Volume 10 Number 8 1982 Nucleic Acids Research

Adenine-guanine base pairing in ribosomal RNA

Wolfie Traub and Joel L.Sussman

Structural Chemistry, Weizmann Institute of Science, Rehovot, Israel

Received 26 January 1982; Revised and Accepted 25 March 1982

SUMMARY

Analyses of secondary structures proposed for ribosomal RNA's show that, of the different kinds of base pairs directly adjoining the ends of postulated double-helical regions, only A-G with A at the 5' end significantly exceeds the number expected for a random base distribution. An A(syn)-G(trans) hydrogen-bonded basepair is proposed. This could fit at the end of an undistorted double helix, but would prevent further base stacking, thus favoring a break in the double helix to produce a non-linear tertiary structure.

INTRODUCTION

Some 15 years ago it was suggested that RNA' may contain a small proportion of adenine-guanine base pairs with the adenine moeity in the syn conformation(1). Model building showed that, in such an A-G base pair, the separation and orientation of the glycosyl bonds are closely similar to those of the Watson-Crick G-C and A-U base pairs (2), but that incorporation in a double-helical structure would imply short contacts between atoms of adenine and the sugar of the adjacent nucleotide in the 5' direction (1). Hence, it was proposed that an A-G pair could fit onto a double-helical region of RNA, without distortion of the phosphate-sugar backbone, but only with adenine at a 5' end of the helical sequence.

Recently, secondary structures have been proposed for several ribosomal RNA's from different organisms (3-7). These structures have been inferred from a variety of chemical, enzymatic and computational studies, including isolation of double-helical nuclease-resistant fragments, modification by single-strand specific reagents, comparison of homologous sequences for conservation of base-pairing, and the use of algorithms to search for conformations with maximum base-pairing. However, in the absence of direct structure determinations by X-ray diffraction, these postulated structures may very well not be completely correct, particularly regarding the exact ends of helical

0305-1048/82/1008-2701$2.00/0

regions (6), short stretches of double helix, and the incorporation of various unusual base interactions such as have been found in tRNA (8,9).

Probably the best studied ribosomal RNA is that of E. coli, particularly that of the small subunit, for which closely similar structures have been proposed by three different groups (3,4,5). We have analysed the most recently published structures for this 16S rRNA (5) and for the 23S rRNA of E. coli (6) for possible sites of the proposed A-G base pairs. We noted all the different kinds of base pairs directly adjoining the ends of the postulated double-helical regions, allowing at least two bases at terminal loops (Fig 1). In a few cases these were U-G, G-C or A-U pairs, which were not shown as hydrogen-bonded in the proposed structures, so we included these in the double-helical regions and noted the next base pair. We then compared the numbers we had counted for the various end pairs with those expected if the bases in the postulated single-stranded regions were randomly distributed (Table I).

RESULTS

It is clear from Table I that for both 16S and 23S E. coli ribosomal RNA there are significantly more potential A-G base pairing sites than would be expected for a random distribution of bases in single-stranded regions. They occur at one quarter of the helix ends in 16S rRNA and about one third in 23S rRNA. None of the nine other non-standard base pairs appear to be significantly favored, and it is particularly noteworthy that A-G sites greatly outnumber G-A sites, which are statistically equivalent for a random distribution. We would also emphasise that A-G pairing was not among the criteria taken into account for predicting these RNA secondary structures, and that these could no doubt be modified to increase the number of A-G pairs.

The A-G sites are quite evenly distributed among terminal-loop, middle-exclusion, and open-end portions of double-helical regions (illustrated in Fig. 1). However, as shown in Table II, there appears to be a marked preference for the adenine of the A-G pair to adjoin the purine of the last base pair of the double helix.

We have also examined secondary structures proposed for several other ribosomal subunits. The 16S rDNA from Z. mays chloroplasts shows 74% homology with the E. coli 16S rRNA (10), and indeed the secondary structure proposed for this Z. mays rRNA (5) has 25 potential A-G base-pairing sites, compared with 14 G-A, only one less in each case than the analogous structure proposed for the E. coli 16S rRNA (Table I). The structure proposed for Z.

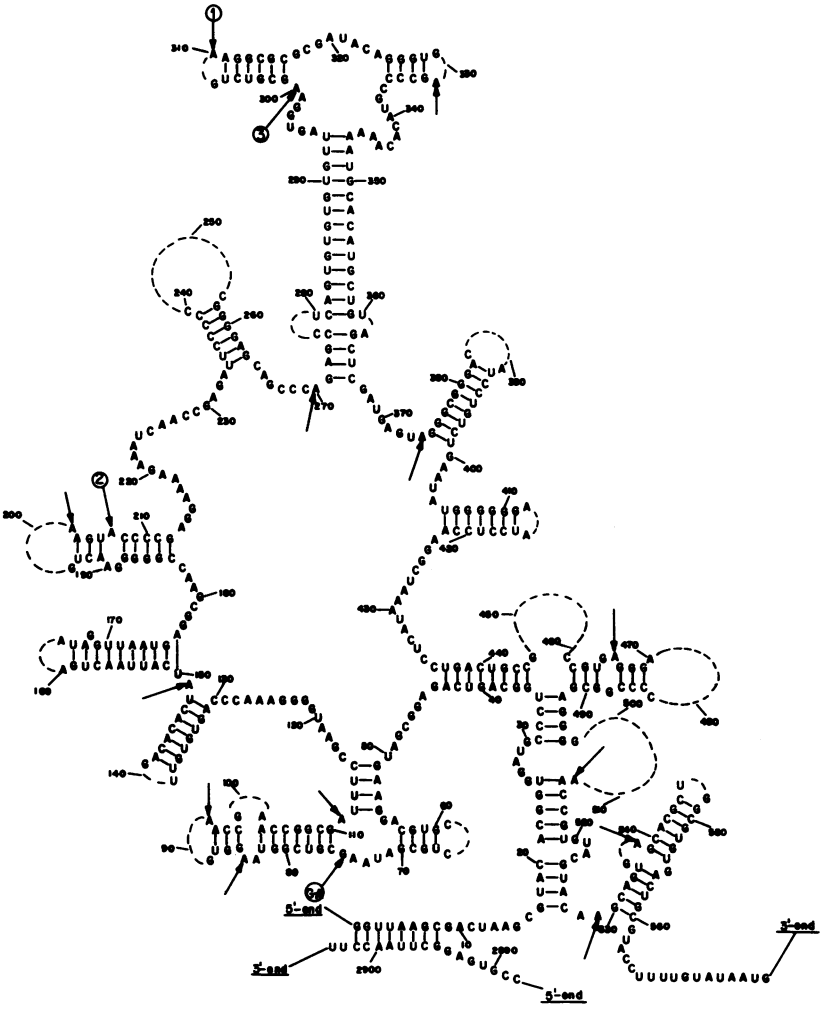

Fig. 1 Secondary structure proposed by Glotz et al (6), for first 577 bases of E. coli 23S rRNA. Arrows indicate 15 potential A-G base pairing sites and one G-A site (GA). (1), (2), and (3) indicate respectively, terminal-loop, middle-exclusion and open-end portions of double-helical regions.

mays 23S rRNA (6), which has 71% homology with E. coli 23S rRNA (11), shows 46 potential A-G base-pairing sites compared with 11 G-A. However,secondary structures proposed for some other ribosomal RNA's, with sequences quite different from those of E. coli, show few A-G sites. These include S. cerevisiae 18S rRNA (5)(15 A-G, 14 G-A), human mitochondrial 12S rRNA (5)(8 A-G, 5 G-A),

TABLE I

Comparison of the numbers of the various base pairs observed at helix ends, in structures proposed for E. coli 16S and 23S rRNA, with those expected for a random base distribution in the postulated single-stranded regions.

BASE PAIR 5'end 3'end	16S RANDOM	16S OBSERVED	23S RANDOM	23S OBSERVED
A-G	16.0	26	30.7	64
G-A	16.0	15	30.7	14
A-C	10.1	11	18.3	16
C-A	10.1	12	18.3	15
U-C	4.5	1	8.0	7
C-U	4.5	5	8.0	5
A-A	22.7	17	44.1	33
U-U	4.5	5	8.5	13
G-G	11.2	9	21.4	18
C-C	4.5	3	7.6	15

The 1541 bases of E. coli 16S rRNA include 488 G, 389 A, 351 C and 313 U. Given 234 G-C, 127 A-U and 70 U-G pairs in the proposed structure, this leaves 184 (27.1%) G, 262 (38.6%) A, 117 (17.2%) C and 116 (17.1%) U among 679 bases in single-stranded regions. For a random distribution, one would expect relative proportions of .386 X .271=.105 for A-G and for G-A pairs, .386 X .386=.149 for AA pairs etc., totalling .682 for pairs other than GC, AU, and UG: so some .105/.682 of the 104 helix ends (i.e. 16.0) should have AG pairs. The 2904 bases of E. coli 23S rRNA include 913 G, 764 A, 640 C and 587 U. There are 418 G-C, 230 A-U and 123 U-G pairs in the proposed structure, leaving 372 (27.3%) G, 534 (39.2%) A, 222 (16.3%) C and 234 (17.2%) U in postulated single-stranded regions. Therefore, for a random base distribution, one would expect .107/.682 of the 196 helix ends (i.e. 30.7) to have A-G pairs.

and human mitochondrial 16S rRNA (6)(7 A-G, 6 G-A).

A generalized secondary structure has recently been derived from the base sequences of seventeen different prokaryotic 5S rRNA's (7). This structure could incorporate an appropriately oriented A-G base pair at 4 of the 9 helix ends for E. coli and other gram negative bacteria; these include A15-G69, A59-G24, A78-G98 and A104-G72. However, for gram positive bacteria this same structure would incorporate only two or three of these A-G pairs.

Detailed X-ray structure determinations have been reported for four different transfer RNA's (8,9,12,13,14). These do not show A-G base-pairing, though in two cases (8,9,14) the sequences do indicate, by the criteria described above, a potential A9-G26 pair at the beginning of the D stem. In

TABLE II

Base pairs in double-helical regions adjoining A-G pairs in structures proposed for E. coli 16S and 23S rRNA's.

BASE PAIR	16S	23S
A-U	3	14
U-A	2	4
G-C	10	23
C-G	6	11
U-G	2	2
G-U	3	10

Thus for the 90 A-G pairs, adenine is adjacent to a purine in 63 cases and to a pyrimidine in only 27.

fact in $tRNA^{Phe}$, the most refined of these structures (8,9),dimethyl guanine 26 was found to make a non-planar trans-trans interaction with adenine 44 at the beginning of the anticodon stem, but with the dimethyl guanine at the 5' end of the double-helical region, that is with the opposite polarity to the A-G pairing we have been considering. There are, however, many tRNA sequences with adenine in position 26 and guanine at 44 (15), not yet investigated by X-ray diffraction, where the proposed A-G pair might occur.

We have investigated how well an A(syn)-G(trans) base pair can fit onto the end of a standard RNA-A double helix (16). We started with standard conformations for a double helix of dinucleotide strands, and then rotated the adenine by 180° about the glycosyl bond into the syn conformation. Using the CORELS model-building and refinement procedure (17), we were then able with fairly small adjustments (maximum change of 15° in dihedral angles) to arrive at a sterically acceptable model structure. In this the A-G pair makes two hydrogen bonds, $(A)N_6H...O_6(G)$ = 2.9Å and $(A)N_7...HN_1(G)$ = 2.9Å, the same 10.8Å distance between glycosyl C_1' atoms as for the Watson-Crick pairs, and no short van der Waals contacts (Fig. 2). This model also indicates that for such an A-G pair the guanine would make more favorable stacking interactions with a purine adjacent to A than with a pyrimidine, in accordance with the observations shown in Table II.

An A-G base pair with these dimensions could be accomodated with A(syn) at either the 5' or 3' end of a helical region, or indeed in the middle of the region, unless prevented by unfavorable steric interactions with adjacent nucleotides in the sequence. Our studies indicate that the short contacts be-

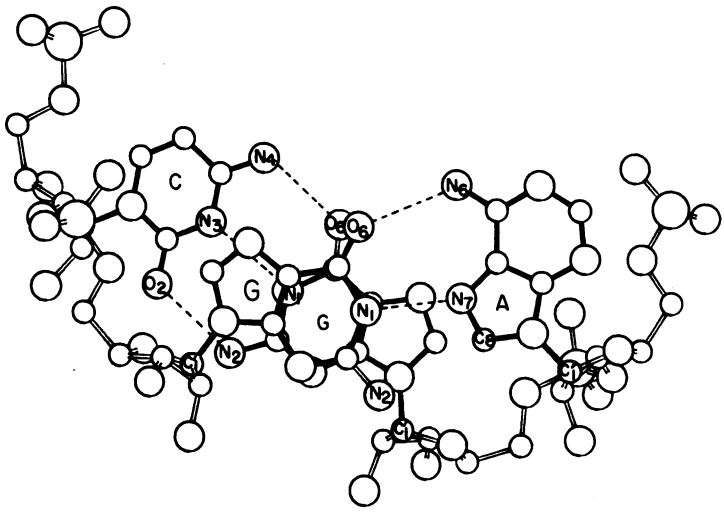

Fig. 2 Suggested A(syn)-G(trans) base pair, with adenine at the 5' end of a double-stranded region of RNA-A. Note that the guanine of the A-G pair overlaps the purine on the 3' side of the adenine.

tween atoms of adenine and the $C_2^{\prime}$ sugar atom of the adjacent nucleotide in the 5' direction found for a DNA structure (1) are less serious for RNA, and can be relieved by quite small adjustments leading to an H-bonded structure similar to that shown in Fig. 2. However, whereas in the CORELS refined model, with adenine at the 5' end, both hydrogens attached to the N_2 atom of guanine could make hydrogen bonds to water after rotation about the N_2-C_2 single bond, this is not possible for the model with adenine at the 3' end, due to short contacts between the water and either C_8 of adenine or the sugar $C_1^{\prime}$ of the nucleotide adjacent to guanine.

Conceivably an A(trans)-G(trans) pair with two hydrogen bonds as found in $tRNA^{Phe}$ might occur in rRNA. However, this would require about 12.7 Å between glycosyl $C_1^{\prime}$ atoms implying considerable distortion of the sugar-phosphate backbone. We have found no reason why such a base pair should be energetically preferred to G(trans)-A(trans) or other non-standard base pairs in accordance with the observations shown in Table I.

DISCUSSION

The proposal for the occurrence of A-G base-pairing at the ends of helices with standard Watson-Crick base pairs was originally put forward to

account for the widely observed tendency in RNA base compositions for the 6-amino bases (A+C) to roughly equal the 6-keto bases (G+U) and for purines (A+G) generally to exceed pyrimidines (U+C)(1, 18). Indeed, for E.coli 16S rRNA 6-amino/6-keto = 0.92 and purine/pyrimidine = 1.32, and for the 23S rRNA these ratios are 0.94 and 1.37 respectively. However, the degree of potential A-G pairing we have noted does not adequately explain these figures. The original suggestion did not take account of U-G base-pairing, which would tend to lower the 6-amino/6-keto ratio below 1.0, as is in fact generally observed, but only to a small degree (1). Furthermore, there was an implicit assumption that the compositional features were mainly a reflection of the base-pairing, implying that single-stranded regions either represented a rather small proportion of the structure or did not differ appreciably from the overall base composition. In fact, in the proposed secondary structures for both 16S rRNA and 23S rRNA of E. coli, close to half the total bases lie in single-stranded regions, and these have considerably more adenine and less of the other bases than the double-stranded regions (see footnote to Table I). In order to explain the overall base composition in terms of base-pairing, the secondary structures would need to have many more A-U and A-G pairs than have been proposed and possibly A-A pairs as well. It is perhaps premature to draw firm conclusions on this question from current predictions of RNA secondary structures, which may well still be extended and modified by the application of more powerful analytical methods.

Nevertheless, the statistical data we have presented in Table I and the text above does appear to provide very strong evidence for the existence of some kind of adenine-guanine association at the ends of double-helical regions at least in rRNA's of E. coli and related species. The paucity of A-G sites found in yeast and mitochondrial rRNA's may indicate that A-G pairing is highly species dependent or, alternatively, may reflect on the accuracy of the proposed secondary structures. The inclusion of A-G base-pairing as one of the criteria for choosing plausible secondary structures may help to resolve this question, and indeed may prove a useful mechanism for improving structure predictions.

The A(syn)-G(trans) base pair illustrated in Fig. 2 appears to offer an attractive structural explanation for our statistical findings, and it should be possible to test this hypothesis by X-ray diffraction analyses of appropriate RNA or oligonucleotide structures.

Finally, we have considered the possible function of A-G pairs in RNA's, and conceivably also in some DNA's. The A-G structure we have suggested would

seal off the end of a double-helical region and prevent further base stacking at that point. It would thus introduce a break, or at least a kink, in a linear array of double-helical stems or, in other words, favor a non-linear tertiary structure where this is required for biological function.

ACKNOWLEDGEMENT

We are grateful to Dr. David Elson for helpful and stimulating discussions.

REFERENCES

1. Traub, W. and Elson, D. (1966) Science 153, 178-180
2. Donohue, J. and Trueblood, K.N. (1960) J. Mol. Biol. 2, 363-371
3. Noller, H.F. and Woese, C.R. (1981) Science 212, 403-411
4. Stiegler, P., Carbon, P., Zucker, M., Ebel, J.P. and Ehresmann, C. (1981) Nucleic Acids Res. 9, 2153-2172
5. Zwieb, C., Glotz, C. and Brimacombe, R. (1981) Nucleic Acids Res. 9, 3621-3640
6. Glotz, C., Zwieb, C., Brimacombe, R., Edwards, K. and Kössel, H. (1981) Nucleic Acids Res. 9, 3287-3306
7. Studnicka, G.M., Eiserling, F.A. and Lake, J.A. (1981) Nucleic Acids Res. 9, 1885-1904
8. Sussman, J.L., Holbrook, S.R., Warrant, R.W., Church, G.M. and Kim, S.H. (1978) J. Mol. Biol. 123, 607-630; Holbrook, S.R., Sussman, J.L., Warrant, R.W. and Kim, S.H. (1978) J. Mol. Biol. 123, 631-660
9. Jack, A., Ladner, J.E. and Klug, A. (1976) J. Mol. Biol. 108, 619-649; Hingerty, B., Brown, R.S. and Jack, A. (1978) J. Mol. Biol. 124, 523-534
10. Schwarz, A. and Kössel, H. (1980) Nature (Lond.) 283, 739-742
11. Edwards, K. and Kössel, H. (1981) Nucleic Acids Res. 9, 2853-2869
12. Schevitz, R.W., Podjarny, A.D., Krishnamachari, N., Hughes, J.J., Sigler, P.B. and Sussman, J.L. (1979) Nature (Lond.) 278, 188-190
13. Woo, N., Roe, B and Rich, A. (1980) Nature (Lond.) 288, 346-351
14. Moras, D., Comarmond, M.B., Fischer, J., Weiss, R., Thierry, J.C., Ebel, J.P. and Giegé, R. (1980) Nature (Lond.) 286, 669-674
15. Singhal, R.P. and Fallis, P.A.M. (1979) Prog. in Nucleic Acids Res. and Mol. Biol. 23, 228-290
16. Arnott, S., Hukins, D.W.L. and Dover, S.D. (1972) Biochem. Biophys. Res. Comm. 48, 1392-1399
17. Sussman, J.L., Holbrook, S.R., Church, G.M. and Kim, S.H. (1977) Acta Cryst. A33, 800-804
18. Elson, D. and Chargaff, E. (1955) Biochim. Biophys. Acta 17 367-376