SOBRY, R., FONTAINE, F. & LEDENT, J. (1994). *J. Appl. Cryst.* **27**, 482–496.
SOBRY, R., LEDENT, J. & FONTAINE, F. (1991). *J. Appl. Cryst.* **24**, 516–525.
WILSON, A. J. C. (1949). *X-ray Optics*. London: Methuen.
WILSON, A. J. C. (1969). *J. Appl. Cryst.* **2**, 181–183.
WILSON, A. J. C. (1970). *Elements of X-ray Crystallography*, ch. 10. Reading, MA: Addison-Wesley.
WU, H. & SCHMIDT, P. W. (1971). *J. Appl. Cryst.* **4**, 224–231.
WU, H. & SCHMIDT, P. W. (1974). *J. Appl. Cryst.* **7**, 131–146.

Acta Cryst. (1995). A**51**, 69–80

Restrained Real-Space Macromolecular Atomic Refinement using a New Resolution-Dependent Electron-Density Function

BY MICHAEL S. CHAPMAN

Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, *USA, and Department of Chemistry* and Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida* 32306, *USA*

(*Received* 6 *April* 1994; *accepted* 21 *June* 1994)

Abstract

A new atomic electron-density function is derived by Fourier transformation of resolution-truncated atomic scattering factors. It forms the basis of a new real-space refinement method, *RSREF*, that is a substantial improvement on prior implementations that did not formally consider the absence of high-resolution terms in a typical macromolecular electron-density map. Real-space refinement is further improved through the simultaneous refinement of stereochemical restraints analogous to reciprocal-space methods. Parallel refinements of a viral capsid structure show that real-space refinement produces models that are at least as good as those refined in reciprocal space, by either restrained or molecular-dynamics methods, and that refinement cycles are ~50 times faster. Real-space refinement will not replace reciprocal-space methods for proteins, where, without the high noncrystallographic symmetry of viruses, experimental phases and electron-density maps are not of the same high quality. However, applied to local regions, it can be used to speed up and improve the quality of interactive model building before a full refinement is started.

1. Notation

a	Atomic radius
A, B	Real and imaginary parts of a structure factor
B	Temperature factor
d^*	Reciprocal-space distance from origin
$d^*_{\min}, d^*_{\max}$	Reciprocal-space resolution limits
$F^{\rm obs}, F^{\rm calc}$	Observed and calculated structure factors
f	Form factor
g	Scattering factor with thermal motion, $g = f(d^*)\exp(-Bd^{*2}/4)$
G	Fourier transform of a solid sphere
h^*	Reciprocal-space distance
h	Reflection index
H	Total number of reflections
k	Scaling constant (reciprocal space) or threshold (real space)
M	Number of atoms
$\mathcal{R}$	Overall residual minimized in least squares
$\mathcal{R}_{\rm geom}$	Residual difference between model and ideal stereochemistry
$\mathcal{R}_\rho$	Residual difference between observed and calculated electron density
$\mathcal{R}_{\rm X\text{-}ray}$	Residual difference between observed and calculated structure factors
$R^{\rm conv}$	Conventional reciprocal-space R factor
$R^{\mathcal{ED}}$	Real-space R factor
$R_T^{\rm free}$	Reciprocal-space free R factor (Brünger, 1992)
r	Radial distance from the center of an atom
$r^{\rm max}_{\rm calc}$	Maximum r for calculation of electron density
$r^{\rm max}_{\rm ref}$	Maximum r for calculation of derivatives
ρ_m	Electron density for the mth atom
ρ_c	Electron density calculated from all atoms of a model
ρ_o	Observed electron density
S	Scale constant to bring ρ_o to an absolute scale
T	Generic Fourier transform
w	Figure of merit
Z	Number of electrons in an atom

2. Introduction

Although real-space methods of refinement were applied successfully to some of the first protein structures (Diamond, 1974; Deisenhofer & Steigemann, 1975;

* Current address.

Watenpaugh, Sieker, Herriott & Jensen, 1973), they have now been supplanted by reciprocal-space methods (Hendrickson, 1985; Brünger, Kuriyan & Karplus, 1987). The main advantage of reciprocal-space methods is that the minimization of the difference between observed and calculated structure factors, $\sum_h(|F_h^{obs}| - |F_h^{calc}|)^2$, is independent of poorly determined experimental phases. In real-space refinement, a difference between observed and calculated electron densities, $\int(\rho_o - \rho_c)^2 dv$, is minimized. ρ_o is calculated using the experimental phases (isomorphous replacement, symmetry averaging *etc.*) or from a preliminary atomic model, in which case refinement may be biased towards the preliminary model.

There are several niches for which real-space refinement is well suited. Prior to reciprocal-space refinement and between rounds of the refinement, much time is spent manually optimizing the fit of the model to the electron density to ensure that it is within the convergence radius of automatic refinement. Real-space refinement has previously been used within interactive molecular modeling programs (Jones, 1978; Jones, Zou, Cowan & Kjeldgaard, 1991), to hasten modeling by reducing the interactive input required. The methodological improvements described here may extend greatly such applications of real-space refinement.

The convergence radius of reciprocal-space refinement is sometimes increased through the use of molecular dynamics (Brünger *et al.*, 1987) or through relaxation of the stereochemistry of restrained refinement (Hendrickson, 1985). Misuse of such methods has increased the possibility of errors in structure determinations (reviewed by Brändén & Jones, 1990). Although the R_T^{free} factor of Brünger (1992) should help to detect such errors, some might be avoided by starting refinement in real space before moving to reciprocal space. Real-space refinement has a large convergence radius (Diamond, 1985) even with good stereochemical restraints (see below) and it also incorporates more observational restraints through the implicit use of phases in addition to structure-factor magnitudes. Real-space refinement is also the method of choice for bootstrapping the phases of partial models where it has been shown that under-restrained reciprocal-space refinement can bias supposedly bias-free omit maps (Shreuder, Curmi, Cascio & Eisenberg, 1990; Hodel, Kim & Brünger, 1992).

The phases of virus crystals, refined through the application of 20- to 60-fold noncrystallographic symmetry, are of unusually high quality. Arnold & Rossmann (1988) showed that reciprocal-space refinement was improved through the addition of explicit phase restraints. It is likely that the implicit use of phases in real-space refinement will also be beneficial. Owing to their large size, the refinement of virus structures is a daunting task. There are often more than 10^6 reflections and each calculated structure factor depends on all the atoms of each of the noncrystallographically symmetry related capsid proteins. Thus, it has become common practice to refine virus structures against subsets of the data (Arnold & Rossmann, 1988). By contrast, in real-space refinement, all of the data can be implicitly used in a symmetry-averaged map, but only one of the symmetry equivalents needs to be refined. A 50-fold increase in speed has been realized when comparing real-space refinement on a modest workstation with vectorized reciprocal-space refinement on a Cyber 205 computer.

Real-space refinement has not achieved its full potential for two reasons. Firstly, the function previously used to model the electron density of each atom is appropriate only at very high resolution. Secondly, good stereochemistry has been imposed through constraints, which, in reciprocal-space refinement, has a smaller convergence radius than the addition of stereochemical 'observational' restraints (Waser, 1963; Hendrickson, 1985).

Diamond (1971) modeled the electron density at a distance r from the center of an atom as a spherical Gaussian:

$$\rho_c(r) = (Z/a^3)\exp(-\pi r^2/a^2), \qquad (1)$$

where Z is the number of electrons in the atom and a is the atomic radius, which is related to the atomic temperature factor by

$$a = (B/4\pi)^{1/2} \quad \text{if} \quad d^* = 2\sin\theta/\lambda. \qquad (2)$$

Low-resolution cases, typical of macromolecular crystallography, were modeled using inflated atomic radii/B factors, either by uniformly increasing the radii of all atoms (Deisenhofer & Steigemann, 1975), or by refining them independently (Fermi, 1975). This presents two problems. Firstly, refined temperature factors no longer correspond, in a simple way, to the actual atomic vibration. More importantly, no B factor can model a sharp drop in the scattering due to a resolution limit imposed during data collection or processing. Algorithms for pseudo-real-space refinement (Tronrud & Ten Eyck, 1992), which calculate the derivatives from a resolution-truncated difference map, avoid this and other problems described below. However, the Fourier transformations of pseudo-real-space methods are calculated using the entire contents of the unit cell and are not well suited to fast refinement of local regions of the unit cell. By contrast, the approach here is to derive a resolution-dependent electron-density function that could be used for rapid refinement of a small part of a unit cell. The local region might contain a few residues of a protein (during interactive modeling) or one of several equivalent proteins related by noncrystallographic symmetry. It is this second application that is demonstrated here using a virus structure refinement as an example.

Previous implementations of real-space refinement have used constraints to impose standard stereochemistry. Diamond (1971) reduced the average number of positional parameters per atom to 0.75 by allowing

only torsion angles to vary. Real-space refinement was alternated with energy refinement (Levitt, 1974). Jones & Liljas (1984) searched for an optimal fit of rigid-body fragments of amino acids, alternated with geometrical regularization of bonded geometry (Hermans & McQueen, 1974). In the implementation described here, *TNT*'s program *SHIFT* (Tronrud, Ten Eyck & Matthews, 1987) is used to calculate shift vectors from derivatives calculated by *TNT*'s stereochemical module, *GEOMETRY*, and a real-space module described below. In this protocol, the fit to the map and the stereochemistry are simultaneously refined in a manner analogous to current reciprocal-space refinement methods (Tronrud *et al.*, 1987; Hendrickson, 1985). The stereochemical restraints include both bonded and nonbonded interactions.

3. Theory

3.1. *Electron-density calculation*

The atomic scattering factor $[f(d^*)]$ is, by definition, the Fourier transform of an isolated atom. The electron density of an atom can therefore be calculated through an inverse three-dimensional transform of a scattering factor. It is assumed that the electron density of a molecule is the sum of the electron densities of the constituent atoms. It is also assumed that the atomic scattering factors, electron-density distribution and thermal vibration are spherically symmetric (isotropic). Let the scattering factor, $g(d^*)$, of a vibrating atom be approximated by a set of n finitely thin resolution elements, Δd^*, of uniform scattering power, $g(\bar{d}_i^*)$:

$$g(d^*) = \begin{cases} g(\bar{d}_i^*) & \text{if} \quad d_{i-1}^* \le d^* < d_i^* \\ 0 & \text{if} \quad d^* < d_{\min}^* \\ 0 & \text{if} \quad d^* > d_{\max}^* \end{cases}. \tag{3}$$

Note that, beyond the resolution limits ($d_{\max}^*$, $d_{\min}^*$), $g(d^*)$ is zero. Through this resolution truncation of the scattering factor, it is possible to accurately represent the electron density of a map calculated using data between fixed resolution limits. Equation (3) approximates the scattering power as a set of discrete spherical shells each with weight $g(\bar{d}_i^*)$. In a compromise of precision for computational efficiency, $g(\bar{d}_i^*)$ is sampled at a single point rather than integrated between $d^*{}_1$ and d_2^*. From integration in the limiting case, $\Delta d^* \rightarrow 0$, it can be shown that g is best sampled at the root mean square:

$$\bar{d}_i^* = \left[\left(d_i^{*2} + d_{i-1}^{*2}\right)/2\right]^{1/2}. \tag{4}$$

Although not required, it is convenient if the d_i^* are evenly spaced:

$$d_i^* = d_{\min}^* + (i-1)\Delta d^*; \quad n = \left(d_{\max}^* - d_{\min}^*\right)/\Delta d^*; \quad \text{integer } n. \tag{5}$$

The Fourier transform of a solid sphere in reciprocal space, of radius h^* and reciprocal-space density w, can be calculated analytically (*cf.* Rossmann & Blow, 1962):

$$\begin{aligned} T(r) &= w\left(\tfrac{4}{3}\pi h^{*3}\left\{3[\sin\left(2\pi h^* r\right) \right.\right. \\ &\quad \left.\left. - 2\pi h^* r\cos\left(2\pi h^* r\right)]\left(2\pi h^* r\right)^{-3}\right\}\right) \\ &= 4\pi w G\left(r, h^*\right), \end{aligned} \tag{6}$$

where

$$\begin{aligned} G\left(r, h^*\right) &= h^{*3}\left\{[\sin\left(2\pi h^* r\right) - 2\pi h^* r\cos\left(2\pi h^* r\right)]\right. \\ &\quad \left. \times \left(2\pi h^* r\right)^{-3}\right\}. \end{aligned} \tag{7}$$

As the transform of a sum is the sum of the transforms, the transform of a spherical shell of radii $h_2^* > h_1^*$ is

$$T(r) = 4\pi w\left[G\left(r, h_2^*\right) - G\left(r, h_1^*\right)\right]. \tag{8}$$

Similarly, the transform of the sum of n spherical shells is the sum of the individual transforms:

$$T(r) = 4\pi\sum_{i=1}^{n} w_i\left[G\left(r, h_i^*\right) - G\left(r, h_{i-1}^*\right)\right]. \tag{9}$$

The electron density $\rho(r)$ at a distance r from the atom's center can be calculated by a similar Fourier transformation of (3):

$$\rho(r) = 4\pi O\sum_{i=1}^{n} g\left(\bar{d}_i^*\right)\left[G\left(r, d_i^*\right) - G\left(r, d_{i-1}^*\right)\right], \tag{10}$$

where O is the fractional atomic occupancy. Equation (10) is the atomic electron-density function for an individual atom, which, to distinguish it from other functions, is termed the FT(f) function (Fourier transform of a form factor).

There are several equivalent forms and derivations of (10). The three-dimensional Fourier transform of an isotropic resolution-truncated scattering factor can be reduced to a one-dimensional integral through the Jacobian transformation:

$$\rho(r) = 4\pi O\int_{d_{\min}^*}^{d_{\max}^*} g\left(d^*\right) d^{*2}\left[\sin\left(2\pi r d^*\right)/2\pi r d^*\right] \mathrm{d}d^*. \tag{11}$$

The integral is evaluated in pieces, with the assumption that $g(d^*)$ is approximately constant within each integral:

$$\rho(r) = 4\pi O\sum_{i=1}^{n} g\left(\bar{d}^*\right)\int_{d_{i-1}^*}^{d_i^*} d^{*2}\left[\sin\left(2\pi r d^*\right)/2\pi r d^*\right] \mathrm{d}d^* \tag{12}$$

$$\begin{aligned} \rho(r) &= 4\pi O\sum_{i=1}^{n} g\left(\bar{d}^*\right)[\{\sin\left(2\pi r d^*\right) \\ &\quad - 2\pi r d^*\sin\left(2\pi r d^*\right)\}/\left(2\pi r d^*\right)]_{d_{i-1}^*}^{d_i^*} \\ &= 4\pi O\sum_{i=1}^{n} g\left(\bar{d}^*\right)[\{(\pi/2)/\left(2\pi r d^*\right)^3\}^{1/2} \\ &\quad \times J_{3/2}2\pi r d^*]_{d_{i-1}^*}^{d_i^*}. \end{aligned} \tag{13}$$

Note that the definite integral contains either the Fourier transform of a sphere [*cf.* (10) and (7)] or a half-order Bessel function, $J_{3/2}$.

The electron density at any point in an M-atom structure is derived from

$$\rho(xyz) = \sum_{m=1}^{M} \rho(r_m) = 4\pi \sum_{m=1}^{M} O_M \sum_{i=1}^{n} g_m\left(\bar{d}_i^*\right) G[(r_m, d_i^*) - G(r_m, d_{i-1}^*)]. \quad (14)$$

3.2. *Derivatives*

The derivatives of (14) with respect to refinable atomic parameters are required for least-squares optimization. Methods with faster convergence require the second derivatives (Tronrud, 1992). All of these can be calculated analytically. First, the derivatives of the calculated density are derived for each of the atomic parameters and then it is shown how these can be used to calculate the derivatives of the least-squares residual. Let P_q be one of the atomic parameters of atom q; occupancy O_q, position (x_q, y_q, z_q) or temperature factor B_q. Note that the derivative of the calculated electron density, with respect to an atomic parameter, depends only on that parameter, *i.e*:

$$\partial\rho_{xyz}/\partial P_q = \partial\rho(r_q)/\partial P_q. \quad (15)$$

From (14),

$$\frac{\partial\rho}{\partial O_m} = 4\pi \sum_{i=1}^{n} g_m\left(\bar{d}_i^*\right)(G_i - G_{i-1}); \quad \frac{\partial^2\rho}{\partial O_m^2} = 0; \quad (16)$$

$$\frac{\partial\rho}{\partial B_m} = 4\pi O_m \sum_{i=1}^{n} \frac{\partial g\left(\bar{d}_i^*\right)}{\partial B_m}(G_i - G_{i-1}). \quad (17)$$

From the definition of g,

$$g = f\left(d^*\right)\exp\left(-B_m d^{*2}/4\right); \quad (18)$$

$$\frac{\partial g}{\partial B_m} = \left(-d^{*2}/4\right) f\left(d^*\right)\exp\left(-B_m d^{*2}/4\right). \quad (19)$$

Thus,

$$\frac{\partial\rho}{\partial B_m} = 4\pi O_m \sum_{i=1}^{n}\left(-d^{*2}/4\right) g_m\left(\bar{d}_i^*\right)(G_i - G_{i-1}) \quad (20)$$

and, similarly,

$$\frac{\partial^2\rho}{\partial B_m^2} = 4\pi O_m \sum_{i=1}^{n}\left(d^{*4}/16\right) g_m\left(\bar{d}_i^*\right)(G_i - G_{i-1}). \quad (21)$$

As the electron-density function is spherically symmetric, it is only the radial components of the positional derivatives that are nonzero. Let

$$\varphi = 2\pi r d^*. \quad (22)$$

Then,

$$G = \left(d^{*3}/\varphi^3\right)(\sin\varphi - \varphi\cos\varphi); \quad (23)$$

$$\partial G/\partial r = (\partial G/\partial\varphi)(\partial\varphi/\partial r); \quad (24)$$

$$\partial G/\partial r = \left[\left(d^{*3}/\varphi^3\right)(\cos\varphi + \varphi\sin\varphi - \cos\varphi) - \left(3d^{*3}/\varphi^4\right)(\sin\varphi - \varphi\cos\varphi)\right]; \quad (25)$$

$$\partial G/\partial r = (1/r)\left\{\left[\left(d^{*3}\sin\varphi\right)/\varphi\right] - 3G\right\} \equiv G' \quad (26)$$

$$\frac{\partial^2 G}{\partial r^2} = \left(-1/r^2\right)\left\{\left[\left(d^{*3}\sin\varphi\right)/\varphi\right] - 3G\right\} + \left(d^{*3}/r\right)2\pi d^*\left[(\cos\varphi/\varphi) - \left(\sin\varphi/\varphi^2\right)\right] - (3/r)G'; \quad (27)$$

$$\frac{\partial^2 G}{\partial r^2} = \left(1/r^2\right)\left\{\left(3 - \varphi^2\right)G - 3rG' - \left[\left(d^{*3}\sin\varphi\right)/\varphi\right]\right\}. \quad (28)$$

The derivative of (14) with respect to the distance from atom r_m is

$$\frac{\partial\rho}{\partial r_m} = 4\pi O \sum_{i=1}^{n} g_m\left(\bar{d}_i^*\right)\left(\frac{\partial G_i}{\partial r_m} - \frac{\partial G_{i-1}}{\partial r_m}\right). \quad (29)$$

Substitution of (26) into this equation gives $\partial\rho/\partial r$, and $\partial^2\rho/\partial r^2$ can be similarly calculated. [These derivatives can also be calculated by differentiation of (11):

$$\frac{\partial\rho}{\partial r} = (4\pi O/r)\left[\int_{d^*_{\min}}^{d^*_{\max}} g\left(d^*\right)d^{*2}\cos\left(2\pi r d^*\right)\mathrm{d}d^* - \rho\right], \quad (30)$$

where the integrals are evaluated piecewise with constant $g(d^*)$ as for (12).]

The radial derivatives [(29)] must be decomposed into their Cartesian components. Consider the x component:

$$\partial\rho/\partial x = (\partial\rho/\partial r)(\partial r/\partial x), \quad (31)$$

where

$$r = \left(x^2 + y^2 + z^2\right)^{1/2}, \quad (32)$$

$$(\partial\rho/\partial x) = (x/r)(\partial\rho/\partial r), \quad (33)$$

$$\frac{\partial^2\rho}{\partial x^2} = \frac{\partial r}{\partial x}\left\{\left(\frac{x}{r}\right)\frac{\partial^2\rho}{\partial r^2} + \frac{\partial\rho}{\partial r}\left[\left(\frac{-x}{r^2}\right) + \left(\frac{1}{r}\right)\frac{\partial x}{\partial r}\right]\right\}, \quad (34)$$

$$\frac{\partial^2\rho}{\partial x^2} = \left(\frac{1}{r}\right)\left[\left(\frac{x^2}{r^2}\right)\left(r\frac{\partial^2\rho}{\partial r^2} - \frac{\partial\rho}{\partial r}\right) + \frac{\partial\rho}{\partial r}\right]. \quad (35)$$

In the restrained least-squares refinement, a residual $\mathcal{R}$ is minimized that is the sum of electron-density and stereochemical terms:

$$\mathcal{R} = \mathcal{R}_\rho + \mathcal{R}_{\rm geom}, \quad (36)$$

$$\mathcal{R}_\rho(P) = \sum_{xyz} \{[S\rho_o(xyz) + k] - \rho_c(P, xyz)\}^2, \quad (37)$$

where ρ_o and ρ_c are the observed and calculated electron densities to be fitted by varying the atomic parameters P and where S and k are refinable scaling parameters required to bring the observed map to an absolute scale. To optimize the fit of the model to the electron density, the derivatives with respect to the atomic parameters for each atom, $\partial\mathcal{R}_\rho/\partial P_m$, are required.

$$\frac{\partial\mathcal{R}_\rho}{\partial P_m} = \sum_{xyz} -2[S\rho_o(xyz) + k - \rho_c(xyz)]\frac{\partial\rho_c(xyz)}{\partial P_m}, \quad (38)$$

$$\frac{\partial^2\mathcal{R}_\rho}{\partial P_m^2} = \sum_{xyz} -2\left\{[S\rho_o(xyz) + k - \rho_c(xyz)]\frac{\partial^2\rho_c(xyz)}{\partial P_m^2} - \left(\frac{\partial\rho_c(xyz)}{\partial P_m}\right)^2\right\}. \quad (39)$$

The partial derivatives $\partial\rho/\partial P_m$ and $\partial^2\rho/\partial P_m^2$ with respect to each of the atomic parameters have already been derived [(33), (35), (29), (16), (20) and (21)] and can be substituted into these last two equations. Note that, as shown in (15), the partial derivatives $\partial\rho/\partial P_m$ and $\partial^2\rho/\partial P_m^2$) contain only terms from the mth atom, but the calculated electron density itself, $\rho_c(xyz)$, contains contributions from all neighboring atoms. The residual $\mathcal{R}_\rho$ is therefore calculated by summing over all xyz grid points of the observed map that are close to any atom.

3.3. *Real-space R factor*

To assess the quality of the fit of a model to electron density, a normalized real-space R factor, $R^{\mathcal{ED}}$, is defined as

$$R^{\mathcal{ED}} = \sum_{xyz} [S\rho_o(xyz) + k - \rho_c(P, xyz)]^2 \times \left\{\tfrac{1}{2}\sum_{xyz} [S\rho_o(xyz) + k + \rho_c(P, xyz)]^2\right\}^{-1}. \quad (40)$$

For an overall measure of quality, all xyz map grid points close to atoms are included in the summation. This is similar in concept to the R factor of Jones *et al.* (1991), except for the factor of $\frac{1}{2}$ in the denominator, making it somewhat more compatible with conventional crystallographic reciprocal-space R factors but leading to values double those of Jones *et al.* (1991).

4. Methods and results

4.1. *Accuracy of electron-density calculation*

The electron density for a single C atom ($B = 15$ Å^2) was calculated at several different resolutions by three different methods:

(1) Structure factors were calculated (Ten Eyck, 1977) for an atom at the center of a 10 Å cubic unit cell, using analytical approximations to the standard scattering factors (Lee & Pakes, 1969; Tronrud & Ten Eyck, 1992). These were used for a resolution-truncated Fourier synthesis (Ten Eyck, 1973).

(2) The FT(f) function (10) was used with $\Delta d^* = 0.1$ Å^{-1}.

(3) The spherical Gaussian function of Diamond [1971; equations (1) and (2)] was used. A search was made for the additional B factor that would give the best agreement of this method with method (1).

The agreement of methods (2) and (3) with method (1) is good (Fig. 1). Comparison of the electron densities to those of method (1) shows that method (2) [FT(f); $R^{\mathcal{ED}} = 0.007$] is slightly better than method (3) (spherical Gaussian; $R^{\mathcal{ED}} = 0.042$) for the 2 Å data shown in Fig. 1 and at 3 Å ($R^{\mathcal{ED}} = 0.005$ *cf.* 0.016). However, the principal advantage of FT(f) is the absence of a requirement to smear the calculated electron density. For the spherical Gaussian, the optimum additional B factors are 28 Å^2 at 2 Å, 73 Å^2 at 3 Å and 139 Å^2 at 4 Å.

The differences between the FT(f) and spherical Gaussian methods are highlighted in Fig. 2. The spherical Gaussian has a more extended tail whereas the

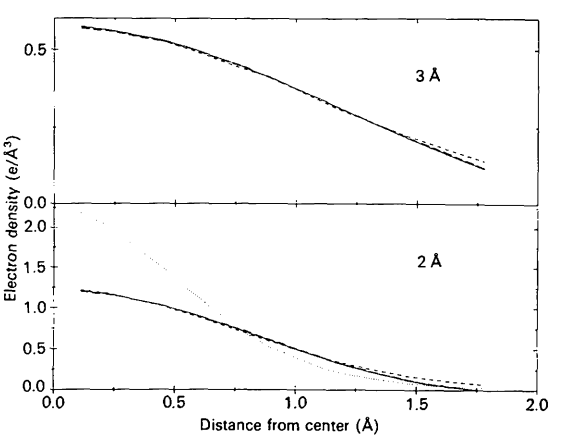

Fig. 1. Calculation of electron density. Three different methods of calculation are compared at two resolutions: (1) structure-factor calculation and resolution-truncated Fourier inversion (long dashes); (2) the new resolution-dependent function (solid lines); (3) spherical Gaussian smeared with optimized B factors of 73 Å^2 (top) and 28 Å^2 (bottom) (short dashes). The dotted line shows the electron density calculated at infinitely high resolution.

FT(f) dies down to a small ripple beyond 2.4 Å. While the difference is modest, the overall contribution of an atom throughout a calculated electron-density map is the volume integral of the atomic density function. The volume of radial shells increases as the square of the radius and Fig. 2 demonstrates that the 'tails' of the spherical Gaussian increase the overall electron density more than the FT(f).

4.2. *Implementation*

Real-space refinement has been introduced as an additional module (*RSREF*) for *TNT* (Tronrud *et al.*, 1987). The refinement thereby makes use of a particularly user-friendly stereochemical description that has been used extensively for both proteins and nucleic acids. *RSREF* provides *TNT*'s control program *SHIFT* with a list of the derivatives [(38) and (39)] for the agreement of calculated and observed electron densities with respect to atomic parameters.

RSREF contains 1350 lines of ANSI C and calls C and Fortran77 object libraries for general crystallographic, vector and matrix operations and for input/output. Storage for the observed and calculated electron densities is allocated dynamically, so that the resources required for a local refinement are minimal. Memory is allocated for a grid within a parallelepiped that encloses, with a small margin, those atoms that are to be refined. Observed electron-density values are read from dsn6-type files that are used by the programs *FRODO* (Jones, 1978) and *O* (Jones *et al.*, 1991) using a Fortran interface. Through their direct-access structure and also through their organization as $8 \times 8 \times 8$ 'bricks' of electron density (rather than as sections), input is minimized. Use of dsn6 files also facilitates the development of these methods for interactive model-building applications. Other files are based on the token-driven free-format system of *TNT*.

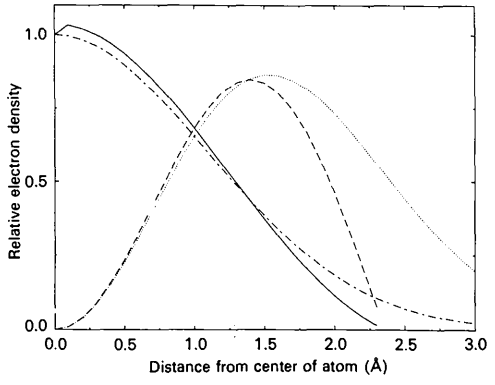

Fig. 2. New [FT(f)] and spherical Gaussian functions compared. Calculated electron densities relative to that at the center of the atom are shown for the new FT(f) (solid line) and spherical Gaussian (Diamond, 1971; dash-dotted line). The resolution range is 20–3 Å and an optimal additional B factor of 73 Å^2 was used for the spherical Gaussian. (The dip in electron density at zero radius is due to truncation ripples from the 20 and 3 Å resolution limits both destructively interfering to their maximum extents at zero radius.) The dashed and dotted lines show, respectively, the relative electron densities of the new and spherical Gaussian methods weighted by the volume of the shell at that radius.

On a first pass, grid points are flagged if they are found to be within $r_{\rm calc}^{\rm max}$ of any atom that is being refined. The electron densities of these grid points are calculated according to (14). The grid of the map need not be orthogonal. The optimal $r_{\rm calc}^{\rm max}$ is about 3.4 Å, as determined below. Form factors could be interpolated from a table. Here, they are evaluated from the multiple Gaussian approximation used by *TNT* (Tronrud *et al.*, 1987). Other analytical approximations (Vand, Eiland & Pepinsky, 1957; Lee & Pakes, 1969; Cromer & Waber, 1974) that estimate the scattering to within 1% would be equally acceptable. It is computationally more efficient to calculate an atom's contribution to all local grid points before proceeding to the next atom than to search for all atoms that contribute to a given grid point. This means that the g coefficients [(14)] can be used repeatedly until either the atom type or the B factor changes. ρ_o is scaled to ρ_c by linear least-squares determination of S and k by minimization of the residual $\mathcal{R}_\rho$ [(37)]. On a second pass, the first [(38)] and optionally the second [(39)] derivatives are calculated as required for the mode in which *TNT*'s *SHIFT* is to be run. Only grid points within $r_{\rm ref}^{\rm max}$ (see below) of the center of an atom are used in the calculation of its derivatives. Evaluations of the algebraically derived partial derivatives have been cross-checked by comparison with finite difference methods.

TNT is designed to be run as a series of 'long' cycles, each of which contains a number of 'short' cycles. On the first short cycle of each long cycle, *TNT*'s control program *SHIFT* expects to be given derivatives from each of the modules such as *RSREF* and *TNT*'s *GEOMETRY*. *SHIFT* then calculates a shift vector specifying the directions and relative sizes of shifts to all of the atomic parameters. *SHIFT* calculates the shift vector by one of four methods depending on the number of previous long cycles and the derivatives supplied. The steepest-descent and conjugate-gradient methods require only first derivatives but the gradient-over-curvature and conjugate-direction methods require second derivatives (Tronrud & Ten Eyck, 1992; Tronrud, 1992). A shift of one atom affects the electron-density fit and stereochemistry of its neighbors. Over the subsequent short cycles, a search is made for an optimal scale factor for the shift vector that minimizes the residual [(36)]. In each of the short cycles, a trial shift is applied and each of the modules is polled for its residual (function value in *TNT* terminology). *RSREF* can do this in a single-pass mode, specified by a command-line flag, in which no derivatives are calculated.

Real-space refinement, as implemented here, will be most advantageous when it is applied to a fraction of the full unit cell. Whether one of several symmetry-related identical protomers or a small region of a protein is being refined, atoms from neighbors related by crystallographic

Table 1. *Optimization of the cut-off radii*

Real-space R factors, $R^{\mathcal{ED}}$, are tabulated for the match of a nucleotide structure to a 3 Å test map. The FT(f) method and the spherical Gaussian method (Diamond, 1971) are compared using different cut-off radii for contribution to the calculated electron density $r_{\rm calc}^{\rm max}$ and for scaling/residual calculation $r_{\rm ref}^{\rm max}$.

Refinement			Cut-off for electron-density calculation [$r_{\rm ref}^{\rm max}$ (Å)]							
$r_{\rm ref}^{\rm max}$ (Å)	Volume (Å^3)	Method	1.00	1.33	1.73	2.20	2.74	3.38	4.10	4.91
1.00	4.2	New: FT(f)	0.274	0.144	0.081	0.054	0.048	0.023	0.021	0.014
		Gaussian	0.259	0.140	0.084	0.063	0.069	0.070	0.070	0.070
1.33	11	New: FT(f)		0.192	0.093	0.075	0.063	0.030	0.026	0.021
		Gaussian		0.190	0.092	0.077	0.088	0.090	0.090	0.090
1.72	24	New: FT(f)			0.149	0.123	0.097	0.043	0.037	0.037
		Gaussian			0.142	0.117	0.135	0.138	0.138	0.138
2.20	49	New: FT(f)				0.203	0.163	0.067	0.057	0.067
		Gaussian				0.205	0.218	0.220	0.220	0.220
2.74	97	New: FT(f)					0.243	0.092	0.075	0.104
		Gaussian					0.362	0.365	0.365	0.365
3.38	180	New: FT(f)						0.123	0.102	0.142
		Gaussian						0.450	0.450	0.450
4.10	322	New: FT(f)							0.152	0.193
		Gaussian							0.496	0.496
4.91	557	New: FT(f)								0.249
		Gaussian								0.547

or noncrystallographic symmetry must be considered for correct stereochemical refinement and because their electron densities may overlap those of the atoms that are being refined. Support programs provide the following functions:

EXPCOORD calculates the atomic parameters for crystallographically and noncrystallographically symmetry related residues that fall within a given box or within a given distance from any atom that is being refined.

SEQUENCE is a parser, written in the language Lex, that prepares the input for *TNT*'s *GEOMETRY* so that the coordinate file can contain disconnected peptide fragments generated by *EXPCOORD*.

BOX & FIX calculates the dimensions of a box that surrounds a zone of residues and flags fixed atoms in a surrounding buffer region.

RMDERIV collates the output of two runs of *TNT*'s *GEOMETRY* to create a list of derivatives for one of the symmetry equivalents, which includes the nonbonded interactions with neighboring molecules.

REEXPAND updates the parameters of 'fixed' atoms with those from symmetry-equivalent atoms that have been refined.

4.3. *Optimization of parameters*

Two limiting radii are used by *RSREF*: $r_{\rm calc}^{\rm max}$ is the distance beyond which the contribution of an atom to the calculated electron density of a pixel will be assumed to be zero; and $r_{\rm ref}^{\rm max}$ ($\leq r_{\rm calc}^{\rm max}$) defines a sphere enclosing grid points around each atom that are used for refinement, namely in scaling, derivative and residual calculations. *RSREF* is provided with command-line options to facilitate the optimization of these parameters, which may depend on the maximum resolution. An example survey of $r_{\rm ref}^{\rm max}$ and $r_{\rm calc}^{\rm max}$ for a test case is shown in Table 1. The test coordinates are those of a small nucleotide. A 3 Å test map was calculated from the coordinates using method (1) (§4.1). The $R^{\mathcal{ED}}$ values of Table 1 compare the test map with electron densities calculated by *RSREF* using either method (2) [FT(f)] or method (3) (spherical Gaussian), using all atoms within $r_{\rm calc}^{\rm max}$ of any given pixel. These calculated electron densities were least-squares scaled to the test map using only those grid points within $r_{\rm ref}^{\rm max}$ of any of the test atoms. The real-space R factor, $R^{\mathcal{ED}}$, was calculated using only these grid points. For the spherical Gaussian, the electron-density calculation and scaling were repeated for each pair of radii to search for the best additional B factor in steps of 10 Å^2. (A local search with a finer step indicated that the minimal $R^{\mathcal{ED}}$ for the spherical Gaussian is within 0.03 of the true minimum.)

Calculation radius r_{calc}^{max}. The accuracy improves as $r_{\rm calc}^{\rm max}$ increases, with substantial improvements up to about 3.4 Å from the atom centers. The larger radii (~3 Å) are important for several reasons:

(1) The overall contribution of each radial shell to the map is weighted by the square of the radius (proportional to the shell volume; Fig. 2).

(2) The average number of atoms that are within a sphere surrounding a grid point and therefore contribute to its density shell is proportional to the cube of the radius.

Refinement radius r_{ref}^{max}. $R^{\mathcal{ED}}$ increases steadily as $r_{\rm ref}^{\rm max}$ increases. Again, there are several likely causes:

(1) The scaling of calculated and test electron densities has reduced degrees of freedom as the radial cut-off increases and more grid points are used.

(2) If only the central regions are used, fitting is less sensitive to the exact shape of the calculated electron-density function.

In the choice of an appropriate $r_{\rm ref}^{\rm max}$, several other factors should be considered:

(1) If a large volume is used for partial-derivative calculation, partials will be calculated using regions where other atoms dominate. There will be a tendency for atomic parameters to shift to compensate for the errors of their neighbors. This would lead to slow convergence.

(2) Computational complexity increases with the number of grid points [$\propto r_{\rm ref}^{\rm max})^3$].

(3) An experimental map contains errors. The accumulated contribution of error to the residual and partial derivatives (noise) will increase approximately as the square root of the number of grid points [$(r_{\rm ref}^{\rm max})^{3/2}$]. For radii $\leq$ 1.5 Å, the total volume-weighted density of radial shells (signal) also increases (Fig. 2). Beyond the maximum, the signal-to-noise ratio must decrease.

(4) The radius of convergence of real-space refinement is proportional to the overlap between observed and calculated electron-density peaks (Diamond, 1985). The radius of convergence is therefore proportional to $r_{\rm ref}^{\rm max}$.

(5) For computational expediency, Diamond (1971) mimimized a volume integral: $\int_V (\rho_o - \rho_c)^2 {\rm d}V$. The integrals needed for the normal matrix were precise only with gross oversampling and grid sum corrections [equation (9) of Diamond (1971)]. The current method minimizes a simple difference at grid points [(37)], avoiding these problems. The maps need to be oversampled, but only to the extent that the five parameters per atom are overdetermined several-fold.

The choices of $r_{\rm calc}^{\rm max}$ and especially $r_{\rm ref}^{\rm max}$ are compromises that depend on the resolution, grid spacing and type of refinement: quick and crude or slow and precise. For refinements at ~3 Å, $r_{\rm ref}^{\rm max}$ = 1.6 Å and $r_{\rm calc}^{\rm max}$ = 3.4 Å have worked well, with smaller values sometimes being used for crude initial refinement. Δd^* has usually been set to give n = 10 evaluations of the form factor [(14)]. With a small loss of precision (increasing $R^{\mathcal{ED}}$ by $\leq$ 0.02), refinement can be accelerated through the use of n = 5 terms.

4.4. *Test* 1: *isolated atoms*

A 3 Å electron-density map was calculated from four nonoverlapping atoms. For repeated tests, coordinates were first randomly displaced and then refined without the use of geometrical restraints. Displacements of various root-mean-square (r.m.s.) magnitudes were tested. The convergence radius was about 1.0 Å using either the FT(f) function or spherical Gaussians, and the residual positional error was 0.06 Å for spherical Gaussians and 0.03 Å for the FT(f) function. Convergence was reached in four to eight cycles. Thus, for isolated atoms, both functions appear acceptable.

4.5. *Test* 2: *overlapping atoms*

The second test was much more stringent because the test map was calculated from a nucleotide with atoms separated by ~1.5 Å whose electron densities overlap at 3 Å resolution. Furthermore, no geometrical restraints were used and shifts were not dampened. Thus, this is an attempt at free-atom refinement at subatomic resolution. Refinement using the spherical Gaussian function diverges, even with a r.m.s. displacement of only 0.1 Å. Refinement with the FT(f) function diverges only with initial displacements $\geq$0.5 Å (Fig. 3). Although refinement is not precise, with initial displacements of 0.2 to 0.3 Å, refinement moves the coordinates in the right directions, approximately halving the initial error. This modest success is probably due to better representation of the outer 'tails' of the atomic electron densities (see §5). The improved conditioning of free-atom refinement using the FT(f) function is probably the reason why stereochemical restraints can be used. Methods using spherical Gaussians require tight rigid-body constraints (Diamond, 1971; Jones & Liljas, 1984).

4.6. *Test* 3: *canine parvovirus capsid*

The empty capsid structure of canine parvorirus (CPV) provided a good test example. Its structure had initially been determined by molecular replacement, starting with an unrefined model of the full capsid that contains DNA (Tsao *et al.*, 1991). The empty structure was refined (Wu, Keller & Rossmann, 1993) from a conventional R factor of 35% to one of 21%, in six rounds of reciprocal-space refinement, using either restrained methods (*PROLSQ*; Hendrickson, 1985) or molecular dynamics (*X-PLOR*; Brünger *et al.*, 1987). Between rounds 1 and 2, parts of the model were rebuilt manually. Between rounds 3 and 4, the oscillation data

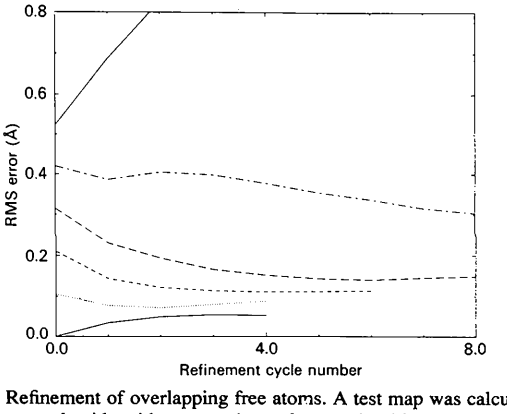

Fig. 3. Refinement of overlapping free atoms. A test map was calculated from a nucleotide with atoms whose electron densities overlap at 3 Å resolution. Random displacements were introduced before refinement (cycle 0) and the progress of real-space refinement [using the FT(f) function] is shown. Each line shows the progress of refinement started with a different initial displacement whose size is shown by the y intercept.

Table 2. *CPV empty capsid model statistics at selected stages of refinement*

The 'final' model is that of Wu *et al.* (1993), except that the 85 water molecules per protein have been removed. Following the r.m.s. contact error is the number of poor contacts per protomer. The refinement round (†) follows the numbering scheme of Wu *et al.* (1993).

			R.m.s. bond/contact errors				R.m.s. differences	
Model/round†	R^{conv} (%)	Data	Length (Å)	Angle (°)	Torsion angle (°)	Contact (Å)	*Cf.* input	*Cf.* final
A: unrefined	34.7	3 Å/3σ	0.022	2.6	20.4	0.28 (216)	–	1.04
B: after *PROLSQ*/1	30.6	3 Å/3σ	0.026	4.2	21.5	0.25 (167)	0.77	0.97
C: *RSREF* from *A*	29.1	3 Å/3σ	0.020	2.4	16.0	0.04 (234)	0.75	0.88
D: input to *X-PLOR*	27.9	3.2 Å/4σ	0.021	3.9	21.0	0.26 (164)	–	0.91
E: after *X-PLOR*/3	25.2	3.2 Å/4σ	0.027	4.5	20.3	0.60 (15)	0.47	0.88
F: *RSREF* from *D*	25.3	3.2 Å/4σ	0.026	4.5	16.8	0.12 (330)	0.61	0.87
G: input to *PROLSQ*	26.2	3 Å/3σ	0.027	4.5	20.1	0.67 (23)	–	0.64
H: after *PROLSQ*/4	24.4	3 Å/3σ	0.022	4.3	20.5	0.26 (86)	0.32	0.54
I: *RSREF* from *G*	24.0	3 Å/3σ	0.018	4.6	17.4	0.16 (155)	0.48	0.61
J: Final *PROLSQ*/6	23.1	3 Å/3σ	0.023	4.4	19.6	0.29 (94)	–	–

were postrefined (Rossmann, 1985), again using slightly different cell dimensions, and the phases were re-refined by symmetry averaging (Rossmann *et al.*, 1992). The starting phases were calculated from the atomic model of round 3, but past experience has shown that the power of 30-fold averaging is sufficient to remove model bias (see, for example, Kim *et al.*, 1989).

Both *PROLSQ* and *X-PLOR* have been adapted for virus structure refinement. Arnold & Rossmann (1988) have shown that, with the accurate phases available for viruses, refinement against complex structure factors (rather than magnitudes) is effective. For the CPV empty structure, a compromise crystallographic residual was used (Wu *et al.*, 1993):

$$\mathcal{R}_{\text{X-ray}} = \sum_h \Big\{ (1 - w_h)\big(|F_h^{\text{obs}}| - |F_h^{\text{calc}}|\big)^2 + w_h \Big[\big(A_h^{\text{obs}} - kA_h^{\text{calc}}\big)^2 + \big(B_h^{\text{obs}} - kB_h^{\text{calc}}\big)^2 \Big] \Big\}, \quad (41)$$

where k is a scaling constant and (A_h, B_h) are the real and imaginary parts of the structure factor F_h of reflection h that has figure of merit w_h. If all w_h were zero, this would be conventional crystallographic refinement.

As was done by Arnold & Rossmann (1988), to reduce computing time, each cycle of *PROLSQ* was run with alternating subsets of reflections. The R factors given by Wu *et al.* (1993) were calculated using the same subset of data as was used for the immediately preceding refinement. These R factors might be artificially low because refinement can overfit a model to the diffraction data (Brünger, 1992). The R factors quoted in Table 2 were recalculated using data that were selected randomly from the whole data set. These R factors are slightly higher, showing that some overfitting had occurred in the preceding cycle of *PROLSQ*. However, they are still expected to be lower than the R_T^{free} proposed by Brünger (1992) because the reflections had been used in previous cycles of refinement. The stereochemical statistics were also recalculated using *TNT*'s *GEOMETRY* (Tronrud *et al.*, 1987) to ensure that the same standard parameters and weights were used to compare all atomic models.

To reduce the computational requirements, *X-PLOR* was run using a simulation at 2000 K rather than the usual 3000–4000 K (Wu *et al.*, 1993). Refinement at higher temperatures has been shown to be more effective and it might have been possible to reduce the R factor by an additional 1% (Brünger *et al.*, 1987).

RSREF was compared to the three rounds of refinement (1, 3 and 4) during which greatest improvement had been seen by Wu *et al.* (1993). Round 1 refined the initial model using *PROLSQ*. Round 3 used *X-PLOR* and round 4 was the final *PROLSQ* refinement prior to the inclusion of isotropic temperature factors and water molecules. In all cases, *RSREF* was run using the final symmetry-averaged map of Wu *et al.* (1993) and using weights adjusted to give stereochemical statistics similar to those of the corresponding rounds of Wu *et al.* (1993). Refinement and model statistics are given in Table 2. In judging the quality of the refinements, several statistics should be compared as there are biases in each. R_T^{free} (Brünger, 1992) could not be calculated because all of the data had been used in previous rounds of *PROLSQ* or *X-PLOR*. A conventional R factor (R^{conv}) can give an unrealistically good estimate of model quality following reciprocal-space refinement (Brünger, 1992) because the differences $||F^{\text{obs}}| - |F^{\text{calc}}||$, used to calculate the R factor have been minimized directly (see §5). Equally, $R^{\mathcal{ED}}$ should be lower for a real-space-refined model than a reciprocal-space-refined model of similar quality. The nonlinear least-squares technique is guaranteed only to find a local minimum in the residual. A 'final' structure with $R^{\text{conv}} \simeq 0.2$ retains significant errors and is not the 'correct' structure of the global minimum. Models from the final cycles of refinement will contain similar errors

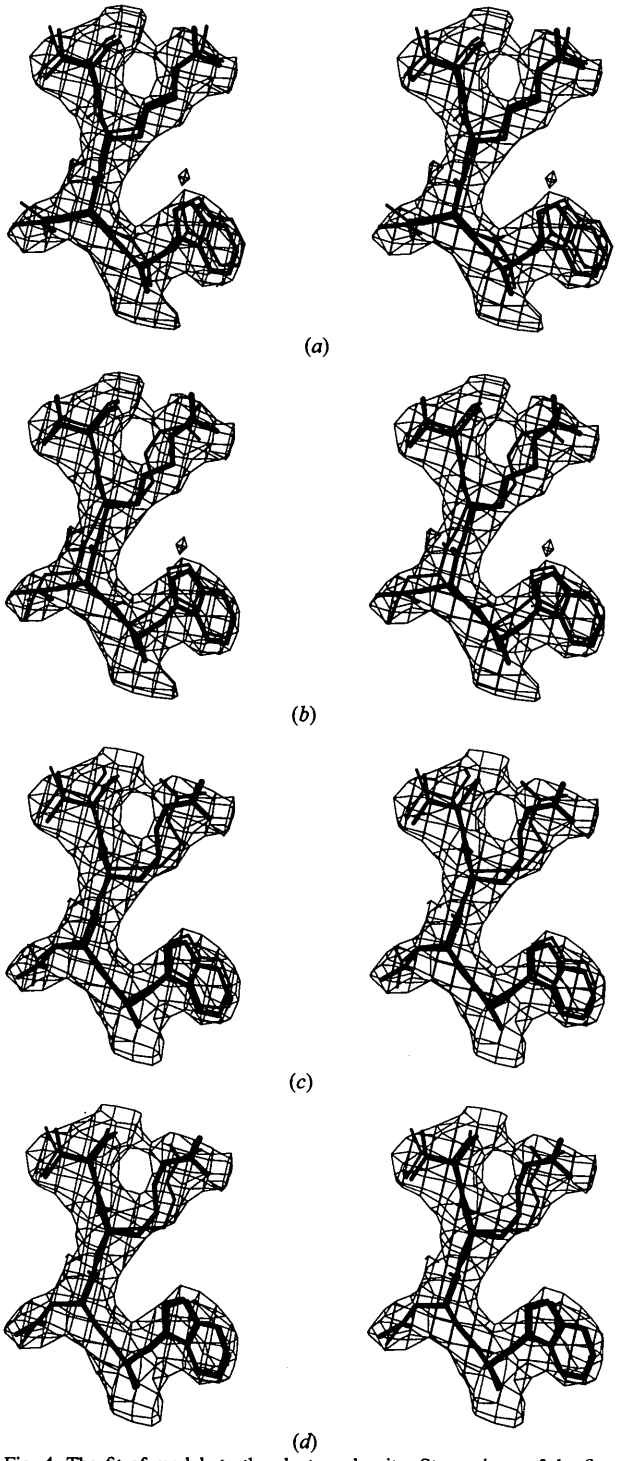

Fig. 4. The fit of models to the electron density. Stereoviews of the fits of residues Trp 214 (bottom), Asp 215, Arg 216 and Ser 217 (top) to the symmetry-averaged electron density. (*a*) *RSREF* real-space-refined model *C* (thick lines) and unrefined model *A* (thin lines). (*b*) Initial refinements: *RSREF* model *C* (thick lines) and *PROLSQ* model *B* (thin lines). (*c*) *RSREF* model *F* (thick lines) and *X-PLOR*-refined model *E* (thin lines). (*d*) *RSREF* model *I* (thick lines) and later run of *PROLSQ* (model *H*; thin lines). The figure was drawn with the program *O* (Jones *et al.*, 1991) using the electron density and reciprocal-space-refined models of Wu *et al.* (1993).

and the difference in coordinates (compared to the final 'yardstick') underestimates the errors of the models. As the biases of these statistics are different, the quality of refinement should be judged by considering them all.

For the early steps in refinement, comparison of models *B* and *C* (Table 2) shows that *RSREF* performs slightly better. The r.m.s. difference between models *B* and *C* is 0.75 Å. The *R* factor is lower (even with tighter geometry) and the model has moved closer to the final model. At an intermediate stage of refinement, the *R* factors of models *E* and *F* suggest that *RSREF* is comparable to *X-PLOR*. The r.m.s. movements are slightly larger for *RSREF* but the models have similar r.m.s. differences with the final model. Between models *E* and *F*, the r.m.s. difference is of similar magnitude, 0.68 Å. Towards the final stages of refinement, comparison of models *H* and *I* suggests that *RSREF* performs slightly better than *PROLSQ*. Models *H* and *J* are highly correlated so r.m.s. positional comparisons are not a good indicator.

Fig. 4 compares the fit of each of the models to the electron density. In subjective comparisons using *O* (Jones *et al.*, 1991), without knowing which was which, crystallographic colleagues picked real-space-refined models over the corresponding reciprocal-space-refined models as fitting the electron density slightly better. The differences are more obvious at the start of refinement (model *B cf.* model *C*) than at the end of refinement (model *H cf.* model *I*). Comparison of the unrefined model *A* to model *C* shows that *RSREF* has centered the backbone and pulled all of the side chains further into the electron density. It is less obvious from this figure that it has also moved all of the backbone carbonyls towards slight bulges in the electron density.

Convergence of *RSREF* was monitored using the real-space *R* factor $R^{\mathcal{ED}}$ [(40)]. Its asymptotic decrease to convergence for each of the three rounds of refinement is shown in Fig. 5. Weights were kept constant through these cycles, except for a slight tightening of geometry after cycle 4 of the refinement of model *G* to model

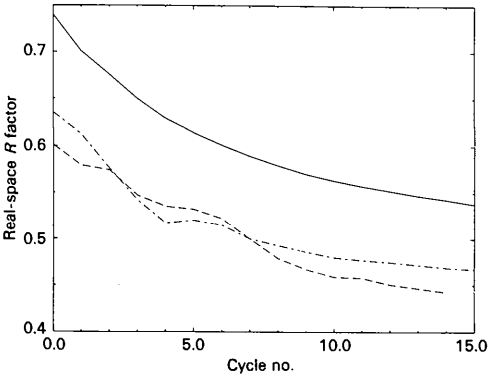

Fig. 5. Convergence of real-space refinement. The refinement of model *A* to model *C* is shown by a solid line, that of model *D* to model *F* by dash-dotted line and that of model *G* to model *I* by a dashed line.

I. A few additional cycles were used to adjust the weights to give stereochemical statistics similar to the rounds of Wu *et al.* (1993), yielding $R^{\mathcal{ED}} = 0.526$, $R^{\mathcal{ED}} = 0.464$ and $R^{\mathcal{ED}} = 0.419$ for models *C*, *F* and *I*, respectively. Convergence is slightly faster if second derivatives are used, except for stereochemically very poor models when it is easier to find suitable weights if only first derivatives are used. Alternate radical loosening and tightening of geometrical restraints can sometimes enable a crude model to be refined by conventional reciprocal-space methods. With *RSREF*, this is no more effective than smooth convergence to the target values, perhaps because the convergence radius of real-space refinement is already quite large.

For the ~600 residues of the CPV noncrystallographic asymmetric unit and ~ 600 residues of symmetry-equivalent neighbors, full derivative calculation by *RSREF* takes 7 min of CPU time on a MIPS 4000 processor. Its memory requirements are modest, about 2.5 Mbytes of virtual memory, and it runs optimally if at least 1.5 Mbytes of physical memory are available. For a typical long refinement cycle that includes two short cycles, the turnaround time is about 25 min per long cycle.

5. Discussion

5.1. *Electron-density functions compared*

The FT(f) electron-density function [(10)] can be used to calculate the electron-density function of a multi-atom structure with about three times the accuracy of spherical Gaussians (Table 1). The slowly asymptotic form of the Gaussian tail (Fig. 2) is a poor approximation to the electron density of atoms as calculated from resolution-truncated sets of structure factors. Low-resolution maps are modeled with the spherical Gaussian method by adding an arbitrary atomic radius constant or equivalent B factor to all atoms (Deisenhofer & Steigemann, 1975). The optimal constant is very dependent on the resolution (see §4.1 above) and on the integration volume: optimal B factors for the data of Table 1 ranged between 30 and 80 Å^2. There is no *a priori* reason to expect the most appropriate constant to be the same for atoms of different elements and thermal vibrations. The FT(f) function avoids these problems.

5.2. *Reciprocal-space R factors*

There are several reasons why a real-space-refined model might have a higher R factor than a conventionally refined model of equal quality. Minimization of $\int(\rho_o - \rho_c)^2 \mathrm{d}v$ is equivalent to minimization of $(1/V)\sum_h(|\mathbf{F}_h^{\mathrm{obs}} - \mathbf{F}_h^{\mathrm{calc}}|)^2$ (Diamond, 1985). Compare this to the residual minimized in conventional crystallographic refinement (Hendrickson, 1985):

$$\mathcal{R}_{\text{X-ray}} = \sum_h [\sigma^2(F_h)]^{-1}\left(\left|F_h^{\mathrm{obs}}\right| - \left|F_h^{\mathrm{calc}}\right|\right)^2, \quad (42)$$

where $\sigma(F_h)$ is the estimated standard deviation of F. Real-space refinement can therefore be considered to be a phased analog of reciprocal-space refinement in which all reflections are given equal weight. Weights of $1/\sigma^2(F_h)$ are used in reciprocal-space refinement because they should yield a solution with minimal error in F and therefore a low R factor. Other weighting schemes, including the uniform weighting of real-space refinement, will yield slightly higher R factors.

Brünger (1992) has demonstrated that the the conventional R factor (R^{conv}) may be artificially lowered by overfitting a model to structure factors. R^{conv} is unlike standard statistical tests of goodness of fit in that no account is taken of the degrees of freedom [number of independent atomic parameters; see Hamilton (1964)]. This explains Brünger's observation that the difference between R^{conv} and R_T^{free} is greatest at low resolution where the ratio of data to parameters is low. The crystallographic asymmetric units of viruses have many identical copies of the same molecule, giving a high ratio of total data to parameters, and it is expected that R^{conv} should be only slightly lower than R_T^{free}. However, if the model is refined against a subset of the data and R^{conv} is calculated with the same subset, then R^{conv} is likely to be underestimated.

As real-space refinement implicitly uses all data, but previous viral refinements have used partial data sets, for comparative purposes, it is useful to know by how much R^{conv} can be lowered. R factors were recalculated for CPV and other examples. For initial refinement cycles, R^{conv} calculated from the subset of data used for refinement could be compared with data that had not been used in refinement. For later cycles, it was compared with a random selection from the whole data set from which different subsets had been used in previous refinement cycles. With > 150 000 reflections (A and B coefficients), there was no appreciable lowering. With 45 000 to 60 000, R^{conv} was lowered by about 1.2%, and with 14 000 to 30 000, R^{conv} was lowered by about 4.5%.

5.3. *Real-space refinement and virus structures*

The statistics of Table 2 show that, while *RSREF* and *X-PLOR* are comparable, the performance of *RSREF* can exceed that of *PROLSQ*. The differences in model quality are relatively small. *RSREF* does not, for example, alleviate the need for manual rebuilding. There are several reasons why *RSREF* might be slightly better:

(1) *PROLSQ* was used with subsets of the data, whereas real-space refinement is equivalent to use of all of the data simultaneously.

(2) Real-space refinement might have a larger convergence radius.

(3) Real-space refinement has the potential for being much better conditioned than reciprocal-space refinement. In reciprocal space, the structure factors depend

on all atoms and therefore the derivatives are highly interdependent. In real space, the interdependency is limited to the number of atoms whose electron densities overlap. Good conditioning is apparent not only in the resulting model but also in the number of iterations required for refinement. It was rare that more than two short cycles were required for optimization of the shift vectors.

The major attraction of *RSREF* is that it is much faster than either *PROLSQ* or *X-PLOR*. For a typical virus, each cycle of *PROLSQ* requires several hours of supercomputer CPU time and generally has a turn-around time of about 24 h. Each round of *X-PLOR* takes many days on a supercomputer. Each (long) cycle of *RSREF–TNT* for a viral protomer typically turns around in less than 25 min on a desktop graphics workstation. Each round of refinement (~15 cycles) can be completed overnight, compared to several weeks for *PROLSQ* or *X-PLOR*. With *RSREF*, such refinements will no longer be daunting.

5.4. *RSREF and protein structures*

Owing to high noncrystallographic redundancy, viruses have phases that are unusually accurate (Arnold & Rossmann, 1988). For viruses, it is beneficial to implicitly use the phases in real-space refinement. Real-space refinement is unlikely to replace reciprocal-space refinement for proteins, for which it is desirable that refinement be independent of poorly determined experimental phases, but it may be a useful addition. Much effort is expended in manually improving the fit of models to the electron density to ensure that they are within the convergence radius of reciprocal-space refinement. With consideration of this, a macro has been written to run *RSREF* from within the graphics program *O* (Jones *et al.*, 1991). Within a couple of minutes, several cycles of refinement can be run for a short zone of residues and their neighbors. The macro has been successfully applied to examples with good maps and tests will be extended to a systematic survey of structures with maps of poorer quality (Zhou & Chapman, work in progress). The *RSR* tool that is available in *O* is able to search for the best fit of rigid-body fragments to the density. The *RSREF*-based macro is a substantial improvement in that it is a least-squares refinement incorporating full bonded/nonbonded geometrical restraints and uses a much improved electron-density function. Improvement of the starting model will enable a more conservative approach to refinement, lessening the chance that an incorrect model could be found with an acceptable R factor.

This work was supported by an NSF grant to Michael Rossmann (while the author was on leave of absence from FSU to complete a postdoctoral fellowship) and by grants from the Lucille P. Markey Foundation for the development of structural biology at Purdue and Florida State Universities. I am grateful to Hao Wu for use of CPV tetragonal data prior to publication; to Kyung Kim, Robert McKenna and V. J. Reddy for R-factor calculations with rhinovirus 14, flock house virus and bacteriophage $\varphi X174$, respectively; and to Michael Rossmann for encouragement and support. I am also grateful to Don Caspar and to one of the referees for pointing out alternative forms and derivations for some of the equations.

References

ARNOLD, E. & ROSSMANN, M. G. (1988). *Acta Cryst.* A**44**, 270–288.
BRÄNDÉN, C.-I. & JONES, T. A. (1990). *Nature (London)*, **343**, 687–689.
BRÜNGER, A. T. (1992). *Nature (London)*, **355**, 472–475.
BRÜNGER, A. T., KURIYAN, J. & KARPLUS, M. (1987). *Science*, **235**, 458–460.
CROMER, D. T. & WABER, J. T. (1974). *International Tables for X-ray Crystallography*, Vol. IV, edited by J. A. IBERS & W. C. HAMILTON. Birmingham: Kynoch Press. (Present distributor Kluwer Academic Publishers, Dordrecht.)
DEISENHOFER, J. & STEIGEMANN, W. (1975). *Acta Cryst.* B**31**, 238–250.
DIAMOND, R. (1971). *Acta Cryst.* A**27**, 436–452.
DIAMOND, R. (1974). *J. Mol. Biol.* **82**, 371–391.
DIAMOND, R. (1985). *Methods Enzymol.* **115**, 237–252.
FERMI, G. (1975). *J. Mol. Biol.* **97**, 237–256.
HAMILTON, W. C. (1964). *Statistics in Physical Science*. New York: Ronald Press.
HENDRICKSON, W. W. (1985). *Methods Enzymol.* **115**, 252–270.
HERMANS, J. JR & MCQUEEN, J. E. (1974). *Acta Cryst.* A**30**, 730–739.
HODEL, A., KIM, S.-H. & BRÜNGER, A. T. (1992). *Acta Cryst.* A**48**, 851–858.
JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.
JONES, T. A. & LILJAS, L. (1984). *Acta Cryst.* A**40**, 50–57.
JONES, T. A., ZOU, J.-Y., COWAN, S. W. & KJELDGAARD, M. (1991). *Acta Cryst.* A**47**, 110–119.
KIM, S., SMITH, T. J., CHAPMAN, M. S., ROSSMANN, M. G., PEVEAR, D. C., DUTKO, F. J., FELOCK, P. J., DIANA, G. D. & MCKINLAY, M. A. (1989). *J. Mol. Biol.* **210**, 91–111.
LEE, J. D. & PAKES, H. W. (1969). *Acta Cryst.* A**25**, 712–713.
LEVITT, M. (1974). *J. Mol. Biol.* **82**, 393–420.
ROSSMANN, M. G. (1985). *Methods Enzymol.* **114**, 237–280.
ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24–31.
ROSSMANN, M. G., MCKENNA, R., TONG, L., ZIA, D., DAI, J.-B., WU, H., CHOI, H.-K. & LYNCH, R. E. (1992). *J. Appl. Cryst.* **25**, 166–180.
SHREUDER, H. A., CURMI, P. G. M., CASCIO, D. & EISENBERG, D. S. (1990). *Proceedings of the CCP4 Study Weekend, 26–27 January* 1990, pp. 73–82, edited by K. HENRICK, D. S. MOSS & I. J. TICKLE. SERC Daresbury Laboratory, Warrington, England.
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. (1992). *Acta Cryst.* A**48**, 912–916.
TRONRUD, D. E. & TEN EYCK, L. F. (1992). *TNT Refinement Package*, Release 5-A. Univ. of Oregon, Eugene, OR, USA.
TRONRUD, D. E., TEN EYCK, L. F. & MATTHEWS, B. W. (1987). *Acta Cryst.* A**43**, 489–501.
TSAO, J., CHAPMAN, M. S., AGBANDJE, M., KELLER, W., SMITH, K., WU, H., LUO, M., SMITH, T. J., ROSSMANN, M. G., COMPANS, R. W. & PARRISH, C. (1991). *Science*, **251**, 1456–1464.
VAND, V., EILAND, P. F. & PEPINSKY, R. (1957). *Acta Cryst.* **10**, 303–306.
WASER, J. (1963). *Acta Cryst.* **16**, 1091–1094.
WATENPAUGH, K. D., SIEKER, L. C., HERRIOTT, J. R. & JENSEN, L. H. (1973). *Acta Cryst.* B**29**, 943–956.
WU, H., KELLER, W. & ROSSMANN, M. G. (1993). *Acta Cryst.* D**49**, 572–579.