Acta Cryst. (1981). A**37**, 662–668

Phasing Low-Resolution Macromolecular Structure Factors by Matricial Direct Methods

BY ALBERTO D. PODJARNY

Department of Biophysics and Theoretical Biology, The University of Chicago, Chicago, Illinois 60637, *USA and Laboratorio de Rayos-X, Departamento de Fisica, UNLP, CC*67, *La Plata* 1900, *Argentina*

AND RICHARD W. SCHEVITZ AND PAUL B. SIGLER

Department of Biophysics and Theoretical Biology, The University of Chicago, Chicago, Illinois 60637, *USA*

(*Received* 26 *February* 1980; *accepted* 18 *March* 1981).

Abstract

Matricial direct methods were used to phase 28 strong low-resolution (100–19 Å) structure factors of crystalline yeast $tRNA_f^{Met}$ ($P6_422$, $z = 12$) which had not been phased by multiple isomorphous replacement (MIR). The starting phase set was composed of 107 terms in the resolution range 32–14 Å which had been phased by MIR. Extending the phase set to the strong low-resolution terms significantly improved the electron-density map. The goal of establishing a well defined molecular boundary was clearly achieved and provided the basis of a successful structure determination to 4·0 Å resolution. The phases determined by direct methods deviated from the phases subsequently calculated from the refined atomic coordinates by an unweighted average value of 73°; 36% of these were determined with a figure of merit greater than 0·75 and showed a discrepancy of only 44°. The accuracy of the phases determined by matricial methods compared favorably with those of the starting MIR phase set. An analysis of the resolution dependence of the intensities suggests plausible substructures as the basis of the normalization leading to the successful extension of phases to very low resolution.

Introduction

Direct methods, long a valuable tool in the crystal structure determination of small molecules, have been increasingly applied in the past decade to crystallographic studies of biological macromolecules (de Rango, Mauguen & Tsoucaris, 1975; Sayre, 1974; Podjarny & Yonath, 1977; Collins, Brice, La Cour & Legg, 1976; Bricogne, 1974). These methods have been used principally to improve the phases obtained by multiple isomorphous replacement (MIR)* and/or to extend the phase determination to resolutions higher than those obtainable by MIR. In a rather special case, direct methods have also been used to map the distribution of heavy atoms in isomorphous derivatives of protein crystals (*cf.* Wilson, 1978, and references therein).

Direct methods operate in either real or reciprocal space. Historically the first approaches were applied in reciprocal space and exploited either algebraic or probabilistic relationships between structure factors derived from *a priori* assumptions about the nature of the scattering density. For example, the Sayre equation (Sayre, 1952) gives an algebraic relationship between structure factors based on the constraint that the density function is organized in non-overlapping approximately equal atoms of known shapes. In cases where exact algebraic expressions cannot be used, probabilistic relationships among a small number of structure factors can be derived by assuming random distributions of scattering units in the cell (Cochran & Woolfson, 1955; Karle & Karle, 1966). Prior knowledge of the distribution of the scattering units can be used to improve these methods (Main, 1976).

When applied to macromolecular studies, these methods have produced their most impressive results at high resolution where the scattering units are the individual atoms and, therefore, the underlying assumptions are most correct. This is particularly true of the algebraic relationships such as Sayre's convolution equation (Sayre, 1972) and of algebraic matricial methods (Knosow, de Rango, Mauguen, Sarrazin & Tsoucaris, 1977; Navaza & Silva, 1979). Statistical matricial methods (Tsoucaris, 1970; Castellano, Podjarny & Navaza, 1973) were also originally applied to protein crystals at high resolution (de Rango, Mauguen & Tsoucaris, 1975; Podjarny, Yonath & Traub, 1976). Recently, Podjarny & Yonath (1977) have shown that, at least in one favorable case, these methods can be successfully applied at medium resolution (5–3 Å) to extend the phases of crystalline $tRNA^{Phe}$ to higher resolution. This was possible, however, only after the MIR phases had been improved by real-space techniques.

* In principle, direct methods could be applied to improve approximate phases obtained by molecular replacement, a technique in which a known molecular structure is placed in the same orientation and position as the same or similar structure in different crystals.

0567-7394/81/050662-07$01.00 © 1981 International Union of Crystallography

The second approach is applied in real space and addresses itself directly to a provisional electron density by modifying it according to chemically reasonable constraints – often equivalent to those underlying certain reciprocal-space techniques. New and presumably improved phases are obtained from the Fourier transform of the modified density and combined in some weighted way with the previous phases and the observed amplitudes. With the advent of 'fast Fourier' algorithms these techniques often proved computationally more efficient and flexible and in certain cases have proven a valuable complement to their reciprocal-space counterparts. They are discussed in the following paper (Schevitz, Podjarny, Zwick, Hughes & Sigler, 1981) where real-space direct methods were used to improve the low- and medium-resolution structure factors of crystalline yeast $t\mathrm{RNA}_f^{\mathrm{Met}}$.

Whether direct methods are applied to macromolecular crystal structures in real or reciprocal space, either alone or in combination, the trend has been either to improve the phases or to extend them to *higher* resolution. In this paper we report the successful application of reciprocal-space direct methods to very-low-resolution (100–19 Å) structure factors of yeast initiator *t*RNA that were not phased by MIR (Schevitz, Navia, Bantz, Cornick, Rosa, Rosa & Sigler, 1972; Schevitz, Podjarny, Krishnamachari, Hughes, Sigler & Sussman, 1979).

A novel element of this structure analysis was the use of statistical matricial methods to extend the phase determination to a resolution *lower* than the MIR phase set. Our effort was motivated by our inability to define confidently the molecular boundary in the MIR map which we attributed to the fact that 28 of the strong low-resolution (100–19 Å) terms – some of the very strongest in the diffraction pattern – were not phased and therefore omitted from the map. Normally, the omission of such terms is compensated by a large number of well phased terms of comparable or higher resolution. Our MIR phases were apparently not sufficiently accurate to compensate for the absence of the very strong low-resolution terms. Indeed, phases later calculated from the refined structure in the range 100–14 Å deviated from the MIR phases by a mean value of 77°. The absence of a clearly discernible molecular envelope not only confounded a detailed interpretation of the map, but also prevented us from exploiting the fact that over 80% of the crystal structure was solvent and therefore potentially amenable to phase improvement by real-space 'solvent leveling' (Hendrickson, 1981; Bricogne, 1974; Nixon & North, 1976). Our failure to phase these lower-resolution terms by MIR resulted from the poorly ordered crystal lattice ($B = 150$ Å^2) which reduced the average intensity at 4 Å resolution to less than 1% of the value at 20 Å resolution. To record accurately the very weak high-resolution intensities, each of the oscillation photographs was obtained by exhaustively irradiating a separate crystal. In so doing, the stronger low-resolution reflections were over-exposed and far exceeded the dynamic range of even the most weakly exposed film. Even if relatively accurate amplitudes had been obtained for the low-resolution terms of the parent and derivatives, it is likely that the heavy-atom contribution would not have been large enough to provide accurate difference amplitudes to phase the strongest of these reflections. To circumvent this problem the inner core of intense parent intensities was determined from a special set of weakly exposed precession photographs and the phases of 28 of the very strongest low-resolution reflections were determined by applying matricial methods to the amplitudes and MIR phases of 107 other reflections. The details of this procedure are explained below; however, it is evident from Fig. 2 that there was substantial enhancement of the contrast between molecule and solvent. The improved map enabled us to define confidently 65% of the unit cell as solvent.

The map derived from the application of reciprocal-space procedures was improved enough to serve as a basis for the construction of a molecular model. This model was constrained by the requirement that interpretation must accommodate the position of four heavy-atom markers covalently linked to specific nucleotide residues (Rosa & Sigler, 1974; Pasek, Venkatappa & Sigler, 1973; Tropp & Sigler, 1979) and that the molecule was composed of helical domains of the clover-leaf hydrogen-bonded scheme. The model ultimately served as a basis for a constrained restrained least-squares refinement (Sussman, Holbrook, Church & Kim, 1977; Sussman & Podjarny, 1981) in which elastically linked helical segments of the structure were refined as rigid bodies against the low-resolution amplitudes. The size of the rigid bodies was progressively reduced to nucleotide bases, ribose and phosphate groups and the entire structure was refined to 4 Å resolution with a residual of 25% (Schevitz, Podjarny, Krishnamachari, Hughes, Sigler & Sussman, 1979).

Methods and results

Multivariate analysis was used to predict 28 reflections in the resolution range 100–19 Å. The details of the theory are described elsewhere (Tsoucaris, 1970; Castellano, Podjarny & Navaza, 1973; Podjarny, Yonath & Traub, 1976). Here we only describe the basic assumptions underlying the method and the most important equations used in applying it.

(*a*) *Normalizing structure factors at low resolution*

The structure amplitude $F(\mathbf{h})$ for reflection $\mathbf{h}$ is obtained from its intensity $I(\mathbf{h})$ by

$$|F(\mathbf{h})|^2 = A\, I(\mathbf{h}) \exp(2B_t s^2), \quad (1)$$

where A is a scale factor, B_t is a temperature factor, and s is $\sin\theta/\lambda$. We obtain normalized structure factors $E(\mathbf{h})$ (Main, 1976) from

$$|E(\mathbf{h})|^2 = A\, I(\mathbf{h})/\langle |F(\mathbf{h})|^2 \rangle \exp(-2B_t s^2),$$

where $\langle |F(\mathbf{h})|^2 \rangle$ is the mean amplitude for the zone of resolution defined by $|\mathbf{h}|$. If the structure is assumed to consist of N_g group scatterers, each with n_i atoms, then, following Main (1976), the group scattering factor $g_i(\mathbf{h})$ is defined for correctly positioned atoms as

$$g_i(\mathbf{h}) = \sum_{j=1}^{n_i} f_j \exp[2\pi i \mathbf{h}.(\mathbf{r}_j - \mathbf{R}_i)], \quad (2)$$

where $\mathbf{R}_i$ is the position of the center of mass of the ith group, so that

$$F(\mathbf{h}) = \sum_{i=1}^{N_g} g_i(\mathbf{h}) \exp(2\pi i \mathbf{h}.\mathbf{R}_i)$$

and, for randomly positioned groups,

$$\langle |F(\mathbf{h})|^2 \rangle = \sum_{i=1}^{N_g} |g_i(\mathbf{h})|^2. \quad (3)$$

From (2)

$$|g_i(\mathbf{h})|^2 = \sum_{j=1}^{n_i} \sum_{l=1}^{n_l} f_j f_l \exp 2\pi i \mathbf{h}.(\mathbf{r}_j - \mathbf{r}_l). \quad (4)$$

By spherically averaging over all group orientations, we obtain at low resolution*

$$\langle |g_i(\mathbf{h})|^2 \rangle = \sum_{j=1}^{n_i} \sum_{l=1}^{n_l} f_j f_l [1 - (2/3)\pi^2 |\mathbf{h}|^2 |\mathbf{r}_j - \mathbf{r}_l|^2],$$

making $\langle |g_i(\mathbf{h})|^2 \rangle$ an approximately spherically symmetric function of $|\mathbf{h}|$. If we further assume equal groups, with isotropic, real, positive group scattering factors

$$g(|\mathbf{h}|) = \langle |g_i(\mathbf{h})|^2 \rangle^{1/2}, \quad (5)$$

then from (1), (3) and (5)

$$\langle I(\mathbf{h}) \rangle = A^{-1} N_g g^2(|\mathbf{h}|) \exp(-2B_t s^2). \quad (6)$$

* Expanding the Debye formula (Debye, 1915; Main, 1976) for spherically averaged groups

$$\langle |g_i(\mathbf{h})|^2 \rangle = \sum_{j=1}^{ni} \sum_{l=1}^{ni} f_j f_l \frac{\sin(2\pi |\mathbf{h}||\mathbf{r}_j - \mathbf{r}_l|)}{2\pi |\mathbf{h}||r_j - \mathbf{r}_l|},$$

we obtain

$$\langle |g_i(\mathbf{h})|^2 \rangle = \sum_{j=1}^{ni} \sum_{l=1}^{ni} f_j f_l (1 - \tfrac{2}{3}\pi^2 |\mathbf{h}|^2 |\mathbf{r}_i - \mathbf{r}_l|^2 + \text{terms of order 4 and higher})$$

and at very low resolution the terms of order 4 or higher become very small for almost all of the atomic pairs.

Assuming that the groups contain roughly the same number of equal atoms n, extrapolating (4) to $|\mathbf{h}| = 0$ gives

$$(g^0)^2 = n^2 (f^0)^2,$$

where the superscript ^{0} denotes extrapolation to $|\mathbf{h}| = 0$. Therefore the mean intensity $\langle I^0 \rangle$ obtained from (6) as $|\mathbf{h}|$ goes to zero is given by

$$\langle I^0 \rangle = N_g n^2 (f^0)^2 A^{-1}$$

or

$$\langle I^0 \rangle = n N_a (f^0)^2 A^{-1},$$

where $N_a = N_g n$ is the total number of atoms in the asymmetric unit.

Fig. 1 shows the I curve, a plot of $\ln\langle I \rangle^{1/2}$ as a function of d^*, the reciprocal of the resolution expressed as the normal interplaner spacing in Å. $\langle I^0 \rangle_L$ is obtained by extrapolating the *low*-resolution part of this I curve ($d \geq 14$ Å) to $d^* = 0$. Assuming the relationship

$$\langle |F(\mathbf{h})|^2 \rangle = \sum_{j=1}^{N_a} f_j^2(\mathbf{h})$$

holds at higher resolution ($d < 10$ Å), we obtain from (1) for equal atoms as $|\mathbf{h}|$ goes to zero

$$\langle I^0 \rangle_H = N_a (f^0)^2 A^{-1},$$

where $\langle I^0 \rangle_H$ is obtained by extrapolating the high-resolution part of the I curve to $d^* = 0$. By combining the above we obtain n, the number of atoms per group.

$$n = \frac{\langle I^0 \rangle_L}{\langle I^0 \rangle_H}. \quad (7)$$

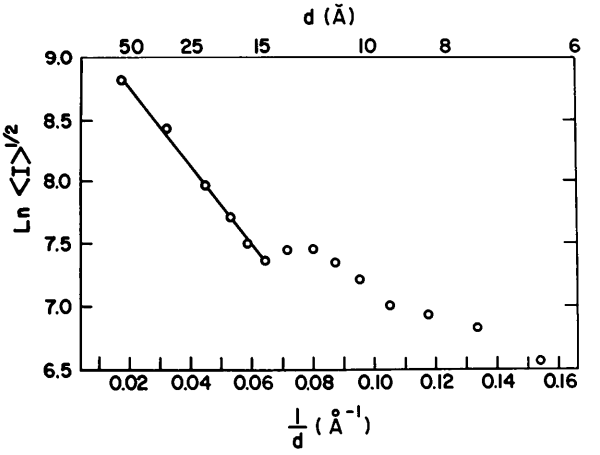

Fig. 1. The variation of the mean intensity $\langle I \rangle^{1/2}$ with resolution d^* or I curve. The straight line represents the least-squares fit to these data expressed in equation (6). The exceedingly strong term 100 which arises from the tight molecular clustering around the 6_4 screw axis was omitted.

Application of (7) to both the high- and low-resolution parts of the I curve gives a value for n in the the range of 30 to 36.

Assuming randomly positioned and oriented equal groups we use (3) to write the normalized structure factor

$$|E(\mathbf{h})|^2 = \frac{|F(\mathbf{h})|^2}{\langle |F(\mathbf{h})|^2 \rangle} = \frac{|F(\mathbf{h})|^2}{N_g g^2(|\mathbf{h}|)}.$$

The normalization factor $N_g g^2(|\mathbf{h}|)$ is evaluated by applying (6) to the value of $\langle I \rangle$ obtained from the least-squares fit of the observed intensities to the expression (Karle, 1976)

$$\ln \langle I \rangle^{1/2} = a - b|\mathbf{h}|^c. \tag{8}$$

Since $N_g = N_a/n$, where n is derived from (7), we can evaluate $E_{000} = N_g^{1/2}$ and therefore the unitary structure factor

$$U(\mathbf{h}) = \frac{E(\mathbf{h})}{E_{000}}.$$

The data between 100 and 15 Å are best fit by least squares to a linear function

$$\ln \langle I \rangle^{1/2} = 9{\cdot}383 - 31{\cdot}63d^*, \tag{9}$$

where $\langle I \rangle$ is the mean value of the intensity for a zone of resolution having mean value d^*, and $c = 1$. $E(\mathbf{h})$ is then given by

$$E(\mathbf{h}) = F(\mathbf{h})/[\langle I \rangle^{1/2} A^{1/2} \exp(B_t s^2)],$$

where $\langle I \rangle^{1/2}$ is derived from the parameterized relationship given in (9). The implications of this normalization are discussed below.

(*b*) *Building the matrix*

From a set of M structure factors $E(\mathbf{h}_1) \ldots E(\mathbf{h}_2) \ldots E(\mathbf{h}_M)$ we select a subset $E_1, \ldots, E_N$ where $N \leq N_g$. We assume these to be a set of random variables, and calculate all pairwise correlations, which are nonzero, since the structure factors arise from a common structure, and are therefore not independent. In space group P_1, the correlation of using E_1 and E_2 is a third structure factor $U(\mathbf{h}_1 - \mathbf{h}_2) = E(\mathbf{h}_1 - \mathbf{h}_2)N_g^{1/2}$, which can be obtained from the original set of M structure factors. In general, the correlation σ_{ij} of two structure factors E_i and E_j is a sum over the unitary structure factors $U(\mathbf{h})$, given by

$$\sigma_{ij} = \sum_s \langle U(\mathbf{h}_i G_s - \mathbf{h}_i) \rangle \sum_t \exp(2\pi i \mathbf{h}_i . \mathbf{g}_{s,t}),$$

where G_s and $\mathbf{g}_{s,t}$ are the symmetry matrix and translation factor corresponding to the s symmetry operation (Castellano, Podjarny & Navaza, 1973, equation 2.4). Considering now the whole set of generating random variables $E_1, \ldots, E_n$; E_i and E_j are correlated by the element σ_{ij} of the Goedkoop matrix (Goedkoop, 1950). The σ matrix is used to correlate the phase of the last term E_N with the amplitudes and phases of the rest of the generating variables, E_1 through E_{N-1}. The 'regression' equation we obtain is

$$(\bar{E}_N) = (-D_{NN})^{-1} \sum_{j=1}^{N-1} D_{Nj} E_j,$$

where D_{Nj} is the last row of D_{ij}, which is the matrix obtained by inverting σ.

We are interested in the phase, φ_N, of $\bar{E}_N$, the predicted value of the structure factor E_N. As the prediction of φ_N is of statistical nature, it is accompanied by a figure of merit, *fm*, defined by

$$fm = I_1(B)/I_0(B),$$

where I_1 and I_0 are modified Bessel functions and $B = 2|E_N||\bar{E}_N|D_{NN}$. The quality of the phase prediction thus depends on the predicted and observed structure factor amplitudes.

To predict other phases (from the set of M reflections) we sequentially set each desired structure factor to E_N, and perform the corresponding regression on the set $E_1, \ldots, E_{N-1}$. We have to recalculate D_{Nj} in every instance; however, we do not vary the set $E_1, \ldots, E_{N-1}$ but only the last row and column of the matrix σ keeping constant rows and columns to order $N - 1$. This simplifies considerably the task of matrix inversion (Podjarny, Yonath & Traub, 1976).

From a total set of M structure factors, we take the $N - 1$ largest E values as 'generating reflections'. The matrix is then defined by the variables N and N_g which determines the scaling of the U's. Previous experience (Podjarny, Yonath & Traub, 1976; Knosow, de Rango, Mauguen, Sarrazin & Tsoucaris, 1977) shows that N should vary between $N_g/10$ and $N_g/3$. A matrix was evaluated by its capacity to predict correctly phases determined by MIR. Furthermore, the 'occupancy' of the matrix, which tells the percentage of sites in the matrix occupied by reflections of known phase (reflections of unknown phase being set to zero), measures how much external phase information is incorporated into the matrix. For test cases, such as triclinic lysozyme (Podjarny, Yonath & Traub, 1976), an occupancy of 70% was effective whereas occupancies less than 30% were clearly deficient.

N and N_g were varied to find the most suitable matrices for phase determination. We first assessed the matrix's capacity to predict phases already determined by MIR by following the discrepancy as a function of the matricial figure of merit (Table 1). Phases predicted with a high figure of merit should agree best with the MIR values. If there were too few reflections to build a suitable matrix, centric reflections were incorporated with the sign which best predicted the known MIR phases.

In order to establish the optimal matrix order and occupancy we initially addressed ourselves to only 107 reflections for which MIR phases were available in the resolution zone 100–14 Å and found that the matricially determined phases deviated from the MIR phases by an unacceptable mean value of 84°. Moreover, many highly erroneous phases were determined with a high matricial figure of merit. This problem reflected the fact that three very-low-order intense reflections, 100, 300 and 003, were omitted as they had not been phased by MIR. Being centrosymmetric, these highly interactive terms were readily included by employing the sign which led to a positive definite matrix and the best phase predictions.* Table 1 evaluates the results of the matricial phasing of the 107 terms which were phased by MIR using 110 terms (107 plus the three very intense centrosymmetric terms). The matrix of order 4 having 30 group scatterers gave the minimum deviation from MIR phases for terms with a figure of merit greater than 0·75.

(c) Determining the phase of 28 strong low-resolution terms

Table 2 summarizes the results of the phase determination for all 28 reflections phased only by direct methods – including the three very strong low-order centrosymmetric terms. We used matrices of order 4 and 44% occupancy employing 30 group scatterers. These parameters were indicated as optimal by the test case described above. The correctness of this choice was subsequently borne out by two criteria. Firstly, the matricially determined phases deviated from phases calculated from the refined model by a mean value of 73°. This compares favorably with a deviation of 77° for the 107 MIR phases. More importantly, this case shows a favorable inverse correlation between figure of merit and phase error in that 36% of the previously unknown phases were determined with a figure of merit greater than 0·75 and a mean phase error of only 44°.

The effect of matricial direct methods is convincingly seen in Fig. 2 which contrasts the MIR map with a map containing the same structure factors plus the 28 structure factors phased with matricial direct methods and weighted with their respective figures of merit. Large regions of the map have been voided of significant density and are clearly solvent; thus the molecular envelope has been roughly determined. More objective indications of improvement were the fact that density wrongly located on dyad axes (including the 6_4 screw) was considerably attenuated, and that the remaining density encompassed the sites marked by heavy atoms covalently linked to specific residues.

* The phases of the 100 and 300 reflections were later shown to correspond to the phases calculated from the refined model, whereas the phase of the 003 reflection proved to be wrong – despite the fact that it was determined with a figure of merit of 0·99. This almost certainly reflects a systematic error in the MIR phases which was discovered after the structure was solved.

Conclusions

After the identification of the molecular boundary, the electron density map was more interpretable in terms of

Table 1. *Comparison of* 107 *low-resolution* (100–14 Å) *phases determined by both matricial direct methods* (φ^{MAT}) *and MIR* (φ^{MIR})

Number of terms in parentheses.

			$\varphi^{MAT} - \varphi^{MIR}$ (°)					
				Range of figure of merit (*fm*)				
$N-1$	Matrix occupancy (%)	Number of groups	Overall	0·0–0·42	0·42–0·75	0·75–0·84	0·84–0·87	0·87–1·0
4	44	20	77·8 (107)	90 (45)	60 (32)	91 (15)	67 (4)	64 (10)
4	44	30	76·2 (107)	78 (58)	88 (30)	37 (7)	37 (8)	117 (3)
5	37	30	72·3 (107)	73 (60)	80 (28)	45 (10)	101 (2)	95 (2)
5	37	40	71·6 (107)	71 (61)	73 (34)	71 (7)	27 (2)	174 (1)

Table 2. *Comparison of* 28 *low-resolution* (100–19 Å) *phases determined only by matricial direct methods* (φ^{MAT}) *and phases calculated from the atomic coordinates of a refined model* (φ^{MOD})

Number of terms in parentheses.

			$\varphi^{MAT}-\varphi^{MOD}$ (°)			
				Range of figure of merit (*fm*)		
$N-1$	Matrix occupancy (%)	Number of groups	Overall	0–0·42	0·42–0·75	0·75–1·0
4	44	20	73 (28)	89 (8)	45 (6)	76 (14)
4	44	30	73 (28)	83 (11)	100 (7)	44 (10)
5	37	30	77 (28)	88 (10)	83 (9)	47 (9)
5	37	40	78 (28)	93 (13)	80 (7)	50 (8)

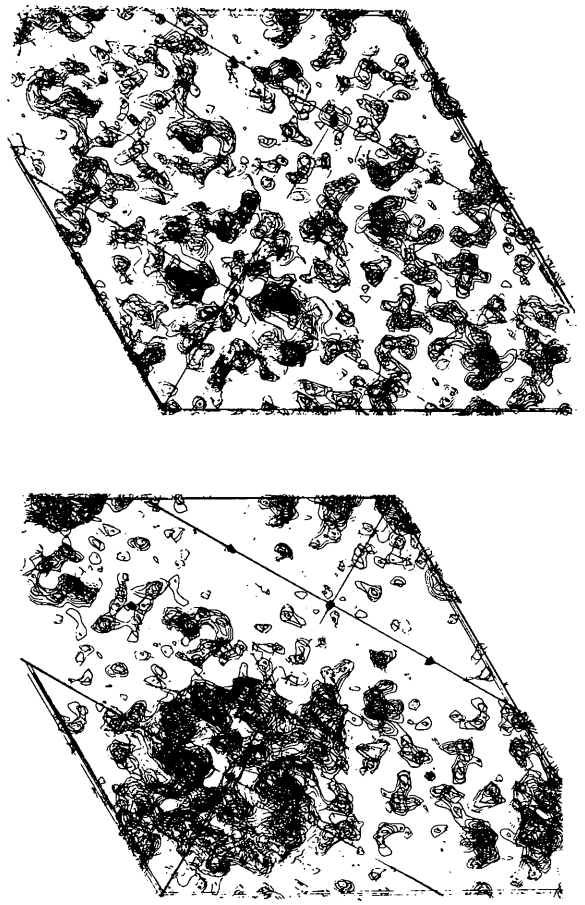

Fig. 2. Comparison of ten sections of the 4·5 Å map of yeast tRNA$_f^{Met}$ before (above) and after (below) the inclusion of 28 intense low-resolution terms whose phases were determined by matricial methods. The maps are contoured at the same intervals.

a molecular model. The immediate clues to the correctness of the new map were the deletion of electron density on the dyad axes (including the 6_4 screw) and the coincidence of the remaining density with heavy-atom markers which had been previously established as attached to known residues of the molecule. The map was further interpreted in terms of a model and this model refined to an R factor of 25% (Schevitz, Podjarny, Krishnamachari, Hughes, Sigler & Sussman, 1979).

It is clear that identifying the molecular envelope and improving the electron density distribution by matricial phases of low-resolution structure factors played a crucial role in the production of an interpretable map. It is therefore worth while identifying and discussing the unique features of this procedure.

Firstly, both real- and reciprocal-space direct methods have been applied with success almost exclusively at high resolution where the near-atomicity of the structures provides a firm basis for the underlying assumptions.† The first well-documented successful departure from high resolution was the work of Podjarny & Yonath (1977) to extend the phases of yeast tRNA^{Phe} from 5 to 3 Å resolution after starting with the phases of a modified MIR map in which the negative density had been strongly attenuated. They attribute part of their success at medium resolution to the fact that the electron density of tRNA can be organized into substructures representing the base, sugar and phosphate groups and in the resolution range 5–3 Å the density can be interpreted in terms of these scattering groups. A more complete treatment of the relationship between scattering groups and direct methods will be presented elsewhere (Podjarny, 1981); however, it is worth analyzing the group scattering properties of tRNA in connection with the current work.

It can be shown that the linear fit shown in Fig. 1 and expression (9) is a noticeably better statement of the resolution dependence of the intensity than the more commonly encountered d^{*2} dependency and that at this very low resolution the contribution of 'thermal' disorder to this expression is negligible (Podjarny, 1981). The Fourier transform of the exponential of a linear function shows that the spherically averaged scattering group is not Gaussian in character but a rather flatter function

$$g(\mathrm{r}) \propto 1/[(b/2\pi)^2 + r^2]^2,$$

where r is the radius of the group, $b/2\pi$ represents the radius at 0·25 of the maximum density, and b is taken from (9).

Taking $2{\cdot}5(b/2\pi)$ as the effective packing radius‡ the molecular volume of crystalline yeast tRNA$_f^{Met}$ (Johnson, Adolph, Rosa, Hall & Sigler, 1970) can account for 30 to 35 overlapping scattering elements which closely approximate the 30 scattering groups that produce the optimal matrix. Since the diameter of an RNA helix and thus the full reach of a base pair is roughly 25 Å and there are 30 secondary and tertiary base pairing interactions, it is tempting to speculate that the scattering groups that dominate the resolution dependency used to best normalize the very-low-resolution intensities are base pairs. As a base pair contains 42 or 43 nonhydrogen atoms, this speculation is roughly consistent with the value of 30–36 derived for n, the number of atoms per group (equation 7).

This speculation can be extended to the nature of the peak at 13·5 Å in Fig. 1, which when inserted in the

† Direct methods have successfully solved the distribution of heavy atoms in isomorphous heavy-atom derivatives of proteins at low resolution ($d \geq 6$–8 Å). However, a low-resolution study of a heavy-atom constellation composed of a small number of widely separated strong scattering elements situated in a unit cell large enough to accommodate a protein is the effective equivalent of a high-resolution study of a small structure.

‡ In terms of $g(r)/g(0)$, this is equivalent to $r = 2{\cdot}81\sigma$ in a Gaussian curve.

Debye formulation suggests that within the dominant low-resolution scattering elements there is a characteristic spacing of strong electron density at about 16 Å. This would correspond to the spacing of the opposing phosphate groups in an RNA helix. By a similar analysis the linear dependency of intensity on resolution between 10 and 6 Å would correspond to a group scattering density of roughly 3–4 Å radius and which probably represents the individual bases, sugars and phosphates, of which the phosphates are the strongest scattering elements.

In summary, the normalization of the intensities may well be rationalized in the form of scattering groups (Main, 1976) and may form the basis for the successful application of matricial methods to the low-resolution 'inward' extension of phase information. This idea is currently under study (Podjarny, 1981). In any case, our experience suggests that we should continue to apply reciprocal-space direct methods, where appropriate, to improve the low-resolution imaging of macromolecular crystal structures.

This research was supported by grants from the USPHS (GM 15225), the NSF (Int 78-21875), and CONICET (Joint Project No. 18). ADP is a fellow of the CONICET, Argentina.

References

BRICOGNE, G. (1974). *Acta Cryst.* A**30**, 395–405.

CASTELLANO, E. E., PODJARNY, A. D. & NAVAZA, J. (1973). *Acta Cryst.* A**29**, 609–615.

COCHRAN, W. & WOOLFSON, M. M. (1955). *Acta Cryst.* **8**, 1–12.

COLLINS, D. M., BRICE, M. D., LACOUR, T. F. M. & LEGG, M. J. (1976). In *Crystallographic Computing Techniques*, edited by F. R. AHMED, pp. 330–335. Copenhagen: Munksgaard.

DEBYE, P. (1915). *Ann. Phys. (Leipzig)*, **46**, 809.

GOEDKOOP, J. A. (1950). *Acta Cryst.* **3**, 374–378.

HENDRICKSON, W. A. (1981). *Structural Aspects of Biomolecules*, edited by R. SRINIVASAN & V. PATTABHI, pp. 31–80. New Delhi: Macmillan.

JOHNSON, C. D., ADOLPH, K., ROSA, J. J., HALL, M. D. & SIGLER, P. B. (1970). *Nature (London)*, **226**, 1246–1247.

KARLE, I. L. (1976). In *Crystallographic Computing Techniques*, edited by F. R. AHMED, pp. 27–67. Copenhagen: Munksgaard.

KARLE, J. & KARLE, I. L. (1966). *Acta Cryst.* **21**, 849–859.

KNOSOW, M., DE RANGO, C., MAUGUEN, Y., SARRAZIN, M. & TSOUCARIS, G. (1977). *Acta Cryst.* A**33**, 119–125.

MAIN, P. (1976). In *Crystallographic Computing Techniques*, edited by F. R. AHMED, pp. 97–105. Copenhagen: Munksgaard.

NAVAZA, J. & SILVA, A. M. (1979). *Acta Cryst.* A**35**, 266–275.

NIXON, P. E. & NORTH, A. C. T. (1976). *Acta Cryst.* A**32**, 325–333.

PASEK, M., VENKATAPPA, M. P. & SIGLER, P. B. (1973). *Biochemistry*, **12**, 4834–4840.

PODJARNY, A. D. (1981). In preparation.

PODJARNY, A. D. & YONATH, A. (1977). *Acta Cryst.* A**33**, 655–661.

PODJARNY, A. D., YONATH, A. & TRAUB, W. (1976). *Acta Cryst.* A**32**, 282–292.

RANGO, C. DE, MAUGUEN, Y. & TSOUCARIS, G. (1975). *Acta Cryst.* A**31**, 227–233.

ROSA, J. J. & SIGLER, P. B. (1974). *Biochemistry*, **13**, 5102–5109.

SAYRE, D. (1952). *Acta Cryst.* **5**, 60–65.

SAYRE, D. (1972). *Acta Cryst.* A**28**, 210–212.

SAYRE, D. (1974). *Acta Cryst.* A**30**, 180–184.

SCHEVITZ, R. W., NAVIA, M. A., BANTZ, D. A., CORNICK, G., ROSA, J. J., ROSA, M. D. & SIGLER, P. B. (1972). *Science*, **177**, 429–431.

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

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

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

SUSSMAN, J. L. & PODJARNY, A. D. (1981). In preparation.

TROPP, J. & SIGLER, P. B. (1979). *Biochemistry*, **24**, 5489–5495.

TSOUCARIS, G. (1970). *Acta Cryst.* A**26**, 492–499.

WILSON, K. S. (1978). *Acta Cryst.* B**34**, 1599–1608.