Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 8 October 2003
Accepted 2 March 2004

A comparison between experimental and theoretical aspherical-atom scattering factors for charge-density refinement of large molecules

Virginie Pichon-Pesme, Christian Jelsch, Benoit Guillot and Claude Lecomte*

LCM^{3}B, CNRS – Faculté des Sciences et Techniques, Université Henri Poincaré, Nancy 1, BP 239, 54506 Vandoeuvre-lès-Nancy, France. Correspondence e-mail: claude.lecomte@lcm3b.uhp-nancy.fr

The differences between two databases describing the polypeptide main chain in terms of charge-density parameters, directly usable in protein structure refinements, are discussed. These databases contain averaged multipole populations of peptide pseudo-atoms obtained from refinement against theoretical simulated data and against high-resolution experimental data on small peptide or amino acid molecules. The main discrepancy becomes apparent when electrostatic properties are calculated.

1. Introduction

Since 1990, the number of reported high-resolution protein structures ($d \leq 1$ Å) has increased dramatically. We have recently shown (Jelsch *et al.*, 2000) that these data deserve a better model than that of the spherical atom, allowing one to model the charge transfer and the deformation of the electron density caused by covalent and intermolecular interactions, as usually performed in small-molecule crystallography (see, for example, Coppens, 1997; Koritsanszky & Coppens, 2001; Lecomte *et al.*, 2003).

In order to model the valence electron density and related properties of protein atoms (net atomic charges, electrostatic potential, interaction energy *etc.*) from X-ray diffraction data, we have built a database of experimental average aspherical-atom scattering-factor parameters for chemically unique peptide pseudo-atoms (Pichon-Pesme *et al.*, 1995). This database is now complete and includes all side chains of biologically active peptides (Pichon-Pesme *et al.*, 2004). This database is referred to as P2JL in the following. We have also demonstrated the transferability of these aspherical scattering factors for accurate protein crystallography by refining large peptide structures (Jelsch *et al.*, 1998) and ultra-high-resolution protein data: the database was successfully tested on a scorpion toxin (Housset *et al.*, 2000), on protein crambin (Jelsch *et al.*, 2000) and on aldose reductase (Muzet *et al.*, 2003).

Furthermore, in order to be able to refine ultra-high-resolution protein data with an aspherical formalism, we have written a new refinement program *MoPro* (Guillot *et al.*, 2001) based on the Hansen–Coppens formalism (Hansen & Coppens, 1978):

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

The first two terms describe a spherically symmetric core and the valence density, while the third term reflects the non-spherical redistribution of the valence-electron density. The $y_{lm\pm}$ functions are the multipolar spherical harmonic angular functions in real form, and $R_l = N_l r^n \exp(-\kappa'\zeta r)$ are Slater-type radial functions, in which N_l is a normalization factor, and n and ζ are parameters chosen according to the recipe given by Hansen & Coppens (1978). $P_{\rm val}$, $P_{lm\pm}$, κ and κ' are the refined parameters for which average values are stored in the database.

Besides aspherical-atom refinement, the *MoPro* program also contains all necessary restraints used for atomic resolution protein refinement and allows, or not, depending on the resolution of the diffraction data, charge-density parameter refinement from starting values transferred from the aspherical scattering-factor database. These aspherical parameters lead to the first experimental determination of the electrostatic potential of a protein (Muzet *et al.*, 2003).

Recently, in parallel to this research, Koritsanszky *et al.* (2002) proposed a database (called KVC in the following) of theoretical aspherical scattering factors for the accurate description of the deformation density and electrostatic properties of large peptides and proteins. These scattering factors have been obtained from valence-only structure factors of isolated tripeptide molecules calculated using the density functional theory (DFT) at the B3LYP level utilizing the 6-31G** standard basis set. The calculations were performed on isolated molecules with the *GAUSSIAN98* program package (Frisch *et al.*, 1998) after geometry optimizations derived from molecular mechanics. The resulting calculated static structure factors $F(\mathbf{H})$ were then used for the charge-density refinements with a modified version of the program *XD* (Koritsanszky *et al*, 1995) to a maximum resolution of $\sin\theta/\lambda \leq 1.15$ Å^{-1}; therefore, no experimental data were used. As shown in Table 1, the multipole expansion was

Table 1
Averaged multipole populations of the peptide backbone of the P2JL database (upper rows) compared with those given by Koritsanszky *et al* (2002) (lower rows).

Bold values are the most significant parameters used for *MoPro* refinement of peptides and proteins (Guillot *et al.*, 2001). Numbers between parentheses are the standard error of the mean. [In the KVC parameters, 0.0 corresponds to insignificant populations and '–' are the constrained symmetry-forbidden spherical harmonics.]

	Kappa κ	Kappa′ κ'	 P_{val}	11^+ x	11^- y	10 z	20 $3z^2-1$	21^+ xz	21^- yz	22^+ x^2-y^2	22^- xy
C′	**0.998 (4)**	**0.941 (16)**	**3.976 (3)**	**0.102 (8)**	0.002 (5)	−0.009 (6)	**−0.304 (11)**	0.002 (6)	−0.007 (4)	**0.102 (11)**	−0.026 (13)
	1.002 (1)	0.850 (3)	4.05 (1)	0.115 (7)	0.0	–	−0.303 (5)	–	–	0.096 (3)	0.020 (5)
Cα	**0.992 (5)**	**0.911 (18)**	**4.119 (42)**	−0.009 (15)	−0.051 (17)	−0.009 (23)	0.021 (26)	−0.006 (13)	0.012 (15)	0.002 (17)	0.023 (13)
	1.006 (12)	0.920 (4)	4.02 (2)	0.0	−0.080 (5)	0.092 (5)	0.048 (5)	−0.045 (7)	0.0	0.023 (4)	0.058 (4)
O	**0.977 (3)**	**0.960 (28)**	**6.307 (16)**	**−0.075 (8)**	0.011 (7)	−0.001 (4)	**−0.078 (7)**	0.002 (5)	−0.009 (4)	**−0.070 (6)**	0.005 (6)
	0.9869 (2)	1.003 (5)	6.15 (1)	−0.098 (3)	−0.008 (1)	–	−0.087 (1)	–	–	−0.095 (1)	−0.009 (2)
N	**0.987 (3)**	**0.867 (15)**	**5.312 (26)**	−0.005 (7)	**−0.072 (10)**	−0.006 (3)	**−0.050 (11)**	0.008 (14)	0.003 (5)	0.000 (8)	0.034 (12)
	0.9972 (4)	1.126 (36)	5.13 (3)	0.0	0.040 (6)	0.0	0.066 (2)	0.0	0.0	0.0	−0.018 (3)
H_N	**1.203 (24)**	**1.051 (27)**	**0.700 (22)**	**0.164 (11)**							
	1.253 (26)	1.537 (40)	0.81 (3)	0.11 (1)							
H_α	**1.182 (16)**	**1.091 (28)**	**0.804 (20)**	**0.107 (9)**							
	1.236 (1)	1.616 (12)	0.825 (5)	0.068 (1)							

	30 $(5z^2-3)z$	31^+ $(5z^2-1)x$	31^- $(5z^2-1)y$	32^+ $(x^2-y^2)z$	32^- $2xyz$	33^+ $(x^2-3y^2)x$	33^- $(3x^2-y^2)y$
C′	0.000 (4)	−0.009 (7)	0.002 (5)	0.005 (6)	−0.002 (4)	**0.411 (15)**	−0.011 (7)
	–	0.0	0.024 (3)	–	–	0.424 (6)	–
Cα	0.017 (11)	**−0.171 (13)**	**−0.268 (15)**	−0.001 (13)	−0.016 (11)	**0.265 (20)**	**0.056 (11)**
	0.0	−0.161 (6)	−0.213 (2)	0.0	0.0	0.195 (3)	−0.062 (4)
O	−0.001 (3)	−0.013 (7)	0.006 (4)	0.002 (3)	0.003 (2)	**0.039 (3)**	−0.014 (5)
	0.0	0.0	−0.004 (1)	–	–	0.0	0.0
N	−0.004 (5)	0.011 (5)	−0.009 (2)	0.004 (2)	0.008 (5)	**0.260 (9)**	−0.018 (8)
	0.0	0.014 (4)	0.0	0.0	0.0	0.12 (1)	0.009 (2)

also limited to $l = 3$ for C, O and N atoms, and only dipolar contributions of the H atoms were deemed necessary. These expansion limits and the chosen local axes systems are identical to those used for the experimental database. The refinements against theoretical data were performed by imposing m symmetry on the carbonyl C′ and O atoms, and on the C and O atoms of the COH group, and $m3$ symmetry on the methyl group. In contrast, fewer constraints were imposed on the experimental parameters: the selection of the P_{lm} parameters was made, *a posteriori*, on a statistical basis during the database building (Pichon-Pesme *et al.*, 1995, 2004). The parameters that fulfil the statistical criteria defined in the above papers were deposited in the experimental database. These parameters are presented in bold in Table 1.

2. Electron-density parameters

In Table 1, the aspherical parameters of the peptide unit of the theoretical database are compared with the experimental parameters. Note that the experimental parameters have been updated since the Pichon-Pesme *et al.* (1995) paper; the averaged parameters are now calculated on 23 samples compared with four in the previous paper. For each of the refined structures, the refinement was performed with the same strategy: first spherical high-order refinement to obtain unbiased positional (x, y, z) and thermal (U_{ij}) parameters, then multipole refinement using all reflections. At the end, all charge-density parameters including kappas were refined together with the positional and thermal parameters. Usually, no large correlation coefficients were observed. As written in the KVC paper, the theoretical non-spherical parameters of Cα, C′ and, to a lesser extent, O atoms are in excellent agreement with experiment (if monopole and kappa parameters are excluded), whereas large differences appear for N and H atoms. Fig. 1 gives the deformation density of the nitrogen atom using the KVC parameters (Fig. 1*a*) compared with that calculated with the experimental parameters (Fig. 1*b*) stored in the database. The large difference observed is, *a priori*, not surprising, because least-squares fitting of static theoretical valence-only structure factors may partition the aspherical density of the central N peptide atom differently from that obtained through refinement of accurate experimental dynamic structure factors. This is, however, not straightforward as the non-spherical parameters of the C′ and Cα neighbouring atoms agree very well. In the experimental case, the N-atom charge density extends more along the bonds, as predicted by its negative charge and its κ' value being smaller than unity (Table 1): $\kappa'_{exp} = 0.87$, $P_{val\ exp} = 5.31$, compared with $\kappa'_{th} = 1.13$, $P_{val\ th} = 5.13$. As the theoretical refinement was performed on isolated peptides, Koritsanszky *et al.* invoke intermolecular interactions occurring in the crystalline state to explain this difference. This seems to be in contradiction with the fact that one observes a much better agreement for the carbonyl O atom which is also involved in

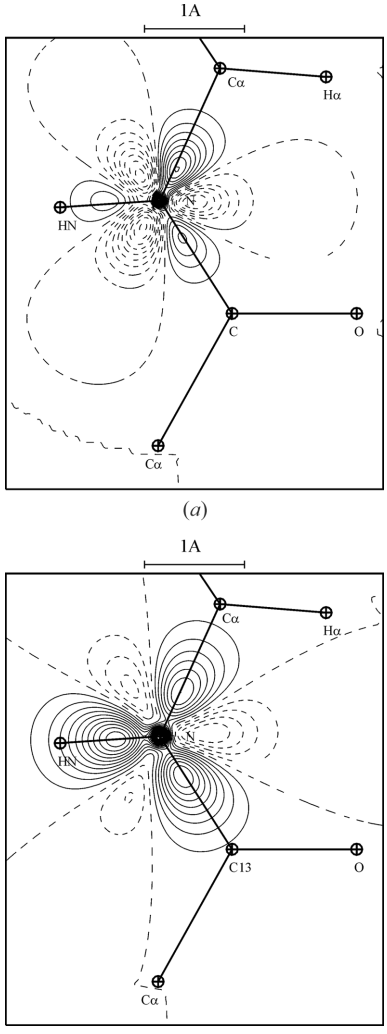

Figure 1
Deformation electron density of the N atom (*a*) using the KVC parameters and (*b*) calculated with the P2JL parameters. Contours are at 0.05 e Å^{-3} with full lines denoting positive contours and dashed lines denoting negative and zero contours.

hydrogen bonding. The electron density of the carbonyl O atom together with the NH group is the most subject to variability depending on the chemical environment within the polypeptide main chain (Fernandez-Serra *et al.*, 2000).

The theoretical and experimental H_N dipolar and Kappa parameters disagree in Table 1; this is partly a consequence of the difference between N-atom population parameters. The theoretical net atomic charges of H_N and H_α are almost equal (+0.2), despite the more acidic character of H_N compared with H_α. The experimental database shows the acidic character of H_N (q_{HN} = +0.3) compared with q_{H_α} = +0.2 for H_α. Furthermore, the H_N dipole directed towards the N atom is lower in the theoretical calculation (0.11) compared with the experimental database (0.16), which reflects a weaker electron transfer from the H atom to the N atom. As intermolecular effects are averaged over 28 H_N atoms of the experimental database, the experimental database therefore provides electron-density parameters of a polarized H_N atom averaged over most of the possible intermolecular configurations in polypeptides, *i.e.* very suitable for an aspherical scattering-factor database to be used in protein refinement, as demonstrated by Muzet *et al.* (2003). These comments are also valid for the H_α and O peptide atoms. These observations are confirmed in Fig. 2, which gives the theoretical (Fig. 2*a*) and experimental (Fig. 2*b*) deformation densities of the NHC′O group and their difference (Fig. 2*c*). Unexpectedly, the bonding deformation density is generally higher in the experimental database and the difference in H dipolar population (P_{11+}) appears as a strong contracted dipole centred on the H atom. The small features around the O atom (Fig. 2*c*) can be attributed to the lack of experimental resolution (∼1 Å^{-1}).

3. Electrostatic properties

The most important application of building a charge-density database, either from experiment or from theory, is its application to electrostatics. As shown in Table 1, the two databases clearly disagree when charges (P_{val}) and Kappa parameters

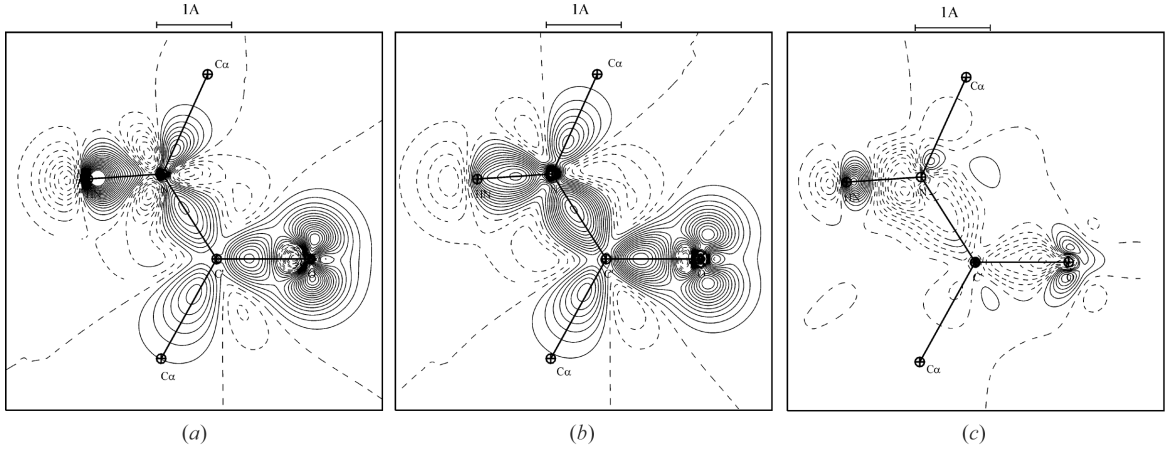

Figure 2
Static deformation electron densities of the NH—C′═O group, (*a*) using the KVC parameters, (*b*) calculated with the P2JL parameters, and (*c*) the difference between (*a*) and (*b*). Contours are as in Fig. 1.

are compared. These parameters are the most important for describing the medium- and long-range interactions. Fig. 3 compares the electrostatic potential obtained from theoretical (Fig. 3*a*) and experimental (Fig. 3*b*) databases for a group of four atoms involved in a peptide bond (NH—C′═O), calculated using the *ELECTROS* program (Ghermani *et al.*, 1998). Qualitatively, the experimental database potential is deeper than the corresponding theoretical potential. This disagreement is partly caused by the net charge of the [C′, N, O, H_N] group: −0.3 e for our database compared with −0.14 e for the KVC parameters. As most experimental charge-density analyses of peptides (Pichon-Pesme *et al.*, 2000; Benabicha *et al.*, 2000) lead to a negative potential of approximately −0.3 e Å^{-1} around the carbonyl O atom, the experimental values clearly reproduce the peptide environment on average. On the other hand, the theoretical potentials obtained for molecules *in vacuo* are similar for the KVC database and in a semi-empirical molecular-orbital calculation (Gopi Mohan *et al.*, 1996) around O═C oxygen atoms in thymine and glycine molecules (−0.15 and −0.18 e Å^{-1}, respectively).

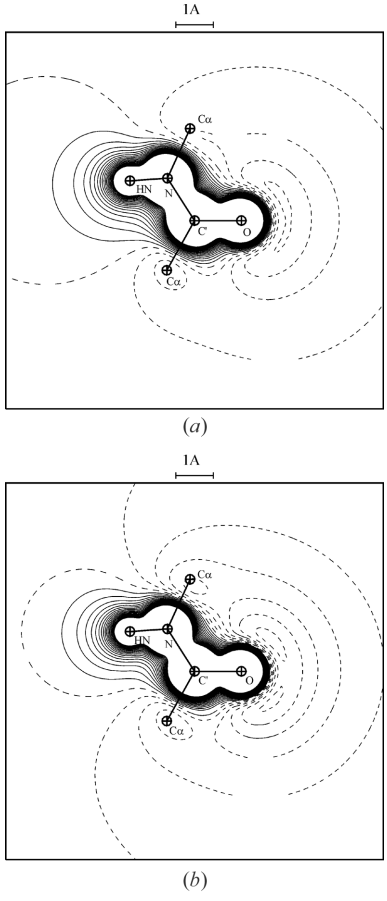

Figure 3
Electrostatic potential generated by the NH—C′═O group obtained from (*a*) the KVC and (*b*) the P2JL database. Contours are at 0.05 e Å^{-1} with full lines denoting positive contours and dashed lines denoting negative and zero contours.

4. Conclusions

The electron-density maps derived from the two methods agree qualitatively only. Most P_{val} parameters differ by 0.1 e, which leads to different electrostatic properties; the charges of the experimental database are in accordance with chemical intuition. Finally, the experimental database has been successfully tested on protein high-resolution X-ray data (Housset *et al.*, 2000; Jelsch *et al.* 2000; Muzet *et al.*, 2003; Blessing *et al.*, 2003).

This research was supported by the Centre National de la Recherche Scientifique (CNRS) and the French Ministère de la Recherche.

References

Benabicha, F., Pichon-Pesme, V., Jelsch, C., Lecomte, C. & Khmou, A. (2000). *Acta Cryst.* B**56**, 155–165.
Blessing, R. H., Guillot, B., Lecomte, C. & Smith, G. D. (2003). American Crystallographic Association Meeting, Cincinnati, OH, USA, 26–31 July 2003. Abstract No. 01.03.03, Series 2, 30, 33.
Coppens, P. (1997). *X-ray Charge Densities and Chemical Bonding*. Oxford University Press.
Fernandez-Serra, M. V., Junquera, J., Jelsch, C., Lecomte, C. & Artacho, E. (2000). *Solid State Commun.* **116**, 395–400.
Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Zakrzewski, V. G., Montgomery, J. A. Jr, Stratmann, R. E, Burant, J. C., Dapprich, S., Millam, J. M., Daniels, A. D., Kudin, K. N., Strain, M. C., Farkas, O., Tomasi, J., Barone, V., Cossi, M., Cammi, R., Mennucci, B., Pomelli, C., Adamo, C., Clifford, S., Ochterski, J., Petersson, G. A., Ayala, P. Y., Cui, Q., Morokuma, K., Salvador, P., Dannenberg, J. J., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Cioslowski, J., Ortiz, J. V., Baboul, A. G., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Gomperts, R., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Andres, J. L., Gonzalez, C., Head-Gordon, M., Replogle, E. S. & Pople J. A. (1998). *GAUSSIAN98* program package. Gaussian, Inc., Pittsburgh, PA, USA.
Ghermani, N. E., Bouhmaida, N.& Lecomte, C. (1992–1998). Internal report, UMR CNRS 7036, Université Henri Poincaré-Nancy 1, France.
Gopi Mohan, C., Kumar, A. & Mishra, P. C. (1996). *Int. J. Quant. Chem.* **60**, 699–708.
Guillot, B., Viry, L., Guillot, R., Lecomte, C. & Jelsch, C. (2001). *J. Appl. Cryst.* **34**, 214–223.
Hansen, N. K. & Coppens, P. (1978). *Acta Cryst.* A**34**, 909–921.
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.
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.
Koritsanszky, T. & Coppens, P. (2001). *Chem. Rev.* **101**, 1583–1628.
Koritsanszky, T., Howard, S. T., Mallinson, P. R., Su, Z. & Hansen, N. K. (1995). *XD, a Computer Program Package for Multipole Refinement and Analysis of Charge Densities from X-ray Diffraction Data*. Free University of Berlin, Germany.
Koritsanszky, T., Volkov, A. & Coppens, P. (2002). *Acta Cryst.* A**58**, 464–472.
Lecomte, C., Souhassou, M. & Pillet, S. (2003). *J. Mol. Struct.* **647**, 53–64.

Muzet, N., Guillot, B., Jelsch, C. & Lecomte, C. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 8742–8747.
Pichon-Pesme, V., Jelsch, C. & Lecomte, C. (2004). In preparation.
Pichon-Pesme, V., Lachekar, H., Souhassou, M. & Lecomte, C. (2000). *Acta Cryst.* B**56**, 728–737.
Pichon-Pesme, V., Lecomte, C. & Lachekar, H. (1995). *J. Phys. Chem.* **99**, 6242–6250.