vectorized SGFFT algorithm is up to 26 times faster than the vectorized direct summation method on the Cray-XMP. Since the SGFFT algorithm is designed in a general and simple way, it is expected that it could be efficiently implemented on supercomputers with new parallel architectures.

The author thanks M. Karplus, J. Kuriyan, G. A. Petsko and W. I. Weis for useful discussions. The work described in this paper was begun while the author was a research associate at Harvard University.

### References

AGARWAL, R. C. (1978). *Acta Cryst.* A**43**, 791-809.
AGARWAL, R. C. (1981). In *Refinement of Protein Structures*, pp. 24-28. SRC Daresbury Laboratory, Warrington, England.
BRIGHAM, E. O. (1974). In *The Fast Fourier Transform*. Englewood Cliffs, NJ: Prentice-Hall.
BRÜNGER, A. T. (1988*a*). *Crystallographic Refinement by Simulated Annealing*. In *Crystallographic Computing* 4: *Techniques and New Technologies*, edited by N. W. ISAACS & M. R. TAYLOR. Oxford Univ. Press.
BRÜNGER, A. T. (1988*b*). *J. Mol. Biol.* In the press.
BRÜNGER, A. T., KARPLUS, M. & PETSKO, G. A. (1989). *Acta Cryst.* A**45**, 50-61.
BRÜNGER, A. T., KURIYAN, K. & KARPLUS, M. (1987). *Science*, **235**, 458-460.
COCHRAN, W. (1948). *Acta Cryst.* **1**, 138-142.
COOLEY, J. W. & TUKEY, J. W. (1965). *Math. Comput.* **19**, 297-301.
FINZEL, B. C. (1987). *J. Appl. Cryst.* **20**, 53-55.
FLETCHER, R. & REEVES, C. M. (1964). *Comput. J.* **7**, 149-154.
HENDRICKSON, W. A. & TEETER, M. A. (1981). *Nature (London)*, **290**, 107-112.
*International Tables for X-ray Crystallography* (1952). Vol. I. Birmingham: Kynoch Press.
*International Tables for X-ray Crystallography* (1974). Vol. IV. Birmingham, Kynoch Press. (Present distributor Kluwer Academic Publishers, Dordrecht.)
ISAACS, N. (1984). In *Methods and Applications in Crystallographic Computing*, edited by S. R. HALL & T. ASHIDA, pp. 193-205. Oxford: Clarendon Press.
JACK, A. & LEVITT, M. (1978). *Acta Cryst.* A**34**, 931-935.
KONNERT, J. H. & HENDRICKSON, W. A. (1980). *Acta Cryst.* A**36**, 344-349.
LUNIN, V. Y. & URZHUMTSEV, A. G. (1985). *Acta Cryst.* A**41**, 327-333.
MOSS, D. S. & MORFFEW, A. J. (1982). *Comput. Chem.* **6**, 1-3.
RAFTERY, J., SAWYER, L. & PAWLEY, G. S. (1985). *J. Appl. Cryst.* **18**, 424-429.
SMITH, D. L., RINGE, D., FINLAYSON, W. L. & KIRSCH, J. F. (1986). *J. Mol. Biol.* **191**, 301-302.
SOWADSKI, J. M., HANDSCHUMACHER, M. D., MURTHY, H. M. K., FOSTER, B. A. & WYCKOFF, H. W. (1985). *J. Mol. Biol.* **186**, 417-433.
SUSSMAN, J. L., HOLBROOK, S. R., CHURCH, G. M. & KIM, S. H. (1977). *Acta Cryst.* A**33**, 800-804.
TEN EYCK, L. F. (1973). *Acta Cryst.* A**29**, 183-191.
TEN EYCK, L. F. (1977). *Acta Cryst.* A**33**, 486-492.
TRONRUD, D. E., TEN EYCK, L. F. & MATTHEWS, B. W. (1987). *Acta Cryst.* A**43**, 489-500.

*Acta Cryst.* (1989). A**45**, 50-61

# Crystallographic Refinement by Simulated Annealing: Application to Crambin

BY AXEL T. BRÜNGER

*The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT* 06511, *USA, and Department of Chemistry, Harvard University, Cambridge, MA* 02138, *USA*

MARTIN KARPLUS

*Department of Chemistry, Harvard University, Cambridge, MA* 02138, *USA*

AND GREGORY A. PETSKO

*Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA* 02139, *USA*

(*Received* 22 *April* 1988; *accepted* 20 *July* 1988)

### Abstract

A detailed description of the method of crystallographic refinement by simulated annealing is presented. To test the method, it has been applied to a 1·5 Å resolution X-ray structure of crambin. The dependence of the success of the simulated annealing protocol with respect to the temperature of the heating stage is discussed. Optimal success is achieved at relatively high temperatures. Regardless of the protocol used, the molecular-dynamics refined structure always yields an improved $R$ factor compared with restrained least-squares refinement without manual re-fitting. The differences between the various refined structures and the corresponding electron density maps are discussed.

0108-7673/89/010050-12$03.00 © 1989 International Union of Crystallography

## Introduction

Conventional refinement of the X-ray structure of biological macromolecules involves a series of steps, each of which consists of a few cycles of least-squares refinement with stereochemical and internal packing restraints (Sussmann, Holbrook, Church & Kim, 1977; Jack & Levitt, 1978; Konnert & Hendrickson, 1980; Moss & Morffew, 1982; Hendrickson, 1985; Tronrud, Ten Eyck & Matthews, 1987), that are followed by re-fitting the model structure to difference electron density maps with interactive computer graphics (Jones, 1978). During the final stages of refinement, solvent molecules are usually included and alternative conformations for some atoms or residues in the protein may be introduced.

The aim of least-squares refinement is to minimize the difference between the observed $[|F_{obs}(\mathbf{h})|]$ and calculated $[|F_{calc}(\mathbf{h})|]$ structure-factor amplitudes, which is usually expressed as a weighted sum of the residual

$$\sum_{\mathbf{h}} W_{\mathbf{h}}[|F_{obs}(\mathbf{h})| - k|F_{calc}(\mathbf{h})|]^2 \qquad (1)$$

where $\mathbf{h} = (h, k, l)$ are indices referring to the reciprocal-lattice points of the crystal.

In the case of macromolecules, the restrained least-squares refinement (RLSQ refinement) procedure is easily trapped in a local minimum and it does not correct the positions of residues that are misplaced by more than about 1 Å so that manual adjustments of the model structure are necessary. Crystallographic refinement can be understood as a nonlinear optimization problem with the aim of finding a minimum close to the global minimum of a target function $f(X) = f(x_1, x_2, \ldots, x_n)$ containing the residual sum [(1)], and stereochemical and other interactions of the macromolecule; the quantities $x_1, x_2, \ldots, x_n$ represent the variables of the system, such as the atomic positions or individual atomic temperature factors. In the past few years considerable progress has been reported in nonlinear optimization problems by the introduction of simulated annealing (Kirkpatrick, Gelatt & Vecchi, 1983). Simulated annealing (SA) is now in widespread use in areas such as electronic circuit design (Soukop, 1981). The method usually consists of simulating the many-parameter system by a Monte Carlo algorithm (Metropolis, Rosenbluth, Rosenbluth, Teller & Teller, 1953) which involves trial moves of $x_1, x_2, \ldots, x_n$. Gradient descent methods, such as RLSQ refinement, accept only moves that reduce $f(X)$ and, therefore, cannot escape from local minima. Monte Carlo algorithms, on the other hand, accept certain moves that increase $f(X)$, as well as those that decrease $f(X)$. A move which increases $f(X)$ is accepted with a probability given by $\exp[-\Delta f(X)/kT]$ where $\Delta f(X)$ is the difference between the values of $f(X)$ before and after the trial move, $T$ is the 'temperature' of the system and $k$ is Boltzmann's constant. In fact, $T$ should be understood not as a physical temperature but rather as a control parameter that determines whether the system can escape certain local minima. Thus, very high values of $T$ may have to be introduced if the barriers between local minima are large. The success of simulated annealing is dependent on the 'annealing' schedule which determines how the temperature is modified during the simulation. Initially, the temperature is kept very high and the system is then 'annealed' by slowly reducing the temperature. In other words, a coarse search is carried out at high temperatures and a local minimum is approached during the cooling stage. This procedure may have to be repeated to reach the global minimum. The major difficulty in applying simulated annealing to specific problems appears to be the choice of an efficient annealing schedule (Bounds, 1987).

A first attempt to introduce simulated annealing into crystallographic refinement has been reported (Brünger, Kuriyan & Karplus, 1987). A major difference from the optimization problems discussed by Kirkpatrick *et al.* (1983) is that one is optimizing a single unit with internal degrees of freedom representing a macromolecule rather than a fluid-like system consisting of many identical subunits. A direct application of the Metropolis algorithm to macromolecules turns out to be inefficient if all degrees of freedom are included, as the covalent bonds of the system will lead to rejection of most steps taken by the algorithm. Instead, molecular dynamics (Karplus & McCammon, 1983) can be used to follow the gradients of a target function and to introduce a temperature into the system to escape from local minima. Refinement by simulated annealing with molecular dynamics proceeds in the same way as with the Monte Carlo algorithm: there is a heating stage and a cooling stage, with heating and cooling possibly repeated several times. The conformational space searched by molecular dynamics is confined to regions allowed by stereochemical and other restraints by the inclusion of an empirical potential energy term (Brooks, Bruccoleri, Olafson, States, Swaminathan & Karplus, 1983) in the target function. The diffraction data are introduced into the target function by including the crystallographic residual [(1)] as an additional term. The resulting effective potential energy is similar to the function used in refinement by least-squares minimization (Jack & Levitt, 1978).

Brünger, Kuriyan & Karplus (1987) showed that SA refinement has a radius of convergence that is significantly larger than that of conventional RLSQ refinement and that the method can reduce the need for manual corrections. In the present paper a detailed description of the SA-refinement method is given with a brief description of the program. The robustness of the method with respect to data resolution and

temperature during the heating stage is demonstrated by application to a 1·5 Å resolution structure of crambin (Hendrickson & Teeter, 1981).

## Methodology

In this section we first describe the form of the potential energy function with emphasis on the changes required to avoid artifacts that could be introduced by the high temperatures used in the simulation. Next, we describe how the potential energy function can be extended to include nonbonded interactions between symmetry-related molecules. We then show how the temperature can be controlled during a molecular dynamics calculation, followed by a description of an effective potential energy that contains information about the diffraction data. We then give a brief description of the program *X-PLOR* used in the SA refinement. Finally, we introduce the system studied in the present paper.

### *Empirical potential energy*

The empirical potential energy $E_i$ is a function of all atomic positions of the system describing internal stereochemical interactions (bond lengths, bond angles, dihedral torsion angles, chiral centers, planarity of aromatic rings) as well as nonbonded (van der Waals and electrostatic) interactions.

$$E_i = \sum_{\text{bonds}} k_b(r-r_0)^2 + \sum_{\text{angles}} k_\theta(\theta-\theta_0)^2 + \sum_{\text{dihedrals}} k_\varphi \cos(n\varphi + d) + \sum_{\text{chiral, planar}} k_\omega(\omega-\omega_0)^2 + \sum_{\text{atom pairs}} (ar^{-12} + br^{-6} + cr^{-1}). \quad (2)$$

The parameters of the empirical potential energy $E_i$ are inferred from experimental as well as theoretical investigations (Lifson & Stern, 1982; Levitt, 1983; Brooks *et al.*, 1983; Némethy, Pottie & Scheraga, 1983; Hermans, Berendsen, van Gunsteren & Postma, 1984; Nilsson & Karplus, 1986; Weiner, Kollman, Nguyen & Case, 1986).

In this work stereochemical and nonbonded parameters for the empirical potential energy were taken from the explicit polar hydrogen parameter set PARAM19 and TOPH19 of *CHARMM* (Brooks *et al.*, 1983) with several modifications that avoid unphysical structural changes induced by the use of very high temperatures in the dynamics. The standard twofold dihedral angle potential energy for rotations around H–N=C–O peptide bonds has two minima, allowing both *trans* and *cis* peptide bonds. It was replaced by a onefold potential energy allowing only *trans* peptide bonds. The force constant was set to 418·4 kJ mol$^{-1}$ rad$^{-2}$. With the standard twofold term, transitions of peptide bonds from *trans* to *cis* can occur at high temperatures. Thus, in the present form of the calculation, *cis* peptide bonds are excluded for most amino acids unless they are specifically introduced by manual intervention. In the case of proline peptide bonds, the twofold term was retained with a reduced force constant of 20·9 kJ mol$^{-1}$ rad$^{-2}$ to allow transitions between *cis* and *trans* prolines. The force constant of the improper torsion angle which specifies the tetrahedral geometry of $C^\alpha$ carbon atoms was increased to 2092 kJ mol$^{-1}$ rad$^{-2}$. This was done in order to avoid transitions from L- to D-amino acids at high temperatures. The force constants of the improper torsion angles which specify the planarity of aromatic rings, such as Phe or Tyr, were increased to 1046 kJ mol$^{-1}$ rad$^{-2}$. In addition, the 1-4 and 1-5 nonbonded interactions in aromatic rings were included in the calculation. Without the increased force constants and the nonbonded interactions, aromatic rings could be twisted into unreasonable conformations in cases where the initial structure contained close contacts around the aromatic ring.

The Lennard-Jones interactions were multiplied by a cubic switching function and the electrostatic interactions were multiplied by a quartic shifting function (Brooks *et al*, 1983); nonbonded interactions were included up to 7·5 Å. The reduced cutoff was used in order to decrease total computational time. A dielectric constant of unity was employed throughout the system.

### *Crystal symmetry interactions*

Experience has shown that the influence of crystal packing interactions can be important in crystallographic refinement. In the absence of crystal packing information, backbone or side-chain atoms of a molecule in the asymmetric unit can penetrate symmetry-related molecules. This can occur in contact regions between molecules where the electron density does not clearly define the molecular boundaries. A potential energy which describes nonbonding interactions between the molecule(s) located in the asymmetric unit and all symmetry-related molecules surrounding the asymmetric unit is given by

$$E_i^{\text{crystal}} = \sum_s \sum_{i \geq j} \text{NB}\{|\mathcal{F}^{-1}\,\text{MinG}(\mathcal{F}\mathbf{r}_i - \mathcal{O}_s\mathcal{F}\mathbf{r}_j + \mathbf{t}_s)|\} \quad (3)$$

where NB{ } and MinG( ) are functions that are defined below, $\mathcal{F}$ is the matrix that converts orthogonal coordinates into fractional coordinates, the first sum extends over all symmetry operators $(\mathcal{O}_s, \mathbf{t}_s)$ of the crystal, and the second sum extends over all pairs of atoms $(i, j)$ for which the argument of the function NB{ } is less than a specified cutoff $r_{\text{cut}}$, and $r_i$ and $r_j$ are the coordinates of atoms $i$ and $j$ respectively. The function MinG(**r**) defines the minimum image distance in fractional coordinate

space. It operates on each component of the three-dimensional vector **r** separately where the operation on each component $x$ is given by

$$\text{MinG}(x) = \text{sign}(-x)\,\text{int}(|x|+\tfrac{1}{2})+x. \quad (4)$$

The function int $(y)$ is defined as the integer part of $y$, and sign $(y)$ is defined as the sign of $y$. For the application of (3) and (4) to compute the crystal symmetry interactions, one has to make the assumption that the cutoff $r_{cut}$ is less than half of any of the edges of a *maximal* orthogonal box fitting into the unit cell of the crystal. In the special case of an orthorhombic space group $r_{cut}$ has to be less than half of any of the unit-cell vectors of the crystal. Thus, the present simplified approach is limited to large molecules such as proteins. NB $\{r\}$ has the form of a nonbonding interaction potential, *i.e.* it is a sum of van der Waals and electrostatic interactions,

$$\text{NB}\{r\} = ar^{-12} - br^{-6} + cr^{-1}, \quad (5)$$

where $a$, $b$, $c$ are taken from Brooks *et al.* (1983). The form of NB $\{r\}$ is chosen to be identical to the potential used for intramolecular interactions. Thus, the nonbonding interaction energy between two atoms is independent of whether it is an intermolecular or an intramolecular interaction. The pairwise search in (3) for interactions less than $r_{cut}$ is computationally expensive. The computational time is greatly reduced by introducing an approximation which consists of storing the atomic pair indices $(i,j)$ of all crystal-symmetry interactions less than $r_{cut}$ in a list that is only updated when any atom has moved by more than 0·5 Å. The pairwise search is carried out by searching for nearest neighbors among all residue centroids and then by searching the atom pairs between the selected residue pairs. Equation (3) includes only intermolecular interactions between different molecules since the nonbonding interaction energy for atoms in the asymmetric unit is already included in the standard potential energy $E_i$. The crystal symmetry interaction energy $E_i^{\text{crystal}}$ is added to the empirical potential energy $E_i$.

*Conjugate gradient minimization*

Energy minimization of the empirical potential energy $E_i$ is carried out by use of a conjugate gradient minimization algorithm (Fletcher & Reeves, 1964; Powell, 1977). The algorithm requires the value of the energy and its first derivatives. It does not require explicit computation of second derivatives since information is being built up about second derivatives during subsequent shifts in different directions. This implies that the conjugate gradient minimization technique requires a few more steps to get 'started' than does the least-squares method used in *PROLSQ* (Konnert & Hendrickson, 1980). (Note that in *PROLSQ* conjugate gradient minimization is used to invert the least-squares matrix, but it is not used actually to minimize the crystallographic residual.)

*Molecular dynamics*

Molecular dynamics simulations involve the simultaneous solution of the classical equations of motion for all atoms of a macromolecule where the forces are derived from the empirical potential energy $E_i$ (Karplus & McCammon, 1983). The solution of the classical equations of motion is carried out numerically by the Verlet (1967) algorithm. The initial velocities are assigned to a Maxwellian distribution at the appropriate temperature. Control of the temperature during the molecular dynamics calculation is obtained by periodic uniform re-scaling of the velocities $v_i$, *i.e.*

$$V_i^{\text{new}} = \text{scale} \times V_i^{\text{old}} \quad (6)$$

for all atoms $i$. The factor scale is given by

$$\text{scale} = \left\langle \sum_i m V_i^{\text{old}}(t)^2 \right\rangle \Big/ nkT \quad (7)$$

where the sum is carried out over all atoms $i$ of the system. The parameter $n$ is the number of degrees of freedom of the system, $k$ is Boltzmann's constant, $T$ is the temperature, $V_i^{\text{old}}(t)$ denotes the velocity of atom $i$ at time $t$, and $\langle\,\rangle$ denotes a trajectory average over the time intervals between the rescaling of the velocities. The rescaling of velocities enforces a constant temperature during the molecular dynamics simulation.

*Structure-factor difference expressed as an effective potential energy*

In SA refinement the empirical potential energy $E_i$ is augmented by an effective potential energy $E_x$ containing information about the diffraction data (Brünger, Kuriyan & Karplus, 1987; Brünger, 1988*a*). The total potential energy $E_{\text{total}}$ is then given by

$$E_{\text{total}} = E_i + E_x. \quad (8)$$

At the present state of empirical potential energy parameterization, a molecular dynamics simulation without the effective potential energy term $E_x$ cannot reproduce the crystal structure to sufficient accuracy (Kuriyan, Petsko, Levy & Karplus, 1986), even when one starts with a solved X-ray structure. It is the combination of the empirical potential energy and the effective potential energy describing the crystallographic residual that makes it possible to use molecular dynamics in SA refinement.

The effective potential energy $E_x$ consists of the weighted differences between observed ($|F_{\text{obs}}|$) and calculated ($|F_{\text{calc}}|$) structure-factor amplitudes,

$$E_x = W_A / N_A \sum_{\mathbf{h}} W_{\mathbf{h}} [|F_{\text{obs}}(\mathbf{h})| - k|F_{\text{calc}}(\mathbf{h})|]^2. \quad (9)$$

which is identical to (1) except for the additional weight $W_A$ which relates $E_x$ to the empirical potential energy $E_i$; $N_A$ is a normalization factor. The purpose of the normalization factor $N_A$ is to make the weight $W_A$ approximately independent of the resolution range during SA refinement. Several choices would be possible, *e.g.* $N_A = \sum_{\mathbf{h}} W_{\mathbf{h}}$ or $N_A = \sum_{\mathbf{h}} W_{\mathbf{h}} |F_{\rm obs}(\mathbf{h})|^2$. By trial and error we found that the latter expression for $N_A$ worked somewhat better than the former. The sum in (9) extends over all observed reflections with indices **h**. $W_{\mathbf{h}}$ are the individual weights for each reflection **h**. The scale factor $k$ is set to the value which makes the derivative of $E_x$ with respect to $k$ zero,

$$k = \sum_{\mathbf{h}} W_{\mathbf{h}} |F_{\rm obs}(\mathbf{h})| |F_{\rm calc}(\mathbf{h})| \Big/ \left[ \sum_{\mathbf{h}} W_{\mathbf{h}} |F_{\rm calc}(\mathbf{h})|^2 \right]. \quad (10)$$

This is a necessary condition to make $E_x$ minimal. The effective potential energy $E_x$ is the same as the function used by Jack & Levitt (1978), except for the treatment of the scale factor $k$. In Jack & Levitt, $k$ is treated as an independently minimized parameter whereas in this work $k$ is eliminated by inserting (10) into (9). If $k$ is not eliminated from (9) it would be necessary to optimize the choice of $k$ at each molecular dynamics step. This would introduce a slight inconsistency between $E_x$ and its first derivatives with respect to the atomic positions.

The structure factors $[F_{\rm calc}(\mathbf{h})]$ of the atomic model can be expressed in the following way:

$$F_{\rm calc}(\mathbf{h}) = \sum_s \sum_i f_i(\mathbf{h}) \exp[-B_i(\mathcal{F}^*\mathbf{h})^2/4] \times \exp[2\pi i \mathbf{h} \cdot (\mathcal{O}_s \mathcal{F} \mathbf{r}_i + \mathbf{t}_s)]. \quad (11)$$

The first sum extends over all symmetry operators $(\mathcal{O}_s, \mathbf{t}_s)$ composed of the matrix $\mathcal{O}_s$ representing a rotation and a vector $\mathbf{t}_s$ representing a translation. The second sum extends over all unique atoms $i$ of the system. The quantity $\mathbf{r}_i$ denotes the orthogonal coordinates of atom $i$ in real space. $\mathcal{F}$ is the $3 \times 3$ matrix that converts orthogonal coordinates into fractional coordinates; $\mathcal{F}^*$ denotes the transpose of it. The columns of $\mathcal{F}^*$ are equal to the reciprocal unit-cell vectors $\mathbf{a}^*$, $\mathbf{b}^*$, $\mathbf{c}^*$. $B_i$ is the individual atomic temperature factor for atom $i$. The atomic scattering factors $f_i(\mathbf{h})$ are approximated by an expression consisting of four Gaussians and a constant,

$$f_i(\mathbf{h}) = \sum_{i=1}^{4} a_{ki} \exp[-b_{ki}(\mathcal{F}^*\mathbf{h})^2/4] + a_{0i}. \quad (12)$$

The constants $a_{ki}$ and $b_{ki}$ were obtained from *International Tables for X-ray Crystallography* (1974). Equation (11) represents the space-group-general form of the 'direct summation' formula to compute the structure factors. Somewhat more efficient expressions can be obtained for each space group separately by appropriate re-summation.

Expressions for derivatives of $F_{\rm calc}(\mathbf{h})$ with respect to the atomic coordinates $\mathbf{r}_i$ are needed for molecular dynamics and energy minimization; they can be obtained from (9), (11), (10), and (12) (*cf.* Brünger, 1989). The functional dependence of $k$ on $f_{\rm calc}(\mathbf{h})$ is included implicitly in the computation of the derivatives of $E_x$ with respect to the atomic model parameters [(10)].

Equation (11) represents a simple way to compute the structure factors and derivatives of the atomic model, but it is computationally very expensive. For example, in the case of the aspartate aminotransferase, a protein consisting of 396 amino acids, the evaluation of the right-hand side of (11) and its derivatives takes 102 s on a Cray-2 at 2·8 Å resolution in space group $C222_1$ whereas the evaluation of the empirical potential energy $E_i$ takes only 2·8 s. In principle, the structure factors and their derivatives should be re-computed at each time step of the numerical integration of the equations of motion. To reduce the computational requirements of SA refinement we have introduced two approximations. The first approximation consists of not computing $F_{\rm calc}(\mathbf{h})$ and its first derivatives at every dynamics step as already suggested by Jack & Levitt (1978). The first derivatives are kept constant until any atom has moved by more than $\Delta_F$ relative to the position at which the derivatives were last computed. At that point all derivatives are updated. Tests have shown that for $\Delta_F$ less than 0·3 Å this approximation does not affect the convergence properties and it speeds up the computation by almost an order of magnitude. For minimizations $\Delta_F$ has to be set to small values (0·0–0·05 Å) in order to ensure that the gradient of $F_{\rm calc}(\mathbf{h})$ is smooth function of the displacement vector used in the minimizing algorithm. This is particularly important during the final cycles of the minimization.

The second approximation consists of computing $F_{\rm calc}(\mathbf{h})$ by numerical evaluation of the atomic electron density on a finite grid followed by fast Fourier transformation (FFT). The FFT method provides a way to speed up the calculation greatly (Ten Eyck, 1973; Agarwal, 1978). The implementation of the FFT method for SA refinement with particular emphasis on supercomputer applications is discussed by Brünger (1989).

An indicator for the progress of the crystallographic refinement is the $R$ factor

$$R = \sum_{\mathbf{h}} \| |F_{\rm obs}(\mathbf{h})| - k|F_{\rm calc}(\mathbf{h})| \| \Big/ \sum_{\mathbf{h}} |F_{\rm obs}(\mathbf{h})| \quad (13)$$

where the summations extend over all observed reflections. A minimum of $E_x$ in (9) corresponds approximately to a minimum of the $R$ factor and *vice versa*. The $R$ factor cannot be used to define the effective energy term $E_x$ since the first derivatives of (13) with respect to the structure factors have a singularity at the origin.

*The program X-PLOR*

All calculations in this work were carried out with the program *X-PLOR*, a general-purpose macromolecular refinement program that uses crystallographic diffraction data or nuclear magnetic resonance interproton or other internuclear distance data in combination with energy minimization or molecular dynamics. *X-PLOR* was developed by ATB primarily at Harvard, working in collaboration with MK, and is partly based on the *CHARMM* program. *X-PLOR* has parameter and topology data structures in common with *CHARMM* (Brooks *et al*, 1983). Its command language and execution control is different from *CHARMM* and more focused on the needs of interactivity and ease of use. It helps the user in both interactive and batch operation. On-line help explains the options available during the run and on-line query of variables shows the current control status of the program. Structuring syntax, such as 'if ... then ... else' and 'do while ... end do' in the command language facilitates self-documentation of operations. References to previously prepared command files shorten set-up time of command sequences and help the user build up procedure libraries.

*X-PLOR* was designed to provide a comprehensive refinement package, user-friendly input, machine-portability, and highly efficient algorithms for the most CPU-time-consuming tasks on vector-parallel as well as conventional scalar machines. Table 1 lists the features of *X-PLOR*. The program is available on request from ATB.

*The system studied*

The system studied to provide a test of the methodology is crambin, a small protein composed of 46 amino acids, for which high-resolution X-ray diffraction data and a refined structure (determined by resolved anomalous phasing and RLSQ refinement with manual re-fitting) were available (Hendrickson & Teeter, 1981). The initial structure for the SA refinement was obtained from an NMR structure determination that used simulated data (Brünger, Campbell, Clore, Gronenborn, Karplus, Petsko & Teeter, 1987). The orientation and the position of the initial structure in the unit cell were obtained by Patterson searches. Rotation and translation function results of several NMR-derived structures were averaged to yield the correct orientation and position. The initial structure that is used in the following is the average structure of the NMR-derived structures (Brünger, Campbell, Clore, Gronenborn, Karplus, Petsko & Teeter, 1987). The diffraction data between 8·0 and 1·5 Å were used as in Hendrickson & Teeter (1981). All reflections were weighted equally, *i.e.* $W_h$ was set to 1·0 for all reflections. The complete SA-refinement protocol discussed in the next section required approximately 30 min of Cray-XMP time.

Table 1. *Features of X-PLOR*

| | |
|---|---|
| Crystallographic refinement | supports all space groups; phase difference restraints; direct summation method or FFT method with memory reduction; restrained, grouped and overall temperature factor optimization |
| NMR structure refinement | interproton distance restraints; choice of bi-harmonic, square-well, or soft-asymptote restraining potential; $R^{-6}$ or center averaging for unresolvable protons; analysis of observed/computed interproton distance |
| Empirical potential energy | bond length, bond angle, dihedral angle, chiral and planar torsion, electrostatic and van der Waals, crystal packing/symmetry interactions |
| Restraints and constraints | fixing of atoms, bond lengths or bond angles; harmonic restraints of atomic positions to another structure; dihedral angle restraints |
| Refinement methods | conjugate gradient minimization, simulated annealing with molecular dynamics |
| Analysis | analysis of stereochemistry; r.m.s. deviations between structures |
| Manipulation | manipulation of atomic properties, structure factors; hydrogen building |
| Command language | free-field; on-line help; structured loop and conditional statements; variable substitution; atom selection parser |
| Implementation | VAX/VMS; all Cray and Convex systems; Stellar GS1000 |

Table 2. *Crambin SA-refinement protocol*

| Stage | Description |
|---|---|
| 1 | Determination of weight $W_A$ |
| 2 | Minimization, 40 conjugate gradient steps, 8·0–1·5 Å, $C^\alpha$ restraints at 83·68 kJ mol$^{-1}$ Å$^{-2}$, $W_A$ = 50 208 kJ mol$^{-1}$, $B$ = 6·0 (10·0)* Å$^2$, $\Delta_f$ = 0·05 Å |
| 3 | Molecular dynamics, 1·5 ps, $T$ = 9000 K, 8·0–1·5 Å, time step = 0·25 fs, velocity scaling every 75 fs, $W_A$ = 50 208 kJ mol$^{-1}$, $B$ = 6·0 (10·0) Å$^2$, $\Delta_f$ = 0·3 Å |
| 4 | Molecular dynamics, 1·0 ps, $T$ = 300 K, 8·0–1·5 Å, time step = 1 fs, velocity scaling every 100 fs, $W_A$ = 50 208 kJ mol$^{-1}$, $B$ = 6·0 (10·0) Å$^2$, $\Delta_f$ = 0·3 Å |
| 5 | Minimization, 80 steps, 8·0–1·5 Å, $W_A$ = 50 208 kJ mol$^{-1}$, $B$ = 6·0 (10·0) Å$^2$, $\Delta_f$ = 0·0 Å |

* Temperature factors are specified as '$a(b)$' where '$a$' refers to $C^\alpha$, N, C atoms and '$b$' refers to other atoms.

## Results and discussion

Table 2 shows the protocol that worked quite successfully for crambin. The SA refinement proceeded in five stages. The first stage determined the weight $W_A$ in (9) that relates the effective energy $E_x$ to the empirical potential energy $E_i$. It was chosen such that the norm of the gradient of $E_x$ was comparable to the norm of the gradient of the empirical potential energy $E_i$ for the structure of crambin obtained after a 0·1 ps molecular dynamics simulation with $W_A$ set to zero, starting with the initial NMR-derived structure. The molecular dynamics simulation is preceded by several cycles of conjugate gradient minimization to relieve bad contacts. It is necessary to use a molecular dynamics structure rather than the initial X-ray or energy-minimized structure to compute the gradients. The initial structure could be strained and thus would artificially increase the gradient, and energy minimization alone with $W_A$ set to zero cannot be used since a perfectly energy-minimized structure

has a gradient of $E_i$ close to zero. A molecular dynamics structure has a non-zero gradient since the atoms are fluctuating around their equilibrium positions, thus making possible a comparison of the gradients of $E_i$ and $E_x$.

Experience has shown that this procedure is relatively insensitive to the choice of resolution and initial structure. The comparison of the gradients suggested that a value of the relative weight $W_A$ equal to about 50 208 kJ mol$^{-1}$ should be used in the case of crambin. $W_A$ was kept constant throughout the SA refinement. This choice of the weight factor resulted in good stereochemistry for the final structure without having to readjust the weight factor during the SA refinement. The same value of $W_A$ appears to be useful also for conventional RLSQ refinement. It eliminates the initial guess and subsequent adjustments of $W_A$ which are quite common in RLSQ refinement where the choice of $W_A$ is based on the current stereochemistry of the structure.

During the second stage in Table 2 conjugate gradient minimization was employed to relieve any initial strain or close nonbonded contracts in the structure that could pose a problem when starting the molecular dynamics calculation. During that stage, $C^\alpha$ backbone atoms are harmonically restrained to their initial positions. This avoids a large drift of the structure if any bad nonbonded contacts have to be relieved.

During the third stage, the heating stage, molecular dynamics was carried out at 9000 K for 1·5 ps; the initial velocities were assigned to a Maxwellian distribution at 9000 K. The velocities were re-scaled every 75 fs in order to keep the temperature at 9000 K. Owing to the high temperature, the time step of the molecular dynamics integration (0·25 fs) had to be chosen smaller than what can be used at physiological temperatures (typically 1 fs). In the fourth stage, the system was cooled to temperatures near 300 K by rescaling the velocities every 100 fs. The fifth stage consisted of conjugate gradient minimization which further optimizes the $R$ factor and stereochemistry of the system.

This annealing schedule was developed by trial and error. It differs from the one described by Kirkpatrick *et al.* (1983) in the way the temperature is controlled during annealing; *i.e.* for Kirkpatrick *et al.* the cooling rate is slowed down if the system is near a critical point whereas here the system is cooled at a constant rate. The major difference between SA refinement and RLSQ refinement consists of the heating stage and the cooling stage. If these stages were bypassed the protocol would be similar to the Jack & Levitt (1978) refinement.

It should be pointed out that the high temperature of the SA refinement and the motions during the SA refinement have no physical meaning since the total energy function is a hybrid of an empirical potential energy and an artificial effective energy describing the crystallographic residual. The temperature is merely a parameter that determines the height of local energy barriers that can be overcome during SA refinement.

Table 3. *Comparison of refinement protocols for crambin*

| Temperature (K) | Dynamics $R$ factor*† | Final $R$ factor* | $\Delta_{bonds}$ (Å) | $\Delta_{angles}$ (°) | ($C^\alpha$, C, N) difference‡ (Å) |
|---|---|---|---|---|---|
| 0 (initial) | - | 56·1 | 0·031 | 3·7 | 1·06 |
| 0 (RLSQ minimized)§ | - | 37·2 | 0·017 | 3·8 | 0·70 |
| 0 (manually refined)¶ | - | 25·5 | 0·013 | 2·5 | 0·05 |
| 300 | 35 | 33·7 | 0·017 | 3·0 | 0·57 |
| 1000 | 36 | 35·1 | 0·017 | 3·4 | 0·58 |
| 3000 | 37 | 31·0 | 0·015 | 3·0 | 0·54 |
| 4500 | 41 | 33·0 | 0·016 | 3·6 | 0·35 |
| 6000 | 43 | 30·7 | 0·017 | 3·3 | 0·29 |
| 7500 | 45 | 28·6 | 0·014 | 2·9 | 0·24 |
| 9000(*a*) | 48 | 27·2 | 0·013 | 2·8 | 0·23 |
| 9000(*b*) | 49 | 28·2 | 0·015 | 3·0 | 0·09 |
| 10 500 | 51 | 30·1 | 0·014 | 2·9 | 0·32 |

* The $R$ factor for crambin was computed with constant temperature factors ($B$ = 6 Å for $C^\alpha$, N, C atoms, $B$ = 10 Å for other atoms) and without solvent at 8·0–1·5 Å resolution.

† The dynamics $R$ factor refers to the average $R$ factor during the last 100 ps of the heating stage (stage 3 in Table 2).

‡ Atomic r.m.s. differences from the manually refined structures (Hendrickson & Teeter, 1981).

§ 'RLSQ minimized' refers to 40 cycles conjugate gradient minimization of $E_{total}$ at 4 Å resolution, followed by 40 cycles at 8·0–3·0 Å, 40 cycles at 8·0–2·3 Å, 40 cycles at 8·0–1·9 Å and 40 cycles at 8·0–1·5 Å resolution.

¶ For comparison, the manually refined strcture (Hendrickson & Teeter, 1981) was subjected to conjugate gradient minimization without solvent and with constant temperature factors.

Table 3 compares SA refinements of crambin at various temperatures, the initial structure, the manually refined structure (Hendrickson & Teeter, 1981), and a structure that was obtained by minimization without re-fitting (simply referred to as the 'RLSQ-minimized structure'). The refinement of the latter structure was carried out starting at low resolution (4·0 Å) and then extending the resolution in several stages to 1·5 Å resolution. This is the usual procedure for RLSQ refinement where a larger radius of convergence is achieved at lower resolution. It appears that the conjugate gradient minimization method (Powell, 1977) that was employed in this paper produced comparable results to *PROLSQ* which resulted in an $R$ factor of 38·1% at 2 Å resolution [*cf.* Table 1 of Brünger, Kuriyan & Karplus (1987)]. In contrast to the RLSQ-minimized structure and in contrast to the work of Brünger, Kuriyan & Karplus (1987), the SA refinements here were carried out only at high resolution. This was done in order to demonstrate the large radius of convergence of SA refinement even at high resolution.

The two structures at 9000 K differ by random number seeds for the initial velocities during stage 2 (Table 2). It appears that all SA-refined structures are closer to the manually refined structure than is the RLSQ-minimized structure. The RLSQ-minimized structure has a root-mean-square (r.m.s.) difference of 0·7 Å for backbone atoms whereas the

r.m.s. difference for the SA-refined structures ranges from 0·09 to 0·58 Å. This is also reflected in the final $R$ factor at 1·5 Å resolution, which is 37·2% for the RLSQ-minimized structure and ranges from 35·1 to 28·2% for the SA-refined structures. The $R$ factor of the manually refined structure is 25·5% at 1·5 Å resolution with constant temperature factors and without solvent. The stereochemistry (*i.e.* the deviations of bond lengths and bond angles from ideality, $\Delta_{\text{bonds}}$ and $\Delta_{\text{angles}}$) is good for all SA-refined structures.

It appears that the results in Table 3 are comparable to the SA refinement in Brünger, Kuriyan & Karplus (1987). In fact, the structures at 6000, 7500, 9000 and 10 500 K have a smaller r.m.s. difference from the manually refined structure than the structure of Brünger, Kuriyan & Karplus (1987); the latter had a r.m.s. difference for backbone atoms of 0·34 Å. The results in Table 3 suggest that the smaller radius of convergence at high resolution can be avoided by executing the SA refinement at high temperatures. This proves, at least in the case of the high-resolution structure of crambin, that SA refinement is less sensitive to the initial choice of the resolution than RLSQ refinement.

The SA refinement of crambin was carried out with constant temperature factors and without solvent. Despite this, some of the SA-refined structures are very close to the manually refined crambin structure that included individual atomic temperature factors and solvent. The closest structures [9000($a$), 9000($b$), 7500 K] have the lowest $R$ factors. This suggests that the effects of solvent and individual atomic temperature factors on the atomic coordinates of crambin are relatively small.

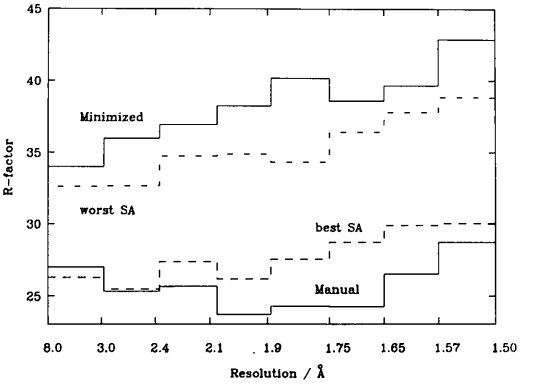


Fig. 1. $R$ factor as a function of resolution for the RLSQ-minimized, the manually refined (Hendrickson & Teeter, 1981), the worst, and the best SA-refined crambin structure. The $R$ factor is plotted in equal $1/r^3$ intervals where $r$ is the resolution in Å.

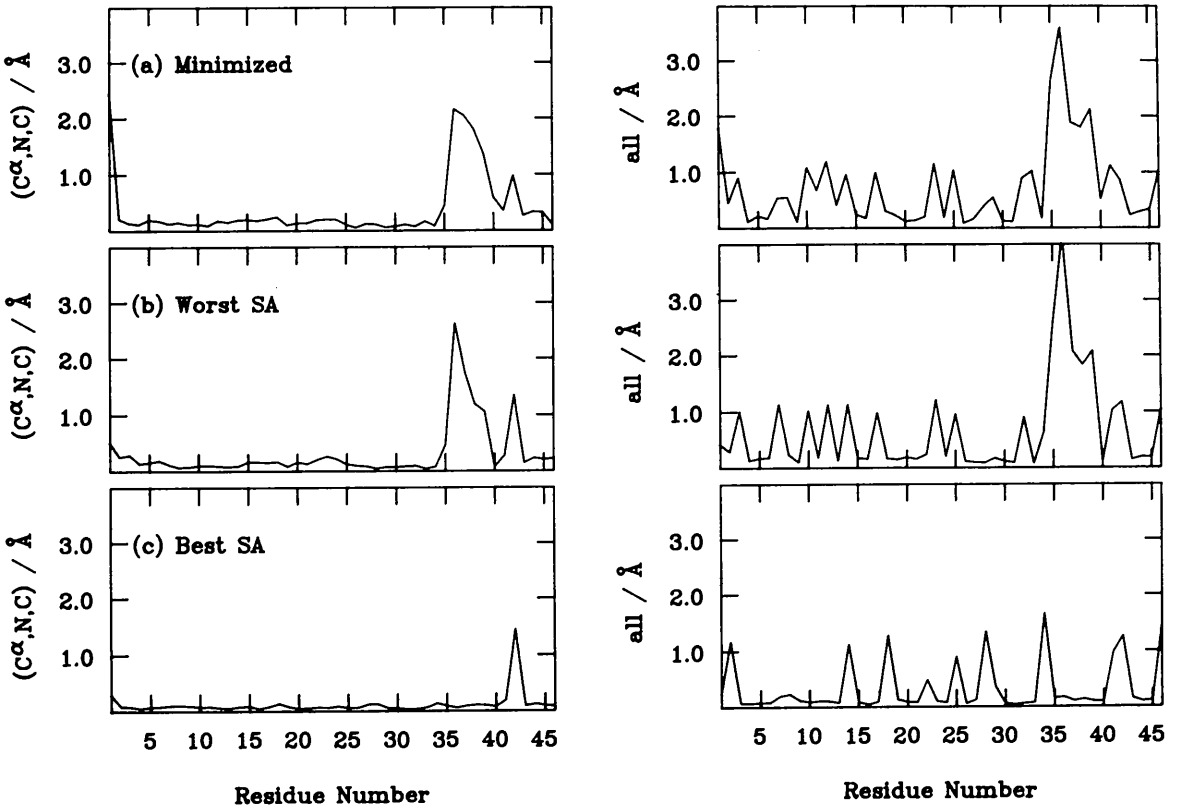


Fig. 2. R.m.s. differences from the manually refined crambin structure (Hendrickson & Teeter, 1981). The r.m.s. differences are plotted as function of residue number for backbone atoms and all atoms in the case of ($a$) the RLSQ-minimized, ($b$) worst SA-refined, and ($c$) the best SA-refined structure.

As shown in Table 3, the efficiency of SA refinement is dependent on the temperature at which the heating stage is carried out. The SA refinement of crambin is optimal at 9000 K. When the temperature is increased even further the results get worse. The increase in temperature is correlated with the value of the average $R$ factor during the heating stage. This $R$ factor is always going to be larger than the final $R$ factor after cooling since the atoms are undergoing large fluctuations around their equilibrium positions. However, in the limit of very high temperatures, the effective energy $E_x$ no longer influences the dynamics of the system. Clearly, this is not useful for SA refinement. From Table 3 it appears that the $R$ factor during the heating stage should be kept below 50%. Further investigations of this point are necessary for other systems to determine the optimum temperatures.

Fig. 1 shows the final $R$ factor as a function of resolution for the RLSQ-minimized, the worst SA-refined (at 1000 K), the best SA-refined [at 9000 K, labelled ($a$)], and the manually refined structure. The attributes 'best' and 'worst' are defined in terms of the final $R$ factor. Fig. 2 indicates r.m.s. differences for backbone atoms with respect to the manually refined structure, and Fig. 3 shows the deviations of bond lengths and bond angles from ideality as a function of residue number for these structures. Both the RLSQ-minimized and the worst SA-refined structure fit the high-resolution shells of the data less well than the manually refined and the best SA-refined structure.

The r.m.s. differences for residue positions of the initial NMR-derived structure are as large as 3·5 Å, with particularly large differences for residues 35 to 40 [see Fig. 1 of Brünger, Kuriyan & Karplus (1987)]. The r.m.s. differences for the RLSQ-minimized structure (Fig. 3$a$) and the worst SA-refined structure (Fig. 3$b$) are still quite large for residues 35 to 40. The r.m.s. differences are reduced in the case of the best SA-refined structure (Fig. 3$c$). There appears to be

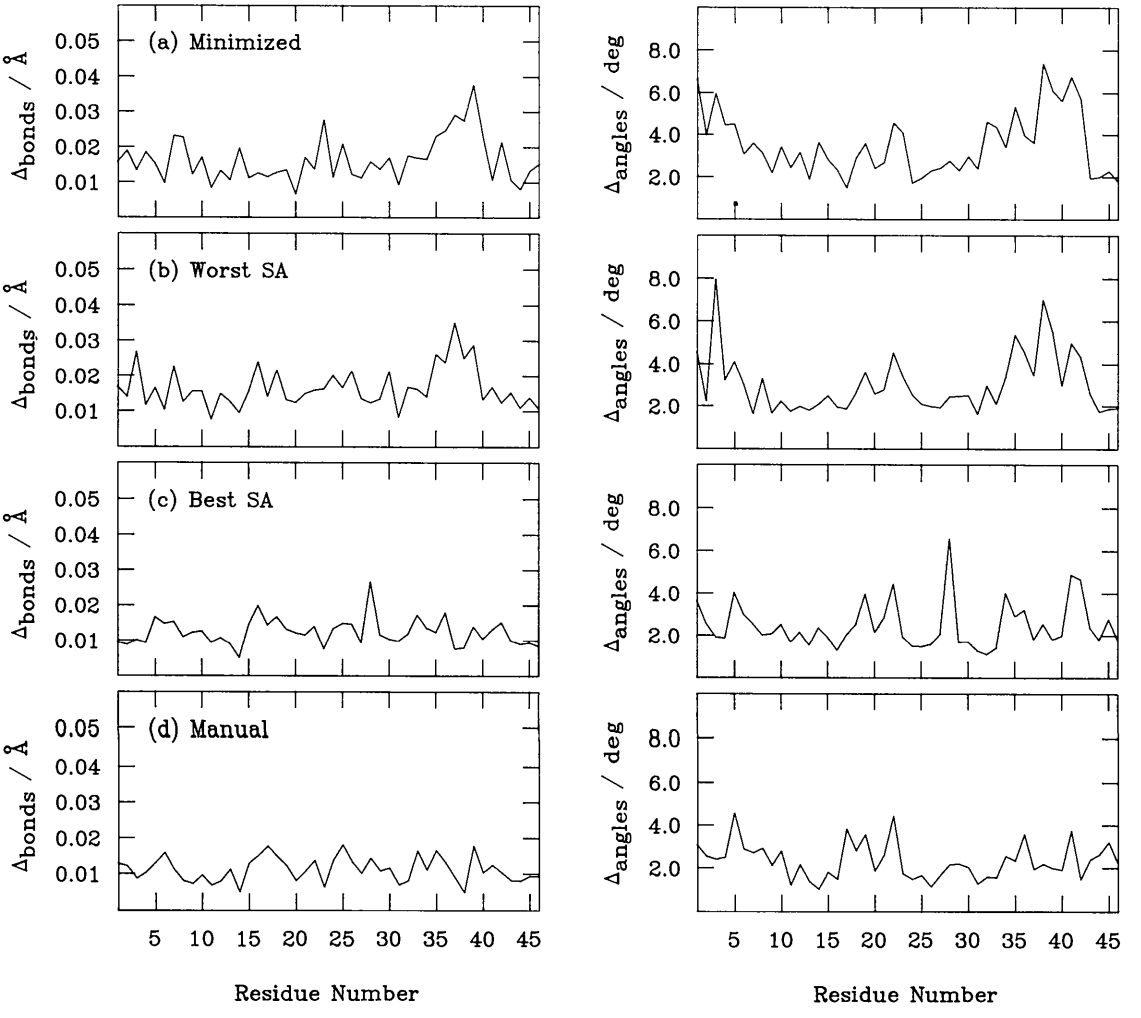


Fig. 3. R.m.s. deviations of bond lengths and bond angles from ideality as a function of residue number for ($a$) the RLSQ-minimized, ($b$) the worst SA-refined, ($c$) the best SA-refined and ($d$) the manually refined (Hendrickson & Teeter, 1981) structure of crambin.

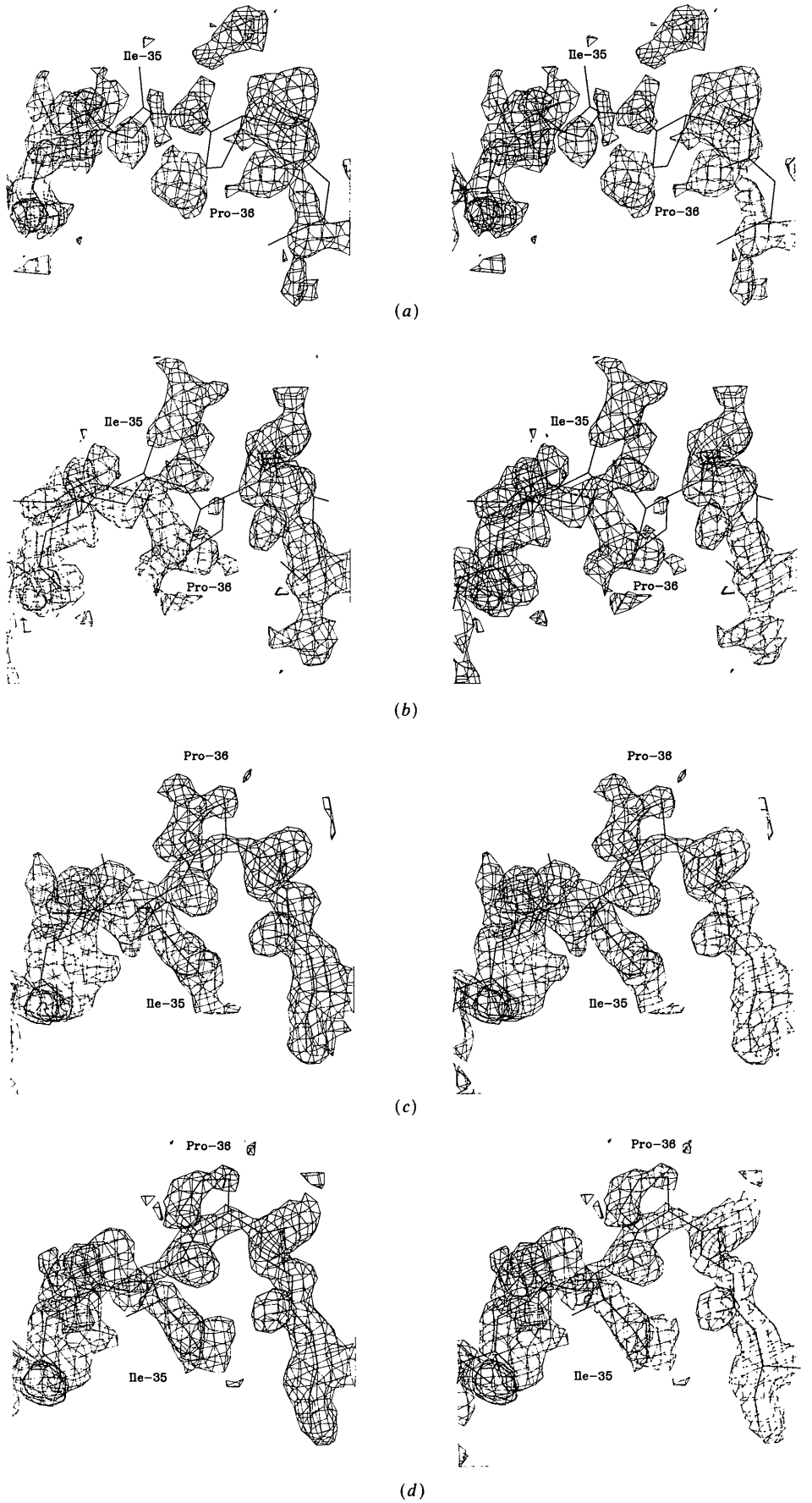


Fig. 4. Environment of Pro-36 in crambin. In (*a*) the RLSQ-minimized, in (*b*) the worst SA-refined, in (*c*) the best SA-refined, and in (*d*) the manually refined structure (Hendrickson & Teeter, 1981) are shown. The atomic models are shown as thick lines. Electron density maps are superimposed as thin lines. The maps were calculated to a resolution of 1·5 Å with coefficients $(F_{obs} - F_{calc})$ and phases from the calculated structure factors of the corresponding atomic models. Residues Cys-32 to Ala-38 have been omitted for the calculation of $F_{calc}$. Maps were contoured at $1{\cdot}5\sigma$ where $\sigma$ is the r.m.s. electron density in the unit cell. Display of the atomic models and the electron density maps was carried out by using *FRODO* (Jones, 1978).

some correlation between the magnitude of the r.m.s. differences and the r.m.s. deviations from ideality of either bond lengths or bond angles or both; for example, in the case of the RLSQ-minimized structure the large deviations of bond angles from ideality around residues 35–40 correspond to large r.m.s. differences.

Table 3 and the results of Brünger, Kuriyan & Karplus (1987) show that SA refinement always produces better results than RLSQ refinement without manual re-fitting, regardless of the choice of protocol. However, some SA-refinement protocols appear to work better than others. From experience with SA refinement of aspartate aminotransferase, a protein larger than crambin (Brünger, 1988*b*), one has to conclude that the choice of the 'best' protocol actually depends on the particular system studied. Therefore, a particular SA refinement for a new system may not optimally converge in all parts of the structure.

To investigate what the crystallographer would have gained in the worst case using SA refinement, Fig. 4 shows the environment of residue Pro-36 for the RLSQ-minimized, the worst SA-refined, the best SA-refined and the manually refined structure. Superimposed on the structures are electron density maps using $(F_{obs} - F_{calc})$ amplitudes and $F_{calc}$ phases corresponding to the particular structures, where residues Cys-32 to Ala-38 have been omitted for the calculation of $F_{calc}$. The SA refinement has not converged to the manually refined structure in the case of the RLSQ-minimized (Fig. 4*a*) and the worst SA-refined (Fig. 4*b*) structure. Some of the Pro-36 atoms penetrate the correct Ile-35 electron density whereas the $C^{\gamma 2}$ atom of Ile-35 penetrates the Pro-36 electron density. Despite the fact that the worst SA-refined structure has not converged to the manually refined structure, the quality of the electron density map is superior to the map of the RLSQ-minimized structure. The map in (Fig. 4*b*) outlines the correct positions of the atoms (*cf.* Fig. 4*d*) and is continuous except for a break between Pro-36 and Gly-37. The map of the RLSQ-minimized structure (Fig. 4*a*) shows less clearly defined features for residues 36–40 and manual re-fitting in this case would be difficult. The conventional RLSQ refinement carried out by Brünger, Campbell, Clore, Gronenborn, Karplus, Petsko & Teeter (1987) starting with the same initial structure required several re-fitting sessions and many cycles of RLSQ refinement to correct this region. Since the map in Fig. 4(*b*) of the worst SA-refined structure is relatively easy to interpret, even poorly converged SA refinement has simplified the manual re-fitting of the crambin structure.

In the case of the best SA-refined structure, all atoms shown in Fig. 4(*c*) have converged to the manually refined structure except for the sidechain of Ile-37 which is rotated 180° around the $\chi_1$ bond. The unaccounted density patches in Figs. 4(*c*) and (*d*) indicate possible water molecules which are not included in this work. However, the inclusion of water molecules into SA refinement is relatively straightforward and will be discussed elsewhere.

## Concluding remarks

SA refinement by simulated annealing has been shown to be superior to RLSQ refinement. This superiority is reflected in many cases by a larger radius of convergence, *i.e.* SA refinement can move atoms that are displaced by more than 1·0 Å. On the other hand SA refinement does not always converge in all regions of the structure. The method is inherently stochastic in nature by the introduction of initial random velocities assigned to a Maxwellian distribution. There is no guarantee that a particular transition from one local minimum into another will happen. On average, however, the results are always improved over conventional RLSQ refinement. It appears that even in cases where SA refinement does not converge in a particular region it produces electron density maps in that region of the structure that allow better interpretation of atomic positions than maps produced by RLSQ refinement.

We have also discussed the robustness of the method in terms of the resolution range and the temperature used for the SA refinement. Particular choices of resolution range and temperature may result in an optimal refinement that converges to the manually refined structure without any re-fitting. However, even in cases where the SA refinement does not fully converge, the SA-refined structure has improved with respect to the $R$ factor, stereochemistry, and the quality of the electron density maps.

The SA refinement of Brünger, Kuriyan & Karplus (1987) employed a different set of parameters than used in this work. In particular, all hydrogens were simulated by Brünger, Kuriyan & Karplus (1987) whereas here aliphatic hydrogens were neglected and implicitly taken into account by 'extended' carbons. The success of both SA refinements indicates that SA refinement is insensitive to the particular choice of empirical potential energy parameters. The effective energy term $E_x$ describing the diffraction information determines the outcome of the SA refinement to a large extent. In fact, the r.m.s. differences of the SA-refined structures compared with the manually refined structure are much smaller than has been reported for any free (*i.e.* without $E_x$) molecular dynamics simulation of macromolecules including solvent and crystal contacts (van Gunsteren, Berendsen, Hermans, Hol & Postma, 1983). This suggests further uses of the effective energy term $E_x$ that go beyond crystallographic refinement. In particular, more precise simulations of anisotropic temperature factors and alternative conformations can now be

attempted that do not suffer from the inherent drift of free molecular dynamics calculations (Kuriyan & Brünger, work in progress).

We thank M. M. Teeter for providing the 1·5 Å resolution data of crambin and the protein coordinates of the refined structure. We thank J. Kuriyan and W. Hendrickson for useful discussions. ATB acknowledges support by the Pittsburgh Supercomputer Center of the National Science Foundation (Grant PSCA 226).

### References

AGARWAL, R. C. (1978). *Acta Cryst.* A**43**, 791-809.
BOUNDS, D. G. (1987). *Nature (London)*, **329**, 215-219.
BROOKS, B. R., BRUCCOLERI, R. E., OLAFSON, B. D., STATES, D. J., SWAMINATHAN, S. & KARPLUS, M. (1983). *J. Comput. Chem.* **4**, 187-217.
BRÜNGER, A. T. (1988*a*). In *Crystallographic Computing* 4: *Techniques and New Technologies*, edited by N. W. ISAACS & M. R. TAYLOR. Oxford Univ. Press.
BRÜNGER, A. T. (1988*b*). *J. Mol. Biol.* In the press.
BRÜNGER, A. T. (1989). *Acta Cryst.* A**45**, 42-50.
BRÜNGER, A. T., CAMPBELL, R. L., CLORE, G. M., GRONENBORN, A. M., KARPLUS, M., PETSKO, G. A. & TEETER, M. M. (1987). *Science*, **235**, 1049-1053.
BRÜNGER, A. T., KURIYAN, K. & KARPLUS, M. (1987). *Science*, **235**, 458-460.
FLETCHER, R. & REEVES, C. M. (1964). *Comput. J.* **7**, 149-154.
GUNSTEREN, W. F. VAN, BERENDSEN, H. J. C., HERMANS, J., HOL, W. G. J. & POSTMA, J. P. M. (1983). *Proc. Natl Acad. Sci. USA*, **80**, 4315-4319.
HENDRICKSON, W. A. (1985). *Methods Enzymol.* **115**, 252-270.
HENDRICKSON, W. A. & TEETER, M. A. (1981). *Nature (London)*, **290**, 107-112.
HERMANS, J., BERENDSEN, H. J. C., VAN GUNSTEREN, W. F. & POSTMA, J. P. M. (1984). *J. Mol. Biol.* **23**, 1513-1518.
*International Tables for X-ray Crystallography* (1974). Vol. IV. Birmingham: Kynoch Press. (Present distributor Kluwer Academic Publishers, Dordrecht.)
JACK, A. & LEVITT, M. (1978). *Acta Cryst.* A**34**, 931-935.
JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268-272.
KARPLUS, M. & MCCAMMON, J. A. (1983). *Annu. Rev. Biochem.* **52**, 263-300.
KIRKPATRICK, S., GELATT, C. D. JR & VECCHI, M. P. (1983). *Science*, **220**, 671-680.
KONNERT, J. H. & HENDRICKSON, W. A. (1980). *Acta Cryst.* A**36**, 344-349.
KURIYAN, J., PETSKO, G. A., LEVY, R. M. & KARPLUS, M. (1986). *J. Mol. Biol.* **190**, 227-254.
LEVITT, M. (1983). *J. Mol. Biol.* **168**, 595-620.
LIFSON, S. & STERN, P. (1982). *J. Chem. Phys.* **77**, 4542-4550.
METROPOLIS, N., ROSENBLUTH, M., ROSENBLUTH, A., TELLER, A. & TELLER, E. (1953). *J. Chem. Phys.* **21**, 1087-1092.
MOSS, D. S. & MORFFEW, A. J. (1982). *Comput. Chem.* **6**, 1-3.
NÉMETHY, G., POTTIE, M. S. & SCHERAGA, H. A. (1983). *J. Phys. Chem.* **87**, 1883-1887.
NILSSON, L. & KARPLUS, M. (1986). *J. Comput. Chem.* **7**, 591-616.
POWELL, M. J. D. (1977). *Math. Programming*, **12**, 241-254.
SOUKOP, J. (1981). *Proc. IEEE*, **69**, 1281-1304.
SUSSMAN, J. L., HOLBROOK, S. R., CHURCH, G. M. & KIM, S. H. (1977). *Acta Cryst.* A**33**, 800-804.
TEN EYCK, L. F. (1973). *Acta Cryst.* A**29**, 183-191.
TRONRUD, D. E., TEN EYCK, L. F. & MATTHEWS, B. W. (1987). *Acta Cryst.* A**43**, 489-500.
VERLET, L. (1967). *Phys. Rev.* **159**, 98-105.
WEINER, S. J., KOLLMAN, P. A., NGUYEN, D. T. & CASE, D. A. (1986). *J. Comput. Chem.* **7**, 230-252.

*Acta Cryst.* (1989). A**45**, 61-63

# An Alternative Convention for Solving the Ambiguity Problem of (3+1) Superspace Group Symbols

BY V. PETŘÍČEK

*Institute of Physics, Czechokslovak Academy of Sciences, Na Slovance* 2, 180 40 *Praha* 8, *Czechoslovakia*

(*Received* 20 *December* 1987; *accepted* 1 *August* 1988)

### Abstract

An alternative convention is proposed for solving the ambiguity problem of (3+1) superspace group symbols described by Yamamoto, Janssen, Janner & de Wolff [*Acta Cryst.* (1985), A**41**, 528-530] based on the requirement that the condition for equivalence of modulation vectors to be independent on a selection of basis vectors is satisfied.

The ambiguity of (3+1) superspace group symbols was discussed by Yamamoto, Janssen, Janner & de Wolff (1985) (hereafter referred to as I) and their solution consists of making a specific choice of symbols for basic space groups summarized in their Table 2.

In order to make clearer our alternative solution we will present below a simple derivation of transformation properties of a supersymmetry operator in $(3+d)$ superspace for the case of replacing modulation vectors.

A translational periodicity in $(3+d)$ superspace is characterized by a lattice $\Lambda$ spanned by $\mathbf{b}_1, \ldots, \mathbf{b}_{3+d}$ (de Wolff, 1974; Janner, Janssen & de Wolff, 1983):

$$\begin{aligned} \mathbf{b}_i &= \mathbf{a}_i - \sum_{j=1}^{d} \mathbf{e}_j \sigma_{ji} \quad (i=1,2,3) \\ \mathbf{b}_{i+3} &= \mathbf{e}_i \qquad\qquad (i=1,\ldots,d) \end{aligned} \tag{1}$$

0108-7673/89/010061-03$03.00 © 1989 International Union of Crystallography