research papers

Polarizable atomic multipole X-ray refinement: application to peptide crystals

Michael J. Schnieders,[a] Timothy D. Fenn,[b,c] Vijay S. Pande[a]* and Axel T. Brunger[b,c]*

[a]Department of Chemistry, Stanford, CA 94305, USA, [b]Department of Molecular and Cellular Physiology, Stanford, CA 94305, USA, and [c]Howard Hughes Medical Institute, USA

Correspondence e-mail: pande@stanford.edu, brunger@stanford.edu

Recent advances in computational chemistry have produced force fields based on a polarizable atomic multipole description of biomolecular electrostatics. In this work, the Atomic Multipole Optimized Energetics for Biomolecular Applications (AMOEBA) force field is applied to restrained refinement of molecular models against X-ray diffraction data from peptide crystals. A new formalism is also developed to compute anisotropic and aspherical structure factors using fast Fourier transformation (FFT) of Cartesian Gaussian multipoles. Relative to direct summation, the FFT approach can give a speedup of more than an order of magnitude for aspherical refinement of ultrahigh-resolution data sets. Use of a sublattice formalism makes the method highly parallelizable. Application of the Cartesian Gaussian multipole scattering model to a series of four peptide crystals using multipole coefficients from the AMOEBA force field demonstrates that AMOEBA systematically underestimates electron density at bond centers. For the trigonal and tetrahedral bonding geometries common in organic chemistry, an atomic multipole expansion through hexadecapole order is required to explain bond electron density. Alternatively, the addition of interatomic scattering (IAS) sites to the AMOEBA-based density captured bonding effects with fewer parameters. For a series of four peptide crystals, the AMOEBA–IAS model lowered R_{free} by 20–40% relative to the original spherically symmetric scattering model.

1. Introduction

The number of X-ray crystal structures in the Protein Data Bank (PDB) with a resolution of higher than 1.0 Å continues to increase rapidly (Berman *et al.*, 2000). In late 2002, there were already over 100 structures available at subatomic resolution (Afonine & Urzhumtsev, 2004), while as of early 2009 the number had more than tripled to well over 300. Examples include the proteins lysozyme at 0.65 Å (Wang *et al.*, 2007), aldose reductase at 0.66 Å (Howard *et al.*, 2004) and serine protease at 0.78 Å (Kuhn *et al.*, 1998), as well as nucleic acid structures such as B-DNA at 0.74 Å (Kielkopf *et al.*, 2000), Z-DNA at 0.60 Å (Tereshko *et al.*, 2001) and an RNA tetraplex at 0.61 Å (Deng *et al.*, 2001). Crystals that diffract to high resolution are ideal for studying valence-electron distributions (Jelsch *et al.*, 2000; Muzet *et al.*, 2003; Zarychta *et al.*, 2007; Volkov *et al.*, 2007; Coppens & Volkov, 2004) that dictate the electrostatic properties of macromolecules. Electrostatics, in turn, is one of the driving forces in protein and nucleic acid folding, which should be understood in detail in order to

predict biomolecular thermodynamics and kinetics (Snow *et al.*, 2002, 2005; Sorin & Pande, 2005; Pande *et al.*, 2003). In this work, we contribute an improved theory and algorithm for computing the anisotropic and aspherical valence-electron density of molecules for X-ray crystal structure refinement.

Calculation of structure factors is generally based on scattering factors defined by the isolated-atom model (IAM), which assumes that the electron density around each atom is spherically symmetric. However, subatomic resolution diffraction data capture aspherical features of the electron density that result from bonding and the local chemical environment. The difference between the IAM and the true electron density is defined as the deformation density. For example, aspherical electron-density models of diamond, silicon and germanium developed by DeMarco and Weiss and later by Dawson explained the peaks of deformation density at bond midpoints observed in the experimental data (Dawson, 1967*a,b,c*; DeMarco & Weiss, 1965). In these works, the IAM was augmented by atom-centered spherical harmonic expansions, whose physical consequence was to redistribute electron density from nonbonding lobes into the tetragonal arrangement of bond centers.

A variety of radial functions have been used in combination with atom-centered spherical harmonic expansions. Modified Gaussians were promoted by Dawson (1967*a*), a set of harmonic oscillator wavefunctions by Kurki-Suonio (1968) and more recently a formalism based on Slater-type orbitals (STO) was described by Stewart and coworkers (Epstein *et al.*, 1977; Cromer *et al.*, 1976; Stewart, 1979, 1977) and by Hansen & Coppens (1978), which represents the current standard (Jelsch *et al.*, 2005; Zarychta *et al.*, 2007; Volkov *et al.*, 2007; Coppens, 2005). However, spherical harmonics are not the only basis set available to describe the angular dependence of the deformation density.

We first present a formulation of anisotropic and aspherical atomic densities based on Cartesian Gaussian multipoles, which leads to much simpler formulae for the calculation of structure factors *via* direct summation in reciprocal space than the STO-based theory of Hansen & Coppens (1978). We also demonstrate that Cartesian Gaussian multipoles allow the computation of structure factors *via* fast Fourier transformation (FFT) of the real-space electron density (Cooley & Tukey, 1965). The latter approach, originally proposed by Ten Eyck (1973, 1977), is the basis of the efficient macromolecular refinement algorithms (Brünger, 1989; Afonine & Urzhumtsev, 2004; Afonine *et al.*, 2007; Agarwal, 1978) implemented in programs such as *CNS* (Brünger *et al.*, 1998; Brunger, 2007) and *PHENIX* (Adams *et al.*, 2002). The sublattice method implemented in *CNS* lends itself to efficient parallelization (Brünger, 1989).

Boys originally proposed Cartesian Gaussian functions as basis functions to solve the many-electron Schrödinger equation (Boys, 1950). The advantage of Gaussians over STOs in this context is that two-electron integrals have analytic forms, which has led to the adoption of Gaussian basis sets for many *ab initio* calculations (Hehre *et al.*, 1969, 1970). We note that the equivalence of spherical harmonics and Cartesian

tensors is well known, with key relationships having been presented by Stone (1996) and Applequist (1989, 2002).

We apply Cartesian Gaussian multipoles to restrained crystallographic refinements based on the Atomic Multipole Optimized Energetics for Biomolecular Applications (AMOEBA) force-field electrostatic model (Ponder & Case, 2003; Ren & Ponder, 2002, 2003, 2004; Schnieders *et al.*, 2007; Schnieders & Ponder, 2007). The AMOEBA electrostatic model is based on the superposition of permanent atomic multipoles truncated at quadrupoles and induced dipoles. Permanent electrostatics represents the electron density of a group of atoms in the absence of interactions with the environment, which may include other parts of the molecule or solvent. Groups are chosen to be relatively rigid in order to avoid conformational variability in the permanent multipole moments. Conversely, the induced dipoles of AMOEBA represent polarization, the response of the electron density to the local electric field.

Force fields are widely used to restrain macromolecular refinement by contributing forces to local optimizations and molecular dynamics (Brünger *et al.*, 1987), with the latter used within simulated-annealing algorithms to promote global optimization (Brünger, 1988, 1991; Brünger *et al.*, 1989, 1990, 1997; Kuriyan *et al.*, 1989; Adams *et al.*, 1997; Brünger & Rice, 1997). Up to now, force fields in crystallography have been largely limited to the geometric and repulsive terms and have had no influence on the atomic scattering factors. Therefore, refinement using a scattering model based on AMOEBA electrostatics is novel and lends insight into the progress being made in the development of precise, transferable force fields. Another limitation of the use of force fields for restraining X-ray refinement has been the lack of proper treatment of long-range electrostatic interactions, which is overcome in this work *via* use of particle-mesh Ewald summation (PME; Darden *et al.*, 1993; Essmann *et al.*, 1995; Sagui *et al.*, 2004).

In addition to AMOEBA, polarizable force fields are being studied by a number of other groups. Maple and coworkers have pursued a model similar to AMOEBA, but with the permanent moments truncated at dipole order, which has shown promising results for protein–ligand complexes (Friesner *et al.*, 2005; Maple *et al.*, 2005). As an alternative to induced dipoles, Patel and Brooks employed a fluctuating-charge model of polarization (Patel & Brooks, 2006), while Lamoureux and Roux have demonstrated success using classical Drude oscillators (Lamoureux *et al.*, 2006; Lamoureux & Roux, 2003). In addition to polarization, Gresh and coworkers have developed a methodology to include nonclassical effects such as electrostatic penetration and charge transfer (Gresh, 2006; Gresh *et al.*, 2007; Piquemal *et al.*, 2006, 2007).

Although classical potentials can be validated against a range of experimental observables, for example small-molecule solvation energies (Shirts *et al.*, 2003; Shirts & Pande, 2005), high-resolution diffraction data can pinpoint deficiencies in an electrostatics model with high precision. For example, we show that truncation of permanent atomic multipoles at quadrupole order limits the ability of the AMOEBA model to place charge density at bond midpoints.

research papers

We use an efficient solution to this limitation by refining partial charges at bond centers as originally proposed by Afonine *et al.* (2007).

2. Theory

2.1. Subgrid fast Fourier transform

The starting point for this work is the subgrid fast Fourier transform algorithm (SGFFT), which will be briefly summarized (Brünger, 1989). In FFT-based methods, the electron density is computed over a lattice chosen to be fine enough to avoid aliasing effects at a given resolution. This computation can be made more efficient by an artificial increase in the atomic displacement parameters (ADPs) of all atoms. The optimum choice in *CNS* v.1.2 (Brunger, 2007) for the ADP offset and grid size follows the work of Bricogne (2001). An important point is that the electron density is only computed within a cutoff radius around each atom. As the resolution increases, the cutoff is increased based on an empirical scheme to maintain agreement between direct-summation structure factors and derivatives and the SGFFT calculation (Brunger, 2007).

Structure factors are computed by FFT of the electron density of an asymmetric unit of atoms (Agarwal, 1978). The SGFFT is based on factorizing this computation into smaller FFTs that are computed separately on sublattices, which allows efficient parallelization since these tasks are independent (Brünger, 1989; Kay Diederichs, private communication). *CNS* v.1.21 has implemented this approach *via* an OpenMP environment (courtesy Kay Diederichs, University of Konstanz; available at http://cns-online.org). Crystallographic symmetry is then applied to the structure factors, and the target function and its derivatives with respect to structure factors are evaluated. Symmetry operators are applied to the derivatives of the target function with respect to the structure factors followed by inverse Fourier transform. Using the chain rule, derivatives of the target function with respect to atomic parameters are then computed by multiplication and summation over the local neighborhood around each atom of the derivatives of the electron density with respect to atomic parameters.

Although the original SGFFT method was developed with an isolated-atom description of electron density and isotropic ADPs, it is generalizable to aspherical Cartesian Gaussian multipoles and anisotropic ADPs. All that is needed are formulae for the electron density and the derivatives of the electron density with respect to atomic parameters, which then can be inserted into equations (29) and (40) of Brünger (1989). In the following sections, we develop these necessary formulae.

2.2. Isolated-atom Gaussian density

The key mathematical property of Gaussians with respect to efficient calculation of structure factors is that they are an eigenfunction of the Fourier transform (FT). In other words, a Gaussian in real space transforms to a Gaussian in reciprocal space and *vice versa*. Consider the canonical spherically symmetric Gaussian atomic scattering factor (Agarwal, 1978),

$$f^{(n,\kappa)}(\mathbf{r}) = \kappa^3 (4\pi)^{3/2} \sum_{i=1}^{n} \frac{a_i}{b_i^{3/2}} \exp\left(-\frac{4\pi^2\kappa^2|\mathbf{r}|^2}{b_i}\right), \qquad (1)$$

where a_i and b_i are constant parameters fitted to *ab initio* calculations on isolated atoms (this work is based on a sum of six Gaussians; $n = 6$; Su & Coppens, 1998), κ is an expansion/contraction parameter used to adjust the width of the density and $\mathbf{r}$ is a position vector relative to the center of the atom. Its FT is given by

$$\hat{f}^{(n,\kappa)}(\mathbf{s}) = \sum_{i=1}^{n} a_i \exp\left(-\frac{b_i|\mathbf{s}|^2}{4\kappa^2}\right), \qquad (2)$$

where $\mathbf{s}$ is the reciprocal-lattice vector and we have used the FT definition given in Appendix *A*. The reciprocal-lattice vector is $\mathbf{s} = \mathbf{h}^t\mathbf{A}^{-1} = (\mathbf{A}^{-1})^t\mathbf{h}$, where $\mathbf{h}$ is a column vector with the Miller indices of a Bragg reflection and $\mathbf{A}$ is the fractionalization matrix that transforms coordinates $\mathbf{r}$ with respect to a Cartesian basis to fractional coordinates $\mathbf{r}_{\text{frac}}$ as defined in a crystallographic basis set. The Debye–Waller factor (Waller, 1923) is given by

$$\hat{t}(\mathbf{s}) = \exp(-2\pi^2\mathbf{s}^t\mathbf{U}\mathbf{s}) \qquad (3)$$

in reciprocal space, where each element of the symmetric positive-definite matrix $\mathbf{U}$ is defined *via* a Cartesian basis consistent with PDB ANISOU records (Trueblood *et al.*, 1996; Grosse-Kunstleve & Adams, 2002). Multiplication of (3) by the atomic form factor from (2) gives the scattering factor

$$\hat{\rho}^{(n,\kappa)}(\mathbf{s}) = \hat{f}^{(n,\kappa)}(\mathbf{s})\hat{t}(\mathbf{s}) = \sum_{i=1}^{n} a_i \exp(-2\pi^2\mathbf{s}^t\mathbf{U}_i\mathbf{s}) \qquad (4)$$

based on $\mathbf{U}_i$ that are defined by

$$\mathbf{U}_i = \begin{pmatrix} U_{11} & U_{12} & U_{13} \\ U_{21} & U_{22} & U_{23} \\ U_{31} & U_{32} & U_{33} \end{pmatrix} + I_3\left(\frac{b_i}{8\pi^2\kappa^2} + U_{\text{add}}\right), \qquad (5)$$

where U_{add} is the artificial isotropic increase or decrease in the ADP discussed above and I_3 is a 3×3 identity matrix. Removal of U_{add} analytically from each structure factor after the FT is straightforward. The only difference, therefore, between each $\mathbf{U}_i$ is the isolated-atom scattering parameter b_i.

Application of the inverse FT to (4) gives the real-space anisotropic electron density

$$\rho^{(n,\kappa)}(\mathbf{r}) = (2\pi)^{-3/2} \sum_{i=1}^{n} a_i |\mathbf{U}_i|^{-1/2} \exp(-\tfrac{1}{2}\mathbf{r}^t\mathbf{U}_i^{-1}\mathbf{r}), \qquad (6)$$

where $|\mathbf{U}_i|$ is the determinant of matrix $\mathbf{U}_i$ and $\mathbf{U}_i^{-1}$ is its inverse. This expression can also be viewed as the convolution of the Gaussian form factor of (1) with the inverse Fourier transform of the Debye–Waller factor of (3). Although the underlying isolated-atom scattering factor is spherically symmetric, convolution with anisotropic ADPs can lead to an angular dependence in $\rho^{(n,\kappa)}(\mathbf{r})$. Using the relationship that $B = 8\pi^2 U$, one can show that (6) reduces to the isotropic density expression reported by Brünger in equation (16) of

Brünger (1989) if all diagonal elements of $\mathbf{U}_i$ are equal to $U_{\text{iso}} + b_i/8\pi^2 + U_{\text{add}}$ with zero off-diagonal components.

2.3. Polarizable atomic multipole electron density

For the derivation of an atomic multipole expansion from a collection of point charges we begin with the Taylor expansion of the electric potential $V(\mathbf{r})$ at $\mathbf{r}$ arising from n partial point charges that represent the electron density of an atom,

$$
\begin{aligned}
V(\mathbf{r}) &= \sum_{i=1}^{n} \frac{c_i}{|\mathbf{r} - \Delta_i|} \\
&= \sum_{i=1}^{n} c_i \left[\frac{1}{r} + \Delta_{i,\alpha} \left(\frac{\partial}{\partial \Delta_{i,\alpha}} \frac{1}{|\mathbf{r} - \Delta_i|} \right)_{\Delta_i = 0} \right. \\
&\quad \left. + \frac{1}{2} \Delta_{i,\alpha} \Delta_{i,\beta} \left(\frac{\partial}{\partial \Delta_{i,\alpha}} \frac{\partial}{\partial \Delta_{i,\beta}} \frac{1}{|\mathbf{r} - \Delta_i|} \right)_{\Delta_i = 0} + \cdots \right] \\
&= \sum_{i=1}^{n} c_i \left[\frac{1}{r} - \Delta_{i,\alpha} \frac{\partial}{\partial r_\alpha} \frac{1}{r} + \frac{1}{2} \Delta_{i,\alpha} \Delta_{i,\beta} \frac{\partial}{\partial r_\alpha} \frac{\partial}{\partial r_\beta} \frac{1}{r} - \cdots \right] \\
&= \sum_{i=1}^{n} \left(1 - \Delta_{i,\alpha} \nabla_\alpha + \frac{1}{2} \Delta_{i,\alpha} \Delta_{i,\beta} \nabla_\alpha \nabla_\beta - \cdots \right) \frac{c_i}{r},
\end{aligned} \tag{7}
$$

where Δ_i is the position of partial charge c_i, $\nabla_\alpha = \partial/\partial r_\alpha$ is one component of the del operator, $\alpha \in \{x, y, z\}$ and the Greek subscripts $\{\alpha, \beta\}$ represent the use of the Einstein summation convention for summing over tensor elements (Stone, 1996). We omit the constant factor of $1/4\pi\varepsilon_0$ throughout for compactness. Let the monopole, dipole and traceless quadrupole moments be defined as

$$
\begin{aligned}
q &= \sum_{i=1}^{n} c_i, \\
d_\alpha &= \sum_{i=1}^{n} c_i \Delta_{i,\alpha}, \\
\Theta_{\alpha\beta} &= \frac{3}{2} \sum_{i=1}^{n} c_i (\Delta_{i,\alpha} \Delta_{i,\beta} - \frac{1}{3} \Delta_i^2 \delta_{\alpha\beta}),
\end{aligned} \tag{8}
$$

where removal of the trace in the definition of the quadrupole moment is allowed because the potential satisfies the Laplace equation (i.e. $\nabla^2 V = 0$). Substitution of the relationships in (8) into the final expression of (7) gives the electric potential in terms of a Cartesian multipole expansion, which we truncate at quadrupole order

$$
V(\mathbf{r}) = (q - d_\alpha \nabla_\alpha + \tfrac{1}{3} \Theta_{\alpha\beta} \nabla_\alpha \nabla_\beta) \frac{1}{r}. \tag{9}
$$

We now replace the Coulomb potential of (9) with the potential from the sum of Gaussians from (1), which is given by

$$
\varphi^{(n,\kappa)}(r) = \sum_{i=1}^{n} a_i \frac{\text{erf}(2\pi\kappa r/b_i^{1/2})}{r} \tag{10}
$$

and find

$$
\Phi(\mathbf{r}) = (q - d_\alpha \nabla_\alpha + \tfrac{1}{3} \Theta_{\alpha\beta} \nabla_\alpha \nabla_\beta) \varphi^{(n,\kappa)}(r). \tag{11}
$$

We now introduce unique superscripts on the charge, dipole and quadrupole Gaussian basis sets, denoted by $\{n_q, n_d, n_\Theta\}$ and $\{\kappa_q, \kappa_d, \kappa_\Theta\}$, to allow them to differ in number and width.

$$
\Phi(\mathbf{r}) = q\varphi^{(n_q,\kappa_q)}(r) - d_\alpha \nabla_\alpha \varphi^{(n_d,\kappa_d)}(r) + \tfrac{1}{3} \Theta_{\alpha\beta} \nabla_\alpha \nabla_\beta \varphi^{(n_\Theta,\kappa_\Theta)}(r). \tag{12}
$$

The potential of the charge density of (12) quickly approaches the Coulomb potential as r increases since the error function goes to unity such that at large r this potential satisfies the Laplace equation and the use of a traceless quadrupole tensor is still justified. Application of the Laplace operator to both sides of (12) gives the negative of a continuous charge density based on Cartesian Gaussian multipoles,

$$
\rho(\mathbf{r}) = -qf^{(n_q,\kappa_q)}(\mathbf{r}) + d_\alpha \nabla_\alpha f^{(n_d,\kappa_d)}(\mathbf{r}) - \tfrac{1}{3} \Theta_{\alpha\beta} \nabla_\alpha \nabla_\beta f^{(n_\Theta,\kappa_\Theta)}(\mathbf{r}). \tag{13}
$$

In crystallography the convention is that electron density is positive, so we will keep the negative sign. Therefore, a negative partial charge equates to positive scattering density.

Inclusion of ADPs is described by convolution of (13) with the real-space temperature factor,

$$
\rho_{\text{ADP}}(\mathbf{r}) = \rho(\mathbf{r}) * \text{t}(\mathbf{r}). \tag{14}
$$

Based on the convolution differentiation rule

$$
[\nabla_\alpha \rho(\mathbf{r})] * t(\mathbf{r}) \equiv \nabla_\alpha [\rho(\mathbf{r}) * t(\mathbf{r})] \tag{15}
$$

the solution to (14) is given by substituting for $f(\mathbf{r})$ in (13) with the corresponding $\rho(\mathbf{r})$ from (6) to give

$$
\begin{aligned}
\rho_{\text{ADP}}(\mathbf{r}) &= -q\rho^{(n_q,\kappa_q)}(\mathbf{r}) + d_\alpha \nabla_\alpha \rho^{(n_d,\kappa_d)}(\mathbf{r}) \\
&\quad - \tfrac{1}{3} \Theta_{\alpha\beta} \nabla_\alpha \nabla_\beta \rho^{(n_\Theta,\kappa_\Theta)}(\mathbf{r}).
\end{aligned} \tag{16}
$$

However, since q only represents partial atomic charges, the contributions from valence and core electrons need to be added. Additionally, the AMOEBA force field divides each atomic dipole moment into permanent ($\mathbf{d}$) and induced ($\mathbf{u}$) contributions to account for polarization. Therefore, we construct the total atomic electron density at a location $\mathbf{r}$ relative to the center of atom j by adding the contribution of core and valence electron density to (16) and splitting the dipole into permanent and induced components to give

$$
\begin{aligned}
\rho_{\text{ADP},j}(\mathbf{r}) &= P_j^{(c)} \rho_j^{(6,1)}(\mathbf{r}) + [P_j^{(v)} - q_j] \rho_j^{(6,\kappa_v)}(\mathbf{r}) \\
&\quad + (d_{j,\alpha} + u_{j,\alpha}) \nabla_\alpha \rho_j^{(1,\kappa_d)}(\mathbf{r}) - \tfrac{1}{3} \Theta_{j,\alpha\beta} \nabla_\alpha \nabla_\beta \rho_j^{(1,\kappa_\Theta)}(\mathbf{r}),
\end{aligned} \tag{17}
$$

where $P_j^{(c)}$ is the integer number of core electrons (carbon has two) and $P_j^{(v)}$ is the integer number of valence electrons (carbon has four). The superscripts on the anisotropic Gaussian form factors $\rho_j^{(n,\kappa)}(\mathbf{r})$ have been made explicit for our model. We make the reasonable choice of using the isolated-atom scattering parameters for both core and valence electron densities. The width of the core electron density is frozen at the isolated-atom description ($\kappa = 1$) based on the observation that chemical bonding does not perturb it significantly (Hansen & Coppens, 1978). On the other hand, the width of the valence electron density expands or contracts relative to the isolated-atom model owing to a gain or reduction, respectively, of electron density from or to covalently bonded atoms. This effect is modeled by the width parameter of the valence density κ_v. In this work, the dipole and quadrupole

densities are described by a single Gaussian ($n_d = n_\Theta = 1$) based on a and b parameters set to unity. The widths of the dipole and quadrupole densities are controlled by the κ_d and κ_Θ parameters. In this work, the width parameters $\{\kappa_v, \kappa_d, \kappa_\Theta\}$ are optimized against the diffraction data for each AMOEBA multipole type. The multipole moments are fixed by the AMOEBA force field and are not refined against the data.

The partial derivatives through second order of the aniso-tropic and aspherical density defined in (6), which are required for the real-space multipolar density given in (17), are

$$
\nabla_\alpha \rho^{(n,\kappa)}(\mathbf{r}) = -(2\pi)^{-3/2} \sum_{i=1}^{n} a_i |\mathbf{U}_i|^{-1/2}
$$
$$
\times \exp(-\tfrac{1}{2} \mathbf{r}^t \mathbf{U}_i^{-1} \mathbf{r})(\mathbf{r}^t \mathbf{U}_i^{-1} \mathbf{u}_\alpha)
$$
$$
\nabla_\alpha \nabla_\beta \rho^{(n,\kappa)}(\mathbf{r}) = (2\pi)^{-3/2} \sum_{i=1}^{n} a_i |\mathbf{U}_i|^{-1/2}
$$
$$
\times \exp(-\tfrac{1}{2} \mathbf{r}^t \mathbf{U}_i^{-1} \mathbf{r})[(\mathbf{r}^t \mathbf{U}_i^{-1} \mathbf{u}_\alpha)(\mathbf{r}^t \mathbf{U}_i^{-1} \mathbf{u}_\beta) - U_{i,\alpha\beta}^{-1}], \tag{18}
$$

where $\mathbf{u}_\alpha$ is a unit vector in the α direction with $\alpha \in \{x, y, z\}$. In addition, the third-, fourth- and fifth-order terms of the expansion are presented as supplementary information along with a Mathematica notebook.[1]

To the best of our knowledge, (17) is the first expression reported in the literature for a real-space form factor that is the convolution of an atomic multipolar electron density with anisotropic ADPs. This equation opens the door to exploring precise polarizable atomic multipole refinements in tandem with efficient computation of structure factors *via* FFT.

Given a molecular conformation, the AMOEBA perma-nent multipole moments for each atom in the global coordi-nate frame (q, $\mathbf{d}$, Θ) are converted *via* rotation from a local frame. For example, as shown in Fig. 1, the z axis of the local frame for the carbonyl O atom of the peptide bond is in the direction of the bond to the carbonyl C atom. Its positive x axis is located in the O=C−C$^\alpha$ plane in the direction of the C$^\alpha$ atom and the y axis is chosen to give a right-handed coordinate system (Ren & Ponder, 2002). The induced dipole ($\mathbf{u}$) on each atom is determined *via* a self-consistent field (SCF) calculation, where the field is a sum of contributions from the permanent atomic multipoles and induced dipoles. The AMOEBA polarization model is described in greater detail in work by Ren & Ponder (2002).

2.4. Derivatives of the electron density

2.4.1. Atomic coordinates. As a simplification, the deriva-tion up to this point has assumed that the atomic center was the origin of the coordinate system. However, for this section on the derivatives with respect to atomic coordinates we place atom j at $\mathbf{r}_j$ in the global frame. In order to keep the derivation manageable, we split the total electron density into that produced by permanent charges ρ_{perm} and that of induced charges ρ_{ind},

$$
\rho_{\text{total}}(\mathbf{r}) = \sum_{j=1}^{n} \rho_{\text{perm},j}(\mathbf{r} - \mathbf{r}_j) + \rho_{\text{ind},j}(\mathbf{r} - \mathbf{r}_j). \tag{19}
$$

The derivative of the permanent multipole electron density of atom j with respect to the α coordinate of atom j is given by

$$
\frac{\partial \rho_{\text{perm},j}(\mathbf{r} - \mathbf{r}_j)}{\partial r_{j,\alpha}} = P_j^{(c)} \frac{\partial \rho_j^{(6,1)}(\mathbf{r} - \mathbf{r}_j)}{\partial r_{j,\alpha}}
$$
$$
+ \left[P_j^{(v)} - q_j \right] \frac{\partial \rho_j^{(6,\kappa_v)}(\mathbf{r} - \mathbf{r}_j)}{\partial r_{j,\alpha}} + d_{j,\beta} \frac{\partial [\nabla_\beta \rho_j^{(1,\kappa_d)}(\mathbf{r} - \mathbf{r}_j)]}{\partial r_{j,\alpha}}
$$
$$
+ \frac{\partial d_{j,\beta}}{\partial r_{j,\alpha}} \nabla_\beta \rho_j^{(1,\kappa_d)}(\mathbf{r} - \mathbf{r}_j) - \frac{1}{3} \Theta_{j,\beta\gamma} \frac{\partial [\nabla_\beta \nabla_\gamma \rho_j^{(1,\kappa_\Theta)}(\mathbf{r} - \mathbf{r}_j)]}{\partial r_{j,\alpha}}
$$
$$
- \frac{1}{3} \frac{\partial \Theta_{j,\beta\gamma}}{\partial r_{j,\alpha}} \nabla_\beta \nabla_\gamma \rho_j^{(1,\kappa_\Theta)}(\mathbf{r} - \mathbf{r}_j), \tag{20}
$$

where the derivative of the dipole and quadrupole densities are each composed of two terms owing to the chain rule. As described above, the dipole and quadrupole moments of each atom are implicitly a function of its coordinates and the coordinates of a few of its bonded neighbors (atoms k) that define the local frame of the multipole. Therefore, the deri-vative of the permanent multipole electron density of atom j with respect to the α coordinate of atoms k must also be considered,
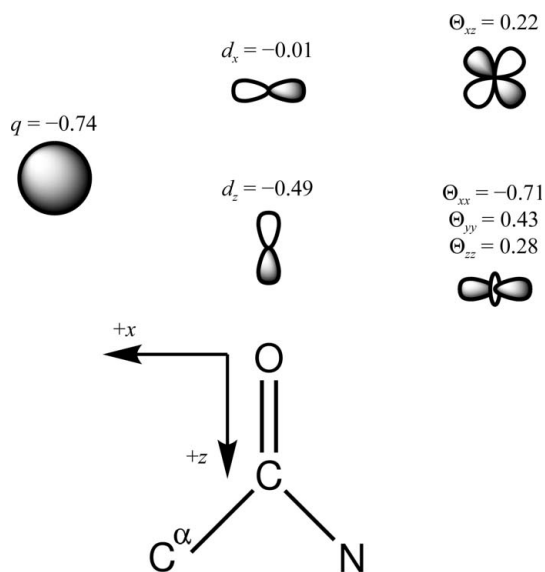
Figure 1
The local multipole frame of the carbonyl O atom of the peptide backbone is shown. The positive z axis is along the C=O bond and the x axis is chosen in the O=C−C$^\alpha$ plane in the direction of the C$^\alpha$ atom. The y axis is directed into the page in order to achieve a right-handed coordinate system. Also shown are the nonzero multipole moments of the O atom and a qualitative representation of their shape. The d_z Cartesian Gaussian dipole (in Debye units) places electron density along the C=O bond, while the trace of the Cartesian Gaussian quadrupole (in Buckingham units) positions electron density approximately at lone-pair positions.

$$\frac{\partial \rho_{\mathrm{perm},j}(\mathbf{r} - \mathbf{r}_j)}{\partial r_{k,\alpha}} = \frac{\partial d_{j,\beta}}{\partial r_{k,\alpha}} \nabla_\beta \rho_j^{(1,\kappa_d)}(\mathbf{r} - \mathbf{r}_j)$$
$$- \frac{1}{3} \frac{\partial \Theta_{j,\beta\gamma}}{\partial r_{k,\alpha}} \nabla_\beta \nabla_\gamma \rho_j^{(1,\kappa_\Theta)}(\mathbf{r} - \mathbf{r}_j), \quad (21)$$

where the derivatives of spherically symmetric terms are zero with respect to the coordinates of atom k because they have no dependence on the orientation of the local frame. Note that the partial derivative of an anisotropic and aspherical density tensor with respect to an atomic coordinate is the negative of the partial derivatives given in (18), simply due to the negative sign on $\mathbf{r}_j$. The derivatives of the polarizable density with respect to atomic coordinates are very specific to the AMOEBA electrostatic model and are discussed in Appendix B. However, we note that computing the derivatives of a polarizable density with respect to atomic coordinates is $O(n^2 \log n)$ using PME, which quickly becomes the most expensive part of the overall calculation.

2.4.2. ADPs. The derivative of the anisotropic electron density of atom j with respect to an anisotropic displacement parameter $U_{j,\tau\upsilon}$ is given by

$$\frac{\partial \rho_j(\mathbf{r})}{\partial U_{j,\tau\upsilon}} = P_j^{(c)} \frac{\partial \rho_j^{(6,1)}(\mathbf{r})}{\partial U_{j,\tau\upsilon}} + [P_j^{(v)} - q_j] \frac{\partial \rho_j^{(6,\kappa_v)}(\mathbf{r})}{\partial U_{j,\tau\upsilon}}$$
$$+ (d_{j,\alpha} + u_{j,\alpha}) \frac{\partial [\nabla_\alpha \rho_j^{(1,\kappa_d)}(\mathbf{r})]}{\partial U_{j,\tau\upsilon}}$$
$$- \frac{1}{3} \Theta_{j,\alpha\beta} \frac{\partial [\nabla_\alpha \nabla_\beta \rho_j^{(1,\kappa_\Theta)}(\mathbf{r})]}{\partial U_{j,\tau\upsilon}} \quad (22)$$

and requires the partial derivatives of the Cartesian Gaussian tensors with respect to ADP components. Introducing a few relationships facilitates their presentation. Firstly, based on the equality

$$\frac{\partial |\mathbf{U}|}{\partial U_{\tau\upsilon}} = |\mathbf{U}| U_{\tau\upsilon}^{-1} (2 - \delta_{\tau\upsilon}) \quad (23)$$

we have

$$\frac{\partial |\mathbf{U}|^{-1/2}}{\partial U_{\tau\upsilon}} = -\frac{1}{2} |\mathbf{U}|^{-1/2} U_{\tau\upsilon}^{-1} (2 - \delta_{\tau\upsilon}), \quad (24)$$

where the Kronecker delta $\delta_{\tau\upsilon}$ is unity for diagonal elements of $\mathbf{U}$ and zero otherwise. Differentiating an identity from matrix algebra $\mathbf{U}^{-1}\mathbf{U} = \mathbf{I}$ gives the following relationship

$$\frac{\partial \mathbf{U}^{-1}}{\partial U_{\tau\upsilon}} = -\mathbf{U}^{-1} \frac{\partial \mathbf{U}}{\partial U_{\tau\upsilon}} \mathbf{U}^{-1}, \quad (25)$$

which makes it possible to differentiate $\mathbf{U}$ instead of its inverse. This is preferred since only one or two elements of $\partial \mathbf{U}/\partial U_{\tau\upsilon}$ are equal to unity and the rest are zero. Specifically, a single element is equal to unity if τ equals υ, while two elements are equal to unity otherwise, since $U_{\tau\upsilon}$ and $U_{\upsilon\tau}$ represent the same variable in this case. For convenience, we define a 3×3 matrix $\mathbf{J}^{(\tau\upsilon)}$,

$$\mathbf{J}^{(\tau\upsilon)} = -\mathbf{U}^{-1} \frac{\partial \mathbf{U}}{\partial U_{\tau\upsilon}} \mathbf{U}^{-1}, \quad (26)$$

and based on the chain rule we have

$$\frac{\partial}{\partial U_{\tau\upsilon}} \exp(-\tfrac{1}{2}\mathbf{r}^t \mathbf{U}^{-1}\mathbf{r}) = -\tfrac{1}{2}\mathbf{r}^t \mathbf{J}^{(\tau\upsilon)}\mathbf{r} \exp(-\tfrac{1}{2}\mathbf{r}^t \mathbf{U}^{-1}\mathbf{r}). \quad (27)$$

Differentiating (6) with respect to $\mathbf{U}_{\tau\upsilon}$ and using (24), (27) and the product rule gives

$$\frac{\partial \rho^{(n,\kappa)}(\mathbf{r})}{\partial U_{\tau\upsilon}} = (2\pi)^{-3/2} \sum_{i=1}^{n} a_i |\mathbf{U}_i|^{-1/2}$$
$$\times \exp(-\tfrac{1}{2}\mathbf{r}^t \mathbf{U}_i^{-1}\mathbf{r})\tfrac{1}{2}[-\mathbf{r}^t \mathbf{J}_i^{(\tau\upsilon)}\mathbf{r} - U_{i,\tau\upsilon}^{-1}(2 - \delta_{\tau\upsilon})]$$
$$\frac{\partial [\nabla_\alpha \rho^{(n,\kappa)}(\mathbf{r})]}{\partial U_{\tau\upsilon}} = -(2\pi)^{-3/2} \sum_{i=1}^{n} a_i |\mathbf{U}_i|^{-1/2} \exp(-\tfrac{1}{2}\mathbf{r}^t \mathbf{U}_i^{-1}\mathbf{r})$$
$$\times \{\tfrac{1}{2}[-\mathbf{r}^t \mathbf{J}_i^{(\tau\upsilon)}\mathbf{r} - U_{i,\tau\upsilon}^{-1}(2 - \delta_{\tau\upsilon})](\mathbf{r}^t \mathbf{U}_i^{-1}\mathbf{u}_\alpha) + (\mathbf{r}^t \mathbf{J}_i^{(\tau\upsilon)}\mathbf{u}_\alpha)\}$$
$$\frac{\partial [\nabla_\alpha \nabla_\beta \rho^{(n,\kappa)}(\mathbf{r})]}{\partial U_{\tau\upsilon}} = (2\pi)^{-3/2} \sum_{i=1}^{n} a_i |\mathbf{U}_i|^{-1/2} \exp(-\tfrac{1}{2}\mathbf{r}^t \mathbf{U}_i^{-1}\mathbf{r})$$
$$\times \{\tfrac{1}{2}[-\mathbf{r}^t \mathbf{J}_i^{(\tau\upsilon)}\mathbf{r} - U_{i,\tau\upsilon}^{-1}(2 - \delta_{\tau\upsilon})][(\mathbf{r}^t \mathbf{U}_i^{-1}\mathbf{u}_\alpha)(\mathbf{r}^t \mathbf{U}_i^{-1}\mathbf{u}_\beta) - U_{i,\alpha\beta}^{-1}]$$
$$+ [\mathbf{r}^t \mathbf{J}_i^{(\tau\upsilon)}\mathbf{u}_\alpha](\mathbf{r}^t \mathbf{U}_i^{-1}\mathbf{u}_\beta) + (\mathbf{r}^t \mathbf{U}_i^{-1}\mathbf{u}_\alpha)[\mathbf{r}^t \mathbf{J}_i^{(\tau\upsilon)}\mathbf{u}_\beta] - J_{i,\alpha\beta}^{(\tau\upsilon)}\}. \quad (28)$$

2.4.3. Gaussian width. The Gaussian width parameter κ controls radial expansion and contraction of the Cartesian Gaussian multipoles. Analogous parameters are used to optimize the STOs within the Hansen and Coppens scattering model (Hansen & Coppens, 1978). The derivative of the electron density with respect to this parameter is similar to the gradient for the ADP parameters. Two chain-rule terms are necessary. Firstly, the gradient of the normalizing term

$$\frac{\partial}{\partial \kappa}(|\mathbf{U}_i|^{-1/2}) = -\frac{1}{2}|\mathbf{U}_i|^{-3/2}\frac{\partial |\mathbf{U}_i|}{\partial \kappa}, \quad (29)$$

where

$$\frac{\partial |\mathbf{U}_i|}{\partial \kappa} = -\frac{3b_i^3}{256\pi^6\kappa^7} - \frac{b_i^2(U_{11} + U_{22} + U_{33} + 3U_{\mathrm{add}})}{16\pi^4\kappa^5}$$
$$+ \{b_i[U_{12}^2 + U_{13}^2 + U_{23}^2 - U_{11}U_{22} - U_{11}U_{33} - U_{22}U_{33}$$
$$- 2U_{\mathrm{add}}(U_{11} + U_{22} + U_{33}) - 3U_{\mathrm{add}}^2]\}/4\pi^2\kappa^3. \quad (30)$$

Secondly, the gradient of the inverse ADP matrix is most conveniently expressed using the gradient of the original ADP matrix,

$$\frac{\partial \mathbf{U}_i^{-1}}{\partial \kappa} = -\mathbf{U}_i^{-1}\frac{\partial \mathbf{U}_i}{\partial \kappa}\mathbf{U}_i^{-1}, \quad (31)$$

where

$$\frac{\partial \mathbf{U}_i}{\partial \kappa} = -\frac{b_i}{4\pi^2\kappa^3}\mathbf{I}_3. \quad (32)$$

For convenience the matrix $\mathbf{J}_i^{(\kappa)}$ is defined to more compactly represent this result,

$$\mathbf{J}_i^{(\kappa)} = \frac{b_i}{4\pi^2\kappa^3}\mathbf{U}_i^{-1}\mathbf{U}_i^{-1}. \quad (33)$$

Differentiating (6) with respect to κ and using (29), (33) and the product rule gives

$$\frac{\partial \rho^{(n,\kappa)}(\mathbf{r})}{\partial \kappa} = (2\pi)^{-3/2} \sum_{i=1}^{n} a_i |\mathbf{U}_i|^{-1/2} \exp(-\tfrac{1}{2}\mathbf{r}^t \mathbf{U}_i^{-1} \mathbf{r})$$

$$\times \frac{1}{2}\left[-\mathbf{r}^t \mathbf{J}_i^{(\kappa)} \mathbf{r} - |\mathbf{U}_i|^{-1} \frac{\partial |\mathbf{U}_i|}{\partial \kappa} \right],$$

$$\frac{\partial [\nabla_\alpha \rho^{(n,\kappa)}(\mathbf{r})]}{\partial \kappa} = -(2\pi)^{-3/2} \sum_{i=1}^{n} a_i |\mathbf{U}_i|^{-1/2} \exp(-\tfrac{1}{2}\mathbf{r}^t \mathbf{U}_i^{-1} \mathbf{r})$$

$$\times \left\{ \frac{1}{2}\left[-\mathbf{r}^t \mathbf{J}_i^{(\kappa)} \mathbf{r} - |\mathbf{U}_i|^{-1} \frac{\partial |\mathbf{U}_i|}{\partial \kappa} \right] \right.$$

$$\left. \times (\mathbf{r}^t \mathbf{U}_i^{-1} \mathbf{u}_\alpha) + [\mathbf{r}^t \mathbf{J}_i^{(\kappa)} \mathbf{u}_\alpha] \right\},$$

$$\frac{\partial [\nabla_\alpha \nabla_\beta \rho^{(n,\kappa)}(\mathbf{r})]}{\partial \kappa} = (2\pi)^{-3/2} \sum_{i=1}^{n} a_i |\mathbf{U}_i|^{-1/2} \exp(-\tfrac{1}{2}\mathbf{r}^t \mathbf{U}_i^{-1} \mathbf{r})$$

$$\times \left\{ \frac{1}{2}\left[-\mathbf{r}^t \mathbf{J}_i^{(\kappa)} \mathbf{r} - |\mathbf{U}_i|^{-1} \frac{\partial |\mathbf{U}_i|}{\partial \kappa} \right] \right.$$

$$\times [(\mathbf{r}^t \mathbf{U}_i^{-1} \mathbf{u}_\alpha)(\mathbf{r}^t \mathbf{U}_i^{-1} \mathbf{u}_\beta) - U_{i,\alpha\beta}^{-1}]$$

$$+ [\mathbf{r}^t \mathbf{J}_i^{(\kappa)} \mathbf{u}_\alpha](\mathbf{r}^t \mathbf{U}_i^{-1} \mathbf{u}_\beta)$$

$$\left. + (\mathbf{r}^t \mathbf{U}_i^{-1} \mathbf{u}_\alpha)[\mathbf{r}^t \mathbf{J}_i^{(\kappa)} \mathbf{u}_\beta] - J_{i,\alpha\beta}^{(\tau\upsilon)} \right\}, \quad (34)$$

together with the third- and fourth-order terms available as supplementary information.

2.5. Fourier transform of the polarizable atomic multipole electron density

Remarkably, the FT of the anisotropic and aspherical density given in (17) is simply

$$\hat{\rho}_{\mathrm{ADP},j}(\mathbf{s}) = \{ P_j^{(c)} \hat{f}_j^{(6,1)}(\mathbf{s}) + [P_j^{(\nu)} - q_j] \hat{f}_j^{(6,\kappa_\nu)}(\mathbf{s})$$

$$- (d_{j,\alpha} + u_{j,\alpha}) 2\pi i s_\alpha \hat{f}_j^{(1,\kappa_d)}(\mathbf{s})$$

$$+ \tfrac{1}{3} \Theta_{j,\alpha\beta} 4\pi^2 s_\alpha s_\beta \hat{f}_j^{(1,\kappa_\Theta)}(\mathbf{s}) \} \hat{t}_j(\mathbf{s}), \quad (35)$$

where the dipole and quadrupole terms in (35) depend on the FT of the partial derivatives defined in (18). Through fifth order the reciprocal-space tensors are

$$F[\nabla_\alpha \rho^{(n,\kappa)}(\mathbf{r})](\mathbf{s}) = -2\pi i s_\alpha \hat{f}^{(n,\kappa)}(\mathbf{s})\hat{t}(\mathbf{s})$$

$$F[\nabla_\alpha \nabla_\beta \rho^{(n,\kappa)}(\mathbf{r})](\mathbf{s}) = -4\pi^2 s_\alpha s_\beta \hat{f}^{(n,\kappa)}(\mathbf{s})\hat{t}(\mathbf{s})$$

$$F[\nabla_\alpha \nabla_\beta \nabla_\gamma \rho^{(n,\kappa)}(\mathbf{r})](\mathbf{s}) = 8\pi^3 i s_\alpha s_\beta s_\gamma \hat{f}^{(n,\kappa)}(\mathbf{s})\hat{t}(\mathbf{s})$$

$$F[\nabla_\alpha \nabla_\beta \nabla_\gamma \nabla_\delta \rho^{(n,\kappa)}(\mathbf{r})](\mathbf{s}) = 16\pi^4 s_\alpha s_\beta s_\gamma s_\delta \hat{f}^{(n,\kappa)}(\mathbf{s})\hat{t}(\mathbf{s})$$

$$F[\nabla_\alpha \nabla_\beta \nabla_\gamma \nabla_\delta \nabla_\varepsilon \rho^{(n,\kappa)}(\mathbf{r})](\mathbf{s}) = -32\pi^5 i s_\alpha s_\beta s_\gamma s_\delta s_\varepsilon \hat{f}^{(n,\kappa)}(\mathbf{s})\hat{t}(\mathbf{s}) \quad (36)$$

and in compressed tensor notation the general expression for order $u + v + w$ is

$$F[\nabla_x^u \nabla_y^v \nabla_z^w \rho^{(n,\kappa)}(\mathbf{r})](\mathbf{s}) = (-2\pi i)^{u+v+w} s_a^u s_b^v s_c^w \hat{f}^{(n,\kappa)}(\mathbf{s})\hat{t}(\mathbf{s}). \quad (37)$$

This expression is considerably more compact than any reported previously for an aspherical scattering factor in reciprocal space, particularly the formulation based on STOs and spherical harmonics (Hansen & Coppens, 1978). Notably, our formulation has no dependence on cumbersome Fourier Bessel transforms of Slater-type functions (Dawson, 1967a; Hansen & Coppens, 1978; Su & Coppens, 1990). Our equation

(35) has been implemented by 'direct summation' for comparison to the performance of the FFT algorithm.

3. Scattering models

Four scattering models were implemented by modifying and combining the *CNS* (Brünger *et al.*, 1998) and *TINKER* (Ponder, 2004) code bases. The scattering models were added to the *CNS* code base, while *TINKER* was used to compute AMOEBA chemical forces and to supply *CNS* with polarizable multipoles in the global frame.

3.1. Isolated atom

The first scattering model ('IAM') is the conventional IAM based on the relativistic elastic scattering factors described by Su & Coppens (1998).

3.2. Isolated atom with inter-atomic scattering

The second scattering model ('IAM–IAS') augments the IAM with inter-atomic scattering sites at bond centers (Afonine *et al.*, 2007). Unlike the model of Afonine and coworkers, our implementation does not include IAS sites at lone pairs or at the center of aromatic rings. We have neglected these sites based on the rationale that the AMOEBA electrostatic model is sufficient to capture these details of the electron density, which we provide further evidence for below when discussing the refinement of a Tyr-Gly-Gly tripeptide.

In our approach, chemically equivalent bonds are constrained to use the same IAS parameters. Charge density that is added to or removed from bond centers is exactly balanced by changing the net charge of the bond-defining atoms. For example, a bond charge of -0.2 e requires atomic charge increments that sum to 0.2 e. In this way, all molecules retain their original net charge. Each bond type requires three parameters: the charge increments of both atoms and the Gaussian width of the scattering site. Bond types are defined based on the concatenation of the AMOEBA force-field atom types.

3.3. AMOEBA

The third scattering model ('AMOEBA') is based on the polarizable atomic multipoles of the AMOEBA force field. Each chemically unique multipole type requires three Gaussian width parameters as described in §2. The induced dipoles were iterated to self-consistency using PME whenever any atomic coordinates were changed during refinement (Darden *et al.*, 1993; Sagui *et al.*, 2004; Essmann *et al.*, 1995).

### 3.4. AMOEBA with inter-atomic scattering

The final scattering model ('AMOEBA–IAS') augments AMOEBA electrostatics with inter-atomic scattering sites. It became clear during the course of this study that an atomic multipole expansion truncated at quadrupole order is insufficient to capture bond charge density for most molecular geometries. This is consistent with theoretical observations by

Stone and coworkers that the convergence of a distributed multipole analysis (DMA) may be improved by using both atoms and bond centers as expansion sites (Stone & Alderton, 1985; Stone, 2005). Furthermore, experimental data from the X-ray scattering of diamond and silicon, simple examples of tetrahedral bonding geometry, are explained by the super-position of one atomic octopole moment and one atomic hexadecapole moment (Dawson, 1967a,b). The characteristics of the four scattering models are further clarified below with respect to four peptide test cases.

The following computational details were constant across all of the refinements. The isotropic ADP offset U_{add} was set to $1/(4\pi^2)$, which is equivalent to $B_{add} = 8\pi^2 U_{add} = 2$, the FFT grid factor to 0.33 (as appropriate for crystal structures at sub-atomic resolution), and the electron-density cutoff around each atom was 18 (specified by the E_{lim} parameter in *CNS*). These conservative parameters led to close agreement between direct summation and FFT computation of structure factors. The *CNS* parameter w_A that controls the weighting of X-ray target function relative to the force-field energy was set to 1.0, although we also tested 0.2.

$$E_{total} = w_A E_{X\text{-}ray} + E_{force\ field}. \quad (38)$$

This raised R_{free} values by less than 0.1% and lowered the AMOEBA potential energy differences between refinements
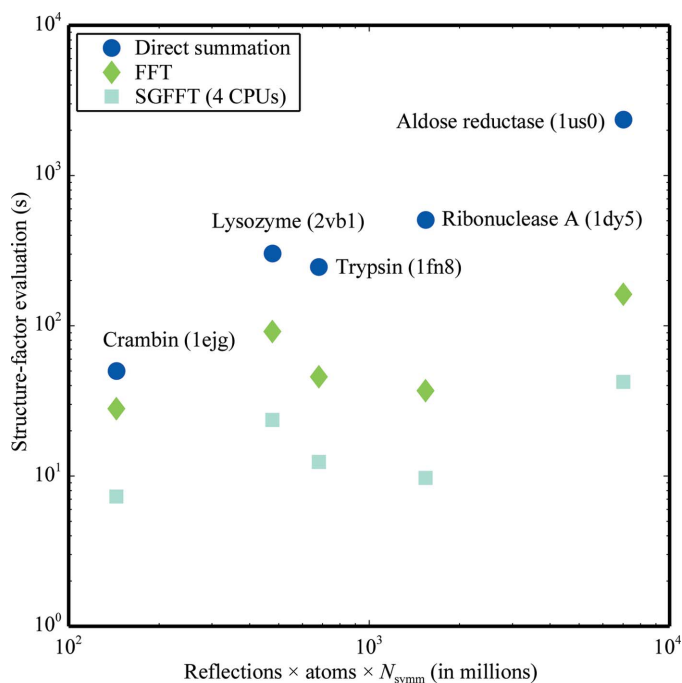
Figure 2
The scaling of the Cartesian Gaussian multipole model, truncated at quadrupole order, is plotted on a log–log scale for computation of the intensity-based maximum-likelihood target function (MLI) for direct summation, FFT and SGFFT. Direct summation scales linearly with the product of the number of atoms, the number of reflections and the number of symmetry operators. Computation of the crystallographic target function by FFT of the Cartesian Gaussian multipole electron density shows a speedup of a factor of between 1.8 and 14.5 compared with direct summation. A further speedup factor of nearly four is achieved using the SGFFT method on a four-processor machine.

presented below, but did not alter any trends or our conclusions. It should be noted that force-field restraints are not necessarily required for refinement at subatomic high resolution. However, their use in this study gives an insight into the relative energetic cost of the structural changes arising from differences in the four scattering models. A modified version of the `refine.inp` *CNS* task file was used for all refinements using the MLI target function.

4. Applications

To demonstrate the behavior of X-ray refinements based on Cartesian Gaussian multipoles, we present two sets of applications. The first set is simply to illustrate the performance of direct summation *versus* FFT and SGFFT computation of structure factors as a function of system size. The second set describes refinements on a series of four peptide crystals that diffract to 0.59 Å resolution or better. All examples use the AMOEBA force field for chemical forces, instead of the default *CNS* force field based on Engh & Huber parameters (Engh & Huber, 1991). Although the refinements were performed in the native space group of each crystal, AMOEBA energies and gradients as computed using the *TINKER* code base required expanding to $P1$. This did not increase the number of refined variables, but suggests the need for an AMOEBA code that takes advantage of crystal symmetry.

4.1. Runtime scaling on protein data sets

Evaluation of the target function and its derivatives by direct-summation calculation of structure factors *via* (35) and (36) is $O(N_{atoms} \times N_{reflections} \times N_{symm})$. Alternatively, the FFT algorithm based on (17) and (18) is $O(N_{grid} \times \log N_{grid})$, where the number of grid points N_{grid} depends on the resolution of the diffraction data. Aspherical refinements based on the Hansen–Coppens formalism are currently limited to direct summation, since the real-space form of the electron density convolved with ADPs is unknown. Therefore, the performance of X-ray refinements based on Cartesian Gaussian multipoles and FFT is of particular interest. The results are summarized in Table 1 and are plotted in Fig. 2. Although the performance difference is only about a factor of two for the small protein crambin, over an order of magnitude improvement is achieved for both ribonuclease A and aldose reductase. Parallelization with the SGFFT method results in a further significant speedup (a speedup of a factor of nearly four relative to a single processor on a four-processor machine).

4.2. Refinement of peptide crystals

In principle, a more precise scattering model based on Cartesian Gaussian multipoles with coefficients from the AMOEBA electrostatics model should improve the quality of refinements relative to the IAM as judged by both R_{free} and the potential energy of the asymmetric unit. Furthermore, the quality of the AMOEBA potential energy function can also be

Table 1
Comparison of computational efficiency of direct-summation, FFT and SGFFT methods for the computation of the Cartesian Gaussian multipole scattering factors.

The permanent multipole expansion was truncated at atomic quadrupoles and polarization was included *via* induced dipoles. The FFT method shows a speedup factor of 1.8–14.5 relative to direct summation. Parallelization by SGFFT provided an additional factor of 3.7–3.9 using four processors. All calculations were performed on a MacPro workstation with 2×2.66 GHz Dual Core Intel Xeon processors.

PDB code	Atoms	Reflections	N_{symm}	Atoms $\times$ reflections $\times$ $N_{symm} \times 10^{-6}$	Direct (s)	FFT (s)	Direct/FFT	SGFFT (s)	Direct/SGFFT
1ejg	642	112209	2	144.1	49.9	28.1	1.8	7.3	6.8
2vb1	2544	187165	1	476.1	301.8	91.5	3.3	23.6	12.8
1fn8	4294	158550	1	680.8	245.1	45.8	5.4	12.4	19.8
1dy5	4835	159422	2	1541.6	505.6	37.0	13.7	9.7	52.1
1us0	6865	511265	2	7019.7	2346.2	162.3	14.5	42.3	55.5

Table 2
Refinement systems.

Molecule	Space group and unit-cell parameters (Å, °)	Non-H atoms	H atoms	Bonds	d_{min} (Å)	Reflections
YG$_2$	$P2_12_12_1$, $a = 7.98$, $b = 9.54$, $c = 18.32$	22	19	40	0.43	4766
P$_2$A$_4$	$P2_12_12_1$, $a = 10.13$, $b = 12.50$, $c = 19.50$	35	36	72	0.37	24878
AYA + 3 waters	$P2_1$, $a = 8.12$, $b = 9.30$, $c = 12.53$, $\beta = 91.21$	26	27	50	0.59	5019
AYA + ethanol	$P2_1$, $a = 8.85$, $b = 9.06$, $c = 12.36$, $\beta = 94.56$	26	27	52	0.59	5258

assayed, since it is reasonable to expect that potential energy and R_{free} should be correlated.

The peptide crystals studied include YG$_2$ (Pichon-Pesme *et al.*, 2000), cyclic P$_2$A$_4$ (Dittrich *et al.*, 2002) and AYA with three waters or with an ethanol molecule (Chęcińska, Forster *et al.*, 2006; Chęcińska, Mebs *et al.*, 2006). Detailed descriptions of the unit-cell parameters, number of atoms, resolution and measured reflections are given in Table 2. The refinement results are summarized in Table 3 and compared with previous work below.

4.2.1. YG$_2$. The R_{free} values of the IAM and IAM–IAS refinements of YG$_2$ (4.60 and 3.86%, respectively) are slightly lower than those reported by Afonine and coworkers (4.72 and 4.06%, respectively; Afonine *et al.*, 2007). The R_{free} value of the AMOEBA–IAS refinement (3.50%) is a significant improvement. The R_{work} value (3.17%) of the AMOEBA–IAS refinement is also lower and is comparable to multipolar refinements reported by Volkov and coworkers using transferred or refined multipole coefficients (3.66% and 3.42%, respectively; Volkov *et al.*, 2007). Cross-validation-based comparisons are unavailable in this case. We note that the AMOEBA–IAS refinement used a reflections-to-parameters ratio of 11.1, which is slightly higher than the value of 10.6 reported by Volkov and coworkers using refined multipole coefficients. This is computed based on the number of reflec-
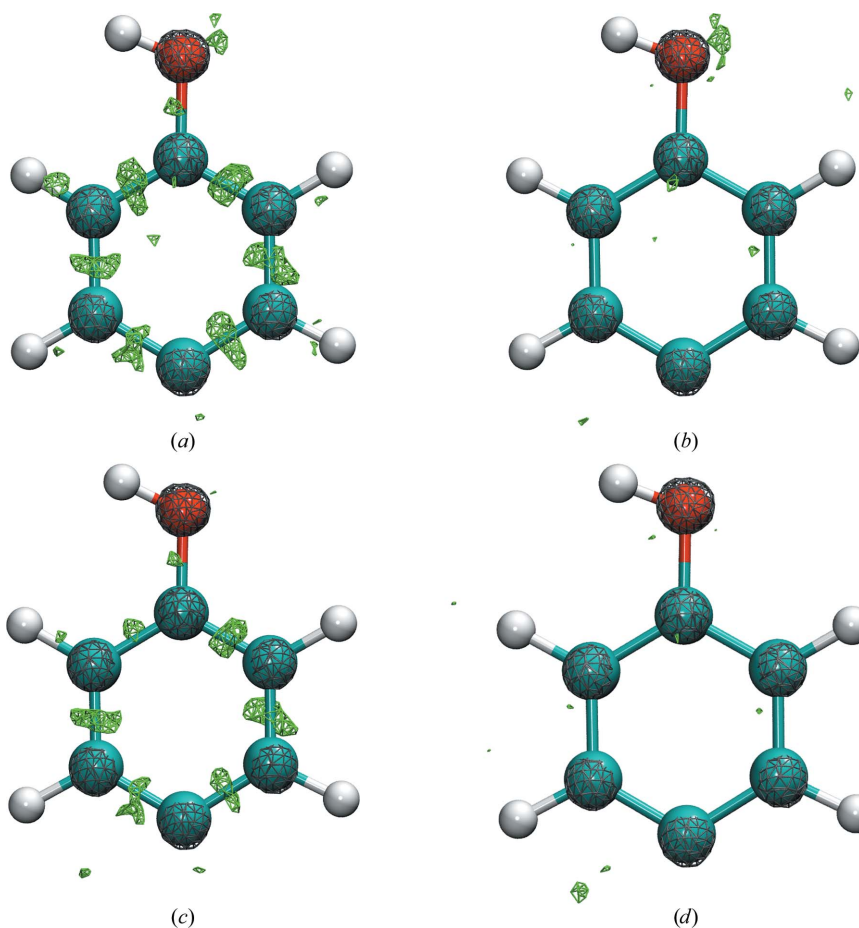
Figure 3
(a) IAM, (b) IAM–IAS, (c) AMOEBA and (d) AMOEBA–IAM refinements, respectively, for GY$_2$. The $F_o - F_c$ and $2F_o - F_c$ σ_A-weighted electron-density maps are contoured at 3.5σ and shown in green and gray, respectively. Both the IAM and AMOEBA models fail to explain the electron density at bond centers seen in the data. In addition, the IAM model does not account for lone-pair density on the O atom.

Table 3
Refinement statistics and the relative AMOEBA potential energy per asymmetric unit are given for four small peptide crystals using the IAM, IAM–IAS, AMOEBA and AMOEBA–IAS scattering models.

In all cases, the lowest R_{free} was found using the AMOEBA–IAS scattering model. Furthermore, the structure with the lowest AMOEBA potential energy per asymmetric unit also corresponded to AMOEBA–IAS refinement.

Molecule	Scattering model	N_{param}	N_{data}/N_{param}	R_{work}/R_{free} (%)		Energy† (kcal mol^{-1})
				$I_{obs}/\sigma(I_{obs}) > 0$	$I_{obs}/\sigma(I_{obs}) > 3$	
YGG	IAM	274	17.4	4.73/4.74	4.41/4.60	36.5
	IAM–IAS	349	13.7	3.93/4.01	3.59/3.86	7.2
	AMOEBA	355	13.4	4.50/4.56	4.16/4.37	6.8
	AMOEBA–IAS	430	11.1	3.54/3.72	3.17/3.50	0.0
PPAAAA	IAM	339	73.4	4.25/4.22	3.65/3.73	32.2
	IAM–IAS	417	59.7	3.56/3.48	3.00/3.01	18.3
	AMOEBA	417	59.7	4.24/4.23	3.69/3.77	12.9
	AMOEBA–IAS	495	50.3	3.42/3.42	2.86/2.94	0.0
AYA + 3 waters	IAM	342	14.7	2.75/2.79	2.67/2.71	17.5
	IAM–IAS	411	12.2	2.24/2.47	2.16/2.39	4.1
	AMOEBA	423	11.9	2.40/2.55	2.31/2.47	4.7
	AMOEBA–IAS	492	10.2	1.72/2.03	1.64/1.95	0.0
AYA + ethanol	IAM	342	15.4	3.30/3.50	3.20/3.33	23.1
	IAM–IAS	423	12.4	2.32/2.66	2.21/2.49	14.8
	AMOEBA	435	12.1	3.42/3.75	3.32/3.58	3.7
	AMOEBA–IAS	516	10.2	1.90/2.25	1.79/2.08	0.0

† 1 kcal mol^{-1} = 4.186 kJ mol^{-1}.

tions reported in Table 2 and the number of parameters given in Table 3.

Electron-density maps of the tyrosine ring for the four scattering models are shown in Fig. 3, which lend visual insight into their properties. The non-H atom positions are apparent
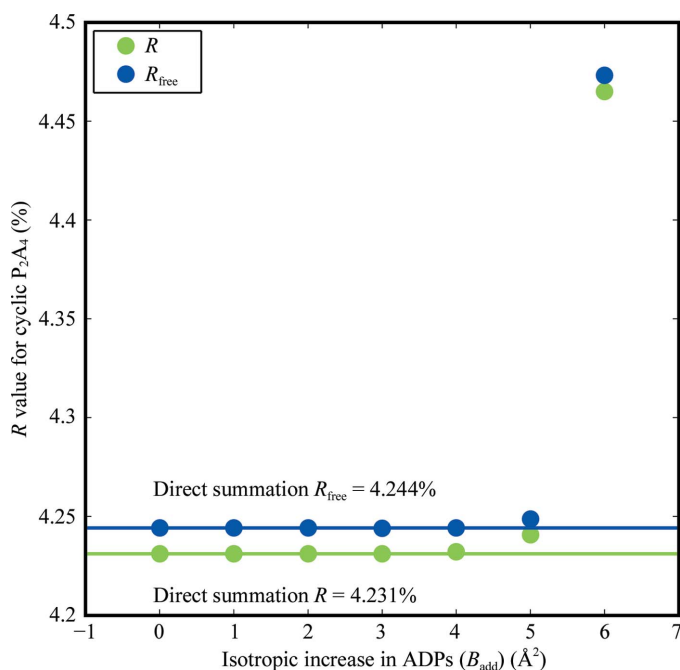
Figure 4
The precision of numerical computation of the R_{work} and R_{free} values *via* FFT is compared with analytic direct summation as a function of the isotropic increase B_{add} in ADP parameters for P_2A_4 under the AMOEBA scattering model. Note that $B_{add} = 8\pi^2 U_{add}$.

in the $2F_o - F_c$ contours for each refinement. The standard IAM scattering model underestimates the electron density at bond centers and at the oxygen lone-pair sites, as shown by the $F_o - F_c$ contours. Our IAM–IAS scattering model explains the electron density at bond centers, but does not capture lone-pair electron density. Conversely, the AMOEBA model places electron density approximately at the lone-pair positions but not at bond centers. Finally, the AMOEBA–IAS model explains much of the lone-pair and bonding electron densities.

4.2.2. P_2A_4. The R_{free} values of our IAM and IAM–IAS refinements of P_2A_4 (3.73 and 3.01%, respectively) agree closely with the values of Afonine and coworkers (3.63 and 3.23%, respectively; Afonine *et al.*, 2007). The R_{free} value of the AMOEBA–IAS refinement (2.94%) is lower by 0.07%, which is the least amount of improvement seen for AMOEBA–IAS relative to IAM–IAS in this study. The R_{work} value (2.86%) of the AMOEBA–IAS refinement is slightly higher, but comparable to those reported by Volkov and coworkers using transferred or refined multipole coefficients (2.60% and 2.53%), although this work uses a higher reflections-to-parameters ratio (50.3 compared with 43.6; Volkov *et al.*, 2007). As for YG$_2$, cross-validation was not performed. The similarity of the R values for YG$_2$ and P_2A_4 between the AMOEBA–IAS refinements and the multipolar refinements of Volkov and coworkers is consistent with the principle that bond scattering sites capture density that is represented by higher order atomic moments missing in the AMOEBA model (octopole and hexadecapole).

In Fig. 4 the precision of the R_{work} and R_{free} values computed using discrete FTs are compared with analytic direct summation for P_2A_4 under the AMOEBA scattering model. Agreement to four decimal places is seen for B_{add} values between 0 and 3 Å^2, which serves as validation of the correctness of (17) and (35). These results support the conclusion that FFT-based computation of structure factors is appropriate for anisotropic and aspherical scattering models.
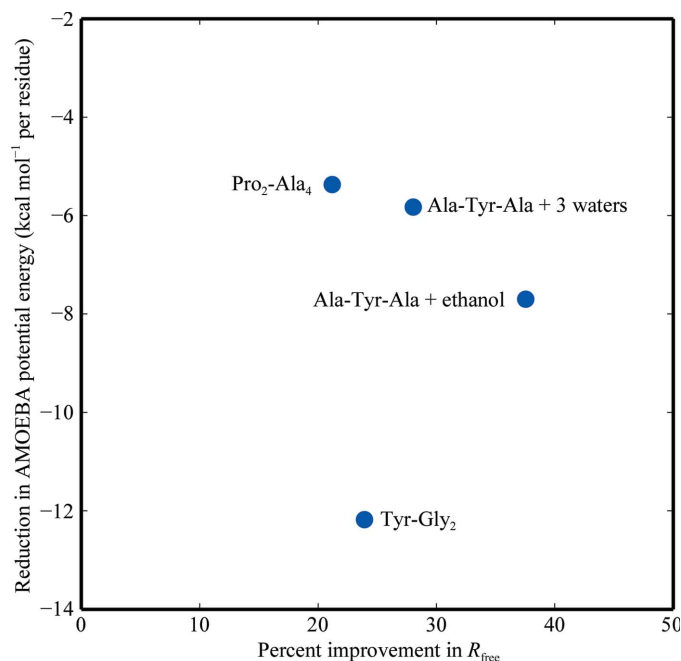
4.2.3. AYA. The AYA data sets were chosen because of the extremely low temperature achieved during the measurement of structure factors (9 K for AYA + three waters and 20 K for AYA + ethanol). For AYA + water, Chęcińska and coworkers (Chęcińska, Forster *et al.*, 2006; Chęcińska, Mebs *et al.*, 2006) originally reported an R value of 2.4%, which is in agreement with the R value of our IAM refinement (2.67%). Addition of IAS lowered the R_{free} statistic from 2.71% to 2.39%, while addition of polarizable atomic multipole electron density showed a further improvement to an R_{free} of 1.95%. For AYA + ethanol the R_{work} value of the IAM (3.20%) is

comparable to that reported originally by Chęcińska and coworkers (2.9%). IAM–IAS lowered R_{free} from 3.33 to 2.49%, while AMOEBA–IAS achieved 2.08%.

4.3. Refinement summary

The results for all four peptide refinements are summarized in Fig. 5. In every case, use of the AMOEBA–IAS scattering model relative to the IAM scattering model lowered both R_{free} and the potential energy of the crystal. When the IAM scattering model is used, molecular conformations are highly strained to compensate. For example, H—C atom bonds are too short because the IAM model centers electron density at the hydrogen nucleus. In the crystal structures, this electron density is shifted towards the C atom. As the description of the electron density is improved, the molecular conformation relaxes by approximately 16 kJ mol^{-1} per residue. The precise amount of relaxation depends on the weighting between the crystallographic target and the force field. Unrestrained refinements with an IAM scattering model could adopt even more unphysical conformations. This suggests that accurate chemical restraints are necessary even for ultrahigh-resolution refinements unless an anisotropic and aspherical scattering model is used.
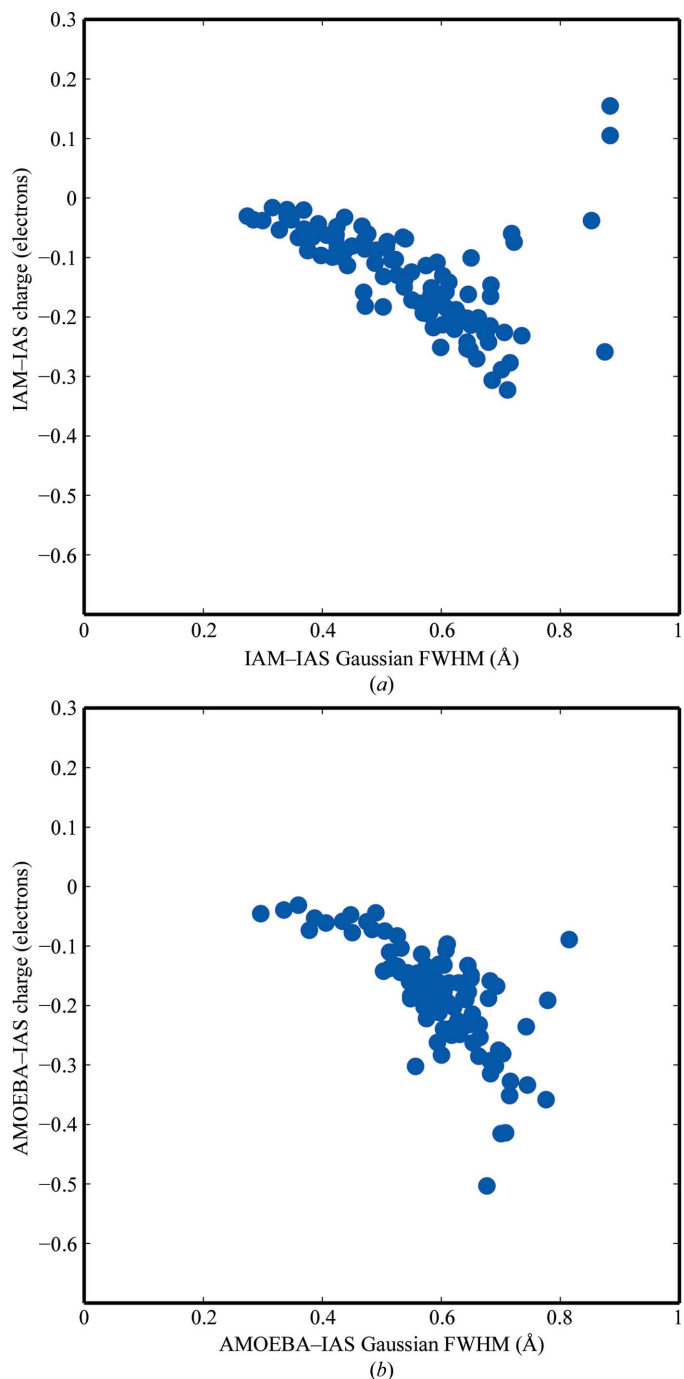
In Fig. 6, we present plots of the IAS sites that were refined for each peptide system. Their Gaussian full-width at half-maximum (FWHM) is plotted against charge magnitude for both the IAM–IAS and the AMOEBA–IAS models. The majority of the charges under the IAM–IAS model and all of the charges under the AMOEBA–IAS model refined to negative partial charge values (or positive scattering density), which is consistent with the physical concentration of charge density at chemical bonds. The similarity of the refined charges between the IAM–IAS and the AMOEBA–IAS models suggests that an atomic multipole description of electron density truncated at quadrupole order underestimates density at trigonal and tetrahedral bond centers.

Figure 5
The improvement arising from the AMOEBA–IAS scattering model, relative to the IAM model, is plotted as a function of relative percentage improvement in R_{free} value and the relative AMOEBA potential energy per residue. For all data sets, the best R_{free} value and lowest potential energy per residue were achieved using the AMOEBA–IAS scattering model. 1 kcal mol^{-1} = 4.186 kJ mol^{-1}.

Figure 6
For the inter-atomic scattering sites of the IAM–IAS (a) and AMOEBA–IAS (b) scattering models, the refined Gaussian full-width at half-maximum (FWHM) is plotted versus partial charge magnitude. The majority of charges for the IAM–IAS model and all charges for the AMOEBA–IAS are negative. The sub-angstrom FWHM values are consistent with very localized bond densities.

5. Conclusions

Cartesian Gaussian multipoles offer an efficient alternative to the Hansen and Coppens formulation of aspherical scattering. They eliminate the use of Slater-type functions and allow structure factors to be computed by FFT. Numerical tests show that that FFT and direct-summation implementations of Cartesian Gaussian multipoles agree to high precision. For subatomic resolution biomolecular data sets such as ribonuclease A and aldose reductase, parallelized computation of structure factors using the SGFFT method results in a speedup of one to two orders of magnitude compared with direct summation.

Ideally, a force-field electrostatics model should be accurate enough to explain the electron density observed in X-ray diffraction experiments. Although the AMOEBA polarizable multipole force field energetic model shows promise, truncation of the permanent moments at quadrupole order systematically underestimates electron density at bond centers. Our results suggest that the added computational expense of including hexadecapole moments in the atomic scattering factor computation is justified. As supplementary information we have provided a Mathematica notebook and formulae that allow computation of Cartesian Gaussian multipoles up to the fourth order in anticipation of further improvements to force fields.

In the near future, we will present the results of applying our polarizable atomic multipole refinement methodology to macromolecules. For ultrahigh-resolution macromolecular data sets, such as HEWL at 0.65 Å (Wang *et al.*, 2007), our scattering model significantly improves refinement statistics, as it does for the simpler peptide cases presented here. Equally exciting will be the use of the AMOEBA force field and in particular the electrostatic forces to orient water molecules in the absence of clear H-atom electron density. We anticipate that refinement of hydrogen-bonding networks will enhance the usefulness of X-ray crystallography experiments with respect to explaining pK_a shifts, ligand-binding affinities and enzymatic mechanisms.

APPENDIX A
Fourier transform definition

The definition and notation for the Fourier transform as used in this work is given by

$$\hat{f}(k) = \int_{-\infty}^{\infty} f(t)\exp(2\pi ikt)\,dt$$
$$= F[f(t)](k) \tag{39}$$

and the corresponding inverse Fourier transform by

$$f(t) = \int_{-\infty}^{\infty} \hat{f}(k)\exp(-2\pi ikt)\,dk$$
$$= F^{-1}[\hat{f}(k)](t). \tag{40}$$

APPENDIX B
Derivative of the polarizable electron density with respect to atomic coordinates

The total polarizable electron density arising from the induced dipole of all atoms is given by

$$\rho_{\text{ind}}(\mathbf{r}) = \sum_{i=1}^{n} \mathbf{u}_{i,\beta} \nabla_{i,\beta} \rho_i^{(1,\kappa_d)}(\mathbf{r}). \tag{41}$$

The gradient of this density with respect to the α component of atom j is

$$\frac{\partial \rho_{\text{ind}}(\mathbf{r})}{\partial r_{j,\alpha}} = \frac{\partial}{\partial r_{j,\alpha}} \sum_{i=1}^{n} u_{i,\beta} \nabla_{i,\beta} \rho_i^{(1,\kappa_d)}(\mathbf{r})$$
$$= \sum_{i=1}^{n} \left[\frac{\partial u_{i,\beta}}{\partial r_{j,\alpha}} \nabla_{i,\beta} \rho_i^{(1,\kappa_d)}(\mathbf{r}) \right] + \delta_{ij} u_{i,\beta} \nabla_\alpha \nabla_\beta \rho_i^{(1,\kappa_d)}(\mathbf{r}). \tag{42}$$

The second term is nonzero only for $i = j$ and is simple to calculate. The first term, however, depends on $\partial u_{i,\beta}/\partial r_{j,\alpha}$ which is the derivative of a component of the induced dipole of atom i with respect to the α component of atom j. In other words, perturbing the position of atom j affects not only its own scattering but that of all polarizable atoms. The induced dipole on atom i arises from the self-consistent crystal field multiplied by the polarizability,

$$\mathbf{u}_i = \alpha_i \mathbf{E}_i$$
$$= \alpha_i \left[\sum_{k \neq i} \mathbf{T}_{ik}^{(1)} \mathbf{M}_k + \sum_{k \neq i} \mathbf{T}_{ik}^{(11)} \mathbf{u}_k \right], \tag{43}$$

where α_i is the atomic polarizability of atom i, $\mathbf{T}_{ik}^{(1)}$ is a matrix of tensors that produces the field at site i

$$\mathbf{T}_{ik}^{(1)} = \begin{pmatrix} \dfrac{\partial}{\partial x_i} & \dfrac{\partial^2}{\partial x_i \partial x_k} & \dfrac{\partial^2}{\partial x_i \partial y_k} & \dfrac{\partial^2}{\partial x_i \partial z_k} & \cdots \\ \dfrac{\partial}{\partial y_i} & \dfrac{\partial^2}{\partial y_i \partial x_k} & \dfrac{\partial^2}{\partial y_i \partial y_k} & \dfrac{\partial^2}{\partial y_i \partial z_k} & \cdots \\ \dfrac{\partial}{\partial z_i} & \dfrac{\partial^2}{\partial z_i \partial x_k} & \dfrac{\partial^2}{\partial z_i \partial y_k} & \dfrac{\partial^2}{\partial z_i \partial z_k} & \cdots \end{pmatrix} \dfrac{1}{r_{ik}} \tag{44}$$

owing to the multipole $\mathbf{M}_k$ at site k

$$\mathbf{M}_k = (q_k, d_{k,x}, d_{k,y}, d_{k,z}, \Theta_{k,xx}, \Theta_{k,xy}, \Theta_{k,xz}, \ldots, \Theta_{k,zz})^t \tag{45}$$

and $\mathbf{T}_{ik}^{(11)}$ is the matrix of tensors that produces the field at site i

$$\mathbf{T}_{ik}^{(11)} = \begin{pmatrix} \dfrac{\partial^2}{\partial x_i \partial x_k} & \dfrac{\partial^2}{\partial x_i \partial y_k} & \dfrac{\partial^2}{\partial x_i \partial z_k} \\ \dfrac{\partial^2}{\partial y_i \partial x_k} & \dfrac{\partial^2}{\partial y_i \partial y_k} & \dfrac{\partial^2}{\partial y_i \partial z_k} \\ \dfrac{\partial^2}{\partial z_i \partial x_k} & \dfrac{\partial^2}{\partial z_i \partial y_k} & \dfrac{\partial^2}{\partial z_i \partial z_k} \end{pmatrix} \dfrac{1}{r_{ik}} \tag{46}$$

owing to the induced dipole $\boldsymbol{u}_k$ at site k. For simplicity, we have not formulated (43) using PME electrostatics. Therefore, the sum over k includes all atoms in the crystal except atom i. The derivative of (43) with respect to coordinate $r_{j,\alpha}$ is given by

$$\frac{\partial u_i}{\partial r_{j,\alpha}} = \alpha_i \left\{ \sum_{k\neq i}^{n} \left[\frac{\partial \mathbf{T}_{ik}^{(1)}}{\partial r_{j,\alpha}} \mathbf{M}_k + \mathbf{T}_{ik}^{(1)} \frac{\partial \mathbf{M}_k}{\partial r_{j,\alpha}} \right] \right.$$

$$\left. + \sum_{k\neq i}^{n} \left[\frac{\partial \mathbf{T}_{ik}^{(11)}}{\partial r_{j,\alpha}} \mathbf{u}_k + \mathbf{T}_{ik}^{(11)} \frac{\partial \mathbf{u}_k}{\partial r_{j,\alpha}} \right] \right\}$$

$$= \alpha_i \left[\delta_{jk} \frac{\partial \mathbf{T}_{ik}^{(1)}}{\partial r_{j,\alpha}} \mathbf{M}_k + \sum_{k=\{j\}} \mathbf{T}_{ik}^{(1)} \frac{\partial \mathbf{M}_k}{\partial r_{j,\alpha}} \right.$$

$$\left. + \delta_{jk} \frac{\partial \mathbf{T}_{ik}^{(11)}}{\partial r_{j,\alpha}} \mathbf{u}_k + \sum_{k\neq i}^{n} \mathbf{T}_{ik}^{(11)} \frac{\partial \mathbf{u}_k}{\partial r_{j,\alpha}} \right]. \qquad (47)$$

The first three terms on the right-hand side are not difficult to compute. However, the fourth term shows that the gradients of the polarizable scattering are $O(n^3)$ without use of PME. Specifically, there are $3n \times 3n$ induced dipole density derivatives, each of which is the sum of $3n$ terms. In this work, we have computed these derivatives by finite differences using PME, which is $O(n^2 \log n)$.

References

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1948–1954.

Adams, P. D., Pannu, N. S., Read, R. J. & Brünger, A. T. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 5018–5023.

Afonine, P. V., Grosse-Kunstleve, R. W., Adams, P. D., Lunin, V. Y. & Urzhumtsev, A. (2007). *Acta Cryst.* D**63**, 1194–1197.

Afonine, P. V. & Urzhumtsev, A. (2004). *Acta Cryst.* A**60**, 19–32.

Agarwal, R. C. (1978). *Acta Cryst.* A**34**, 791–809.

Applequist, J. (1989). *J. Phys. A*, **22**, 4303–4330.

Applequist, J. (2002). *Theor. Chem. Acc.* **107**, 103–115.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Boys, S. F. (1950). *Proc. R. Soc. London Ser. A*, **200**, 542–554.

Bricogne, G. (2001). *International Tables for Crystallography*, Vol. *B*, edited by U. Schmueli, pp. 25–98. Dordrecht: Kluwer Academic Publishers.

Brünger, A. T. (1988). *J. Mol. Biol.* **203**, 803–816.

Brünger, A. T. (1989). *Acta Cryst.* A**45**, 42–50.

Brünger, A. T. (1991). *Annu. Rev. Phys. Chem.* **42**, 197–223.

Brünger, A. T. (2007). *Nature Protoc.* **2**, 2728–2733.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Brünger, A. T., Adams, P. D. & Rice, L. M. (1997). *Structure*, **5**, 325–336.

Brünger, A. T., Karplus, M. & Petsko, G. A. (1989). *Acta Cryst.* A**45**, 50–61.

Brünger, A. T., Krukowski, A. & Erickson, J. W. (1990). *Acta Cryst.* A**46**, 585–593.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458–460.

Brünger, A. T. & Rice, L. M. (1997). *Methods Enzymol.* **277**, 243–269.

Checińska, L., Förster, D., Morgenroth, W. & Luger, P. (2006). *Acta Cryst.* C**62**, o454–o457.

Checińska, L., Mebs, S., Hubschle, C. B., Forster, D., Morgenroth, W. & Luger, P. (2006). *Org. Biomol. Chem.* **4**, 3242–3251.

Cooley, J. W. & Tukey, J. W. (1965). *Math. Comput.* **19**, 297–301.

Coppens, P. (2005). *Angew. Chem. Int. Ed.* **44**, 6810–6811.

Coppens, P. & Volkov, A. (2004). *Acta Cryst.* A**60**, 357–364.

Cromer, D. T., Larson, A. C. & Stewart, R. F. (1976). *J. Chem. Phys.* **65**, 336–349.

Darden, T., York, D. & Pedersen, L. (1993). *J. Chem. Phys.* **98**, 10089–10092.

Dawson, B. (1967a). *Proc. R. Soc. London Ser. A*, **298**, 264–288.

Dawson, B. (1967b). *Proc. R. Soc. London Ser. A*, **298**, 379–394.

Dawson, B. (1967c). *Proc. R. Soc. London Ser. A*, **298**, 395–401.

DeMarco, J. J. & Weiss, R. J. (1965). *Phys. Rev.* **137**, A1869.

Deng, J. P., Xiong, Y. & Sundaralingam, M. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 13665–13670.

Dittrich, B., Koritsánszky, T., Grosche, M., Scherer, W., Flaig, R., Wagner, A., Krane, H. G., Kessler, H., Riemer, C., Schreurs, A. M. M. & Luger, P. (2002). *Acta Cryst.* B**58**, 721–727.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Epstein, J., Bentley, J. & Stewart, R. F. (1977). *J. Chem. Phys.* **66**, 5564–5567.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. & Pedersen, L. G. (1995). *J. Chem. Phys.* **103**, 8577–8593.

Friesner, R. A., Baldwin, R. L. & Baker, D. (2005). *Adv. Protein Chem.* **72**, 79–104.

Gresh, N. (2006). *Curr. Pharm. Des.* **12**, 2121–2158.

Gresh, N., Cisneros, G. A., Darden, T. A. & Piquemal, J. P. (2007). *J. Chem. Theory Comput.* **3**, 1960–1986.

Grosse-Kunstleve, R. W. & Adams, P. D. (2002). *J. Appl. Cryst.* **35**, 477–480.

Hansen, N. K. & Coppens, P. (1978). *Acta Cryst.* A**34**, 909–921.

Hehre, W. J., Ditchfie, R., Stewart, R. F. & Pople, J. A. (1970). *J. Chem. Phys.* **52**, 2769–2773.

Hehre, W. J., Stewart, R. F. & Pople, J. A. (1969). *J. Chem. Phys.* **51**, 2657–2664.

Howard, E. I., Sanishvili, R., Cachau, R. E., Mitschler, A., Chevrier, B., Barth, P., Lamour, V., Van Zandt, M., Sibley, E., Bon, C., Moras, D., Schneider, T. R., Joachimiak, A. & Podjarny, A. (2004). *Proteins*, **55**, 792–804.

Jelsch, C., Guillot, B., Lagoutte, A. & Lecomte, C. (2005). *J. Appl. Cryst.* **38**, 38–54.

Jelsch, C., Teeter, M. M., Lamzin, V., Pichon-Pesme, V., Blessing, R. H. & Lecomte, C. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 3171–3176.

Kielkopf, C. L., Ding, S., Kuhn, P. & Rees, D. C. (2000). *J. Mol. Biol.* **296**, 787–801.

Kuhn, P., Knapp, M., Soltis, S. M., Ganshaw, G., Thoene, M. & Bott, R. (1998). *Biochemistry*, **37**, 13446–13452.

Kuriyan, J., Brünger, A. T., Karplus, M. & Hendrickson, W. A. (1989). *Acta Cryst.* A**45**, 396–409.

Kurki-Suonio, K. (1968). *Acta Cryst.* A**24**, 379–390.

Lamoureux, G., Harder, E., Vorobyov, I. V., Roux, B. & MacKerell, A. D. (2006). *Chem. Phys. Lett.* **418**, 245–249.

Lamoureux, G. & Roux, B. (2003). *J. Chem. Phys.* **119**, 3025–3039.

Maple, J. R., Cao, Y. X., Damm, W. G., Halgren, T. A., Kaminski, G. A., Zhang, L. Y. & Friesner, R. A. (2005). *J. Chem. Theory Comput.* **1**, 694–715.

Muzet, N., Guillot, B., Jelsch, C., Howard, E. & Lecomte, C. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 8742–8747.

Pande, V. S., Baker, I., Chapman, J., Elmer, S. P., Khaliq, S., Larson, S. M., Rhee, Y. M., Shirts, M. R., Snow, C. D., Sorin, E. J. & Zagrovic, B. (2003). *Biopolymers*, **68**, 91–109.

Patel, S. & Brooks, C. L. (2006). *Mol. Simul.* **32**, 231–249.

Pichon-Pesme, V., Lachekar, H., Souhassou, M. & Lecomte, C. (2000). *Acta Cryst.* B**56**, 728–737.

Piquemal, J. P., Chelli, R., Procacci, P. & Gresh, N. (2007). *J. Phys. Chem. A*, **111**, 8170–8176.

Piquemal, J. P., Cisneros, G. A., Reinhardt, P., Gresh, N. & Darden, T. A. (2006). *J. Chem. Phys.* **124**, 104101.

Ponder, J. W. (2004). *TINKER: Software Tools for Molecular Design*, v.4.2.

Ponder, J. W. & Case, D. A. (2003). *Adv. Protein Chem.* **66,** 27–85.

Ren, P. Y. & Ponder, J. W. (2002). *J. Comput. Chem.* **23**, 1497–1506.

Ren, P. Y. & Ponder, J. W. (2003). *J. Phys. Chem. B*, **107**, 5933–5947.

Ren, P. Y. & Ponder, J. W. (2004). *J. Phys. Chem. B*, **108**, 13427–13437.

Sagui, C., Pedersen, L. G. & Darden, T. A. (2004). *J. Chem. Phys.* **120**, 73–87.

Schnieders, M. J., Baker, N. A., Ren, P. Y. & Ponder, J. W. (2007). *J. Chem. Phys.* **126**, 124114.

Schnieders, M. J. & Ponder, J. W. (2007). *J. Chem. Theory Comput.* **3**, 2083–2097.

Shirts, M. R. & Pande, V. S. (2005). *J. Chem. Phys.* **122**, 134508.

Shirts, M. R., Pitera, J. W., Swope, W. C. & Pande, V. S. (2003). *J. Chem. Phys.* **119**, 5740–5761.

Snow, C. D., Nguyen, N., Pande, V. S. & Gruebele, M. (2002). *Nature (London)*, **420**, 102–106.

Snow, C. D., Sorin, E. J., Rhee, Y. M. & Pande, V. S. (2005). *Annu. Rev. Biophys. Biomol. Struct.* **34**, 43–69.

Sorin, E. J. & Pande, V. S. (2005). *J. Comput. Chem.* **26**, 682–690.

Stewart, R. F. (1977). *Chem. Phys. Lett.* **49**, 281–284.

Stewart, R. F. (1979). *Chem. Phys. Lett.* **65**, 335–342.

Stone, A. J. (1996). *The Theory of Intermolecular Forces.* Oxford: Clarendon Press.

Stone, A. J. (2005). *J. Chem. Theory Comput.* **1**, 1128–1132.

Stone, A. J. & Alderton, M. (1985). *Mol. Phys.* **56**, 1047–1064.

Su, Z. & Coppens, P. (1990). *J. Appl. Cryst.* **23**, 71–73.

Su, Z. & Coppens, P. (1998). *Acta Cryst.* A**54**, 357.

Ten Eyck, L. F. (1973). *Acta Cryst.* A**29**, 183–191.

Ten Eyck, L. F. (1977). *Acta Cryst.* A**33**, 486–492.

Tereshko, V., Wilds, C. J., Minasov, G., Prakash, T. P., Maier, M. A., Howard, A., Wawrzak, Z., Manoharan, M. & Egli, M. (2001). *Nucleic Acids Res.* **29**, 1208–1215.

Trueblood, K. N., Bürgi, H.-B., Burzlaff, H., Dunitz, J. D., Gramaccioli, C. M., Schulz, H. H., Shmueli, U. & Abrahams, S. C. (1996). *Acta Cryst.* A**52**, 770–781.

Volkov, A., Messerschmidt, M. & Coppens, P. (2007). *Acta Cryst.* D**63**, 160–170.

Waller, I. (1923). *Z. Phys. A*, **17**, 398–408.

Wang, J., Dauter, M., Alkire, R., Joachimiak, A. & Dauter, Z. (2007). *Acta Cryst.* D**63**, 1254–1268.

Zarychta, B., Pichon-Pesme, V., Guillot, B., Lecomte, C. & Jelsch, C. (2007). *Acta Cryst.* A**63**, 108–125.