Acta Cryst. (1983). B39, 495–505

The Use of a Constrained–Restrained Least-Squares Procedure for the Low-Resolution Refinement of a Macromolecule, Yeast $tRNA_f^{Met}$

BY JOEL L. SUSSMAN AND A. D. PODJARNY

Department of Structural Chemistry, Weizmann Institute of Science, Rehovot, Israel and CINDECA, FCE, UNLP, 48 y 115, La Plata 1900, Argentina

(*Received 2 August* 1982; *accepted 1 February* 1983)

Abstract

A method of low-resolution group least-squares refinement for macromolecular structures is described. It has the ability of correcting gross positional errors (~7 Å) using large groups such as helical domains as refinement units. These large domains are then broken apart gradually into smaller groups, providing a smooth transition to conventional macromolecular refinement *via* a constrained–restrained least-squares procedure. The resolution of the data is increased in such a way that the ratio of number of parameters to observables allows for refinement convergence. As an application of the method, the refinement of yeast $tRNA_f^{Met}$ is described. The resolution of the refinement ranged from 12·5 to 4 Å; the groups were the whole molecule, large helical stems, small helical stems (with associated loops), nucleosides, and phosphates in steps of increasing degrees of freedom. The r.m.s. shift attained by the rigid helical refinement was 6·82 Å. The final model converged to an R factor of 26·5%, at 4 Å resolution, with acceptable stereochemistry.

Introduction

The use of real-space restraints combined with the method of least squares in reciprocal space has proven to be an extremely powerful way to refine macromolecular structures after an approximate trial model has been postulated. [For a comprehensive review of this subject see Machin, Campbell & Elder (1981).]

In all cases thus far starting models have been obtained from the interpretation of an electron density map based either on the method of multiple isomorphous replacement (MIR), or on the molecular-replacement method (MRM). With the exception of myohemerythrin (Hendrickson, Klippenstein & Ward, 1975) and demetallized concanavalin A [Shoham, Yonath, Sussman, Moult, Traub & Kalb (Gilboa), 1979], the resolution of the original map was 3 Å or better. The case of myohemerythrin represents the furthest incursion into low resolution so far reported, with an initial model built into a 5·5 Å electron density map and the refinement begun at 4 Å and then carried to 3 Å. The constraints and restraints introduced in the refinement procedures dealt primarily with the known stereochemistry of the biological polymer. In general no *a priori* information about secondary structure conformation of a polymer strand was used. Only in a few cases was additional information used, *e.g.* Fe positions in hemerythrin and the use of Watson–Crick base-pairing interactions between strands for the tRNA structures.

In the present work, we have interpreted a very noisy electron density map of yeast initiator tRNA ($tRNA_f^{Met}$) using a computer-graphics system. After this interpretation, the agreement factor was essentially random except at very low resolution. We had then a very marginal interpretation of an electron density map, and clearly it was not possible to proceed into the kind of high-resolution refinement that was used in the above examples. However, a study of packing conditions and of covalently bound heavy-atom markers, whose particular attachment site on $tRNA_f^{Met}$ were known from independent experiments (Schevitz *et al.*, 1975; Schevitz, Podjarny, Krishnamachari, Hughes, Sigler & Sussman, 1979), led us to believe that the proposed model was *not* completely wrong, though it probably had very large errors, possibly as much as 5 Å (as estimated from model calculations).

As the $tRNA^{Phe}$ molecule has very well defined double-helical domains with known stereochemistry (Holbrook, Sussman, Warrant & Kim, 1978; Jack, Ladner & Klug, 1976), it was possible to reduce drastically the number of degrees of freedom by introducing that information. Furthermore, we worked at very low resolution (20·0–12·5 Å) to increase the radius of convergence, based on test cases, so that the method would overcome the large errors of the original model. Even at this low resolution the number of degrees of freedom was small enough to maintain a reasonable ratio of observables to parameters.

Below, we discuss the procedure, applied to $tRNA_f^{Met}$, by starting with very few degrees of freedom

0108-7681/83/040495-11$01.50 © 1983 International Union of Crystallography

and very-low-resolution data and gradually increasing both to the point where a standard least-squares procedure was applied during the later stages.

The details of the data collection, initial interpretations of the MIR maps, a description of the structure and its biological significance can be found in Schevitz *et al.* (1975, 1979), Sussman, Podjarny, Schevitz & Sigler (1981), Sigler, Schevitz, Gross, Kaplansky, Podjarny & Sussman (1981), Podjarny, Schevitz & Sigler (1981), Schevitz, Podjarny, Zwick, Hughes & Sigler (1981), and Schevitz, Podjarny, Sussman & Sigler (1983).

Methods

The constrained–restrained least-squares refinement program *CORELS* (Sussman, Holbrook, Church & Kim, 1977) used in this work was originally designed to handle *constrained* groups which were the monomer units of a biopolymer such as nucleosides, phosphates or amino acids. A precise definition of a constrained group is a molecular moiety where all bond distances and bond angles have been fixed to respective canonical values, but which can have any number of easily defined rotatable bonds. These constrained groups are then connected together *via* flexible links called *restraint* distances, which correspond to a bond length, a distance between atoms facing a bond angle, a hydrogen bond or a non-bonded contact, *etc.* This distinction between a rigid *constraint* and a looser *restraint* was initially pointed out by Konnert (1976) and then later used by Sussman, Holbrook, Church & Kim (1977).

In order to attempt refinement at low resolution, the *CORELS* program was modified to handle a constrained group with any number of atoms. In this way it was possible to incorporate the known stereochemistry of large portions of the tRNA structure.

The molecule is treated as a single group and refined as a rigid body in a least-squares way to minimize the difference between observed and calculated structure factors. This kind of procedure has been used for some time at higher resolution in rigid-body refinement of smaller structures (see *e.g.* Doedens, 1970). Here it can be considered more like an extension of the molecular-replacement technique, because of the exclusive use of low-resolution data. As there are only six degrees of freedom, the computational problem is small. Then the molecule is divided into a few rigid groups, introducing specific restraints to preserve reasonable stereochemistry at the arbitrary joints between groups. In the case of tRNA we first used two groups corresponding to the two helical domains and later further divided these into four groups consisting of the individual arms and stems of the cloverleaf. We increased the resolution gradually, refining the four groups at each stage. The temperature factors were then refined for the four separate groups, resulting in a total of 28 degrees of freedom. This procedure has a large radius of convergence at low resolution. As the resolution was increased gradually, the refinement did not move the

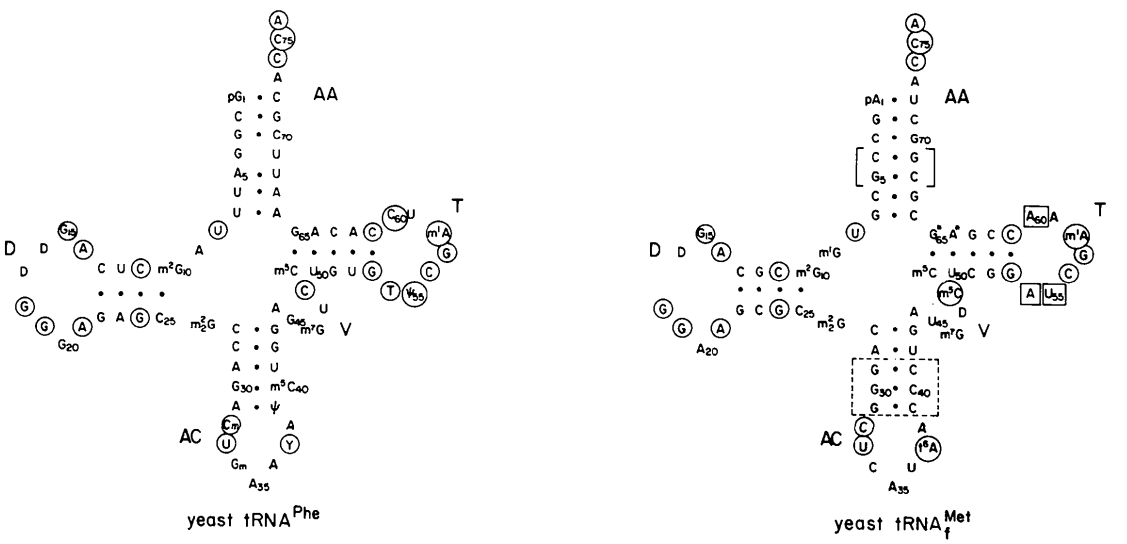


Fig. 1. A comparison of the sequence of yeast $tRNA^{Phe}$, an elongator tRNA (RajBhandary & Chang, 1968), and yeast $tRNA_f^{Met}$, an initiator tRNA (Simsek & RajBhandary, 1972). For simplicity, the numbering scheme for $tRNA^{Phe}$ is used for both sequences. Abbreviations for the various regions of the structure are: AA amino acid stem, D dihydrouracil arm (where an *arm* is a domain consisting of a *stem* and *loop*), AC anticodon arm, V variable arm, T TψC arm. Circles indicate invariant or semi-invariant bases found in almost all tRNA's. Squares indicate the unique features found only in eukaryotic initiator tRNA's. The region in the $tRNA_f^{Met}$ sequence enclosed in dashed lines at the end of the AC stem is a stack of three G-C base pairs found in virtually all initiator tRNA's (both eukaryotic and prokaryotic) at this site and not found *there* for other tRNA's.

structure from the global minimum. Eventually, as the extent of data included in the refinement is increased, the model can be divided into units directly related to local flexibility. In the case of tRNA, we used phosphate and individual nucleosides or nucleoside base pairs. The resolution at which this detailed refinement is carried out should be compatible with the size of the monomer. The refinement then proceeds according to previously described constrained–restrained methods (Sussman *et al.*, 1977).

Low-resolution least-squares refinement of $tRNA_f^{Met}$

The model of $tRNA_f^{Met}$ which was built from its MIR map both in an optical comparator – 'Richards box' (Richards, 1968) – and later on a static (Katz & Levinthal, 1972; Podjarny, 1976) and real-time computer-graphics system (Tsernoglou, Petsko, McQueen & Hermans, 1977) was similar to the structure of $tRNA^{Phe}$, although due to limitations in the model-building procedure it had somewhat poorer stereochemistry. To correct this we decided to start the refinement with a stereochemically more reasonable model so the $tRNA^{Phe}$ structure (Sussman, Holbrook, Warrant, Church & Kim, 1978) was fitted by a least-squares technique (Nyburg, 1974) matching up the P coordinates to those of $tRNA_f^{Met}$. The r.m.s. distance between the 75 P atoms in common for both coordinate sets was 5·4 Å (see Fig. 1 for a comparison of the two sequences). The agreement factor between observed and calculated structure factors in the range 12·5–20 Å resolution for the fitted $tRNA^{Phe}$ model was 58·1% for 143 reflections (random is 67% for this case).

The entire structure was refined as a single rigid body (see Fig. 2), converging to an *R* factor of 49·4% in five cycles for 143 reflections (see Table 1) in the 12·5–20 Å range.

The molecule was then divided into two groups, one containing the amino acid acceptor stem (AA) and the T arm, and the other containing the D, anticodon (AC) and variable (V) arms (see Fig. 2). Using the same data, and in this case 12 degrees of freedom, together with specific restraints between groups (see Fig. 3), the procedure converged in five cycles to an *R* factor of 39·1% (see Table 1 and Fig. 4).

We further divided the molecule into four groups: (1) the AA stem, (2) the T arm, (3) the AC and V arms, and (4) the D arm. With the same conditions as before, the *R* factor dropped to 33·0% in four cycles, with 24 degrees of freedom.

Fig. 4 shows the decrease in the *R* factor as a function of the number of degrees of freedom. The most important result of the procedure outlined above was a shift of the center of mass of the structure along the 6_4-axis direction by 3·7 Å. This corrected a mis-

Table 1. *Low-resolution rigid domain refinement of yeast* $tRNA_f^{Met}$

No. of groups	Degrees of freedom	Structure domain	No. of atoms	Parameter shifts						Resolution (Å)	No. of reflections	No. of cycles	R_i*	R_f*	CC_i†	CC_f†	R.m.s. shift (Å)	
				φ (°)	θ (°)	ρ (°)	x (Å)	y (Å)	z (Å)								From previous structure	Total
1	6	tRNA	1609	−2·65	−2·47	−0·23	−1·75	−1·96	−1·03	12·5–20·0	143	5	58·1	49·4	0·10	0·34	2·53	2·53
2	12	AA stem, T arm, CCA end	741	−5·14	−5·17	−2·07	−0·58	−0·17	−1·17	12·4–20·0	146	5	49·0	39·1	0·35	0·64	3·18	5·02
		D arm, AC arm, V arm	868	1·60	−5·29	−11·63	0·89	−0·81	−2·35									
4	24	AA stem, CCA end	425	−9·92	6·34	−3·02	0·53	0·30	−1·55	12·4–20·0	146	5	39·1	33·0	0·64	0·73	1·95	5·85
		T arm	316	−0·37	−11·57	0·43	0·60	−0·14	−0·27									
		D arm	375	4·86	−3·39	4·90	−0·10	1·03	0·30									
		AC arm, V arm	493	4·02	−3·80	1·70	−0·33	−0·24	−1·26									
4	24	AA stem, CCA end	425	1·53	−3·88	0·87	0·05	0·28	−0·22	10·0–20·0	311	2	43·4	41·8	0·59	0·63	0·83	6·08
		T arm	316	−5·12	−0·54	−0·82	0·66	0·59	−1·13									
		D arm	375	−4·39	1·32	−4·29	0·06	0·25	−0·46									
		AC arm, V arm	493	−0·26	0·69	−4·35	0·35	−0·01	−0·32									
4	24	AA stem, CCA end	425	4·24	−2·57	1·12	−0·25	0·17	0·07	8·0–20·0	634	2	49·3	46·9	0·58	0·62	0·96	6·28
		T arm	316	−1·32	5·87	−0·60	−0·86	−0·34	−0·15									
		D arm	375	−3·05	−0·26	1·66	0·44	0·72	−0·36									
		AC arm, V arm	493	−0·40	−0·10	−0·88	0·70	0·12	−0·50									
4	24	AA stem, CCA end	425	4·02	−1·62	−1·51	−0·40	0·26	0·87	6·0–20·0	1400	4	52·9	49·5	0·61	0·71	1·48	6·82
		T arm	316	6·56	1·32	4·21	−0·31	−0·45	0·47									
		D arm	375	−1·25	0·05	6·75	−0·33	−0·93	−1·02									
		AC arm, V arm	493	−1·23	0·38	4·87	0·10	0·04	−1·74									

* R factor is $\sum_h ||F_o| - |F_c|| / \sum_h |F_o|$. R_i initial, R_f final R factors.
† Correlation coefficient between F_o and F_c is: $\sum [(F_o - \bar{F}_o)(F_c - \bar{F}_c)] / [\sum (F_o - \bar{F}_o)^2 \times \sum (F_c - \bar{F}_c)^2]^{1/2}$. CC_i initial, CC_f final correlation coefficients.

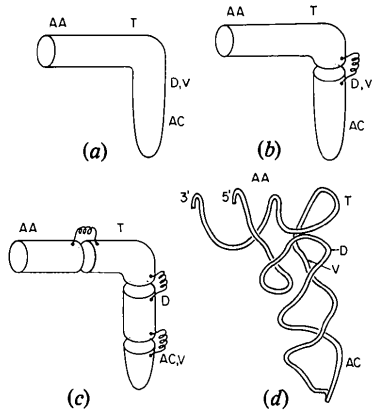


Fig. 2. Schematic representation of how the structure of $tRNA_f^{Met}$ was successively refined as rigid groups at low resolution. (*a*) The entire 'L' shaped structure was treated as a *single* rigid group (six degrees of freedom, *i.e.* three rotational and three translational). (*b*) The structure was divided into the *two* major domains, *i.e.* one consisting of the AA–T arms and the other consisting of the D–V–AC arms (12 degrees of freedom). The spring-like connections between groups represent stereochemical restraints used to maintain reasonable bond lengths and bond angles at the arbitrary division between groups. (*c*) The structure was divided into *four* rigid groups, *i.e.* the AA stem, T arm, AC and V arms and the D arm (24 degrees of freedom). (*d*) For comparison, the backbone of $tRNA^{Phe}$, in approximately the same orientation, is shown.

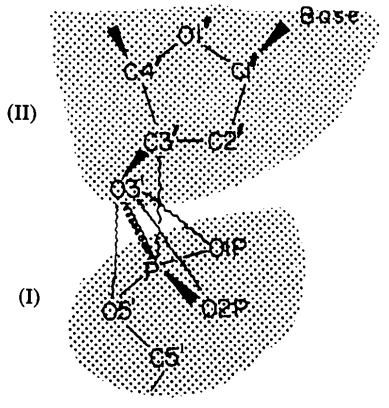


Fig. 3. Schematic representation of restrained distances used in the refinement of $tRNA_f^{Met}$. Restrained distances corresponding to the O(3')—P bond length are shown by a *tight spring*, while restrained distances corresponding to bond angles by *loose* springs.

assignment of the *z* coordinate of a heavy-atom position, which caused most of the error in the MIR map. This is a large movement on an atomic scale, and could only be obtained because it is small compared with the resolution of data we used, 12·5 Å. It is clear that if we had worked at higher resolution the procedure would not have converged. A very simple construction for a one-atom, one-reflection case would hint that the maximum error compatible with convergence of a derivative-driven algorithm is one quarter

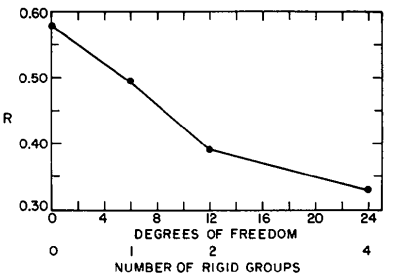


Fig. 4. *R* factor for the low-resolution (12·5–20 Å) domain refinement of $tRNA_f^{Met}$ as the structure was successively divided into one, two and four rigid groups (see Fig. 2).

of the resolution of the reflection. If we consider that we are working with 143 reflections in the 12·5–20 Å range, with most of them near 12·5–15 Å, we can see that we were nearly at the limit of the method. Model cases which we considered at the same resolution with perfect (calculated) amplitudes converged when the displacement was 5·0 Å and diverged when it was 7·0 Å.

For comparison purposes it is interesting to note that the maximum error in small structures for an atomic position which is stable under least-squares refinement conditions is around 0·25–0·3 Å, for a maximum resolution of 1·0 Å. These figures roughly agree with our observation of the permissible error being one quarter to one third of the resolution limit.

It should be mentioned here that another starting model, obtained directly from the MIR interpretation, refined in the opposite direction, moving the T–AA stem toward the twofold axis. This might imply that we were indeed at the resolution limits of the refinement procedure.

Starting from the model obtained from the 12·5 Å refinement with 24 degrees of freedom, we continued the refinement gradually increasing the resolution in

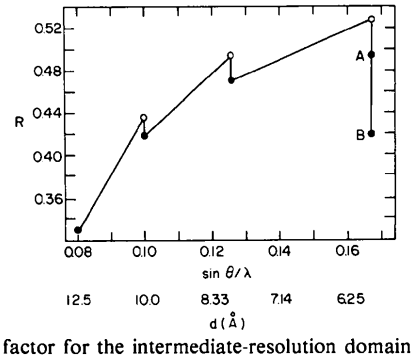


Fig. 5. *R* factor for the intermediate-resolution domain refinement of $tRNA_f^{Met}$. Each open circle corresponds to the *R* factor as higher-resolution data were included and the closed circles to the *R* factor at the same resolution after the structure was refined to convergence as four rigid groups. The vertical drop at 6 Å resolution corresponds to the constrained temperature-factor refinement of each of the four rigid groups (*i.e.* an additional four degrees of freedom).

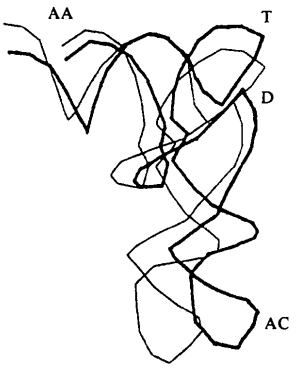


Fig. 6. Change in the $tRNA_f^{Met}$ structure after low- and intermediate-resolution refinement as four rigid groups (see Fig. 2c). The light trace represents the starting coordinates (R = 58%, 12·5–20 Å) and the heavy trace the structure after the initial group refinement (R = 42%, 6–20 Å).

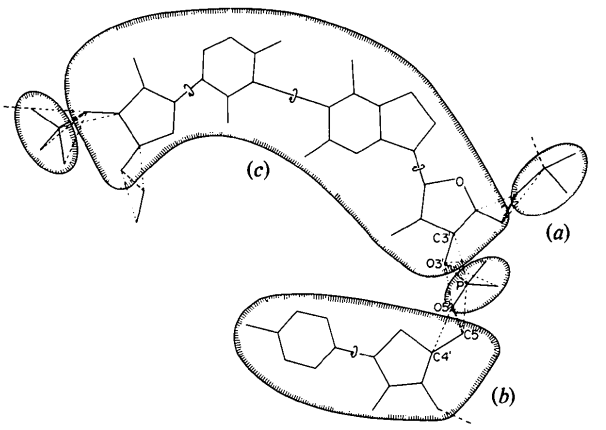


Fig. 7. The three different kinds of constrained groups used in the higher-resolution refinement of $tRNA_f^{Met}$ are shown. (*a*) The smallest is a phosphate group with six positional degrees of freedom and a single temperature factor (a total of seven degrees of freedom). (*b*) The next largest group is a nucleoside with torsional flexibility of the base relative to its ribose, as well as separate subgroup temperature factors for each moiety, *i.e.* ribose and base (a total of nine degrees of freedom). (*c*) The largest group is a constrained base pair (for the double-helical-stem regions). In addition to the torsional flexibility of each ribose relative to its respective base, one of the nucleosides is permitted to twist as a unit about a vector between the two bases (for a total of 13 degrees of freedom). Restrained distances corresponding to bond lengths are shown by dashed lines, and distances corresponding to bond angles by dotted lines.

stages up to 6 Å (see Table 1 and Fig. 5). At the end of this four-rigid-groups refinement at 6 Å resolution the R factor converged to 49·5%. The center of mass of the molecule moved to 0·94 Å further away from the original position. A comparison of this structure with the original model is shown in Fig. 6. The main difference between the two molecules at this level of refinement is in the relative position of the AC arm. This was confirmed by more detailed refinements which also showed that the conformation of the AC loop differs significantly from this model (see below).

Based on a Wilson plot, an initial temperature factor of 100 $Å^2$ was assigned to each atom in the structure. The temperature factors of the four different groups were then refined independently starting from this value and the R factor dropped to 42·3% (see Table 2). Although these temperature factors are quite high (as seen in the Wilson plot) they show large variations between different parts of the molecule that were later confirmed by more detailed refinement.

The base sequence which had been that of $tRNA^{Phe}$ (RajBhandary & Chang, 1968), was then changed to that of $tRNA_f^{Met}$ (Simsek & RajBhandary, 1972). Also, the domain constraints were liberated. For the loop regions the groups consisted of separate phosphates and nucleosides (ribose and base) and for the double-helical-stem regions the groups were phosphates and base-paired nucleosides (see Fig. 7). At this stage *neither* restraints *nor* constraints were imposed on the tertiary interactions, *i.e.* each base not in a helical stem was refined separately.

With this model, a 6·0 Å refinement was attempted, and it failed to converge, especially for the phosphate groups in the loops. It was observed that for these groups the translational refinement was well defined, but the rotational refinement was unstable. This was due to the limited resolution of the data. At 6·0 Å resolution a phosphate group is not anisotropic enough for a rotational refinement to converge. Therefore, a 4·5 Å refinement was started. At this

Table 2. *Group temperature factors of domains of* $tRNA_f^{Met}$

Group	Group temperature factor (B)
AA stem, CCA end	24 $Å^2$
T arm	29
D arm	65
AC arm, V arm	138

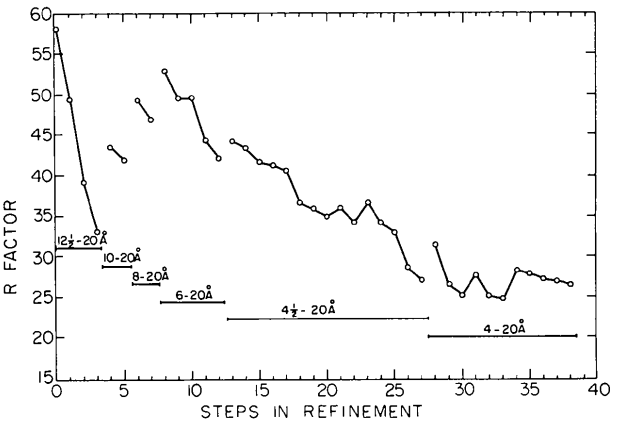


Fig. 8. R factor for the entire *CORELS* refinement of $tRNA_f^{Met}$ (see Table 3 for details of each step).

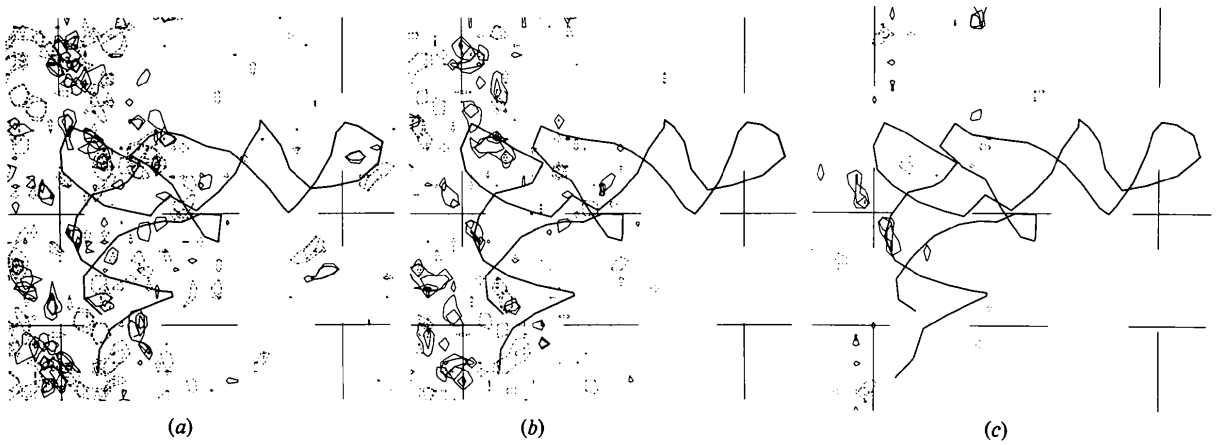
Table 3. *Course of CORELS refinement of the yeast* $tRNA_f^{Met}$ *crystal structure at* 4.0 Å *resolution*

Step	Refinement operation	*R* (%)	Correlation* coefficient	No. of data	Resolution of data (Å)	No. of restraints	No. of variables refined this step	Total† No. degrees of freedom	Ratio No. of data to total No. degrees of freedom	No. of cycles per step	Computer time (min)‡	Comments
(*A*)	Low-resolution large-rigid-group refinement											
0	Generated starting coordinates using residues 1–16, 18–75 of $tRNA^{Phe}$ fitted to computer-graphics model of $tRNA_f^{Met}$	58·1	0·10	143	12·5–20·0	–	–	–	–	–	0·5	
1	Refined entire structure as *single* rigid group	49·4	0·34	143	12·5–20·0	0	6	6	24·3	5	4·9	
2	Refined entire structure as *two* rigid groups	39·1	0·64	146	12·4–20·0	10	12	12	12·2	5	5·4	
3	Refined entire structure as *four* rigid groups	33·0	0·73	146	12·4–20·0	20	24	24	6·1	4	5·4	
4	Calculated *R* factor for 10–20 Å resolution	43·4	0·59	311	10·0–20·0	–	–	–	–	–	1·1	
5	Refined structure as four rigid groups	41·8	0·63	311	10·0–20·0	20	24	24	13·0	2	4·6	
6	Calculated *R* factor for 8–20 Å resolution	49·3	0·58	634	8·0–20·0	–	–	–	–	–	2·2	
7	Refined structure as four rigid groups	46·9	0·62	634	8·0–20·0	20	24	24	26·4	2	9·0	
8	Calculated *R* factor for 6–20 Å resolution	52·9	0·61	1400	6·0–20·0	–	–	–	–	–	7·5	
9	Refined structure as four rigid groups	49·5	0·71	1400	6·0–20·0	20	24	24	58·3	4	42·4	
10	Unrestrained refinement of structure as four rigid groups	49·5	0·72	1400	6·0–20·0	–	24	24	58·3	1	7·0	
11	Refined overall scale factor	44·2	0·72	1400	6·0–20·0	–	1	25	56·0	1	0·3	
12	Refined constrained temperature factors for each group	42·3	0·73	1400	6·0–20·0	–	4	29	48·3	2	23·0	
(*B*)	4½ Å resolution refinement											
13	Fitted sequence of $tRNA_f^{Met}$ to refined structure of step 10 and refined overall scale factor	44·1	0·69	2602	4·5–20·0	–	1	29	89·7	1	9·5	
14	Refined D stem as constrained phosphate groups and base pairs	43·2	0·70	2602	4·5–20·0	1722	84	855	3·0	3	8·3	In addition to covalent restraints between groups, non-bonded terms were included (as restraints) to prevent too close contacts between non-bonded atoms
15	Refined AA stem as in 14	41·5	0·72	2602	4·5–20·0	1722	141	855	3·0	3	14·5	
16	Refined AC stem as in 14	41·1	0·73	2602	4·5–20·0	1722	105	855	3·0	3	10·7	
17	Refined T stem as in 14	40·5	0·74	2602	4·5–20·0	1722	105	855	3·0	3	12·0	
18	Refined D, AC and T loops	36·6	0·78	2602	4·5–20·0	1722	273	855	3·0	3	28·0	Refined each loop in separate blocks and then repeated for a total of three cycles
19	Refined four helical stems in separate blocks	35·8	0·79	2602	4·5–20·0	1722	435	855	3·0	2	35·9	Refined each stem in separate blocks and then repeated for a total of two cycles
20	Refined D, AC and T loops	34·9	0·81	2602	4·5–20·0	1722	273	855	3·0	2	18·2	Refined each loop in separate blocks and then repeated for a total of two cycles

Table 3 (*cont.*)

Step	Refinement operation	*R* (%)	Correlation* coefficient	No. of data	Resolution of data (Å)	No. of restraints	No. of variables refined this step	Total† No. degrees of freedom	Ratio No. of data to total No. degrees of freedom	No. of cycles per step	Computer time (min)‡	Comments
(*C*)	Model building with static computer graphics and continuation of 4½ Å resolution refinement											
21	Model built, the 3′ terminal residues C74 and C75 to a 'chopped' difference Fourier map on static computer graphics	36·0	0·79	2602	4·5–20·0	–	–	–	–	–	1·0	
22	Refined residues 73–75	34·2	0·81	2602	4·5–20·0	2059	39	855	3·0	5	5·0	
23	Added on 3′ terminal A76 using static computer graphics, and then model built residues C75 and A76	36·7	0·79	2602	4·5–20·0	–	–	–	–	–	1·0	
24	Refined residues 73–76	34·2	0·81	2602	4·5–20·0	2082	50	870	3·0	5	7·7	
25	Refined half AA stem, D and T loops, and residues 73–75	32·9	0·82	2602	4·5–20·0	2082	321	871	3·0	2	21·5	Refined each block separately and then repeated
26	Refined constrained group temperature factor for each phosphate, ribose and base group	28·6	0·86	2602	4·5–20·0	–	225	1096	2·4	3	55·5	
27	Refined entire structure in 9 non-overlapping blocks	27·0	0·88	2602	4·5–20·0	1996	870	1096	2·4	2	67·1	Refined each block separately and then repeated
(*D*)	4 Å refinement											
28	Calculate *R* factor for 4–20 Å resolution data and refine overall scale factor	31·4	0·84	3302	4·0–20·0	–	1	1096	3·0	1	14·4	
29	Refined entire structure in 9 non-overlapping blocks	26·5	0·88	3302	4·0–20·0	1996	870	1096	3·0	2	86·9	Refined each block separately and then repeated
30	Refined constrained group temperature factors for each phosphate, ribose and base group	25·2	0·89	3302	4·0–20·0	–	225	1096	3·0	3	31·5	Began to use a faster version of *CORELS*
31	*CORELS* model built to target coordinates to improve stereochemistry	27·7	0·87	3302	4·0–20·0	–	–	–	–	–	12·0	
32	Refined AA, D and T stems, and residues 8-9, D loop, and the T loop in separate blocks	25·1	0·89	3302	4·0–20·0	1996	544	1096	3·0	2	38·5	Refined each block separately and then repeated
33	Refined constrained group temperature factors as in 22	24·7	0·90	3302	4·0–20·0	–	225	1096	3·0	3	30·6	
34	*CORELS* model built to target coordinates to improve stereochemistry, including some base–base tertiary restraints	28·3	0·87	3302	4·0–20·0	–	–	–	–	–	6·4	Explicit tertiary restraints included between residues U8-A14, G15-m^2C48, G19-C56, G22-m^7G46, m_2^2G26-A44, A54-m^1A58.
35	Refined AA, D and T stems in separate blocks	27·8	0·87	3302	4·0–20·0	2007	336	1096	3·0	3	34·5	Refined each block separately and then repeated twice
36	Refined AC stem, 8·9·D loop 25 and V loop, and T loop	27·2	0·88	3302	4·0–20·0	2007	391	1096	3·0	3	34·7	Refined each block separately and then repeated twice
37	Refined AC loop and 3′ terminal residues ACCA	27·0	0·88	3302	4·0–20·0	2023	143	1096	3·0	2	9·0	Refined each block separately and then repeated
38	Refine constrained group temperature factors as in 22	26·5	0·88	3302	4·0–20·0	–	225	1096	3·0	3	31·6	

* Correlation coefficient between F_o and F_c is: $\Sigma\,[(F_o - \bar{F}_o)(F_c - \bar{F}_c)]/[\Sigma(F_o - \bar{F}_o)^2 \times \Sigma\,(F_c - \bar{F}_c)^2]^{1/2}$.
† Total number of degrees of freedom for particular step, *i.e.* allowed flexibility.
‡ On an IBM 370/165 computer.



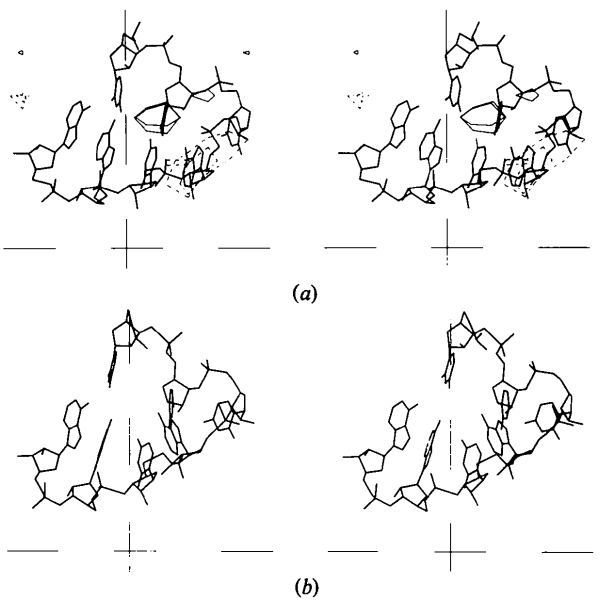
(*a*) (*b*) (*c*)

Fig. 9. Difference electron density maps at 4·5–20 Å resolution at three different stages in the refinement of the crystal structure of $tRNA_f^{Met}$ together with a superimposed trace of the backbone. (*a*) A map at $R = 40{\cdot}5\%$ where only residues in the four double-helical stems were allowed to vary freely during the refinement, while the loop regions were constrained to the $tRNA^{Phe}$ conformation. There is a large amount of unaccounted for electron density near the CCA end, in the region of the T and D loops as well as in the middle of the AC loop. (*b*) A map after several cycles of *CORELS* refinement allowing the residues in the loop region to vary, $R = 34{\cdot}9\%$. (*c*) A map at the end of the refinement having treated the entire structure as composed of the groups shown in Fig. 7, $R = 27\%$. Here the map as a whole is much cleaner than in (*a*) showing how the conformation of the structure changed during the refinement to fit the observed X-ray data better.

resolution, the refinement of the phosphate groups behaved considerably better. The starting point had an R factor of 44·1% at 4·5 Å resolution; for convenience of computing, the refinement proceeded in different steps, where at each step only a part of the molecule was refined. Fig. 8 and Table 3 show the overall behavior of the R factor during the entire refinement.

Fig. 9(*a*) shows a 4·5 Å difference map before the conformation of the loops was allowed to vary. Note the 'peak–hole' pair associated with the AC loop. We used a weighted difference map, where the weight associated with each reflection is $F_c^4/(F_c^4 - F_o^4)$ (Nordman in Stout & Jensen, 1968). The purpose of this weight was to accentuate the negative regions associated with wrongly located groups.

Several cycles of refinement at 4·5 Å resolution (see Table 3) succeeded in correcting the model substantially in several regions. An example is the anticodon loop, where a new difference map (Fig. 9*b*) showed a marked difference. The refinement had moved the loop significantly and the peak–hole pair had disappeared. Fig. 10 shows a close-up of this region; Fig. 10(*a*) shows the difference map and corresponding model before the conformation of the loop was allowed to vary; Fig. 10(*b*) shows the difference map after refining the conformation of the AC loop. The new conformation of the anticodon did not change significantly in subsequent refinement. However, Fig. 9(*b*) shows that there are many features of the difference map unaccounted for by this refinement model. Most of these features are related to a wrongly positioned CCA end, which was still in the $tRNA^{Phe}$ conformation and caused unacceptable close



(*a*)

(*b*)

Fig. 10. Enlarged stereoviews of the AC loop for the difference electron density maps shown in Fig. 9(*a*,*b*).

contacts with symmetry-related molecules. The difference map showed the right direction to move the CCA end. Furthermore, the presence of a covalently bound heavy-atom marker on residues 74 and 75 (Schevitz *et al.*, 1975, 1979) made it easier to position the CCA end. This was done using an interactive static graphics system (Katz & Levinthal, 1972; Podjarny, 1976), and after several trials a solution was found which fitted the difference map, had reasonable

stereochemistry, and agreed with the heavy-atom markers.

After further refinement of this manually corrected model, an unweighted difference map showed a much cleaner region with some remnants of holes near the CCA end. This was traced back to incorrect temperature factors, especially for the last residues of the CCA end. Following temperature-factor refinements, the R factor dropped to 28·6%, and the map was also improved (see Fig. 9*c*). This was followed by several cycles of refinement alternating stems and loops (see Table 3), converging to a final R factor at 4·5 Å resolution of 27%.

We then included the 4 Å data, going from 2602 to 3302 reflections. The starting R factor, using the result from the 4·5 Å refinement, was 31·4%. At 4 Å, we refined the whole structure for two cycles, in the same order that was used at 4·5 Å, to an R factor of 25·2%. A difference map calculated at this point was very clean in the molecular region. Some peaks corresponding to small molecules appear quite clearly. We interpret them as possibly spermines attached to the tRNA molecule.

In this model, over 85% of total bond distances and angles are constrained, and therefore have canonical values (see Fig. 7). However, at this stage we felt that the stereochemistry between groups (see Fig. 3) needed to be improved, as the intergroup average deviations from ideal bond lengths and bond angles were 0·14 Å and 12°. We rebuilt the structure using the current coordinates as guide points and then proceeded with reciprocal-space refinement using larger weights for the distance restraints. This resulted in a structure with intergroup bond-length and bond-angle deviations of only 0·05 Å and 5° respectively and a final R factor of 26·5%. The corresponding average deviations for all bond distances and angles are 0·008 Å and 0·6° respectively.

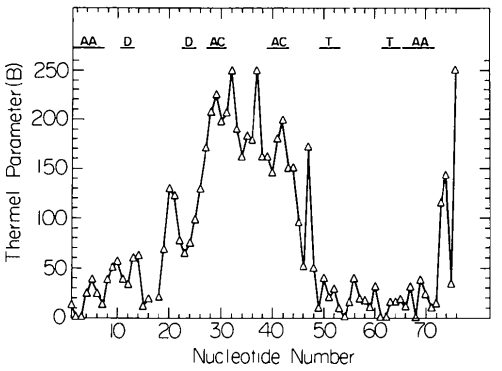


Fig. 11. Group thermal parameters (B's) (Å^2) for tRNA$_f^{Met}$ plotted as a function of nucleotide sequence number. For each residue the average of the group temperature factor for the phosphate, ribose and base moiety are plotted. The horizontal lines inserted near the top of the figure represent the positions of the various double-helical-stem regions for reference purposes (see Fig. 1).

Fig. 11 is a plot of the group temperature factors along the sequence. They are rather low for the acceptor arm, of medium value for the D stem, except for residues 20–21, very high for the AC arm and extra arm, low again for the T arm and very high for the very last residue. This shows that the molecule packed quite tightly near the 6_4 axis (Schevitz *et al.*, 1979), with increasing vibrational or static disorder as we go toward the anticodon arm.

In order to confirm the very high level of disorder shown by the temperature factors in the AC loop, a 'chopped' difference map (Shoham *et al.*, 1979) was calculated for this region (Fig. 12). This map is an $(F_{obs} - F_{calc}) \exp(i\varphi_{calc})$ difference map where the model used in calculating the F_{calc} does not include the AC loop and therefore whatever density is observed in that region is a function of the F_{obs} and the rest of the model. The chopped difference map correlates well with the temperature-factor plot, showing weak and discontinuous density in the AC loop (Fig. 12*b*) and stronger

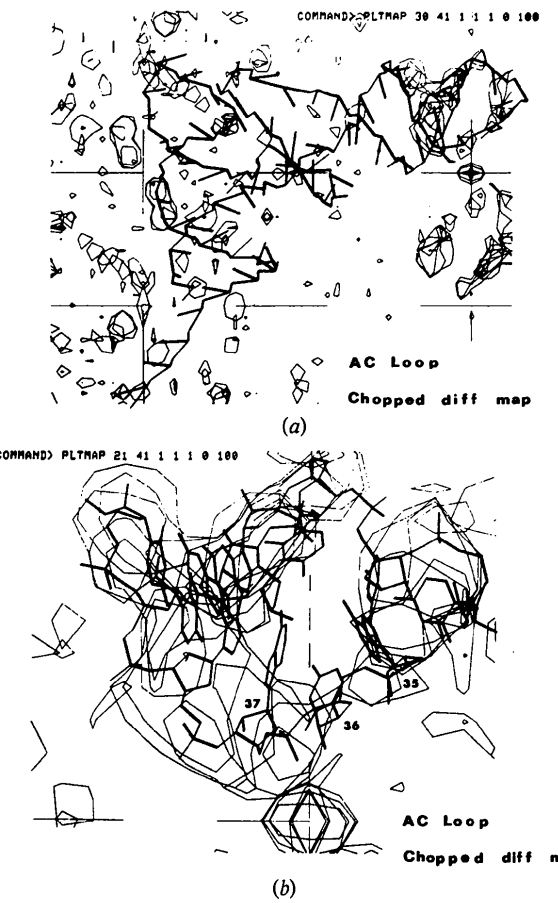


Fig. 12. 'Chopped' difference electron density maps for the 4·0 Å refined structure of yeast tRNA$_f^{Met}$. The coefficients of the Fourier summation are $(F_{obs} - F_{calc}) \exp(i\varphi_{calc})$, with all residues except 31–39 (*i.e.* the AC loop and last base pair of the AC stem) included in the calculation of the structure factor amplitudes and phases. (*a*) The entire structure and (*b*) a close-up of the AC loop.

density in the last base pair of the AC stem, which was also omitted for reference purposes. It should be noted that the density corresponding to the AC loop is above the noise level of the rest of the map as shown in Fig. 12(*a*). Similar chopped difference maps were calculated for different parts of the structure confirming in every case the results of the refinement.

Discussion

The low-resolution refinement described above shows some of the characteristics generally associated with molecular replacement. It combines the use of large portions of the structure which are roughly known, with a least-squares minimization procedure. The rigidity of the original model could easily be relaxed in gradual steps to accommodate the major differences of the original and final structures by introducing a minimal number of degrees of freedom, increasing the resolution in such a way that the number of observations was commensurate with the number of parameters. For example, at 12·5 Å resolution, the *F*'s are mostly a function of the spacing of helical stems, in accordance with the parameters used. (This statement might be difficult to generalize to proteins, because of the lower contrast of the polypeptide backbone and the solvent, in comparison with the polynucleotide backbone and the solvent.) It should be noted that the group-temperature-factor refinement at 6 Å resolution is clearly a gross approximation. Tests are being done on methods to improve the procedure.

Regarding the final refinement, it is clear that the lack of resolution of the data will strongly affect the accuracy of the parameters. Firstly, we should stress that unrestrained atomic refinement is not possible. Consequently, the final result should be interpreted in terms of group positions and not of atomic detail. We can note here that the tertiary interactions which were not restrained are in good agreement with similar ones obtained for $tRNA^{Phe}$ from high-resolution refinement (Sussman *et al.*, 1981). In contrast, other regions show significant departures from the $tRNA^{Phe}$ structure, *e.g.* the anticodon arm.

Some of these changes could be obtained from the refinement directly, *e.g.* the AC loop. In other cases the shifts were outside the radius of convergence of the method and required detailed examination of difference maps together with model building, *e.g.* the CCA end. 'Chopped' difference maps were very useful for verification of the results of the refinement.

An interesting question arises at this step, and that is the validity of the refinement of relatively small groups, like phosphate, at 4·0 Å resolution. As a rule, the final shifts in the translational parameters were very small, usually on the order of 0·1 Å and rarely as large as 0·45 Å (for groups with very high temperature factors). However, the rotational parameters, especially for the phosphate groups, did not always converge smoothly, and for 10 phosphate groups, P14, P21, P29, P30, P35, P36, P37, P45, P57 and P76, the last movement of the rotation angles had some component greater than 30°, associated with very high temperature factors. However, most of the ribose–base groups converged rather smoothly, and in only five cases, RA20, RD47, RC48, RC74 and RA76, did the rotation angles have final shifts greater than 20° and the temperature factors were high. For most cases, especially in the helical stems, the final rotation-angle shift was less than 10°. Phosphate groups are then well defined translationally, and ribose–base groups are well defined both translationally and rotationally. The instability of some phosphate groups in the rotational refinement must be related to its relative lack of anisotropy, even at 4·0 Å resolution. However, if a phosphate group is positioned between two well defined ribose groups and its translation is fixed, then its rotational parameters are quite well defined by restraints. The added stereochemical information should then be enough to position the small isotropic groups with rough accuracy.

It is interesting to try to estimate the error associated with very high temperature factors, such as that of P33, at the tip of the anticodon loop. This phosphate has no density in the chopped difference map. Its final translational shift is 0·45 Å in the refinement, and the positional uncertainty, $(\sqrt{\langle\mu^2\rangle})$, associated with a *B* of 250 Å^2 is 1·76 Å. On the other hand, a well defined group, like P10, has a final translational shift of 0·09 Å, while the $\sqrt{\langle\mu^2\rangle}$ associated with a *B* of 20 Å^2 is 0·5 Å.

In general, we can say that groups with low *B*'s are better defined by the refinement than those with large *B*'s, and that the error imposed by the refinement in the translational sense ranges between 0·09 and 0·45 Å, as estimated by the final shift of the group.

In summary, this work shows how the judicious use of *a priori* known conformational information can be used to carry existing refinement methods to their low-resolution limits. At this resolution the radius of convergence increases greatly, thus making it possible to correct large initial errors. It also opens the possibility of using stereochemical information to refine structures in cases where the resolution of the data is very limited.

We are very grateful to Professor Paul B. Sigler and Dr Richard W. Schevitz for very useful discussions and criticism at every stage of this work, and also for providing the X-ray diffraction data of yeast $tRNA_f^{Met}$. We are also grateful to Professor Wolfie Traub, Dr Ada Yonath and Dr John Moult for helpful comments. This project has been partially supported by the NSF-CONyCET (Argentina) collaboration No. 18, the US–Israel Binational Science Foundation and the Committee of Friends of the Weizmann Institute in

Argentina. We acknowledge the very generous support we received from the Weizmann Institute Computation Center.

References

DOEDENS, R. (1970). *Crystallographic Computing*, pp. 198–200. Copenhagen: Munksgaard.

HENDRICKSON, W. A., KLIPPENSTEIN, G. L. & WARD, K. B. (1975). *Proc. Natl Acad. Sci. USA*, **72**, 2160–2164.

HOLBROOK, S. R., SUSSMAN, J. L., WARRANT, R. W. & KIM, S.-H. (1978). *J. Mol. Biol.* **123**, 631–660.

JACK, A., LADNER, J. E. & KLUG, A. (1976). *J. Mol. Biol.* **108**, 613–643.

KATZ, L. & LEVINTHAL, C. (1972). *Annu. Rev. Biophys. Bioeng.* **1**, 465–504.

KONNERT, J. H. (1976). *Acta Cryst.* A**32**, 614–617.

MACHIN, P. A., CAMPBELL, J. W. & ELDER, M. (1981). *Refinement of Protein Structures*. Proceedings of the Daresbury Study Weekend, Daresbury Laboratory, Daresbury, Warrington, England.

NYBURG, S. C. (1974) *Acta Cryst.* B**30**, 251–253.

PODJARNY, A. D. (1976). PhD thesis, Weizmann Institute of Science, Rehovot, Israel.

PODJARNY, A. D., SCHEVITZ, R. W. & SIGLER, P. B. (1981). *Acta Cryst.* A**37**, 662–668.

RAJBHANDARY, U. L. & CHANG, S. A. (1968). *J. Biol. Chem.* **243**, 598–608.

RICHARDS, F. M. (1968). *J. Mol. Biol.* **37**, 225–230.

SCHEVITZ, R. W., KRISHNAMACHARI, N., HUGHES, J., ROSA, J., PASEK, M., CORNICK, G., NAVIA, M. A. & SIGLER, P. B. (1975). *Structure and Conformation of Nucleic Acids and Protein–Nucleic Acid Interactions*, edited by M. SUNDARALINGAM & S. T. RAO, pp. 85–99. Baltimore Univ. Park Press.

SCHEVITZ, R. W., PODJARNY, A. D., KRISHNAMACHARI, N., HUGHES, J. J., SIGLER, P. B. & SUSSMAN, J. L. (1979). *Nature (London)*, **278**, 188–190.

SCHEVITZ, R. W., PODJARNY, A. D., SUSSMAN, J. L. & SIGLER, P. B. (1983). In preparation.

SCHEVITZ, R. W., PODJARNY, A. D., ZWICK, M., HUGHES, J. J. & SIGLER, P. B. (1981). *Acta Cryst.* A**37**, 669–677.

SHOHAM, M., YONATH, A., SUSSMAN, J. L., MOULT, J., TRAUB, W. & KALB (GILBOA), A. J. (1979). *J. Mol. Biol.* **131**, 137–155.

SIGLER, P. B., SCHEVITZ, R. W., GROSS, M., KAPLANSKY, A., PODJARNY, A. D. & SUSSMAN, J. L. (1981). *Structural Aspects of Recognition and Assembly in Biological Macromolecules*, edited by M. BALABAN, pp. 631–643. Rehovot and Philadelphia: Balaban ISS.

SIMSEK, M. & RAJBHANDARY, U. L. (1972). *Biochem. Biophys. Res. Commun.* **49**, 508–515.

STOUT, G. H. & JENSEN, L. H. (1968). *X-ray Structure Determination. A Practical Guide*, p. 360. New York: Macmillan.

SUSSMAN, J. L., HOLBROOK, S. R., CHURCH, G. M. & KIM, S.-H. (1977). *Acta Cryst.* A**33**, 800–804.

SUSSMAN, J. L., HOLBROOK, S. R., WARRANT, R. W., CHURCH, G. M. & KIM, S.-H. (1978). *J. Mol. Biol.* **123**, 607–630.

SUSSMAN, J. L., PODJARNY, A. D., SCHEVITZ, R. W. & SIGLER, P. B. (1981). *Structural Aspects of Recognition and Assembly in Biological Macromolecules*, edited by M. BALABAN, pp. 597–614. Rehovot and Philadelphia: Balaban ISS.

TSERNOGLOU, D., PETSKO, G. A., MCQUEEN, J. E. & HERMANS, J. (1977). *Science*, **197**, 1378–1381.

Acta Cryst. (1983). B**39**, 505–516

A Description of the Techniques and Application of Molecular Replacement Used to Determine the Structure of Polyoma Virus Capsid at 22·5 Å Resolution

BY IVAN RAYMENT, TIMOTHY S. BAKER AND DONALD L. D. CASPAR

Rosenstiel Basic Medical Sciences Research Center, Brandeis University, 415 *South Street, Waltham, Massachusetts* 02254, *USA*

(*Received* 22 *November* 1982; *accepted* 15 *March* 1983)

Abstract

The electron density map of polyoma virus capsid crystals solved at 22·5 Å resolution by molecular replacement [Rayment, Baker, Caspar & Murakami (1982). *Nature* (*London*), **295**, 110–115] shows that the 72 capsomeres that form the polyoma capsid are all pentamers. An extensive series of refinement calculations were undertaken to demonstrate the validity of this unexpected result. This report describes the details of the data collection, structure determination and the tests of the methods applied. The refinement calculations demonstrate that the refined phases are insensitive to the initial phasing model for a wide variety of models. They also show that it is vital to include in the refinement calculations interpolated values for the unrecorded data. A variety of tests demonstrate that the all-pentamer structure of the polyoma capsid is determined by the diffraction amplitudes and the self-consistent constraint that the 72 capsomeres are arranged with icosahedral symmetry within a well defined limiting envelope.

Introduction

Crystals of the icosahedrally symmetric polyoma virus capsids give clear X-ray diffraction patterns to a

0108-7681/83/040505-12$01.50 © 1983 International Union of Crystallography