Charge-density analysis of a protein structure at subatomic resolution: the human aldose reductase case

Benoît Guillot,[a] Christian Jelsch,[a]* Alberto Podjarny[b] and Claude Lecomte[a]

[a]Laboratoire de Cristallographie et Modélisation des Matériaux Minéraux et Biologiques, LCM3B, CNRS, UMR 7036, Nancy Université, Faculté des Sciences et Techniques, BP 239, 54506 Vandoeuvre-lès-Nancy CEDEX, France, and [b]Département de Biologie et Génomique Structurales, IGBMC, CNRS INSERM ULP, Illkirch, France

Correspondence e-mail:
christian.jelsch@uhp-nancy.fr

The valence electron density of the protein human aldose reductase was analyzed at 0.66 Å resolution. The methodological developments in the software *MoPro* to adapt standard charge-density techniques from small molecules to macromolecular structures are described. The deformation electron density visible in initial residual Fourier difference maps was significantly enhanced after high-order refinement. The protein structure was refined after transfer of the experimental library multipolar atom model (ELMAM). The effects on the crystallographic statistics, on the atomic thermal displacement parameters and on the structure stereochemistry are analyzed. Constrained refinements of the transferred valence populations P_{val} and multipoles P_{lm} were performed against the X-ray diffraction data on a selected substructure of the protein with low thermal motion. The resulting charge densities are of good quality, especially for chemical groups with many copies present in the polypeptide chain. To check the effect of the starting point on the result of the constrained multipolar refinement, the same charge-density refinement strategy was applied but using an initial neutral spherical atom model, *i.e.* without transfer from the ELMAM library. The best starting point for a protein multipolar refinement is the structure with the electron density transferred from the database. This can be assessed by the crystallographic statistical indices, including R_{free}, and the quality of the static deformation electron-density maps, notably on the oxygen electron lone pairs. The analysis of the main-chain bond lengths suggests that stereochemical dictionaries would benefit from a revision based on recently determined unrestrained atomic resolution protein structures.

1. Introduction

Biocrystallography has evolved significantly in recent years, with the introduction of new methods and techniques in each stage of protein structure resolution. Data-collection techniques and data quality have been significantly improved thanks to third-generation synchrotron sources, automated data reduction and methodological improvements (Petrova & Podjarny, 2004; Schmidt & Lamzin, 2002; Dauter *et al.*, 1997; Dauter, 2003). Moreover, protein model building and refinement can now in many cases be performed almost automatically, allowing the scientist to spend more time on structure analysis and interpretation. Some of these improvements arise from so-called high-throughput crystallography, which is dedicated to almost automated and very rapid protein structure determination in the context of genomic projects. All these improvements, but in particular

those in crystallogenesis and synchrotron-radiation sources, have another extremely important consequence: more and more protein or nucleic acid structures are being solved and refined at atomic or subatomic resolution. The record resolutions for proteins are currently crambin at 0.54 Å resolution (Jelsch *et al.*, 2000), an antifreeze protein at 0.62 Å (Ko *et al.*, 2003), triclinic lysozyme at 0.65 Å (Wang *et al.*, 2007), human aldose reductase (hAR) at 0.66 Å (Podjarny *et al.*, 2003; Howard *et al.*, 2004), a rubredoxin at 0.69 Å (Bönisch *et al.*, 2005), a domain of syntenin at 0.73 Å (Kang *et al.*, 2004) and a hydrophobin at 0.75 Å (Hakanpää *et al.*, 2006). Protein structures at resolutions between 1 and 0.5 Å allow detailed analysis of the protein structure on the atomic scale. Most of the H atoms then become visible in difference Fourier $F_{obs} - F_{calc}$ maps. This is of extreme importance for understanding enzymatic catalysis and for protein-structure interpretation. Many recent reviews have indeed described the importance of such resolution in protein-structure analysis (Dauter *et al.*, 1997; Schmidt & Lamzin, 2002; Housset *et al.*, 2000). Moreover, subatomic resolution (∼0.7 Å or higher) has another advantage: the level of detail that can be observed in experimental electron density goes far beyond the localization of atoms and the identification of their chemical species. Indeed, at subatomic resolution it becomes possible to observe the nonspherical character of atomic electron densities or, in other words, their deformation arising fom chemical interactions such as covalent bonding, electrostatic interactions and hydrogen bonds. Indeed, protein structures are generally refined using the standard independent-atom model (IAM), which describes atoms as non-interacting spherically symmetrical species. Such refinement at subatomic resolution leads to residual nonmodelled deformation electron density on covalent bonds and on electron lone pairs (Coppens, 1967; Pichon-Pesme *et al.*, 1995; Jelsch *et al.*, 1998; Afonine *et al.*, 2004).

These significant electron-density peaks correspond to the deviation from the spherical atom approximation and obviously cannot be taken into account by the IAM model. Consequently, structure refinement at subatomic resolution using spherical atoms leads to an imperfect molecular model. The presence of bonding electron density in residual Fourier maps is linked to severe bias on structural parameters, especially on atomic thermal motion and some covalent-bond lengths (Coppens, 1967; Brock *et al.*, 1991; Pichon-Pesme *et al.*, 1995; Jelsch *et al.*, 1998; Afonine *et al.*, 2004; Dittrich *et al.*, 2006; Zarychta *et al.*, 2007; Volkov *et al.*, 2007). These features are well known in small-molecule crystallography, in which structures usually display lower atomic thermal motion and for which subatomic resolution X-ray data can generally be accurately collected. Several nonspherical atom models have therefore been developed, notably the multipole formalism (Stewart, 1969; Hansen & Coppens, 1978), which is used in this study.

In the early 1990s, Brock *et al.* (1991) successfully transferred the deformation electron-density parameters obtained from a charge-density study of a perylene molecule to related naphthalene and anthracene compounds in order to significantly improve the description of their thermal motion (Brock

et al., 1991). This first successful application of the transferability principle is based on the assumption that the deformation electron-density features are similar for a given atom type in different molecules or crystal contexts. This principle was subsequently extended to the possibility of transferring average deformation electron-density parameters obtained from subatomic resolution peptide and amino-acid charge-density studies to larger compounds or protein structures solved at lower resolution (Pichon-Pesme *et al.*, 1995; Jelsch *et al.*, 1998; Volkov *et al.*, 2007). These results and the determination of several precise charge densities of peptides and amino acids prompted the creation of the experimental library multipolar atom model (ELMAM; Pichon-Pesme *et al.*, 1995, 2004; Zarychta *et al.*, 2007). The library contains all the multipolar parameters, in the Hansen & Coppens (1978) formalism, needed to describe the electron density of each chemical moiety that can be found in protein structures. At the same time, in order to apply the multipole formalism to protein structures, the *MoPro* software suite was developed by our group (Guillot *et al.*, 2001; Jelsch *et al.*, 2005), implementing the ELMAM database. This program suite includes the techniques developed by the small-molecule charge-density community, as well as the refinement methods such as stereochemical restraints and conjugate-gradient iterations used in protein crystallography.

Several successful applications based on the transfer of the ELMAM library parameters and the use of the *MoPro* software have been reported to date: studies of the LBZ octapeptide (Jelsch *et al.*, 1998), of the scorpion *Androctonus australis* Hector protein toxin II (Housset *et al.*, 2000), of trypsin (Schmidt *et al.*, 2003) and of a snake-venom phospholipase (Liu *et al.*, 2003). The electrostatic properties of the human aldose reductase active site complexed with $NADP^+$ and inhibitor have also been analyzed (Muzet *et al.*, 2003; Lecomte *et al.*, 2004, 2005; Leherte *et al.*, 2006).

Following our investigations, Afonine *et al.* (2004, 2007) have discussed the modelling of deformation electron density in high-resolution protein structures by using interatomic additional scatterers. The electron density in the crystal structure of cholesterol oxidase at atomic resolution has also been analyzed (Lario & Vrielink, 2003 and it was found that peptide carbonyl groups are more polarized in α-helices than in β-sheets.

Two other research groups have subsequently developed alternative multipolar electron-density databases (Coppens & Volkov, 2004; Volkov, Koritsanszky *et al.*, 2004; Volkov, Li *et al.*, 2004; Li *et al.*, 2006; Dominiak *et al.*, 2007; Dittrich *et al.*, 2005, 2006), both of which are based on theoretical electron densities obtained from first-principles computations.

Successful applications of these alternative libraries to the DL-serine molecule (Dittrich *et al.*, 2006), a Pro-Ala octapeptide (Dittrich *et al.*, 2002; Volkov *et al.*, 2007) and a Tyr-Gly-Gly tripeptide (Volkov *et al.*, 2007) confirmed our work on peptide compounds. Both the experimental and theoretical electron-density parameter libraries clearly show that the application of transferred aspherical scattering factors leads to better refinement statistics, improved anisotropic displace-

ment parameter (ADP) values and, to a lesser extent, better molecular geometries. However, no charge-density refinement has yet been reported on an average-sized protein structure of pharmaceutical interest. This is the aim of the current paper, in which we will describe the application of the transferability principle to an enzyme structure: the 0.66 Å resolution human aldose reductase protein complex (Lamour *et al.*, 1999; Cachau *et al.*, 2000; Howard *et al.*, 2004; Muzet *et al.*, 2003).

Human aldose reductase (hAR; EC 1.1.1.21; Fig. 1) is a 316-amino-acid NADPH-dependent enzyme of molecular weight 36 kDa that is able to reduce a large variety of substrates such as aldehydes or ketones. In particular, AR reduces excess glucose into D-sorbitol as part of a metabolic scheme called the polyol pathway. As the accumulation of sorbitol in specific cells is known to be at the origin of degenerative diseases in diabetes (Yabe-Nishimura, 1998), aldose reductase is considered to be a good pharmacological target to limit sorbitol accumulation and thus to prevent the complications associated with diabetes (Oka & Kato, 2001). Several inhibitor molecules are now in the clinical trials phase (Giannoukakis, 2003, 2006; Hotta *et al.*, 2006; Nakahara *et al.*, 2005). However, the inhibition of hAR as a therapeutic direction may suffer from side effects owing to a lack of selectivity of the drugs against the inhibition of aldehyde reductase, which is involved in aldehyde detoxification in the kidneys (El-Kabbani *et al.*, 2005). To ensure good specificity of inhibitor drugs towards aldose reductase, a good understanding of the enzymatic mechanism is of key importance. There is some controversy in the mechanism of aldose reductase with regard to the identification of the amino acid which acts as the proton donor during the reduction process catalyzed by the enzyme. Directed mutagenesis and modelling studies (Bohren *et al.*, 1994; Tarle *et al.*, 1993; Singh *et al.*, 2000)

have shown that two residues located in the active site are candidates for this proton donation: a histidine residue (His110) and a tyrosine residue (Tyr48). Recent neutron diffraction, high-resolution X-ray diffraction and quantum modelling studies indicate that Tyr48 is the most likely proton donor (Blakeley *et al.*, 2008).

2. Materials and methods

2.1. Multipolar atom model

Several nonspherical atom models have been proposed that are more elaborate than the IAM approximation. Most of them are based on nonspherical scattering centres modelled by a superposition of real spherical harmonic functions (Yáśubnez *et al.*, 1978). The Hansen and Coppens atom model used in this study (Hansen & Coppens, 1978) is based on this multipolar formalism and is now the most widely used in the field of small-molecule charge-density studies:

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

This multipolar model takes into account the atomic electron density by a sum of three terms, the first of which is a core spherical electron density (ρ_{core}) which is not allowed to vary as core electrons are not involved in interatomic interactions. The second term is a valence spherical electron density (ρ_{val}), which is modulated by its valence population P_{val} and by an expansion–contraction parameter κ. Both P_{val} and κ are atomic variable parameters that can be refined in the least-squares procedure. The valence population gives access to atomic charges (*i.e.* the deviation from neutrality) and the
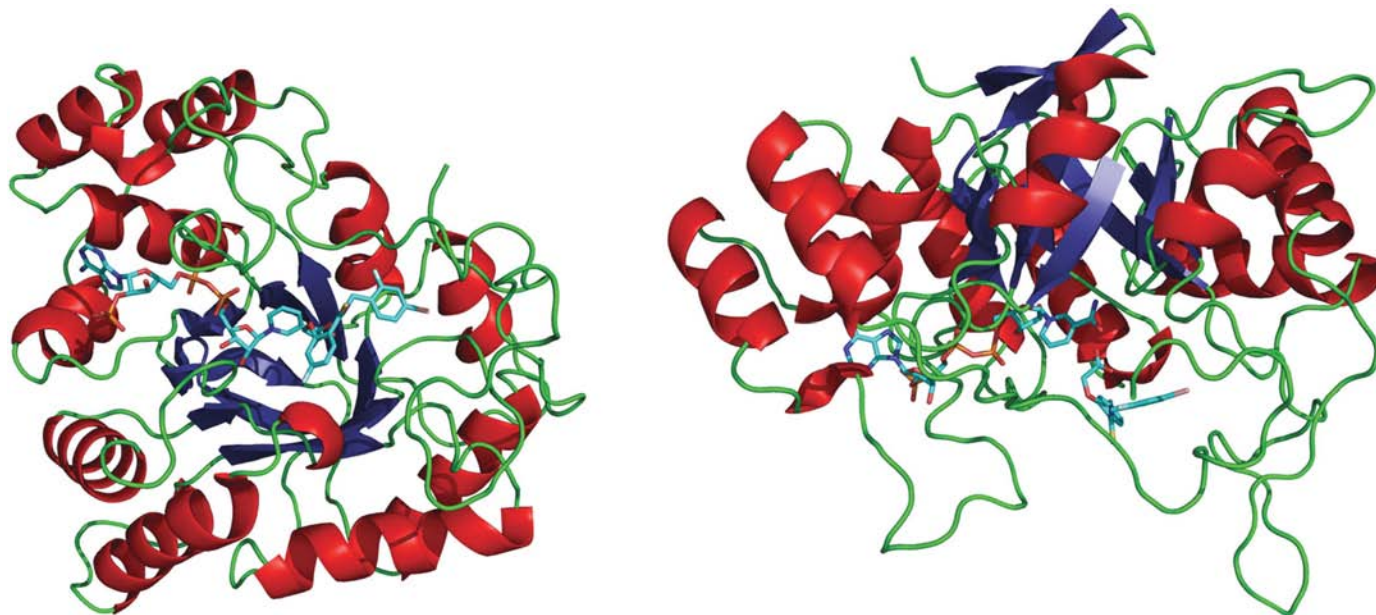

Figure 1
Two orthogonal ribbon views of the hAR structure, with the NADP$^+$ cofactor and the IDD594 inhibitor bound at the bottom of the parallel β-sheet barrel.

Table 1
ζ and n_l coefficients of the radial Slater function (2) used for definition of the multipoles, from dipoles ($l = 2$) to hexadecapoles ($l = 5$).

Atom	ζ (bohr^{-1})	n_2	n_3	n_4	n_5
H	2.26	2			
O	4.50	2	2	3	
N	3.80	2	2	3	
C	3.00	2	2	3	
S	3.85	4	4	6	8

parameter κ allows the spherical valence electron density to become more contracted or more expanded around the atomic position during refinement. The core and valence electron densities are usually obtained from spherically averaged free-atom Hartree–Fock densities (Clementi & Roetti, 1974). The third term in (1) is able to model the deviation from spherical shape of the valence electron density. It is written as a superimposition of real spherical harmonic functions $y_{lm\pm}(\theta, \varphi)$, oriented with respect to a local axes system. This local coordinate system (X, Y, Z) is centred on the atom under consideration and is oriented with respect to the atomic neighbours. For example, in the present study a main-chain carbonyl C atom has its X axis oriented toward the connecting O atom of the C=O moiety. The XY plane is defined by the atoms C, O and N (of the next amino acid along the sequence), the Z axis is then perpendicular to that plane and the Y direction is finally defined by the direct cross product of the Z and X basis vectors (Zarychta et al., 2007). Each of these spherical harmonic functions is parameterized by multipolar populations P_{lm}, which can be refined in order to fit the deformation electron density. The sum over l indices of the real spherical harmonic terms is the level of the multipolar expansion, which determines the number of refined multipolar parameters P_{lm} needed to adequately model the deformation electron density of any chemical species. For instance, in the case of protein H atoms a dipolar expansion ($l = 1$) limited to one dipole oriented along the bond is sufficient, in our estimation, to model the covalent bond directed toward its bonded heavy atom. More generally, one attributes the dipolar level to H atoms, the octupolar level ($l = 3$) to O, C and N atoms and up to the hexadecapolar level ($l = 4$) to S, P and heavier atom types.

Furthermore, the spherical harmonic functions $y_{lm\pm}(\theta, \varphi)$ are multiplied by a radial function of Slater type R_{nl} (Slater, 1932),

$$R_{nl}(\kappa'r) = \zeta^{n_l+3}(\kappa'r)^{n_l}\exp(-\zeta\kappa'r)/(n_l + 2)!, \qquad (2)$$

so that the shape of this radial function modulates the expansion/contraction of the multipoles which model the nonspherical part of the valence electron density. The n_l and ζ values in (2) are fixed parameters with values depending on the atomic species. The values that have been used in this study are reported in Table 1 and are taken from free-atom HF-SCF computations (Clementi & Raimondi, 1963). While the n_l and ζ values are kept fixed, the κ' parameter modulates the radial dependance of the Slater function, allowing adjustment in the least-squares refinement of the contraction/

expansion of the multipolar electron density. To summarize, the P_{val}, P_{lm}, κ and κ' atomic variable of the Hansen and Coppens model can be refined in the least-squares procedure against X-ray data to fit the deformation electron density.

2.2. Electron-density maps

The charge distribution of a molecule is efficiently displayed by the deformation of the electron density $\Delta\rho$. Static deformation maps, $\Delta\rho_{\mathrm{stat}}$, for which the atomic thermal motion is removed, are directly computed from the crystallographic modelling as the atomic superposition sum over the molecule,

$$\Delta\rho_{\mathrm{stat}} = [P_{\mathrm{val}}\kappa^3\rho_{\mathrm{val}}(\kappa r) - N_{\mathrm{val}}\rho_{\mathrm{val}}(r)]$$
$$+ \sum_{l=0}^{l_{\max}} \kappa'^3 R_{nl}(\kappa'r) \sum_{m=0}^{l} P_{lm\pm}y_{lm\pm}(\theta, \varphi). \qquad (3)$$

The valence electron density of spherical neutral (IAM) atoms is represented by the term $N_{\mathrm{val}}\rho_{\mathrm{val}}(r)$. The residual electron-density maps $F_{\mathrm{obs}} - F_{\mathrm{calc}}$ are computed by Fourier synthesis at the various stages of the crystallographic refinement, using all experimental data satisfying $I/\sigma(I) > 0$. The structure-factor amplitudes used for Fourier synthesis include all atoms and all scale factors (overall scale and bulk-solvent modelling) incorporated in the model. Phases used for residual electron-density map computations are obtained using the model corresponding to the refinement stage examined.

2.3. Diffraction data

The X-ray structure of the hAR–IDD-594–NADP$^+$ monomeric complex was recently refined using the independent-atom model (IAM) to the exceptional resolution of 0.66 Å (Lamour et al., 1999; Howard et al., 2004). As the 0.66 Å resolution X-ray data collection and processing have already been described in detail elsewhere (Howard et al., 2004), we only recall the key points of this experiment here. The crystals belong to the monoclinic space group $P2_1$, with unit-cell parameters $a = 49.43$, $b = 66.79$, $c = 47.40$ Å, $\beta = 92.4°$. The X-ray diffraction data were collected from two single crystals of approximate dimensions $0.3 \times 0.4 \times 0.6$ mm on the 19ID beamline of the SBC-CAT at the APS synchrotron. For both crystals, low-, medium- and high-resolution diffraction data were collected separately at 100 K with optimized exposure times. The diffraction intensities were then scaled and merged together to produce the final data set of 511 265 unique Bragg reflections. The average signal-to-noise ratio reaches 14.5 and is 2.4 in the highest resolution shell. The overall completeness is 90%; despite the fact that these data were not originally collected for the purpose of a charge-density study, they nevertheless show an overall redundancy of 2.3.

2.4. Molecular model

The present study is based on the hAR atomic model that has been provided to us by the authors of the subatomic resolution hAR study (Howard et al., 2004). This structure was obtained from standard IAM stereochemically restrained refinement performed against the 0.66 Å resolution data using

Table 2
The fragments of the hAR structure selected for charge-density refinement.

Gly16–Gly25	Val27–Thr28	Ala30–Val31
Val33–Tyr39	His41–Asp43	Ala45–Ile58
Leu72–Tyr82	Thr95–Leu96	Thr103
Asp105–Trp111	Thr113–Phe115	Phe121–Phe122
Gly128–Pro132	Thr140–Trp141	Ala143–Glu145
Leu147–Asp149	Ala155–Phe161	Gln165–Val166
Gly174–Tyr177	Val181–Glu185	Val206–Ser214
Gln246–Ile249	Phe251–Arg255	Leu257–Ser263
Thr265	Leu280–Ser281	Gln283–Ser290
Leu300	Ser302–Cys303	Asp308–Pro310

the *SHELXL*97 program (Sheldrick, 2008). This model reaches statistical factors of $R(F)$ = 9.38% and $R_{free}(F)$ = 9.71% using all available reflections. This is the lowest statistical agreement factor obtained for a protein of such size. The final model is constituted of 316 amino acids, one third of which display double or even triple conformations. Half of the chemically expected H atoms were located in difference Fourier maps. The final model includes 500 water molecules, the NADP$^+$ cofactor, the inhibitor and two citrate ions bound at the protein surface. The definition of the atomic local axes system in the multipolar formalism requires the H atoms. When they were not visible in the electron-density maps, the H atoms were generated according to standard geometry with the program *SHELXL* (Sheldrick, 2008). A rotational degree of freedom remains for H atoms in hydroxyl OH and NH$_3$ or CH$_3$ groups. The rotation of the rigid group was oriented manually when necessary on the basis of putative hydrogen bonds to neighbouring acceptor atoms. The positions and thermal parameters of H atoms were not refined but
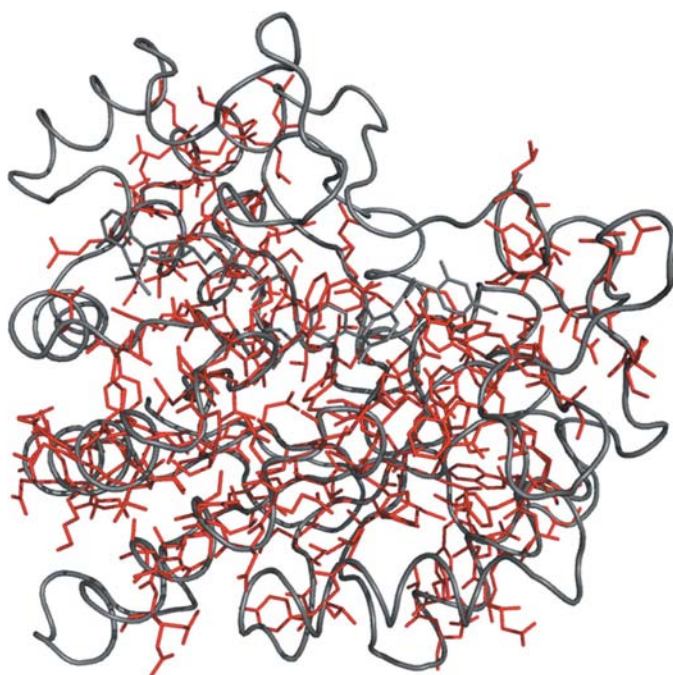
Figure 2
Schematic view of the hAR structure with subsystem atoms selected for refinement explicitly represented. Atoms included in the selection are represented in red.

constrained. The H atoms were assigned isotropic thermal parameters riding on the equivalent thermal parameter of the bound atom: $U_{iso}(H) = 1.2U_{eq}(X)$; for the OH, NH$_3$ or CH$_3$ groups, which have a rotational degree of freedom, the constraint was set to $U_{iso}(H) = 1.5U_{eq}(X)$ (Sheldrick, 2008). In this initial model, the exponential scaling model (Moews & Kretsinger, 1975) of disordered solvent correction was applied during the *SHELXL*97 refinement. In order to improve the quality of the low-resolution data fitting, it was replaced in *MoPro* by the bulk-solvent model (Jiang & Brünger, 1994; Fokine & Urzhumtsev, 2002; Afonine et al., 2005) based on the estimation of the structure factors of the disordered solvent part in the unit cell.

2.5. Subsystem selection

In order to perform the stereochemically unrestrained refinement of structural and multipolar parameters, a subset of the protein structure was selected. Selection of atoms for the hAR molecular subsystem was based on the magnitude of the atomic B_{eq} factors in the initial molecular model (Howard *et al.*, 2004). A high-order refinement, without geometry restraints, is only possible for atoms with low thermal motion. Also, the deformation electron density showed up in initial residual maps in regions of low thermal motion; this is well established in small-molecule crystallography.

As one of our aims is to perform a stereochemically unrestrained structural refinement (atom coordinates and thermal displacement parameters), neither disordered parts of the structure nor residues containing highly agitated atoms were selected. The selected subset was initially based on amino acids with atoms showing an initial thermal factor $B_{eq} < 8$ Å^2. It was then defined as 1156 of the 3042 non-H atoms of the hAR structure. The subset is detailed in Table 2 and represented in Fig. 2. In the part of the sequence described, the O$^{\delta 1}$ and O$^{\delta 2}$ atoms of the carboxylic acid moiety of Asp308 and the main-chain O atoms of Asp149, Gly213 and Gln254, which had B_{eq} factors higher than 8 Å^2, were excluded. The selected subsystem also contained the active-site water molecules numbered 2001 and 2044, for which the O atoms has low thermal motion (3.2 and 3.5 Å^2, respectively) and both H atoms were located in electron-density maps.

The non-H atoms in this substructure had an average thermal factor $\langle B_{eq}\rangle$ of 4.23 Å^2, with a maximum value of 7.98 Å^2 for the main-chain O atom of Pro24. The average ADP anisotropy in the subsystem was 0.55 (defined as the ratio of the smallest to the largest eigenvalue of the U_{ij} tensor); the smallest corresponded to the main-chain O atom of Gln49 (which also had a relatively large B_{eq} factor of 7.68 Å^2). Only seven atoms in the subsystem displayed a large anisotropy (less than 0.2); these all had an atomic B_{eq} factors higher than 7 Å^2.

The composition of this substructure was maintained unchanged during the whole study presented here. The remaining parts of the structure, which were disordered or displayed higher thermal motion, were maintained fixed in the state corresponding to the initial model.

2.5.1. Procedure overview. Here, for the sake of clarity, we present a short summary of the overall refinement strategy applied in this study. The procedure was divided into several major steps, as depicted in Fig. 3, that were applied to the substructure defined from the published model (Howard *et al.*, 2004).

(i) Bulk-solvent correction.

(ii) IAM refinement.

(iii) IAM high-order (HO) refinement.

(iv) H-atom position readjustment according to standard neutron H—X distances.

(v) Transfer of the ELMAM library parameters on the protein atoms.

(vi) Positional and thermal parameter refinement.

(vii) Constrained multipolar refinement (ELMAM refinement).

Alternatively, the two last steps were applied directly to the model obtained after the IAM HO refinement and H-atom position modification [without the transfer procedure in step (v)]. The effects of these successive steps on the statistical crystallographic indices $R(F)$, $R_{free}(F)$ and $wR^2(F)$ are given in Table 3. The stages of the procedure will be described in detail in the following sections.

2.6. IAM refinement

In this section, we present the IAM refinement stages used in the study. A $I/\sigma(I) > 0$ cutoff was applied to the reflections for all refinements. Crystallographic refinement was performed by minimizing the sum $\sum w_H^2 (F_H^{cal} - F_H^{obs})^2$ over the reflections using a standard weighting scheme $w_H^2 = 1/\sigma^2(F_H^{obs})$. Only the non-H atoms of the substructure were refined, while

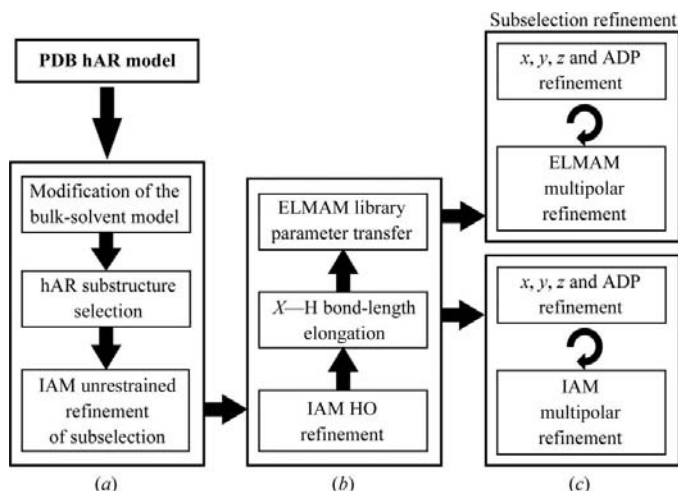**Table 3**
Statistical crystallographic indices after the various steps of the refinement, computed using all data included in the refinement [$I/\sigma(I) > 0$].

Where appropriate, the maximum number of refined parameters included in each refinement step is indicated. The resolution ranges for refinement steps are indicated in the text. $R(F) = \sum(|F_{obs}| - |F_{calc}|)/\sum|F_{obs}|$. $wR_2(F) = \{\sum(|F_{obs}| - |F_{calc}|)^2/\sigma^2_{F_{obs}}]/\sum(F^2_{obs}/\sigma^2_{F_{obs}})\}^{1/2}$.

Steps	Maximum No. of parameters	$R(F)$ (%)	$R_{free}(F)$ (%)	$wR^2(F)$ (%)
Initial IAM		9.70	9.97	14.59
Solvent correction	3	9.63	9.86	14.45
IAM, x, y, z, U_{ij} refinement	10404	9.23	9.78	12.93
IAM HO refinement	10404	9.36	9.71	13.41
X—H elongation		9.54	9.90	13.56
ELMAM database transfer	1	9.11	9.43	13.51
x, y, z, U_{ij} refinement	10405	8.80	9.26	12.74
ELMAM multipolar refinement	10709	8.74	9.23	12.49
IAM multipolar refinement	10709	8.97	9.49	12.68

H-atom geometry was constrained. Neither thermal restraints (limited anisotropy and rigid bonds) nor stereochemical restraints (distances and angles) were applied during substructure refinement. The IAM refinements preceding the introduction of aspherical scattering factors into the procedure were as follows.

Firstly, after introduction of the bulk-solvent correction, a subsystem was selected from the hAR structure based on atomic thermal motion and an IAM stereochemically unrestrained refinement was performed on the substructure atoms.

Secondly, the fractional coordinates and ADP values of the substructure atoms were refined against the high-resolution diffraction data only. Such high-order (HO) refinement allows improvement of the deformation electron density observable in residual maps by achieving partial deconvolution from the atomic thermal motion. Indeed, the contribution of the diffuse valence electron density in reciprocal Fourier space is very low at ultrahigh resolution, where only core electrons contribute significantly (Fig. 4). Early direct least-squares refinement
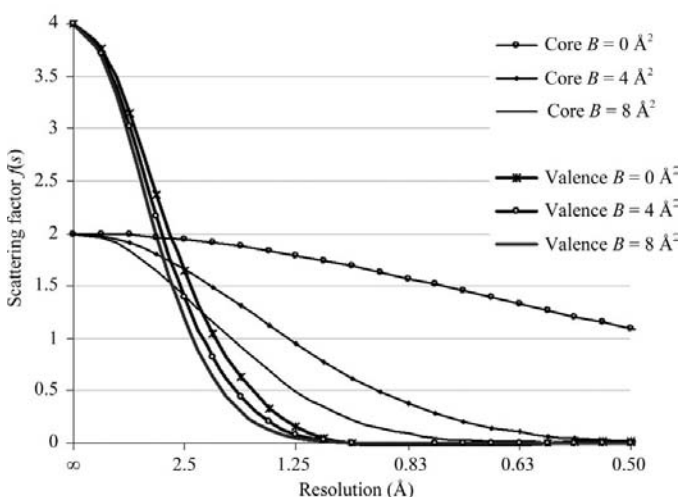

Figure 3
Flow chart describing the three main stages of the hAR refinement. (*a*) The selection of substructure from the PDB model and bulk-solvent modelling followed by IAM unrestrained refinement. (*b*) HO refinement, X—H bond-length modification and the transfer of multipolar parameters from the ELMAM library. (*c*) Structural and constrained multipolar refinement, both performed in a cyclic way, till convergence. One multipolar refinement is performed after ELMAM transfer and the other is performed starting from the IAM structure directly after adjustment of the H-atom positions.

Figure 4
Scattering factors of the core and valence electrons as a function of resolution for a C atom with isotropic thermal parameters B_{eq} of 0, 4 and 8 Å^2.

attempts in the narrow resolution shell $0.66 < d < 0.77$ Å led to very high and abrupt variations in stereochemical parameters, with some ADPs changing to nonpositive-definite values. This may be a consequence of lack of completeness of the diffraction data in this high-resolution range. To avoid this and to allow smaller initial parameter shifts, the high-order refinement was performed using three successive high-resolution ranges of decreasing size: the 1156 non-H atoms of the selected subset were successively refined in the 0.66–1.0, 0.66–0.83 and 0.66–0.77 Å resolution ranges. The refinement strategy used for each of the resolution shells consisted of alternate refinements of the scale factors, of the atomic fractional coordinates and of the ADPs. For each resolution range, after a few cycles and on nearing convergence all the structural parameters were refined together. This refinement sequence was performed until total convergence successively in the three resolution ranges.

This refinement was followed by displacement of H atoms along their covalent bonds to make the bond lengths fit to standard distances as observed in neutron diffraction studies (Allen, 1986).

2.7. Database transfer

The charge-density variables (κ, κ', P_{val} and P_{lm}) of the ELMAM library were transferred to all the protein atoms. For disordered parts of the structure, only the P_{val} valence populations as defined in the data bank were attributed. The transfer was performed automatically based on the atom names in the protein nomenclature. All histidine residues were described as neutral in the protein, with the $N^{\delta 1}$ atoms being protonated. The H atoms of water molecules 2001 and 2044, which were visible in the electron density, were added to the model. The database was therefore also transferred to these ordered waters. The deformation electron density of the $NADP^+$ cofactor was derived from crystallographic analysis of the NAD^+ molecule at subatomic resolution (Guillot et al., 2003). The inhibitor and the citrate ions of the structure were excluded from the transfer. Some of the inhibitor atoms incidentally had no corresponding atom type present in the ELMAM library for proteins. An overall electroneutrality constraint was applied so that the total number of valence electrons of the substructure remained constant during the P_{val} refinement.

2.8. Multipolar atom refinement

In the ELMAM multipolar refinement, the starting hAR molecular model used was the result of IAM HO refinement followed by database transfer. In the IAM multipolar refinement, another charge-density refinement was performed starting from the structure obtained after IAM HO refinement and $X-H$ bond elongation. This refinement, starting from zero P_{lm} multipolar and neutral atoms $P_{val} = N_{val}$ parameters, was nevertheless performed using the fixed κ and κ' values defined in the ELMAM library.

The atomic coordinates and ADPs of the substructure were refined against the whole resolution range. Constrained

refinement of the charge-density parameters P_{val} and P_{lm} was also performed on the H atoms. For both multipolar refinements, the charge-density parameters were refined alternatively with the structural (x, y, z) and thermal U_{ij} parameters of the substructure atoms until total convergence.

Only the multipole parameters that differed from zero in the ELMAM database were refined. The κ and κ' coefficients were kept fixed at the transferred ELMAM database values owing to the difficulty in refining them. In the multipolar refinement of crambin, the κ and κ' coefficients also had to be kept fixed. One can then consider fixing these parameters as a constraint on valence electron-density expansion/contraction (Volkov et al., 2001). Such κ and κ' constrained modelling was also proposed by Abramov et al. (1999) for the charge-density refinement of small molecules in order to obtain results that are physically more meaningful.

3. Results of the IAM refinements

3.1. Flat solvent modelling

In the case of the hAR crystals, the solvent content, 36%, is rather small for a protein (Howard et al., 2004). The bulk-solvent parameters were initially $k_{sol} = 0.35$ e Å^{-3} and $B_{sol} = 50$ Å^2 as suggested by Fokine & Urzhumtsev (2002) and converged during MoPro refinement to $k_{sol} = 0.33$ (1) e Å^{-3} and $B_{sol} = 61$ (4) Å^2. This procedure leads to a significant improvement of statistical indices at low resolution when the solvent structure factors were applied: the $R(F)$ factor dropped from 14.6% to 13.7% and $R_{free}(F)$ from 16.3% to 14.9% for reflections with $d > 2$ Å. The statistical indices using the whole resolution range and an $I/\sigma(I) > 0$ cutoff dropped from 9.70% to 9.63% for $R(F)$ and 9.97% to 9.86% for $R_{free}(F)$.

3.2. Unrestrained IAM refinement

In order to take into account the modifications that are inherent to the new solvent model, which have an impact on the low-resolution structure factors, the atomic fractional coordinates and ADPs of the selected system were first refined versus all diffraction data. In this refinement, the crystallographic indices $R(F)$ and $R_{free}(F)$ dropped from 9.70% and 9.86% to 9.23% and 9.78%, respectively (Table 3). The lack of restraints during this refinement leads to changes in some structural features compared with the SHELXL refinement. At first, the average atomic B_{eq} factors of heavy atoms within the selected subsystem increased from 4.23 to 4.46 Å^2. This increase in thermal motion amplitude may be explained by the suppression in this refinement of the restraints that limit the anisotropy of the ADPs. These restraints may have a deconvolution effect between the ADPs and the deformation electron density. One can conclude from this observation that the use of restraints to refine a macromolecular structure against subatomic resolution X-ray data at the IAM level has the ability to limit, at least partially, the perturbation arising from the nonmodelled bonding electron density.

research papers

3.3. IAM residual deformation density

The deformation electron-density signal was inspected in the residual maps expressed as Fourier difference syntheses after the IAM stereochemically nonrestrained refinement with *MoPro* software. Significant electron-density peaks can be observed on a large proportion of covalent bonds, especially where the atomic B_{eq} factors are sufficiently small, but generally at very varying levels.

The residual electron density after IAM HO refinement is shown in Fig. 5 for the C=O, C—N and C—C$^\alpha$ bonds of the selected subset. The residual peaks globally decrease with the thermal motion of the atoms forming the bonds. Significant bonding electron-density peaks can be observed on most of the main-chain covalent bonds. The exponential fits in Fig. 5 show that the C=O carbonyl double bonds have the strongest electron-density peaks on average, as expected and as previously pointed out by Lario & Vrielink (2003). The C=O peaks heights also decrease the most slowly with thermal motion.

Some amino acids, especially those in or close to the active site, such as Tyr48, present bonding electron density on each covalent bond along the side chain, comparable to the situation one might expect for a small molecule. This is also the case for some long side chains with large conformational freedom, such as lysine residues, for which thermal displacement parameters are often higher.

Some density peaks corresponding to electron lone pairs are also visible, but to a lesser extent than for covalent bonds, especially on the main-chain O atoms (Fig. 6). A striking feature is the inhomogeneous character of the residual electron density observed on the hAR structure when compared with small-molecule cases. Even though residual electron density can be observed on bonds in large parts of the structure, the peak heights are very variable for chemically
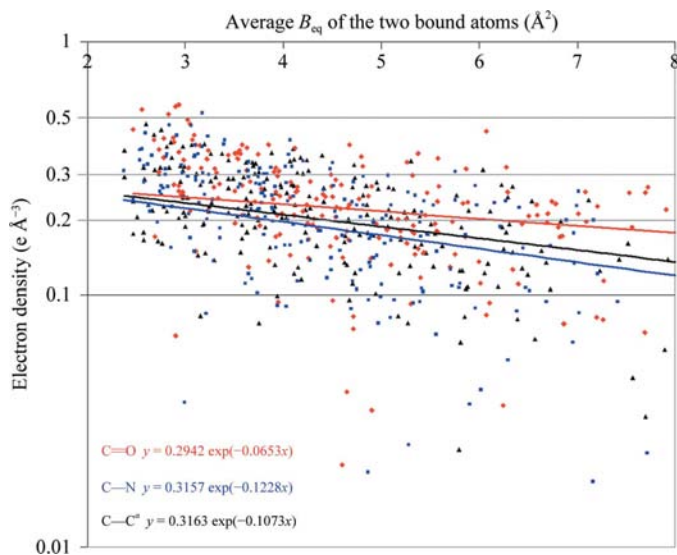
equivalent bonds involving atoms with comparable B_{eq} factors (Fig. 6). For instance, on some bonds with low atomic thermal motion and with the rigid-bond conditions satisfied (Hirshfeld, 1976) either no deformation electron density is visible or it is present to a much lower extent than on neighbouring covalent bonds. Fig. 7 shows the Fourier difference electron-density three-dimensional map truncated at the 0.25 e Å^{-3} isosurface in the region around the Lys262 residue. The smallest electron-density peaks are visible on the peptide bonds rather than on the lysine aliphatic side chains. Bonding density is completely unobserved, for example, on the Pro261 C—Lys262 N bond, while it is present on the neighbouring main-chain covalent bonds. However, the atomic B_{eq} factors in this region are all very small: the lowest value is 2.8 Å^2 for Pro261 C and the largest, as expected, is for the terminal Lys262 N$^\zeta$ atom (4.7 Å^2). The absence of bonding electron density for some covalent bonds with low thermal motion is explained by random noise in the residual Fourier maps resulting from the diffraction data with $R(F)$ = 9.23%.
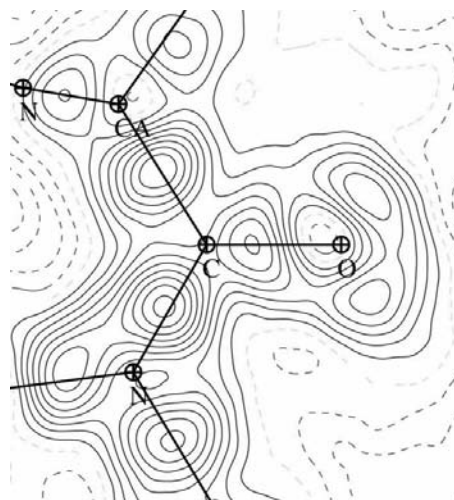

Figure 6
Residual deformation electron density in the peptide 260–261 bond plane, showing peaks corresponding to electron lone pairs on a carbonyl O atom. Contours are ±0.05 e Å^{-3}; black continuous lines, positive; grey dashed lines, negative.

Figure 5
Residual peak heights in difference Fourier $F_{obs} - F_{calc}$ electron-density maps after HO IAM refinement on the C—C$^\alpha$, C—N and C=O covalent bonds of the substructure.

In the plot (Figure 5):
- Average B_{eq} of the two bound atoms (Å^2)
- Electron density (e Å^{-3})
- C=O $y = 0.2942 \exp(-0.0653x)$
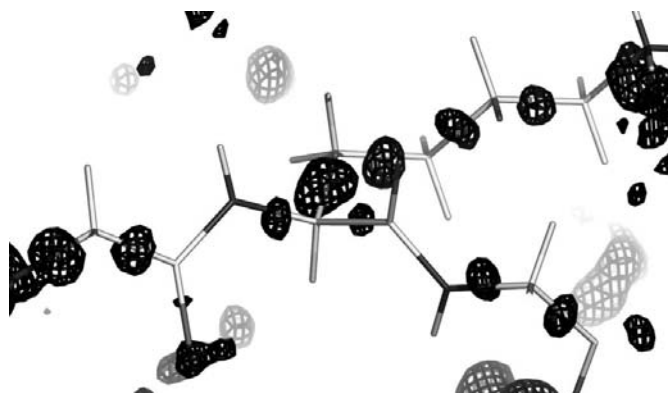- C—N $y = 0.3157 \exp(-0.1228x)$
- C—C$^\alpha$ $y = 0.3163 \exp(-0.1073x)$

Figure 7
0.25 e Å^{-3} isosurface of the initial residual electron-density map around the Lys262 residue. The lysine side chain is visible at the top right corner.

Table 4
Evolution of statistical indices during IAM HO refinement with the three resolution shells used successively.

The last line gives the crystallographic agreement factors for all reflections with $I/\sigma(I) > 0$ before and after the whole procedure of IAM HO refinement.

Resolution range (Å)	No. of reflections Work/free	$R(F)$ (%) Initial	Final	$R_{\text{free}}(F)$ (%) Initial	Final	$wR^2(F)$ (%) Initial	Final
0.66–1.0	311063/16393	9.83	9.57	10.05	9.80	9.83	9.50
0.66–0.83	206132/10852	10.47	10.08	10.72	10.45	10.03	9.58
0.66–0.77	144791/7663	11.75	11.20	11.96	11.68	11.01	10.30
All data	467221/24615	9.23	9.36	9.78	9.71	12.93	13.41

3.4. High-order refinement

High-order refinement (HO) is a method that is typically used in the field of small-molecule charge-density studies. The procedure consists of refinement of the structural and thermal parameters of the non-H atoms against only the high-resolution X-ray data. HO refinement improves the deconvolution of the deformation electron density from the thermal smearing effects. This allows the x, y, z and U_{ij} atomic parameters to be obtained in a less biased manner by the nonmodelled aspherical electron density. Indeed, at the stage of IAM refinement, the structural parameters are usually perturbed by the nonmodelled deformation electron density owing to valence-electron redistribution (Brock *et al.*, 1991; Pichon-Pesme *et al.*, 1995; Jelsch *et al.*, 1998).

As the valence-shell electron density is more diffuse than the core electron density, the scattering factors of the valence electrons in reciprocal space decrease more rapidly to zero at high resolution when compared with the core electrons. This effect on the atomic scattering factor is depicted in Fig. 4 for a free C atom. The carbon core electrons correspond to the $1s^2$ shell and the valence electrons to the $2s^2$ and $2p^2$ shells. In the high-order refinement method, the atomic fractional coordinates and thermal motion parameters are then mostly adjusted against the core electron density. This procedure leads to an initial partial deconvolution of the ADPs in the hAR structure. As a consequence, the residual maps usually show stronger deformation electron-density peaks and the thermal motion modelling of the protein is improved.

To be effective in protein crystallography, the resolution range for the HO refinement should be chosen in order to minimize the influence of valence-electron scattering compared with that of the core electrons. In principle, for C, N and O atomic species, a resolution truncation at $d < 0.7$ Å is sufficient (Fig. 4). In practice, however, the selected resolution range should be chosen with the lower limit as high as possible while still satisfying the parameters/observations ratio criteria. Therefore, the resolution limit for the high-order refinement was chosen at $d < 0.75$ Å with a diffraction data resolution limit of $d < 0.66$ Å. This allows the 10 404 structural parameters of the selected subsystem to be properly refined against more than 144 000 experimental data remaining in this resolution range. This only applies to the non-H atoms of the subsystem. H atoms do not scatter further than $d = 1.0$ Å and

therefore cannot be refined against HO reflections only. Each of these successive refinements of atom positions and thermal motion parameters led to very significant improvements in all the crystallographic indices $R(F)$, $wR^2(F)$ and $R_{\text{free}}(F)$, as depicted in Table 4.

At the end of IAM HO refinement, the positions of the H atoms were modified using *MoPro* distance-constraint functionalities: the X—H bond lengths were adjusted to their corresponding average values observed in neutron diffraction studies (Allen, 1986). This procedure is necessary to properly model the electron density of X—H groups upon ELMAM database transfer. All residual electron-density maps presented subsequently are computed taking into account this X—H bond-length elongation. As a consequence, the electron-density depletion around H-atom nuclei should appear more pronounced in residual Fourier maps. As observed in Table 3, this modification of H-atom positions leads as expected to a slight worsening of the agreement factors: they no longer correspond to X-ray positions and the H—X bonding electron density is not modelled in the IAM formalism.

The agreement factors computed with all experimental data after IAM HO refinement are slightly higher than the previous indices (after stereochemically unrestrained IAM refinement). This is a consequence of the low-order reflections that were omitted in IAM HO refinement. These omitted low-resolution reflections then have higher R factors and this effect is related to the observation that $R_{\text{free}}(F) > R(F)$ (Brünger, 1992). The increase in the R factors during IAM HO refinement in the medium- and low-resolution range can be attributed to the fact that the bonding electron density has not been modelled. In the refinement against the whole resolution range, this deformation electron density was previously partially accounted for by the U_{ij} thermal parameters, although this has no physical meaning (Coppens, 1967; Brock *et al.*, 1991; Pichon-Pesme *et al.*, 1995; Jelsch *et al.*, 1998; Afonine *et al.*, 2004). There is also a notable variation of the overall scale factor between the two resolution ranges, which may have an influence on the agreement factors when they are computed in different shells.

3.5. Residual electron density and HO refinement

Most of the strongest residual electron density which appears in the initial maps after IAM structure refinement (Fig. 7) corresponds to the nonmodelled covalent bonding electron density. HO refinement, when using spherical neutral atom modelling (IAM), leads to a significant enhancement of the bonding electron-density peaks visible in residual maps in many parts of the hAR substructure. Fig. 8 illustrates the effect of IAM HO refinement in the case of the His110 residue involved in the catalytic process. IAM HO refinement allows a drastic sharpening of the bonding electron density visible on main-chain and side-chain atoms, except for the C^α—C^β covalent bond. Furthermore, the resulting electron density clearly confirms the protonation state of the His110 residue: the $N^{\varepsilon 2}$ hydrogen electron density becomes vastly more

noticeable. In the same way, IAM HO refinement allows the electron lone pair of the nonprotonated $N^{\delta 1}$ atom to appear clearly, directed towards the H atom of a water molecule. Surprisingly, no electron density appears for the $C^{\varepsilon 1}—H^{\varepsilon 1}$ bond in either the initial or the IAM HO refinement residual maps.

Another example of signal improvement induced by HO refinement is the case of the Asp43 residue. This residue, which is located in the active site, makes several short hydrogen bonds to Lys77 and to a hydroxyl group of the NADP$^+$ cofactor. As depicted in Fig. 9, HO refinement leads to a significant improvement of the oxygen electron lone pairs on the Asp43 carboxylate moiety, but with striking dissymmetry. This feature was observed for numerous carbonyl-type O atoms in the hAR structure. This behaviour may be explained for Asp43 $O^{\delta 1}$ atom: the visible electron lone pair corresponds to the direction of a short hydrogen bond to a neighbouring water molecule [$d(43\ O^{\delta 1}—2059\ OW) = 2.8$ Å]. The other electron lone pair is involved in a longer hydrogen bond to Gln183 [$d(43\ O^{\delta 1}—183\ N^{\varepsilon 2}) = 3.0$ Å]. It should also be recalled that the lower the thermal motion is, the stronger the bonding electron density and electron lone pairs generally appear in residual maps. In the same way, the $O^{\delta 2}$ atom of Asp43 forms a hydrogen bond to a hydroxyl O atom of the NADP$^+$ cofactor. The visible electron lone pair of Asp43 corresponds to that involved in this interaction. The OD2 atom also makes a hydrogen bond to Lys77, in which it could be partially protonated (Blakeley et al., 2008).

The global enhancement of the deformation electron-density signal by HO refinement is illustrated in Figs. 8 and 9 for several bond types. The average peak height corresponding to the residual electron density in all (non-H) covalent bonds in the subsystem refined region increases slightly and has a narrower shape after HO refinement (Fig. 10).

4. Multipolar refinements

4.1. Charge-density parameters and constraints

Beyond the atomic coordinates, occupancies and description of thermal motion, the Hansen & Coppens (1978) multipolar atom model naturally implies supplementary atomic parameters that are required to take into account the deformation electron-density distribution. Firstly, the spherical valence ρ_{val} (1) is multiplied by a factor P_{val} which corresponds to the atomic valence population. In the IAM

stage, P_{val} was set to the value for the neutral atom (e.g. six electrons for O atoms, four for C atoms . . .). The refinement of P_{val} leads to an experimental estimation of the atomic partial charge by allowing interatomic electron transfer. The total number of electrons in the asymmetric unit was kept constant by applying an electroneutrality constraint. The multipolar populations, P_{lm}, allow deviations of the atomic electron density from spherical symmetry to be taken into account. Both spherical and multipolar valence density are modulated by radial expansion coefficients κ and κ'. A value above or below unity leads to an electron-density contraction or expansion around the atomic nucleus, respectively. The Hansen and Coppens atom model, including structural and thermal motion parameters, leads to a maximum of seven parameters per H atom, 25 parameters for second-row element atoms and 36 parameters for heavier chemical species. The κ and κ' parameters are added to these parameters and can be associated either with each atom of the molecular structure, with groups of atoms types or with chemical species only. As a consequence, the multipolar parameterization of electron density leads to an increase in the number of atomic parameters. However, in the transferability approximation one can apply constraints to charge-density parameters in order to reduce the number of variables effectively refined in the least-squares procedure.
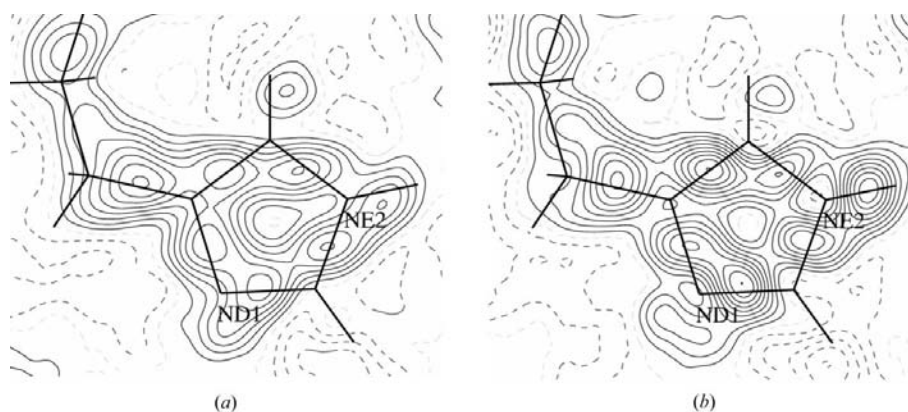

Figure 8
Residual deformation electron density in the catalytic His110 imidazole plane (a) after IAM and (b) after IAM HO refinement. Contours as in Fig. 6.
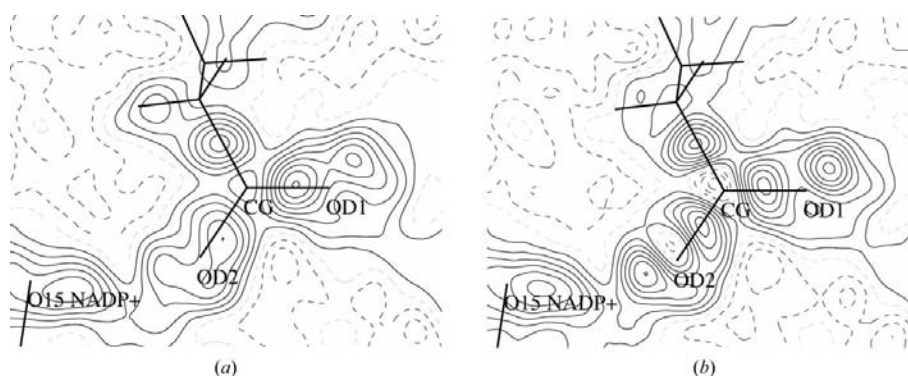

Figure 9
Residual deformation electron density in the plane formed by the $O^{\delta 2}$, C^{γ} and $O^{\delta 1}$ atoms of Asp43 (a) after IAM refinement and (b) after IAM HO refinement. Contours as in Fig. 6.

The first type of constraint is one of chemical equivalence, which is based on the assumption that atoms of the same chemical type with a similar chemical neighbourhood should share an identical charge-density distribution. The equivalence constraints are applied here according to the ELMAM library definition (Pichon-Pesme et al., 1995, 2004; Zarychta et al., 2007). The atoms that are most prone to deviate from the average electron density of the library are polar groups such as N—H and C=O (Fernandez-Serra et al., 2000). The application of such constraints allows a reduction in the number of refined charge-density parameters, as it is sufficient to refine the variables associated with one reference atom amongst all identical constrained atoms. In a protein structure, the identical chemical patterns along the polypeptide main chain, as well as the repetition several times of all amino acids in the sequence, are obviously well adapted to this procedure. In the field of small-molecule charge-density studies, such constraints are often applied in the early stages of refinement in order to improve its stability. They are relaxed in the later stages in order to take into account local electron-density distribution features. However, in the case of the present macromolecular study these constraints were held fixed during the whole refinement process.

The second type of constraints are the local symmetries (mirror planes, threefold axes etc.) applying to a given atom. Their application leads to zero multipoles which are not refined any further. For example, in the case of sp^2-hybridized atoms which have threefold pseudo-symmetry, such as the peptide N atom, only two multipoles are used. The octupole $Ox(x^2 - 3y^2)$, which has threefold symmetry, allows the modelling of most of the deformation electron density of the N atom (Table 5). Consequently, with the application of these constraints the number of refined multipole populations was only 269. In the same way, the number of refined valence populations was lowered to 50.

4.2. Library transfer

After improving the deformation electron-density signal visible in the residual maps through IAM HO refinement, transfer of multipolar parameters from the ELMAM library to
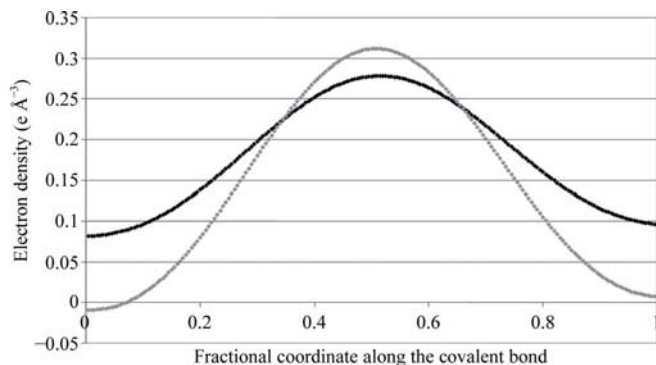

Figure 10
Average residual electron-density profile along covalent bonds between non-H atoms of the subsystem refined against high-resolution data: black curve, after IAM refinement; grey curve, after IAM HO refinement.

the hAR protein atoms was carried out (Zarychta et al., 2007). The transfer procedure, after an overall scale-factor adjustment, leads to an immediate and significant improvement in the statistical indices. The $R(F)$ and $R_{free}(F)$ dropped from 9.54% and 9.90% (values obtained after X—H bond-length elongation) to 9.11% and 9.43%, respectively (Table 3). This immediate drop in the crystallographic indices already indicates an improvement of the protein model. The decrease in the free R factor (Brünger, 1992) indicates that this model improvement does not suffer from the effect of overfitting arising from the introduction of new parameters. Furthermore, as only the overall scale factor was refined at this point, there is no risk of an over-refinement effect.

The amino acids described in the data bank were constructed using average multipolar parameters for atom types, independently of the amino-acid category. Therefore, the residues in the database do not hold a total number of valence electrons corresponding to a strict formal charge of −1, 0 or +1 electrons as in some dictionaries. For instance, formally neutral amino acids such as alanine or methionine actually have a total charge of −0.02 and −0.12 electrons in the database, respectively, while the basic amino acid lysine has a charge of +1.1 electrons (Zarychta et al., 2007). This implies that the transfer of multipolar parameters to the amino acids of a protein structure does not strictly fulfil the electroneutrality constraint.

In the case of hAR, if a formal charge of +1 or −1 is attributed to the charged residues (namely 11 Arg, 25 Lys, 15 Asp and 21 Glu, no protonated His), the global protein charge is zero. The total number of valence electrons was N_e = 17 146 in the IAM case and became N_e = 17 129 electrons after the valence population transfer procedure. Even if the difference is small (less than 0.3%), it is necessary to keep the molecular system electrically neutral for electrostatic potential computations. Redistributed over all non-H and nonsolvent atoms, neutrality of hAR is reached with a charge shift of only 6.10^{-3} electrons per atom. This is much below the standard deviation usually observed for refined valence populations in small molecules and of course is well below those obtained in the case of hAR charge-density refinement. The magnitude of the atomic charge adjustment needed to keep the number of electrons in the protein constant upon transfer is then sufficiently small to validate the multipolar database valence population transfer.

4.3. ELMAM structure refinement

Following the multipolar parameter transfer and scale-factor readjustment, the structural parameters were refined. As expected, this refinement, which makes a structural adjustment after the important modifications induced by the modelling of the deformation electron density, again leads to a significant improvement of all statistical indices, which drop to $R(F)$ = 8.80% and $R_{free}(F)$ = 9.16%. Such an effect has previously been observed in several studies of database transfer to small molecules (Pichon-Pesme et al., 1995; Jelsch et al., 1998, Zarychta et al., 2007; Dittrich et al., 2005, 2006, 2007).

It is interesting to note that this subsequent improvement of agreement factors essentially arises from the modification of the thermal motion parameters in the refinement. This also means that the HO procedure at 0.66 Å resolution was not totally successful in the deconvolution of the thermal motion and the deformation electron density. A final small decrease (10%) in statistical indices occurs when both parameter types are subsequently refined together. Furthermore, few refinement cycles are generally necessary to almost reach the final agreement factors and convergence, indicating a robust least-squares procedure. This is an important point for the generalization of the application of the ELMAM refinement method.

4.4. ELMAM multipolar refinement

The starting model at this stage of the study is the protein structure obtained after ELMAM parameter transfer and structural parameter refinement. At this point in the refinement procedure, the electron-density model is strictly based on the ELMAM database. However, it is possible to go beyond this description by refining the multipolar parameters P_{val} and P_{lm} against the X-ray diffraction data, taking advantage of the symmetry and chemical equivalence constraints. As a result of the local symmetry application and the presence of several copies of a given atom type in the polypeptide, the reduction in the number of multipoles enables the refinement of the charge-density parameters of macromolecules (Jelsch *et al.*, 2000). The constrained multipolar refinement is based on atom chemical equivalence constraints derived from the ELMAM database atom-type definition (Zarychta *et al.*, 2007). Among the selected atom subset of hAR, a reference atom was chosen for each atom type represented. Usually, this reference atom type is the first occurrence of a given atom type within the selected substructure. All other atoms of this atom type are then set to be equivalent, in terms of multipolar parameterization, to the reference atom.

For instance, all the ELMAM library atom types corresponding to polypeptide main-chain $H^\alpha-C^\alpha-(X{=}O)-N-H$ were set to be equivalent, except for glycine or proline residues. A differentiation was made between standard $C^\alpha-H^\alpha$ groups and the $H^{\alpha1}-C^\alpha-H^{\alpha2}$ moiety in the glycine residue. The proline N atom was also differentiated from other N—H groups. This way, a total of 56 atom types were defined. Nearly all of the protein atoms within the substructure are involved in a chemical equivalence constraint. In some cases,

for instance for the protein main chain, the number of atoms that were set to be equivalent reaches 105. Other atom types are less represented within the subset, such as the methionine and cysteine sulfur moieties, which are present only twice and once, respectively. Only the valence populations and the multipolar parameters defined as nonzero in the ELMAM database were refined. This corresponds to a total of 320 refined multipolar parameters.

The agreement factor $R(F)$ shows a further small drop from 8.80% to 8.74% after the atomic position and thermal motion parameter refinement. In the same way, both $wR^2(F)$ and $R_{free}(F)$ improve during the constrained multipolar refinement (Table 3). Fig. 11 depicts the evolution of the residual electron density at various stages of the refinement. It clearly shows that most of the bonding electron-density features are taken into account by database transfer. The subsequent constrained charge-density refinement has a limited effect. Some electron density is still visible, but at a much lower level and without any significant local accumulation.

4.5. IAM multipolar refinement

The influence of the starting electron-density modelling on the multipolar refinement was then assessed. A similar

Table 5
Multipole parameters of the polypeptide main-chain atoms in the ELMAM library (first line) and after refinement *versus* the protein diffraction data with the starting points ELMAM library (second line) or IAM (third line).

The expansion/contraction coefficients κ and κ' were transferred from the ELMAM library in all three cases. Glycine C^α and H^α and proline N and C^α atoms are excluded. Estimated standard uncertainties obtained from the normal matrix inversion of the valence and multipole population refinement are indicated in parentheses. The estimated error of the average value in the library is also given.

Atom	κ, κ'	P_{val}	Dipoles	Quadrupoles		Octupoles	
C^α			y	xz	yz	x^3	y^3
	0.9930 (4)	4.111 (74)	−0.049 (15)	−0.171 (11)	−0.261 (14)	0.260 (17)	0.057 (9)
	0.9160 (15)	4.136 (12)	0.000 (3)	−0.133 (7)	−0.232 (7)	0.213 (7)	0.064 (6)
		4.154 (12)	0.001 (5)	−0.097 (7)	−0.183 (7)	0.168 (7)	0.048 (6)
C			x	z^2	$x^2 - y^2$	x^3	
	0.9980 (3)	3.976 (32)	0.102 (7)	−0.304 (11)	0.102 (10)	0.411 (14)	
	0.9410 (16)	3.973 (11)	0.090 (5)	−0.271 (5)	0.075 (5)	0.388 (7)	
		3.927 (11)	0.019 (5)	−0.133 (5)	0.017 (5)	0.313 (7)	
O			x	z^2	$x^2 - y^2$	x^3	
	0.9770 (3)	6.307 (16)	−0.075 (8)	−0.078 (7)	−0.070 (6)	0.039 (2)	
	0.9600 (27)	6.270 (8)	−0.070 (3)	−0.074 (4)	−0.078 (4)	0.031 (5)	
		6.190 (8)	−0.023 (3)	−0.019 (4)	−0.047 (4)	0.026 (5)	
N			y	z^2		x^3	
	0.9870 (3)	5.312 (55)	−0.072 (9)	−0.050 (10)		0.260 (8)	
	0.8670 (14)	5.213 (10)	−0.069 (4)	−0.054 (4)		0.233 (6)	
		5.095 (9)	−0.017 (4)	0.001 (4)		0.188 (5)	
H0			x				
	1.2080 (23)	0.680 (22)	0.166 (11)				
	1.0170 (26)	0.686 (7)	0.178 (6)				
		0.746 (7)	0.175 (6)				
H^α			x				
	1.1820 (14)	0.804 (46)	0.107 (8)				
	1.0910 (24)	0.873 (7)	0.142 (5)				
		0.910 (8)	0.139 (5)				

multipolar refinement strategy was applied, but starting from the molecular model obtained immediately following the IAM

HO refinement: *i.e.* without the transfer procedure. Interestingly, the model converges towards a higher $R(F)$ factor than
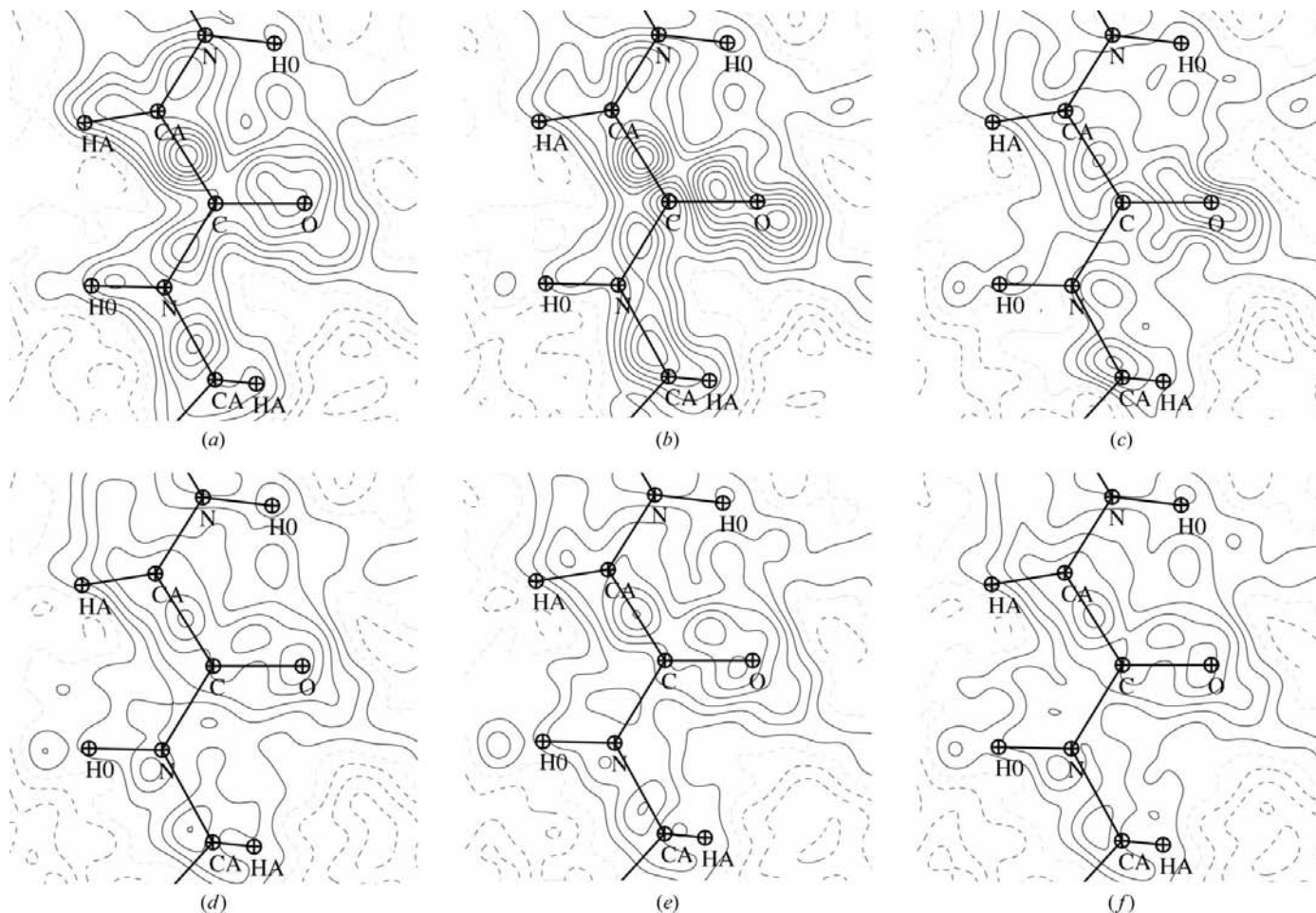

Figure 11
Evolution of residual electron density on the Ala45–His46 peptide bonds of the substructure during the various refinement steps. The stages of refinement are (*a*) IAM unrestrained, (*b*) IAM HO, (*c*) after database-transfer procedure, (*d*) after positional and thermal motion parameter refinement to readjust the transfer, (*e*) after IAM multipolar refinement and (*f*) after ELMAM multipolar refinement. Contours are as in Fig. 6.
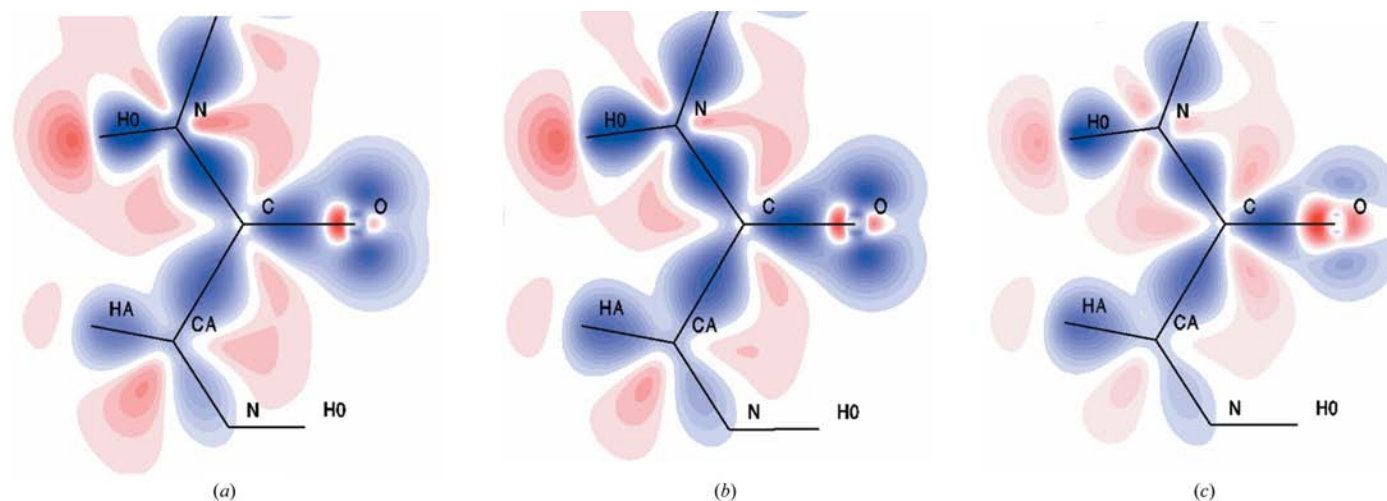
Figure 12
Static deformation electron density in the plane of the peptide bond. (*a*) As transferred from ELMAM library. (*b*) After ELMAM constrained multipolar refinement. (*c*) After multipolar refinement starting from IAM spherical neutral atoms. Contours are ± 0.05 e Å^{-3}, with positive values in blue and negative values in red. Electron-density values between -0.05 and $+0.05$ e Å^{-3} are not represented.

Guillot *et al.* · Charge-density analysis of human aldolase reductase   **579**

that obtained after the ELMAM constrained refinement. The $R(F)$ and $R_{\text{free}}(F)$ values are 8.97% and 9.49%, respectively (Table 3). The higher $R_{\text{free}}(F)$ factor demonstrates that the constrained charge-density refinement starting from IAM is trapped in a false minimum. However, the R factors are lower compared with the IAM refinement stage, which indicates that the IAM constrained multipolar refinement is better than the IAM refinement.

4.6. Main-chain static electron density

Atom equivalence constraints on the electron-density parameters act as an implicit averaging of the deformation electron density amongst chemical moieties. Hence, as expected, the constrained ELMAM multipolar refinement leads to deviations from the library that are linked to both the number of atoms included in a given constraint and to the atomic thermal motion of constrained moieties. Usually, the larger the number of constrained atoms, the smaller the

Figure 13
Static deformation electron density in the plane of the carboxylate moieties of glutamic acid and aspartic acid residues. (*a*) As transferred from ELMAM database. (*b*) After constrained refinement. Contours as in Fig. 6.

Figure 14
Static deformation electron density in the hydroxyl plane of a tyrosine residue. (*a*) As transferred from ELMAM library. (*b*) After constrained multipolar refinement. Contours as in Fig. 6.

discrepancy from the starting parameters. This is obviously the case for the polypeptide main-chain atoms, which correspond to the largest occurrences of a given atom type within the substructure. The static deformation electron densities in the peptide plane obtained in the multipolar refinements are shown in Fig. 12. 17 multipoles are used to describe the standard main-chain group $H0-N-(C=O)-C^{\alpha}-H^{\alpha}$. In the ELMAM multipolar refinement, the average deviation observed for multipoles is only 1.2 times the standard error $\sigma = 0.01$ of the average values in the ELMAM library (Table 5).

The multipoles that deviate from the library by more than 2σ in the peptide moiety are the yz^2 and x^3 multipoles of the C^{α} atom (3.3σ for both). Not surprisingly, the dipoles attributed to the main-chain H0 and H^{α} atoms also show a relatively large variation (2.5σ). The most significant deviations lead to a decrease (in absolute value) in the multipolar populations, indicating an attenuation of the static deformation electron density compared with small molecules. This is in accordance with an average thermal motion in the hAR structure which is larger than that usually found in the small molecules that were used for the database building. Higher atomic B_{eq} factors usually lead to thermally smeared deformation electron density, as previously observed for the protein crambin (Jelsch *et al.*, 2000).

In the IAM multipolar refinement performed starting from neutral and spherical atoms, the same scheme of constraints was applied. The resulting static deformation electron densities show an even more attenuated morphology. The static deformation electron density in the peptide-bond plane is shown in Fig. 12. Electron-density peaks are in this case 30% weaker on average on the covalent bonds. The attenuation is dramatic in the case of the O-atom electron lone pairs. However, the observation of the deformation electron density starting from the IAM stage clearly demonstrates that the nonspherical features of the electron density are intrinsically present in the hAR diffraction data. In this case, it cannot be a bias introduced in the model by the transfer of the ELMAM multipolar parameters.

4.7. Side-chain static electron density

The multipolar parameters of the protein side chains were also refined, although with fewer equivalence constraints available compared with the main chain. Only starting multipolar parameters from the ELMAM database
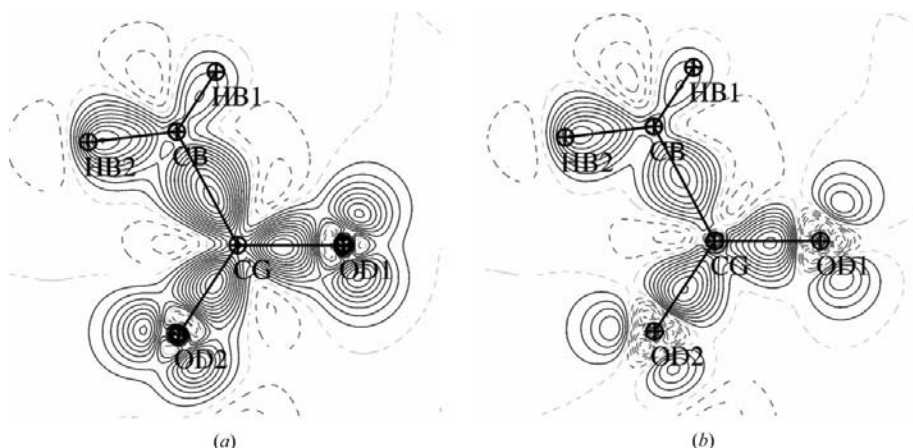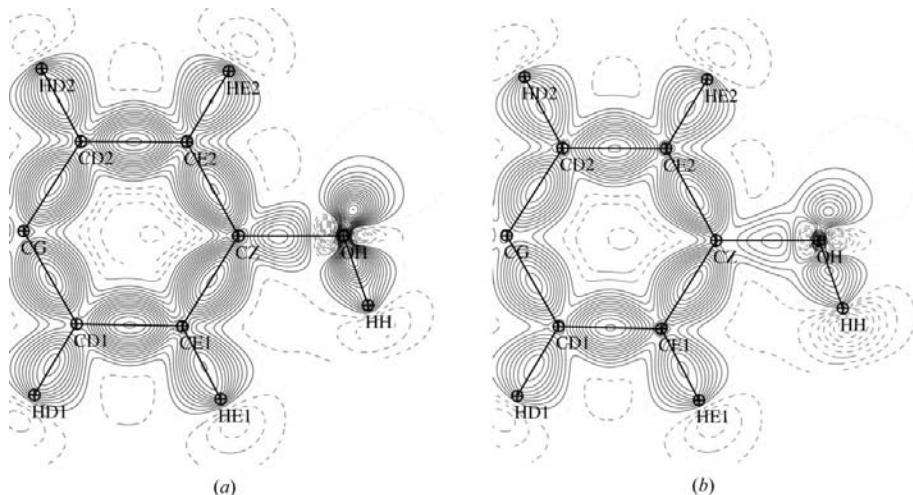
were used here. O atoms present some difficulty in charge-density refinement owing to the weakness of the electron lone-pair signal. This is confirmed by an examination of the static electron densities of Asp and Glu carboxylate moieties (Fig. 13). The lone-pair peaks in the static electron-density maps of the carboxylate O atoms ($N = 5$ copies) become 30% lower after multipolar refinement compared with the ELMAM values.

The same behaviour can be observed for the tyrosine hydroxyl O atom ($N = 7$ copies). Fig. 14 shows the static electron density in the $C^{\zeta}-O^{\eta}-H^{\eta}$ plane for the ELMAM database and after constrained multipolar refinement. In agreement with the general increase in atomic thermal motion along the side chain, there is only a 20% decrease in the bonding electron density along the ring C—C bonds, whereas a 40% decrease is observed for the $C^{\gamma}-O^{\eta}$ bond and a 50% decrease for hydroxyl O-atom electron lone pairs. This can also be explained by the smaller number of chemical equivalences available for O^{η} compared with the C—H atoms of the aromatic ring. A very striking feature can be seen for the tyrosine H^{η} atom, which shows a clear amplification of the dipolar electron density. This may be the consequence of a peculiarity of the tyrosine residue in the ELMAM database (Zarychta *et al.*, 2007), in which the $O^{\eta}-H^{\eta}$ bonding electron density is modelled mostly by the O-atom multipoles and by an unusually weak hydrogen dipole of +0.06. After the constrained multipolar refinement this dipole increases to a higher value of +0.13, while the multipole variations for the O^{η} atom are smaller. The increase in the H-atom dipole can be explained by the fact that the multipolar modelling principle allows an electron-density accumulation to be taken into account by several multipoles attributed to several distinct neighbouring atoms.
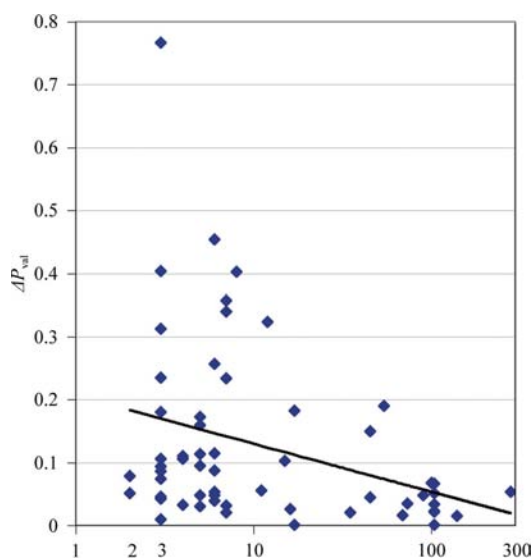

Figure 15
Scatter plot of the valence population differences $|\Delta P_{val}|$ (e) between the ELMAM values and the ELMAM constrained multipolar refinement. The absolute values of the differences are plotted *versus* the number of occurrences of the atom type in the subselection.

Aromatic and aliphatic neutral residues such as phenyl-alanine, isoleucine, leucine and valine are made up of few different atom types. In the case of phenylalanine, for instance, all sp^2 C—H C atoms have the same atom type and were set to be chemically equivalent in the constrained refinement. Again, the most significant feature observed is a 25% attenuation of static deformation electron density on the covalent bonds.

The overall largest P_{lm} multipole discrepancies are observed for the sulfur-containing residues methionine and cysteine, for which the final values are sometimes of opposite sign when compared with the multipolar database values and reach unrealistic values. This is a consequence of the fact that these moieties are involved in a limited number of equivalence constraints ($N = 2$ and 1). Furthermore, the S^{δ} atom of the methionine residue in the multipolar database is described by a larger number of multipoles up to the hexadecapole level. This points out the difficulty in refining multipolar parameters of protein atoms when the number of representatives present in the (selected) structure is small, especially for heavy atom types. In the case of hAR, the two S atoms require the availability of more numerous atom equivalences in order to properly refine their electron density.

4.8. Refined valence populations

The valence populations P_{val} of the substructure atoms were refined using the same constraint scheme as for the multipolar populations. In the hAR constrained refinement, atomic valence populations converge towards realistic values, with the exception of the sulfur-containing moieties mentioned above. An overall electroneutrality constraint on the substructure was applied during this refinement, but no limitations were imposed to keep the number of electrons constant within a given residue or a specific moiety. This implies that this refinement allowed the atomic valence populations to adjust freely with respect to experimental data. As the electroneutrality constraint introduces a correlation between all involved P_{val} parameters (the sum of all P_{val} values being constrained to remain constant), the substructure valence-population refinement was therefore performed excluding the S atom types.

Fig. 15 plots the valence-population shifts after constrained refinement *versus* the number of times a given atom type is involved in its corresponding equivalence constraint. It shows again, as mentioned before, that the deviation of the parameter from the database value in the constrained refinement is always small when the atom type is involved in a large number of constraints. As for the multipolar parameters, the valence populations of the main-chain atoms involved in 105 equivalences are among the most stable during refinement. The largest shift of 0.07 electrons with respect to the database values occurs for the N atom (Table 5). This shift is only twice the standard error on the mean obtained from the multipolar database building. In the database, the total number of electrons for the polypeptide main-chain atoms does not correspond to a neutral peptide (Zarychta *et al.*, 2007). The

Guillot *et al.* · Charge-density analysis of human aldolase reductase **581**

ELMAM charges are therefore relevant for use as an experimental force field.

There is a global attenuation ($\times 0.85$) of atomic charges during the ELMAM constrained multipolar refinement (Fig. 16). The initial charge (sum of $N_{val} - P_{val}$) of the main-chain repetitive moiety $H^\alpha - C^\alpha - (C = O) - N - H0$ is $q = -0.19\ |e|$ in the ELMAM database. The constrained main-chain refinement leads to a slight attenuation of this negative charge, which becomes $q = -0.15\ |e|$. However, in the refinement, when one starts from neutral spherical atoms (IAM) the global charge of the peptide moiety remains close to zero ($q = -0.02\ |e|$; Table 5). This can be explained by the global zero charge in the starting model and by the reduced charge of the N atom in the IAM multipolar refinement. The global charge is slightly dissimilar to the $-0.10\ |e|$ value found in the University of Buffalo library based on theoretical calculations (Pichon-Pesme et al., 2004; Volkov, Koritsanszky et al., 2004; Volkov, Li et al., 2004; Li et al., 2006). The discrepancy may be attributed to experimental and theoretical uncertainties and to the differing chemical environments and atomic compositions of the peptides.

5. Comparison of the various steps

5.1. Residual electron-density maps

Residual Fourier electron-density maps for a representative peptide bond, Ala45–His46, are given in Fig. 11 after the main steps of the refinement procedure. As previously mentioned, it can be seen that IAM HO refinement enhances the deformation electron-density signal visible in Fig. 11(b).

At the step after $X - H$ bond elongation, database transfer and scale-factor refinement, it is obvious that the major part of
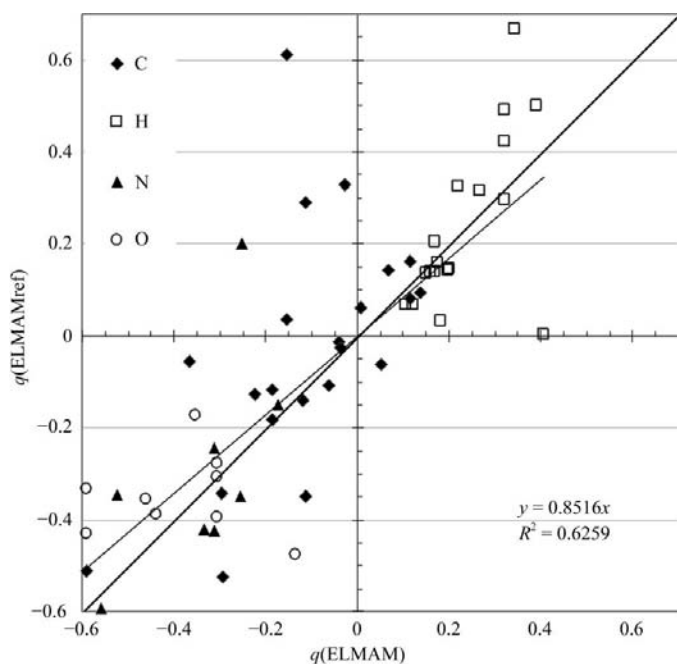

Figure 16
Scatter plot of atomic charges ($|e|$) from the ELMAM multipolar constrained refinement versus ELMAM values.

the bonding electron-density features have disappeared from the map shown in Fig. 11(c). Some attenuated deformation electron-density features seem to remain, notably around the O atom and on the $C^\alpha - C$ and $N - C^\alpha$ bonds. This effect is generally observed in OMIT maps: when modifications are applied to a molecular model, they are partially cancelled out in the Fourier $F_{obs} - F_{calc}$ difference maps (Hodel et al., 1992). A subsequent global refinement of the structure can then reduce the bias towards the initial structure.

In the next step, the refinement of the positional and thermal motion atomic parameters therefore leads to a significant reduction of the deformation electron-density features in the residual map (Fig. 11d). The subsequent constrained ELMAM multipolar refinement with x, y, z, U_{ij} parameters refined alternatively leads to a very similar residual map when compared with the previous step (Figs. 11e and 11f).

The residual electron-density map obtained after IAM multipolar refinement is, as expected, quite similar, with a slightly higher level of deformation electron density compared with that computed after ELMAM multipolar refinement. Although performed to convergence, this refinement leads to an attenuation of deformation electron-density modelling compared with that achieved in ELMAM multipolar refinement.

In both multipolar refinement cases, a significant level of residual electron density remains in a quite random distribution and corresponds to the R factor, which is around 9%. The bonding electron-density peak remaining here on the $C^\alpha - C$ bond of Ala45 is not generally observed on other residues and might be attributed to random noise in the Fourier residual map. The final multipolar model does not show the non-modelled deformation electron density in residual maps which is systematically visible in the starting steps of the procedure. The ELMAM transferred electron-density modelling is therefore a real improvement.

5.2. Anisotropic thermal displacement parameters

As expected for crystals that diffract to such high resolution, the atomic thermal displacement parameters, especially in the inner core of the protein, show exceptionally low values for a protein structure. Fig. 17 depicts 50% thermal ellipsoids for the whole starting model. Examination of atomic B_{eq} factors in the substructure shows that the lowest value is 2.41 $Å^2$ for the backbone C atom of Ser76, followed by 2.49 $Å^2$ for the C^α atom of Asp43 located in the active site. The highest atomic B_{eq} factor does not exceed 6 $Å^2$ in the protein active site, excluding the adenine part of the $NADP^+$ cofactor which is located near the protein surface. Based on the subatomic resolution and on the moderate thermal motion, one can expect limited thermal smearing, making it possible to observe significant deformation electron-density peaks at least on the covalent bonds in the residual maps.

The transfer of ELMAM parameters and subsequent refinement of the structural parameters led to a slight decrease in the dynamic parameters compared with the initial values.

The average atomic B_{eq} factor of the refined substructure dropped from 4.23 to 4.13 $\mathring{A}^2$ after ELMAM refinement. It has to be noted that this drop in B_{eq} factors is small but almost systematic for non-H atoms included in the selection. Indeed, only 6.5% of the refined non-H atoms show an increase in B_{eq} factor after ELMAM refinement (of up to +0.39 $\mathring{A}^2$). The remaining 93.5% of atoms show a decrease in the B_{eq} factor, with the largest decrease being -0.41 $\mathring{A}^2$ for the main-chain O atom of Ser214. Furthermore, it can be seen in Fig. 18 that the largest variations in atomic B_{eq} factors are observed for atoms that exhibit the largest thermal motion within the substructure. For instance, the largest discrepancy occurs for the $C^{\gamma 1}$ atom of Val166, the B_{eq} factor of which increases from 7.21 to 7.60 $\mathring{A}^2$. For atoms with initial moderate thermal motion ($2 < B_{eq} < 4.5$ $\mathring{A}^2$), the decrease in thermal motion amplitude is systematic. This clearly indicates an overestimation of the ADPs owing to the nonmodelled bonding electron density. However, the average decrease observed is 0.1 $\mathring{A}^2$, which is smaller than expected (-0.4 $\mathring{A}^2$) when compared with the results obtained from a database transfer on an oligopeptide (Jelsch *et al.*, 1998). This is a characteristic of protein refinement when compared with small molecules or even with oligopeptides.

The atomic B_{eq} factors at the end of the IAM multipolar refinement have an average value of 4.29 $\mathring{A}^2$, *i.e.* slightly higher than after the transferred ELMAM refinement. This is in accordance with the attenuated deformation electron density obtained from that refinement. The IAM multipolar refinement seems to be trapped in a false minimum and obviously indicates the necessity of ELMAM transfer before an optimal charge-density refinement.
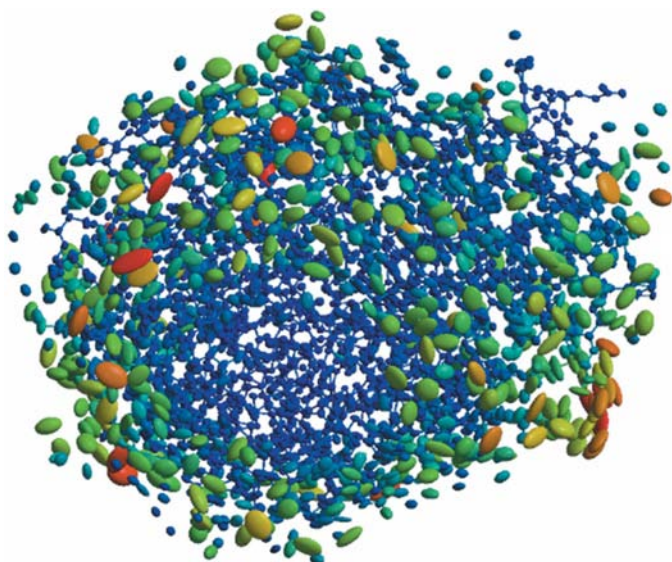
Differences in mean-square displacement amplitude (DMSDA) between atoms linked by a covalent bond are shown in Fig. 19. The rigid-bond test (Hirshfeld, 1976) gives an initial average value for DMSDA of 0.0023 $\mathring{A}^2$. This relatively low value can be explained by the use of the rigid-bond restraint (DELU instruction) with a 0.01 $\mathring{A}^2$ allowed standard deviation in refinement with *SHELXL*97 (Sheldrick, 2008). Hence, rigid-bond tests (Hirshfeld, 1976) were performed after the unrestrained IAM refinement with *MoPro* and after the ELMAM multipolar refinement on the covalent bonds of the protein substructure. There is a slight decrease in the DMSDA: the average absolute value drops from 0.0025 $\mathring{A}^2$
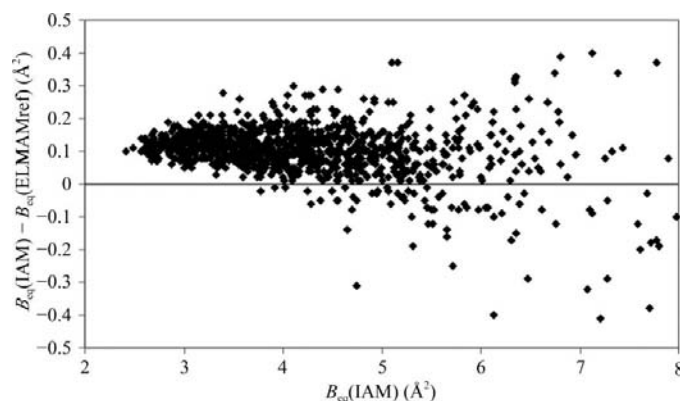

Figure 18
Differences of equivalent thermal factors B_{eq} of the substructure atoms between the initial model and after the final ELMAM multipolar refinement.

Figure 17
50% thermal ellipsoids with colouring scheme based on the equivalent atomic B factor for the full protein structure including water molecules at the surface. Lowest atomic B factors are shown in blue, while the highest are depicted in colours ranging from green to red. This picture was created with the program *Raster3d* (Merritt & Murphy, 1994).
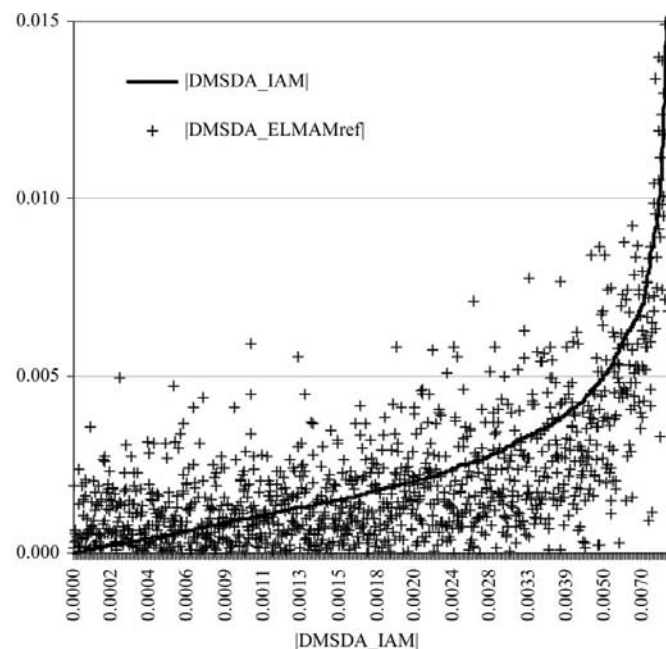
Figure 19
Analysis of the differences in mean-square displacement amplitudes (DMSDAs). The DMSDAs (Hirshfeld rigid-bond test) after the ELMAM constrained multipolar refinement are plotted as a function of the DMSDAs in the IAM unrestrained refinement. The line drawn corresponds to the reference DMSDA_ELMAMref values sorted in increasing order.

Guillot *et al.* · Charge-density analysis of human aldolase reductase   **583**

after the unrestrained IAM refinement to 0.0020 Å^2 after the ELMAM multipolar refinement. In a minority of cases, a slight increase in the DMSDA value can be observed: 224 atoms of the 1156 atoms of the selection show higher DMSDA values.

It has to be noted that the average anisotropy (U_{min}/U_{max}) of atomic ADPs surprisingly remains rather constant, with a 0.55 value in the initial model and a 0.54 value at the end of the ELMAM refinement.

5.3. Stereochemistry

The lengths of several bond types in the refined protein substructure after IAM unrestrained structure refinement and after ELMAM multipolar refinement are shown in Table 6. In

order to focus on the covalent bond types which correspond to the largest sample in the substructure, only the main-chain bonds are analyzed. They are depicted in Fig. 20 as ELMAM *versus* IAM unrestrained refined distance scatter plots. A good linear relationship appears between the two steps for each of the covalent bond types analysed with a 83–86% correlation. The slopes of the linear fits are in the 0.81–0.92 range, indicating a general narrowing of the distance distribution in the ELMAM multipolar refined structure.

These distributions are shown in Fig. 21 for the C—N and C=O peptide-bond lengths at the initial, IAM unrestrained and final ELMAM stages. For both bond types, the initial distribution shows a narrower shape (r.m.s.d. = 0.008 Å) than at the end of the IAM unrestrained refinement (r.m.s.d. =
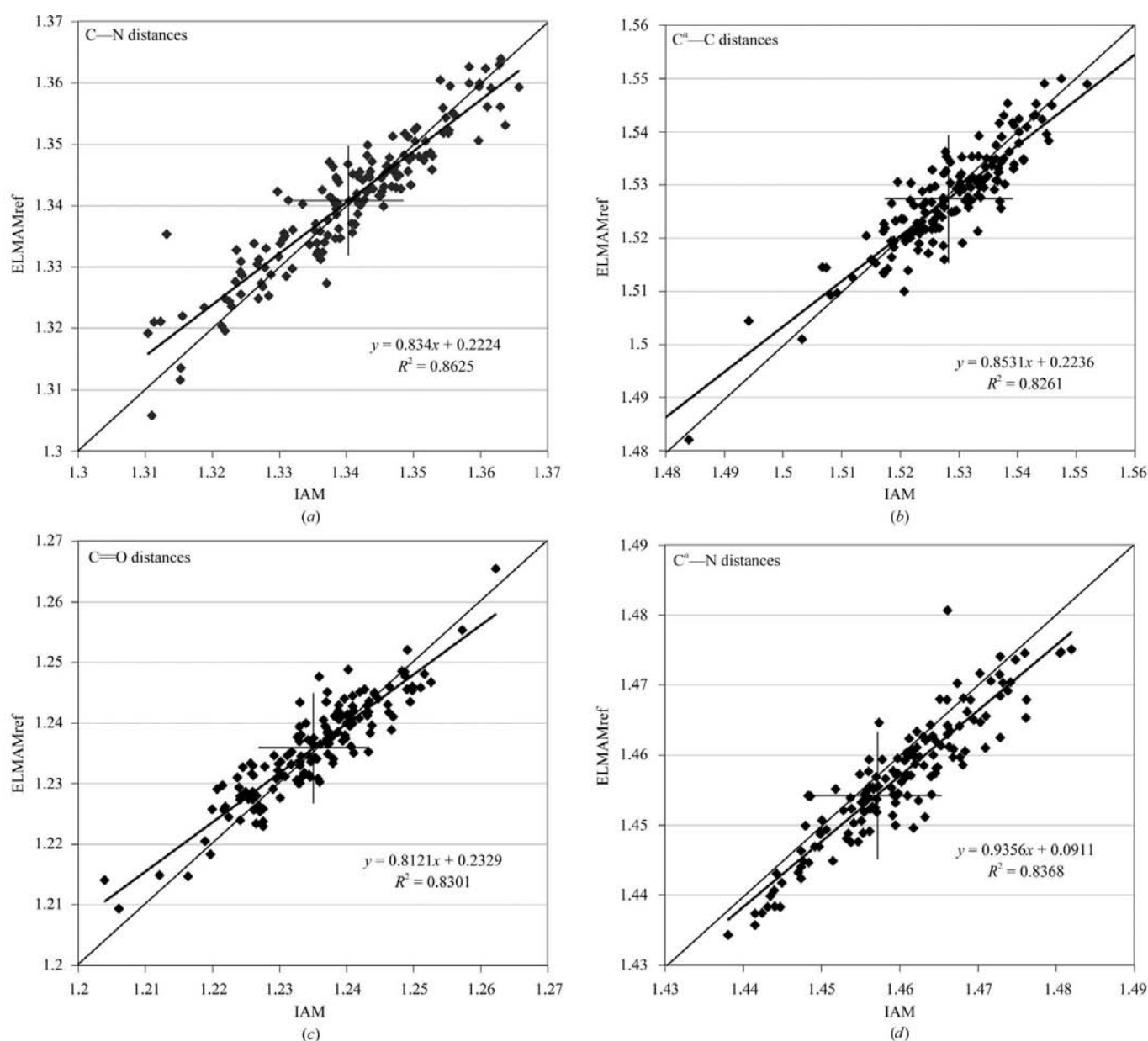

Figure 20
Main-chain interatomic distances (Å): (a) C—N, (b) C—C$^\alpha$, (c) C=O and (d) C$^\alpha$—N. The distances obtained after the ELMAM multipolar refinement are plotted *versus* the distances from the IAM stereochemically unrestrained structure. The linear fit and diagonal lines are also shown. The cross corresponds to the average values.

Table 6
Selection of bonds lengths (Å) in the hAR structure after IAM stereochemically nonrestrained refinement and after ELMAM multipolar nonrestrained refinement.

The values can be compared with those of the Engh and Huber stereochemical dictionary (Engh & Huber, 1991). The revisited values of Jaskolski *et al.* (2007) are from ten ultrahigh-resolution ($d < 0.8$ Å) protein structures in the PDB (UHRP) or from peptide structures in the Cambridge Structural Database (CSD) with $R(F) < 5\%$. The root-mean-square deviation (r.m.s.d.) in the sample is given in parentheses. The number of bond occurrences in the hAR subselection is also indicated.

Bond type	C=O	C−N†	C^{α}−N‡	C^{α}−C§
ELMAMref mean	1.236 (8)	1.341 (11)	1.457 (8)	1.528 (10)
IAM mean	1.235 (9)	1.340 (12)	1.459 (8)	1.529 (10)
Engh and Huber	1.231 (20)	1.329 (14)	1.458 (19)	1.525 (21)
Jaskolski and coworkers UHRP	1.234 (12)	1.334 (13)	1.455 (12)	1.527 (13)
Jaskolski and coworkers CSD	1.231 (9)	1.332 (8)	1.455 (7)	1.523 (11)
No. of bonds	143	125	124	128

† Proline N atoms are excluded. ‡ Proline and glycine residues excluded. § Glycine residues excluded.

0.011 Å). This is a consequence of the application of stereochemical restraints in the IAM initial restrained refinement.

In all cases, the C−N bond-length distribution is shifted towards larger values than the target value 1.329 Å of the Engh and Huber dictionary (Engh & Huber, 1991). It is interesting to note that even the distribution after the initial stereochemically restrained refinement is centred on a 1.343 Å C−N bond length, which is significantly larger than the target value (Fig. 21*a*).

The initial distribution of the C=O bonds in Fig. 21(*d*) clearly reflect the restraints applied in the refinement, with a target $d_{C=O} = 1.231$ Å from the Engh and Huber dictionary (Engh & Huber, 1991). The refinement of structural parameters without stereochemical restraints had the consequence of slightly increasing the r.m.s.d. of the bond lengths, corresponding to a wider distribution than when distance restraints are used (Fig. 21*e*). For instance, in the initial protein model the average C=O bond length in the ordered main chain was 1.237 Å with an r.m.s.d. of 0.006 Å. The unrestrained IAM refinement leads to an average value of 1.235 Å with an r.m.s.d. of 0.009 Å, with some distances being located beyond a 2σ deviation from the ideal value. The subsequent unrestrained ELMAM refinement allows the distribution to become slightly narrower, with an r.m.s.d. of 0.008 Å.

When average values are analysed, it appears that the bond-length differences between the IAM and the ELMAM refinements are rather small and do not exceed 0.002 Å for C^{α}−N bonds (Table 6). In all cases, variations are contained within the sample r.m.s.d. and are smaller than the usual standard uncertainties computed for bond lengths (~0.005 Å).

The use of a spherical atom model for the carbonyl moiety might be expected to lead to an overestimation of the C=O bond length owing to the nonmodelled electron lone pairs. On the other hand, the nonmodelling of the C=O bonding electron density could lead to a bond shortening. The IAM and the ELMAM unrestrained multipolar refinements lead to a very similar average C=O covalent bond length (Table 6). This can be explained by the high resolution of the diffraction

data. Indeed, the higher the resolution is, the less the structure is biased by the nonmodelled deformation electron density (Afonine *et al.*, 2004).

Ideal distances for stereochemical restraints were recently revisited (Jaskolski *et al.*, 2007) by the analysis of several protein structures refined at atomic resolution and of peptide compounds extracted from the CSD (Cambridge Structural Database; Allen, 2002). It appears for all four peptide-bond types in Table 6 that the sample r.m.s.d. computed from the hAR substructure is very close to the r.m.s.d. values obtained by Jaskolski *et al.* (2007) from the CSD peptides and from ultrahigh-resolution protein structures.

The average bond lengths and associated r.m.s.d. values obtained in our study fit especially well to those obtained by Jaskolski and coworkers from the protein structures. In their study, disordered regions and atoms with high thermal motion were also excluded. Our results confirm the slight overestimation of the bond-length variability in the Engh & Huber (1991) dictionary pointed out by Jaskolski and coworkers and in the recent 0.65 Å resolution analysis of the triclinic lysozyme structure (Wang *et al.*, 2007). The main discrepancy occurs for the C−N peptide bond. In the present study, the C−N bond [$d = 1.340$ (12) Å] is however slightly longer than in Jaskolski *et al.* (2007) [$d = 1.334$ (14) Å] and in the Engh and
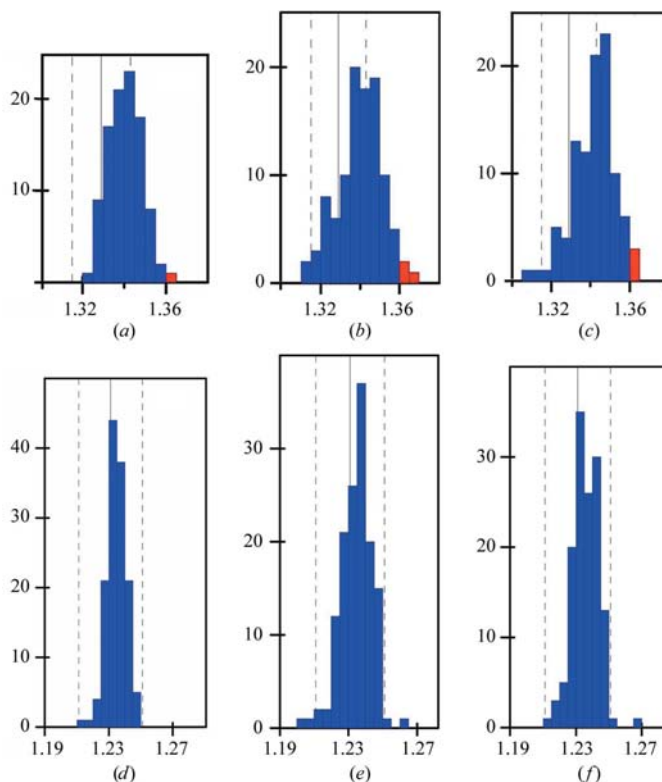
Figure 21
Distribution of C−N (*a–c*) and C=O (*d–f*) peptide-bond lengths in the refined substructure: (*a*), (*d*) initial stereochemically restrained model; (*b*), (*e*) IAM unrestrained model; (*c*), (*f*) ELMAM refinement model. The plain vertical line is the target value and the dashed lines correspond to the 2 r.m.s.d. deviation from the Engh and Huber dictionary (Engh & Huber, 1991). Figures were produced with *PROCHECK* (Laskowski *et al.*, 1993).

Huber dictionary [$d = 1.329\,(13)$ Å]. These C—N bond lengths values are nevertheless all compatible within the r.m.s.d. found in the samples.

5.4. Phase differences

In the case of small molecules, the phase change $|\Delta\varphi|$ between the IAM and the multipolar modelling has an impact of up to 0.2 e Å^{-3} on the electron density (Souhassou *et al.*, 1991, 1992). The evolution of calculated structure-factor phases was therefore analysed during the various refinement steps. The average absolute values of phase variations are weighted by the structure-factor amplitudes and are plotted in Fig. 22 in different resolution ranges. The nonweighted phase differences generally display the same tendencies but with a translation of the curves of +1° to +2°. When phases are compared between the initial IAM refinement and the final ELMAM multipolar refinements, it appears that the main phase variations are rather uniform *versus* resolution, with an average value of 2.9°. An interesting point here concerns the database-transfer procedure itself when phases are compared with the preceding IAM HO refinement step. It can be seen (Fig. 22) that the phase variations are largest in the low-resolution shells and show a general decrease with the resolution. The influence of the multipoles on the atomic scattering factors is more important at low resolution: the average value $|\Delta\varphi|$ is larger than 2° when $d > 0.8$ Å. In the highest resolution shells, the mean magnitude of phase variations upon transfer is very small, with $|\Delta\varphi| \simeq 1°$.

6. Conclusions and perspectives

A constrained multipolar refinement has been performed on a hAR substructure selected on the basis of residues with low thermal motion generally located in the protein core or close to the active site. The procedure applied is based both on macromolecular crystallography and on methods commonly used in charge-density analyses of small molecules. A dis-
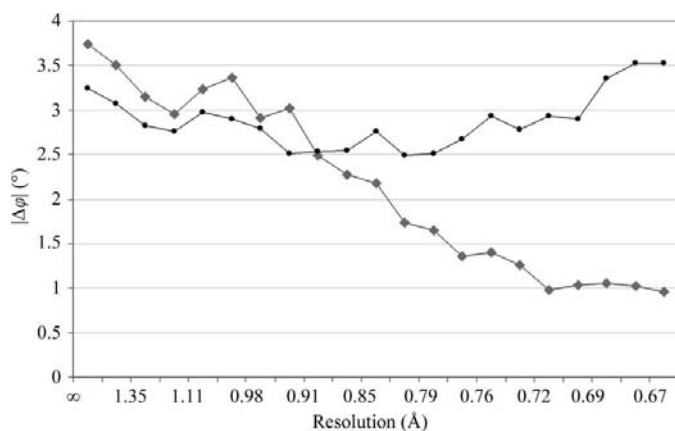
ordered solvent correction has been applied followed by a HO refinement of structural parameters. Refinement *versus* high-order reflections led to significant enhancement of the deformation electron density visible in residual Fourier maps. After the ELMAM library transfer to the atoms of the substructure, a constrained multipolar refinement was performed. Taking advantage of the repetition of chemical functional groups along the protein sequence, a chemical equivalence constraints scheme was defined, allowing a reduction of the number of refined parameters in the Hansen and Coppens multipole formalism.

The structural, thermal and multipolar atomic parameters were refined alternately until convergence. The resulting final molecular model showed excellent crystallographic agreement factors for a protein: $R(F) = 8.74\%$, $R_{\mathrm{free}}(F) = 9.23\%$ for all reflections with $I/\sigma(I) > 0$ and $R(F) = 7.89\%$, $R_{\mathrm{free}}(F) = 8.34\%$ when an $I/\sigma(I) > 4$ cutoff was applied. Analysis of the phases revealed average differences of 2–4° between the steps of the refinement procedure, demonstrating an obvious influence of the multipolar electron-density modelling.

After ELMAM multipolar refinement, the residual Fourier maps do not display any systematic remaining bond density. There is still significant residual electron density, but it is more randomly distributed and arises from the error in the measured diffraction intensities and from other systematic errors. The overall transfer and refinement procedure can be applied rather easily with the recent improvements of *MoPro* (Jelsch *et al.*, 2005). Protein crystallographers who happen to refine a structure at ultrahigh resolution have a general tendency to release the stereochemical restraints in the well ordered parts of the molecular structure in order to derive unbiased structural results. The application of the multipolar database approach, beyond the standard protein structure refinement, is recommended for accurate stereochemical analyses (Stec, 2007).

The inhibitor IDD594 was not included in the atom selection and was not involved in database transfer and multipolar refinement. It was therefore held fixed in the state obtained after the IAM HO refinement step. The modelling and refinement of the deformation electron density for the inhibitor deserves to be analyzed more thoroughly. Partially unconstrained refinements of the electron density performed on active-site residues, IDD594 inhibitor and NADP$^+$ cofactor will be the next stage of the hAR charge-density study. This investigation will be the subject of a forthcoming paper and will be performed using the result of the ELMAM multipolar refinement as the starting molecular model.

Figure 22
Absolute value of weighted phase variations as a function of resolution examined between the following steps: HO refinement and the transfer procedure (grey diamonds) and IAM and ELMAM multipolar refinements (black circles).

References

Abramov, Y. A., Volkov, A. V. & Coppens, P. (1999). *Chem. Phys. Lett.* **311**, 81–86.

Afonine, P. V., Grosse-Kunstleve, R. W. & Adams, P. D. (2005). *Acta Cryst.* D**61**, 850–855.

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

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

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

Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.

Blakeley, M., Federico Ruiz, F., Cachau, R., Hazemann, I., Meilleur, F., Mitschler, A., Ginell, S., Afonine, P.,Ventura, O. N., Cousido-Siah, A., Haertlein, M., Joachimiak, A., Myles, D. & Podjarny, A. (2008). *Proc. Natl Acad. Sci.* **105**, 1844–1848.

Bohren, K. M., Grimshaw, C. E., Lai, C. J., Harrison, D. H., Ringe, D., Petsko, G. A. & Gabbay, K. H. (1994). *Biochemistry*, **33**, 2021–2032.

Bönisch, H., Schmidt, C. L., Bianco, P. & Ladenstein, R. (2005). *Acta Cryst.* D**61**, 990–1004.

Brock, C. P., Dunitz, J. D. & Hirshfeld, F. L. (1991). *Acta Cryst.* B**47**, 789–797.

Brünger, A. T. (1992). *Nature (London)*, **355**, 472–474.

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. IV*, **10**, Pr10-3–Pr10-13.

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

Clementi, E. & Roetti, C. (1974). *At. Data Nucl. Data Tables*, **14**, 177–478.

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

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

Dauter, Z. (2003). *Methods Enzymol.* **368**, 288–337.

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

Dittrich, B., Hübschle, C. B., Luger, P. & Spackman, M. A. (2006). *Acta Cryst.* D**62**, 1325–1335.

Dittrich, B., Hübschle, C. B., Messerschmidt, M., Kalinowski, R., Girnt, D. & Luger, P. (2005). *Acta Cryst.* A**61**, 314–320.

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

Dittrich, B., Munshi, P. & Spackman, M. A. (2007). *Acta Cryst.* B**63**, 505–509.

Dominiak, P. M., Volkov, A., Li, X. & Coppens, P. (2007). *J. Chem. Theory Comput.* **3**, 232–247.

El-Kabbani, O., Carbone, V., Darmanin, C., Oka, M., Mitschler, A., Podjarny, A., Schulze-Briese, C. & Chung, R. P. T. (2005). *J. Med. Chem.* **48**, 5536–5542.

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

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

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

Giannoukakis, N. (2003). *Curr. Opin. Investig. Drugs*, **4**, 1233–1239.

Giannoukakis, N. (2006). *Curr. Opin. Investig. Drugs*, **7**, 916–923.

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

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

Hakanpää, J., Linder, M., Popov, A., Schmidt, A. & Rouvinen, J. (2006). *Acta Cryst.* D**62**, 356–367.

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

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

Hodel, A., Kim, S.-H. & Brünger, A. T. (1992). *Acta Cryst.* A**48**, 851–858.

Hotta, N., Akanuma, Y., Kawamori, R., Matsuoka, K., Oka, Y., Shichiri, M., Toyota, T., Nakashima, M., Yoshimura, I., Sakamoto, N. & Shigeta, Y. (2006). *Diabetes Care*, **29**, 1538–1544.

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

Jaskolski, M., Gilski, M., Dauter, Z. & Wlodawer, A. (2007). *Acta Cryst.* D**63**, 611–620.

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.* D**54**, 1306–1318.

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

Jiang, J. S. & Brünger, A. T. (1994). *J. Mol. Biol.* **243**, 100–115.

Kang, B. S., Devedjiev, Y., Derewenda, U. & Derewenda, Z. S. (2004). *J. Mol. Biol.* **338**, 483–493.

Ko, T.-P., Robinson, H., Gao, Y. G., Cheng, C. H., DeVries, A. L. & Wang, A. H. (2003). *Biophys. J.* **84**, 1228–1237.

Lamour, V., Barth, P., Rogniaux, H., Poterszman, A., Howard, E., Mitschler, A., Van Dorsselaer, A., Podjarny, A. & Moras, D. (1999). *Acta Cryst.* D**55**, 721–723.

Lario, P. I. & Vrielink, A. (2003). *J. Am. Chem. Soc.* **125**, 12787–12794.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Lecomte, C., Guillot, B., Jelsch, C. & Podjarny, A. (2005). *Int. J. Quant. 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., Guillot, B., Vercauteren, D. P., Pichon-Pesme, V., Jelsch, C., Lagoutte, A. & Lecomte, C. (2006). *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*, edited by R. J. Boyd & E. Matta, pp. 285–316. Weinheim: Wiley-VCH.

Li, X., Volkov, A. V., Szalewicz, K. & Coppens, P. (2006). *Acta Cryst.* D**62**, 639–647.

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

Merritt, E. A. & Murphy, M. E. P. (1994). *Acta Cryst.* D**50**, 869–873.

Moews, P. C. & Kretzinger, R. H. (1975). *J. Mol. Biol.* **91**, 201–225.

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

Nakahara, M., Miyata, K., Otani, S., Miyai, T., Nejima, R., Yamagami, S. & Amano, S. (2005). *Br. J. Ophthalmol.* **89.** 266–268.

Oka, M. & Kato, N. (2001). *J. Enzyme Inhib.* **16**, 465–473.

Petrova, T. & Podjarny, A. (2004). *Rep. Prog. Phys.* **67**, 1565–1605.

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

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

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

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. Opin. Struct. Biol.* **12**, 698–703.

Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.

Singh, S. B., Malamas, M. S., Hohman, T. C., Nilakantan, R., Carper, D. A. & Kitchen, D. (2000). *J. Med. Chem.* **43**, 1062–1070.

Guillot *et al.* · Charge-density analysis of human aldolase reductase **587**

Slater, J. C. (1932). *Phys. Rev.* **42**, 33–43.

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

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.

Stec, B. (2007). *Acta Cryst.* D**63**, 1113–1114.

Stewart, R. F. (1969). *J. Chem. Phys.* **51**, 4569–4577.

Tarle, I., Borhani, B. W., Wilson, D. K., Quiocho, F. A. & Petrash, J. M. (1993). *J. Biol. Chem.* **268**, 25687–25693.

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

Volkov, A., Koritsanszky, T., Li, X. & Coppens, P. (2004). *Acta Cryst.* A**60**, 638–639.

Volkov, A., Li, X., Koritsanszky, T. & Coppens, P. (2004). *J. Phys. Chem. A*, **108**, 4283–4300.

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

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

Yabe-Nishimura, C. (1998). *Pharmacol. Rev.* **50**, 21–34.

Yáśubnez, M., Stewart, R. F. & Pople, J. A. (1978). *Acta Cryst.* A**34**, 641–648.

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