Towards the charge-density study of proteins: a room-temperature scorpion-toxin structure at 0.96 Å resolution as a first test case

Dominique Housset, Farid Benabicha, Virginie Pichon-Pesme, Christian Jelsch, Andreas Maierhofer, Sylvain David, Juan-Carlos Fontecilla-Camps, Claude Lecomte

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Towards the charge-density study of proteins: a room-temperature scorpion-toxin structure at 0.96 Å resolution as a first test case

Dominique Housset, Farid Benabicha, Virginie Pichon-Pesme, Christian Jelsch, Andreas Maierhofer, Sylvain David, Juan Carlos Fontecilla-Camps and Claude Lecomte

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Towards the charge-density study of proteins: a room-temperature scorpion-toxin structure at 0.96 Å resolution as a first test case

Dominique Housset,[b] Farid Benabicha,[a,c] Virginie Pichon-Pesme,[a] Christian Jelsch,[a] Andreas Maierhofer,[b] Sylvain David,[b] Juan Carlos Fontecilla-Camps[b] and Claude Lecomte[a]*

[a]Laboratoire de Cristallographie et Modélisation des Matériaux Minéraux et Biologiques, UPRESA CNRS 7036, Université Henri Poincaré, Nancy I, Faculté des Sciences, BP 239, F-54506 Vandoeuvre-lès-Nancy CEDEX, France, [b]Laboratoire de Cristallographie et Cristallogenèse des Protéines, Institut de Biologie Structurale 'Jean-Pierre Ebel', CEA-CNRS, 41 Rue Jules Horowitz, F-38027 Grenoble CEDEX 1, France, and [c]Département de Physique, Université Moulay Ismail, Faculté des Sciences, Meknès, Morocco

Correspondence e-mail: lecomte@lcm3b.u-nancy.fr

Received 26 July 1999
Accepted 19 November 1999

The number of protein structures refined at a resolution higher than 1.0 Å is continuously increasing. Subatomic structures may deserve a more sophisticated model than the spherical atomic electron density. In very high resolution structural studies (d < 0.5 Å) of small peptides, a multipolar atom model is used to describe the valence electron density. This allows a much more accurate determination of the anisotropic thermal displacement parameters and the estimate of atomic charges. This information is of paramount importance in the understanding of biological processes involving enzymes and metalloproteins. The structure of the scorpion *Androctonus australis* Hector toxin II has been refined at 0.96 Å resolution using synchrotron diffraction data collected at room temperature. Refinement with a multipolar electron-density model in which the multipole populations are transferred from previous peptide studies led to the observation of valence electrons on covalent bonds of the most ordered residues. The refined net charges of the peptide-bond atoms were of the correct sign but were underestimated. Such protein-structure refinements against higher resolution data collected at cryogenic temperature will enable the calculation of experimental atomic charges and properties such as electrostatic potentials.

1. Introduction

Ten years ago, the pioneering protein structures refined at atomic resolution (1 Å) were of very limited number and were confined to small proteins such as rubredoxin (Watenpaugh *et al.*, 1979), crambin (Teeter & Hendrickson, 1979), insulin (Sakabe *et al.*, 1981) or BPTI (Wlodawer *et al.*, 1984) and gramicidin (Langs, 1988). Quite recently, crystals of several proteins, some of which have more than 500 residues, have been obtained that diffract to 1 Å or less (Dauter *et al.*, 1995; Dauter, Lamzin *et al.*, 1997; Longhi *et al.*, 1998 for reviews). The reasons for this are essentially the improvement of detectors, the use of synchrotron radiation and the cryo-cooling of the crystals, allowing collection of a complete high-resolution diffraction data set from a single crystal in a day or so. The atomic resolution data will obviously provide several methodological advances for protein structures such as automated refinement (Lamzin & Wilson, 1993) and *ab initio* structure determination, which has been successfully performed on the *A. australis* Hector scorpion toxin II (AaHII; Smith *et al.*, 1997) and on lysozyme (Deacon *et al.*, 1998). The biologically relevant benefits of the very high resolution have, however, also to be pointed out. To date, the protein structures that have been refined at a resolution of 1.0 Å or less have provided much more detailed protein

models. A much better agreement between observed and calculated structure factors is obtained (R factor around 10% or less), while the high observation-to-parameter ratio allows a much more accurate estimate of the atomic coordinates (typical errors of 0.02 to 0.03 Å for protein atoms, 0.003 Å for S or heavy atoms; Dauter, Wilson *et al.*, 1997). As the resolution of the crystal structure increases, many side-chain alternate conformations appear and can be modelled (Genick *et al.*, 1998). For instance, up to 30% of the residues present such a discrete disorder in the structure of the protein crambin at 0.83 Å resolution (Stec *et al.*, 1995). Conformations with occupancy as low as 25% can be accurately determined. The chemical types (C, N, O, S) of atoms can be identified more precisely in the electron density. H atoms show up in the electron density (Kuhn *et al.*, 1998), allowing the determination of the protonation state for well ordered residues (histidines, carboxylate groups *etc.*) and the analysis of specific hydrogen bonds. This allows the determination of the exact conformation of residues such as histidines, asparagines or glutamines. Finally, the refinement of ADPs has provided information on the correlated movement of rigid bodies inside the protein. These results obtained from atomic resolution protein structures give us a taste of the details that can be observed in the electron density. To go further and seek a more quantitative analysis of the electron density, a more sophisticated model of the atomic electron density is needed. It should take full advantage of the large amount of data provided by an accurate atomic resolution data set and help to deconvolute the ADPs from the electron-density deformation. Hence, when using a spherical model of the atomic electron density, the ADPs account for all non-spherical features present in the electron density (hybridization of valence electrons, electron lone pairs, d-orbital population *etc.*).

At subatomic resolutions, typically 0.6 Å (or $\sin\theta/\lambda > 0.8$ Å^{-1}), the atomic charge and the non-spherical character of the electron density are observable and can be quantified (Coppens, 1997, 1998; Lecomte, 1995). Owing to chemical bonding and atom–atom interactions, the atomic electron density is not spherical and this deformation density can be mapped from accurate low-temperature X-ray diffraction experiments. Furthermore, taking into account the asphericity of the electron density should, in principle, allow the deconvolution of ADPs ($U_{ij} = B_{ij}/8\pi^2$) from the atomic deformation density (see §2).

The scope of this study is to model aspherical features of the atomic electron density arising from chemical bonding for the first time on room-temperature data in a small protein by transferring the electron-density multipole parameters obtained from accurate X-ray diffraction experiments on amino acids and small peptides (Pichon-Pesme *et al.*, 1995). Previous tests on an octapeptide with a helical structure (Jelsch *et al.*, 1998) have shown the limitations and the feasibility of such a refinement: one can expect improvement in the refinement if the resolution is better than 0.9 Å for the atoms that have an equivalent B factor of lower than 5 Å^2. The scorpion-toxin data collected at 0.96 Å resolution at room temperature seem to be a limiting case for the application of the method for the first time to a protein and for the identification of the fine details that can be observed in the electron density (B factors are on average 8 Å^2 for the non-disordered parts of the molecule, with a minimum of 5 Å^2).

2. Materials and methods

2.1. Purification and crystallization

The protein (molecular weight 7141 Da) was purified following the protocol described by Miranda *et al.* (1970). The crystals were grown by slow evaporation at 277 K of a solution containing 5–10 mg ml^{-1} toxin II in 0.2 M ammonium acetate (pH 6.8). The crystals belong to the orthorhombic space group $P2_12_12_1$ (unit-cell parameters $a = 45.9$, $b = 40.7$, $c = 30.1$ Å) and have one toxin molecule per asymmetric unit (Fontecilla-Camps *et al.*, 1988).

2.2. Data collection and processing

The crystallographic data were collected at the DESY synchrotron (Hamburg, Germany). Two crystals (0.4 × 0.4 × 0.6 mm) were mounted in glass capillaries and used for data collection at room temperature (289 K). One data set was collected from one crystal, using the X11 beamline ($\lambda = 0.92$ Å) and a 30 cm MAR Research imaging-plate detector, to a resolution of 1.11 Å. Two data sets (with different exposure times in order to obtain both low- and high-resolution shells) were collected on another crystal, using the X31 beamline ($\lambda = 0.72$ Å) and a 22 cm MAR Research imaging-plate detector, to a resolution of 0.96 Å. The data processing was performed with the *MOSFLM* package version 5.1 (Leslie, 1992) and the *CCP*4 (Collaborative Computational Project, Number 4, 1994) program suite (*ROTAVATA* and *AGROVATA* programs). The three data sets have been merged to provide a 90.9% complete set of unique reflections in the resolution range 10.0–0.96 Å. The anomalous dispersion of the S atom was not taken into account. The data statistics are summarized in Table 1. No absorption correction for the protein crystal, the capillary or the solvent was performed.

2.3. Refinement of the structure

2.3.1. Refinement with *SHELXL*97. The refinement of the structure was initiated with the *SHELXL*97 program (Sheldrick & Schneider, 1997), using the structure refined at 1.3 Å resolution (Housset *et al.*, 1994) as a starting model. In order to assess the quality of the refinement by computation of the free R factor, a randomly selected 10% of the reflections were also omitted during the refinement (Brünger, 1992). Major model modifications were performed manually on a graphics workstation with the program *O* (Jones *et al.*, 1991). No van der Waals antibumping restraints were applied. In order to obtain a good estimate of the structural information contained in this high-resolution crystallographic data, the refinement was performed in three successive steps.

Step I is a standard macromolecular refinement with x, y, z and the isotropic temperature factor as atomic parameters and stereochemical restraints. H atoms were not included. The

Table 1
Data-collection statistics.

Values for the last shell are given in parentheses.

	Resolution (Å)	R_{merge}	$\langle I/\sigma(I)\rangle$	No. of measurements	No. of reflections	Completeness (%)
First crystal (X11, λ = 0.92 Å)	3.5–1.11	0.052 (0.133)	8.3 (4.9)	72171	20089	91 (81)
Second crystal	2.8–0.96	0.102 (0.543)	6.0 (1.3)	96456	30451	89 (80)
(X31, λ = 0.72 Å)	15.8–1.75	0.091 (0.224)	4.5 (3.0)	18174	5774	97 (87)
Merged data	10.0–0.96	0.073	—	—	32890	90.9 (75.6)

model consists of the entire polypeptide chain as well as 85 solvent sites. The disulfide bridge Cys12–Cys63, the carbonyl group of Gly17 and, to a lesser extent, the Asp9, Glu24 and Arg56 side chains showed clear two-state static disorder in the electron-density maps. Accordingly, these residues were modelled with two conformations (Fig. 1*d*). Three water molecules also had alternate sites. A bulk-solvent model was used as implemented in *SHELXL*97 (SWAT option). At this stage, although no extra water molecules could be unambiguously identified, some signal remained in the residual $(F_{obs} - F_{sph}^{iso}, \varphi_{sph}^{iso})$ electron-density map.

Step II. According to the residual electron-density map, explicit H atoms were then added to the model resulting from refinement step I. The H-atom positions were calculated using the 'riding-atom' model as implemented in the program *SHELXL*97. These H atoms used for structure-factor calculation did not increase the number of parameters to be refined.

Step III. Using the model resulting from the previous step, a refinement including x, y, z and the anisotropic displacement parameters (ADPs) as variables with stereochemical restraints and explicit H atoms was performed. A rigid-bond restraint (Hirshfeld, 1976) that renders the mean-square displacement along a bond direction similar for two covalently bonded atoms ($\sigma_{\Delta Z}$ = 0.001 Å^2) was applied. The significant improvement of the residual electron-density map allowed the modelling of a few more solvent sites and the occupancies of water molecules with low electron density were refined. There is a total of 121 water molecules in the structure; all except four are visible in the $(2mF_{obs} - DF_{sph}, \varphi_{sph})\sigma_A$ (Read, 1986) electron density with a 1σ contour level; the residual $(mF_{obs} - F_{sph}, \varphi_{sph})$ map has no peaks above 0.34 e Å^{-3} (Table 2). The final model was further refined using the full-matrix method without restraints on the B factor or on the stereochemistry in order to assess the errors in the estimated parameters.

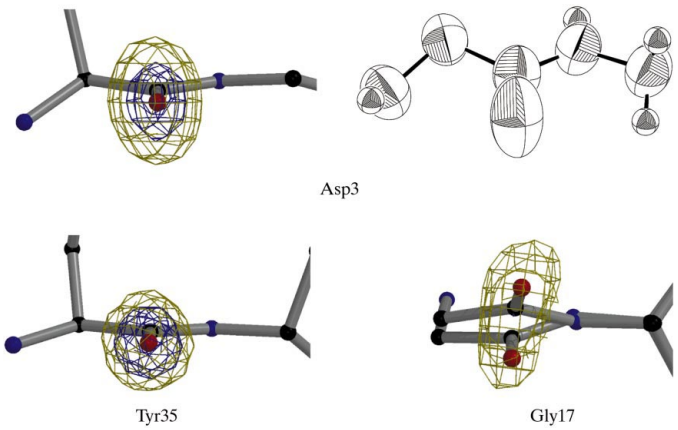

Figure 1
Anisotropic displacement of the O atom of the peptide-bond carbonyl group. Motion perpendicular to the amide plane is observed for most peptide bonds. The $2mF_{obs} - DF_{calc}$ electron density is shown contoured in blue at the 4σ level and in green at the 1.2σ level. For Asp3, this displacement is well described by an anisotropic B factor. The thermal ellipsoid is shown. Tyr35 has the lowest B value. For Gly17, the displacement has the largest amplitude and is modelled by two distinct conformations.

2.3.2. Refinement with *MOLLY*. The refinement was then continued with *MOLLY* in order to apply the non-spherical atomic electron-density model to the toxin structure. This multipole formalism was developed 20 years ago for small-molecule accurate high-resolution X-ray data to describe precisely the atomic electron density ρ_{atom} (Hansen & Coppens, 1978),

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

The term ρ_{core} represents the spherically symmetric Hartree–Fock core-electron density and ρ_{val} represents the spherically averaged free-atom Hartree–Fock valence-electron density. The last term describes the non-spherical electron density arising from bonded and intermolecular interactions. κ and κ' are the expansion/contraction coefficients of the spherical and multipolar valence densities, respectively. A κ value greater than one means a contraction of the density and therefore a smaller atomic volume; a value smaller than one corresponds to an expansion of the electron's shell. In this density model, the refined valence population P_{val} gives an estimation of the net atomic charge q in the atomic volume with respect to the number of electrons N_{val} in the free-atom valence orbitals: $q = N_{val} - P_{val}$. The y_{lm} are spherical harmonic functions of order l in real form and R_l are Slater-type radial functions. The P_{lm} are the multipolar population parameters of each associate y_{lm} spherical harmonic.

This model implies the least-squares estimation of 18 supplementary parameters per atom [P_{val}, κ, κ', three dipoles, five quadrapoles and seven octapoles (if l_{max} = 3)] which cannot be obtained from our restricted number of protein stucture factors even in a low-temperature and higher resolution experiment. Therefore, the multipole density coefficients are obtained by applying the notion of transferability. A database of experimental electron-density parameters (P_{lm}, P_{val}) is being built from accurate X-ray analyses of peptide crystals in the LCM3B (Nancy) (Pichon-Pesme *et al.*, 1995). At present, the database includes results from ultra

Table 2
Refinement statistics.

ADPs are given as B factors for comparison with other protein structures ($B_{ij} = 8\pi^2 U_{ij}$).

(a) *SHELXL* program

	Step I	Step II	Step III
No. of protein atoms	513	513 + 446 H	560 + 483 H
No. of solvent atoms	85	85	121
No. of reflections used	32847	32827	32827
d (Å)	16.0–0.96	16.0–0.96	16.0–0.96
No. of parameters	2400	2438	5730
No. of restraints	2164	2174	2588
R factor (F)†			
All data	0.188	0.181	0.118
Free	0.205	0.197	0.136
1.00–0.96 Å	0.319	0.317	0.263
$F > 4\sigma(F)$	0.170	0.163	0.101
Free	0.185	0.178	0.118
wR factor (I)‡			
All data	0.475	0.465	0.302
Residual map ($F_o - F_c$) (e Å^{-3})			
Min	−0.52	−0.51	−0.33
Max	0.60	0.59	0.34
σ	0.11	0.10	0.07
Stereochemistry (r.m.s. deviation from ideal values)			
Bond distances (Å) (σ = 0.013 Å)	0.008	0.009	0.011
Angle distances (Å) (σ = 0.025 Å)	0.024	0.024	0.026
ADP rigid-bond restraints (Å^2) (σ_B = 0.08 Å^2)	—	—	0.032

(b) *MOLLY* program

	Step IV	Step V	Step VI
No. of reflections used, $I > 3\sigma(I)$	24269	24269	24269
d (Å)	3.3–0.96	3.3–0.96	3.3–0.96
No. of parameters	4330	4330	15
No. of restraints	1872	1872	—
R factor (F)	0.090	0.091	0.096
wR factor (F)	0.093	0.095	0.102
Free R factor (F)	0.107	0.108	0.114

† $R(F) = \sum |F_{obs} - F_{calc}| / \sum |F_{obs}|$, $wR(I) = \{\sum[w(F^2_{obs} - F^2_{calc})^2] / \sum[w(F^2_{obs})^2]\}^{1/2}$.
‡ $wR(F) = [\sum w(|F_{obs}| - |F_{calc}|)^2 / \sum w(F_{obs})^2]^{1/2}$, $w = 1/\sigma^2(F_{obs})$.

high resolution ($d < 0.45$ Å) studies of amino-acid and peptide crystals. The database analyses have shown that, as we would expect from the quasi-constancy of properties of chemical functional groups, there is a high degree of transferability of the experimental electron-density multipole parameters P_{lm} ($l \neq 0$) for atoms of the same chemical type. For the structure-factor computations, aspherical scattering factors were derived from the database and the ρ_{val} density was calculated from Clementi & Raimondi (1963) wave functions and from Stewart *et al.* (1965) for the non-H and H atoms, respectively. The real and imaginary dispersion corrections to the form factors were given by Cromer (1974). The H atoms were constrained to their ideal position as in *SHELXL*97.

To ensure control of the geometry of the structure, some stereochemical restraints have been implemented in *MOLLY* (Jelsch *et al.*, in preparation). Distance restraints were applied to neighbouring atoms separated by one or two covalent bonds. The target ideal distances are the same as in the *SHELXL*97 stereochemistry dictionary. The rigid-bond restraint (Hirshfeld, 1976) was added to *MOLLY* with the same target as in step III. Three successive refinement steps were carried out.

Step IV. Firstly, a refinement with *MOLLY* using a spherical neutral atom model ($P_{val} = N_{val}$, $P_{lm} = 0$) was performed where the x, y, z coordinates and anisotropic displacement parameters factors U_{ij} for all non-H atoms were refined. This step was basically equivalent to step III, the major difference being the program used to perform it; this allowed a comparison of the two programs and served as a reference for the comparison with the next steps, avoiding possible misinterpretation arising from slightly different refinement procedures.

Step V. A non-spherical description of the atomic density was then introduced by transfer of the multipole parameters from our database. All atom types present in the neurotoxin were modelled from the database, except for the histidine imidazole ring. The H-atom positions were adjusted by moving the H atoms outward along the C—H, N—H and O—H bond directions to bond lengths equal to average values from neutron-diffraction studies (C—H, 1.08 Å; N—H, 1.03 Å; O—H, 0.96 Å; Allen, 1986). Only the coordinates and ADPs of the non-H atoms were refined. The multipole parameters P_{lm} from the database (Pichon-Pesme *et al.*, 1995) were transferred but not refined. The atomic charges were kept neutral ($P_{val} = N_{val}$) and no expansion/contraction refinement was made.

Step VI. As the atomic positions and ADPs obtained at the end of the refinement step V should be the best estimates, these parameters were used and were kept fixed during refinement step VI. A charge P_{val}, κ refinement [$P_{lm} = 0$, equation (1)] was then performed on chemically equivalent atoms. This calculation permits an estimation of the net charge of each individual atom without taking into account the non-spherical shape of the valence-electron distribution (Coppens *et al.*, 1979).

2.4. Electron-density maps

Two kinds of electron-density maps were used for our analysis of the different models. The residual electron density was calculated as

$$\Delta\rho_{res}(r) = V^{-1} \sum_h [k^{-1} F_{obs}(h) - F_{sph}(h)] \exp i(\varphi_{sph} - 2\pi \mathbf{h} \cdot \mathbf{r}), \quad (2)$$

where F_{sph} and φ_{sph} are the calculated structure factor and phase, respectively, derived from the structure refined with the spherical neutral atom density model. The same electron-density map after multipolar refinement (step V) was computed with the coefficients ($F_{obs} - F_{mult}$, φ_{mult}). The residual maps reveal the agreement between the observed and calculated structure factors and also the failure of the structural model, including its atomic electron density.

The deformation electron-density maps were computed at the end of refinement step V using the equation

$$\Delta\rho_{\rm def}(r) = V^{-1}\sum_h [k^{-1}F_{\rm obs}(h)\exp(i\varphi_{\rm mult}) - F_{\rm sph}(h)\exp(i\varphi_{\rm sph})]\exp(-2i\pi\mathbf{h}\cdot\mathbf{r}). \quad (3)$$

The phases $\varphi_{\rm mult}$ were calculated using the multipolar atom model, whereas spherical atomic scattering factors were used in the calculation of $F_{\rm sph}$ and $\varphi_{\rm sph}$. This electron-density map shows the difference between the best estimate of the protein electron density ($F_{\rm obs}$, $\varphi_{\rm mult}$) based on the multipolar refinement and the density calculated with the spherical atomic electron-density model ($F_{\rm sph}$, $\varphi_{\rm sph}$). This map is a depiction of the rearrangement of the electron density owing to interatomic bonding.

2.5. Rigid-body TLS analysis of ADPs

The analysis of ADPs has been performed using the program *THMA*11 (Trueblood, 1990), which we adapted to protein PDB files. The rigid-body motion is described by three tensors *T, L* and *S* accounting for translation, libration and correlation between translation and libration, respectively, of the rigid group defined by the user. These tensors are determined by applying a least-squares fit between calculated and observed U_{ij} (Shomaker & Trueblood, 1968). The analysis has been performed on all (rigid) aromatic residues of the toxin and the results may be checked by calculating an *R* factor,

$$R_2 = (\sum_{ij}|U_{ij} - U_{ij}^{\rm TLS}|^2 / \sum_{ij}|U_{ij}|^2)^{1/2}.$$

A low value of this *R* factor indicates that the rigid-body behaviour of the selected group of atoms is well described by the refined ADPs.

3. Results

3.1. Refinement with *SHELXL*97

Refinement with *SHELXL*97 (Sheldrick & Schneider, 1997) has been performed in three steps (see §2). At the end of step I, the whole polypeptide chain as well as most of the solvent sites were determined. Many positive peaks of the residual map ($F_{\rm obs} - F_{\rm sph}^{\rm iso}$, $\varphi_{\rm sph}^{\rm iso}$) were located near expected H-atom positions. Some other weakly positive peaks (2σ) were observed in the middle of conjugated bonds in the less flexible regions, possibly corresponding to valence electrons. However, the fairly noisy residual map (Table 2) is a consequence of the inadequacy of the isotropic model of thermal motion to fit the observed data, especially at high resolution. This fact is well supported by a significantly high *R* factor. The very small difference between the *R* factor and the free *R* factor illustrates the high observation-to-parameter ratio.

Refinement step II confirmed the contribution of the large number of H atoms that were observed in the residual electron density of step I. Adding all H atoms in the model using the 'riding-atom' method (*i.e.* no additional parameters in the refinement) improved both the *R* factor and the free *R* factor (Table 2). The electron density also supports well the use of explicit H atoms in the structure-factor calculation.

Table 3
Analysis of ADPs after *SHELXL*97 refinement step III.

Hirschfeld's rigid-bond criteria: $\Delta Z_B = 8\pi^2\Delta Z_U < 0.08$ Å^2. $R_2 = (\sum_{ij}|U_{ij} - U_{ij}^{\rm TLS}|^2 / \sum_{ij}|U_{ij}|^2)^{1/2}$.

	Without U_{ij} restraints			With rigid-bond U_{ij} restraints		
Residue	R_2 (%)	$\langle B\rangle$ (Å^2)	$\langle\Delta Z_B\rangle$ (Å^2)	R_2 (%)	$\langle B\rangle$ (Å^2)	$\langle\Delta Z_B\rangle$ (Å^2)
Tyr5	7.1	8.90	0.95	5.9	8.87	<0.08
Tyr14	8.6	9.88	1.03	7.8	9.87	<0.08
Phe15	12.2	13.54	3.71	10.0	13.58	<0.08
Tyr21	11.1	10.65	2.37	8.0	10.63	<0.08
Tyr35	12.2	8.24	1.89	8.0	8.21	<0.08
Trp38	17.7	12.38	2.61	14.5	12.38	<0.08
Tyr42	14.2	13.90	2.92	12.1	13.77	<0.08
Tyr47	8.3	7.76	0.55	7.8	7.73	<0.08
Tyr49	7.4	7.35	1.18	6.5	7.35	<0.08

The purpose of refinement step III was to obtain the best estimate of the atomic anisotropic *B* factors using a spherical atomic electron-density model. A dramatic decrease in both the *R* and free *R* factors was observed (Table 2), supporting the relevance of such a refinement at this resolution. Some main-chain carbonyl O atoms show a clear anisotropic thermal ellipsoid, with the longest axis perpendicular to the peptide-bond plane (Fig. 1). The improvement of the electron-density maps allowed the localization of 36 extra solvent sites. The few residual electron-density peaks that were observed in the middle of conjugated bonds after refinement step I are no longer present. When the refinement was performed without rigid-bond restraints on ADPs, the 'rigid-bond' test (Hirshfeld, 1976) was no longer satisfied and no significant decrease in the *R* factor was observed. TLS analysis (Trueblood & Dunitz, 1983) of the nine aromatic rings confirms that the anisotropic *B* factors quite fairly reflect the dynamic behaviour of these groups. The disagreement factors R_2 (see §2) between U_{ij} and $U_{ij}^{\rm TLS}$ fall in the range 5.9 –14.5%. If the rigid-bond restraints on the U_{ij} are not used, the R_2 value increases slightly (Table 3). In this case, it appears that the ADP values refined do not only account for the atomic displacements. The error associated with the ADP estimate may be a consequence of a limited signal-to-noise ratio in the atomic resolution shell and of systematic errors in the diffraction data. The ADPs also incorporate non-spherical features in the electron density; an aspherical electron-density model was therefore tested.

3.2. Structure refinement with the program *MOLLY*

The possibility of transferring average charge-density parameters has already been applied successfully to 100 K and room-temperature data from a tripeptide structure (Pichon-Pesme *et al.*, 1995) and to 100 K data from an octapeptide helix (Jelsch *et al.*, 1998). Table 2 gives the disagreement factors of the refinements. As the *R* factors are only slightly better, but are for a smaller subset of data (Table 2), refinements IV and V are statistically close (see §2). Thus, it can be concluded that the more realistic non-spherical model of atomic electron density does not provide a significant overall improvement in the agreement between $F_{\rm obs}$ and $F_{\rm calc}$ for the toxin structure.

This is a consequence of both the limited resolution (0.96 Å) and the high average B factor (8 Å^2) compared with small-molecule structures (Jelsch *et al.*, 1998). However, it is important to notice that in the most ordered regions of the structure (Fig. 2) where the B factor is in the 5–6 Å^2 range, aspherical features of the atomic electron density are being observed. These regions include Tyr47 and peptide bonds 47–48, on which we have focused our further analysis.

3.2.1. Residual electron-density maps [equation (2)]. Refinement step IV confirms what was observed with *SHELXL*97 in step III. The residual electron-density map shows randomly distributed peaks rather than significant residual deformation density, as shown for Tyr47 in Fig. 3(*a*). The same map, computed by averaging the electron density over the 19 most ordered peptide planes of the toxin, does not show any significant signal at the 0.02 e Å^{-3} level either. This is likely to be because of the least-squares fit of the x, y, z and U_{ij} parameters, which takes into account most of the bonding density by overestimating the U_{ij} parameters (Coppens, 1967). The approximation of the spherical atom model is widely compensated by the too powerful least-squares fit of the six U_{ij} parameters per atom. The fraction of non-spherical electron density not taken into account by the U_{ij} is too small to be seen in the residual map. Consequently, for the toxin structure, the deconvolution of the deformation electron density and of the thermal motion should be performed only by transfer of multipole parameters.

3.2.2. Deformation electron-density maps [equation (3)]. After step V, the transfer-deformation density (see §2) of the protein may provide a more sensitive tool for detecting the valence electrons. In the regions of lowest ADP, the deformation density shows peaks in the middle of covalent bonds as shown for the Tyr47 ring in Fig. 3(*b*). On this map, all density peaks can be located and centred on the atomic bonds; the deformation density peaks on the C—C bonds of the ring vary between 0.04 and 0.18 e Å^{-3}. This is small compared with the 0.50 e Å^{-3} peaks observed in the LBZ helix octapeptide at cryogenic temperature after multipoles transfer (Jelsch *et al.*, 1998). A lack of deformation density on the covalent bond C$^{\zeta}$—O$^{\eta}$H can be noticed; this is a consequence firstly of a higher thermal motion at the end of the tyrosine side chain and secondly of the definition of the deformation density: as the O atom has more valence electrons, the superposition of C and O free-electron densities leads to more density being located in the bond direction than in the case of two free C atoms. Therefore, the reference state is different and the σ-deformation density is smaller in C—O bonds than in C—C bonds for the same interaction distance. This was initially observed by Dunitz & Seiler (1983) and also shown on the tyrosine residues of different peptide molecules (Pichon-Pesme *et al.*, 1992; Dahaoui *et al.*, 1998).

The peptide bond 47–48 is shown in Fig. 3(*c*). The deformation density peaks are visible on all covalent bonds, reaching a height of 0.18 e Å^{-3} on the C═O bond. The deformation density is less well defined on the N—H bond. The electron lone pairs, which require better resolution than covalent bonds because they are contracted in direct space, are also not resolved in the toxin room-temperature data set.

An average deformation density over the 19 less mobile peptide bonds of the toxin has also been computed and is shown in Fig. 3(*d*). The noise which was present in the previous individual deformation maps has completely vanished after the averaging process. The average deformation density peaks reach 0.12 e Å^{-3} on the C═O bond and the density is slightly expanded on the carbonyl O atom, which is reminiscent of the two electron lone pairs.

In conclusion, although the deformation density observed in these two considered planes is much lower than that usually observed in charge-density studies of peptides (Souhassou *et al.*, 1992), these results nevertheless show that at 0.96 Å resolution, even at room temperature, valence electrons can be observed in the most ordered regions.

3.2.3. Charges and κ refinement. Atomic charges are not directly observable physical quantities and must be obtained through a model (Mulliken, 1955), fitting of the electrostatic potential (Bayly *et al.*, 1993), Bader analysis (Bader & Essen, 1984) or X-ray diffraction analysis (Coppens *et al.*, 1979; Bouhmaida *et al.*, 1999). One of the advantages of the charge-density formalism is its ability to estimate these atomic charges with $q = N_{\rm val} - P_{\rm val}$, where $N_{\rm val}$ is the number of valence electrons in the neutral atom and $P_{\rm val}$ is the refined valence-shell population in the bound pseudoatom (1). In the case of a multipolar atom model, these charges cannot be considered as the total atomic charge, as the multipoles also contribute to the electron transfers from one atom to the other. An alternative charge-density partitioning is the charge and κ refinement of spherical atoms (Coppens *et al.*, 1979). We have carried out such a refinement (step VI) in which only the κ dilatation coefficients and the valence populations $P_{\rm val}$ are allowed to vary (see §2), the

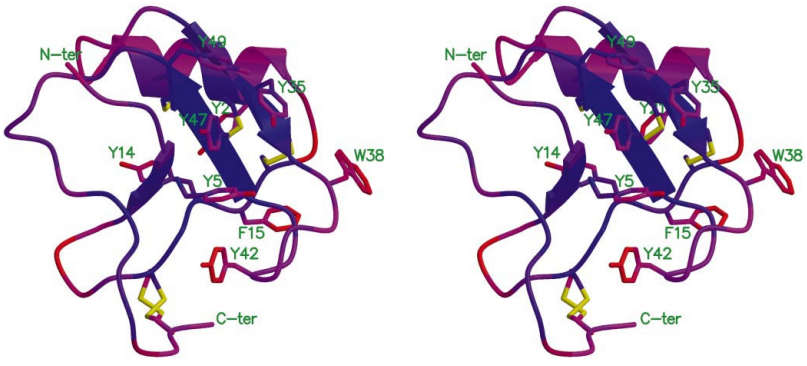

Figure 2
Stereoview of the overall structure of the toxin II. The backbone is coloured according to the temperature factor (colour ramp from blue for the lowest B values, 4 Å^2, to red for the highest B values, 15 Å^2). The aromatic residues shown in ball-and-stick are those used for the analysis of the most ordered regions. Disulfide bridges are shown in green. The β-sheet region (residues 37–37 and 45–49) appears to have a low flexibility, being well stabilized by two disulfide bridges.

positional and temperature factors being fixed. We estimate the atomic charges on the peptide group by taking advantage of the averaging over the 63 peptide bonds present in the structure. The number of parameters has been reduced to a minimum by refining κ and P_{val} for only eight different atom types: the H atom κ was fixed to 1.16 (Stewart *et al.*, 1965). The refined atomic charges (Table 4) are generally of the sign expected from chemical knowledge: H atoms are positively charged, whereas O and N atoms are negatively charged. However, the charges obtained for the polar moieties of the peptide bond (C═O and N—H) are underestimated compared with standard values found in small peptides (Pichon-Pesme & Lecomte, 1998) and in charge dictionaries. This result demonstrates that the multipolar formalism is able

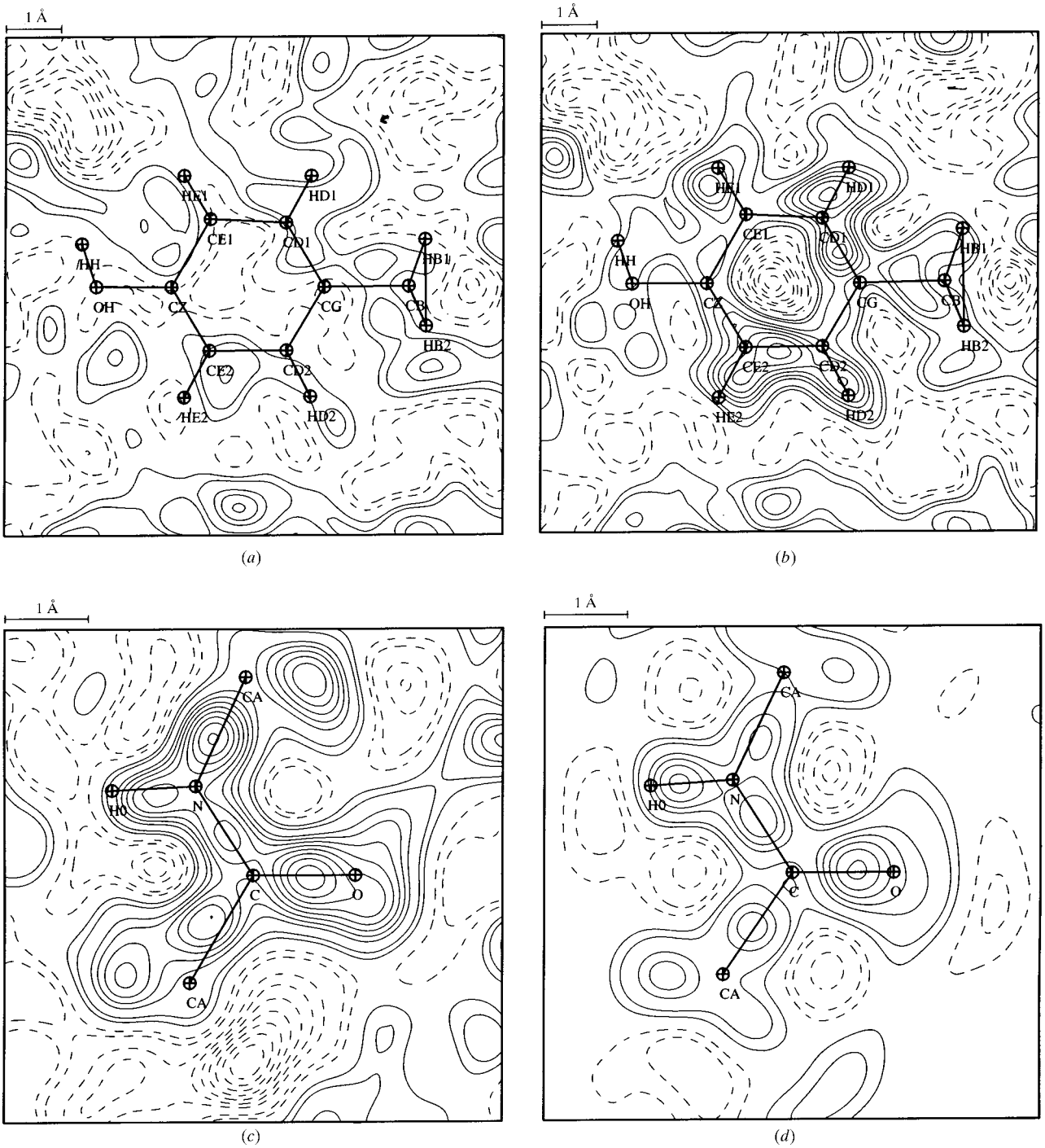

Figure 3
(*a*) Residual electron density in the Tyr47 plane after refinement step IV. Experimental deformation electron-density maps after refinement step V (*b*) in the Tyr47 plane, (*c*) in the peptide plane Tyr47–Cys48 and (*d*) averaged over the 19 peptide-bond planes with lowest thermal motion. The bonding density of H atoms is well defined in the deformation maps. Contour interval 0.02 e Å^{-3}; positive, solid line; negative, dashed line; zero contour omitted.

Table 4
Net charges (in electrons) obtained from P_{val} refinement (step VI) and their standard uncertainty for the eight atom types.

Atom	Type	Net charges
N	N—H	−0.20 (2)
H	H—N	+0.07 (2)
O	O═C	−0.11 (2)
C	C═O	+0.20 (2)
C	C^{α}	−0.23 (2)
H	H^{α}	+0.04 (1)
C	C(sp^3)	−0.10 (2)
C	C(sp^2)	+0.28 (3)

to provide some qualitative information on the atomic charges of the peptide bond for the toxin structure. It also supports the significance of the deformation density observed. Finally, the *R* factor obtained in refinement step VI (atomic charge refinement using a spherical atom model) is higher than in refinement steps IV and V. This is because of the smaller number of parameters refined: the thermal displacement parameters are fixed in that refinement and cannot model the aspherical features of the electron density (Coppens, 1967) in this high-resolution protein structure.

3.3. Analysis of ADP: rigid-bond test and TLS

As we have previously observed at 0.9 Å resolution and with an average *B* factor of 8 Å^2, the estimate of the ADPs is not significantly changed with a multipolar *versus* a spherical atom model. One of the reasons for this is that the expected difference between *B* factors at the end of steps IV and V ($\Delta B \simeq 0.4$ Å^2; Jelsch *et al.*, 1998) is small in comparison with the lowest toxin *B* factor (5 Å^2) and its associated error (0.3 Å^2). After step V, most *B* factors of atoms belonging to peptide bonds and aromatic rings show, as expected, a general slight decrease of their values in comparison with the previous step; however, the change does not exceed twice the standard deviation.

The 'rigid-bond' test (Hirshfeld, 1976) is also a useful tool to show whether the ADPs account only for dynamic and static disorder or include other features. For the test to be satisfied, the difference of the ADP in the direction of the chemical bond between two neighboring atoms should be low ($\Delta Z_B > 0.08$ Å^2). When no rigid-bond restraints are applied, the rigid-bond test is not satisfied: at the end of refinement III and when restraints on ADPs are removed, the r.m.s.d. of ΔZ_B value is much too high (4.2 Å^2 for the most ordered protein atoms and 3.1 Å^2 for all the main-chain atoms). However, when using strict rigid-bond restraints to satisfy this test in refinement III, no significant increase of the *R* factor is observed.

A TLS analysis (Shomaker & Trueblood, 1968; Trueblood & Dunitz, 1983) of all aromatic rings after refinement III was performed and the agreement between B_{anis} and B_{TLS} is shown in Table 3. The discrepancy between the crystallographic and TLS thermal tensors is about 10%, indicating that the ADPs are representative of the effective thermal motion of the protein atoms. When the rigid-bond restraint is applied, the TLS reliability index shows a 20% improvement. The TLS analysis of the same groups was also carried out after refinement V, but no significant variations were observed in ADPs.

4. Discussion

As in other protein structures at similar resolution, refinement with *SHELXL*97 has shown the relevance of the anisotropic ADP refinement. However, the use of the more sophisticated model for atomic electron density on the 0.96 Å resolution toxin structure leads to mixed results. The deformation density map derived from the multipolar model clearly shows that valence electrons are observed in the covalent bonds located in the most ordered regions of the protein. Thus, even if the usual statistical criteria (*R* factors) do not clearly distinguish between the spherical and the non-spherical atomic electron-density model used for the refinement, the benefit of a more sophisticated model appears in the regions where the ADPs are low; the deformation density obtained is mostly a consequence of the improvement of positional and thermal parameters when using aspherical atom form factors.

As observed in low-resolution small-molecule crystallography, because of the limited resolution of the toxin data and thermal smearing, no significant residual electron density can be observed after refinement IV (Fig. 3*a*). As described above, when a spherical atom model is used, the ADPs are able to 'absorb' most of the aspherical density features. Consequently, the weakness of the deformation density was rather predictable. The average ADP is $\langle B_{eq} \rangle = 8$ Å^2 for the toxin atoms, whereas it has been observed that the deformation density peaks already start to be diluted with *B* values as low as $B_{eq} = 4$ Å^2 (Jelsch *et al.*, 1998).

The charge refinement has allowed qualitative determination of the net atomic charges of the peptide bond by averaging on 19 of these bonds. Thus, even if we are far from being able to determine the charge of a whole protein structure, this shows that electrostatic information is present in room-temperature 0.96 Å resolution data which the usual free-atom model is unable to give. Accurate higher resolution protein data for structures at low temperature will allow charge refinement on the less mobile residues without the need for averaging over similar chemical groups.

Concerning the *B*-factor analysis, no significant improvement was obtained when the multipolar model was used. The only feature that differentiates the two refinements is the smaller overall *B* factor observed with the multipolar refinement (Jelsch *et al.*, 1998). No significant difference appears with either the rigid-bond test or the TLS analysis. In fact, the expected ADP difference between the two types of refinement is small and compares with the errors estimated by full-matrix refinement (0.2–0.4 Å^2 for *B* factors). In relative values, the error reaches 10%, a value similar to the R_2 residual calculated in the TLS analysis (Table 3). This lack of accuracy of *B* factors is also a consequence of the limited resolution and to the fact that this is really the limit in terms of signal-to-noise ratio for

the crystallographic data. In order to improve the precision of ADPs, it will be necessary to collect more accurate crystallographic data with an overall lower B factor, higher resolution, cryo-cooled crystals and a more intense and less divergent X-ray beam (third-generation synchrotrons).

In order to show the relationship between the diffraction resolution limit and the thermal motion, the diffraction power of a C atom with increasing B factors has been represented as a function of resolution in Fig. 4. The square of the scattering factor has been represented as it is closely related, after Lorentz polarization correction, to the measurable quantity of the diffraction intensities. The scattering power of the atoms at high resolution (d < 0.9 Å) arises from the core electrons as they are located close to the nucleus, contrary to the valence electrons, which are more diffuse in real space.

For a C atom with no thermal motion at all (B = 0 Å^2), the diffraction would actually decrease very slowly with resolution (Fig. 4). If the B factor reaches 2 Å^2, the diffraction of the crystal is still expected to extend to subatomic resolution (0.6–0.5 Å). In the case of the toxin structure, the atoms with the lowest B factor (B = 5 Å^2) have a squared scattering factor which decreases by a factor of 300 from infinite to 0.9 Å resolution. These atoms are mainly responsible for the diffraction power at atomic resolution of the toxin crystals, since those with an ADP of 8 Å^2 have a squared scattering factor lower by one order of magnitude at 0.9 Å resolution. However, we have to keep in mind that the toxin data have been collected at room temperature. It is expected that collection of diffraction data at cryogenic temperature (100 K or even 10 K) will lower the ADPs and improve the quality of both the crystallographic data and the final model (Walsh *et al.*, 1998). The static disorder might be much more difficult to reduce. Additionally, the use of more sophisticated cryogenic mother liquors and freezing techniques than the traditional flash-cooling of the crystal at room temperature is an area of research *per se* which could lead to crystalline protein structures with reduced static and dynamic disorder. The reduction of the atomic B factors to values as low as 2–4 Å^2 enables an extension of the field of charge-density studies from small molecules to protein crystallography, as shown in the experimental results obtained on 100 K crambin crystals (Jelsch *et al.*, 1999) at 0.54 Å resolution. In the case of crambin, the averaging of residual electron density over the peptide bonds of the protein clearly shows some non-modelled bonding density when using a spherical atom model. Also, quantum-mechanics calculations are now being performed on whole protein structures such as crambin (Van Alsenoy *et al.*, 1998; Fernandez-Serra *et al.*, 1999). This joint approach will enable the calculation of reliable protein electrostatics in the near future.

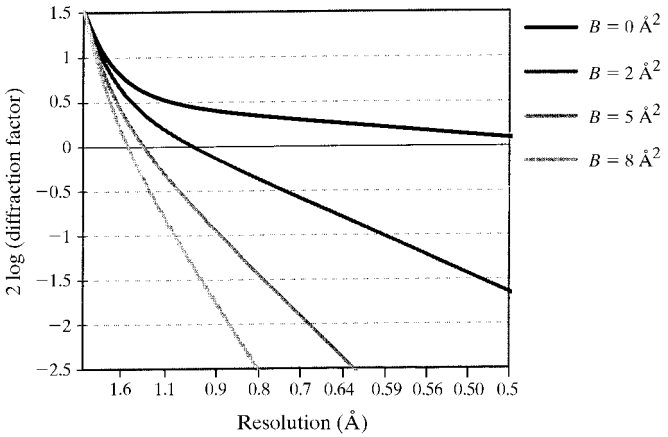

Figure 4
Diffraction power of a C atom as a function of resolution for several isotropic temperature factors. The square of the scattering factor has been chosen as it is closely related to the measured diffraction intensities (after Lorentz polarization correction).

References

Allen, F. H. (1986). *Acta Cryst.* B**42**, 515–522.
Bader, R. F. W. & Essen, H. (1984). *J. Chem. Phys.* **80**, 1943–1960.
Bayly, C. I., Cieplack, P., Cornell, W. D. & Kollman, P. A. (1993). *J. Phys. Chem.* **97**, 10269–10280.
Bouhmaida, N., Ghermani, N.-E., Lecomte, C. & Thalal, A. (1999). *Acta Cryst.* A**55**, 729–738.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–474.
Clementi, E. & Raimondi, D. L. (1963). *J. Chem. Phys.* **41**, 2686–2689.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Coppens, P. (1967). *Science*, **158**, 1577.
Coppens, P. (1997). *X-ray Charge Densities and Chemical Bonding.* Oxford: IUCr/Oxford University Press.
Coppens, P. (1998). *Acta Cryst.* A**54**, 779–788.
Coppens, P., Guru Row, T. N., Leung P., Stevens, E. D., Becker, P. J. & Yang, Y. W. (1979). *Acta Cryst.* A**35**, 63–72.
Cromer, D. T. (1974). *International Tables for X-ray Crystallography*, edited by J. A. Ibers & W. E. Hamilton, Vol. IV, pp. 148–151. Birmingham: Kynoch Press.
Dahaoui, S., Jelsch, C., Howard, J. & Lecomte, C. (1998). *Acta Cryst.* B**55**, 226–230.
Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1995). *Curr. Opin. Struct. Biol.* **5**, 784–790.
Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1997). *Curr. Opin. Struct. Biol.* **7**, 681–688.
Dauter, Z., Wilson, K. S., Sieker, L.C., Meyer, J. & Moulis, J. M. (1997). *Biochemistry*, **36**, 16065–16073.
Deacon, A. M., Weeks, C. M., Miller, R. & Ealick, S. E. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 9284–9289.
Dunitz, J. D. & Seiler, P. (1983). *J. Am. Chem. Soc.* **105**, 7056–7058.
Fernandez-Serra, M. V., Junquera, J., Artacho, E., Jelsch, C. & Lecomte, C. (1999). Proceedings of the Second European Charge Density Meeting, Sitges, Spain.
Fontecilla-Camps, J.-C., Habersetzer-Rochat, C. & Rochat, H. (1988). *Proc. Natl Acad. Sci. USA*, **85**, 7443–7447.
Genick, U. K., Soltis, M., Kuhn, P., Canestrelli, I. L. & Getzoff, E. D. (1998). *Nature (London)*, **392**, 206–209.
Hansen, N. K. & Coppens, P. (1978). *Acta Cryst.* A**34**, 909–921.
Hirshfeld, F. L. (1976). *Acta Cryst.* A**32**, 239–244.
Housset, D., Habersetzer-Rochat, C., Astier, J. P. & Fontecilla-Camps, J. C. (1994). *J. Mol. Biol.* **238**, 88–103.
Jelsch, C., Pichon-Pesme, V., Lecomte, C. & Aubry, A. (1998). *Acta Cryst.* D**54**, 1306–1318.
Jelsch, C., Teeter, M. M., Pichon-Pesme, V., Blessing, R. H. & Lecomte, C. (1999). Abstract No. M11.BB.004. XVIIIth IUCR Congress.
Jones, T. A., Cowan, S., Zou, J. Y. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kuhn, P., Knapp, M., Soltis, S. M., Ganshaw, G., Thoene, M. & Bott, R. (1998). *Biochemistry*, **37**, 13446–13452.
Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–147.
Langs, D. A. (1988). *Science*, **241**, 188–191.
Lecomte, C. (1995). *Advances in Molecular Structure Research*, edited by I. Hargittai & M. Hargittai, Vol. 1, pp. 261–302. Greenwich, CT, USA: JAI Press Inc.
Leslie, A. G. W. (1992). *CCP4 ESF–EACMB Newslett. Protein Crystallogr.* **26**.
Longhi, S., Czjzek, M. & Cambillau, C. (1998). *Curr. Opin. Struct. Biol.* **8**, 730–737.
Miranda, F., Kupeyan, C., Rochat, H. & Lissitzky, S. (1970). *Eur. J. Biochem.* **17**, 477–484.
Mulliken, R. S. (1955). *J. Chem. Phys.* **23**, 1833–1846.
Pichon-Pesme, V. & Lecomte, C. (1998). *Acta Cryst.* B**54**, 485–493.
Pichon-Pesme, V., Lecomte, C. & Lachekar, H. (1995). *J. Phys. Chem.* **99**, 6242–6250.
Pichon-Pesme, V., Lecomte, C., Wiest, R. & Bénard, M. (1992). *J. Am. Chem. Soc.* **114**, 2713–2715.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Sakabe, N., Sasaki, K. & Sakabe, K. (1981). *Acta Cryst.* A**37**, C50–C53.
Sheldrick, G. M. & Schneider, T. (1997). *Methods Enzymol.* **277**, 319–343.
Shomaker, V. & Trueblood, K. N. (1968). *Acta Cryst.* B**24**, 63–76.
Smith, G. D., Blessing, R. H., Ealick, S. E., Fontecilla-Camps, J. C., Hauptman, H. A., Housset, D., Langs, D. A. & Miller, R. (1997). *Acta Cryst.* D**53**, 551–557.
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, S., Zhou, R. & Teeter, M. M. (1995). *Acta Cryst.* D**51**, 663–681.
Stewart, R. F., Davidson, E. R. & Simpson, W. T. (1965). *J. Chem. Phys.* **43**, 175–187.
Teeter, M. M. & Hendrickson, W. A. (1979). *J. Mol. Biol.* **127**, 219–223.
Trueblood, K. N. (1990). *THMA*11. University of California at Los Angeles, USA.
Trueblood, K. N. & Dunitz, J. D. (1983). *Acta Cryst.* B**39**, 120–133.
Van Alsenoy, C., Yu, C. H., Peeters, A., Martin, J. M. L. & Schäfer, L. (1998). *J. Phys. Chem.* **102**, 2246–2251.
Walsh, M. A., Schneider, T. R., Sieker, L. C., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1998). *Acta Cryst.* D**54**, 522–546.
Watenpaugh, K. D., Sieker, L. C. & Jensen, L. H. (1979). *J. Mol. Biol.* **131**, 509–522.
Wlodawer, A., Walter, R., Huber, R. & Sjolin, L. (1984). *J. Mol. Biol.* **180**, 301–329.