Hydrogen bonding in yeast phenylalanine transfer RNA

(electron density map/base stacking/base pairing/ion interactions)

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

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

Contributed by Alexander Rich, September 29, 1975

ABSTRACT   Further analysis of the three-dimensional electron density map of yeast phenylalanine tRNA is presented. Attention is focused on the several types of unique hydrogen bonding that are found in the molecule and a number of sections of the electron density map are presented. These sections are compared with an electron density map of a dinucleoside phosphate. The bases in the helical stem regions are all involved in Watson-Crick hydrogen bonding interactions with the exception of the guanine-uracil base pair. Several additional tertiary hydrogen bonding interactions are described.

The specificity in the structure and biological function of the nucleic acids is largely due to the highly specific hydrogen bonding that is found in these molecules. This is especially true for transfer RNA (tRNA). The nucleotide sequences of over seventy tRNAs have been determined, and all of them can be organized in the familiar cloverleaf diagram, which is composed of stems and loops, the stems containing regions in which sequences have complementary bases. Three years ago, the x-ray diffraction analysis of orthorhombic crystals of yeast phenylalanine tRNA at 4 Å resolution showed that this tRNA molecule contained the anticipated right-handed, antiparallel double helical regions corresponding to the stems (1). The acceptor and TΨC stems of the molecule were found to be approximately colinear and oriented along one arm of an L-shaped molecule, while the dihydrouridine (D) stem and anticodon stems were oriented nearly at right angles to it. These segments were found joined by a complex intertwining of the D loop and the TΨC loop. At 4 Å resolution the ribose-phosphate backbone could be traced, but it was not possible to resolve the additional tertiary hydrogen bonding. That information was revealed last year at 3 Å resolution both for the orthorhombic (2) and the monoclinic (3) crystal forms of yeast phenylalanine tRNA. These two analyses were very similar, except that the electron density map of the monoclinic crystal was not sufficiently resolved to interpret the tertiary interactions that hold the D-loop and TΨC loop together. We have continued our analysis of the orthorhombic crystals of yeast phenylalanine tRNA and presently have diffraction data to a resolution of 2.5 Å. We have examined in considerable detail the electron density map that was constructed using the multiple isomorphous replacement (MIR) method. Here we document a large number of the hydrogen bonding interactions in yeast phenylalanine tRNA by presenting the MIR electron density maps at 3 Å resolution together with the interpretations that we have placed on this map. In the course of this analysis we have confirmed the hydrogen bonding in-

teractions that were described approximately a year ago (2) and have also discovered some additional interactions that further stabilize the three-dimensional structure of this tRNA.

MATERIALS AND METHODS

The methods for preparing crystals of orthorhombic yeast phenylalanine tRNA have been described (4, 5). The electron density maps were calculated using phases from a number of isomorphous derivatives which have also been described (1, 6). The initial fitting of the electron density map was carried out by building the model in an optical comparitor using the MIR electron density map and partial structure Fourier maps (7). Subsequent to this, refinement procedures were carried out that have aided in some of the interpretation. The details of refinement will be described elsewhere.

Sections of the electron density map were obtained using the coordinates of three points, usually in the plane of the hydrogen bonding bases. Electron density maps were then prepared in the plane of the base pair as well as in two planes at a distance of 1 Å on either side of this plane. These three sets of contours were plotted on the figure and the idealized skeletal model of that part of the molecule was then superimposed.

RESULTS

Yeast phenylalanine transfer RNA is a flattened, L-shaped molecule approximately 25 Å thick and 76 Å between the two ends. Two side views of the molecule are shown schematically in Fig. 1, in which the continuously coiled tube represents the ribose-phosphate backbone and the cross bars represent secondary and tertiary hydrogen bonding interactions between bases. Tertiary interactions are shown in black. All but four of the bases in the molecule are involved in two large stacking domains along the vertical and horizontal axes in Fig. 1.

Small molecule diffraction studies usually have a resolution of 1 Å or smaller and produce an electron density map showing each atom as an individual peak. However, in macromolecular investigations, the resolution is generally not this high and consequently the electron density map usually shows unresolved clusters of atoms. An example of this can be illustrated using data from the crystal structure of a dinucleoside phosphate adenylyl-3',5'-uridine (ApU) (8) in which the phasing is nearly perfect. The structure was solved to 0.8 Å resolution; the molecule forms a right-handed, antiparallel, double helical fragment with hydrogen bonding between adenine and uracil residues. Fig. 2a shows the electron density map of the ApU molecule as seen at a resolution of 3 Å and drawn as described in *Materials and Methods*. Individ-
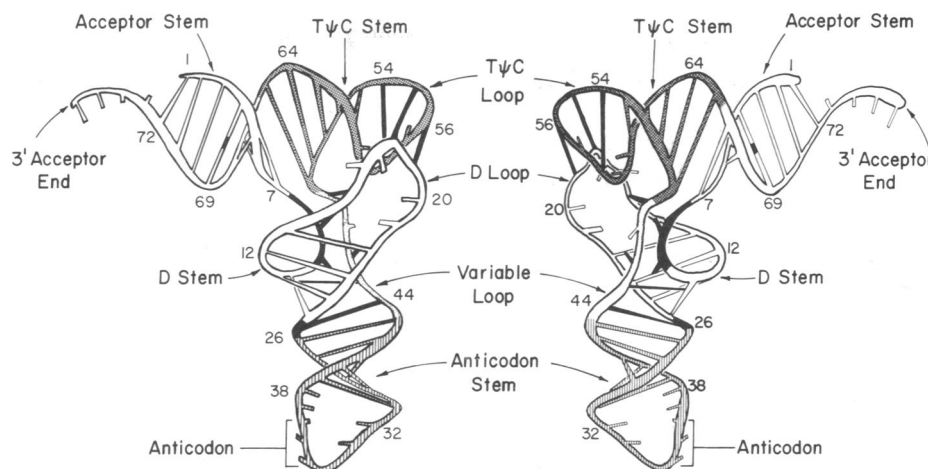
FIG. 1.   A schematic diagram showing two side views of yeast phenylalanine tRNA (2, 9). The ribose-phosphate backbone is depicted as a coiled tube, and the numbers refer to nucleotide residues in the sequence (16). Hydrogen bonding interactions between bases are shown as cross-rungs. Tertiary interactions between bases are shown as solid black lines which indicate either 1, 2, or 3 hydrogen bonds between them as described in the *text*. Those bases that are not involved in hydrogen bonding to other bases are shown as shortened rods attached to the coiled backbone.

ual atoms are not visible, but groups of atoms can be seen. The uracil and adenine bases show up as irregular peaks, and other peaks are associated with the ribose and the phosphate groups. A number of other details are known to be due to water molecules (W) or to sodium ions (8). Fig. 2a represents nearly the best electron density map that one can obtain of an adenosine-uridine Watson-Crick base pair at 3 Å resolution. Fig. 2b–f show sections of the electron density map of yeast phenylalanine tRNA taken through base pairs in the double helical stems. These sections were obtained from an MIR electron density map and are therefore much less perfect. One can see that the A·U pairs in Fig. 2b–d are

not as smoothly contoured as those seen in Fig. 2a; nonetheless they are similar. In general, peaks are found for the base, ribose, and phosphate groups. The A·U bases tend to remain separate from each other; the G·C pairs as seen in Fig. 2e and f are often somewhat less clearly resolved from each other with the electron density tending to merge across the hydrogen bonded region. A number of additional peaks are seen in the sections. Some of these are associated with other parts of the molecule and some of them are due to ions and solvent which may be resolved more clearly at higher resolution and in the refinement analysis. In general the peaks associated with pyrimidines tend to be somewhat

(a)   (b)   (c)
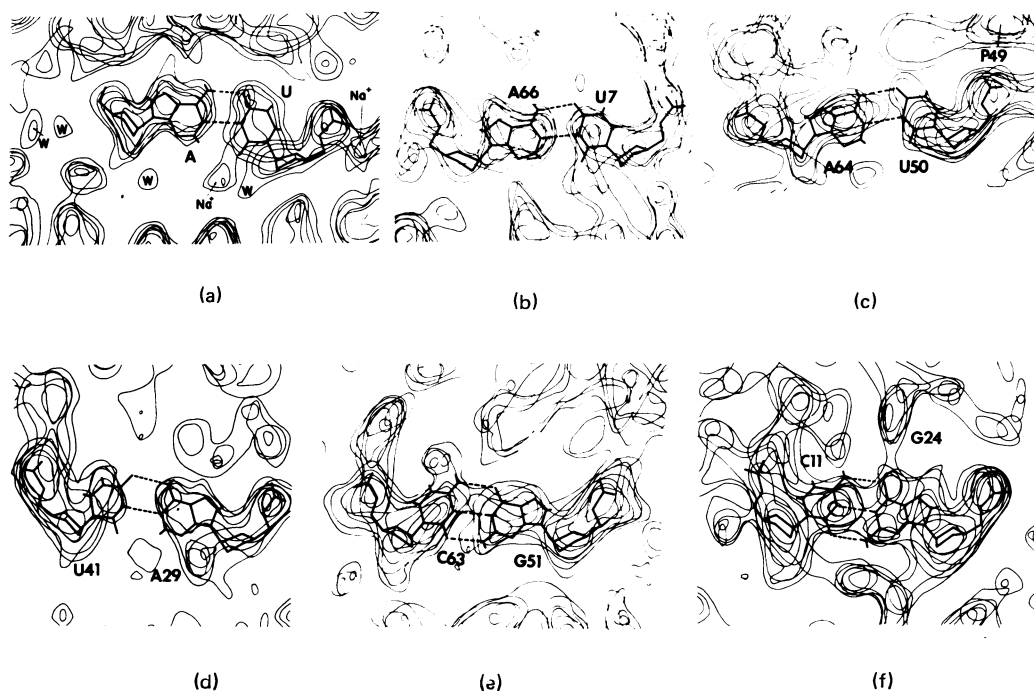
(d)   (e)   (f)

FIG. 2.   Electron density maps showing base pairing. (a) A section of the electron density map of the single crystal structure of ApU (8). Although the structure was solved to atomic resolution, the electron density map was plotted at 3 Å resolution. The position of sodium ions and water molecules (W) are indicated. (b–f) Sections of the electron density maps from the helical stem regions of yeast phenylalanine tRNA. The numbers correspond to the residue number in the nucleotide sequence (16), as used in Fig. 1. In general, A·U base pairs are more clearly resolved from each other at this resolution than are the G·C base pairs.
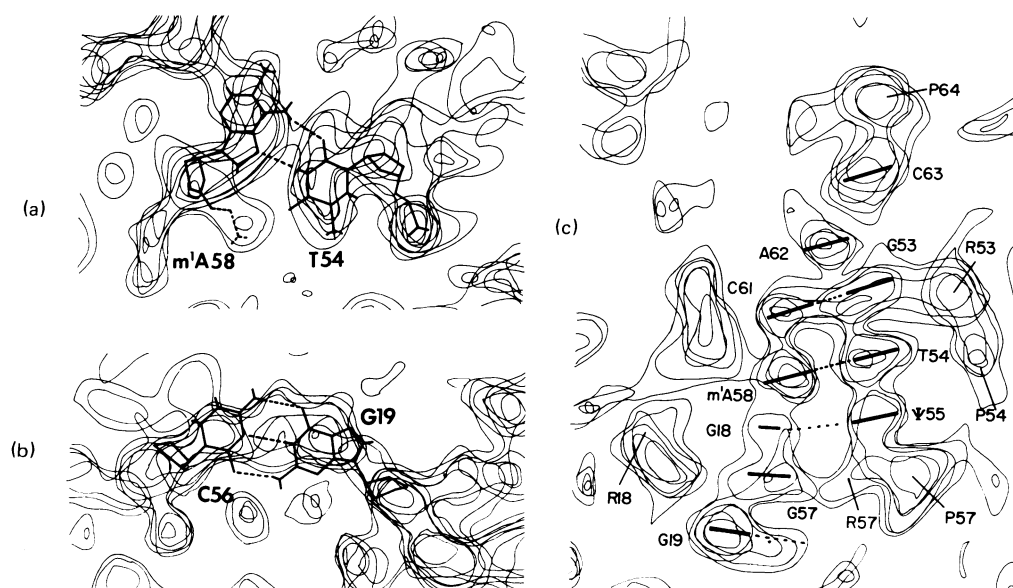
FIG. 3. Electron density map from the TΨC loop. (a) and (b) show two of the base pairs in the TΨC loop. (c) cuts through the plane of the base pair T54·m¹A58 and is a composite of two parallel planes 1 Å apart. The numbers indicate the various residues. In this section the bases are seen approximately edge-on, and therefore are shown schematically as a single solid line, with dashes indicating hydrogen bonding. P stands for the phosphate group and R for ribose.

smaller than those associated with purines. This observation serves to confirm that the correct assignments are being made.

As described previously (2, 9), the TΨC loop is stabilized by four different levels of stacking interactions involving the hydrogen bonded pair T54 and m¹A58, Ψ55 and G18 (of the D-loop), a stacking interaction involving G57, and finally the hydrogen bonded pair C56 and G19 (of the D-loop). Further details of the interaction between the TΨC and D loops are shown in Fig. 3a and b. As pointed out (2), there is a reversed Hoogsteen interaction between T54 and m¹A58 and a normal Watson-Crick interaction between G19 and C56. Fig. 3c shows a plane cutting through the plane of base pair T54·m¹A58. The electron density map shows the stacked interaction of the hydrogen bonding pairs in the TΨC loop as well as some bases of the D loop. The section cuts through the last base pair of the TΨC stem, C61·G53, and the adjacent pair T54·m¹A58. These are shown diagrammatically by the heavy lines, with dots indicating hydrogen bonds. The pairing between Ψ55 and G18 is indicated in the diagram even though the section is no longer passing through the center of this base pair as a result of the helical twist of the loop. The peak due to G57 is indicated, as well as that due to G19. The dotted line at the bottom right extending from G19 is going towards C56, which is not in the plane of the section. This figure shows the four levels of stacking interaction that stabilize the end of the TΨC loop and also shows the stacking interaction of three guanosine residues, which appears to play a significant role in stabilizing the interaction of the D and TΨC loops. Most of the bases shown in this corner of the molecule in Fig. 3c are common to all tRNAs involved in polypeptide chain elongation. G18 and G19 are the two constant guanosines that are found in the D loop of all tRNA sequences, and residue 57 is always a purine. Thus it is likely that these stacking interactions will be found in all tRNA molecules (9).

The interaction between Ψ55 and G18 is of an unusual nature. It was previously stated (2) that these bases are held together by either one or two hydrogen bonds. Further anal-

ysis suggests that there may be only one hydrogen bond between these two bases, involving the N2 or N1 of G18, forming a hydrogen bond to O4 of Ψ55. In addition, the N3 of Ψ55 is probably forming a single hydrogen bond to the oxygen of phosphate P58 across the loop. The Ψ thus appears to be hydrogen bonding with both the backbone on the other side of the loop as well as with the guanine base. It is interesting to note that phosphate 57 is lying below the plane of Ψ55, which terminates the stacking interaction in that direction (Fig. 3c). Further refinement will be necessary to specify all of the interactions in this interesting arrangement.

Some of the most unusual hydrogen bonding occurs in the center of the molecule (2, 3), where the ribose-phosphate chain is elongated and makes a very sharp bend in the region of residues 7 to 10 (Fig. 1). This folding of the polynucleotide chain is stabilized by several hydrogen bonding interactions involving nucleotides that are constant in all tRNAs, including hydrogen bonding to the wide groove of the D stem. Fig. 4 shows electron density sections of four consecutive levels of the molecule containing residues 15, 14, 13, and 12. The sections in Fig. 4 are drawn so that the bottom of each section is in an orientation approximately on the front surface of the molecule as seen in the left-hand side of Fig. 1. Fig. 4a shows the *trans* pairing of C48 from the extra loop with G15 of the D loop. The two chains are running parallel to each other at that point (Fig. 1), and although these are not constant nucleotides, there is always a purine found in position 15 and usually a complementary pyrimidine found in position 48, which is the first position in the extra loop (10). Fig. 4b shows the reversed Hoogsteen pairing between constant residues U8 and A14. This pairing leaves the uridine O4 position free. Residue A21 is found in the same plane as the U8·A14 pair. However, it can be seen in Fig. 4b that it is not in a position to form hydrogen bonds with the bases, but rather is in a position to form hydrogen bonds with the ribose of residue U8 (2). A portion of G20 intersects that plane; however, the guanosine ring is oriented almost at right angles to the other bases shown in section 4b. As shown earlier (11), this guanosine residue is easily modi-

Biochemistry: Quigley *et al.*
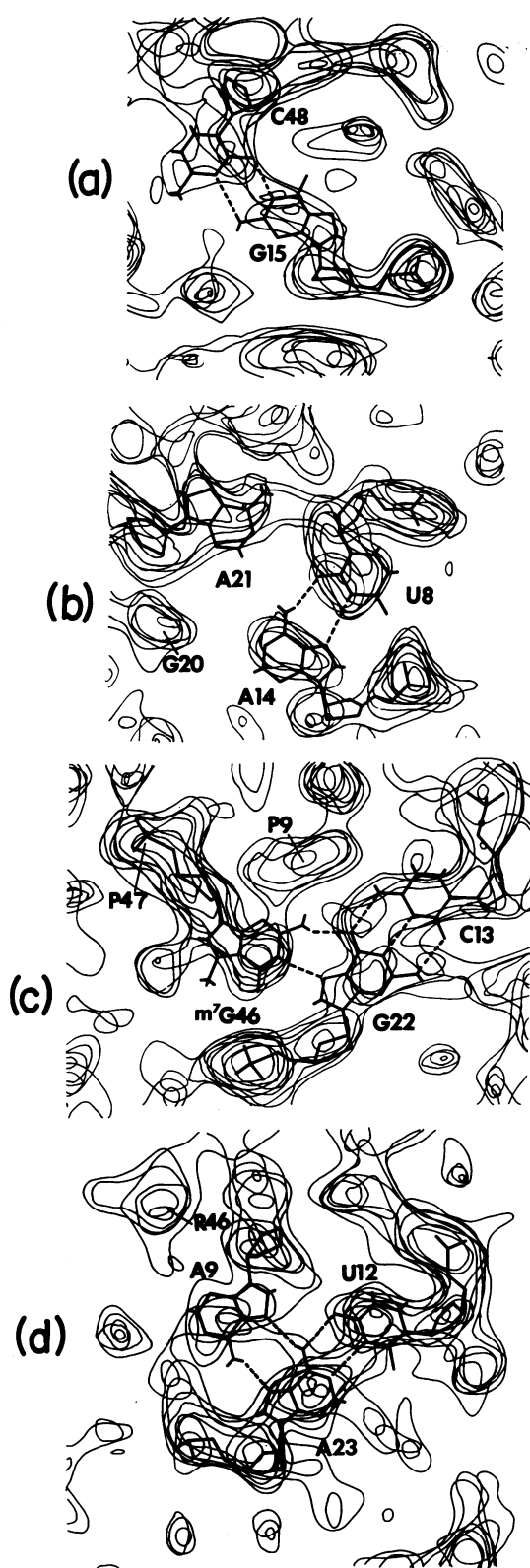
Proc. Nat. Acad. Sci. USA 72 (1975)   4869

FIG. 4. A series of sections taken from the center of the molecule showing the base pairing in the D stem and adjoining regions. (a) and (b) are base planes immediately adjoining the D stem, while (c) and (d) show base pairs of the D stem together with hydrogen bonding of bases in the major groove of the D stem helix. (b) shows adenosine 21 in a position to hydrogen bond with the ribose of uridine 8 although the hydrogen bonding is not drawn in.

fied by kethoxal. Thus it is not surprising that G20 is among the few bases not involved in stacking or hydrogen bonding interactions with the other bases of the molecule. However, it is involved in crystal packing by stacking on C56 of an adjacent molecule.

Fig. 4c shows the D stem base pair C13·G22, with the additional hydrogen bonding of m^7G46 from the extra loop interacting with G22 in the wide groove of the double helix. It can be seen that the positively charged m^7G46 is shown with several phosphate groups, including the phosphate of A9. Fig. 4d shows the D stem pair A23·U12, together with the hydrogen bonding of residue A9, which is pairing with A23 in a manner similar to that found in double helical polyriboadenylic acid (12).

An additional hydrogen bonding interaction is found in the wide groove of the D stem. In our earlier paper we described residue G45 as being directed towards the wide groove of the D stem at an angle to its base pairs (2). Further analysis showed that G45 forms a single hydrogen bond from its amino group N2 to O6 of G10. Thus, three of the four base pairs in the D stem of yeast phenylalanine tRNA are involved in hydrogen bonding interactions with other bases, two with bases from the variable loop and one with A9, near the point where the polynucleotide chain turns an abrupt corner.

As shown in Fig. 1, the ribose-phosphate chains come very close together in the center of the molecule just before they join the TΨC and the acceptor stems. At this point the ribose-phosphate chains are running in opposite directions and ribose 8 is very close to ribose 48. This interaction is stabilized by hydrogen bonding between these two ribose residues.

A number of ions are seen in the electron density map; two of them are of special interest and will be described here. These are ion sites (13) that are occupied by the samarium ions, and are presumably normally occupied by magnesium ions. Both of these ions are positioned in a manner so that they can interact with a string of phosphate groups where the ribose-phosphate chain curves around a corner. One of these atoms [position 2 (13)] is found in the position where the ribose-phosphate chain turns a sharp corner, near residues 8, 9, 10, and 11. The samarium ion is found in that curve, where it can interact primarily with phosphates 8 and 9. A similar position is found for another samarium ion [position 3 (13)], where it interacts with the phosphates of residue 17 of one molecule and the phosphates of residues 20 and 21 of an adjacent molecule at a point in the D loop where the ribose-phosphate chain has a sharp bend. It is likely that the folding of the polynucleotide chain is stabilized by the cations found in these positions.

Studies are being carried out on the conformation of the ribose residues in this molecule. Even though the electron density map does not fix the positions of the atoms of the ribose ring, the conformation can be inferred by the relative positions of the peaks due to the two phosphate groups attached to the ribose as well as the base. In the normal C3′ endo conformation (14), these three peaks have a different configuration about the ribose peak than they have in the C2′ endo conformation. The C3′ endo conformation is found in the vast majority of the residues in the yeast phenylalanine tRNA, particularly in the double helical regions; however, a significant number of C2′ endo conformations have also been discovered. Their nature and distribution will be discussed in association with the refinement procedures carried out on this molecule.

DISCUSSION

In this paper we show the fitting of the model for some of the hydrogen bonding interactions found in orthorhombic crystals of yeast phenylalanine tRNA (2, 9) to the MIR electron density map obtained at 3 Å resolution. A number of new features of the molecule are described in addition to those which have been discussed earlier (2, 9).

It is of considerable interest to compare the hydrogen bonding observed in the orthorhombic crystals of yeast phenylalanine tRNA with those which have been described for the monoclinic crystal form. Unfortunately the published interpretation of the monoclinic electron density map was incomplete in that the nature of the vital interactions between the D loop and the TΨC loop at the corner of the molecule were not reported (3). Most of the other interactions, however, are similar in the two crystal forms. There appear to be some differences in the type of hydrogen bonding that is seen at the joining of the anticodon stem and the D stem, and the molecular packing constraints in the two unit cells suggest some differences in the anticodon stem or loop (15). The similarity of the molecule in these two different crystal forms suggests that there is indeed one stable conformation of the molecule observed here.

Further work will be necessary to complete the characterization of tRNA. At present, we are continuing to collect diffraction data out to the limit of the diffraction pattern, which is approximately 2 Å. Refinement will be continued with the aim of accurately defining the three-dimensional conformation of this species of tRNA. As pointed out earlier (9), it is likely that the three-dimensional structure of yeast phenylalanine tRNA will serve as a useful model for interpreting the three-dimensional structure of all transfer RNA species. Thus this information may provide the useful groundwork necessary for understanding the biological mechanism of transfer RNA function.

Note Added in Proof. The coordinates of yeast tRNA[Phe] in the monoclinic cell have recently been published [Ladner, J. E., Jack, A., Robertus, J. D., Brown, R. S., Rhodes, D., Clark, B. F. C. & Klug, A. (1975) *Nucleic Acids Res.* 2, 1629–1637]. It appears that the joining of the D loop and the TΨC loop in the monoclinic cell is essentially the same as that which we described for the orthorhombic cell (2) which is further discussed above. Furthermore, Ladner *et al.* have reinterpreted the region where the D stem and anticodon stems join, so that instead of having m_2^2G26 "partially intercalated" between A44 and G45 (3), they now have a conformation similar to that reported for the orthorhombic lattice (2), with hydrogen bonding distances between m_2^2G26 and A44. Coordinates from the orthorhombic lattice obtained by two independent methods of refinement have been submitted for publication (Quigley, G. J., Seeman, N. C., Wang, A. H.-J., Suddath, F. L. & Rich, A., *Nucle-*

ic Acids Res., in press; Sussman, J. L. & Kim, S. H., *Biochem. Biophys. Res. Commun.*, in press), and they show that the conformation of yeast tRNA[Phe] is essentially indistinguishable in the orthorhombic and the monoclinic forms at this resolution, with the exception of a few poorly resolved residues, especially at the 3′-OH ends of the molecules. Some further observations are presented (Quigley *et al.*, *Nucleic Acids Res*, in press) about the structural role of the constant residues in tRNA sequences.

1. 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.
2. 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.
3. Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. S., Clark, B. F. C. & Klug, A. (1974) *Nature* 250, 546–551.
4. Kim, S. H., Quigley, G., Suddath, F. L. & Rich, A. (1971) *Proc. Nat. Acad. Sci. USA* 68, 841–845.
5. Kim, S. H., Quigley, G., Suddath, F. L., McPherson, A., Sneden, D., Kim, J. J., Weinzierl, J. & Rich, A. (1973) *J. Mol. Biol.* 75, 421.
6. Kim, S. H., Quigley, G., Suddath, F. L., McPherson, A., Sneden, D., Kim, J. J., Weinzierl, J., Blattmann, P. & Rich, A. (1972) *Proc. Nat. Acad. Sci. USA* 69, 3746–3750.
7. Sussman, J. L. & Kim, S. H. (1975) *Eighth Jerusalem Symposium: Environmental Effects on Molecular Structure and Properties*, ed. Pullman, B., in press.
8. Rosenberg, J. M., Seeman, N. C., Kim, J. J. P., Suddath, F. L., Nicholas, H. B. & Rich, A. (1973) *Nature* 243, 150–154.
9. Kim, S. H., Sussman, J. L., Suddath, F. L., Quigley, G. J., McPherson, A., Wang, A. H. J., Seeman, N. C. & Rich, A. (1974) *Proc. Nat. Acad. Sci. USA* 71, 4970–4974.
10. Levitt, M. (1969) *Nature* 224, 759–763.
11. Litt, M. (1969) *Biochemistry* 8, 3249—3256.
12. Rich, A., Davies, D. R., Crick, F. H. C. & Watson, J. D. (1961) *J. Mol. Biol.* 3, 71–86.
13. Suddath, F. L., Quigley, G. J., McPherson, A., Sneden, D., Kim, J. J., Kim, S. H. & Rich, A. (1974) *Nature* 248, 20–24.
14. Seeman, N. C., Sussman, J. L., Berman, H. M. & Kim, S. H. (1971) *Nature New Biol.* 233, 90–92.
15. Quigley, G. J., Suddath, F. L., McPherson, A., Kim, J. J., Sneden, D. & Rich, A. (1974) *Proc. Nat. Acad. Sci. USA* 71, 2146–2150.
16. RajBhandary, U. L. & Chang, S. A. (1968) *J. Biol. Chem.* 243, 598–608.