

Acta Crystallographica Section A

**Foundations of  
Crystallography**

ISSN 0108-7673

Editor: **D. Schwarzenbach**

**On the application of an experimental multipolar pseudo-atom library  
for accurate refinement of small-molecule and protein crystal structures**

**Bartosz Zarychta, Virginie Pichon-Pesme, Benoît Guillot, Claude Lecomte and  
Christian Jelsch**

Copyright © International Union of Crystallography

Author(s) of this paper may load this reprint on their own web site provided that this cover page is retained. Republication of this article or its storage in electronic databases or the like is not permitted without prior permission in writing from the IUCr.

# On the application of an experimental multipolar pseudo-atom library for accurate refinement of small-molecule and protein crystal structures

Bartosz Zarychta,<sup>a,b</sup> Virginie Pichon-Pesme,<sup>a</sup> Benoît Guillot,<sup>a</sup> Claude Lecomte<sup>a</sup> and Christian Jelsch<sup>a\*</sup>

<sup>a</sup>Laboratoire de Cristallographie et Modélisation des Matériaux Minéraux et Biologiques, LCM3B, CNRS, UMR 7036, UHP, Faculté des Sciences et Techniques, BP 239, 54506 Vandoeuvre-lès-Nancy CEDEX, France, and <sup>b</sup>Institute of Chemistry, University of Opole, Oleska 48, 45-052 Opole, Poland. Correspondence e-mail: christian.jelsch@lcm3b.uhp-nancy.fr

With an increasing number of biomacromolecular crystal structures being measured to ultra-high resolution, it has become possible to extend to large systems experimental charge-density methods that are usually applied to small molecules. A library has been built of average multipole populations describing the electron density of chemical groups in all 20 amino acids found in proteins. The library uses the Hansen & Coppens multipolar pseudo-atom model to derive molecular electron density and electrostatic potential distributions. The library values are obtained from several small peptide or amino acid crystal structures refined against ultra-high-resolution X-ray diffraction data. The library transfer is applied automatically in the *MoPro* software suite to peptide and protein structures measured at atomic resolution. The transferred multipolar parameters are kept fixed while the positional and thermal parameters are refined. This enables a proper deconvolution of thermal motion and valence-electron-density redistributions, even when the diffraction data do not extend to subatomic resolution. The use of the experimental library multipolar atom model (ELMAM) also has a major impact on crystallographic structure modelling in the case of small-molecule crystals at atomic resolution. Compared to a spherical-atom model, the library transfer results in a more accurate crystal structure, notably in terms of thermal displacement parameters and bond distances involving H atoms. Upon transfer, crystallographic statistics of fit are improved, particularly free *R* factors, and residual electron-density maps are cleaner.

## 1. Introduction

The electron clouds around atoms in molecules are deformed primarily due to chemical bonding and secondarily by non-bonded interactions such as hydrogen bonds. Accurate electron-density distributions in the crystalline state can be derived from ultra-high-resolution X-ray diffraction experiments (Coppens, 1967, 1997, 1998; Lecomte, 1995; Koritsanszky & Coppens, 2001).

Charge-density analyses of peptide and amino acid crystals have been performed in our laboratory for some years, and compilation of all the charge-density parameters for the different chemical atom types resulted in the construction of our electron-density library (Pichon-Pesme *et al.*, 1995, 2004). The library of multipolar pseudo-atom parameters presently describes the electron density of all chemical groups found in peptides and proteins.

During the last decade, several methodological and technical improvements in macromolecular crystallography have led to the appearance of several protein structures at resolutions better than 1 Å (Schmidt *et al.*, 2003; Petrova & Podjarny, 2004). This leads to the possibility of analysing the structure at a very fine level of detail: one can refine multiple conformations, measure the deviations from the standard stereochemistry, visualize H atoms, and determine the protonation states of catalytic residues of enzymes (Minasov *et al.*, 2002). In many cases, this information is directly related to the function of proteins. The availability of crystallographic data at atomic resolution opens wide perspectives for theoretical modelling and experimental charge-density analyses.

Our laboratory has been involved in charge-density studies on molecules of increasing size: first enkephalin (Pichon-Pesme *et al.*, 1992), then the helical octapeptide LBZ (Jelsch *et al.*, 1998) and a scorpion toxin with diffraction data measured

at room temperature up to 0.96 Å (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 better than 1.0 Å is increasing continuously. The least-squares computer program *MoPro* linked to the current library was thus developed for the charge-density analysis of proteins (R. Guillot *et al.*, 2001; Jelsch *et al.*, 2005).

The latter development allowed the charge-density analysis of the 46 amino acid protein crambin (Jelsch *et al.*, 2000) using diffraction data measured at 100 K up to 0.54 Å resolution. The crystal structure was refined with a model for charged, non-spherical, multipolar atoms to describe accurately the molecular electron-density distribution. The initial multipoles and charges were transferred from our library to calculate non-spherical atomic form factors. The average electron-density parameters of the protein main chain were then refined against the crambin diffraction data and revealed significant electron density in the covalent bonds.

The transferability principle postulates that the deformation charge density due to chemical bonding and electron lone pairs in a given chemical group is similar in different molecules and crystal contexts. In this approximation, the polarization of atoms due to the chemical environment like in the hydrogen-bonding case is neglected. The very first application of the transferability of deformation densities among related molecules was published by Brock *et al.* (1991). The atomic multipole parameters refined from perylene diffraction data were used to improve the estimation of molecular vibrations in naphthalene and anthracene crystals.

A multipolar database has been proposed from theoretical computations (Volkov, Koritsanszky *et al.*, 2004; Volkov, Li *et al.*, 2004; Li *et al.*, 2006; Dominiak *et al.*, 2007), and the feasibility of building a database of theoretical atomic deformation density parameters applicable to the construction of the densities of biomacromolecules and to the interpretation of their X-ray diffraction data was demonstrated (Koritsanszky *et al.*, 2002). The pseudo-atoms are highly transferable and fairly invariant with respect to rotations around single bonds in the peptide framework.

Comparison between the two databases for the polypeptide backbone was discussed by Pichon-Pesme *et al.* (2004) and Volkov, Koritsanszky *et al.* (2004). More generally, the interplay between experiment and theory in the field of charge-density modelling is analysed by Coppens & Volkov (2004). The charge-density studies obtained from experimental diffraction data and quantum calculations generally agree well (Volkov *et al.*, 2000; Pichon-Pesme *et al.*, 2004), although some dissimilarities appear for the polar atoms and valence populations of some atoms and when electrostatic properties are calculated. The disagreement may partly be due to the fact that the experiment is performed on the crystalline state while the computations are performed *in vacuo*, leading to different results due to polarization effects (Spackman *et al.*, 1999; Fernandez-Serra *et al.*, 2000).

Recently, the name of invarioms (invariant atoms) was proposed by the group of P. Luger for transferable charge-density parameters derived from theoretical computations.

The invarioms were applied at several resolutions to the crystal structures of DL-serine at three temperatures, and the results (Dittrich *et al.*, 2005) confirm ours (Pichon-Pesme *et al.*, 1995; Jelsch *et al.*, 1998) and show a significant improvement of the crystallographic statistics results. In the studies of tri(L-alanine) (Rödel *et al.*, 2006), tri(L-valine) and terbogrel (Dittrich *et al.*, 2004), the transferability of invarioms was investigated. The results showed the interest of database approaches for modelling larger molecular systems or structures determined at lower resolution. Moreover, the examinations of tri(L-valine) and terbogrel demonstrated that the transfer of the invariom multipole parameters enables evaluation of detailed non-spherical electron density and electrostatic potential distributions simultaneously with improving the refinement of  $x$ ,  $y$ ,  $z$  and  $U_{ij}$  parameters even for room-temperature data sets. It was also demonstrated that invarioms are helpful in the absolute structure determination of light-atom crystal structures (Dittrich, Strumpel *et al.*, 2006). The theoretical database of invarioms for structural refinement of amino acid, oligopeptide and protein molecules was published by Dittrich, Hübschle *et al.* (2006). Such results encourage further development of the transferable charge-density-parameters database.

## 2. Materials and methods

### 2.1. Multipolar atom model

In the peptide refinements with program *MOLLY* (Hansen & Coppens, 1978) or *MoPro* (B. Guillot *et al.*, 2001), the atomic electron density is modelled as the sum of the core-plus-valence spherical electron density and the non-spherical valence density, which is described as a sum of multipoles (Hansen & Coppens, 1978):

$$\rho_{\text{atom}}(\mathbf{r}) = \rho_{\text{cor}}(r) + P_{\text{val}}\kappa^3\rho_{\text{val}}(\kappa r) + \sum_{l=0}^{l_{\text{max}}} \kappa'^3 R_{nl}(\kappa' r) \sum_{m=0}^l P_{lm\pm} y_{lm\pm}(\theta, \varphi). \quad (1)$$

$P_{\text{val}}$  is the valence population parameter and gives an estimation of the net atomic charge  $q = N_{\text{val}} - P_{\text{val}}$ , where  $N_{\text{val}}$  is the number of valence electrons for the neutral atom. The  $y_{lm}$  represent spherical harmonic functions of the order  $l$  in real form,<sup>1</sup> the  $R_{nl}$  are Slater-type radial functions and the  $P_{lm\pm}$  are the multipole populations. The normalization of the  $y_{lm\pm}$  functions implies that a  $P_{lm\pm}$  value of +1 transfers one electron from the negative lobe of the  $y_{lm\pm}$  function to its positive lobe. The coefficients  $\kappa$  and  $\kappa'$  describe the contraction–expansion for the spherical and multipolar valence densities, respectively.

For the structure-factor computations, the form factor for the H atoms was taken from Stewart *et al.* (1965); the form factors for non-H atoms were calculated from Clementi & Raimondi (1963) wavefunctions. The real and imaginary anomalous-dispersion corrections to the form factors were from Cromer (1974).

<sup>1</sup> Supplementary data for this paper are available from the IUCr electronic archives (Reference: SH5051). Details for accessing these data are given at the back of the journal.

**Table 1**

Charge-density analyses of peptides used for building the electron-density library.

Peptide	Reference
<i>N</i> -Ac-L-Trp-NHMe	Souhassou <i>et al.</i> (1991)
( <i>Z</i> )- <i>N</i> -acetyl- $\alpha,\beta$ -dehydro-Phe-NHMe	Souhassou <i>et al.</i> (1992)
Leu-enkephaline	Wiest <i>et al.</i> (1994); Pichon-Pesme <i>et al.</i> (1992)
L-Arginine phosphate	Espinosa <i>et al.</i> (1996)
Triglycine	Pichon-Pesme & Lecomte (1998)
<i>N</i> -Acetyl-L-tyrosine ethyl ester	Dahaoui, Jelsch <i>et al.</i> (1999)
L-Cystine	Dahaoui, Pichon-Pesme <i>et al.</i> (1999)
Gly-L-Thr	Benabicha <i>et al.</i> (2000)
Gly-L-Asp; L-Tyr-Gly-Gly	Pichon-Pesme <i>et al.</i> (2000)
DL-Alanylmethionine	R. Guillot <i>et al.</i> (2001)
<i>N</i> -Acetyl-L-Gln	
L-Cysteine	Dahaoui (in preparation)
L-Histidinephosphate	Mata <i>et al.</i> (2006)
<i>tert</i> -Bu-L-Pro-His-NHMe	

The multipole expansion is truncated at the hexadecapole level for S atoms, at the octapole level for O, N and C atoms, and one dipole along the H–*X* axis for H atoms. The coefficients of the Slater functions are described in the supplementary material.<sup>2</sup>

## 2.2. Library building

The experimental-electron-density library has been constructed from the compounds listed in Table 1. The experimental temperatures during data collection were cryogenic, generally 100 K, and the diffraction data were collected up to subatomic resolution  $d = 0.45 \text{ \AA}$  (or with  $\sin \theta_{\max}/\lambda$  reaching  $1.1 \text{ \AA}^{-1}$ ). All the crystallographic  $wR^2(F)$  residual factors were below 4% after multipolar refinement. In the initial peptide structures, all the H atoms were located in difference Fourier maps.

The peptide charge-density refinements were performed according to a common general strategy. At first, the positional and anisotropic thermal displacement parameters of the non-H atoms were refined against the high-order reflections with  $(\sin \theta)/\lambda > 0.7\text{--}0.8 \text{ \AA}^{-1}$ . The positional and isotropic thermal displacement parameters of the H atoms were then refined against low-order data with  $(\sin \theta)/\lambda < 0.7 \text{ \AA}^{-1}$ . The H-atom positions were then extended along the experimental *X*–H direction to the average bond distances obtained from neutron diffraction experiments (Allen, 1986). When neutron diffraction data were available for the compound, as for L-arginine phosphate (Espinosa *et al.*, 1996), the H-atom positions were taken from the neutron refinement. The different types of charge-density variables, *viz*  $P_{\text{val}}$ ,  $P_{lm\pm}$ ,  $\kappa$  and  $\kappa'$ , were refined separately and consecutively. These variables were introduced progressively in the refinements to ensure a good convergence to physically meaningful charge densities.  $\kappa'$  parameters of some H atoms were kept fixed to unity when their refinement was divergent or led to unrealistic values. All parameters were not refined simultaneously, even in the final step, as this would have led to divergence. At each stage of

<sup>2</sup> See deposition footnote.

refinement, deformation and residual density maps were checked to ensure a realistic deformation density.

To a first approximation, the charge density of a chemical group obeys local symmetries applying to the group. Chemically equivalent groups in the crystal asymmetric unit are also expected to show similar charge densities. In the initial steps of the multipolar refinements, symmetry and chemical equivalence constraints were applied to reduce the number of variables and improve the stability of the refinement, and were progressively released. In the last stage of the refinement, no multipolar parameter was fixed to zero. Some chemical equivalence constraints were maintained in the final refinement only in the case of the relatively large peptide Leu-enkephalin (Pichon-Pesme *et al.*, 1992).

Within atom groups of the same chemical nature and in the same geometrical environment, multipole parameters were observed to be similar (Pichon-Pesme *et al.*, 1995). For instance, in the case of the N-terminus N atom (Table 2), the r.m.s. uncertainty of a given multipole, within the sample of peptides, is on average 0.04. For the multipoles with largest magnitude, the uncertainty represents 22% in the case of  $P_{yz^2} = -0.201$  and 14% in the case of  $P_{x^3} = 0.197$ . The library parameters were obtained by averaging the values over the equivalent atoms.

The standard uncertainty  $\sigma_x$  of the parameters within the sample was obtained by the equation

$$\sigma_x = \left[ \sum_{i=1,N} (x_i - \langle x \rangle)^2 / (N - 1) \right]^{1/2}. \quad (2)$$

An estimation of the uncertainty of the average parameter may be obtained by the equation

$$\sigma_{\langle x \rangle} = \left\{ \sum_{i=1,N} (x_i - \langle x \rangle)^2 / [N(N - 1)] \right\}^{1/2}, \quad (3)$$

where  $x_i$  are the parameter values within the sample,  $\langle x \rangle$  is the mean value of parameter  $x$  and  $N$  is the sample size.

The different atom types are not represented by the same number  $N$  of experimental atoms in the library. The atoms of the polypeptide backbone and the ammonium and carboxylate termini are more numerous than the atoms of side chains (up to 23 times against once for some atoms).

The experimental standard uncertainty for refined multipole populations is generally about 0.02, as described for example in the ultra-high-resolution electron-density analysis of triglycine (Pichon-Pesme & Lecomte, 1998). When the standard uncertainty in the sample is of this order of magnitude and the sampling large enough ( $N > 3$ ), the mean value can be considered as determined with a good accuracy. For the largest multipolar parameters, the ratio of the average over the standard uncertainty of the mean is larger than 20.

The  $P_{lm\pm}$  multipole populations recorded in the experimental library were set to zero if the average value in the sample was smaller than three times its estimated standard uncertainty. In addition, the average multipole populations that turned out to be lower than 0.03 in absolute value were set equal to zero in the library, as they do not contribute

**Table 2**

Multipolar parameters of  $-\text{NH}_3^+$  N atoms in the compounds used for the library construction and their average values.

The local  $X$  axis is along the  $\text{NT}-\text{C}\alpha$  direction. The  $Y$  direction is along the  $\text{NT}-\text{H1}$  bond. Multipoles antisymmetric with respect to the  $Z$  axis are expected to be small as the  $-\text{NH}_3^+$  group obeys a mirror orthogonal to  $Z$  ( $m_{\perp z}$ ). The multipoles marked in the last line are expected to be *a priori* small as they are antisymmetric with respect to the  $m_{\perp z}$  mirror (atoms following the threefold symmetry around the  $X$  axis are also involved). Only the values underlined in the line 'mean' are recorded in the library, others are set to zero.

Parameter	Dipoles			Quadrupoles					Octapoles					$\kappa$	$\kappa'$			
	$P_{\text{val}}$	11+	11-	10	20	21+	21-	22+	22-	30	31+	31-	32+			32-	33+	33-
	$x$	$y$	$z$	$z^2$	$xz$	$yz$	$x^2-y^2$	$xy$	$z^3$	$xz^2$	$yz^2$	$z(x^2-y^2)$	$xyz$	$x^3$	$y^3$			
Leu-enkephaline	5.516	-0.092	-0.005	-0.012	0.020	-0.004	0.083	0.076	0.083	-0.016	-0.170	-0.151	0.107	0.049	0.251	0.029	0.973	0.812
Triglycine	5.750	-0.116	0.001	-0.002	0.058	0.032	-0.006	-0.035	0.022	0.001	-0.214	-0.312	-0.011	0.059	0.232	0.084	0.968	0.814
Triglycine	5.668	-0.075	-0.010	-0.038	0.039	0.053	0.000	-0.018	-0.024	-0.044	-0.170	-0.213	-0.058	-0.016	0.213	0.067	0.969	0.814
L-Tyr-Gly-Gly	5.504	-0.061	-0.012	0.028	0.073	-0.001	0.017	-0.075	0.016	0.030	-0.171	-0.242	0.018	-0.025	0.151	0.137	0.986	0.718
Gly-L-Asp	5.677	-0.123	-0.027	0.046	-0.114	0.109	0.055	-0.091	-0.052	-0.070	-0.214	-0.204	0.007	-0.008	0.191	0.022	0.967	0.763
L-Cysteine	5.397	-0.092	0.020	0.020	0.052	-0.036	0.010	-0.023	-0.013	0.021	-0.108	-0.165	-0.039	0.038	0.174	0.035	0.979	0.929
Gly-L-Thr	5.649	-0.058	-0.011	-0.053	0.074	-0.023	-0.016	-0.024	-0.060	0.107	-0.145	-0.180	0.118	-0.012	0.199	0.000	0.978	0.825
Dl-Ala-Met	5.391	-0.057	0.035	0.018	-0.016	0.007	-0.030	-0.022	0.000	0.019	-0.125	-0.156	0.012	-0.013	0.171	0.038	1.012	0.901
L-Cysteine	5.487	-0.036	0.076	0.005	-0.101	0.105	0.064	0.024	-0.042	-0.059	-0.137	-0.190	-0.100	0.035	0.189	0.046	0.989	0.954
Mean	<u>5.560</u>	<u>-0.079</u>	<u>0.007</u>	<u>0.001</u>	<u>0.009</u>	<u>0.027</u>	<u>0.020</u>	<u>-0.021</u>	<u>-0.008</u>	<u>-0.001</u>	<u>-0.162</u>	<u>-0.201</u>	<u>0.006</u>	<u>0.012</u>	<u>0.197</u>	<u>0.051</u>	<u>0.980</u>	<u>0.837</u>
$\sigma_{\text{mean}}$	<u>0.043</u>	<u>0.010</u>	<u>0.011</u>	<u>0.011</u>	<u>0.024</u>	<u>0.018</u>	<u>0.013</u>	<u>0.016</u>	<u>0.015</u>	<u>0.018</u>	<u>0.012</u>	<u>0.017</u>	<u>0.024</u>	<u>0.011</u>	<u>0.010</u>	<u>0.013</u>	<u>0.005</u>	<u>0.026</u>
Expected to be small			$3x$	$m_{\perp z}$		$m_{\perp z}$	$m_{\perp z}$		$3x$	$m_{\perp z}$		$m_{\perp z}$	$m_{\perp z}$					

significantly to the electron density. Thus, as the number of parameters for each atom type is reduced, fewer parameters need be transferred to large molecules, which is very convenient for least-squares refinement.

### 2.3. Local axes definition

As described by Hansen & Coppens (1978), a local axial system applying to the multipolar harmonic functions is defined for each atom. The axes, namely  $X$ ,  $Y$ ,  $Z$  orthogonal directions, are chosen in order to be able to apply geometrical symmetry constraints on the multipoles. The *MoPro* local axial system of an atom is derived from two directions  $X$  and  $Y'$ , which are defined by the original origin atom and two neighbouring atoms.

For instance, the definition of the axes for a water molecule is the following:  $\mathbf{X}$  is defined along  $\text{OW}-\text{H1}$ ;  $\mathbf{Y}'$  is defined along  $\text{OW}-\text{H2}$ ; and  $\mathbf{Z}$  and  $\mathbf{Y}$  are consequently derived as vectorial products:  $\mathbf{Z} = \mathbf{X} \wedge \mathbf{Y}'$  and  $\mathbf{Y} = \mathbf{Z} \wedge \mathbf{X}$ .<sup>3</sup> In the current library, all local axes are of  $\mathbf{XY}$  type, although the *MoPro* software allows the use of dummy atoms or of bisecting directions (Jelsch *et al.*, 2005).

Local symmetry mirror planes may apply to an atom. In the current paper, the notations  $m_{\perp z}$  or  $mz$  refer to a mirror orthogonal to the local  $z$  axis. For instance, a  $C_{2v}$  symmetry may be imposed on phenyl ring atoms (two mirrors  $m_{\perp z}$  and  $m_{\perp y}$ ). For  $sp^2$  N and C atoms, the  $z$  direction is by convention taken orthogonal to the hybridization plane. At the beginning of the charge-density refinement of the library peptides, the multipoles were constrained to obey local atom symmetries in order to get a good start. To average multipole populations values for the library, the local axes for a given chemical atom

<sup>3</sup> A figure showing the definition of the atomic local orthogonal system is included in the supplementary material. See deposition footnote.

type were always defined in the same way in the different peptides analysed.

### 2.4. Electron-density maps

Static maps correspond to the electron density that would be observed for a molecule with no thermal motion ( $B = 0 \text{ \AA}^2$ ) and are computed directly using the refined values of the  $P_{lm\pm}$ ,  $P_{\text{val}}$ ,  $\kappa$  and  $\kappa'$  pseudo-atom parameters describing the molecular charge density in equation (1).

The charge distribution of a molecule is efficiently displayed by the deformation of the electron density  $\Delta\rho$ . Deformation maps are computed from the crystallographic modelling as the atomic superposition-sum over the molecule:

$$\Delta\rho = \sum_{\text{atoms}} (\rho_{\text{mul}} - \rho_{\text{IAM}}), \quad (4)$$

where  $\rho_{\text{mul}}$  is the total electron density as in equation (1) and  $\rho_{\text{IAM}}$  is the density of spherical neutral atoms (independent-atom model, IAM).

The contour figures in two dimensions are directly generated with the *VMoPro* part of the *MoPro* package. To draw iso-contour surfaces, three-dimensional grid maps in *X-PLOR/CNS* (Brünger *et al.*, 1998) or cube formats are calculated with *VMoPro* and used as input in other programs like *Pymol* (DeLano, 2002).

## 3. Results and discussion

### 3.1. Construction and analysis of the library

The compounds included in the library are representative amino acids or small peptides with up to five residues. Each compound was studied by X-ray diffraction at a high level of resolution and precision using a nitrogen cryostream. For details of the experimental electron-density determinations, see the references given in Table 1. In the orthorhombic form

**Table 3**

Experimental library parameters of some  $sp^2$  atoms: C, O, N atoms in the peptide backbone, aromatic C atom ( $C_\gamma$  of residue Phe), CH aromatic atoms ( $C\delta 1$ ,  $C\delta 2$ ,  $C\epsilon 1$ ,  $C\epsilon 2$ ,  $C\zeta$  of residue Phe) and C atoms of flat five-membered ring ( $C_\gamma$  of Hip and His).

The local symmetries of atoms are indicated in the last column. The multipoles that are antisymmetric with respect to mirrors  $m_{\perp y}$  and  $m_{\perp z}$  are indicated as such. The multipolar populations in italics are expected to be *a priori* small as they are antisymmetric with respect to a mirror applying to the atom.

Atom type	Axis X	Axis Y'	P <sub>val</sub>	Dipoles			Quadrupoles					Octapoles					κ	κ'	Atom symmetry		
				x	y	z	z <sup>2</sup>	xz	yz	x <sup>2</sup> -y <sup>2</sup>	xy	z <sup>3</sup>	xz <sup>2</sup>	yz <sup>2</sup>	z(x <sup>2</sup> -y <sup>2</sup> )	xyz				x <sup>3</sup>	y <sup>3</sup>
C	O	N	3.976	0.102	0	0	-0.304	0	0	0.102	0	0	0	0	0	0.411	0	0.998	0.941	<i>m</i> <sub>⊥z</sub>	
O	C	N	6.307	-0.075	0	0	-0.078	0	0	-0.070	0	0	0	0	0	0.039	0	0.977	0.960	<i>m</i> <sub>⊥z</sub> <i>m</i> <sub>⊥y</sub>	
N	C	Cα	5.312	0	-0.072	0	-0.050	0	0	0	0	0	0	0	0	0.260	0	0.987	0.867	<i>m</i> <sub>⊥z</sub>	
<i>C</i> <sub>γ</sub> _Phe	<i>C</i> <sub>ζ</sub>	<i>C</i> <sub>δ2</sub>	4.040	0	0	0	-0.250	0	0	0	0	0	0	0	0	-0.330	0	0.994	0.926	<i>m</i> <sub>⊥z</sub> <i>m</i> <sub>⊥y</sub>	
<i>C</i> <sub>δ1</sub> _Phe	<i>C</i> <sub>ε2</sub>	<i>C</i> <sub>γ</sub>	4.155	0	0	0	-0.247	0	0	0	0	0	0	0	0	-0.334	0	0.989	0.929	<i>m</i> <sub>⊥z</sub> <i>m</i> <sub>⊥y</sub>	
<i>C</i> <sub>γ</sub> _His	<i>C</i> <sub>δ2</sub>	<i>N</i> <sub>δ1</sub>	4.028	0.070	-0.030	0	-0.070	0	0	0	0.060	0	0	0	0	0.240	0.100	1.021	0.968	<i>m</i> <sub>⊥z</sub>	
<i>C</i> <sub>γ</sub> _Hip	<i>C</i> <sub>δ2</sub>	<i>N</i> <sub>δ1</sub>	3.780	0.058	0	0	-0.126	0	0	0.091	0.063	0	0.033	0.059	0	0	0.219	0.113	1.045	0.994	<i>m</i> <sub>⊥z</sub>
Antisymmetries							<i>m</i> <sub>⊥z</sub>					<i>m</i> <sub>⊥z</sub>							3z		
				<i>m</i> <sub>⊥y</sub>			<i>m</i> <sub>⊥z</sub>			<i>m</i> <sub>⊥y</sub>		<i>m</i> <sub>⊥z</sub>			<i>m</i> <sub>⊥y</sub>		<i>m</i> <sub>⊥z</sub>				

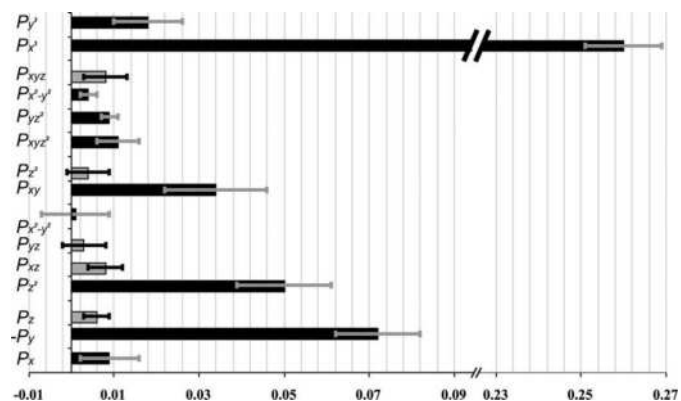
of L-cysteine (Kerr & Ashmore, 1973), as the H atom on the thiol group is disordered, an additional neutron diffraction analysis was carried out.

For each defined atom type, the experimental multipolar parameters obtained from the amino acid and peptide crystals were averaged to build the library. The averaged library parameters can be applied to large peptide and protein molecules built of the chemical groups found in the 20 amino acids.

The library contains different atom types defined by several characteristics:

- (i) the chemical element;
- (ii) the hybridation state ( $sp^2$ ,  $sp^3$ );
- (iii) the number and nature of the first atom neighbours (covalent bond connectivity);
- (iv) the nature of the covalent bonds with the neighbours (single, double, aromatic);

The charge density of atoms belonging to the same type is considered as transferable, *i.e.* similar if environment effects such as hydrogen bonding and local crystal electric field are



**Figure 1** Multipole populations  $P_{lm\pm}$  in absolute value for the peptide N atom stored in the library. The grey bars relate to the spherical harmonic functions which are antisymmetric with respect to  $z$  and are therefore expected to be small. The standard uncertainties in the sample of size 21 are shown as error bars.

neglected. The protein atoms that are considered as chemically equivalent in the library are as follows.

(i) The main-chain polypeptide  $-(N-HO)-(C\alpha H\alpha)-(C=O)-$  atoms are considered equivalent for all residues, except for the  $H\alpha_1-C\alpha-H\alpha_2$  of glycine and N atom of proline which are specific.

(ii) The  $-NH_3^+$  and  $-COO^-$  termini were averaged for all the peptides studied. The ammonium group of Lys and the carboxylate groups of Asp/Glu are defined to be identical to the  $-NH_3^+$  and  $-COO^-$  termini, respectively.

Several atoms or groups of atoms in side chains are considered equivalent:

- (i) all aromatic C-H atoms in residues Phe, Tyr and the six-membered ring of Trp;
- (ii) the aromatic  $C_\gamma$  atoms of Phe and Tyr;
- (iii) the  $>CH-$  groups in Ile, Leu and Val side chains;
- (iv) the  $>CH_2$  groups bound to two C atoms, as in Asp, Asn, Gln, Lys, Arg, Phe, Tyr, Trp and Met;
- (v) the methyl  $-CH_3$  groups, bound to a C atom in side chains of residues Ile, Leu, Val, Ala and Thr;
- (vi) the two  $NH_2$  groups of Arg;
- (vii) the two amide groups of Asn and Gln;
- (viii) the hydroxyl groups C-OH of Ser and Thr.

Some atom types remain specific to a single atom in one amino acid: the C-OH hydroxyl group of Tyr, some Arg side-chain atoms, some Trp indole atoms, and the S atoms in Cys (C-S-H, C-S-S) and Met (C-S-C).

Some amino acids are present in two states. The oxidized cysteine (Cys) forming disulfide bridges and the reduced cysteine (Cyh) are different. Some acidic residues (Glu and Asp) are sometimes in their uncharged form (COOH) when they are buried beneath the protein surface. The protonated aspartic and glutamic acids (residues called Ash and Glh) are also present in the library, as several peptide crystals had a COOH group either on a side chain (GlyAsp) or as a C-terminus (N-acetyl-L-Gln).

The neutral imidazole groups can be singly protonated on either the  $N\epsilon 2$  atom (residue His) or the  $N\delta 1$  atom (residue Hid). The atoms of the two residues are considered equivalent after taking into account the interchange of the atoms. The

positively charged histidine Hip, doubly protonated on the two N atoms, is also modelled in the library (Espinosa *et al.*, 1996).

### 3.2. Comparison of multipole populations for some $sp^2$ atoms

At the level of effective transferability, the charge density of the peptide C, O and N atoms obeys mirror symmetry about the plane of the  $sp^2$  hybridization. The multipoles which are antisymmetric with respect to the  $mz$  mirror have small values of both signs in different molecules, and their average values were all below the  $3\sigma$  threshold, as shown for the peptide N atom in Fig. 1, and were set to zero in the library (Table 2).

The number of significant  $P_{lm\pm}$  parameters for an atom depends particularly on the geometrical arrangement of the atom's first neighbours and the local symmetries.

The peptide C atom and several  $sp^2$  C atoms form covalent bonds with three other atoms. This is also the case for the carbonyl C atom of the polypeptide main chain and for aromatic C atoms. The local atomic geometry of these atoms

follows an approximate local threefold axis, as the three bond angles are about  $120^\circ$ . The octapole  $P_{x^3}$  (defined as  $P_{33+}$  in Table 2) for which the harmonic function  $x(x^2 - 3y^2)$  obeys a threefold symmetry axis and models the three bonds is, as expected, generally large for these  $sp^2$  atoms (Table 3). The octapole  $P_{y^3}$  also has triad symmetry but it is small as it is rotated  $30^\circ$  with respect to the  $P_{x^3}$  octapole.

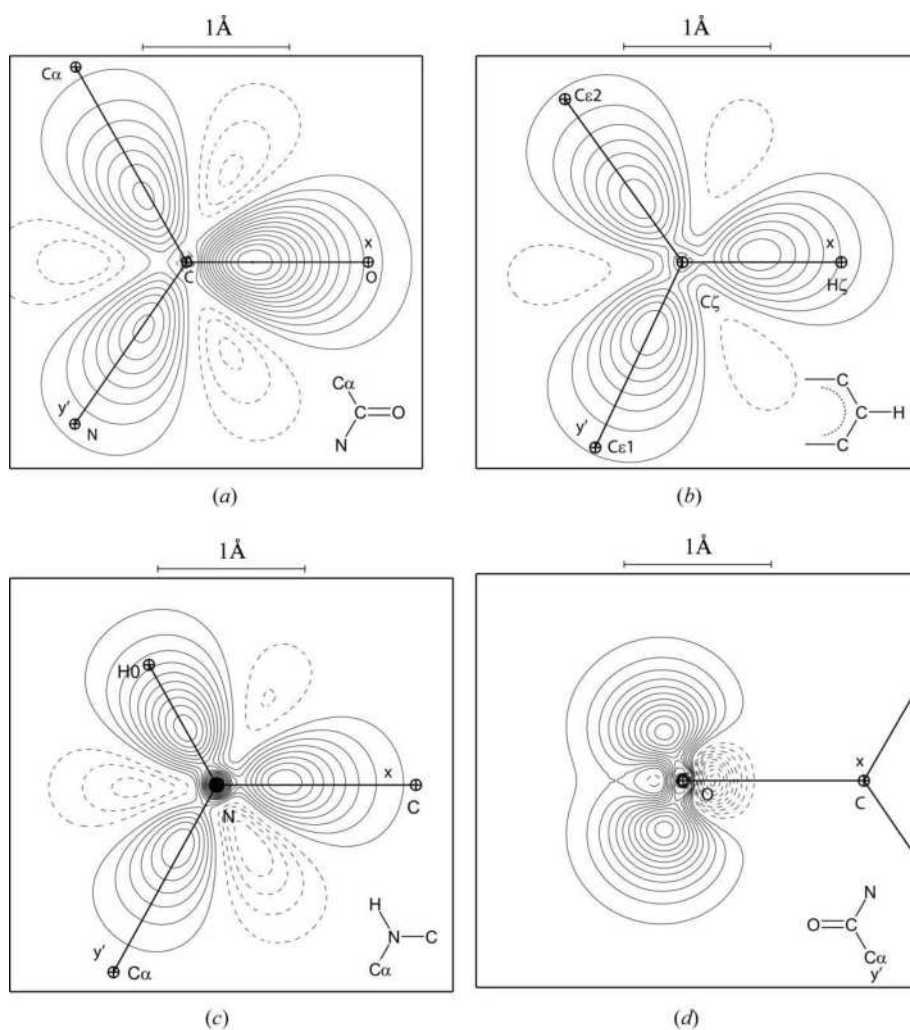
For the peptide N atom, the largest multipolar parameter is the threefold symmetric octapole  $P_{x^3}$  which models the main characteristic of the deformation density, *viz* positive deformation density peaks in the three bond directions (Fig. 2). The  $sp^2$  character of the N atom is due to the resonance within the H–N–C=O peptide group. The quadrupole  $P_{z^2}$  has a smaller value, as the peptide N atom displays less  $\pi$ -electron delocalization than the C atom. The N-atom dipole  $P_y$  parameter takes a significantly negative value ( $-0.072$ ), the  $X$  direction being directed towards the carbonyl C atom. As a result, the pseudo-threefold symmetry of the N atom is broken by the  $P_y$  dipole. In the library, the N atom has only three significant multipoles, which

obey mirror symmetry about the hybridization plane.

As shown in Table 3, for the polypeptide backbone carbonyl  $sp^2$  C atoms, the  $P_{z^2}$  quadrupole ( $-0.304$ ) and the  $P_{x^3}$  octapole ( $0.411$ ) are the only parameters that take large values; they model respectively the  $\pi$ -electron delocalization and the electron accumulation in the three covalent bonds. The  $\pi$ -electron delocalization is largest for the peptide C atom as can be seen in Fig. 3, which shows a strong electron depletion above and below the  $sp^2$  plane, and is also visible for the O atom ( $P_{z^2} = -0.078$ ).

Depending on the nature of the bonds and of the atomic neighbours, mirror planes other than  $mz$  may apply to the local electron density. In the case of aromatic C(H) atoms with at least two C atoms as neighbours, a second mirror, denoted  $my$  in *MoPro*, perpendicular to the  $sp^2$  plane and parallel to the C–H bond, applies.

The C $\gamma$  atom of Phe approximately obeys a threefold axis along  $z$ , therefore only the two multipoles  $P_{z^2}$  and  $P_{x^3}$  turn out to be different from zero. This is also true for the C(H) aromatic atoms of the phenyl group for

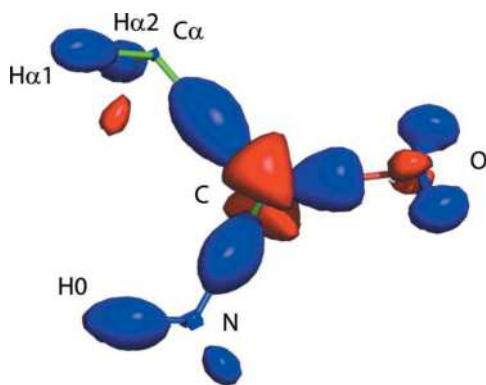


**Figure 2**

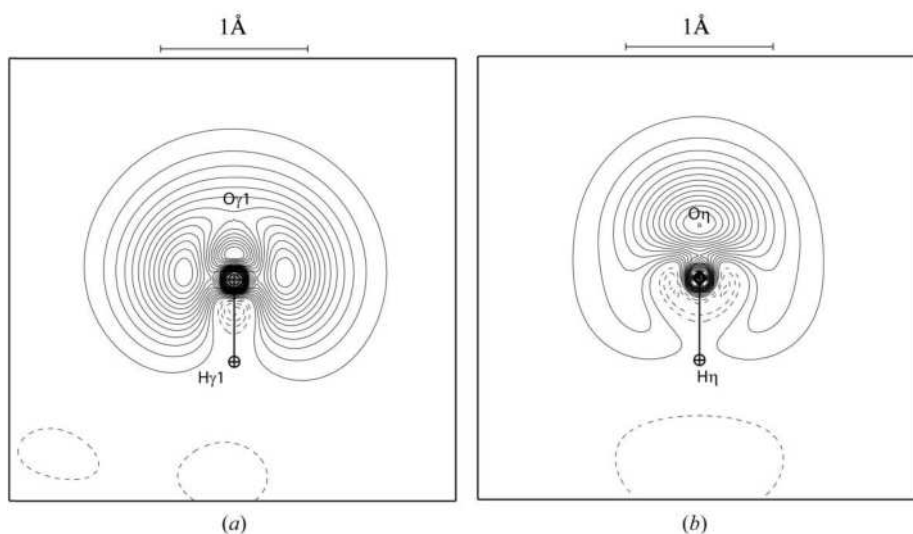
Static multipole electron density generated by some  $sp^2$  atoms: (a) the C atom of the peptide group; (b) the aromatic phenyl C atom; (c) the N atom of the peptide group; (d) the O atom of the peptide group. The  $x$  and  $y$  symbols are related to the local axis system. Contours are at  $\pm 0.05 \text{ e } \text{\AA}^{-3}$  with full lines denoting positive density and dashed lines denoting negative density.

which the  $P_{x^3}$  octapole value also has a large magnitude of  $-0.334$ . The opposite sign of the  $P_{x^3}$  parameters is due to a different definition of the local axes: for the aromatic C atoms, the  $X$  direction is directed towards the opposite atom on the phenyl ring and not towards a neighbour atom.

The number of parameters necessary to describe the C-atom electron density is larger for a planar five-membered ring (histidine, tryptophan) than for a six-membered aromatic ring, as some multipoles follow threefold symmetry. In the five-membered rings, the strongest contributors to the deformation density are still the same  $P_{x^3}$  and  $P_{z^2}$  multipoles. The octapole  $P_{x^3}$  models three peaks at  $120^\circ$  while the bond angles in a five-membered ring are  $108^\circ$  instead.



**Figure 3**  
Three-dimensional static deformation density map of the (H $\alpha$ 1–C $\alpha$ –H $\alpha$ 2)–(C=O)–N–H0 glycine moiety in triglycine after library transfer. Blue surfaces denote positive density and red denote negative density in this *Pymol* view (DeLano, 2002). The cut-offs for positive and negative surfaces are  $+0.4$  and  $-0.1 \text{ e } \text{\AA}^{-3}$ , respectively.



**Figure 4**  
Deformation electron density of the hydroxyl O atom: (a) O $\gamma$  in serine/threonine and (b) O $\eta$  in tyrosine. Contours are at  $\pm 0.05 \text{ e } \text{\AA}^{-3}$  with full lines denoting positive density and dashed lines denoting negative density. The figure is shown in the plane perpendicular to the C–O–H group and bisecting the (O–C, O–H) directions.

The peptide carbonyl O atom displays also three positive deformation density lobes located on the C=O double bond and on the two non-bonded lone-pair electron sites. The O atom has only three significant multipoles, all with relatively small  $P_{lm\pm}$  values, including a particularly small value for the  $P_{x^3}$  parameter, which does not have  $3m$  symmetry. The O-atom multipoles describe essentially the two electron lone pairs (Fig. 2*d*), while the C=O bonding density is essentially described by the C-atom multipoles.

### 3.3. Comparison of multipole populations of $sp^3$ atoms

A C or N atom with four neighbours in a tetragonal geometry follows symmetries if some of the neighbour atoms are identical. For instance,  $-\text{CH}_3$  and  $-\text{NH}_3^+$  groups display typically a threefold axis and three mirror planes. For instance, in the C– $\text{NH}_3^+$  moiety, the threefold axis is along the C–N bond and the mirrors are in the plane of C, N, H atoms. The multipole populations of the different  $-\text{NH}_3^+$  terminal N atoms are given in Table 2.

According to the experimental charge-density determination in nine peptides, the number of significant  $P_{lm\pm}$  parameters for the ammonium N atom turns out to be small (Tables 2 and 4). This is in accordance with the large number of local symmetries ( $3m$ ) applying to the ammonium N atom. With the definition of local axes used for the ammonium group, a mirror plane formed by H1–NT–C $\alpha$  atoms perpendicular to  $z$  and parallel to  $x$  applies. Owing to the threefold symmetry along the  $x$  axis, the dipole along any direction orthogonal to the  $z$  axis should be zero. Therefore, the dipole along the  $y$  axis is *a priori* null and so are more generally the multipoles of the form  $x^n y$  ( $n \geq 0$ ), e.g. the quadrupole  $P_{xy}$ . The number of significant multipoles is then limited for the ammonium N atom:  $P_x$ ,  $P_{xz^2}$  and  $P_{yz^2}$  display large values. The combination of these multipole functions describes most of the electron density in the four bonds formed by the N atom.

The multipole populations of all the tetrahedral C and N atoms are very similar, the three strongest multipoles being the same (Table 4). For instance, the static electron deformation densities generated by the  $-\text{CH}_3$  and  $-\text{NH}_3^+$  groups display a correlation coefficient of 0.96. The polypeptide main-chain C $\alpha$  atom, which loses both the  $mz$  mirror and the threefold symmetry, needs more multipoles to represent its electron density.

The hydroxyl O atoms in the library are of two types. The Ser and Thr hydroxyl groups are considered equivalent, whereas



**Table 4**  
Library parameters for some C and N tetrahedral  $sp^3$  atoms.

Multipoles antisymmetric with respect to the Z axis are marked  $m_{\perp z}$  in the last line. Atoms obeying  $m_{\perp z}$  mirror symmetry or threefold axis are marked as such in the last column. The multipoles in italics are expected to be small *a priori* as they are antisymmetric with respect to  $m_{\perp z}$ , this concerns atoms following  $m_{\perp z}$  or threefold symmetry.

Atom type	Axis X	Axis Y'	$P_{\text{val}}$	Dipoles			Quadrupoles				Octapoles						Atom symmetry							
				$x$	$y$	$z$	$z^2$	$xz$	$y^3$	$yz$	$x^2-y^2$	$xy$	$z^3$	$xz^2$	$yz^2$	$z(x^2-y^2)$		$xyz$	$x^3$	$y^3$	$\kappa$	$\kappa'$		
$\geq\text{CH}$	C $\alpha$ -Ala	C	N	4.111	0	-0.049	0	0	0	0	0	0	0	0	0	-0.171	-0.261	0	0	0.057	0.057	0.993	0.916	
$>\text{CH}-$	C $\beta$ -Ile	C $\alpha$	C $\gamma$ 1	4.097	-0.079	0	0	0	0	0	0	0	0	0	-0.193	-0.231	0	0	0	0	0	0.992	0.929	
$>\text{CH}_2$	C $\alpha$ -Gly	C	N	4.224	0	0	0	0	0	0	0.035	0	0	-0.170	-0.215	0	0	0	0.064	0.064	0.991	0.947	$m_{\perp z}$	
$-\text{CH}_3$	C $\beta$ -Ala	C	H	4.295	-0.093	0	0	0	0	0	0	0	0	-0.226	-0.140	0	0	0	0.054	0.054	0.984	0.883	$3x$	
$-\text{CH}_2-$	C $\beta$ -Arg	C $\alpha$	C $\gamma$	4.185	-0.053	0	0	0	0	0	0	0	0	-0.165	-0.226	0	0	0	0.065	0.065	0.989	0.911	$m_{\perp z}$	
$-\text{NH}_3^+$	NTerm	C $\alpha$	H1	5.560	-0.079	0	0	0	0	0	0	0	0	-0.162	-0.201	0	0	0	0.051	0.051	0.980	0.837	$3x$	
Antisymmetries							$m_{\perp z}$	$m_{\perp z}$	$m_{\perp z}$			$m_{\perp z}$						$m_{\perp z}$						

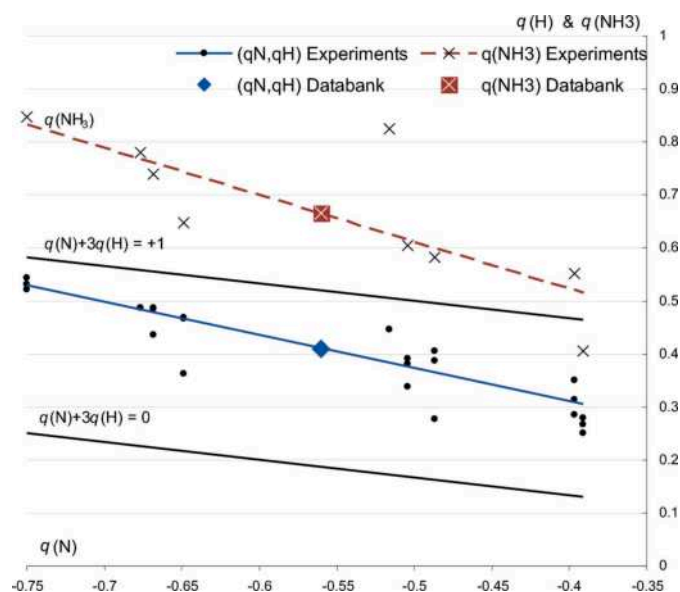
the tyrosine C—OH remains specific. This is related to the obvious difference between them: the Tyr hydroxyl group is conjugated with the aromatic ring and the topology of its lone pairs is dissimilar (Fig. 4). The lone pairs around the Tyr O atom are nearly merged. The conjugation between the ring and the Tyr hydroxyl group is confirmed by the smaller bond distances. The length of the Tyr C—O bond is 1.38 (2) Å, while for Ser/Tyr, the bond length is 1.42 (2) Å in the Engh & Huber (1991) stereochemistry dictionary. The same C—O bond-length model (Tyr < Ser ~ Thr) is found for average data obtained from neutron diffraction results [1.36 (2) and 1.43 (1) Å for Tyr and Ser/Thr, respectively (Allen, 1986)]. Resonance in the structure of Tyr suggests the hybridization

state of the Tyr O atom is between  $sp^2$  and  $sp^3$ , whereas for Ser and Thr the O atom is fully  $sp^3$ .

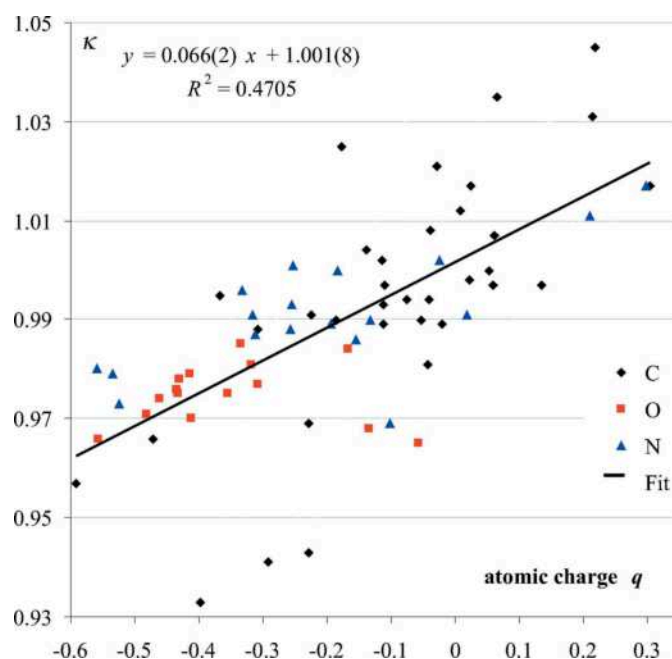
### 3.4. Atomic charges

The average net atomic charges  $q = N_{\text{val}} - P_{\text{val}}$  of the amino-acid residues generally accord with their electrostatic nature (Table 5). The deprotonated forms of the acidic residues Asp and Glu display negative charges of -0.837 and -0.670 |e|, respectively. Similarly, the basic residues show positive charges for their protonated form: Arg (+0.902 |e|), Lys (+1.145 |e|) and His (+0.655 |e|).

The ‘neutral’ residues do not have, as expected, an exactly zero charge in the library. Most of these residues have a global charge lower than 0.1 |e| in absolute value. The neutral resi-



**Figure 5**  
Atomic charge  $q_{\text{H}}$  of H atoms versus the atomic charge  $q_{\text{N}}$  of N atoms in the different  $-\text{NH}_3^+$  ammonium groups found in the peptides. The solid blue line is the least-squares fit of the  $(q_{\text{N}}, q_{\text{H}})$  points. The average value in the library is shown as a blue lozenge. The lower of the two parallel lines corresponds to electroneutrality  $q(\text{NH}_3) = 0$  of the ammonium group, whereas the upper line represents an  $-\text{NH}_3^+$  group with a formal charge of +1 |e|. The global charge  $q(\text{NH}_3)$  of the ammonium group and its least-squares fit (dashed red line) are also shown as a function of  $q_{\text{N}}$ .



**Figure 6**  
Relationship between  $\kappa$  expansion/contraction coefficients and atomic charges  $q = N_{\text{val}} - P_{\text{val}}$  for 27 C-atom types, 10 O-atom types and 11 N-atom types in the library. The linear fit for all atoms is shown as a black line.

**Table 5**

Global charges of the protein amino acids  $(N-HO)-(C\alpha-H\alpha)-R-(C=O)$  in the library.

The water molecule HOH, which has an initial charge of 0.050 |e|, has been neutralized.

Neutral residues	Charge ( e )	Acidic residues	Charge ( e )
Ala	-0.002	Asp	-0.837
Asn	+0.018	Glu	-0.670
Cystine	+0.113		
Cysteine	+0.070	Basic residues	
Gln	+0.185	Arg	+0.902
Gly	-0.097	Lys	+1.145
His	-0.398	Hip	+0.655
Ile	+0.437		
Leu	+0.437	-COO <sup>-</sup> term	-0.814
Met	+0.116	-NH <sub>3</sub> <sup>+</sup> term	+0.667
Phe	+0.012	HOH	0.000
Pro	+0.080		
Ser	-0.269	ASH (Asp + H <sup>+</sup> )	-0.294
Thr	-0.164	GLH (Glu + H <sup>+</sup> )	-0.127
Trp	+0.111		

dues with a polar side chain like His, Ser and Thr have a negative charge while hydrophobic residues have more positive charges. Molecular modelling and crystallographic software usually use a zero charge for neutral amino acids, while basic and acidic residues are assigned +1 and -1 |e| charge, respectively (Bayly *et al.*, 1993; Brünger *et al.*, 1998). The valence population is a quantity that is less transferable than multipoles of order  $l > 0$ . This is due to the electroneutrality constraint: the valence population  $P_{val}$  of a given atom depends definitely on the global molecular and asymmetric unit composition.

The repetitive group of atoms within the polypeptide main-chain motif denoted by  $-(N-HO)-(C\alpha-H\alpha)-(C=O)-$  has a negative charge: -0.190 |e|. This global charge is slightly different from the -0.10 |e| value found in the library based on theoretical calculations (Li *et al.*, 2006). The discrepancy may be attributed to experimental and theoretical uncertainties, and to the different chemical environments and atomic compositions of the peptides. The experimental charge-density analyses are performed on peptides in the crystalline state while molecules are *in vacuo* in the computations performed for the theoretical library (Pichon-Pesme *et al.*, 2004; Volkov, Koritsanszky *et al.*, 2004). The presence or absence of hydrogen bonding, through atomic polarization, has an impact on the experimental deformation density of polar groups (Spackman *et al.*, 1999; Fernandez-Serra *et al.*, 2000) and therefore affects the valence populations.

For the glycine residue, which bears a second positively charged H atom on the C $\alpha$  atom, the experimental charge of the global peptide is less negative: -0.097 |e|. The H atoms on -CH<sub>3</sub>, >CH<sub>2</sub> and >CH- (e.g. C $\beta$  of Val) groups have increasing positive charges of +0.161, +0.176 and +0.191 |e|, respectively. The C atoms of these same groups have negative charge of -0.295, -0.185 and -0.097 |e|, respectively, which also increase with the number of bound H atoms.

The atomic charges of the ammonium -NH<sub>3</sub><sup>+</sup> groups in all the peptides are analysed in Fig. 5. The  $q_H$  charges of H atoms are between +0.35 and +0.55 |e|, and the corresponding  $q_N$  charges of N atoms are between -0.39 and -0.75 |e|. All the experimental ( $q_N$ ,  $q_H$ ) data points as well as the least-squares fit are located in the zone between the two parallel lines  $q(NH_3) = 0$  and  $q(NH_3) = +1$  |e| in Fig. 5.

The relationship between the charges of the N and H atoms in different -NH<sub>3</sub><sup>+</sup> groups is clearly visible, as they are anti-correlated by -88%. N atoms with a strong negative charge have the tendency to bear H atoms with a high positive charge. As a result, the correlation coefficient between  $q_H$  and  $q(NH_3)$  is +92% while that between  $q_N$  and  $q(NH_3)$  is -79%. The charge of the ammonium group is between +0.40 and +0.84 |e| with an average library value of +0.67 |e|. No correlation was found between the thermal factors  $U_{eq}$  and the charges of the N atoms.

During library transfer, the atomic charges of a neutral residue may be left as such or may be modified to fit the residue electroneutrality constraint. For example, when applied to glycine, the charge of the residue resulting from the library is  $q = -0.097$  |e|, which may be corrected by subtracting  $q/7 = -0.014$  |e| from the valence populations of all seven atoms.

When the library electron density is transferred to a neutral molecule, the electroneutrality can be preserved by distributing the charge difference over all the atoms of the molecule. For example, the library charge of the Leu-enkephalin molecule is +0.361 |e| upon transfer; then a trifling value of +0.004 |e| is added to the 86 atoms of the molecule to achieve electroneutrality. An automatic procedure, during the library transfer (B. Guillot *et al.*, 2001; Jelsch *et al.*, 2005), permits charge neutralization for the whole molecule or for selected amino acids expected to be electrically neutral. The amount of redistributed fraction-of-an-electron charge on each atom is small, and decreases with the number of atoms in the molecule; which is an important point for protein structure ELMAM refinement (see §3.6).

Atomic charges can also be obtained by integration of the electron density over the atomic basins (Bader, 1990). These topological charges, computed after database transfer on several peptides (Leherte *et al.*, 2006), turn out to be generally larger than the charges derived from the  $N_{val\%} - P_{val}$  difference.

### 3.5. $\kappa$ parameter versus net atomic charge

The relation between expansion-contraction coefficient  $\kappa$  of the valence shell and net atomic charge  $q$  is shown in Fig. 6. As reported earlier by Coppens *et al.* (1979) and by Volkov *et al.* (2001),  $q$  correlates well with the  $\kappa$  parameter derived from multipole refinements versus both experimental and density-functional theory (DFT) crystal structure factors for the C, N and O atoms. The same correlation is found for average data derived from experimental ultra-high-resolution crystallographic studies of peptides and in the present multipole library. The fitted linear equation  $\kappa = 1.001(8) + 0.066(2)q$

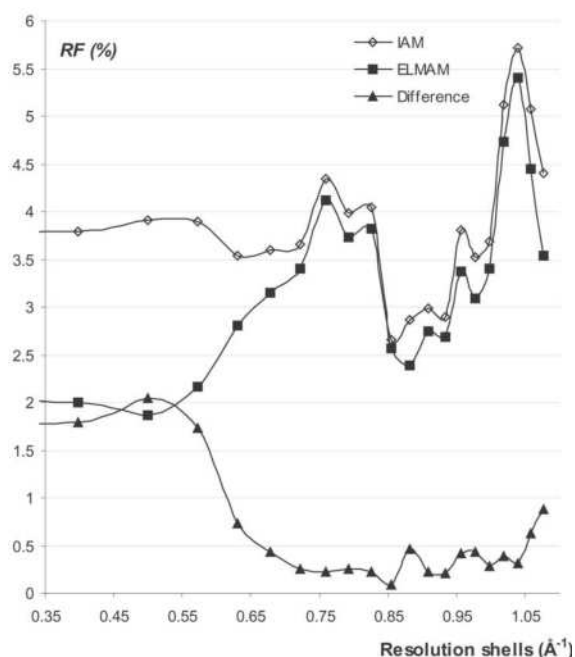
**Table 6**

Final agreement statistics in the IAM and ELMAM refinements for selected peptide structures.

Targets for  $X-H$  distance restraints were as in *SHELXL* (Sheldrick & Schneider, 1997) in the IAM case and from neutron diffraction in the ELMAM case.  $RF = \sum |F_o - F_c| / \sum F_o$ ;  $wR^2F = \{ \sum [(F_o - F_c) / \sigma F_o]^2 / \sum [(F_o / \sigma F_o)^2] \}^{1/2}$ .

$X-H$ distance restraints applied		Yes	Yes	No	No	
Refinement		IAM	ELMAM	IAM	ELMAM	
Compound	Temperature (K)	$RF$ (%)	$wR^2F$ (%)	$\Delta RF$	$RF$ (%)	
Resolution ( $\text{\AA}$ )					$wR^2F$ (%)	
L-Alanyl-L-threonine		3.08	1.63	1.45	3.08	1.61
0.62	105 (2)	3.40	1.60		3.39	1.55
L-Seryl-L-valine		3.63	2.79	0.84	3.58	2.68
0.73	105 (2)	4.29	3.05		4.14	2.91
L-Threonyl-L-alanine		3.76	2.49	1.27	3.76	2.42
0.76	105 (2)	3.11	1.81		3.06	1.72
L-Seryl-L-phenylalanine		4.55	3.95	0.60	4.49	3.91
0.79	105 (2)	4.33	3.62		4.24	3.57
Pseudophomin A		3.91	2.99	0.99	3.79	2.97
0.85	100 (2)	4.37	3.25		4.28	3.22
Triglycine		3.75	2.74	1.01	3.76	2.72
0.46	123 (2)	4.53	3.16		4.52	3.14

gives a  $\kappa$  parameter of 1.0015 for a neutral atom. This linear fit is in perfect agreement with value  $\kappa = 0.999(1) + 0.063(2)q$  published for DFT crystal structure factors (Volkov *et al.* 2001). The correlation coefficient between  $\kappa$  and  $q$  is 65% for C, N and O atoms; the scatter of data included in the database calls for further expansion of the transferable molecular electron-density base. No such correlation between the experimental expansion/contraction  $\kappa$  and the net atomic charge  $q$  was found for H atoms, presumably due to a deficiency of the isotropic Debye–Waller refinement.



**Figure 7**

Analysis of  $RF$ -factor improvement upon library transfer as a function of reciprocal resolution, in the case of the triglycine peptide at  $\sin \theta / \lambda = 1.05 \text{ \AA}^{-1}$

### 3.6. Refinement of peptides with and without library transfer

We compare in the following two types of refinement models: the usual independent-atom model (IAM) and the experimental library multipolar atom model (ELMAM). To confirm that the ELMAM refinements improve significantly the crystal structure and result in more accurate and physically meaningful thermal displacement parameters, we have analysed several peptide compounds published in *Acta Crystallographica* as their diffraction data are available.

The influence of multipolar population parameters transfer on the improvement of the principal refinement parameters was investigated on the compounds L-alanyl-L-threonine

(Netland *et al.*, 2004), L-seryl-L-phenylalanine (Helle *et al.*, 2004), L-threonyl-L-alanine (Görbitz, 2005), L-seryl-L-valine (Moen *et al.*, 2004) and triglycine (Pichon-Pesme & Lecomte, 1998). A pseudo-peptide of significantly larger molecular weight ( $1158 \text{ g mol}^{-1}$ ), pseudophomin A (Quail *et al.*, 2002) with primary structure hydroxydecanoyl-L-Leu-D-Glu-D-*allo*-Thr-D-Ile-D-Leu-D-Ser-L-Leu-D-Ser-L-Ile monohydrate was also tested. The resolution of the diffraction data for all these compounds is summarized in Table 6.

The refinement for all these compounds was done in two stages. In the IAM refinements, a conventional spherical neutral atom model was applied. Scale factor, atomic positions and thermal displacement parameters for all atoms were refined using the *MoPro* program until convergence. For the H atoms, the isotropic  $B$  factors and the  $X-H$  bond distances were restrained. A restraint gives a target to a structural parameter or a derived quantity and allows a standard uncertainty from the target. The target values of  $X-H$  distance restraints were set to the same values as in the *SHELXL* program (Sheldrick & Schneider, 1997), *i.e.*  $-CH$   $0.98 \text{ \AA}$ ;  $-CH_2$   $0.97 \text{ \AA}$ ;  $-CH_3$   $0.96 \text{ \AA}$ ;  $-NH_3^+$   $0.89 \text{ \AA}$ ;  $C_{ar}-H$   $0.93 \text{ \AA}$ ;  $-OH$   $0.82 \text{ \AA}$ , with allowed standard uncertainty of  $0.005 \text{ \AA}$ . URATIO restraints (riding-motion thermal parameters in the *MoPro* syntax) were set to 1.5 for  $-CH_3$ ,  $-OH$ ,  $-NH_3^+$  groups with a remaining rotation degree of freedom and 1.2 for  $>CH_2$ ,  $>C_{ar}-H$ ,  $C-H$ ,  $-NH_2$ ,  $>NH$  groups with deviation of  $0.005 \text{ \AA}^2$ .

In the ELMAM refinements, the same parameters were varied, but a multipolar charged atom model was applied. The electron-density parameters (valence and multipole populations, expansion/contraction coefficients) were transferred from the library and subsequently kept fixed. The URATIO thermal restraints on H-atom  $B$  factors were applied similarly to IAM refinement. The target values for  $X-H$  distance restraints were set to the values expected from neutron diffraction studies [*e.g.*  $-CH$   $1.099(4) \text{ \AA}$ ;  $-CH_2$   $1.092(12) \text{ \AA}$ ;

**Table 7**

Crystallographic  $wR^2F$  and  $wR^2F^{\text{free}}$  factors for the L-alanyl-L-threonine and triglycine peptides after IAM and ELMAM refinements.

The improvement of the statistics upon transfer is also shown as the difference (IAM-ELMAM),  $wR^2F = \{ \sum[(F_o - F_c)/\sigma F_o]^2 / \sum[(F_o/\sigma F_o)^2] \}^{1/2}$ .

Compound	Refinement	$wR^2F$	$\langle wR^2F^{\text{work}} \rangle$	$\langle wR^2F^{\text{free}} \rangle$
Ala-Thr	IAM	3.40	3.39 (2)	3.8 (5)
	ELMAM	1.60	1.60 (1)	1.8 (3)
	IAM – ELMAM	1.8	1.79	2.0
Gly-Gly-Gly	IAM	4.53	4.52 (4)	4.8 (7)
	ELMAM	3.16	3.16 (4)	3.3 (7)
	IAM – ELMAM	1.37	1.36	1.5

–CH<sub>3</sub> 1.059 (30) Å; –NH<sub>3</sub><sup>+</sup> 1.033 (22) Å; >C<sub>ar</sub>–H 1.083 (11) Å; –OH 0.967 (10) Å (Allen, 1986)] with deviation of 0.005 Å. The IAM and ELMAM refinements were also repeated with no X–H distance restraints. All refinements were done against all the available resolutions applying a zero  $I/\sigma(I)$  cut-off.

The agreement statistics of both refinements are compared for all the peptides in Table 6. The ELMAM refinement shows dramatic improvement in statistical indexes. For example, the *R* factor (*RF*) is reduced from 3.08 to 1.63% for L-alanyl-L-threonine at 0.62 Å resolution. The improvement was smallest for L-seryl-L-phenylalanine with *RF* decreasing from 4.55 to 3.95% at 0.79 Å resolution. This demonstrates that ELMAM refinement is statistically better than IAM refinement. As expected, the resolution limit of the diffraction data does not radically influence the improvement of refinement statistics  $\Delta RF \simeq 1\%$ , since valence scattering contributes significantly only at low and medium resolutions. Consequently, the ELMAM refinement notably improves the crystallographic figures, not only for high resolution but also for usual crystal structures. This is confirmed by the analysis of *RF* values in different resolution shells for the 0.46 Å structure of

triglycine (Fig. 7). The *R*-factor improvement, in this case, is important essentially in the lower resolution range ( $s < 0.7 \text{ \AA}^{-1}$ ), where the contribution of the deformation density part to the atomic scattering factors is significant.

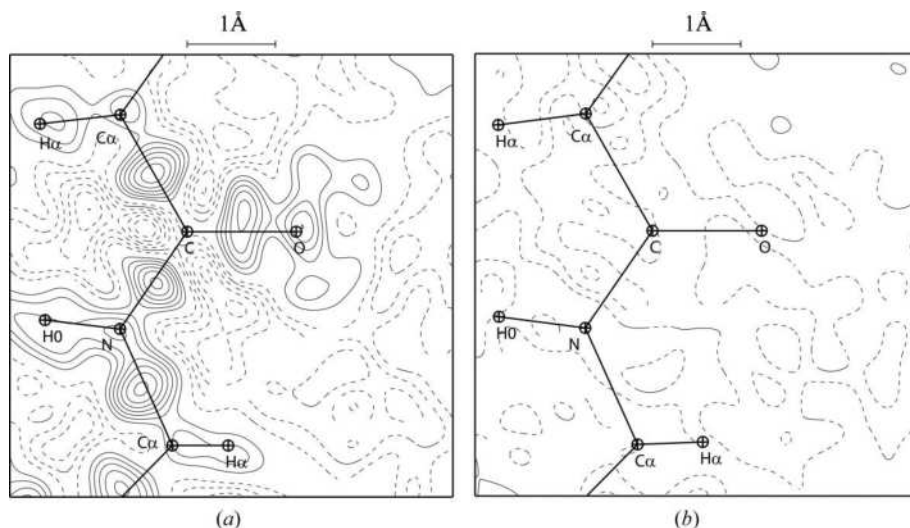
Even for pseudophomin A (Cu *K*α diffraction source, 180 symmetrically independent atoms), the drop of *RF* was a noteworthy 24%. This encourages use of the ELMAM refinement for Cu *K*α diffraction data, for lower resolution and for larger molecules.

Since the conventional *RF* factor is not a totally objective indicator of model quality, we have performed some *R*-free refinements (Brünger, 1992, 1993). The general principle of such a refinement consists in dividing the diffraction data into a large ‘working’ set (96% of reflections in our case) and a small ‘test’ set (4%), which contains the free reflections. Free reflections are not used in the refinement, but the  $wR^2F^{\text{free}}$  value is computed on that subset and is expected to decrease along with the conventional *RF* factor. We have applied this so-called cross-validated refinement method for the L-alanyl-L-threonine and triglycine structures using the following procedure.

The *hkl* reflections were subdivided to define 25 working sets and their corresponding 25 complementary test sets. This subdivision allowed us to execute  $N = 25$  refinements with a different test set for each. For instance, in refinement 1, the first free reflections are reflections 1, 26, 51, ... In refinement 2, the test set consists of reflections 2, 27 etc. Using this strategy, we obtained a set of 25 refined  $wR^2F^{\text{work}}$  and 25  $wR^2F^{\text{free}}$  factors; their average values are listed in Table 7. The averaged  $wR^2F^{\text{free}}$  factor can be expected to be more precise than a single  $wR^2F^{\text{free}}$  factor and less dependent on the chosen test set. The crystallographic factors for L-alanyl-L-threonine and triglycine are in line with the expected order  $wR^2F^{\text{free}} > wR^2F > wR^2F^{\text{work}}$ , for both IAM and ELMAM refinements. For the L-alanyl-L-threonine structure, the  $wR^2F^{\text{work}}$  and

$wR^2F^{\text{free}}$  factors decreased upon transfer by 1.8 and 2.0 percentage points, respectively, which represents a 50% drop in relative value. The  $wR^2F^{\text{free}}$  factors decrease slightly more than the  $wR^2F^{\text{work}}$  factors, which clearly indicates that the ELMAM refinement is superior.

The improvement of the refined model is also shown by the residual electron-density maps depicted in Fig. 8. For the same diffraction data set, the flatter the maps are, the better the scattering factor model is. After IAM refinement, some residual density remains systematically on the covalent bonds, up to  $0.35 \text{ e \AA}^{-3}$ , whereas after ELMAM refinement very little such residual density remains.



**Figure 8**

Fourier difference map of L-alanyl-L-threonine after (a) IAM and (b) ELMAM refinements. Contours are at  $\pm 0.05 \text{ e \AA}^{-3}$  with full and dashed lines denoting positive and negative density, respectively.

**Table 8**

$X-H$  distances (Å) for IAM and ELMAM refinements of L-alanyl-L-threonine, compared with average values obtained from neutron diffraction (Allen, 1986).

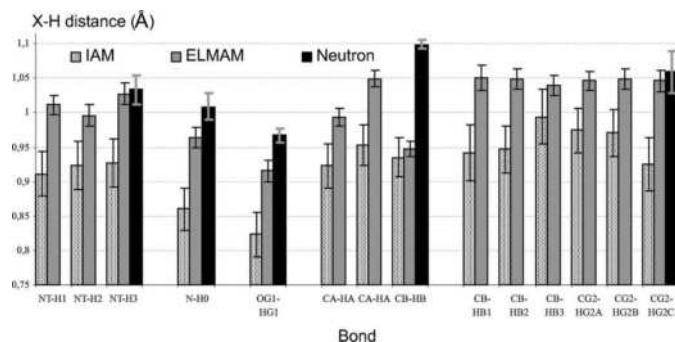
No  $X-H$  distance restraints were applied.

Residue	$X-H$ bond	IAM refinement	ELMAM refinement	Neutron diffraction
Ala	NT–H1	0.911 (32)	1.011 (14)	1.033 (22)
Ala	NT–H2	0.923 (35)	0.996 (16)	
Ala	NT–H3	0.927 (35)	1.027 (16)	
Thr	N–H0	0.860 (31)	0.964 (14)	1.009 (19)
Ala	$C\beta-H\beta1$	0.942 (41)	1.050 (18)	1.059 (30)
Ala	$C\beta-H\beta2$	0.947 (34)	1.048 (15)	
Ala	$C\beta-H\beta3$	0.994 (40)	1.039 (17)	
Thr	$C\gamma2-H\gamma2A$	0.974 (32)	1.046 (14)	
Thr	$C\gamma2-H\gamma2B$	0.971 (34)	1.049 (15)	
Thr	$C\gamma2-H\gamma2C$	0.925 (38)	1.046 (16)	
Ala	$C\alpha-H\alpha$	0.923 (32)	0.994 (13)	1.099 (7)
Thr	$C\alpha-H\alpha$	0.953 (30)	1.049 (12)	
Thr	$C\beta-H\beta$	0.935 (29)	0.947 (11)	
Thr	$O\gamma1-H\gamma1$	0.823 (32)	0.915 (15)	0.967 (10)

### 3.7. Effect of transfer on the structure

The effect of the ELMAM refinement is significant on  $X-H$  bond distances when  $X-H$  distance restraints are not applied, as analysed in the case of the peptide L-alanyl-L-threonine. The improvement of  $X-H$  distances towards the average values obtained from neutron diffraction (Allen, 1986) is clearly visible (Table 8, Fig. 9). The differences between H–N bond lengths in  $-NH_3^+$  groups obtained from the ELMAM refinement and from neutron diffraction do not exceed 0.037 Å, compared to 0.122 Å for the IAM refinement. For the  $-CH_3$  groups, the largest discrepancy is 0.020 and 0.134 Å with the ELMAM and IAM structure refinements, respectively.

The tertiary C–H bond lengths from the ELMAM refinement do not agree as well with the neutron diffraction literature (discrepancy from 0.050 up to 0.152 Å). The H-atom coordinates after ELMAM refinement are however still closer to those obtained from a neutron diffraction experiment


**Figure 9**

Values of  $X-H$  distances in the structure of L-alanyl-L-threonine. Light grey column values are obtained from the IAM refinement, grey columns values obtained from the ELMAM refinement. The black columns refer to average values obtained from neutron diffraction (Allen, 1986). Error bars are indicated for all values.

compared to the IAM refinement. This was foreseen in Pichon-Pesme *et al.* (1995) and analysed also by Dittrich *et al.* (2005) on the amino acid DL-Ser by transferring from a theoretical database.

The effect on other  $X-Y$  bonds (involving C, O and N atoms) is small. The bond lengths in all the studied peptides remained similar within two standard uncertainties  $\sigma$  after the ELMAM refinement, as observed previously (Pichon-Pesme *et al.*, 1995). The largest bond-length change over uncertainty  $\Delta d/\sigma$  was observed for covalent bonds involving the O atom in the serine residue of L-seryl-L-valine and of L-seryl-L-phenylalanine ( $\Delta d/\sigma = 1.8$  and 1.7, respectively). The r.m.s. (root mean square) value of  $\Delta d$  is 0.0026 Å for all the bonds in L-seryl-L-phenylalanine. This small value confirms that there is no notable change in the  $x$ ,  $y$ ,  $z$  parameters after IAM and ELMAM refinements for C, O and N atoms.

### 3.8. Effect of transfer on thermal parameters

The effect of the transfer on the thermal parameters has been discussed in some of our latest publications (*e.g.* Pichon-Pesme *et al.*, 1995; Jelsch *et al.*, 1998). In the present paper, we have chosen for the thermal motion analysis the dimethylstilbene molecule (Ogawa *et al.*, 2006), as the effect of the atom model is expected to be dramatic for a diffraction measurement at ultra-low temperature  $T = 16$  K (Jelsch *et al.*, 2005). The anisotropic displacement parameters (ADPs) of both the IAM and ELMAM refined structures were analysed using the program *THMA14* (Schomaker & Trueblood, 1998) from the *WinGX* package (Farrugia, 1999). Half of the molecule, which is also the asymmetric unit content, was treated as a single rigid group.

The rigid-body motion is described by three tensors **T**, **L** and **S** accounting for translation, libration and screw motion (correlation between translation and libration), respectively. These tensors are determined by applying a least-squares fit between calculated and observed  $U_{ij}$ 's (Schomaker & Trueblood, 1968). The results may be checked by calculating an  $R_U$  residual factor, which is

$$R_U = \left[ \frac{\sum_{ij} |U_{ij} - U_{ij}^{\text{TLS}}|^2}{\sum_{ij} |U_{ij}|^2} \right]^{1/2}. \quad (5)$$

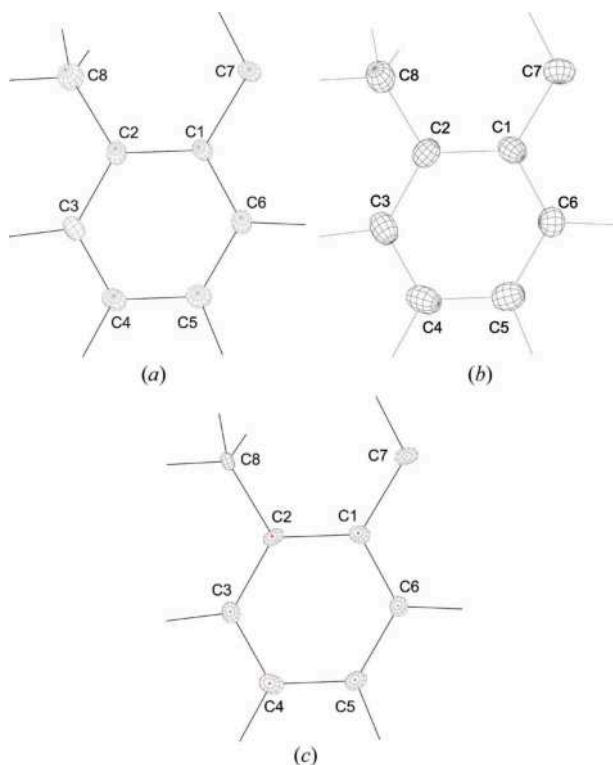
The  $R_U$  factor for the IAM structure amounts to 0.151 for all  $U_{ij}$ 's and to 0.117 for diagonal elements  $U_{ii}$  only. The ELMAM refinement results in decreased values  $R_{Uij} = 0.092$  and  $R_{Uii} = 0.063$ , respectively. A low value of the  $R_U$  factor indicates that the rigid-body behaviour of the selected group of atoms is well described by the modelled ADPs. Figs. 10(a), (b), where the atomic thermal ellipsoids after IAM and ELMAM refinements are depicted, are good illustrations of the  $U_{ij}$  changes. The atomic  $U_{eq}$  values, after ELMAM refinement, show a reduction ranging between 26 and 33%.

The  $U_{ij}$ 's differences are represented in Fig. 10(c) with program *PEANUT* (Hummel *et al.*, 1990), which allows the representation of non-positive-definite differences of thermal tensors. The  $U_{ij}$  values are larger in the spherical-atom case in the plane of the aromatic ring as they try to incorporate the

non-modelled bonding deformation electron density. In the direction perpendicular to the ring, the  $U_{ij}$  difference is on the other hand slightly negative. This is reminiscent of the electron depletion above and below the aromatic C atoms which is modelled by negative  $P_z^2$  quadrupole populations (Table 2).

To determine if the anisotropic thermal displacement parameters  $U_{ij}$  account only for dynamic and static disorder and are physically realistic, we have performed the rigid-bond test (Hirshfeld, 1976). The general idea of the test is to analyse the differences between the r.m.s. displacements along the bonds between neighbouring atoms. The discrepancy should be small:  $\Delta Z^2 < 10^{-3} \text{ \AA}^2$  (Hirshfeld, 1976). For dimethylstilbene, the average  $|\Delta Z^2|$  after IAM refinement is  $9.8 \times 10^{-4} \text{ \AA}^2$ , whereas after ELMAM refinement it drops to  $5.5 \times 10^{-5} \text{ \AA}^2$ , which represents a 94% improvement. This analysis undoubtedly illustrates that the result of the refinement is physically acceptable and the vibration parameters are properly derived from the ELMAM method.

The rigid-bond test was also applied to the other studied molecules. The r.m.s. values of the rigid-bond test after IAM and ELMAM refinements are presented in Table 9. The results of the test are generally better for the structures with the highest diffraction resolution (triglycine and dimethylstilbene). The  $\Delta Z^2$  values have a systematic tendency for all the molecules to be improved (about 30% reduction on average) after ELMAM compared to IAM refinements.

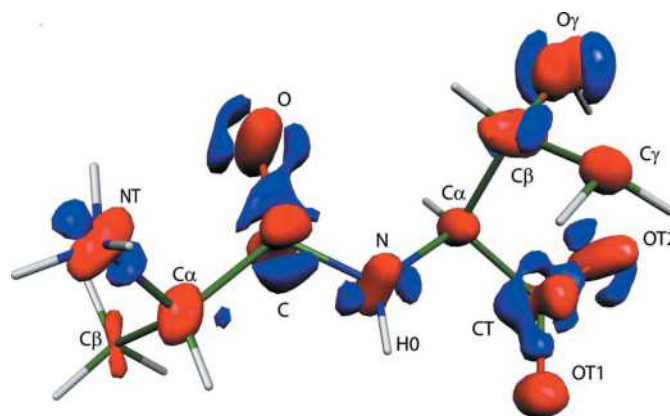


**Figure 10** View of thermal  $U_{ij}$  ellipsoids for dimethylstilbene after (a) ELMAM, (b) IAM structure refinements, (c)  $\Delta U_{ij}$  difference IAM – ELMAM. Representation of 90% r.m.s.u. with program *PEANUT* (Hummel *et al.*, 1990). Red refers to non-positively defined thermal tensor  $U_{ij}$  differences.

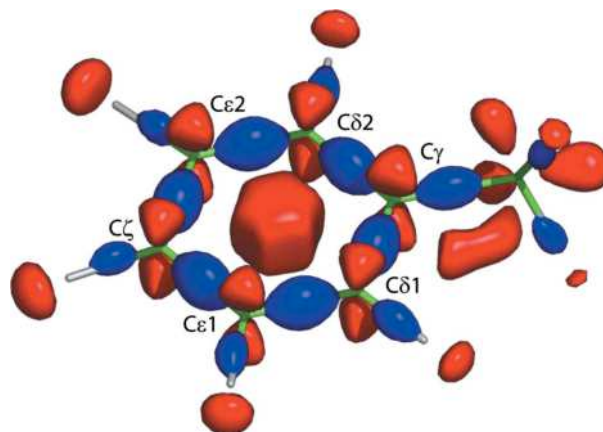
**Table 9** Root-mean-square values of the rigid-bond test for selected peptides after IAM and ELMAM refinements.

$\Delta Z^2$  is the difference between the two atomic  $U_{ij}$ 's projected on the bond direction.

Compound	$\Delta Z^2 [\times 10^4] (\text{\AA}^2)$	
	IAM	ELMAM
Dimethylstilbene	9.8	0.5
L-Alanyl-L-threonine	9.8	6.3
L-Seryl-L-valine	22.0	16.1
L-Threonyl-L-alanine	26.9	16.8
L-Seryl-L-phenylalanine	31.4	29.0
Pseudophomin A	29.4	20.6
Triglycine	8.6	4.7

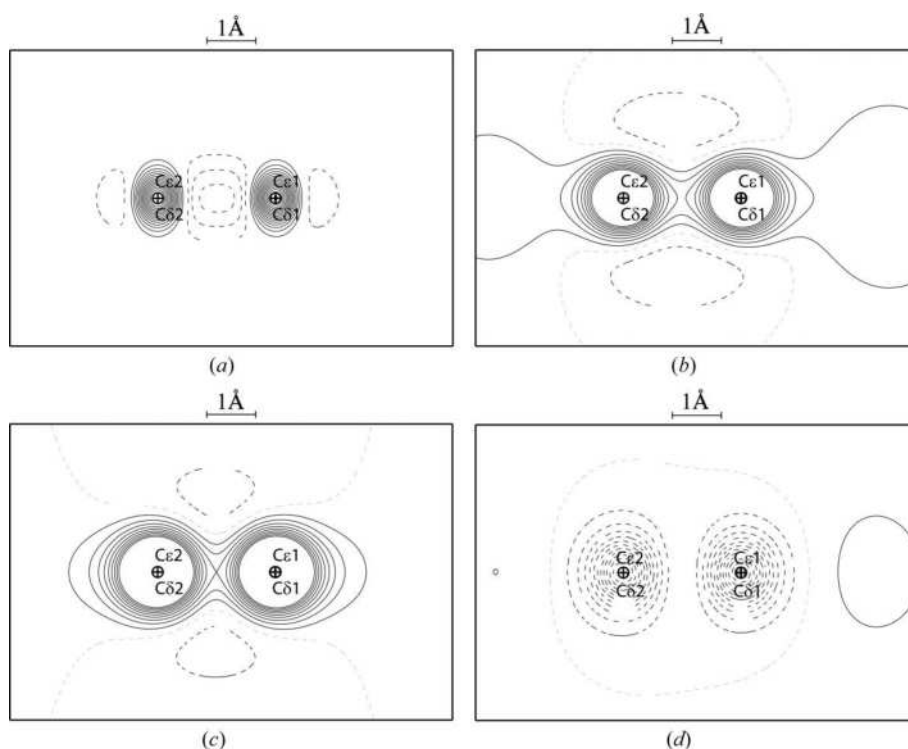


**Figure 11** Three-dimensional electron-density view representing the  $U_{ij}$  difference for L-alanyl-L-threonine. Such representation is obtained by computing the Fourier transform of the difference between the IAM structure factors and those computed from the ELMAM refinement with multipolar parameters reset to zero after the refinement. The grid map was computed with reflections up to  $0.4 \text{ \AA}$  resolution. The view was created with program *Molekel* (<http://www.cscs.ch/molekel>). The red and blue surfaces describe negative and positive density, respectively. The cut-offs for positive/negative surfaces are  $\pm 0.02 \text{ e \AA}^{-3}$ .



**Figure 12** Three-dimensional static deformation-density map of the phenyl group  $\text{C}_6\text{H}_5\text{-CH}_2\text{-}$  after library transfer. The view was created with program *Pymol* (DeLano, 2002). Blue surfaces denote positive density and red denote negative density. The cut-offs for positive and negative surfaces are  $+0.4$  and  $-0.1 \text{ e \AA}^{-3}$ , respectively.

The improvement of anisotropic thermal parameters after ELMAM refinement can also be visualized by calculating  $U_{ij}$  difference-Fourier electron-density maps. The first set of structure factors in the difference-Fourier synthesis is obtained from the IAM refined molecular structure. The second set is computed from the molecular structure (coordinates and ADPs) after ELMAM refinement but with multipole parameters reset to zero (IAM). This  $U_{ij}$  difference map includes mostly detailed information about the ADP's improvement due to the presence of non-spherical scattering factors. An example of such a map is shown for L-alanyl-L-threonine in Fig. 11, where the blue and red surfaces describe positive and negative values of the difference electron density, respectively. It is clear that some positive  $U_{ij}$  difference electron density is distributed around O atoms for instance and corresponds to electron lone pairs and bonding density. Some covalent bonds (NT—C $\alpha$ , N—C, N—C $\alpha$  and C=O) also display significant positive electron density. Fig. 11 shows generally negative difference densities located on all heavy-atom positions. This is a consequence of the overall decrease of the thermal-motion amplitudes when ELMAM refinement is performed. In the case of the N and NT atoms, the positive  $U_{ij}$  difference density peaks try to reproduce the covalent bonds but they are displaced with respect to the bonds. This is due to the atomic thermal ellipsoids, which are constrained to be centrosymmetric around each atom as a result of the harmonic approximation.



**Figure 13** Electrostatic potential of the phenylalanine residue in the plane passing through the middle and perpendicular to the C $\delta$ 1—C $\epsilon$ 1 and C $\delta$ 2—C $\epsilon$ 2 bonds. (a) Static deformation electron density; (b) corresponding electrostatic potential obtained from transferred parameters; (c) electrostatic potential obtained from point charges of the AMBER dictionary; (d) difference potential ELMAM — AMBER. Solid black lines: positive contour; dashed black lines: negative contour, dashed grey lines: zero contour.

This significant  $\Delta U_{ij}$  electron density is an illustration that the ADPs from the spherical-atom model suffer a significant bias due to the non-modelled deformation electron density. This difference-density representation, to our knowledge not described before, is then a clear indicator of real-space localization of thermal-motion bias in IAM refinements against high-resolution X-ray data. In conclusion, the ELMAM refinements lead clearly to a better estimation of the atomic thermal ellipsoids (Pichon-Pesme *et al.*, 1995; Jelsch *et al.*, 1998; Dittrich *et al.*, 2005).

### 3.9. Electrostatic potential

An important application of the library is the electrostatic potential, which is a fundamental quantity for understanding the reactivity and intermolecular properties. The potential can be directly computed from the charge-density library (Muzet *et al.*, 2003). The aromatic ring of the phenylalanine residue from Leu-enkephalin (Pichon-Pesme *et al.*, 1992) was selected to compare the ELMAM-derived electrostatic potential to that computed from the AMBER atomic point charges (Bayly *et al.*, 1993).

A three-dimensional view of the phenyl deformation electron density can be seen in Fig. 12. In Fig. 13, the deformation density and electrostatic potentials are shown in the plane perpendicular to the C $\delta$ 1—C $\epsilon$ 1 bond at its midpoint. The bonding electron density on the C—C aromatic covalent bonds is clearly elliptical, elongated in the direction perpendicular to the ring in Fig. 13(a). This is due to the  $\pi$  character of C—C phenyl bonds.

Each aromatic C atom follows a local  $mm2$  symmetry. As a consequence, the C atoms have only two multipoles which are significantly different from zero in the library:  $P_z^2$  and  $P_x^3$  (Table 2).

The absence of horizontal symmetry of the potential in Fig. 13(b) is due to the presence of H $\beta$ 1 and H $\beta$ 2 atoms of the CH<sub>2</sub> group connected to the aromatic ring, which are twisted towards the C $\delta$ 1 atom direction (Fig. 12). The bond critical points (BCPs) of the C—C bonds in the aromatic ring are located at the midpoints; the ellipticity is 0.16, which is comparable to the value of 0.22 obtained by Koritsanszky & Coppens (2001). The total electron density  $2.24 \text{ e } \text{\AA}^{-3}$  and the Laplacian  $-21.96 \text{ e } \text{\AA}^{-5}$  at the BCP correlate well with typical experimental values

given in the same review ( $2.15 \text{ e } \text{\AA}^{-3}$  and  $-16.9 \text{ e } \text{\AA}^{-5}$ , respectively).

Fig. 14 shows the deformation density and electrostatic potential in the plane perpendicular to the aromatic ring and parallel to the  $\text{C}\zeta\text{--H}\zeta$  bond. The centre of the cycle is an electron-depleted region (Figs. 13*a* and 14*a*). The regions located around the H atom, as well as above and below the aromatic ring plane near the C-atom positions, are also electron depleted in Fig. 14(*a*). The last electron depletion zone is modelled by the negative  $P_z^2$  quadrupole populations of the aromatic C atoms (Table 2) and is easily visible in Fig. 12.

The phenyl H atoms carry a positive charge of  $+0.17 \text{ |e|}$  in the library. The AMBER charges of the H atom range from  $+0.17$  to  $+0.183 \text{ |e|}$ , which is quite comparable. The  $\text{C}\gamma$  atom is nearly neutral in both cases:  $-0.04 \text{ |e|}$  in the multipolar library and  $+0.01 \text{ |e|}$  in the AMBER dictionary. The C(H) atoms carry a negative charge of  $-0.155 \text{ |e|}$  in the multipolar library, which is in good agreement with the charge in the AMBER library in the  $-0.12$  to  $-0.17 \text{ |e|}$  range.

There is good agreement between the two charge representations with regard to the zero potential contour separating electronegative and electropositive regions in Figs. 13(*b*), (*c*) and 14(*b*), (*c*). The electrostatic potentials derived from the

ELMAM transfer and from the AMBER point charges show the largest discrepancies on the C—C bond (Fig. 13*d*) and in the C—H bond regions in the difference maps (Fig. 14*d*). These bonding regions are more electronegative in the multipolar model due to the proper representation of the bonding electron density. The most important regions for electrostatic potential consideration with respect to interatomic interactions are however located outside the van der Waals surface which show less disagreement between the two charge models.

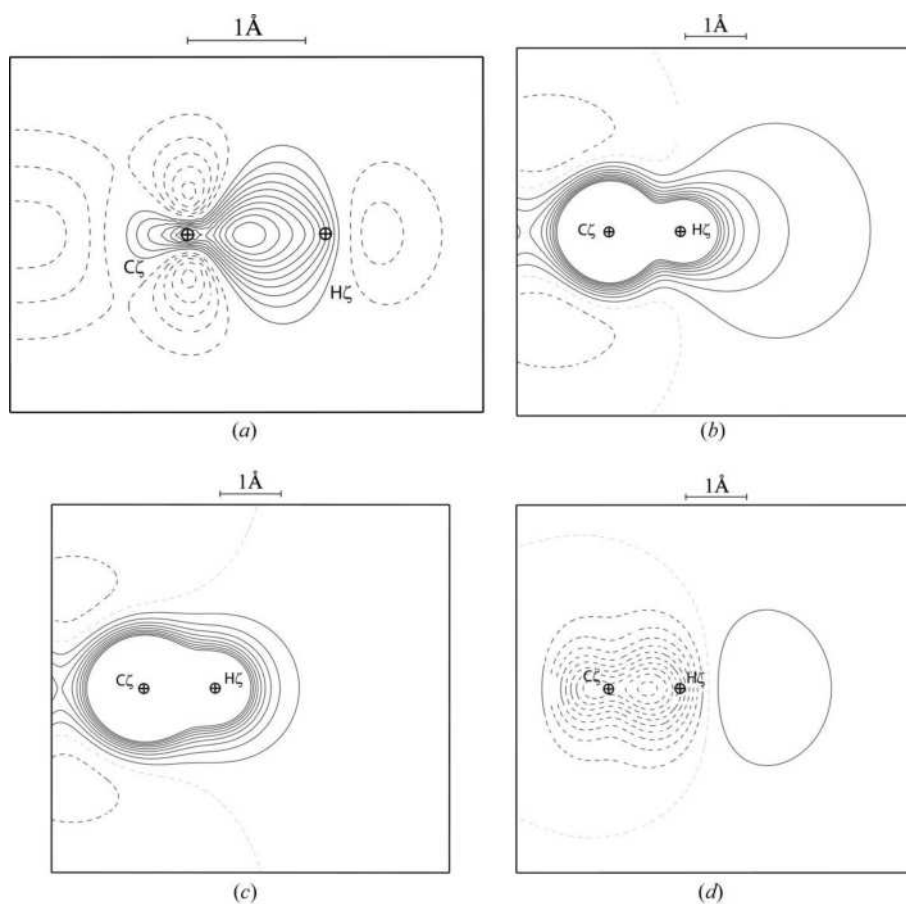
The electronegative potential on both sides of the aromatic ring, due to  $\pi$  bonding, is slightly stronger in the ELMAM case compared to the AMBER modelling. In the phenyl ring plane, when the regions of interactions at van der Waals distance from atoms are considered, the potential is more electropositive in the ELMAM case (Figs. 13*d*, 14*d*). The impact of the hydrogen dipole is obvious here. Despite similar charges in both models, the electropositive potential around the H atoms extends to a longer range in the ELMAM case.

### 3.10. Transfer to proteins

Since 1990, the amount of high- or ultra-high-resolution protein X-ray diffraction data has increased considerably, as demonstrated in the Protein Data Bank (Berman *et al.*, 2002) records and by recent reviews (Longhi *et al.*, 1998; Schmidt & Lamzin, 2002; Petrova & Podjarny, 2004). As of September 2006, the number of independent protein structures refined at an atomic resolution better than  $1.0 \text{ \AA}$  was 251, and there were 13 structures at a resolution better than  $0.7 \text{ \AA}$ .

We describe briefly an application on human aldose reductase. Diffraction data were collected to  $0.66 \text{ \AA}$  resolution at the APS synchrotron for the enzyme complexed with  $\text{NADP}^+$  and the inhibitor IDD 594, a candidate therapeutic agent for treatment of diabetic complications (Cachau *et al.*, 2000; Howard *et al.*, 2004; Ruiz *et al.*, 2004). The complete analysis of the protein electron-density distribution with the *MoPro* package will be published in a forthcoming paper.

The structure factors generated by the bulk solvent were added to the protein structure factors, as described in Jelsch *et al.* (2005), with a solvent scale factor and a thermal factor



**Figure 14**

Electrostatic potential of the phenylalanine side chain in the plane perpendicular to the aromatic cycle and parallel to the  $\text{C}\zeta\text{--H}\zeta$  bond. (*a*) Static deformation electron density; (*b*) corresponding electrostatic potential obtained from transferred parameters; (*c*) electrostatic potential obtained from point charges of the AMBER dictionary; (*d*) potential difference ELMAM – AMBER.



refined. The multipolar library was transferred to the whole protein structure. The average charge density for a selection of atoms with low thermal motion along the polypeptide main chain was then refined *versus* the protein diffraction data (Podjarny *et al.*, 2003; Lecomte *et al.*, 2004, 2005). The same atoms had their positions and thermal motion refined *versus* reflections at high order  $s > 0.6 \text{ \AA}^{-1}$ . The resulting deformation density is very close to that computed from the library transfer (Figs. 15*a, b*).

The  $RF$  and  $RF^{\text{free}}$  factors (Brünger, 1992, 1993) of the crystal structure after spherical-atom refinement were 9.28 and 9.45%, respectively. After library transfer and structure refinement, they decreased to 8.79 and 9.16%, respectively. The refinement of the polypeptide main-chain charge density leads to a further small drop of the  $RF$  factor down to 8.71%, while the  $RF^{\text{free}}$  factor decreased to 9.12%. This further  $RF$ -factor decrease is related to the addition of a few parameters in the refinement and possibly to a better modelling of the electron density. The very small  $RF^{\text{free}}$  factor drop after multipolar refinement indicates that the library electron density is improved very modestly. Most of the improvement in the modelling occurred during the ELMAM transfer step as judged by the drop of  $RF^{\text{free}}$  factor.

The library transfer in the case of this enzyme results in improved crystallographic statistics, although to a lesser extent in relative value than in the case of small molecules. A direct constrained charge-density refinement of the protein, starting from spherical atoms, results in statistical indices  $RF = 8.86$  and  $RF^{\text{free}} = 9.30\%$ . The deformation density is comparable to the two previous ones with the exception of the electron lone pairs of the O atom where it is strongly attenuated (Fig. 15*c*). The same result was observed for the charge-density analysis of the protein crambin and, in that study, it was decided not to refine the multipoles of the O atom. The O-atom lone pairs are features finer than the bonding electron density and are

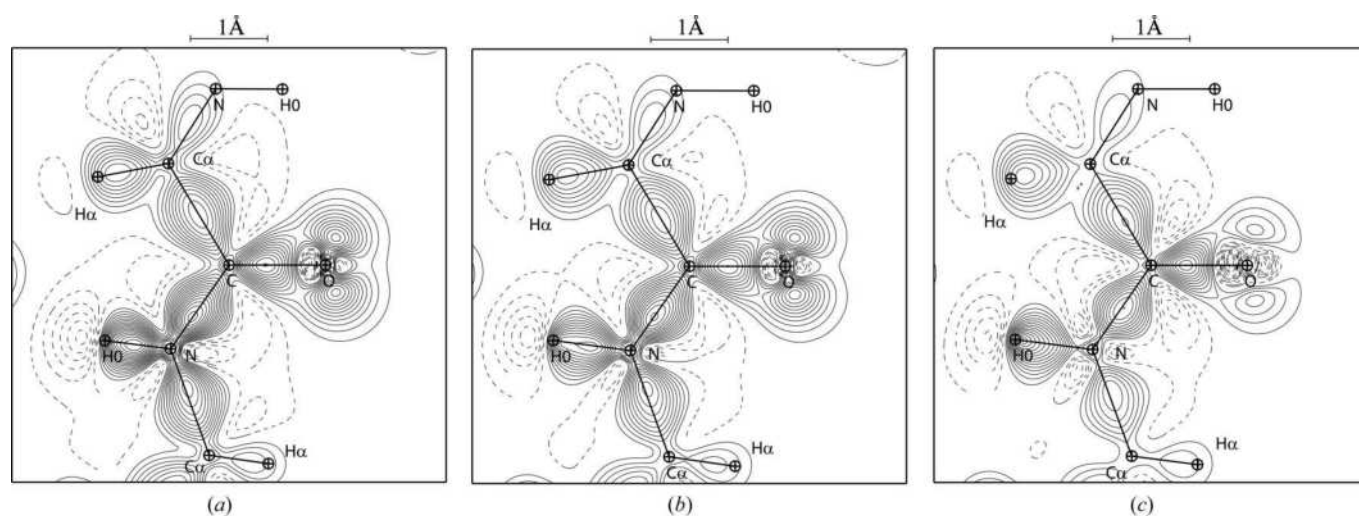
moreover in a region of high density generated by the core electrons. They require therefore a higher resolution in the diffraction data and lower thermal motion in the crystal structure in order to be well observed in an experimental charge density.

The deformation density visible when no initial library transfer is performed demonstrates that there is some significant bonding deformation density information present in the diffraction data of the protein. The statistical factors, notably the free  $RF$  factor, are less favourable for the direct charge-density refinement compared to the library transfer followed by a constrained charge-density refinement. This demonstrates that the initial transfer of the library charge density is the best starting point for a crystallographic refinement of such a protein at subatomic resolution.

#### 4. Conclusions and perspectives

By defining transferable pseudo-atoms, one can replace, without significant computing cost, the spherical independent-atom model (IAM) by the experimental library multipolar atom model (ELMAM). Transferable multipole populations are not refined and the total molecular electron density is obtained by summation of their contributions. The application of this method leads to improved molecular geometry (atomic positions), especially for H atoms, and thermal-motion description. Library transfer, owing to an improved atomic model, results in better figures of merit like a lower crystallographic  $R$  factor and in more accurate phases than the IAM refinement. As the ELMAM description of the atomic electron density allows for non-sphericity, the atomic thermal motion description is improved.

Extension of the library from proteins to nucleic acids (DNA, RNA) is underway; for instance, the charge-density analysis of nicotinamide adenine dinucleotide has already



**Figure 15**

Static deformation density in the peptide bond plane of human aldose reductase. The atoms (N–H)–(C $\alpha$ –H $\alpha$ )–(C=O)– chemically equivalent along the polypeptide main chain are constrained to have the same charge density. Contour interval  $\pm 0.05 e \text{ \AA}^{-3}$ ; positive: continuous line; negative: dotted line. (a) After library transfer, (b) after library transfer and constrained charge-density refinement of the protein main chain, (c) after constrained charge-density refinement without transfer.

been carried out (Guillot *et al.*, 2003). A generalization of the library to common chemical functions is also in progress as a limited number of atom types need to be defined to describe the first- and second-row elements in the Periodic Table.

An additional improvement of the current library would be a different definition of some local axes with the use of bisecting directions in order to decrease the number of multipoles describing an atom (Jelsch *et al.*, 2005). For example, with the current definition of local axes, only one of the two mirror planes of the water molecule HOH is taken into account.

The multipolar library presented in this paper is derived from experimental ultra-high-resolution crystallographic studies of peptides. Meanwhile, two other databases describing the electron density in peptide chemical groups have been developed from simulated data obtained from *ab initio* calculations of peptides *in vacuo* (Volkov, Koritsanszky *et al.*, 2004; Volkov, Li *et al.*, 2004; Dittrich, Hübschle *et al.*, 2006). The main advantage of the experimental database is the inclusion of chemical environment effects such as average hydrogen bonding.

Analyses of the precise electronic distribution have been performed up to now on small molecules, which generally satisfy the necessary conditions of subatomic resolution and low thermal motion. In suitable cases, electron-density analysis of macromolecules has now become feasible. The information contained in the amino acids and oligopeptides library (Pichon-Pesme *et al.*, 1995, 2004) can be transferred to protein atoms. The ELMAM refinement of a protein charge density at ultra-high resolution (0.5–0.7 Å) leads to a stronger deformation density and more accurate results compared to spherical neutral atoms (IAM) as the starting point. Structure files in several formats (PDB, CIF, *SHELXL*) containing the atomic coordinates and thermal displacement parameters can be read by the *MoPro* software suite. The information transferred is the local axes definitions and the atomic charge-density parameters ( $\kappa$ ,  $\kappa'$ ,  $P_{\text{val}}$ ,  $P_{\text{lm}\pm}$ ).

These improvements, in conjunction with the implementation of numerous restraints/constraints necessary in macromolecular crystallography, make the program *MoPro* a reliable tool for protein structure refinement, taking into account the precise charge density. The transferred molecular electron density can be used to evaluate a host of physical properties, notably the molecular electrostatic potential and interaction energies.

This work was partly funded by ANR grant NT05-3\_41509. Bartosz Zarychta is grateful for a scholarship from the Foundation for Polish Science (Domestic Grant for Young Scientists) and to the ANR. We would also like to thank Professor Robert H. Blessing for helpful discussions and Professor K. Ogawa for the diffraction data for dimethylstilbene.

## References

- Allen, F. (1986). *Acta Cryst.* **B42**, 515–522.
- Bader, R. F. W. (1990). *Atoms in Molecules: a Quantum Theory. The International Series: Monographs in Chemistry*. Oxford: Clarendon Press.
- Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. (1993). *J. Phys. Chem.* **97**, 10269–10280.
- Benabicha, F., Pichon-Pesme, V., Jelsch, C., Lecomte, C. & Khmou, A. (2000). *Acta Cryst.* **B56**, 155–165.
- Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B. *et al.* (2002). *Acta Cryst.* **D58**, 899–907.
- Brock, C. P., Dunitz, J. D. & Hirshfeld, F. L. (1991). *Acta Cryst.* **B47**, 789–797.
- Brünger, A. T. (1992). *Nature (London)*, **355**, 472–474.
- Brünger, A. T. (1993). *Acta Cryst.* **D49**, 24–36.
- 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.* **D54**, 905–921.
- 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. Phys.* **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.
- Coppens, P. (1998). *Acta Cryst.* **A54**, 779–788.
- Coppens, P., Guru Row, T. N., Lung, P., Stevens, E. D., Becker, P. & Yang, Y. N. (1979). *Acta Cryst.* **A35**, 63–79.
- Coppens, P. & Volkov, A. (2004). *Acta Cryst.* **A60**, 357–364.
- Cromer, D. T. (1974). *International Tables for X-ray Crystallography*, edited by J. A. Ibers & W. E. Hamilton, pp. 18–151. Birmingham: Kynoch Press.
- Dahaoui, S., Jelsch, C., Howard, J. A. K. & Lecomte, C. (1999). *Acta Cryst.* **B55**, 226–230.
- Dahaoui, S., Pichon-Pesme, V., Howard, J. A. K. & Lecomte, C. (1999). *J. Phys. Chem.* **A103**, 6240–6250.
- DeLano, W. L. (2002). *The PyMOL Molecular Graphics System*, <http://www.pymol.org>.
- Dittrich, B., Hübschle, C. B., Luger, P. & Spackman, M. A. (2006). *Acta Cryst.* **D62**, 1325–1335.
- Dittrich, B., Hübschle, C. B., Messerschmidt, M., Kalinowski, R., Girnt, D. & Luger, P. (2005). *Acta Cryst.* **A61**, 314–320.
- Dittrich, B., Koritsanszky, T. & Luger, P. (2004). *Angew. Chem. Int. Ed. Engl.* **43**, 2718–2721.
- Dittrich, B., Strumpel, M., Schäfer, M., Spackman, M. A. & Koritsanszky, T. (2006). *Acta Cryst.* **A62**, 217–223.
- Domiñak, P. M., Volkov, A., Li, X. & Coppens, P. (2007). *J. Chem. Theory Comput.* **3**, 232–247.
- Engh, A. R. & Huber, R. (1991). *Acta Cryst.* **A47**, 392–400.
- Espinosa, E., Lecomte, C., Molins, E., Veintemillas, S., Cousson, A. & Paulus, W. (1996). *Acta Cryst.* **B52**, 519–534.
- Farrugia, L. J. (1999). *J. Appl. Cryst.* **32**, 837–838.
- Fernandez-Serra, M. V., Junquera, J., Jelsch, C., Lecomte, C. & Artacho, E. (2000). *Solid State Commun.* **116**, 395–400.
- Görbitz, C. H. (2005). *Acta Cryst.* **E61**, o2012–o2014.
- Guillot, B., Muzet, N., Artacho, E., Lecomte, C. & Jelsch, C. (2003). *J. Phys. Chem.* **B107**, 9109–9121.
- Guillot, B., Viry, L., Guillot, R., Lecomte, C. & Jelsch, C. (2001). *J. Appl. Cryst.* **34**, 214–223.
- Guillot, R., Muzet, N., Dahaoui, S., Lecomte, C. & Jelsch, C. (2001). *Acta Cryst.* **B57**, 567–578.
- Hansen, N. K. & Coppens, P. (1978). *Acta Cryst.* **A34**, 909–921.
- Helle, I. H., Løkken, C. V., Görbitz, C. H. & Dalhus, B. (2004). *Acta Cryst.* **C60**, o771–o772.

- Hirshfeld, F. L. (1976). *Acta Cryst.* **A32**, 239–244.
- Housset, D., Pichon-Pesme, V., Jelsch, C., Benabicha, F., Maierhofer, A., David, S., Fontecilla-Camps, J. C. & Lecomte, C. (2000). *Acta Cryst.* **D56**, 151–160.
- Howard, E. R., 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.
- Hummel, W., Hauser, J. & Burgi, H. B. (1990). *J. Mol. Graph.* **8**, 214–20.
- Jelsch, C., Guillot, B., Lagoutte, A. & Lecomte, C. (2005). *J. Appl. Cryst.* **38**, 38–54.
- Jelsch, C., Pichon-Pesme, V., Lecomte, C. & Aubry, A. (1998). *Acta Cryst.* **D54**, 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.
- Kerr, K. A. & Ashmore, J. P. (1973). *Acta Cryst.* **B29**, 2124–2127.
- Koritsanszky, T. & Coppens, P. (2001). *Chem. Rev.* **101**, 1583–1628.
- Koritsanszky, T., Volkov, A. & Coppens, P. (2002). *Acta Cryst.* **A58**, 464–472.
- Lecomte, C. (1995). *Advances in Molecular Structure Research*, Vol. 1, edited by I. & M. Hargittai, pp. 261–302. Greenwich, CT, USA: JAI Press Inc.
- Lecomte, C., Guillot, B., Jelsch, C. & Podjarny, A. (2005). *Int. J. Quantum Chem.* **101**, 624–634.
- Lecomte, C., Guillot, B., Muzet, N., Pichon-Pesme, V. & Jelsch, C. (2004). *Cell Mol. Life Sci.* **61**, 774–782.
- Leherte, L., Vercauteren, D. P., Pichon-Pesme, V., Guillot, B., Jelsch, C., Lagoutte, A. & Lecomte, C. (2006). *Recent Advances in the Quantum Theory of Atoms in Molecules*, pp. 285–316. Weinheim: Wiley-VCH.
- Li, X., Volkov, A. V., Szalewicz, K. & Coppens, P. (2006). *Acta Cryst.* **D62**, 639–647.
- Longhi, S., Czjzek, M. & Cambillau, C. (1998). *Curr. Opin. Struct. Biol.* **8**, 730–737.
- Mata, I., Espinosa, E., Molins, E., Veintemillas, S., Maniukiewicz, W., Lecomte, C., Cousson, A. & Paulus, W. (2006). *Acta Cryst.* **A62**, 365–378.
- Minasov, G., Wang, X. & Shoichet, B. K. (2002). *J. Am. Chem. Soc.* **124**, 5333–5340.
- Moen, A., Frøseth, M., Görbitz, C. H. & Dalhus, B. (2004). *Acta Cryst.* **C60**, o564–o565.
- Muzet, N., Guillot, B., Jelsch, C. & Lecomte, C. (2003). *Proc. Natl Acad. Sci. (USA)*, **15**, 8742–8747.
- Netland, K. A., Andresen, K., Görbitz, C. H. & Dalhus, B. (2004). *Acta Cryst.* **E60**, o951–o953.
- Ogawa, K., Noda, Y., Lüthi, T. & Büergi, H.-B. (2006). Private communication.
- Petrova, T. & Podjarny, A. (2004). *Rep. Prog. Phys.* **67**, 1565–1605.
- Pichon-Pesme, V., Jelsch, C., Guillot, B. & Lecomte, C. (2004). *Acta Cryst.* **A60**, 204–208.
- Pichon-Pesme, V., Lachekar, H., Souhassou, M. & Lecomte, C. (2000). *Acta Cryst.* **B56**, 728–737.
- Pichon-Pesme, V. & Lecomte, C. (1998). *Acta Cryst.* **B54**, 485–493.
- 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.
- Podjarny, A., Howard, E., Mitschler, A., Chevrier, B., Lecomte, C., Guillot, B., Pichon-Pesme, V. & Jelsch, C. (2003). *Europhys. News*, **33**, 113–117.
- Quail, J. W., Ismail, N., Pedras, M. S. C. & Boyetchko, S. M. (2002). *Acta Cryst.* **C58**, o268–o271.
- Rödel, E., Messerschmidt, M., Dittrich, B. & Luger, P. (2006). *Org. Biomol. Chem.* **4**, 475–481.
- Ruiz, F., Hazemann, I., Mitschler, A., Joachimiak, A., Schneider, T., Karplus, M. & Podjarny, A. (2004). *Acta Cryst.* **D60**, 1347–1354.
- Schmidt, A., Jelsch, C., Ostergaard, P., Rypniewski, W. & Lamzin, V. S. (2003). *J. Biol. Chem.* **278**, 43357–43362.
- Schmidt, A. & Lamzin, V. S. (2002). *Curr. Opinion Struct. Biol.* **12**, 698–703.
- Schomaker, V. & Trueblood, K. N. (1968). *Acta Cryst.* **B24**, 63–76.
- Schomaker, V. & Trueblood, K. (1998). *Acta Cryst.* **B54**, 507–514.
- Sheldrick, G. M. & Schneider, T. (1997). *SHELXL: High-Resolution Refinement. 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.
- Souhassou, M., Lecomte, C., Blessing, R. H., Aubry, A., Rohmer, M. M., Wiest, R., Bénard, M. & Marraud, M. (1991). *Acta Cryst.* **B47**, 253–266.
- Souhassou, M., Lecomte, C., Ghermani, N.-E., Rohmer, M.-M., Wiest, R., Bénard, M. & Blessing, R. H. (1992). *J. Am. Chem. Soc.* **114**, 2371–2382.
- Spackman, M. A., Byrom, B. G., Alfredsson, M. & Hermansson, K. (1999). *Acta Cryst.* **A55**, 30–47.
- Stewart, R. F., Davidson, E. R. & Simpson, W. T. (1965). *J. Chem. Phys.* **42**, 3175–3187.
- Volkov, A., Abramov, Y. A. & Coppens, P. (2001). *Acta Cryst.* **A57**, 272–282.
- Volkov, A., Gatti, C., Abramov, Y. & Coppens, P. (2000). *Acta Cryst.* **A56**, 252–258.
- Volkov, A., Koritsanszky, T., Li, X. & Coppens, P. (2004). *Acta Cryst.* **A60**, 638–639.
- Volkov, A., Li, X., Koritsanszky, T. & Coppens, P. (2004). *J. Phys. Chem.* **A108**, 4283–4300.
- Wiest, R., Pichon-Pesme, V., Bénard, M. & Lecomte, C. (1994). *J. Phys. Chem.* **98**, 1351–1362.