Acta Cryst. (1995). A51, 691–708

Holographic Methods in X-ray Crystallography. IV. A Fast Algorithm and its Application to Macromolecular Crystallography

BY JOHN R. SOMOZA*

Department of Chemistry and Lawrence Berkeley Laboratory, University of California at Berkeley, Berkeley, CA 94720, USA

HANNA SZÖKE, DENNIS M. GOODMAN AND PIERRE BÉRAN†

Lawrence Livermore National Laboratory, Livermore, CA 94550, USA

DAGMAR TRUCKSES

Department of Biochemistry, University of Wisconsin, Madison, WI 53711, USA

SUNG-HOU KIM

Department of Chemistry and Lawrence Berkeley Laboratory, University of California at Berkeley, Berkeley, CA 94720, USA

AND ABRAHAM SZÖKE‡

Lawrence Livermore National Laboratory, Livermore, CA 94550, USA

(Received 17 *October* 1994; *accepted* 15 *February* 1995)

Abstract

The holographic method makes use of partially modeled electron density and experimentally measured structure-factor amplitudes to recover electron density corresponding to the unmodeled part of a crystal structure. This paper describes a fast algorithm that makes it possible to apply the holographic method to sizable crystallographic problems. The algorithm uses positivity constraints on the electron density and can incorporate a 'target' electron density, making it similar to solvent flattening. The potential for applying the holographic method to macromolecular X-ray crystallography is assessed using both synthetic and experimental data.

Introduction

Holographic methods for X-ray crystallography were introduced by Szöke (1993) (hereafter paper II) and by Maalouf, Hoch, Stern, Szöke & Szöke (1993). For a brief description of the method, let us assume that the electron density is known in part of the unit cell of a crystal. The complex amplitude of the wave diffracted from that part can then be calculated. This wave is analogous to a reference wave in holography. Similarly, the (unknown) wave diffracted from the unknown part of the unit cell is analogous to the object wave of holography. The intensity of the wave diffracted by the full crystal, which is proportional to the square of the observed structure factors in X-ray diffraction, is then analogous to a recorded hologram. It contains the diffracted intensities of the known and the unknown parts of the electron density separately and, in addition, a term corresponding to the interference of the waves scattered from the known and unknown parts of the crystal. The interference term contains 'phase information' that can 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.

Our previous papers contained a detailed description of the holographic method and derived some of the mathematical properties of the resulting equations. As discussed previously, the method has several attractive properties. Its most important advantage over other methods currently used for calculating electron density is that it accommodates in a natural way additional constraints on the electron density. It is well known that constraints on the electron density, such as its positivity or the presence of known solvent regions, are essential for the faithful recovery of the crystal structure. Our mathematical derivation highlighted the similarity of the crystallographic phase problem to a class of inverse

*Present address: Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, CA 94143, USA.

†Present address: Institut de Physique, Université de Neuchatel, 1 Rue Breuget, 2000 Neuchatel, Switzerland.

‡ To whom correspondence should be addressed.

problems in mathematics that include those encountered in image processing. Following those theoretical guidelines, we have succeeded in producing a computer program based on an accurate, fast and practical algorithm that will be applicable to problems in macromolecular crystallography. In its present form, the program contains no knowledge of protein chemistry.

As opposed to most traditional methods, which are expressed in Fourier space, the holographic method was designed to solve for the electron density in physical space, and it uses known electron density in part of the unit cell as a 'target' density. This strategy was shown to be successful in a feasibility study by Béran & Szöke (1995). The deviations of the recovered electron density from the target density are used in a cost function that is minimized in parallel with the 'standard' holographic cost function (paper II), which depends on the deviations of the calculated structure-factor amplitudes from the measured ones. The relative magnitude of the two cost functions can be controlled by a Lagrange multiplier. When the known density is in a solvent region, the procedure is similar to solvent flattening.

A description of the computer program and some tests of its application to macromolecular crystallography are the subject of the present paper. In their present form, the programs 'stand alone', *i.e.* they are not part of any crystallographic program package, but they communicate readily with known crystallographic programs. The first section of this paper discusses the basic holographic equations, the addition of constraints, the actual algorithms used and their implementation as the program *EDEN*. The second section describes the application of *EDEN* to some synthetic but realistic problems, in order to test the limits and accuracy of the method. A third section reports on the application of *EDEN* to the solution of the structure of a mutant of staphylococcal nuclease. After summarizing our results, we give some programming details about *EDEN* in the Appendix.

1. Fast holographic algorithm

The general description of the holographic algorithm will be presented in this section. A more detailed description of the algebra and some programming details are relegated to the Appendix.

1.1. *The holographic equations and their solution*

The notation in this paper is the same as in paper II, to which the reader is referred for more precise definitions (Szöke, 1993). The electron density in the unit cell of a crystal is divided 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 of spacing $\mathcal{F}\Delta\mathbf{r} = (\Delta x, \Delta y, \Delta z)$ in fractional coordinates, with an unknown number of electrons, n_p, in each Gaussian blob (voxel). 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:

$$\rho_{\text{unknown}}(\mathbf{r}) \simeq (\pi\eta\Delta r^2)^{-3/2} \sum_{p=1}^{P} n_p \exp(-|\mathbf{r} - \mathbf{r}_p|^2/\eta\Delta r^2). \qquad (2)$$

The value of η determines the widths of the Gaussian basis functions. It is chosen to represent a continuous electron density optimally. In the algorithm to be described below, the unknowns 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 §2.1.) A derivation presented in paper II results in the following formula for the structure factors of the unknown part, $O(\mathbf{h})$:

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

The square of the absolute magnitude 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. We will assume that the appropriate normalization has been done and, therefore, the values of $|F(\mathbf{h})|^2$ are known from experiment. They satisfy the equation

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

When the expansion of the unknown density is substituted from (3), equation (4) becomes a set of quadratic equations in the unknowns, n_p. In our current algorithm, the equations are linearized by initially neglecting the $|O(\mathbf{h})|^2$ term. The linearized equations are then solved using a (linear) conjugate gradient algorithm (Goodman, Johansson & Lawrence, 1993). After the linearized solution is obtained, the equations are iterated by adding the recovered $O(\mathbf{h})$ to the known part. This is an iterative solution of (4). The number of equations, N_h, is usually not equal to the number of unknowns, P; also, the equations are ill conditioned (see paper II). Under these conditions, the best (quasi-)

* The notation of $R(\mathbf{h})$ for the structure factors of the known part of the structure and $O(\mathbf{h})$ for the structure factors of the unknown part of the structure is derived from holography theory. In holography, $R(\mathbf{h})$ and $O(\mathbf{h})$ denote the reference and object wave, respectively.

solution is the one that minimizes the discrepancy function,

$$f_{\text{eden}} = \sum_{\mathbf{h}} w'(\mathbf{h})^2 \left[\sum_{p=1}^{P} n_p M_p(\mathbf{h}) - H(\mathbf{h}) \right]^2, \quad (5)$$

where $H(\mathbf{h}) = |F(\mathbf{h})|^2 - |R(\mathbf{h})|^2$, the set of positive weights $w'(\mathbf{h})^2$ will be determined below and $M_p(\mathbf{h})$ is the linear 'encoding' operator, corresponding to the set of linearized equations (4). $M_p(\mathbf{h})$ is given by

$$M_p(\mathbf{h}) = \exp[-\eta(\pi \Delta r |\mathcal{F}^T \mathbf{h}|)^2][R(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathcal{F}\mathbf{r}_p) + R^*(\mathbf{h}) \exp(2\pi i \mathbf{h} \cdot \mathcal{F}\mathbf{r}_p)]. \quad (6)$$

The weights $w'(\mathbf{h})^2$ are determined by two considerations. First, we divide each term of the discrepancy function f_{eden} in (5) by $[|F(\mathbf{h})| + |R(\mathbf{h})|]$; we note that $|F(\mathbf{h})|^2 - |R(\mathbf{h})|^2$ divided by this factor yields a first-order term, $|F(\mathbf{h})| - |R(\mathbf{h})|$. This division, therefore, increases the curvature of the discrepancy function around its minimum. Accordingly, we set $w'(\mathbf{h})^2 = w(\mathbf{h})^2/[|F(\mathbf{h})| + |R(\mathbf{h})|]^2$. Second, values of $\mathbf{h}$ for which $|R(\mathbf{h})| \ll |F(\mathbf{h})|$ correspond to small singular values of the operator in (4), giving rise to numerical instability of the algorithm. The singularity is eliminated by using the weights $\{2|R(\mathbf{h})|/[|F(\mathbf{h})| + |R(\mathbf{h})|]\}$ in the discrepancy function. Theoretical reasons for the above considerations were discussed in detail in paper II.

In summary, the resulting discrepancy function used in our calculations is

$$f_{\text{eden}} = \sum_{\mathbf{h}} w(\mathbf{h})^2 \{[R(\mathbf{h})O^*(\mathbf{h}) + R^*(\mathbf{h})O(\mathbf{h})] \times [|F'(\mathbf{h})| + |R(\mathbf{h})|]^{-1} - [|F'(\mathbf{h})| - |R(\mathbf{h})|]\}^2, \quad (7)$$

with regularizing weights

$$w(\mathbf{h})^2 = \{2|R(\mathbf{h})|/[|F'(\mathbf{h})| + |R(\mathbf{h})|]\}^2. \quad (8)$$

In addition, the measured structure-factor amplitudes are modified to

$$|F'(\mathbf{h})| = |F(\mathbf{h})| \exp[-\delta\eta(\pi \Delta r |\mathcal{F}^T \mathbf{h}|)^2], \quad (9)$$

where the parameter δ is of order unity. Such 'apodization' adjusts the resolution of the measured diffraction pattern to the finite resolution of the Gaussian basis set used in the solution. It is equivalent to an appropriate smearing of the electron density of the protein. Such smearing is needed in order to be able to fit the high-resolution reflections in a mathematically stable manner.

1.2. *Addition of constraints*

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 the encoding operator, $M_p(\mathbf{h})$ in (6), can be added to any vector $\mathbf{n}_p$ that minimizes the cost function (5). However, it is well known that additional information, in the form of constraints or restraints, can reduce the arbitrariness of the solution. Although both constraints and restraints will be incorporated into the algorithms, we will use the term constraints in all cases.

Three different types of constraint will be discussed. The first constraint is the positivity of the electron density. There are two different ways to make use of positivity. If the electron density is believed to be substantially correct in the 'known' part of the unit cell, one can demand that only a positive electron density be added everywhere; this will be denoted 'completion' mode. In contrast, the program can be used to correct the electron density even in the 'known' region. In this 'correction' mode, only the total electron density (the sum of the known density and the recovered density) is required to be positive. A second kind of constraint is the knowledge of the electron density in certain regions. It can be, for example, the known part of the molecule, or the position and the density of the solvent. In either case, a 'target' density can be specified in part of the unit cell. This is similar to Béran's considerations (Béran & Szöke, 1995). Such constraints are incorporated into the algorithm with the help of a cost function, f_{space}, that is proportional to the square deviation of the calculated electron density from the target density and a relative weight, or Lagrange multiplier, λ_{space}. Third, the projection of n_p on the null space of the encoding operator (6) can be minimized by adding a term proportional to $|\text{Im}[R^*(\mathbf{h})O(\mathbf{h})]|^2$ with appropriate weights and a Lagrange multiplier, λ_{null}. Such a term produces the usual quasi-solution of ill conditioned equations with minimum norm.

The first type of constraint, the positivity of the electron density, is incorporated directly into the conjugate-gradient optimizer by stipulating that the solution vector is bounded from below by the negative of the known initial electron density (correction mode) or by zero (completion mode).

The second type of constraint, the 'target' density, is expressed in terms of the amplitudes of the same basis functions as used by the main program. They will be denoted by $n_{p,\text{target}}$. The cost function, f_{space}, is expressed as

$$f_{\text{space}} = \lambda_{\text{space}} P \sum_{p=1}^{P} \tilde{w}_p{}^2 (n_p - n_{p,\text{target}})^2. \quad (10)$$

The Lagrange multiplier, λ_{space}, and the weights, $\tilde{w}_p{}^2 \leq 1$, express the 'strength of our belief' in the correctness of the target density: the weights, $\tilde{w}_p$, are used to emphasize or de-emphasize different regions of the target density, while λ_{space} determines the relative importance f_{space} with respect to f_{eden}. The cost function, f_{space}, is in a different space from f_{eden}; therefore, it is important to normalize them relative to each other. A

simple calculation, using the fact that the Patterson function is the Fourier transform of $|F(\mathbf{h})|^2$, Parseval's theorem and the approximate locality of the basis functions in (2), yield the result that the number of grid points, P, is the proper normalization if all the weights, $\tilde{w}_p{}^2$, are unity. Although our target density covers only part of the unit cell, this is still a good approximation. The value of the Lagrange multiplier, λ_{space}, that gives best results in practical problems is quite small, 0.001–0.01, independent of problem size.

Finally, in the absence of other constraints, a non-zero value of λ_{null} produces an electron density that is very similar to the result of 'Gábor's reconstruction' or to the difference Fourier method. In our test cases, a non-zero λ_{null} did not make much difference, as the conjugate-gradient method also tends to find a solution of the electron density of minimal norm; indeed, we found that, in the absence of constraints, our $\mathbf{n}_p$ tends to the difference Fourier solution. The cost function, f_{null}, used is

$$f_{\text{null}} = \lambda_{\text{null}} \sum_{\mathbf{h}} w(\mathbf{h})^2 (\{\text{Im}[R^*(\mathbf{h})O(\mathbf{h})]\}/|R(\mathbf{h})|)^2. \quad (11)$$

The actual cost function used in the computer program is the sum of f_{eden} (7), f_{space} (10) and f_{null} (11).

$$f_{\text{total}} = f_{\text{eden}} + f_{\text{space}} + f_{\text{null}}. \quad (12)$$

1.3. *A fast algorithm*

The discrepancy function (5) depends on the discretized electron density, $\mathbf{n}_p$. We will show that it can be calculated using fast Fourier transforms. First, we rewrite (3) as

$$O(\mathbf{h}) = \exp[-\eta(\pi \Delta r |\mathcal{F}^T \mathbf{h}|)^2] \sum_{p=1}^{P} n_p \exp(2\pi i \mathbf{h} \cdot \mathcal{F} \mathbf{r}_p). \quad (13)$$

If the points $\mathbf{r}_p$ are on a regular symmetry-adapted grid in the unit cell,

$$\mathbf{r}_p = p_a \mathbf{a}/P_a + p_b \mathbf{b}/P_b + p_c \mathbf{c}/P_c. \quad (14)$$

It can easily be seen that

$$\exp(2\pi i \mathbf{h} \cdot \mathcal{F} \mathbf{r}_p) = \exp[2\pi i (hp_a/P_a + kp_b/P_b + lp_c/P_c)], \quad (15)$$

where the components of $\mathbf{h}$ are denoted by $\{h,k,l\}$. The Fourier transform of the solution vector, denoted by $n(\mathbf{h})$, can therefore be calculated by the fast discrete Fourier transformation (DFT). Defining DFT^+ in the obvious way, we get

$$\begin{aligned} n(\mathbf{h}) &= \text{DFT}^+(\mathbf{n}_p) \\ &= \sum_{p=1}^{P} n_p \exp(2\pi i \mathbf{h} \cdot \mathcal{F} \mathbf{r}_p) \\ &= \sum_{p_a=1}^{P_a} \sum_{p_b=1}^{P_b} \sum_{p_c=1}^{P_c} n_p \exp[2\pi i (hp_a/P_a \\ &\quad + kp_b/P_b + lp_c/P_c)]. \end{aligned} \quad (16)$$

It is easy to see that $O(\mathbf{h})$ can now be calculated by a single vector multiplication of length $N_{\mathbf{h}}$, the number of structure factors used. Similarly, the other parts of the discrepancy function, (10), (11), and their gradients can be calculated by fast discrete Fourier transforms. The whole operation can therefore be carried out in a time of order $P \log P$. The matrix, $M_p(\mathbf{h})$, of (6) (that is of length $N_{\mathbf{h}}P$) is never calculated and therefore only vectors of length P and $N_{\mathbf{h}}$ have to be stored. This is the basis of the fast algorithm. Its details are presented in the Appendix. Mathematically, the fast algorithm outlined above depends on the shift invariance of the holographic kernel and is equivalent to it.

Our computer programs minimize the discrepancy function f_{total} of (12) using a conjugate-gradient algorithm developed by one of the authors (Goodman, Johansson & Lawrence, 1993). This algorithm is especially advantageous in the presence of non-linear constraints. By using a 'bending' search method in a given direction, several components of the solution vector can be constrained in a single search. This is very important for the efficient solution of the large problems that are routinely encountered in crystallography. The algorithm is very robust; we have never observed it to fail. Each component of the solution vector, $\mathbf{n}_p$, can be unconstrained, constrained (from above and/or from below), or held fixed. As the calculation proceeds from one linearized step to another, the constraining bounds may be changed by the program.

Multiple minima of the cost function (12) are expected, especially in the presence of constraints. In such a case, conjugate-gradient minimizers may be far from optimal. Our general method lends itself easily to other methods of searching for global minima, such as simulated annealing, but these methods have not been pursued so far.

1.4. *Programs and options*

There is a suite of three programs, *BACK*, *EDEN* and *REGRID* (together with some utility programs); their general properties are described in this section. More details are given in the Appendix. The programs readily communicate with known crystallographic programs *via* standard input and output files. The inputs to the main program are the measured structure-factor amplitudes, as well as complex structure-factors calculated from the partial model. Also, the standard crystal parameters have to be defined, together with certain computational parameters. For example, the user should define the desired resolution of the data to be read and the fineness of the grid on which the electron density is to be calculated. In fact, three different resolutions can be

distinguished. One is the inherent resolution of the data, equivalent to an effective crystallographic B value. It can be adjusted by using an appropriate value of δ in (9). The second is the resolution of the solution grid and the corresponding value of η in (3). The third is the maximum resolution of the structure factors used in the summation for the determination of the discrepancy function (12). The programs' final output is an electron-density file. Also, all three of our suite of programs employ an intermediate binary file format for convenience in studying intermediate results.

The preprocessor program, called *BACK*, produces a vector, $\mathbf{n}_p$, of size $P = P_a P_b P_c$, corresponding to the electron-density expression in (2). This is an optimum set of positive Gaussian electron densities, given a set of complex structure factors, F_{known}. This program also uses Goodman's (Goodman, Johansson & Lawrence, 1993) conjugate-gradient algorithm but, instead of minimizing the discrepancy function in (12), it minimizes

$$f_{\text{back}} = \sum_{\mathbf{h}} |F_{\text{known}} \exp[-\delta\eta(\pi\Delta r|\mathcal{F}^T\mathbf{h}|)^2] - O(\mathbf{h})|^2, \tag{17}$$

where the symbols are as defined above. The main use for *BACK* is to transform the known electron-density information into Gaussian basis function amplitudes, $\mathbf{n}_p$, in order to use them for setting limiting constraints on the magnitude of the solution in the main holographic program and for setting up a 'target' density for f_{space} in (10). In fact, *BACK* can replace inverse Fourier transforms (for whatever purpose those are used) and because positivity is ensured everywhere, the well known problems caused by missing data and by the termination of Fourier series are avoided. Also, given an electron density in terms of Gaussians on a grid, the complex structure factors can be calculated without additional error, using *FORTH*, a program that is the inverse of *BACK*.

The main program, *EDEN* (for electron density) solves the holographic equations by minimizing f_{total} of (12). The program reads in the reflection data sets to the prescribed resolution. It reconciles the reflection data, taking only reflections that are present in the measured data. As discussed in paper II, the gridding resolution has to be close to that of the data, otherwise either resolution is lost or the equations become underdetermined. There are two options for establishing the solution grid: a simple grid or a body-centered grid. The default value of η in (2) and (6) is 0.36 for a simple grid and 0.28 for a body-centered grid. Both these values correspond roughly to the Rayleigh resolution criterion. The advantage of a body-centered grid is that, at a cost of only twice the number of grid points, it provides a more uniform coverage of space and it has less preference for the directions of the crystal axes. (In trigonal or hexagonal grids, a hexagonal close-packed grid would have to be used. This has not yet been implemented in *EDEN*.) Ideally, the same grid spacing should be used in all crystal dimensions. In order to approximate this equality, we have used a FFT based on products of powers of 2, 3 and 5, in place of the usual 2-based FFT. The program calculates the actual grid spacing used in each crystal direction by dividing the length of the unit cell by the desired grid spacing, $P_a \simeq |\mathbf{a}|/\Delta r$ *etc.* and setting it to be the nearest integer that can be factored by 2, 3 and 5. For relatively high resolution studies, the grid spacing in any of the three crystal directions is within $\sim 10\%$ of the desired resolution.

EDEN can be used to complete a structure or it can be used to correct a structure. In completion mode, the program attempts to minimize the discrepancy, (12), using only positive Gaussians, $n_p \geq 0$. In correction mode, a 'known' set of Guassian amplitudes, $n_{p,\text{known}}$, has to be supplied and the program forces the sum of $n_p + n_{p,\text{known}}$ to be positive. This allows the program to correct the known electron density and also satisfy the positivity criterion. In addition, any number of variables can be further constrained to have an unchangeable value by supplying masks in real (grid) space. Note the distinction between the absolute constraints mentioned here and the 'soft' ones introduced in (10).

Once the problem is set up, the 'inner loop' that uses the constrained conjugate-gradient algorithm is started. It solves the linearized form of (4) by minimizing f_{total} of (12). The inner loop has its own stopping criteria, described by Goodman, Johansson & Lawrence (1993). In practice, the algorithm most often stops when the value of the gradient decreases 'enough' (to less than a preset fraction, usually 1%, of its starting value). When the inner loop stops, the recovered density is added to that found previously, the values of the minimum constraints are changed accordingly, the density found is Fourier transformed, apodized [see (3)] and added to the reference, $R(\mathbf{h})$. The progress of the 'outer loop' of the solution is monitored by the value of the discrepancy, f_{total}, and by the crystallographic R factor. The solution is deemed to be complete if the value of the discrepancy stops decreasing or if the crystallographic R factor has fallen to a predetermined value. It should be noted that *EDEN* uses the full electron density to calculate the structure factors. Consequently, the crystallographic R factor in *EDEN* is different from the conventional R factor, which is calculated using structure factors computed from a model.

When the value of λ_{space} is set to be non-zero, a 'target' electron density has to be supplied. Such a target density can be that of a solvent region or the density in part of the protein that is particularly well known. Most commonly, we used as target the region of the crystal that was known to contain only disordered solvent and we assigned it a target electron density of 0.33 Å^{-3}. In order to prepare a sufficiently conservative solvent region, the (tentative) model of the protein is transformed to physical space using *BACK*, with $\delta = 4$ [equation (9)], for example.

Such a high δ smears the protein very strongly and all regions whose electron density was less that $\sim 10\%$ of the maximum could be considered to be solvent regions. The same procedure was used to define the solvent masks.

The third program, *REGRID*, is a post processor. It uses (2) to calculate the electron density on a fine grid, in the desired part of the unit cell, from a given solution, n_p. Insofar as a body-centered grid is used in *EDEN*, the simple and intercalated grid points are reconciled on the fine grid. *REGRID* also uses fast Fourier transforms to accomplish its task.

2. Testing the applicability of the holographic method to macromolecular crystallography using synthetic data

Our first tests of the holographic method, as implemented by *EDEN*, used synthetic data, *i.e.* data computed from a protein model. Calculated structure-factor amplitudes that are used in place of experimental data will be referred to as F_{true} data. Using F_{true} data allows us to test the holographic method on realistic data, but under very controlled conditions. It allows us to study the sensitivity of the holographic method to well defined sources of error in the data and to problems in treating experimental data. Also, the use of F_{true} data circumvents the phase biasing problems often associated with omit maps (Hodel, Kim & Brünger, 1992). Finally, the phases corresponding to the F_{true} amplitudes are known, and therefore the results of tests using the holographic method can be quantitatively evaluated either in reciprocal space or in real space.

All of the initial testing of the holographic method was carried out using the structure of the 207 residue sweet-tasting protein thaumatin (Ogata, Gordon, de Vos & Kim, 1992). Structure-factor amplitudes were calculated from the thaumatin model placed in a $P2_12_12_1$ cell with cell parameters of $a = 74.44$, $b = 53.77$, $c = 52.32$ Å. An individual isotropic temperature factor of 14.0 Å^2 was specified for all atoms of the thaumatin model. The primary goals of the synthetic data experiments were to understand the inherent inaccuracies of the method due to expansion of the electron density into a finite number of basis functions, to make sure the holographic method could cope with imprecision in experimental data and in the atomic positions of the 'known' model, and to give us some understanding of how accurately the $F(000)$ term and scale factor need to be determined.

As mentioned above, for the tests using synthetic data, we unambiguously knew the 'true' structure factors, and were able to assess quantitatively the quality of the recovery either in reciprocal space or in real space. In reciprocal space, the results were evaluated by calculating the weighted average phase difference between the calculated ('*c*') structure factors produced by *EDEN* and the 'true' structure factors corresponding to the ideal structure:

$$\Delta\varphi = \left\{ \sum_{\mathbf{h}} |F_{\text{true}}(\mathbf{h})| \{\varphi_{\text{true}}(\mathbf{h}) - \varphi_c(\mathbf{h})\}^2 \Big/ \sum_{\mathbf{h}} |F_{\text{true}}(\mathbf{h})| \right\}^{1/2}. \tag{18}$$

In real space, the electron density resulting from the holographic reconstruction was evaluated by calculating the variance of the electron-density error in the unit cell,

$$\Delta\rho = \left\{ \sum_p (n_p - n'_p)^2 \Big/ \sum_p (n_p)^2 \right\}^{1/2}, \tag{19}$$

where n_p and n'_p are the two sets of electron densities (in units of electrons/voxel) to be compared.

All of the electron-density maps described in §§2.2 and 2.3 were prepared using the *EDEN* software package in 'completion' mode, as described above. The electron density recovered for the omitted parts of the model was also qualitatively evaluated by direct viewing of the density using the *PSFRODO* software package (Jones, 1985; Pflugrath, Saper & Quiocho, 1984).

2.1. *Limitations due to grid resolution*

One basic limitation on the achievable accuracy of the recovered electron density is its representation in terms of basis functions, of a fixed shape, on a fixed grid. In the present work, we used a Gaussian basis function expansion (2) and its Fourier transform (3). While other basis functions could also be used, the obvious advantage of Gaussians is that the positivity of the electron density is easily ensured by making the amplitude of each Gaussian positive. To test the effect of describing the electron density in terms of these basis functions, the structure factors calculated from the thaumatin model were apodized to a 2.5 Å data resolution, *BACK* was used to calculate the corresponding electron density on a body-centered grid and *FORTH* was then used to recalculate structure factors. We then compared the starting and ending structure factors by calculating the average phase difference and the crystallographic R factor. These results are shown in Fig. 1(*a*) as a function of the grid spacing, at a constant data resolution. The phase error is seen to be approximately proportional to the grid spacing. A similar study was made on the influence of η in (2), while keeping the product, $\delta\eta$ [of (9)], constant. The results of this study are shown in Fig. 1(*b*). It is clear from these results that the discrete grid size limits the accuracy of the phase recovery quite severely, while the actual value of η is much less important, at least within reasonable limits. We were quite surprised by the large values of the phase error obtained.

Two other tests were carried out to explore further the implications of solving for the electron density on a discrete grid. In the first test, we generated an electron

density that is a sum of a set of positive Gaussians centered on the grid points, with random positive amplitudes. The structure factors corresponding to this density were calculated using *FORTH* and the original density was then recovered using *BACK*. As expected, the resulting density error was small ($\Delta\rho \simeq 1\%$). We then changed the electron density on the grid points by adding additional random densities of 10 and 20%, respectively. The additional density led to phase changes in the structure factors of $\sim$5.7 and 12° ($\sim$0.1 and 0.2 rad), respectively. This observation verified that $\Delta\varphi$ is indeed proportional to $\Delta\rho$, at least for modest changes, and that the weighting of $\Delta\varphi$ by the amplitudes of the structure factors, $|F_{\text{true}}|$, is indeed the proper one. In the second test, we displaced the Gaussians from the grid points by $\sim$10 and 20% of the grid spacing, respectively. Again, we calculated the structure factors (directly) and recovered the original density using *BACK*. This time, there were large phase errors generated, $\sim$16 and 33°, respectively, while $\Delta\rho$ between the original density and the recovered density was only $\sim$10 and 20%. It was also observed that the recovered electron density tends to be attracted to the grid points. These results suggest that the main source of the phase errors in Figs. 1(*a*) and (*b*) is the failure of the basis functions on the grid to represent a general electron-density function. This property of basis functions on a discrete grid is well described in the literature (Stark, 1987); it carries over even to wavelet expansions (Daubechies, 1992.) Note also that in the second test the proportionality constant between the phase error, $\Delta\varphi$, and the density error, $\Delta\rho$, is much larger than in the first test. The second test is a more realistic simulation of reality, since electron density does not usually lie on a grid, and we conclude that a phase error of $\sim$30° corresponds to an acceptable density error of $\sim$10%.

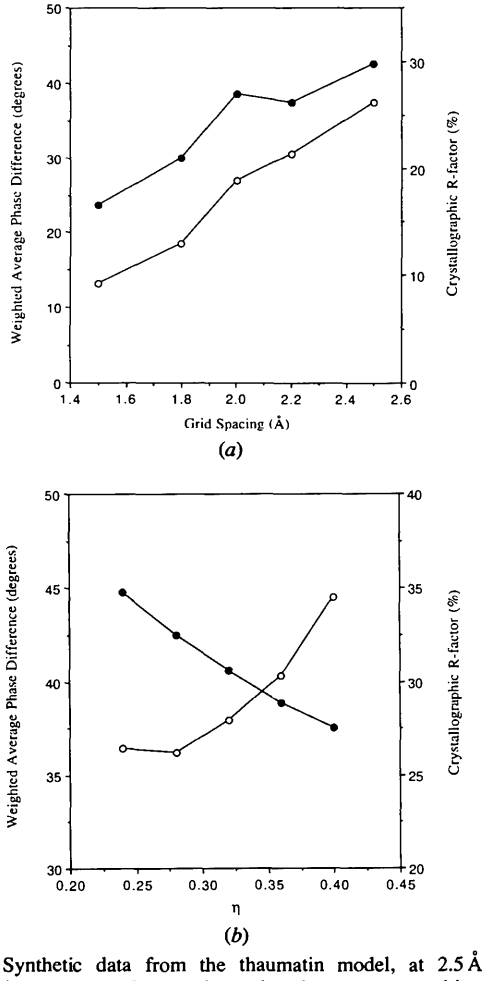

Fig. 1. Synthetic data from the thaumatin model, at 2.5 Å data resolution, were used to evaluate the phase errors resulting from modeling electron density on a finite grid. The weighted average phase difference (dark circles) and the crystallographic *R* factor (light circles) are shown as (*a*) a function of the grid spacing, and (*b*) as a function of the width of the basis function, η.

2.2. *Recovery of part of the structure*

The first set of tests using the synthetic data consisted in deleting residues from the N terminus of the thaumatin model and using the holographic method to recover the electron density corresponding to this omitted region. In all tests, a body-centered solution grid was used, with a grid spacing of 1.8 Å and $\eta = 0.28$. Data between infinity and 2.0 Å resolution were used. Fig. 2 shows the weighted average phase difference between the structure factors calculated from the intact thaumatin model and the structure factors calculated from the truncated thaumatin model before (open circles) and after (close circles) *EDEN* was used to recover the omitted portion of the model. As can be seen in this figure, the

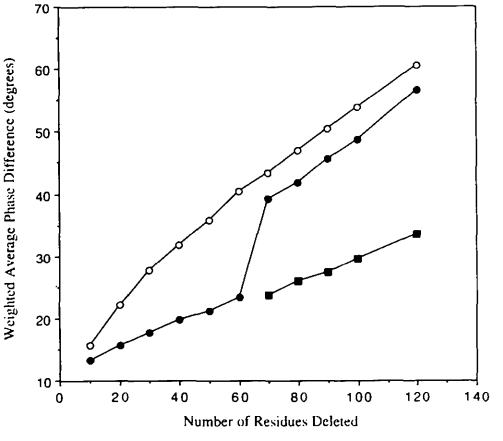

Fig. 2. The recovery of missing residues from the thaumatin model was evaluated by calculating the weighted average phase difference between the F_{true} data and the structure factors corresponding to the truncated model, before and after *EDEN* was used to recover the missing electron density. Open 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; closed squares show the phase difference after an *EDEN* run with a solvent mask.

use of the holographic method results in a clear improvement in the phases, especially when the omitted region represents approximately one third or less of the scattering matter. This phase improvement corresponds to improvements in the electron density recovered for the omitted parts of the model. Fig. 3(*a*) shows a representative sample of the electron density recovered for an omitted portion of the thaumatin model. For comparison purposes, an unweighted $(F_{true} - F_c)$ difference Fourier map is also shown in Fig. 3(*b*). The map produced by *EDEN* has fewer breaks in electron density, as well as fewer pieces of spurious density, than the corresponding $(F_{true} - F_c)$ Fourier map. Although the goal of this exercise was not to compare the holographic maps with difference Fourier maps, it is clear from these results that the holographic technique has the potential of producing useful electron-density maps. Fig. 3(*a*) also illustrates an interesting characteristic of the holographic maps. Frequently, the electron-density map producd by *EDEN* is slightly displaced with respect to the model. We believe that this phenomenon is due to the difficulties in gridding electron-density maps that were described in §2.1.

We have also tested the effects of solvent masks on the recovery of omitted regions of thaumatin. A solvent mask was generated as described in §1.4 and *EDEN* was run in completion mode at a data resolution of 2.0 Å and gridding resolution of 1.8 Å. The effect on the electron-density recovery brought about by the use of solvent masks is illustrated by the squares in Fig. 2. Clearly, the solvent mask is very helpful in the density recovery. A representative sample of the density recovered after deleting residues 1–120 of thaumatin model is shown in Fig. 4(*a*). The corresponding unweighted difference Fourier map (using no solvent mask) is shown in Fig. 4(*b*).

2.3. *Sensitivity of the method to errors*

In preparation for applying the holographic technique to problems involving experimentally derived data, the F_{true} data from the thaumatin model were used to test the tolerance of the method to errors in scaling the data, to errors in the magnitude of the $F(000)$ term, to Gaussian noise in the structure-factor amplitudes and to positional errors in the 'known' part of the model.

The use of the holographic method requires that the measured structure-factor amplitudes be placed on an absolute scale. The sensitivity of the holographic method to the scale factor was tested by varying the scaling of the F_{true} data. For each re-scaled data set, *EDEN* was used to recover electron density corresponding to an omitted part (30, 60 or 90 residues) of the thaumatin structure. Fig. 5(*a*) summarizes the effects on the electron-density

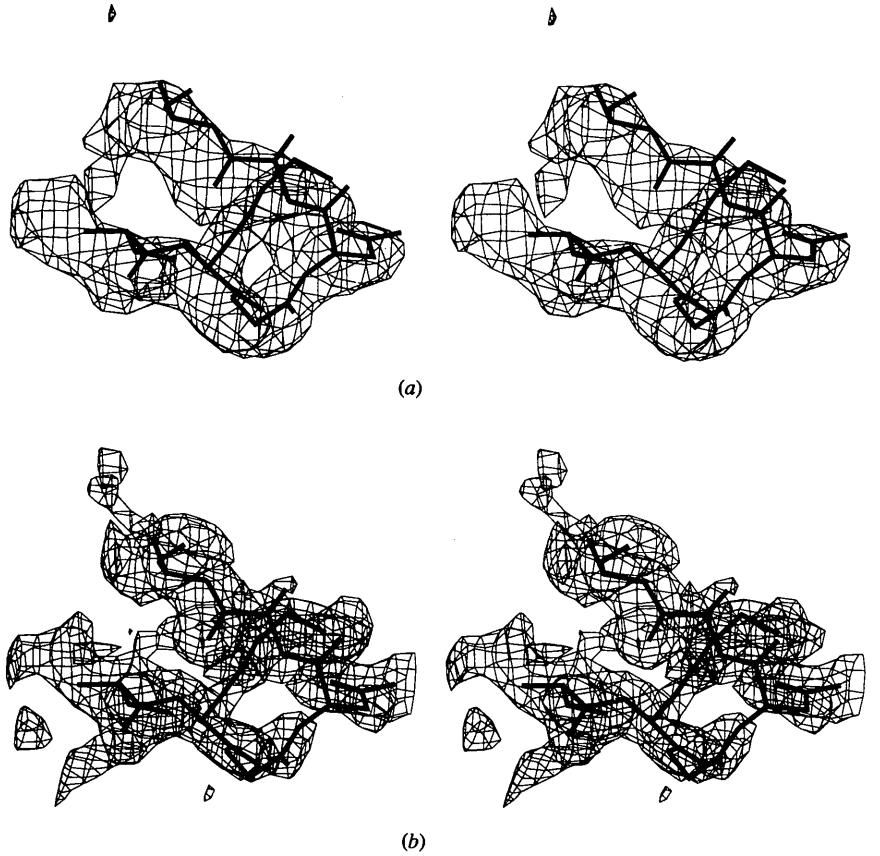

Fig. 3. A representative sample (residues 18–23) of the electron density recovered for the thaumatin model after omitting residues 1–60 (out of 207). (*a*) Electron density recovered using the holographic method. (*b*) The corresponding electron density calculated with standard Fourier techniques, using unweighted $(F_{true} - F_{calc})$ coefficients, and contoured at 1.5σ.

recovery brought about by errors in the scaling of the data. From these results, it is clear that an incorrect estimate of the scale factor can seriously compromise the quality of the recovery and, given the difficulties in accurately placing a data set on an absolute scale, this sensitivity to scaling may prove to be an important limitation to the use of the holographic technique. In our experience, if high-resolution data are available, a Wilson-plot analysis (Wilson, 1949) of the data produces a sufficiently accurate estimate of the scale factor for purposes of applying the holographic method. Similarly, if a model of the structure being worked on is available, the experimentally measured structure factors can be scaled to structure factors calculated from the model. However, in the absence of high-resolution data or a reliable model, placing the data on an absolute scale can be a difficult problem. Under these conditions, it might be necessary to systematically try a range of scale factors and empirically choose the scale factor based on the quality of the recovered electron density.

The holographic method also requires an estimate of the $F(000)$ term and the effects on the electron-density recovery brought about by changes in the magnitude of this term are shown in Fig. 5(*b*). The $F(000)$ term specifies the number of electrons in the unit cell. Consequently, estimating the magnitude of $F(000)$ requires knowledge of the molecular formula, the number of molecules in the unit cell and an estimate of the solvent content. Our ability to estimate $F(000)$ seems adequate for applying the holographic method. Interestingly, *EDEN* is far less tolerant of an overestimate of $F(000)$ than of an underestimate. In fact, when a relatively large amount of density needs to be retrieved, an underestimated $F(000)$ term actually improves the

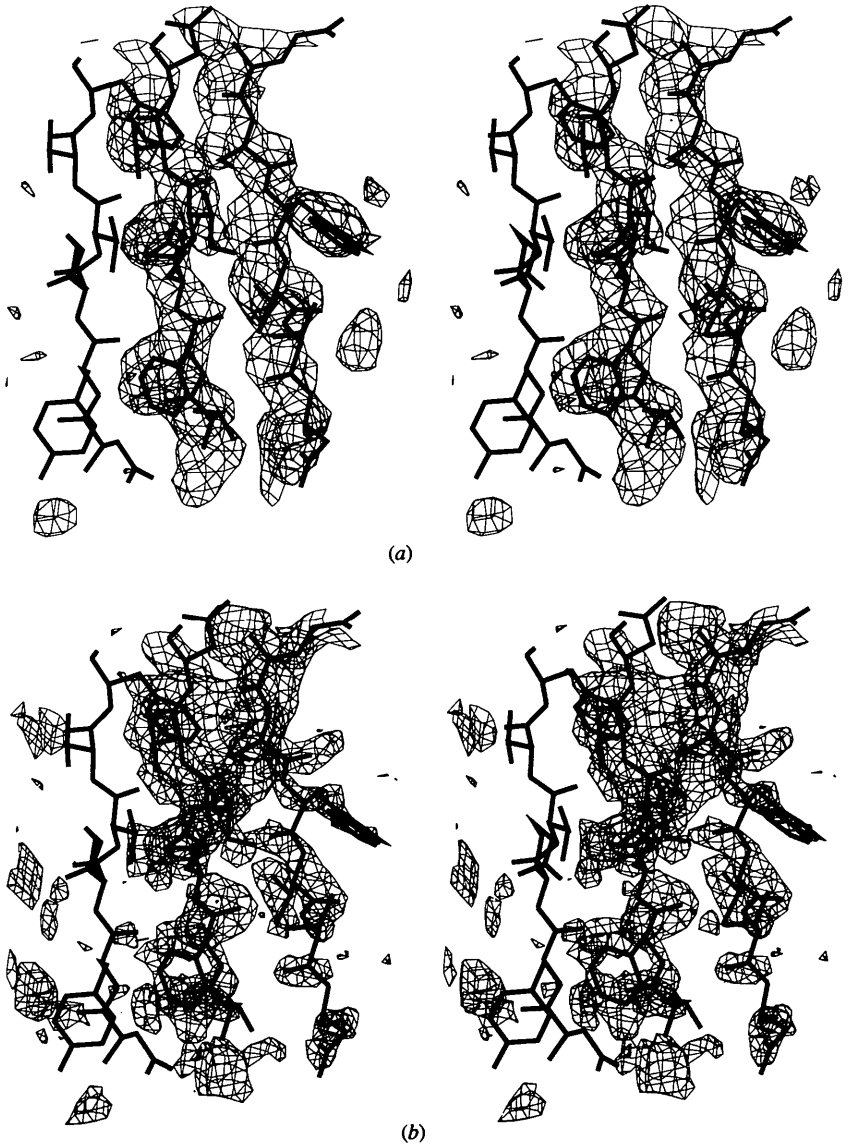

Fig. 4. A sample (residues 2–7, 35–40) of the electron density recovered for the thaumatin model after omitting residues 1–120. (*a*) Electron density recovered using the holographic method, using a solvent mask. Note that the leftmost of the three strands is part of the known part of the model and, consequently, electron density for this region is not recovered. (*b*) The corresponding electron density calculated with standard Fourier techniques, using unweighted $(F_{true} - F_{calc})$ coefficients, and contoured at 1.5σ.

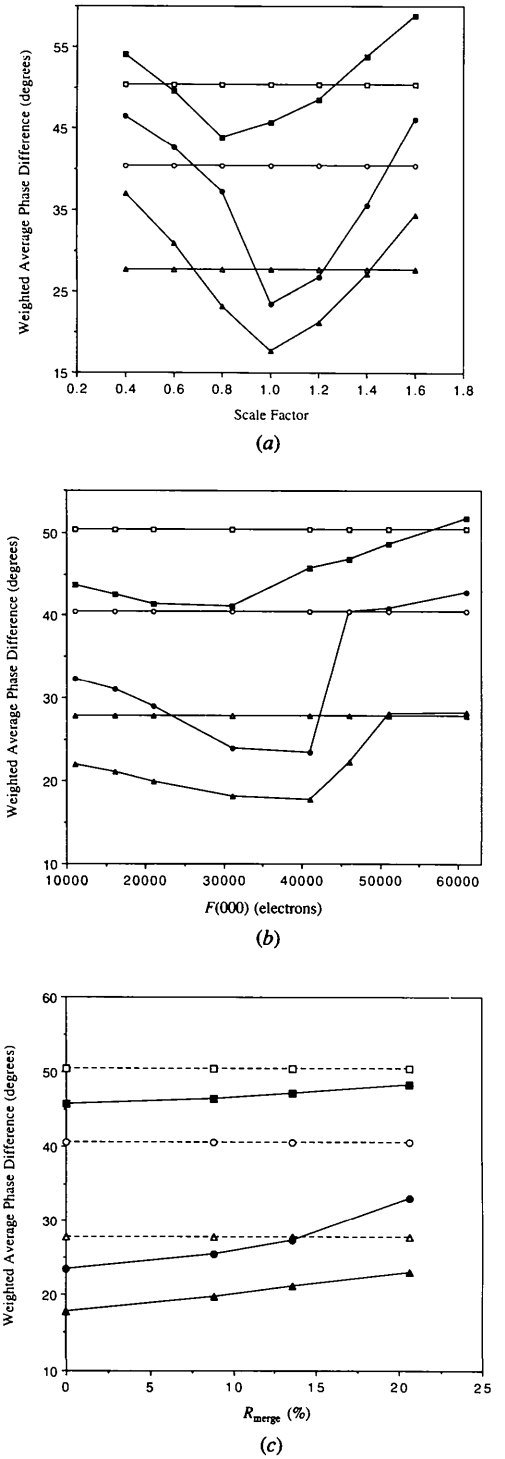

Fig. 5. The effect on the recovery of electron density corresponding to 30 (triangles), 60 (circles), and 90 (squares) residues, brought about by artificially induced errors in the data. The open and closed symbols show the weighted average phase difference before and after recovery with *EDEN*, respectively. The weighted average phase difference is used to quantify the effects brought about by (*a*) errors in the factor used to place the data on an absolute scale, (*b*) errors in the magnitude of the $F(000)$ term, and (*c*) noise in the structure-factor amplitudes.

recovery. This is probably because, by lowering the $F(000)$ term, the positivity constraint helps the recovery, at least for a model protein. We surmise that a low $F(000)$, together with the positivity constraint, prevents the recovery of spurious density.

Noise was added to the data in such a way that the noisy intensities would form a normal distribution around the true intensity. To do this, the F_{true} structure-factor amplitudes were squared to obtain intensities, and a Box–Muller transformation (Press, Flannery, Teuolsky & Vetterling, 1989) was used to obtain random deviates with a normal (Gaussian) distribution around each of the correct intensities. The amount of noise in the data was adjusted by controlling the width of the normal distribution. The noise in the data was assessed by computing two noisy data sets using the same parameters and calculating the '$R_{merge}(|F|^2)$' between them:

$$R_{merge}(|F|^2) = \sum_{\mathbf{h}} \left| |F_1|^2 - |F_2|^2 \right| \Big/ \left(\sum_{\mathbf{h}} |F_2|^2 \right). \quad (20)$$

The effects of noisy F_{true} amplitudes on the recovery of electron density is shown in Fig. 5(*c*). As can be seen in Fig. 5(*c*), the recovery of electron density is fairly insensitive to Gaussian noise in the data, at least within the level of noise that might be expected from experimental data.

Finally, we have tested the ability of the holographic method to cope with errors in the atomic positions of the 'known' part of the model. Positional errors were introduced by subjecting the thaumatin model to constrained molecular dynamics, using the program *X-PLOR*, with the *CHARMM* force field (Brooks *et al.*, 1983), and by periodically writing coordinates during the simulation. Parts of these coordinates were then used as the known region of thaumatin, and used to recover density corresponding to omitted parts of the structure. The coordinates derived from the dynamics simulation were compared with the unaltered coordinates by calculating the root-mean-square (r.m.s.) deviation between each pair of atoms in the two coordinate sets.

As expected, the accuracy of the 'known' part of the structure influences the recovery of the unknown electron density. Figs. 6(*a*)–(*c*) shows the effect of positional errors in the known part of the structure on the phase correction brought about by running *EDEN*. Note that, because of the errors in the known part, a perfect recovery is not possible. Thus, Fig. 6 also shows the maximum possible improvement in phases obtainable, given the errors in the structure. As expected, as the number of deleted residues increases, the recovery is increasingly sensitive to errors in the known part. Significant correction of the average phases was possible only when fewer than 60 residues were deleted and when the positional error in the known part was less than 0.2 Å.

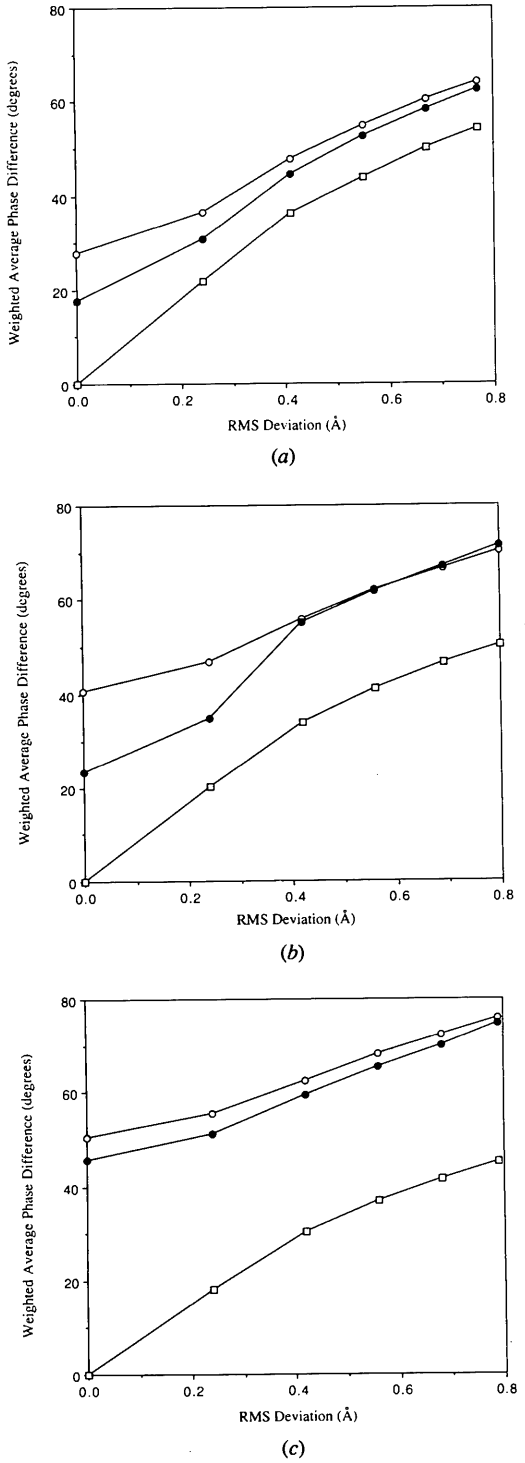

Fig. 6. The effect of errors in the atomic coordinates of the 'known' part of the model was evaluated after deleting (*a*) 30, (*b*) 60, and (*c*) 90 residues from the N terminus of the thaumatin model. The average weighted phase difference before (open circles) and after (closed circles) recovery with *EDEN* is plotted against the root-mean-square deviation (Å) between the correct thaumatin model and the model with the positional errors. Given the errors in the known part of the model, a perfect recovery is not possible. The open squares show the best possible recovery given these errors.

3. Experimenting with real data: the structure determination of H124L, P117G, P47G staphylococcal nuclease

Staphylococcal nuclease is a 149-residue enzyme that catalyzes the hydrolysis of the phosphate backbone of DNA and RNA; this protein has been used extensively to study enzyme kinetics and thermodynamics. We have obtained a mutant of staphylococcal nuclease with the following single-site substitutions: H124L, P117G and P47G. The two proline to glycine mutations occur in loops and were expected to lead to substantial local changes in conformation. As a preliminary test of the holographic method using experimental data, we determined the structure of this triple mutant, using *EDEN* to calculate all of the electron-density maps. We also used the experimental data from this protein to test *EDEN*'s ability to recover the electron density corresponding to large omitted regions.

3.1. *Data collection and initial phasing by molecular replacement*

Crystals of H124L, P117G, P47G staphylococcal nuclease were obtained using conditions and methods similar to those described by Loll & Lattman (1989). This triple mutant crystallizes in the tetragonal space group $P4_1$, with cell parameters $a = b = 49.18$, $c = 63.51$ Å.

Data to approximately 1.9 Å resolution were collected on a Rigaku R-AXIS II C area detector and reduced using the *R-AXIS* data-processing software. 30773 observations of 10604 independent reflections were measured. $R_{\rm merge}(I)$* for the resulting data set was 4.74% [based on all observations where $I/\sigma(I) > 0.0$]. The data are 94.5% complete to 1.95 Å resolution and 90.0% complete in the resolution shell between 2.0 and 1.95 Å resolution.

The structure was solved by molecular replacement using *X-PLOR* (Brünger, 1992*a*), using the crystal structure of native staphylococcal nuclease (Hynes & Fox, 1991) as the search model. The orientation of the model was determined by a rotation search followed by Patterson correlation refinement and the position of the model in the unit cell was determined by a translation search. The top peak from the translation search was 10σ above the mean, with an R factor of 43.1% (for data between 8.0 and 3.5 Å resolution) and showed reasonable packing interactions with symmetry-related molecules. The position and orientation of the model was further optimized by rigid-body refinement, which brought the R factor down to 30.0% (data between 8.0 and 3.0 Å).

$$*R_{\rm merge}(I) = 100.0 \sum_{h,k,l} \sum_{i} |\bar{I}(h,k,l) - I_i(h,k,l)| \Big/ \sum_{h,k,l} \sum_{i} I_i(h,k,l),$$

where the sums are taken over all reflections and over all observations of the same reflection.

3.2. *Preparation of electron-density maps using the holographic method*

All of the electron-density maps used for manual positioning of the model were obtained using the holographic method as implemented by the *EDEN* software package. The factor that places the observed structure factors on an absolute scale was found to be approximately 1.86, as determined from a Wilson plot of the data between 3.0 and 1.95 Å. This estimate is generally consistent with the scale factor that was calculated by comparing the observed structure factors with the structure factors calculated from the model ($k = 2.07$).

The $F(000)$ term was estimated based on the number of each type of atom in the model, and based on the estimate that around 44% of the unit cell is occupied by the protein. The $F(000)$ term used was 60 157 e.

Essentially all of the maps used in the course of the refinement were prepared in the 'completion' mode, as described above. In the final stages of the refinement, we made use of a crude solvent mask when using *EDEN* to recover missing electron density. The solvent region was defined as starting 3.0 Å away from the protein model and this region was assigned an electron density of 0.33 $Å^{-3}$.

It should be noted that many of the electron-density maps were calculated using data greater than 7 Å resolution. In retrospect, this was a mistake, and the recovery of the missing density would undoubtedly have benefited from the information included in the low-resolution reflections.

3.3. *Refinement of the staphylococcal nuclease model*

The refinement of the staphylococcal nuclease triple mutant was carried out using the *X-PLOR* versions 3.0 and 3.1 software packages (Brünger, 1992*a*,*b*) using the parameters described by Engh & Huber (1991). The refinement, excluding the initial rigid-body refinement, was done with approximately 9% of the data omitted for future calculations of the free R factor (Brünger, 1992*c*). During the initial refinement, the individual B factors from the native staphylococcal nuclease model were retained and the following residues were omitted: 123–125, 45–50, 116–118 and 104. The first three of these omissions correspond to regions surrounding the three residues that differ between native staphylococcal nuclease and the triple mutant being refined. Residue 104 is a valine buried in the core of the protein and is therefore expected to be defined by clear electron density This residue was omitted as a control.

A typical round of refinement consisted of the following sequence of steps: positional refinement, simulated annealing, positional refinement, individual atomic B-factor refinement and manual refitting of the model using the electron-density maps generated by *EDEN*. In the final stages of the refinement, water molecules were added to the model. At the current stage of refinement, the staphylococcal nuclease model consists of residues 6 through 44 and 51 through 141 and includes 30 water molecules. The crystallographic R factor for this model is 18.0% for the data greater than $2\sigma(F)$ and between 6.0 and 1.95 Å resolution, with $R_{free} = 25.5\%$. The root-mean-square deviations from ideal bond lengths and angles are 0.010 Å and 1.55°, respectively.

3.4. *Results of the refinement of* H124L, P47G, P117G *staphylococcal nuclease*

The structure determination of the H124L, P47G, P117G staphylococcal nuclease triple mutant was the first realistic test of the holographic method's ability to recover missing electron density using experimentally determined structure-factor amplitudes. Our primary concerns were whether we would be able to estimate correctly the $F(000)$ term for an experimentally measured data set, whether we would be able to place the data accurately on an absolute scale, and generally to test *EDEN*'s ability to recover electron density based on experimentally derived data.

In practice, *EDEN* was able to recover the density corresponding to the mutated regions of the staphylococcal nuclease structure well enough to correctly model these regions. Figs. 7(*a*),(*b*) show some sample electron density corresponding to the regions around leucine 124 and around glycine 117. This map was calculated early in the refinement, before either of the two regions was modeled. Very little electron density was recovered for the region surrounding glycine 47. However, even in highly refined models of staphylococcal nuclease this region is characterized by poor or non-existent electron density.

3.5. *Recovery of parts of the staphylococcal nuclease using target functions*

In §2.2, we showed that a large fraction of a protein model can be recovered to fairly high accuracy by the use of the positivity constraint and, even more, by the use of solvent masks. In real proteins, there are no voids; the interstices among the molecules are largly filled with solvent. The holographic method takes advantage of this fact by incorporating solvent targets. As a test of *EDEN*'s ability to recover large amounts of electron density, 40 residues were deleted from the N terminus of the refined staphylococcal nuclease model, the model was submitted to simulated annealing, and *EDEN*, with the use of a solvent target, was used to recover the electron density corresponding to the omitted part of the model. The solvent target function was prepared as described in §1.4. Shown in Fig. 8 is a sample of the recovered electron density. The continuity of the recovered electron density is quite good and there is relatively little spurious electron density.

4. Discussion and concluding remarks

In macromolecular X-ray crystallography, the electron density corresponding to the missing part of a model is traditionally recovered by 'omit maps', using the difference Fourier method. In paper II, we rederived the well known fact that a difference Fourier map, computed with unweighted $(F_o - F_c)$ coefficients, contains an equal superposition of the missing electron density and of its holographical dual image (with respect to the known part of the protein). The holographic method, in contrast, is capable of changing the phases of the structure factors. Conseqently, if there is sufficient additional knowledge or, in other words, if there are

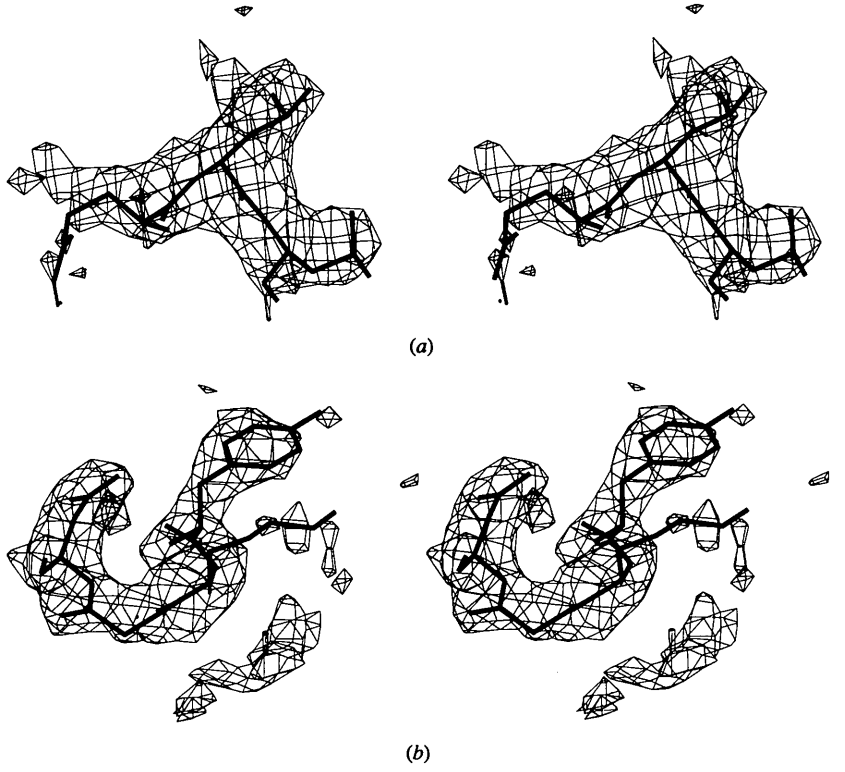

Fig. 7. Electron density recovered for (*a*) residues 123–125, and (*b*) residues 115–118 of staphylococcal nuclease.

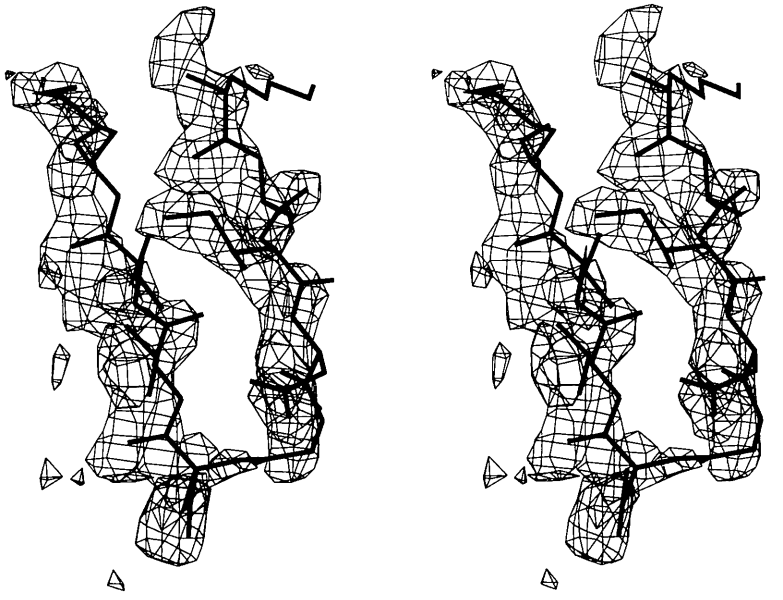

Fig. 8. Representative density (residues 16–24) recovered for staphylococcal nuclease after deleting residues 1–40. The *EDEN* recovery was carried out using a low-resolution solvent target.

sufficient constraints on the electron density, the holographic method is capable of eliminating the dual image entirely. An omit map calculated by the holographic method has the potential of recovering the missing electron density in its entirety with none of the dual image.

We have implemented a fast algorithm for applying the holographic method, and we have described some preliminary tests to evaluate its potential for use in the field of macromolecular X-ray crystallography. Our experience with synthetic data calculated from a thaumatin model is very promising. The application of the holographic method, after deleting regions of the thaumatin model, consistently led to an improvement in the phases, and a corresponding improvement in the electron density recovered for the omitted region of the structure. For example, after deleting 60 residues from the N terminus of thaumatin (approximately 30% of the model), the application of the holographic method led to a near-perfect recovery of this region's electron density, with a 17° average improvement in the phases (Figs. 2 and 3). Our work with the synthetic thaumatin data also demonstrates the power inherent in using knowledge of the solvent region. The application of a solvent mask dramatically improves the recovery of the electron density (Fig. 2). For example, we were able to completely recover the electron density corresponding to 120 residues deleted from the N terminus of thaumatin ($\sim$ 58% of the protein), with a 27° improvement in the phases (see Figs. 2 and 4). These results are somewhat comparable to those of Béran & Szöke (1995).

We have also presented the results of tests using experimentally derived data. Clearly, working with real data presents problems that do not exist when working with synthetic data. There are inaccuracies in the structure-factor amplitudes and in the 'known' part of the structure, as well as potential errors in estimating the $F(000)$ term or the scale factor that places the data on an absolute scale. Another important difference between real crystals and our synthetic models is that in real crystals the solvent volume is not empty but instead contains an electron density of roughly 0.33 $Å^{-3}$. The presence of solvent has several consequences. First, the contrast (measured in terms of the difference in average electron densities) between the protein and the solvent region is decreased by about a factor of three compared to the model. Second, the positivity of the electron density is not an effective constraint in the solvent region. Third, while the structure in the solvent region is not as pronounced as in the interior of the protein, it is still present. Density variations come from ordered water molecules, from the presence of counterions, and from the different adherence of the solvent to hydrophilic and hydrophobic residues. The electron density of a crystal that is truly based on experiment should have all of these features.

Our experience with experimentally measured diffraction data from a mutant of staphylococcal nuclease was encouraging. We were successfully able to refine the structure of this triple mutant using *EDEN* to calculate all of the electron-density maps. While this was a relatively simple exercise, it confirmed what was suggested by our experience with the synthetic thaumatin data: that the holographic method, as implemented by *EDEN*, is robust enough to work with the inaccuracies of real data and real models. These results also show that we are able to estimate the $F(000)$ term and the scale factor accurately enough to recover meaningful electron density.

The holographic method readily lends itself to the incorporation of a solvent mask and, more generally, to restraining the electron density of specific regions of the unit cell to specified values. This ability to 'suggest' electron density can be applied to regions of suspected solvent or to parts of the crystal that are particularly well known. We have illustrated the use of a solvent target using the experimental data from the staphylococcal nuclease mutant. Specifically, we used a target solvent function to aid in the recovery of electron density corresponding to 40 residues that were deleted from the N terminus of the staphylococcal nuclease model. Although this example is somewhat contrived, in the sense that it assumed *a priori* knowledge of the solvent region, it clearly illustrates the power of the solvent target function.

To conclude, we would like to mention an unsolved problem and sketch some of the directions we plan to take in the near future. A major unsolved problem is the absence of a good overall measure of the quality of the recovered electron density. In our studies of synthetic data, we found that our crystallographic R factor does not correlate well with the quality of the recovered density, as measured by $\Delta\varphi$ and $\Delta\rho$ [(18) and (19)]. Other measures of the correlation between F_{obs} and F_{calc} fare no better. (Of course, $\Delta\varphi$ and $\Delta\rho$ cannot be applied to real proteins.) Moreover, when the grid resolution is increased with respect to the data resolution, *EDEN* readily reduces the cost function, f_{total}, in (12) to near zero even though the electron density may be meaningless. One possibility is that the R factor is decreased by trading off the correctness of the electron density in the volume of the protein with that of the solvent region. While the free R factor (Brünger, 1992*c*) may be of greater reliability, it would be extremely valuable to find a simple, reliable and automatic way to evaluate the quality of electron-density maps.

Several additions will be made in our implementation of the holographic method. First, the quadratic term that was neglected in (4) will be added. This should allow us to set all the weights in (7) to unity and to use our algorithm for phase extension, *i.e.* to recover the protein structure at successively higher resolutions. Also, since we suspect that a major source of the remaining error in the electron-density recovery is the tendency of the

algorithm to pull the electron density towards the grid points (§2.1), the quadratic term, by allowing us to use higher gridding resolution (at a given data resolution), should lead to a more accurate recovery. We also plan to implement an algorithm for multiple isomorphous replacement (MIR). While the traditional MIR algorithm is equivalent, in principle, to its holographic version (see paper II), the incorporation of positivity constraints and the possibility of using solvent 'targets' make our algorithm sufficiently different to be worth exploring. One promising avenue is to use MIR phases as a starting point for the holographic algorithm. Finally, the manner in which *EDEN* solves for electron density lends itself to the incorporation of additional constraints on the recovered electron density. This should allow us to make use of any non-crystallographic symmetry that might be present in the asymmetric unit. In addition, we will attempt to incorporate chemical knowledge in the form of partial molecular models, to be automatically fitted into the electron density.

The authors would like to thank Jim Brase and Rick Twogood for their support of the computational effort. Work at the Lawrence Livermore National Laboratory was performed under the auspices of the US Department of Energy under Contract no. W-7405-ENG-48. The work at the University of California at Berkeley has been supported by NIH grant DC 00145 to SHK, and by the Director, Office of Energy Research, Office of Health Effects Research, of the US Department of Energy under Contract no. DE-AC 03-76SF00098.

APPENDICES

In these Appendices, the program *EDEN* is described. This is followed by some information on space groups and symmetry, intermediate file formats, auxiliary programs and other computational issues. The programs are copyright© of the Regents of the University of California. Enquiries about the programs should be directed to the authors (e-mail address szoke2@llnl.gov).

*A*1. Description of *EDEN*

The program *EDEN* is discussed below in terms of initializations and four aspects of the outer loop that solves the linearized holographic equation: set-up of arrays; the inner loop that minimizes the cost function; reports and output of the current solution; checks to determine outer-loop continuation.

Initializations

Parameters. There are about 30 input parameters, including: unit-cell parameters (a, b, c, angles and symmetry group); file information (names of .fobs and .fcalc files, as well as names of files containing physical space models, targets, weights and masks); computational parameters [grid spacing, data resolution for reading data, values for η and for δ in (9), limiting crystallographic R value, Lagrange multipliers, criteria for quitting the conjugate-gradient solver and a factor for scaling the .fobs to the .fcalc values].

Arrays in physical space. All of these are generally identified on a double grid: a simple plus a body-centered grid. The electron/voxel arrays are: the known model, $n_{p,\text{known}}$; the electrons found in a particular outer-loop iteration, n_p; and the total electrons found in all outer-loop iterations, $n_{p,\text{sum}}$. At the start of the run, n_p and $n_{p,\text{sum}}$ are initialized to zero, while $n_{p,\text{known}}$ is set equal to the known input model (in correction mode) or to 0 (completion mode). There are also arrays required by the solver, including predetermined minima and maxima for the solver limit, and $n_{p,\text{type}}$, characterizing the permitted changes in the electron density at each point in space. At the start of the *EDEN* run, $n_{p,\min} = -n_{p,\text{known}}$, $n_{p,\max}$ = (a large number) and $n_{p,\text{type}}$ = 'bounded'. Wherever a mask is in effect, $n_{p,\text{type}}$ = 'fixed' (which precludes the accumulation of n_p). If λ_{space} is non-zero, $n_{p,\text{target}}$ and $\tilde{w}_p$ are also read in for use according to (10).

Arrays in reciprocal space. These include $F(\mathbf{h})$, read from the .fobs file, scaled to the model, if necessary, and apodized according to (9) to yield $F'(\mathbf{h})$, and $R_0(\mathbf{h})$ (complex), read from the .fcalc file. R_0 is the initial value of R, corresponding to (1). If the model from which R_0 is derived is of higher resolution than is needed for the recovery (*i.e.* the grid resolution), an appropriate apodization, δ_{calc}, is applied to R_0. Both of these files must have an $F(000)$ term, whose value is the total number of electrons and the number of electrons in the known part of the unit cell, respectively. Both $F(000)$ terms include all appropriate solvent electrons. The symmetry-forbidden ($\mathbf{h}$) terms are explicitly set to 0 in $F'(\mathbf{h})$. Only those $R_0(\mathbf{h})$ for which there is a corresponding $F'(\mathbf{h})$ are used in *EDEN*. See also Appendix *A*2, below. Other arrays set up for use in the FFT's are exponential factors efac($\mathbf{h}$) defined below [(*A*4)] and a crystallographic weight factor $w_{\text{crys}}(\mathbf{h})$ that is 0.5 for the $F(000)$ term and 1.0 for all other terms (in a $P1$-expanded set of structure factors). However, to simplify the equations below, we shall assume that a crystallographic weight factor has already been included in $w(\mathbf{h})$, defined in (8).

FFT initializations. We have used the routine *cfft*99 from Cray Research (written by Clive Temperton, ECMWF, 11/78, revised by Russ Rew, 9/80 and Dave Fulker, 11/80).

Set up of arrays

At the start of each outer-loop iteration, the current value of $R(\mathbf{h})$ is recalculated, using the sum of $R_0(\mathbf{h})$ and the properly apodized Fourier transform of $n_{p,\text{sum}}$. $R(\mathbf{h})$ is

written as the current most complete reciprocal-space model. For the following discussion, it is useful to combine (13), (16) and (7) into the following form:

$$f_{\text{eden}} = 1/2 \sum_{\mathbf{h}} w(\mathbf{h})^2 \{2\text{Re}[R\exp^*(\mathbf{h})\text{DFT}^+(n_p)] - H_{\text{diff}}(\mathbf{h})\}^2, \quad (A1)$$

where the variable $R\exp(\mathbf{h})$ is defined as

$$R\exp(\mathbf{h}) = \text{efac}^*(\mathbf{h})R(\mathbf{h})/[|F'(\mathbf{h})| + |R(\mathbf{h})|] \quad (A2)$$

$$H_{\text{diff}}(\mathbf{h}) = [|F'(\mathbf{h})| - |R(\mathbf{h})|]. \quad (A3)$$

$w(\mathbf{h})$ is defined in (8) [but includes the crystallographic weight factor, $w_{\text{crys}}(\mathbf{h})$] and

$$\text{efac}(\mathbf{h}) = \exp[-\eta(\pi\Delta r|\mathcal{F}^T\mathbf{h}|)^2]\exp[2\pi i\,\text{off}(\mathbf{h})]. \quad (A4)$$

The offset array, off(**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,

$$\text{off}(\mathbf{h}) = 1/2(h/P_a + k/P_b + l/P_c). \quad (A5)$$

It is clear from this formulation that the arrays $R\exp(\mathbf{h})$ and $H_{\text{diff}}(\mathbf{h})$ may be updated prior to the inner-loop iteration, leaving only n_p as the variable array.

Inner loop, minimizing the cost function

The linearized conjugate-gradient solver, *lgetsol*, searches the solution space for a minimum. Each time a step is taken in solution space, producing a new set of n_p, *lgetsol* calls an *EDEN*-specific function. This, the heart of *EDEN*, does the FFT's to compute the components of the cost function f_{eden} from (A1) and calculates f_{space} from (10). [The f_{null} term of (11) was explored in an earlier version of *EDEN* and is not currently in the program.]

The gradients of the current solution are also required by the solver. For this purpose, it is useful to write f_{eden} in yet another form:

$$f_{\text{eden}} = 1/2 \sum_{\mathbf{h}} [w(\mathbf{h})^2 r(\mathbf{h})^2], \quad (A6)$$

where

$$r(\mathbf{h}) = [R^*(\mathbf{h})\text{efac}(\mathbf{h})\text{DFT}^+(n_p) + R(\mathbf{h})\text{efac}^*(\mathbf{h})\text{DFT}^-(n_p)]/[|F'(\mathbf{h})| + |R(\mathbf{h})|] - |F'(\mathbf{h})| + |R(\mathbf{h})|. \quad (A7)$$

Using the fact that the derivative of $\text{DFT}^+(n_p)$ with respect to a particular n_p is $\exp[2\pi i\mathbf{h}\cdot\mathbf{x}_p]$, where $\mathbf{x}_p$ is the fractional coordinate of the point p in the unit cell, we get

$$\text{d}r(\mathbf{h})/\text{d}n_p = \{R^*(\mathbf{h})\text{efac}(\mathbf{h})\exp[2\pi i\mathbf{h}\cdot\mathbf{x}_p] + R(\mathbf{h})\text{efac}^*(\mathbf{h})\exp[-2\pi i\mathbf{h}\cdot\mathbf{x}_p] \times |F'(\mathbf{h})| + |R(\mathbf{h})|^{-1}\} \quad (A8)$$

and from this, the full gradient of f_{eden}, g_{eden} is

$$g_{\text{eden}} = 2\text{Re}(\text{DFT}^-\{[w(\mathbf{h})^2 r(\mathbf{h})R(\mathbf{h})\text{efac}^*(\mathbf{h})] \times [|F'(\mathbf{h})| + |R(\mathbf{h})|]^{-1}\}), \quad (A9)$$

or, more concisely from (A2),

$$g_{\text{eden}} = 2\text{Re}\{\text{DFT}^-[w(\mathbf{h})^2 R\exp(\mathbf{h})r(\mathbf{h})]\}. \quad (A10)$$

From (10), we have the gradient of f_{space} :

$$g_{\text{space}} = \lambda_{\text{space}} 2P \sum_{p=1}^{P} \tilde{w}_p^2 (n_p - n_{p,\text{target}}). \quad (A11)$$

The full gradient is $g_{\text{eden}} + g_{\text{space}}$. From among a large variety of possible conditions for stopping, the conjugate-gradient solver most commonly quits because the gradient in solution space has gone down sufficiently. (This condition is not mentioned by Goodman, Johansson & Lawrence, 1993.) Typically, the criterion is that the gradient has gone down by two orders of magnitude. Other causes for stopping the search are that the cost function has fallen enough during this call to the solver or that *lgetsol* determines that it has reached a minimum.

Reports and output of current solution

On returning from the solver, *EDEN* interprets return codes and monitors the progress of the cost function. The array $n_{p,\text{sum}}$ is updated by adding in n_p, the electrons found in the iteration just completed, while $n_{p,\text{min}}$ is updated by subtracting n_p from it (thus enabling the solver to correct for overshoot). *EDEN* writes the sum of $n_{p,\text{known}}$ and $n_{p,\text{sum}}$. It also calculates the structure factors corresponding to the present solution and recalculates and reports the crystallographic R factor as well as other figures of merit.

Checks to determine loop continuation

After each inner-loop completion, *EDEN* decides whether to continue iterating or to terminate. The usual condition that causes *EDEN* to terminate is that the cost function is 'stuck' (it does not decrease over five iterations). It is also common for a user to terminate *EDEN* by force if current figures of merit suggests that it is not progressing. *EDEN* quits gracefully. In some cases, *EDEN* terminates if the crystallographic R factor has fallen below an input value, R_{stop} (typically 3% but subject to change as an input parameter). Other terminating conditions have been built into *EDEN* but are not encountered in practice. These include: the conjugate-gradient solver returns an error condition; a maximum of 50 outer-loop iterations has been completed (typically there are 5–15 outer-loop iterations). In trivial test problems only, *EDEN* terminates when the residual set of electrons/voxel recovered in an iteration is essentially zero.

*A*2. Symmetry and space groups

The space groups supported by *EDEN* are $P1$, $P2_1$, $P2_12_12_1$, $P4_1$ and $P4_3$. It would require only a small effort to extend this set to include all but the hexagonal and trigonal crystal systems. The unit cell need not be orthorhombic, although large deviations from angles of 90° will affect the shape of the Gaussian basis functions significantly.

In order to enforce and preserve crystallographic symmetry during the operation of *EDEN*, the following procedures are observed: the minimum region of reciprocal space is read into the program; an octant for $P2_12_12_1$, $P4_1$ and $P4_3$; a quadrant for $P2_1$ and a half-sphere for $P1$, with the requisite edges. The set of structure factors is expanded to the half-sphere, using the appropriate symmetry relations. When reading the experimental structure factors, symmetry-related hard zeros are added to the data, where appropriate. With regard to model structure factors, only those values of (hkl) for which there are experimental readings are used.

Calculations are carried out for the full unit cell. Whenever electron densities are read in or recalculated, the resulting array is 'symmetrized' by averaging over equivalent points in the asymmetric units. Such averaging should not be necessary and any significant deviations of the symmetrized from the incoming values are thus reported as warning messages. Both in theory and in practice, asymmetric density values are the (rare) consequences of numerical instabilities. However, the check is useful for highlighting gross anomalies, such as garbage data in files, for example. At present, non-crystallographic symmetry is not included in *EDEN*.

*A*3. Intermediate file structure

It is only *REGRID* that converts electron densities to the output format of *X-PLOR* (the so-called .sol files). All other electron-density maps in physical space are stored as binary data in a format designed for use with the signal processing program, *VIEW* (Brase, Miller & Wieting, 1988). The purpose of this usage is (*a*) to retain the compactness and precision of binary data and (*b*), more importantly, to enable the analysis and display of data using this program. Some of the capabilities of *VIEW* that have been exploited in our holographic method studies are: clipping and thresholding; searching for peaks; comparing density maps; performing statistical analysis of maps. Parts of the unit cell are visualized by 'slicing' the cell perpendicular to an axis (typically the *c* axis) and displaying the slices, using false color to indicate density at each position. While these display methods are much less sophisticated than those of *FRODO*, for example, they do enable quantitative analysis of electron densities: there are no 'chicken wire' cutoffs and actual (non-negative) electron densities, in units of electrons/voxel or, after *REGRID*, in units of $Å^{-3}$, may be read off at any point within the unit cell.

*A*4. Auxiliary programs

There are a number of auxiliary programs used in conjunction with *EDEN*. *ADD* adds or subtracts comparable entries in two reciprocal-space model files. *APODIZE* supplies the best-fit slope of $\log(F^2)$ *vs* $(1/\Delta r^2)$, using either experimental diffraction intensities or reciprocal-space model data. This is similar to a Wilson plot but does not use atomic scattering factors. *BACK* converts a reciprocal-space model into a best-fit physical space model in electrons/voxel using (17). *DPHASE* finds the average phase differences between two reciprocal-space models by (18). *DRHO* calculates the variance between two sets of *VIEW* electron/voxel data by (19). *FORTH* transforms electron/voxel data into reciprocal space, inverting the effect of *BACK*. *REGRID* converts data in electrons/voxel to $Å^{-3}$ and spreads data onto a finer grid. It also uses FFTs.

*A*5. Computational issues

Although higher-level code in *EDEN* and the auxiliary programs is written in C, some of the critical underlying code (*lgetsol*, cost-function calculations, the FFT's) is all Fortran. The programs have been run on a variety of platforms, including SGI Iris and Indigo, IBM 6000 and HP 9000 machines. On the SGI Iris, *EDEN* typically runs for 15 min to 2 h, depending on the gridding resolution; *BACK* and *REGRID* run for a few minutes each. Other utilities are completed within a minute. The IBM 6000 and HP 9000 machines are typically four times faster than the older SGI Iris. In all practical cases, about 80% of running time is spent on the FFT routine. Regridding large electron density maps or using a fine regrid factor (4 rather than 2) may cause *REGRID* to encounter memory problems. These are circumvented by handling the unit cell on a piecewise basis, producing two ore more *X-PLOR* .sol files for successive sections of the unit cell.

References

BÉRAN, P. & SZÖKE, A. (1995). *Acta Cryst.* A**51**, 20–27.

BRASE, J. M., MILLER, V. J. & WIETING, M. G. (1988). *The VIEW Signal and Image Processing System.* Report UCID-21368. Lawrence Livermore National Laboratory, Livermore, CA 94550, USA.

BROOKS, B., BRUCCOLERI, R., OLAFSON, B., STATES, D., SWAMINATHAN, S. & KARPLUS, M. (1983). *J. Comput. Chem.* **4**, 187–217.

BRÜNGER, A. T. (1992*a*). *X-PLOR. A System for Crystallography and NMR.* Version 3.0. New Haven: Yale University.

BRÜNGER, A. T. (1992*b*). *X-PLOR. A System for Crystallography and NMR.* Version 3.1. New Haven: Yale University.

BRÜNGER, A. T. (1992*c*). *Nature (London)*, **355**, 472–475.

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

ENGH, R. A. & HUBER, R. (1991). *Acta Cryst.* A**47**, 392–400.

GOODMAN, D. M., JOHANSSON, E. M. & LAWRENCE, T. W. (1993). *Multivariate Analysis: Future Directions*, edited by C. R. RAO, ch. 11. Amsterdam: Elsevier.
HODEL, A., KIM, S.-H. & BRÜNGER, A. T. (1992). *Acta Cryst.* A**48**, 851–858.
HYNES, T. R. & FOX, R. O. (1991). *Proteins: Struct. Funct. Genet.* **10**, 92–105.
JONES, T. A. (1985). *Methods Enzymol.* **115**, 157–171.
LOLL, P. J. & LATTMAN, E. E. (1989). *Proteins: Struct. Funct. Genet.* **5**, 183–201.
MAALOUF, G. J., HOCH, J. C., STERN, A. S., SZÖKE, H. & SZÖKE, A. (1993). *Acta Cryst.* A**49**, 866–871.
OGATA, C. M., GORDON, P. F., DE VOS, A. M. & KIM, S.-H. (1992). *J. Mol. Biol.* **228**, 893–908.
PFLUGRATH, J. W., SAPER, M. A. & QUIOCHO, F. A. (1984). *Methods and Applications in Crystallographic Computing*, edited by S. HALL & T. ASHIAKA, p. 407. Oxford: Clarendon Press.
PRESS, W. H., FLANNERY, B. P., TEUOLSKY, S. A. & VETTERLING, W. T. (1989). *Numerical Recipes: the Art of Scientific Computing (FORTRAN version)*. Cambridge Univ. Press.
STARK, H. (1987). Editor. *Image Recovery: Theory and Application.* Orlando: Academic Press.
SZÖKE, A. (1993). *Acta Cryst.* A**49**, 853–866.
WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.

Acta Cryst. (1995). A**51**, 708–716

Solution of the Phase Problem in Crystallography and Application to Dynamical Electron Diffraction

BY WILLIAM F. TIVOL

Wadsworth Center for Laboratories and Research and the School of Public Health, Empire State Plaza, PO Box 509, *Albany, NY* 12201-0509, *USA*

(*Received* 5 *September* 1994; *accepted* 6 *March* 1995)

Abstract

Unitarity, a fundamental principle of scattering theory, leads to the prediction of an essentially unique set of phases for the scattering amplitude from a complete knowledge of the differential cross section or, in the case of a crystal, from the diffracted intensities. The Sayre equation and all the direct methods of phasing following therefrom are derived as a special case of unitarity for zero excitation error. Dynamical and kinematical scattering are considered, and the relationship between them, $\hat{S} = \exp(i\tau n_z \hat{K})$, is obtained. Applications to the case of electron diffraction including for non-zero excitation error are discussed.

Introduction

When diffraction was first used to calculate molecular structures, it was realized that in addition to the intensities, which were directly measured, phases had to be determined for each of the reflections. Many schemes were devised to ascertain these phases, such as comparing isomorphous crystals, one of which had one or more heavy atoms that were lacking in the other, or looking at an unknown molecule which had as part of its structure a molecule whose structure was already known (Argos & Rossmann, 1980). These methods were very successful; however, not all materials of interest could be crystallized with and without heavy atoms or described by a known part plus an unknown part.

An alternative procedure is the derivation of the phases from the values of the measured intensities. All such techniques of using the known intensities to provide information about the unknown phases are collected under the category of direct methods of phase determination. Many of these methods are based on an equation first derived by Sayre (1952), who calculated the diffraction amplitudes of an arrangement of equal non-overlapping atoms and of the same arrangement of 'squared atoms'. By comparing the Fourier expansions of these two expressions, he was able to relate one (phased) amplitude to a convolution of all other (phased) amplitudes:

$$F(\mathbf{H}) = (\theta/V) \sum_{\mathbf{K}} F(\mathbf{K}) F(\mathbf{H} - \mathbf{K}), \qquad (1)$$

where $\mathbf{H}$ and $\mathbf{K}$ are sets of Miller indices, F is the complex amplitude, V is the unit-cell volume and θ is a constant of proportionality.

Direct methods of phasing diffraction patterns have been used very successfully in *ab initio* structure solutions in both X-ray (Ladd & Palmer, 1980; Day & Pendry, 1993; Glusker, 1993) and electron crystallography (Dorset, 1993; Dorset, Tivol & Turner, 1991, 1992, 1993). Phase extension, where initially the low-order phases are determined by some means and direct methods are used to relate the higher-order phases to the low-order ones, have also been quite successful (Dorset, 1993; Dorset, Kopp, Fryer & Tivol, 1995).

It has been stated many times that the phases can be extracted from the measured intensities because the electron density is everywhere positive and the unit cell of a crystal consists of equal non-overlapping point-like atoms. It is also stated that the fact that the atoms are not

ISSN 0108-7673