

- FREI, V. & DEUS, P. (1984). *Phys. Status Solidi B*, **125**, 121–130.
- FREI, V. & POLČÍK, M. (1984). *Czech. J. Phys.* **B34**, 883–886.
- FREI, V., MANDULA, M. & SLANINA, F. (1985). *Czech. J. Phys.* **B35**, 95–98.
- FUKSA, J. (1986). Thesis. The Institute of Physics, Czechoslovak Academy of Sciences, Prague, Czechoslovakia. (In Czech.)
- International Tables for Crystallography* (1987). Vol. A. Dordrecht: Kluwer.
- International Tables for X-ray Crystallography* (1959). Vol. II. Birmingham: Kynoch Press. (Present distributor Kluwer Academic Publishers, Dordrecht.)
- KOPSKÝ, V. (1988*a*). *Comput. Math. Appl.* **16**, 493–505.
- KOPSKÝ, V. (1988*b*). Private communication.
- VLACHAVAS, D. S. (1985). *Acta Cryst.* **A41**, 530–540.

Acta Cryst. (1990). **A46**, 915–922

An Envelope-Based Approach for Direct Phase Determination of Macromolecular Structures

BY DOUGLAS C. REES

Division of Chemistry and Chemical Engineering, 147–75CH, California Institute of Technology, Pasadena, California 91125, USA

(Received 14 December 1989; accepted 12 June 1990)

Abstract

An initial electron density distribution for a crystal structure may be directly derived from observed diffraction data by maximizing the product of the observed and calculated Patterson functions with respect to the electron density values within an envelope. This maximization problem may be formulated as an eigenvalue equation, in which potential electron density distributions are obtained as eigendensities (eigenvectors) of a symmetric matrix. Elements of this matrix depend only on the indices and intensities of the observed reflections, and on the coordinates of grid points inside the envelope. Eigendensities are calculated for a set of small envelopes (enclosing about 20% of the molecular volume) covering a unique region of the unit cell whose points are unrelated by space-group operations, origin shifts or changes in enantiomorph. On the basis of correlation coefficients between the observed and calculated values for both the Patterson function and structure-factor amplitudes, a small set of eigendensities are selected for combination into a final electron density distribution. This electron density distribution may be Fourier transformed to yield calculated structure factors. Test calculations on lysozyme indicate that phase errors of less than 60° may be obtained for strong low-resolution reflections by this procedure. An extension of this approach to handle crystal structures containing non-crystallographic symmetry is described.

Introduction

A molecular envelope divides the interior of macromolecular crystals into two regions of roughly equal volumes, consisting of either the molecule or

the solvent. Owing to positional disorder of the solvent molecules, the electron density in the solvent region is approximately constant. Enforcement of this constraint on the solvent density ('solvent flattening') has provided a powerful phase refinement method, in both the presence (Rossmann, 1972) and absence (Wang, 1985) of non-crystallographic symmetry. In principle, knowledge of the molecular envelope also provides an approach to direct phase determination, but this has not yet been routinely achieved in practice. In this note, a constrained symmetric quadratic function of the electron densities of grid points within a molecular envelope is formulated that depends only on the indices and intensities of observed reflections. Electron density distributions derived from the maximization of this function may be used to directly obtain phase information.

Although crystallographic calculations may be formally accomplished in either real or reciprocal space, a real-space emphasis will be adopted in this work since the distinction between molecular and solvent regions has a particularly simple real-space interpretation. Assuming a molecular envelope has been defined (the determination of the envelope will be discussed more fully below), the problem to be addressed is to find density values for grid points within the envelope that are consistent with the observed Patterson function. This may be accomplished by finding electron density values that maximize Φ :

$$\Phi = \sum_k P_o(u_k) P_c(u_k) \quad (1)$$

where Φ is the product of the multiply corrected observed and calculated Patterson functions, $P_o(u_k)$ and $P_c(u_k)$, respectively. The sum is over points at

positions u_k contained within the Patterson asymmetric unit. $P_c(u_k)$ may be calculated from the (as yet unknown) electron density values of points within the molecular envelope. As seen from the form of (1), Φ is related to the correlation coefficient between the observed and calculated Patterson functions, differing only by the presence of terms related to the variances of the two Patterson functions in the denominator of the correlation coefficient.

In matrix notation, the dependence of Φ on the electron density of points within the molecular envelope may be written in the following fashion:

$$\Phi = \rho^T A^T B A \rho, \quad (2)$$

where ρ is the vector of electron density values of grid points within the envelope, and the A and B matrices transform ρ values into the sum of the product of the observed and calculated Patterson functions. The elements of A and B have the form:

$$\begin{aligned} A_{2i-1,j} &= a^c(h_i, x_j) \\ A_{2i,j} &= b^c(h_i, x_j) \\ B_{2i-1,2i-1} &= B_{2i,2i} = \sum_k P_o(u_k) a^p(h_i, u_k) \\ B_{i,j} &= 0, \quad i \neq j, \end{aligned} \quad (3)$$

where a^c , b^c and a^p are the real (a) and imaginary (b) parts of the structure-factor expression for the crystal space group (superscript c) and the Patterson space group (superscript p), respectively; h_i are the Miller indices of the observed reflections; and x_j and u_k are positions of points within the molecular envelope and the Patterson asymmetric unit, respectively. Significantly, the A and B matrices depend only on the indices and intensities of the observed reflections [since $P_o(u_k)$ is the Fourier transform of the observed intensities] and are independent of the electron density values in ρ .

Maximization of Φ with respect to ρ requires introduction of a constraining relationship on the elements of ρ :

$$\rho^T \rho = 1. \quad (4)$$

This condition must be imposed since otherwise the maximization problem is not well defined as multiplication of ρ by a constant, S , increases Φ by S^2 . The constraining relation may be incorporated into the maximization of Φ by the use of Lagrange multipliers (Hestenes, 1975). The resulting function, Ψ , to be maximized is

$$\begin{aligned} \Psi &= \Phi - \lambda(\rho^T \rho - 1) \\ &= \rho^T A^T B A \rho - \lambda(\rho^T \rho - 1) \end{aligned} \quad (5)$$

where λ is the Lagrange multiplier.

At extrema points (including maxima), the derivative of Ψ with respect to ρ vanishes. Since $A^T B A$ is a symmetric matrix, ρ values defining these points

satisfy the eigenvalue equation

$$A^T B A \rho = \lambda \rho. \quad (6)$$

In general, if there are N grid points in the molecular envelope, there will be N different sets of coupled eigenvalue, eigenvector solutions to (6). Eigendensities ρ_i associated with the largest values of Φ are the eigenvectors corresponding to the largest eigenvalues of the symmetric matrix $A^T B A$. Since Φ is not exactly the correlation coefficient between observed and calculated Patterson functions, the ρ_i that maximize Φ need not correspond exactly to the true electron density values, even with the correct molecular envelope. One may hope (a word that occurs frequently in the molecular-replacement literature), however, that the true density may be expressed as a sum of only a relatively few eigendensities ρ_i . This expectation will be explored in greater detail below.

Methods

The general approach to an envelope-based phase determination consists of a sequence of three steps:

1. definition of the envelope;
2. calculation of eigendensities;
3. selection and combination of eigendensities.

Implementation of these steps will be illustrated by a specific application to the lysozyme crystal structure. Chicken egg-white lysozyme crystallizes in space group $P4_32_12$, with cell dimensions $a = b = 79.1$, $c = 37.9$ Å (Blake, Koenig, Mair, North, Phillips & Sarma, 1965). Diffraction data from lysozyme crystals were collected on a multiwire area detector (Cork, Hamlin, Vernon & Xuong, 1985), and intensities of the unique reflections were scaled and merged with the *ROCKS* crystallographic package (Reeke, 1984; Bethge, 1984). Phases for the observed structure-factor amplitudes were calculated from atomic coordinates (Kundrot & Richards, 1987) in the 2LYM set of the Brookhaven Protein Data Bank (Bernstein *et al.* 1977). Calculations in this paper were performed to a limiting resolution of 6 Å, with an observed data set containing 348 reflections (199 acentric, 149 centric, 94% of the theoretical number). Electron density and Patterson maps calculated from these structure factors were sampled at approximately 3 Å intervals, with 28, 28 and 12 grid points along the crystallographic a , b and c axes, respectively. Fast Fourier transform methods (Ten Eyck, 1973, 1977) were used for all map calculations.

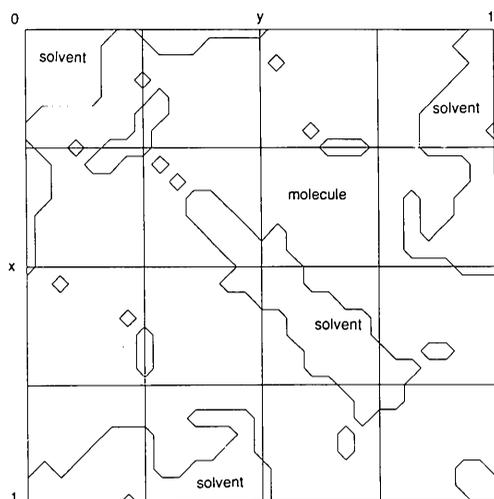
Step 1: Definition of envelope

A molecular envelope is described by specifying the shape, size and position. The actual molecular envelope for lysozyme in the section $z = 0/12$ is illustrated in Fig. 1(a). The outline of the envelope in the section $z = 0/12$ is also illustrated in Fig. 1(b) for a

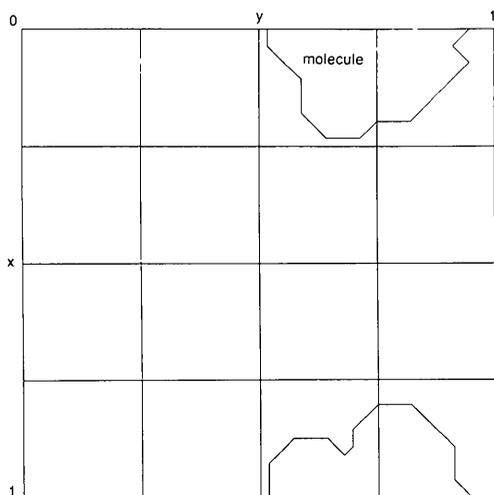
particular lysozyme molecule with molecular center near $(0, \frac{3}{4}, 0)$. As seen from these figures, the precise molecular envelope is rather complex, which complicates efforts to specify exactly the molecular envelope for an unsolved structure.* The following approach was used to approximate an envelope:

Shape: Simple shapes such as spheres and cubes are often used to model unknown envelopes. Cubic

* While this manuscript was in review, an experimental determination of a molecular envelope from intensity measurements by the method of solvent contrast variation was described (Carter, Crumley, Coleman, Hage & Bricogne, 1990).



(a)



(b)

Fig. 1. Molecular envelopes in the section $z=0$ for (a) the lysozyme crystal, and (b) one of the crystallographically equivalent lysozyme molecules in the crystal, centered near $(0, \frac{3}{4}, 0)$. The envelopes were generated from the lysozyme coordinates by assigning a van der Waals radius of 3 Å to each atom.

envelopes were used in this work since different (trial) envelopes could be stacked together like bricks so as to fill uniformly a given region of space.

Size: The molecular envelope of lysozyme encloses a volume of approximately $1.7 \times 10^4 \text{ \AA}^3$, or about 650 ($\sim 9^3$) grid points at a sampling interval of 3 Å. While this sets an upper limit on the size of an envelope, for computational reasons it is desirable to determine the smallest envelope volume that will allow acceptable phasing. Average phase differences were determined between the exact phases (α_{model} , computed from the atomic model) and phases obtained from inversion of $(F_{\text{obs}}, \alpha_{\text{model}})$ electron density maps masked with cubes containing 3, 5, 7, 9 and 11 grid points on a side. The cubes were centered at $(0, \frac{3}{4}, 0)$, which is near the molecular center of one of the crystallographically equivalent lysozyme molecules. Overall absolute phase differences, as well as absolute phase differences for classes of reflections defined on the basis of resolution and structure-factor amplitude, are presented in Table 1. Rather unexpectedly, acceptable phase errors ($< 50^\circ$) were obtained for strong reflections (defined as $F_{\text{obs}} > F_{\text{ave}}$) using cubic envelopes containing only 5^3 grid points, or about 20% of the molecular volume.

Position: For envelopes of sufficiently high symmetry, the search region for the envelope position may be restricted to that fraction of the asymmetric unit containing a unique set of points unrelated by changes in origin or enantiomorph. In space group $P4_32_12$, structures related by origin shifts of $(0, 0, 0)$, $(\frac{1}{2}, \frac{1}{2}, 0)$, $(0, 0, \frac{1}{2})$, $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ have identical diffraction patterns (Rogers, 1980). The origin shifts imply that points with the following coordinates unrelated by crystallographic symmetry in space group $P4_32_12$ give rise to the same set of Patterson vectors:

$$\begin{aligned} (a) & \quad x \quad y \quad z \\ (b) & \quad \frac{1}{2} + x \quad \frac{1}{2} + y \quad z \\ (c) & \quad -x \quad -y \quad z \\ (d) & \quad \frac{1}{2} - x \quad \frac{1}{2} - y \quad z. \end{aligned} \quad (7)$$

Additional equivalent positions appear in the sections $z=0, \frac{1}{8}, \dots$ due to the presence of crystallographic rotation and screw axes in these planes. As a consequence, only $\frac{1}{4}$ of the asymmetric unit needs to be searched for the envelope center. There is no enantiomorph ambiguity in this space group, since a change in hand requires a change in space group to $P4_12_12$.

One choice of asymmetric unit in space group $P4_32_12$ is included in the volume $0 \leq x < 1$, $0 \leq y \leq x$, $0 \leq z \leq \frac{1}{8}$. With only 12 sections along the entire c axis (for calculations at 6 Å resolution), and a cubic envelope with 5 points along an edge, the search for the envelope center in the lysozyme calculation was restricted to the one section $z = 0/12$. Since this region

Table 1. Comparison between true structure factors for lysozyme and structure factors calculated to 6 Å resolution by inversion of the true electron density map, masked with cubes of the indicated size

The unit cell was sampled at 28, 28 and 12 divisions along the crystallographic a , b and c axes, respectively. All cubes were centered at (0, 21/28, 0), near the center of mass of one of the crystallographically equivalent copies of a lysozyme molecule. The correlation coefficients are between the true and calculated values for both the Patterson function (omitting the origin and the nearest-neighbor points) and the structure-factor amplitudes. The absolute phase differences (in degrees) between the true and calculated phases are presented for all reflections, as well as four mutually exclusive categories of reflections (WL , SL , WH , SH). W and S indicate weak and strong reflections, as defined by $F_{\text{obs}} < F_{\text{ave}}$ or $F_{\text{obs}} > F_{\text{ave}}$, respectively. F_{ave} values are calculated in shells of 0.001 \AA^{-2} in $\sin^2 \theta/\lambda^2$. L and H indicate 'low'- and 'high'-resolution values, defined as $0.002 (\sim 11 \text{ \AA}) < \sin^2 \theta/\lambda^2 < 0.0045 \text{ \AA}^{-2}$ ($\sim 7.5 \text{ \AA}$) and $0.0045 < \sin^2 \theta/\lambda^2 < 0.0069 \text{ \AA}^{-2}$ (6 \AA), respectively.

Cube edge (grid points)	Correlation coefficients		Overall	Absolute phase differences (°)			
	Patterson	Structure factor		WL	SL	WH	SH
3	0.174	0.129	82	92	66	93	70
5	0.294	0.288	73	95	48	98	43
7	0.357	0.476	56	86	36	75	27
9	0.673	0.741	35	54	9	54	14
11	0.894	0.917	18	37	7	23	7

is significantly larger in the x and y directions than the trial envelopes, it is necessary to try multiple envelope positions to sample the entire search region.

Step 2: Calculation of eigendensities

The crystallographic coordinates of the 125 grid points in a $5 \times 5 \times 5$ cubic envelope may be determined once the center position of the envelope is specified. These coordinates, coupled with the indices of the observed reflections and the observed Patterson function, may be used to calculate elements of the A and B matrices (3). The relevant crystal and Patterson space groups for lysozyme are $P4_32_12$ and $P4/mmm$, respectively. The summation for elements of the B matrix is over points in the Patterson asymmetric unit of lysozyme ($0 \leq u \leq \frac{1}{2}$, $0 \leq v \leq u$, $0 \leq w \leq \frac{1}{2}$). The origin of the Patterson function and the nearest-neighbor points were excluded from the summation. The Patterson function was not modified for this calculation, although sharpening could be performed to enhance the contribution of higher-resolution reflections. All trigonometric expressions were evaluated with a lookup algorithm described by Burnett & Nordman (1974) and Hoard & Nordman (1979).

Eigenvalues and vectors of the $A^T B A$ matrix were evaluated with the *EIGRS* routine of the IMSL library. The resulting eigendensities were characterized by calculating correlation coefficients between the observed and calculated (1) electron densities, (2) structure-factor amplitudes and (3) Patterson functions, as well as the absolute difference between phases calculated from the atomic model and from the eigendensity values. Of these indicators, only the structure-factor amplitude and Patterson-function correlation coefficients could be calculated for an unknown structure. As examples to be explored more fully below, a listing of these indicators for the first ten eigendensities calculated for cubic envelopes centered at (2/28, 18/28, 0) and (0, 22/28, 0) are presented in Table 2. Since the eigendensities are specified only to within a constant (which can be

either positive or negative), both positive and negative correlation coefficients between the observed electron density and the eigendensities may be observed.

Step 3: Selection and combination of eigendensities

Eigendensities were calculated for $5 \times 5 \times 5$ cubic envelopes centered in the $z = 0$ section and spaced at two grid intervals in the x and y crystallographic directions. The average sum of the Patterson function and structure-factor amplitude correlation coefficients for the first ten eigendensities of each envelope are plotted in Fig. 2(a) as a function of the envelope center position. Encouragingly, regions exhibiting above average correlation coefficients agree well with allowed regions of a packing function (Fig. 2b) illustrating permissible positions for the center of an 11 Å radius sphere in this space group (Hendrickson & Ward, 1976).

Since the lysozyme structure is known, the agreement between the true electron density and the eigendensities can be directly assessed. While no eigendensity exactly reproduces the true density within the envelope, typically, several eigendensities exhibit significant correlation coefficients (magnitude > 0.3) with the true density (Table 2). This suggests that a better approximation to the true electron density could be obtained by combining multiple eigendensities from different envelopes into a single density distribution.

Envelopes centered at (2/28, 18/28, 0) and (0, 22/28, 0) were selected for further analysis, since: (1) the maximum value for the average correlation coefficients plotted in Fig. 2(a) occurs at the point (2/28, 18/28, 0), and (2) the envelope centered at (0, 22/28, 0) is in the middle of the same contiguous allowed region of Fig. 2(a) as the first envelope, but does not overlap with it. A projection of the region occupied by these envelopes in the asymmetric unit of lysozyme is illustrated in Fig. 3. These particular envelope positions were selected in preference to other equivalent points in Fig. 2(a) for consistency

Table 2. Characterization of eigendensities calculated for $5 \times 5 \times 5$ envelopes centered at $(2/28, 18/28, 0)$ and $(0/28, 22/28, 0)$

Eigendensities are ordered in decreasing value of Φ calculated from equation (1). The correlation coefficients (designated *ED*, *Patt* and *SF*) are between the true and calculated values of the electron density inside the cube, the Patterson function and the structure-factor amplitudes, respectively. Phase differences are as described in the legend to Table 1. Perfectly anticorrelated (sign-reversed) electron density sets will have an absolute phase difference of 180° . Eigendensities flagged with an asterisk (*) were selected for combination into the final calculated structure factors on the basis of the Patterson function and structure-factor amplitude correlation coefficients.

Eigenvector	Correlation coefficients			Overall	Absolute phase differences ($^\circ$)			
	<i>ED</i>	<i>Patt</i>	<i>SF</i>		<i>WL</i>	<i>SL</i>	<i>WH</i>	<i>SH</i>
$5 \times 5 \times 5$ cube centered at $(2/28, 18/28, 0)$								
1	-0.159	0.189	0.219	91	74	106	87	100
*2	0.493	0.490	0.405	90	92	77	97	86
3	-0.027	0.207	0.349	89	93	105	88	86
4	0.003	0.320	0.319	89	86	87	91	91
5	0.077	0.183	0.169	93	104	85	91	93
6	-0.135	0.324	0.292	97	85	85	104	102
7	0.045	0.462	0.351	88	96	94	85	85
*8	-0.331	0.487	0.406	97	86	103	98	100
*9	-0.126	0.472	0.376	93	92	111	87	93
10	0.017	0.160	0.158	95	92	97	93	99
$5 \times 5 \times 5$ cube centered at $(0/28, 22/28, 0)$								
1	-0.275	0.168	0.211	91	101	112	91	75
*2	-0.112	0.557	0.391	90	91	106	87	86
3	-0.259	0.115	0.191	93	99	89	90	93
4	-0.109	0.218	0.235	92	84	97	91	95
5	-0.098	0.181	0.279	90	87	91	87	94
*6	0.247	0.502	0.369	89	95	85	95	80
*7	0.427	0.426	0.320	88	101	64	94	84
8	0.130	0.363	0.266	90	87	85	99	85
9	0.016	0.280	0.290	94	90	96	94	94
10	-0.132	0.437	0.269	94	84	81	98	101

with the known lysozyme structure (Fig. 1*b*), which facilitates comparison of observed and calculated phases. From the sum of the Patterson function and structure-factor amplitude correlation coefficients, the top three eigendensities (Table 1) from each of

these two envelopes were selected for combination into a single electron density distribution.

Combination of individual eigendensities was achieved by the following method. Let F_1 and F_2 be structure factors calculated from multiple

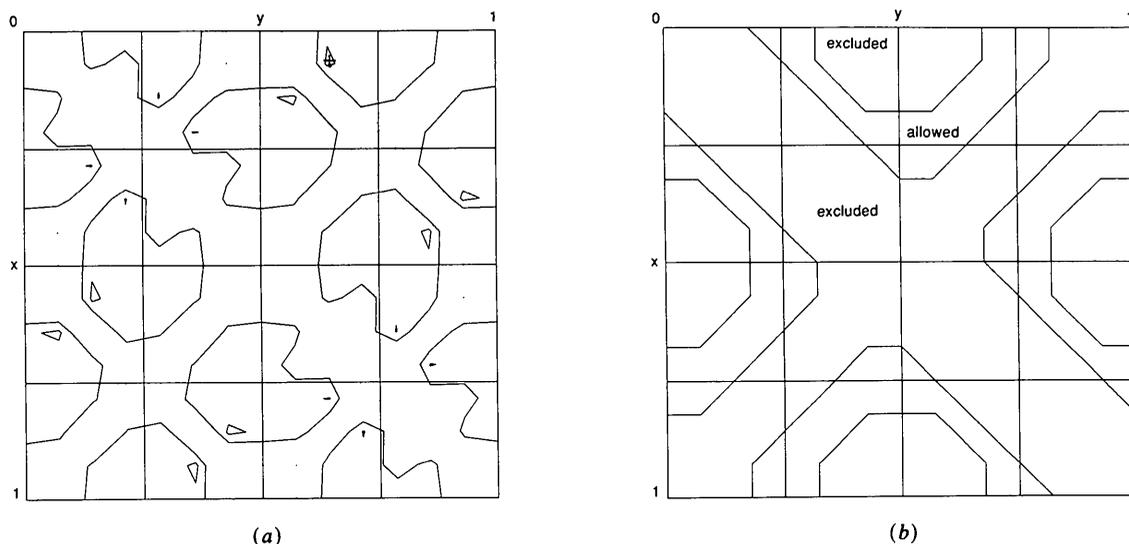


Fig. 2. (a) Contour plot of the average value of the sum of the correlation coefficients between observed and calculated values for both the Patterson function and the structure-factor amplitudes for the top ten eigendensities generated from envelopes centered in the section $z = 0$. Envelopes contained 5^3 grid points, and were spaced at intervals of 2 grid units in the x and y coordinates. Contour levels are at ± 0.5 and $\pm 1.0\sigma$ from the mean of the observed values for the correlation coefficients. Negative contours are represented by dotted lines. The maximum in this section occurs at $(2/28, 18/28, 0)$ and equivalent points, and is indicated by +. (b) Packing function diagram illustrating the allowed and excluded positions for the center of an 11 \AA sphere in the section $z = 0$ of the lysozyme space group and unit cell.

eigendensities associated with envelopes 1 and 2, respectively. F_1 and F_2 are obtained by summing the structure factors calculated from the individual eigendensities through the relationships:

$$\begin{aligned} F_1 &= \sum_i \beta_{1,i} f_{1,i} \\ F_2 &= \sum_j \beta_{2,j} f_{2,j} \end{aligned} \quad (8)$$

where $f_{1,i}$ and $f_{2,j}$ are structure factors calculated from the i th and j th eigendensity of envelopes 1 and 2, respectively, and the β 's are the corresponding weighting factors. Values for the β 's may be determined by maximizing the quantity C , which is related to the correlation coefficient between F_1 and F_2 (Srinivasan & Parthasarathy, 1976). C is defined as

$$C = \langle \text{Re} (F_1 F_2^*) \rangle \quad (9)$$

where Re indicates the real component of this complex function, and the average is taken over all reflections. Expansion of C shows that this expression is a symmetric quadratic in $\beta_{1,i}, \beta_{2,j}$, so that the maximization may be accomplished by the same method described previously to evaluate the eigendensities. Specifically, the weighting coefficients may be determined by solving for the eigenvectors corresponding to the largest eigenvalues of the matrix:

$$\begin{bmatrix} 0 & \sum_h f_{1,i} f_{2,j} \\ \sum_h f_{1,i} f_{2,j} & 0 \end{bmatrix}. \quad (10)$$

Maximization of this correlation coefficient is equivalent to finding the linear combinations of the

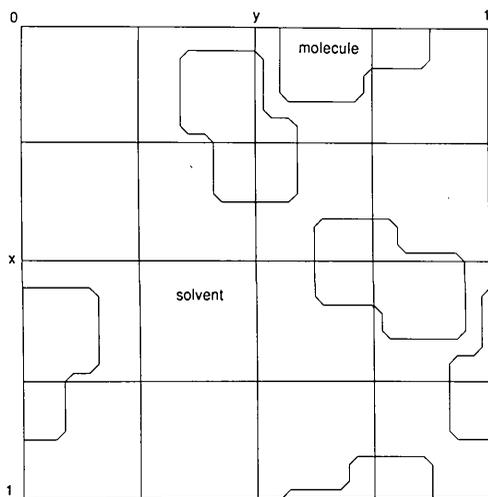


Fig. 3. Superposition down the z axis ($0 \leq z \leq \frac{1}{2}$) of the regions contained within the $5 \times 5 \times 5$ grid point envelopes centered at $(2/28, 18/28, 0)$ and $(0, 22/28, 0)$ and the crystallographically equivalent positions.

f 's that yield the most similar values for the complex structure factors F_1 and F_2 .

Since test calculations suggested that phases were most reliably determined for strong reflections ($F_{\text{obs}} > F_{\text{ave}}$) at low resolution (Tables 1 and 2), the β 's were evaluated using only strong reflections in the resolution range $0.002 < \sin^2 \theta / \lambda^2 < 0.0045 \text{ \AA}^{-2}$. Two sets of coefficients were obtained (Table 3), corresponding to the top two eigenvalues of (10), which had correlation coefficients > 0.15 between F_1 and F_2 . These weights were used to determine a combined set of calculated structure factors, F_c , through the relationship:

$$F_c = \sum_{j=1}^2 \sum_{i=1}^3 \beta_{j,i} f_{j,i}. \quad (11)$$

Three sets of calculated structure factors ($SF1$, $SF2$ and $SF12$) were generated from the two individual sets of weighting coefficients, and a combination of the two sets, respectively. For strong low-resolution reflections, the absolute phase differences between the model and calculated reflections were 75, 61 and 54°, respectively. These phase differences are significantly smaller than random (90°), and suggest that useful phase information may be generated by this approach.

Electron density maps of lysozyme calculated with envelope-derived phases are illustrated in Fig. 4. This figure shows the superposition of sections covering the asymmetric unit of lysozyme for maps calculated with Fourier coefficients ($F_{\text{obs}}, \alpha_{\text{model}}$; Fig. 4a), and ($F_{\text{obs}}, \alpha_{\text{calc}}$; Fig. 4b) where α_{calc} is from the combined $SF12$ Fourier set. Reflections with ($F_{\text{obs}} > F_{\text{ave}}$) in the resolution range $0.002 < \sin^2 \theta / \lambda^2 < 0.0045 \text{ \AA}^{-2}$ were used in the Fourier calculation. The correlation coefficient between the ($F_{\text{obs}}, \alpha_{\text{model}}$) and the ($F_{\text{obs}}, \alpha_{\text{calc}}$) maps is 0.60, demonstrating a significant correlation between the calculated and true phases.

Discussion

Electron density distributions may be directly derived from observed diffraction data by maximizing the product of the observed and calculated Patterson functions with respect to the electron density values within an envelope. The maximization problem may be formulated as an eigenvalue equation (6), in which the potential electron density distributions are obtained as eigendensities (eigenvectors) of a symmetric matrix. The approach to direct phase determination adopted in this work is to calculate eigendensities for multiple small envelopes that cover a minimal search region of the unit cell. Promising eigendensities are selected from several non-overlapping envelopes, on the basis of agreement with the observed Patterson function and structure-factor amplitudes, and combined into the final electron

Table 3. Values of the weighting coefficients β calculated from the eigenvectors of equation (10) for combining eigendensities into calculated structure-factor sets

Eigendensities a, b, c, d, e and f correspond to eigendensities 2, 8 and 9 of the envelope centered at $(2/28, 18/28, 0)$ and eigendensities 2, 6 and 7 of the envelope centered at $(0/28, 22/28, 0)$ respectively (Table 1). Coefficients for the structure-factor set $SF12$ were obtained by summing the coefficients for sets $SF1$ and $SF2$ and normalizing.

Structure-factor set	Weighting coefficients (β) for eigendensities					
	a	b	c	d	e	f
$SF1$	-0.04	-0.46	-0.53	-0.62	0.19	0.27
$SF2$	0.66	0.16	-0.19	0.33	0.42	0.46
$SF12$	0.44	-0.21	-0.51	-0.21	0.43	0.52

density distribution. Phases are then calculated by Fourier transformation of this distribution. Test calculations with lysozyme indicate that phase errors of less than 60° may be obtained for strong low-resolution reflections. These phases could be used to image the structure directly (Fig. 4), to locate heavy atoms by difference-Fourier analyses, or perhaps to serve as starting values for various phase refinement methods.

Clearly, the critical step in this process is the selection of eigendensities to be combined into the final electron density distribution. Since no eigendensities were found in the lysozyme calculation that exactly reproduced the true electron density within an envelope, it is not necessarily evident which eigendensities should be selected for further combination. The magnitudes of the correlation coefficient between observed and calculated values for both the Patterson function and structure-factor amplitudes provide an objective criterion for the initial selection of eigendensities. These choices may be further refined by examination of the magnitudes of the weighting coefficients, β , which provide an estimate of the con-

tribution of the associated eigendensity to the calculated structure factors (8). An iterative process for eigendensity selection can be envisioned, in which eigendensities with small β 's are replaced by new eigendensities, until a final set is selected.

Although not relevant to the lysozyme calculation, combination of eigendensities from structures in polar space groups may be more complicated than treated here. In this case, it is quite possible that individual eigendensities may correspond to different choices of origins along the polar axis. The initial combination of eigendensities might be treated by using only seminvariant reflections (which are insensitive to choice of origin) to evaluate β 's from (10), followed by a systematic shifting along the polar axis to find the positions of highest correlation coefficient.

Crystal structures with non-crystallographic symmetry may be handled by a straightforward extension of this procedure. Provided the non-crystallographic symmetry relationships have been determined, the A matrix can be easily modified to incorporate this information. Specifically, if the subscript l designates

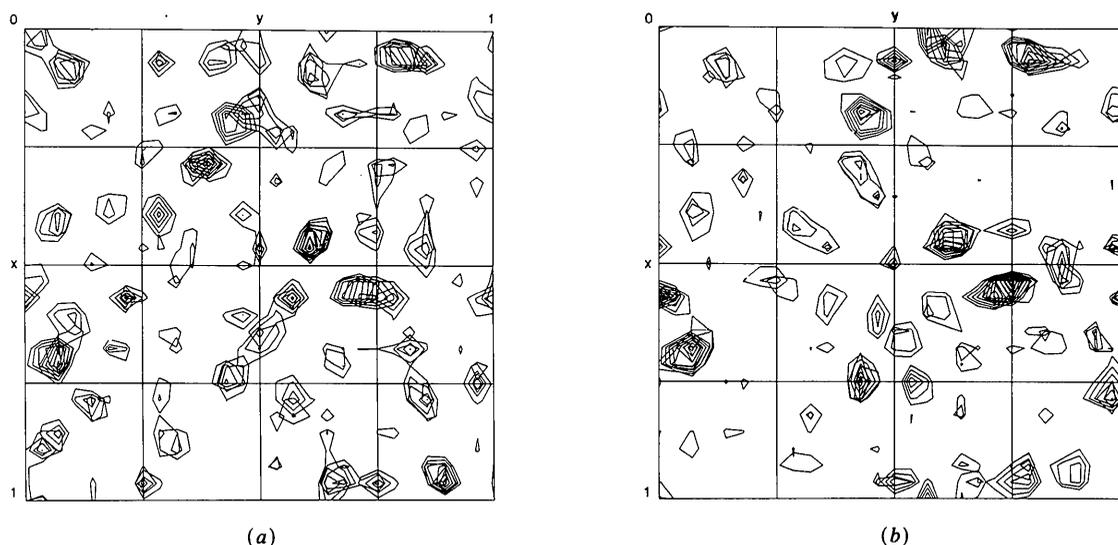


Fig. 4. Superposition down the z axis ($0 \leq z \leq \frac{1}{2}$) of the electron density in the lysozyme asymmetric unit, calculated with Fourier coefficients (a) ($F_{\text{obs}}, \alpha_{\text{model}}$) and (b) ($F_{\text{obs}}, \alpha_{\text{calc}}$). α_{calc} was calculated from the $SF12$ structure-factor set.

coordinates related by the non-crystallographic symmetry, then A may be written:

$$\begin{aligned} A_{2i-1,j} &= \sum_l a^c(h_i, x_{j,l}) \\ A_{2i,j} &= \sum_l b^c(h_i, x_{j,l}). \end{aligned} \quad (12)$$

In this manner, the size of the envelope may be effectively increased without increasing the size of the eigendensity vector, which should enhance the ability of a few eigendensity vectors to represent the structure.

Crowther (1967, 1969) has presented an elegant linear formulation of the non-crystallographic symmetry problem that provided a basis set of eigendensities to describe an unknown structure. If the Fourier transform of the j th eigendensity is represented by $u_j(h_i)$, then the observed intensity $|F_{\text{obs}}^2(h_i)|$ may be written as the symmetric quadratic:

$$|F_{\text{obs}}^2(h_i)| = \sum_j \sum_k \mu_j \mu_k u_j(h_i) u_k(h_i), \quad (13)$$

where the μ_j are the weighting coefficients of the corresponding u_j . A difficulty in the implementation of Crowther's formulation was the quadratic dependence of the observed intensities on the (unknown) μ_j . If the problem is recast to maximize the sum of the product between the observed and calculated intensities with respect to the μ_j , however, then the μ_j may be obtained directly by the eigenvalue method discussed in this paper.

Direct methods of phase determination involve the application of constraints on the electron density (such as non-negativity and/or atomicity in the case of small-molecule structures) to provide useful relationships between structure factors. A characteristic feature of macromolecular structures is the relatively large solvent content, and the existence of an envelope that divides the crystal into regions of molecule and solvent. In this paper, an envelope-based approach has been developed for deriving phase information directly from observed diffraction data. While the ultimate utility of this method can

only be established by the successful solution of unknown structures, test calculations suggest that phases of promising quality may be obtained by this procedure.

We thank B. T. Hsu, R. E. Marsh, V. Schomaker, J. N. Franklin, T. O. Yeates and D. E. Malerba for discussions and programs. This work was supported by grants from the NIH, NSF and the Joseph Irvine Equipment Fund. DCR is an A. P. Sloan Research Fellow.

References

- BERNSTEIN, F., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F., BRICE, M. D., RODGERS, J. R., KENNARD, O., SHIMANOCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535-542.
- BETHGE, P. H. (1984). *J. Appl. Cryst.* **17**, 215.
- BLAKE, C. C. F., KOENIG, D. F., MAIR, G. A., NORTH, A. C. T., PHILLIPS, D. C. & SARMA, V. R. (1965). *Nature (London)*, **206**, 757-761.
- BURNETT, R. M. & NORDMAN, C. E. (1974). *J. Appl. Cryst.* **7**, 625-627.
- CARTER, C. W., CRUMLEY, K. V., COLEMAN, D. E., HAGE, F. & BRICOGNE, G. (1990). *Acta Cryst.* **A46**, 57-68.
- CORK, C., HAMLIN, R., VERNON, W. & XUONG, N. H. (1985). *Methods Enzymol.* **114**, 452-472.
- CROWTHER, R. A. (1967). *Acta Cryst.* **22**, 758-764.
- CROWTHER, R. A. (1969). *Acta Cryst.* **B25**, 2571-2580.
- HENDRICKSON, W. A. & WARD, K. B. (1976). *Acta Cryst.* **A32**, 778-780.
- HESTENES, M. R. (1975). *Optimization Methods*, pp. 138-176. Wiley: New York.
- HOARD, L. G. & NORDMAN, C. E. (1979). *Acta Cryst.* **A35**, 1010-1015.
- KUNDROT, C. E. & RICHARDS, F. M. (1987). *J. Mol. Biol.* **193**, 157-170.
- REEKE, G. N. (1984). *J. Appl. Cryst.* **17**, 125-130.
- ROGERS, D. (1980). In *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. LADD & R. A. PALMER, pp. 23-92. New York: Plenum.
- ROSSMANN, M. G. (1972). Editor. *The Molecular Replacement Method*. New York: Gordon and Breach.
- SRINIVASAN, R. & PARTHASARATHY, S. (1976). *Some Statistical Applications in X-ray Crystallography*, pp. 221-222. Oxford: Pergamon.
- TEN EYCK, L. F. (1973). *Acta Cryst.* **A29**, 183-191.
- TEN EYCK, L. F. (1977). *Acta Cryst.* **A33**, 486-492.
- WANG, B. C. (1985). *Methods Enzymol.* **115**, 90-112.