A Theoretical Databank of Transferable Aspherical Atoms and Its Application to Electrostatic Interaction Energy Calculations of Macromolecules

Paulina M. Dominiak,* Anatoliy Volkov,* Xue Li, Marc Messerschmidt, and
Philip Coppens*

Department of Chemistry, State University of New York at Buffalo, New York 14260

Received June 14, 2006

Abstract: A comprehensive version of the theoretical databank of transferable aspherical pseudoatoms is described, and its first application to protein−ligand interaction energies is discussed. The databank contains all atom types present in natural amino acid residues and other biologically relevant molecules. Each atom type results from averaging over a family of chemically unique pseudoatoms, taking into account both first and second neighbors. The spawning procedure is used to ensure that close transferability is obeyed. The databank is applied to the syntenin PDZ2 domain complexed with four-residue peptides and to the PDZ2 dimer. Analysis of the electrostatic interactions energies calculated by the exact-potential/multipole-moment-databank method stresses the importance of the P_0 and P_{-2} residues of the peptide in establishing the interaction, whereas the P_{-1} residue is shown to play a much smaller role. Unexpectedly, the charged P_{-3} residue contributes significantly to the interaction. The class I and II peptides are bound with the same strength by the syntenin PDZ2 domain, though the electrostatic interaction energy of the P_{-2} residue is smaller for class I peptides. There is no difference between the interaction energies of the peptides with PDZ2 domains from single-domain protein fragments and those from PDZ1−PDZ2 tandems.

Introduction

In a preceding paper, we have described a databank of transferable aspherical atoms to be used in the evaluation of electrostatic interactions between large macromolecules and in the refinement of high-resolution structures of molecules such as proteins, polysaccharides, and nucleic acids.[1] The databank was shown to give an excellent reproduction of the electron density in a number of amino acids when compared with those calculated with conventional ab initio methods at the B3LYP/6-31G** level, while requiring only a small fraction of the computational time. It was shown that the databank allows calculation of the electrostatic interaction energies of amino acid dimers with an accuracy

of ~3 kJ per mole.[2] Its merits relative to the experimental databank pioneered by Pichon-Pesme et al.[3] have been the subject of a previous discussion.[4,5] In a more recent article, the databank has been used to evaluate the electrostatic component of the interaction between fragments of the antibiotic vancomycin and substrate molecules, with the nonelectrostatic contribution evaluated with perturbation-theory-based pairwise atom−atom potentials.[6]

The databank may be compared with the invariom approach, pursued by Dittrich and co-workers, in which aspherical atoms are defined by one atom in an as close as possible bonding environment.[7−9] In our procedure, a large number of molecules is analyzed and chemical similar atoms are grouped into families with closely related multipole population and expansion−contraction parameters. Standard deviations are calculated from the spread in the parameters within one family. A new family is spawned when significant

* Corresponding authors. Fax: (716) 645-6948 (P.C.). E-mail: pd24@buffalo.edu (P.M.D.), volkov@buffalo.edu (A.V.), coppens@buffalo.edu (P.C.).

outliers occur. This procedure ensures that close transferability is obeyed and implicitly accounts for differences in bond order in the atom's coordination sphere.

We note that reconstructing the electron density from transferable aspherical molecular fragments stored in a databank has been pursued by other research groups using quantum-chemical partitioning quite different from that employed here. The fuzzy density fragmentation method, for example, has been introduced by Mezey et al.[10,11]

We describe here the extension of the databank by systematic application of the spawning procedure. The new version includes all atom types encountered in amino acid residues. In a first application to a large protein molecule, the databank is applied to the interactions between the PDZ2 domain of the scaffolding protein syntenin and peptides. The results shed light on the nature of the interaction between the protein and the various residues, which has been the subject of recent discussions.[12,13]

Methodology

Theoretical Calculations. Theoretical calculations were performed as described by Volkov et al.[1] To construct the databank, single-point calculations on selected small molecules (Table S1, Supporting Information) were performed with the Gaussian 98 and Gaussian 03 programs[14] using density functional theory (DFT) with a standard split-valence double-exponential 6-31G** basis set including polarization functions.[15] The DFT calculations were based on Becke's three-parameter hybrid method[16] combined with the nonlocal correlation functional of Lee, Yang, and Parr (B3LYP).[17] Experimental molecular geometries as recorded in the Cambridge Structural Database were used.[18] Hydrogen positions were obtained by extending X$-$H distances to their standard neutron diffraction values.[19] Complex static valence-only structure factors in the range $0 < \sin\theta/\lambda < 1.1$ Å^{-1} were obtained by the analytic Fourier transform of the molecular charge densities for reciprocal lattice points corresponding to a pseudo-cubic cell with 30 Å edges.[20]

Aspherical Atom Refinement of the Theoretical Structure Factors. The data were fitted with the Hansen$-$Coppens pseudoatom formalism,[21] using version 4.12 of the XD program suite.[22] Both radial screening factors (κ and 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. 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.

Atomic scattering factors were based on the atomic wave functions of Clementi and Roetti[23] and were, in the course of the refinement, modified by the κ parameter. For the deformation functions single-ζ exponents corresponding to weighted averages over the s- and p-shell values given by

Table 1. Slater Radial Function Coefficients, n_l and ζ, Applied in the Refinements of Theoretical Structure Factors

element	n_l	ζ [bohr^{-1}]
C	2, 2, 3, 4	3.176
N	2, 2, 3, 4	3.839
O	2, 2, 3, 4	4.466
S	2, 4, 6, 8[a]	3.851
Cl	4, 4, 4, 4	4.259
H	1, 2	2.000

[a] As suggested by Dominiak and Coppens.[25]

Clementi and Raimondi[24] were used, with powers of r (n_l) given in Table 1. They were multiplied by the refinable κ' parameters.

The variables$-P_{val}$, the valence electrons population (in the model used equivalent to the spherical monopole population); P_{lm}, populations of the deformation functions; κ and κ', the expansion$-$contraction parameters$-$were averaged over each atom type and constitute the prime information stored in the databank.

Local Coordinate System Assignment. Averaging of the multipole populations of equivalent pseudoatoms requires that the spherical harmonic functions centered at these atoms be expressed in a common local frame. The program LSDB[1] includes a fully automatic definition of unique local coordinate systems based on the coordination environment of each atom. It is an essential tool in the charge density analysis of any large molecule, for which the manual assignment of coordinate systems becomes prohibitively cumbersome, and in the building of the pseudoatom databank based on a large number of small molecules.

Consistency in selecting the local coordinate system is crucial for defining the atom types and subsequent averaging. In the first step, the local coordinate system is oriented such as to allow local symmetry constraints. Whenever the local symmetry allows more than one possibility of selecting a particular local coordinate axis, the following procedure is used to choose the neighboring atom 1, which defines axis 1:[26]

(1) Atoms with the highest atomic number are selected, and if there is only one such atom, it is chosen as atom 1.

(2) If step 1 is inconclusive, atoms with the highest atomic number are ordered according to their hybridization state (sp first, then sp^2 and sp^3) and then by their valency (number of attached atoms; the smallest number first). The first atom from the list is chosen as atom 1.

(3) If the preceding procedure is not conclusive, that is, there is more than one atom at the top of the list with the same element type, hybridization, and valency, atoms are sorted according to their distance to the central atom, and the one with the shorter distance is chosen as atom 1.

After atom 1 is selected, atom 2 is chosen from the remaining neighboring atoms following the rules defined above. In some cases, it is necessary to define a dummy atom to make use of the symmetry, as for example for the central C atom of the $-C-CH_2-C-$ group, for which the dummy atom will be placed at the midpoint between the heavier carbon atoms, and axis 1 will be oriented along the local 2-fold symmetry axis. Such use of local symmetry allows
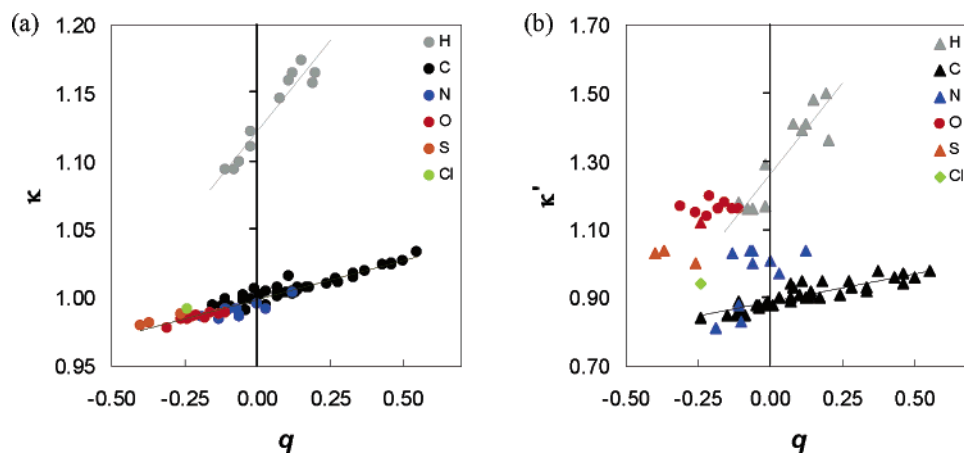
Figure 1. Correlation between the monopole-derived charge, q, and the expansion–contraction parameter (a) κ or (b) κ' for hydrogen (grey), carbon (black), nitrogen (blue), oxygen (red), sulfur (orange), or chlorine (green) atom types.

minimization of the number of multipole parameters. In case of atoms belonging to a planar ring, axis 1 is always oriented toward the center of the ring and mirror plane symmetry is always applied even if higher symmetry is possible. Usually, a right-handed coordinate system is defined, except in the case of chiral atoms, for which both right- and left-handed systems are allowed. Chirality is defined locally, that is, only by the character of the nearest neighbors.

In the initial stages, very low symmetry or no symmetry at all was applied in refinement. In case symmetry was evident after averaging over an atom type, that symmetry was adopted in the final refinement of the molecules. Residual density maps were examined carefully to ensure that the symmetry constraint was justified.

Atom Type Definitions. The key step in the construction of any databank is the choice of atom types, that is, the selection of atoms which are sufficiently similar to be treated identically. A compromise must be struck between achieving the most accurate representation of the charge density of many molecules and minimizing the number of atom types.

In the construction of the databank, a new atom type is introduced whenever the average value of any deformation density parameter over a group of atoms differs by an amount greater than one standard deviation of the sample consisting of the remaining atoms. This approach leads to the following set of general criteria for the definition of an atom type: (1) element type, (2) the number of attached atoms (atom valence, number of the nearest neighbors), (3) nearest neighbor type, which may be affected by the next-nearest neighbors, (4) aromaticity, and (5) local symmetry.

Atoms that are part of planar ring systems are given an aromaticity flag and treated as separate types. Usually, only element type and hybridization are taken into account to characterize any neighboring atom, except for neighboring sp^3 or sp^2-hybridized nitrogen atoms, which are additionally subdivided into $sp^3(4)$, $sp^3(3)$, $sp^2(3)$, and $sp^2(2)$, where the number in parentheses represents the number of neighbors of the nitrogen atom. Hybridization states of neighboring atoms are derived solely from the number of attached atoms and, in the case of nitrogens, planarity of the group. The atom types in the current version of the databank are listed

in Table 2, whereas a list of neighbor types used in construction of the databank is given in Table 3.

In some cases, the criteria described above are modified by including the effect of next-nearest neighbors or, on the other hand, by ignoring differences between nearest neighbors. An example of the first situation is oxygen atom types O101–O104[27] (Table 2), while the second occurs for carbon atoms such as C405 for which the type of neighboring C atom does not affect the central atom's population parameters. A simpler definition of an atom type is sometimes used when the number of occurrences of a intended atom type in the molecular sample is too small to get statistically meaningful average values of the deformation-density parameters.

For each atom, the LSDB program analyzes the coordination environment and assigns the appropriate atom type. The corresponding charge-density parameters are then transferred from the databank for the subsequent analysis.

Atom Types in the Databank. At the time of writing, the databank contained 74 atom types (Table 2). (Note added in proof: 104 atom types are incorporated in the version made available on the Web in October 2006.) It includes all atom types encountered in peptides, proteins, and some other biologically interesting molecules.

As noted before,[28,29] the monopole-derived net atomic charges correlate with the expansion–contraction parameters. The κ parameters are strongly related to the charges (Figure 1a) and form two distinctive groups: those describing hydrogen atom types with a fitted linear equation $\kappa = 1.122(3) + 0.27(3)q$, $R = 0.96$, and those for non-hydrogen atoms which are fitted by $\kappa = 0.9987(5) + 0.056(2)q$, $R = 0.96$. A somewhat less clear linear dependence of κ' on q[30] is observed for hydrogen and carbon atom types (Figure 1b). For the other elements, the range of charges is insufficient to examine any correlation, and unlike for κ/q, the points in the graph for N and O do not fall on the line derived for carbon (Figure 1b).

The net atomic charges of the pseudoatoms do not always reflect accepted electronegativity concepts, as they depend crucially on the division of the charge in the bonding regions. For the hydrogen atoms, for example, refinement of both bond-directed dipoles and quadrupoles effectively allows

Table 2. Atom Types in the Databank[a]

Carbon

C301	**C302**	**C303**	**C304**	**C306**
C351	**C352**	**C355**	**C357**	**C358**
C359	**C360**	**C361**	**C362**	**C363**
C370	**C371**	**C402**	**C403**	**C405**
C409	**C411**	**C412**	**C414**	**C415**
C418	**C419**	**C420**	**C421**	**C424**
C425 (6)	**C426** (5)	**C433** (9)	**C434** (6)	**C436** (23)
C437	**C438**	**C439**	**C440**	**C443**

Nitrogen

N402	**N405**	**N301**	**N302**	**N303**
N307	**N309**	**N310**	**N311**	**N201**

Table 2 (Continued)

Oxygen

O101	O102	O103	O104	O201
0202	0204	0206	0207	

Sulfur

S206	S207	S208

Chlorine

Cl01

Hydrogen

H101	H102	H103	H104	H105
H107	H108	H109	H110	H111

H112

[a] The symbol D represents dummy atoms, required for coordinate system definition.

Table 3. Labeling of the Nearest and Next-Nearest Neighbor Atom Types[a]

C	any Csp^2 or Csp^3 carbon
C_1	Csp^2 aromatic and nonaromatic
C_2	Csp^3
Car	Csp^2 in planar ring
N_1	$Nsp^2(2)$
N_2	$Nsp^2(3)$
N_3	$Nsp^3(3)$
N_4	$Nsp^3(4)$
O_1	$Osp^2(1)$
O_2	$Osp^3(2)$
S_2	$Ssp^3(2)$
X	any non-hydrogen atom

[a] Numbers in parentheses represent number of bonds to the atom.

assignment of the X−H bond density to the H atom, which becomes slightly negatively charged when bonded to carbon atoms. In general, the electron population on atoms linked to hydrogens decreases as the number of bonded hydrogen atoms increases. The effect is clearly visible for X−Csp^3−

R_3 carbon atom types (with X = C, $Nsp^2(3)$, $Nsp^3(3)$, $Nsp^3(4)$, $Osp^3(2)$, or $Ssp^3(2)$ and R = C or H, where the number in parentheses represents the number of neighboring atoms) when atom types with the same X substituent are compared, as shown in Figure 2. The spherical hydrogen-atom population represented by the monopoles correlates with the density of the even and odd multipoles on the H atoms,[31] which give a total density more strongly concentrated along the bond axis and toward the bonded X atom (Figure 3). A similar situation occurs for the monopole populations of X−Csp^3−R_3 carbon-atom types and the octupole and hexadecapole populations (Figure 4).[32]

Carbon Atoms. The databank contains 40 carbon atom types, 23 of which are sp^3 carbons (numbers from C402 to C443). The deformation density parameters for Csp^3 are given in Table S2 (Supporting Information). The corresponding deformation density maps are presented in Figure 5. All sp^3-carbon types are positive, with the monopole-defined charge ranging from +0.03(3) *e* for the C434 type to +0.55(6) *e* for C409.
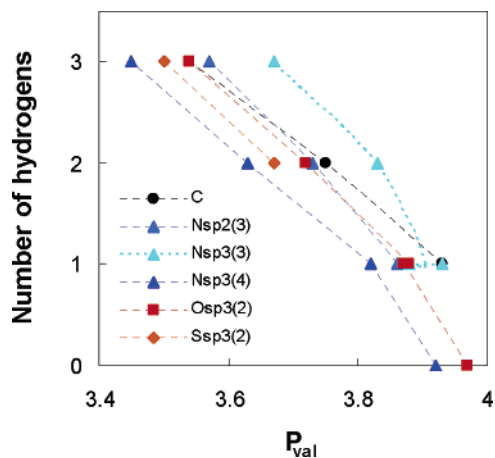
Figure 2. Dependence of the monopole population, P_{val}, of carbon on the number of hydrogen atoms attached in X−Csp^3−R_3 atom types, where X stands for C, $Nsp^2(3)$, $Nsp^3(3)$, $Nsp^3(4)$, $Osp^3(2)$, or $Ssp^3(2)$ atom types (see Table 3) and R stands for C or H. The numbers in parentheses represent the number of bonded atoms.

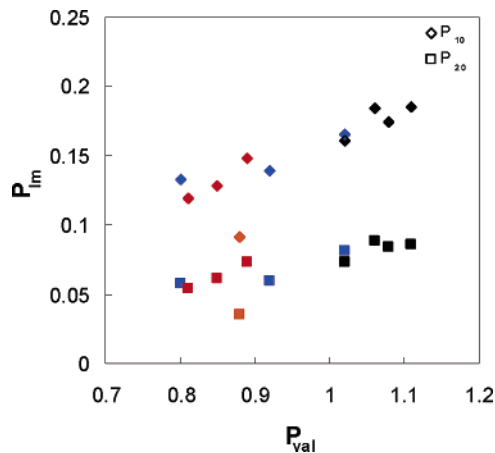**Figure 3.** Correlations between the monopole population, P_{val}, and P_{lm} for hydrogen atoms connected to carbon (black), nitrogen (blue), oxygen (red), or sulfur (orange).

For C atoms in X−Csp^3−R_3 bonded to the same number of hydrogens, where X stands for C, $Nsp^2(3)$, $Nsp^3(3)$, $Nsp^3(4)$, $Osp^3(2)$, or $Ssp^3(2)$ and R stands for C or H, the tetravalent nitrogens {$Nsp^3(4)$} induce the largest positive charge, followed by divalent sulfur {$Ssp^3(2)$}, oxygen {$Osp^3(2)$}, trivalent sp^2-hybridized nitrogen {$Nsp^2(3)$}, carbon, and trivalent sp^3-hybridized nitrogen {$Nsp^3(3)$}, as illustrated in Figure 2. Moreover, the more populated the d_{11+} dipole (oriented along the C−X bond), the more populated the d_{20} and d_{22+} quadrupoles become (Figure 6).[33] The populations of d_{11+}, d_{20}, and d_{22+} decrease in the following order of the substituents X: $Nsp^3(4)$, $Osp^3(2)$, $Nsp^2(3)$, $Ssp^3(2)$, $Nsp^3(3)$, and C.

With the exception of the S atom, the order of the substituent effect agrees well with the concept of electronegativity. According to Huheey's group electronegativity concept,[34] the tetravalent sp^3-hybridized nitrogen is the most electronegative group among X (3.65 for −NH_3^+ on the Huheey scale), followed by $Osp^3(2)$ (3.51 for −OH), $Nsp^3(3)$ (2.61 for −NH_2), $Ssp^3(2)$ (2.32 for −SH), and
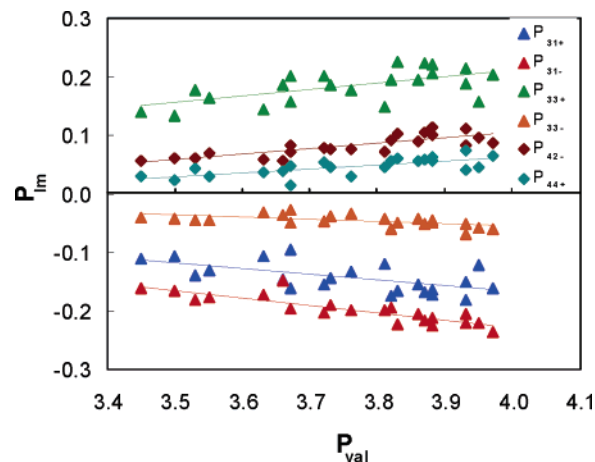

Figure 4. Correlation between the monopole population, P_{val}, and selected P_{lm} in X−Csp^3−R_3 carbon atom types, where X stands for C, $Nsp^2(3)$, $Nsp^3(3)$, $Nsp^3(4)$, $Osp^3(2)$, or $Ssp^3(2)$ (as defined in Table 3) and R stands for C or H. The local coordinate system for C402, C403, C405, C409, C415, C421, C433, C434, and C439 is rotated to coincide with that of other X−Csp^3−R_3 carbon atom types such as C411.

$Csp^3(4)$ (2.27 for −CH_3). It is not surprising, then, that $Nsp^3(4)$ generates the biggest positive charge on carbon (i.e., the smallest values of P_{val} among carbon types with the same number of hydrogens). The number of data points for the sulfur atom is less than those for the other atoms included in the databank, so a discussion of this exception appears premature.

The influences of hydrogen and Csp^3 carbon on the aspherical deformation density of the central carbon atom are very comparable, leading to mirror plane symmetry relating H and Csp^3 with each other in the chiral Csp^3 atoms C414, C420, C426, and C438, as evident from the group-averaged values of the P_{lm} coefficients (Table S2, Supporting Information).

The difference between sp^3 and sp^2 of the nearest neighboring carbon appears to have only a minor influence on the central carbon atom. For example, the C405 atom type was originally divided into two subgroups, with two Csp^3 atoms as first neighbors (*mm*2 local symmetry) and with Csp^2 and Csp^3 atoms as neighbors (*m* symmetry). However, after appropriate coordinate system rotation, the two subtypes became statistically indistinguishable.

There are currently 12 aromatic carbon atom types in the databank. Their deformation density parameters are given in Table S3 (Supporting Information), and deformation density maps are presented in Figure 7. The atoms are close to neutral and vary from +0.11(6) (C363) to −0.15(4) (C360) electron units. In the first 10 atom types, representing atoms participating only in a single ring, the multipoles d_{33+}, d_{20}, and d_{11+} are dominant. They describe respectively the deformation density along the three bonds, the π density above and below the ring, and the difference between the density in the exocyclic and cyclic bonds.

Nonaromatic sp^2 carbon atoms are represented by five different types in the current version of the databank. They are described in Table S4 (Supporting Information); the graphical representation of the deformation densities is shown
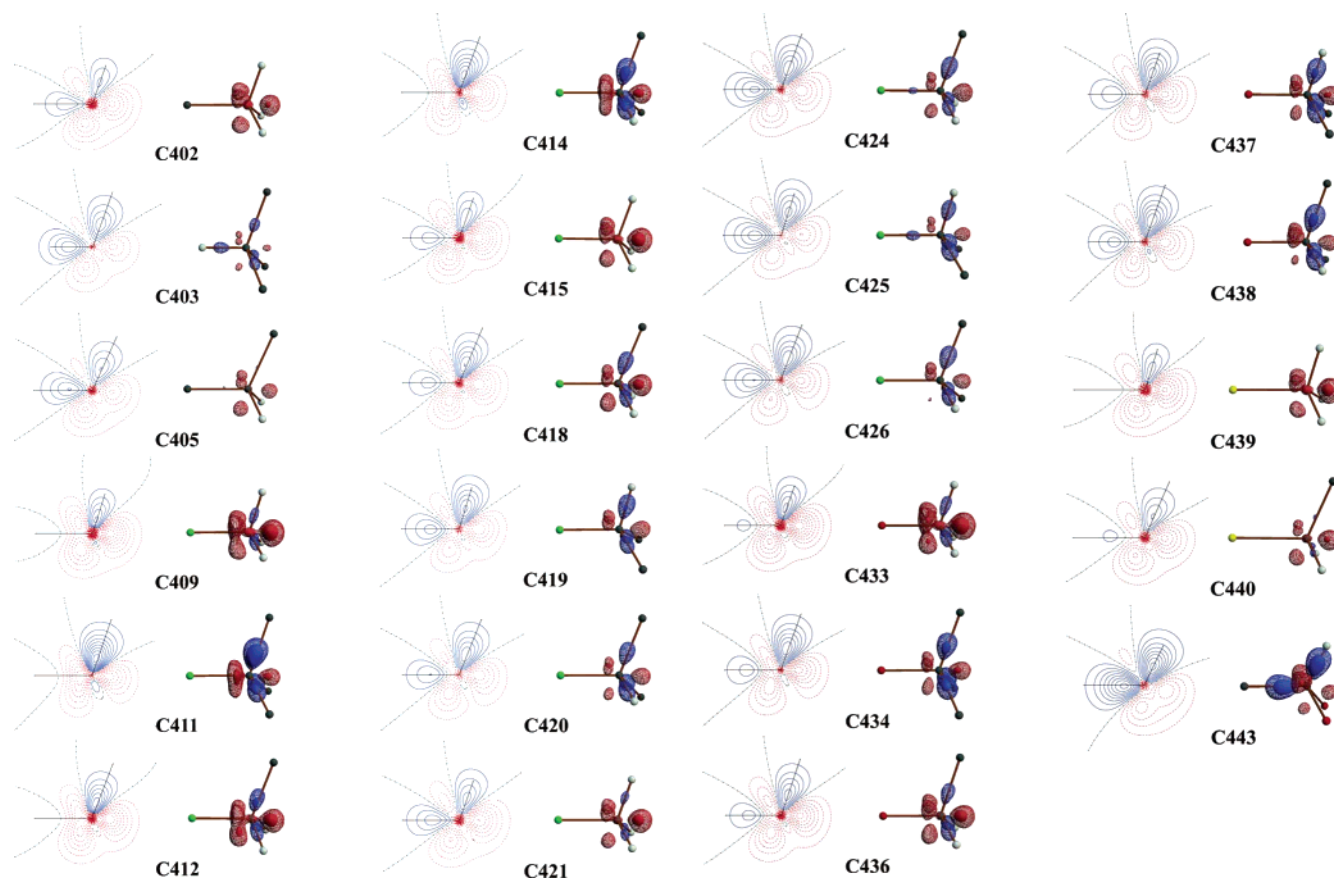
Figure 5. Deformation densities of the sp³ carbon atom types. Left: In the plane of two neighboring atoms, orientation as in Table 2, contour levels at 0.05 $e/\text{Å}^3$. Right: 3D representation, contour levels at ± 0.2 and ± 0.3 $e/\text{Å}^3$. Positive contours, blue; negative contours, red; zero contour, black. Color codes for atoms: C, dark gray; H, white; N, green; O, red; S, yellow.
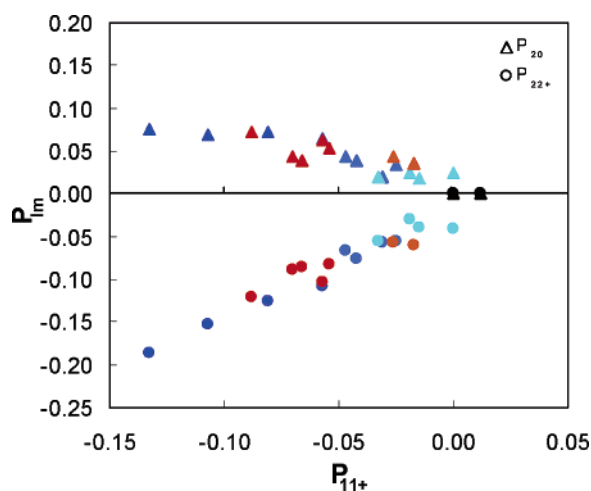

Figure 6. Correlation between P_{11+} and P_{20} (triangles) or P_{22+} (circles) of X−Csp³−R₃ carbon atom types, where X stands for C (black), Nsp²(3) (blue), Nsp³(3) (light blue), Nsp³(4) (dark blue), Osp³(2) (red), or Ssp³(2) (orange) atom types (as defined in Table 3) and R stands for C or H. The local coordinate system for C402, C4035, C405, C409, C415, C421, C433, C434, and C439 is rotated to coincide with that of other X−Csp³−R₃ carbon atom types such as C411.

in Figure 8. Four of the types describe carbonyl carbon in, respectively, carboxylate anions, carboxylic acids, esters, and amides. The last type represents carbon atoms in the

guanidine group. To distinguish between carboxylic acid and ester type C atoms, the atom type definition includes second neighbors to preserve uniqueness leading to atom types C302 and C303 (Table 2).

Nitrogen Atoms. On the basis of the deformation density, the 10 nitrogen atom types stored in the databank can be divided into two groups: those dominated by lone pair electrons (N301, N302, N303, and N201) and those which do not carry a lone electron pair (N402, N405, N307, N309, N310, and N311; Table S5 of the Supporting Information, Figure 9). The former have large positive deformation densities in the lone pair region and very small positive deformation densities in the bond directions. Additionally, trivalent sp³-hybridized nitrogens have relatively large electron deficiencies between bond regions and, in the case of aromatic divalent nitrogen (N201), below and above the aromatic ring plane.

Characteristic features for planar trivalent nitrogen atom types (N307, N309, and N310) are charge excess lobes above and below the plane, extending toward the neighboring Csp² atom.

Deformation densities for tetravalent nitrogen types are similar to those for tetravalent carbons connected only to C or H atoms. The only difference is that the nitrogen charge is more negative [−0.10(4) and − 0.19(8) e for N402 and N405 respectively], causing the electron-charge-accumulating lobes to be more pronounced.

Databank of Transferable Aspherical Atoms

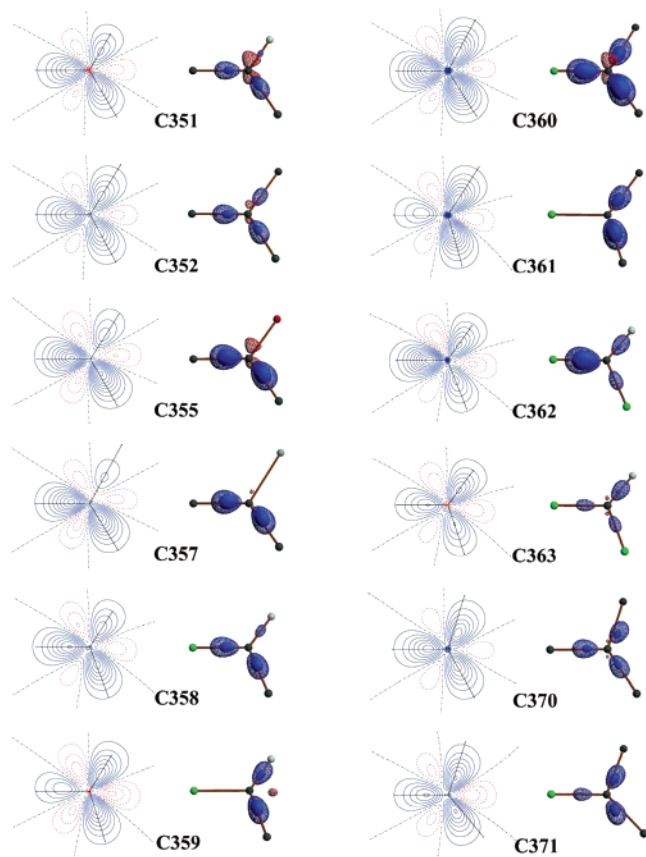*J. Chem. Theory Comput., Vol. 3, No. 1, 2007* **239**

Figure 7. Deformation densities of the aromatic sp² carbon atom types. Left: In the plane of two neighboring atoms, orientation as in Table 2, contour levels at 0.05 $e/Å^3$. Right: 3D representation, contour levels at ±0.2 and ±0.3 $e/Å^3$. Positive contours, blue; negative contours, red; zero contour, black. Color codes for atoms: C, dark gray; H, white; N, green; Cl, white.
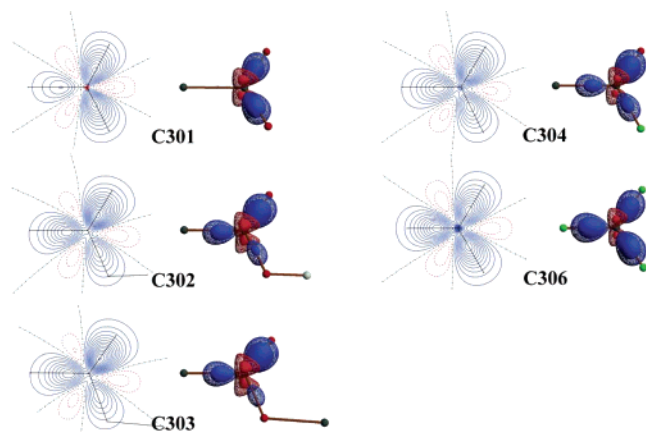

Figure 8. Deformation densities of the nonaromatic sp² carbon atom types. Left: In the plane of two neighboring atoms, orientation as in Table 2, contour levels at 0.05 $e/Å^3$. Right: 3D representation, contour levels at ±0.2 and ±0.3 $e/Å^3$. Positive contours, blue; negative contours, red; zero contour, black. Color codes for atoms: C, dark gray; H, white; N, green; O, red.

The aromatic trivalent nitrogen type, N311, differs from aromatic carbons by not having electron deficiency lobes above and below the plane (d$_{20}$ is not populated).
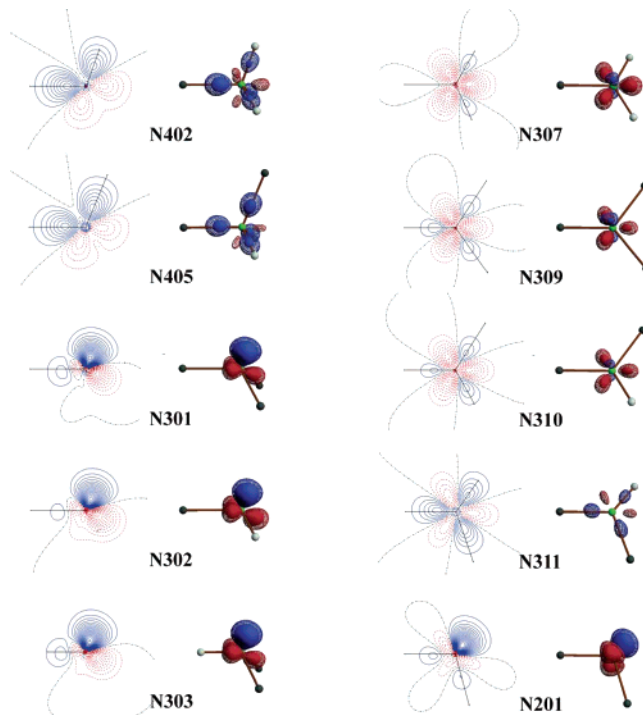

Figure 9. Deformation densities of the nitrogen atom types. Left: In the plane of two neighboring atoms, orientation as in Table 2, contour levels at 0.05 $e/Å^3$. Right: 3D representation, contour levels at ±0.2 and ±0.3 $e/Å^3$. Positive contours, blue; negative contours, red; zero contour, black. Color codes for atoms: C, dark gray; H, white; N, green.

Oxygen Atoms. The deformation densities for all nine oxygen atom types stored in the databank are dominated by the lone pair electron region (Table S6 of the Supporting Information, Figure 10). For monovalent oxygens, two lone pairs are clearly visible; in divalent ones, the lone pairs do not split in two disjunctive regions. The electron deficiency lobe is always oriented perpendicular to the lone pair plane and embraces the bond region. There is no excess density in the bonds originating from the oxygen atoms. All oxygen types bear negative charges ranging from −0.11(3) e for carbonyl oxygen in esters to −0.31(3) e for carboxylate oxygens.

Sulfur Atoms. There are only three sulfur types in the databank at the moment (Table S7 of the Supporting Information, Figure 11). The deformation of density around them is considerably less pronounced than that for oxygens. Still, the lone pair region dominates, but the concentration of electrons is much smaller, and additionally, there is an electron concentration in the bonding regions. Like oxygens, all sulfur types are negative, in the range from −0.26(2) to −0.40(3).

Chlorine Atoms. Currently, only one type of chlorine atom is stored in the databank, corresponding to Cl connected to aromatic rings. The deformation density has *mm*2 symmetry, see C*l*01 in Table S7 of the Supporting Information and Figure 11. Moreover, with a coordinate system with the *z* axis pointing to the carbon, only one multipole component, the quadrupole 22+ with population −0.013(2), violates cylindrical symmetry and flattens the density toward the
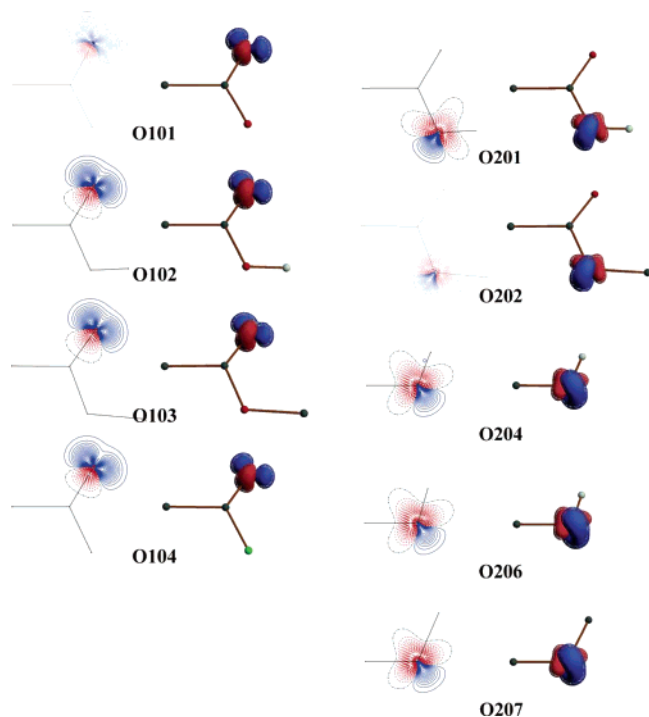
Figure 10. Deformation densities of the oxygen atom types. Left: In the plane of two neighboring atoms, orientation as in Table 2, contour levels at 0.05 $e/\text{Å}^3$. Right: 3D representation, contour levels at ± 0.2 and ± 0.3 $e/\text{Å}^3$. Positive contours, blue; negative contours, red; zero contour, black. Color codes for atoms: C, dark gray; H, white; N, green; O, red.
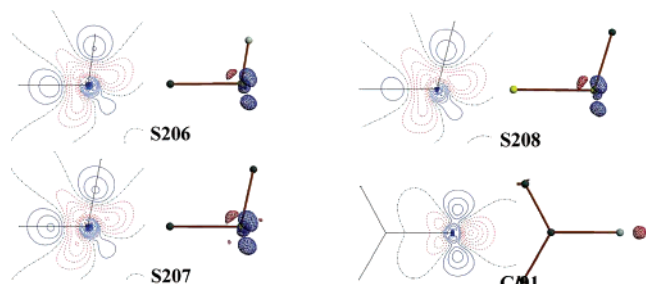

Figure 11. Deformation densities of the sulfur and chlorine atom types. Left: In the plane of two neighboring atoms, orientation as in Table 2, contour levels at 0.05 $e/\text{Å}^3$. Right: 3D representation, contour levels at ± 0.2 and ± 0.3 $e/\text{Å}^3$. Positive contours, blue; negative contours, red; zero contour, black. Color codes for atoms: C, dark gray; H, white; S, yellow; Cl, white.

aromatic plane. The population of the monopole is 7.24(3), indicating the Cl to be negative.

Hydrogen Atoms. The databank contains 11 hydrogen types. In the case of hydrogens connected to sp^3-hybridized carbons and oxygens, the definition of atom type has to be extended beyond the first neighbor type defined in Table 3. As a result, hydrogens connected to tetravalent carbons are divided into three groups on the basis of the total number of hydrogens attached to the carbon, whereas hydrogens connected to oxygens are split into those in alcohols (H109), carboxylic acids (H110), and phenols (H111). As discussed above, all deformation density parameters for hydrogens

correlate with the monopole populations (Figure 3, Table S8 of the Supporting Information).

Application

In a first macromolecular application, the databank is applied to the 0.7-Å-resolution crystal structure of the second PDZ domain (PDZ2) from the scaffolding protein syntenin[35] and subsequently to all PDZ−peptide complex structures available in the Protein Data Bank (PDB).[36,10,11]

PDZ domains are the most ubiquitous protein−protein interaction domains.[37,38] They are structurally conserved modules of about 90 amino acids in length and with a distinct fold of six β strands and two α helices.[39,40] Their function is achieved by binding of the C terminus of the target protein as an antiparallel β strand in a highly conserved hydrophobic groove between the $\beta 2$ sheet and $\alpha 2$ helix. The side chains of the P_0 and P_{-2} ligand residues point into the groove and account for the specificity (P_0 and P_{-n} denote the C-terminal residue of the bound ligand and the nth upstream amino acid residue, respectively). The PDZ binding motifs have been divided into three major classes: type I with the consensus sequence $-(S/T)X\Phi$, type II with $-\Phi X\Phi$, and type III with $-(D/E)X\Phi$, where X is any amino acid, Φ is a residue with hydrophobic side chains, and S/T is either serine or threonine and D/E is either aspartic or glutamic acid, respectively.[38]

However, it has been pointed out that this simple classification is unable to explain the increasing number of PDZ-mediated interactions that do not conform to this canonical type of recognition.[11] Some PDZ domains are able to bind with internal, rather than C-terminal, peptide sequences. Others recognize noncanonical sequences or more than one class of ligands. The syntenin PDZ2 domain, for example, demonstrates degenerate specificity. It was shown to bind template peptides modeling IL5Rα (SVF, class I), syndecan-4, neurexin, and ephrin B (FYA, YYV, and YKV, respectively, all class II) protein ligands (see Table 4 for nomenclature). Additionally, it has been established that interactions of syntenin with proteins are often mediated cooperatively by both of its PDZ domains.

Preparation of the Structures. The high-resolution structure of the syntenin's PDZ2 domain (PDB code: 1R6J, called 1R6J in the following), including hydrogen atom coordinates, was kindly provided by Prof. Z. S. Derewenda. For disordered parts of the protein, only major conformers were taken into account. The missing hydrogen of the thiol group of Cys239 was generated by the program Reduce[41] from the MolProbity Web service.[42] Two tautomers of His208 were tested with the proton at ND1 (original position) and at NE2. The latter, proposed by Reduce, was chosen for further calculations.

Structures of the PDZ2 domains in the complexes with peptides were taken from the protein data bank (PDB codes: 1OBX, 1OBY, 1OBZ, 1YBO, 1W9O, 1W9E, 1W9Q, and 1V1T). The Reduce program was used to add hydrogen atoms to the protein−peptide complexes by optimization of the hydrogen-bond network. All histidines were treated as singly protonated. The side chains of the Arg, Lys, Asp, and Glu residues were assigned as ionized by the program. Only major conformers of side chains were taken into account in

Databank of Transferable Aspherical Atoms

J. Chem. Theory Comput., Vol. 3, No. 1, 2007 **241**

Table 4. Electrostatic Energy [kJ/mol] of Interaction between the Syntenin PDZ2 Domain and Truncated Peptides[a]

			P_{-3}	P_{-2}	P_{-1}	P_0	total ES
			Class I				
			X(S/T)XΦ[b]				
			DSVF				
1OBX:	A	single domain	−172(−24)	−37(−32)	4(13)	−249(−235)	−454(−278)
1OBZ:	A	tandem	−168(−64)	−45(−32)	−1(8)	−266(−233)	−480(−321)
			Class II				
			XΦXΦ				
			EFYA				
1OBY:	A	single domain	−199(−48)	−106(−94)	27(11)	−193(−194)	−471(−325)
	B	single domain	−206(−50)	−113(−101)	5(8)	−197(−201)	−512(−344)
1YBO:	A	tandem	−169(−60)	−121(−107)	0(−10)	−170(−171)	−460(−348)
	B	tandem	−178(−20)	−87(−69)	−12(10)	−188(−185)	−464(−264)
			EYYV				
1W9O:	A	tandem	−163(−47)	−85(−56)	−25(10)	−138(−138)	−411(−231)
	B	tandem	−213(−65)	−125(−100)	5(11)	−143(−86)	−476(−240)
			EYKV				
1V1T:	A	tandem	−163(−49)	−100(−81)	20(12)	−176(−180)	−419(−298)
	B	tandem	−173(−50)	−134(−103)	−15(3)	−181(−191)	−502(−341)
			EFYF				
1W9E:	A	tandem	−180(−49)	−104(−94)	−5(1)	−247(−204)	−536(−346)
	B	tandem	−199(−60)	−169(−159)	−20(−5)	−237(−218)	−626(−442)
			EFAF				
1W9Q:	A	tandem	−193(−56)	−120(−107)	16(17)	−283(−246)	−580(−392)

[a] Columns labeled "P_{-3}" to "P_0": individual peptide residues interacting with the protein. Column labeled "total ES": total electrostatic energy of four-residue peptide interacting with the protein. Numbers in parenthesis correspond to interactions for main-chain atoms of peptides only.[b] X is any amino acid, Φ is a residue with hydrophobic side chains, S/T is either serine or threonine, D is aspartic acid, V is valine, F is phenylalanine, E is glutamic acid, Y is tyrosine, A is alanine, and K is lysine.

the calculations. All protein chains were truncated to 73-residue-long fragments starting from Thr196 and finishing with the Thr268 residue, because the N and C termini of the PDZ2 domain are not in the vicinity of the peptide-binding groove. The N and C termini of the protein chains were capped with the neutral acetyl and methylamino blocking groups, respectively. All peptides were truncated to C-terminus four-residue fragments, as the residues upstream from P_{-3} are solvent-exposed and not involved in the protein−peptide interaction. The C termini of peptides were deprotonated. Ideally, the NH_2-terminal group of P_{-3} residues should be capped with the neutral acetyl group. However, the lack of coordinates for the P_{-4} residue in some structures did not allow construction of the capping groups. Therefore, the neutral form of the peptide N termini was used to avoid introducing artificial positive charges, which would have biased the calculation.

In all cases, ions and water molecules were omitted in the calculations. The type and position of the ions vary among the structures considered, which prevents a direct analysis of their relevance for the electrostatic interactions. The variability among the ions in the structures results most likely from crystallization conditions and does not reflect the native structure of the syntenin PDZ2 domain. Only a few water molecules have similar positions from one structure to another, but none of them is present in all structures. The position of the water molecules suggests that their contribution to the binding of the ligand is minor.

Results

Charge Density Reconstruction. Less than 1 min on an AthlonMP 1800+ CPU is required by LSDB to assign atom types to more than 1000 protein atoms. The CPU-time scales approximately linearly with the number of atoms. The sum of the valence populations obtained with the databank differs by only 0.08% from the target value, with 2.6 and 2.4 electrons missing for the assembly of 1100−1200-atom 1R6J and truncated PDZ domains, respectively. This demonstrates the excellent self-consistency of the databank. To ensure the electroneutrality, the monopole populations of the pseudo-atoms were scaled a posteriori, using the Faerman and Price[43] scaling algorithm implemented in LSDB.[1] The deformation densities shown in Figure 12 illustrate that all bonding and electron pair features are very well reconstructed.

Electrostatic Potentials. Electrostatic potentials for un-complexed PDZ2 domains are computed from the charge density distributions according to the method of Su and Coppens[44] available in the XDPROP module of the XD package. An example of the electrostatic potential +0.05 and −0.05 e/Å isosurfaces and the electrostatic potential mapped on the 0.01 e isodensity surface (for 1W9E:B) is shown in Figure 13. All PDZ2 domains exhibit the strong polarization evident in Figure 13a. The binding groove starts on the border between negative and positive potentials and extends into the positive potential region. The electrostatic potentials of the PDZ2 domain and the peptide (both calculated for the nonbonded molecules) mapped on the molecular surface reveals as expected that the PDZ2−peptide complex is
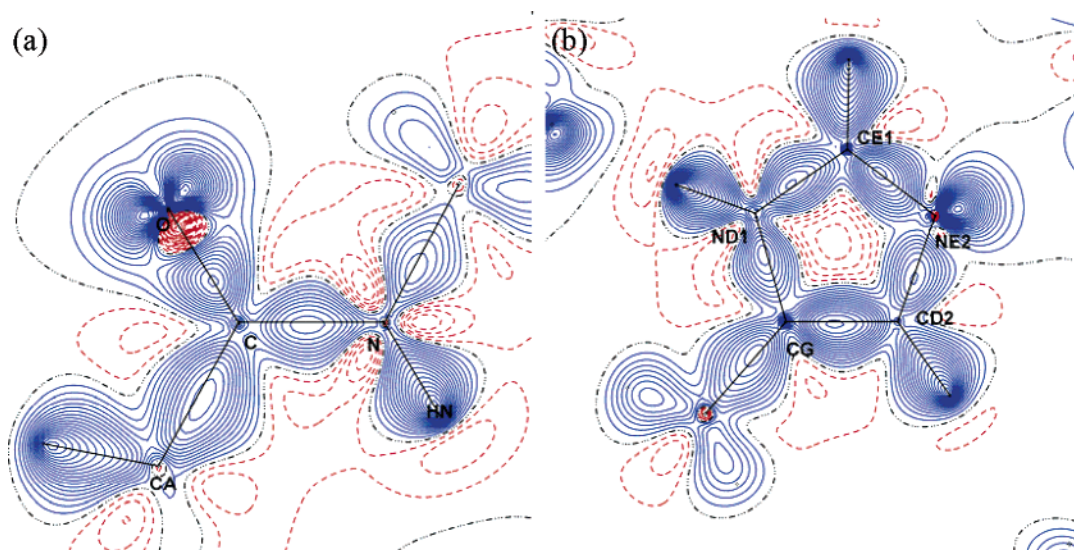
Figure 12. Deformation densities in the syntenin PDZ2 domain in the plane of (a) the peptide bond between Gly210 and Phe211 and (b) the imidazole ring of His208 with ND1 nitrogen protonated. Contour levels at 0.05 e/Å^3; positive contours, blue; negative contours, red; and zero contour, black.
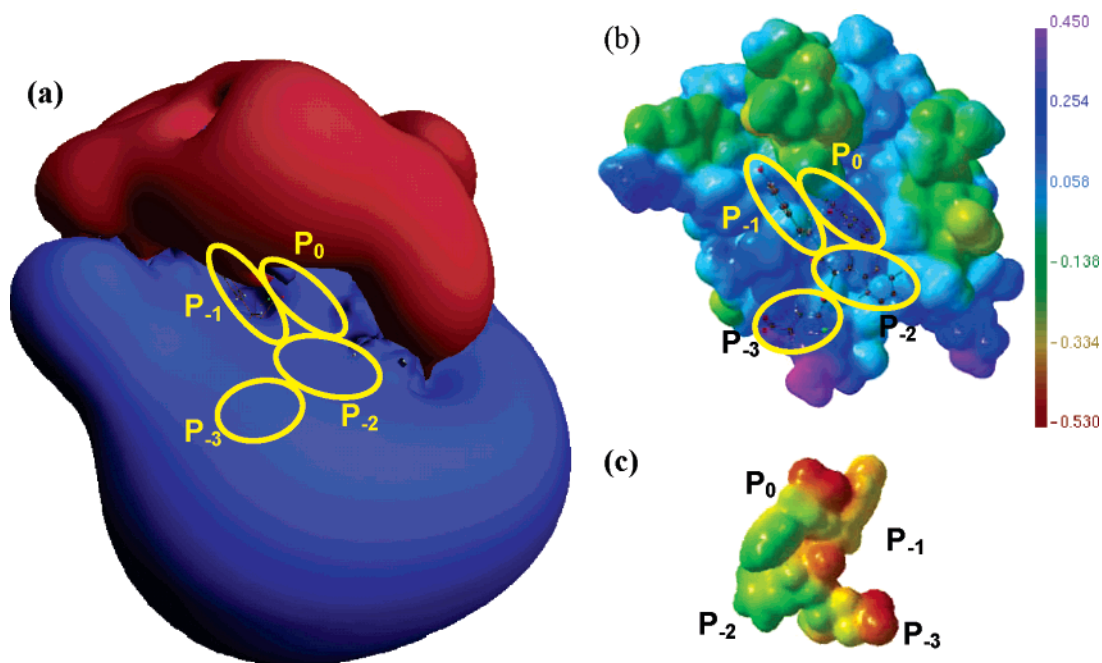

Figure 13. Electrostatic properties of the uncomplexed syntenin PDZ2 domain exemplified by the 1W9E:B structure. (a) Electrostatic potential of the PDZ2 domain calculated from the databank-generated charge density; isosurfaces: blue +0.05 e/Å, red −0.05 e/Å. Electrostatic potential [e/Å] of the PDZ2 domain (b) and peptide (c) mapped on the 0.01 e isodensity surface. Orientation of the protein as in Figure 16a; the peptide is rotated by 180°. The peptide molecule in the ball-and-stick model is presented only to illustrate the binding site and was not included in the electrostatic potential calculations.

formed between surfaces whose electrostatic potentials are complementary (Figure 13b and c).

Electrostatic Interactions. Electrostatic interactions are evaluated with the exact potential/multipole moment (EPMM) method of Volkov et al.[2,45] implemented in the XDPROP module of the XD package. The EPMM method combines numerical evaluation of the exact Coulomb integral for the short-range interactions between atoms within a certain limiting distance (here, taken as 4.5 Å) with the Buckingham-type multipole approximation for the long-range interatomic interactions between atoms for which no charge density

overlap occurs. The computational effort of the calculation of the interaction between distant atoms within the Buckingham approximation is negligible when compared to the numerical evaluation, the latter scales linearly with the number of the short-range interactions. It takes 78 s to process 100 interactions on an AthlonMP 1800+ CPU.

PDZ2 Dimer. In the 1R6J structure, the peptide binding groove is occupied by the C-terminal tails of the neighboring molecule, mimicking the recognition of the peptide ligand. The PDZ2 dimer serves as an example of protein−protein interactions. Application of the EPMM/databank method to

the 1R6J structure gives a value of −526 kJ/mol (−388 kJ/mol for the HisD tautomer) for the electrostatic energy of interaction between the PDZ2 domain and its symmetry-related ligand-mimicking counterpart. For comparison, the MMFF94 force field[46] implemented in Sybyl 7.1[47] gives a value of −742 kJ/mol for the electrostatic interaction (−345 kJ/mol for the HisD tautomer).

The 138 kJ/mol difference between the electrostatic interaction energies of the PDZ2 dimer with different tautomeric forms of His208 originates almost entirely from the difference in the His208−Glu235 interaction, which is +97 kJ/mol for the D tautomer and −33 kJ/mol for the E tautomer. The NE2 nitrogen from the imidazole ring of His208 is only 2.64 Å from the OE2 oxygen from the side-chain carboxylic group of Glu235 of the ligand-mimicking molecule. Such a close distance can only be explained by formation of a NE2−HE2···OE2 hydrogen bond, which is clearly supported by the above results. Therefore, the PDZ2 domain with His208 protonated at NE2 was chosen for further calculations.

The individual electrostatic contributions of each residue of the receptor molecule interacting with the whole ligand-mimicking molecule in the PDZ2 complex are presented in Figure 14a−c. Only a small number of receptor protein residues have large contributions to the binding energy, both attractive and repulsive. Among them are Lys203, Asp204, Thr206, His208, Val209, Gly210, Phe211, and Lys214 from the $\beta 2$ strand and the preceding conserved glycine-rich loop. Additionally, isolated strongly attractive and repulsive electrostatic interactions with the ligand-mimicking molecule are observed for the charged residues (Lys223 and Asp224, Lys250, Asp251, and Arg229). Three PDZ domain fingerprint residues, Val209, Gly210, and Phe211, interact electrostatically with the ligand protein, mainly through main-chain atoms (Figure 14b). The remaining interactions are with side chains of the interacting residues (Figure 14c).

Comparison of the individual residue−residue electrostatic interactions reveals that the larger absolute values are observed for interactions between charged residues (in the range from 162 to −126 kJ/mol, on average ±26 kJ/mol). However, interactions between these almost cancel each other, resulting in only a −86 kJ/mol contribution to the total electrostatic interaction energy.

PDZ2−Peptide Complexes. PDZ2 domains complexed with short peptides are an excellent example of protein−ligand interactions. We analyzed all eight PDB-available structures of syntenin domains interacting with short peptides. Six of these structures (1OBY, 1YBO, 1W9O, 1W9E, 1W9Q, and 1V1T) represent tandems of PDZ1 and PDZ2 domains. The remaining two (1OBX and 1OBY) consist solely of the PDZ2 domain. Only the PDZ2 domains were analyzed in this study; the PDZ1 domains were removed from the tandem structures. The structures of complexes superimposed on each other are shown in Figure 15a. The class II peptides (EFYA, EYYV, EYKV, EFYF, and EFAF) interact with the PDZ2 domain as the canonical model predicts, see Figure 15b. The DSVF (class I) peptides are not fully inserted in the binding groove.
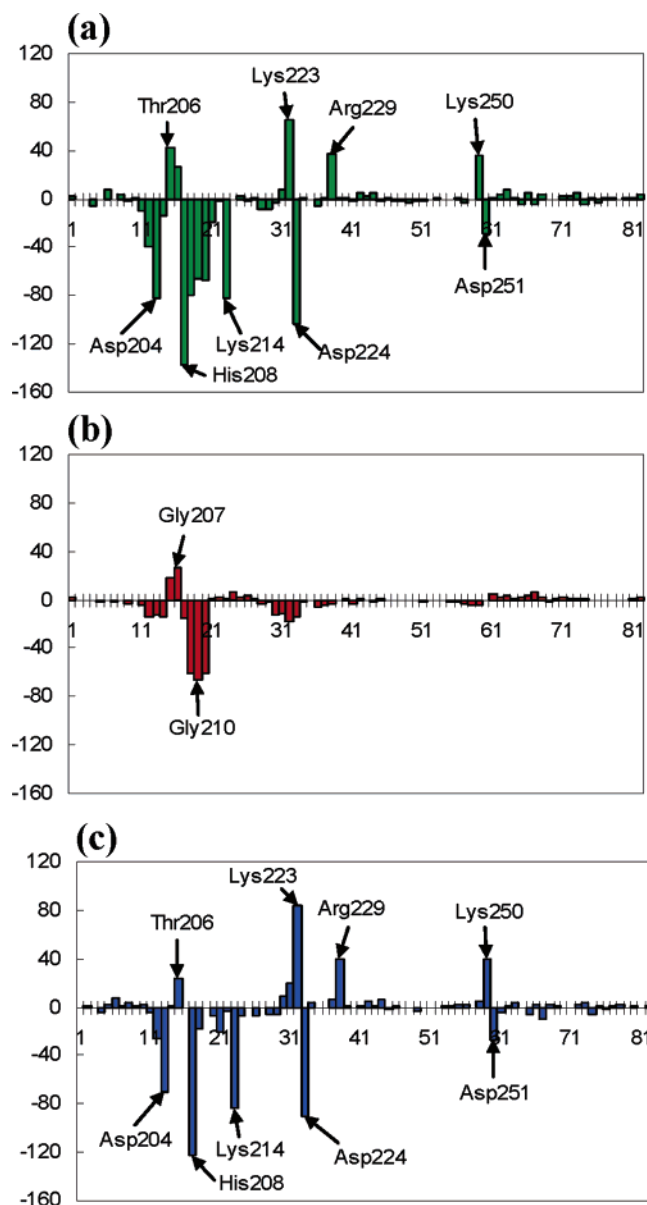

Figure 14. Electrostatic contributions (kJ/mol) of individual amino acid residues of the PDZ2 domain in the 1R6J structure interacting with the symmetry-related domain, mimicking the native ligand/PDZ2 interaction (a) all interactions, (b) main-chain/PDZ2 interactions, and (c) side-chain/PDZ2 interactions.

For the PDZ2 domains, the electrostatic energy of interaction between the protein and the truncated peptide is found to range from −411 to −626 kJ/mol (Table 4).

Several of the peptides contain Phe as P_0. It is the strongest interacting residue out of all of the residues examined, with the electrostatic interaction energy averaged over all occurrences equal to 256(18) kJ/mol. The main-chain atoms of P_0 (Phe) are the major contributors to the interaction energy, whereas 11% (29 kJ/mol) is provided by side-chain atoms. When P_0 is Ala or Val, the interaction energies are lower by about 80 kJ/mol. With the exception of the 1W9O:B structure, no significant contributions of the side-chain atoms of Ala or Val are observed. Interestingly, the extra interaction energy of Phe is almost equally distributed between the side-chain and main-chain atoms, suggesting that the side-chain phenyl ring enforces a better fit of the main chain. The results
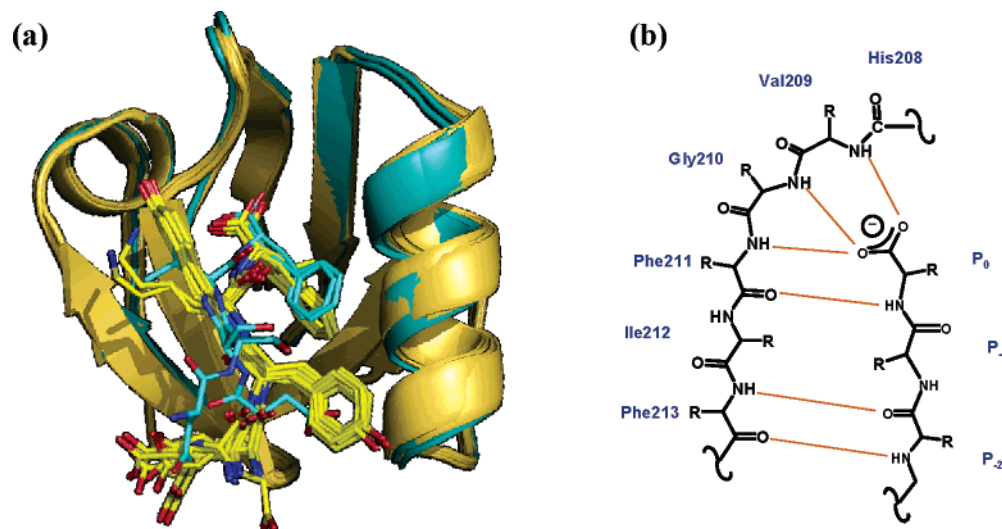
Figure 15. Structural details of the syntenin PDZ2 domain complexes. (a) Superposition of PDZ2 domains interacting with class I (blue; 1OBY and 1OBZ) or class II peptides (yellow; 1OBY, 1YBO, 1W9O, 1VLT, 1W9E, 1W9Q). (d) The canonical hydrogen bonds formed between the protein (left) and the peptide (right) in the syntenin PDZ2/peptide complex.

indicate stronger binding for phenylalanine than for alanine, as noted by Grembecka et al.,[12] but are not in agreement with the experimental dissociation constants for the interaction of the TNEFYA (with alanine at P_0) and TNEFYF (with phenylalanine at P_0) hexapeptides with PDZ2 in tandem, which are 0.16(1) and 0.79 (3) mM, respectively.[10] Grembecka et al. attributed this discrepancy to an energetically unfavorable rearrangement of the PDZ2 domain to accommodate the large phenyl group, which is reflected in the dissociation constants but not in the interaction energies.

It is interesting that residues P_{-1} do not contribute to electrostatic binding energy [av 0(15) kJ/mol]. Studies have shown that this residue is not conserved among the C termini of ligand proteins that interact with PDZ domains and is considered unspecific. However, this does not explain the fact that the replacement of Tyr (Y) at the P_{-1} position by Lys (K) is observed to decrease the binding. Dissociation constants for interaction of the TNEYYV and TNEYKV hexapeptides with PDZ2 in tandem have been found to be equal to 0.10(1) and 1.15(9) mM, respectively.[10]

The P_{-2} residue of the ligand is the major cause of PDZ specificity. The Ser P_{-2} residue common to class I peptides interacts much weaker with the protein [av −41(6) kJ/mol] than the aromatic P_{-2} residues [Phe and Tyr, av −115(24) kJ/mol] of class II peptides. There is no clear difference in interaction strength between Phe and Tyr at the P_{-2} position.

It is noteworthy that the P_{-3} residues [negatively charged Asp (D) or Glu (E)] significantly strengthen the electrostatic interaction [av −183(17) kJ/mol], even though they do not have short contacts with the protein and are solvent-exposed. The interactions are mainly maintained by side-chain atoms, while the main chains contribute only in 27% (−49 kJ/mol in average).

Although the interaction energies of the individual residues vary, there is no substantial difference between the electrostatic energies of class I and class II peptides interacting with the syntenin PDZ2 domain. This result is in agreement with fluorometric dissociation constants for the interactions of

IL5Rα- and syndecan-derived dansylated peptides with PDZ2 in tandem, which are very similar [1.9(3) and 2.3(5) μM for EDSVF and NEFYA, respectively].[48] Although the interaction strength of the P_{-2} residue is indeed weaker for class I peptides, this loss of interaction energy is compensated by stronger binding of the P_0 and P_{-3} residues.

There is no difference between the interaction energies of the short peptides with PDZ2 domains from single-domain protein fragments and those from tandem protein fragments from which PDZ1 has been removed (see single domains vs tandems, Table 4).

A more detailed analysis of the individual atom−atom interactions gives information on the types of electrostatic interactions which contribute most to peptide binding by the syntenin PDZ2 domain. Major contributions come from hydrogen bonds between main-chain atoms of the Val209−Phe211 fragment and main-chain atoms of P_0 peptide residues (including the COO⁻-terminal group; Table 5). The two strongest electrostatic interactions are Val209 H···OXT P_0 and Gly210 H···O P_0 hydrogen bonds, which average −53(10) and −49(6) kJ/mol, respectively. Hydrogen bonds between Phe213 and P_{-2} also contribute to the binding. Equally important are charge−charge interactions between side chains of Lys214 and the P_{-3} residue [−80(18) kJ/mol on average], though they are less numerous than the hydrogen bonds.

Conclusions

A comprehensive version of the theoretical databank of transferable aspherical pseudoatoms has been developed and applied for the first time to protein−ligand interaction energies. The databank consists of all atom types encountered in natural amino acid residues and a number of other biologically relevant molecules. Each atom type results from averaging over a family of chemically unique pseudoatoms, taking into account both first and second neighbors. A new atom type is spawned when one of the parameters for a subgroup of the family deviates more than one standard

Databank of Transferable Aspherical Atoms

J. Chem. Theory Comput., Vol. 3, No. 1, 2007 **245**

Table 5. Most Important Atom−Atom Contributions [kJ/mol] to the Total Electrostatic Interaction Energies for the Syntenin PDZ2 Domain Complexed with Truncated Peptides

interacting atoms	1OBX: AB	1OBZ: AP[a]	1OBY: AB	1OBY: BQ	1YBO: AC	1YBO: BD[a]	1W9O: AT	1W9O: BS	1VLT: AT	1VTL: BS	1W9E: AT	1W9E: BS	1W9Q: BS
					In H Bonds[b]								
Val209 H···OXT P_0	−60	−57	−58	−52	−41	−35	−40	−45	−63	−52	−55	−56	−72
Val209 H···O P_0	−16	−18	−17	−17	−16	−17	−16	−19	−15	−15	−15	−17	−18
Gly210 H···OXT P_0	−16	−16	−16	−17	−16	−15	−16	−16	−19	−18	−17	−17	−19
Gly210 H···O P_0	−63	−51	−51	−52	−40	−52	−42	−46	−45	−46	−49	−48	−53
Phe211 H···O P_0	−37	−38	−32	−32	−33	−29	−28	−27	−24	−28	−33	−35	−30
Phe211 O···HP_0	−31	−37	−38	−38	−42	−70	−31	−18	−36	−37	−46	−38	−69
Phe213 H···O P_{-2}	−8	−8	−30	−33	−30	−33	−22	−29	−27	−32	−36	−79	−38
Phe213 O···H P_{-2}	−4	−2	−32	−30	−34	−16	−19	−36	−25	−34	−25	−34	−37
					In Ionic Interactions								
Lys214 NH_3^+···⁻OOC P_{-3}	−109	−86	−105	−102	−54	−87	−62	−86	−62	−72	−65	−70	−77

[a] OXT and O oxygens of P_0 residue are interchanged compared to labeling scheme in original PDB file. [b] Only interactions between hydrogen and acceptor atoms are taken into account because interactions with other atoms are very weak, 0(1) kJ/mol on average.

deviation of the sample from the average of the remaining atoms. The algorithm for atom-type definition ensures that close transferability is obeyed.

The pseudoatoms as derived with the Hansen−Coppens formalism from theoretical wave functions show several regularities. The monopole-derived net atomic charge correlates strongly with the expansion−contraction parameter κ and somewhat less with κ', as found before.[26,27] In general, the atomic charges reflect the group-electronegativity concept as defined by Huheey. However, hydrogen pseudoatoms, for which the division of the charge in the bonding regions is particularly difficult, behave differently. Their populations significantly affect charges on the neighboring atoms. The charge of the atoms linked to hydrogens becomes more and more positive as the number of bonded hydrogen atoms increases, while the H-atom charge shifts toward negative values. The deformation density of atoms possessing lone pairs is dominated by the lone pair electrons, while the bonding regions contribute very little.

Application of the databank to the syntenin PDZ2 domain complexes allows a detailed analysis of the electrostatic interaction energies. The results stress the importance of the P_0 and P_{-2} residues of the peptide in establishing the interaction, whereas the P_{-1} residue is shown to play a much smaller role. Unexpectedly, the charged P_{-3} residue contributes significantly also.

Class I and class II peptides are bound with the same strength by the syntenin PDZ2 domain. Although the electrostatic interaction energy of the P_{-2} residue is smaller for class I peptides, this loss is compensated by stronger binding of the P_0 and P_{-3} residues.

In agreement with previous conclusions,[10] our results confirm that there is no difference between the interaction energies of short peptides with PDZ2 domains from single-domain protein fragments and those from PDZ1−PDZ2 tandem protein fragments. The experimentally observed cooperation of PDZ1 and PDZ2 domains upon binding the target protein has to be attributed to interactions in other regions of the complexes.

Although the electrostatic energy is the dominant component in the hydrogen-bonded systems studied, other contributions such as dispersion, exchange-repulsion, and induction terms must be considered. We are developing a set of symmetry-adapted perturbation-theory-based pairwise atom−atom potentials which will be applied in a subsequent study. Entropic effects, such as solvation and configurational changes in solution, must also be taken into account in a comprehensive treatment.

Supporting Information Available: Table S1 with the CSD codes, chemical diagrams, names, formulas, and references for all compounds used for the construction of the databank and Tables S2−S8 with deformation density parameters for atom types present in the current version of the databank. This material is available free of charge via the Internet at http://pubs.acs.org. The LSDB program and the full databank are available at http://harker. chem.buffalo.edu.

References

(1) Volkov, A.; Li, X.; Koritsánszky, T.; Coppens, P. *Ab Initio* Quality Electrostatic Atomic and Molecular Properties Including Intermolecular Energies from a Transferable Theoretical Pseudoatom Databank. *J. Phys. Chem.* **2004**, *108*, 4283−4300.

(2) Volkov, A.; Koritsánszky, T.; Coppens, P. Combination of the Exact Potential and Multipole Methods (EP/MM) for Evaluation of Intermolecular Electrostatic Interaction Energies with Pseudoatom Representation of Molecular Electron Densities. *Chem. Phys. Lett.* **2004**, *391*, 170−175.

(3) Pichon-Pesme, V.; Lecomte, C.; Lachekar, H. On Building a Data Bank of Transferable Experimental Electron Density Parameters Applicable to Polypeptides. *J. Phys. Chem.* **1995**, *99*, 6242−6250.

(4) Pichon-Pesme, V.; Jelsch, C.; Guillot, B.; Lecomte, C. A Comparison between Experimental and Theoretical Aspherical-Atom Scattering Factors for Charge-Density Refinement of Large Molecules. *Acta Crystallogr., Sect. A* **2004**, *60*, 204−208.

(5) Volkov, A.; Li, X.; Koritsánszky, T.; Coppens, P. Response to the Paper A Comparison between Experimental and Theoretical Aspherical-Atom Scattering Factors for Charge-Density Refinement of Large Molecules, by Pichon-Pesme, Jelsch, Guillot & Lecomte 2004. *Acta Crystallogr., Sect. A* **2004**, *60*, 638−639.

(6) Li, X.; Volkov, A. V.; Szalewicz, K.; Coppens, P. Interaction Energies between Glycopeptide Antibiotics and Substrates in Complexes Determined by X-ray Crystallography: Application of a Theoretical Databank of Aspherical Atoms and a Symmetry-Adapted Perturbation Theory-Based Set of Interatomic Potentials. *Acta Crystallogr., Sect. D* **2006**, *62*, 639−647.

(7) Dittrich, B.; Koritsánszky, T.; Luger, P. A Simple Approach to Nonspherical Electron Densities by Using Invarioms. *Angew. Chem., Int. Ed.* **2004**, *43*, 2718−2721.

(8) Dittrich, B.; Hübschle, C. B.; Messerschmidt, M.; Kalinowski, R.; Girnt, D.; Luger, P. The Invariom Model and Its Application: Refinement of D,L-Serine at Different Temperatures and Resolution. *Acta Crystallogr., Sect. A* **2005**, *61*, 314−320.

(9) Dittrich, B.; Strumpel, M.; Schäfer, M.; Spackman, M. A.; Koritsánszky, T. Invarioms for Improved Absolute Structure Determination of Light-Atom Crystal Structures. *Acta Crystallogr., Sect. A* **2006**, *62*, 217−223.

(10) Walker, P. D.; Mezey, P. G. Molecular Electron Density Lego Approach to Molecule Building. *J. Am. Chem. Soc.* **1993**, *115*, 12423−12430.

(11) Szekeres, Zs.; Exner, T. E.; Mezey, P. G. Fuzzy Fragment Selection Strategies, Basis Set Dependence, and HF − DFT Comparisons in the Applications of the ADMA Method of Macromolecular Quantum Chemistry. *Int. J. Quantum Chem.* **2005**, *104*, 847−860.

(12) Grembecka, J.; Cierpicki, T.; Devedjiev, Y.; Derewenda, U.; Kang, B. S.; Bushweller, J. H.; Derewenda, Z. S. The Binding of the PDZ Tandem of Syntenin to Target Proteins. *Biochemistry* **2006**, *45*, 3674−3683.

(13) Kang, B. S.; Cooper, D. R.; Devedjiev, Y.; Derewenda, U.; Derewenda, Z. S. Molecular Roots of Degenerate Specificity in Syntenin's PDZ2 Domain: Reassessment of the PDZ Recognition Paradigm. *Structure* **2003**, *11*, 845−853.

(14) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.

(15) Hariharan, P. C.; Pople, J. A. The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies. *Theor. Chim. Acta.* **1973**, *28*, 213−222.

(16) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(17) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.

(18) Allen, F. H. The Cambridge Structural Database: A Quarter of a Million Crystal Structures and Rising. *Acta Crystallogr., Sect. B* **2002**, *58*, 380−388.

(19) C−H: 1.099 Å for primary and 1.092 Å for secondary carbons, 1.059 Å for methyl groups connected to carbon and 1.066 Å for methyl groups connected to a non-carbon atom, 1.083 Å for aromatic carbons and 1.077 Å for nonaromatic sp^2 carbons. N−H: 1.033 Å for tetravalent and 1.009 Å for trivalent nitrogens. O−H: 1.015 Å for carboxylic acids and 0.967 for other hydroxyl groups. S−H 1.338 Å for thiols. Wilson, A. J. C.; Prince, E. *International Tables for Crystallography*; Kluwer Academic Publishers: Dordrecht, The Netherlands1992; Vol. C.

(20) Koritsánszky, T. S.; Volkov, A. V.; Coppens, P. Aspherical-Atom Scattering Factors from Molecular Wave Functions. 1. Transferability and Conformation Dependence of Atomic Electron Densities of Peptides within the Multipole Formalism. *Acta Crystallogr., Sect. A* **2002**, *58*, 464−472.

(21) Hansen, N. K.; Coppens, P. Testing Aspherical Atom Refinements on Small-Molecule Data Sets. *Acta Crystallogr., Sect. A* **1978**, *34*, 909−921.

(22) Koritsánszky, T. S.; Howard, S.; Macchi, P.; Gatti, C.; Farrugia, L. J.; Mallinson, P. R.; Volkov, A.; Su, Z.; Richter, T.; Hansen, N. K. *XD, a Computer Program Package for Multipole Refinement & Analysis of Electron Densities from Diffraction Data*, version 4.12, October 2004; Free University of Berlin: Berlin, Germany; University of Wales: Cardiff, U. K.; Universitá di Milano: Milano, Italy; CNR−ISTM: Milano, Italy; University of Glasgow: Glasgow, U. K.; State University of New York: Buffalo, NY; University of Nancy: Nancy, France, 2004. Home page: http://xd.chem. buffalo.edu/ (accessed May 2006).

(23) Clementi, E.; Roetti, C. Roothan-Hartree−Fock Atomic Wave Functions. *At. Data Nucl. Data Tables* **1974**, *14*, 177−478.

(24) Clementi, E.; Raimondi, D. L. Atomic Screening Constants from SCF Functions. *J. Chem. Phys.* **1963**, *38*, 2686−2692.

(25) Dominiak, P. M.; Coppens, P. Finding Optimal Radial-Function Parameters for S Atoms in the Hansen−Coppens Multipole Model through Refinement of Theoretical Densities. *Acta Crystallogr., Sect. A* **2006**, *62*, 224−227.

(26) In the XD procedure, two vectors are defined by neighboring atoms. The first from the original atom to atom 1 defines vector 1, which is also axis 1; vector 2 is defined by atoms 2 and 3, atom 2 being usually identical to the original atom. Axis 2 is defined as v1 × v2, and axis 3 is the cross product of axes 1 and 2. For full details, see the XD manual.

(27) The first number of the atom-type designation represents the number of bonded atoms; the following numbers are a sequential number.

(28) Coppens, P.; Guru Row, T. N.; Leung, P.; Stevens, E. D.; Becker, P. J.; Yang, Y. W. Net Atomic Charges and Molecular Dipole Moments from Spherical-Atom X-ray Refinements, and the Relation between Atomic Charge and Shape. *Acta Crystallogr., Sect. A* **1979**, *35*, 63−72.

(29) Volkov, A.; Abramov, Y. A.; Coppens, P. Density-Optimized Radial Exponents for X-ray Charge-Density Refinement from *ab initio* Crystal Calculation. *Acta Crystallogr., Sect. A* **2001**, *57*, 272−282.

(30) Fitted linear equation for H: $\kappa' = 1.26(2) + 1.1(2)q$, $R = 0.90$; for C: $\kappa' = 0.888(3) + 0.17(1)q$, $R = 0.89$.

(31) Fitted linear equation for P_{10}/P_{val}: $P_{10} = -0.06(4) + 0.22(5)P_{val}$, $R = 0.85$; for P_{20}/P_{val}: $P_{20} = -0.04(3) + 0.12(3)P_{val}$, $R = 0.81$.

(32) Fitted linear equation for P_{31+}/P_{val}: $P_{31+} = 0.2(1) - 0.10(3)P_{val}$, $R = -0.60$; for P_{31-}/P_{val}: $P_{31-} = 0.27(6) - 0.12(2)P_{val}$, $R = -0.86$; for P_{33+}/P_{val}: $P_{33+} = -0.2(1) + 0.11(3)P_{val}$, $R = 0.62$; for P_{33-}/P_{val}: $P_{33-} = 0.10(4) - 0.04(1)P_{val}$, $R = -0.60$; for P_{42-}/P_{val}: $P_{42-} = -0.27(6) + 0.09(2)P_{val}$, $R = 0.81$; for P_{44+}/P_{val}: $P_{44+} = -0.21(6) + 0.07(1)P_{val}$, $R = 0.72$.

(33) Fitted linear equation for P_{20}/P_{11+}: $P_{20} = 0.014(4) - 0.57(7)P_{11+}$, $R = -0.88$; for P_{22+}/P_{11+}: $P_{22+} = -0.018(4) + 1.125(7)P_{11+}$, $R = 0.97$.

(34) Huheey, J. E. The Electronegativity of Groups. *J. Phys. Chem.* **1965**, *69*, 3284−3291.

(35) Kang, B. S.; Devedjiev, Y.; Derewenda, U.; Derewenda, Z. S. The PDZ2 Domain of Syntenin at Ultrahigh Resolution: Bridging the Gap between Macromolecular and Small Molecule Crystallography. *J. Mol. Biol.* **2004**, *338*, 483−493.

(36) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242. http://www.pdb.org/ (accessed Apr 10, 2006).

(37) Hung, A. Y.; Sheng, M. PDZ Domains: Structural Modules for Protein Complex Assembly. *J. Biol. Chem.* **2002**, *277*, 5699−5702.

(38) Sheng, M.; Sala, C. PDZ Domains and the Organization of Supramolecular Complexes. *Annu. Rev. Neurosci.* **2001**, *24*, 1−29.

(39) Doyle, D. A.; Lee, A.; Lewis, J.; Kim, E.; Sheng, M.; MacKinnon, R. Crystal Structures of a Complexed and Peptide-Free Membrane Protein-Binding Domain: Molecular Basis of Peptide Recognition by PDZ. *Cell* **1996**, *85*, 1067−1076.

(40) Morais Cabral, J. H.; Petosa, C.; Sutcliffe, M. J.; Raza, S.; Byron, O.; Poy, F.; Marfatia, S. M.; Chishti, A. H.; Liddington, R. C. Crystal Structure of a PDZ Domain. *Nature* **1996**, *382*, 649−652.

(41) Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. Asparagine and Glutamine: Using Hydrogen Atom Contacts in the Choice of Side-Chain Amide Orientation. *J. Mol. Biol.* **1999**, *285*, 1735−1747.

(42) Lovell, S. C.; Davis, I. W.; Arendall, W. B., III; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. Structure Validation by Cα Geometry: Φ, Ψ, and Cβ Deviation. *Proteins: Struct., Funct., Genet.* **2003**, *50*, 437−450.

(43) Faerman, C. H.; Price, S. L. A Transferable Distributed Multipole Model for the Electrostatic Interactions of Peptides and Amides. *J. Am. Chem. Soc.* **1990**, *112*, 4915−4926.

(44) Su, Z. W.; Coppens, P. On the Mapping of Electrostatic Properties from the Multipole Description of the Charge Density. *Acta Crystallogr., Sect. A* **1992**, *48*, 188−197.

(45) Volkov, A.; King, H. F.; Coppens, P. Dependence of the Intermolecular Electrostatic Interaction Energy on the Level of Theory and the Basis Set. *J. Chem. Theory Comput.* **2006**, *2*, 81−89.

(46) Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.

(47) *SYBYL 7.1.*; Tripos Inc.: St. Louis, Missouri.

(48) Kang, B. S.; Cooper, D. R.; Jelen, F.; Devedjiev, Y.; Derewenda, U.; Dauter, Z.; Otlewski, J.; Derewenda, Z. S. PDZ Tandem of Human Syntenin: Crystal Structure and Functional Properties. *Structure* **2003**, *11*, 459−468.

(49) Hübschle, C. B. *MolIso−A Program for Color Mapped Iso-surfaces*, 2005; http://userpage.chemie.fu-berlin.de/ ∼chuebsch/moliso/ (accessed Apr 7, 2006).

(50) DeLano, W. L. *The PyMOL Molecular Graphics System*; DeLano Scientific: San Carlos, CA, 2002. http:/www. pymol.org. (accessed Jun 6, 2005).