research papers

Advances in protein and small-molecule charge-density refinement methods using *MoPro*

Christian Jelsch,* Benoit Guillot, Angelique Lagoutte and Claude Lecomte

Laboratoire de Cristallographie et Modélisation des Matériaux Minéraux et Biologiques, CNRS, UMR 7036, Faculté des Sciences BP 239, 54506 Vandoeuvre-lès-Nancy, France. Correspondence e-mail: christian.jelsch@lcm3b.uhp-nancy.fr

With an increasing number of biological macromolecule structures solved at ultra-high resolution and with the advances of supramolecular chemistry, it becomes necessary to extend to large systems experimental charge-density study methods that are usually applied to small molecules. The latest developments in the refinement program *MoPro* (*Molecular Properties*), dedicated to the charge-density refinement at (sub)atomic resolution of structures ranging from small molecules to biological macromolecules, are presented. *MoPro* uses the Hansen & Coppens [*Acta Cryst.* (1978), **A34**, 909–921] multipolar pseudo-atom model for the electron-density refinement. Alternative methods are also proposed, such as modelling bonding and lone-pair electron density by virtual spherical atoms. For proteins at atomic resolution, a charge-density database developed in the laboratory enables the transfer of multipolar parameters. The program allows complex refinement strategies to be written and has numerous restraints, constraints and analysis tools for use in the structure and electron-density analysis. New kappa and multipolar parameter restraints/constraints are also implemented and discussed. Furthermore, constraints on the electron density, such as local symmetry and atom equivalence, are easily defined. Some examples of applications, from small molecules to large unit cells (including the enzyme aldose reductase), are given in order to guide the *MoPro* user and to show the large field of applicability of this code.

1. Introduction

In the usual least-squares treatment of X-ray diffraction data, the continuous electron density is subdivided into independent atomic charge densities, and the basic assumptions are that atoms are neutral and of spherical shape with a radial dependence equal to that of free atoms in the gas phase. However, due to atom–atom interactions such as chemical bonding, hydrogen bonding and van der Waals effects, the atomic electron density is obviously not spherical. Indeed, accurate low-temperature X-ray data, collected up to subatomic resolution (typically $d < 0.6$ Å or $\sin\theta/\lambda > 0.8$ Å^{-1}), allow the non-spherical character of the atomic electron density to be represented in deformation electron-density maps, and experimentally quantified by using a non-spherical model of the atomic electron density (Coppens, 1967, 1997; Lecomte, 1995). In this case, the deconvolution of the atomic deformation density from the thermal motion can be achieved, and a crystallographic refinement taking into account the asphericity of the electron density allows an unbiased description of the structure stereochemistry and thermal motion.

Charge-density refinement studies of small molecules now constitute a very mature field, in which non-spherical models of atomic electron density are used almost routinely. However, recent advances in macromolecular crystallography techniques, synchrotron sources and crystallogenesis methods allow the solution of numerous protein and nucleic acid structures at resolutions ranging from atomic to subatomic (Schmidt & Lamzin, 2002). Some recent communications have shown that in theses cases, even with data resolution close to 0.8 Å, information on valence electron-density distribution may be obtained for proteins if the thermal displacement parameters are small enough (Jelsch *et al.*, 1998; Guillot, Muzet *et al.*, 2001). For example, in the aldose reductase structure at 0.66 Å resolution, 54% of the hydrogen atoms were identified (Howard *et al.*, 2004; Cachau *et al.*, 2000) and most of the deformation-electron-density peaks were visible in the bonds of the protein active site, where the equivalent thermal factors B_{iso} of the non-hydrogen atoms were about 4 Å^2 (Podjarny *et al.*, 2001). Similarly, Afonine *et al.* (2004) have recently confirmed that bonding-electron-density peaks are observable in an RD1 antifreeze protein structure at 0.62 Å resolution, and have shown from refinements using theoretical structure factors that bonding density can also be seen at atomic resolution (around 0.9 Å), but again this presence is strongly correlated with atomic thermal motion.

Hence, if information related to the deformation electron density is present in experimental maps, it becomes necessary
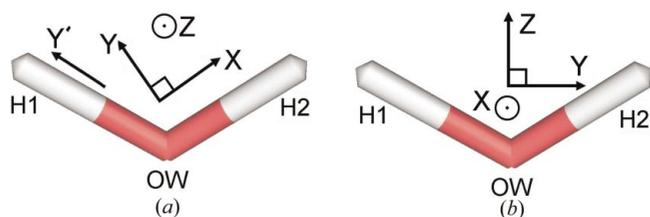
Figure 1
Two examples of definition for the atomic local axes system (**X**,**Y**,**Z**) in the case of a water oxygen atom.

to adapt charge-density methods usually applied to small molecules to such larger systems. For this purpose, we have developed the macromolecular refinement software *MoPro*, which takes into account the deformation of the electron density in the refinement, and implements specific methods derived from small-molecule and biological macromolecular crystallography. Beyond an improvement of the structural and dynamic molecular model, the use of a non-spherical-atom model allows the extraction of derived properties, such as atomic charges and electrostatic potential. As a consequence, the model density obtained with these methods can be used to analyse electrostatic properties of macromolecules, such as protein cofactor/inhibitor interactions (Muzet *et al.*, 2003).

2. Refinement algorithms

2.1. The multipolar atom

The *MoPro* least-squares refinement software (Guillot, Muzet *et al.*, 2001) has been developed by extensive modifications of *MOLLY*, a program dedicated to the charge-density refinement of small compounds (Hansen & Coppens, 1978). The atomic electron-density model is based on the multipolar-atom formalism:

$$\rho_{atom}(\mathbf{r}) = \rho_{core}(\mathbf{r}) + P_{val}\kappa^3 P_{val}(\kappa r) + \sum_{l=0,l_{max}} \kappa'^3 R_l(\kappa'\mathbf{r}) \sum_{|m|\leq l} P_{lm}y_{lm}(x_c, y_c, z_c). \quad (1)$$

In this formalism, the total density is decomposed into core electrons and valence electrons. The term ρ_{core} represents the spherically symmetric core electron density and the term ρ_{val} the spherically averaged free-atom valence electron density. The second term gives an estimate of the atomic charge ($N_{val} - P_{val}$), where N_{val} is the number of valence electrons. The carbon, nitrogen and oxygen atoms have for instance two core electrons and respectively two, three and four valence electrons; hydrogen atoms have only one valence electron. The form factors of the core ρ_{core} and of the spherical valence $\rho_{val}(\kappa r)$ may be specified in *MoPro* as tables as a function of $s = \sin\theta/\lambda$, or computed from Clementi & Raimondi (1963) wavefunctions.

The third term describes the multipolar part of the valence electron distribution as a multipole density. The R_l are generally exponential Slater-type radial functions:

$$R_l(r) = \varsigma_l^{n_l+3} r^{n_l} \exp(-\varsigma_l r)/(n_l + 2)! . \quad (2)$$

Table 1
List of constraints available in *MoPro*.

Keyword	Description
CONDIS	Hydrogen or virtual atom is constrained to be at a given distance from another atom
CONANG	Defines the angular position of 1 to 3 hydrogen atoms belonging to a same chemical group (Table 4)
CONURA	Hydrogen atom is constrained to have temperature factor U_{iso} proportional to U_{eq} of neighbouring atom
CONUIJ	Virtual atom is constrained to have temperature factor U_{ij} proportional to U_{ij} of neighbouring atom
CONPLM	Two or more atoms have same multipole populations; the program stops if multipoles are unequal initially
AVEPLM	Two or more atoms have same multipole populations; the multipoles are averaged if unequal initially
CONVAL	Two or more atoms have same valence and multipole populations; the program stops if parameters are unequal initially
AVEVAL	Two or more atoms have same valence and multipole populations; the parameters are averaged if unequal initially
CONKAP	Two or more atoms have same κ and κ' values
SYMPLM	Multipoles obey a local or crystallographic symmetry: inversion, mirrors, twofold, threefold, fourfold, sixfold rotation or roto-inversion axes

The radial functions may also be given as tables as a function of the reciprocal resolution s, or may be derived from Clementi wavefunctions. The kappa coefficients κ and κ' describe the expansion/contraction of the perturbed valence density (spherical and multipolar part, respectively).

The y_{lm} are spherical harmonic functions in real form (Hansen & Coppens, 1978; Brown *et al.*, 1995). The x_c, y_c, z_c coordinates are direction cosines expressed in a local axis system centred on the atom and oriented according to neighbouring atoms. This facilitates the application of chemical symmetry constraints as well as the comparison atom by atom. The normalization factors of the multipolar density function are based on $\int |y_{lm}| \, dV = 2$, except for the case $l = 0$, where $\int |y_{lm}| \, dV = 1$. This normalization of the density function implies that with $P_{lm} = 1$, one electron has moved from the negative to the positive lobes of the functions with $l \geq 1$.

Multipolar populations P_{lm} correspond to spherical harmonics lobes oriented with respect to a local axis system. An appropriate definition of these axes ensures the possibility for optimal application of symmetry constraints on the deformation density. Two simple examples of local axes for multipole orientation with respect to local atomic geometry are described in Fig. 1 for a water molecule H1—OW—H2. In the first case, the axes definition allows (but does not impose) the application of only one mirror symmetry with respect to the H1—OW—H2 plane (SYMPLM multipole constraint, Table 1). In the second example, in which the **Z** and **X** axes are defined as bisecting the OW—H1 and OW—H2 directions, two mirror symmetries can easily be imposed on the multipole development with respect to the *YOZ* and *ZOX* planes. A few rules permit the definition of standard local axes for the database of multipole parameters (Pichon-Pesme *et al.*, 1995).

2.2. Bulk-solvent correction

For macromolecular refinements, the solvent-scattering contribution is large at low resolution. The absence of bulk-

solvent modelling leads to an overestimation of structure-factor moduli at very low resolution. Two methods of disordered solvent correction are implemented in *MoPro*.

The bulk-solvent model (Fokine & Urzhumtsev, 2002) needs supplementary information, namely the structure factors F_{sol}^{cal} of the diffuse solvent. They can be obtained by computing a Fourier transform of the mask encompassing bulk-solvent regions in the unit cell. The structure-factor computation of the molecular model F_{mol}^{cal} includes the water molecules of the protein first-hydration shells, which are visible in the electron-density map. The solvent structure factors F_{sol}^{cal}, modulated by a Gaussian function of the reciprocal resolution s, are then added in *MoPro* as complex quantities to the structure factors of the molecular model:

$$F_{tot}^{cal} = F_{mol}^{cal} + F_{sol}^{cal} k_{sol} \exp(-B_{sol}s^2). \qquad (3)$$

The moduli of the resulting structure factors are then compared with experimental data, and can be adjusted through the refinement of parameters k_{sol} and B_{sol}.

The second method implemented in *MoPro* is the so-called exponential scaling model. This method is based on an empirical correction of the model structure-factor moduli, which are scaled by a function of the reciprocal resolution s, which compensates the overestimation of model structure factors at low resolution if the solvent is not corrected:

$$F_{tot}^{cal} = [1 - k_{sol} \exp(-B_{sol}s^2)]F_{mol}^{cal}. \qquad (4)$$

Here, again, the parameters k_{sol} and B_{sol} can be refined against experimental data. This modelling is more readily applied, but is less efficient than the first method; it may be useful if the unit cell contains a very small proportion of disordered solvent. The meanings of the parameters k_{sol} and B_{sol} in equations (3) and (4) are different and should not be confused.

2.3. Other parameters describing the structure

MoPro was designed to combine methods from small-molecule charge-density and protein refinement. The atomic thermal motion is generally modelled by anisotropic displacement parameters (ADPs), while for hydrogen atoms an isotropic thermal parameter is used. The thermal displacement modelling based on an anharmonic description of atoms with Gram–Charlier tensors of a given order (Maslen *et al.*, 1995) can also be used in the analysis of crystals with small unit cells. Concerning the extinction correction, an isotropic Gaussian correction is available (Becker, 1977). Anomalous scattering factors may be calculated automatically at the experiment wavelength using the tables created by Kissel *et al.* (1995).

MoPro offers the possibility to use a scale factor that is a polynomial function of the reciprocal resolution $s = 1/2d$:

$$k = k_0 + k_1 s + k_2 s^2 + \ldots + k_n s^n. \qquad (5)$$

In this expression, the coefficients k_i are refinable parameters, and the k_i with $i > 0$ are the corrections applied to the standard scale factors, namely k_0 here. This is useful when, for instance, low- and high-resolution data have been collected separately, which is often the case for ultra-high-resolution protein data collection.

In addition, the use of imaging plates as detectors in X-ray crystallography causes a non-uniform diffraction measurement at oblique incidence of the diffracted beam on the plate (Zaleski *et al.*, 1998). The correction becomes important for wavelengths shorter than about 1 Å.

Indeed, as valence and core scattering have varying influence at different resolutions, a difference in the scale factors at high and low resolution may lead to artifacts in the refined deformation electron density and thermal parameters. Both of these resolution effects on the diffracted intensities can be partly corrected by applying such a polynomial scale factor. Inclusion of such a nonlinear intensity scaling based on equation (5) should be treated with care as there will be a strong correlation with a global isotropic temperature factor. Whatever the oblique factor is in the intensity data, it is always better to correct it during data collection rather than treating it as a parameter.

The default reflection weighting scheme used in *MoPro* is $W_{\mathbf{H}} = 1/\sigma_I^2$, where σ_I is the uncertainty on the reflection intensity measurement. Alternatively, a unitary weighting scheme $W_{\mathbf{H}} = 1$ can be applied. A more complex weighting scheme, $W_{\mathbf{H}} = 1/\sigma_I'^2$, taking into account the reflection intensity $I_{\mathbf{H}}$ and the resolution d (Å), may be applied:

$$\sigma_I' = (a\sigma_I^2 + bI_{\mathbf{H}}^2)^{1/2}d^{-c}. \qquad (6)$$

The appropriate choice of coefficients a, b, c, which is suggested by the program, enables the average $|F_{obs} - F_{cal}|$ discrepancy to be in accordance with the corrected uncertainty σ_I', resulting in a goodness of fit that is approximately equal to unity in all resolution and intensity shells. The application to the standard deviation of a d^{-c} factor enables more weight to be given smoothly to either the high- or the low-resolution reflections. For protein structures at resolution of around 0.6 to 0.9 Å, a weighting scheme where reflections at high resolution are over-weighted, *i.e.* a 'high-order refinement' commonly applied in electron-density analysis of small molecules at a resolution of 0.5 Å or better, can be approached. After such a refinement using a spherical-atom model, the bonding density was partially retrieved for regions of the phospholipase PLA2 with a low thermal motion (Liu *et al.*, 2003).

2.4. Least-squares refinement

The function minimized in the least-squares refinement is composed of the crystallographic error term and restraint components. The function is a summation over the reflections and over all applied restraints:

$$E = \sum_{\mathbf{H}} W_{\mathbf{H}}(|F|_{obs} - |F|_{cal})^2 + \sum_{restraints} W_R(R_{target} - R_{cal})^2. \qquad (7)$$

The summation over the reflections $\mathbf{H}$ can be performed on the structure-factor moduli difference or alternatively over the intensity difference $I_{obs} - I_{cal}$.

In usual macromolecular biological crystallography ($\sim$2 Å resolution), due to the limited resolution and the low variables/observations ratio, it is necessary to incorporate stereochemical and dynamical knowledge into the refinement.

The restraints reflect the target geometry as anticipated from high-precision small-molecule structures, or, in the future, as built from high-resolution small-compound and protein structures (Blessing *et al.*, 2003). Restrained reciprocal-space least-squares refinement (Konnert, 1976; Konnert & Hendrickson, 1980) is very effective and increases the robustness of the convergence.

Stereochemical restraints or constraints are still necessary in the case of macromolecules at ultra-high resolution, as some parts of the protein (often long side chains at the molecular surface) display high thermal motion or static disorder. The atoms with the lowest ADPs are those contributing most to the diffraction at high resolution. The atoms with higher thermal parameters (typically $B = 10$–30 Å^2), despite the resolution of the diffraction data, still have their position and thermal parameters determined with a high uncertainty (Fig. 2).

In the case of small-unit-cell crystals diffracting at subatomic resolution, stereochemical restraints will be necessary to refine properly the hydrogen-atom parameters when no accurate neutron data are available. Restraints application in the context of small-molecule and protein refinement is discussed in §3.

2.5. Solving the normal equations system

In a cycle of least-squares minimization of the function E [equation (7)], the n-dimensional vector $\mathbf{dP}$ of parameter shifts is obtained by solving a system of n linear equations (the so-called 'normal equations'), of the form

$$\mathbf{A}\,\mathbf{dP} = \mathbf{V}, \qquad (8)$$

where $\mathbf{A}$ is the n^2 symmetric positive definite matrix of normal equations and $\mathbf{V}$ is a vector of dimension n.

The normal matrix element A_{ij} concerning the refined parameters p_i and p_j is obtained from the weighted summation of the structure-factor derivative products over the reflections $\mathbf{H}$ and restraints R:

X coordinate e.s.d (10^{-4} Å)
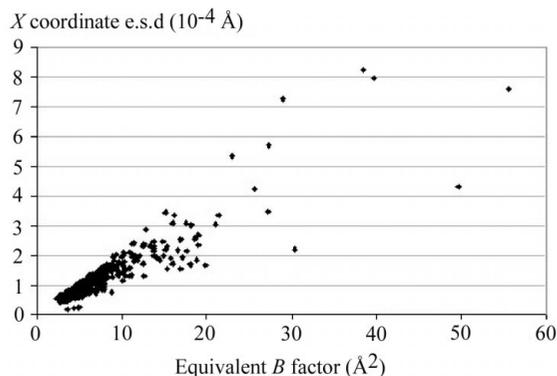
Figure 2
Standard uncertainties of X atom coordinates as a function of equivalent isotropic B factors in the case of aldose reductase at 0.66 Å resolution.

$$A_{ij} = \sum_{\mathbf{H}} W_{\mathbf{H}} \partial F_{\mathrm{cal}}(\mathbf{H})/\partial p_i \partial F_{\mathrm{cal}}(\mathbf{H})/\partial p_j$$
$$+ \sum_{R} W_R \partial R_{\mathrm{cal}}(\mathbf{H})/\partial p_i \partial R_{\mathrm{cal}}(\mathbf{H})/\partial p_j. \qquad (9)$$

The vector $\mathbf{V}$ is obtained from the summation over the reflections and restraints:

$$V_i = \sum_{\mathbf{H}} W_{\mathbf{H}} \partial F_{\mathrm{cal}}(\mathbf{H})/\partial p_i (F_{\mathrm{cal}} - F_{\mathrm{obs}})$$
$$+ \sum_{R} W_R \partial R_{\mathrm{cal}}(\mathbf{H})/\partial p_i (R_{\mathrm{cal}} - R_{\mathrm{obs}}). \qquad (10)$$

This system of normal equations is solved generally in all small-molecule least-squares programs and in *MOLLY* (Hansen & Coppens, 1978) by inverting the normal equations matrix $\mathbf{A}$, which is computationally prohibitive for large systems with many variables. However, one advantage of the full normal matrix-inversion method is the possibility to obtain the standard deviations of the final parameters (Hamilton, 1964) after a refinement cycle.

The matrix inversion can be replaced by the conjugate gradient algorithm: an iterative procedure, first described by Hestenes & Stiefel (1952) and Fletcher & Reeves (1954), which avoids the inversion of the normal matrix and is less sensitive to normal matrix singularities.

The solution of the normal equation system is approximated by successive displacements along $\mathbf{A}$-conjugate directions in the parameter space until convergence. The convergence is usually reached in a number of iterations that is much smaller than the number of variables n. The rate of convergence of the conjugate gradient algorithm is related to the ratio between the largest and the smallest eigenvalue, also called the condition number of the normal matrix $\mathbf{A}$ (Tronrud, 1992).

To ensure a faster and more robust convergence of the conjugate gradient iterations, the normal matrix $\mathbf{A}$ can be preconditioned (Tronrud, 1992) in order to have a condition number closer to unity, *i.e.* a narrower range of eigenvalues. Using this property, it is legitimate to transform the normal matrix $\mathbf{A}$ into a normalized matrix $\mathbf{A}'$ with diagonal elements equal to unity. The preconditioner used in *MoPro* is simply the diagonal part of the matrix $\mathbf{A}$; the normal equation (7) then becomes

$$\mathbf{A}'\,\mathbf{D}\,\mathbf{dP} = \mathbf{D}^{-1}\,\mathbf{V}, \qquad (11)$$

where the matrix $\mathbf{A}'$ is obtained by normalizing the elements with respect to the diagonal values:

$$A'_{ij} = A_{ij}/(A_{ii}A_{jj})^{1/2}. \qquad (12)$$

The matrix $\mathbf{A}'$ is then not far from the identity matrix, especially with high-resolution diffraction data as the normal matrix is very sparse (Jelsch, 2001).

2.6. Normal matrix sparsity

If n is the number of refined parameters, the (symmetric) normal matrix $\mathbf{A}$ contains $n(n + 1)/2$ independent elements. In the case of macromolecules containing several thousands of atoms, it is recommended to omit the off-diagonal elements with small values, as storing the full normal matrix can become

Table 2
List of restraints available in *MoPro*.

Each restraint is weighted by a coefficient $W_R = 1/\sigma_r^2$ depending on the allowed standard deviation σ_r from the target.

Keyword	Description	Minimized function
Targets for atomic parameters		
XYZRES	Coordinates $X\ Y\ Z$	$(X - X_r)^2$
UIJRES	Thermal parameters U_{ij}	$(U_{ij} - U_{ijr})^2$
KAPPA1	Expansion/contraction of valence	$(\kappa - \kappa_r)^2$
KAPPA2	Expansion/contraction of multipoles	$(\kappa' - \kappa_r')^2$
Stereochemistry		
DISTAN	Distance between two atoms	$(d - d_r)^2$
ANGLER	Angle θ between three atoms	$(\theta - \theta_r)^2$
PLANAR	Planarity	$\lambda_1 \lambda_2 \lambda_3 / (\lambda_1 + \lambda_2 + \lambda_3)^3$
SIMDIS	Similarity of several distances	$\sum_i (d_i - \langle d \rangle)^2$
SIMANG	Similarity of several angles	$\sum_i (\theta_i - \langle \theta \rangle)^2$
Thermal motion		
RIGIDB	Hirshfeld rigid-bond criterion	$(Z_U - Z_U')^2$
URATIO	Isotropic thermal factor proportional (k factor) to that of bonding atom	$(U_{\text{iso}} - kU_r)^2$
UIJRAT	Anisotropic thermal factor proportional (k factor) to that of bonding atom	$\sum_{ij} (U_{ij} - kU_{ijr})^2$
SIMUIJ	Similarity of anisotropic thermal factors of two atoms	$\sum_{ij} (U_{ij} - U_{ij}')^2$
ISOTRO	Thermal ellipsoid limited anisotropy	$\sum (U_i - U_j)^2 / \angle U_{ii})^2 = 3 \left[\sum_i (U_{ii} - \langle U_{ii} \rangle)^2 + \sum_{i \neq j} U_{ij}^2 \right] / \langle U_{ii} \rangle^2$
Charge density		
KP1VAL	Correlation of $q = N_{\text{val}} - P_{\text{val}}$ atomic charge and of κ expansion/contraction	$\left[\sum_k (\kappa - \kappa_{\text{target}})^2 \right]^{1/2}$
SIMKAP	Similarity of contraction/expansion	$\sum_{i \neq j} (\kappa_i - \kappa_j)^2$ and $\sum_{i \neq j} (\kappa_i' - \kappa_j')^2$
SIMVAL	Valence and multipole populations of two or more atoms are similar	$\sum_{i \neq j} \left[(P_{\text{val}_1} - P_{\text{val}_2})^2 + \sum_l \sum_m (P_{lm_i} - P_{lm_j}')^2 \right]$
SIMPLM	Multipole populations of two or more atoms are similar	$\sum_{i \neq j} \sum_l \sum_m (P_{lm_i} - P_{lm_j}')^2$
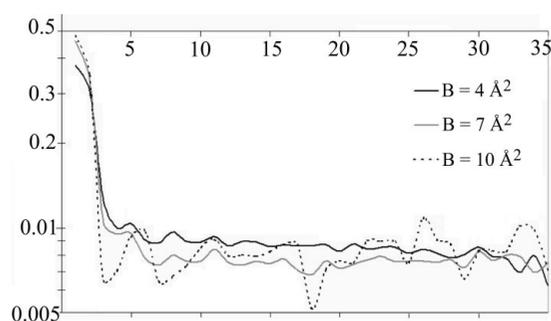
Figure 3
Magnitude of the normalized normal matrix elements in the case of protein trypsin at 0.80 Å resolution (Schmidt *et al.*, 2003) as a function of interatomic distance d_{ij}. The x coordinates of all non-hydrogen atoms with $B_{\text{eq}} < 12$ Å^2 were refined. The matrix elements A_{ij} are normalized with respect to the corresponding diagonal elements: $A_{ij}' = A_{ij}/(A_{ii}A_{jj})^{1/2}$. The elements $A_{X_i X_j}'$ represented here correspond to pairs of (x_i, x_j) coordinates. The interatomic distance d_{ij} is relative to the two atoms concerning the parameters x_i and x_j. Only pairs of atoms with similar B_{eq} factors are taken into account: the three curves correspond to $B_{\text{eq}} = 4, 7$ and 10 ± 0.5 Å^2. The root mean square values of $A_{X_i X_j}$ were computed from shells of 1 Å width in distance and 1 Å^2 width in B_{eq}.

computationally prohibitive. In all the crystallographic refinements of macromolecules at atomic and subatomic resolution performed in the laboratory, the matrix $\mathbf{A}'$ turns out to be very sparse, as nearly all of the off-diagonal elements have very small values compared with unity (Jelsch, 2001).

A normal matrix element A_{ij} associated with a pair of atomic parameters p_i and p_j is consequently related to a pair of atoms a_i and a_j. The elements of the normal matrix have a global tendency to decrease rapidly with the Patterson vector length between the two atoms concerned, as already observed by Templeton (1999).

The diagonal elements of $\mathbf{A}'$ are all equal to unity and the magnitude of an off-diagonal element A_{ij}' can then be assessed by comparison with unity. For example, in the case of crambin at 0.54 Å resolution (Jelsch *et al.*, 2000), the positional parameters of the non-hydrogen protein atoms were refined. The matrix elements are, in their vast majority, very small ($|A_{ij}'| < 0.03$) for interatomic distances d_{ij} longer than 2 Å (Jelsch, 2001). Matrix elements with a large magnitude correspond generally to parameter pairs concerning the same atom, alternative positions for disordered atoms, or atoms linked by a covalent bond.

To analyse the effect of the atomic temperature factors on the sparsity of the normal matrix, a refinement of x coordinates for all non-hydrogen atoms was carried out on trypsin at 0.80 Å resolution (Schmidt *et al.*, 2003). The root mean square (r.m.s.) value of A'_{ij} is around 0.008 for interatomic distances larger than 3 Å (Fig. 3). The r.m.s. magnitude is smaller in the trypsin case compared with crambin (0.004) at 0.54 Å resolution, despite the lower resolution. This is due to the larger number of atoms in the asymmetric unit (388 residues for trypsin) compared with crambin (46 residues), resulting in a larger number of reflections (165 000 *versus* 112 000). A larger number of reflections results in a better canceling out of weakly correlated terms in the summation of equation (9).

The average magnitude of the diagonal elements A_{ii} decreases globally with the B factor of the atoms concerned, resulting in larger standard deviations $\sigma_i = (\mathbf{A}^{-1})_{ii}^{1/2}$ for the refined parameters, as can be seen in Fig. 3. The magnitude of the A'_{ij} ratio [equation (12)] does not seem to be affected by the equivalent B factor of the concerned atoms, as can be seen in the three curves displayed in Fig. 3. The diagonal and non-diagonal elements are dependent on the B factor in a similar way.

The *a priori* knowledge that most of the matrix elements are negligible, together with the application of the conjugate gradient method in the software *MoPro*, avoids the storage, the computation and the inversion of the full normal matrix.

3. Restraints and constraints

The need for restraints and/or constraints in a crystallographic refinement arises from quite different reasons in the two contexts of small-molecule and biological macromolecule crystallography. In *MoPro*, restraints and constraints concern the molecular stereochemistry or dynamics as well as the electron-density description.

In most cases, stereochemical information is not necessary for the non-hydrogen atoms in small-molecule crystallography: diffraction data and the resolution are good enough to obtain an unambiguous description of atomic positions and ADPs. In that case, using significant external information in the refinement may introduce artificial perturbations of the molecular structure. For proteins refined at usual macromolecular resolution (d around 2 Å), such restraints are necessary to overcome a poor parameter to observation ratio, and to ensure convergence towards a stereochemically meaningful structure. However, for medium-sized molecules and macromolecules at atomic resolution, stereochemical information is only necessary in disordered regions, or where the thermal motion is too high. An option in *MoPro* permits the restriction of the application of any type of restraint to selected atoms with thermal motions B_{eq} higher than a user-defined threshold. For example, in the structure of aldose reductase at 0.66 Å, stereochemical restraints are not necessary to refine active-site atoms and were only applied to atoms with $B_{eq} > \sim 8$ Å^2.

3.1. Stereochemical and dynamical restraints

The geometric restraints implemented in *MoPro* concern interatomic distances, bond angles and planarity (Table 2). Most restraints use a simple quadratic function ($R_{cal} - R_{target}$)^2 and seek to minimize the squared difference between them.

Planarity restraints were programmed by considering the eigenvalues λ_i of the 3×3 matrix $\mathbf{V}$ (Haneef *et al.*, 1985):

$$V_{ij} = \sum_{\text{atoms}} (X_i - \langle X_i \rangle)(X_j - \langle X_j \rangle), \qquad (13)$$

where X_1, X_2, X_3 define respectively the x, y, z coordinates of the atoms belonging to the plane. This makes it possible to refine the optimum plane orientation at the same time as the atomic coordinates. The function minimized in *MoPro* is a dimensionless quantity: a function of the λ_i eigenvalues product and sum (Table 2).

The anisotropy of thermal motion ellipsoids can be limited for atoms with high thermal motion (ISOTRO keyword). The Hirshfeld rigid-bond criterion allows restraint of the ADPs (RIGIDB keyword): the thermal ellipsoids of two bonded atoms should be similar along the bond direction. A hydrogen-atom isotropic thermal factor B_{iso} may be restrained to be proportional to the equivalent B factor of the bound heavy atom (URATIO).

The special XYZRES and UIJRES restraints allow specific targets to be set for positional and anisotropic thermal motion parameters, respectively. Such restraints may be useful for the refinement of hydrogen atoms when a neutron diffraction structure is available. The neutron coordinates and thermal displacement parameters with their estimated errors can be directly used as targets for x, y, z and U_{ij}. The ADPs may also be obtained by other means (spectroscopic values, quantum chemical computations).

The Hirshfeld (1976) rigid-bond criterion is sometimes used as a restraint in small-molecule charge-density analysis, as it helps in the high-order refinement of thermal parameters U_{ij} and allows a better deconvolution of deformation electron density and thermal vibrations. This method has recently been used in protein refinement [DELU option in *SHELXL* for instance (Sheldrick & Schneider, 1997)], but may find greater importance in the context of a subatomic-resolution protein multipolar refinement, where the effects of the ADPs on bonding-density modelling are critical.

The isotropic B_{iso} factors of hydrogen atoms can also be constrained to be proportional to the bound heavy-atom displacement parameters. Typically, coefficient factors of 1.2 and 1.5 are applied if the hydrogen-atom position is respectively unique or has a degree of freedom (as in –OH, –NH$_3^+$ and –CH$_3$ groups). Such proportionality restraints are also available for anisotropic thermal parameters and may be applied to virtual atoms, which are used to model the accumulation of electron density on covalent bonds.

3.2. Stereochemical and dynamical constraints on hydrogen atoms

In the macromolecular case, many hydrogen atoms are usually not visible in the electron-density maps, the probability

Table 3
Example of some constraints applicable to the alanine residue of the dipeptide Ala-Met: distances (CONDIS), angles (CONANG), B-factor proportionality (CONURA), atom equivalence (CONVAL, CONPLM, CONKAP) for charge density and local symmetries applying to multipoles (SYMPLM).

CONDIS	ALA	CB	HB1	1.059	!neutron distance			
CONDIS	ALA	CB	HB2	1.059				
CONDIS	ALA	CB	HB3	1.059				
CONDIS	ALA	CA	HA	1.092				
CONANG	AXHHH	ALA	NT	CA	H1	H2	H3	
CONANG	AXYZH	ALA	CA	CB	NT	C	HA	
CONANG	AXHHH	ALA	CB	CA	HB1	HB2	HB3	
CONURA	ALA	CB	HB1	1.5				
CONURA	ALA	CB	HB2	1.5				
CONURA	ALA	CB	HB3	1.5				
CONURA	ALA	CA	HA	1.2				
CONVAL	ALA	HB1	HB2	HB3				
CONVAL	ALA	H1	H2	H3				
CONPLM	ALA	HB1	HA					
CONKAP	ALA	HB1	HB2	HB3				
CONKAP	ALA	H1	H2	H3				
SYMPLM	3m	ALA	NT	! 3m axis along z				
SYMPLM	3m	ALA	CB					
SYMPLM	my	ALA	CA	! mirror $\perp y$				
SYMPLM	mz	ALA	O	! mirror $\perp z$				
SYMPLM	mzmy	ALA	CT	! 2 mirrors $\perp z$ and $\perp y$				

Table 4
Definition of the stereochemical angular constraints CONANG.

A is the central atom of the constrained moiety containing at least one hydrogen atom. X, Y, Z represent the non-hydrogen atoms covalently bound to atom A, while H are the hydrogen atoms (up to three).

Option applied to CONANG keyword	Description
AXHHH 5_atoms	Tetrahedral A atom, XAH = 109.47°, HAH = 109.47°
	Tetrahedral A atom, XAY and HAH planes perpendicular
AXYHH 5_atoms (θ)	HAH optional angle, default 109.47°
AXYZH 5_atoms	Tetrahedral A atom, **HA** $\parallel$ (**AX**/AX + **AY**/AY + **AZ**/AZ)
AXYH 4_atoms	Trigonal A atom, **HA** $\parallel$ (**AX**/AX + **AY**/AY)
AXHH 4_atoms (θ)	Trigonal A atom, HAH optional angle θ, default 120°
AXH 3_atoms (θ)	XAH optional angle θ, default 109.5° (*cf.* COH)
AHH 3_atoms (θ)	HAH optional angle θ, default 104.5° (*cf.* HOH)

of locating hydrogen atoms being strongly dependent on the local thermal motion and on the diffraction data resolution. One commonly applies geometric stereochemical constraints in order to generate ideal geometries for chemical groups containing hydrogen atoms. The geometry of groups containing hydrogen atoms can be regularized in *MoPro* using CONANG and CONDIS constraints (Tables 1 and 3). The idealization of the bond lengths and angles is thus dissociated in two steps. Any group, planar or tetrahedral, containing from one to three hydrogen atoms can be idealized upon application of the constraint by displacement of the hydrogen atoms (Table 4). Such a description of ideal distances and angles enables the compilation of stereochemistry dictionaries for common chemical moieties, such as amino acids in proteins for instance.

The proportionality restraints of isotropic and anisotropic thermal parameters are also available as constraints.

The case of hydrogen atoms is noteworthy as their refined coordinates in X-ray crystallography do not correspond to the true proton position and are strongly affected by the $X-$H covalent bond. The application of standard $X-$H distances obtained from neutron diffraction studies (Allen, 1986) is deemed necessary for both small compounds (restraints) and macromolecules (constraints). This elongation of the $X-$H distance has the effect of 'releasing' the hydrogen-atom bonding electron density, which is then modelled by a dipole directed along the bond.

The treatment of hydrogen atoms is different in small-molecule charge-density and in protein refinement. In most cases of small-molecule crystals, hydrogen atoms can be located in electron-density maps when the resolution is truly subatomic (around 0.5 Å) and the thermal motion is around B_{eq} = 1 or 2 Å^2. In these favourable cases, an $X-$H distance and temperature factor proportionality restraint is recommended for a proper charge-density refinement with a limited number of variables (P_{val}, one dipole Dz, κ, κ').

In the macromolecular case, when the hydrogen atoms are hardly visible or not visible in the electron density, their positions are conveniently fixed using distance and angle constraints. Furthermore, the hydrogen thermal isotropic motion should be set proportional to that of the bound non-H atom, with a ratio of 1.2 or 1.5, depending on whether the hydrogen atom has a rotation degree of freedom ($-$CH$_3$, $-$OH) or not. Such constraints are also necessary in the case of small molecules when the thermal motion is larger or when static disorder occurs.

3.3. Charge-density parameters

Some important constraints defined in *MoPro* are related to the charge-density parameters. Due to limitations in the

quality of the diffraction data (resolution, Fourier series truncation effects, signal to noise ratio, *etc.*), it is generally not realistic to refine all charge-density parameters together, since there is not necessarily a single minimum in the least-squares refinement and multipolar parameters can be very strongly correlated (valence population and kappa parameter for example) or with ADPs (U_{ij} and quadrupoles for example). Constraints of atomic chemical equivalence and of local symmetries are useful to guide the refinement in its first steps and may be released later on. Multipole refinements of electron-density distributions may lead to ambiguous or even meaningless results for non-centrosymmetric crystal structures, where most phases are unknown. The application of constraints on the electron-density models improves this situation (El Haouzi *et al.*, 1996).

In the macromolecular case, the transfer of database multipole parameters (Pichon-Pesme *et al.*, 1995; Pichon-Pesme *et al.*, 2004) ensures a better convergence of a constrained charge-density refinement to a physically meaningful solution.

The use of external information obtained from accurate small-molecule crystallographic analyses, such as stereochemical restraints in a macromolecular refinement, can be extended to charge-density parameters. The case of the parameters κ and κ' describing the radial expansion/contraction of the valence electron density is noteworthy, these features being often the most difficult to estimate in electron-density analyses, even for small molecules. Pérès *et al.* (1999) and Volkov *et al.* (2000) focus particularly on the radial expansion κ' of the multipolar deformation density and
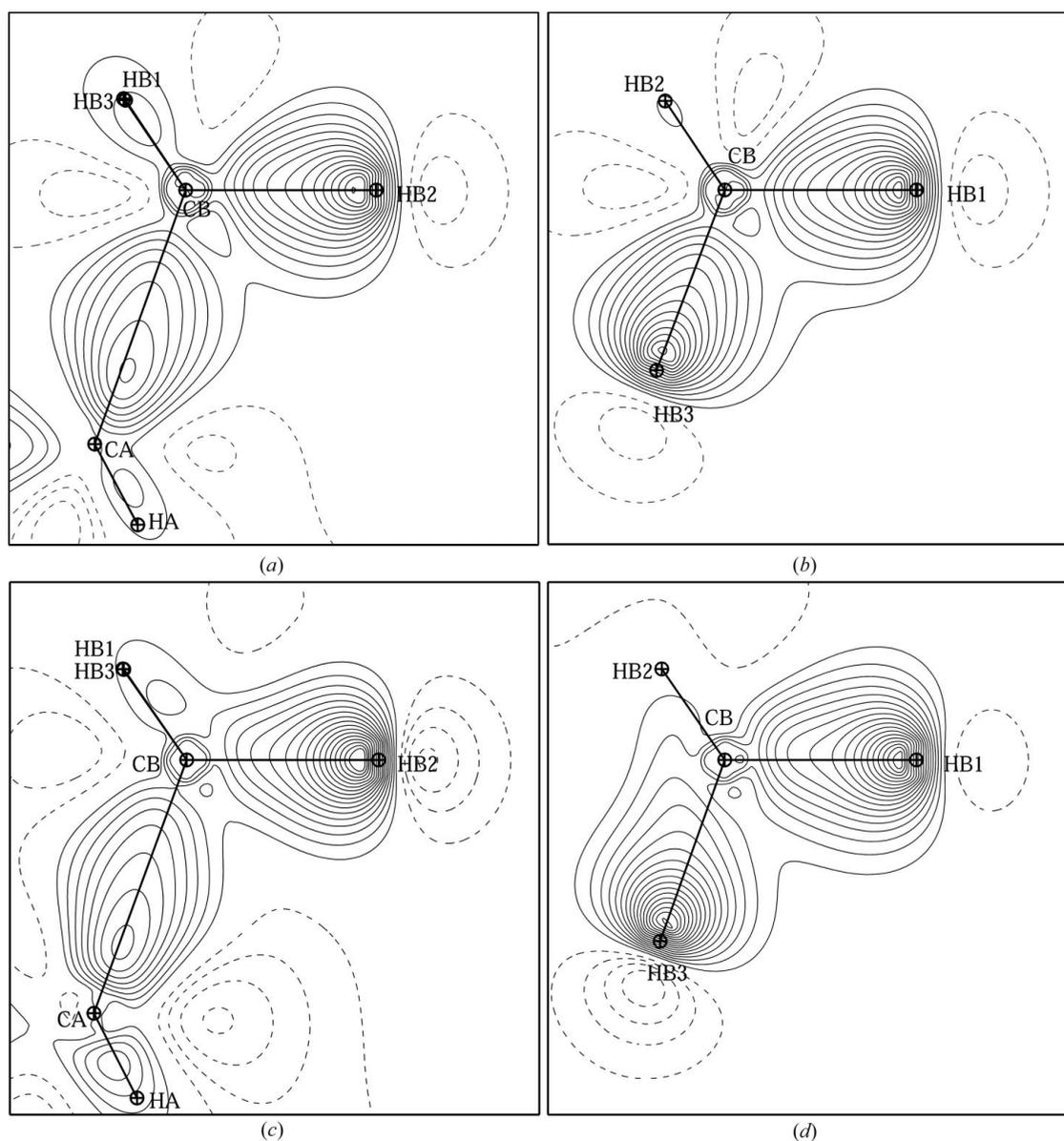
Figure 4
Deformation of the static electron density of the methyl group for the dipeptide DL-Ala-Met. Top: constrained; bottom: unconstrained. Contours ±0.05 e Å^{-3}, continuous lines $\delta\rho > 0$, broken lines $\delta\rho < 0$.

Table 5
List of structure geometry analysis commands available in *MoPro*.

When a refinement of the concerned parameters is performed before the analysis, estimated standard deviations (e.s.d.s) are given for some stereo-chemical values. A symmetry code may be added in the case of contact and hydrogen-bond analysis to restrict the search to a given symmetry-related neighbouring molecule in the crystal packing. The symmetry nomenclature is similar to that developed for the program *ORTEP* (Burnett & Johnson, 1996).

Command	Information obtained
WRIT AXES	Atomic local axes
WRIT DIST	All bond distances and e.s.d.s
WRIT ANGL	All angles and e.s.d.s
WRIT DIHE	All dihedral angles and e.s.d.s
WRIT CONT d_{max} (SYM)	All packing contacts with distance < d_{max}
WRIT HBON d_{max} (SYM)	All hydrogen bonds in structure with distance (acceptor–donor) < d_{max}
WRIT U_{ij}	Thermal U_{ij} ellipsoid eigenvalues and expression of U_{ij} in an orthogonal axes system
WRIT RIGB	All Hirshfeld rigid-bond values
SHOW DIST at1 at2 (SYM)	Specified distance, angle or hydrogen-bond information; the last atom may be symmetry related if specified
SHOW ANGL at1 at2 at3 (SYM)	
SHOW HBON atD atH atA (SYM)	
SHOW PLAN 4 ASP CB CG OD1 OD2	Equation of mean squares plane formed by the atoms and distance of atoms to the plane

suggest constraining that parameter. The expansion/contraction parameters can be restrained in *MoPro* during the multipolar refinement (Table 2). For instance, the selected target may be values in the database of experimental multipole parameters (Pichon-Pesme *et al.*, 1995; Pichon-Pesme *et al.*, 2004). Targets for the κ and κ' variables can also be deduced from theoretical calculations (Volkov *et al.*, 2000; Volkov *et al.*, 2001). Unlike a constrained refinement, where some values are fixed, the refinement with restraints allows parameters to adjust with respect to the specific chemical environment of the considered atom. For instance, restraining the κ coefficient of the pyrophosphate atoms was necessary in the multipolar refinement of NAD^+, the oxidized form of the nicotinamide adenine dinucleotide molecule (Guillot *et al.*, 2003). This was the first application of a charge-density restraint in a small-molecule multipolar refinement.

The refined expansion/contraction κ coefficients of atoms turn out to vary almost linearly with the atomic charge, $q = N_{val} - P_{val}$ (Volkov *et al.*, 2001). As a consequence the refinement of these two classes of parameters may be restrained for a group of atoms of one or several chemical types. Such a linearity restraint is particularly useful in the case of hydrogen atoms for which κ are not well defined or for large-size molecules. This restraint aims at obtaining a correlation coefficient $\rho(\kappa, q)$ between the atomic charges and κ coefficients close to unity.

The minimization of the quadratic quantity $(1 - \rho)$ with a zero target value could be a valid way to apply this correlation restraint, but it leads to unstable refinements. Therefore, a target value is computed for each atomic charge q in order to fulfil this correlation. The discrepancy of each charge q from its target value is minimized separately (Table 2):

$$\kappa_{target} - \langle \kappa \rangle = (q - \langle q \rangle) \sigma_\kappa / \sigma_q. \tag{14}$$

In that way, the standard deviation within the κ_{target} sample is the same as the standard deviation σ_κ of the κ sample. At the beginning of the refinement, when atoms are neutral, the sample standard deviation σ_q is zero and the restraint is meaningless and not applied. The restraints may be applied to all the atoms, or to one or several specified chemical atom types: for example, hydrogen atoms which display higher correlation between refined parameters in the normal matrix and are more difficult to refine.

Atomic chemical equivalence uses the *a priori* knowledge that identical chemical moieties should share, to a first approximation, the same charge-density distribution. Chemical-equivalence constraints are especially appropriate in the case of biological macromolecules (proteins, nucleic acids), when advantage can be taken of the repetition of identical chemical motifs.

Local or crystallographic symmetry (Kurki-Suonio, 1977) constraints, applied to the multipolar charge density, are easily defined with an appropriate system of atomic local axes. For instance, *mm* symmetry may be applied to a water oxygen atom, as the HOH moiety displays two perpendicular mirror planes (Fig. 1). With these symmetry constraints, the number of parameters refined is divided by 4 and the charge-density parameters that are expected to be small are not refined initially. The symmetry constraints could alternatively be applied by never refining the multipoles breaking the symmetries, setting them at zero values. For many chemical groups, local symmetries can be exploited to limit the number of refined multipoles and to improve the refinement convergence. For instance, in the case of $-NH_3^+$ and $-CH_3$ groups, the SYMPLM 3m keywords allow a 3m symmetry axis parallel to the Z axis to be imposed on the specified atoms; the number of refined multipoles (up to the octapole level) is consequently decreased from 15 to 4.

The example of the $C\alpha-C\beta H_3$ group in the DL-Ala-Met dipeptide (Guillot, Viry *et al.*, 2001) is depicted in Fig. 4. This non-polar group is expected to be little influenced by the chemical environment. The side chain of alanine is therefore highly transferable and follows closely a local 3m symmetry, with the three hydrogen atoms being equivalent. The release of the local symmetry (3m axis on atom CB) and atom-equivalence (H1, H2 and H3) constraints in the refinement leads to differences in the deformation electron density between the three hydrogen atoms which are not significant. The diffraction data were measured for this peptide at 100 K using a laboratory X-ray source and a two-dimensional CCD detector up to a resolution of 0.43 Å. The degree of release of the symmetry and chemical-equivalence constraints has to be decided in a charge-density analysis depending on the quality of the data in terms of resolution, signal to noise ratio and systematic errors.

MoPro is designed to apply many symmetries and chemical equivalencies, which is especially convenient for large systems using a 'constraints dictionary' with atom names following the standard nomenclature (Tables 1, 3 and 4). *MoPro* allows a

large flexibility in restraints and constraints, as they can be applied or released when the program encounters specific keywords in the input command file during execution.
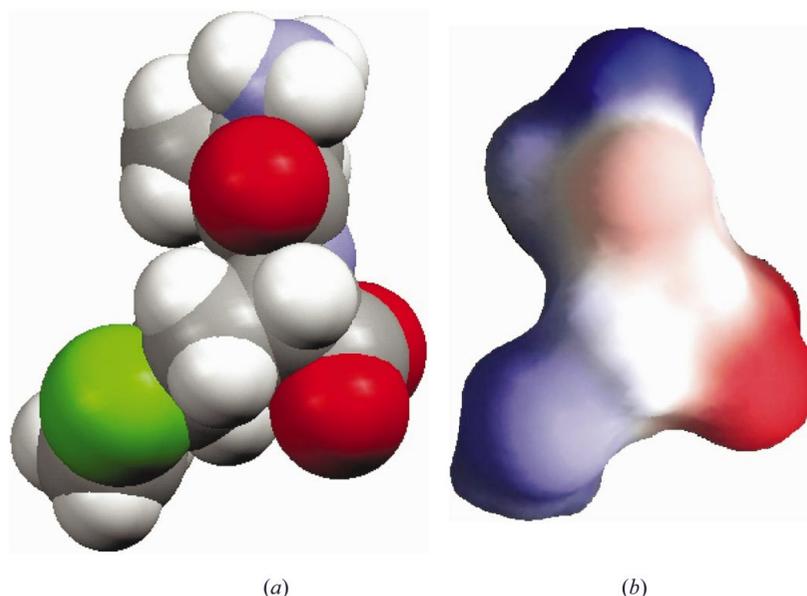
3.4. Similarity restraints

A new type of restraint has been introduced in *MoPro*: similarity restraints. They concern both stereochemical (distance and angles) and charge-density parameters (κ, κ', P_{lm} and P_{val}). They are based on the assumption that equivalent chemical groups should, to a first approximation, share common properties. Similarity restraints avoid the use of information that is 'external' to the diffraction data. They are particularly useful in molecules with repeated motifs, like polypeptides or polymers, as they can be applied to a large number of atoms simultaneously. In the similarity restraints, the quadratic function to be minimized is summed over all concerned atoms and the target is the average of the parameters (Table 2).

The development of similarity restraints arises from our current attempts to extract deformation electron-density information from 0.9–0.6 Å resolution protein structures. They are also being tested for multipolar parameters at subatomic resolution (0.6–0.7 Å) as an interesting and smoother alternative to chemical-equivalence constraints.

4. *VMoPro*: a visualization tool

The program *VMoPro* has been developed to allow easy computation and graphical representation of electron-density maps and of derived properties. *VMoPro* is an interactive program, which many types of electron-density map to be

drawn step by step or by automated batch execution. After a crystallographic refinement (notably a high-order refinement of coordinates and thermal parameters of non-hydrogen atoms), difference electron-density maps using the Fourier structure-factor difference ($F_{obs} - F_{cal}$) can be computed:

$$\Delta\rho_{res}(\mathbf{r}) = V^{-1} \sum_{\mathbf{H}} \left[k^{-1}|F_{obs}(\mathbf{H})| - |F_{cal}(\mathbf{H})| \right] \exp i(\varphi_{cal} - 2\pi\mathbf{H} \cdot \mathbf{r}). \tag{15}$$

After a refinement at ultra-high resolution using a spherical-atom model, the residual maps reveal, for instance, the non-modelled bonding electron density.

The deformation of the electron density reveals the valence electron redistribution due to covalent and non-bonded interactions between atoms. The 'experimental electron deformation density' maps are computed using the following equation:

$$\Delta\rho_{def}(\mathbf{r}) = V^{-1} \sum_{\mathbf{H}} \left[k^{-1}|F_{obs}(\mathbf{H})| \exp(i\varphi_{mult}) - |F_{sph}(\mathbf{H})| \exp(i\varphi_{sph}) \right] \exp(-2i\pi\mathbf{H} \cdot \mathbf{r}). \tag{16}$$

In this case, the φ_{mult} phases are calculated from the structure after multipolar refinement. Spherical atomic scattering factors are used in the calculation of F_{sph} and φ_{sph}. The dynamic deformation differs from the static map, as it is smeared by the atomic thermal motion and more directly affected by the truncation of the Fourier summation over the measured diffraction data.

Static electron-density maps representing the total or the deformation electron density are computed from the charge-density parameters and do not make implicit use of the diffraction intensities. The word 'static' means that the charge density is computed for atoms at rest. The contribution of the whole electron-density modelling or of a specified level of multipoles may be selected:

$$\delta\rho(\mathbf{r}) = \sum_{j=1,Nat} \left[\rho_{multipolar}(\mathbf{r}-\mathbf{r}_j) - \rho_{free-atoms}(\mathbf{r}-\mathbf{r}_j) \right]. \tag{17}$$

VMoPro also computes the total static electron density, its gradient norm and Laplacian. These functions of the electron density allow the observation of topological features, such as critical points, where $\nabla\rho = 0$, and atomic basins limits (Bader, 1990; Souhassou & Blessing, 1999).

The electrostatic potential plays a key role in molecular recognition processes, including drug–receptor interactions, and is an important property for the evaluation of lattice energy in small-unit-cell crystals (Aubert *et al.*, 2004; Spackman, 1992; Volkov & Coppens, 2004).

The use of a precise electron-density model is essential for accurate electrostatic potential computation. The difference between potential maps derived from multipolar- and spherical-

| (a) | (b) |

Figure 5
Electrostatic potential of the DL-Ala-Met dipeptide derived from the refined structure using a multipolar-atom model (Guillot, Viry *et al.*, 2001). The potential is computed with *VMoPro* in a three-dimensional grid. The program *GRASP* (Honig & Nicholls, 1995) reads this grid and enables the electrostatic potential at the van der Waals molecular surface to be displayed.

Table 6
Example of some commands available in the input file 'mopro.inp'.

First, input files and options are defined. In the second part, the charge density and the structural/thermal parameters of hydrogen atoms are refined using all reflections. In the third part, a high-order refinement is performed on the coordinates and thermal factors of non-hydrogen atoms. Finally, the refinement part is followed by a stereochemical analysis and structure-factor computations.

Command example	Description
PARA peptide.01	Molecular structure and charge-density parameters file
DATA peptide.Ihkl	Diffraction data file
SIGC 0.	Intensity over σ cut-off.
VERB SHEL 10	Show refinement statistics in ten resolution shells
RESO 0.5 100.	Use reflections in resolution shell specified in Å
SELE SCA XYZ UIJ HYD	Select parameters to refine: scale factor, coordinates and thermal factors of hydrogen atoms
REFI CG 2	Perform two cycles of least-squares refinement with conjugate gradient algorithm
SELE PLM	Select parameters to refine: multipoles
REFI CG 2 DAMP 0.8	Perform two conjugate-gradient refinement cycles, applying a damping shifts factor
RESO 0.5 0.7	Use reflections in specified high-resolution shell
SELE XYZ UIJ NOH	Select parameters to refine: coordinates and thermal factors of non-hydrogen atoms
REFI CG LS 2	Perform two cycles of least-squares refinement with full matrix inversion
WRITE DIHE	Analyse all dihedral angles in structure
WRITE FOUR mul.Fhkl	Write structure factors to specified file
FREE	Set atoms free and neutral: $\kappa = 1$, $\kappa' = 1$, $P_{val} = N_{val}$, $P_{lm} = 0$
WRITE FOUR sph.Fhkl	Write structure factors of free-atom structure
STOP	Interrupt instruction reading in 'mopro.inp'

Table 7
Keywords in *MoPro* allowing easy selection of parameters to refine.

Keyword	Description
XYZ	Coordinates X, Y, Z
UIJ	Thermal displacement parameters U_{ij} or U_{iso}
OCC	Occupation factors (only those lower than unity)
VAL	Valence populations
PLM	Multipole populations P_{lm}
SCA	Scale factors
SW1	Solvent coefficient K_{sol}
SW2	Solvent coefficient B_{sol}
UOV	Overall thermal factor
EXT	Extinction (Gaussian isotropic)
ANH	Anharmonic coefficients
WAT	Water atoms only
DIS	Disordered atoms only
VIR	Virtual atoms only
NOW	No water atoms
NOD	No disordered atoms
NOV	No virtual atoms
ISO	Atoms with isotropic thermal parameters only
ANI	Atoms with anisotropic thermal parameters only
B>##	Atoms with B_{eq} larger than a given value
B<##	Atoms with B_{eq} smaller than a given value

atom models reaches ~28% (Jelsch *et al.*, 1998). Electrostatic potential map computations in *VMoPro* are derived from the program *ELECTROS* (Bouhmaida *et al.*, 1999) and display isocontour lines in a chosen plane section. The *GRASP* software (Honig & Nicholls, 1995) is widely used to compute and display the potential at the molecular surface and to analyse the electrostatic properties of protein structures. A three-dimensional grid file suitable for *GRASP* can be easily produced with the program *VMoPro* (Fig. 5).

5. Strategy for charge-density refinement of small-unit-cell crystals

The deconvolution of the atomic thermal displacement from the deformation electron density is usually performed by using different resolution shells depending on the parameters refined. A standard procedure may be written in the 'mopro.inp' input command file of *MoPro* for the global refinement of the molecular charge-density refinement (see Table 6). The variables to be refined are easily designated and combined using keywords (see Table 7). Hydrogen, virtual or water atoms may be selected or excluded in the parameter selection. Other criteria, such as atom (an)isotropy, disorder, *B*-factor value, may be used as selection criterion. A standard charge-density refinement strategy may be as follows.

(i) First, the scale factors SCA and coordinates XYZ and thermal parameters UIJ of all non-hydrogen atoms are refined using all reflections.

(ii) Then a refinement of XYZ and UIJ of non-hydrogen atoms is performed at high order (high resolution), for example using reflections with $s = \sin\theta/\lambda > 0.7$ Å^{-1}.

(iii) Then SCA and the hydrogen-atom XYZ and UIJ are refined using low-order data (*e.g.* $s < 0.7$ Å^{-1}). Restraints on distance, angles and planarity are applied to hydrogen-atom positions, while their isotropic thermal factors may be restrained to be proportional to that of the neighbouring atom (Table 2).

(iv) The valence and multipoles populations VAL, PLM, are refined using all reflections. These parameters are progressively and successively introduced in the refinement while steps (ii) and (iii) are continued.

(v) The expansion/contraction coefficients KP1 and finally KP2 are also refined *versus* all reflections. Steps (ii), (iii), (iv) and (v) are recycled until convergence of the refinement. At each stage of the refinement, residual maps and dynamic deformation maps are calculated to monitor the refinement quality.

The refinement strategy may be rendered more complex with the adjunction and progressive release of constraints and restraints. The order of introduction of VAL, PLM, KP1 and KP2 charge-density parameters into the refinement may be changed depending on the size of the molecule and the nature of the atoms. However, KP2 is generally refined last, as this parameter may be the most delicate to refine. The initial KP1 values of hydrogen atoms can be set to 1.16 instead of unity (Stewart *et al.*, 1965), as this chemical species is expected to be electron depleted and the electron density therefore

contracted. Depending on the quality of the diffraction data and on the availability of neutron data, the contraction/expansion coefficients for hydrogen atoms can be fixed at recommended values (Coppens, 1997; Volkov *et al.*, 2001).

The program enables the analysis of the structure's stereochemistry (Table 5) and retrieves notably all bond lengths, angles and dihedral angles, as well as a list of intermolecular contacts. If a refinement with full normal matrix inversion of the appropriate structural parameters is performed before the analysis, the estimated standard deviations are also given (Hamilton, 1964; Fig. 2).

6. Virtual atoms on bonds and electron lone pairs

The electron density can also be refined in *MoPro* using virtual atoms in addition to the atoms constituting the molecule. These additional atoms are placed on the covalent bonds and on the oxygen/nitrogen/sulfur lone-pair regions. The virtual atoms can be described as the second term of the second member of Hansen & Coppens equation: a spherical valence population P_{val} and an expansion/contraction coefficient are refined:

$$\rho_{atom}(\mathbf{r}) = P_{val}\kappa^3\rho_{val}(\kappa\mathbf{r}). \qquad (18)$$

Such a modelling of the molecular electron density was described over 25 years ago in several papers (*e.g.* Dietrich & Scheringer, 1978; Scheringer *et al.*, 1978). It was reapplied recently in macromolecular refinement (Afonine *et al.*, 2002, 2004) and in materials science, on AsGa crystals (Pietsch & Hansen, 1996).

The scattering factor $f(s)$ of the virtual atom placed on a covalent bond or at a lone-pair site can be set by the user. A model of a virtual atom is, for example, the hydrogen atom with the starting number of valence electrons N_{val} set to 0

instead of 1. A virtual atom with a negative valence population may also be placed on the electron depletion region of a C=O bond near the oxygen atom (Fig. 6).

The coordinates of the bonding virtual atoms can be efficiently guided to remain along a covalent bond using an angular restraint. The position of electron lone pairs can be stabilized with angular, distance and planarity restraints. The anisotropic thermal displacement parameters of the bonding virtual atoms are restrained to be similar to the parameters of the two atoms forming the covalent bond. The resulting target value is then the average U_{ij} over the two atoms. Electron lone pairs may have thermal factors riding on the real atom. The use of virtual atoms, compared with multipole modelling of the electron density, results in a reduced number of parameters to refine, but is not a better modelling of the 'true' electron density. It is desirable that such bonding-density modelling fulfils the topological properties of electron densities (Bader, 1990).

7. Multipolar library and transferability

High-resolution X-ray diffraction studies have been performed in our laboratory on several peptides in order to analyse the electron-density distribution of all natural amino acids (*e.g.* Souhassou *et al.*, 1991; Pichon-Pesme *et al.*, 1992; Benabicha *et al.*, 2000; Guillot, Viry *et al.*, 2001). These studies have led to the creation of a database (Pichon-Pesme *et al.*, 1995, 2004) of atomic charge-density parameters P_{val}, P_{lm}, κ and κ'. These parameters were shown to be transferable to amino acid residues in oligopeptides (Pichon-Pesme *et al.*, 1995; Jelsch *et al.*, 1998) and proteins (Housset *et al.*, 2000; Jelsch *et al.*, 2000). Fig. 7 gives, for instance, the static deformation electron density calculated from this multipolar-parameter library for the side chain of an alanine residue. The

(a) (b)
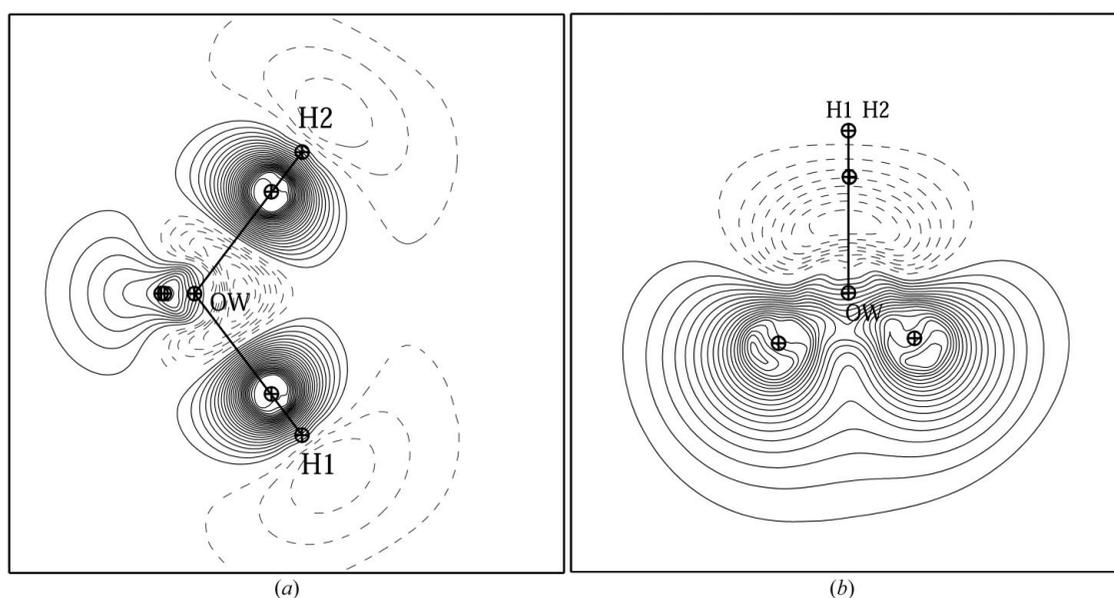
Figure 6
Static deformation density of the water molecule in crystals of Gly-L-Thr dihydrate. Left: the H1–OW–H2 plane. Right: plane formed by the oxygen atom and the two virtual atoms modelling the two electron lone pairs. The two lone-pair virtual atoms are constrained to have identical charge-density parameters (P_{val} and κ) as well as the two OW–H bond virtual atoms. Contours ± 0.05 e Å^3, continuous lines $\delta\rho > 0$, broken lines $\delta\rho < 0$.
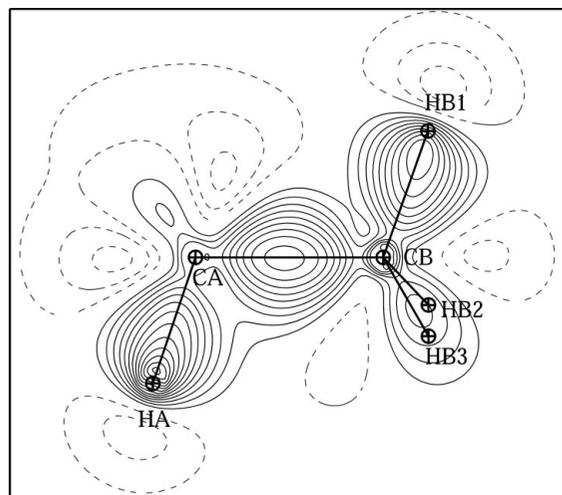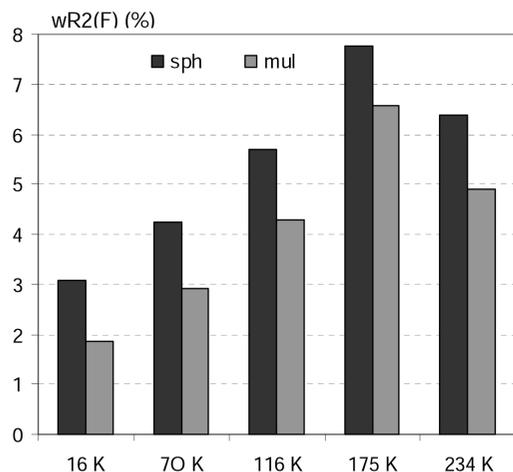
Figure 7
Deformation electron density in the HA–CA–CB plane of alanine calculated from the multipolar library (Pichon-Pesme *et al.*, 1995, 2004). Contours ±0.05 e Å^3, continuous lines $\delta\rho > 0$, broken lines $\delta\rho < 0$.

electrons build up in the bonding region, while the hydrogen atoms are electron depleted.

The experimental deformation density of small-unit-cell crystals is also used to calibrate theoretical electron-density calculations, such as in the case of small blocked peptides (Souhassou *et al.*, 1991; Pichon-Pesme *et al.*, 1992) and the NAD$^+$ oxydo-reduction cofactor (Guillot *et al.*, 2003). After calibration, first-principles computations can be performed on large systems like protein active sites, *e.g.* aldose reductase complexed with NAD$^+$ and an inhibitor (Muzet *et al.*, 2003).

Density functional calculations have also been performed on crambin *in vacuo* (Fernandez-Serra *et al.*, 2000) in order to analyse the polypeptide main-chain variability in particular. The electron deformation density in its peptide bonds
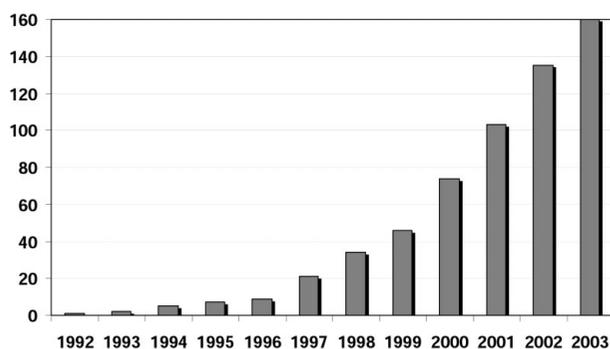

Figure 8
Crystallographic residual factor when using a spherical- and multipolar-atom model for the (*E*)-2,2′-dimethylstilbene compound. The five different temperatures of the diffraction data sets are indicated. $wR2(F) = \left[\sum_H W_H (F_{obs} - F_{cal})^2 / \sum_H W_H F_{obs}^2\right]^{1/2}$, where W_H is the reflection weighting scheme applied.

compares well with the electron distribution obtained from ultra-high-resolution X-ray crystallography. Oxygen lone pairs are generally stronger in theoretical maps compared with the experiment; these fine features are presumably attenuated due to the limited diffraction resolution and thermal motion smearing. The theoretical analysis of several peptide bonds suggests that the carbonyl oxygen and N–H regions display a significant variability of their electron density. The presence of polar interactions like hydrogen bonds seems to be an environmental effect of primary importance that perturbs the electron cloud of these moieties. The transferability of electron densities between chemically similar groups ends when such polarization effects start to be taken into account. This has been observed in experimental studies of small compounds: carbonyl oxygen atoms can display very dissimilar electron lone pairs due to hydrogen bonding, as observed in the case of DL-Ala-Met (Guillot, Viry *et al.*, 2001).

A new example of successful application of the transferability principle is the study of the (*E*)-2,2′-dimethylstilbene molecule (Nyffeler *et al.*, 2002). Diffraction data sets of this small molecule were collected at five different temperatures ranging from 16 to 234 K (Ogawa *et al.*, 1992). The main purpose of this study was the analysis of the molecular dynamics and stereochemistry from multi-temperature anisotropic displacement parameters (Bürgi & Capelli, 2000; Nyffeler *et al.*, 2002). After a conventional spherical-atom refinement, the multipoles were initially transferred from the charge-density library and subsequently refined using a maximal number of constraints: atom equivalences and local mirror symmetries. Indeed, this procedure allows a quick and satisfactory modelling of the deformation density, and thus a proper deconvolution between thermal motion and deformation electron density. This is confirmed by the significant improvement of crystallographic residual (Fig. 8). The crystallographic residual factor drops by about 1% in absolute value at all temperatures; in relative value the improvement is highest at the lowest temperatures. The ΔB_{ij} difference ($B_{ij} = 8\pi^2 U_{ij}$) between the spherical- and multipolar-atom modelling is expected to be about 0.4 Å^2 as analysed previously on an octapeptide (Jelsch, *et al.*, 1998). The difference $\Delta B_{ij} = \left[\sum_{ij}(B_{ij} - U'_{ij})^2/3\right]^{1/2}$ is 0.25 and 0.31 Å^2 for (*E*)-2,2′-dimethylstilbene at the lowest and highest temperature, respectively. The ratio of the equivalent B factor using the multipolar- *versus* the spherical-atom model is on average $B_{mul}/B_{sph} = 0.80$ and 0.87 at 16 and 234 K, respectively. The impact of multipolar modelling on the refined anisotropic U_{ij} values is very significant at the lowest temperature, where the average B_{eq} value is only 0.6 Å^2. Therefore, such a procedure should be routinely applied in crystallographic refinement of small molecules.

8. Applications to proteins

For a charge-density study with *MoPro*, a protein structure can be imported from *SHELXL* (Sheldrick & Schneider, 1997) or from a PDB format (Table 8). The local axes system and the charge-density parameters of the protein atoms, with

Figure 9
Cumulated number of macromolecular structures deposited at the Protein Data Bank (PDB; Berman *et al.*, 2000) at a resolution better than 1 Å as a function of year.

Table 8
Some data import/export possibilities in *MoPro*.

| Action	Command	Description
Export structure file	WRITE CIF	Crystallographic Information File
	WRIT PDB	Protein Data Bank format (Berman *et al.*, 2000)
	WRIT XYZ	*XYZ* format
	WRIT MOLL	*MOLLY* format (Hansen & Coppens, 1978)
Export reflections file	WRIT FOUR	For Fourier map computations with *VMoPro*
	WRIT FCNS	*CNS* format (Brünger *et al.*, 1998)
	WRIT FCF3	*SHELXL*97 formats
	WRIT FCF6	(Sheldrick & Schneider, 1997)
Import structure files	PDB	Protein Data Bank format
	SHELXL	*SHELXL*97 .ins or .res format
	CIF	Crystallographic Information File
	MOLLY	*MOLLY* format

the standard atom name nomenclature, are automatically transferred from the multipolar library (Pichon-Pesme *et al.*, 1995, 2004).

Our laboratory has been involved in charge-density studies on molecules of increasing size: the octapeptide LBZ of helical structure (Jelsch *et al.*, 1998) and a scorpion toxin (Housset *et al.*, 2000). With the combined use of synchrotron radiation sources and of crystal cryo-cooling, the number of protein structures refined at a resolution higher than 1.0 Å is increasing continuously (Fig. 9; Dauter *et al.*, 1995, 1997; Longhi *et al.*, 1998; Schmidt & Lamzin, 2002).

The charge density of the protein crambin (46 amino acids) has been analysed (Jelsch *et al.*, 2000) at ultra-high resolution (0.54 Å). The crystal structure was refined with a multipolar-atoms model to describe the molecular electron-density distribution accurately. The initial multipoles and charges were transferred from the database of average parameters (Pichon-Pesme *et al.*, 1995, 2004) derived from the analysis of several crystals of amino acids and small peptides. The average electron density for the polypeptide main chain refined against the crambin diffraction data is illustrated in Fig. 10 (Jelsch *et al.*, 2000).

A more recent application is the case of the phospholipase PLA2 from *Agkistrodon acutus* snake venom, for which the crystal structure was determined at 0.80 Å resolution (Liu *et al.*, 2003). The bonding electron density is visible in the residual electron density ($F_{obs} - F_{cal}$) maps for most of the structure, which displays moderate thermal motion.

In the case of trypsin from *Fusarium oxysporum*, diffraction data were measured to 0.80 Å resolution (Schmidt *et al.*, 2003). The electron density in the active site of the ultra-high-resolution trypsin structure shows that the catalytic residue His-56 is definitely deprotonated, despite its interaction with
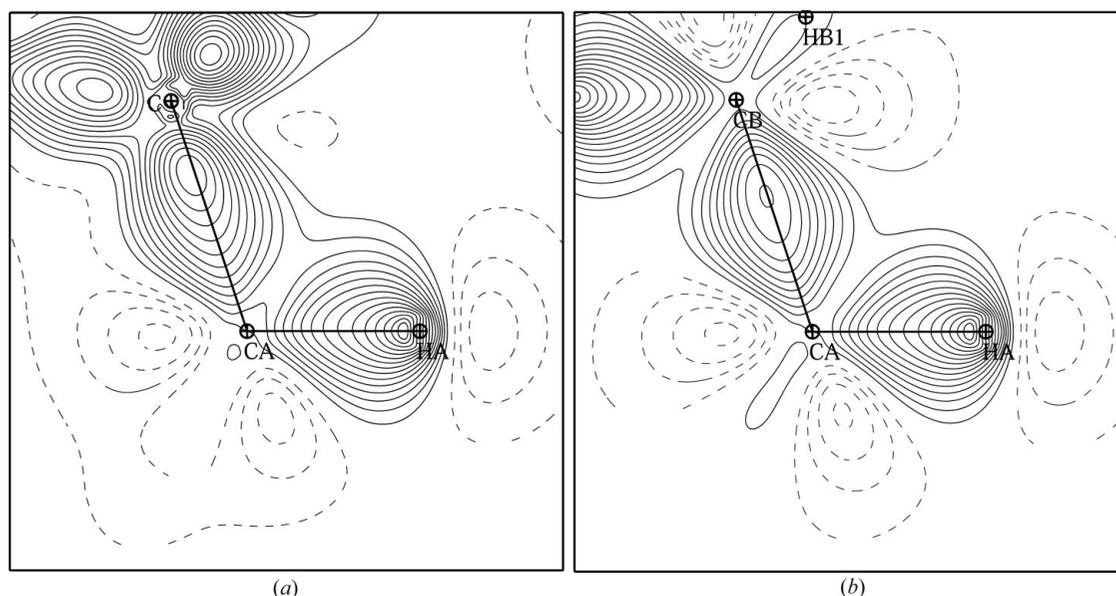
Figure 10
Deformation of the static electron density in the Cα–Hα–C and Cα–Hα–Cβ planes of the polypeptide refined *versus* the diffraction data of protein crambin (Jelsch *et al.*, 2000). Contours ±0.05 e Å^{-3}, black lines $\delta\rho > 0$, grey dotted lines $\delta\rho < 0$.
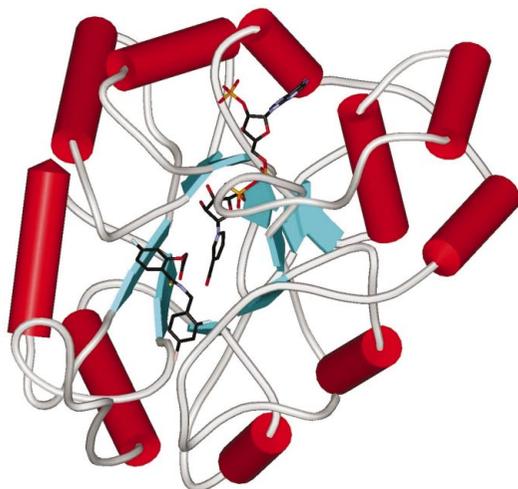
Figure 11
Schematic view displaying the three-dimensional α,β barrel structure of human aldose reductase. The complexed cofactor NADP$^+$ and the inhibitor IDD-594 are at the top and bottom of the active site, respectively. Helices are represented as red tubes and the β strands of the antiparallel β sheet as blue arrows.

the negatively charged Asp-99. Another application is performed with the allosteric insulin hexameric structure, solved at atomic resolution (Smith *et al.*, 2003).

9. The aldose reductase enzyme

The analysis of the electron-density distribution is also underway for the human protein aldose reductase–NADP$^+$–inhibitor ternary complex (Fig. 11). Crystals diffracting to very high resolution have been grown for this enzyme of 315 amino acids and diffraction data have been collected to 0.66 Å resolution at the APS Synchrotron (Cachau *et al.*, 2000; Howard *et al.*, 2004).

This exceptional system, in terms of resolution and thermal motion parameters, is used to test and improve specific refinement strategies of large biological complexes. Some general considerations about multipolar refinement strategies, based on the experience acquired in the analysis of this aldose
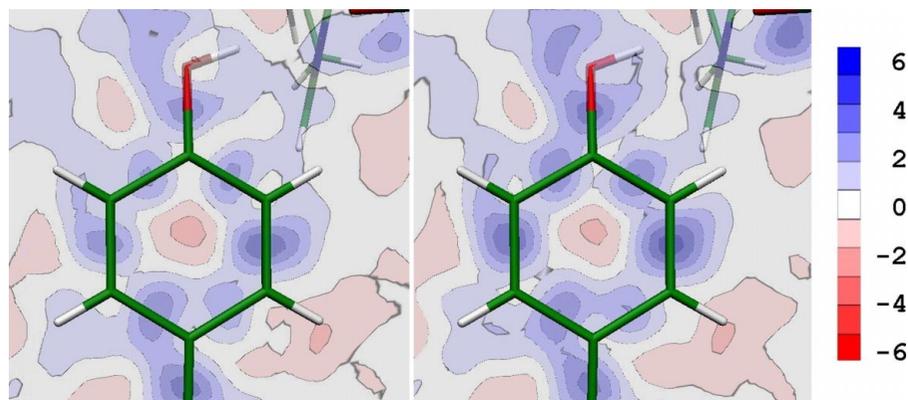
reductase complex, are presented here. A strategy which has been shown to be well adapted to this kind of system is described below.

9.1. Spherical-atom refinement

At first, starting position and thermal motion parameters are taken from a completed usual (spherical-atom) refinement, performed with any crystallographic refinement software. The refinement of solvent parameters, namely B_{sol} and K_{sol}, is not straightforward, mainly due to the strong correlation between both parameters and the overall scale factor. The best procedure is to perform a systematic search by slight increment of these parameters around the average values K_{sol} = 0.35 and B_{sol} = 50 Å^2 (Fokine & Urzhumtsev, 2002).

Subsequently to the solvent modelling, a structural refinement using all reflections is carried out. Then, for the deconvolution of the atomic thermal displacement and of the deformation of the electron density, a high-order (HO) refinement is performed, using, or not, stereochemical and dynamical restraints. The resolution range used for this high-order refinement cannot be restricted too quickly to the very high resolution shells due to refinement instability. We have shown that a gradual decrease of the resolution range allows this problem to be overcome. For instance, in the aldose reductase case, three resolution ranges were successively applied ($\sin\theta/\lambda = 1/2d = 0.5$–0.78 Å^{-1}, 0.6–0.78 Å^{-1}, 0.65–0.78 Å^{-1}), refining non-hydrogen structural and thermal motion parameters until convergence at each step. A minimal number of necessary structural or dynamical restraints were applied, mostly depending on the atoms thermal motion. This procedure improves the signal of the residual electron-density maps: the bonding-density peaks and the electron lone pair are generally stronger after the HO refinement (Fig. 12), which correlatively is a check of the quality of such a refinement (Fig. 12).

The solvent modelling and the high-order refinement lead to a drastic sharpening of bonding electron density. The deformation-density features are present in almost all regions of the enzyme where thermal motion parameters are sufficiently low. An upper limit can be defined around $B_{eq} = 6$–7 Å^2, above which thermal motion smearing seems to be too high. This value confirms those recently obtained by Afonine *et al.* (2004).

9.2. Multipolar-atom refinement

After convergence of structural parameters in the spherical-atom approximation, and correct treatment of hydrogen-atom positions (see above), the next step of the procedure is the multipolar refinement. One can either directly refine the non-spherical part of the electron density, or use a transfer of starting multipolar parameters from the multipoles library (Pichon-Pesme *et al.*,

Figure 12
Residual electron-density map ($F_{obs} - F_{cal}$) of aldose reductase in the aromatic plane of residue tyrosine 48. Left: after classical refinement. Right: after high-order refinement.

1995, 2004). The database transfer has been shown to be almost necessary to obtain a more meaningful description of the charge density, especially in terms of valence populations. This use of starting multipolar parameters can be compared with the application of stereochemical restraints in protein structure refinement at usual resolutions (∼2 Å). *A priori* experimental knowledge of the model parameters helps to overcome refinement instability or to avoid false-minimum trapping. Furthermore, the multipolar-parameter transfer leads to an improvement of crystallographic agreement factors, *e.g.* the crystallographic residual $R(F) = \sum ||F_{\mathrm{obs}}| - |F_{\mathrm{cal}}||/ \sum |F_{\mathrm{obs}}|$ decreases from 9.3 to 8.8% in the case of aldose reductase.

Multipolar refinement itself is also far from being straightforward. The charge-density database takes advantage of the chemical moiety repetition along the protein polypeptidic chain with a minimal number of multipolar parameters being transferred. Hence, numerous symmetry and chemical-equivalence constraints can be applied in the first part of the refinement. As described above, these constraints can be slightly and progressively relaxed to converge to a charge-density description, taking into account specific chemical environment of some sites of interest, *i.e.* the enzyme active site.

10. Perspectives

In the human aldose reductase complex, an inhibitor IDD-594 is bound in the active site. All of its atoms are not described in the multipole library. As the active site region, where the inhibitor is fixed, shows very low atomic thermal motion, typically less than 4 Å^2, it is possible to analyse the protein–ligand interactions from an experimental charge-density point of view, traditionally limited to small-molecule crystallography. This will be described in a forthcoming paper.

The multipolar database will be generalized to nucleic acids (RNA, DNA) and to most common biological cofactors, the NAD$^+$ molecule having already been analysed (Guillot *et al.*, 2003), and to common chemical groups. The use of transferred multipolar scattering factors on ultra-high-resolution protein structures will enable the precise analysis of their stereochemistry.

The electrostatic potential is an essential tool for understanding the molecular reactivity and the modelling of intermolecular interactions. This fundamental property is obtained either from quantum chemical calculations or from high-resolution X-ray diffraction experiments. Both the multipolar- and virtual-atom modelling of the total electron density provide far more precision than placing a single point charge at the atomic nuclei for the computation of the electrostatic potential. The computation of electrostatic interaction energies in host–guest systems (zeolites, protein–ligand) will be implemented in *MoPro*, using either a numerical integration of the total electrostatic energy (Aubert *et al.*, 2004) or a distributed multipole interaction (Stone, 1996; Volkov & Coppens, 2004).

Applying the multipolar-library transfer for electrostatic potential computation is meaningful for enzyme structures at atomic resolution, as soon as all the active-site atoms, including hydrogen, have well defined coordinates. It should be remembered here that most hydrogen atoms, with the exception of those in hydroxyl, ammonium and methyl groups, can be modelled in proteins without ambiguity, even if they are not visible in the electron-density maps.

The program and a detailed description are freely available from the authors upon request.

References

Afonine, P. V., Lunin, V. Y., Muzet, N. & Urzhumtsev, A. (2004). *Acta Cryst.* D**60**, 260–274.

Afonine, P. V., Pichon-Pesme, V., Muzet, N., Lecomte, C. & Urzhumtsev, A. (2002). *CCP4 Newslett. Protein Crystallogr.* No. 41. http://www.ccp4.ac.uk/newsletter41/00_contents.html.

Allen, F. H. (1986). *Acta Cryst.* B**42**, 515–522.

Angyan, J. G., Chipot, C., Dehez, F., Hättig, C., Jansen, G. & Millot, C. (2003). *J. Comput. Chem.* **24.** 997–1008.

Aubert, E., Porcher, F., Souhassou, M. & Lecomte, C. (2004). *J. Phys. Chem. Solids.* In the press.

Bader, R. F. W. (1990) *Atoms in Molecules: a Quantum Theory. The International Series: Monographs in Chemistry*. Oxford: Clarendon Press.

Becker, P. (1977). *Acta Cryst.* A**33**, 667–671.

Benabicha, F., Pichon-Pesme, V., Jelsch, C., Lecomte, C. & Khmou, A. (2000). *Acta Cryst.* B**56**, 155–165.

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

Blessing, R. H., Guillot, B., Lecomte, C. & Smith, G. D. (2003). American Crystallographic Association Meeting, Cincinnati (Ohio, USA), 26–31 July 2003.

Bouhmaida, N., Ghermani, N. E., Lecomte, C. & Thalal, M. (1999). *Acta Cryst.* A**55**, 729–738.

Brown, P. J., Fox, A. G., Maslen, E. N., O'Keefe, M. A., Sabine, T. M. & Willis, B. T. M. (1995). *International Tables for X-ray Crystallography*, Vol. C, edited by A. J. C. Wilson, pp. 475–516. Dordrecht: Kluwer Academic Publishers.

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.

Bürgi, H. B. & Capelli, S. C. (2000). *Acta Cryst.* A**56**, 403–412.

Burnett, M. N. & Johnson, C. K. (1996). *ORTEPIII*, Report ORNL-6895, Oak Ridge National Laboratory, Tennessee, USA.

Cachau, R., Howard, E., Barth, P., Mitschler, A., Chevrier, B., Lamour, V., Joachimiak, A., Sanishvili, R., Van Zandt, M., Sibley, E., Moras, D. & Podjarny, A. (2000). *J. Physique*, **10**, 3–13.

Clementi, E. & Raimondi, D. L. (1963). *J. Chem. Phys.* **41**, 2686–2689.

Coppens, P. (1967). *Science*, **158**, 1577–1579.

Coppens, P. (1997). *X-ray Charge Densities and Chemical Bonding*. IUCr/Oxford University Press.

research papers

Coppens, P., Guru Row, T. N., Leung, P., Stevens, E. D., Becker, P. J. & Yang, Y. W. (1979). *Acta Cryst.* A**35**, 63–72.

Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1997). *Curr. Opin. Struct. Biol.* **7**, 681–688.

Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1995). *Curr. Opin. Struct. Biol.* **5**, 784–790.

Dietrich, H. & Scheringer, C. (1978). *Acta Cryst.* B**34**, 54–63.

El Haouzi, A., Hansen, N. K., Le Hénaff, C. & Protas, J. (1996). *Acta Cryst.* A**52**, 291–301.

Fernandez-Serra, M. V., Junquera, J., Jelsch, C., Lecomte, C. & Artacho, E. (2000). *Solid State Commun.* **116**, 395–400.

Fletcher, R. & Reeves, C. M. (1954). *Comput. J.* **7**, 149–154.

Fokine, A. & Urzhumtsev, A. G. (2002). *Acta Cryst.* D**58**, 1387–1392.

Guillot, B., Muzet, N., Artacho, E., Lecomte, C. & Jelsch, C. (2003). *J. Phys. Chem. B*, **107**, 9109–9121.

Guillot, R., Muzet, N., Dahaoui, S., Lecomte, C. & Jelsch, C. (2001). *Acta Cryst.* B**57**, 567–578.

Guillot, B., Viry, L., Guillot, R., Lecomte, C. & Jelsch, C. (2001). *J. Appl. Cryst.* **34**, 214–223.

Hamilton, W. C. (1964). *Statistics in Physical Science, Estimation Hypothesis Testing and Least Squares*. New York: The Ronald Press.

Haneef, I., Moss, D. S., Stanford, M. J. & Borkakoti, N. (1985). *Acta Cryst.* A**41**, 426–433.

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

Hestenes, M. R. & Stiefel, E. (1952). *J. Natl Bur. Stand.* **49**, 409–436.

Hirshfeld, F. L. (1976). *Acta Cryst.* A**32**, 239–244.

Honig, B. & Nicholls, A. (1995). *Science*, **268**, 1144–1149.

Housset, D., Pichon-Pesme, V., Jelsch, C., Benabicha, F., Maierhofer, A., David, S., Fontecilla-Camps, J. C. & Lecomte, C. (2000). *Acta Cryst.* D**56**, 151–160.

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 Struc. Funct. Genet.* **55**, 792–804.

Jelsch, C. (2001). *Acta Cryst.* A**57**, 558–570.

Jelsch, C., Pichon-Pesme, V. & Lecomte, C. & Aubry, A. (1998). *Acta Cryst.* D**54**, 1306–1318.

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

Kissel, L., Zhou, B., Roy, S. C., Sen Gupta, S. K. & Pratt, R. H. (1995). *Acta Cryst.* A**51**, 271–288.

Konnert, J. H. (1976). *Acta Cryst.* A**32**, 614–617.

Konnert, J. H. & Hendrickson, W. A. (1980). *Acta Cryst.* A**36**, 344–350.

Koritsanszky, T. & Coppens, P. (2001). *Chem. Rev.* **101**, 1583–1628.

Koritsanszky, T., Volkov, A. & Coppens, P. (2002). *Acta Cryst.* A**58**, 464–472.

Kurki-Suonio, K. (1977). *Isr. J. Chem.* **16**, 115–123.

Lamzin, V. S., Morris, R. J., Dauter, Z., Wilson, K. S. & Teeter, M. M. (1999). *J. Biol. Chem*, **274**, 20753–20755.

Lecomte, C. (1995). *Advances in Molecular Structure Research*, Vol. 1, pp. 261–302. Greenwich, CT: JAI Press.

Lecomte, C., Guillot, B., Jelsch, C., Pichon-Pesme, V. & Muzet, N. (2004). *Commun. Mol. Life Sci.* **61**, 1–9.

Liu, Q., Teng, Q. H. M., Weeks, C. M., Jelsch, C., Zhang, R. & Niu, L. (2003). *J. Biol. Chem.* **278**, 41400–41408.

Longhi, S., Czjzek, M. & Cambillau, C. (1998). *Curr. Opin. Struct. Biol.* **8**, 730–737.

Maslen, E. N., Fox, A. G. & O'Keefe, M. A. (1995). *International Tables for X-ray Crystallography*. Dordrecht: Kluwer Academic Publishers.

Moews, P. C. & Kretsinger, R. H. (1975). *J. Mol. Biol.* **91**, 201–228.

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

Nyffeler, T. L., Jelsch, C., Ogawa, K. & Bürgi, H.-B. (2002). *Acta Cryst.* A**58**(Supplement), C141.

Ogawa, K., Sano, T., Yoshimura, S., Takeuchi, Y. & Toriumi, K. (1992). *J. Am. Chem. Soc.* **114**, 1041–1051.

Pérès, N., Bouhkris, A., Souhassou, M., Gavoille, G. & Lecomte, C. (1999). *Acta Cryst.* A**55**, 1038–1048.

Pichon-Pesme, V., Jelsch, C., Guillot, B. & Lecomte, C. (2004). *Acta Cryst.* A**60**, 204–208.

Pichon-Pesme, V., Lecomte, C. & Lachekar, H. (1995). *J. Phys. Chem.* **99**, 6242–6250.

Pichon-Pesme, V., Lecomte, C., Wiest, R. & Bénard, M. (1992). *J. Am. Chem. Soc.* **114**, 2713–2715.

Pietsch, U. & Hansen, N. K. (1996). *Acta Cryst.* B**52**, 596–604.

Podjarny, A., Howard, E., Mitschler, A., Chevrier, B., Lecomte, C., Guillot, B. & Jelsch, C. (2001). *Bull. Soc. Fr. Phys.* **131**, 12–15.

Podjarny, A., Howard, E., Mitschler, A., Chevrier, B., Lecomte, C., Guillot, B., Pichon-Pesme,V. & Jelsch, C. (2003). *Europhys. News*, **33**, 113–117.

Raymond, K. N. (1972). *Acta Cryst.* A**28**, 163–166.

Scheringer, C., Kutoglu, A., Mullen, D. & Hellner, E. (1978). *Acta Cryst.* A**34**, 475–476.

Schmidt A., Jelsch, C., Rypniewski, W. & Lamzin, V. S. (2003). *J. Biol. Chem.* **278**, 43357–43362.

Schmidt, A. & Lamzin, V. S. (2002). *Curr. Op. Struct. Biol.* **12**, 698–703.

Sheldrick, G. M. & Schneider, T. (1997). *SHELXL: High-Resolution Refinement*, in *Methods in Enzymology*, Vol. 276, *Macromolecular Crystallography*, Part B, edited by C. W. Carter Jr & R. M. Sweet, pp. 319–343. New York: Academic Press.

Smith, G. D., Pangborn, W. A. & Blessing, R. H. (2003). *Acta Cryst.* D**59**, 474–482.

Souhassou, M. & Blessing, R. H. (1999). *J. Appl. Cryst.* **32**, 210–217.

Souhassou, M., Lecomte, C., Blessing, R. H, Aubry, A., Rohmer, M. M., Wiest, R., Bénard, M. & Marraud, M. (1991). *Acta Cryst.* B**47**, 253–266.

Spackman M. A. (1992). *Chem. Rev.* **92**, 1769–1797.

Stewart, R. F., Davidson, E. R. & Simpson, W. T. J. (1965). *Chem. Phys.* **42**, 3175–3187.

Stone, A. J. (1996). *The Theory of Intermolecular Forces*, edited by J. S. Rowlinson, *International Series of Monographs on Chemistry*. Oxford: Clarendon Press.

Templeton, D. H. (1999). *Acta Cryst.* A**55**, 695–699.

Tronrud, D. E. (1992). *Acta Cryst.* A**48**, 912–916.

Urzhumtsev, A. G. (1991). *Acta Cryst.* A**47**, 723–727.

Volkov, A., Abramov, Y. A. & Coppens, P. (2001). *Acta Cryst.* A**57**, 272–282.

Volkov, A., Abramov, Y., Coppens, P. & Gatti, C. (2000). *Acta Cryst.* A**56**, 332–339.

Volkov, A. & Coppens, P. (2004). *J. Comput. Chem.* **25**, 921–934.

Volkov, A., Wu, G. & Coppens P. (1999). *J. Synchrotron Rad.* **6**, 1007–1015.

Zaleski, J., Wu, G. & Coppens, P. (1998). *J. Appl. Cryst.* **31**, 302–304.