

A PRELIMINARY REFINEMENT OF YEAST *t*RNA^{Phe} AT 3 Å RESOLUTION

JOEL L. SUSSMAN and SUNG-HOU KIM

Dept. of Biochemistry, Duke University Medical School, Durham, N.C. 27710, U.S.A.

1. Introduction

The crystal structure of yeast *t*RNA^{Phe} (in the orthorhombic form) is being refined at 3 Å resolution in order to obtain a set of atomic coordinates that:

- (1) fit the X-ray diffraction data and
- (2) are consistent with the geometry of standard nucleotides and do not violate non-bonded interactions.

Two directions are being taken to obtain the above goal. One is the 'real space refinement' procedure of Diamond pursued by the M.I.T. group and the other is the procedure which we have been developing at Duke University, which is described here.

In the course of this refinement we have applied some crystallographic techniques that are not commonly used by macromolecular crystallographers, and have developed a convenient way to handle nucleic acid structures in a computer.

It must be stressed at the outset that compared to the state of refinement achieved in some protein structures at high resolution (e.g., rubredoxin [1], high potential iron protein [2], and the complex of bovine trypsin and bovine pancreatic trypsin inhibitor [3]), we are at a very early stage. Our present stage of refinement might be considered as model building guided by the electron density map.

2. Improving the Electron Density Map by the Partial Structure Fourier Method

During the course of the structure determination of yeast *t*RNA^{Phe}, it was relatively easy to trace the flow of the backbone chain of the polynucleotide in the 4 Å electron density map calculated on the basis of the multiple isomorphous replacement (MIR) method [4]. However, it was very difficult to recognize either bases or the accurate positions of sugars and phosphates, even though the connectivity of the electron density was quite clear above the background noise level. In the 3 Å electron density map, it became relatively easy to locate individual phosphate, ribose, and base groups in the double-stranded regions. However, it was much more difficult to trace the course of the chain in the single-stranded loop regions [5, 6].

To resolve the ambiguities in interpretation of the single-stranded regions in the electron density map, we used a variation of a technique which is very common in small molecule crystallography. Phases are calculated from model coordinates of a fragment of the structure which has been seen by the heavy atom method or the direct

methods, then the successive application of the Fourier method or tangent formula method [7] reveals the remainder of the structure [8]. We used a similar approach except our initial fragment was based on the electron density map calculated from the MIR phases. It has been shown for proteins that phases calculated from model atomic coordinates are far superior to those based on the MIR method [9]. The Fourier method was applied initially to the 'partial group structure' and then later to the 'partial atomic structure'.

Structural features of nucleic acids, along with stereochemical constraints based on the crystal structures of nucleotides [10], dinucleosides [11, 12, 13], and a trinucleoside [14], aided in the interpretation of the MIR map. An RNA structure is made up

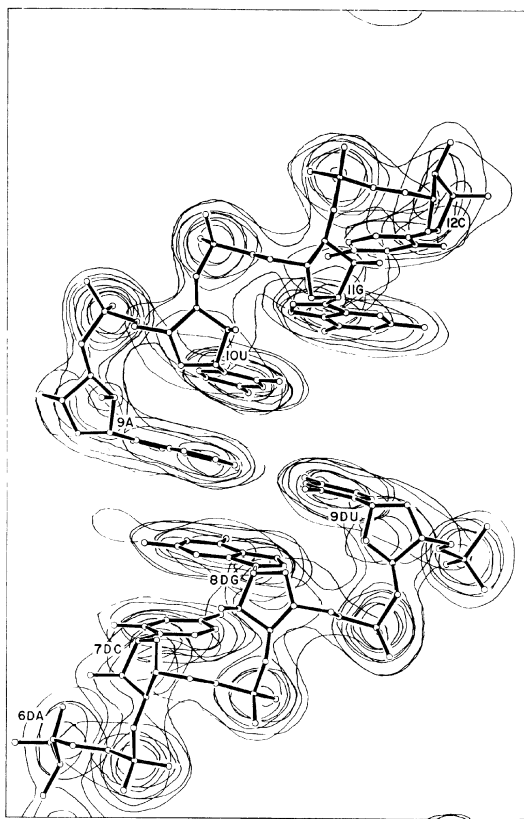


Fig. 1. Calculated electron density map of a 5 Å thick section of a hypothetical RNA structure at 3 Å resolution. A skeletal model of the structure is superimposed to aid in the interpretation of the map. At 3 Å resolution the sugars and phosphates both appear as individual peaks, with the phosphate group almost spherical and the sugar more ellipsoidal in shape. The bases are easily resolved, appear planar, and it is possible to distinguish purine bases from pyrimidines.

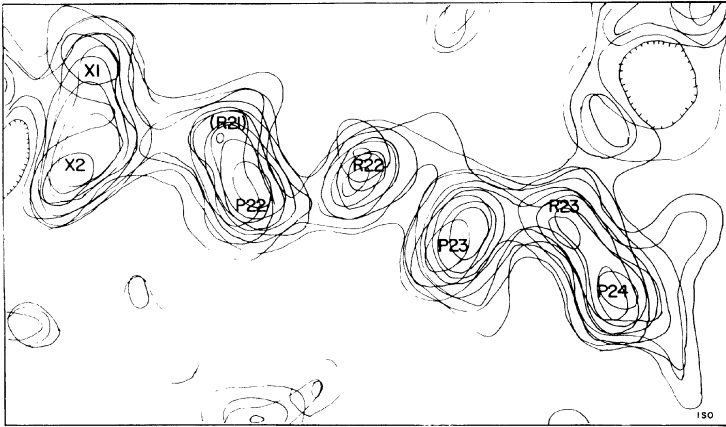


Fig. 2. A 2.5 Å thick section of the MIR electron density map of yeast *t*RNA^{Phe} at 3 Å resolution which corresponds to the DHU stem portion of the structure. Peaks for Ribose 21 (R21) through Phosphate 24 (P24) can be individually distinguished.

of a repeating backbone consisting of two groups of atoms; a phosphate group of five atoms and a ribose group of seven atoms. A relatively large side chain group, a purine or pyrimidine, is attached to each sugar. These groups show up very clearly as individual peaks in a 3 Å resolution electron density map, as can be seen in Figure 1 and Figure 2.

After an initial atomic skeletal model was built on an optical comparator to fit the 3 Å MIR map, we applied the partial group structure method (see Figure 3) to resolve the ambiguities in the chain tracing of the single-stranded loops (for a more detailed description of the method, see Reference 15). This was accomplished with two partial group structure Fourier maps, as shown in Figure 4. Following the calculation of each partial structure Fourier synthesis, the portion of the structure *not* included in the calculation was built in the optical comparator to fit the portion of the new map. Finally the coordinates from the Kendrew skeletal model were measured and least-square fitted to the group centers which had previously been measured from the electron density maps.

3. Refinement of the Coordinates using Chemical Constraints and the Electron Density Map

At this stage we wanted to adjust the coordinates to fit both the geometry of the standard nucleotides and the electron density map. This was achieved by the use of a program REFIN 2 written by Hermans and McQueen [16, 17], as modified by Sussman [unpublished results] for nucleic acid structures. The objective of this program is to fit a flexible polymer made up of a series of residues (amino acids or

nucleotides) to a series of guide points or targets, while maintaining the standard stereochemical constraints of bond lengths, bond angles, and fixed dihedral angles, as well as non-bonded interactions. This objective is accomplished by means of a criterion function which is minimized for each atom sequentially in the structure. Using Hermans' notation [16], for the i th atom this function is

$$F_i = w_l \sum (l - l_0)^2 + w_t \sum (\theta - \theta_0)^2 + w_r \sum (\rho - \rho_0)^2 + w_e \sum_j E_{ij} + w_m (\mathbf{x} - \mathbf{x}_0)^2. \quad (1)$$

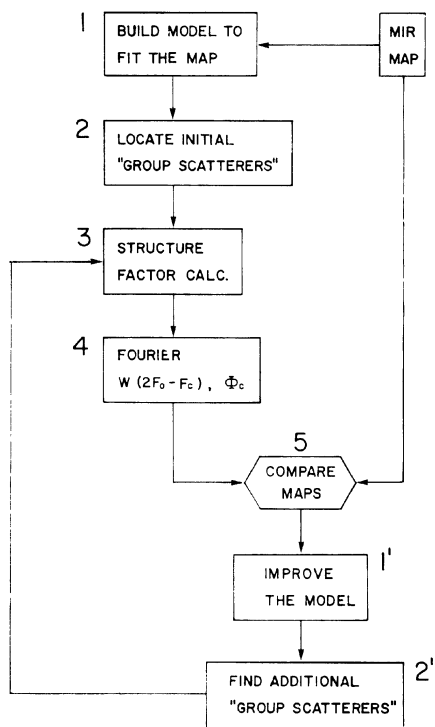


Fig. 3. Flow diagram of the 'partial group structure method'. The method begins with an electron density map calculated by the method of multiple isomorphous replacement (MIR). (1) Build a Kendrew skeletal atomic model to fit the electron density of as much of the structure as possible using an optical comparator ('Richards' box). (2) Obtain the coordinates of as many 'group centers' as can be unambiguously identified. (3) Calculate amplitudes (F_c) and phases (ϕ_c) for the partial group structure using 'group scattering factors' (spherically-averaged scattering factors for all the atoms in the group). (4) Calculate a 'sum function' Fourier synthesis, where amplitudes are $W(2F_o - F_c)$ and phases are ϕ_c (the W is the Sim weighting coefficient). (5) Compare the 'sum function' map to the MIR map to improve the model (1') so as to fit the electron density maps better and to locate additional group centers (2') that previously could not be unambiguously identified. Repeat steps 1' through 5 until all the groups can be identified. In later stages, atomic positions are located and used rather than group center coordinates in steps 2, 3, and 2'.

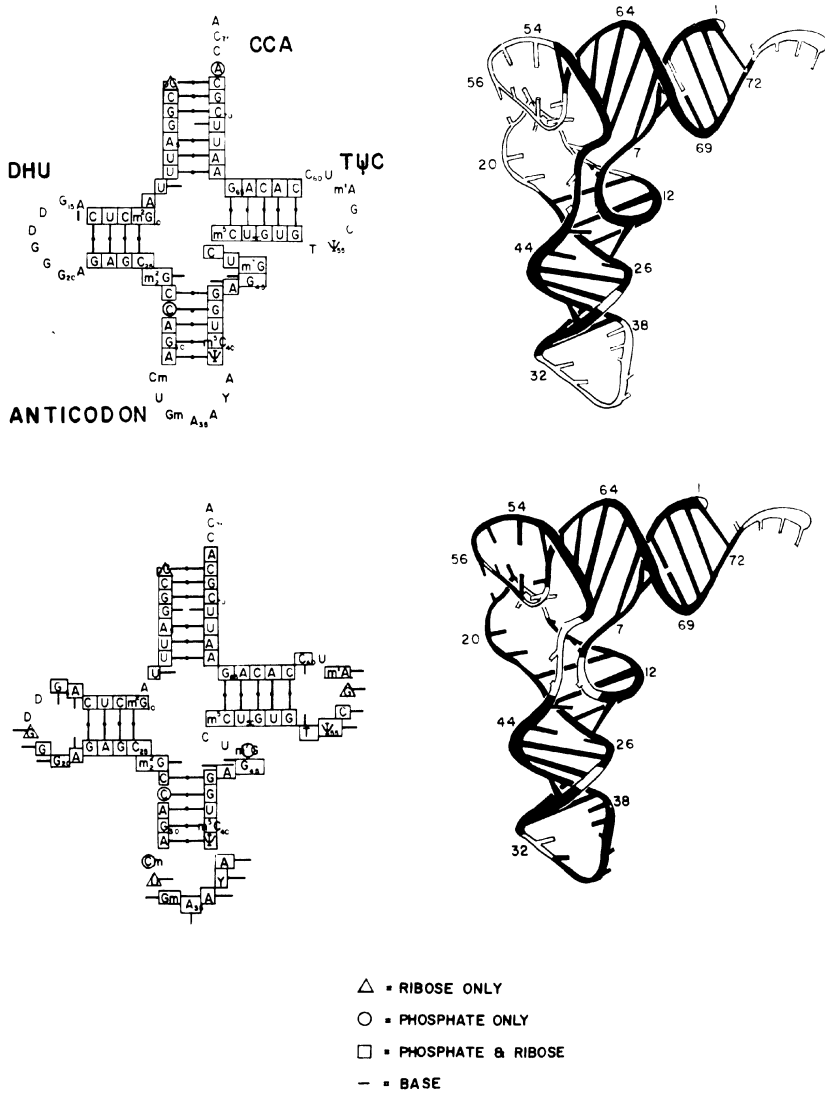


Fig. 4. The stages in the unravelling of the backbone of yeast *tRNA^{Phe}* by the partial group structure method. The top two drawings show the initial 145 groups that were derived from the MIR map (n.b. that most of these groups correspond to the stem regions of the structure, as can be seen easily in the cloverleaf drawing). The bottom two drawings show the 192 groups that were used in the second Fourier synthesis. Here many of the residues in the loop regions were included. Groups included in the structure factor calculations are marked by various symbols.

The first three summations are carried out for all bond lengths l , bond angles θ , and all 'fixed' torsion angles ϱ which are affected by the position of atom i . We allowed some of the torsion angles to be free, based on observations from di- and trinucleoside crystal structures [11–14] (see Figure 5). The subscript 'o' for these summations refers to standard values derived from X-ray diffraction work on polynucleotide fibers, which are supplied by the user of the program in a dictionary for the particular kind of polymer. The standard geometry we used was derived from the work of Arnott *et al.* [18].

The fourth summation corresponds to an energy function for all atoms j not directly bonded to atom i , and represented by a van der Waals attractive and repulsive energy term. This term is essential for nucleic acid model building in order to maintain a reasonable base stacking such that the stacked base planes are approximately parallel and separated by about 3.4 Å. In addition to this, the fourth summation is used to constrain certain specified pairs of atoms at hydrogen bond distances. Thus, we initially specified in the refinement the base-paired secondary structure as well as the tertiary structure hydrogen bonds which could be seen in the MIR and partial group structure Fourier maps.

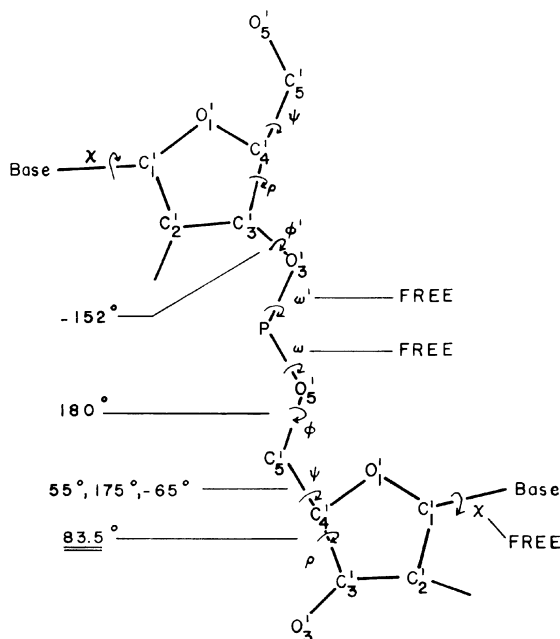


Fig. 5. Torsion angle constraints used in the refinement. ϱ , which determines the sugar pucker, has been held fixed at 83.5° corresponding to a C_s' endo sugar conformation. ϕ and ϕ' were allowed to vary approximately $\pm 20^\circ$ from the angles listed. ψ was allowed to vary $\pm 20^\circ$ from each of the three angles listed. ω , ω' , and χ were not constrained at all.

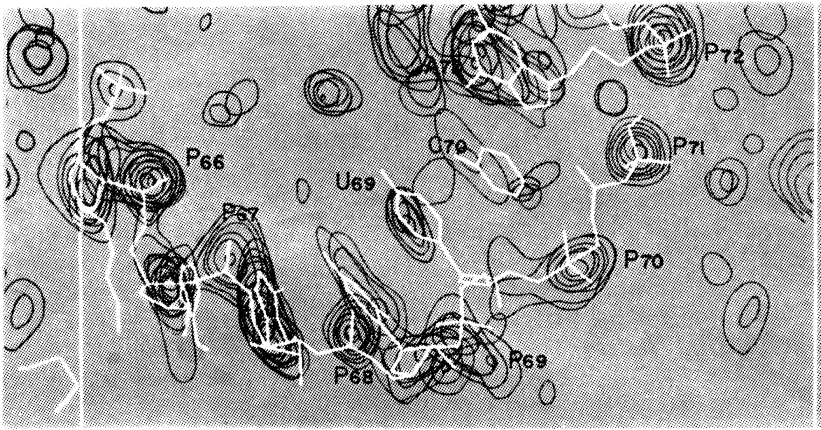


Fig. 6. A section of our electron density map at an early stage of the partial structure Fourier procedure. 'P' is for phosphate and 'R' for ribose. The skeletal model drawing is incomplete for R₇₀ and R₇₁. Electron densities for R₆₉, R₇₀, and C₇₀ are out of the section chosen. P₆₉, R₆₉, U₆₉, and P₇₀ were not included in the calculation of the phases but came up in the electron density map. After the preliminary fitting as shown here, the refinement procedure optimizes the fitting to the 'target positions' and the chemical constraints.

The fifth term represents the distance between the current position of an atom and its target position (indicated by ' σ '). The targets were the centers of the phosphate and ribose groups as measured from the maps. This term prevents the groups from moving outside their electron density.

The w_l , w_r , w_p , w_e , and w_m are weights associated with each summation and can be adjusted by the user to emphasize one of the summations relative to another.

A portion of an electron density map at an intermediate stage of the 'partial structure Fourier procedure' is shown in Figure 6.

4. Results of the Refinement

A preliminary refinement has been done for the entire structure of yeast *t*RNA^{Phe}. For the four stem regions, the average error in bond length is 0.009 Å and the average error in bond angle is 0.44°. The average distance between the centers of all the phosphate groups and the ribose groups to their respective target positions in the electron density map for all 76 nucleotides is 0.55 Å, while the rms deviation is 0.68 Å.

The 'fixed' torsion angles of the structure vary slightly from their standard values (Figure 7a), while the torsion angles which were allowed to be completely free provided a great deal of the flexibility in the structure (Figure 7b). The *R* factor and correlation coefficient are summarized in Table I.

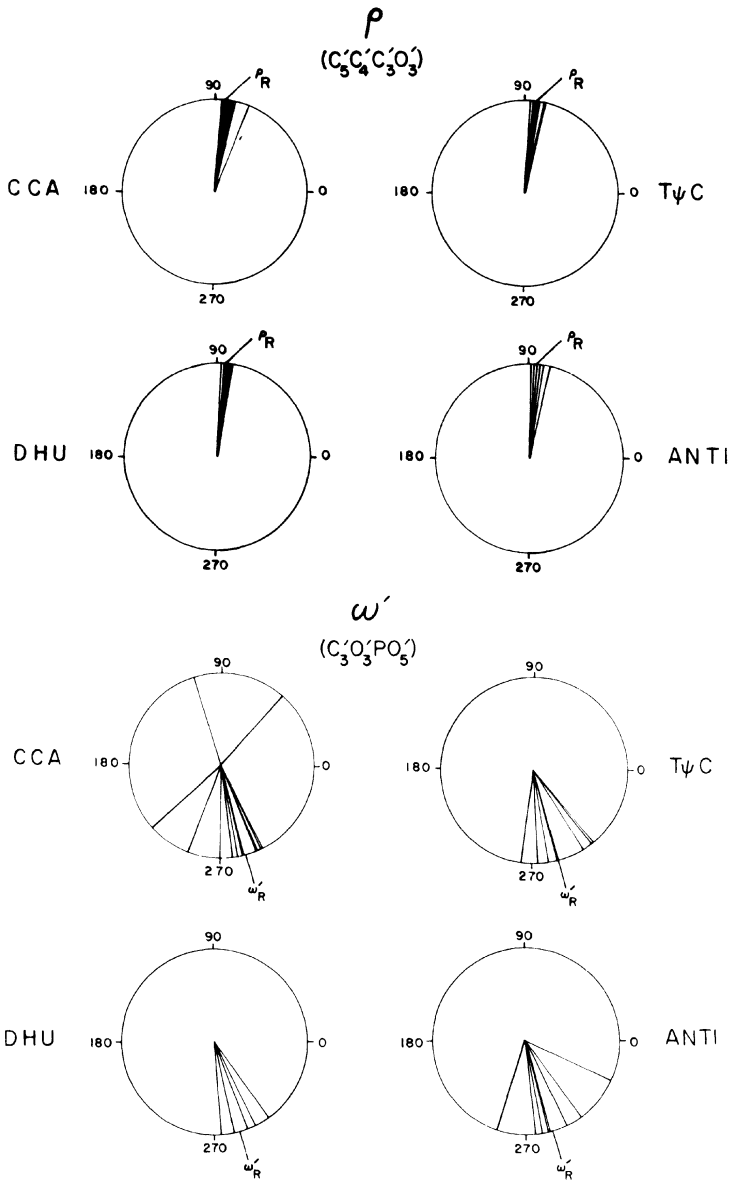


Fig. 7. Distribution of stem torsion angles at the current stage of the refinement of yeast tRNA^{Phe} for (a) a 'fixed' torsion angle ρ and (b) an unconstrained torsion angle ω' (see Figure 5 for definition of the torsion angle).

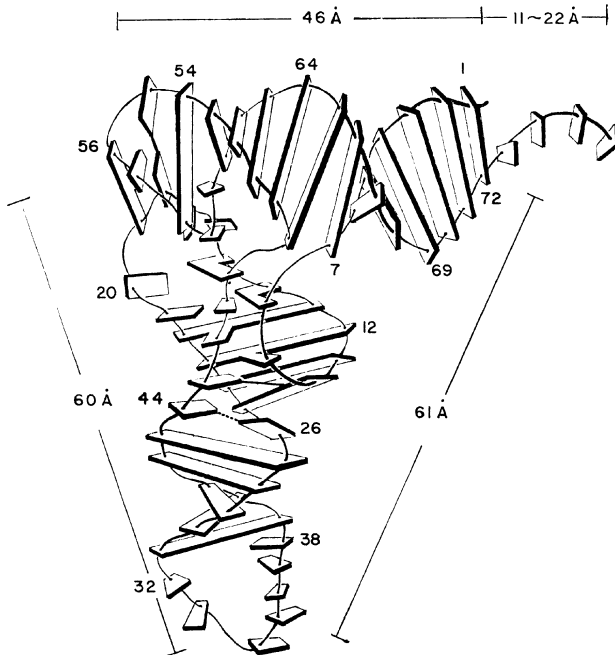


Fig. 8. An artistic drawing of the structure of yeast *tRNA*^{Phe} at the present stage of refinement. The backbone phosphate ribbon is indicated by a thin line, while the bases are indicated by slabs. The whole molecule is flat with an average thickness of about 20 Å. Secondary and tertiary hydrogen-bond interactions between bases are indicated by either fusing two slabs together or by connecting the slabs with a line. Other types of tertiary interactions are not shown, for clarity. Distances shown are between the four bases A₃₅, C₅₆, C₇₂, and A₇₆. The range given for the C₇₂—A₇₆ distance corresponds to the stacked and stretched structure of that single-stranded region.

TABLE I

	<i>N</i> = 6200	<i>N</i> = 4800
$R = \frac{\sum F_0 - F_c }{\sum F_0}$	0.39	0.38
$CC = \frac{N \sum F_0 F_c - \sum F_0 \sum F_c}{\{N \sum F_0^2 - (\sum F_0)^2\} \{N \sum F_c^2 - (\sum F_c)^2\}^{1/2}}$	0.73	0.74

Based on the refinement thus far, we have been able to confirm or clarify all of the tertiary interactions we saw previously [5, 19], as well as to see several new tertiary interactions. The following additional hydrogen bonds have been interpreted to be present based on the current state of refinement (some of these interpretations may be subject to revision as the refinement progresses):

- (1) *Hydrogen bonds between the bases.* In addition to those published [5, 6, 19], there is one possible hydrogen bond between G₁₀ and G₄₅. This explains the fact that G₄₅ is protected from chemical modification. G₄ and G₆₉ are both stacked within the acceptor stem making at least one hydrogen bond. This G–U ‘base pair’ introduces a slight over-winding of the stem helix. The hydrogen bonding between the bases is shown in Figure 8.
- (2) *Hydrogen bonds between base and ribose.* Two likely such interactions are between O₂' of U₈ and N₁ of A₂₁; O₂' of A₉ and N₇ of G₁₀.
- (3) *Hydrogen bonds between phosphate and base.* A possible example is between phosphate oxygen of A₅₈ and N₁ of ψ_{55} .
- (4) *Hydrogen bonds between ribose and ribose.* The most common hydrogen bond is between O₂' and O₁' of the following nucleotide in the sequence. There are many candidates for such an interaction.

As we continue the refinement of yeast *tRNA*^{Phe}, we plan to modify the criterion function (I) to include the electron density map itself in a way analogous to the Diamond II refinement procedure [20, 21], while still maintaining the stereochemical and non-bonded constraints.

Errata: A72 in Figure 6 should be labeled G71.

Acknowledgements

We would like to acknowledge our collaborators in Alex Rich's laboratory at M.I.T., especially F. L. Suddath, Gary Quigley, and Alex McPherson, for the X-ray diffraction data used in this work and for helpful discussions.

We would like to thank Marilyn Warrant and Mei-Chien Huang for their excellent help.

This work was supported by grants from the U.S. Public Health Service (CA-15802) and the National Science Foundation (GB-40814). J.L.S. is a fellow of the Arthritis Foundation.

References

1. Watenpugh, K. D., Sieker, L. C., Herriott, J. R., and Jensen, L. H.: *Acta Cryst.* **B29**, 943–956 (1973).
2. Freer, S. T., Alden, R. A., Carter, C. W., and Kraut, J.: *J. Biol. Chem.* **250**, 46–54 (1975).
3. Huber, R., Kukla, D., Wolfram, B., Schwager, P., Bartels, K., Deisenhofer, J., and Steigemann, W.: *J. Mol. Biol.* **89**, 73–101 (1974).
4. Kim, S. H., Quigley, G. J., Suddath, F. L., McPherson, A., Sneden, D., Kim, J. J., Weinzierl, J., and Rich, A.: *Science* **179**, 285–288 (1973).
5. Kim, S. H., Suddath, F. L., Quigley, G. J., McPherson, A., Sussman, J. L., Wang, A. H. J., Seeman, N. C., and Rich, A.: *Science* **185**, 435–440 (1974).
6. Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. S., Clark, B. F. C., and Klug, A.: *Nature* **250**, 546–551 (1974).
7. Karle, J.: *Acta Cryst.* **B24**, 182–186 (1968).
8. Stout, G. H. and Jensen, L. H.: *X-Ray Structure Determination: A Practical Guide*, The MacMillan Company, London, 1968, Chapter 11.
9. Jensen L. H.: *Annual Review of Biophysics and Bioengineering* **3**, 81–93 (1974).

10. Sundaralingam, M.: *Biopolymers* **7**, 821–860 (1969).
11. Sussman, J. L., Seeman, N. C., Kim, S. H., and Berman, H. M.: *J. Mol. Biol.* **66**, 403–421 (1972).
12. Rosenberg, J. M., Seeman, N. C., Kim, J. J. P., Suddath, F. L., Nicholas, H. B., and Rich, A.: *Nature* **243**, 150–154 (1973).
13. Day, R. O., Seeman, N. C., Rosenberg, J. M., and Rich, A.: *Proc. Nat. Acad. Sci. U.S.A.* **70**, 849–853 (1973).
14. Suck, D., Manor, P. C., Germain, G., Schwalbe, C. H., Weimann, G., and Saenger, W.: *Nature New Biology* **246**, 161–165 (1973).
15. Sussman, J. L. and Kim, S. H.: submitted to *Acta Cryst.* (1974).
16. Hermans, J. and McQueen, J. E.: *Acta Cryst.* **A30**, 730–739 (1974).
17. Hermans, J.: in this volume (1975).
18. Arnott, S., Hukins, D. W. L., and Dover, S. D.: *Biochem. Biophys. Res. Commun.* **48**, 1392–1399 (1972).
19. Kim, S. H., Sussman, J. L., Suddath, F. L., Quigley, G. J., McPherson, A., Wang, A. H. J., Seeman, N. C., and Rich, A.: *Proc. Nat. Acad. Sci. U.S.A.* **71**, 4970–4974 (1974).
20. Diamond, R.: *Acta Cryst.* **A27**, 436–452 (1971).
21. Diamond, R.: *J. Mol. Biol.* **82** 371–391 (1974).

DISCUSSION

Honig: How well did the model builders do in predicting the three dimensional structure of *t*RNA?

Sussman: None of the many proposed models were correct. I think this is because many of the model builders tried to line up the helical stem portions of the structure along a common axis in order to agree with the low angle X-ray scattering data, while the single crystal structure shows that there are two major axes in the structure at approximately 90°. One of the better *t*RNA models was the one proposed by Michael Levitt. Although his model had the stems aligned along a common axis, he did predict correctly several of the tertiary interactions that we have found, while some of his tertiary interaction predictions were incorrect.

Saenger: Did you employ phase refinement techniques to improve the phase angle?

Sussman: No, although we are considering trying them in the future. One of the problems is that few workers have tried phase refinement at such low resolution as 3 Å, so we are somewhat sceptical as to whether they would in fact improve our maps.

Saenger: What are the main differences between your model and the model of the Cambridge, England group?

Sussman: Although the two *t*RNA PHE's crystallize in different forms their overall structures are remarkably similar. Based on the Cambridge (MRC) groups published structure it is difficult to judge the differences between the interactions of the *D* loop and *T* loop as they stated it was still difficult to make a unique chain tracing for these regions. There does seem to be a possible real difference near the 'hinge' region of $m_2^2G_{26}$. In our maps $m_2^2G_{26}$ comes near A44, possibly hydrogen bonded, while in the MRC model $m_2^2G_{26}$ seems to be intercalated between A₄₄ and G₄₅. This may represent a *real difference* in the structures and point to a region of flexibility in *t*RNA's, i.e. allowing the anti codon stem to flex about this point.