Proc. Nat. Acad. Sci. USA
Vol. 72, No. 11, pp. 4414–4418, November 1975
Biophysics

Structure of yeast phenylalanine transfer RNA at 2.5 Å resolution

(x-ray diffraction/tRNA conformation/hydrogen bonding)

J. E. LADNER, A. JACK, J. D. ROBERTUS*, R. S. BROWN, D. RHODES, B. F. C. CLARK†, AND A. KLUG

MRC Laboratory of Molecular Biology, Cambridge CB2 2QH, England

Communicated by F. H. C. Crick, August 11, 1975

ABSTRACT **The x-ray analysis of the monoclinic form of yeast $tRNA^{Phe}$ has been taken to a resolution of 2.5 Å by the method of isomorphous replacement. The model proposed at 3 Å has been confirmed and extended to reveal additional features of the tertiary structure and of the stereochemistry. An extensive hydrogen bonding network is described involving specific interactions between bases and the ribose-phosphate backbone. The structure of a G-U base pair has been solved.**

In this paper we present the second stage of the x-ray analysis of the monoclinic form of yeast phenylalanine tRNA. A year ago we proposed a molecular model based on the interpretation of an electron density map at 3 Å resolution calculated with phases obtained from isomorphous replacement (1). A similar model was also proposed by Kim *et al.* (2) for the closely related orthorhombic form. Our map was of sufficient quality to trace the ribose-phosphate chain and to assign all except five nucleotides to peaks in the electron density. These nucleotides lay in a corner of the molecule where the density was strong, but could not be interpreted with as much confidence as the rest of the structure because of tight intra- and intermolecular packing, and two alternative chain tracings were given. In the rest of the molecule we were able to define unambiguously a number of base-base interactions involved in maintaining the tertiary structure, and to describe the extensive stacking interactions present. Many of these interactions could be related to invariances or semi-invariances in the generalized nucleotide sequence of tRNA (Fig. 1). We showed how other species of tRNA could be accommodated in the same tertiary structure by coordinated changes of sequence (1, 3). The results of a companion study of the chemical reactivity of yeast $tRNA^{Phe}$ were found to be in good accord with the model (4, 5).

The x-ray analysis has now been extended to 2.5 Å by the method of isomorphous replacement using six heavy atom derivatives to produce an improved electron density map. The structure reported previously (1) has been confirmed and the tracing of the chain in the uncertain region is now unambiguous. The crystallographic discrepancy index, R, has been reduced from 46 to 39%, a value as good as that for protein structure determinations at a comparable stage. The atomic coordinates are unlikely to suffer any large changes in subsequent refinement, and are therefore being published (8).

The 2.5 Å model allows us to complete the detailed description of base-base interactions in the molecule (Fig. 2). It has also revealed a large number of base-backbone interactions and a new semi-invariant base pair. A hydrogen-bonding network runs through most of the molecule, apart from the exposed amino-acid and anticodon stems. A number of the ribose groups in the nonhelical regions are of the 2′-*endo* type rather than the 3′-*endo* found in RNA helices. We now have a picture of the ordered complexity of a folded RNA molecule, a complexity as great as that of a protein. An interesting by-product is a detailed picture of the stereochemistry of a G-U base pair in a double helical stem, where the pairing is that predicted by the "wobble" hypothesis (9).

X-ray analysis

The resolution of the isomorphous replacement map was extended from 3 Å to 2.5 Å by collecting the strongest 40% of the x-ray reflections. Five of the heavy atom derivatives are as listed before (1) and the sixth was made by soaking in hydroxymercuri-hydroquinone-*OO*-diacetate (10). Heavy atom refinements, computation of the map, and building of the wire model were carried out as before (1). Some data on the quality of the x-ray analysis are given in Table 1. Atomic coordinates, measured by hand from the wire model, gave an R index of 41%. Bad contacts and major measurement errors were removed by the energy refinement method of Levitt (11), thereby reducing the R index to 39%.

General features of structure at 2.5 Å resolution

In our first account of the tertiary structure (1), we focussed attention on the interactions between bases. We mentioned only a few of the other H-bonding interactions, deferring discussion until they could be checked at 2.5 Å. We have now traced an extensive network of H-bonds, including ribose and phosphate groups as well as bases, in which almost half the bonds involve 2′-OH groups of riboses as acceptors, donors, or both. A prominent set of H-bonds stabilizes the T-join made in the structure by the two long double helices which meet approximately at right-angles. Another set reinforces the links between bases at the bends of the D and TΨC loops.

In the 3 Å model all sugar conformations were assumed to be C3′-*endo*, as found in RNA double helices (12). In the more detailed model building occasioned by fitting the 2.5 Å map, we have changed the pucker of ten of the sugars to the C2′-*endo* conformation (13). Although the resolution of the map is still not sufficient to show the shape of a sugar unequivocally, the conformation can be deduced from the restrictions imposed on it by the relative dispositions of the two phosphate groups on either side and of the base that emanates from it. A good example is that of nucleotide m^1A58 whose sugar is C2′-*endo* in the 2.5 Å model. In the 3 Å model the phosphate and base densities were well fitted, but the C3′-*endo* ribose was not. Fitting the ribose density moved the base of 58 too close to its base paired partner, T54, which was well fixed. Now, the C2′-*endo* sugar resides in its proper density and the base pair is comfortably made (Fig. 4).

Abbreviation: R index, crystallographic discrepancy index.

* Present address: Department of Chemistry, University of Texas, Austin, Texas 78712.
† Present address: Department of Chemistry, University of Aarhus, Denmark.

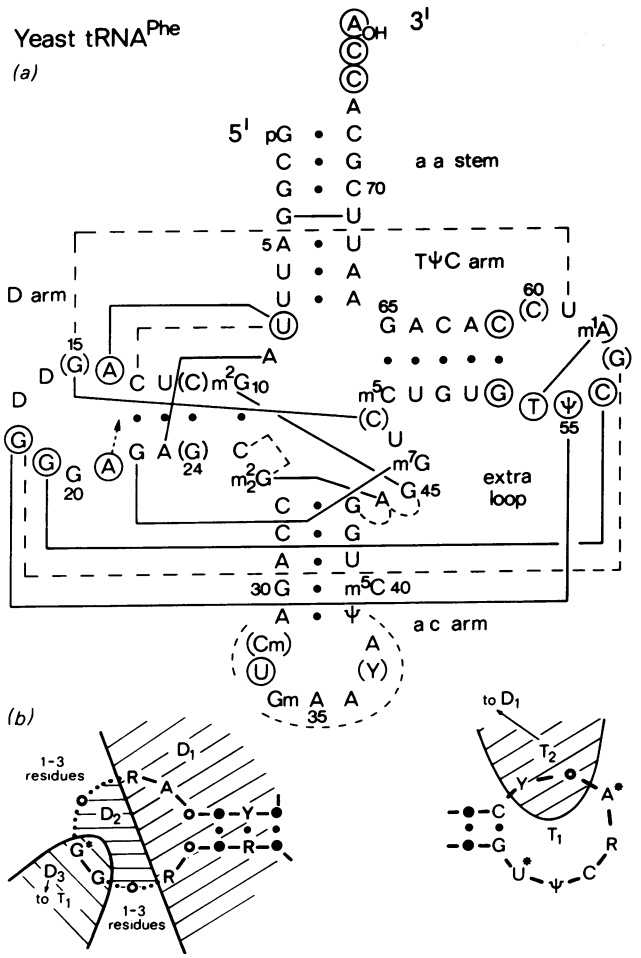


FIG. 1. (a) The nucleotide sequence of yeast $tRNA^{Phe}$ (6) arranged in the cloverleaf formula. Bases invariant in all tRNA sequences (7) are circled; semi-invariants, i.e., purines or pyrimidines exclusively, are in parentheses. Bases paired in the tertiary structure are joined by solid lines, and those which stack on each other by dashed lines. A21 (dashed arrow) is discussed in the *text*.

(b) Generalized formulae for the D-arm (left) and TΨC loop (right), with divisions indicated by the tertiary structure: R, purine; Y, pyrimidine.

G-U base pair

Yeast $tRNA^{Phe}$ has a base pair G4-U69 in the amino-acid stem and hence affords the first opportunity for observing directly what happens when a G and U face each other on opposite strands of a nucleic acid double helix. In his "wobble" theory, Crick (9) predicted that the glycosyl bonds would suffer a relative displacement to allow the two bases to make two H-bonds. The electron density for the bases G4 and U69 is quite good (Fig. 3). Their relative position indicates that two H-bonds are formed between them and that no tautomeric shifts are necessary. The backbone density, although weak toward the CCA end, is sufficient to show that the "wobble" in forming the G-U pair is accommodated locally with acceptable rotations in the backbone bonds (Fig. 3, inset). These rotations produce *trans* conformations (rather than the normal *gauche* +) for the C4′-C5′ bonds of the two sugars on either side of phosphate 5, which is itself displaced by about 2 Å from its normal position in a regular double helix.

G-U base pairs are quite frequent in the cloverleaf structure (14). They are often found in the middle of the aa stem,

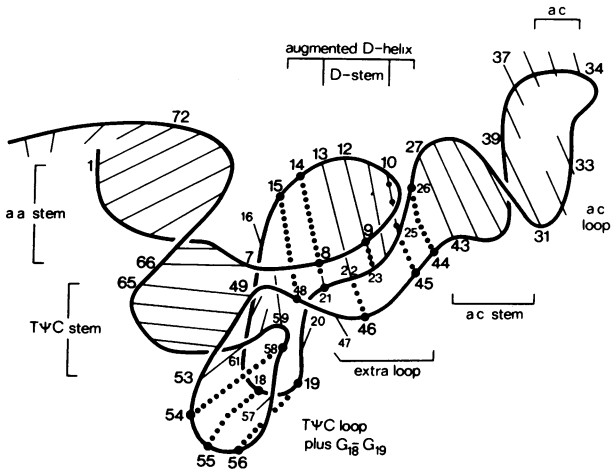


FIG. 2. A schematic diagram of the chain folding and tertiary interactions between bases in yeast $tRNA^{Phe}$. The ribose-phosphate backbone is shown as a continuous line. Base pairs in the double helical stems are represented by long light lines and nonpaired bases by shorter lines. Base pairs additional to those in the cloverleaf formula are represented by dotted lines.

as in yeast RNA^{Phe}, and may function as a recognition site in this exposed stem. The commonest position for a G-U pair is at the junction of aa and TΨC helices, i.e., at 49–65 in Fig. 2. Here its role could be structural, to lessen the strain at the sharp bend between residues 48 and 49. Indeed in our model, phosphate 49 is displaced quite appreciably from its helical position to avoid the bad contacts which would otherwise be made.

Invariant and semi-invariant bases

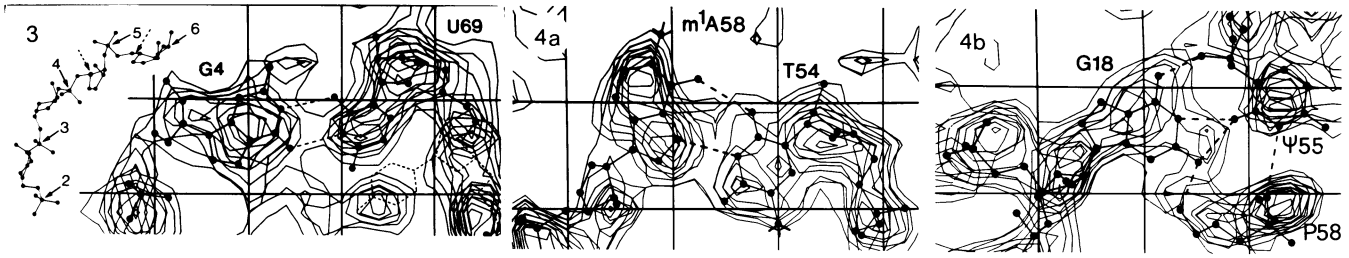
A further role for invariant or semi-invariant bases in the tertiary structure is to fix the conformation of the backbone in tight regions. The cases of A21, G18, and Ψ55 are discussed below. A further example is the base in position 61 in yeast $tRNA^{Phe}$ which is a C in all known sequences. The N4 of C61 hydrogen bonds to an oxygen of phosphate 60 and thus is part of the H-bonding network which anchors this part of the TΨC loop at the end of the augmented D helix.

A new semi-invariance has come to light. The 2′OH of ribose 9 is the acceptor of a hydrogen bond from N4 of base C11, a bond made possible by a 2′-*endo* conformation of the ribose ring. This 2′OH could just as easily make an H-bond

Table 1. X-ray refinement statistics to 2.5 Å

Derivative	rms f_H	rms E	$P = f_H/E$	n
trans-$Cl_2(NH_3)_2Pt$	55.3	42.4	1.31	5505
Lu^{3+}	70.9	53.2	1.33	3843
Sm^{3+}	50.8	43.0	1.18	6035
Os-ATP complex	40.4	41.4	0.98	5870
Pt + Lu	58.0	48.5	1.19	5242
Hg (HQDA)	43.2	48.8	0.88	5595

f_H is the calculated heavy atom structure factor. E is the closure error, f_H (obs:) − f_H (calc.). P is the "phasing efficiency" of a derivative. n is the number of reflections used in the refinement. HQDA is hydroxymercuri-hydroquinone-*OO*-diacetate. The previous refinements at 3 Å (1) were done using the most probable phase angles, but in repeating and extending these computations, "best" phase angles were used, as it was found that this reduced the dominance of the most strongly occupied derivatives on the phasing. The final figure of merit for 6139 reflections from 20 to 2.5 Å was 0.59.



FIGS. 3 AND 4. (*3*) The G-U base pair. A composite of three map sections 1 Å apart is shown, but the superposed atomic skeleton is shown over a greater depth to provide continuity. Grid squares are of side 5 Å. *Inset:* the backbone around the G residue showing the movement of phosphate 5; dashed arrows mark riboses with *trans* configurations. (*4*) Two of the base pairs involved in forming the "conserved cluster." G18 donates hydrogens from both N1 and N2 to the O4 of Ψ55 and from N2 to O1′ of m¹A58. Its O6 receives from the 2′-OH of ribose 55. In addition, but not shown here, ribose 2′OH of G18 donates to an O of phosphate 19 and receives from N2 of G57.

to a U in position 11 by donating to atom O4 of the U, and, indeed, one finds that all known tRNA sequences (7) do have either a C or a U at position 11. It may be relevant that the pair 11-24 is the site of the mutation in the tryptophan suppressor tRNA from U-G to U-A (15).

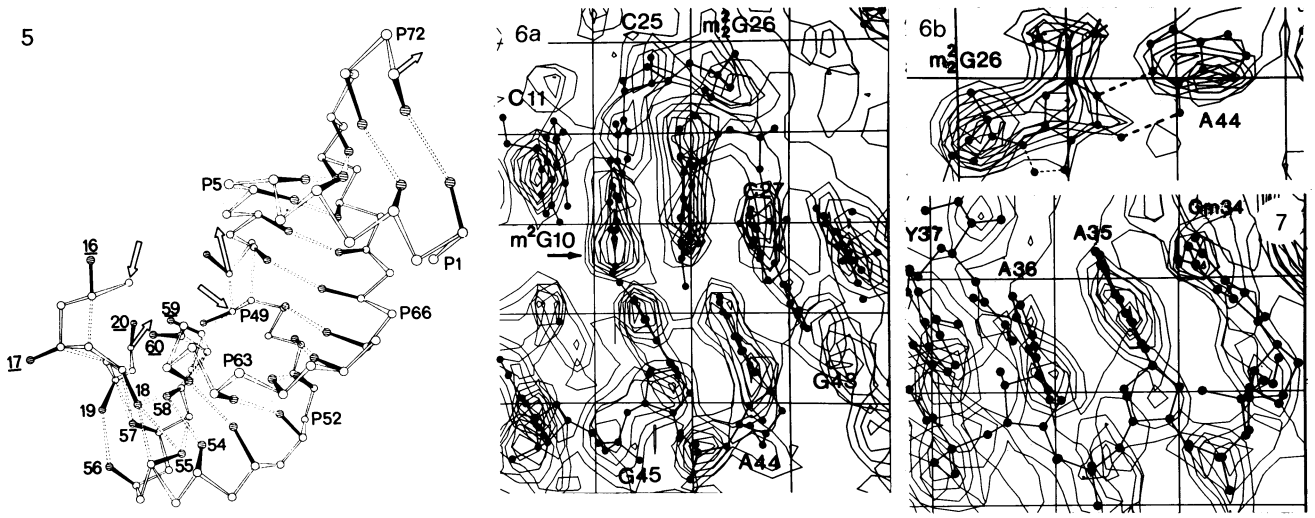
Augmented D-helix, T-join, and "variable pocket"

We have already described the major skeletal substructures in the molecule (1). The amino-acid stem and the TΨC stem are stacked to form a long, double helix with a "break" between residues 7 and 49 (Fig. 2). Approximately at right angles to this break lies the augmented D helix (1) forming the central part of the molecule. This consists of the D stem, extended in length by the residues 14, 15, and 59 (to the left in Fig. 2) and by 26 (to the right), and augmented laterally by interactions with the extended sequence 8-9 and with residues 48 and 46 of the extra loop.

The augmented D helix contains the base triples A9-U12-A23, C13-G22-m⁷G46, and the non-standard pairs U8-A14 and G15-C48. There was a puzzle about A21 which, although making an H-bond to the ribose of U8, appeared not to form a pair of stereochemically acceptable H-bonds to U8-A14 (1). Now the best fit to the new map still has the base A21 turned so that its active groups do not point to the bases U8 and A14, but rather make H-bonds to riboses 8 and 48. The latest report of the orthorhombic structure (16) now also has A21 hydrogen bonded to ribose 8, rather than to the base pair 8-14 as first stated (2).

The axis of the augmented D helix is roughly perpendicular to the amino-acid + TΨC helix and joins it near its center, so that the two are arranged like the letter T. On the near side (as viewed in Fig. 2) of this T-join there is an intricate hydrogen bonding network crosslinking the backbone at residues 59 of the TΨC loop, 48 of the extra loop, and 7 and 49 on each side of the "break". A21 also participates, as just mentioned.

On the far side of the T-join there is, by contrast, less H-bonding. Here the residues D16 and D17 arch out from the molecular surface and together with G20 form a patch of variable and exposed bases (1). Residues U59 and C60, which are, so to speak, expelled from the TΨC loop, also lie in this neighborhood, but somewhat deeper into the molecular body, so that the five bases together form a pocket (Fig. 5). This grouping of the variable bases of the D and TΨC loops suggests that they form a protein recognition site, or part of some discrimination mechanism for different tRNAs.



FIGS. 5–7. (*5*) The segregation of invariant and variable bases in the D and TΨC loops. The aa + TΨC arms and part of the D loop are shown as viewed looking down from a point to the left of the aa stem in Fig. 2. The variable bases (underlined) fall above and to the left of the conserved cluster of invariant bases 18, 19, and 54-58. Bases are represented by hatched circles and hydrogen bonds by dashed lines. Empty arrows indicate points where the chain would continue to the other half of the molecule. (*6a*) Edge-on views of the bases at the junction of the D and anticodon stems. A composite of four map sections, 2 Å apart, normal to the *a* axis is shown. (*6b*) A slice in a plane at right angles to (a), showing one of the base pairs in the junction. (*7*) Edge-on views of the bases of the anticodon. Only part of the density of the adjoining Y base is visible. The contours at the top right belong to the TΨC stem of an adjacent molecule. The slab of density is 8 Å deep and cut normal to the *a* axis.

Interaction of TΨC and D loops

The 2.5 Å map shows that the five uncertain residues follow the first of the alternative paths described (1) and appear to be in positions similar to those proposed for the orthorhombic form (2). Our knowledge of this one tertiary structure and of common features in tRNA sequences prompt us to divide the D loop into three parts (Fig. 1b). The first division, D_1, contains the invariant or semi-invariant residues 14, 15, and 21 and augments the D stem. The second division, D_2, is variable in number and in yeast $tRNA^{Phe}$ contains D16, D17, and G20. These bases together with U59 and C60 form the variable pocket described above. The third division, D_3, consists of the two invariant Gs, numbered 18 and 19 in yeast $tRNA^{Phe}$. In the structure they become integrated with five residues of the TΨC loop (T_1 in Fig. 1b), the two stretches reciprocally stabilizing one another. This cluster of conserved bases forms a short stack which extends below the TΨC stem (Figs. 2 and 5).

The base G18 is used to almost its full hydrogen bonding capacity. It is H-bonded to both the base and ribose of Ψ55, and also to ribose 58, a residue on the opposite side of the TΨC loop from 55 (Fig. 4). The base G18 of the D loop thus spans the two sides of the TΨC loop, helping the chain make the tight turn at residue C56 at the bottom of the loop. The base C56 is itself held in place by three hydrogen bonds to G19 in the only Watson–Crick pair found outside the helices. Further, O2′ of ribose 18 and O1′ of ribose 19 are H-bonded to the amino N2 of the base G57, which is intercalated between G18 and G19 (1). Thus, reciprocally, through G57, the TΨC loop helps the D loop make its turn.

In the TΨC loop itself, T54 is hydrogen bonded to m^1A58 (Fig. 4). The base m^1A58 stacks on top of the three bases G18, G57, and G19 to form a stack of four purines. The interactions of Ψ55, C56, and G57 have been described above. It may be significant that the unique potential H-bonding site N1 of Ψ does not appear to be used in the tertiary structure. Finally, the bases U59 and C60 are oriented quite differently from the rest of the bases in the loop. U59 stacks on the pair G15-C48, and together with C60, appears to point towards an ion, but otherwise remains accessible in the arch formed by D16 and D17. Ribose 59 reaches across to ribose 48 to form an H-bond between the two 2′O's. Phosphate 60 is firmly anchored to the 2′OH of m^1A58 and to N4 of C61.

Junction of anticodon and D helices

The anticodon stem is not simply continued on from the D stem, as is the amino-acid stem from the TΨC stem. Unlike the situation at the latter junction, there is always a lone base here in the sequence, m_2^2G26 in yeast $tRNA^{Phe}$. In the tertiary structure it lies between the base C27 at the end of the anticodon stem and the pair m^2G10-C25 at the end of the D stem (Fig. 6a), but whereas the stacking on 27 continues the helical succession, the screw relationship to residue 25 is different. The anticodon helix is tipped by about 20° to the D helix (1) and its large groove is thus narrowed. The base A44 is not coplanar with m_2^2G26 but the distances between the N1's and between O6 of G and N6 of A indicate that they form an H-bonded pair (Fig. 6b). Likewise, although G45 is not coplanar with the pair m^2G10-C25, its N2 is within H-bonding distance of N7 and O6 of m^2G10, so that these bases probably form a base triple.

When building the 3 Å model, we did not make the H-bonds just described, because the members of each potential base pair lay in distinct planes (Fig. 6a). The 2.5 Å map, however, leaves us in little doubt that the active groups have favorable mutual orientations and separation distances. In the orthorhombic crystal structure, the bases m_2^2G26 and A44 were thought to be H-bonded (2), probably by a single bond (16), but there appears to have been further uncertainty on this point (cf. Fig. 3 of ref. 16, where they are shown as unbonded).

Anticodon loop

The anticodon loop resembles the model suggested by Fuller and Hodgson (17) in having five bases stacked on the 3′ side and two on the 5′ side. Our earlier description (1) stands, but further fine details have emerged. Thus, on the 5′ side, the two bases C32 and U33 are well stacked on each other but not quite parallel to A31. The N3 of U33 makes an H-bond across the loop to phosphate 36. There is a sharp bend between residues U33 and Gm34, the nucleotide at the extreme tip of the anticodon. On the 3′ side, the bases from G34 to A38 are well stacked (Fig. 7) and exhibit more overlap than is found in a normal double helix. The chain may well have adopted a conformation characteristic of single stranded RNA, and the five purine bases approach maximum overlap, unrestrained by the presence of a complementary strand. The density for the aromatic part of the Y base is quite clear, but the long sidechain does not appear to be well fixed. The chain is fixed with only a few H-bonds and the anticodon appears to be readily available for interaction with messenger RNA.

Concluding remarks

The stereochemical improvements and the new features revealed in the second round of model building were achieved by closely following the density in the 2.5 map, so that their occurrence increases our confidence in the map and its interpretation. The model is now of sufficient accuracy to proceed to the next stage of crystallographic refinement, which involves replacing the isomorphous replacement map with one computed from phases calculated from the model (18). This procedure introduces an element of circularity, since such maps may show peaks at positions of wrongly sited atoms which are little lower than those due to atoms correctly placed. We have therefore concentrated on obtaining as good a start as possible. Indeed, the refinement is already under way and first results are very promising.

Our conviction, that essentially the correct distribution of groups in the molecule has been found, is strengthened by the various features described above, for example, the intricate relation between the TΨC and D loops. Here the structure is highly cross-strutted and there is a spatial segregation of variable and invariant residues into two distinct regions (Fig. 5), which speaks for some biochemical function. The variable pocket formed by bases 16, 17, and 20 together with 59 and 60 may form part of a recognition system for different tRNAs, perhaps for sorting into classes for synthetase discrimination, whereas the "conserved cluster" formed by the bases 18, 19, and 54 to 58 may constitute a ribosomal recognition site common to all tRNAs (1). The sequence G-T-Ψ-C-G, which has been implicated in binding to the ribosome, through the 5S RNA (19), is tightly tethered in the tertiary structure, and would have to be opened out before interacting with a complementary sequence. What is striking, however, is that although the base Ψ55 has two of its H-bonding groups involved in the tertiary structure, it has two more, an NH and O, unused and pointing outward from the molecule (Fig. 4). The adjoining base T54 has its methyl group in the same vicinity, and one wonders whether the

two together might not help form some signal for a preliminary binding step. It is clear, however, that one is only at the beginning of relating the structure of tRNA and its possible conformational changes to actual biochemical processes, but a knowledge of the features outlined above and earlier provides a necessary framework for this understanding.

We thank our colleagues, Drs. J. T. Finch, T. LaCour, and M. Levitt for their help and NATO grant 893 for travel support.

1. 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.
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. Klug, A., Ladner, J. E. & Robertus, J. D. (1974) *J. Mol. Biol.* **89,** 511–516.
4. Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. S., Clark, B. F. C. & Klug, A. (1974) *Nucleic Acids Res.* **1,** 927–932.
5. Rhodes, D. (1975) *J. Mol. Biol.* **94,** 449–460.
6. RajBhandary, U. L. & Chang, S. A. (1968) *J. Biol. Chem.* **243,** 598–608.
7. Barrell, B. G. & Clark, B. F. C. (1974) *Handbook of Nucleic Acid Sequences* (Joynson-Bruvvers, Oxford).
8. Ladner, J. E. Jack, A., Robertus, J. D., Brown, R. S., Rhodes, D., Clark, B. F. C. & Klug, A. (1975) *Nucleic Acid Res.* **2,** 1629–1637.
9. Crick, F. H. C. (1966) *J. Mol. Biol.* **19,** 548–555.
10. Fiskin, A. M. & Beer, M. (1965) *Biochim. Biophys. Acta* **108,** 159–163.
11. Levitt, M. (1974) *J. Mol. Biol.* **82,** 393–420.
12. Arnott, S. (1970) in *Progress in Biophysics and Molecular Biology* (Pergamon Press, Oxford), Vol. 21, pp. 265–319.
13. Sundaralingam, M. (1969) *Biopolymers* **7,** 821–869.
14. Clark, B. F. C. & Klug, A. (1975) *Proceedings of 10th Congress of FEBS, Paris,* in press.
15. Hirsh, D. (1971) *J. Mol. Biol.* **58,** 439–458.
16. Kim, S. J., 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.
17. Fuller, W. & Hodgson, A. (1967) *Nature* **215,** 817–821.
18. See, for example, Diesenhofer, J. & Steigemann, W. (1975) *Acta Crystallogr. Sect. B,* **31,** 238–250.
19. Erdmann, V. A., Sprinzl, M. & Pongs, O. (1973) *Biochem. Biophys. Res. Commun.* **54,** 942–948.