Proc. Nat. Acad. Sci. USA
Vol. 71, No. 12, pp. 4970–4974, December 1974

The General Structure of Transfer RNA Molecules

(base stacking/hydrogen bonding/tRNA sequences/tRNA conformation)

S. H. KIM*, J. L. SUSSMAN*, F. L. SUDDATH, G. J. QUIGLEY, A. McPHERSON, A. H. J. WANG, N. C. SEEMAN, AND ALEXANDER RICH

* Department of Biochemistry, Duke University School of Medicine, Durham, North Carolina 27710; and Department of Biology, Massachusetts Institute of Technology, Cambridge, Mass. 02139

Contributed by Alexander Rich, October 10, 1974

ABSTRACT **The three-dimensional structure of yeast phenylalanine tRNA serves as a useful basis for understanding the tertiary structure of all tRNAs. A large number of tRNA sequences have been surveyed and some general conclusions are drawn. There are only a few regions in the molecule in which there are differences in the number of nucleotides; and the structure of yeast phenylalanine tRNA can accommodate these differences by forming or enlarging protuberances on the surface of the basic framework molecule. The nature and distribution of the differences in number of nucleotides are surveyed and possible hydrogen bonding interactions are discussed for a number of tRNA classes. The two most significant features of the molecule are the large number of stacking interactions which are seen to include most of the nucleotides in the molecule and the system of specific hydrogen bonding interactions. It is likely that these stabilizing elements are preserved in all tRNA structures.**

Until recently the most striking feature of transfer RNA (tRNA) sequences has been the fact that they could all be arranged in the familiar cloverleaf diagram with complementary hydrogen bonding between the bases in the stem regions (1). In addition, some positions are always occupied by constant nucleotides. Almost 2 years ago, the x-ray diffraction analysis of yeast phenylalanine tRNA ($tRNA^{phe}$) at 4-Å resolution showed that this molecule not only contains the double helical stems implicit in the cloverleaf diagram, but also was found to have an L shape with the anticodon loop at one end of the L, the acceptor stem at the other end, and the dihydrouracil (D) and TψC loops forming the corner of the molecule (2). More recently the x-ray diffraction analysis of this molecule has been extended to 3-Å resolution for both the orthorhombic (3, 4) and monoclinic crystal forms (5) and the tertiary interactions in both crystal forms appear very similar. The most striking feature of the 3-Å analysis is the extent to which a large number of the bases which are constant to all tRNAs are used in the tertiary hydrogen bonding interactions. These, together with base stacking, maintain the three-dimensional form of the molecule. This suggests rather directly that the three-dimensional structure seen in yeast $tRNA^{phe}$ may be generalized to understand the structure of all tRNAs. The idea that all tRNAs have a common or similar structure is not surprising in view of the fact that all tRNAs involved in protein synthesis must go through the ribosomal machinery. Here we discuss the manner in which the three-dimensional structure of yeast $tRNA^{phe}$ may serve as a framework for understanding the structure of all tRNA molecules. We do this by comparing tRNA sequences (6) and suggesting plausible ways in which structural components in this molecule may be modified to fit in other tRNA sequences.

Structural features of yeast phenylalanine tRNA

The preliminary details of yeast $tRNA^{phe}$ tertiary interactions have been published (4, 5). Further improvement of the phases using a "partial structure method" (Sussman and Kim, in preparation) has reinforced most of the tertiary interactions previously published and reveals more clearly regions which were previously uncertain (manuscript in preparation). Fig. 1 shows the familiar cloverleaf sequence for yeast $tRNA^{phe}$ (7), and the solid lines indicate tertiary hydrogen bonding interactions. In addition, positions occupied by constant bases are indicated. It can be seen that many of the tertiary hydrogen bonding interactions involve nucleotides which are constant in all tRNA sequences. The three-dimensional form of the molecule is shown schematically in Fig. 2 in which the ribose-phosphate backbone is represented as a continuous tube and the rodlike crossbars indicate secondary interactions while the black bars represent tertiary interactions. In addition, the dotted lengths on the backbone represent regions in which there are a variable number of nucleotides in other

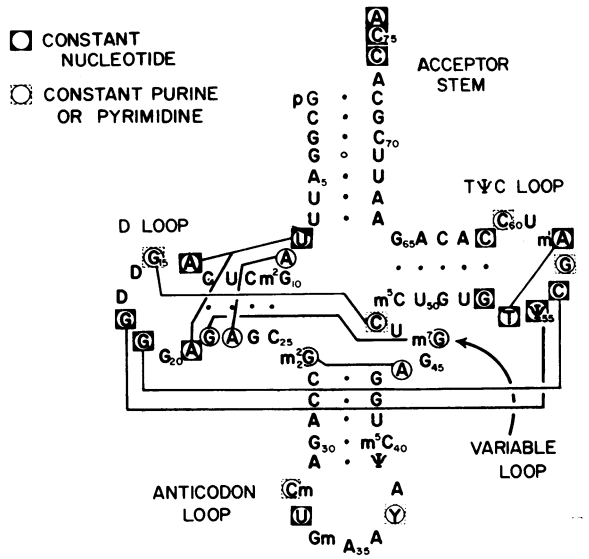

FIG. 1. The nucleotide sequence of yeast $tRNA^{phe}$. Constant nucleotides as well as constant purine or pyrimidine positions for all tRNA sequences are shown. These do not all apply to eukaryotic initiator tRNAs, however. The *solid lines* connecting nucleotides indicate tertiary hydrogen bonding interactions consisting of one, two, or three hydrogen bonds. A21 appears to be hydrogen bonding to the ribose of U8. Y in position 37 stands for a modified purine nucleotide.

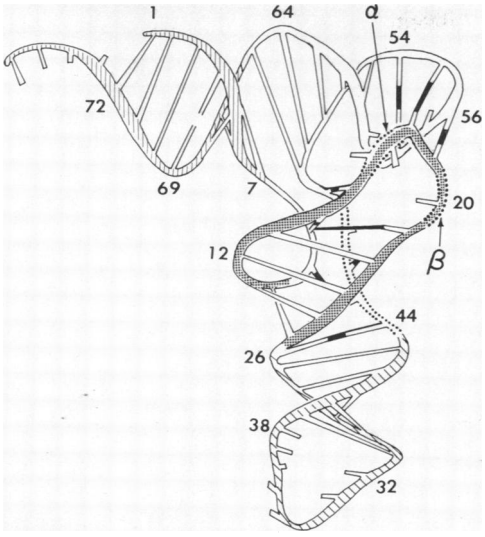

FIG. 2. A schematic diagram illustrating the folding of the yeast tRNA^{phe} molecule. The ribose phosphate backbone is drawn as a continuous cylinder and *unshaded crossbars* indicate the secondary hydrogen bonds in double helical stems. Tertiary interactions are shown by *crossbars which have black segments* in them. Positions of single bases are indicated by *rods* which are intentionally shortened. The variable regions in terms of the number of nucleotides in different tRNA molecules are shown in *dotted outline* and two of the variable regions in the D loop, α and β, are labeled. The anticodon is at the bottom of the figure while the amino acid acceptor is at the upper left. The numbering of nucleotides is the same as in Fig. 1. The D loop and stem is heavily *stippled* whereas the TψC loop is *unshaded*.

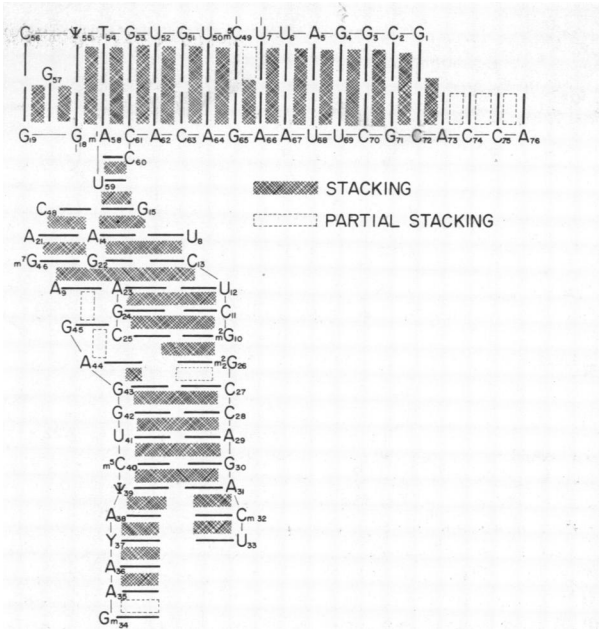

FIG. 3. A schematic diagram illustrating the hydrophobic stacking interactions between the nucleotides of yeast tRNA^{phe}. Full stacking and partial stacking are indicated. In this orientation the anticodon is at the bottom, the CCA acceptor stem is at the *upper right* and the TψC loop is shown at the *upper left* corner. Where adjacent stacking nucleotides are connected by a ribose phosphate chain, the linkage is noted by a *thin line*. The *heavy solid lines* attached to the nucleotide symbols represent in schematic fashion the purine or pyrimidine base while the hydrophobic interaction is indicated by the *blocks* between the bases. The connectivity of the molecule is indicated only by the numbering scheme. This accounts for the fact that many nucleotides appear to be unconnected. Bases not involved in stacking (bases 16, 17, 20, 47) are omitted from the figure.

tRNAs. All the stems in the molecule except the D stem have helical parameters similar to RNA-A (8), but differ in details.

Fig. 3 is a schematic diagram which shows the extensive system of base stacking interactions. A remarkable feature of the molecule is the fact that it is made up of a horizontal and a vertical stacking unit. Each of these units contains not only the base stacking associated with the helical stem regions, but also the additional stacking associated with the tertiary interactions in the TψC and D loops, as well as in the variable loop. Four of the five bases in the variable loop are involved in stacking interactions. In fact, the only nucleotides not involved in considerable stacking interactions are the base U47 in the variable loop as well as bases 16, 17, and 20 in the D loop. The vertical stacking interactions (Fig. 3) are continued down into the anticodon loop (5) in a manner similar to that proposed by Fuller and Hodgson (9) although differing in detailed conformation.

Fig. 4 shows a number of hydrogen bonding interactions, some of which are seen in yeast tRNA^{phe} and are also directly applicable to other tRNA sequences. For example, in the U8-A14 pair involving two constant bases, (Fig. 4a), U8 is stacked parallel to and overlapping C13. The relative orientation of U8 and C13 is very close to that predicted by Bergstrom and Leonard for photodimerization (10). In *Escherichia coli* tRNAs, position 8 is 4-thio uracil and photodimer cross-linking has been reported for six different *E. coli* tRNAs (11, 12). This directly implies a structure similar to that seen in yeast tRNA^{phe} in this region. Another pairing involves the hydrogen bonding of the positively charged m^{7}G46 to G22 (Fig. 4b). In this position, m^{7}G46 is close to the phosphate group of nucleotide 9 implying electrostatic stabilization in the center of the molecule.

The D loop

It is apparent from inspection of the sequences of all tRNAs (6) that there are two regions in the D loop in which the number of nucleotides vary from one tRNA to another. In yeast tRNA^{phe}, this includes nucleotides D16 and D17, as well as G20. Nucleotides 16 and 17 bulge out from the molecule in such a way that the bases are exposed and the backbone is folded on itself. These two regions contain from 1 to 3 nucleotides in different tRNA sequences; thus it seems likely that the size of the bulge may vary in different tRNAs. These two regions, which we call α(position 16–17) and β(position 20) are separated by the two constant guanine nucleotides 18 and 19 which are hydrogen bonded to the TψC loop. Fig. 5 is a histogram showing the number of nucleotides in the D loop for 56 tRNA sequences. In tRNAs with three base pairs in the D stem we have arbitrarily excluded from the loop the first two bases next to the paired stem. The numbers in parentheses in the histogram show the number of nucleotides in the α and β regions. It can be seen that the commonest group are those with two nucleotides in the α region and one in the β region, as in yeast tRNA^{phe}. Although there are a large number of tRNAs with one nucleotide each in the α and β regions, none are found with three nucleotides each in the α and β regions. The α and β regions contain mostly pyrimidines, with

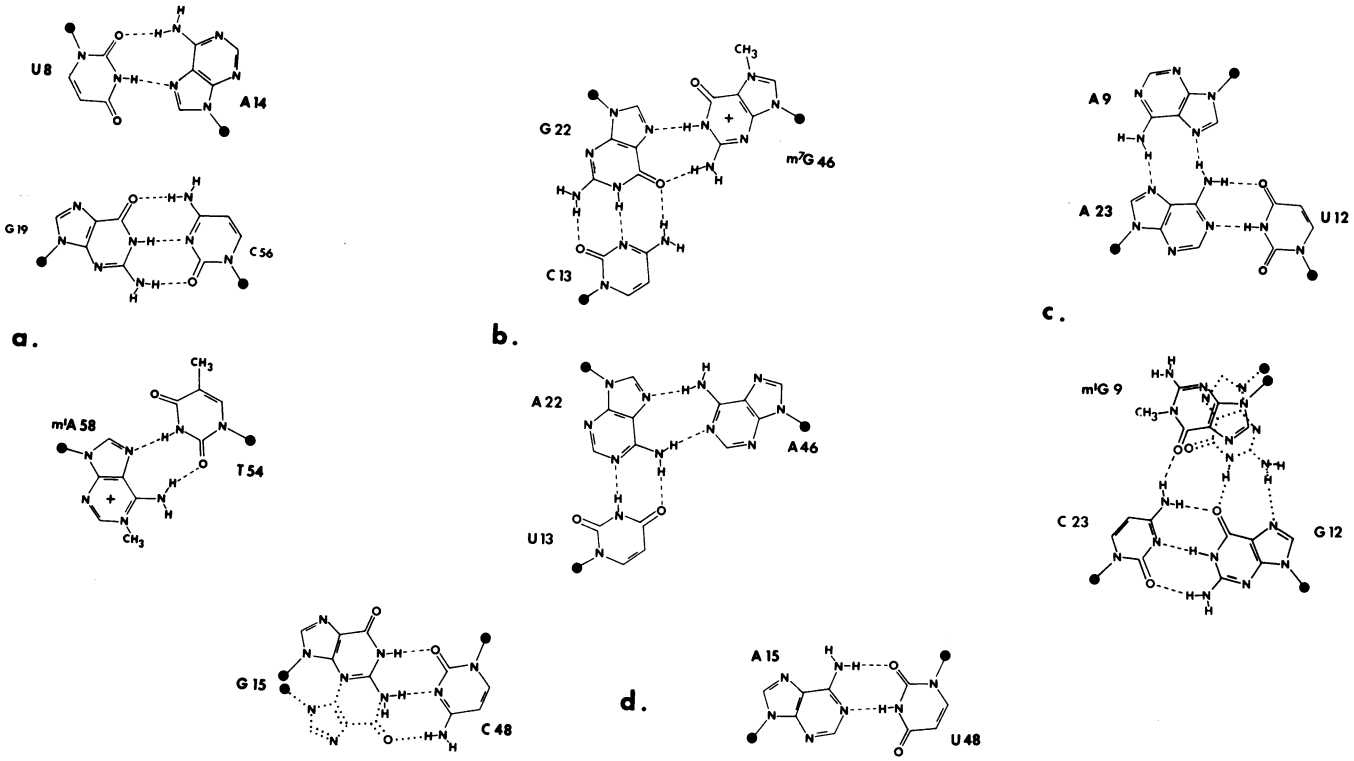

FIG. 4. Diagrams showing various types of hydrogen bonding that are seen in yeast $tRNA^{phe}$ and postulated for other tRNAs. The *filled black circles* represent the C1′ of the ribose residues. In two cases (c and d) alternative positions of guanine residues are illustrated by a dotted outline.

D accounting for over 70% of the bases. It has been suggested recently (5) that the differences in regions α and β may constitute an enzyme recognition site for different tRNAs. However, this seems unlikely since at least three different *E. coli* tRNAs have 2 D's in the α region and one D in the β region.

Classification of tRNA sequences

From the cloverleaf sequences, most tRNAs can be grouped into four major classes based on the number of base pairs in the D stem and the number of nucleotides in the variable loop (13). For example, class D4V5 has four base pairs in the D stem and five nucleotides in the variable loop. This is the largest single class of tRNAs and accounts for almost half of the approximately 60 tRNAs which have been sequenced. Other classes are D3V5, D(3 or 4)V4, and D3VN, where N varies from 13 to 21 nucleotides.

Class D4V5

Yeast $tRNA^{phe}$ belongs to this class. Fig. 6 shows a cloverleaf diagram with generalizations that can be seen by studying 28 published sequences. The nucleotides which are common to all 28 sequences are shown in this figure, with a few exceptions that are listed in the figure legend. The numbering system for yeast $tRNA^{phe}$ is used throughout in the descriptions of this and all other classes of tRNA.

The hydrogen bonding shown in Fig. 4a shows the interactions which are seen at 3-Å resolution for G19-C56, m^1-A58-T54, and U8-A14. It should be noted that rabbit liver $tRNA^{phe}$ has m^1A14 (6) which is consistent with the observed hydrogen bonding. Since these pairs involve bases constant in all tRNAs, except for the eukaryotic initiator tRNAs, it can be assumed that they are represented in all classes. In yeast $tRNA^{phe}$ the relationship between G15 and C48 is shown in Fig. 4d where the glycosidic bonds are on opposite sides (*trans*) of the base pair. Occasionally there are coordinate changes in these two sites, yielding A15 and U48 which could have the hydrogen bonding shown in Fig. 4d if the *trans* configuration is preserved (5). An exception to this generalization is found in the sequence of *E. coli* and *S. typh.* $tRNA^{gly}$ which has A15 and C48 (6). A *trans* pairing between A and C can be easily accommodated here. In yeast $tRNA^{phe}$, m^7G46 hydrogen bonds to G22 as shown in Fig. 4b. In 25 out of 28 sequences, position 46 is occupied by the positively charged m^7G. Unmodified G is found in one sequence whereas yeast $tRNA_3^{arg}$ has A46 with U13-A22 (6). Its probable hydrogen bonding is shown in Fig. 4b (5).

An interesting variation occurs in the interaction between the purine in position 9 with the 12–23 base pair in the D stem. The commonest arrangement is A9 pairing to A23 which in turn is paired to U12 (Fig. 4c). However, a frequent sequence is one in which the pair G12-C23 is associated with either G or m^1G in position 9. A possible pairing for this arrangement is shown in Fig 4c where two positions are shown for G9. The most likely interaction for m^1G is shown in the solid line figure with a single hydrogen bond between O6 of m^1G and the amino group of C23. It is intriguing to note that the three electronegative atoms N7 of m^1G9 and O6 and N7 of G12 form an arrangement which could easily bind a metal cation. A *syn* conformation is possible for the guanine residues which do not have a methyl group on N1 as shown in the dotted line figure.

A21 is in the same plane as the base pair U8-A14 and appears to be hydrogen bonded to ribose U8. In yeast $tRNA^{phe}$ there may be an interaction between m_2^2G26 and A44 involv-

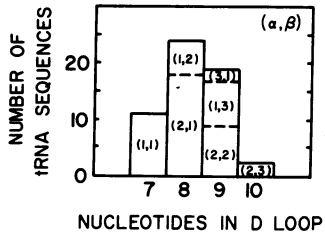

FIG. 5. A histogram showing the number of tRNA sequences which have different numbers of nucleotides in the D loop. For tRNAs with four base pairs in the D stem, the D loop consists of all of the remaining nucleotides. The *asterisk* on D loop indicates that for tRNAs with three base pairs in the D stem, the two nucleotides which would correspond to the fourth base pair are not included in the D loop. The *figures in parentheses* indicate the number of nucleotides in the α or the β regions as defined in the *text*.

ing a single hydrogen bond (4). G26 and A44 are the most common pairs of residues found in other members of this class although in several cases the positions of A and G are reversed between 26 and 44.

In the eukaryotic initiator $tRNA_f^{met}$, the sequence (T54, ψ55) becomes (A54, U55) while position 60 has an A instead of a pyrimidine. From studying the conformation of the TψC loop, it is likely that A can be inserted in position 54 in such a way that it could form a pair of hydrogen bonds with A58 using the hydrogen bonding found in helical polyadenylic aicd (14), whereas position 60 could accommodate an adenine instead of a pyrimidine.

All the hydrogen bonded interactions and the base stacking system described for Class D4V5 are assumed to be applicable for the remaining tRNA classes with the exceptions noted below.

Class D3V5

Compared to class D4V5, there are fewer constant nucleotides in the eight sequences of this class. For example, there are departures from regularity at positions, 9, 44, and 46 and the noncomplementary bases in positions 13 and 22 show great variety. Nonetheless, it is not unreasonable to assume that hydrophobic stacking interactions which dominate the structure of yeast $tRNA^{phe}$ will be a feature of all tRNA molecules. Accordingly, it is likely that the bases in positions 13 and 22 will be maintained in a stacked configuration. All of the nucleotides in positions 13 and 22 could pair with each other in the *trans* orientation if nucleoside 22 were in the *syn* conformation. This would maintain the stacking interactions, even though the separation between the ribose phosphate chains would vary in that region from one tRNA to the next. It is possible to postulate plausible but unsystematic hydrogen bonding interactions between the residues found in position 46 and 22 which might nonetheless remain stacked. Likewise, the interactions between the nucleotide in position 9 with the base pair 12–23 are not uniform; single tertiary hydrogen bonds can always be formed, but stacking again may be more important for stability.

Class D3/4V4

This class contains only five sequences and few generalities can be made. From the sequences of the four bases in the variable loop it is likely that the pyrimidine found in position 47 of class D4V5 is eliminated from the variable loop since

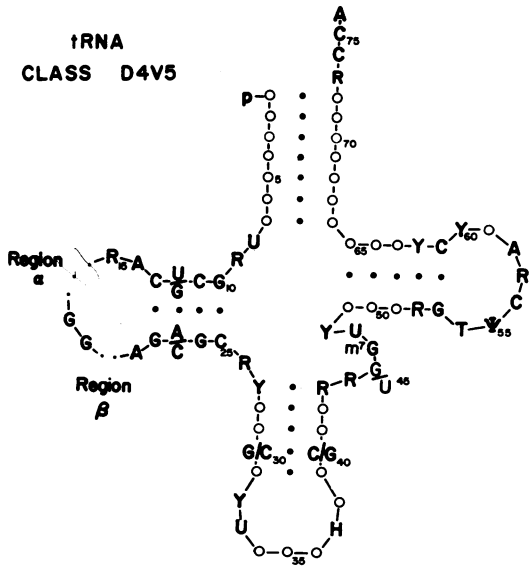

FIG. 6. A cloverleaf diagram illustrating the constant nucleotides, purines or pyrimidines which are found in the class of tRNAs with four base pairs in the D stem and five nucleotides in the variable loop (D4V5). In this diagram R stands for purine, Y for pyrimidine, and H stands for a hypermodified purine. Twenty-eight sequences are included in this diagram (5). Exceptions and comments to the generalizations include the following: R9 is usually A or m^1G; C11-G24 is A11-U24 for *E. coli* $tRNA_f^{met}$ and U11-A24 for *E. coli* $tRNA^{asp}$ and *E. coli* $tRNA^{trp}$ (su^+); C13-G22 is U13-A22 for yeast $tRNA_3^{arg}$; R15 and Y48 are G15 and C48 in 23 sequences, A15 and U48 in 5 sequences; C25 is U in *E. coli* $tRNA_1^{arg}$, *E. coli* $tRNA_1^{asp}$, and in mammalian initiator $tRNA_f^{met}$; R26 is a pyrimidine in 3 of the 28 sequences; Y27-R43 is G27-C43 in *E. coli* $tRNA^{phe}$; U33 is C33 in mammalian initiator $tRNA_f^{met}$; H37 is usually a hypermodified purine; R44 is C in *E. coli* $tRNA_1^{arg}$, U in *Salmonella typhurium* and *E. coli* $tRNA_1^{his}$; m^7 G46 is A in yeast $tRNA_3^{arg}$ and G in yeast $tRNA_1^{ala}$; U47 is always uracil or a uracil derivative; Y62-R52 is A62-U52 for yeast and wheat germ $tRNA^{phe}$, but usually C62-G52; T54, ψ55 and Y60 are A54, U55, and A60 in mammalian initiator $tRNA_f^{met}$; R73 is C for *S. typh.* $tRNA_1^{his}$ and U for haploid yeast $tRNA_1^{lys}$ and *E. coli* $tRNA_3^{gly}$.

position 46 always contains an A and the next residue is analogous to pyrimidine 48 in being complementary to purine 15. The pyrimidine in position 47 of yeast $tRNA^{phe}$ bulges out from the molecule with its base unstacked where it is readily subjected to chemical modification (15). It is not unreasonable to believe that the A in position 46 can form hydrogen bonds with the base in position 22, which is G in four of the sequences. There are no systematic relations involving position 9 and the base may either hydrogen bond with the base pair 12–23 or interact with the A24 found in all sequences.

Class D3VN

There are 15 sequences in this class with variable loops containing 13 to 21 bases. The variable arm probably forms a hydrogen bonded stem region (3 to 7 base pairs) with a small loop at its tip. An interesting correlation can be made between the nucleotides which are found at the top of the anticodon stem. Position 26 is always a purine whereas position 44 is always a pyrimidine; however, they are always anticomplementary and one finds either G-U or A-C. These are structurally identical if there is *trans* base pairing which could occur if purine 26 were in the *syn* conformation with an

altered pucker in its ribose ring. Alternatively, *cis* pairing is also possible here. In agreement with this, the pyrimidine in position 44 is not subjected to chemical modification (15). Such a postulated anti-complementary *trans* pairing may introduce an element of stability compensating for the loss of interactions associated with positions 45 and 46 in yeast $tRNA^{phe}$. It is interesting that position 22 in class D3VN is always occupied by an adenine whereas position 13 is always occupied by a purine. Thus we could have purine·purine hydrogen bonding involving A-A or A-G pairs and, as discussed above, these could be in the *trans* configuration if nucleoside 22 is in the *syn* conformation. At the same time, it is possible that the constant A22 may be involved in hydrogen bonding with the base in position 9. No simple relationship exists between the base in position 9 and the 12–23 pair.

Two sequences are reported with three nucleotides in the variable loop (6). An examination of the structure of yeast $tRNA^{phe}$ makes it rather unlikely that the form of the molecule could be conserved with only three nucleotides in the variable loop. It would be useful if the accuracy of these sequences could be reexamined in order to confirm the existence of only three nucleotides.

DISCUSSION

Elucidation of the three-dimensional tertiary interactions of yeast $tRNA^{phe}$ has revealed the essential role played by many of the constant bases. In addition, a number of other bases which are not constant are involved in tertiary interactions. By surveying the sequences of all tRNAs, we conclude that it is likely that the structure of yeast $tRNA^{phe}$ is a good prototype for understanding the structure of all tRNAs. One of the reasons for believing this is that the variations in the number of nucleotides are spatially localized and can easily be accommodated by the same three dimensional framework. Two of these sites discussed above are the α and β regions in the D loop. Another region of variability is the fourth base pair in the D stem (positions 13–22) and the interactions associated with the nucleotide in position 9 with a base in the D stem. The fact that the major stabilizing interaction in the molecule is the extensive system of stacked purines and pyrimidines suggests to us that this feature of the structure is maintained in those molecules in which the pattern of hydrogen bonding differs somewhat from that found in the yeast $tRNA^{phe}$. It is interesting that changes in the number of nucleotides either in the α or β regions of the D loop or in the length of the variable loop can be accommodated by modifying bulges on the surface of the molecule, thereby allowing the base stacking pattern of the molecule to be maintained despite these alterations in nucleotide number.

It is likely that the form of the molecule seen in the crystal lattice is similar to that seen in a tRNA solution that is biologically active. However, this question is being subjected to many experimental approaches including nuclear magnetic resonance spectroscopy, oligonucleotide binding, and laser raman scattering. An interesting paradox is presented by the work of Erdmann and his associates (16) which strongly suggests that the TψC sequence of tRNAs active in polypeptide chain elongation interacts with a complementary sequence in the 5S RNA in the ribosome. In the crystal structure it is clear that this loop is unavailable for hydrogen bonding; however, the TψC corner of the molecule may open up when it interacts with ribosomal components. The important goal underlying structural analysis of tRNA is the insight which it may ultimately provide into the biological function of the molecule. We are just at the beginning of this process.

A detailed description of the tertiary interactions in the orthorhombic crystals of yeast $tRNA^{phe}$ as described in this paper and in ref. 4 was presented at the American Chemical Society Regional Meeting at Purdue University on June 3, 1974 in a Symposium organized by Professor Struther Arnott.

We wish to thank Karyl Irion, George Church, George Harris, and Sue Bock for their enthusiastic help. This research was supported by grants from the U.S. Public Health Service (CA-15802 and CA-04186), the National Science Foundation (GB40814 and GB30688), the American Cancer Society and the National Aeronautics and Space Administration. J.L.S. is a fellow of the Arthritis Foundation, G.J.Q. is a fellow of The Medical Foundation, and N.C.S. is a fellow of the National Institutes of Health.

1. Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marguisse, S. H., Merrill, J., Penwick, R. & Zamir, A. (1965) *Science* **147**, 1462–1465.
2. Kim, S. H., Quigley, G. J., Suddath, F. L., McPherson, A., Sneden, D., Kim, J. J., Weinzierl, J. & Rich, A. (1973) *Science* **179**, 285–288.
3. Suddath, F. L., Quigley, G. J., McPherson, A., Sneden, D., Kim, J. J., Kim, S. H. & Rich, A. (1974) *Nature* **248**, 20–24.
4. Kim, S. H., Suddath, F. L., Quigley, G. J., McPherson, A., Sussman, J. L., Wang, A. H. J., Seeman, N. C. & Rich, A. (1974) *Science* **185**, 435–440.
5. Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. D., Clark, B. F. C. & Klug, A. (1974) *Nature* **250**, 546–551.
6. A convenient compilation of tRNA sequences can be found in Barrell, B. G. & Clark, B. F. C. (1974) *Handbook of Nucleic Acid Sequences* (Joynson-Bruvvers Ltd., Oxford, England). A large number of sequences are also listed in Holmquist, R., Jukes, T. H. & Pangburn, S. (1973) *J. Mol. Biol.* **78**, 91–116. Some tRNA sequences have been omitted from our compilations.
7. RajBhandary, U. L. & Chang, S. A. (1968) *J. Biol. Chem.* **243**, 598–608.
8. Arnott, S., Hukins, D. W. L. & Dover, S. D. (1972) *Biochem. Biophys. Res. Commun.* **48**, 1392–1399.
9. Fuller, W. & Hodgson, A. (1967) *Nature* **215**, 817–821.
10. Bergstrom, D. E. & Leonard, N. J. (1972) *Biochemistry* **11**, 1–8.
11. Yaniv, M., Favre, A. & Barrell, B. G. (1969) *Nature* **225**, 1331–1333.
12. Chafin, L., Omilianowski, D. R. & Bock, R. M. (1971) *Science* **172**, 854–855.
13. Levitt, M. (1969) *Nature* **224**, 759–763.
14. Rich, A., Davies, D. R., Crick, F. H. C. & Watson, J. D. (1961) *J. Mol. Biol.* **3**, 71–86.
15. Chemical modifications are reviewed by: Cramer, F. (1971) *Progr. Nucleic Acid Res. Mol.* **11**, 391–422; Chang, S. E., Ish-Horowicz, D. (1974) *J. Mol. Biol.* **84**, 375–388.
16. Erdmann, V. A., Sprinzl, M. & Pongs, O. (1973) *Biochem. Biophys. Res. Comm.* **54**, 942–948.