Volume 5 Number 11 November 1978 Nucleic Acids Research

Stacking of Crick Wobble pair and Watson-Crick pair: stability rules of G-U pairs at ends of helical stems in tRNAs and the relation to codon-anticodon Wobble interaction

H.Mizuno and M.Sundaralingam

Department of Biochemistry, College of Agricultural and Life Sciences, University of Wisconsin-Madison, Madison, WI 53706, USA

Received 17 July 1978

ABSTRACT

The occurrence of the noncomplementary G-U base pair at the end of a helix is found to be governed by stacking interactions. As a rule, a G-U pair with G on the 5'-side of a Watson-Crick base pair exhibits strikingly greater stacking overlap with the Watson-Crick base pair than a G-U pair on the 3'-side of a Watson-Crick base pair. The former arrangement is expected to be more stable and indeed is observed 29 times out of 32 in the known transfer RNA molecules. In accordance with this rule, the major wobble base pairs G-U or I-U in codon-anticodon interactions have G or I on the 5'-side of the anticodon. Similarly, in initiator tRNAs, this rule is obeyed where now the G is the first letter of the codon (5'-side). In the situation where U is in the wobble position of the anticodon, it is usually substituted at C(5) and may also have a 2-thio group and it can read one to four codons depending on its modifications. A G at the wobble position of the anticodon can recognize the two codons ending with U or C and modification of G (unless it is I) does not change its reading properties.

INTRODUCTION

The occurrence of the noncomplementary G-U base pair was initially envisioned by Crick in his wobble hypothesis for codon-anticodon interaction (1). In a number of tRNA molecules G-U oppositions have been observed both within and at the stems of the cloverleaf structure (2). The internal G-U opposition is found in the amino acid stem of yeast $tRNA^{Phe}$, whose three-dimensional x-ray structure has been established for both the monoclinic (3,4) and orthorhombic (5,6) polymorphic forms. The structures have revealed for the first time that the G and U bases are involved in the wobble hydrogen pair as predicted by Crick (1). We find that this unusual base pairing leads to remarkably different base stacking interactions with the Watson-Crick base pairs situated on either sides of it (7). The G-U base pair exhibits greater stacking interactions with the Watson-Crick base pair following it on the 3'-side of G (or 5'-side of U) than the Watson-Crick base pair preceding it on the 5'-side of G (or 3'-side of U). This observation has important implications on the occurrence of the G-U base pair at the end of a helical

stem. Following the direction of the chain from the 5'- to the 3'-end, the terminal G-U base pair (5'-end G-U pair) is more stabilized by stacking interactions with the Watson-Crick base pair following it than the terminal U-G pair (3'-end G-U pair). In accordance with this prediction, an examination of the known tRNA secondary structures reveals that when a terminal G-U pair occurs, it is predominantly on the 5'-side of the Watson-Crick base pair. A similar observation has been made previously (8,9), but the stacking attributes of the 5'-side G-U versus the 3'-end G-U have not been recognized. In this paper, we examine also the reading properties of the 5'-end G-U pair and of the 3'-end G-U pair in codon-anticodon interactions.

DISCUSSION

G-U base pair in yeast $tRNA^{Phe}$

In the structure of yeast phenylalanine tRNA the G-U base pair is located in the middle of the amino acid stem helix. These bases are held together by two hydrogen bonds as shown in Fig. 1. This hydrogen bonding scheme results in unequal angles at the glycosyl bonds; i.e. 59° at U and 38° at G, in contrast to the values of 52° in the Watson-Crick base pairs. Despite the displacement of the bases in forming the unusual base pair, the helical conformation of both backbone chains and glycosyl torsion angles are perturbed only slightly and are within their preferred ranges (10). However, a striking difference is observed in the base stacking interactions with the adjacent base pairs. The base pair in the 3'-side of G (and 5'-side of U)

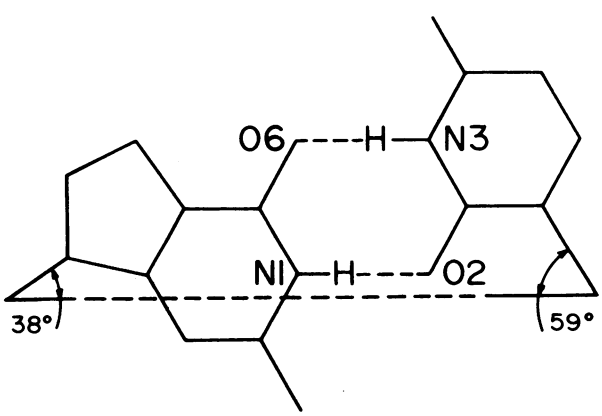
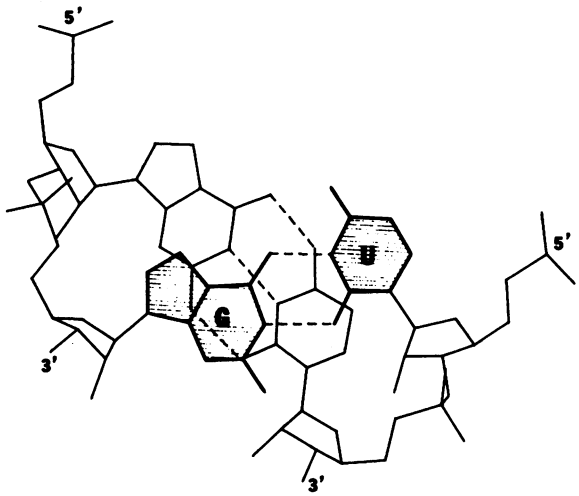
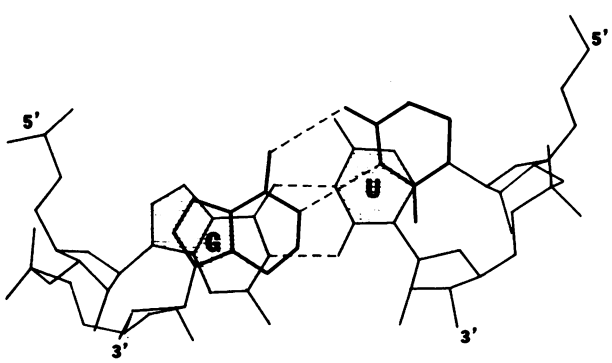


Fig. 1. Geometry of the G-U base pair showing the unequal angles at the glycosyl bonds.

shows a greater base overlap with the G-U base pair than the base pair on the 5'-side of G (Fig. 2). Therefore, it appears that the enhanced stacking overlap on the 3'-side of G may have an important stabilizing effect on the G-U base pair. In Fig. 2 is shown the stacking of the Crick Wobble pair with the two adjacent Watson-Crick pairs of yeast $tRNA^{Phe}$. In the alternative situations, if G and C or A and U are interchanged, the extent of stacking overlap changes somewhat but not significantly.



a



b

Fig. 2. Stacking of the G4-U67 base pair in the middle of the amino acid stem of yeast $tRNA^{Phe}$ (a) with G3-C68 base pair (b) with A5-U66 base pair. The former corresponds to the "3'-end G-U" base pair type, while the latter to the "5'-end pair G-U" base pair type. For nomenclature see ref. 12 and Fig. 3.

G-U base pair at the end of a helix

As long as the G-U base pair is within a helix, the above stacking interactions will always be observed. However, for the G-U base pair at the end of a helix, considerable overlap of the bases is observed only when the G is on the 5'-side (Fig. 2b), and not if it is on the 3'-side (Fig. 2a). Hence the former situation (hereafter referred to as "5'-end G-U") may be expected to be considerably more stable than the latter ("3'-end G-U"). To test this idea we examined the distribution of the G-U (including G-ψ) base pairs at the ends of the helical stems of the cloverleaf secondary structure of the approximately 80 known transfer tRNAs. The predicted stable "5'-end G-U" base oppositions are shown on the tRNA cloverleaf structure in Fig. 3.

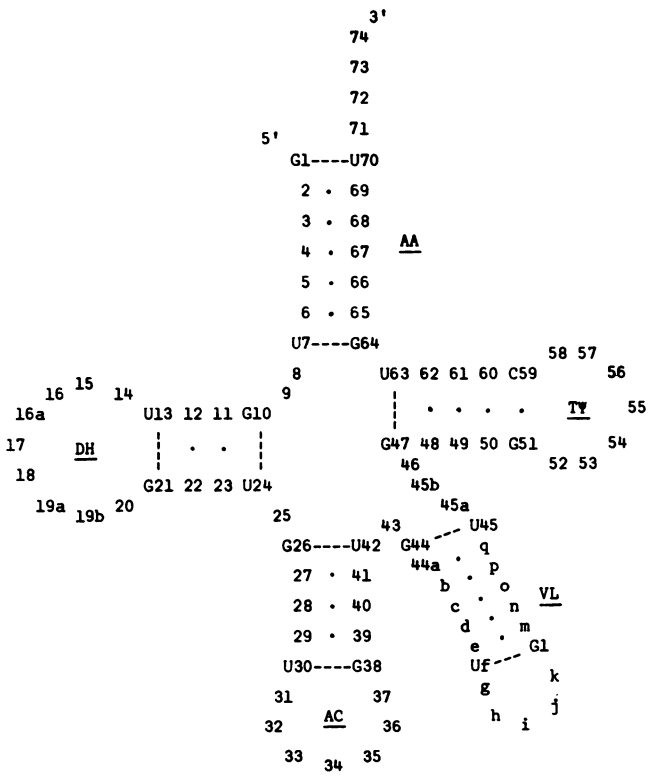


Fig. 3. Generalized numbering scheme for the tRNA cloverleaf structure based on a minimal tRNA chain length of 74 nucleotides with four insertion regions, two in the D-loop and two in the V-loop (14). The predicted stable 5'-end G-U base pairs at the ends of the stems are indicated by dashed lines. Switching the G and U in these will result in the less stable 3'-end G-U base pairs. G51-C59 is an invariant base pair.

It is seen from the statistics given in Table 1 that besides the 53 G-U base pairs located within one or more of the helical stems of tRNAs, there are 32 G-U base pairs at the ends of the cloverleaf helical stems. It is striking that 29 of the latter G-U base pairs are of the "5'-end G-U" type. There are only three exceptions: ψ26-G42 of baker's yeast $tRNA^{Glu}$ (11), G7-U64 of bacteriophage T4 coded $tRNA^{Pro}$ (12) and U47-G63 of E. coli $tRNA_{1A}^{Ala}$ (13). Although the latter two cases are "3'-end G-U" base pairs, they exhibit good stacking because they are located at the interface of the continuous helix formed by the AA and Tψ stems, assuming that their three-dimensional structures are analogous to the known $tRNA^{Phe}$ structure.

We further note that the Watson-Crick pair following a terminal "5'-end G-U" pair is of the pyrimidine-purine base pair type, i.e. the G stacks over the pyrimidine and the U over the purine.

The terminal G-U base pair rule can be expected to be obeyed in secondary structures of other RNA molecules, such as ribosomal RNAs and viral RNAs. A number of possible secondary structural models have been proposed for some of these RNAs from available sequence data. The G-U base pair rule can serve as a useful adjunct to construct more plausible secondary structural models of such RNA molecules.

Codon-Anticodon Triplet Interactions

In codon-anticodon triplet interactions, standard Watson-Crick base pairs are strictly required at the first two positions of the codon while some wobbling can be allowed at the third position (3'-end) of the codon (1). The first base of the anticodon (5'-end) is therefore designated the "wobble" position. In elongation tRNAs, G at the 5'-end of the anticodon gives "5'-end G-U" base pair when the third letter of the codon is U:

```
(5')  G . .  (3')        anticodon
      : | |
(3')  U . .  (5')          codon
```

This representation has been chosen for convenience to conform to the cloverleaf structure where the 5' side of the anticodon is on the left and the 3'-side on the right. The alternative nomenclature with the 5'-side of the message on the left and the 3'-side on the right would conform to the table of the genetic code. For example, phenylalanine tRNA from yeast (15) containing Gm (2' O-methyl-guanosine) at the 5'-end of the anticodon can form 5'-end G-U wobble base pair when the phenylalanine codon is UUU. It is interesting to note that all of the ten known phenylalanine tRNAs having different nucleotide sequences contain either G or Gm at the end of their

Table 1. The distribution of G-U base pairs in tRNAs. The numbering is based on Fig. 3. The number of "3'-end G-U" base pairs are in parentheses.

AA stem		VL stem	
1 - 70	1	44 - 45	1
2 - 69	6	44a - q	-
3 - 68	4	b - p	-
4 - 67	10	c - o	-
5 - 66	7	d - n	1
6 - 65	7	e - m	-
7 - 64	1 (1)	f - 1	1
AC stem		**Tψ stem**	
26 - 42	5 (1)	47 - 63	12 (1)
27 - 41	-	48 - 62	7
28 - 40	-	49 - 61	3
29 - 39	3	50 - 60	-
30 - 38	-	51 - 59	-
DH stem			
10 - 24	6		
11 - 23	5		
12 - 22	-		
13 - 21	5		

total	85
within helix	53
5'-end G-U	29
3'-end G-U	3

anticodons (16), suggesting not only support of the wobble hypothesis, but also that one anticodon triplet sequence is able to translate the two phenylalanine codons UUC or UUU. Similarly, the known tRNAs of Tyr, His, Asn, Asp, Cys and Ser with doubly degenerate codons ending in U or C contain always G or a derivative of it at the 5'-end of the anticodon. Thus the stability of a 5'-end G-U base pair appears to be similar to that of a G-C base pair at the wobble position.

Initiator tRNAs respond to AUG as well as to GUG codons. In the latter case, a G-U base pair is at the first position of the codon, i.e. G is now on the first position of the codon. This "double reversal" again results in a "5'-end G-U" base pair.

```
(5')  C  A  U  (3')      anticodon
      |  |  :
(3')  G  U  G  (5')        codon
```

However, in elongation tRNAs U at the 5'-end of the anticodon results in a "3'-end G-U" base pair:

```
(5')  U  .  .  (3')      anticodon
      :  |  |
(3')  G  .  .  (5')        codon
```

In such cases, with the exception of a few (17,18), the uridine is modified either at the C(5)- position or at both the C(5)- position and the 2-keto oxygen (19). According to the nature of the modification of the uridine at the wobble position, there are two types of codon-anticodon recognition (see Table 2). The presence of an unmodified uridine (20) or of a 5-oxyacetic acid uridine (and 5-methoxy uridine) leads to an extension of the wobble recognition: such uridine derivatives recognize three or four codons (21,22). On the other hand, the presence of a 2-thiouridine derivative (and 5-acetic acid methyl ester uridine (22)) restricts the wobble recognition to A with only a weak or no recognition of G: the thiouridine derivatives recognize only two codons (19). Thus, these two types of recognition correlate with the number of codons specifying an amino acid. The 2-thiouridine derivatives occur only in tRNAs recognizing codons where the second letter is A (with the exception of Phage T4 coded $tRNA^{Gly}$ (24)), thereby apparently preventing mispairing with U or C in the cases of Glu, Lys, and Gln. Such codons ending in U or C (with the second letter A) are recognized by tRNAs containing heavily substituted G residues at the first position of the anticodon (Q, man Q, gal Q) (21). For Arg (with AG as the first two letters of the codons), such miscoding seems to be prevented by the 5-methoxycarbonylmethyl uridine (23); while, for Leu (with UU as the first two letters of the codons), the nucleotide at the first position of the anticodon has not been identified in the only reported sequence, but is known to be modified. So far, all the substituted uridines of the latter group, which recognize at most two codons, are 5-methyluridine derivatives. On the contrary, unmodified uridine, as well as cmo^5U and mo^5U (where C(5) is directly linked to an oxygen), exhibit strong "wobbling" as I whenever an amino acid is specified by four codons.

Several models have been suggested for explaining the restrictive "wobble" recognition of the substituted uridines at the first position of the anticodon:

- unlikely N-H ...S hydrogen bond between G and 2-thiouridine (25);
- stacking interactions between the S atoms and the middle base of the triplet (26);
- hydrogen bonding of the C(5)-substituent to the backbone (27);
- unusual nucleotide conformations (28).

It is not clear which of these models is correct. It could be that some or all of these mechanisms play a role. We note further that, for both groups of Table 2, the U with longer substituents are usually followed by a

Table 2. Anticodons of tRNAs containing uridine derivatives at the wobble position (2, 29).

Group 1 (two codons)			Group 2 (two codons)		
Yeast $tRNA_3^{Arg}$	:	mcm^5UCU	B. subtilis $tRNA^{Ala}$	:	mo^5UGC
E. coli K12 $tRNA_1^{Gln}$	:	N*UG	E. coli $tRNA_{1a}^{Ala}$	:	cmo^5UGC
Phage T4 $tRNA^{Glu}$	:	NUG	E. coli $tRNA^{Gly}$	:	NCC
E. coli B $tRNA^{Glu}$	:	mam^5s^2UUC	1A S. Epidermitis $tRNA^{Gly}$	:	UCC
E. coli $tRNA_2^{Glu}$	:	mam^5s^2UUC	1B S. Epidermitis $tRNA^{Gly}$	:	UCC
Yeast $tRNA_3^{Glu}$	:	mcm^5s^2UUC	Phage T4 $tRNA^{Gly}$	:	mam^5s^2UCC
Phage T4 $tRNA^{Leu}$	:	NAA	B. subtilis $tRNA_1^{Gly}$	:	$cmam^5UCC$ (30)
E. coli B $tRNA^{Lys}$	:	mam^5s^2UUU	Yeast $tRNA^{Leu}$	:	UAG
B. subtilis $tRNA^{Lys}$	:	U $(cmam^5s^2U)UU$ (30)	Phage T4 $tRNA^{Pro}$	:	NGG
Yeast $tRNA_2^{Lys}$	:	mcm^5s^2UUU	E. coli $tRNA_1^{Ser}$	:	cmo^5UGA
			Phage T4 $tRNA^{Ser}$	:	NGA
			B. subtilis $tRNA^{Thr}$	:	mo^5UGU
			E. coli B $tRNA^{Val}$	:	cmo^5UAC
			Yeast $tRNA_{2A}^{Val}$	:	NAC
			B. subtilis $tRNA^{Val}$	:	mo^5UGU

N : derivative of U; N* : derivative of 2-thiouridine

mcm^5U : 5-methoxycarbonylmethyluridine ($-CH_2-CO_2-CH_3$)

mcm^5s^2U : 5-methoxycarbonylmethyl-2-thiouridine

mam^5s^2U : 5-methylaminomethyl-2-thiouridine ($-CH_2-NH-CH_3$)

$cmam^5s^2U$: 5-carboxymethylaminomethyl-2-thiouridine ($-CH_2-NH-CH_2-CO_2H$)

mo^5U : 5-methoxyuridine ($-OCH_3$)

cmo^5U : 5-oxyacetic acid uridine ($-OCH_2-CO_2H$)

pyrimidine base while mo^5U and cmo^5U are usually followed by a purine in the anticodon.

CONCLUSIONS

G-U oppositions at helix termini are invariably of the "5'-end G-U" type. It is further found that the "5'-end G-U" pair is followed by a pyrimidine-purine Watson-Crick base pair. However, at the wobble position, both the "5'-end" and the "3'-end" terminal G-U base pair occur as required by the wobble hypothesis. This seemingly indicates that the environment at the wobble pair is different from that of a helix terminus. The 5'-end G translates only two codons regardless of the fact that G is modified or not. Thus, the role of a G modification is not clear. For the 5'-end U, the stacking overlap to the middle codon-anticodon base pair is diminished. The loss of stacking probably allows the U base at the 5'-end of the anticodon more freedom to engage in various wobble pairs. Thus, in addition to pairing with G, the 5'-end U can also participate in the U-U and U-C short wobble pairs. Further, the 5'-end uridine is usually modified. While an unmodified U can read four codons, a modified U can read one to four codons depending on the modification. It appears that the nature of the middle base (pyrimidine or purine) may be correlated with the nature of the substitution: the longer substituents at C(5) and the 2-thio modified uridines generally have a pyrimidine as the middle base while the shorter substituents generally have a purine base.

The sharp turn at the invariant uridine before the anticodon bases could be partly responsible for the loss of geometrical constraints of the third codon-anticodon base pair. This might also be the case when a Watson-Crick base pair is at the wobble position. This model assumes that the conformational changes in the anticodon loop, if any, upon binding at the receptor site of the ribosome do not greatly perturb the anticodon bases from the 3'-stacked conformation (31) found in the crystal structures (3-6). The flipping of the anticodon bases into a 5'-stacked conformation is unlikely for the reason that this would also involve flipping of the tilts of the anticodon stem base pairs from their characteristic orientation found in RNA helices to the reverse direction with respect to the helical axis of the anticodon stem (32,33). This point has generally been overlooked by advocates of the "flip-flop" hypothesis (34,35).

It has been observed (14,36) that the tertiary base interactions in the tRNA structures show variations around the dihydrouridine stem and its

interface with the anticodon stem. Besides possible variations in the secondary helical structures brought by base sequence differences, these changes resulting from tertiary base interactions might be related to the nature of the anticodon triplets and therefore to anticodon-codon recognition. In addition, these changes are expected to be important for the recognition of the different tRNAs by their cognate synthetases.

ACKNOWLEDGEMENTS

We gratefully thank Dr. E. Westhof for his generous help in the preparation of the manuscript. Research supported by grants (GM-18455 and GM-17378) from the National Institutes of Health of the United States Public Health Service.

REFERENCES

1. Crick, F.H.C. (1966) J. Mol. Biol. 19, 548-555.
2. Barrell, B.G. and Clark, B.F.C. (1974) Handbook of Nucleic Acid Sequences. Joynson-Bruvvers, Oxford.
3. Stout, C.D., Mizuno, H., Rubin, J., Brennan, T., Rao, S.T. and Sundaralingam, M. (1976) Nucl. Acids Res. 3, 1111-1123.
4. Ladner, J.E., Jack, A., Robertus, J.D., Brown, R.S., Rhodes, D., Clark, B.F.C. and Klug, A. (1975) Nucl. Acids Res. 2, 1629-1637.
5. Quigley, G.J., Seeman, N.C., Wang, A.H.-J., Suddath, F.L. and Rich, A. (1975) Nucl. Acids Res. 2, 1629-1637.
6. Sussman, J.L. and Kim, S.H. (1976) Biochem. Biophys. Res. Commun. 68, 89-96.
7. Mizuno, H. (1977) Am. Cryst. Assn. Meetings, Abstract F5, East Lansing, Michigan, August 7-12.
8. Ninio, J. (1973) Prog. Nucleic Acid Res. Mol. Biol. 13, 301-337.
9. Clark, B.F.C. (1978) Ibid. 20, 1-9.
10. Stout, C.D., Mizuno, H., Rao, S.T., Swaminathan, P., Rubin, J., Brennan, T. and Sundaralingam, M. (1978) Acta Cryst. B34, 1529-1544.
11. Kobayashi, T., Irie, T., Yoshida, M., Takeishi, K. and Ukita, A. (1974) Biochem. Biophys. Acta 366, 168-181.
12. Barrell, B.G., Seidman, J.G., Guthrie, C. and McClain, W.H. (1974) Proc. Nat. Acad. Sci. USA 71, 413-416.
13. Williams, R.J., Nagel, W., Roe, B. and Dudock, B. (1974) Biochem. Biophys. Res. Commun. 60, 1215-1221.
14. Brennan, T. and Sundaralingam, M. (1976) Nucl. Acids Res. 3, 3235-3251.
15. RajBhandary, U.L. and Chang, S.H. (1968) J. Biol. Chem. 243, 598-608.
16. Keith, G. and Dirheimer, G. (1978) Biochim. Biophys. Acta 517, 133-149.
17. Roberts, R.J. (1974) J. Biol. Chem. 249, 4787-4796.
18. Randerath, K., Chia, L.S.Y., Gupta, R.C., Randerath, E., Hawkins, E.R., Brum, C.K. and Chang, S.H. (1975) Biochem. Biophys. Res. Commun. 63, 157-163.
19. Nishimura, S. (1972) Prog. Nucleic Acid Res. Mol. Biol. 12, 49-85.
20. Weissenbach, J., Dirheimer, G., Falcoff, R., Sauceau, J. and Falcoff, E. (1977) FEBS Letters 82, 71-76.
21. McCloskey, J.A. and Nishimura, S. (1977) Acc. Chem. Res. 10, 403-410.
22. Mitra, S.K., Lustig, F., Akesson, B., Lagerkvist, U. and Strid, L. (1977) J. Biol. Chem. 252, 471-478.
23. Weissenbach, J. and Dirheimer, G. (1978) Biochim. Biophys. Acta 518, 530-534.

24. Stahl, S., Paddock, G.V. and Abelson, J. (1974) Nucleic Acids Res. 1, 1287-1304.
25. Sekiya, T., Takeishi, K. and Ukita, T. (1969) Biochim. Biophys. Acta 182, 411-426.
26. Mazumdar, S., Saenger, W., Scheit, K. (1974) J. Mol. Biol. 85, 213-229.
27. Berman, H.M., Marcu, D. and Narayanan, P. (1978) Nucleic Acids Res. 5, 893-903.
28. Grosjean, H.J., De Henau, S. and Crothers, D.M. (1978) Proc. Nat. Acad. Sci. USA 75, 610-614.
29. Sprinzl, M., Gruter, F. and Gauss, D.H. (1978) Nucleic Acids Res. 5, r15-r27.
30. Ishikura, H. (1978) personal communication.
31. Fuller, W. and Hodgson, A. (1967) Nature 215, 817-819.
32. Sundaralingam, M., Brennan, T., Yathindra, N. and Ichikawa, T. (1975) In Structure and Conformation of Nucleic Acids and Protein-Nucleic Acid Interactions, pp. 101-115. Edited by M. Sundaralingam and S.R. Rao, University Park Press, Baltimore, Maryland.
33. Sundaralingam, M. (1978) In Proceedings of the International Symposium on Biomolecular Structure, Function, Conformation and Evolution (Madras), Pergamon Press.
34. Woese, C.R. (1970) Nature 226, 817-819.
35. Crick, F.H.C., Brenner, S., Klug, A. and Pieczenik, G. (1976) Origins of Life 7, 389-397.
36. Sundaralingam, M. and Bott, R., in preparation and presented at the Cold Spring Harbor Meeting on tRNA, Long Island, N.Y., August 23-27, 1978.