Transferability of Atomic Volumes and Charges in the Peptide Bond Region in the Solid State

Birger Dittrich,*,† Stephan Scheins,† Carsten Paulmann,‡ and Peter Luger†

Institut für Chemie/Kristallographie, Freie Universität Berlin, Takustrasse 6, D-14195 Berlin, Germany, and Mineralogisches Institut der Universität Hamburg, Grindelallee 48, D-20146 Hamburg, Germany

Received: December 12, 2002; In Final Form: March 17, 2003

Bader's partitioning scheme that makes use of the zero flux surfaces (ZFS) in the electron density gradient vector field has been applied to the peptide bond regions of three oligopeptides (two dipeptides, one hexapeptide), yielding eight peptide bond regions to compare. From integration over the atomic volume, very reproducible atoms in molecules (AIM) charges were calculated which agree within the given atom types by 0.04−0.08 *e*. The polarization of the peptide bond atoms is high compared to the charges normally used in force field parametrization. The positive charges of the C_α, C′, and H atoms sum up to $\approx +1.6$ *e*, while the negative charges of N and O amount to ≈ -1.85 *e*, so that for each peptide bond region an excess of −0.25 *e* has to be compensated by the C_α hydrogen and the side chains.

Introduction

In Bader's theory of atoms in molecules (AIM)[1] the experimentally obtainable charge density $\rho(r)$ plays a central role, in that major properties of a chemical system are functionally related to its distribution of charge.[2] One important aspect of this theory is the partitioning of a molecular structure into submolecular regions, functional groups or single atoms. Having, on one hand, the tools to obtain well-defined submolecular fragments, it is, on the other hand, an important question whether the atomic properties can be used to construct or predict molecular ones, that is, to what extent submolecular topological properties are additive and transferable to larger systems. The procedure of partitioning a molecule into atomic regions makes use of the zero-flux surfaces (ZFS) in the electron density gradient vector field $\nabla\rho(r)$.[1] Surfaces of this type establish atomic basins around nuclear attractors of the corresponding trajectories of $\nabla\rho(r)$ and uniquely define atomic volumes. They can be used to evaluate a number of atomic or functional group properties; for instance, atomic charges can be obtained by integration over the charge density in the given atomic volume. Algorithms for the calculation of these AIM charges from experimental charge densities have become available recently, for example, by VALRAY[3] and through the TOPXD program;[4] however, applications have been restricted mainly to smaller molecules.[5−7]

Bader and co-workers have studied the possibility of a theoretical construction of polypeptides[8,9] and other, chemically more complicated molecules.[10] They have shown that the properties of a number of oligopeptides can be predicted from those corresponding to the constituent amino acid fragments or functional groups.

Koritsanszky et al. have investigated several tripeptides by calculation of structure factors from wave functions for isolated molecules followed by a multipole refinement.[11] Their approach, building a database of transferable pseudoatoms for improved refinement and prediction of electrostatic properties for peptides, follows Pichon Pesme et al.[12] but uses only theoretical data.

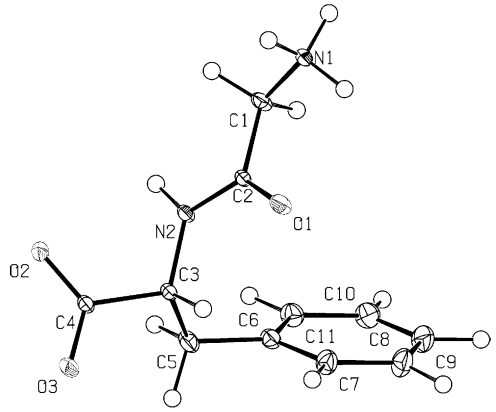

Figure 1. ORTEP [15] representation (left) with atomic numbering scheme (50% probability).

Our complementary approach to that question is based on experimental charge density determinations on some oligopeptides. Here we report on the results of such a study on the centrosymmetric dipeptide Gly-DL-Phe (see molecular structure and atomic numbering in Figure 1), where a full topological analysis to derive bonding properties was performed and in addition atomic volumes and charges were derived from the experimental charge density. Atomic partitioning was also applied to the dipeptide Gly-L-Thr dihydrate and the hexapeptide cyclo-(L-Ala)$_4$-(D,L-Pro)$_2$ monohydrate. Their experimental charge densities were determined earlier, but at the time of their publication[13,14] the partitioning algorithm was not applied. From these data, atomic properties of eight peptide groups of three different oligopeptides are available, so that some comparative results can be obtained.

Experimental Section

For Gly-DL-Phe (see Table 1), the charge density was determined from a high-resolution synchrotron/CCD area detector experiment at 100 K following a similar experimental

* Fax: (+49) 30 83853464. E-mail: birger@chemie.fu-berlin.de.
† Freie Universität Berlin.
‡ Mineralogisches Institut der Universität Hamburg.

10.1021/jp022584s CCC: $25.00
Published on Web 08/21/2003

TABLE 1. Selected Crystal Data and Experimental Conditions for Gly-DL-Phe

empirical formula	$C_{11}H_{14}N_2O_3$
formula weight	222.24
measurement temperature	100 K
sample detector distance	8 cm (low order), 6 cm (medium and high order)
time per frame	2 s (low order), 4 s (medium), 8 s (high order)
increment	0.2 in ω or ϕ
crystal size	$0.42 \times 0.16 \times 0.12$ mm^3
crystal system, space group	*Pbca*
wavelength λ	0.540 Å
lattice constants (Å)	$a = 9.213(1)$, $b = 28.108(1)$, $c = 8.619(1)$
unit cell volume	2231.96 Å^3
Z	8
density (calc)	1.323 g cm^{-1}
absorption coefficient	0.06 mm^{-1}
$F(000)$	944 e
min/max. hkl (before merging)	$-21 < h < 20$, $-48 < k < 64$, $-14 < l < 19$
total number of reflections	77231
unique reflections	14106
$F_o > 2.5\sigma(F_o)$	11404
resolution $(\sin\theta/\lambda)_{max}$	1.16 Å^{-1}
$R_{int}(F^2)$	0.035
$R_1(F)$	0.027
$R_w(F)$	0.029
R_{all}	0.041
GoF	1.77
overall redundancy	4.85
completeness	97%
diffractometer	Huber Eulerian Cradle
beamline (HASYLAB)	F1
detector	Smart CCD

procedure as described in ref 13. The data were interpreted with the Hansen−Coppens multipole formalism.[16] The multipole refinement (starting atomic parameters taken from Marsh et al.[17]) was carried out with the full-matrix LSQ program (XDLSM) of the XD program package.[18] In the multipole formalism the core and the spherical valence density of the heavy atoms were composed of Hartree−Fock wave functions expanded over Slater type basis functions. A total number of 481 variables, of which 280 were multipolar parameters (including 12 κ-parameters), were refined. To reduce the number of variables, m symmetry was used for the carbon atom C(4) and the aromatic carbon atoms. They were, like the hydrogen atoms of the phenyl ring, of the amino group, and of the methyl group, chemically constrained. A list of all multipole parameters as well as maps of the deformation density and the residual electron density in the planes chosen for the gradient vector field can be found in the Supporting Information. Bond distances to hydrogen atoms were normalized to standard neutron diffraction distances. After multipole refinement using hexadecapoles for C, N, and O and bond directed dipoles for hydrogen atoms, the R-factor was 2.7%.

For Gly-L-Thr dihydrate and cyclo-(L-Ala)$_4$-(D,L-Pro)$_2$ monohydrate, similar procedures to derive the charge density have been applied and recently been published.[13,14]

For the calculation of the theoretical electron density, the programs Gaussian 98[19] and AIMPAC[20] were used.

Results and Discussion

In Table 2 bond critical points are given for Gly-DL-Phe for some chosen covalent bonds and compared to the results from HF and B3LYP quantum chemical calculations on the experimental geometry as well as to mean values obtained for 13 of the naturally occurring amino acids.[21] The agreement of the experimental topological descriptors with the mean values of the literature[22] is 0.1 e Å^{-3} for $\rho(r_{cp})$ and 3 e Å^{-5} for $\nabla^2\rho(r_{cp})$, confirming, as stated earlier,[14,22] that this is the range where transferability of topological parameters can be considered. The agreement between experiment and theory is in the same range as that given above, with the exception for the Laplacians of the C−O bonds, which was also found earlier and attributed to the inflexibility of the deformation radial functions.[21,23]

TABLE 2. Selected Topological Bond Descriptors in Gly-DL-Phe[a]

bond	$\rho(r_{cp})$	$\nabla^2\rho(r_{cp})$	d	method
O(2)−C(4)	2.76(3)	−32.4(2)	0.7613	experimental
	2.65	−17.7	0.8348	HF/6-311++G(2d,2p)
	2.60	−19.2	0.8159	B3LYP/6-311++G(2d,2p)
	2.83(11)	−35.6(36)	0.772(15)	lit.[21]
O(3)−C(4)	2.62(3)	−31.6(2)	0.7843	exp
	2.59	−17.6	0.8408	HF
	2.55	−19.4	0.8208	B3LYP
	2.71(9)	−33.6(45)	0.791(17)	lit.[21]
N(1)−C(1)	1.82(3)	−15.4(1)	0.8562	exp
	1.62	−11.4	0.9925	HF
	1.62	−14.4	0.9171	B3LYP
	1.68(5)	−10.8(21)	0.853(17)	lit.[21]
C(3)−C(4)	1.68(2)	−12.7(1)	0.7570	exp
	1.75	−16.4	0.8018	HF
	1.64	−12.7	0.7968	B3LYP
C−C$_\alpha$	1.75(6)	−13.0(22)	0.759(19)	lit.[21]
C(3)−C(5)	1.62(2)	−7.0(1)	0.7785	exp
	1.63	−13.4	0.7855	HF
	1.55	−10.7	0.7838	B3LYP
C$_\alpha$−C$_\beta$	1.68(8)	−11.7(21)	0.778(16)	lit.[21]
C(2)−N(2)	2.28(3)	−23.2(1)	0.7910	exp
	2.38	−32.6	0.8447	HF
	2.31	−26.6	0.8065	B3LYP

[a] $\rho(r_{cp})$ denotes the electron density, and $\nabla^2\rho(r_{cp})$, the Laplacian at the bond critical point, r_{cp}. d is the distance from the first atom defining the bond to the cp.

Figure 2 shows gradient vector fields of Gly-DL-Phe together with the molecular graph: one in the plane of the peptide bond and the other one in the plane of the phenyl group. They illustrate the shape of the individual atoms.

Atomic volumes and charges in the peptide bond region are summarized in Figure 3. The averages of comparable quantities show that the internal consistency for volumes is <1 Å^3 for N but better than 0.2 Å^3 for C$_\alpha$. The atomic volumes at the C$_\alpha$ and C′ atoms are equal within the statistical error, except if C$_\alpha$ belongs to a Gly residue. The average volumes of the (non Gly) C$_\alpha$ atoms and the C′ atoms are by 1.5 Å^3 smaller than those of the glycine C$_\alpha$ atoms, where the second hydrogen allows the carbon atom to expand. The high spread of the N atomic volumes is mainly caused by the nitrogens in the proline residues of the hexapeptide, where they are part of the five membered ring and bonded to a second carbon atom instead of hydrogen, which reduces the volume by about 2 Å^3 analogous to the above quoted volume expansion for the glycine C$_\alpha$.

The AIM charges (averages; see also Figure 3) agree within the given atom types by 0.04−0.08 e, which is a surprisingly small spread. The C$_\alpha$ atoms carry a small positive charge, the hydrogens of the peptide N−H carry a medium positive charge, and the C′ atoms carry a high positive charge, while strong negative charges of almost 1 e are seen on the N and O atoms. These experimental results indicate that the polarization of Bader atoms is much higher than that obtained, for example, from theoretical orbital methods (NBO or Mulliken charges)[24] or than that used in force field parametrization. The AMBER[25] force field, for example, uses charges of −0.5 for oxygen, +0.5 for

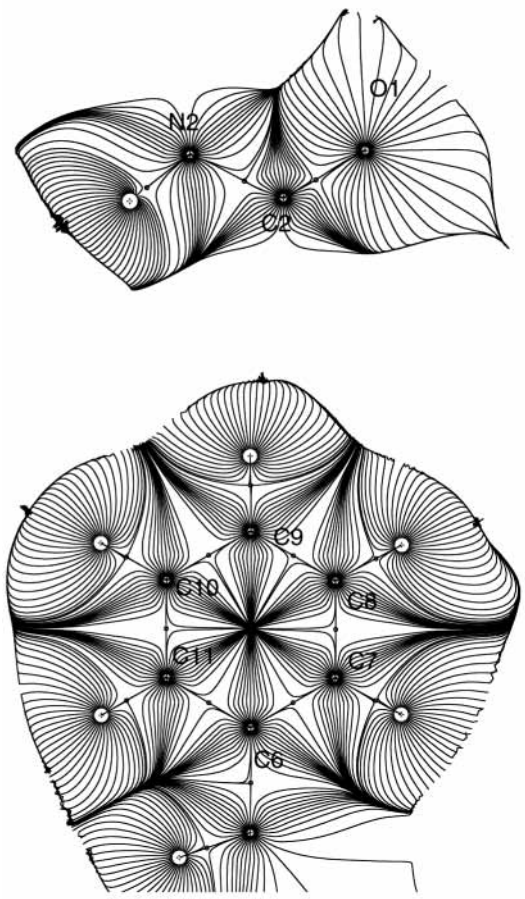

Figure 2. Gradient vector field in the peptide bond plane (top) and the phenyl ring plane (bottom) of Gly-DL-Phe.

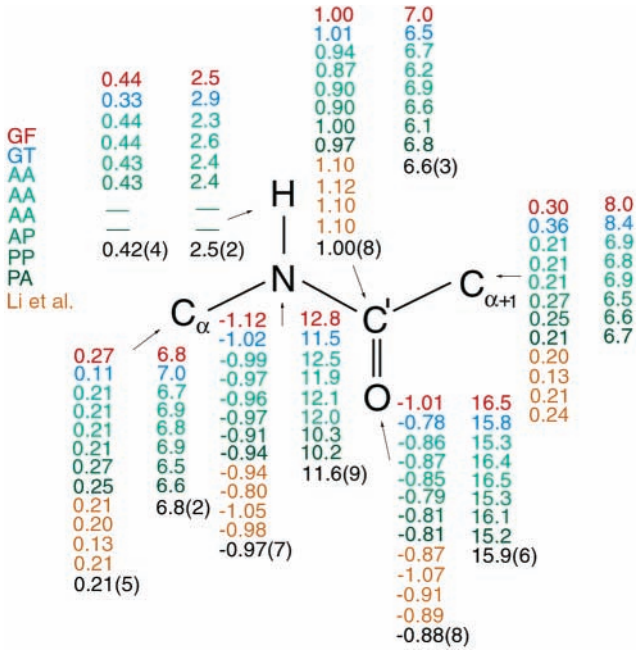

Figure 3. Comparison of AIM charges and volumes for different peptide bonds: The values correspond to the residues given in the colored two letter code. Brown values were taken from Li et al.[26] Left columns, AIM charges (*e*); right ones, volumes (Å^3); last entries, averages with esd's (from the mean values) in parentheses. Mean values for $C_{\alpha+1}$ are not given, as they are mostly contained in the C_α column.

carbon, −0.57 for nitrogen, and +0.37 for hydrogen atoms in the peptide bond. The definition and determination of atomic charges has been under controversial discussion in the last years,[24] and charges derived from different methods may differ significantly. AIM charges are based on well-defined atomic segments of the electronic charge density and can be derived from an experiment.

Koritsanszky et al.[11] compare AIM charges directly obtained from the isolated molecule DFT wave function for the atoms in the peptide bond region with charges obtained from the multipole fit and discuss the differences. Their results support our experimental findings, although differences exist, mainly at the C′ atom.

Very recently Li et al.[26] have published AIM atomic charges for a pseudopentapeptide of which a structural study has been published before.[27] They make a comparison to AIM charges from a molecular theoretical DFT calculation and find that the theoretical AIM charges almost consistently exceed the experimental AIM charges, indicating a possible systematic difference apart from the effect of intermolecular interaction.

Their experimental AIM charges for the atoms in the peptide bond (volumes were not published) are included in Figure 3 and fit very well to the charges derived for our three substances. Two of their peptide bonds, the one in the glutamine rest group and the bond to the boc protected residue, were, due to the different chemical environment, not included.

As the three molecules studied here are quite large, a comparison to periodical theoretical calculations exceeds our actual computational capacities. The interesting question of the origin of the discrepancies between theoretical isolated molecule or crystal environment AIM charges on one hand and the experimental values on the other, as reported in ref 26, requires further investigations.

A preliminary conclusion of this study for the peptide bond is that very reproducible charges for the contributing atoms were derived. The positive charges of the C_α, C′, and H atoms sum up to approximately +1.6 *e*, while the negative charges of N and O amount to ≈ −1.85 *e*, so that for each peptide bond region the excess of −0.25 *e* has to be compensated by the neighboring atoms, the C_α hydrogen and C_β of the side chain or, in the case of a glycine residue, another hydrogen atom. Although our results summarized in Figure 3 are based on three different charge density investigations under (in parts) different experimental conditions with eight peptide bonds involved, the sample is too small for a general conclusion based on a sound statistic, more so, since the three investigated molecules contain only 5 of the 20 usually occurring amino acid residues. Hence, further studies on this subject are required.

Acknowledgment. We thank the Deutsche Forschungsgemeischaft (Lu 222/22-1, 22-2) for financial support and M. Messerschmidt for help with using TOPXD.

Supporting Information Available: List of all multipole parameters as well as maps of the deformation density and the residual electron density in the planes chosen for the gradient vector field. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Bader, R. F. W. *Atoms in Molecules—A Quantum Theory*; Clarendon Press: Oxford, 1990; *Angew. Chem., Int. Ed. Engl.* **1994**, *33*, 620.
(2) Bader, R. F. W.; Bayles, D. *J. Phys. Chem. A* **2000**, *104*, 5579.
(3) Flensburg, C.; Madsen, D. *Acta Crystallogr.* **2000**, *A56*, 24.
(4) Volkov, A.; Gatti, C.; Abramov, Y.; Coppens, P. *Acta Crystallogr.* **2000**, *A56*, 252.

(5) Messerschmidt, M.; Wagner, A.; Wong, M. W.; Luger, P. *J. Am. Chem. Soc.* **2002**, *124*, 732.

(6) Bach, A.; Lentz, D.; Luger, P.; Messerschmidt, M.; Oelsch, C.; Patzschke, M. *Angew. Chem.* **2002**, *114*, 311; *Angew. Chem., Int. Ed.* **2002**, *l41*, 296.

(7) Zhurova, E. A.; Tsirelson, V. G.; Stash, A. I.; Pinkerton, A. A. *J. Am. Chem. Soc.* **2002**, *124*, 4574.

(8) Chang, C.; Bader, R. F. W. *J. Phys. Chem.* **1992**, *96*, 1654.

(9) Popelier, P. L. A.; Bader, R. F. W. *J. Phys. Chem.* **1994**, *98*, 4473.

(10) Matta, C. F. *J. Phys. Chem. A* **2001**, *105*, 11088.

(11) Koritsanszky, T.; Volkov, A.; Coppens, P. *Acta Crystallogr.* **2002**, *A58*, 464.

(12) Pichon Pesme, V.; Lecomte, C.; Lachekar, H. *J. Phys. Chem.* **1995**, *99*, 6242.

(13) Dittrich, B.; Flaig, R.; Koritsánszky, T.; Krane, H.-G.; Morgenroth, W.; Luger, P. *Chem. Eur. J.* **2000**, *14*, 2582.

(14) Dittrich, B.; Koritsánszky, T.; Grosche, M.; Scherer, W.; Flaig, R.; Wagner, A.; Krane, H. G.; Kessler, H.; Riemer, C.; Schreurs, A. M. M.; Luger, P. *Acta Crystallogr.* **2002**, *B58*, 721.

(15) Burnett, M. N.; Johnson, C. K. ORTEP-III, Oak Ridge Thermal Ellipsoid Plot Program for Crystal Structure Illustrations; Oak Ridge National Laboratory Report ORNL-6895; Oak Ridge, TN, 1996.

(16) Hansen, N. K.; Coppens, P. *Acta Crystallogr.* **1978**, *A34*, 909.

(17) Marsh, R. E.; Ramakumar, S.; Venkatesan, K. *Acta Crystallogr.* **1976**, *B32*, 66.

(18) Koritsánszky, T.; Howard, S.; Mallinson, P. R.; Su, Z.; Richter, T.; Hansen, N. K. XD—A computer program package for multipole refinement and analysis of electron densities from diffraction data. User manual; Freie Universität Berlin: 1997.

(19) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; 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. *Gaussian 98*, revision A.7; Gaussian, Inc.: Pittsburgh, PA, 1998.

(20) AIMPAC program package; McMaster University: Hamilton, Ontario, 1992.

(21) Flaig, R.; Koritsánzky, T.; Dittrich, B.; Wagner, A.; Luger, P. *J. Am. Chem. Soc.* **2002**, *124*, 3407.

(22) Pichon Pesme, V.; Lachekar, H.; Souhassou, M.; Lecomte, C. *Acta Crystallogr.* **2000**, *B56*, 728.

(23) Volkov, A.; Coppens, P. *Acta Crystallogr.* **2001**, *A57*, 395.

(24) Wiberg, K. B.; Raben, P. R. *J. Comput. Chem.* **1993**, *14*, 1504.

(25) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(26) Li, X.; Abramov, A. V.; Coppens, P. *Proc. Natl. Acad. Sci.* **2002**, *99*, 12132.

(27) Abramov, Y.; Volkov, A.; Wu, G.; Coppens, P. *Acta Crystallogr.* **2000**, *A56*, 585.