ways; (i) in ascending order of separation between complementary stretches, and (ii) in ascending order of location.

**Step 3:** A closed loop is formed by choosing maximally separated complementary sequences $I, J$ where $I$ and $J$ are locations of the first nucleotide in each of the two complementary subsequences from 5′-end.

**Step 4:** Loop thus formed is bisected such that $|I' - J'| \simeq L/2$, where $|I - J| = L$

and $I$ and $J$ are the locations of the first nucleotides in the new complementary subsequences.

**Step 5:** The bisection of large loops into smaller ones are continued till $L \simeq 100$, using the same constraints mentioned in Step 4.

**Step 6:** Constraints arising out of *inter*molecular complementation are imposed. Thus, for example, in a typical structure (figure 1 and table 1), subsequences 891–899 and 1195–1203 in 16S rRNA and the corresponding complementary subsequences 1912–1904 and 768–760 in 23S rRNA are disallowed from forming any *intra*molecular base pairing. However, no such condition was imposed if single nucleotide in the *inter*molecular base pairs were already involved in *intra*molecular base pairing.

**Step 7:** Similar constraints are imposed by disallowing base pair formation through those C, U, A and G bases which, from experimental studies, were found to be in single-stranded regions.

**Step 8:** In a given loop, maximal base pairing is carried out using a procedure described by Nussinov and Jacobson (1980) but disallowing the base pairing between complementary neighbours, as such base pair formation, if allowed, will exert considerable conformational strain. In the Nussinov and Jacobson procedure a matrix $M(i,j)$ of maximum base pairing is obtained by:

$$M(i,j) = \max \begin{cases} M(i,k-1) + M(k+1,j-1) + 1 \\ M(i,j-1) \end{cases} \qquad i \leqslant k < j$$

where $i$ and $j$ are the locations of the nucleotides which are complementary to each other in a loop under consideration and vary in the limit $I$ to $J$ (which close that loop) and $k$ varies between $i$ and $j$ taking all possible integral values.

**Step 9:** Matrix $K(i,j)$ was also constructed where $K(i,j) = k$ corresponding to $M(i,j)$ and contains numerical position of base $B_k$ that allows maximal base pairing within each segment $B_i \ldots . B_j$ which form closed loop.

**Step 10:** The initial loops obtained are now bisected in perpendicular direction when compared to the bisection carried out in Step 4.

**Step 11:** Step 5 to 9 are repeated.

## Results and discussion

In the present study several structures of 16S and 23S rRNAs were obtained. One of the many such structures for 16S–23S rRNA complex is shown in figure 1. Various loops are indicated by numbers and the nucleotide stretches in corresponding loops are enlisted in table 2. It can be seen from figure 1 that there are 28 loops in 16S rRNA and 33 loops in 23S rRNA. Each of these loops, however, consists of several residues which are base paired. Details of such base pairing, for example, is given in figure 2 which is a small part (loops 3, 4, 27 and 28 in 16S rRNA) of figure 1. It can be seen from figure 2

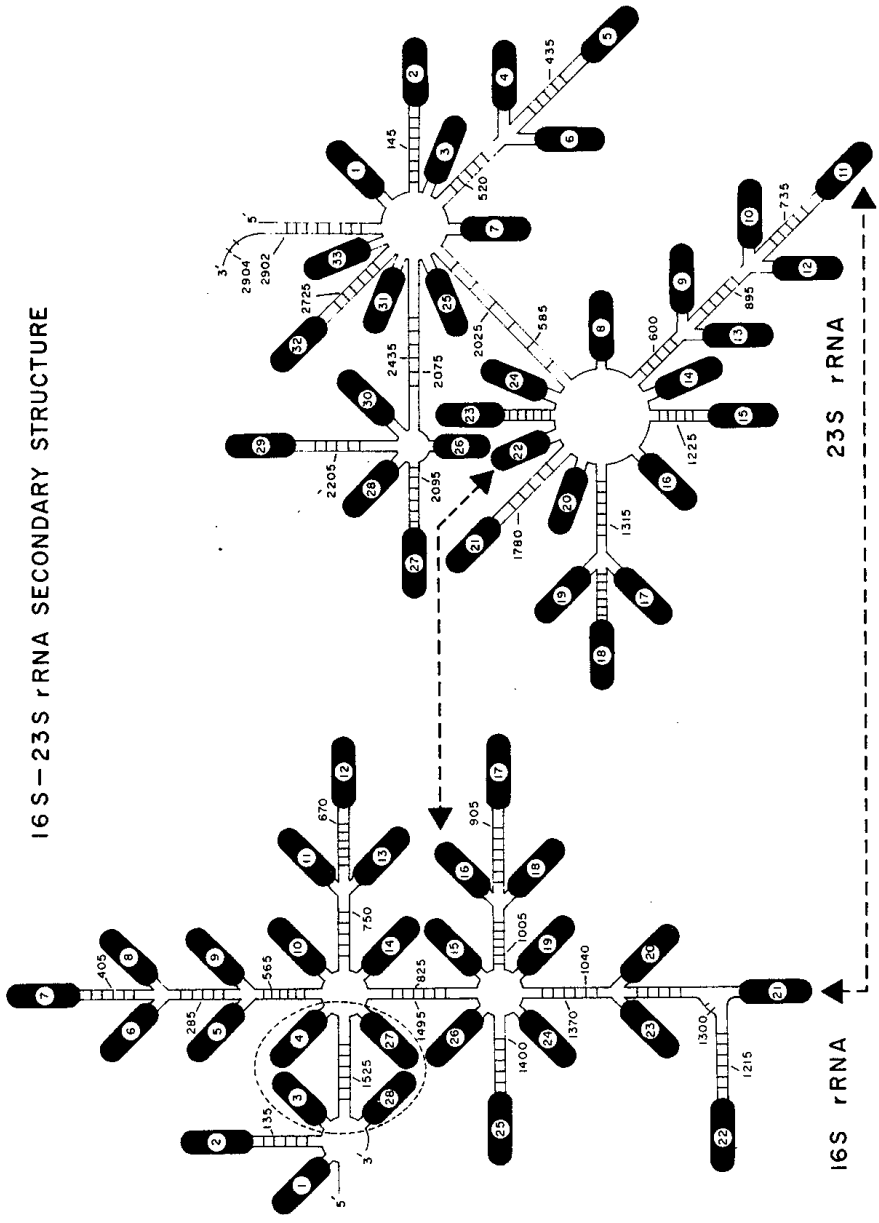


**Figure 1.** Secondary structure of the complex of 16S and 23S rRNA. Circled numbers in each of the individual loops refer to the corresponding loop in table 2. Dotted lines ending with arrows indicate the *inter*molecular nanonucleotide base pairing. Refer to figure 2 for the part of the secondary structure of 16S rRNA enclosed in a dotted enclosure. Note that energy minimization is carried out in each loop such that maximal base pairs are formed within the loop.

**Table 2.** Subsequences involved in the loops in 16S and 23S rRNAs as indicated in figure 1.

| Loop No. | 16S rRNA | 23S rRNA |
|---|---|---|
| 1 | 1–42 | 9–141 |
| 2 | 48–133 | 149–281 |
| 3 | 139–173* | 289–381 |
| 4 | 180–231* | 387–431 |
| 5 | 238–282 | 437–494 |
| 6 | 288–326 | 500–516 |
| 7 | 332–402 | 522–580 |
| 8 | 408–451 | 587–597 |
| 9 | 457–563 | 603–711 |
| 10 | 570–596 | 717–732 |
| 11 | 602–664 | 738–807 |
| 12 | 671–734 | 813–893 |
| 13 | 741–749 | 899–954 |
| 14 | 755–822 | 960–1099 |
| 15 | 829–852 | 1105–1222 |
| 16 | 859–899 | 1228–1310 |
| 17 | 906–976 | 1317–1395 |
| 18 | 983–1000 | 1402–1564 |
| 19 | 1007–1035 | 1571–1664 |
| 20 | 1042–1134 | 1671–1775 |
| 21 | 1140–1211 | 1781–1897 |
| 22 | 1218–1293 | 1903–1945 |
| 23 | 1307–1366 | 1952–1997 |
| 24 | 1373–1399 | 2004–2021 |
| 25 | 1405–1456 | 2028–2070 |
| 26 | 1462–1491 | 2077–2093 |
| 27 | 1498–1521* | 2100–2182 |
| 28 | 1528–1542* | 2189–2202 |
| 29 | — | 2208–2381 |
| 30 | — | 2387–2432 |
| 31 | — | 2439–2558 |
| 32 | — | 2565–2723 |
| 33 | — | 2730–2894 |

* The secondary structure of these loops is indicated in figure 2.

that each of the loops contain varied number of bases and they are maximally hydrogen bonded to give required energy stabilization. It should be mentioned here that we have not yet incorporated base stacking energy term in our energy minimization algorithm; however, we hope to be able to do this in near future.

The points of contact between 16S and 23S rRNA as envisaged through our search for *inter*molecular complementary subsequences are given in table 1. As an example of *inter*molecular contacts, the nanonucleotides from 16th and 21st loops of 16S rRNA which are predicted to be interacting with 22nd and 11th loops of 23S rRNA, respectively, are shown in figure 1. These interactions would necessitate the tertiary folding of corresponding structural units to bring them in the vicinity of each other.

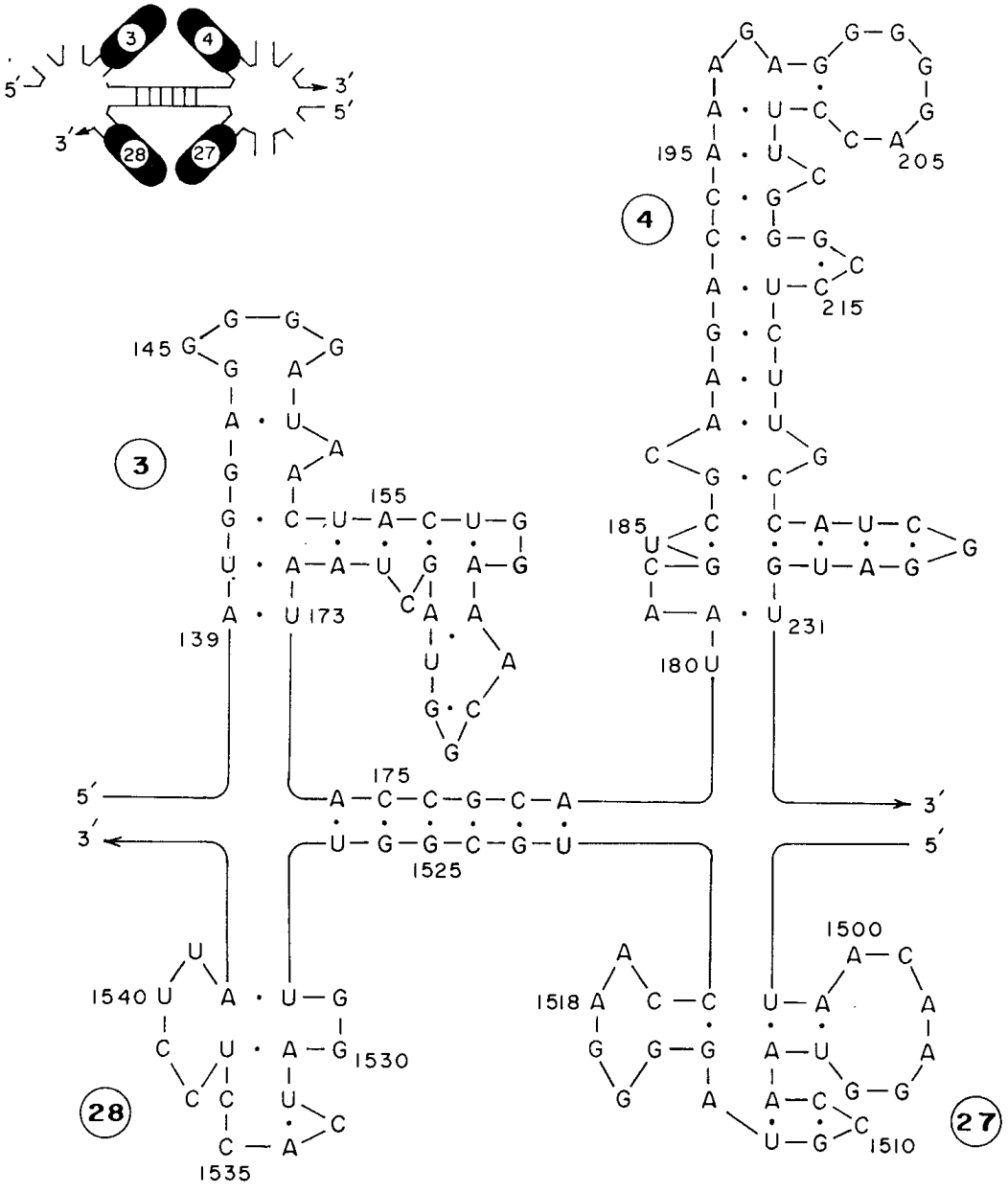


**Figure 2.** The secondary structure of the part of 16S rRNA consisting of loops 3, 4, 27 and 28, showing the details of maximal base pairing.

Information obtained through the experiments on the crosslinking between various regions of 16S and 23S rRNA could be incorporated at this stage in order to choose the meaningful structure of the complex between 16S and 23S rRNA molecules. It would be interesting to note that in a secondary structure such as shown in figure 1, 65 % of the bases are involved in the base pair formation. The extent of base pairing observed in several experimental studies is also of the same order. In fact, to our knowledge this is the first report in which the secondary structure with such an extensive base pairing has been observed.

These studies, therefore, clearly point out that a simple approach discussed above gives a meaningful secondary structures which are automatically subdivided into domains. The interactions between such domains can then make total structure rigid and energetically stable to a certain extent. As observed in earlier studies there are hardly any double helical regions in which nucleotides of length more than 8–9 are base paired. Detailed comparison of our structure with other structures will be discussed elsewhere. However, we would like to stress that our structure will be different from those reported earlier mainly because we have, for the first time, incorporated (1) long range *inter*domain and *inter*molecular interactions, and (2) the constraints arising out of the experimental observations even before the maximization of base pairing.

We, therefore, now have a workable algorithm which when used judiciously, particularly after the incorporation of experimental data, can give a realistic structure of even large nucleic acid molecules. In addition, our algorithm can be handled by a small computer and thus certainly has an advantage over other ones.

## References

Branlant, C., Krol, A., Machatt, M. A., Pouyet, J., Ebel, J. P., Edwards, K. and Kossel, H. (1981) *Nucleic Acids Res.*, **9**, 4303.

Brosius, J., Dull, T. J. and Noller, H. F. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 201.

Brosius, J., Palmer, M. L., Kennedy, P. J. and Noller, H. F. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 4801.

Burma, D. P., Nag, B. and Tewari, D. S. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 4875.

Fox, G. and Woese, C. R. (1975) *Nature (London)*, **256**, 505.

Glotz, C., Zwieb, C., Brimacombe, R., Edwards, K. and Kossel, H. (1981) *Nucleic Acids, Res.*, **9**, 3287.

Maly, P. and Brimacombe, R. (1983) *Nucleic Acids Res.*, **11**, 7263.

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.

Noller, H. F. and Woese, C. R. (1981) *Science*, **212**, 403.

Novotny, J., Bruccoleri, R. and Karplus, M. (1984) *J. Mol. Biol.*, **177**, 787.

Nussinov, R. and Jacobson, A. B. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 6309.

Nussinov, R. and Pieczenik, G. (1984) *J. Theor. Biol.*, **106**, 245.

Stiegler, P., Carbon, P., Zuker, M., Ebel, J. P. and Ehresmann, C. (1981) *Nucleic Acids Res.*, **9**, 2153.

Zwieb, C., Glotz, C. and Brimacombe, R. (1981) *Nucleic Acids Res.*, **9**, 3621.