Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 16 April 2002
Accepted 19 June 2002

Aspherical-atom scattering factors from molecular wave functions. 1. Transferability and conformation dependence of atomic electron densities of peptides within the multipole formalism

Tibor Koritsanszky,* Anatoliy Volkov and Philip Coppens

Department of Chemistry, State University of New York at Buffalo, NY 14260-3000, USA. Correspondence e-mail: chem9986@acsu.buffalo.edu

In this study, the feasibility of building a database of theoretical atomic deformation density parameters applicable to the construction of the densities of biomacromolecules and to the interpretation of their X-ray diffraction data is discussed. The procedure described involves generation of valence-only structure factors of tripeptides calculated from theoretical densities at the B3LYP level and the refinement of multipole parameters against these simulated data. Our results so far indicate that the backbone pseudoatoms extracted in such a way are highly transferable and fairly invariant with respect to rotations around single bonds in the peptide framework. The ultimate goal is to use the aspherical-atom database for improved macromolecular refinements that are based on high-resolution data and for prediction of electrostatic properties of larger molecules.

1. Introduction

With the revolutionary advances of macromolecular crystallography over the past decade, the time needed to collect an extended data set has been reduced from months to a few days or less, while accuracy has been enhanced because of the high level of redundancy achievable and the now common practice of low-temperature data collection.

Improved statistical methods for data reduction, refinement and structure validation have led to better figures of merit to evaluate the absolute quality of the structure associated with a given set of diffraction data. These developments are increasingly producing data of such a quality that application of scattering models beyond the spherical-atom approximation is becoming a realistic possibility (Fernández-Serra *et al.*, 2000).

A scattering model widely applied in the course of experimental charge-density determination is the 'pseudoatom' or multipole formalism (Hansen & Coppens, 1978). The atomic partitioning that is inherent in the pseudoatom definition makes it possible to explore the degree of transferability of atomic density functions in large biomolecules with similar repeating subunits. A plausible reasoning suggests that chemically equivalent atoms make close to identical contributions to the total density, provided their local connectivity is similar. Such chemical symmetry is often imposed to reduce the number of variables in charge-density refinements. But it can also be used to construct molecular densities from a limited number of known pseudoatomic densities. This idea was first realized in a study of displacement amplitudes, in the course of which deformation-density parameters (Hirshfeld, 1971) extracted from perylene high-resolution data were used to improve the static scattering model of the naphthalene and anthracene molecules (Brock *et al.*, 1991). The same strategy has been adopted for building the charge distribution of peptide backbones from experimental pseudoatom densities of small peptides and amino acids (Pichon-Pesme *et al.*, 1995). The simulated deformation densities were found to agree within 0.1 e $Å^{-3}$ in the bonding regions with those extracted directly from X-ray data. An experimental 'deformation density databank' of averaged multipole populations of chemically unique peptide pseudoatoms has been reported (Pichon-Pesme *et al.*, 1995). Application of the database to the thyrotropin-releasing hormone analog tripeptide *p*Glu–Phe–D-Pro–ϕ[CN_4]–Me resulted in a significant improvement of the fit, as compared to that of the conventional model. This improvement is reflected not only in terms of statistical figures of the fit but also in terms of enhanced physical significance of the thermal displacement parameters (Pichon-Pesme *et al.*, 1995). The treatment of a relatively high resolution (0.82 Å) X-ray data set on the 3_{10} helix octapeptide Ac–Aib_2–L-Lys(Bz)–Aib_2–L-Lys(Bz)–Aib_2–NHMe likewise gave a considerable improvement in the crystallographic *R* factor, accumulation of electrons in the lone-pair and bond regions in the Fourier deformation maps and a significant change in the temperature factors (Jelsch *et al.*, 1998). Efforts have been made toward aspherical-atom refinement using the limited model. The study on the scorpion *Androctonus australis*

Hector toxin II (room-temperature data of 0.96 Å resolution) also revealed bonding effects, but led to an underestimation of the net charges on the peptide-bond atoms (Housset *et al.*, 2000). In the case of the protein crambin (0.54 Å resolution, low-temperature data), the refinement resulted in multipole parameters within 25% of those from the transferable database but lower bonding features than predicted by the database (Fernández-Serra *et al.*, 2000).

An alternative route to a pseudoatom databank is offered by systematic analysis of theoretical densities that can be readily calculated for small and medium-size molecules given the power of modern computers. Model studies using X-ray data generated from known densities have unlimited potential. The simulated static data are free of random and systematic errors as well as thermal motion effects, can be calculated to arbitrary resolution and with a uniform reciprocal-space sampling, at different levels of theory, for any molecule (limited only by the number of basis functions required), for an arbitrary conformation, either isolated or embedded in the crystal. Effects such as the size of the basis set, level of theory and molecular conformation can be studied separately and the model validation can be based on a sufficiently large sample obtained under systematically varied conditions. The known phases can be imposed as constraints in the refinement of the theoretical data or used as independent observations.

X-ray charge densities, on the other hand, include crystal matrix and electron correlation effects, but the extent to which these effects can be extracted from experimental structure factors remains to be explored. A comparative analysis of crystalline densities, derived *via* periodic *ab initio* methods and multipole densities projected from the corresponding structure factors, revealed that the pseudoatom model is capable of retrieving the effect of intermolecular interactions (Spackman *et al.*, 1999). However, a closely related model study on the non-centrosymmetric structure of urea concluded that random errors introduced into the simulated structure factors influenced the multipole density extracted from the data to such an extent that the effect of intermolecular interactions could not be retrieved (de Vries *et al.*, 2000). Correlation effects are also difficult to evaluate separately since the experimental density does not correspond to a quantum state.

On the other hand, the theoretical approach is compromised by the level of approximation and the size of the basis set applied. While the density functional theory in principle includes electron correlation, the basis set utilized is necessarily limited if periodic calculations are to be considered. Our choice of the B3LYP method utilizing the 6-31G** basis set relies on the results of recent studies demonstrating that the density obtained by this method for small molecules compares well with those extracted from high-resolution X-ray data (Volkov, Abramov & Coppens, 2001; Abramov *et al.*, 1999). This is also a feasible level of theory to be applied in periodic calculations.

The first step towards a theoretical database of pseudoatoms is to explore the chemical transferability and the conformational invariance of deformation density parameters. These aspects of the theoretical approach to protein electrostatics have been analyzed by Koch *et al.* (1995) in terms of multipole moments based on Stone's distributed multipole method (Stone, 1981). The conformational dependence of topological atoms, defined within the framework of the theory of 'atoms in molecules' (AIM) (Bader, 1990), has been studied in amino acids by Matta & Bader (2000). The most recent related study is concerned with building the electrostatic potential of macromolecules from those of fragments (Zhong *et al.*, 2002).

2. Scattering model, data simulation and density refinement

In the Hansen–Coppens (1978) formalism, the molecular (crystalline) density is taken as a superposition of atomic densities:

$$\rho(\mathbf{r}) = \rho_c(r) + P_v\kappa^3\rho_v(\kappa r) + \sum_{l=0}^{l_{\max}} R_l(\kappa' r) \sum_{m=0}^{l} P_{lmp} d_{lmp}(\theta, \varphi), \quad (1)$$

where ρ_c and ρ_v are the spherical core and valence densities, respectively, while the aspherical (deformation) part is given as a linear combination of density-normalized real spherical harmonics (d_{lmp}, $p = \pm$) expressed in local frames attached to the nucleus. The core and spherical valence functions are Hartree–Fock densities (Clementi & Roetti, 1974), while the radial functions of the deformation part are simple Slater functions with energy-optimized exponents (Clementi & Raimondi, 1963). In addition to the conventional parameters, the number of valence electrons (P_v), the multipole populations (P_{lmp}) and radial screening parameters (κ and κ') are optimized in the least-squares refinement against structure-factor data. The number of independent multipole populations can be restricted by imposing local symmetries.

The X-ray data were generated from theoretical densities of isolated tripeptide molecules at the B3LYP level of theory utilizing the 6-31G** standard basis set using the *GAUSSIAN98* program package (Frisch *et al.*, 1998). Complex structure factors were calculated for reciprocal-lattice points corresponding to a cubic cell with 15 Å edges. The Fourier transform of the Gaussian orbital products was derived analytically as described by Chandler & Spackman (1978). Though the quantum-chemical calculations were done at the all-electron level, the contribution of the core orbitals was not included in the structure-factor generation. The pseudoatom model is a frozen-core approximation and the scattering factors corresponding to the core densities are derived from more extended calculations than those applied here for the molecular wave functions. While the difference between Slater-type and the six-term Gaussian basis functions is expected to be negligible, the exclusion of the core density avoids the possible source of bias associated with inadequate fitting of core features. An additional advantage of the approach is that data resolution is less of an issue, as the valence-only structure factor falls of more rapidly as a function of $\sin\theta/\lambda$ than the total structure factor. At a resolution of

Table 1
Relevant geometrical parameters of the peptide backbone of the central serine residue A = Ala, G = Gly, S = Ser.

C′═O 1.2214, C−O(H) 1.3309, Cα−N 1.4605, Cα−C′ 1.5097, Cα−Cβ 1.5501, C′−N 1.3534 Å, N−Cα−C′ 107.4, Cα−C′−N 117.5, Cα−C′═O 119.4, N−C′═O 123.1°. Units are Å and °. ψ and ϕ are the dihedral angles about the Cα−C′ and N−Cα bonds, respectively.

	ASA	ASG	ASS	SSA	SSG	ASSA
ϕ_2	−60.0	−61.5	−59.8	−59.3	−61.2	−62.1
ϕ_3	−66.1	−69.0	−67.5	−66.1	−69.2	−53.2
ϕ_4						−58.2
ψ_1	−52.3	−52.1	−52.6	−59.4	−59.4	−51.5
ψ_2	−53.6	−55.3	−53.0	−53.3	−54.8	−49.6
ψ_3						−49.0

Table 2
Local coordinate systems assigned to the peptide backbone atoms.

The Axis 1–Axis 2 plane is defined by two vectors; one pointing from Atom to Atom 1, the other from Atom 1 to Atom 2. The third axis is chosen to be perpendicular to this plane completing a right-handed system.

Atom	Atom 1	Axis 1	Atom 2	Axis 2
Cα	C′	*X*	N	*Y*
C′	O	*X*	N	*Y*
Cβ	Cα	*X*	O(H)	*Y*
N	C″	*X*	H(N)	*Y*
O	C′	*X*	N	*Y*

1.15 Å^{-1} – a cutoff value applied in the course of this study – the scattering power of the molecules studied drops below 0.1% of $F_{\rm val}(000)$ (the number of valence electrons). Furthermore, the crystallographic *R* value calculated for the valence-only structure factors provides a more sensitive figure-of-merit of the fit than the *R* factor for the total structure.

All refinements were done with a locally modified version of the *XDLSM* module of the *XD* program suite (Koritsanszky *et al.*, 1995). The theoretical phases ($\varphi_{\rm theo}$) were imposed as constraints during the fit by replacing, at each cycle, the model phases with the 'true' phases in the calculation of the structure-factor derivatives:

$$F = A\cos\varphi_{\rm theo} + B\sin\varphi_{\rm theo} \qquad (2a)$$

$$\partial F/\partial p = (\partial A/\partial p)\cos\varphi_{\rm theo} + (\partial B/\partial p)\sin\varphi_{\rm theo}, \qquad (2b)$$

where *p* stands for a valence-density parameter. This restriction resulted in significant changes in odd-order multipole populations compared to those derived *via* the corresponding unconstrained fit, without a noticeable deterioration in the convergence of the refinement. Unit weights were used throughout.

3. Transferability of the pseudoatoms of the peptide backbone

The structure factors were derived for molecular structures obtained by geometry optimizations at the molecular-mechanics level using the *MMFF* force field (Halgren, 1996) implemented in *Spartan5.0* (1997). Relevant geometrical parameters are listed in Table 1. The bond distances and angles are in close agreement with those widely used in constrained protein refinements and the ψ and ϕ torsion angles (dihedral angles about the Cα−C′ and N−Cα bonds, respectively)

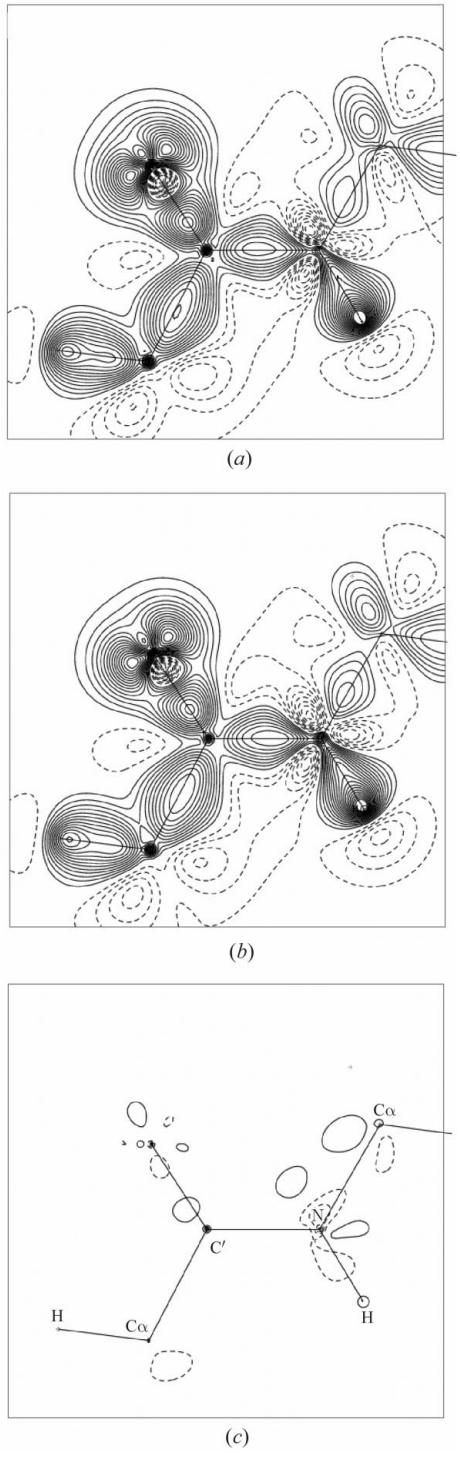


Figure 1
Deformation densities in the peptide plane of Ala–Ser–Ala. (*a*) Multipole model density fitted to the B3LYP theoretical density. (*b*) Multipole density constructed from the pseudoatom database. (*c*) The difference between (*a*) and (*b*). Contour intervals are at 0.05 e Å^{-3}.

Table 3
Averaged multipole populations of the peptide backbone atoms in the central serine residue of the tripeptides Ala–Ser–Ala, Ala–Ser–Gly, Ala–Ser–Ser, Ser–Ser–Ala and Ser–Ser–Gly.

Ordered pairs in the first column ($lm\pm$) designate spherical harmonics ($d_{lm\pm}$). Entries corresponding to the symmetry-forbidden spherical harmonics are left blank. Zero entries correspond to insignificant populations: $|P_{lmp}| < 3\sigma(P_{lmp})$. Numbers in parentheses are standard uncertainties based on unit weight (model errors). Bold-face entries in lower lines are the experimental populations reported earlier (Pichon-Pesme *et al.*, 1995).

P_{lmp}	Cα	C′	N	O	Hα	H(N)
0 0	4.02 (2)	4.05 (1)	5.13 (3)	6.15 (1)	0.825 (5)	0.81 (3)
1 1+	0.0	0.115 (7)	0.0	−0.098 (3)		
		0.12		**−0.10**		
1 1−	−0.080 (5)	0.0	0.040 (6)	−0.008 (1)		
1 0	0.092 (5)		0.0		0.068 (1)	0.11 (1)
						0.15
2 0	0.048 (5)	−0.303 (5)	0.066 (2)	−0.087 (1)		0.0
		−0.32		**−0.06**		
2 1+	−0.045 (7)		0.0			
2 1−	0.0		0.0			
2 2+	0.023 (4)	0.096 (3)	0.0	−0.095(1)		
		0.13		**−0.06**		
2 2−	0.058 (4)	0.020 (5)	−0.018 (3)	−0.009 (2)		
3 0	0.0		0.0	0.0		
3 1+	−0.161 (6)	0.0	0.014 (4)	0.0		
	−0.16					
3 1−	−0.213 (2)	0.024 (3)	0.0	−0.004 (1)		
	−0.21					
3 2+	0.0		0.0			
3 2−	0.0		0.0			
3 3+	0.195 (3)	0.424 (6)	0.12 (1)	0.0		
	0.25	**0.43**	**0.27**			
3 3−	−0.062 (4)	0.0	0.009 (2)	0.0		
κ	1.006 (12)	1.002 (1)	0.9972 (4)	0.9869 (2)	1.236 (1)	1.253 (26)
κ'	0.920 (4)	0.850 (3)	1.126 (36)	1.003 (5)	1.616 (12)	1.537 (40)

correspond to α-helix-type secondary structure. Each non-hydrogen pseudoatom was represented up to the octupole level ($l_{max} = 3$) in the multipole expansion in local frames given in Table 2 for the atoms in the central peptide residue. Local mirror symmetry was imposed on the carbonyl C′ and O atoms of the peptide backbone, on the C and O atoms of the —COH group in the side chain of the serine residue and $m3$ symmetry on the methyl group of the alanine residue. The deformation density of the hydrogen atoms in the C—H bonds was described by a bond-directed dipole ($l = 1, m = 0$; one for each type of hydrogen atom), while a quadrupole of cylindrical symmetry ($l = 2, m = 0$) was added to the H atoms of the N—H and O—H bonds. The choice of the site coordinate systems and the local symmetries imposed are consistent with those used in building the experimental databank (Pichon-Pesme *et al.*, 1995), thus allowing direct comparison of the experimental and theoretical results. All atoms of the peptide backbone were treated independently but chemically equivalent side-chain atoms were constrained to have the same deformation density. Both κ and κ' radial screening parameters were refined, even for hydrogen atoms. Each tripeptide was constrained to remain neutral during the refinement. Final valence R values were in the range 0.010–0.012 with reflection-to-variable ratios of over 70.

Table 3 shows multipole populations for the atoms of the peptide bond of the central serine residue obtained by averaging the results of multipole refinements of theoretical X-ray data of tripeptides in their neutral form (Ala–Ser–Ala, Ala–Ser–Gly, Ala–Ser–Ser, Ser–Ser–Ala and Ser–Ser–Gly). The relatively small standard deviations (calculated with respect to the sample's mean) indicate a reasonable invariance of the multipole populations with change of the terminal residues of the tripeptides. These results suggest that the density asphericity at the Cα-, C′-, N- and O-atom sites can be accounted for by, respectively, 10, 6, 6 and 6 extra parameters (with respect to the spherical-atom model). Of these, 5, 4, 3 and 3 are the minimum number of multipoles needed to describe the dominant features of the valence deformation density in each of the atom types, respectively. Bold-faced entries in Table 3 are the experimental multipole populations reported earlier (Pichon-Pesme *et al.*, 1995). The two methods lead to practically identical carbon pseudoatoms of the peptide unit in terms of the dominant terms of the multipole expansion (P_{31+}, P_{31-} and P_{33+} for Cα and P_{11+}, P_{20}, P_{22+} and P_{33+} for C′), while for the peptide N atom a considerable difference is found for the 33+ spherical harmonic density function, which is the major contributor to its asphericity. The aspherical density of the experimental O and N pseudoatoms is represented, respectively, by three (d_{11+}, d_{20} and d_{22+}) and one (d_{33+}) significantly populated terms. The extra functions in Table 3 may be needed for an accurate description of the valence density.

In Fig. 1, the multipole deformation densities in the plane of the central peptide group in Ala–Ser–Ala obtained by direct fit (Fig. 1*a*) are compared with those constructed from the averaged pseudoatoms (Fig. 1*b*). The two maps are practically equivalent, the differences (Fig. 1*c*) being in the range of experimental accuracy of X-ray charge densities. The quality

Table 4
Critical-point properties of selected bonds in the peptide backbone of Ala–Ser–Ala.

First line: B3LYP/6-31G** theoretical density; second line: multipole density fitted to the theoretical density; third line: density based on averaged multipole populations. Units are electrons and Å.

Bond	ρ_{BC}	$\nabla^2\rho_{BC}$	λ_1	λ_2	λ_3
C′═O	2.74	1.3	−25.7	−23.9	50.3
	2.83	−31.3	−24.5	−22.5	15.8
	2.78	−26.9	−24.1	−22.1	19.3
N−Cα	1.77	−17.8	−12.5	−11.8	6.5
	1.67	−6.5	−11.6	−10.7	15.7
	1.66	−6.8	−11.4	−10.5	15.1
C′−N	2.21	−23.0	−17.5	−15.5	10.0
	2.18	−18.8	−17.4	−14.6	13.2
	2.17	−19.8	−17.2	−14.4	11.8
Cα−C′	1.80	−16.4	−13.2	−12.5	9.3
	1.78	−13.2	−12.9	−11.4	11.0
	1.77	−12.9	−12.5	−11.3	10.9
Cα−Cβ	1.64	−13.5	−11.3	−10.9	8.6
	1.57	−9.2	−10.5	−9.9	11.3
	1.57	−9.0	−10.4	−9.9	11.4

ρ_{BC} and $\nabla^2\rho_{BC}$ are the density and the Laplacian, respectively, at the bond critical point (where $\nabla\rho = 0$), λ_1 and λ_2 are bond-perpendicular and λ_3 bond-parallel curvatures ($\nabla^2\rho_{BC} = \lambda_1 + \lambda_2 + \lambda_3$).

of the constructed density was evaluated also by comparing the corresponding structure factors with those derived directly from the wave function. Since the superposition of databank pseudoatoms does not necessarily give rise to a neutral pseudomolecule, the valence populations (P_v) were readjusted prior to the structure-factor calculation to obtain the correct number of electrons. For the Ala–Ser–Ala molecule, an R value of 1.1% is obtained with the databank densities, that is, they fit the structure factors as well as the model based on the direct refinement. The same conclusion is reached in the analysis of the Ala–Ser–Ser–Ala tetrapeptide.

3.1. Bond-topological properties

Table 4 lists bond-topological properties (Bader, 1990) for Ala–Ser–Ala obtained directly from the theoretical density, from the multipole density fitted to it and from the density constructed on the basis of the theoretical databank. While the density and the bond-perpendicular curvatures (λ_1, λ_2) at the bond critical points (BCP's) agree well, large discrepancies are found between the bond-parallel curvatures (λ_3) derived from the primary density based on the wave function and those obtained from the multipole density, especially for the polar bonds (C′═O and C−N). This observation is in line with a previous analysis, and attributed to the bias introduced by inadequate radial functions of the deformation model (Volkov, Abramov, Coppens & Gatti, 2000; Volkov & Coppens, 2001). As both are based on the same radial functions, the bond topological indices of the databank-constructed density are in almost perfect agreement with values derived from the fitted multipole density. Fig. 2 displaying the Laplacian distribution in the peptide bond of the central serine residue demonstrates that this equivalence is not only satisfied at the BCP: both maps show the same features characteristic to bonded and non-bonded valence-shell charge concentrations.

Table 5
Topological atomic net charges (q), dipole (d_{lmp}) and quadrupole moments (Q_{lmp}) in spherical tensor form for the atoms in the central peptide group in Ala–Ser–Ala.

Upper line: calculated from the B3LYP density; middle line: calculated from the multipole density fitted to the theoretical data; lower line: calculated from the databank parameters. $|\mathbf{d}| = \{\sum(d_{1m})^2\}^{1/2}$, $Q = \{\sum(Q_{2m})^2\}^{1/2}$. Entries are in atomic units.

	Cα	C′	N	O	Hα	H(N)
q	0.41	1.55	−1.25	−1.20	0.05	0.43
	0.24	1.30	−0.89	−1.03	0.07	0.32
	0.22	1.31	−0.91	−1.04	0.09	0.35
d_{11+}	0.44	0.50	−0.20	−0.53	0.02	−0.01
	0.40	0.45	−0.10	−0.29	0.01	0.0
	0.39	0.49	−0.08	−0.31	0.01	0.0
d_{11-}	−0.03	−0.52	−0.09	0.02	−0.01	0.0
	0.07	−0.29	−0.01	−0.03	0.0	0.0
	0.01	−0.35	−0.09	−0.02	0.0	0.0
d_{10}	0.02	0.03	0.03	−0.04	0.12	0.18
	0.12	0.02	0.03	−0.01	0.13	0.14
	0.11	0.02	0.02	−0.01	0.13	0.14
$\|d\|$	0.45	0.72	0.23	0.53	0.12	0.18
	0.43	0.53	0.11	0.30	0.13	0.14
	0.41	0.60	0.13	0.32	0.14	0.14
Q_{20}	−0.20	−0.79	−1.09	−0.09	0.16	0.02
	−0.12	−0.70	−1.22	−0.37	0.20	0.10
	−0.18	−0.71	−1.25	−0.36	0.20	0.09
Q_{21+}	−0.03	−0.02	−0.06	−0.02	−0.02	0.01
	0.16	−0.01	−0.05	−0.01	−0.02	0.0
	0.08	−0.01	−0.06	−0.01	−0.02	0.0
Q_{21-}	0.04	0.02	0.01	−0.02	0.01	0.0
	0.15	0.02	−0.01	0.0	0.0	0.0
	0.13	0.02	0.02	0.01	0.0	0.0
Q_{22+}	0.35	0.25	−0.21	0.21	0.0	−0.01
	0.25	0.26	−0.09	0.21	0.01	−0.01
	0.24	0.26	−0.09	0.18	0.01	−0.01
Q_{22-}	0.01	0.38	0.05	−0.01	0.0	0.0
	0.11	0.30	0.17	−0.03	0.01	0.0
	0.08	0.34	0.17	−0.01	0.01	0.0
Q	0.40	0.91	1.11	0.23	0.16	0.03
	0.37	0.81	1.24	0.42	0.20	0.11
	0.35	0.83	1.27	0.40	0.20	0.09

3.2. Atomic electrostatic moments

The transferability of the pseudoatoms was analyzed also in terms of electrostatic moments based on the AIM partitioning scheme (Bader, 1990). In Table 5, atomic net charges (q), components and magnitudes of dipole (d) and quadrupole moments (Q) are listed for the atoms in the central serine moiety of the ASA molecule. The integration of the theoretical density was performed with the *PROMEGA* module of the *AIMPACK* program suite (Cheeseman *et al.*, 1992), while the multipole densities were integrated with *TOPXD* (Volkov, Gatti, Abramov & Coppens, 2000).

For direct comparison of the results, the spherical tensor atomic moments (Price *et al.*, 1984) in Table 5 are expressed in local frames used in the refinements (Table 2). As expected on the basis of bond-topological data (Table 4), there are significant discrepancies between theoretical moments and those derived from the fitted density. While the B3LYP charges indicate more polarized bonds (larger charge separations)

Table 6
Geometrical data of the central serine residue in Ala–Ser–Ala.

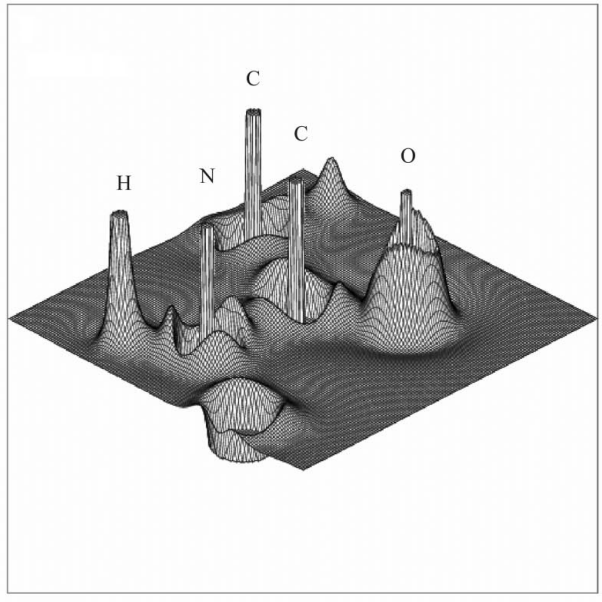
ψ and ϕ refer to dihedral angles about the Cα—C′ and N—Cα bonds, respectively. Units are Å and °. The last column gives the number of structures with a particular backbone conformation over which average distances and angles were calculated.

ϕ	ψ	N—Cα	Cα—C′	N—Cα—C′	No. of structures
−45	−15	1.434	1.513	114.33	5
−45	−45	1.459	1.512	112.84	7
−75	165	1.455	1.530	109.45	6
−75	135	1.463	1.515	109.62	7
−75	105	1.496	1.532	106.89	1
−75	45	1.464	1.516	111.64	2
−75	−15	1.459	1.528	113.14	24
−75	−45	1.462	1.522	111.36	41
−105	165	1.460	1.517	110.43	3
−105	135	1.454	1.536	110.16	3
−105	105	1.462	1.525	107.71	7
−105	45	1.464	1.516	111.64	2
−105	−15	1.456	1.523	114.82	8
−105	−30	1.457	1.519	113.52	3
−105	−45	1.457	1.519	113.52	1
−105	−60	1.457	1.519	113.52	1
−135	105	1.467	1.537	107.41	2
−135	75	1.466	1.541	108.32	6
−135	15	1.463	1.523	113.20	2
−165	165	1.452	1.533	105.86	3

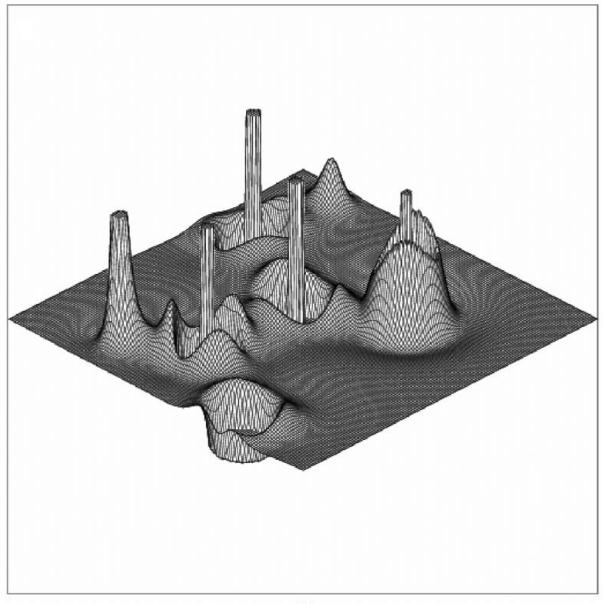
than those calculated from the multipole density, the peptide unit is neutral in both representations. At each atom, the bond-directed component ($d_x = d_{11+}$) and the magnitude ($|\mathbf{d}|$) of the local dipole moment is underestimated by the multipole model. The quadrupole moments appear to be better preserved by the multipole projection, except for the N atom. The important observation is that the dipole and quadrupole moments derived from the simulated density are practically equal in value to those corresponding to the fitted multipole density.

4. Conformation dependence of the backbone pseudoatoms

The conformational flexibility of a peptide backbone arises from the low-energy-barrier rotations about the Cα—C′ and Cα—N single bonds. To examine the conformation dependence of the theoretically derived multipole parameters, structure-factor data sets were generated for Ala–Ser–Ala from B3LYP/6-31G** wave functions calculated at different geometries. The ψ(Cα—C′)–ϕ(N—Cα) conformation space associated with the serine residue was represented by 20 points given in Table 6. The coordinates cover regions characteristic to both the α-helix- and β-sheet-type secondary structures. The dependence of the backbone Cα—C′ and N—Cα bond lengths and the N—Cα—C′ bond angles on the ψ(Cα—C′) and ϕ(N—Cα) torsion angles was accounted for as described by Jiang *et al.* (1997). Their analysis includes high-resolution structural data of 43 oligopeptides retrieved from the CSD and arranged according to (ψ,ϕ) regions of 30 × 30°. The Cα—C′ and N—Cα distances and the N—Cα—C′ bond angles assigned to each (ψ,ϕ) interval are average values over the peptide population found in that interval. Simulated data sets derived for each molecular geometry were fitted using the model described above, after which the resulting multipole populations for chemically equivalent atoms were averaged. Fig. 3(*a*) displays the evolution of the net charge of each pseudoatom in the serine peptide group with a change in the ψ(Cα—C−) torsion angle from −60 to +165° and ϕ(N—Cα) = −105°. The two C atoms appear to be the most sensitive to the changes in the torsion angle and a small charge transfer seems to be taking place between these sites. To a good approximation, the overall charge of the peptide group remains unchanged when the conformation is changed from the



(*a*)



(*b*)

Figure 2
Relief map of the Laplacian distribution in the peptide plane of Ala–Ser–Ala (*a*) from the multipole model density fitted to the theoretical density, (*b*) constructed from the pseudoatom database

β-sheet to the α-helix. Higher-order populations fluctuate somewhat as $\psi(C\alpha-C')$ increases. The changes displayed for the octupole populations ($P31-$ and $P31+$) of the Cα atom are typical and do not seem to reveal any systematic trend (Fig. 3*b*).

Table 7 lists multipole populations and radial scaling parameters (κ and κ') for the atoms in the serine residue averaged over 20 conformations. As the standard deviations in Tables 3 and 7 are comparable, variations in the populations upon conformational change (amounting typically to 10% with respect to mean values) are in the range of those found upon changing the sequence along the peptide backbone.

5. Summary

As pointed out earlier (Pichon-Peshme *et al.*, 1995), there are three important points to consider when building a database of pseudoatoms applicable to model macromolecular densities: (i) chemical transferability; (ii) conformation invariance; and (iii) the effect of intermolecular interaction. The former two aspects were explored in this study, while the third one will be the objective of a future study in which results of periodic crystal calculations will be compared to those obtained for the corresponding isolated molecules.

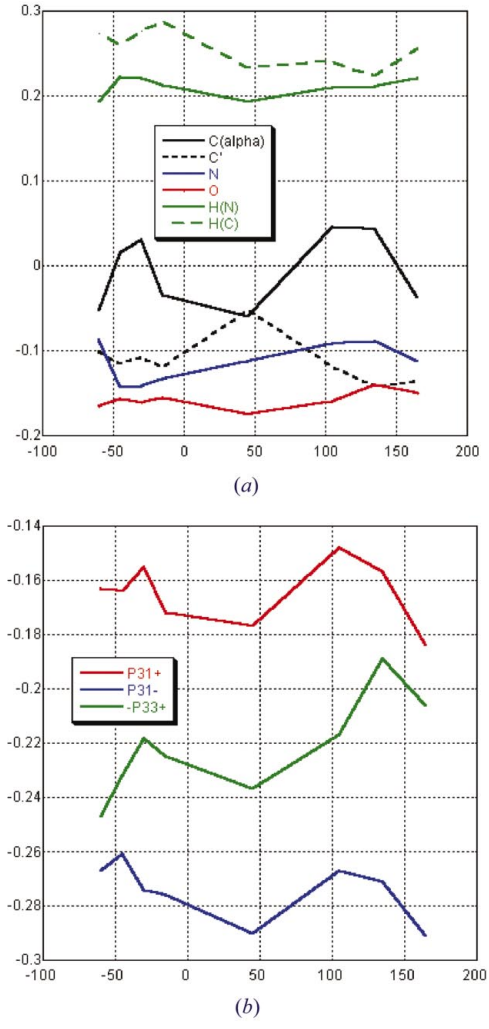


Figure 3
(*a*) The variation of the atomic net charges with the $\psi(C\alpha-C')$ torsion angle of the central serine residue in Ala–Ser–Ala [$\phi(N-C\alpha) = -105°$]. (*b*) The variation of the dominant multipole populations of the central Cα atom with the $\psi(C\alpha-C')$ torsion angle of the central serine residue in Ala–Ser–Ala [$\phi(N-C\alpha) = -105°$].

Chemical transferability seems to be an intrinsic property of the pseudoatoms in the peptide backbone for the tripeptides considered here. The density of the Ala–Ser–Ala tripeptide, as built from the limited databank, resembles quantitatively that derived from refinement of the theoretical structure factors, not only in terms of deformation features and associated structure factors but also in terms of bond topological figures, the Laplacian distribution and AIM moments. For the C atoms (Cα and C$'$), the results clearly support the experimental method of deriving transferable multipole populations. However, noticeable discrepancies between theoretical and experimental deformation parameters were found for the N and H atoms. In the analysis of these deviations, several factors should be considered. First, the well known difficulty of the experimental method in retrieving the deformation density at the hydrogen sites should be mentioned. Furthermore, the error free, static, valence-only structure factors analyzed in this study allow refinement of the radial screening parameters of the deformation density (κ'), even for the H atoms, while in X-ray charge-density refinements the same procedure often fails. Since the N and H atoms for which discrepancies occur are involved in hydrogen bonding, the effect of intermolecular interactions on the density, which is not accounted for in our current calculations, is a possible source of disagreement. Last but not least, basis-set inadequacies have to be explored before a final conclusion can be drawn.

The small charge redistribution with geometry change is reflected in the multipole density, but the dependence of the pseudoatom populations on the $\psi(C\alpha-C)$ and $\phi(N-C\alpha)$ torsion angles of the central residue in ASA is too complex to be described in terms of a few parameters. The multipole populations corresponding to the 20 points of the Ramachandran domain that have been considered are scattered in a relatively narrow range of not more than 10–15%, their standard deviations are as small as those found upon variation of the terminal residue. It follows that the conformational dependence of the pseudoatoms is relatively small but not completely negligible for the backbone carbon atoms.

Relevant to the above discussion is the degree of locality of the pseudoatomic density and the flexibility of the deformation radial functions. The much discussed question to what extent individual multipole populations can be considered as 'atomic properties' should be revisited if better transferability is to be achieved. Model inadequacies mentioned above are being analyzed in our laboratory. The possibility of multipole representation of pseudoatoms derived *via* the stockholder partitioning scheme (Hirshfeld, 1977) and the design of radial functions capable of retrieving bond-parallel curvatures is being investigated.

Table 7
Conformational averages of the multipole population of the peptide backbone atoms in Ala–Ser–Ala.

See headnote of Table 3.

P_{lmp}	Cα	C′	N	O	Hα	H(N)
0 0	3.97 (2)	4.12 (1)	5.124 (8)	6.159 (4)	0.770 (9)	0.786 (4)
1 1+	−0.055 (4)	0.099 (5)	0.0	−0.099 (1)		
1 1−	−0.089 (5)	−0.049 (10)	0.030 (2)	−0.006 (2)		
1 0	0.082 (10)		−0.011 (2)		0.070 (2)	0.113 (1)
2 0	0.085 (6)	−0.292 (7)	0.090 (2)	−0.068 (1)		0.039 (2)
2 1+	0.0		0.032 (1)			
2 1−	−0.034 (4)		0.021 (1)			
2 2+	0.026 (3)	0.096 (2)	−0.007 (1)	−0.078 (1)		
2 2−	0.043 (7)	0.018 (4)	−0.022 (2)	0.0		
3 0	0.0		−0.037 (2)	0.0		
3 1+	−0.147 (6)	0.0	0.016 (2)	−0.009 (2)		
3 1−	−0.264 (4)	0.0	−0.012 (1)	0.0		
3 2+	−0.017 (4)		−0.007 (2)			
3 2−	0.028 (5)		0.034 (2)			
3 3+	0.212 (5)	0.448 (8)	0.181 (3)	0.025 (1)		
3 3−	−0.028 (7)	0.0	−0.008 (1)	−0.005 (1)		
κ	1.005 (1)	0.988 (1)	0.990 (1)	0.986 (1)	1.258 (4)	1.207 (1)
κ'	0.902 (2)	0.915 (6)	1.158 (3)	1.592 (19)	1.490 (10)	

Nevertheless, the results reported here are promising as they support 'atomicity' as a practical concept to be applied in charge-density studies of macromolecules. We have described a strategy that can be adopted to identify chemically unique atoms in biomacromolecules. The simplest approach considers only the primary structure and defines atoms having the same first neighbor to be equivalent. At this level, a division into peptide groups and side groups emerges naturally. Since all but the Cα atom in a peptide backbone have the same connectivity, the transferability of the peptide groups is expected to break down only locally, *i.e.* at the Cα sites. It is anticipated that the effect of a side group on the Cα atom is determined by the chemical nature and connectivity of the Cβ atom. This argument implies five types of Cα atoms corresponding to the five distinct Cβ atoms [H(GLY), CH_3, CH_2, CH and CH_2(PRO)] found in naturally occurring proteins.

The same strategy can be followed for defining transferable side-chain atoms. In the side groups of the 20 amino acids found in genetically encoded proteins, there are, all together, 206 atoms (118 H, 68 C, 9 N, 9 O and 2 S) that can be reduced, according to the above approach, to 36 types (7 H, 19 C, 4 N, 4 O and 2 S). More atom types are to be introduced to account for ionized sites.

We conclude that construction of the electron density of peptides and proteins on the basis of atom-centered multipole formalism is a realistic goal. Such a density can be used to improve high-resolution protein refinement. It has the potential of predicting electrostatic properties of large molecules and estimating the Coulomb interactions between macromolecules and substrates.

Support of this work by the National Institutes of Health (GM56829) and the National Science Foundation (CHE9981864) is gratefully acknowledged.

References

Abramov, Yu., Volkov, A. & Coppens, P. (1999). *Chem. Phys. Lett.* **311**, 81–86.
Bader, R. F. W. (1990). *Atoms in Molecules – a Quantum Theory*. Oxford University Press.
Brock, C. P., Dunitz, J. D. & Hirshfeld, F. L. (1991). *Acta Cryst.* B**47**, 789–797.
Chandler, G. S. & Spackman, M. A. (1978). *Acta Cryst.* A**34**, 341–343.
Cheeseman, J., Keith, T. A. & Bader, R. F. W. (1992). *AIMPACK Program Package*. McMaster University, Hamilton, Ontario, Canada.
Clementi, E. & Raimondi, D. L. (1963). *J. Chem. Phys.* **38**, 2686–2692.
Clementi, E. & Roetti, C. (1974). *At. Data Nucl. Data Tables*, **14**, 177.
Fernández-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., Stratmann, R. E. Jr, 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., 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., Gonzalez, C., Challacombe, M., Gill, P. M. W., Johnson, B. G., Chen, W., Wong, M. W., Andres, J. L., Head-Gordon, M., Replogle E. S. & Pople, J. A. (1998). *Gaussian98* (Revision A.***x***). Gaussian, Inc., Pittsburgh, PA, USA.
Halgren, T. A. (1996). *J. Comput. Chem.* **17**, 490–519.
Hansen, N. K. & Coppens, P. (1978). *Acta Cryst.* A**34**, 909–921.
Hirshfeld, F. L. (1971). *Acta Cryst.* B**27**, 769–781.
Hirshfeld, F. L. (1977). *Theor. Chim. Acta*, **44**, 129–138.
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.
Jiang, X., Yu, C.-H., Cao, M., Newton, S. Q., Paulus, E. F. & Schäfer, L. (1997). *J. Mol. Struc.* **403**, 83–93.

Koch, U., Popelier, P. L. A. & Stone, A. J. (1995). *Chem. Phys. Lett.* **238**, 253–260.
Koritsanszky, T., Howard, S. T., Richter, 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.
Matta, C. F. & Bader, R. F. W. (2000). *Prot. Struct. Funct. Genet.* **40**, 310–329.
Pichon-Pesme, V., Lecomte, C. & Lachekar, H. (1995). *J. Phys. Chem.* **99**, 6242–6250.
Price, S. L., Stone, A. J. & Alderton, M. (1984). *Mol. Phys.* **52**, 987–1001.
Spackman, M. A., Byrom, P. G., Alfredsson, M. & Hermansson, K. (1999). *Acta Cryst.* A**55**, 30–47.
Spartan5.0 (1997). Wavefunction, Inc., 18401 Von Karman Ave, 370, Irvine, CA 92612, USA.
Stone, A. J. (1981). *Chem. Phys. Lett.* **83**, 233–239.
Volkov, A., Abramov, Y. A. & Coppens, P. (2001). *Acta Cryst.* A**57**, 272–282.
Volkov, A., Abramov, Y., Coppens, P. & Gatti, C. (2000). *Acta Cryst.* A**56**, 332–339.
Volkov, A. & Coppens, P. (2001). *Acta Cryst.* A**57**, 395–405.
Volkov, A., Gatti, C., Abramov, Y. & Coppens, P. (2000). *Acta Cryst.* A**56**, 252–258.
Vries, R. Y. de, Feil, D. & Tsirelson, V. G. (2000). *Acta Cryst.* B**56**, 118–123.
Zhong, S., Dadarlat, V. M., Glaeser, R. M., Head-Gordon, T. & Downing, K. H. (2002). *Acta Cryst.* A**58**, 162–170.