let P represent a point on the exit surface of the crystal. The integrated intensity, $I_q(P)$, emerging from P is

$$I_q(P) = \int \mathrm{d}\varepsilon_1 \int \mathrm{d}\varepsilon_2 |D_q(\varepsilon_1, \varepsilon_2)|^2. \quad (32)$$

The integrated power, P_q, is then obtained by integrating (32) over the exit surface

$$P_q = \int_P \mathrm{d}\mathbf{P}.\hat{s}_q I_q(P). \quad (33)$$

The surface element d**P** points in the direction of the outward drawn normal.

It is difficult to perform analytically the integrations in (31), (32) and (33) and only numerical solutions are feasible for arbitrary crystal shapes.

Concluding remarks

The Takagi–Taupin equations have been examined for three coupled waves in the case of Laue diffraction. When the coupling constant $1/\tau_{\pm g}$ or $1/\tau_{\pm(h-g)}$ becomes negligible it is possible to use integral equation techniques to obtain the boundary-value Green functions for the wave fields. The solutions, excluding any singular terms, may be taken as approximate expressions for the fields excited by a transversally limited wave packet when keeping away from the surface region and the $s_0 s_h$ and the $s_0 s_g$ planes. To the author's knowledge no experimental study of the fringes generated by three indirectly coupled waves has been published. In principle, it should be possible to confirm the spatial behaviour of the intensities $I_q \sim D_q D_q^*$ by intercepting a beam with the aid of a small pin hole or, alternatively, by locating a point source close to the surface of the crystal. In the case of an incident wave with an extended wave front, the intensity can be calculated by integrating the Green functions over the entrance surface allowing for the proper variation in amplitude of the incident wave.

References

AUTHIER, A., MALGRANGE, C. & TOURNARIE, M. (1968). *Acta Cryst.* A**24**, 126–136.

AUTHIER, A. & SIMON, D. (1968). *Acta Cryst.* A**24**, 517–526.

BREMER, J. (1984). *Acta Cryst.* A**40**, 283–291.

CHANG, S.-L. (1984). *Multiple Diffraction of X-rays in Crystals.* Berlin: Springer.

HART, M. & LANG, A. R. (1961). *Phys. Rev. Lett.* **7**, 120–121.

HATTORI, H. & KATO, N. (1966). *J. Phys. Soc. Jpn*, **21**, 1772–1777.

HØIER, R. & AANESTAD, A. (1981). *Acta Cryst.* A**37**, 787–794.

HØIER, R. & MARTHINSEN, K. (1983). *Acta Cryst.* A**39**, 854–860.

KATO, N. (1961). *Acta Cryst.* **14**, 526–532, 627–636.

KATO, N. (1974). In *X-ray diffraction*, edited by L. V. AZAROFF, pp. 176–438. New York: McGraw-Hill.

KATO, N. (1976). *Acta Cryst.* A**32**, 453–457.

PINSKER, Z. G. (1978). *Dynamical Scattering of X-rays in Crystals.* Berlin: Springer.

POST, B. (1983) *Acta Cryst.* A**39**, 711–718.

TAKAGI, S. (1962). *Acta Cryst.* **15**, 1311–1313.

TAKAGI, S. (1969). *J. Phys. Soc. Jpn*, **26**, 1239–1253.

TAUPIN, D. (1964). *Bull. Soc. Fr. Minéral. Cristallogr.* **87**, 469–511.

WERNER, S. A. (1974). *J. Appl. Phys.* **45**, 3246–3254.

Acta Cryst. (1986). A**42**, 197–202

Statistical Geometry. IV. Maximum-Entropy-Based Extension of Multiple Isomorphously Phased X-ray Data to 4 Å Resolution for α-Lactalbumin

BY STEPHEN W. WILKINS*

Laboratory of Molecular Biophysics, Department of Zoology, University of Oxford, Oxford OX1 3QU, England and CSIRO, Division of Chemical Physics, PO Box 160, *Clayton, Victoria* 3168, *Australia*

AND DAVID STUART

Laboratory of Molecular Biophysics, Department of Zoology, University of Oxford, Oxford OX1 3QU, England

(*Received* 26 *July* 1985; *accepted* 20 *December* 1985)

Abstract

The simplest level of the statistical geometric (SG) or maximum-entropy (ME) approach to X-ray structure refinement is applied to the task of trying to extend the resolution of electron-density maps for a small protein (α-lactalbumin). The refinement was started from X-ray structure factor data to 4 Å resolution, which had been phased by multiple isomorphous replacement (MIR), and it was found that, even at this simple level, the ME-based approach yields a significant improvement in the maps and gives encouragement to the more general applications of these methods.

* Permanent address: CSIRO, Australia.

0108-7673/86/030197-06$01.50

1. Introduction

As a step toward the implementation of maximum-entropy (ME)-based methods of X-ray structure refinement and determination (Gull & Daniell, 1978; Wilkins, Varghese & Lehmann, 1983; Bricogne, 1984), we have investigated how far structural information can be extended by a single pass with ME-based refinement, starting from multiple isomorphously (MIR) phased data for a small protein, α-lactalbumin. Our present work is principally aimed at understanding the power of one element of the statistical geometric (SG) or ME method outlined by Wilkins, Varghese & Lehmann (1983, hereafter termed SG1), rather than at trying to obtain a definitive 'best map' by, say, a combination of elements and methods, such as by using an iterative process involving successive passes of ME followed by phase recombination or by combining several elements of the ME method.

ME-based methods of structure refinement have been described by various authors. The essence of these is as follows. For a discrete and normalized density $\mathbf{p}$, the configurational entropy of the map relative to an initial map $\mathbf{q}$ is given by (see *e.g.* Steenstrup & Wilkins, 1984; Bricogne, 1984; Wilkins, Steenstrup & Varghese, 1985)

$$S = -n \sum_{j=1}^{N} p_j \ln [p_j/q_j], \qquad (1)$$

where p_j and q_j are the densities of the current and prior map at the jth pixel, while n is a parameter which relates to the quantity and quality of the data (Wilkins, Steenstrup & Varghese, 1985) but in the present context plays no role. The available information (and data) is assumed to be given in the form of constraint equations, which for the case of a set, D_1, of imperfectly phased unitary (*i.e.* F's scaled by F_{000}) structure factors, U_k, we have taken to have the form

$$f_1(\mathbf{p}) = (2N_1)^{-1} \sum_{k \in D_1} |P_k - U_k|^2/\sigma_{k,1}^2 = C_1, \qquad (2)$$

corresponding to a reduced χ^2 distribution with P_k the Fourier transform of $\mathbf{p}$, and σ_k the estimated magnitude of the standard deviation in U_k. The summation in (2) is over all Friedel pairs ($2N_1$ structure factors) and the expected value of f_1 over a large number of sets of data should ideally be $C_1 = 1$. A value of C_1 greater than 1 implies that we are not extracting all the information contained in the data, while a value of $C_1 < 1$ implies that we are fitting to noise as well as signal. Equation (2) represents the so-called first constraint; if the reduced χ^2 is based on amplitudes only, then we obtain the second constraint (SG1). The mathematical and numerical task of the ME-based methods of structure refinement is to find the map, $\mathbf{p}^S$, which maximizes S subject to all the available information (Gull & Daniell, 1978; SG1). At its lowest and simplest level, this information is taken to be (2).

2. Application of the ME method to data for α-lactalbumin

The present work differs from earlier ME-based calculations on three-dimensional data from proteins as follows. In the work of Collins (1982), basic equations apparently similar to (1) and (2) above were used to treat 2 Å resolution MIR data for rubredoxin, but the guiding principles of the ME method were obscured somewhat by:

(i) duplication of information by its inclusion *via* both the initial map $\mathbf{q}$ and the constraint equation;

(ii) constant updating of the initial map $\mathbf{q}$ in the iterative process of attempting to solve the ME equations numerically;

(iii) lack of any proper check on the convergence of the numerical method for treating the ME equations.

More specifically, information that is included in $\mathbf{q}$ should not in principle (at least not without some justification) also be included *via* constraints, since there is a presumption in the ME method that these sources of information are independent. This also means that $\mathbf{q}$ should not in principle be modified unless it is to preserve old information as new information is introduced. The lack of any proper check on the convergence of the solution to the ME equations raises questions as to the nature of the maps presented by Collins and any claim to their being ME maps. Such doubts are heightened when it is appreciated that, by continually transferring phases from the iterated solution to the measured amplitudes, Collins was effectively treating the case of the second constraint (reduced χ^2 based on amplitudes) alone, which is fraught with the problems of non-convexity and slow convergence (Gull & Daniell, 1978; SG1; Bricogne, 1984).

More recently, a conceptually clean application of the ME method with first constraint alone has been described by Bricogne (1984), but this was for the somewhat artificial case of measured amplitudes with *refined-model phases* for 3 Å data for crambin. While giving strong support to the relevance of ME methods to macromolecular structure refinement and determination, his calculations were based on very small $\sigma_{k,1}$'s, since the phases were derived from an exceptionally accurate model (Hendrickson & Teeter, 1981), and so leaves open the question as to the power of this component of the method when MIR phases, typically with much larger random and systematic errors, are used. For this reason we here present a brief report on our findings obtained when the ME method with first constraint alone was applied to 4 Å resolution data with MIR phases for α-lactalbumin

(unit cell 33·6×69·6×47·0 Å and space group $P2_12_12$), a small protein of 123 residues with strong structural (but not functional) homology with the c-type lysozymes (Smith, 1982). This structure is currently being refined in Oxford (Stuart, Phillips, Lewis & Smith, 1985) and we were therefore in the fortunate position of having a fairly accurate independent set of model phases against which we could compare phases produced by ME extrapolation. It should be emphasized that these model phases were used solely for the purposes of comparison and formed no part of the ME calculations.

Before applying ME to the structure factor data, a form of correction for the effect of solvent was applied to these data which does not depend on a knowledge of the position of the solvent–protein boundary and is based on Babinet's principle (Stuart, 1985). The net effect of applying this correction was to rescale the data by a **k**-dependent factor which boosts the values of the low-angle reflections (say above about 5 Å resolution) and leaves the high-angle reflections virtually unchanged, thereby effectively removing the solvent contribution from the low-angle data, at least to a crude first approximation. The data were then placed on an absolute scale appropriate to protein alone. These steps were felt to be helpful to the ME-based refinement because the method partly relies on positivity of the map.

Assignment of estimates for the errors, $\sigma_{k,1}$, was made *via* the relation

$$\sigma_F^2 = \sigma_{|F|}^2 + |F|^2\sigma_\varphi^2, \qquad (3)$$

which ignores correlations in errors between phase and amplitude of the given reflection. Values of $\sigma_{|F|}$ were taken based on counting errors and consistency between equivalent reflections, while σ_φ was estimated from the MIR-determined figures of merit, m, *via* the relation

$$\sigma_\varphi = \arccos(m_\varphi). \qquad (4)$$

From the set of 3839 MIR-phased reflections inside 4 Å, only those reflections with initial contribution to χ^2 greater than 2 (*i.e.* $|F| > \sqrt{2}\sigma$, where σ includes the estimated error in the phase φ) when starting with a flat map for **p** were selected for inclusion in the first constraint as well as those reflections with $|F| < 0{\cdot}004$. Thus only 1669 reflections (about $\frac{1}{2}$ in total $\sum |F_k|$ for data inside 4 Å) were regarded as being strong contributors to the initial structure refinement process.

3. Results

The numerical methods used in the present study are based on the work of references SG1 and Wilkins (1983) and a skeletal outline of the method and algorithm is presented in Appendix 1. With these methods, a ME map was calculated on a grid 32×48×40 over the unit cell based on a subset of the 4 Å data as selected above, and some selected regions of the ME map are presented in Figs. 1 and 2. Convergence of the solution to a ME solution was established *via* the Skilling criterion (Skilling, 1983) involving a test for parallelism between the directions of ∇S and $\nabla\chi^2$ to a cosine magnitude better than 0·995. A target value for $C_1 = 0{\cdot}84$ was chosen in the refinement, as it was deemed better to err slightly on the side of

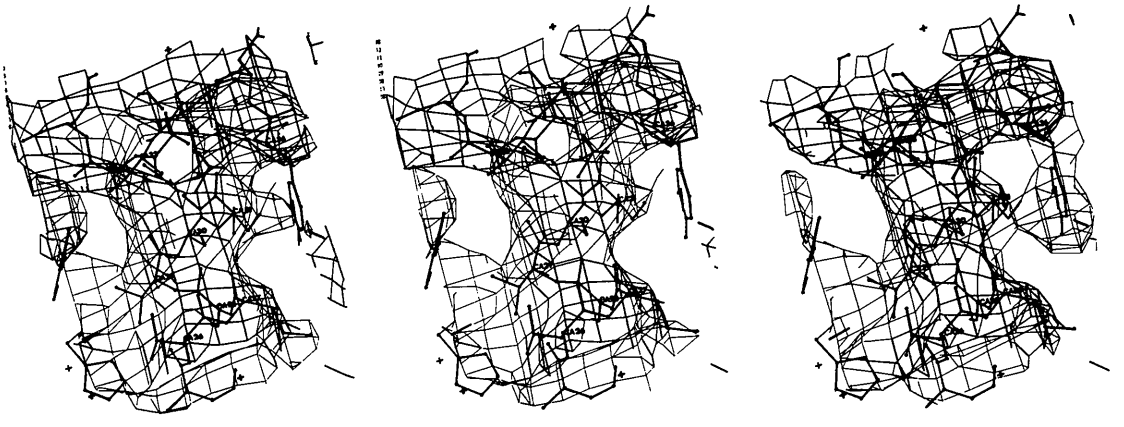
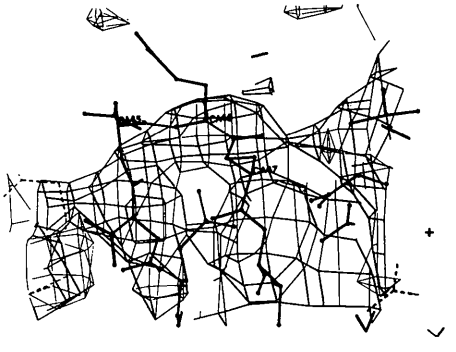


Fig. 1. Equivalent portions of the figure-of-merit-weighted MIR map and the maximum-entropy distribution. The heavy lines join bonded atoms in the partially refined ($R = 0{\cdot}27$ for data from ∞ to 1·7 Å) model of α-lactalbumin. The distributions are shown as meshes enclosing areas of density above the given threshold. These contour levels were chosen to present as fair a comparison as possible. Exactly equivalent volumes of the map (and viewpoints) are shown for the two maps. (*a*), (*b*) and (*c*) are centred on residue 30 with (*a*) corresponding to the MIR map with contour level 1750, (*b*) the MIR map at contour level 2000 and (*c*) the ME map at contour level 1500. Note that the density for the ME distribution is generally sharper, producing an effect of somewhat better resolution. That this is not simply an effect of the choice of contour level may be seen from the density around residue Tyr 36.

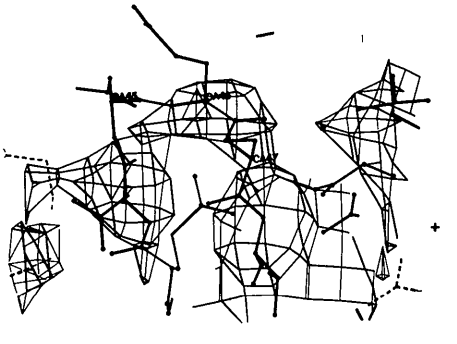
including some noise in the map rather than losing possible signal. These ME maps may be compared with the corresponding figure of merit (FoM)-weighted maps (Figs. 1 and 2) also presented. In each case a general sharpening of the ME map relative to the FoM maps can be seen. Contour levels were chosen as those appropriate for each map and are listed in the figure captions. Moreover, when the various maps are compared with the current best model for α-lactalbumin, several improvements in the ME maps can be seen, in addition to the general sharpening, involving a redistribution of the density. In particular, it is very encouraging to see (Figs. 2*a*, *b*, *c*) that some of the least well defined portions of the map are improved since some previous attempts at structure refinement have, on the contrary, led to the stronger features being enhanced at the expense of the weaker ones (Blundell & Johnson, 1976). We consider that, although the differences between the ME map and the appropriately contoured FoM map are subtle, they might nevertheless be of considerable significance when attempting interpretation of the maps in terms of a molecular structure.

In order to help give a quantitative guide as to the extent of the refinement process in reciprocal space, we introduce the following correlation function ('figure of merit').

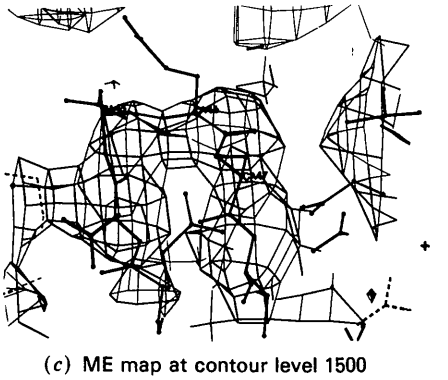
$$\Gamma^{AB} = \sum_k |F_k^{\rm obs}| \exp\{i(\varphi_k^A - \varphi_k^B)\} \Big/ \sum_k |F_k^{\rm obs}| \quad (5)$$



(*a*) FoM weighted map at contour level 1750

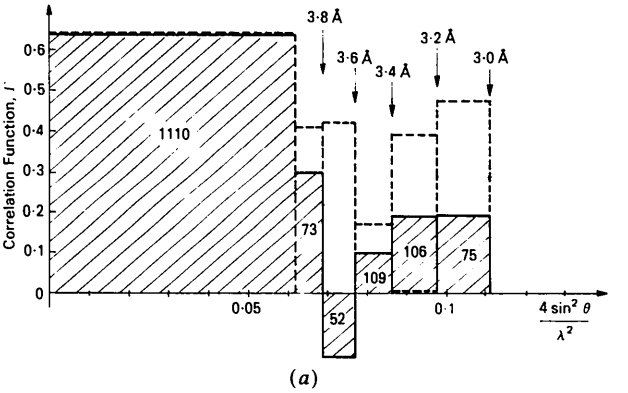


(*b*) FoM weighted map at contour level 2000

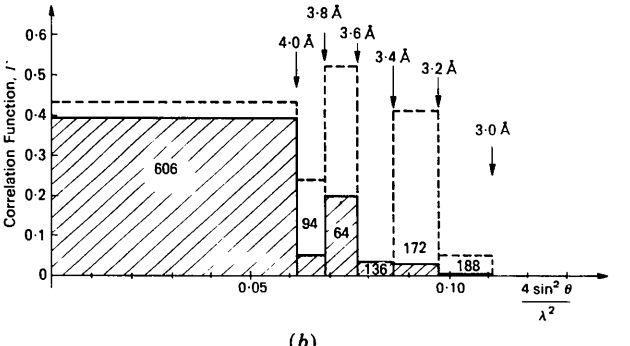


(*c*) ME map at contour level 1500

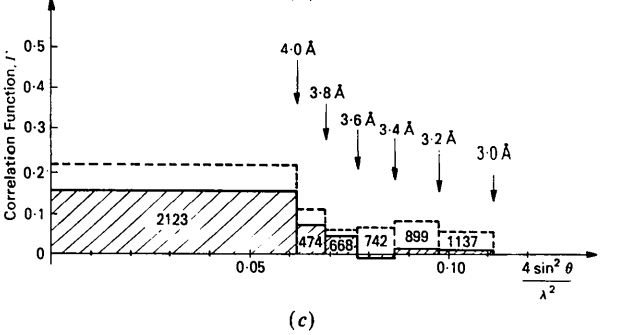
Fig. 2. (*a*), (*b*) and (*c*) correspond to Figs. 1(*a*), (*b*) and (*c*) respectively but are centred on residue 46. Note that the external loop is much more clearly defined in the ME map.



(*a*)



(*b*)



(*c*)

Fig. 3. Correlation function, Γ^{AB}, for agreement in phase between ME and model phases (solid-line histogram) and MIR and model phases (broken-line histogram) as a function of scattering angle. (*a*) is for high-FoM reflections, while (*b*) and (*c*) are essentially for intermediate- and low-FoM reflections respectively.

between sets A and B of corresponding phases. This function is clearly +1 if corresponding phases in A and B are equal and tends to zero if corresponding phases are uncorrelated. In Fig. 3 this correlation function is plotted as a function of scattering angle for the correlation taken between the refined model phases and:

(i) the corresponding phases obtained from Fourier transforming the ME map (solid-line histogram);

(ii) the corresponding MIR phases (broken line histogram).

The three histograms relate to all reflection data inside 3 Å subdivided according to the following categories: (a) those reflections having figures of merit greater than or equal to 0·85 (Fig. 3a); (b) those reflections not included in (a) but with individual contribution to χ^2 greater than 2 relative to the flat map or with $|F| < 0{\cdot}004$ (Fig. 3b); and (c) the remainder (Fig. 3c). From these results we observe that: (α) the MIR phases included in the χ^2 constraint are well fitted by ME refinement; (β) considerable phase information is generated by the ME refinement for reflections *not* included in the χ^2 constraint. This information is found both within the 4 Å shell (phase interpolation) and beyond it (phase extension); (γ) there is very little correlation between the information (β) and the MIR phases. This last observation is particularly encouraging since it suggests that we could treat these two sources of phase information as independent and, in principle at least, successively repeat the ME refinement with each reflection in turn eliminated from the χ^2 constraint in order to obtain a set of ME phase estimates which would then be combined with the MIR phase probability distributions. In principle, this would overcome one of the limitations of the single-pass ME methods, namely

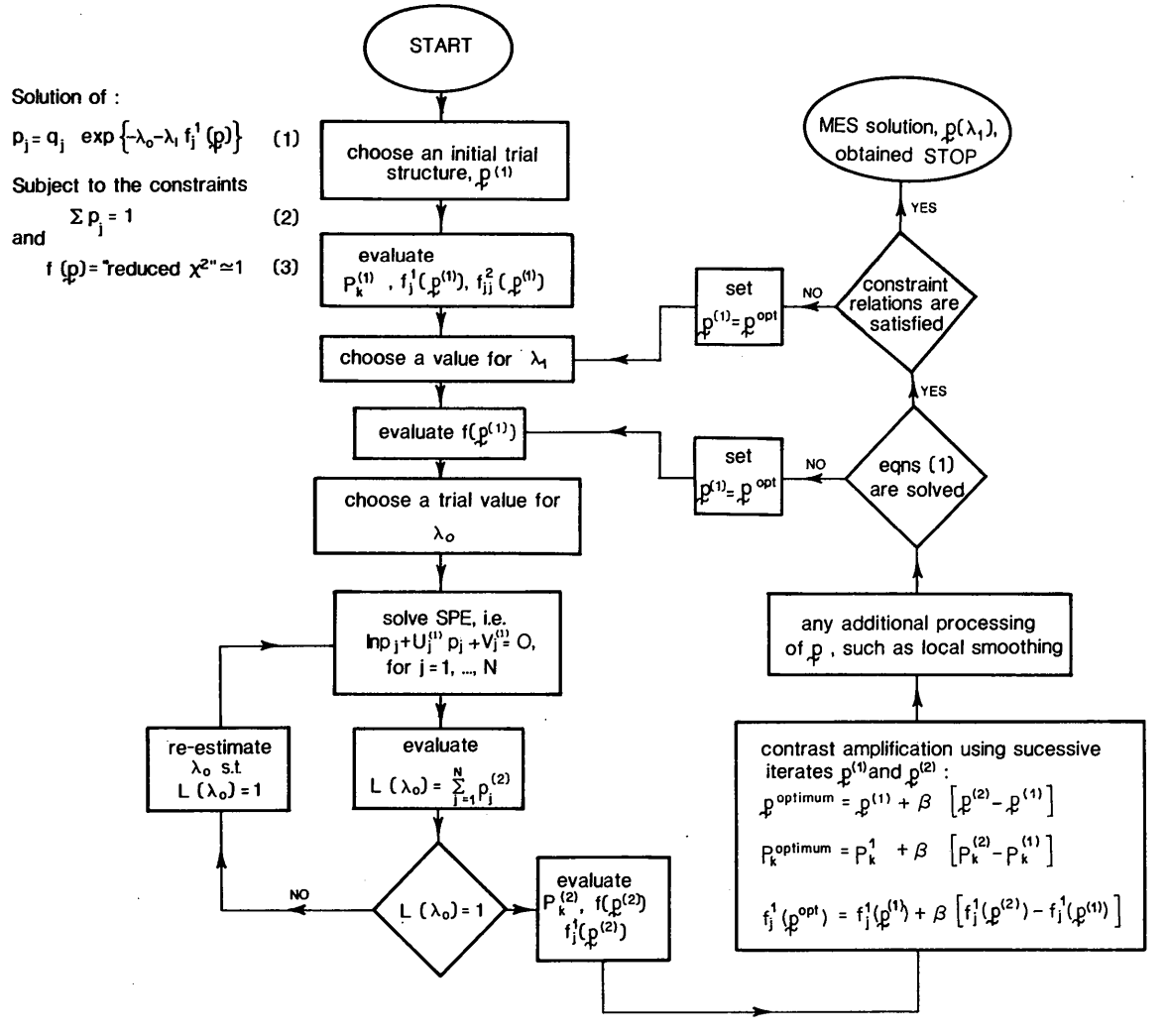


Fig. 4. Flow chart of the algorithm used in the present maximum-entropy refinement of data for α-lactalbumin using only the first constraint [equation (2) of text]. Equation numbers inside flow chart refer to those in upper-left-hand corner of the figure.

the representation of the, frequently bimodal, MIR phase probability distributions as unimodal.

More generally, other types of information, such as those listed in SG1 and including non-convex constraints, may be incorporated into the maps *via* ME and using the present map as a starting solution. The more information that is introduced, the smaller should be the range of allowable solutions, even when uniqueness cannot be guaranteed.

We are grateful to various colleagues for helpful discussion and encouragement including Drs S. Steenstrup, J. N. Varghese, P. Colman, A. McL. Mathieson, W. Hendrickson and J. Skilling. One of us (SWW) is grateful to the Royal Society for the award of a Visiting Fellowship and to Professor Sir David Phillips for his encouragement of this work and for extending the hospitality of his department for a stay during which this work was largely undertaken.

APPENDIX 1

The algorithm used in deriving the present ME structures is essentially that outlined in our previous work (SG1; Wilkins, 1983), but with only the first constraint [equation (2)] operative. With the same notation as in our earlier works, a flow chart of the algorithm is presented in Fig. 4 and includes indications as to where additional processing (such as local smoothing) may be carried out, although such processing was not actually carried out in the present study.

References

BLUNDELL, T. L. & JOHNSON, L. N. (1976). *Protein Crystallography*, p. 438 (and references therein). London: Academic Press.
BRICOGNE, G. (1984). *Acta Cryst.* A**40**, 410-445.
COLLINS, D. M. (1982). *Nature (London)*, **298**, 49-51.
GULL, S. F. & DANIELL, G. J. (1978). *Nature (London)*, **272**, 686-690.
HENDRICKSON, W. A. & TEETER, M. M. (1981). *Nature (London)*, **290**, 107-113.
SKILLING, J. (1983). Paper presented at Opt. Soc. Am. Meet. on Signal Recovery and Synthesis with Incomplete Information and Partial Constraints, Incline Village, Nevada.
SMITH, S. G. (1982). D. Phil. thesis, Univ. of Oxford. Unpublished.
STEENSTRUP, S. & WILKINS, S. W. (1984). *Acta Cryst.* A**40**, 163-164.
STUART, D. (1985). Unpublished results.
STUART, D., PHILLIPS, D. C., LEWIS, M. & SMITH, S. G. (1985). Unpublished.
WILKINS, S. W. (1983). *Acta Cryst.* A**39**, 892-896, 896-898.
WILKINS, S. W., STEENSTRUP, S. & VARGHESE, J. N. (1985). In *Crystallographic Statistics*, edited by A. J. C. WILSON, pp. 113-124. New York: Adenine Press.
WILKINS, S. W., VARGHESE, J. N. & LEHMANN, M. S. (1983). *Acta Cryst.* A**39**, 47-60.

SHORT COMMUNICATIONS

Contributions intended for publication under this heading should be expressly so marked; they should not exceed about 1000 *words; they should be forwarded in the usual way to the appropriate Co-editor; they will be published as speedily as possible.*

Acta Cryst. (1986). A**42**, 202-203

Green's tensor for a cubic crystal.

By MUSTAFA DİKİCİ, *Department of Physics, İnönü University, Malatya, Turkey*

(*Received* 12 *September* 1985; *accepted* 20 *December* 1985)

Abstract

The equilibrium equations of classical elasticity for a cubic crystal are solved and Green's tensor for elastic displacements is derived.

1. Introduction

The point force technique provides a powerful tool for the solution of problems in the continuum theory of elasticity, particularly in the continuum theory of dislocation (Hirth & Lothe, 1968). The only difficulty arising is in representing Green's tensor. It is known explicitly for isotropy and transverse isotropy, but not general anisotropy. The aim of this paper is to obtain Green's tensor for an infinite three-dimensional body with cubic symmetry.

2. Formulation of the problem

The usual suffix notation will be employed. The summations should be carried out on repeated indices. This convention is adopted throughout this paper.

The equilibrium equation of classical elasticity is most easily obtained in a coordinate system whose bases parallel the cubic axes of the matrix. The cubic matrix has three independent elastic constants, the Voigt constants c_{11}, c_{12} and c_{44} (Hirth & Lothe, 1968). With the definitions

$$c_{12}=\lambda, \qquad c_{44}=\mu, \qquad c_{11}-c_{12}-2c_{44}=\lambda_1,$$

the fourth-order elastic tensor may be written in the following form:

$$C_{ijkm}=\lambda\delta_{ij}\delta_{km}+\mu(\delta_{ik}\delta_{jm}+\delta_{im}\delta_{jk}) + \lambda_1\delta_{ij}\delta_{jk}\delta_{km}, \qquad (1)$$

0108-7673/86/030202-02$01.50