

Ab Initio Quality Electrostatic Atomic and Molecular Properties Including Intermolecular Energies from a Transferable Theoretical Pseudoatom Databank

Anatoliy Volkov,^{*,†} Xue Li,[†] Tibor Koritsanszky,[‡] and Philip Coppens^{*,†}

Department of Chemistry, State University of New York at Buffalo, Buffalo, New York 14260-3000, and Department of Chemistry, Middle Tennessee State University, MTSU Box 0395, 1301 East Main Street, Murfreesboro, Tennessee 37132

Received: December 22, 2003; In Final Form: February 17, 2004

The development of a theoretical databank of transferable pseudoatoms for fast prediction of the electron densities and related electronic properties of proteins is described. Chemically unique pseudoatoms identified on the basis of common connectivity and bonding are extracted from *ab initio* molecular densities of a large number of small molecules using a least-squares projection technique in Fourier transform space. The performance of the databank is evaluated by comparison of the electron densities and electrostatic properties of the amino acids GLN, SER, and LEU and their dimers with those obtained from molecular calculations on the same test compounds. It is found that deformation density bond peaks are reproduced to within 0.02–0.10 e/Å³, whereas electrostatic potentials, bond critical point indices, atomic charges, and molecular moments show differences with results from calculations performed directly on the test molecules which are comparable with or smaller than the spread of the values between different *ab initio* methods (Hamiltonian, basis set, etc.). The order of intermolecular electrostatic interaction energies for selected dimers of the test compounds are well reproduced, though the results are always smaller, by about 25 kJ/mol on average, than electrostatic energies from Morokuma–Ziegler decomposition of the total interaction energy evaluated with the ADF program. The difference is attributed to the limitations of the Buckingham-type approximation for electrostatic interactions, used in the current study, which assumes nonoverlapping charge densities. The consistency achieved by the pseudoatom databank is much better than that obtained with the AMBER99, CHARMM27, MM3, and MMFF94 force fields, which sometime overestimate, sometimes underestimate, the electrostatic interaction energy. The electrostatic component of the binding energies (directly related to the enthalpy of sublimation) of molecules in crystals, calculated based on the databank parameters, agree within 25–60 kJ/mol with the total binding energies evaluated *ab initio* at the Density Functional level of theory, even though the exchange–repulsion and dispersion terms have not been taken into account in the databank values.

1. Introduction

Electrostatic forces play an important role in the process of protein folding and binding.¹ While significant progress in the field of molecular mechanics toward improving the accuracy of prediction of electrostatic properties of proteins has been made in recent years, these methods are still inferior to *ab initio* type calculations.² On the other hand, the simplicity of the model offered by a force-field approach and the minimal computational effort required for calculations of systems of thousands of atoms and larger still make these methods an attractive choice, especially when *ab initio* type approaches are not applicable because of exceedingly high computational cost. Hence, one of the main issues in the development of a new generation of force fields is the development of a more reliable model for the calculation of electrostatic interactions.

The electrostatic component in the traditional force field is calculated on the basis of an atom-centered point-charge model.

Although extremely simple, this model is not flexible enough to describe aspherical features of molecular charge distributions.²

An alternative is to supplement point charges with higher electrostatic moments. A number of methods for calculation of atomic moments have been explored. These include the following:

(1) Partitioning based on wave function (for example, the distributed multipole analysis^{3,4} and cumulative atomic multipole moments⁵) or the total molecular density (quantum theory of atoms in molecules,⁶ stockholder concept,⁷ etc.).

(2) Fitting to the one-electron density matrix⁸ or electrostatic potential^{9,10} obtained from high-quality *ab initio* calculations.

The combination of the multipole-based electrostatic model with force-field repulsion, dispersion, and torsion components may significantly improve the predicting power of the method.¹¹

The success of a partitioning scheme rests on its predicting power, that is, on the degree of transferability of the atomic properties obtained. Because perfect transferability is unachievable, any method must compromise between mathematical rigor and conceptual simplicity. The latter aspect is a basic require-

* Corresponding authors. Fax: +17166456948. E-mail addresses: volkov@chem.buffalo.edu (A.V.); coppens@buffalo.edu (P.C.).

[†] State University of New York at Buffalo.

[‡] Middle Tennessee State University.

ment for molecular modeling. Wave-function-based schemes offer the widest variety of atomic properties but at the expense of computational cost involved in the evaluation of these properties. In addition, wave functions derived via the variational principle do not necessarily yield electron-density-based molecular properties of the same accuracy as that achieved for the energy.

In view of the Hohenberg–Kohn theorem,¹² the one-electron density, a physical observable, is the property the ultimate partitioning could and should be based on. Indeed, the proof that the external potential (which is unambiguously defined by the nuclear structure) is a unique functional of the electron density, is an explicit statement for the existence of atoms in molecules.

Bader's quantum-topological theory⁶ allows a unique definition of discrete-boundary atoms through their density-based observable properties. The transferability of topological atoms has been analyzed by a number of authors. In earlier studies on a series of carbocations¹³ and hydrocarbons,¹⁴ properties of quantum atoms, such as net charges, dipole moments, energies, and volumes were found to be fairly transferable. The theoretical construction of a polypeptide using topological peptide groups also yielded accurate total molecular properties.¹⁵ In another study,¹⁶ the extent of transferability was measured in terms of a similarity index, approaching unity/zero for identical/different moieties. Values in the range of 0.95–0.99 were obtained for this figure for chemically equivalent atoms in simple hydro- and fluorocarbons. The conformation dependence of quantum-atom properties has been shown to be marginal.^{17,18}

Unfortunately, the discrete partitioning inherent in the topological theory does not allow for a straightforward construction of continuous real-space functions such as the density itself. This is due to mismatches in the interatomic surfaces of transferred atoms, fragments that may leave open holes (or gaps) in the density to be reconstructed.

The total crystalline density can, in principle, be reconstructed from experimental X-ray structure factor amplitudes. However, to account for thermal motion and the missing phases, the interpretation of Bragg data necessitates atomic level modeling of the crystalline density to be measured. This inherent connection between measurability of the density and its description in terms of atoms is the conceptual basis of our approach. The conventional elucidation of X-ray data invokes the concept of the promolecule, the superposition of isolated atomic densities that satisfy maximal transferability but neglect deformations due to bonding. An extension of this fuzzy-type partitioning¹⁹ must account for density deformations but with a minimal loss of locality if transferability is to be maintained.

In the Hansen–Coppens formalism^{20,19} adopted in this work, the static density is described by a superposition of aspherical *pseudoatoms* composed of nucleus-centered density units:

$$\rho_{\mathbf{k}}(\mathbf{r}) = P_c \rho_c(r) + P_v \kappa^3 \rho_v(\kappa r) + \kappa'^3 \sum_{l=1}^4 R_l(\kappa' r) \sum_{m=1}^l P_{lm\pm} d_{lm\pm}(\mathbf{r}/r)$$

where ρ_c and ρ_v are spherically averaged free-atom Hartree–Fock core and valence densities normalized to one electron, $d_{lm\pm}$ are density-normalized real spherical harmonics, and R_l are radial density functions of Slater-type:

$$R_l(\kappa' r) = \kappa'^3 \frac{\zeta^{n_l+3}}{(n_l + 2)!} (\kappa' r)^{n_l} \exp(-\kappa' \zeta r)$$

with energy-optimized ζ exponents.²¹ The κ and κ' dimensionless expansion–contraction parameters, along with the populations P_v and $P_{lm\pm}$, can be refined in the fitting procedure against experimental structure factor amplitudes.

The fair transferability of experimental density parameters in small peptides found by Lecomte et al.^{22–24} made it possible to build a databank of pseudoatoms commonly occurring in proteins. A number of successful applications of the databank to the refinement of high-resolution X-ray data of large systems have been reported.^{24–26} The results indicate that density modeling beyond the spherical-atom formalism can lead not only to significantly improved atomic displacement parameters and molecular geometry, but also to chemically meaningful electrostatic potentials of peptide residues. But we note that no attempt has been made to systematically analyze the accuracy of electrostatic properties evaluated with the experimental databank, as it was always used as a starting point for the experimental charge density determination, rather than as a stand-alone tool for calculation of electrostatic properties.

In recent work by Koritsanszky et al.²⁷ an alternative route to databank building was suggested. The parameters of chemically unique pseudoatoms were derived directly from *ab initio* theoretical densities of tripeptides with a systematically varied central residue. This procedure parallels that applied in the experimental databank, however, it involves the fitting of simulated structure factors obtained via Fourier transform of the wave-function-based density. The method can lead to parameters free of bias, which is practically unachievable for experimental estimates because of the lack of phase information, the inadequate treatment of thermal smearing, disorder, and systematic experimental errors. Furthermore, the simulation allows a great variety of atom types and systems to be studied as the incorporation of new atom types into the theoretical databank requires much less effort than is the case for the experimental approach. The pseudoatoms obtained for a sample consisting of a limited number of peptides were shown to be highly transferable and fairly invariant under conformational changes.

Following this approach, we have built an extended database for C, H, N, and O pseudoatoms applicable to construction of the electron density of proteins. Atoms exhibiting the same local structure (connectivity and bonding) were considered to be chemically equivalent and their parameters were averaged. Here we explore the transferability of the database atoms in terms of the statistical distribution of the density parameters and their performance in density prediction. The validation of the library involves the construction of amino acid molecules not included in the databank building and the comparison of their densities with those directly derived from wave functions or extracted from the corresponding theoretical structure factors. This comparison is done in terms of local and integrated topological properties, molecular moments, electrostatic potentials, electrostatic interaction energies in molecular dimers, and electrostatic binding energies of molecules in crystals. The electrostatic components of the interaction energies are also compared to those obtained with several commonly used force-field approaches.

2. Pseudoatom Databank

Theoretical Calculations. To construct the databank, single-point calculations on selected small molecules (Table A.1, Appendix A) were performed with the Gaussian 98 (G98) program²⁸ at the density functional level of theory (DFT) using a standard split-valence double-exponential 6-31G** basis set

TABLE 1: Pseudoatom Parameters for Oxygen Atoms in the CO₂⁻ Group in Compounds Used in the Construction of the Databank

ID ^a	P_v	κ	κ'	P_{11+}	P_{11-}	P_{20}	P_{22+}	P_{22-}	P_{31+}	P_{33+}	P_{33-}	P_{44+}
1	6.337	0.976	1.17	-0.088	0.005	-0.039	-0.078	0.011	-0.009	0.014	0.003	0.008
1	6.352	0.976	1.16	-0.085	0.013	-0.041	-0.072	0.013	-0.006	0.019	0.003	0.010
2	6.306	0.978	1.16	-0.091	0.007	-0.040	-0.071	0.009	-0.009	0.019	0.002	0.008
2	6.321	0.977	1.15	-0.088	0.012	-0.040	-0.075	0.006	-0.006	0.019	0.003	0.007
4	6.333	0.976	1.19	-0.082	0.005	-0.035	-0.069	0.001	-0.008	0.019	-0.001	0.007
4	6.334	0.976	1.16	-0.091	0.005	-0.038	-0.075	0.005	-0.007	0.021	0.005	0.007
5	6.316	0.978	1.19	-0.082	0.006	-0.040	-0.071	0.007	-0.008	0.016	0.002	0.010
5	6.348	0.976	1.16	-0.089	0.011	-0.038	-0.072	0.005	-0.006	0.016	0.005	0.012
6	6.349	0.976	1.19	-0.079	0.008	-0.040	-0.062	0.007	-0.007	0.021	0.005	0.007
6	6.306	0.978	1.15	-0.089	0.001	-0.032	-0.075	0.003	-0.006	0.019	-0.001	0.006
7	6.383	0.974	1.18	-0.082	0.003	-0.032	-0.077	0.005	-0.012	0.020	-0.004	0.007
7	6.395	0.974	1.17	-0.083	0.011	-0.042	-0.075	0.013	-0.004	0.024	0.007	0.009
8	6.351	0.977	1.16	-0.084	0.013	-0.038	-0.067	0.002	-0.009	0.021	0.004	0.008
8	6.339	0.976	1.17	-0.087	0.005	-0.032	-0.068	0.005	-0.007	0.018	0.001	0.008
9	6.321	0.979	1.13	-0.084	0.017	-0.035	-0.078	0.006	-0.008	0.022	0.002	0.010
9	6.313	0.977	1.18	-0.091	0.003	-0.048	-0.078	0.006	-0.008	0.018	0.002	0.007
10	6.319	0.977	1.18	-0.085	0.004	-0.027	-0.065	0.004	-0.008	0.020	-0.001	0.006
10	6.330	0.976	1.18	-0.085	0.008	-0.037	-0.068	0.007	-0.007	0.021	0.003	0.007
11	6.327	0.977	1.19	-0.084	-0.001	-0.021	-0.064	0.003	-0.009	0.017	-0.001	0.007
11	6.334	0.977	1.21	-0.076	0.007	-0.042	-0.062	0.009	-0.008	0.020	0.005	0.007
12	6.335	0.976	1.19	-0.074	0.013	-0.028	-0.069	-0.002	-0.006	0.021	0.004	0.007
12	6.321	0.977	1.15	-0.094	0.001	-0.049	-0.077	0.006	-0.009	0.024	0.004	0.008
14	6.308	0.978	1.18	-0.083	0.013	-0.026	-0.063	0.002	-0.003	0.019	0.005	0.009
16	6.304	0.978	1.17	-0.092	0.007	-0.045	-0.079	0.011	-0.006	0.014	0.005	0.007
16	6.302	0.979	1.19	-0.078	0.018	-0.035	-0.058	0.003	-0.010	0.018	0.002	0.008
17	6.304	0.979	1.16	-0.087	0.021	-0.027	-0.060	0.004	-0.006	0.020	0.004	0.008
17	6.321	0.977	1.17	-0.087	0.005	-0.042	-0.076	0.003	-0.008	0.019	0.001	0.008
14	6.288	0.979	1.18	-0.088	0.007	-0.043	-0.073	0.011	-0.009	0.018	0.003	0.007
15	6.301	0.979	1.19	-0.081	0.016	-0.035	-0.060	0.004	-0.006	0.019	0.004	0.008
15	6.284	0.980	1.15	-0.092	0.009	-0.057	-0.079	0.010	-0.011	0.016	0.005	0.008
18	6.316	0.977	1.17	-0.097	0.009	-0.044	-0.071	0.009	-0.003	0.015	0.000	0.009
18	6.319	0.978	1.15	-0.084	0.021	-0.040	-0.074	0.009	-0.011	0.014	0.006	0.010
19	6.283	0.979	1.16	-0.087	0.013	-0.039	-0.069	0.007	-0.014	0.021	0.009	0.008
19	6.256	0.980	1.16	-0.093	0.006	-0.043	-0.077	0.007	-0.005	0.016	-0.002	0.007
20	6.285	0.978	1.12	-0.091	0.023	-0.034	-0.076	0.006	-0.005	0.022	0.003	0.008
20	6.276	0.979	1.14	-0.103	0.009	-0.049	-0.072	0.007	-0.007	0.024	0.003	0.008
av	6.320	0.977	1.17	-0.087	0.009	-0.038	-0.071	0.006	-0.007	0.019	0.003	0.008
esd	0.028	0.001	0.02	0.006	0.006	0.007	0.006	0.003	0.002	0.003	0.003	0.001

^a Serial number of the reference compound in Table A.1.

with polarization functions.²⁹ The DFT calculations were based on Becke's three-parameter hybrid method³⁰ combined with the nonlocal correlation functional of Lee, Yang, and Parr³¹ (B3LYP keyword in Gaussian 98). Unlike in the previous study, we used experimental molecular geometries retrieved from the Cambridge structural database³² so that a sample representative for actually occurring solid-state conformations was obtained. Complex static valence-only structure factors in the range of $0 < \sin \theta/\lambda < 1.1 \text{ \AA}^{-1}$ were obtained by analytic Fourier transform of the molecular charge densities for reciprocal lattice points corresponding to a pseudocubic cell with 30 Å edges. These data were fitted in terms of pseudoatom parameters described above, using the XD program suite.³³ Both radial screening factors (κ , κ') were refined independently for each atom, with the exception of the chemically equivalent hydrogen atoms which shared the same κ and κ' parameters. The multipolar expansion was truncated at the hexadecapolar level ($l_{\max} = 4$) for the non-hydrogen atoms and at the quadrupolar level ($l_{\max} = 2$) for hydrogen atoms, for which only bond-directed functions of $l, m = 1, 0$ and $2, 0$ were refined. In order to reduce the number of least-squares variables, local-symmetry constraints were imposed for some atoms. A molecular electroneutrality constraint was applied in all refinements. The phase ambiguity problem was solved by keeping the phases fixed at theoretically calculated values.

Local Coordinate System Assignment, Program LSDB.

Averaging of the multipole populations of equivalent pseudoatoms becomes straightforward if the spherical harmonics

centered at these sites are expressed in a common local frame, as is the case in the XD package. A computer program (LSDB) has been developed to set up the local coordinate systems, equivalency, and symmetry constraints according to the scheme adopted in XD. It is based on the analysis of neighboring atoms of the first coordination sphere and determination of chemical equivalency of first neighbors using information from the second and sometimes the third coordination spheres of the "central" atom. A more detailed description of procedures implemented in LSDB is given in Appendix B. This program has been extensively tested on a large number of organic molecules and was found to work well for all the cases. It should be noted that procedures implemented in LSDB can also be applied to organometallic and inorganic systems without any significant modification.

Averaging of Pseudoatom Parameters. To illustrate the spread of density parameters extracted from simulated data of the model compounds, oxygen atoms of the carboxylate CO₂⁻ group and the oxygen atom O(=C) of the carboxyl COOH group are selected. As the entries in Tables 1 and 2 show, the multipole populations (arranged column wise) scatter in a narrow range. This is also indicated by their standard deviations σ calculated as the square root of the bias-corrected variances:³⁴

$$\sigma(\bar{x}) = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

TABLE 2: Pseudoatom Parameters for the Oxygen Atom O(=C) in the COOH Group in Compounds Used in Construction of the Databank

ID ^a	P_v	κ	κ'	P_{11+}	P_{11-}	P_{20}	P_{22+}	P_{31+}	P_{33+}	P_{40}	P_{42+}	P_{44+}
21	6.090	0.989	1.15	-0.098	-0.011	-0.075	-0.079	-0.004	0.016	-0.006	0.002	0.005
22	6.178	0.986	1.17	-0.098	-0.008	-0.070	-0.079	-0.002	0.017	-0.004	0.001	0.008
5	6.089	0.990	1.16	-0.096	-0.003	-0.079	-0.074	-0.005	0.013	-0.005	0.004	0.007
14	6.117	0.988	1.16	-0.094	-0.007	-0.067	-0.068	-0.002	0.014	-0.002	0.002	0.005
18	6.097	0.989	1.17	-0.087	0.000	-0.075	-0.073	-0.006	0.013	-0.007	0.005	0.007
16	6.107	0.989	1.13	-0.106	-0.006	-0.084	-0.072	-0.004	0.014	-0.004	0.003	0.006
19	6.149	0.986	1.16	-0.097	-0.006	-0.064	-0.073	-0.007	0.013	-0.003	0.001	0.005
av	6.118	0.988	1.16	-0.097	-0.006	-0.074	-0.074	-0.004	0.014	-0.005	0.003	0.006
esd	0.034	0.002	0.02	0.006	0.003	0.007	0.004	0.002	0.002	0.002	0.001	0.001

^a Serial number of the reference compound in Table A.1.**TABLE 3: Pseudoatom Parameters of the H(O) Atom in All Compounds Used in Construction of the Databank^a**

ID ^b	P_v	κ	κ'	P_{10}	P_{20}
21	0.816	1.140	1.32	0.171	0.078
22	0.821	1.164	1.43	0.133	0.060
5	0.792	1.168	1.57	0.104	0.044
14	0.858	1.138	1.46	0.127	0.061
18	0.799	1.165	1.50	0.113	0.046
16	0.829	1.155	1.46	0.124	0.058
19	0.790	1.210	1.58	0.111	0.047
23	0.830	1.183	1.48	0.126	0.056
23	0.894	1.132	1.38	0.142	0.087
3	0.877	1.168	1.44	0.135	0.065
4	0.902	1.158	1.41	0.148	0.075
5	0.877	1.161	1.37	0.156	0.078
6	0.850	1.171	1.38	0.152	0.075
7	0.877	1.168	1.47	0.131	0.062
8	0.913	1.150	1.40	0.142	0.067
9	0.902	1.157	1.35	0.158	0.086
10	0.863	1.149	1.49	0.130	0.069
11	0.898	1.158	1.41	0.142	0.077
12	0.884	1.163	1.35	0.157	0.083
13	0.949	1.135	1.32	0.178	0.097
14	0.860	1.169	1.38	0.153	0.068
15	0.830	1.191	1.48	0.129	0.058
16	0.850	1.186	1.48	0.129	0.056
17	0.910	1.135	1.38	0.148	0.083
18	0.880	1.162	1.36	0.156	0.074
19	0.890	1.173	1.46	0.128	0.056
20	0.930	1.147	1.36	0.163	0.082
21	0.827	1.199	1.57	0.108	0.046
av	0.864	1.163	1.43	0.139	0.068
esd	0.042	0.019	0.07	0.019	0.014

^a The first seven entries correspond to the carboxyl (COOH) group, whereas others correspond to the COH group in alcohols. ^b Serial number of the reference compound in Table A.1.

where x_i is the i th observation of the pseudoatom parameter x , \bar{x} is the mean value of parameter x , and N is the total number of observations. The overall precision obtained, especially for the κ' parameters ($\sigma(\kappa') = 0.02$), exceeds that reachable in experimental charge density studies. Similar results were obtained for all other atom types. The two oxygen sites present an example in which the nearest-neighbor scheme cannot be applied automatically. The valence charges ($P_v(\text{CO}_2^-) = 6.32(3)$, $P_v(\text{COOH}) = 6.12(3)$), the associated expansion-contraction coefficients ($\kappa(\text{CO}_2^-) = 0.977(1)$, $\kappa(\text{COOH}) = 0.988(2)$), and some of the multipole populations (for example, $P_{20}(\text{CO}_2^-) = -0.038(7)$, $P_{20}(\text{COOH}) = -0.074(7)$) are significantly different. The current version of LSDB can recognize "critical" sites that should be treated independently. Hydrogen atoms, on the other hand, are averaged

according to the first-neighbor approximation. For example, the carboxyl and hydroxyl hydrogen atoms are averaged, even though the two oxygen atoms they are attached to are treated as chemically different. The entries in Table 3 justify this approach. A simple statistical analysis suggests that only those populations that satisfy the $|P_{lm\pm}| > 0.002$ and $P_{lm\pm} > \sigma(P_{lm\pm})$ conditions should be included in the databank.

Atom Types in the Current Databank. The current version of the databank was originally intended for studies of the binding process of glycopeptide antibiotics (such as vancomycin) to peptide models of cell wall receptors.^{35,36} The types of atoms included in the current version are shown in Table 4. Although a much larger number of atom types are present in the model compounds (Table A.1), there are often not enough occurrences of a particular type to yield statistically meaningful average values for their parameters. However, because of automated procedures implemented for construction and maintenance of the databank, the extension of the library to include more atom types is straightforward.

Electroneutrality Scaling. To ensure that a density constructed from the library is correctly normalized, monopole populations of the constituent pseudoatoms have to be scaled. Program LSDB offers three scaling schemes. In the first method the correction is proportional to the magnitude of the valence population P_v :

$$P_{v,i}^{\text{scaled}} = \frac{\sum_i Z_i}{\sum_i P_{v,i}}$$

where Z_i is the number of valence electrons in the free atom i . The net charge

$$q_i = Z_i - P_{v,i}$$

can never change its sign and the largest correction is applied to atoms with the largest populations. In the second method it is assumed that all populations are determined at the same level of precision; therefore, the same changes are applied to all valence charges:

$$P_{v,i}^{\text{scaled}} = P_{v,i} + \left(\frac{\sum_i Z_i - \sum_i P_{v,i}}{N} \right)$$

where N is the number of atoms in the neutral molecule

TABLE 4: Atoms Currently Included in the Databank

Oxygen					
Nitrogen					
Carbon					
Hydrogen					
Chlorine					

(fragment). In this scaling the sign of a net atomic charge can change if the correction is large enough and has an opposite sign.

The third method, suggested by Faerman and Price³⁷ and adapted in this study, takes into account the precision of individual atomic charges:

$$P_{v,i}^{\text{scaled}} = P_{v,i} + \left(\frac{\sum_i Z_i - \sum_i P_{v,i}}{\sum_i \sigma(P_{v,i})} \right) \sigma(P_{v,i})$$

The largest correction is thus applied to the less precisely determined valence populations in the databank.

3. The Performance of the Pseudoatom Databank

In order to evaluate the performance of our databank, the constructed densities and related electronic properties are compared to those (a) calculated from molecular wave functions

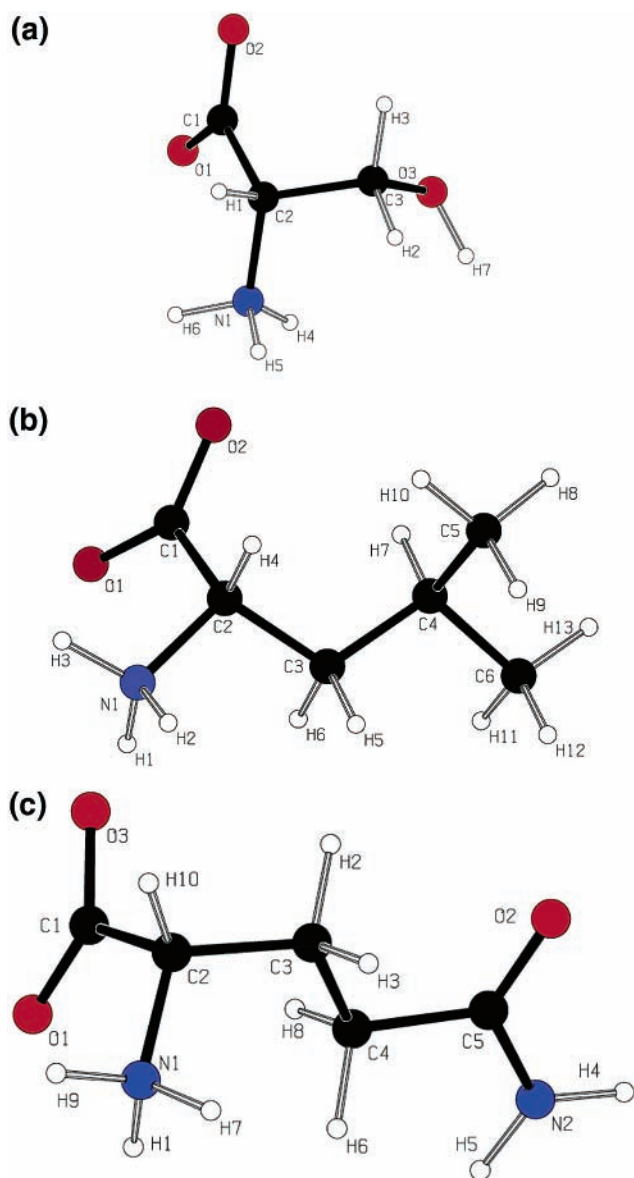
at G98/B3LYP/6-311++G(3df,3pd),³⁸⁻⁴¹ ADF/BLYP/TZP,⁴²⁻⁴⁴ and G98/B3LYP/6-31G** levels of theory (abbreviated as G98-1, ADF, and G98-2 respectively) and (b) based on pseudoatoms extracted from model structure factors of the G98/B3LYP/6-31G** calculation. Pseudoatoms from this last procedure are referred to below as the “model”. The comparison includes the analysis of the following:

- (1) Bonding features in the deformation density maps.
- (2) Electrostatic potential mapped on the 3D isosurfaces of the charge density.
- (3) Local and integrated topological properties of charge densities.
- (4) Intermolecular electrostatic interaction energies in dimers and crystals evaluated for model and pseudoatom parameters within the Buckingham-type approximation.^{45,46}

Because this analysis results in a large amount of numeric data, we present only the most representative results in the main body of the paper and include the rest as Supporting Information. More detailed information on the software and procedures used in this study is given in Appendix C.

TABLE 5: Geometrical Parameters of Dimers in Crystal Structures of SER, LEU, and GLN

dimer	D	H	A	symmetry code	D–H (Å)	H···A (Å)	D···A (Å)	D–H···A (deg)
L-Serine								
SER1	N(1)	H(5)	O(1)	$x, y, 1+z$	1.03	2.27	3.12	138
	N(1)	H(5)	O(2)		1.03	1.89	2.87	158
SER2	N(1)	H(4)	O(2)	$1/2+x, 1/2-y, -z$	1.03	1.87	2.89	167
SER3	N(1)	H(6)	O(1)	$1/2-x, -y, 1/2+z$	1.03	1.84	2.84	161
SER4	O(3)	H(7)	O(3)	$1/2-x, 1-y, 1/2+z$	0.97	2.03	2.92	152
	C(3)	H(2)	O(3)		1.09	2.56	3.21	117
L-Leucine								
LEU1	N(1)	H(1)	O(2)	$x, 1+y, z$	1.03	1.87	2.90	174
L-Glutamine								
GLN1	N(1)	H(1)	O(3)	$x, y, -1+z$	1.03	1.86	2.87	163
	C(4)	H(6)	O(3)		1.09	2.57	3.46	139
GLN2	N(2)	H(4)	O(2)	$1/2-x, -y, -1/2+z$	1.01	2.08	2.94	141
GLN3	N(2)	H(5)	O(1)	$1-x, -1/2+y, 1/2-z$	1.01	1.92	2.91	167
GLN4	N(1)	H(7)	O(2)	$1/2-x, 1-y, -1/2+z$	1.03	1.93	2.95	167
GLN5	N(1)	H(9)	O(3)	$1-x, 1/2+y, 3/2-z$	1.03	1.76	2.77	164

**Figure 1.** Molecular structures of SER (a), LEU (b), and GLN (c).

Test Systems. Three amino acids, L-serine (SER), L-leucine (LEU), and L-glutamine (GLN), not included in the construction of the databank, were selected as test compounds. All calculations were done using the experimental geometries (Figure 1).

TABLE 6: Sum of Valence Populations and Their Relative Errors for Test Molecules as Determined from the Pseudoatom Databank

	SER	LEU	GLN
$\sum_i P_{v,i}$ (electrons)	41.91	53.45	57.74
$\sum_i Z_i$ (electrons)	42.00	54.00	58.00
Δ (electrons) ^a	0.09	0.55	0.25
$\Delta/\sum_i Z_i$ (%)	0.2	1.0	0.4

$$^a \Delta = \sum_i Z_i - \sum_i P_{v,i}.$$

The crystal structure of L-serine ($C_3H_7N_1O_3$, space group $P2_12_12_1$, CSD code: LSERINE01) is based on an accurate room-temperature X-ray study.⁴⁷ In three out of the four hydrogen-bonded serine dimers in the crystal, the leading intermolecular interactions are strong N–H···O hydrogen bonds between the NH_3^+ and COO^- groups. The remaining dimer is formed via an O–H···O interaction between the hydroxyl groups. The structural parameters for L-leucine (Figure 1b) ($C_6H_{13}N_1O_2$, space group $P2_1$, CSD code: LEUCIN02) were taken from a recent accurate 120 K X-ray study by Gorbitz et al.⁴⁸ The asymmetric unit is composed of two independent molecules L-leucine(A) and L-leucine(B). The dimer of our choice is composed of two translation-related L-leucine(A) molecules held together via N–H···O interaction of their NH_3^+ and COO^- groups. The L-glutamine structural data ($C_5H_{10}N_2O_3$, space group $P2_12_12_1$, CSD code: GLUTAM01) were taken from a neutron diffraction study.⁴⁹ This crystal structure is also stabilized via a three-dimensional network of N–H···O hydrogen bonds. The X–H bonds in all test compounds were set at the standard neutron diffraction distances: $C_{methyl-H} = 1.059$ Å, $C_{primary-H} = 1.092$ Å, $C_{secondary-H} = 1.099$ Å, $N_{NH_3-H} = 1.035$ Å, $N_{NH_2-H} = 1.010$ Å, $O_{alcohol-H} = 0.967$ Å.⁵⁰ Further details on the dimer structures are given in Table 5.

Molecular Electroneutrality. Table 6 lists the total valence charges obtained for the test molecules composed from the library pseudoatoms. The databank error in predicting the molecular electroneutrality is very small: 1% for LEU and under 0.5% for SER and GLN. Similar accuracy has been obtained for other compounds tested (for example, 0.06% for L-dopa⁵¹ and 0.3% for Leu-Enkephalin⁵²). This result, taken as a validation test, suggests excellent internal consistency of the databank parameters. A posteriori scaling of valence population parameters to satisfy the total electroneutrality for all three molecules was done using the third scaling method due to Faerman and Price,³⁷ implemented in LSDb.

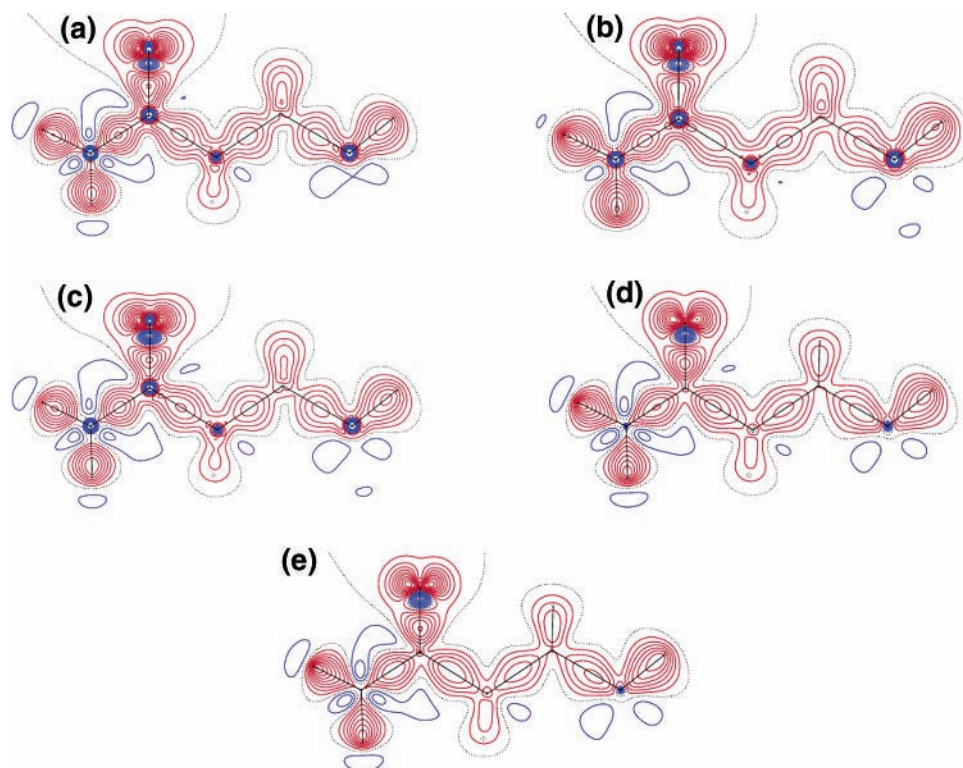


Figure 2. Deformation density in the GLN molecule from (a) G98/B3LYP/6-311++G(3df,3pd), (b) ADF/BLYP/TZP and (c) G98/B3LYP/6-31G** calculations, (d) from pseudoatom parameters after refinement of the G98/B3LYP/6-31G** structure factors and (e) from the databank. Contour levels at $0.1 \text{ e}/\text{\AA}^3$; positive contours, red, negative contours, blue, zero contour, black.

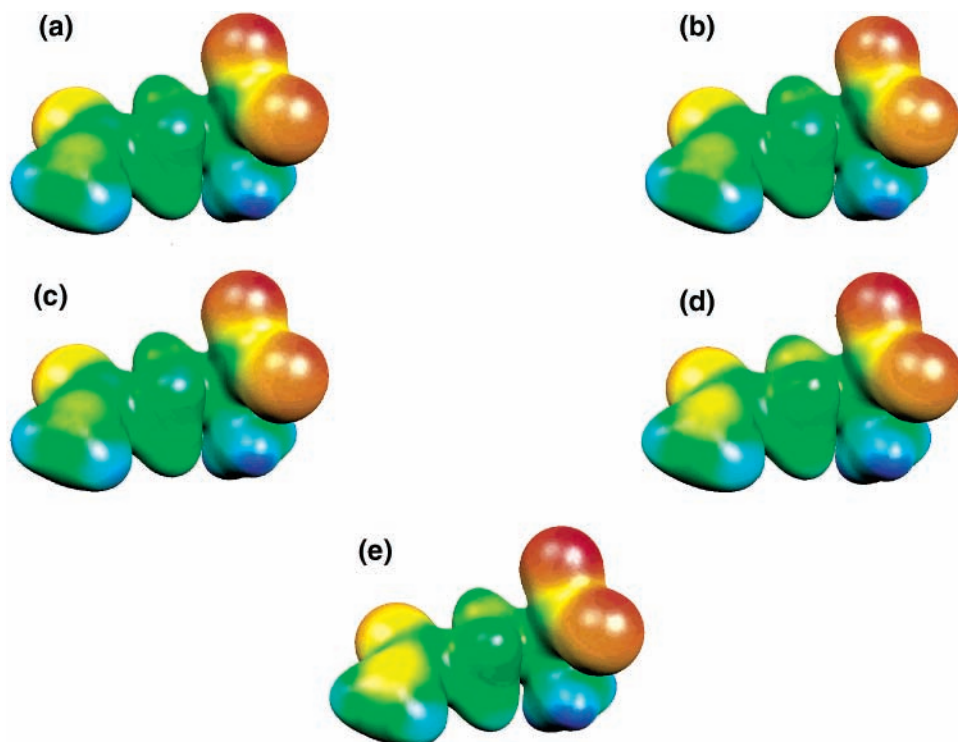


Figure 3. Electrostatic potential mapped on the 0.02 au density isosurface in the GLN molecule from (a) G98/B3LYP/6-311++G(3df,3pd), (b) ADF/BLYP/TZP and (c) G98/B3LYP/6-31G** calculations, (d) from pseudoatom parameters after refinement of the G98/B3LYP/6-31G** structure factors and (e) from the databank. Electrostatic potential is color coded as follows: the deep red color corresponds to the value of -0.15 au and deep blue color corresponds to $+0.35 \text{ au}$, whereas orange, yellow, green and cyan colors represent intermediate values.

Deformation Densities. Figure 2 shows the deformation densities calculated by different methods for the GLN molecule in the plane of the amide group. The corresponding maps for SER and LEU are given in the Supporting Information (Figures

S1 and S2). Inspection of the contour levels reveals that all bonding features are correctly reproduced in the databank density. The location and the magnitude of bond densities compare quite well with those obtained by other methods: the

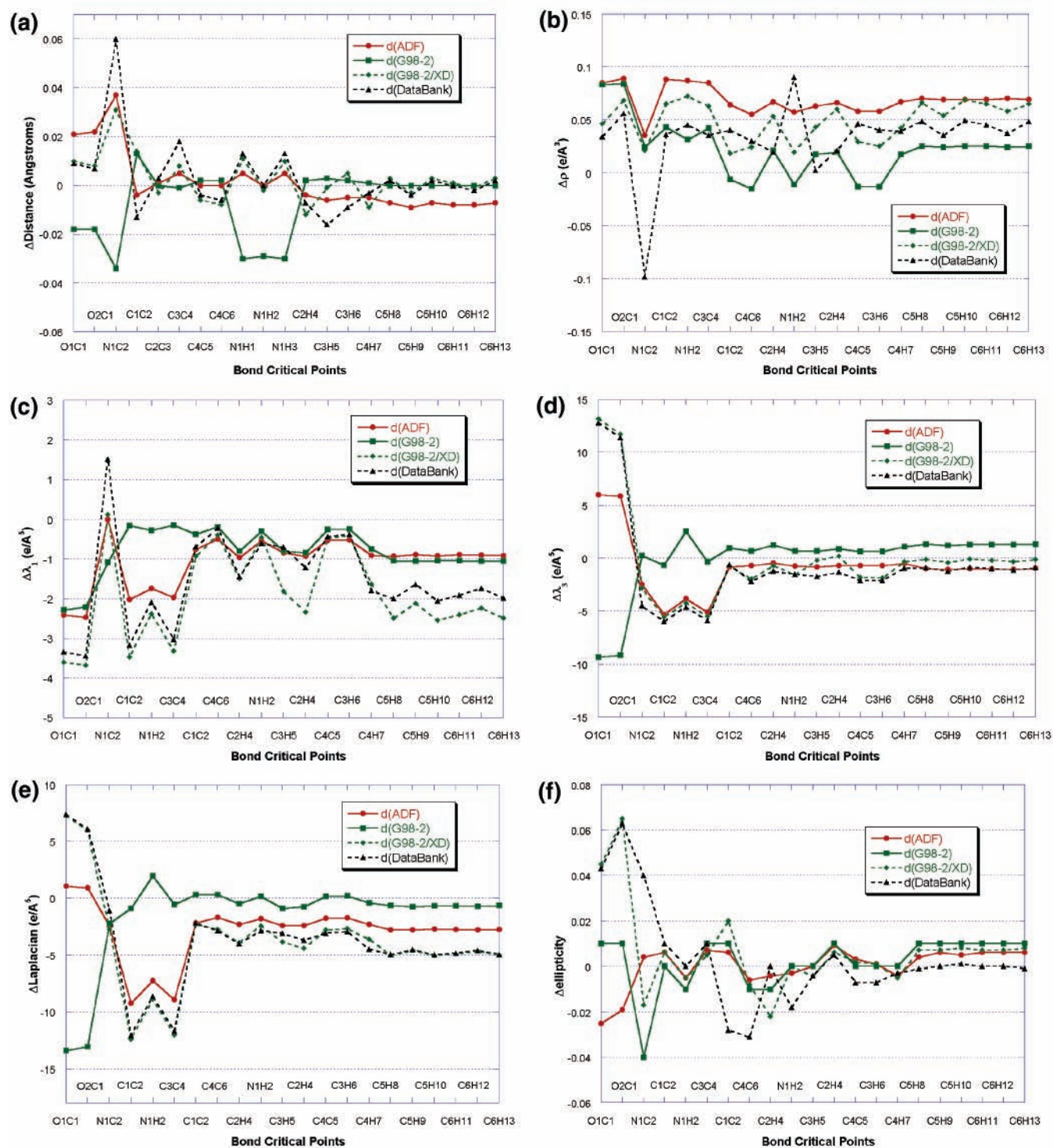


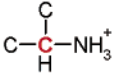
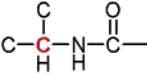
Figure 4. Differences for selected properties of ρ at intramolecular bond critical points in LEU relative to the reference G98/B3LYP/6-311++G-(3df,3pd) calculation: (a) displacement of the critical point from the bond midpoint (positive if displaced toward the second atom); (b) value of the charge density; (c) λ_1 curvature; (d) λ_3 curvature; (e) Laplacian; (f) ellipticity.

bond peaks at the C=O, C–N, C–C, and C–H bonds agree within $0.1\text{--}0.2\text{ e}/\text{\AA}^3$ with those of both G98 (G98-1 and G98-2) and ADF calculations. Even the lone pair peaks of the carbonyl oxygen atoms agree within $0.2\text{--}0.3\text{ e}/\text{\AA}^3$. The comparison of the deformation density obtained directly from the B3LYP/6-31G** wave function (Figure 2c) with its pseudoatom representation (Figure 2d) illustrates the bias introduced by the least-squares projection which leads to slightly different peak shapes in the maps. Features in the vicinity of the nuclei are markedly different because of

the frozen core approximation used in the model and databank pseudoatoms but not relevant to the intended application of the databank. Overall, the performance of the databank in representing the deformation densities is quite satisfactory.

Electrostatic Potential (EP). For the EP we used a color-coded representation projected onto isodensity surfaces (Figure 3 for GLN; Figures S3 and S4 in the Supporting Information for SER and LEU). In all plots the isodensity surface is drawn at 0.02 au and the potential is color coded as follows: the deep

TABLE 7: Pseudoatom Parameters for Two Types of Secondary Carbon Atoms (Marked in Red) as a Function of the Type of the Neighboring Nitrogen Atom^a

			Averaged values in the databank
P_V	3.79(3)	3.85(6)	3.82(5)
κ	1.009(2)	1.008(3)	1.008(3)
κ'	0.94(1)	0.93(1)	0.93(2)
P_{11+}	-0.10(2)	-0.06(1)	-0.08(3)
P_{20}	0.08(2)	0.05(2)	0.07(3)
P_{22+}	-0.15(2)	-0.08(3)	-0.12(4)
P_{31+}	-0.12(2)	-0.15(1)	-0.13(2)
P_{31-}	-0.20(1)	-0.20(2)	-0.20(1)
P_{33+}	0.15(1)	0.19(2)	0.17(3)
P_{33-}	-0.05(1)	-0.05(1)	-0.05(1)
P_{40}	0.02(1)	0.03(1)	0.03(1)
P_{42+}	-0.04(1)	-0.05(1)	-0.04(1)
P_{42-}	0.07(1)	0.09(1)	0.08(1)
P_{44+}	0.04(1)	0.05(1)	0.05(1)
P_{44-}	0.02(1)	0.02(1)	0.02(1)

^a The only significant discrepancy is for quadrupole P_{22+} .

red color corresponds to the value of -0.15 au, the deep blue color corresponds to $+0.35$ au, whereas orange, yellow, green, and cyan colors represent intermediate values. All methods correctly represent the sign of the EP in all regions of the molecule. The agreement between the results of all methods is quite satisfactory. Although the databank results resemble very closely those of the G98-2 and the model, all three show a more pronounced EP on the oxygen atoms of the carboxylate and hydroxyl groups compared to the more extensive G98-1 and ADF calculations.

It should be mentioned that the shapes of the isodensity surfaces calculated from the databank are very close to those from *ab initio* calculations, as may be expected given the agreement among the deformation density maps

Bond Critical Point (BCP) Indices. Figure 4 shows the ρ_{BCP} properties according to the ADF, G98-2, model, and databank methods relative to the G98-1 results for all bonds in LEU (Figure S5 in the Supporting Information for SER and GLN). It is noteworthy that the pattern found for the databank/model indices closely follows that obtained for ADF. The discrepancies usually have the same sign but are somewhat smaller for the last than the first two methods. For the polar C–O, N–H, and C–N bonds the ADF/model/databank predict the BCPs to be located closer to the more electronegative atom relative to the G98-1 calculation, whereas the G98-2 results show the opposite trend.

The ρ_{BCP} magnitudes given by the ADF/model/databank are usually smaller by about 0.02 – 0.1 $e/\text{\AA}^3$ than the reference G98-1 values. Exceptions are (1) the O3–H7 bond in SER for which the ADF ρ_{BCP} is smaller by 0.14 $e/\text{\AA}^3$ and (2) the N–C bonds in all three test systems for which the databank ρ_{BCP} is larger by ~ 0.1 $e/\text{\AA}^3$ (whereas the discrepancies for ADF are about 0.03 $e/\text{\AA}^3$). The latter exception is traced to a small dependence of the population of the quadrupolar function P_{22+} of the secondary carbon atom on the type of the bonded nitrogen atom: $-0.15(2)$ and $-0.08(3)$ when bonded to the ammonium group and nitrogen atom in the peptide bond (Table 7), respectively. Because all other parameters are essentially independent of the type of bonded nitrogen atom, we have averaged them and created a single entry in the

databank with $P_{22+} = -0.12(4)$. In any case, the overall agreement in ρ_{BCP} for all methods is fairly good and is well within 0.1 $e/\text{\AA}^3$.

As anticipated, the properties related to the second derivatives of ρ at the BCPs, (the three principal curvatures; λ_1 , λ_2 , and λ_3 and their algebraic sum, the Laplacian) spread in a wider range. The bias introduced by the pseudoatom projection is clearly revealed by the model versus G98-2 data. On the other hand, the agreement between the ADF and model and databank values, especially for the Laplacian and the curvature along the bond (λ_3), is outstanding. This correlation is especially pronounced in the highly polar C–O bonds for which the G98-2 density overestimates the curvatures.

This may come as a surprise unless we recall that both the pseudoatom model and ADF use the same type of Slater radial functions in their basis set expansions. In the previous studies the differences in properties of ρ at BCPs between the pseudoatom model and densities calculated from the Gaussian wave functions have been attributed to the limited flexibility of the single-exponential Slater-type radial functions used in the pseudoatom model.^{53–57} However, the current study clearly indicates this conclusion to be only partially correct. The real cause of such discrepancies can be the different behavior of Gaussian and Slater radial functions in the vicinity of the bond critical points. In any case, the observed differences for the databank are very much comparable to those found between different *ab initio* calculations, for example ADF/BLYP/TZP and G98/B3LYP/6-311++G(3df,3pd). Thus, for the properties of ρ at bond critical points the bias introduced by the pseudoatom model is not significant since the main differences between the databank and Gaussian 98 calculations originate from the different types of the radial functions used in the density representation.

In general, both the model and the databank slightly underestimate the ellipticity of the density at the BCP for strong polar C–O bonds (as does the G98-2 calculation) compared to the reference G98-1 calculation which in this case is somewhat different from the trend observed for ADF. However, the differences in bond ellipticities for the databank are comparable in magnitude with differences observed for both ADF and G98-2 calculations relative to the reference G98-1 calculation.

Topological Electrostatic Moments. Figure 5a shows the AIM net atomic charges in GLN based on different densities and relative to those of G98-1. ADF charges show the largest scatter, especially for the non-hydrogen atoms. Their correlation with databank and model values is evident. The largest differences between databank and G98-1 charges are observed for the carbon atoms of the carboxylate groups (~ 0.13 electrons) and the carbon atoms bonded to the hydroxyl group (~ 0.17 electrons), whereas for ADF for the same atoms the differences are as large as 0.22 electrons. The overall pattern obtained for the G98-2 density is considerably different from those derived by other methods.

The performance of the databank in reproducing the first- and higher-order atomic moments is demonstrated in parts b–e of Figure 5 for GLN and Figure S6 in the Supporting Information for SER and LEU. For each element of the atomic moments the pattern is somewhat different, and it is not as easy to interpret as that of the charge. For the dipoles the trends revealed by databank and model densities closely

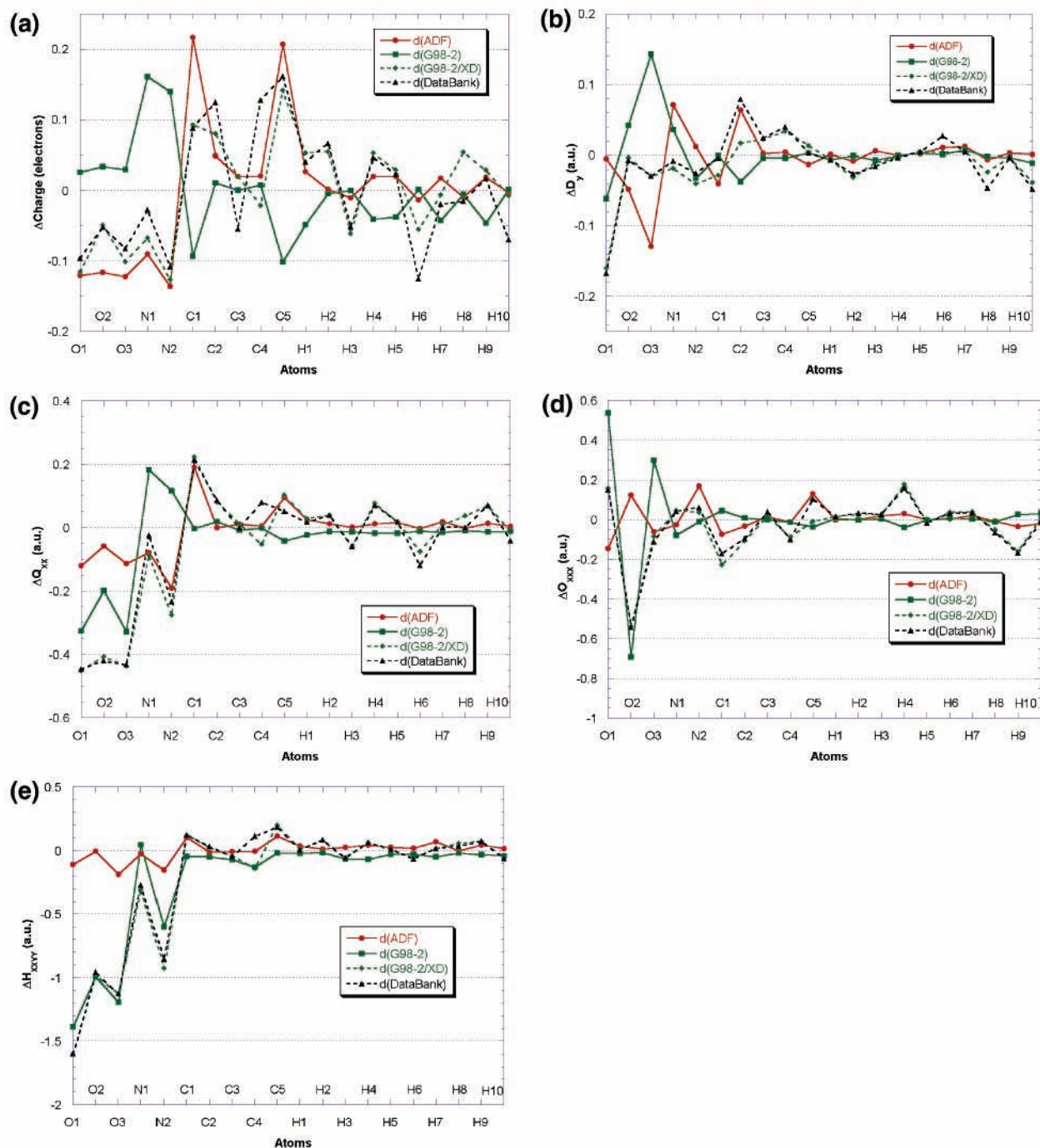


Figure 5. Differences between topological (a) net atomic charges and selected components of higher atomic moments, (b) D_Y , (c) Q_{XX} , (d) O_{XXX} , and (e) H_{XYY} , in GLN and values from the reference G98/B3LYP/6-311++G(3df,3pd) calculation.

resemble each other, except for a few cases in which the differences for the databank are somewhat larger, but they do not follow the pattern found for either the ADF or G98-2 densities. However, the higher-order moments of the pseudo-atom densities appear to reproduce those obtained at the G98-2 level. The largest discrepancies are always observed for the oxygen atoms and for some atomic moments of the nitrogen atoms of the NH_3^+ and NH_2 groups. However, with only very few exceptions, the observed differences for the databank are very close to those observed for the model. In general, the accuracy of atomic moments predicted by the databank is comparable with that obtained by the model density and only slightly worse than that derived from the G98-2 calculation.

Molecular Moments. The total molecular moments were calculated from individual atomic moments rather than by integration of the total molecular densities (Figure 6 for LEU and Figure S7 in the Supporting Information for SER and GLN). Given a good overall performance of the databank in predicting individual atomic moments, one could expect a comparable accuracy for the total molecular moments. Indeed, the accuracy achieved is of the same order as the differences between various *ab initio* methods (ADF, G98-1, and G98-2). Although the difference of 53 D-Å³ between the databank and the G98-1 hexadecapolar molecular moment (H_{xxx}) in LEU is considerable, it amounts only to ~4% because of the large absolute value of this moment (~1200 D-Å³).

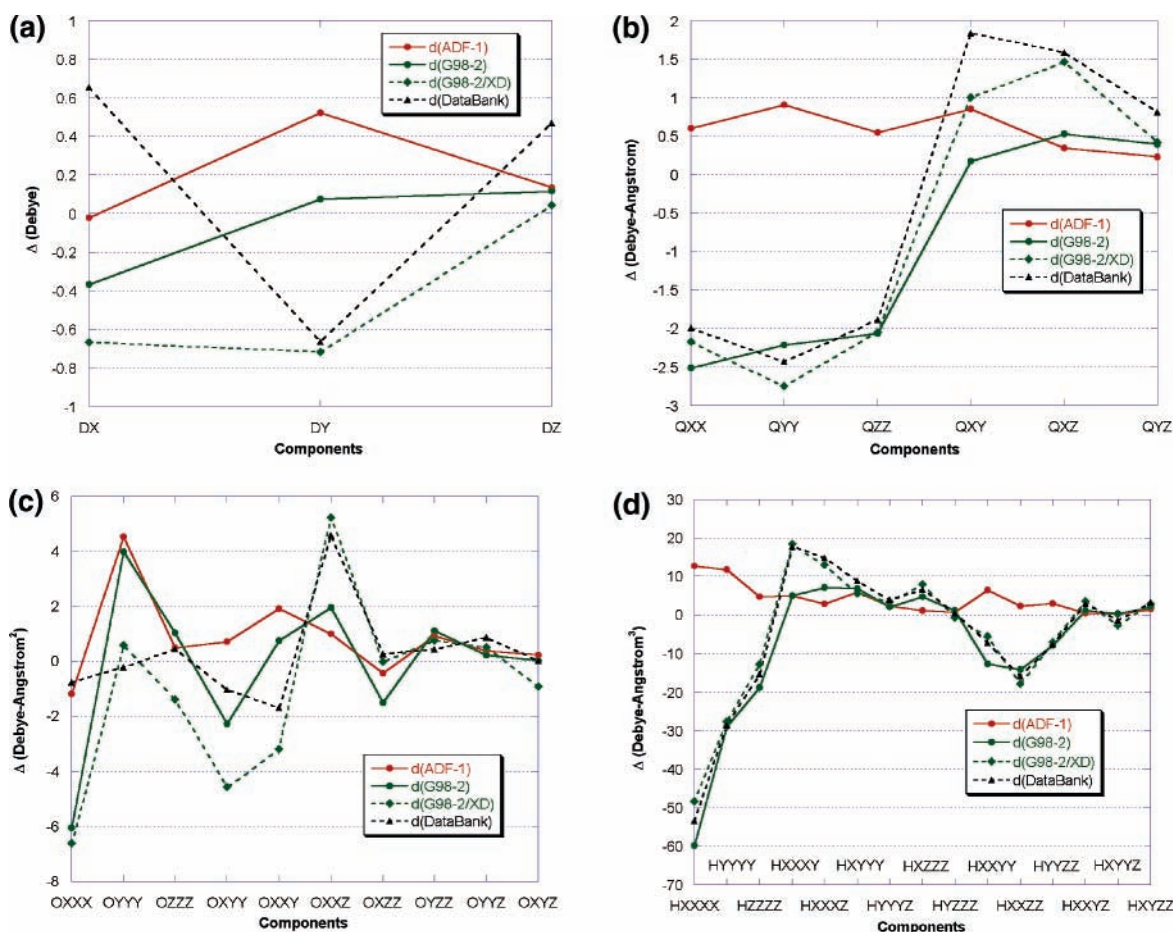


Figure 6. Differences between components of the total molecular moments ((a) first, (b) second, (c) third, and (d) fourth) in LEU and values from the reference G98/B3LYP/6-311++G(3df,3pd) calculation.

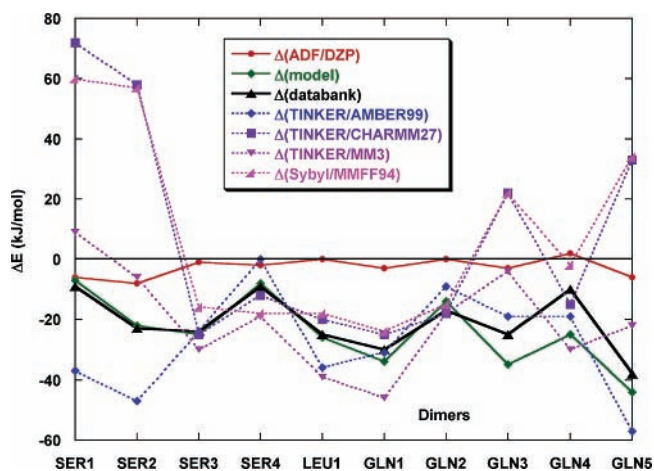


Figure 7. Differences between electrostatic interaction energies in test molecular dimers and values from the Morokuma–Ziegler energy decomposition of the ADF/TZP results.

Intermolecular Electrostatic Interaction Energies of Molecular Dimers. Because Gaussian 98 does not allow for the direct evaluation of the electrostatic interaction energy (E_{es}) in molecular dimers, our analysis relies on comparison with ADF calculations performed at the BLYP level of theory with DZP and TZP basis sets. The E_{es} of the pseudoatom models was calculated within the Buckingham-type approximation,⁴⁵ whereas that based on the ADF wave functions was derived using the Morokuma–Ziegler energy-partitioning

TABLE 8: Root-Mean-Square (rms) Deviation of ΔE_{es} (kJ/mol) from ADF/TZP Reference Values for Ten Amino Acid Dimers for Each of the Methods Examined

	rms
ADF/DZP	4
model	27
databank	23
TINKER/AMBER99	32
TINKER/CHARMM27	35
TINKER/MM3	26
Sybyl/MMFF94	32

scheme, which partitions the energy into electrostatic, orbital interaction, and exchange–repulsion components.^{58–60} As seen in Figure 7, the model and databank predicted values are practically equivalent except for the GLN4 dimer, for which a difference of 15 kJ/mol was found. However, both methods underestimate the ADF electrostatic interaction energy by about 25 kJ/mol on average. These results are in line with our previous findings for E_{es} in molecular dimers of α -glycine, *N*-acetylglycine, and *L*-lactic acid,⁶¹ yet the agreement with the reference ADF energies is much better in the current study, likely because of improvements implemented in the new version of the databank. Much of the discrepancy is due to the use of the Buckingham-type approximation which assumes nonoverlapping densities and underestimates the interaction for short-range contacts. We are developing a new method which eliminates this drawback.⁶² Preliminary results show much reduced discrepancies between the databank and primary values.

TABLE 9: Binding Energies of Molecules in Crystals: Total from CRYSTAL03 Calculations (Corrected for BSSE) and the Electrostatic Component of the Binding Energy Calculated Based on the Pseudoatom Expansion (all in kJ/mol)

	$E_{\text{total}}^{\text{CRYSTAL03}}$	$E_{\text{electrostatic}}^{\text{pseudoatom}}$ only	
		model	databank
SER	-251	-221	-220
LEU	-402	-350	-342
GLN	-235	-176	-210

The comparison of the databank energies with those calculated with force-field methods (Figure 7 and Table 8) shows that the improvement over point-charge only AMBER99, CHARMM27, and MMFF94 calculations is quite pronounced (especially for the first two dimers of serine) and is slightly better than the MM3 model which includes both partial charges and dipoles. The behavior of the databank is also more consistent than that exhibited by the force fields. The databank always gives a lower electrostatic interaction energy compared to the ADF calculation, whereas CHARMM27, MM3, and MMFF94 overestimate the energy for some of the dimers (sometimes by large amounts) and underestimate it for others. It is also remarkable that the performance of the databank is better than that of the AMBER99 force field, the parameters of which were specifically adjusted for each of the amino acids examined, that is, SER, LEU, and GLN. In contrast, in the databank the pseudoatom parameters are averaged over all species examined, which did not include any of the test amino acids. Yet, the agreement in E_{es} for the databank (rms = 23 kJ/mol) is much better than for AMBER99 (rms = 32 kJ/mol). In general, the performance of the databank in the calculation of the electrostatic interaction energy in molecular dimers can be considered satisfactory and more consistent than that calculated with the current force-field methods.

Binding Energies of Molecules in Crystals. Table 9 contains the binding energy of molecules in crystals calculated *ab initio* at the density functional level of theory (fully periodic and single-molecule calculations) and its electrostatic component from the pseudoatom approach using both model and databank parameters. For the binding energy obtained from the pseudoatom approach to be comparable with *ab initio* results, the electrostatic component should be supplemented with atom–atom potential terms which describe repulsion and dispersion forces.^{63–68} Unfortunately, these potentials cannot be defined unambiguously and application of different types of atom–atom potentials can result in significantly different energies. Thus, in this study it was decided to omit dispersion and repulsion terms in the pseudoatom approach and compare only the electrostatic component with the *total* binding energies calculated *ab initio*. Given the strong polarity of molecules in the test crystals, the electrostatic component should be the dominant contribution to the total energy, especially as dispersion and repulsion forces have opposite signs and often approximately cancel each other in crystals of hydrogen-bonded polar molecules. Indeed, the electrostatic binding energies calculated with pseudoatoms predict the correct *relative* order of the total binding energies, but the absolute values can be off by as large a value as 60 kJ/mol. The databank performance is very close to that of the model except for GLN, for which the data-

bank energy is closer to the *ab initio* value by 34 kJ/mol. In any case, incorporation of proper atom–atom potentials should bring the molecular binding energies calculated with the pseudoatom approach much closer to the *ab initio* values.

It should also be noted that the calculation of binding energies based on the pseudoatom approach has a great speed advantage over *ab initio* methods. Even when all the interactions up to and including the hexadecapole–hexadecapole terms were taken into account, it took only several minutes on a 1.8 GHz AthlonXP CPU to complete the calculation for all the test structures.

4. Conclusions

The theoretical pseudoatom databank allows rapid evaluation of the electrostatic interaction energies of complex molecules. It can be used to reproduce both the interaction energies and the molecular electron density distribution and its properties at a fraction of the time required to perform *ab initio* calculations. Although in the current study the interaction energies are evaluated with the Buckingham nonoverlapping density approximation, the databank readily lends itself for use in more sophisticated approaches, such as the pixel-by-pixel method recently developed by Gavezzotti,^{69,70} which also allows evaluation of other contributions to the interaction energy from the electron density.^{71,72} The electrostatic energies calculated with the databank in the Buckingham approximation are always lower (by ~ 25 kJ/mol) than those from the Morokuma–Ziegler energy decomposition of theoretical DFT values, unlike the empirical force fields tested, which sometimes overestimate, other times underestimate, the interaction energy. The electrostatic binding energies of the amino acid molecules in the crystals calculated with the databank within the nonoverlapping densities approximation show the correct trend, though they are somewhat smaller than the *ab initio* values which also include repulsion and dispersion terms. Inclusion of such terms is a logical extension of the databank approach.

The properties of the density are very well reproduced. Deviations from the more advanced theoretical values are equal or smaller than differences between various methods of computation.

Finally, we note that the pseudoatoms do not incorporate the effect of polarization of the density due to intermolecular interactions, which depends on the packing of the molecules in the crystal and is thus variable among solids. However, X-ray scattering factors based on the databank pseudoatoms can be used as a starting point for aspherical-atom refinement of accurate diffraction data on peptides and macromolecules, to give more accurate phases of the experimental structure amplitudes and more accurate structural parameters.

Extension of the databank to include additional pseudoatoms for a comprehensive treatment of both organic and bioinorganic systems is under consideration.

Acknowledgment. We thank Dr. Irina Novozhilova for her help with the ADF calculations and plotting the electrostatic potential. Support from the National Institutes of Health (GM56829) and the National Science Foundation (CHE0236317) is gratefully acknowledged.

Appendix A.

TABLE A.1: Compounds Used for Construction of the Databank

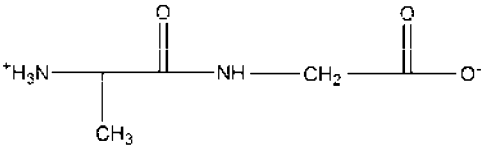
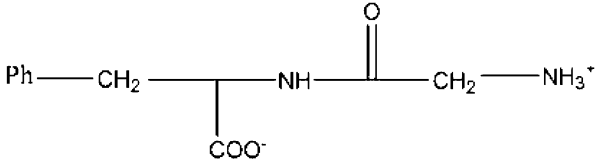
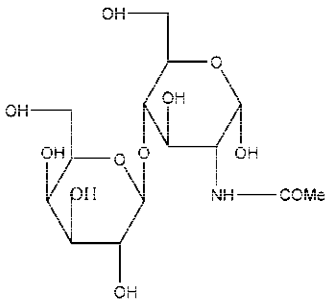
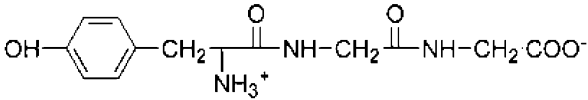
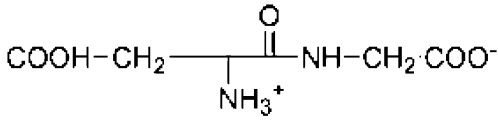
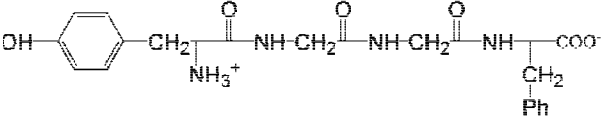
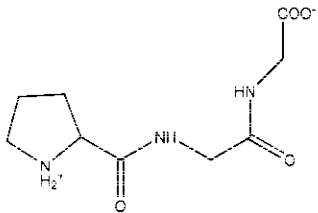
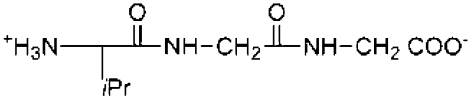
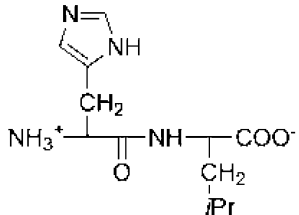
Chemical Diagram	Name; Formula	CSD Code
	L-Alanylglycine C ₅ H ₁₀ N ₂ O ₃	ALAGLY ⁷³
	Glycyl-DL-phenylalanine C ₁₁ H ₁₄ N ₂ O ₃	GLDLPA ⁷⁴
	N-Acetyl-lactosamine monohydrate C ₁₄ H ₂₅ NO ₁₁ .H ₂ O	ACLACT ⁷⁵
	Tyrosyl-glycyl-glycine monohydrate C ₁₃ H ₁₇ N ₃ O ₅ .H ₂ O	LTYRGG01 ⁷⁶
	α-L-Aspartyl-glycine monohydrate C ₆ H ₁₀ N ₂ O ₅ .H ₂ O	ASPGLY ⁷⁷
	Tyrosyl-glycyl-glycyl- phenylalanine trihydrate dimethylsulfoxide solvate C ₂₂ H ₂₆ N ₄ O ₆ .C ₂ H ₆ OS.3(H ₂ O)	TGGPDH10 ⁷⁸
	L-Prolyl-glycyl-glycine C ₉ H ₁₅ N ₃ O ₄	FABXUB10 ⁷⁹
	L-Valyl-glycyl-glycine C ₉ H ₁₇ N ₃ O ₄	COPBIS10 ⁸⁰
	L-Histidyl-leucine C ₁₂ H ₂₀ N ₄ O ₃	JUKMOR ⁸¹

TABLE A.1 (Continued)

Chemical Diagram	Name; Formula	CSD Code
	Sarcosyl-glycylglycine C ₇ H ₁₃ N ₃ O ₄	SARGLY10 ⁸²
	Glycyl-L-alanyl-L-isoleucine hemihydrate C ₁₁ H ₂₁ N ₃ O ₄ ·0.5H ₂ O	JORMIM ⁸³
	L-Tryptophyl-glycyl-glycine dihydrate C ₁₅ H ₁₈ N ₄ O ₄ ·2(H ₂ O)	FIZWOA01 ⁸⁴
	t-Butoxycarbonyl-glycyl- L-valyl-L-tryptophan methyl ester C ₂₄ H ₃₄ N ₄ O ₆	PUVSAA ⁸⁵
	N-Methyl-D-aspartic acid monohydrate C ₅ H ₉ NO ₄ ·H ₂ O	KEWGUO ⁸⁶
	N-Methyl-L-tryptophan C ₁₂ H ₁₄ N ₂ O ₂	WAJBIS ⁸⁷
	(2S,3S,4S)-3-Carboxymethyl- 4-(2-methoxyphenyl) pyrrolidine-2-carboxylic acid C ₁₄ H ₁₇ NO ₅	NEGBUW ⁸⁸
	allo-4-Hydroxy-L-proline dihydrate C ₅ H ₉ NO ₃ ·2(H ₂ O)	AHLPRO ⁸⁹
	DL- α -Methylglutamic acid hemihydrate C ₆ H ₁₁ NO ₄ ·0.5(H ₂ O)	MGLUAH ⁹⁰
	(-)-4-Amino-2- oxabicyclo(3.1.0)hexane- 4,6-dicarboxylic acid C ₇ H ₉ NO ₅	COKJER ⁹¹

TABLE A.1 (Continued)

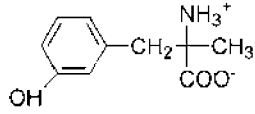
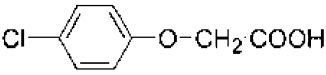
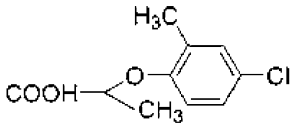
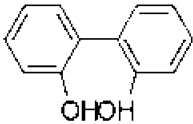
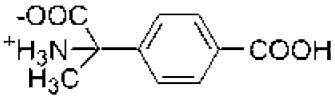

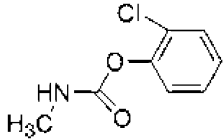
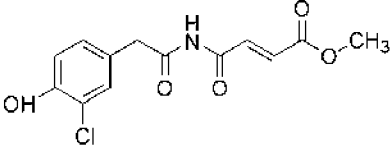
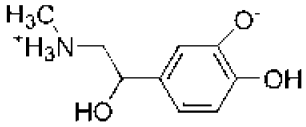
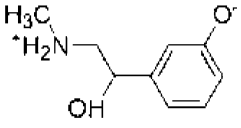
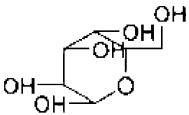
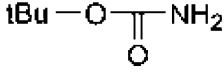
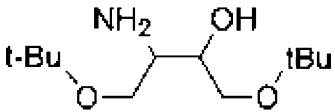
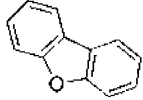
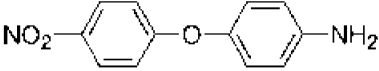
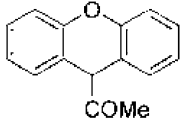
Chemical Diagram	Name; Formula	CSD Code
	DL-α-Methyl-m-tyrosine C ₁₀ H ₁₃ NO ₃	MEMTYR ¹⁰²
	(4-Chlorophenoxy)- acetic acid C ₈ H ₇ ClO ₃	CLPOAC ⁹³
	(+)-2-(4-Chloro- 2-methylphenoxy)- propionic acid C ₁₀ H ₁₁ ClO ₃	CMPXPA ⁹⁴
	2,2'-Biphenol C ₁₂ H ₁₀ O ₂	NUTSUQ ⁹⁵
	(+)-α-Methyl-4- carboxyphenylglycine C ₁₀ H ₁₁ NO ₄	REMVEK ⁹⁶
	Monochlorobenzene C ₆ H ₅ Cl	MCBENZ ⁹⁷
	2-Chlorophenyl-N-methyl- carbamate C ₈ H ₈ ClNO ₂	CLPMCB ⁹⁸
	N-(3-Chloro- 4-hydroxyphenylacetyl) fumarate C ₁₃ H ₁₂ ClNO ₅	JUHHAV ⁹⁹
	(-)-Adrenaline C ₉ H ₁₃ NO ₃	ADRENL ¹⁰⁰
	(-)-Phenylephrine C ₉ H ₁₃ NO ₂	PHEPHR ¹⁰¹
	α-D-Galactose C ₆ H ₁₂ O ₆	ADGALA10 ¹⁰²
	t-Butyl carbamate C ₅ H ₁₁ NO ₂	BIFPEL ¹⁰³
	1,4-bis(t-Butoxy)- 3(R)-amino-2(S)-butanol C ₁₂ H ₂₇ NO ₃	BIDBAR ¹⁰⁴

TABLE A.1 (Continued)

Chemical Diagram	Name; Formula	CSD Code
	Dibenzofuran C ₁₂ H ₈ O	DBZFUR02 ¹⁰⁵
	4-Aminophenyl 4'-nitrophenyl ether C ₁₂ H ₁₀ N ₂ O ₃	FUQBOI ¹⁰⁶
	Methyl 9-xanthenyl ketone C ₁₅ H ₁₂ O ₂	SOYTEF ¹⁰⁷
Et — O — Et	Diethylether hydrogen fluoride C ₄ H ₁₀ O. HF	GOTTEO ¹⁰⁸

Appendix B. LSDB, Program for Automatic Determination of the Local Coordinate System and Interface to Pseudoatom Databank

The procedure for determination of the atomic local coordinate system implemented in LSDB starts by constructing the first coordination sphere of a selected atom using the covalent or user-supplied radii. The neighbors of the first coordination sphere are then tested for chemical equivalency. For non-hydrogen atoms two atoms are defined as equivalent if they pass the following sequence of tests:

- (i) Same atomic number.
- (ii) Same number of atoms among first neighbors.
- (iii) Same number of atom types among first neighbors.
- (iv) Same number of atoms for each atom type among first neighbors.

Once all the chemically equivalent atoms (if any) among the first neighbors are determined, the program chooses the local symmetry elements that relate the chemically equivalent atoms taking into account the total number of first neighbors and following the rules of Kara and Kurki-Suonio.¹⁰⁹ If there are several choices for the definition of the local axes (for example, for $3m$ local symmetry) only the first one is always used in the program. For example, a 4-coordinated "main" atom in a pseudotetrahedral environment can have one of the following conformations among the first neighbors:

(1) All neighbors are chemically different (i.e., no chemically equivalent neighbors), no local symmetry, X -axis is chosen toward the closest neighbor, Y -axis is toward the second closest neighbor.

(2) Two atoms are chemically equivalent and the other two are chemically different from the first two and from each other (i.e., $2 + 1 + 1$), m local symmetry is assigned with a mirror plane passing through the main atom and the two nonequivalent neighbors (define X and Y local axes).

(3) Two pairs of chemically equivalent atoms (i.e., $2 + 2$), $mm2$ local symmetry with the 2-fold axis defined along the main atom and the midpoint between the chemically equivalent atoms of one of the pairs.

(4) Three atoms are chemically equivalent, $3m$ local symmetry with the 3-fold axis defined along the main atom and the neighbor which is chemically different from the other three.

(5) All four atoms are chemically equivalent, in an ideal case this site should be described by Kubic harmonic functions, but

for simplicity the highest possible symmetry of spherical harmonics, that is, $42m$, is selected.

Once the local symmetry of an atom is determined, the program automatically picks up all the symmetry-allowed pseudoatom functions.

As one can see, this method does not use any of the geometrical parameters of the first coordination sphere (i.e., bond lengths and angles); the bond distances are only used to identify the neighbors and are not used in determination of the chemical equivalency of these neighbors. At first glance this may seem to be a shortcoming of the method, but on the other hand, the distances and angles in large protein structures may not be defined accurately enough to make a decision on the chemical equivalency of the atoms. However, the incorporation of chemical-equivalency tests based on bond lengths and angles can be easily added to the current version of the program.

Atoms which belong to a single planar ring (the ring is defined to be planar¹¹⁰ if σ_{plane} is below 0.1\AA) are treated differently: they are assigned m local symmetry with the X -axis pointing toward the center of the ring for all atoms that belong to that ring and the Y -axis pointing toward one of the neighbors. If an atom belongs to a ring which is not planar or belongs to several planar rings, it is treated as if it does not belong to a planar ring at all, which helps eliminate cases in which the two rings to which such an atom belongs are not coplanar.

A different procedure is applied to the hydrogen atoms. They are always assigned cylindrical symmetry with the Z -axis pointing toward the parent atom. The chemical-equivalency test for hydrogen atoms is different from that described above for other atoms and is rather simple. The two hydrogen atoms are defined to be chemically equivalent if they are connected to an atom with (a) the same atomic number, (b) same total number of neighbors, and (c) same number of hydrogen atoms among neighbors.

Appendix C. Calculation of Electrostatic Properties

Local and integrated topological properties of charge densities based on the pseudoatom model were calculated with the program TOPXD,¹¹¹ whereas the corresponding analyses of the Gaussian 98 wave functions were done using the AIMPAC¹¹² suite of programs. AIM moments were generated with the program PROAIMV,¹¹³ locally modified to integrate Cartesian unbridged (instead of traceless) moments up to $l = 4$. The AIM analysis of ADF densities was performed with a recently

developed program TOPADF, which is based on programs TOPOND98¹¹⁴ and TOPXD. As in the other programs in TOPADF, the first and second derivatives of the density are evaluated analytically and the Cartesian unbridged moments with $l \leq 4$ are integrated.

Intermolecular electrostatic interaction energies based on pseudopotential parameters (both model and databank) were calculated with program MIN16^{115–117} which now has a direct interface with XD. Single-point molecular mechanics calculations of electrostatic interaction energies in dimers were performed with the TINKER^{118–121} package using the AMBER99,¹²² CHARM27,¹²³ and MM3¹²⁴ force fields and with Sybyl¹²⁵ using the MMFF94¹²⁶ force field.

Binding energies of molecules in crystals were calculated (a) *ab initio* with CRYSTAL03¹²⁷ at the B3LYP/6-31G** level of theory and corrected for the basis set superposition error (BSSE) using the counterpoise method¹²⁸ and (b) based on the pseudopotential expansion with program XDINTER.^{67,68}

Molecular graphics have been made with programs PLATON¹²⁹ and MOLEKEL.^{130,131}

Supporting Information Available: Deformation densities in SER and LEU; electrostatic potential mapped isosurfaces in SER and LEU; differences for selected properties of ρ at BCPs for SER and GLN relative to the G98/B3LYP/6-311++G(3df,3pd) calculation; differences for the AIM atomic moments for SER and LEU relative to the G98/B3LYP/6-311++G(3df,3pd) calculation; differences for the total AIM molecular moments for SER and GLN relative to the G98/B3LYP/6-311++G(3df,3pd) calculation. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- Xu, D.; Lin, S. L.; Nussinov, R. *J. Mol. Biol.* **1997**, *265*, 68.
- Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27.
- Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233.
- Stone, A. J.; Alderton, M. *Mol. Phys.* **1985**, *56*, 1047.
- Sokalski, W. A.; Poirier, R. A. *Chem. Phys. Lett.* **1983**, *98*, 86.
- Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Clarendon Press: Oxford, 1990.
- Hirshfeld, F. L. *Theor. Chim. Acta* **1977**, *44*, 129.
- Lee, J. G.; Friesner, R. A. *J. Phys. Chem.* **1993**, *97*, 3515.
- Williams, D. E. *J. Comput. Chem.* **1988**, *9*, 745.
- Breneman, C. M.; Wilberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.
- Dudek, M. J.; Ponder, J. W. *J. Comput. Chem.* **1995**, *16*, 791.
- Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, 864.
- Bader, R. F. W. *Can. J. Chem.* **1986**, *64*, 1036.
- Bader, R. F. W.; Larouche, A.; Gatti, C.; Carroll, M. T.; MacDougall, P. J.; Wiberg, K. B. *J. Chem. Phys.* **1987**, *87*, 1142.
- Chang, C.; Bader, R. F. W. *J. Phys. Chem.* **1992**, *96*, 1654.
- Cioslowski, J.; Nanayakkara, A. *J. Am. Chem. Soc.* **1993**, *115*, 11213.
- Popelier, P. L. A.; Bader, R. F. W. *J. Phys. Chem.* **1994**, *98*, 4473.
- Matta, C. F.; Bader, R. F. W. *PROTEINS: Struct., Funct., Genet.* **2000**, *40*, 310.
- Coppens, P. *X-ray Charge Densities and Chemical Bonding*; Oxford University Press: New York, 1997.
- Hansen, N. K.; Coppens, P. *Acta Crystallogr., Sect. A* **1978**, *34*, 909.
- Clementi, E.; Roetti, C. *At. Data Nucl. Data Tables* **1974**, *14*, 177.
- Wiest, R.; Pichon-Pesme, V.; Benard, M.; Lecomte, C. *J. Phys. Chem.* **1994**, *98*, 1351.
- Pichon-Pesme, V.; Lecomte, C.; Lachkar, H. *J. Phys. Chem.* **1995**, *99*, 6242.
- Jelsch, C.; Pichon-Pesme, V.; Lecomte, C.; Aubry, A. *Acta Crystallogr., Sect. D* **1998**, *54*, 1306.
- Muzet, N.; Guillot, B.; Jelsch, C.; Howard, E.; Lecomte, C. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 8742.
- Jelsh, C.; Teeter, M. M.; Lamzin, V.; Pichon-Pesme, V.; Blessing, R. H.; Lecomte, C. *PNAS* **2000**, *97*, 3171.
- Koritsanszky, T.; Volkov, A.; Coppens, P. *Acta Crystallogr., Sect. A* **2002**, *58*, 464.
- Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94*, revision E.2; Gaussian, Inc.: Pittsburgh, PA, 1995.
- Hariharan P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213.
- Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
- Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
- Allen F. H. *Acta Crystallogr., Sect. B* **2002**, *58*, 380.
- Koritsanszky, T.; Howard, S. T.; Richter, T.; Macchi, P.; Volkov, A.; Gatti, C.; Mallinson, P. R.; Farrugia, L. J.; Su, Z.; Hansen, N. K. *XD—A Computer Program Package for Multipole Refinement and Topological Analysis of Charge Densities from Diffraction Data*; 2003.
- Kenney, J. F.; Keeping, E. S. *Mathematics of Statistics*; Van Nostrand: Princeton, NJ, 1962; p 77.
- Loll, P. J.; Kaplan, J.; Selinsky, B. S.; Axelsen, P. H. *J. Med. Chem.* **1999**, *42*, 4714.
- Loll, P. J.; Axelsen, P. H. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 265.
- Faerman, C. H.; Price, S. L. *J. Am. Chem. Soc.* **1990**, *112*, 4915.
- Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650.
- Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; von R. Schleyer, P. J. *Comput. Chem.* **1983**, *4*, 294.
- Gill, P. M. W.; Johnson, B. G.; Pople, J. A.; Frisch, M. J. *Chem. Phys. Lett.* **1992**, *197*, 499.
- Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys.* **1984**, *80*, 3265.
- te Velde, G.; Bickelhaupt, F. M.; van Gisbergen, S. J. A.; Fonseca Guerra, C.; Baerends, E. J.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931.
- Guerra, C.; Snijders, J. G.; te Velde, G.; Baerends, E. *Theor. Chem. Acc.* **1998**, *99*, 391.
- ADF2003.01; SCM, Theoretical Chemistry; Vrije Universiteit: Amsterdam, The Netherlands. <http://www.scm.com>.
- Buckingham, A. D. *Adv. Chem. Phys.* **1967**, *12*, 107.
- Stone, A. J. *The Theory of Intermolecular Forces*; Oxford University Press: New York, 1997.
- Kistenmacher, T. J.; Rand, G. A.; Marsh, R. E. *Acta Crystallogr., Sect. B* **1974**, *30*, 2573.
- Gorbitz, H.; Dalhus, B. *Acta Crystallogr., Sect. C* **1996**, *52*, 1754.
- Koetzle, T. F.; Frey, M. N.; Lehmann, M. S.; Hamilton, W. C. *Acta Crystallogr., Sect. B* **1973**, *29*, 2571.
- International Tables for X-ray Crystallography*; Kluwer Academic Publishers: Dordrecht, Netherlands, 1992; Vol. C.
- Howard, S. T.; Hursthouse, M. B.; Lehmann, C. W.; Poyner, E. A. *Acta Crystallogr., Sect. B* **1995**, *51*, 328.
- Weist, R.; Pichon-Pesme, V.; Benard, M.; Lecomte, C. *J. Phys. Chem.* **1994**, *98*, 1.
- Swaminathan, S.; Craven, B. M.; Spackman, M. A.; Stewart, R. F. *Acta Crystallogr., Sect. B* **1984**, *40*, 398.
- Volkov, A.; Gatti, C.; Abramov, Yu.; Coppens, P. *Acta Crystallogr., Sect. A* **2000**, *56*, 252.
- Volkov, A.; Abramov, Yu. A.; Coppens, P.; Gatti, C. *Acta Crystallogr., Sect. A* **2000**, *56*, 332.
- Volkov, A.; Coppens, P. *Acta Crystallogr., Sect. A* **2001**, *57*, 395.
- Iversen, B. B.; Larsen, F. K.; Figgis, B. N.; Reynolds, P. A. *J. Chem. Soc., Dalton Trans.* **1997**, 2227.
- Bickelhaupt, F. M.; Baerends, E. J. *Rev. Comput. Chem.* **2000**, *15*, 1.
- Ziegler, T.; Rauk, A. *Theor. Chim. Acta* **1977**, *46*, 1.
- Ziegler, T.; Rauk, A. *Inorg. Chem.* **1979**, *18*, 1755.
- Volkov, A.; Coppens, P. *J. Comput. Chem.* **2004**, *25*, 921.
- Volkov, A.; Koritsanszky, T.; Coppens, P. Submitted.
- Spackman, M. A. *J. Chem. Phys.* **1986**, *85*, 6579.
- Spackman, M. A. *J. Chem. Phys.* **1986**, *85*, 6587.
- Spackman, M. A. *J. Phys. Chem.* **1987**, *91*, 3179.
- Spackman, M. A.; Weber, H. P.; Craven, B. M. *J. Am. Chem. Soc.* **1988**, *110*, 775.
- Abramov, Yu. A.; Volkov, A. V.; Wu, G.; Coppens, P. *J. Phys. Chem. B* **2000**, *104*, 2183.
- Abramov, Yu. A.; Volkov, A.; Wu, G.; Coppens, P. *Acta Crystallogr., Sect. A* **2000**, *56*, 585.
- Gavezzotti, A. *J. Phys. Chem. B* **2002**, *106*, 4145.
- Gavezzotti, A. *J. Phys. Chem. B* **2003**, *107*, 2344.
- Gavezzotti, A. *CrystEngComm* **2003**, *5*, 429.
- Gavezzotti, A. *CrystEngComm* **2003**, *5*, 439.
- Koch, M. H. L.; Germain, G. *Acta Crystallogr., Sect. B* **1970**, *26*, 410.

- (74) Marsh, R. E.; Ramakumar, S.; Venkatesan, K. *Acta Crystallogr., Sect. B* **1976**, *32*, 66.
- (75) Longchambon, F.; Ohanessian, J.; Gillier-Pandraud, H.; Duchet, D.; Jacquinet, J.-C.; Sinay, P. *Acta Crystallogr., Sect. B* **1981**, *37*, 601.
- (76) Pichon-Pesme, V.; Lachekar, H.; Souhassou, M.; Lecomte, C. *Acta Crystallogr., Sect. B* **2000**, *56*, 728.
- (77) Eggleston, D. S.; Valente, E. J.; Hodgson, D. J. *Acta Crystallogr., Sect. B* **1981**, *37*, 1428.
- (78) Prange, T.; Pascard, C. *Acta Crystallogr., Sect. B* **1979**, *35*, 1812.
- (79) Lalitha, V.; Subramanian, E.; Parthasarathy, R. *Int. J. Pept. Protein Res.* **1986**, *27*, 223.
- (80) Lalitha, V. Subramanian, E.; Parthasarathy, R. *Int. J. Pept. Protein Res.* **1986**, *27*, 472.
- (81) Krause, J. A.; Baures, P. W.; Eggleston, D. S. *Acta Crystallogr., Sect. B* **1993**, *49*, 123.
- (82) Glusker, J. P.; Carrell, H. L.; Berman, H. M.; Gallen, B.; Peck, R. M. *J. Am. Chem. Soc.* **1977**, *99*, 595.
- (83) Go, K.; Chaturvedi, S.; Parthasarathy, R. *Biopolymers* **1992**, *32*, 107.
- (84) Wu, S.; Tinant, B.; Declercq, J. P.; van Meerssche, M. *Bull. Soc. Chim. Belg.* **1987**, *96*, 361.
- (85) Banumathi, S.; Velmurugan, D.; Subramanian, E.; Katti, S. B.; Haq, W. *Acta Crystallogr., Sect. C* **1998**, *54*, 1681.
- (86) Sawka-Dobrowolska, W.; Glowiak, T.; Kozlowski, H.; Mastalerz, P. *Acta Crystallogr., Sect. C* **1990**, *46*, 1679.
- (87) Seetharaman, J.; Rajan, S. S.; Srinivasan, R. *J. Crystallogr. Spectrosc. Res.* **1993**, *23*, 167.
- (88) Baldwin, J. E.; Bamford, S. J.; Fryer, A. M.; Rudolph, M. P. W.; Wood, M. E. *Tetrahedron* **1997**, *53*, 5255.
- (89) Shamala, N.; Row, T. N. G.; Venkatesan, K. *Acta Crystallogr., Sect. B* **1976**, *32*, 3267.
- (90) Derricott, C.; Trotter, J. *Acta Crystallogr., Sect. B* **1979**, *35*, 2230.
- (91) Monn, J. A.; Valli, M. J.; Massey, S. M. *J. Med. Chem.* **1999**, *42*, 1027.
- (92) Satyshur, K. A.; Rao, S. T. *Acta Crystallogr., Sect. C* **1983**, *39*, 1672.
- (93) Kennard, G. H. L.; Smith, G.; White, A. H. *Acta Crystallogr., Sect. B* **1981**, *37*, 1317.
- (94) Smith, F.; Kennard, C. H. L.; White, A. H.; Hodgson, P. G. *Acta Crystallogr., Sect. B* **1980**, *36*, 992.
- (95) Byrne, J. J.; Chavant, P. Y.; Averbuch-Pouchot, M.-T.; Vallee, Y. *Acta Crystallogr., Sect. C* **1998**, *54*, 1154.
- (96) Wilson, C.; Howard, J. A. K.; Jane, D. E.; Sunter, D. C.; Watkins, J. C. *Acta Crystallogr., Sect. C* **1997**, *53*, 909.
- (97) Andre, D.; Fourme, R.; Renaud, M. *Acta Crystallogr., Sect. B* **1971**, *27*, 2371.
- (98) Czugler, M.; Kalman, A. *Cryst. Struct. Commun.* **1975**, *4*, 531.
- (99) Krohn, K.; Franke, C.; Jones, P. G.; Aust, H. J.; Draeger, S.; Schulz, B. *Liebigs Ann. Chem.* **1992**, *8*, 789.
- (100) Andersen, A. M. *Acta Chem. Scand. B* **1975**, *29*, 239.
- (101) Andersen, A. M. *Acta Chem. Scand. B* **1976**, *30*, 193.
- (102) Sheldrick, B. *Acta Crystallogr. B* **1976**, *32*, 1016.
- (103) Wang, J. L.; Tang, C.-P.; Hseu, T.-H.; Lan, S.-L. *Proc. Natl. Sci. Counc., Repub. China, Part B* **1982**, *6*, 162.
- (104) Cheng, W.-L.; Shaw, Y.-J.; Yeh, S.-M. *J. Org. Chem.* **1999**, *64*, 532.
- (105) Reppart, W. J.; Gallucci, J. C.; Lundstedt, A. P.; Gerkin, R. E. *Acta Crystallogr., Sect. C* **1984**, *40*, 1572.
- (106) Zhao, B.; Chen, C.; Zhou, Z.; Cao, Y.; Li, M. *J. Mater. Chem.* **2000**, *10*, 1581.
- (107) Rochlin, E.; Rappoport, Z. *J. Am. Chem. Soc.* **1992**, *114*, 230.
- (108) Wiechert, D.; Mootz, D. Private communication. **1999**.
- (109) Kara, M.; Kurki-Suonio, K. *Acta Crystallogr., Sect. A* **1981**, *37*, 201.
- (110) Stout, G. H.; Jensen, L. H. *X-ray Structure Determination: A Practical Guide*; Wiley-Interscience: New York, 1989.
- (111) Volkov, A.; Gatti, C.; Abramov, Yu.; Coppens, P. *Acta Crystallogr., Sect. A* **2000**, *56*, 252.
- (112) Biegler-König, F. W.; Nguyen-Dang, T. T.; Tal, Y.; Bader, R. F.; Duke, A. J. *J. Phys. B* **1981**, *14*, 2739.
- (113) Biegler-König, F. W.; Bader, R. F.; Tang, T.-H. *J. Comput. Chem.* **1982**, *13*, 317.
- (114) Gatti, C. *TOPOND98 Users' Manual*; CNR-CSR SRC: Milano, Italy, 1999.
- (115) Buckingham A. D.; Fowler, P. W. *J. Chem. Phys.* **1983**, *79*, 6426.
- (116) Buckingham A. D. In *Intermolecular Interactions—From Diatomics to Biopolymers*; Pullman, B., Ed.; Wiley: New York, 1978; Chapter 1.
- (117) Kisiel, Z. *PROSPE—Programs for Rotational Spectroscopy*. <http://info.ifpan.edu.pl/~kisiel/prospe.htm>.
- (118) Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933.
- (119) Pappu, R. V.; Hart, R. K.; Ponder, J. W. *J. Phys. Chem. B* **1998**, *102*, 9725.
- (120) Hodsdon, M. E.; Ponder, J. W.; Cistola, D. P. *J. Mol. Biol.* **1996**, *264*, 585.
- (121) Kundrot, C. E.; Ponder, J. W.; Richards, F. M. *J. Comput. Chem.* **1991**, *12*, 402.
- (122) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
- (123) MacKerell, A. D.; Wiorkiewicz-Juczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946.
- (124) Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551.
- (125) SYBYL, version 6.7; Tripos Inc.: St. Louis, MO.
- (126) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490.
- (127) Saunders, V. R.; Dovesi, R.; Roetti, C.; Orlando, R.; Zicovich-Wilson, C. M.; Harrison, N. M.; Doll, K.; Civalieri, B.; Bush, I.; D'Arco, Ph.; Llunell, M. *CRYSTAL2003 User's Manual*; University of Torino, Torino, Italy, 2003.
- (128) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
- (129) Spek, A. L. *J. Appl. Crystallogr.* **2003**, *36*, 7.
- (130) Flükiger, P. F. Development of the Molecular Graphics Package MOLEKEL and its Application to Selected Problems in Organic and Organometallic Chemistry. Ph.D. Thesis, Département de Chimie Physique, Université de Genève, Genève, 1992.
- (131) Portmann, S.; Lüthi, H. P. *Chimia* **2000**, *54*, 766.