Portland State University

## PDXScholar

1987

# Improving Crystallographic Macromolecular Images: The Real-Space Approach

A. D. Podjarny
*University of Chicago*

T. N. Bhat
*National Institutes of Health*

Martin Zwick
*Portland State University*, zwick@pdx.edu

### Citation Details

# IMPROVING CRYSTALLOGRAPHIC MACROMOLECULAR IMAGES: THE REAL-SPACE APPROACH[1]

## A. D. Podjarny[2]

Department of Biochemistry and Molecular Biology, University of Chicago, 920 East 58th Street, Chicago, Illinois 60637

## T. N. Bhat

Laboratory of Molecular Biology, National Institutes of Health, NIDDK, Building 2, Bethesda, Maryland 20892

## M. Zwick

Systems Science PhD Program, Portland State University, Portland, Oregon 97207

CONTENTS

# PERSPECTIVES AND OVERVIEW

Macromolecular crystallography is a unique tool for imaging the structures of proteins and nucleic acids. Images are obtained from the Fourier transform of the diffraction pattern of the crystal by use of X-ray, neutron, and/or electron scattering. When X-ray and neutron scattering are used only diffraction amplitudes are experimentally measured, and phases have to be obtained. Multiple isomorphous replacement (MIR) has been the technique of choice for solving this phase problem in the determination of most macromolecular structures. Unfortunately, the method is extremely time consuming, especially when compared with the solution techniques available for small molecules; moreover, structure solution by MIR, even after many years of work, is hardly guaranteed. These drawbacks have stimulated efforts to enhance MIR as a phasing technique. The methods discussed in this paper (without exhaustive coverage, owing to space limitations) have so far been used to refine and/or to extend MIR phases, and also to open up the possibility of ab initio phase determination.

Following the early fundamental work of Karle & Hauptman (34, 35) and Sayre (60), reciprocal-space direct methods were applied to solve the structure of the majority of small molecules (via widely used packages, e.g. MULTAN and SHELX). These methods are used to derive phases statistically from the atomic character of the density. The extension of these methods to macromolecular crystallography is beyond the scope of this review.

Macromolecules present a more difficult problem. The diffraction data are rarely obtained at high enough resolution for the application of the atomicity constraint. Also, the accuracy of the phase predictions by reciprocal-space direct methods decreases with the size of the molecule. However, there are other a priori physical constraints applicable to macromolecular density functions, e.g. continuity and solvent flatness. These constraints are more readily expressed in real space than in reciprocal space. Procedures that exploit such physical constraints in real space are commonly known as "density modification" (DM) methods. These techniques do not merely consist of the real space imposition of a priori physical constraints, but also include reciprocal-space steps of comparable importance. These mixed real- and reciprocal-space DM algorithms are the main subject of this review.

# ALGORITHMS

DM methods are aimed at improving the agreement of a density map with a set of physically meaningful constraints. Density maps must agree simultaneously with experimental data such as the observed amplitudes and the MIR phase distribution and with physical constraints based on a priori knowledge of the characteristics of the density function. These physical constraints include positivity, atomicity (at high resolution), boundedness, uniformity of solvent regions, continuity of the biopolymer chain, and known noncrystallographic symmetry of electron density distribution.

A starting map to be processed by DM, such as an MIR map, is usually obtained from, and thus will fully agree with, the experimental data. If the map also agrees with the set of physical constraints, there is no room for DM techniques, and map interpretation should be attempted. It is more often the case, however, that the map does not fully agree with all the physical constraints. In this case there is opportunity to improve the agreement by the DM technique.

The major features of the DM procedure were originally proposed for both small and large molecules by Hoppe and coworkers (30–32), under the name "phase correction." This approach was initially used for small molecules (24, 25), and was first applied to macromolecular MIR phases by Barrett & Zwick (4). The procedures are outlined in the following diagram:
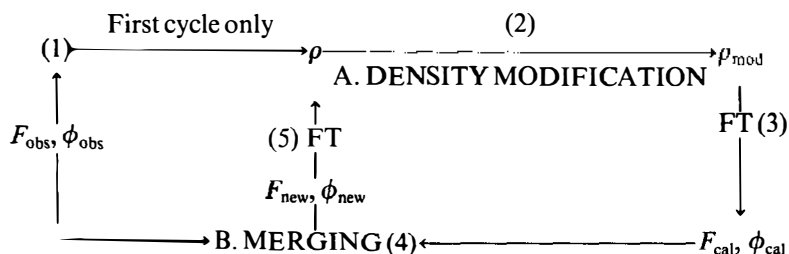


The steps are:

1. The electron density map is calculated from the experimental data $F_{obs} \exp(i\phi)$.
2. A modified electron density is then obtained by the use of some known physical constraint, e.g. $\rho_{mod} = \text{maximum}(0, \rho)$ for negative density truncation, where $\rho$ is the electron density value at a specific grid point.

3. Structure factors $[F_{cal} \exp(i\phi_{cal})]$ are calculated from $\rho_{mod}$.
4. The calculated structure factors are merged with the experimental structure factors to produce a new set of structure factors, $F_{new} \exp(i\phi_{new})$.
5. A density map is calculated from these merged structure factors. The map is either interpreted or used as input to step 2.

The method thus consists of two substantive steps, density modification and structure-factor merging, and two Fourier transform steps (forward and inverse) per cycle. The phases are judged convergent if the phase change (before and after modification, before and after merging, between $\phi_{cal}$ of consecutive cycles, or between $\phi_{new}$ of consecutive cycles) is less than an arbitrary threshold. Such phase convergence only means that the new phase set has incorporated the DM restraints, and does not imply that the set is correct. Convergence can also be based on the agreement between $F_{obs}$ and $F_{cal}$. The usefulness of the results should always be judged from the quality of the resulting map.

If partial model information is available, the constraint of step 2 might include the generation of a composite map consisting of electron densities from both the partial model and the yet uninterpreted portions of the map (8, 9, 41).

The availability of fast Fourier transforms (3a, 4, 64) has greatly reduced the necessary computing time for steps 3 and 5. These steps may be further speeded up by the use of superfast Fourier transforms (3).

# DENSITY MODIFICATION

Density maps are modified by use of one or more of the following physical constraints. In general, the phasing power of a constraint increases with the number of density points it affects and the magnitude of the change it imposes on the density value. The following density modification techniques obviously do not exhaust all possible physical constraints that can be applied to the density. As new structures in different resolution ranges are studied, new methods of modification will be developed, especially since the easy availability of computing power has facilitated the widespread application of these techniques.

## Positivity

Electron density maps are always positive except in the presence of series termination errors. Positivity is thus a useful phasing constraint when the starting map has deep negative holes (such as heavy-atom "ghosts") and/or extended negative regions due to phase errors. To impose positivity, negative density is set to zero or multiplied by an attenuation factor, e.g.

0.1. Sharp boundaries are undersirable and have been avoided in the suppression of negative regions through appropriate shaping of the density by power series modification (21) and by attenuation (61). Proper implementation of this constraint requires a good estimate of the value of $F_{000}$, which can be obtained by adjusting the observed solvent level to a physically meaningful value. The attenuation of negative regions may cause the appearance of excessively high positive peaks; this problem is solved by positive density truncation (see section on boundedness, below). It must be pointed out that since many neutron scattering amplitudes (notably that of deuterium) are negative, the positivity constraint does not apply to neutron diffraction studies. Also, electron density maps at low resolution may have negative regions even with correct phases because of series termination errors. Therefore, the use of this constraint is limited at low resolution. Phase extension by this means was first proposed by Kartha (36).

## Atomicity

This constraint can be applied when very high-resolution data are available and series termination errors are negligible. Sayre (60) observed that at atomic resolution and for equal Gaussian atoms an atomic density is roughly proportional to its square, i.e. $\rho_{app} = K\rho^2$. In reciprocal space the equation $F = s(F^*F)$ holds exactly, where the asterisk indicates a convolution product and $s$ is a resolution-dependent shape factor. The close relation between this formula and the widely applied tangent formula is discussed by Barrett & Zwick, who implemented this convolution by rapid-density squaring in real space (4, 73).

Moreover, in this DM approach negative density can be easily removed before squaring; this avoids the squaring of negative density, which is implicit in the application of convolutional equations. Podjarny & Yonath (52) have shown that removal of negative density is necessary for the success of reciprocal-space matrical implementation of the squaring relation.

Squaring, however, sometimes produces increasingly large densities. To solve this problem, Collins et al (21) and Hoppe & Gassman (31) proposed the "3 2" rule, in which a density function normalized to a maximum of 1 is modified first by the imposition of positivity and then by $\rho_{new} = 3\rho^2 - 2\rho^3$. This modification has been successful for improving maps with resolution better than 2.5 Å (see section on results). Electron-density values of heavy atoms should be excluded from this modification (21, 53). The atomicity constraint can in principle also be adapted to medium- or low-resolution data by defining groups of atoms that diffract as single units; Podjarny & Yonath (52) described a reciprocal-space example.

## Boundedness

The electron density has a true maximum value related to the chemical composition of the structure. This maximum value can be calculated once the density is expressed on an absolute scale. Densities above the maximum value can be truncated to this value (32). In medium- and low-resolution structures the application of positivity and solvent-flatness constraints may lead to an increase of the high-density values. To avoid this problem, Cannillo et al (17) have devised a density modification technique in which the density is allowed to have only two values: a maximum value, which corresponds to the protein region, and a minimum value, which corresponds to the solvent. This technique has been successful in tests of phase extension from 4.0  1.8 Å in the known structure of myoglobin.

Bentley et al (5) encountered a similar problem while improving a 25-Å neutron diffraction map of the nucleosome histone core. The original map phased by the solvent contrast-variation technique had extensive negative regions. Attenuation of these regions by a multiplicative factor of 0.1 improved the map but accentuated peaks of positive density beyond reasonable limits. Therefore, another constraint was added, by which density values above $\rho_{threshold}$ were modified according to the equation $\rho_{mod} = \rho_{threshold} + 0.1(\rho - \rho_{threshold})$.

The optimal value of $\rho_{threshold}$ was found by trial and error to be 0.7 of the maximum density of the original map. Although this modification affected only 1.5% of all points in the unit cell, it prevented the excessive growth of positive peaks during density-modification cycling.

## Solvent Flatness

The existence of flat solvent regions in a crystal structure places strong constraints on the structure factor phases (11, 27, 69). Hendrickson & Lattman (28) showed that the smaller the molecular volume, the larger the number of structure factors that are algebraically related, i.e. the constraints are more powerful if a larger portion of the crystal volume is occupied by solvent.

In macromolecular crystals, between 30 and 85% of the crystal volume consists of solvent. Solvent flatness is implemented for density modification by identifying the molecular boundaries and replacing the density in the solvent region by its mean value. The following procedures have been used for boundary definition (listed roughly in chronological order).

HAND DIGITALIZATION    A minimap can be hand digitalized with the aid of a graphic tablet (27, 61). This procedure is labor intensive, prone to subjective judgmental errors, and difficult to repeat, but early applications

demonstrated the potential of the method and stimulated further developments. Figure 1 shows one of those applications to a medium-resolution case. A series of methods for automatic envelope definition followed.

LINKED HIGH DENSITY APPROACH    On the assumption that the densities are higher in the protein region than in the solvent region, the molecular envelope may be defined as a region of linked high density points; the solvent region is the complementary volume. This method allowed for the first time the automatic and fast calculation of a molecular envelope (8). The method is also applicable for implementing the constraint of connectivity within the molecular volume (see below).

CONVOLUTION TECHNIQUES    A series of methods were then developed that essentially involve a convolution of a modified map with a predefined sampling volume. Every grid point is replaced by a suitable integral over its neighborhood, and the map is thus effectively blurred.

*High mean density approach*    Wang (68) defined the molecular envelope by eliminating negative densities and then obtaining for each point the weighted integral of the density values within a large radius (about 10 Å) of the point. The weights decline linearly from 1.0 at the center to 0 at the edge of the sphere. This procedure is also automatic, but is quite computer intensive. It is nearly equivalent to the simple lowering of the resolution to see the molecular region. For cases of very low resolution [25 Å, solution of aspartyl tRNA-synthetase complex (49)] and medium resolution [4.0 Å, solution of tRNA fmet structure (50)] it has been observed that structure factors of a resolution three to five times less than the measured limit are important for proper definition of the molecular mask. In these cases, the availability of the very-low-resolution amplitudes was a limiting factor. However, B. C. Wang (personal communication) has indicated that his algorithms have been used even without low-resolution data.

*High mean absolute value approach*    Reynolds et al (55) defined molecular volume based on the expectation that densities inside this volume have greater excursions, both positive and negative, from the mean value than densities in the solvent region. An integral of the absolute value of $\rho$ around a given point is therefore an indicator of whether the point belongs to the molecular volume.

The above convolutions can be performed much faster by reciprocal-space multiplication. Several different weighting schemes, including the use of simple Gaussian weights (equivalent to the imposition of a temperature factor) have been tried (38, 51; E. Westbrook, personal communication).

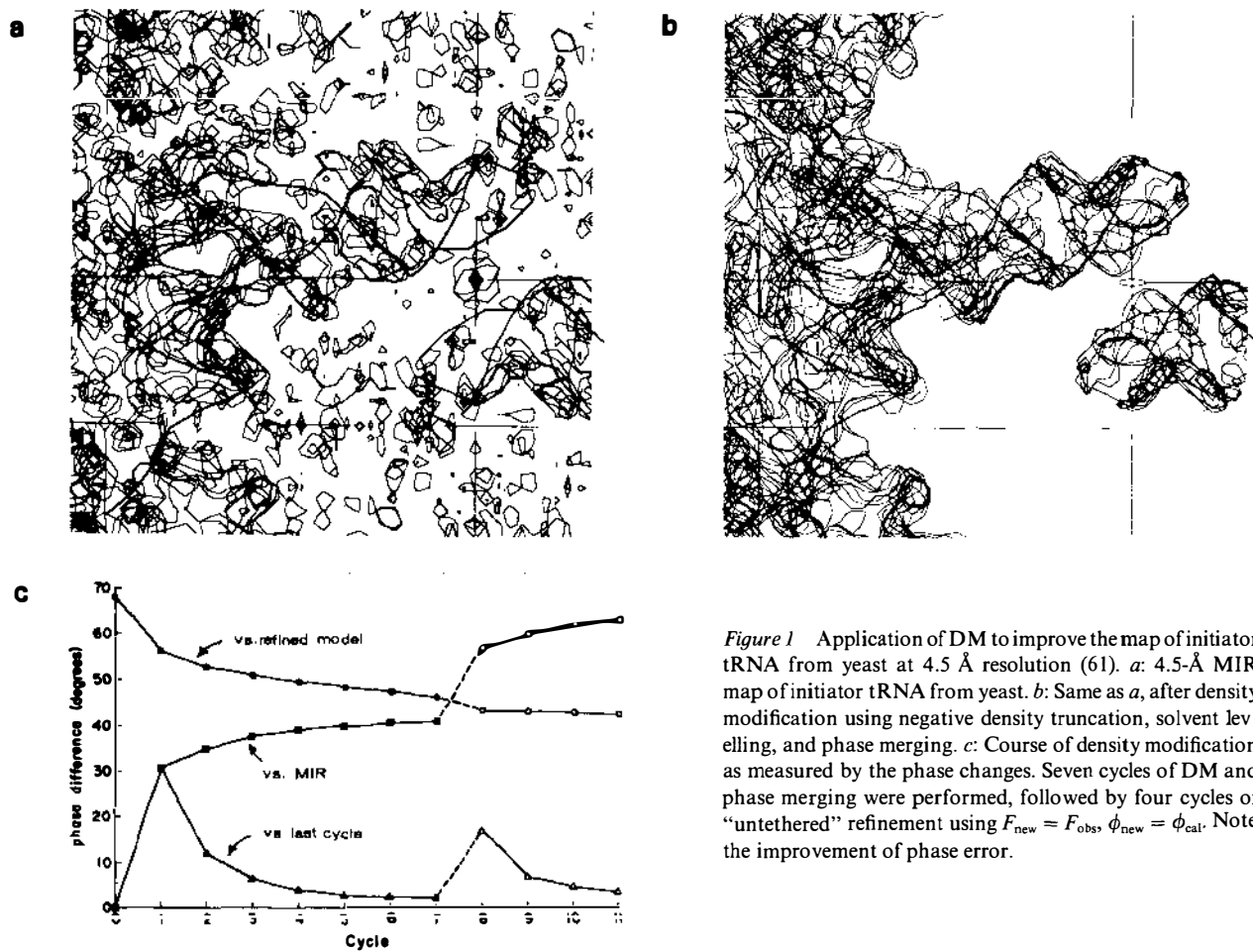Brown (14) observed that the automatic boundary obtained by con-

**a**



**b**



**c**



*Figure 1* Application of DM to improve the map of initiator tRNA from yeast at 4.5 Å resolution (61). *a*: 4.5-Å MIR map of initiator tRNA from yeast. *b*: Same as *a*, after density modification using negative density truncation, solvent levelling, and phase merging. *c*: Course of density modification as measured by the phase changes. Seven cycles of DM and phase merging were performed, followed by four cycles of "untethered" refinement using $F_{new} = F_{obs}$, $\phi_{new} = \phi_{cal}$. Note the improvement of phase error.

volution methods lacked high-resolution detail, e.g. the boundary cut off protruding loops and did not exclude buried parts of the solvent region. B. C. Wang (personal communication) has suggested that a possible solution for this problem may be to diminish the radius of integration as new masks are computed. The method of Bhat & Blow (8) is not subject to this deficiency, since the sampling volume checked for high linked densities is allowed to have any shape and can therefore follow the high-resolution boundary contour. The integral about each grid point is obtained with a smaller radius to establish the connectivity of the electron density values. If the connected grid points are numerous enough they become part of the molecular volume. The integral of the density value over this volume is evidently high.

LOW-RESOLUTION MODELING APPROACH     The molecular boundary can be obtained from a low-resolution model consisting of Gaussian spheres positioned by low-resolution translation searches. The combination of initial phase determination by low-resolution pseudo-atom searches and phase refinement by density modification provides a possible method for ab initio phasing (using a DM package developed by B. Rees) without MIR information (49). Figure 2 shows the improvement of a low-resolution map using this technique.
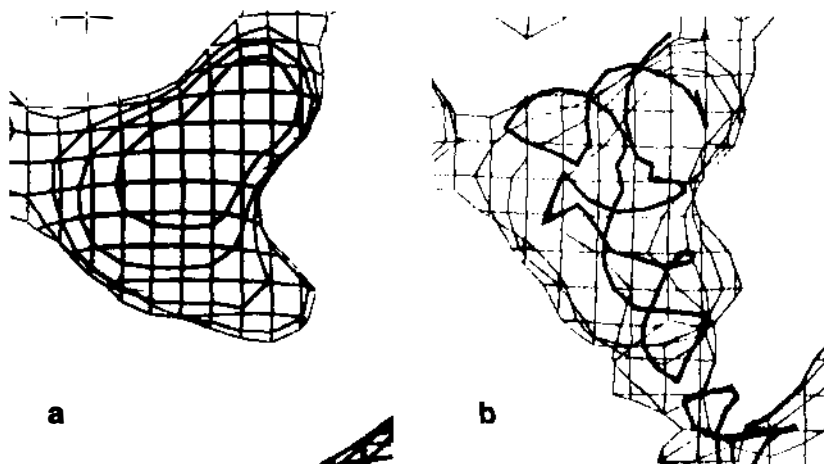


*Figure 2* Result of the application of density modification to a neutron scattering map at 25 Å resolution of the complex of Asp tRNA synthetase from yeast (49). *a*: Part of the map before density modification. Phases were obtained from a model consisting of four Gaussian spheres. *b*: Same part of the map after density modification, with the backbone of a tRNA molecule superimposed on the map. Note the shape difference.

The evolution of solvent-flattening techniques from visual identification to automatic definition of envelopes has been a major focus in the current development of density modification algorithms. Density envelopes have also been used for least square refinement (9). In another approach (1, 41), dummy atoms were inserted and refined inside a molecular region that was not interpreted in terms of a detailed model.

## Map Continuity and Use of a Partial Model

An important property of any density map corresponding to a biological macromolecule is that at medium resolution the density displays single-chain connectivity. Such connectivity corresponds to the single-chain stereochemistry of proteins and nucleic acids. This feature is essential for the interpretation of a map that does not display atomic resolution. Often the MIR map shows long connected chains easily interpretable as a partial molecular model, but the continuity breaks and side chains are not evident; therefore the sequence cannot be aligned with an interpretation.

To make full use of the available restraints from the partial model, Bhat & Blow (8) obtain a molecular mask as described above for the linked high density approach to boundary definition for modification by solvent flatness. They add to this mask high electron density regions of the partial model volume. The method was used to include both the missing atoms in the ordered domain and the density corresponding to the disordered domain in tyrosyl tRNA synthetase, and a more readily interpretable map was obtained (Figure 3).

## Noncrystallographic Symmetry

When noncrystallographic symmetry is present it is a very powerful constraint that relates different density points within a map. In some cases (e.g. virus structures) the positioning of the noncrystallographic axis is very accurate and the redundancy of the data is very high; thus a successful phase extension and refinement is practically assured. Phasing relations arising from this constraint can be formulated either in reciprocal space (59) or in real space (2, 11, 16), where the method follows the overall scheme of density modification. This constraint has been extensively applied and merits a review of its own. Therefore we only cite here some recent examples.

Virus crystals are typically amenable to application of this constraint. In particular, rhinovirus (57), poliovirus (29), and mengo virus (M. Luo, U. Boege, G. Vriend, G. Kamer, I. Minor, E. Arnold, M. G. Rossmann & D. G. Scraba, manuscript in preparation) are the latest structures for
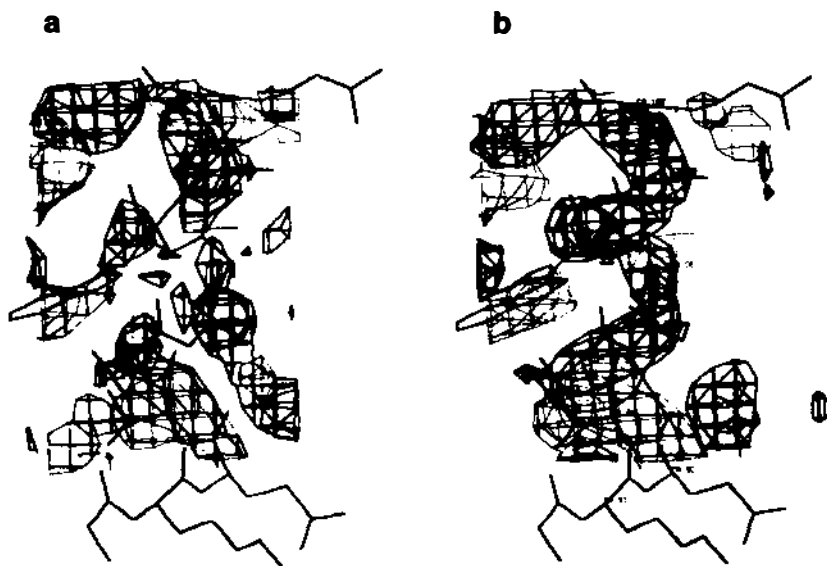
*Figure 3* Application of density modification using map continuity and a partial model. The figure shows a comparison of the electron density in a helical segment of the tyrosyl tRNA synthetase molecule (8) as found in the MIR map (*a*), and using the phases obtained from density modification (*b*).

which this technique has been used for phase extension even without MIR phase information. Mengo virus is a most interesting case; Luo et al have been able to extend phase information from 8 to 3 Å using only the 60-fold noncrystallographic symmetry. They obtained the starting point from molecular replacement by positioning the molecule of rhinovirus in the unit cell of mengo virus and measuring the agreement with the observed data. This agreement was 52%, which should be compared with a random value of 56%. Even with this poor starting point, the final map at 3 Å was extremely clear. This is the only case in which useful phase extension has been reported in this crucial resolution range.

## Multiple Crystal Forms

The structure factors of multiple crystal forms of the same molecule are related by equations similar to those of molecular replacement (44). Isomorphous replacement phases can therefore be improved by the cyclic averaging of the electron densities of two crystal forms (e.g. 40, 66). The

procedure is similar to noncrystallographic symmetry averaging, with the particularity that the two maps to be averaged come from different crystals.

# MERGING OF CALCULATED AND OBSERVED AMPLITUDES AND PHASES

After the modification step, an inverse Fourier transform generates calculated structure factor amplitudes and phases from the modified map. This information is now combined with the observed amplitude and original phase to obtain new values and to synthesize a new improved map. Since the error in phase is normally much larger than the error in amplitude, the observed amplitudes can be considered correct.

## Replacing the Phases

The first merging procedure simply merged the observed amplitude and the calculated phase (4, 31). This simple approach ignored the experimental and calculated phase probability distributions as well as the calculated amplitudes, and produced maps that were biased toward the calculated phases.

## Merging the Amplitudes and Replacing the Phases

To diminish the bias of simple phase replacement the coefficients $2F_{obs} - F_{cal}$ and $\phi_{cal}$ were used (21, 53, 74). Main (43) analyzed in detail the characteristics of the resulting map and showed that this synthesis was superior to other alternatives.

Read (54) has suggested the use of the expression $2mF_{obs} - DF_{par}$ $\exp(i\phi_{cal})$, where $F_{obs}$ is the native amplitude and $F_{par}\exp(i\phi_{cal})$ is the calculated structure factor from a partial model with errors. The term $D$ is a function of the coordinate error, and tends to 1 when the coordinate error tends to zero. The term $m$ is a figure of merit dependent on the agreement of $F_{obs}$ and $F_{par}$.

A conservative variation of this general $2F_{obs} - F_{cal}$ approach is to choose the maximum of $2F_{obs} - F_{cal}$ and $KF_{obs}$; positive $K$ prevents large $F_{cal}$ terms from causing 180° changes in phase (74).

In this approach still only the centroid and figure of merit of the original phase distribution are utilized, i.e. full use is not made of the experimental phase probability distribution. This is especially relevant to cases in which the probability distribution is multimodal and is thus not derived simply from the centroid and figure of merit, e.g. the single isomorphous replacement (SIR) bimodal phase probability distributions (68).

## Merging the Phases

To overcome the problems of multimodal phase probability distributions and excessive bias toward the calculated phase, Hendrickson et al (27) represented each phase by a probability distribution rather than a specific value and combined the phases by multiplying the experimental and calculated phase probability distributions. The combined phase, $\phi_{mer}$, and its figure of merit are obtained from the product distribution and are used for calculating the new electron density map. To obtain the distribution of the calculated phase, Hendrickson and coworkers (26, 27) followed the formula suggested by Sim (62). This is an extension to density modification of the method used by Rossmann & Blow (58) to combine the phases from MIR with that of a partial model. A similar approach has been used by Bricogne (12).

The combination of the phase distributions was reduced to the addition of four Fourier coefficients (28) by representing the phase distribution curve as $\text{Prob}\,\phi = K\,\exp(A\,\cos\,\phi + B\,\sin\,\phi + C\,\cos\,2\phi + D\,\sin\,2\phi)$, in which only these coefficients are saved.

This approach has been used successfully in various density-modification phase-improvement and extension schemes (e.g. 8, 37, 61, 68, 71). It has the advantage of using all the available information, including the original MIR phase probability distribution and the calculated amplitudes.

## Merging Both the Amplitudes and the Phases

Podjarny et al (49) used both amplitude and phase merging by calculating the new Fourier coefficients as $(2F_{obs} - F_{cal})\exp(i\phi_{mer})$; this leads to better results than merging only the phases (see sections on low-resolution modeling approach, above, and results, below).

Utilization of this type of coefficient was previously reported in the context of MIR and model phase calculation. While solving the structure of phosphoglycerate kinase, Rice (56) compared different ways of combining experimental and model structure factors, and suggested the use of the $2F_{obs} - F_{cal}$, $\phi_{mer}$ coefficient. Stuart & Artymiuk (63) further modified the coefficient as $F_{new} = \text{fmerit}[F_{obs} + Q_{mer}(F_{obs} - F_{cal})]$; $\phi_{new} = \phi_{mer}$, where $Q_{mer}$ is a function of the MIR figure of merit, calculated figure of merit, and number of correctly placed atoms. $F_{new}$ tends to $2F_{obs} - F_{cal}$ when most of the atoms are known and the calculated figure of merit is close to 1. For other cases it tends to produce a more accurate map.

A new map is computed using the merged Fourier coefficients and the entire procedure is iterated, starting from step 2, to obtain convergence. Optimally, the final map agrees more fully with the physical constraints because of the modification step, and therefore the phase error is reduced.

The merging step is needed to restore, at least partially, the experimental information; this step thus reduces the bias introduced by the modification.

## SOME NEW DEVELOPMENTS

### Combined Omit Map Technique

The map obtained through density modification is normally an improvement over the original map calculated from experimental phase data. However, to the degree that the original phases arc in error, noise will persist in the modified map. For ab initio phase determination, when initial phases do not exist, this problem becomes crucial; therefore the application of DM has been limited to cases for which reasonable initial phases have been established by other methods. In an attempt to use DM for ab initio phase determination, Bhat (6, 7) designed a consistent electron density (CED) technique to solve the problem of noise transfer to the modified map.

Density modification tends to produce electron density distributions consistent with real-space constraints. The CED method does the same and also achieves consistency between the electron densities in different parts of the asymmetric unit. New electron density values are calculated for each grid point using the Fourier coefficients from a density function in which the grid point and surrounding grid points are set to solvent level. The objective is to obtain an electron density distribution not biased by the original distribution at or around that grid point. Bhat & Cohen (10) adopted a related approach to detect errors in an atomic model at the final stages of refinement.

The CED method produces an electron density map consistent both within itself and with the external restraints (hence its name). The procedure can be initiated from an arbitrary set of reflections, selected either by ranges of resolution and amplitude or just as origin- and enantiomorph-defining reflections. The procedure involves a series of phase extensions to include additional reflections.

The CED method has been applied to bovine heart creatine kinase (7); an ab initio phase set obtained at 4 Å resolution was accurate enough to phase a heavy-atom difference map, as shown in Figure 4.

### Minimization Techniques Using Density Modification Constraints

In most DM algorithms, constraints are imposed alternatively in real and reciprocal space, and a result that agrees with constraints in both spaces is sought. However, the result may not agree simultaneously with all
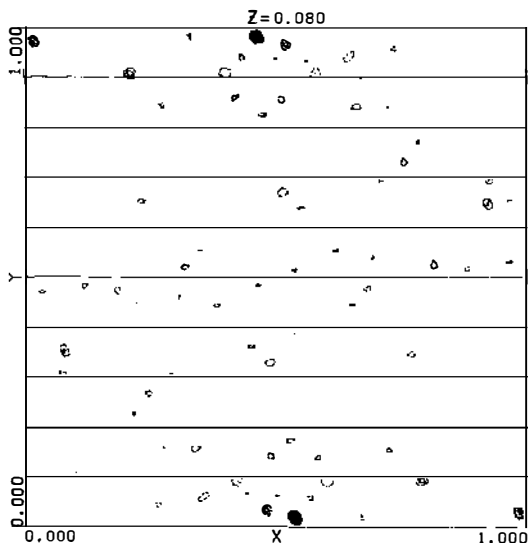
*Figure 4* Section of heavy atom difference Fourier map for bovine heart creatine kinase. Amplitudes are $F_{heavy}$   $F_{native}$ and phases were obtained from the CED method. A peak found at the correct heavy atom position is twice the next largest peak.

constraints. To solve this problem, minimization algorithms were developed. In a first approach, Navaza et al (46) proposed a variational technique using Lagrange multipliers; agreement of observed and calculated amplitudes was maximized subject to the physical constraints in the electron density. This minimization approach is more strictly defined analytically than the fixed-point iteration used in standard DM methods. However, for weak constraints even this minimization may not generate a unique solution.

When there may be multiple solutions that agree with the observed data and with the weak physical constraints, a selection criterion must be defined. This criterion is generally some function of the electron density that is optimized. A description of the several approaches to this problem, generally known as "maximum entropy methods" (e.g. 13, 15, 19a, 20, 22a, 39, 45, 72), is beyond the scope of this review. However, some of these approaches (20 22a, 45) are closely related to DM, and given their potential, we examine one as an example.

Navaza (45) has proposed the following probabilistic approach. All maps that satisfy the physical constraints define an ensemble, and probability values are assigned to each of these maps. An estimate of the true

electron density is calculated using these probabilities as weights, and its transform is required to match the experimental data. When weak constraints underdetermine the probability distribution, a maximum entropy distribution is chosen. This maximum entropy distribution makes the weights as equal as possible.

The implementation of this maximization algorithm uses Lagrange multipliers and forward and inverse Fourier transforms. The sequence of computational steps closely resembles the alternating real- and reciprocal-space structure of the usual DM technique and has the same flexibility for imposing a large variety of real-space constraints. However, it also includes as a specific constraint the agreement between observed and calculated structure factors. If the problem is underdetermined, this agreement may be substantially better than the experimental error would suggest possible, and the remaining degrees of freedom will be absorbed by the maximum entropy condition. If the problem is overdetermined, this agreement is approximate; the maximum entropy condition is then not exactly satisfied, but it influences the apportionment of the disagreement of the observed and calculated structure factors.

The maximum entropy method has been applied to very-low-resolution model calculations in macromolecules (47). For example, the macromolecular model calculations showed that the method can accurately refine and extend phases in the 30–10 Å resolution range.

## RESULTS AND CONCLUSIONS

Table 1 shows the results of several DM applications. The emphasis has shifted from high-resolution structures, where atomicity was the most important constraint, to medium- and low-resolution structures, where solvent flatness is the principal constraint. It is quite clear that DM can improve a MIR map over a wide range of resolution when a suitable protein boundary can be defined. DM appears to be the method of choice when a MIR map cannot be interpreted in terms of a molecular model but shows some recognizable features. In many cases (68) DM has also proved helpful in resolution of the phase ambiguity inherent in the SIR method. When a map can be interpreted to obtain a nearly complete model, DM can improve the map but is not necessary since conventional refinement and model-building techniques are highly effective.

Two DM applications in the medium-resolution range are shown in Figures 1 (61) and 3 (8). Here density modification enabled the researchers to successfully follow the polypeptide chain and solve the structure.

Conventional DM techniques are not geared toward overcoming the

**Table 1** Applications of density modification[a]

| Result | Constraints | Merging | Reference (year) |
|---|---|---|---|
| High resolution: | | | |
| Maximum resolution below 2.5 Å | | | |
| Improved initial map obtained from heavy atom site for many small molecules | Atomicity | $F_{obs}, \phi_{cal}$ | 31 (1968) |
| Extended phases in myoglobin from 3 to 2 Å. Mean error: 78° | Atomicity (squaring), positivity | $E_{obs}, \phi_{cal}$ | 4 (1971) |
| Extended phases in rubredoxin, 2 to 1.5 Å. Mean error: 30°. Refined phases of staphylococcus nuclease at 2 Å; map improved | Atomicity (3 2 rule), positivity, boundedness | $2F_{obs} - F_{cal}, \phi_{cal}$ | 21, 22 (1975, 1976) |
| Phase refinement from 85 to 32° error in 32 atom structure | Atomicity (3 2 rule), positivity | Maximum $(2F_{obs} - F_{cal}, KF_{obs}), \phi_{cal}$ | 74 (1976) |
| Human lysozyme phase refinement. 2.5 Å resolution. Found no improvement after single cycle without merging | Atomicity, positivity, boundedness | None | 48 (1976) |
| Extended 2.8 Å MIR phases in chymotrypsin to 1.8 Å. Map improvement | Atomicity (3 2 rule), positivity | $(2F_{obs} - F_{cal}), \phi_{cal}$ | 53 (1979) |
| Diminished error in 2.5 Å MIR phospholipase A2 phases from 55 to 52° | Solvent flatness, positivity, boundedness | $F_{obs}, \phi_{mer}$ | 37 (1981) |
| Extended 4-Å model phases (with errors) in myoglobin to 1.8 Å with an error of 22° in strongest $F$s | Solvent flatness | $F_{obs}, \phi_{cal}$ | 17 (1983) |
| Refinement of 2.2 Å SIR phases from Dead Sea ferredoxin. Increased density map detail | Solvent flatness, positivity, boundedness | $F_{obs}, \phi_{mer}$ | 51 (1984) |

**Table 1**  *continued*

| Result | Constraints | Merging | Reference (year) |
|---|---|---|---|
| Medium resolution: | | | |
| Maximum resolution between 2.5 and 8 Å | | | |
| Refinement of 5.5 Å myohemerythrin phases. Mean figure of merit improved from 0.89 to 0.92 | Solvent flatness, positivity | $F_{obs}, \phi_{mer}$ | 27 (1975) |
| Refinement of 5 Å tRNA Phe phases. Provided a correct starting phase set for matricial phase extension | Positivity | $F_{obs}, \phi_{cal}$ | 52 (1977) |
| Refinement of 4-Å initiator tRNA MIR phases. Mean phase error (as measured against final refined model) diminished from 68 to 43° | Solvent flatness, positivity | $F_{obs}, \phi_{mer}$ first, $\phi_{cal}$ later | 61 (1981) |
| Refinement of 2.7 Å Tyr tRNA synthetase MIR phases. Overall map improvement led to correct chain tracing. Automated molecular mask | Solvent flatness, positivity, continuity | $F_{obs}, \phi_{mer}$ | 8 (1982) |
| Refinement of 3-Å Bence Jones protein SIR phases. Overall phase error improved from 51 to 32°. Real space automatic masking | Solvent flatness, positivity | $F_{obs}, \phi_{mer}$ | 23 (1983) |
| 6-Å refinement of ketosteroid isomerase MIR phases. Improved protomer boundaries and map interpretability. Reciprocal space automatic masking | Solvent flatness, positivity | $F_{obs}, \phi_{mer}$ | 71 (1984) |
| 4 Å ab initio determination of creatine kinase phases. Phase set obtained accurate enough to phase correctly a heavy atom difference map. | Solvent flatness, positivity, consistent electron density | $F_{obs}, \phi_{cal}$ | 7 (1985) |

| | | | |
|---|---|---|---|
| Phase extension from 5 to 3 Å for rhinovirus. Started from MIR phases. Very accurate and interpretable map obtained | Solvent flatness, positivity, 20-fold noncrystallographic symmetry | $F_{obs}$, $\phi_{cal}$ | 57 (1985) |
| Phase extension from 5 to 2.9 Å for poliovirus phases. Started from averaged SIR phases and proceeded in shells of 0.05 Å. Very clear map obtained | Solvent flatness, positivity, 30-fold noncrystallographic symmetry | $F_{obs}$, $\phi_{cal}$ | 29 (1985) |
| Phase extension between 4.0 and 3.2 Å resolution for *Panulirus interruptus* hemocyanin | 6-fold noncrystallographic symmetry | $F_{obs}$, $\phi_{cal}$ | 67 (1986) |
| Phase extension from 8 to 3 Å for mengovirus phases. Started from an appropriate molecular replacement solution. Very accurate and interpretable map obtained | Solvent flatness, positivity, 30 fold noncrystallographic symmetry | $F_{obs}$, $\phi_{cal}$ | M. Luo, U. Boege, G. Vriend, G. Kamer, I. Minor, E. Arnold, M. G. Rossmann, D. G. Scraba, manuscript in preparation |
| Phase refinement at 3-Å resolution for the data of rubisco from tobacco. Started with MIR phases with a 0.39 figure of merit. Changed phases by 39° and increased the figure of merit to 0.75. 80% of the backbone can be interpreted from the resulting map. Combination of manual and automatic masks. Reciprocal space mask calculation using Butterworth low-pass filter | Solvent flatness, positivity | $F_{obs}$, $\phi_{mer}$ | 19 (1986) |

**Table 1** *continued*

| Result | Constraints | Merging | Reference (year) |
|---|---|---|---|
| Low resolution: Maximum resolution above 8 Å | | | |
| 16-Å refinement of neutron diffraction phases of the histone core of the nucleosome particle, originating from the solvent contrast variation technique and the DNA model. Increased definition of histone dimer boundaries | Solvent flatness, positivity | $F_{obs}$, $\phi_{mer}$ | 5 (1984) |
| 25 Å refinement of neutron diffraction phases of the Asp tRNA synthetase complex, originated from a model of 4 gaussian spheres. Increased interpretability and allowed recognition of tRNA shape. Phase change: 32° | Solvent flatness, positivity, boundedness | $F_{obs}$, $\phi_{cal}$ | 49 (1986) |
| 10 Å refinement of X ray diffraction phases of the Asp tRNA synthetase complex, originated from a low resolution model including the tRNA phosphate backbone. Increased continuity in synthetase region and tRNA defined more clearly | Solvent flatness, positivity, boundedness, noncrystallographic symmetry | $2F_{obs} - F_{cal}$, $\phi_{mer}$ (using Sim distribution for model $\phi$ | 49 (1986) |
| Phase refinement at 20-Å resolution and extension to 15 Å. Starting phases with error calculated from a tRNA synthetase complex model. Improved phase error from 29 to 17° at 20 Å and extended phases to 15 Å with an error of 60° | Positivity, boundedness, maximum entropy | R factor minimization | 47 (1986) |

[a] This list is not exhaustive, but is compiled to illustrate the application of the method. For details consult the original papers; note that "error" may have different meanings in the different papers cited. Applications of the DM method have also recently been reviewed by Tulinsky (65) and Wang (68).

noise introduced by initial phases; therefore their application has been limited to cases in which tentative phases have been determined, most commonly with MIR methods. However, new developments suggest the possibility of ab initio solutions to the phase problem. The work of Bhat (6, 7) shows that the incorporation into the DM algorithm of the composite omit map step is powerful enough to force the phases toward the correct solution even from a random set of initial phases. The work of Podjarny et al (49) shows that phases obtained from very-low-resolution modeling (25 Å) may be refined by DM. This work has led to a reasonable map for the Asp tRNA–Asp tRNA synthetase complex at 10-Å resolution. In this case, previous knowledge of the overall structure of the tRNA component was very helpful for map interpretation at low resolution. The work of Luo et al (manuscript in preparation) opens up the possibility of using DM for further phase extension at a crucial resolution range, namely 8 3 Å. These developments point toward the possibility that MIR will eventually be replaced as the main phasing method for difficult cases in which it is not easily applicable.

## ACKNOWLEDGMENTS

*Literature Cited*

1. Agarwal, R. C., Isaacs, N. W. 1977. *Proc. Natl. Acad. Sci. USA* 74: 2835 39
2. Argos, P., Ford, G. C., Rossmann, M. G. 1974. *Acta Crystallogr.* A31: 499 506
3. Auslander, L., Feig, E., Winograd, S. 1981. In *Computational Crystallography*, ed. D. Sayre, pp. 451–69. Ottawa: Cla rendon
3a. Bantz, D., Zwick, M. 1974. *Acta Crys tallogr.* A30: 257 60
4. Barrett, A. N., Zwick, M. 1971. *Acta Crystallogr.* A27: 6 11
5. Bentley, G. A., Lewit Bentley, A., Finch, J. T., Podjarny, A. D., Roth, M. 1984. *J. Mol. Biol.* 176: 55 75
6. Bhat, T. N. 1984. *Acta Crystallogr.* A40s: C-15
7. Bhat, T. N. 1985. Presented at Ann. Meet. Am. Crystallogr. Assoc., Stan ford, Calif.
8. Bhat, T. N., Blow, D. M. 1982. *Acta Crystallogr.* A38: 21 29
9. Bhat, T. N., Blow, D. M. 1983. *Acta Crystallogr.* A39: 166 70
10. Bhat, T. N., Cohen, G. H. 1984. *J. Appl. Crystallogr.* 17: 244 -48
11. Bricogne, G. 1974. *Acta Crystallogr.* A30: 395 405
12. Bricogne, G. 1976. *Acta Crystallogr.* A32: 832 47
13. Bricogne, G. 1984. *Acta Crystallogr.* A40: 410 45

14. Brown J. H. 1985. MS thesis. Weizmann Inst. Sci., Rehovot, Israel
15. Bryan, R. K., Bansal, M., Folkhard, W., Nave, C., Marvin, D. A. 1983. *Proc. Natl. Acad. Sci. USA* 80: 4728 31
16. Buehner, M., Ford, G. C., Olsen, K. W., Moras, D., Rossmann, M. G. 1974. *J. Mol. Biol.* 901: 25 49
17. Cannillo, E., Oberti, R., Ungaretti, L. 1983. *Acta Crystallogr.* A39: 68 74
18. Deleted in proof
19. Chapman, M. S., Smith, W. W., Suh, S. W., Cascio, D., Howard, A., et al. 1986. *Philos. Trans. R. Soc. London* In press
19a. Collins, D. M. 1978. *Acta Crystallogr.* A34: 533 41
20. Collins, D. M. 1982. *Nature* 298: 49 51
21. Collins, D. M., Brice, M. D., La Cour, T. F. M., Legg, M. J. 1976. In *Crystallographic Computing Techniques*, ed. F. R. Ahmed, K. Huml, B. Sedlacek, pp. 330 35. Copenhagen: Munksgaard
22. Collins, D. M., Cotton, F. A., Hazen, E. E. Jr., Meyer, E. F., Morimoto, C. N. 1975. *Science* 190: 1047 53
22a. Collins, D. M., Mahar, M. C. 1983. *Acta Crystallogr.* A39: 777–83
23. Furey, W. Jr., Wang, B. C., Yoo, C. S., Sax, M. 1983. *J. Mol. Biol.* 1673: 661 92
24. Gassmann, J. 1976. See Ref. 21, pp. 144 54
25. Gassmann, J., Zeckmeister, K. 1972. *Acta Crystallogr.* A28: 270 80
26. Hendrickson, W. A. 1981. In *Structural Aspects of Biomolecules*, ed. R. Srinivasan, V. Pattabhi, pp. 31 80. New Delhi: Macmillan
27. Hendrickson, W. A., Klippenstein, G. L., Ward, K. B. 1975. *Proc. Natl. Acad. Sci. USA* 72: 2160–64
28. Hendrickson, W. A., Lattman, E. E. 1970. *Acta Crystallogr.* B26: 136-43
29. Hogle, J. M., Chow, M., Filman, D. J. 1985. *Science* 229: 1358-65
30. Hoppe W., Gassmann, J. 1964. *Ber. Bunsenges. Phys. Chem.* 68: 808 10
31. Hoppe, W., Gassman, J. 1968. *Acta Crystallogr.* B24: 97 107
32. Hoppe, W., Gassmann, J., Zechmeister, K. 1970. See Ref. 21, pp. 26 36
33. Deleted in proof
34. Karle, J., Hauptman, H. 1950. *Acta Crystallogr.* 3: 181
35. Karle, J., Hauptman, H. 1956. *Acta Crystallogr.* 9: 635
36. Kartha, G. 1969. *Acta Crystallogr.* A25: S87
37. Keith, C., Feldman, D., Deganello, S., Glick, J., Ward, K., et al. 1981. *J. Biol. Chem.* 256: 8602 7
38. Leslie, A. G. W. 1987. *Acta Crystallogr.*

A43: In press
39. Livesey, A. K., Skilling, J. 1985. *Acta Crystallogr.* A41: 113–22
40. Loebermann, H., Tokuoka, R., Deisenhofer, J., Huber, R. 1984. *J. Mol. Biol.* 177: 531 56
41. Lunin, V. Yu., Urzhumtsev, A. G., Vernoslova, E. A., Chirgadze, Yu. N., Nevskaya, N. A., Fomenkova, N. P. 1985. *Acta Crystallogr.* A41: 166 71
42. Deleted in proof
43. Main, P. 1979. *Acta Crystallogr.* A35: 779–85
44. Main, P., Rossmann, M. G. 1966. *Acta Crystallogr.* 21: 167
45. Navaza, J. 1985. *Acta Crystallogr.* A41: 232 44
46. Navaza, J., Castellano, E. E , Tsoucaris, G. 1983. *Acta Crystallogr.* A39: 622 31
47. Navaza, J., Podjarny, A. D., Moras, D. 1986. Presented at Ann. Meet. Am. Crystallogr. Assoc., Hamilton, Ontario
48. Nixon, P. E., North, A. C. T. 1976. *Acta Crystallogr.* A32: 325
49. Podjarny, A. D., Rees, B., Thierry, J. C., Cavarelli, J., Roth, M., et al. 1986. Presented at Ann. Meet. Am. Crystallogr. Assoc., Hamilton, Ontario
50. Podjarny, A. D., Schevitz, R. W., Sigler, P. B. 1981. *Acta Crystallogr.* A37: 662 68
51. Podjarny, A. D., Sussman, J. L., Bhat, T. N., Westbrook, E. M., Harel, M., et al. 1984. *Acta Crystallogr.* A405: C 14
52. Podjarny, A. D., Yonath, A. 1977. *Acta Crystallogr.* A33: 655 61
53. Raghavan, N. V., Tulinsky, A. 1979. *Acta Crystallogr.* B35: 1776-85
54. Read, R. J. 1986. *Acta Crystallogr.* A42: 140–49
55. Reynolds, R. A., Remington, S. J., Weaver, L. H., Fisher, R. G., Anderson, W. F., et al. 1985. *Acta Crystallogr.* B41: 139–47
56. Rice, D. W 1981 *Acta Crystallogr.* A37: 491–500
57. Rossmann, M. G., Arnold, E., Erickson, J. W., Frankenberg, E. A., Griffith, J. P., et al. 1985. *Nature* 317(6033): 145 53
58. Rossmann, M. G., Blow, D. M. 1961. *Acta Crystallogr.* 14: 641 47
59. Rossmann, M. G., Blow, D. M. 1962. *Acta crystallogr.* 15: 24 31
60. Sayre, D. 1952. *Acta Crystallogr.* 5: 60 65
61. Schevitz, R. W., Podjarny, A. D., Zwick, M., Hughes, J. J., Sigler, P. B. 1981. *Acta Crystallogr.* A37: 669 77
62. Sim, G. A. 1959. *Acta Crystallogr.* 12: 813 15

63. Stuart, D., Artymiuk, P. 1985. *Acta Crystallogr.* A40: 713 16
64. Ten Eyck, L. F. 1973. *Acta Crystallogr.* A29: 183–91
65. Tulinsky, A. 1985. *Methods Enzymol.* 115: 77–89
66. Varghese, J. N., Laver, W. G., Colman, P. M. 1983. *Nature* 303: 35 44
67. Volbeda, A., Gaykema, W. P. J., Hol, W. G. J. 1984. *Acta Crystallogr.* A40s: C 13
68. Wang, B. C. 1985. *Methods Enzymol.* 115: 90–112
69. Ward, K. B., Hendrickson, W. A., Klippenstein, G. L. 1975. *Nature* 257: 818
70. Deleted in proof
71. Westbrook, E., Piro, O., Sigler, P. B. 1984. *J. Biol. Chem.* 259: 9096 103
72. Wilkins, S. W., Varghese, J. N., Lehmann, M. S. 1983. *Acta Crystallogr.* A39: 47–60
73. Zwick, M. 1968. PhD thesis. MIT, Cambridge, Mass.
74. Zwick, M., Bantz, D., Hughes, J. 1976. *Ultramicroscopy* 1: 275 77

# CONTENTS

viii    CONTENTS (*continued*)