Acta Cryst. (1997). A**53**, 291–313

Holographic Methods in X-ray Crystallography. V. Multiple Isomorphous Replacement, Multiple Anomalous Dispersion and Non-crystallographic Symmetry†

ABRAHAM SZÖKE,[a]* HANNA SZÖKE[a] AND JOHN R. SOMOZA[b]

[a]*Lawrence Livermore National Laboratory, Livermore, CA 94550, USA, and* [b]*Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, CA 94143, USA. E-mail: szoke1@llnl.gov*

(Received 16 October 1996; accepted 23 December 1996)

Abstract

The holographic method for the recovery of the electron density of macromolecules is based on the expansion of the electron density into Gaussian basis functions. The technique makes consistent use of real- and reciprocal-space information to produce electron-density maps. It enforces positivity of the recovered electron density and makes effective use of previously known information about the electron density, such as knowledge of a solvent region or knowledge of a partial structure. In this paper, we summarize the theory underlying the holographic method, and describe how we extend the range of information that can be used by the method to include information from multiple-isomorphous-replacement (MIR) data, multiple-anomalous-dispersion (MAD) data and knowledge of non-crystallographic symmetry. The convergence properties and the limiting accuracy of the method are discussed. Its power for synthetic problems is demonstrated and the method is applied to experimentally measured MIR data from kinesin, a motor protein domain that has recently been solved. Appendix *A* gives a detailed description of the algorithms and the equations used in *EDEN*, the computer program that implements the holographic method.

1. Introduction

In previous publications of this series [Szöke (1993, hereafter paper II); Maalouf, Hoch, Stern, Szöke & Szöke (1993, paper III); Somoza *et al.* (1995, hereafter paper IV)], we began to examine X-ray crystallographic computations of macromolecules from a long neglected point of view (Bragg, 1950; Tollin, Main, Rossmann, Stroke & Restrick, 1966). We called the results of these considerations the holographic method. Related ideas were explored in the work of Béran (Béran & Szöke, 1995).

The most important limitation on the power of X-ray crystallography is that, as a consequence of Bragg's law, the electron density of a crystal cannot be fully recovered from its diffraction pattern alone. This is the well known phase problem. Many years ago, the eminent mathematician Lánczos (1961) pointed out that no mathematical trickery can remedy lack of information. The holographic method does not attempt to circumvent Lánczos's dictum. Our principal claim is that the holographic method is a simple and effective way of using all available information simultaneously, consistently and sometimes optimally. In our previous papers, two kinds of information were utilized: the positivity of the electron density and the location of contiguous regions of disordered solvent. In the present paper, more kinds of information are added: the presence of isomorphous derivatives, anomalous dispersion and non-crystallographic symmetry.

In paper II, a somewhat abstract analysis of the mathematical structure of the holographic method was given. Our discussion started from the analogy of an X-ray diffraction pattern and a hologram. We assumed, for example, that in part of the unit cell of a crystal the electron density is known. This is the situation in molecular replacement and also during the solution of crystal structures. The complex amplitude of the wave diffracted from the known part can then be calculated and identified as a holographic reference wave. Similarly, the wave diffracted from the unknown part of the unit cell is analogous to an object wave in holography. The pattern of intensities observed in X-ray diffraction from a crystal, when reduced to the absolute squares of the structure factors, is then analogous to a recorded hologram. It contains the sum of the intensities of the waves scattered from the known and the unknown parts of the electron density of the crystal and also their interference. The interference term depends on the cosine of the relative phase of the reference wave with respect to the object wave, which, therefore, can be determined to within a sign ambiguity. This phase information can then be used to recover the unknown part of the electron density. In the language of holography, the unknown wave can be reconstructed and its source can be found. In paper II, the reconstruction was shown to reduce to a standard inverse problem, similar to those encountered in image processing and phase recovery. Our algorithm, built on the above observations, searches (in real space) for an electron density that minimizes the deviation of the magnitudes of the calculated structure

† Part of this work was performed under the auspices of the US Department of Energy under contract no. W-7405-ENG-48.

factors from the measured ones (in reciprocal space). We found that the ubiquitous holographic 'dual image' also appears in X-ray crystallography, and that its presence is equivalent to the aforementioned sign ambiguity. We argued that, under favorable conditions, forcing the recovered electron density to be positive can eliminate the dual image.

A practical algorithm for X-ray crystallography was developed by recognizing that the holographic kernel is shift invariant. This enabled us to use fast Fourier transforms (FFT) and a conjugate-gradient optimizer, developed by Dennis Goodman (Goodman, Johansson & Lawrence, 1993), which is capable of incorporating various constraints. The result was the development of a suite of computer programs (*EDEN*) for the solution of crystallographic problems of current interest. The main solver program runs in $N \log N$ time, where N is the total number of resolution elements in the unit cell. Workstations (IBM 6000, HP 9000, SGI Iris or equivalents) are adequate to treat realistic problems.

Paper IV presented the basic algorithm and some applications. The method was first exercized on a model of thaumatin, a 207-residue protein. Various fractions of the protein were deleted and successfully recovered by the holographic method. We could recover deleted fractions even after varying the scaling of the data and adding noise. An important advantage of such studies was that goodness of recovery could be measured quantitatively. As a real example of protein crystallographic work, we reported briefly on the solution of the structure of a staphylococcal nuclease mutant. As the holographic method recovers electron density, it changes the phases of the structure factors. Also, there was a strong hint of diminished phase bias, compared to traditional Fourier methods.

Another significant step was taken by Béran (Béran & Szöke, 1995). In that work, a model protein was completely recovered from the knowledge of the electron density in about half the volume of the unit cell, even though the known density was entirely in the solvent region. This gives strong support to the notion that, using an appropriate algorithm, phase recovery under these conditions is an 'easy' computational problem. In paper IV, we continued to build on Béran's results and incorporated them into the holographic algorithm. In particular, we used the known electron density in part of the unit cell either as a mask or as a 'target' density. A cost function in real space was constructed from the deviations of the recovered electron density from the target density. It was minimized in parallel with the standard holographic cost function, which measures the deviations of the calculated structure factors from the measured ones. The relative importance of the two cost functions could be controlled by a relative weight (Lagrange multiplier). When the known density is in a solvent region, the procedure is similar to solvent flattening. When the known density is that of a molecular fragment, we expect the results to be similar to molecular replacement.

We will start this paper with a brief summary of the theory of the holographic method and our experience with it. The central theme of the present paper is the consistent use of available information. In *EDEN*, each kind of additional information becomes a new term in the cost function in the form of constraints or restraints. (Although both constraints and restraints will be incorporated into the algorithms, we will use the term constraints in all cases.) The holographic method is extended to multiple isomorphous replacement (MIR), multiple anomalous dispersion (MAD) and non-crystallographic symmetry (NCS).

After the derivation of two different versions of the pertinent equations, the algorithm incorporating MIR is tested on an artificial example, by adding 'heavy atoms' to a model of staphylococcal nuclease. We verify that such an algorithm works: given enough derivatives and heavy enough atoms, it recovers a perfect electron density. Using the same model, we explore the global convergence of the algorithm, as well as the difficulties of MIR when there are too few derivatives or heavy atoms that are too weak. We close the section by applying *EDEN* to the solution of kinesin using MIR data. This molecule has recently been solved by Kull, Sablin, Lau, Fletterick & Vale (1996). We used their data and were able to compare the holographic method with other available solutions.

Our treatment of non-crystallographic symmetry (NCS) is somewhat different from traditional methods. We do not interpolate in either real or reciprocal space. The inherent accuracy of our method is limited by the basis-function representation of the electron density. The expected advantage of the holographic method is that it is not too sensitive to exact *a priori* knowledge of the volume and shape of the monomers, *i.e.* the units repeated by the NCS operation, or of their possible differences. We have not yet tested our NCS algorithm on real problems.

EDEN is available free of charge to qualified collaborators. Please contact HS by e-mail at szoke2@llnl.gov.

2. Theory and algorithms

2.1. *Summary of the holographic algorithm*

The notation in this paper is the same as in our previous papers. For precise definitions, the reader is referred to paper II (Szöke, 1993) and for more details to Appendix *A* of this paper.

The electron density in the unit cell of a crystal is divided, perhaps artificially, into a known and an unknown part. The structure factors of the known part are denoted by $R(\mathbf{h})$. They are given by

$$R(\mathbf{h}) = \int_{\text{unit cell}} \rho_{\text{known}}(\mathbf{r}) \exp(2\pi i \mathbf{h} \cdot \mathcal{F}\mathbf{r})\, d\mathbf{r}, \qquad (1)$$

where we use standard crystallographic notation. The unknown part of the electron density is described as a sum of Gaussian basis functions of equal widths, centered on a grid that divides the unit cell into P_a, P_b, P_c equal parts along the crystallographic axes **a**, **b**, **c**, respectively. The grid points are denoted by $\mathbf{r}_p$, $p = 1, \ldots, P$, where $P = P_a P_b P_c$. Each Gaussian blob (voxel) contains an unknown number of electrons, $n(p)$:

$$\rho_{\text{unknown}}(\mathbf{r}) \simeq (\pi^{3/2}/d_{\text{res}}^3) \sum_{p=1}^{P} n(p) \times \exp[-(\pi^2|\mathbf{r} - \mathbf{r}_p|^2)/d_{\text{res}}^2]. \quad (2a)$$

Such a representation of the electron density carries with it an 'intrinsic' data resolution. The resolution, d_{res}, corresponds to the lattice spacing for which the structure factors decrease to $1/e$ of their peak value or, equivalently, $d_{\text{res}}^2 = B/4$, where B is an average crystallographic B factor. From the formula $B = 8\pi^2\langle u^2\rangle$, where $\langle u^2\rangle$ is the mean square amplitude of the atomic motion, we conclude that $(d_{\text{res}}/\pi)^2 = 2\langle u^2\rangle$. In the program, we use an input resolution $d_{\text{inp}} = 0.53 d_{\text{res}}$, whose value is more in accord with our intuitive notions of data resolution. In our formulas, we will use the more general, if cumbersome, notation that was used in our previous papers:

$$\rho_{\text{unknown}}(\mathbf{r}) \simeq [1/(\pi\eta\Delta r^2)^{3/2}] \times \sum_{p=1}^{P} n(p) \exp(-|\mathbf{r} - \mathbf{r}_p|^2/\eta\Delta r^2), \quad (2b)$$

where Δr is the mean grid spacing and η determines the width of the Gaussians relative to the grid spacing. The two forms of (2) are equivalent if $(d_{\text{res}}/\pi)^2 = \eta\Delta r^2$. As discussed in Appendix *B*, if the grid spacing is sufficiently fine, the electron density of the unknown part of the molecule can be well approximated by such a superposition of Gaussians. We also show that the choice of $\Delta r = d_{\text{res}}/\pi$, where $\Delta r \sim |\mathbf{a}|/P_a \sim |\mathbf{b}|/P_b \sim |\mathbf{c}|/P_c$ and $\eta = 1$ are a good choice for the lattice spacing in a simple lattice. In the algorithm to be described below, the unknowns $\mathbf{n} = \{n(p)\}$ are obtained by minimizing a cost function that measures the error between the calculated and measured structure-factor amplitudes. (The residual error that results from the finite mesh size will be discussed in §2.3.) When (2*b*) is extended periodically over the repetitions of the unit cell, a derivation presented in paper II results in the following formula for the structure factors of the unknown part, $O(\mathbf{h})$:

$$O(\mathbf{h}) = \exp[-\eta(\pi\Delta r|\mathcal{F}^T\mathbf{h}|)^2] \sum_{p=1}^{P} n(p) \exp(2\pi i\mathbf{h} \cdot \mathcal{F}\mathbf{r}_p). \quad (3)$$

The notation $R(\mathbf{h})$ for the structure factors of the known part of the structure and $O(\mathbf{h})$ for those of the unknown part of the structure is adopted from holographic theory, where $R(\mathbf{h})$ and $O(\mathbf{h})$ denote the reference and object wave, respectively. The squares of the absolute magnitudes of the structure factors of the crystal, $|F(\mathbf{h})|^2$, can be related to the measured X-ray diffraction intensities in a known way and satisfy the equation

$$\begin{aligned} |F(\mathbf{h})|^2 &= |R(\mathbf{h}) + O(\mathbf{h})|^2 \\ &= |R(\mathbf{h})|^2 + R(\mathbf{h})O^*(\mathbf{h}) + R^*(\mathbf{h})O(\mathbf{h}) \\ &\quad + |O(\mathbf{h})|^2. \end{aligned} \quad (4)$$

When the representation of the unknown density is substituted from (3), (4) becomes a set of quadratic equations in the unknowns, $n(p)$. The number of equations, N_h, is usually not equal to the number of unknowns, P; the equations are also ill conditioned and therefore their solutions are extremely sensitive to noise in the data. Under these conditions, a (quasi)solution of (2*b*) is obtained by minimizing the discrepancy or cost function (see *e.g.* Dainty & Fienup, 1987)

$$f_{\text{eden}} = \tfrac{1}{2}\sum_{\mathbf{h}} w'(\mathbf{h})^2[|R'(\mathbf{h}) + O(\mathbf{h})| - |F'(\mathbf{h})|]^2, \quad (5)$$

where $R'(\mathbf{h})$ and $F'(\mathbf{h})$ are modified forms of $R(\mathbf{h})$ and $F(\mathbf{h})$, defined below. The weights, $w'(\mathbf{h})^2$ are normally set to unity; however, in $P1$ symmetry, $w'(0,0,0)^2 = 1/2$. The weights $w'(\mathbf{h})^2$ could also be set to be proportional to the reliability of individual measured reflections, *i.e.* inversely proportional to their σ^2. Such a weighting scheme has not yet been implemented.

The effective or intrinsic resolution of the observable structure-factor amplitudes is determined by atomic structure factors, by atomic motions and by crystalline disorder. It can be measured by the slope of the curve of $\log|F(\mathbf{h})|^2$ *vs* $|\mathcal{F}^T\mathbf{h}|^2$, a curve similar to a Wilson plot. The intrinsic resolution of the Gaussian basis set (2*b*) is $(d_{\text{res}}/\pi)^2 = \eta\Delta r^2$. If the effective resolution of the measured structure-factor amplitudes is higher than that of the Gaussian basis functions, they have to be modified to

$$|F'(\mathbf{h})| = |F(\mathbf{h})| \exp[-\delta\eta(\pi\Delta r|\mathcal{F}^T\mathbf{h}|)^2], \quad (6a)$$

using a non-negative parameter δ. This procedure, usually called 'apodization', adjusts the resolution of the measured diffraction pattern to that of the Gaussian basis set used in the solution. It is equivalent to an appropriate smearing of the electron density of the protein, using a Gaussian smearing function. The smearing is essential for a mathematically stable fitting of the high-resolution reflections (see §2.3). An analogous procedure is used to adjust the intrinsic resolution of the known part to give

$$|R'(\mathbf{h})| = |R(\mathbf{h})| \exp[-\delta'\eta(\pi\Delta r|\mathcal{F}^T\mathbf{h}|)^2]. \quad (6b)$$

Two points should be noted about the cost function (5). First, the summation includes only available experimental data, *i.e.* we do not include values of $R(\mathbf{h})$ for which the corresponding $F(\mathbf{h})$ is missing. Thus, the cost function (5) does not make unwarranted assumptions about unobserved reflections: their values are indeterminate, as they should be. Therefore, truncation errors of Fourier inversions are absent and the consistent use of non-negative basis functions is fully justified. Second, the summation over $\mathbf{h}$ in (5) is taken only up to a certain resolution, which is determined by Δr. This is not a severe limitation, as the high-resolution structure factors are small to begin with, because of molecular motion and disorder, and are further apodized according to (6*a*,*b*). In order to simplify the algorithms for Fourier transforms, we expand the structure factors and the electron densities to $P1$ symmetry. This way, crystallographic symmetry is built into the starting set of $R(\mathbf{h})$ and $\{n(p)\}$. If needed, crystallographic symmetry can be maintained during the solution of (5) by the addition of a real-space cost function (Appendix *A*4.)

The equations are solved using a conjugate-gradient algorithm that is very efficient in the presence of nonlinear constraints (Goodman, Johansson & Lawrence, 1993). Although in previous versions of *EDEN* we linearized the cost function and solved it iteratively, in our current version of the program we use the full quadratic formula, (5). This eliminates the need for the complicated weights that appeared in paper IV. (Weights are still needed in the 'asymmetric' algorithm for MIR; see Appendix *A*1.) A basic constraint, non-negativity of the electron density, is incorporated directly into the conjugate-gradient optimizer by stipulating that all elements of the solution vector $n(p)$ be non-negative. The constraint can be used in two different ways: in 'completion mode', the added density itself is non-negative everywhere, while, in 'correction mode', the sum of the known initial electron density plus the added electron density is non-negative everywhere.

A second type of constraint, which we call a 'target' density, is expressed in terms of the amplitudes of the basis functions used in the main program. They will be denoted by $n(p)_{\text{target}}$. The cost function, f_{space}, is expressed as

$$f_{\text{space}} = \tfrac{1}{2}\lambda_{\text{space}}P\sum_{p=1}^{P}\tilde{w}_p^2[n(p) - n(p)_{\text{target}}]^2. \quad (7)$$

The overall relative weight (Lagrange multiplier), λ_{space}, and the individual weights at each point, $\tilde{w}_p^2 \le 1$, express the 'strength of our belief' in the correctness of the target density: the weights, $\tilde{w}_p$, may be used to emphasize or de-emphasize different regions of the target density (although generally they are set to 1 or 0), while λ_{space} determines the relative importance of f_{space} with respect to f_{eden}. In the presence of a target density, the actual cost function used in the computer program is the sum of f_{eden} (5) and f_{space} (7):

$$f_{\text{total}} = f_{\text{eden}} + f_{\text{space}}. \quad (8)$$

The fast algorithm described in paper IV is applicable to the quadratic cost function; in fact, fewer iterations are needed for convergence. For completeness, the actual formulas that are used by *EDEN* are listed in Appendix *A*.

If there is additional information, additional terms are added to the cost function (8). Such terms are discussed in §§3 and 4 of this paper.

2.2. *Basic properties of the holographic algorithm*

When we first proposed the holographic algorithm, we emphasized the close analogy of X-ray diffraction to holography. In the following paragraphs, we will examine the holographic algorithm from three different viewpoints. Let us start from the traditional difference Fourier analysis (or omit maps) for the completion of the electron density of the unit cell when part of the structure is known. In terms of the complex structure factors of acentric reflections (see Fig. 1), the unweighted difference-Fourier solution for the missing part is the set of vectors $O(\mathbf{h})$, each of which connects $R(\mathbf{h})$ to the closest point, O_2, on the circle with radius $F(\mathbf{h})$. Omit maps sometimes use weights for the measured structure factors that change the magnitude but not the phase of O:

$$|O(\mathbf{h})| = \big|w_{\text{Sim}}(\mathbf{h})|F(\mathbf{h})| - |R(\mathbf{h})|\big|. \quad (9)$$

These weights were first proposed by Sim and have been generalized many times since. [For recent publications, see Read (1986, 1990); also a very clear derivation by Viterbo in Giacovazzo (1992, p. 393).] In general, $O(\mathbf{h})$ is not collinear with $R(\mathbf{h})$ for acentric reflections. The Sim weights give an $O(\mathbf{h})$ along $R(\mathbf{h})$ that is closest to the statistically most probable solution (*e.g.* O_1 in Fig. 1).

The solution of (5) is not unique: this is an expression of the well known phase problem of crystallography (paper II). The equivalent mathematical statement is that an arbitrary element of the null space of an 'encoding' operator can be added to any vector $n(p)$ that minimizes the cost function (5). A simple geometric representation of this lack of uniqueness for acentric reflections, in the plane of complex numbers, is shown in Fig. 1. Since the phase of $|F(\mathbf{h})|$ is unknown, any $O(\mathbf{h})$ that connects the tip of $R(\mathbf{h})$ to any point on the circle with radius $|F(\mathbf{h})|$ results in the same value of the cost function. Thus, the difference between any two of the vectors $O(\mathbf{h})$ that satisfies this condition belongs to the null space of the encoding operator. However, constraints on the resulting electron density or the existence of additional diffraction data (*e.g.* MIR or MAD data) reduces the arbitrariness of the solution.

It can be seen from Fig. 1 that the unweighted difference Fourier solution is equivalent to the solution of (5) when, in addition, the null-space contribution is also minimized.* If there are no other constraints, the same solution will be found if f_{eden} in (5) is minimized by a gradient search, starting at $R(\mathbf{h})$. Indeed, as described in paper IV, the addition of a term that minimizes the null-space contribution [f_{null} of equation (11) of that paper] did not seem to make any difference and, therefore, is no longer used.

Second, from the point of view of holography, it was shown in paper II that the unweighted difference-Fourier solution is an equal superposition of the correct electron density of the missing part and its holographic dual image. These two solutions, O_1 and O_3, are shown in Fig. 1 for an acentric reflection. As mentioned above, a consequence of the curvature of the circle of radius $|F(\mathbf{h})|$ is that the midpoint between the correct solution and the dual solution is given more accurately by the weighted formula (9). It can also be seen from Fig. 1 that the weights are important only when the magnitudes of $|F(\mathbf{h})|$ and $|R(\mathbf{h})|$ are comparable. When $|F(\mathbf{h})| \gg |R(\mathbf{h})|$, the known part gives very little phase information. When the opposite is true, $|F(\mathbf{h})| \ll |R(\mathbf{h})|$, the phase of $O(\mathbf{h})$ is almost perfectly defined and the weights are not needed. In all cases but the last one, the main limitation on the reconstruction of the true electron density of macromolecules is the presence of the dual image.

From a third point of view, it was shown by Béran (Béran & Szöke, 1995) that the difference-Fourier solution is equivalent to the assumption that the electron density is equally well known in all parts of the unit cell. In other words, it is assumed that the missing electron density is distributed about equally in the region of the known part of the structure and the unsolved part of the structure.

Several conclusions can be drawn from this discussion; some are well known, some are less well known. First, if there is no other information, the (weighted) difference Fourier solution is indeed the best solution. This agrees with Lánczos's dictum. In fact, such is almost never the case: for instance, it is always known that the electron density is non-negative everywhere in the unit cell. We have noted above that algorithms that restrict $O(\mathbf{h})$ to be parallel to $R(\mathbf{h})$ do not necessarily satisfy non-negativity of the electron density.

Second, it was shown above that the null space of the holographic encoding operator, for acentric reflections, consists of all the vectors that connect points on the circle of radius $|R(\mathbf{h})|$ in Fig. 1. This null space has N_h dimensions, but it is of finite measure: its measure in each dimension is the circumference of the circle. There is a point on the circle that corresponds to the correct electron density (the correct image) and a second point, related symmetrically to it with respect to the direction of $R(\mathbf{h})$, that corresponds to the holographic dual image. Let us ask now what happens to the null space and to the dual image when additional information becomes available. In general, the additional information limits the region of the null space that remains accessible to $O(\mathbf{h})$. As additional information is added, the correct solution remains accessible while the probability that the dual image remains accessible decreases. When the known part is a large fraction of the molecule, on the average, the tip of $R(\mathbf{h})$ gets close to the circle and the magnitude of $O(\mathbf{h})$ gets smaller. This results in an apparent paradox, as the increase of the known part of the electron density does not change the circumference of the circle or the degeneracy of the cost function on it. Therefore, the measure of the arbitrariness of the solution does not seem to be smaller when a larger fraction of the molecule is known. The resolution of the paradox is that, as the unknown part gets smaller, $|O(\mathbf{h})|$ gets shorter on the average and less of the circle is effectively available.

A third conclusion is that, since the cost function (5) contains terms only for the measured intensities, the calculated structure factors that correspond to unmeasured reflections may assume any value. This is the correct mathematical expression of those being unknown. When Fourier syntheses are treated in the usual way, such unknown values are automatically assigned the value zero. This is the chief reason for truncation artifacts when electron-density maps are calculated by (inverse) Fourier transformation. A known way of dealing with these problems is through maximum-entropy techniques.* The holographic method does not have these problems but it

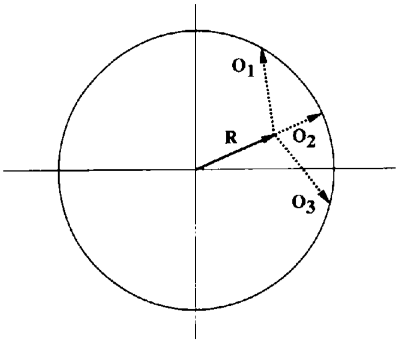


Fig. 1. Geometric representation of equation (4) in the complex plane for acentric reflections. $R(\mathbf{h})$ is the vector representing the structure factor of the known part of the electron density. The circle around the origin has the radius $|F(\mathbf{h})|$. Any vector $O(\mathbf{h})$ that starts at the tip of $R(\mathbf{h})$ and ends on the circle satisfies equation (4). The unweighted difference Fourier solution is $O_2(\mathbf{h})$. If the 'correct' solution is $O_1(\mathbf{h})$, the dual image is represented by $O_3(\mathbf{h})$. The Sim-weighted difference Fourier solution is the midpoint between the tips of $O_1(\mathbf{h})$ and $O_3(\mathbf{h})$.

* Such a solution is equivalent to the (Moore–Penrose) generalized inverse of the (singular) encoding operator (Bertero, 1989).

* The maximum-entropy solution selects the flattest non-negative electron density that is compatible with the experimental data (Gull & Daniell, 1978; Skilling & Gull, 1985).

is not completely equivalent to the maximum-entropy reconstruction.

2.3. *Limiting accuracy of the holographic algorithm*

The representation of the unknown density, presented in (2*b*), uses an overcomplete set of Gaussian basis functions that are not mutually orthogonal. The usefulness and accuracy of the holographic algorithm are determined by the answers to the following mathematical questions. How well can the electron density of an arbitrary protein be approximated by the superposition of such basis functions (with appropriate coefficients)? Is there a well defined algorithm to find such a set of coefficients, given the electron density? Is the set of coefficients unique? And, finally, if two sets of coefficients are similar (close to each other), are the resulting electron densities close to each other? Fortunately, mathematicians have done extensive research on such non-orthogonal redundant basis sets: they are called frames. Excellent discussions can be found in a book by Daubechies (1992) and in a review by Heil & Walnut (1989). Some of their important results are described in Appendix *B* and are summarized below.

The mathematicians' answer to the first question is that electron densities can indeed be approximated well by such representations, if the electron density does not vary too wildly. Restated in technical language, the requirements are that the diffraction pattern and the basis set should have similar intrinsic resolutions and that the grid spacing should be about twice as fine as required by the corresponding Nyquist criterion. In our algorithm, this is achieved by the choices outlined in §2.1. It is also true that two representations with similar coefficients do yield similar electron densities. On the other hand, a given electron density can be represented by several different sets of coefficients. In fact, there are many possible algorithms to find a set of coefficients that approximate the electron density of the crystal equally well, of which the algorithm used by *EDEN* is one (good) example. Conversely, two similar electron densities produce similar sets of coefficients in our algorithm, which is therefore mathematically stable.

In the rest of the section, we investigate the accuracy of the representation of the electron density by *EDEN* quantitatively: first by representing a single Gaussian electron density in a general position, using Gaussians on a grid, then by recovering a model protein.

In the first set of tests, we placed a single Gaussian electron density onto a generic position (*i.e.* not necessarily on a grid point). We calculated its diffraction pattern and we found the best representation of the diffraction pattern in terms of Gaussians on a grid, using the program *BACK* (see Appendix *A*5.3). We then calculated the diffraction pattern of the latter and compared the phase and amplitude accuracy of the result to the accurate diffraction pattern. Alternatively, we obtained an approximate representation of the original Gaussian electron density in terms of Gaussians on the closest grid points by equating their moments. The calculations are presented in terms of a unit grid spacing, *i.e.* the grid points are (0, 0, 0) (1, 0, 0) *etc.* in a simple grid and the points (1/2, 1/2, 1/2), (−1/2, 1/2, 1/2) *etc.* are added to them in a body-centered grid. The value of η is 1.0 for the simple grid and 0.75 in the body-centered grid, in accordance with Appendix *A*1. The results are plotted in Fig. 2 for both a simple and a body-centered grid. As expected, the maximum phase error in a simple grid occurs at the point (1/2, 1/2, 1/2) and in a body-centered grid at the coordinate values (1/2, 0, 0), (0, 1/2, 0), (0, 0, 1/2). In the simple grid, the maximum phase error is 47° and the corresponding amplitude error is 28%. In a body-centered grid, the maximum phase error is 26° and the corresponding amplitude error is 20%. With the analytic approximation (data not shown), the maximum phase and amplitude errors are again 47° and 28% for a simple grid but they are 40° and 26% for a body-centered grid. The average phase errors are 27° and 18° for a simple and a body-centered grid, respectively. The phase error for a complicated molecule that is uniformly distributed in the unit cell is expected to be less. Note that the recovery of the electron density is more accurate with our new values of Δr and η than those shown in §2.1 of paper IV.

In our next set of tests, we repeated the recovery of part of the thaumatin model structure, presented in paper IV. As we changed the algorithm from the iterated linearized system of equations of paper IV to the quadratic cost function presented in this paper, we repeated the calculations for the values Δr = 1.8 Å,

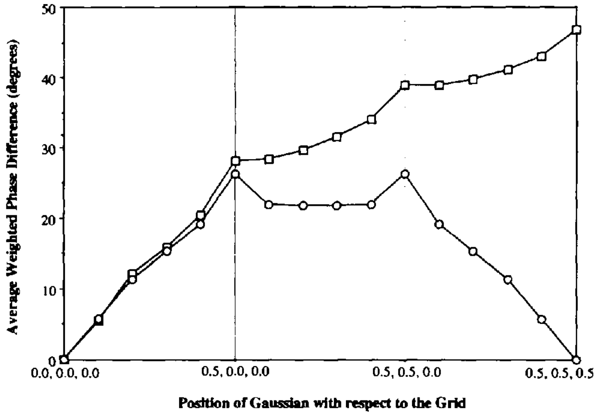
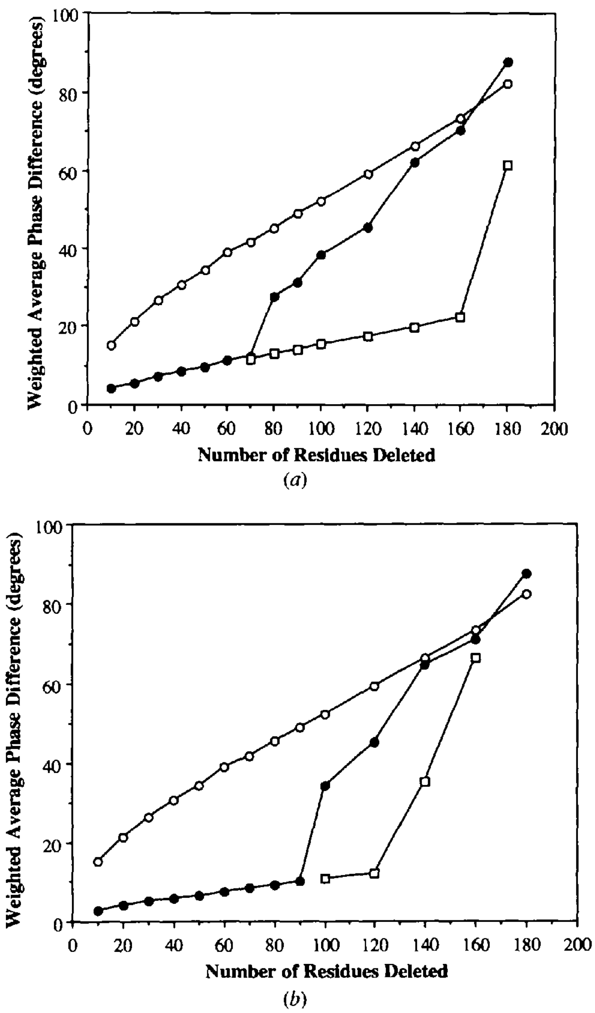


Fig. 2. Phase errors of the diffraction pattern encountered when a Gaussian electron density, in a general position, is approximated by a set of Gaussians on a grid, as determined by a non-linear interpolation using *BACK*. □ Simple lattice; o body-centred lattice. The curves represent the phase error as a function of the position of the Gaussian along straight line segments, whose end points are written on the *x* axes of the curves. In a simple lattice, the linear interpolation (not shown) does equally well but, for the body-centered case, the linear interpolation does worse.

$\eta = 0.28$ and $\delta = 1$ as in paper IV. The results are presented in Fig. 3(*a*). When it is compared to the corresponding Fig. 2 of paper IV, we see that in general the new quadratic algorithm is more accurate and converges somewhat better than the old one. Without a solvent mask, 70 out of the 207 residues could be recovered. When a hard solvent mask that covered half the unit cell was imposed, as many as 160 residues out of the 207 (or 77%) were found essentially perfectly. The recovered electron density was within 10% of that of the original model.



(*a*)

(*b*)

Fig. 3. The recovery of missing residues from the thaumatin model was evaluated by calculating the weighted averaged phase difference between the $F_{\rm true}$ data and the structure factors corresponding to the truncated model before and after *EDEN* was used to recover the missing electron density. Open circles show the phase difference, before recovery, between the truncated model and the complete model. Closed circles show the phase difference after an *EDEN* run without a solvent mask; and squares show the phase difference after an *EDEN* run with a solvent mask. In (*a*), $\Delta r = 1.8$ Å, $\eta = 0.28$ and $\delta = 1$ as in paper IV. In (*b*), $\Delta r = 1.4$ Å, $\eta = 0.75$ and $\delta = 0.6$, which corresponds to an input resolution $d_{\rm inp} = 2.0$ Å.

In Fig. 3(*b*), we used the new set of $\Delta r = 1.4$ Å, $\eta = 0.75$ and $\delta = 0.6$ that corresponds to an input resolution $d_{\rm inp} = 2.0$ Å. As expected, the phase accuracy of the recovery is very good. The convergence of the algorithm without a solvent mask is also better: in fact, 90 out of 207 residues (*i.e.* 43%) were recovered essentially perfectly. With a solvent mask or a solvent target function, we got perfect recovery only up to 120 residues. This is a respectable 58%, but it is less good than using the previous values of Δr and η. We do not understand the reasons for the difference. Fig. 4 is similar to Fig. 4(*a*) of paper IV. Note the absence of the systematic shift of the electron-density contour with respect to the model that was quite noticeable on those figures. Our current understanding, as opposed to what we wrote in paper IV, is that the shifts were caused by a bug in the computer program that we have since eliminated. (We apologize for it.) With the new values of Δr and η, after regridding at a 2:1 ratio, the resulting electron-density map is very similar to the traditional 3:1 finer electron-density maps to which crystallographers are accustomed.

Some additional tests show the power of the positivity constraint in model problems that have no noise or solvent. *EDEN* solved the model of staphylococcal nuclease at 3 Å resolution using a low-resolution (~ 6 Å) solvent target that covered 61% of the unit cell and no other information. (For details of the model used, see §3.3 below.) A similar result was reported by Bricogne (1993) using a very sophisticated algorithm in reciprocal space. We recall that Béran & Szöke (1995) found that the phases of the structure factors of a model protein could be recovered completely when the electron density was given in a little more than half the unit cell. The above results seem to contradict recent conclusions of Millane (1996). We interpret Millane's conclusion as establishing

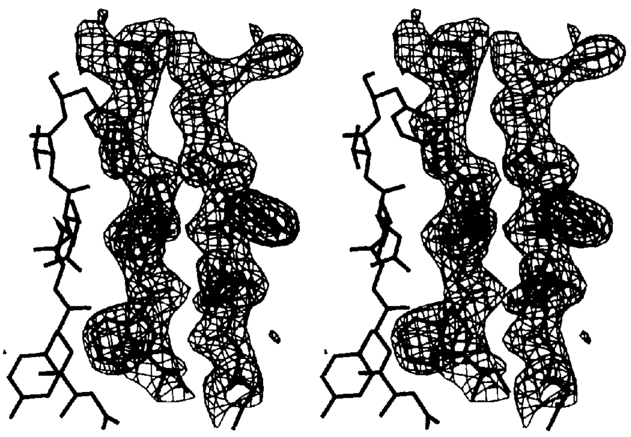


Fig. 4. A sample (residues 18–23) of the electron density recovered for the thaumatin model after omitting residues 1 through 120 using a solvent mask. Note that the leftmost of the three β strands is part of the known part of the model and, consequently, electron density for this region is not recovered.

an upper limit to the additional information needed to solve the crystal structure. In our opinion, he does not exclude the possibility of solving the structure with less information.

2.4. *Convergence of the gradient solver and other methods of solution*

Recovery of the electron density by the holographic method is based on finding the minimum of the cost function (8). The difficulty of this task can vary from the easy to the impossible, depending on the number of minima and the topography of the cost function. In easy cases, the cost function has a single minimum and the conjugate-gradient solver used in *EDEN* finds it. In fact, Figs. 3 and 4 show that in many (supposedly hard) cases the algorithm converges well. In more difficult problems, multiple local minima of the cost function (8), not all of the same depth, are expected. Their presence was clearly seen in Béran's work (Béran & Szöke, 1995). It was found there that a Monte Carlo (simulated-annealing) algorithm, applied in reciprocal space, was able to deal with the difficulty. It was also observed that, when the amount of available information is marginal, the speed of convergence of the algorithm slows down exponentially. We will argue below that the situation in protein crystallography can be even worse: the number of minima of the cost function, with similar measures of goodness, can be so large that it is practically impossible to find the best one.

Let us inquire into the mathematical properties of the holographic method. The value of the cost function (8) is determined by the values of the electron density, $n(p)$, at each lattice point. As there are P lattice points, it is a (real-valued) function defined in a P-dimensional space. Usually, that space has a very large number of dimensions: the number of lattice points, P. In realistic problems, P can be 10^4–10^5. On the other hand, the number of distinguishable densities at each lattice point, N, is limited. In practice, two electron densities that differ by a 10% random variation are indistinguishable to a human observer. In other words, N is of the order of 10. Therefore, the total number of possible electron densities, that is of the order of N^P, is an astronomically large number: $10^{10\,000}$–$10^{100\,000}$. The difference between any two densities, as measured by the integrated root mean square of the density, is of the order of NP, *i.e.* 10^5–10^6. This (Cartesian) distance between any two points in the multidimensional configuration space is thus many orders of magnitude smaller than the number of points in the space. There are other well known systems in nature that exhibit similar mathematical properties. The best known example is the genetic makeup of living organisms. For example, a molecule of DNA with P base pairs is able to encode 4^P different sequences, but only P mutations are needed to transform a given DNA molecule to any other one (Eigen, 1992).

There are other similarities between the mathematical properties of the crystallographic algorithm and the genetic code. The variables of the holographic algorithm are (discretized) electron densities, while measured diffraction intensities that provide constraints are in reciprocal space. The connection between them is given by the Fourier transformation, which has the property that a (positive) change in any $n(p)$ will, in general, increase the magnitude of some structure factors and decrease the magnitude of others in complicated ways. In the genetic code, it is the sequence of the base pairs in DNA that determines the properties of a given protein, while the survival of the individual and its reproductive success depends in very convoluted and not obvious ways on the DNA sequence of a particular protein (Kauffman, 1993). A change in any base pair will have complex effects on various properties of the individual: some of them increasing its average reproductive success, others decreasing it.

In the crystallographic problem, when the information becomes marginal, the cost function may have a very large number of minima. Indeed, if the domain of attraction of the average minimum has radius R, measured in configuration space, the number of expected minima is $(NP/R)^P$. In realistic problems, P is so large that, even if the domain of attraction of the minima is within 1% of the whole space, the number of minima can still be large (2×10^4 for $P = 1000$). The barriers between neighboring minima may form a tree-like structure or may be even less tractable. In most such cases, a simulated annealing (or any other Monte Carlo algorithm) would show critical slowing down, *i.e.* the computer time to solve the structure diverges exponentially. In practical terms, that means that the structure cannot be solved using the information at hand.

We will now present a curious mathematical property of the holographic reconstruction algorithm. Starting from a uniformly filled unit cell [$n(p)$ = constant for all p], there are straight line paths [in the multidimensional $n(p)$ space] on which the cost function is continually decreasing and that lead to each one of the local minima of the cost function, as well as to the global minimum. The derivation of this property is straightforward but tedious. The theorem shows that the ridges that separate local minima of the cost function cannot be higher than the value of the cost function corresponding to the correct number of electrons in the unit cell distributed uniformly. The proof of the theorem assumes that the electron density corresponding to the local minima and to the global minimum of the cost function is known, so the direction of the search is known. Unfortunately, in a space of such high number of dimensions, finding the appropriate directions is not an easier problem than finding the global minimum of the cost function itself.

In §3.2 below, we present studies in which we systematically vary the amount of available information. We did these studies on MIR models in order to be sure that the solutions are unique. When the available information was marginal, we have found many disjoint minima with similar values of the cost function but with large differences in the electron-density distribution. In such situations, especially when experimental deviations are present, it is not clear whether any particular solution of the crystal structure is more reliable than any other one. Such difficulties were discussed by Brändén & Jones (1990), Baker, Krukowski & Agard (1993) and Kleywegt & Jones (1995). A fundamental but usually unstated assumption of protein crystallography is that among all the solutions only one gives a chemically sensible structure. Our studies cannot contradict such an assumption but they cannot support it either.

3. Multiple isomorphous replacement (MIR) and multiple anomalous dispersion (MAD)

3.1. *Derivation of equations for MIR*

Crystal structures can be solved by multiple isomorphous replacement (MIR) if the only change in crystal structure is the addition of heavy atoms. In other words, it will be assumed below that the native protein's structure is unchanged when heavy atoms are added. MIR methods require that individual data sets be taken for each derivative and that the positions of the heavy atoms and their occupancies be found by Patterson or direct methods. Conventional MIR methods then proceed to find the phases of the native protein. Very often, the resulting phase set does not give electron-density maps that are easily interpretable. This is the stage where the holographic method can be of advantage. While in principle the holographic method is equivalent to the conventional method of finding the phases of the structure factors of the native protein, we expect that incorporating all known constraints consistently should improve the attainable accuracy of the solution. In a test case using real data, presented in §3.4, *EDEN* resulted in a clear improvement over conventional methods.

From a mathematical point of view, heavy-atom derivatives (as well as anomalous dispersion) increase the number of independent equations with respect to the unknowns. With a sufficient number of derivatives, the phase problem should therefore be solvable. On the other hand, it is very difficult to predict the convergence properties of different algorithms. We will present two mathematically equivalent versions of MIR: a symmetric and an asymmetric algorithm. In our synthetic test cases, the asymmetric algorithm converges better than the symmetric one does and they both converge less well than traditional methods do. The symmetric algorithm is much better suited to the treatment of real data as it is less sensitive to experimental errors. More will be written about this subject in §3.4.

The relevant equations in *EDEN* are simple generalizations of (1)–(6*a*,*b*). The unknown density (2*b*) is that of the native protein. So is the structure factor, $O(\mathbf{h})$, of (3). Suppose $M+1$ sets of diffraction amplitudes have been measured: one for the native and one each for the M derivatives. Suppose also that the positions of the heavy atoms and their occupancies were found using Patterson or direct methods. The calculated structure factors for the heavy atoms then belong to the known part of the structures. For the mth derivative, they will be designated $R_m(\mathbf{h})$. The measured structure factors of the mth derivative will be denoted $|F_m(\mathbf{h})|$. Then, (4) can be generalized to

$$\begin{aligned}|F_m(\mathbf{h})|^2 &= |R_m(\mathbf{h})+O(\mathbf{h})|^2\\ &= |R_m(\mathbf{h})|^2 + R_m(\mathbf{h})O^*(\mathbf{h})\\ &\quad + R_m^*(\mathbf{h})O(\mathbf{h}) + |O(\mathbf{h})|^2,\end{aligned} \qquad (10)$$

where $m = 0,\ldots,M$ ($m = 0$ designating the native protein). In Fig. 5(*a*), these equations are presented in their well known geometrical form for two derivatives. It can be seen that, for each value of $\mathbf{h}$, solving (10)

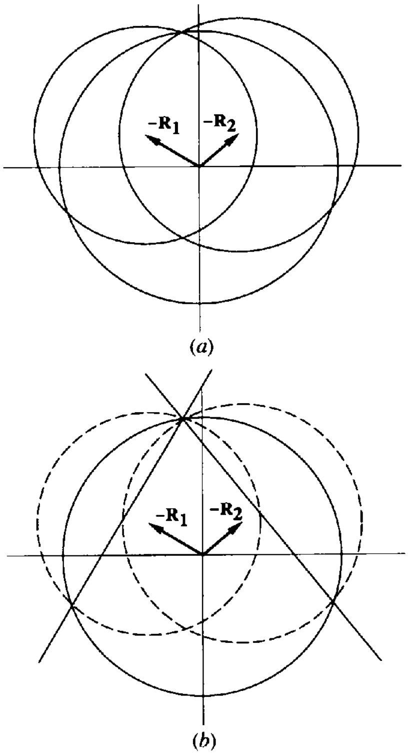


Fig. 5. Geometric construction in the complex plane for an acentric reflection in multiple isomorphous replacement. (*a*) is the traditional construction for the symmetric algorithm, equation (10). The circle centered on the origin has a radius of $|F_o(\mathbf{h})|$, the structure factor of the native. The other circles have radii $|F_m(\mathbf{h})|$, they are centered on $-R_m(\mathbf{h})$. (*b*) is the corresponding construction for the asymmetric algorithm, equation (12). The circle centered on the origin has a radius of $|F_o(\mathbf{h})|$. The equations for the straight lines can be derived from the equations for the derivatives. The points of intersection in (*a*) and (*b*) are the same, as expected.

is equivalent to finding the intersection of the circles that are centered on the negative value of the structure factors of the heavy atoms, $-R_m(\mathbf{h})$, and whose radius is the value of $|F_m(\mathbf{h})|$. This is the same as the (simple minded) solution of the traditional MIR algorithms (see *e.g.* Giacovazzo, 1992). In the symmetric algorithm, (10) are solved by minimizing a cost function that is analogous to (5):

$$f_{\rm eden} = \tfrac{1}{2} \sum_{m=0}^{M} \lambda_m \sum_{\mathbf{h}} w'_m(\mathbf{h})^2 [|F'_m(\mathbf{h})| - |R'_m(\mathbf{h}) + O(\mathbf{h})|]^2. \tag{11}$$

In (11), we introduced weights λ_m that can express the reliability or quality of the measurements of each derivative. The default is $\lambda_m = 1$. If derivatives have different intrinsic resolutions, (6*a*,*b*) can be generalized by allowing each equation in (11) to be apodized separately; see Appendix *A*1.

The asymmetric algorithm is obtained from (10) by subtracting the equation for the native from the equation for each derivative. The equations are

$$\begin{aligned} |F_0(\mathbf{h})|^2 &= |R_0(\mathbf{h})|^2 + R_0(\mathbf{h})O^*(\mathbf{h}) + R_0^*(\mathbf{h})O(\mathbf{h}) \\ &\quad + |O(\mathbf{h})|^2, \end{aligned}$$

$$\begin{aligned} &|F_m(\mathbf{h})|^2 - |F_0(\mathbf{h})|^2 \\ &\quad = |R_m(\mathbf{h})|^2 - |R_0(\mathbf{h})|^2 + [R_m(\mathbf{h}) - R_0(\mathbf{h})]O^*(\mathbf{h}) \\ &\quad + [R_m^*(\mathbf{h}) - R_0^*(\mathbf{h})]O(\mathbf{h}), \quad m = 1, \ldots, M. \end{aligned} \tag{12}$$

Equations (12) are solved by minimizing the cost function

$$\begin{aligned} f_{\rm eden} = \tfrac{1}{2}\Big(&\lambda_0 \sum_{\mathbf{h}} w'_0(\mathbf{h})^2 [|F'_0(\mathbf{h})| - |R'_0(\mathbf{h}) + O(\mathbf{h})|]^2 \\ &+ \sum_{m=1}^{M} \lambda_m \sum_{\mathbf{h}} [w'_m(\mathbf{h})^2 / |R_m(\mathbf{h})|^2] \\ &\times \{|R_m(\mathbf{h})|^2 - |R_0(\mathbf{h})|^2 - |F_m(\mathbf{h})|^2 \\ &+ |F_0(\mathbf{h})|^2 + [R_m(\mathbf{h}) - R_0(\mathbf{h})]O^*(\mathbf{h}) \\ &+ [R_m^*(\mathbf{h}) - R_0^*(\mathbf{h})]O(\mathbf{h})\}^2\Big). \end{aligned} \tag{13}$$

In the asymmetric algorithm, weights similar to those described in paper IV have to be used in order to avoid numerical instability. Equations (12) and (13) can also be written in a more general form that allows for different resolutions for the native and for each of the derivatives. Actual equations used for the cost function, weights and the calculation of the gradient are written out in Appendix *A*1. Equations (12) can also be solved geometrically, similarly to the traditional solution of (10). As presented in Fig. 5(*b*), the solution is obtained by the intersection of a circle (for the native) and straight lines (for each derivative.) The straight line for each derivative goes through the same two points where the analogous circles intersect in Fig. 5(*a*). As expected, the solution does not change but it may have different convergence properties and different sensitivity to imperfect data and noise. We found that, for experimental data on real proteins, the symmetric algorithm is much more stable than the asymmetric one.

Several remarks are in order. First and simplest: if no part of the native is known, $R_0(\mathbf{h}) = 0$ [for $\mathbf{h} \neq (0,0,0)$]. If parts of the native protein are known, *e.g.* from molecular replacement, the known part can be added to $R_m(\mathbf{h})$ in reciprocal space, as well as to the starting set of $n(p)$ in real space. In complete analogy to the case of a single data set, *EDEN* can be run in correction mode. It is thus capable of correcting the density of any part of the molecule that is guessed incorrectly. Second, it is clear from our derivation that solvent (or other) targets, as well as crystalline or non-crystalline symmetry, can be used together consistently. Third, it should be reiterated that the minimization of the cost function (11) or (13) is carried out by changing the density in real space, as opposed to the traditional methods that solve the system of equations (10) for $m = 0, \ldots, M$ for each $\mathbf{h}$ separately.

3.2. *Equations for MAD*

Multiple anomalous dispersion (MAD) can be treated very similarly to MIR. As MAD data sets are taken on a single crystal, the basic assumption of isomorphism is always correct; the main problem with the method is usually the low signal-to-noise ratio. The fundamental assumption in *EDEN*'s treatment of MAD is that the structure amplitudes of the unknown part (which will be called the native) have no anomalous dispersion, *i.e.* f'' for all the unknown atoms is zero and their f' is independent of X-ray energy. In other words, the anomalously scattering atoms are always considered to be 'heavy atoms'. We will start from the point where the anomalously scattering (heavy) atoms have been found by Patterson methods or by direct methods and their structure factors, including the anomalous part, have been calculated. In a $P1$ crystal, the $h \geq 0$ data set can now be treated exactly as a derivative in MIR; both the symmetric and the asymmetric algorithms (10) and (12) are applicable. As is well known, in $P1$ symmetry, the $h \leq 0$ reflections are an independent data set. The easiest way to use them in *EDEN* is to create a 'flipped' data set by negating all the indices of the reflections, $\mathbf{h} \rightarrow -\mathbf{h}$, at the same time flipping the signs of the phases of the heavy atoms and declaring this new data set to be a separate derivative. As shown in Fig. 6, Friedel's relations apply to the structure factors of the native because that part of the structure has no anomalous dispersion. Therefore, the unknowns in this 'flipped' data set are the same as those for the $h \geq 0$ data set. In higher symmetry, similar considerations apply. It is clear that such data sets can be used together with MIR data. The only difficulty one might encounter in this procedure is that the anomalous data sets are weighted too heavily.

3.3. *Characterization of the MIR cost-function surface*

The first test of *EDEN* in its MIR mode was to verify that the program is capable of solving easy problems. The atomic coordinates and B values of staphylococcal nuclease [a 149 residue protein (Somoza *et al.*, 1995)] were used to produce a 'native' diffraction pattern, $|F_0(\mathbf{h})|$, by removing the phases from its calculated diffraction pattern. The model of staphylococcal nuclease itself was modified by running *BACK* and *FORTH* on it (see Appendix *A*5) so that a perfect solution, *i.e.* $f_{\text{eden}} = 0$ of (11) or (13), was attainable. 'Derivatives' were produced by placing heavy atoms into known positions. The calculated structure factors of the heavy atoms were used for $R_m(\mathbf{h})$ in (11) and the magnitudes of $|R_m(\mathbf{h}) + F_0(\mathbf{h})|$ were used for 'derivative data'. The *EDEN* solver was started either from an empty unit cell or from a unit cell with a random electron density, $n_{\text{ran}}(p)$, and from the corresponding structure factors.

We expected that the cost function (11) or (13) would have a single global minimum if there are several derivatives and if the added atoms are heavy enough. If the number of derivatives is diminished and the number of added electrons is reduced, we expected that multiple minima of the cost function would develop. Furthermore, we expected that under less favorable conditions the number of minima might increase considerably, yielding local minima ('solutions') with similar values of the cost function, but very different structures (§2.4). The questions we wanted to answer are: How many minima are there, at least approximately? How are their depths distributed and how high are the barriers between minima? Do the minima have a hierarchical distribution, *i.e.* are there a few large minima surrounded by many small minima? Are there any phase-transition-like singularities on the cost-function surface? As discussed in §2.4, these properties of the cost-function surface determine whether the crystallographic reconstruction problem is easy, hard or impossible.

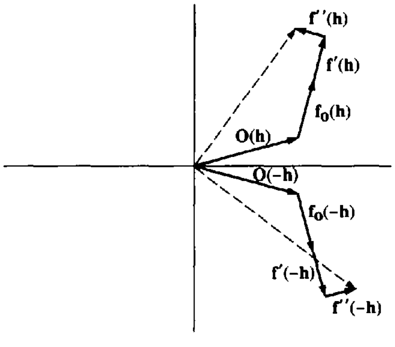


Fig. 6. Geometric construction in the complex plane for an acentric reflection with anomalous dispersion. The structure factor for the 'native' is denoted by $O(\mathbf{h})$. The structure factor of the anomalously diffracting atoms has three parts. The part that is independent of X-ray energy is denoted by f_0. The X-ray energy dependent part has a component that is parallel to f_0, it is denoted by f' and a perpendicular component denoted by f''. The figure is drawn for a Bijvoet pair, $\mathbf{h}$ and $-\mathbf{h}$. The measured structure factors are drawn in broken lines. If the figure for the $-\mathbf{h}$ reflection is reflected about the horizontal axis, the structure factor for the 'native' overlaps that for $\mathbf{h}$.

The goodness of the solution could be judged by several measures: by the root-mean-square deviation of the measured structure factors from the absolute values of the calculated structure factors [this will be referred to as the standard deviation (s.d.)], by the crystallographic R factor and by the distance of any given minimum from the perfect solution. Distances between two densities $n_1(p)$ and $n_2(p)$ were measured by using (54). Of course, the distance from the perfect solution can only be used in artificial tests. On the other hand, when optimization is run from different starting points or when it is started from several random perturbations of a previous minimum, the distance among the various solutions, together with their standard deviations and crystallographic R factors, can be used very advantageously to monitor the progress of the program and to see whether different solutions converge into a single basin of attraction or whether there are many disjoint basins.

All our tests were run at 3 Å resolution. We used three 'derivatives' of the staphylococcal nuclease model with two atoms of Z electrons each added to each asymmetric unit. With three derivatives, the probability of an ambiguous phase determination in a traditional MIR program is much reduced with respect to two derivatives. The program *PHASES* (Furey & Swaminathan, 1990) with $Z = 90$, 60, 30 gave excellent solutions of the structure with phase errors of 22, 22, 14°, respectively.

Our first set of tests was started with an empty model for the native. The empty start is expected to be optimal, as discussed in §2.4. The symmetric algorithm was able to solve the structure with $Z \geq 90$. At $Z = 60$, we did not get convergence. The asymmetric algorithm did much better. It solved the protein for $Z \geq 43$. At $Z = 42$ and less, it abruptly ceased to converge. We tried a 'mild' simulated annealing on $Z = 40$ and did not successfully converge on the correct solution.

In the second set of tests, we started the conjugate-gradient optimizer from a random set of densities in the unit cell. This so-called multistart algorithm is described by Rinnoy Kan & Timmer (1989). At the crudest level, a random start is almost always worse than an empty start. On a less crude, but still qualitative, level, we summarize the results of *EDEN* runs with random starts for various values of Z:

	Number	Number of solutions by class			
Z	of trials	Perfect	Good	Fair	Poor
100	20	6	11	3	0
80	20	5	2	5	8
60	20	1	5	8	6
40	10	0	0	2	8.

The meaning of a perfect solution is clear. Good solutions were 'almost perfect' in the sense that they were clearly recognizable in the electron-density maps and they did not have any gross errors either in the protein part or in the (empty) solvent region. In the usual crystallographic display programs *O* or *FRODO*, such a solution would be indistinguishable from a perfect one. Fair solutions should also be largely traceable, but they would have several regions that were in need of iterative refinement. Finally, poor solutions are, by and large, wrong.

On a more quantitative level, we will discuss the results of the 20 pseudo-random starts with $Z = 100$. Table 1 reports all three measures – standard deviation, distance from the perfect solution and R factors – as well as the qualitative observation, where letters g and f denote good and fair results. (Essentially perfect results have no letter.) The numbers 75–94 in column 2 identify the seeds that were used to generate the initial random distributions. It is clear from this table that both the values of the ending R factor and standard deviation can spot a perfect solution. The same measures are less useful for distinguishing good and fair solutions. For that, a much better measure can be obtained from the $K(K-1)/2$ mutual distances of K solutions. The mutual distances of the 11 good results divide into two clusters, see Tables 2 and 3. The distances within each cluster are of the order of 4500 e, while the distance between the two clusters is ~12 500 e and the distance of each cluster from the perfect solutions is ~9000 e. (For comparison, the distance of the empty model from the correct solution is ~30 000 e and the starting points of the random starts from the correct solution are ~50 000 e.) The three fair solutions (Table 4) are ~18 000 e away from each other and at a similar distance from the perfect solution. It is clear that Table 4 shows three distinct deep basins of attraction in addition to the correct one. However, there are apparently many more shallow local minima of the cost function. Similar tables were produced for $Z = 80$, 60 and 40. As expected, as Z is decreased, the mutual distances even among the good and fair solutions get progressively larger, while the poor solutions get as far as ~40 000 e from each other. This shows that, in these harder cases, the algorithm does not converge at all. We have not found clear evidence for clustering of the solutions into well defined basins of attraction at these lower Z's.

The third set of tests probed how far a solution could be perturbed and still return to the starting point that is a local or global minimum of the cost function. The electron density $n(p)$ of a local optimum or of the perfect solution was perturbed by a set of pseudo-random numbers, uniform in $[0, 1)$, scaled by an appropriate factor. The negative densities were then set to zero and the solver was run to find the closest local minimum. Tables 5 and 6 show the results for $Z = 40$ and $Z = 20$, respectively, when the perfect solution was randomly perturbed with a scale whose distance from the original (perfect) solution is comparable to the total number of electrons (30 000). Surprisingly, these large random perturbations do not take the solution out of the basin of attraction of the global minimum. A similar study was done for $Z = 40$, where the solver converged to a poor local minimum when started from the empty unit cell. We perturbed this local minimum by comparably large random perturbations and observed that even in this case we could not escape its basin of attraction. The latter results are presented in Table 7. An additional surprising feature of this study is the 'roundness' of the basin of attraction, as expressed by the very small variation of the initial standard deviation after the perturbation.

Table 1. *Analysis of trials with Z = 100*

s.d. = standard deviation. Start and end distances are given in electrons.

No.	Seed	Start s.d.	End s.d.	Start distance	End distance	End R factor	Class
1	75	35.4	6.1	50564	8915	5.6	g
2	76	35.4	8.8	51481	13675	9.5	f
3	77	35.8	0.07	50751	136	0.0	
4	78	36.0	0.08	50667	104	0.0	
5	79	35.4	0.07	50836	125	0.0	
6	80	36.2	6.3	50206	9451	6.1	g
7	81	35.2	5.1	49003	9291	5.4	g
8	82	35.4	6.7	51611	11847	7.2	f
9	83	34.9	5.4	51359	9932	5.8	g
10	84	36.0	5.8	50196	10453	6.3	g
11	85	35.6	6.1	50335	8199	5.4	g
12	86	35.7	9.9	51075	21956	12.9	f
13	87	35.5	0.03	51067	95	0.0	
14	88	35.5	6.8	50985	10085	6.7	g
15	89	36.0	6.6	49961	9984	6.5	g
16	90	35.9	6.2	51593	8260	5.4	g
17	91	35.1	0.07	51189	126	0.0	
18	92	36.0	6.7	50443	9768	6.4	g
19	93	35.5	0.08	49642	129	0.0	
20	94	35.9	6.5	50259	9536	6.2	g

One possible reason for the difficulty of escaping local minima is that major basins of attraction are determined mainly by the phases of the low-resolution reflections. We tested this by strongly apodizing the perfect solution, running *BACK* and *FORTH* on the result and starting the solver from the result. Such a procedure is equivalent to the strong smearing of the electron density. On observation, individual features of the protein were completely obliterated. The initial distances of the smeared e voxel^{-1} files from the perfect solution were ~36 000 e ($\delta = 16$) and ~40 000 e ($\delta = 64$). The number of structure factors that were appreciable [greater than 10^{-4} of $F(000)$] and having the correct phases, were 190 and 36, respectively, out of ~3000 for the 3 Å resolution data set. The results were that all our runs with $Z = 40$ or 30 converged to the perfect solution. When the number of electrons was lowered to $Z = 20$, the solutions were not perfect, but the case with 190 correct reflections reached

Table 2. *Mutual distances among 'good' solutions in cluster 1 (electrons)*

	75	80	85	88	89	90	92	94	Average
75	0	4276	3092	5170	4952	3095	5394	4855	4405
80	4276	0	3287	2948	4930	3525	5486	2643	3870
85	3092	3287	0	4494	4620	1681	4671	4037	3698
88	5170	2948	4494	0	4860	4500	5734	3436	4449
89	4952	4930	4620	4860	0	4512	5867	4554	4899
90	3095	3525	1681	4500	4512	0	4981	3836	3733
92	5394	5486	4671	5734	5867	4981	0	5791	5418
94	4855	2643	4037	3436	4554	3836	5791	0	4164

Average = 4329.5

Table 3. *Mutual distances among 'good' solutions in cluster 2 (electrons)*

	81	83	84	Average
81	0	3402	4748	4075
83	3402	0	4521	3962
84	4748	4521	0	4635

Average = 4223.85

Table 4. *Mutual distances of 'fair' solutions (electrons)*

	82	86	76	Average
82	0	17292	16824	17058
86	17292	0	20748	19020
76	16824	20748	0	18786

Average = 18288.1

Table 5. *Perturbations of the perfect solution and subsequent optimization, Z = 40*

s.d. = standard deviation. Start and end distances are given in electrons.

Seed	Scale	Start s.d.	End s.d.	Start distance	End distance	End *R* factor
21	16	19.1	<1.1	29755	972	<1.0
22	16	19.6	0.04	30318	139	0
23	16	19.8	0.04	30410	122	0
24	16	19.4	<0.2	30173	159	0
25	16	19.1	0.04	29988	120	0

Table 6. *Perturbations of the perfect solution and subsequent optimization, Z = 20*

s.d. = standard deviation. Start and end distances are given in electrons.

Seed	Scale	Start s.d.	End s.d.	Start distance	End distance	End *R* factor
21	16	17.6	<0.7	29755	711	<0.7
22	16	18.0	1.6	30318	2946	2.2
23	16	18.3	1.0	30410	1181	1.1
24	16	17.8	<0.3	30173	199	<0.1
25	16	17.5	<0.5	29988	475	<0.5

Table 7. *Perturbations of the poor results of the optimizer started from the empty unit cell at Z = 40*

Distances are given in electrons.

Z	Seed	Scale	Start–perfect distance	End–perfect distance	Start–end distance
40	26	4	36299	37121	3927
40	26	8	37613	37167	5361
40	26	16	42017	36810	15506
40	27	16	42692	37492	16143
40	28	16	42065	37562	14083
40	29	16	43026	38063	15094
40	30	16	42121	36875	15622
50	26	16	42017	35692	14874
60	26	16	42017	34821	14964
70	26	16	42017	33931	14935
80	26	16	42017	33016	15239

within 6500 e of the perfect solution; that puts it into the 'good' category. With 36 correct phases, the distance to the perfect solution was 18 000 e, *i.e.* the solution was 'fair'.

Our experience with the MIR algorithm, applied to test cases, can be summarized as follows. For noiseless synthetic cases, the asymmetric algorithm (13) is usually better than the symmetric one (11). This may be rationalized from the comparison of Figs. 5(*a*) and (*b*): the intersection of two circles seems to be less well defined than the intersection of a circle and a straight line. We have not explored the influence of the relative weights λ_m in (13). Our exploration of the cost-function surface hints at the presence of many minima even at $Z = 100$. The number of minima seems to increase and the radii of their basins of attraction seem to decrease as Z decreases, *i.e.* as the problem gets more and more difficult to solve. Down to $Z = 43$, the empty start is within the basin of attraction of the global minimum. In spite of the straight-line theorem, which informs us that there is always a downhill path from the empty start to the global minimum, the steepest-descent path 'switches' to a different minimum from the global one when Z is lowered to 42 and below. From our studies to 'escape' the basin of attraction of local minima, we conclude that at this level of Z there are many disjoint minima with relatively high barriers among them. Additional tests run with the correct phases for a fairly small number of low-resolution reflections show that a low-resolution electron-density map is of great help for the solution of the MIR problem. This gives a lower bound on the volume of solution space encompassing the basin of attraction of the correct solution, and may indicate that the problem is still manageable. It should be emphasized that the above conclusions are valid only for our algorithm. The traditional MIR algorithm, as well as the one explored by Béran, have markedly different convergence properties.

In practice, the convergence properties of the holographic method when using heavy-atom data place limitations on how it can be used with MIR and MAD data. Unless a more effective minimization algorithm is found for this application, it will not be possible to use this method for solving MIR structures without initial phasing information. However, since the algorithm converges well once the electron density is within the basin of attraction of the correct solution, it should be possible to use conventional approaches to obtain an initial electron-density map, and then use *EDEN* to improve the maps. This possibility has been tested using real data from the protein kinesin. These results are shown in the following section.

3.4. *MIR results using data from kinesin*

To test the effectiveness of the MIR algorithms using real data, we studied the protein kinesin. Kinesin is a microtubule 'motor' protein that functions in intracellular transport and chromosome movement. The data that were used for our tests were collected from a 349-residue piece of the protein that encompasses the motor head. The structure of the kinesin head domain was solved by Kull *et al.* (1996). The original MIR maps were fairly poor, suggesting that they may be improved using the holographic method.

Native data to 1.8 Å were available for this protein, as well as data collected from two derivatives, one containing one I atom and one containing three Hg atoms. The data for each derivative extended to 2.5 Å.

As described in the previous section (§3.3), the *EDEN* implementation of the MIR algorithm suffers from convergence problems if the starting phases are too far from the correct solution. To circumvent this problem, we started from the *MLPHARE* estimate of the phases, assuming that these will place the corresponding electron density within the radius of convergence of our algorithm, and that running *EDEN* will result in an improved map.

The native data were placed on an absolute scale with a Wilson plot program from *CCP*4 using data between 3.0 and 1.8 Å resolution. The utility *APODFO*, which is part of the *EDEN* package, gave scale factors that agreed with those from *CCP*4 to within 3%. The derivatives were placed on an absolute scale by scaling them to the native data set. This was done by apodizing all three data sets to 3.0 Å resolution (and, later, to 2 Å resolution) using *APODFO* and visually aligning the corresponding Wilson-like plots obtained. *EDEN* also requires an estimate of the total number of electrons in the unit cell. We used the formula $F(0,0,0) \simeq (V + N_p)/3$, where V is the volume of the unit cell and N_p is the number of protein electrons in it.

Our first step was to check the occupancies and positions of the heavy atoms. To do this, we worked at a resolution of 3.0 Å. *BACK* was run on the initial MIR phase set to prepare the corresponding electron-density map. *FORTH* was then run to obtain a consistent calculated phase set. The average phase change between the results of the latter and the starting set from *MLPHARE* was 17°. In our experience, such a phase difference is negligible. *SOLVE* was run at 3.0 Å resolution in correction mode and the resulting electron-density maps were visually inspected to see if there were either peaks or holes at the heavy-atom positions. Ideally, there should be no evidence of the heavy atoms in the resulting native electron density. A MIR run that produces a native electron density with the heavy atoms showing through may be the consequence of one of two errors: too low occupancy of the heavy atoms in the derivative or too high a scaling of the derivative with respect to the native. Similarly, if there are holes at the heavy-atom positions, their occupancy may be too high or the scaling of the derivative may be too low with respect to the native. By repeatedly running *EDEN* and inspecting the results, we adjusted the occupancies of the heavy atoms and made slight adjustments to the relative scaling of the derivative and the native data sets.

Preliminary *SOLVE* runs (in MIR correction mode) were done at 3.0 Å resolution. The results were quite encouraging. Before continuing the main MIR solution process, we decided to investigate the isomorphism of the derivatives. In order to do that, we started from the MIR map and ran it in correction mode against the measured structure factors of the native alone. This way the program is not constrained by any of the derivatives. We found that using the starting electron density as a very mild target (using $\lambda_{\rm space} = 0.0003$) prevented the program from straying too far from the original MIR map except where there were real differences. The same procedure was done with each of the derivatives. Pairwise comparisons of the results should reveal lack of isomorphism and local distortions around the heavy atoms. We found that, within our ability to detect differences, the two derivatives of kinesin were isomorphous with the native.

The next step was to obtain an estimate of the solvent envelope. This was done by apodizing the output of the previous 3.0 Å MIR run to 7.0 Å. *BACK* was used to produce the corresponding (smeared) map (on the grid of the 3.0 Å run). The *EDEN* utility *MAKETAR* was used to select the 50% of the grid points with lowest electron density. These were used as the solvent region, and assigned a target electron density of 0.33 e Å^{-3}.

Two full *SOLVE* runs (in MIR correction mode) were done at 3.0 Å resolution using the solvent target and the two derivatives with $\lambda_{\rm space} = 0.003$ and 0.01. The results were very encouraging, and we used the same solvent target to do a full *SOLVE* run (in MIR mode) at 2.0 Å resolution. The resulting electron-density map was compared with that obtained from the original phases derived from *MLPHARE* and with a *DM* modified map (Cowtan & Main, 1993) (Fig. 7). The fully refined kinesin structure was used as a guide for comparing the maps. The *EDEN* map was comparable to the *DM* map everywhere and in some places it was clearly better.

4. Non-crystallographic symmetry

Non-crystallographic symmetry (NCS) is treated in a manner similar to previous sections. In particular, we use a real-space cost function that 'encourages' the symmetry but does not enforce it. Although our method has similarities to well established and successful methods of NCS (Bricogne, 1974; Rossmann *et al.*, 1992; Tsao, Chapman & Rossmann, 1992; Chapman, Tsao & Rossmann, 1992; Zhang, 1993; Cowtan & Main, 1993; *The CCP*4 *Suite*, 1994; Chapman, 1995), there are also differences. Some of these differences are advantageous, at least in theory. First, the exact knowledge of the molecular envelope is not critical. Second, the non-crystallographic constraint is 'soft' and its strength can be varied. Third, we do not interpolate in reciprocal space; instead, we use an expansion into basis functions in physical space. However, this is not an important distinction from other methods. These properties of the method allow the determination of the goodness of the symmetry from the data alone. One should also be able to find out if there are differences in the monomers that are related by non-crystallographic symmetry. The main disadvantage of the method is that it uses basis functions on a grid and therefore it has limited accuracy.

4.1. *Derivation of the equations for NCS*

Non-crystallographic symmetry is characterized by the presence of $N_{\rm NCS}$ monomers within each asymmetric unit of a crystal that have approximately the same electron densities. Our derivation below will first describe a single asymmetric unit and exact NCS. The approximate shapes of the monomers will be described by a weight function, $\tilde{w}_k(\mathbf{r})$, $k = 1, \ldots, N_{\rm NCS}$. The weight function will usually be $\tilde{w}_k(\mathbf{r}) = 1$ if $\mathbf{r}$ is inside the kth monomer and zero if $\mathbf{r}$ is outside it. Let us denote the electron density of the crystal by $\rho(\mathbf{r})$. The expression $\tilde{w}_1(\mathbf{r})\rho(\mathbf{r})$ selects the first monomer [meaning that $\tilde{w}_1(\mathbf{r})\rho(\mathbf{r}) = \rho(\mathbf{r})$ if $\mathbf{r}$ is within the first monomer and $\tilde{w}_1(\mathbf{r})\rho(\mathbf{r}) = 0$ if $\mathbf{r}$ is outside the first monomer.] The kth monomer is related to the first monomer by the coordinate transformation Ω_{k1}. The meaning of this statement is that both the masks and the electron densities are equal at NCS related points:

$$\rho(\Omega_{k1}\mathbf{r}) = \rho(\mathbf{r}) \quad \text{when } \mathbf{r} \text{ is in the first monomer,} \tag{14}$$

$$\tilde{w}_k(\Omega_{k1}\mathbf{r}) = \tilde{w}_1(\mathbf{r}) \quad \text{when } \mathbf{r} \text{ is in the first monomer.} \tag{15}$$

The transformations Ω_{k1} consist of a Cartesian rotation and a translation. In the notation of paper II, they are

$$\mathbf{r}_k = \Omega_{k1}\mathbf{r} = \mathcal{R}_k\mathbf{r} + \tau_k. \tag{16}$$

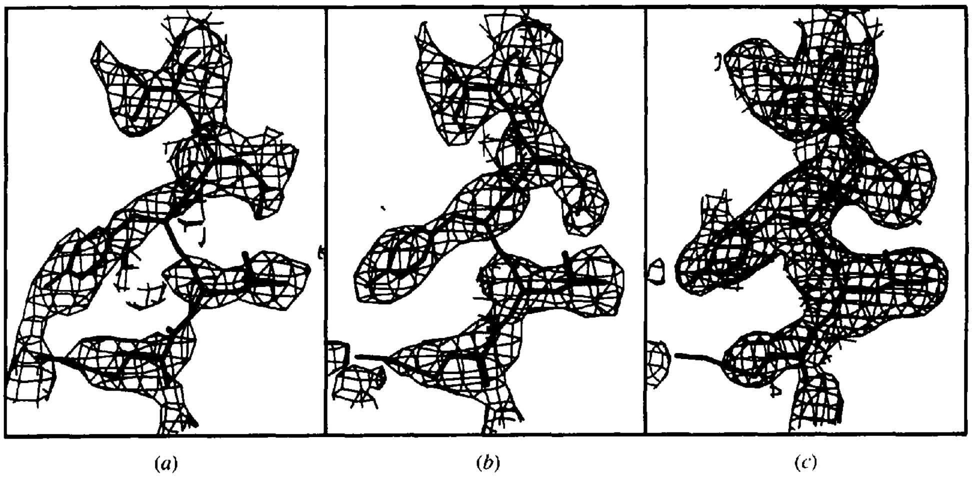


Fig. 7. The electron density corresponding to residues 226–230 of kinesin is shown. In (*a*), the electron density shown is the direct result of phasing with *MLPHARE*. In (*b*), the density has been modified from (*a*) using the program *DM*. In (*c*), the density was modified from (*a*) using *EDEN*.

Note that $\mathbf{r}$ are Cartesian coordinates and $\tilde{w}_k(\mathbf{r}_k) \neq 0$ if and only if $\tilde{w}_1(\mathbf{r}) \neq 0$. The transformations Ω_{k1} do not necessarily form a group, but their inverses are well defined and they satisfy $\Omega_{1k} = \Omega_{k1}^{-1}$. In general, any monomer can be transformed to any other one by the transformation

$$\mathbf{r}_k = \Omega_{kk'}\mathbf{r}_{k'} = \Omega_{k1}\Omega_{1k'}\mathbf{r}_{k'} = \mathcal{R}_k[\mathcal{R}_{k'}^T(\mathbf{r}_{k'} - \boldsymbol{\tau}_{k'})] + \boldsymbol{\tau}_k, \qquad (17)$$

where $\mathcal{R}^T = \mathcal{R}^{-1}$ is the inverse of the (Cartesian) rotation $\mathcal{R}$. We will also demand that the masks do not overlap,

$$\tilde{w}_j(\mathbf{r})\tilde{w}_k(\mathbf{r}) = 0 \quad \text{for} \quad j \neq k, \qquad (18)$$

where j, k now include all asymmetric units.

In the rest of our derivation, we do not demand either exact NCS or an accurate knowledge of the shape of the monomers. Let us define a total mask and a NCS averaged density,

$$\tilde{w}(\mathbf{r}) = \sum_k \tilde{w}_k(\mathbf{r}) \qquad (19)$$

$$\langle\rho(\mathbf{r})\rangle = [1/N_{\rm NCS}] \sum_k \tilde{w}_k(\mathbf{r})\rho(\Omega_{k'k}\mathbf{r}) \qquad (20)$$

for arbitrary k'. In order to 'encourage' but not enforce non-crystallographic symmetry, we will define a NCS cost function,

$$f_{\rm NCS} = \tfrac{1}{2}\lambda_{\rm NCS}N_{\rm CS}V \int_{\rm Asym} \tilde{w}(\mathbf{r})[\rho(\mathbf{r}) - \langle\rho(\mathbf{r})\rangle]^2\,{\rm d}\mathbf{r}, \qquad (21)$$

where $N_{\rm CS}$ is the number of asymmetric units in the unit cell, V is the volume of the unit cell and the integration is restricted to a single asymmetric unit. A formal expansion yields the equivalent formula

$$f_{\rm NCS} = \tfrac{1}{2}\lambda_{\rm NCS}N_{\rm CS}V\left\{\sum_k \int_{\rm Asym} \tilde{w}_k(\mathbf{r})[\rho(\mathbf{r})]^2\,{\rm d}\mathbf{r} - (1/N_{\rm NCS})\sum_k\sum_{k'} \int_{\rm Asym} \tilde{w}_k(\mathbf{r})\rho(\mathbf{r})\rho(\Omega_{k'k}\mathbf{r})\,{\rm d}\mathbf{r}\right\}. \qquad (22)$$

Note that the presence of $\tilde{w}_k(\mathbf{r})$ in the integrals effectively restricts $\mathbf{r}$ to be within the kth monomer.

The above formulation generalizes known symmetry-averaging methods just as our solvent targets generalize known solvent-flattening methods in crystallography. The basis-function expansion of the electron density (2*b*) can be substituted explicitly into (22) and the integrals are carried out in Appendix *A*3.

The NCS algorithm has been implemented in *EDEN*, but it has been tested only on very simple test problems. It is not available in the released version, 2.5.

5. Summary and discussion

In this paper, we have shown that MIR, MAD and NCS information can be incorporated into the holographic method. Using simulated heavy-atom data, we have explored in some detail the convergence properties of our algorithm and the uniqueness of the solution it supplies. These simulations show that the holographic method does not converge as well as traditional reciprocal-space methods, even though the equations are mathematically equivalent. However, once the electron density is within the radius of convergence of the correct minimum, the holographic method quickly and accurately finds the correct structure. Given these findings, we propose that conventional methods should be used to identify an initial MIR solution and that the holographic method should then be able to improve that solution. We have made use of this strategy to determine the structure of the protein kinesin, using experimental MIR data. An initial structure of kinesin was identified using the program *MLPHARE*. Using *EDEN* to optimize this solution led to a clear improvement in the resulting electron-density maps.

The holographic method can be placed in a more global framework. It has been firmly established that the crystal structure of a macromolecule cannot be determined solely from the knowledge of its structure-factor amplitudes (see *e.g.* Bricogne, 1992). However, it seems intuitively clear that a unique structure can be reached if enough additional information is added and if there is an effective way of using that information. We argue that the holographic method provides an excellent framework for the incorporation of additional information into the determination of crystal structures.

In its simplest form, the holographic method was used in cases where a partial structure was known, and it was used to complete the structure. In this implementation, if there are no external constraints, the electron-density maps obtained using the holographic method are very similar to traditional $F_o - F_c$ and $2F_o - F_c$ maps. Traditional Fourier maps are actually marginally more accurate because the holographic method is limited in its accuracy by the (incomplete) basis-function expansion. However, if there are known constraints that must be satisfied by the electron density, the holographic method is able to use that information to recover electrons more accurately than traditional Fourier methods.

The fact that the electron density is always positive is an important constraint; positivity is always enforced in *EDEN*. In addition, often the electron density is known in part of the unit cell, either because the solvent region is known or because a partial structure has been placed in the unit cell. *EDEN* is able to use the localized nature of the known electron density in real space: it can constrain it in some part of the unit cell and not in other parts. It can also use the known electron density as a mild constraint. Therefore, errors in the 'known' part can be both detected and corrected. In the future, we plan to use the holographic method to incorporate chemical constraints into structure determinations (*e.g.* atomicity and connectivity).

On the theoretical side, we scrupulously differentiate between lack of information and tacitly assumed information. For example, we consistently avoid the use of Fourier back transforms. In usual practice, unknown structure factors are given zero value as opposed to keeping them unknown. Similarly, in the presence of non-crystallographic symmetry, some formulations implicitly assume that the electron density is featureless outside the symmetry-related regions. We try to live by Lánczos's dictum: use all the available information and no more. In principle, given a sufficient amount of information, it is possible to recover the crystal perfectly. However, different algorithms may have very different convergence properties and may have very different sensitivity to imperfections in the data. In our opinion, this last point alone is sufficiently important to justify the development of new methods for crystallographic computations.

EDEN can be obtained free of charge by qualified collaborators. Please contact HS by e-mail at szoke2@llnl.gov for details.

We thank Jon Kull and Robert Fletterick of UCSF who provided us with their data on kinesin and helped us to evaluate our results. HS and AS thank Lawrence Livermore National Laboratory for providing them with their physical and computational resources. We thank especially Rick Twogood, Jim Brase and Dennis Goodman for their support and help. AS also thanks Mark Eckart, Richard More and the Physics Department for their support.

APPENDIX *A*

The suite of computer programs *EDEN* consists of the preprocessors *APODFO*, *APODFC*, *EXPANDFO*, *EXPANDFC*, *BACK*, *FORTH*, *MAKETAR*, *SYM*, the main program *SOLVE*, the post processor *REGRID* and evaluation utilities *DPHASE* and *DRHO*. Some of these were described in paper IV. Here we collect some of the 'working equations' that are actually used in *EDEN* Version 2.5 and give some further description of these utilities. *A*1–*A*4 are concerned with *SOLVE* and *A*5 with the other programs.

*A*1. Cost function and gradient for a quadratic algorithm including MIR and MAD

The electron density of the native macromolecule is expressed in terms of the real array $\mathbf{n} = \{n(p)\}$. The quantities $n(p)$ are equal to the total number of electrons in each Gaussian blob that are centered on the lattice points $\mathbf{r}_p$ and have an intrinsic resolution $d_{\rm res}$. The lattice points are obtained by dividing the crystal axes, **a**, **b**, **c**, into P_a, P_b, P_c equal parts, respectively. The natural numbers P_a, P_b, P_c are chosen so that

$$|\mathbf{a}|/P_a \simeq |\mathbf{b}|/P_b \simeq |\mathbf{c}|/P_c \simeq \Delta r. \qquad (23)$$

The widths of the Gaussians in (2*b*) are $\eta(\Delta r)^2$. In our input file, the 'input resolution' $d_{\rm inp}$ must be specified. By default, we then set $\Delta r = 0.6 d_{\rm inp}$ and $\eta = 1$ for a simple lattice and $\Delta r = 0.7 d_{\rm inp}$, $\eta = 0.75$ for a body-centered lattice. The user can also specify a non-default value of η. [In the notation of (2*a*), the widths of the Gaussians are $d^2_{\rm res}/\pi^2$. From the relation $d_{\rm res} = 1.885 d_{\rm inp}$, the default value for the grid spacing in a simple lattice is $\Delta r = d_{\rm res}/\pi$ and $\eta = 1$, while in a body-centered lattice the default values are $\Delta r = 4d_{\rm res}/3\pi$ and $\eta = 0.75$.] The numbers P_a, P_b, P_c also have to factorize into small primes; in the current version of *EDEN*, they have to be multiples of 2, 3 and 5. In addition, they have to be divisible by the orders of all the screw axes of the space group; *e.g.*, for a $P4_1$ space group, P_c has to be divisible by 4. In realistic cases, it is possible to choose them so that the resolutions along the three crystal axes do not differ by more than 10%. The lattice points are represented by

$$\mathbf{r}_p = (p_a/P_a)\mathbf{a} + (p_b/P_b)\mathbf{b} + (p_c/P_c)\mathbf{c},$$
$$0 \le p_a < P_a, \quad 0 \le p_b < P_b, \quad 0 \le p_c < P_c. \qquad (24)$$

Note that $\{(p_a/P_a), (p_b/P_b), (p_c/P_c)\}$ are the fractional coordinates corresponding to the point $\mathbf{r}_p$. Such a 'simple' lattice is used for crystals when one or more of the angles α, β, γ among the crystal axes significantly differs from 90°. For crystals whose angles α, β, γ are close to 90°, a body-centered lattice is used. Body-centered lattices consist of two parts, which we denote the simple lattice and the intercalating lattice. Hexagonal close-packed lattices, suitable for trigonal and hexagonal crystals, are not yet implemented.

The arrays that are needed for the calculation of $O(\mathbf{h})$ will be defined almost the same way as in paper IV. A slight generalization will be incorporated in order to allow for several derivatives that have different effective crystallographic B values. As in §3.1, we assume that there are $M + 1$ sets of independent measurements, *e.g.* M derivatives and the native. The grid spacing and the value of η determine the effective resolution of the gridded Gaussians, as described above. If the resolution of the measurements is higher, the programs *APODFC* and *APODFO* calculate the proper value of the apodization of $R(\mathbf{h})$, $F(\mathbf{h})$, (6*a*,*b*). If the mth derivative has a lower intrinsic resolution, $d_{\rm eff}$, that would correspond to $\delta_m \le$

0, the appropriate value of η_m to use with this derivative is $\eta_m = (d_{\rm res}/d_{\rm eff})\eta$. We define

$$\mathrm{efac}_m(\mathbf{h}) = \exp[-\eta_m(\pi\Delta r|\mathcal{F}^T\mathbf{h}|)^2]\exp[2\pi i\,\mathrm{off}(\mathbf{h})]. \tag{25}$$

This generalization is not yet incorporated into *EDEN*; we use $\eta_m = \eta$ everywhere. The offset array, off($\mathbf{h}$), is 0 for a simple grid and for the simple part of a body-centered grid. For the intercalating part of a body-centered grid,

$$\mathrm{off}(\mathbf{h}) = 1/2(h/P_a + k/P_b + l/P_c). \tag{26}$$

The factor $|\mathcal{F}^T\mathbf{h}|^2 = 1/d(\mathbf{h})^2$ is the reciprocal square of the interplane spacing for the reciprocal-lattice vector $\mathbf{h}$. It is given by

$$\begin{aligned}|\mathcal{F}^T\mathbf{h}|^2 = (1/A^2)[&(h^2/a^2)\sin^2\alpha + (k^2/b^2)\sin^2\beta\\ &+ (l^2/c^2)\sin^2\gamma\\ &+ 2(kl/bc)(\cos\beta\cos\gamma - \cos\alpha)\\ &+ 2(lh/ca)(\cos\gamma\cos\alpha - \cos\beta)\\ &+ 2(hk/ab)(\cos\alpha\cos\beta - \cos\gamma)],\end{aligned} \tag{27}$$

$$\begin{aligned}A^2 = 1 - \cos^2\alpha - \cos^2\beta - \cos^2\gamma\\ + 2\cos\alpha\cos\beta\cos\gamma.\end{aligned} \tag{28}$$

Using the arrays in (25) and (26), the cost function can be calculated by FFT as follows. From (3),

$$O_m(\mathbf{h}) = \mathrm{efac}_m(\mathbf{h})\cdot \mathrm{DFT}^+(\mathbf{n}). \tag{29}$$

This makes the calculation of $f_{\rm eden}$ in (11) possible by a fast Fourier transform, a scalar multiplication and the extraction of a square root. For completeness, we repeat here the generalization of the 'symmetric' algorithm, equation (11):

$$\begin{aligned}f_{\rm eden} = \tfrac{1}{2}\sum_{m=0}^{M}\lambda_m\sum_{\mathbf{h}} w'_m(\mathbf{h})^2[&|R'_m(\mathbf{h}) + O_m(\mathbf{h})|\\ &- |F'_m(\mathbf{h})|]^2.\end{aligned} \tag{30}$$

If there is only the native, $M = 0$. The gradient of the cost function, $g_{\rm eden}$, can be calculated as

$$\begin{aligned}g_{\rm eden}(n) = 2\sum_{m=0}^{M}\lambda_m\mathrm{Re}\Big[\mathrm{DFT}^-\Big(\{&w'_m(\mathbf{h})^2[|R'_m(\mathbf{h})\\ &+ O_m(\mathbf{h})| - |F'_m(\mathbf{h})|]/2|R'_m(\mathbf{h}) + O_m(\mathbf{h})|\}\\ &\times [R'_m(\mathbf{h}) + O_m(\mathbf{h})]\mathrm{efac}^*(\mathbf{h})\Big)\Big];\end{aligned} \tag{31}$$

see also equation (A10) of paper IV.

The derivation of the asymmetric algorithm for MIR will be presented in a slightly more general form than in §3.1. It allows for the possibility that the native and the various derivatives have different resolutions. If the intrinsic resolutions of the data sets are characterized by η_m, the proper $O_m(\mathbf{h})$ for each of them, in parallel with (3), is

$$\begin{aligned}O_m(\mathbf{h}) = &\exp[-\eta_m(\pi\Delta r|\mathcal{F}^T\mathbf{h}|)^2]\\ &\times\sum_{p=1}^{P} n(p)\exp(2\pi i\mathbf{h}\cdot\mathcal{F}\mathbf{r}_p).\end{aligned} \tag{32}$$

Let us define, for $m = 0, \ldots, M$,

$$\alpha_m(\mathbf{h}) = \exp[-\eta_m(\pi\Delta r|\mathcal{F}^T\mathbf{h}|)^2]/\exp[-\eta_0(\pi\Delta r|\mathcal{F}^T\mathbf{h}|)^2]. \tag{33}$$

Obviously, $\alpha_0(\mathbf{h}) = 1$ and the notation $O_0(\mathbf{h})$ for $O(\mathbf{h})$ is superfluous. Usually, the native has the highest intrinsic resolution (the lowest B value) and therefore $\alpha_m(\mathbf{h}) \leq 1$. Equation (10) can then be generalized to

$$\begin{aligned}|F_m(\mathbf{h})|^2 &= |R_m(\mathbf{h}) + O_m(\mathbf{h})|^2\\ &= |R_m(\mathbf{h})|^2 + R_m(\mathbf{h})O_m^*(\mathbf{h}) + R_m^*(\mathbf{h})O_m(\mathbf{h})\\ &\quad + |O_m(\mathbf{h})|^2\\ &= |R_m(\mathbf{h})|^2 + \alpha_m(\mathbf{h})[R_m(\mathbf{h})O_0^*(\mathbf{h})\\ &\quad + R_m^*(\mathbf{h})O_0(\mathbf{h})] + \alpha_m(\mathbf{h})^2|O_0(\mathbf{h})|^2.\end{aligned} \tag{34}$$

Subtracting the equation for the native from each derivative, as in (12), we get

$$\begin{aligned}|F_0(\mathbf{h})|^2 = |R_0(\mathbf{h})|^2 + R_0(\mathbf{h})O_0^*(\mathbf{h}) + R_0^*(\mathbf{h})O_0(\mathbf{h})\\ + |O_0(\mathbf{h})|^2,\end{aligned}$$

$$\begin{aligned}|F_m(\mathbf{h})|^2 &- \alpha_m(\mathbf{h})^2|F_0(\mathbf{h})|^2\\ &= |R_m(\mathbf{h})|^2 - \alpha_m(\mathbf{h})^2|R_0(\mathbf{h})|^2\\ &\quad + \alpha_m(\mathbf{h})\{[R_m(\mathbf{h}) - \alpha_m(\mathbf{h})R_0(\mathbf{h})]O_0^*(\mathbf{h})\\ &\quad + [R_m^*(\mathbf{h}) - \alpha_m(\mathbf{h})R_0^*(\mathbf{h})]O_0(\mathbf{h})\},\\ &\qquad\qquad m = 1, \ldots, M.\end{aligned} \tag{35}$$

The cost function for the asymmetric algorithm can be written as the generalization of (13), with the proper stabilizing weights:

$$\begin{aligned}f_{\rm eden} = \tfrac{1}{2}\Big\{&\lambda_0\sum_{\mathbf{h}} w'_0(\mathbf{h})^2[|R'_0(\mathbf{h}) + O_0(\mathbf{h})| - |F'_0(\mathbf{h})|]^2\\ &+ \sum_{m=1}^{M}\lambda_m\sum_{\mathbf{h}} w'_m(\mathbf{h})^2[\alpha_m(\mathbf{h})U_m(\mathbf{h})O_0^*(\mathbf{h})\\ &+ \alpha_m(\mathbf{h})U_m^*(\mathbf{h})O_0(\mathbf{h}) - H_m(\mathbf{h})]^2\Big\},\end{aligned} \tag{36}$$

where

$$U_m(\mathbf{h}) = [R_m(\mathbf{h}) - \alpha_m(\mathbf{h})R_0(\mathbf{h})]/[|R_m(\mathbf{h}) - \alpha_m(\mathbf{h})R_0(\mathbf{h})|], \tag{37}$$

$$H_m(\mathbf{h}) = [|F_m(\mathbf{h})|^2 - \alpha_m(\mathbf{h})^2|F_0(\mathbf{h})|^2 - |R_m(\mathbf{h})|^2 + \alpha_m(\mathbf{h})^2|R_0(\mathbf{h})|^2]/[|R_m(\mathbf{h}) - \alpha_m(\mathbf{h})R_0(\mathbf{h})|], \tag{38}$$

$$w'_m(\mathbf{h}) = [|R_m(\mathbf{h}) - \alpha_m(\mathbf{h})R_0(\mathbf{h})|^2]/[|R_m(\mathbf{h}) - \alpha_m(\mathbf{h})R_0(\mathbf{h})|^2 + H_m(\mathbf{h})^2]. \tag{39}$$

The $m = 0$ term is the same as in (30). The factors $U_m(\mathbf{h})$ are of unit magnitude; if their denominators are small, the proper limits of $U_m(\mathbf{h})$, $H_m(\mathbf{h})$ and $w'_m(\mathbf{h})$ are used in *EDEN*. The gradient can also be calculated in a straightforward way. It is

$$\begin{aligned} g_{\text{eden}}(\mathbf{n}) = 2\text{Re}\Big[&\lambda_0\text{DFT}^-\Big(\{w'(\mathbf{h})^2[|R'_0(\mathbf{h}) + O_0(\mathbf{h})| \\ &- |F'_0(\mathbf{h})|]/2|R'_0(\mathbf{h}) + O_0(\mathbf{h})|\} \\ &\times [R'_0(\mathbf{h}) + O_0(\mathbf{h})]\text{efac}^*_0(\mathbf{h})\Big) \\ &+ \sum_{m=1}^{M} \lambda_m\text{DFT}^-\{w'_m(\mathbf{h})^2U_m(\mathbf{h}) \\ &\times [\alpha_m(\mathbf{h})U_m(\mathbf{h})O^*_0(\mathbf{h}) \\ &+ \alpha_m(\mathbf{h})U^*_m(\mathbf{h})O_0(\mathbf{h}) - H_m(\mathbf{h})]\text{efac}^*_m(\mathbf{h})\}\Big]. \end{aligned} \tag{40}$$

*A*2. Cost function for molecular replacement and solvent flattening

The target density is described by the set $\{n(p)_{\text{target}}\}$ and the weight array, $\tilde{w}^2_p$. The cost function, (7), is

$$f_{\text{space}} = \tfrac{1}{2}\lambda_{\text{space}}P\sum_{p=1}^{P} \tilde{w}^2_p[n(p) - n(p)_{\text{target}}]^2 \tag{41}$$

and the $n(p)$ component of its gradient is

$$g_{\text{space}}[n(p)] = \lambda_{\text{space}}P\tilde{w}^2_p[n(p) - n(p)_{\text{target}}]. \tag{42}$$

*A*3. Derivation of cost function for non-crystallographic symmetry

The cost function (22) is written for a general continuous density. *EDEN* is written in terms of the basis-function expansion (2*b*), which, if substituted into (22), yields a rather complicated expression

$$\begin{aligned} f_{\text{NCS}} = {} & [\lambda_{\text{NCS}}N_{\text{CS}}V/2(\pi\eta\Delta r^2)^3] \\ & \Big(\sum_k \int_{\text{Asym}} w_k(\mathbf{r})\Big\{\sum_{(p,k)} n(p,k) \\ & \times \exp[-|\mathbf{r} - \mathbf{r}(p,k)|^2/\eta\Delta r^2]\Big\}\Big\{\sum_{(p',k)} n(p',k) \\ & \times \exp[-|\mathbf{r} - \mathbf{r}(p',k)|^2/\eta\Delta r^2]^2\,\mathrm{d}\mathbf{r}\Big\} \\ & - (1/N_{\text{NCS}})\sum_k\sum_{k'}\int_{\text{Asym}}\Big\{w_k(\mathbf{r})\sum_{(p,k)} n(p,k) \\ & \times \exp[-|\mathbf{r} - \mathbf{r}(p,k)|^2/\eta\Delta r^2]\Big\}\Big\{\sum_{(p',k')} n(p',k') \\ & \times \exp[-|\Omega_{k'k}\mathbf{r} - \mathbf{r}(p',k')|^2/\eta\Delta r^2]\Big\}\Big)\,\mathrm{d}\mathbf{r}. \end{aligned} \tag{43}$$

In (43), the function $\tilde{w}_k(\mathbf{r})$ restricts r to be within monomer k. The notation (p,k) denotes a lattice point p within the monomer k. Similarly, $\Omega_{k'k}r$ is within monomer k' and (p',k') denotes a lattice point in that monomer. The summation over (p,k) means a summation over all lattice points within monomer k (and not a summation over the monomers).

The Gaussian integrals can be carried out explicitly if and only if $\tilde{w}_k(\mathbf{r})$ is approximately constant within the region around $r(p,k)$, where the appropriate Gaussian basis function, $\exp[-|\mathbf{r} - \mathbf{r}(p,k)|^2/\eta\Delta r^2]$, is appreciably different from zero. The overlap integrals in the first term of (43) are

$$\begin{aligned} \int_{\text{Asym}} & w_k(\mathbf{r})\exp\{-[|\mathbf{r} - \mathbf{r}(p,k)|^2 \\ & + |\mathbf{r} - \mathbf{r}(p',k)|^2]/\eta\Delta r^2\}\,\mathrm{d}\mathbf{r} \\ & = (\pi\eta\Delta r^2/2)^{3/2}w_k[\mathbf{r}(p,k)] \\ & \quad \times \exp\{-[|\mathbf{r}(p,k) - \mathbf{r}(p',k)|^2/2\eta\Delta r^2]\}. \end{aligned} \tag{44}$$

If the integrals in the second term are transformed by the equality

$$|\Omega_{k'k}\mathbf{r} - \mathbf{r}(p',k')|^2 = |\mathbf{r} - \Omega_{kk'}\mathbf{r}(p',k')|^2, \tag{45}$$

they become similar to those in the first term of (43). After collecting all the terms and using (45) again, we get

$$\begin{aligned} f_{\text{NCS}} = {} & [\lambda_{\text{NCS}}N_{\text{CS}}V/2(2\pi\eta\Delta r^2)^{3/2}] \\ & \times \sum_k\sum_{(p,k)} w_k[\mathbf{r}(p,k)]n(p,k)\Big(\sum_{(p',k)} n(p',k) \\ & \times \exp\{-[|\mathbf{r}(p,k) - \mathbf{r}(p',k)|^2]/2\eta\Delta r^2\} \\ & - (1/N_{\text{NCS}})\sum_{k'}\sum_{(p',k')} n(p',k') \\ & \times \exp\{-[|\Omega_{k'k}\mathbf{r}(p,k) - \mathbf{r}(p',k')|^2]/2\eta\Delta r^2\}\Big). \end{aligned} \tag{46}$$

When $r(p,k)$ is on the periphery of monomer k, $\tilde{w}_k(\mathbf{r})$ is not constant over the region of integration. One remedy is to taper off $\tilde{w}_k(\mathbf{r})$ slowly on a scale of $d_{\rm inp}$. For the actual usage of (46), we note that all the exponentials can be precalculated as they do not depend on $n(p)$. Also, for any given lattice point (p,k), there are only a few other lattice points, (p',k'), for which the exponentials appearing in (46) are appreciable. This makes the calculation of the cost function an order P, rather than an order P^2, calculation. The gradient is very simple:

$$\begin{aligned}\partial f_{\rm NCS}/\partial n(p,k) = {}& [\lambda_{\rm NCS}N_{\rm CS}V/(2\pi\eta\Delta r^2)^{3/2}]w_k[\mathbf{r}(p,k)] \\ & \times\Bigg(\sum_{(p',k)} n(p',k)\exp\{-[\|\mathbf{r}(p,k) \\ & -\mathbf{r}(p',k)|^2]/2\eta\Delta r^2\} - (1/N_{\rm NCS}) \\ & \times\sum_{k'}\sum_{(p',k')} n(p',k')\exp\{-[\|\Omega_{k'k}\mathbf{r}(p,k) \\ & -\mathbf{r}(p',k')|^2]/2\eta\Delta r^2\}\Bigg). \qquad (47)\end{aligned}$$

When there is crystallographic symmetry, the sums in (46), (47) actually go over the whole unit cell.

A4. Cost function for crystal symmetry

When there is crystal symmetry, the grid spacing is always chosen so that all coordinates that are symmetry related to a grid point also fall on grid points. The structure factors and the starting density $\{n(p)\}$ are expanded to the whole unit cell, *i.e.* to $P1$ symmetry. As long as the gradient of the cost function is large, crystal symmetry is maintained automatically by the solver. When the magnitude of the gradient gets small, the program has a tendency to violate crystal symmetry. This is expected and used for diagnostic purposes. At the end of the loop, symmetry-related points are averaged and the corresponding structure factors are calculated. It is interesting to speculate that relaxing crystalline symmetry may make the surface of the cost function smoother and eliminate the separation of some minima.

Crystallographic symmetry can also be maintained during the whole minimization procedure by adding (yet) another cost function. It is easy to see that the complicated equations given above, which describe non-crystallographic symmetry, are satisfied if we ensure that the number of electrons is equal in all symmetry-related grid points. Accordingly, we can define a crystal-symmetry-related average (gridded) 'density'

$$\langle n(p)\rangle = (1/N_{\rm CS})\sum_m n(p(m)), \qquad (48)$$

where the index m runs over the crystal symmetry group of $N_{\rm CS}$ elements and $p(m)$ denotes symmetry-related grid points. Crystal symmetry is established by minimizing the cost function

$$f_{\rm CS} = \tfrac{1}{2}\lambda_{\rm CS}P\sum_{p=1}^{P}\{n(p) - \langle n(p)\rangle\}^2 \qquad (49)$$

and the $n(p)$ component of its gradient is

$$g_{\rm CS}[n(p)] = \lambda_{\rm CS}P\{n(p) - \langle n(p)\rangle\}. \qquad (50)$$

As the measured structure factors, the calculated ones for the reference and the known electron densities all obey crystal symmetry, in principle, only numerical instabilities cause the solution to deviate from it.

A5. Ancillary programs: preprocessors, post processors

A5.1. *APODFO and APODFC*

The two apodization programs, *APODFO* and *APODFC*, carry out an analysis of the structure-factor data that is similar to a Wilson plot. They are used for deriving $F'(\mathbf{h})$ from $F(\mathbf{h})$ and $R'(\mathbf{h})$ from $R(\mathbf{h})$, (6), and for determining the scale factor that places the observed structure-factor amplitudes, $|F'(\mathbf{h})|_{\rm obs}$, on an absolute scale.

APODFO reads structure factors from an input $F_{\rm obs}$ file, while *APODFC* reads structure factors from an input $F_{\rm calc}$ file. Each one generates a set of data points that are mean values of $\ln(|F|^2)$ within shells of equal thickness in a space of $1/d^2$, where $|F|$ stands for $|F|_{\rm obs}$ or $|F|_{\rm calc}$ and

$$1/d^2 = (h^2/a^2) + (k^2/b^2) + (l^2/c^2) \qquad (51)$$

or its generalized form for non-orthogonal crystals, (27).

Each program then finds the slope of the set of data points by fitting a straight line to the data between appropriate resolution limits, given in the input file. The slope is equivalent to an average crystallographic B factor. The program also reports the recommended smearing factor, $\delta_{f\rm obs}$ or $\delta_{f\rm calc}$ in (6) and the y intercept ($y0_{\rm obs}$ or $y0_{\rm calc}$), which can be used for scaling the experimental data. Normally, the programs write out apodized versions of their input using the appropriate $\delta_{f\rm obs}$ or $\delta_{f\rm calc}$. Therefore, $\delta_{f\rm obs}$ or $\delta_{f\rm calc}$ will not need to be used explicitly in the solver.

A5.2. *EXPANDFO and EXPANDFC*

The preprocessors *EXPANDFO* and *EXPANDFC* expand $F_{\rm obs}$ and $F_{\rm calc}$ files to $P1$ symmetry from the set of unique reflections. If there is anomalous scattering, only the symmetry operators appropriate to the space group are used. If there is no anomalous scattering, Friedel's relations are used in addition to the symmetry operators. Unless the crystal symmetry is $P1$ to begin with and anomalous scattering is present, *EDEN* uses only the $h \geq 0$ half ellipsoid.

A5.3. *BACK*

The preprocessor program called *BACK* produces a vector, $n(p)$, of size $P = P_aP_bP_c$, corresponding to an optimum set of positive Gaussian electron densities, given a set of complex structure factors, $F_{\rm known}$. This program uses the same conjugate-gradient algorithm as *SOLVE* but, instead of minimizing the discrepancy function in (5), it minimizes the squared distance between two complex quantities:

$$f_{\rm back} = \sum_{\mathbf{h}} |F_{\rm known} \exp[-\delta\eta(\pi\Delta r|\mathcal{F}^T\mathbf{h}|)^2] - O(\mathbf{h})|^2, \tag{52}$$

where the symbols are as defined in (3) and (6*a*). The main use for *BACK* is to transform known electron-density information into Gaussian basis-function amplitudes, $n(p)$, in order to use them for setting limiting constraints on the magnitude of the solution in the main solver and for setting up a 'target' density for $f_{\rm space}$ in (7). In fact, *BACK* can replace inverse Fourier transforms (for whatever purpose these are used) and, because positivity is ensured everywhere, the well known problems caused by the termination of Fourier series are avoided.

A5.4. *FORTH*

FORTH applies a fast Fourier transform to the electron/voxel information, $n(p)$, multiplying it by the exponential factor $\exp[-\eta(\pi\Delta r|\mathcal{F}^T\mathbf{h}|)^2]$ and converting it to structure factors. *FORTH* may thus be regarded as the inverse of *BACK*. *FORTH* is used to prepare an $F_{\rm calc}$ file that is consistent with a *BACK*-generated set of $n(p)$ data.

A5.5. *MAKETAR*

The utility *MAKETAR* is used to prepare solvent targets and weights for the solver described in (41). *MAKETAR* takes as input a smeared version of the complex $F_{\rm calc}$ of a model, prepared by running *APODFC* at a very low resolution (*e.g.* 7 Å) followed by *BACK* at the regular resolution. Based on their values, points in the resulting $n(p)$ file are redefined in *MAKETAR* as 'low' or 'high' such that a preset fraction (default: 1/2) of them are 'low' and are targeted to contain solvent. 'Low' points are given a weight of 1 and a target that is equivalent to 0.33 e Å^{-3}; 'high' points have a weight of 0 and an unused target.

A5.6. *SYM*

SYM is a utility for manipulating protein data bank (PDB) information directly. It is used for preparing masks used in *SOLVE* with non-crystallographic symmetry.

A5.7. *DPHASE*

DPHASE calculates the phase differences and the cosines of the phase differences between comparable (*hkl*) structure factors in two structure-factor files, $F^1_{\rm calc}$ and $F^2_{\rm calc}$. The phase differences are calculated twice – once weighted by the amplitudes of the first file, $|F^1_{\rm calc}|$, and once weighted by the amplitudes of the second, $|F^2_{\rm calc}|$. In each case, they are averaged and reported over shells of equal $1/d^2$ in (*hkl*) space. *DPHASE* excludes terms for which the amplitude in either file is 0 and it excludes the (000) term.

A5.8. *DRHO*

DRHO measures the normalized variance between two electron/voxel files, $n(p)$ and $n'(p)$.

$$V = \left\{ \sum_{p=1}^{P} [n(p) - n'(p)]^2 / \sum_{p=1}^{P} [n(p)]^2 \right\}^{1/2} \tag{53}$$

or the metric distance

$$D = \left\{ P \sum_{p=1}^{P} [n(p) - n'(p)]^2 \right\}^{1/2} \tag{54}$$

A5.9. *REGRID*

REGRID is a utility that evaluates the electron density in the unit cell from its representation in terms of $n(p)$. It first calculates the structure factors in (3) by a fast Fourier transform, then it uses a back transform to evaluate (2*b*), usually on a twice finer grid. Owing to the exponential factor present in (3), the result of the back transform is always positive and there are no cut-off errors. *REGRID* produces a map file in *X-PLOR* format.

APPENDIX *B*

In this Appendix, we summarize some of the relevant discussion on frames, following Daubechies (1992) and Heil & Walnut (1989) (referred to as D and H&W, respectively). Our interest is in the electron densities of molecules, $\rho(\mathbf{r})$ of (1). They are positive functions of limited resolution or bandwidth. According to (2*b*), they are to be represented by a set of Gaussian basis functions.

A set of functions $\{x_n\}$ is called a frame (with respect to the functions x) if, for any one of the functions x, the sum of the squares of its scalar products with all of the functions x_n is bounded both from above and from below (D 3.2.1, H&W 2.1.1):

$$A\langle x,x\rangle \leq \sum_n (|\langle x,x_n\rangle|^2) \leq B\langle x,x\rangle; \quad 0 < A, B < \infty, \tag{55}$$

where the scalar product $\langle x, y\rangle$ (also called a convolution or projection) is defined as the integral

$$\langle x, y\rangle = \int x(\mathbf{r})y(\mathbf{r})\,\mathrm{d}\mathbf{r}. \tag{56}$$

The existence of a frame ensures that the operator $\mathcal{S}$ that produces $\langle x, x_n\rangle$ from x is a bounded linear operator with a bounded inverse operator $\mathcal{S}^{-1}$. The functions x can then be expanded with the help of the operators $\mathcal{S}$, $\mathcal{S}^{-1}$. Indeed, let us define the operator $\mathcal{S}$ by

$$\mathcal{S}x = \sum_n (\langle x, x_n\rangle)x_n. \tag{57}$$

It follows that the function x can be expanded (represented) in terms of the set of functions x_n by the formula (D 3.2.8, H&W 2.1.5)

$$x = \sum a_n x_n, \quad \text{where} \quad a_n = \langle x, \mathcal{S}^{-1}x_n\rangle. \tag{58}$$

The set of functions $\{\mathcal{S}^{-1}x_n\}$ is also a frame, called the dual frame. It satisfies the bounds (D 3.2.6, H&W 2.1.4)

$$B^{-1}\langle x, x\rangle \le \sum_n (|\langle x, \mathcal{S}^{-1}x_n\rangle|^2) \le A^{-1}\langle x, x\rangle. \tag{59}$$

The significance of (55)–(59) for us is that, if we can show that our 'basis' set of Gaussians, (2*b*), is indeed a frame, we can be assured that any electron density can be represented by them. Moreover, we are assured that the representation is mathematically stable in both directions in the following sense. Given two sets of coefficients, a_n, b_n, that are close, the set of coefficients $c_n = a_n - b_n$, when used in (58), defines a function that is small because in (55) $B < \infty$. The converse is that if the function x is small the set of coefficients is also small because of the right-hand inequality in (59). In fact, (58) supplies an algorithm for the representation of a known electron density.

The basis-function sets that are most familiar are orthonormal ones. Those are the generalizations of orthonormal (Cartesian) coordinate systems to (infinite-dimensional) function spaces. Such basis sets always constitute frames, with $A = B = 1$ in (55), so frames can be thought of as generalizations of orthonormal basis sets. Frames are usually not orthogonal and they are usually redundant in the sense that the representation presented in (58) is not unique. However, there is a connection among all possible representations of x in terms of x_n (H&W 2.1.5). If, in addition to the representation (58), we can find another set of coefficients that represents x (*e.g.* by using the algorithm of *EDEN* or by magic),

$$x = \sum c_n x_n, \tag{60}$$

the following connection exists between the two representations:

$$\sum |c_n|^2 = \sum |a_n|^2 + \sum |a_n - c_n|^2. \tag{61}$$

Equation (61) shows that algorithm (58) always yields a representation with 'minimal norm'.

Two main classes of frames are discussed in D and H&W. The frames in the first class are called Weyl–Heisenberg–Gábor frames; they can also be viewed as windowed Fourier transforms. Those in the second class are wavelet frames. The basis function representation in this paper [(2*b*)] belongs to the former class. The relevant formulas will be shown in one dimension but they apply to three-dimensional lattices as well.

Given a standard Gaussian

$$g(x) = \pi^{-1/4}\exp(-x^2/2), \tag{62}$$

we can define a set of functions

$$g_{m,n}(x) = \exp(im\omega_o x)g(x - nt_o), \quad m, n = \text{integer}. \tag{63}$$

They are the basis functions of the windowed Fourier transform. They measure the frequency content, around $m\omega_o$ in frequency, of a small section of a function centered around nt_o in space. It is shown in D 3.4.4 that if $\omega_o t_o < 2\pi$ the set of functions (63) constitutes a frame. The frame bounds A, B of (55) are estimated in D Table 3.3. The significance of the frame bounds is that, if they are close, the representation of most functions converges rapidly. In our application, we are interested in a representation where only $m = 0$ is kept in (63).

Let us identify the variables in (62), (63) with those of (2*b*) as

$$x = 2^{1/2}\pi r/d_{\rm res}; \quad t_o = 2^{1/2}\pi\Delta r/d_{\rm res}. \tag{64}$$

The experimental data set and the structure factors of the known part are apodized according to (6*a*,*b*) to have the same inherent resolution as the basis set (2*b*). It follows that all possible (positive) electron densities have a maximum resolution $d_{\rm res}$, their Fourier transform falls off in reciprocal space as $\exp(-d^2_{\rm res}|\mathcal{F}^T\mathbf{h}|^2)$ and correspond to a crystallographic $B/4 = d^2_{\rm res}$. A prototypical function with these properties is one of the Gaussians in (63) corresponding to one of the Gaussian basis functions of (2*b*). The Fourier amplitudes of such a function can be calculated by the formula

$$\int_{-\infty}^{\infty} \exp[-(x - nt_o)^2/2]\exp(im\omega_o x)\,\mathrm{d}x = (2\pi)^{1/2}\exp(-m^2\omega_o^2/2)\exp(imn\omega_o t_o). \tag{65}$$

We will choose

$$\Delta r = d_{\rm res}/\pi, \quad \omega_o t_o = \pi. \tag{66}$$

The frame bounds, from D Table 3.3, are close to $A = 1.5$ and $B = 2.5$, their ratio being about 1.7. Thus, the frame is fairly tight and we should expect fairly good convergence of the representation for any electron density. Moreover, from (64), we can calculate the value of the first non-zero Fourier component, $\exp(-\omega_o^2/2) = 0.085$. Therefore, if we neglect all higher Fourier components of the frame, *i.e.* if we restrict our representation to (3), the maximum relative error we make is 8.5%. Similarly, the restriction that all the amplitudes of the Gaussians be non-negative, in order to satisfy the constraint that the electron density be non-negative everywhere, is expected to cause a similarly small error. This is the mathematical basis of our representation of electron densities.

The above discussion can be generalized to a multiresolution representation either in terms of Gábor frames or in terms of wavelets. In either case, satisfying positivity everywhere can be quite difficult.

References

Baker, D., Krukowski, A. E. & Agard, D. A. (1993). *Acta Cryst.* D**49**, 186–192.

Béran, P. & Szöke, A. (1995). *Acta Cryst.* A**51**, 20–27.

Bertero, M. (1989). *Advances in Electronics and Electron Physics*, Vol. 75, pp. 1–120. New York: Academic Press.

Bragg, W. L. (1950). *Nature (London)*, **166**, 399–400.

Brändén, C.-I. & Jones, T. A. (1990). *Nature (London)*, **343**, 687–689.

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

Bricogne, G. (1992). *International Tables for Crystallography*. Vol. B, edited by U. Shmueli. Dordrecht: Kluwer Academic Publishers.

Bricogne, G. (1993). *Acta Cryst.* D**49**, 37–60.

Chapman, M. S. (1995). *Acta Cryst.* A**51**, 69–80.

Chapman, M. S., Tsao, J. & Rossmann, M. G. (1992). *Acta Cryst.* A**48**, 301–312.

Cowtan, K. D. & Main, P. (1993). *Acta Cryst.* D**49**, 148–157.

Dainty, J. C. & Fienup, J. R. (1987). *Image Recovery: Theory and Applications*, edited by H. Stark, Ch. 7. Orlando: Academic Press.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM.

Eigen, M. (1992). *Steps Towards Life*. Oxford University Press.

Furey, W. & Swaminathan, S. (1990). Am. Crystallogr. Assoc. Meet. Abstracts, **18**, 73.

Giacovazzo, C. (1992). Editor. *Fundamentals of Crystallography*. IUCr/Oxford University Press.

Goodman, D. M., Johansson, E. M. & Lawrence, T. W. (1993). *Multivariate Analysis: Future Directions*, edited by C. R. Rao, Ch. 11. Amsterdam: Elsevier.

Gull, S. F. & Daniell, G. (1978). *Nature (London)*, **29**B, 49–51.

Heil, C. E. & Walnut, D. F. (1989). *SIAM Rev.* **31**, 628–666.

Kauffman, S. A. (1993). *The Origins of Order*. Oxford University Press.

Kleywegt, G. J. & Jones, T. A. (1995). *Structure*, **3**, 535–540.

Kull, F. J., Sablin, E. P., Lau, R., Fletterick, R. J. & Vale, R. D. (1996). *Nature (London)*, **380**, 550–555.

Lánczos, C. (1961). *Linear Differential Operators*. London: Van Nostrand.

Maalouf, G. J., Hoch, J. C., Stern, A. S., Szöke, H. & Szöke, A. (1993). *Acta Cryst.* A**49**, 866–871.

Millane, R. P. (1996). *J. Opt. Soc. Am.* A**13**, 725–734.

Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

Read, R. J. (1990). *Acta Cryst.* A**46**, 900–912.

Rinnoy Kan, A. H. G. & Timmer, G. T. (1989). *Handbooks in OR & MS*, Vol. 1, edited by G. L. Nemhauser, Ch. 9. Amsterdam: Elsevier.

Rossmann, M. G., McKenna, R., Tong, L., Xia, D., Dai, J.-B., Wu, H., Choi, H.-K. & Lynch, R. E. (1992). *J. Appl. Cryst.* **25**, 166–180.

Skilling, J. & Gull, S. F. (1985). *Maximum Entropy and Bayesian Methods in Inverse Problems*, edited by C. R. Smith & W. T. Grandy Jr, pp. 83–132. Dodrecht: Reidel.

Somoza, J. R., Szöke, H., Goodman, D. M., Béran, P., Truckses, D., Kim, S.-H. & Szöke, A. (1995). *Acta Cryst.* A**51**, 691–708.

Szöke, A. (1993). *Acta Cryst.* A**49**, 853–866.

The CCP4 Suite (1994). The CCP4 Project, Daresbury Laboratory, Warrington WA4 4AD, England.

Tollin, P., Main, P., Rossmann, M. G., Stroke, G. W. & Restrick, R. C. (1966). *Nature (London)*, **209**, 603–604.

Tsao, J., Chapman, M. S. & Rossmann, M. G. (1992). *Acta Cryst.* A**48**, 293–301.

Zhang, K. Y. J. (1993). *Acta Cryst.* D**49**, 213–222.