JCTC Journal of Chemical Theory and Computation

Dependence of the Intermolecular Electrostatic Interaction Energy on the Level of Theory and the Basis Set[†]

Anatoliy Volkov,* Harry F. King, and Philip Coppens

Department of Chemistry, State University of New York at Buffalo, Buffalo, New York 14260-3000

Received September 2, 2005

Abstract: As electrostatic forces play a prominent role in the process of folding and binding of biological macromolecules, an examination of the method dependence of the electrostatic interaction energy is of great importance. An extensive analysis of the basis set and method dependence of electrostatic interaction energies (E_{es}) in molecular systems using six test dimers of α-glycine is presented. A number of Hartree−Fock, Kohn−Sham, Møller−Plesset, configuration interaction (CI), quadratic CI, and coupled cluster calculations were performed using several double-, triple-, and quadruple-ζ-quality Gaussian- and Slater-type (Kohn−Sham calculations only) basis sets. The main factor affecting E_{es} was found to be the inclusion of diffuse functions in the basis set expansions. Møller−Plesset (even at second order), quadratic CI, and coupled cluster calculations produce the most consistent results. Hartree−Fock and CI methods usually overestimate the E_{es}, while the Kohn−Sham approach tends to underestimate the magnitude of the electrostatic interaction. The combination of the transferable-pseudoatom databank and the exact potential and multipole moment method reproduces Kohn−Sham B3LYP/6-31G** results on which it is based, confirming the excellent transferability of the pseudoatom densities within the systems studied. However, because Kohn−Sham calculations with double-ζ-quality basis sets show considerable deviations from advanced correlated methods, further development of the databank using electron densities from such methods is highly desirable.

Introduction

Electrostatic forces play an important role in the process of protein folding and binding,[1] as the electrostatic interaction energy E_{es} is a major component of the total interaction energy E_{int} of polar molecules. This has long been recognized within the boundaries of the perturbation theory of intermolecular forces[2] in which the electrostatic interaction energy is the leading term in the perturbation expansion of E_{int}:[3]

$$E_{int} = E_{es} + E_{ind} + E_{disp} + E_{ex-rep} \qquad (1)$$

where E_{ind}, E_{disp}, and E_{ex-rep} are the induction, dispersion, and exchange-repulsion energies, respectively. E_{es} describes the electrostatic interaction between two unperturbed charge distributions, E_{ind} originates from the interaction of the unperturbed charge density on one monomer with the induced charge distribution on the other (and visa versa), E_{disp} accounts for instantaneous interactions between fluctuating charge distributions on different monomers, and E_{ex-rep} originates from the antisymmetrization of the wave function as a manifestation of the Pauli principle.[4]

We have recently embarked on a quest for an accurate yet efficient evaluation of electrostatic interaction energies in molecular complexes.[5] In widely used force field approaches, E_{es} is commonly calculated with a multipole or Buckingham-type approximation:[2,6]

$$E_{es} = \sum_i^{N_A}\sum_j^{N_B} \mathbf{T}[\mathbf{r}_{ij}]q_iq_j + \mathbf{T}_\alpha[\mathbf{r}_{ij}](q_i\mu_{\alpha,j} - q_j\mu_{\alpha,i}) + $$
$$\mathbf{T}_{\alpha\beta}[\mathbf{r}_{ij}]\left(\frac{1}{3}q_i\Theta_{\alpha\beta,j} + \frac{1}{3}q_j\Theta_{\alpha\beta,i} - \mu_{\alpha,j}\mu_{\alpha,i}\right) + \dots \quad (2)$$

* Corresponding author e-mail: volkov@chem.buffalo.edu.
[†] This paper is dedicated to the memory of Dr. John Rys.

where q, μ, Θ, etc. are the permanent atomic moments (monopole, dipole, quadrupole, etc.) in the unperturbed molecular charge distributions and parameters $\mathbf{T}_{\alpha\beta\gamma\ldots}[\mathbf{r}_{ij}]$ are the so-called interaction tensors (with the Einstein summation convention for indices α, β, γ, etc. used), which also depend on the separation of atomic centers $\mathbf{r}_{ij}$. Parameters N_A and N_B represent the number of atoms in molecular fragments A and B, respectively. In many cases, only the first point-charge term of expansion 2 is used,[7–10] although the second and part of the third term of expansion 2 (i.e., charge–dipole and dipole–dipole contributions) have been added in some of the force fields.[11]

In the more advanced distributed multipole approach by Stone and co-workers,[12,13] the expansion is extended to higher-order terms but is still subject to the fundamental limitation of the multipole approximation; that is, it is valid only for nonoverlapping charge distributions. This is especially troublesome for strongly bound systems, involving, for example, short H bonds. In such cases, the multipole approach cannot possibly yield accurate results, and the addition of penetration terms,[12,14] the use of off-atom centered[12] and damping functions,[12] etc. have been proposed. This complicates the calculation process and greatly reduces the transferability of atomic properties.

In our recent paper,[15] we have described a novel approach, called the exact potential and multipole moment (EPMM) method, for the fast and accurate evaluation of electrostatic interaction energies (E_{es}) between two molecular charge distributions within the Hansen–Coppens[16,17] pseudoatom electron density formalism. It combines a numerical evaluation of the exact Coulomb integral for the short-range with the Buckingham-type multipole approximation for the long-range interatomic interactions. It was found, for example, that for intermolecular O···H interactions in molecular systems the multipole approximation underestimates the strength of E_{es}(O···H) by as much as 50 kJ/mol for O···H ~ 1.5 Å, while the EPMM method yields almost that exact result.

We have combined the EPMM method with electron densities from our recently developed theoretical databank of transferable aspherical pseudoatoms,[18,19] referred to below as the DB+EPMM approach. The databank consists of chemically unique pseudoatoms, identified on the basis of common connectivity and bonding. They were extracted from B3LYP/6-31G** densities of a large number of small molecules using a least-squares projection technique in Fourier transform space, and show excellent consistency among chemically equivalent atoms in different molecules. The resulting electrostatic interaction energies E_{es} of monomers in molecular dimers were found to be in a very good agreement with those from a Morokuma–Ziegler decomposition[20,21] of double- and triple-ζ energies[15] evaluated at the density functional level of theory (DFT).

The comparison of E_{es} calculated using the databank parameters (derived from Gaussian-type wave functions) with ADF[22–24] results (in which the Slater-type functions are used and only pure DFT functionals, such as BLYP, are available) is not fully convincing because the two levels of theory used are not equivalent. A meaningful comparison should include

intermolecular E_{es} calculated at *exactly* the same level of theory at which the databank parameters were obtained, that is, B3LYP/6-31G**. To this end, a new program, SPDFG, was written for the evaluation of E_{es} from monomer charge distributions expressed in terms of Gaussian-type basis functions. This allows an extensive study of the electrostatic energy of interaction between molecules and its dependence on the orbital basis set for a wide variety of quantum-chemical methods.

Test Systems and Calculations

The current analysis is based on six pairs (dimers) of zwitterionic glycine molecules such as occur in crystals of α-glycine[25] (Figure 1).

Monomer molecular wave functions for Gaussian-type calculations were obtained with the Gaussian03 (G03) suite of programs[26] using methods and basis sets listed in Table 1. The standard Gaussian03 option Output = WFN (and Density = Current for correlated wave functions) generates coefficients of natural orbitals in a primitive basis. For correlated wave functions (MP2, MP4SDQ, CISD, QCISD, and CCSD), generalized densities are based on the Z-vector method.[27–30] All Gaussian03 calculations were performed with the SCF = Tight option, which requests tight self-consistent field convergence criteria.

The new SPDFG program uses the numerical Rys quadrature method[31,32] for the evaluation of one- and two-electron Coulomb integrals. The method is based on a set of orthogonal (Rys) polynomials,[33] which yields a simple general formula for integrals over basis functions, χ, of arbitrarily high angular momentum:

$$\langle\chi_i(1)\chi_j(1)|r_{12}^{-1}|\chi_k(2)\chi_l(2)\rangle = \sum_{\alpha=1}^{N} I_x(u_\alpha)I_y(u_\alpha)I_z^*(u_\alpha)W_\alpha \quad (3)$$

in which u_α and W_α are the roots and weights of the Nth order Rys polynomial and I_x, I_y, and I_z^* are simple two-dimensional integrals, evaluated using efficient and compact recurrence formulas.[32] The program is parallelized using the message-passing interface and can handle basis functions of any angular momentum. E_{es} for Hartree–Fock wave functions evaluated with the SPDFG program are in excellent agreement with those obtained with Morokuma energy decomposition in GAMESS-US.[46]

For Slater-type calculations, E_{es} was obtained using the Morokuma–Ziegler energy decomposition scheme[20,21] implemented in the program ADF,[22–24] which gives electrostatic interaction energies between monomers that are exact within the approximations of the theoretical calculation.

All calculations were performed using our own Linux Beowulf-type cluster equipped with dual-and quad-processor AMD AthlonMP and Opteron nodes.

Results and Discussion

As the electrostatic energy is a major component of the total interaction energy, an analysis of its dependence on the basis set choice and level of theory employed is required for a better understanding of computational results. This is especially important for the evaluation of the performance of the
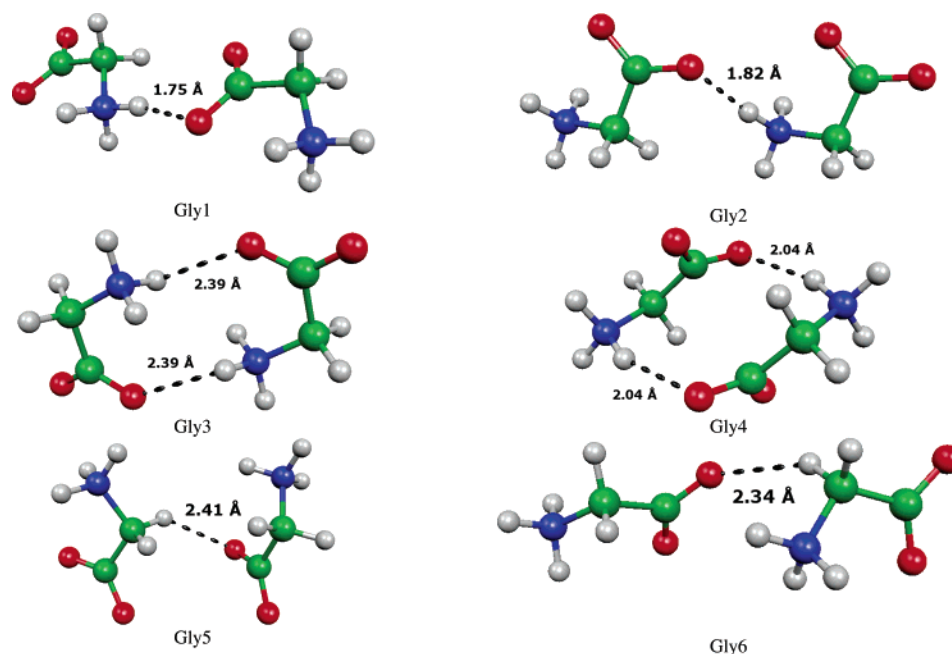
Figure 1. Six dimers in the crystal of α-glycine (oxygen atoms shown in red, nitrogens in blue, carbons in green, and hydrogens in gray).

Table 1. Methods and Basis Sets Used in the Study

	basis sets	
methods	w/o diffuse functions	w/diffuse functions
Gaussian-Type Calculations		
Hartree−Fock (HF)	6-31G**[34]	6-31++G**[35]
DFT with pure BLYP[36] and PBE[37,38] functionals	DZP[39]	DZP+diffuse[40] (DZP+)
DFT with hybrid B3LYP[41] functional	cc-pVDZ[42,43]	aug-cc-pVDZ[42,43]
Møller−Plesset second-order (MP2)	cc-pVTZ[42,43]	aug-cc-pVTZ[42,43]
Møller−Plesset fourth-order with single, double, and quadruple substitutions (MP4SDQ)	cc-pVQZ[42,43,†]	aug-cc-pVQZ[43−45,†]
CI with single and double substitutions (CISD)		
quadratic CI with single and double substitutions (QCISD)		
coupled cluster (CC) with single and double substitutions (CCSD)		
Slater-Type Calculations		
DFT with pure BLYP functional	DZP	
	TZP	
	QZ4P	

† For MP2, HF, and DFT calculations only.

DB+EPMM method, which is to be applied to much larger systems of biological interest to which quantum-mechanical methods are not easily applicable.

1. Effect of Basis Set on the Computed Electrostatic Interaction Energy. *1.1. Comparison of Related Double-, Triple-, and Quadruple-ζ Gaussian and Slater Basis Sets.* The effect of extending the basis set from double-ζ (DZ) to triple-ζ (TZ) is shown in Figure 2a. For Gaussian functions, we report $\Delta E_{es} = E_{es}(\text{cc-pVTZ}) - E_{es}(\text{cc-pVDZ})$, whereas for Slater functions, TZP and DZP are compared. For Gaussians, the energy calculated with the TZ basis is always more negative (more attractive or slightly less repulsive in the case of dimer Gly5) than the DZ value. The most significant changes are observed for DFT calculations. For example, for Gly3 and Gly4 dimers, ΔE_{es} is as large as 10−15 kJ/mol for pure DFT and 9−11 kJ/mol for hybrid B3LYP

functionals. ΔE_{es} at the Hartree−Fock (HF) level is relatively insensitive to the quality of the basis set, the maximum value being just over 4 kJ/mol for the Gly3 dimer. ΔE_{es} values for post-HF calculations are usually intermediate between those for HF and B3LYP.

When considering the dependence of ΔE_{es} on the relative orientation of monomers in dimers, generally, the smallest values are observed for dimers Gly1, Gly2, Gly5, and Gly6 and the largest for dimers Gly3 and Gly4, which are connected by two symmetry-related N−H···O hydrogen bonds.

Slater-type DFT calculations exhibit a different orientational dependence than that observed for Gaussians. For dimer Gly5, the electrostatic energy calculated with the TZ basis is more repulsive by ~4 kJ/mol than that calculated with the DZ basis. No differences between TZ and DZ basis sets are found for dimer Gly4, in marked contrast to results
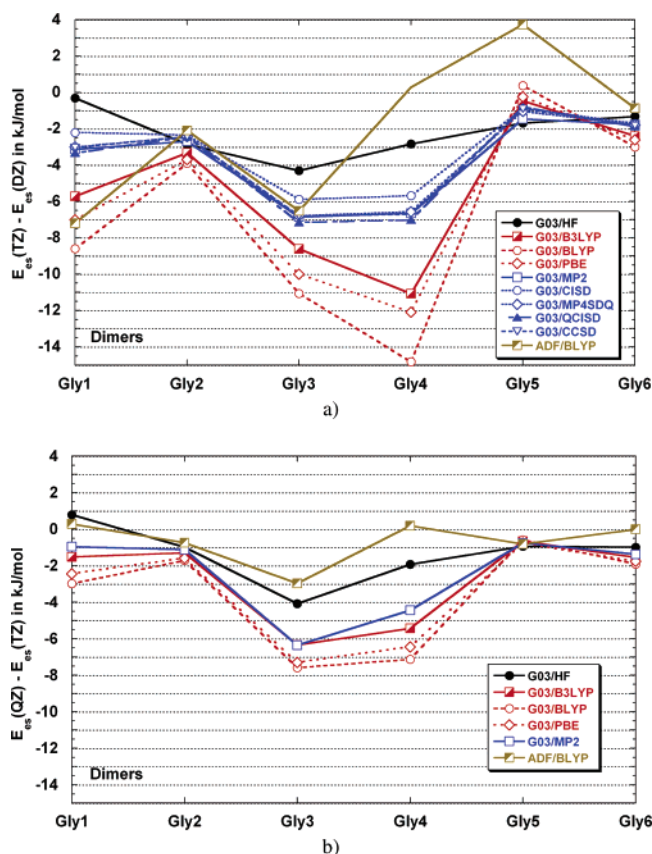
Figure 2. Difference between E_{es} (in kJ/mol) calculated with (a) TZ and DZ basis sets and (b) TZ and QZ basis sets at different levels of theory. For the Gaussian-type calculations, the differences are between (a) cc-pVTZ and cc-pVDZ and (b) cc-pVQZ and cc-pVTZ basis sets. For the Slater-type calculations, (a) TZP-DZP and (b) QZ4P-TZP results are shown.

obtained with Gaussian functions. The largest deviation ($\sim$7 kJ/mol) is observed for dimers Gly1 and Gly3.

Overall, the differences in E_{es} between the DZ and TZ bases are significant, and in the case of α-glycine dimers can reach 15 kJ/mol. All Gaussian-type calculations show approximately the same dependence of ΔE_{es} on the relative orientation of monomers in dimers, which is different from that observed for DFT calculations with Slater functions.

Figure 2b shows the effect of further expansion of the basis set from triple- to quadruple-ζ (QZ). This leads to corrections for Gaussian DFT and MP2 energies which are smaller than the change between DZ and TZ bases. Although the HF corrections are small, they are comparable to those when going from a DZ to a TZ basis. For Slater-type DFT calculations, the QZ/TZ difference is only -3 kJ/mol for dimer Gly3; -1 kJ/mol for dimers Gly2 and Gly5; and essentially zero for dimers Gly1, Gly4, and Gly6.

For Slater-type calculations, the convergence of E_{es} is nearly complete at the QZ level, while even more extended basis sets are needed to achieve a similar convergence in Gaussian-type calculations; that is, quintuple- or perhaps even sextuple-quality basis sets would be required.

1.2. Effect of Inclusion of Diffuse Functions in the Basis Sets. A prominent result obtained in this study is that the inclusion of diffuse functions in *monomer* charge density
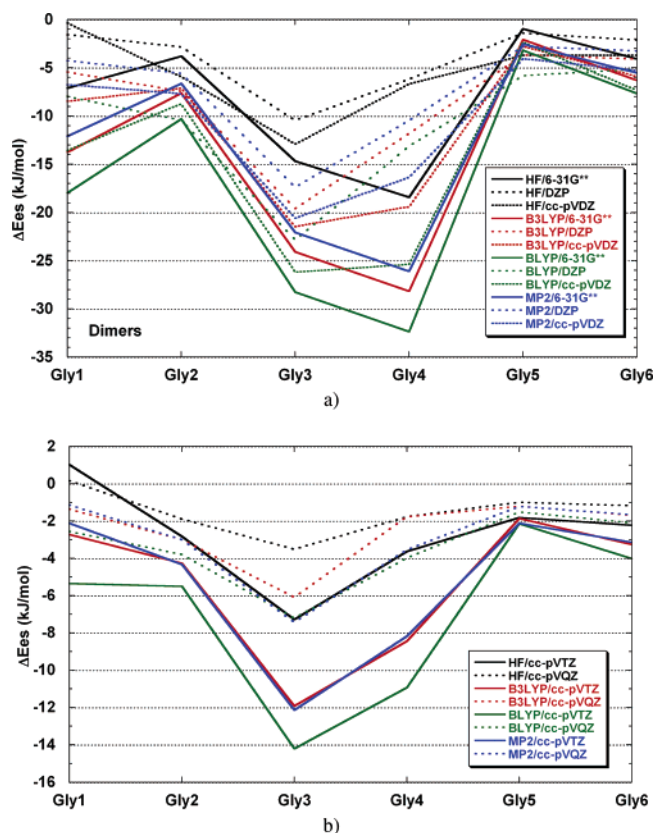

Figure 3. Effect of inclusion of diffuse functions in monomer basis sets on the electrostatic interaction energies in dimers (kJ/mol) for (a) several double-ζ-quality basis sets and (b) triple- and quadruple-ζ basis sets.

calculations has a much more pronounced effect on electrostatic interaction energies than even the change from a simple 6-31G** basis to the cc-pVTZ basis set. Figure 3a shows these effects for DZ-quality basis sets, and Figure 3b shows analogous results for TZ and QZ basis sets. Results are shown only for HF, B3LYP, BLYP, and MP2 calculations, with other methods showing similar behavior.

The inclusion of diffuse functions usually lowers E_{es} (except for a very small positive energy change in HF/cc-pVTZ and cc-pVQZ calculations). Not surprisingly, the 6-31G** basis tends to show a much larger variation in E_{es} upon the inclusion of diffuse functions than any other basis set examined in this study. The change is as small as $2-4$ kJ/mol for dimer Gly5 and as large as $28-32$ kJ/mol for dimers Gly3 and Gly4. The other two DZ-type basis sets (cc-pVDZ and DZP) are somewhat less affected by the inclusion of diffuse functions than 6-31G**. The maximum changes are $\sim$25$-$26 kJ/mol for the cc-pVDZ basis in dimers Gly4 and Gly3 and $\sim$22 kJ/mol for the DZP basis in dimer Gly3. For dimers Gly2, Gly5, and Gly6, the inclusion of diffuse functions does not significantly affect the E_{es} for any of the DZ-quality basis sets: changes are generally under 10 kJ/mol.

As expected, the effect of including diffuse functions diminishes in going from double- to triple- to quadruple-ζ basis sets. The biggest effects are 7, 14, and 32 kJ/mol for QZ-, TZ-, and DZ-quality basis sets, respectively.

Intermolecular Electrostatic Interaction Energy

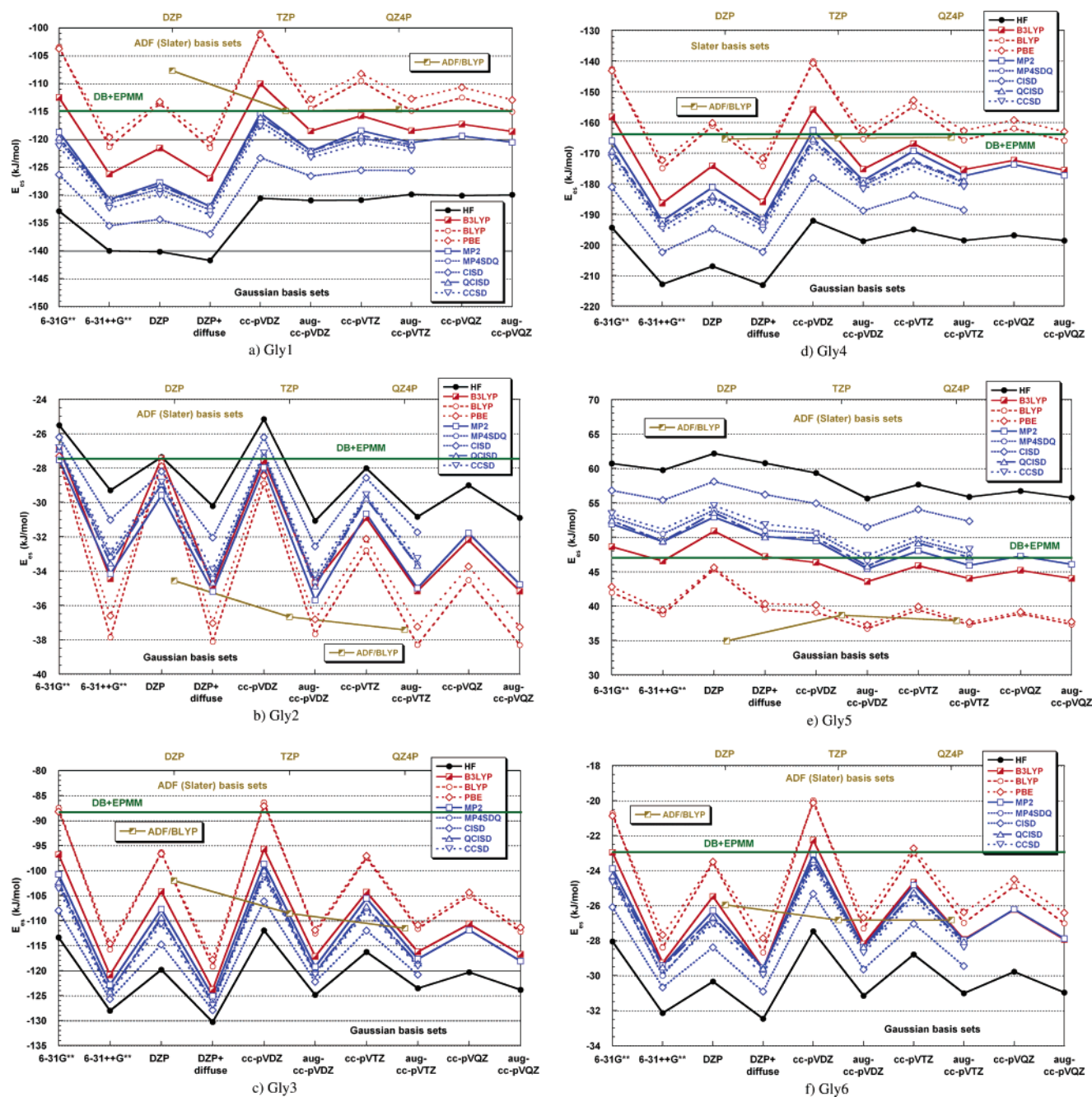*J. Chem. Theory Comput., Vol. 2, No. 1, 2006* **85**

Figure 4. Electrostatic interaction energies (in kJ/mol) in Gly1 (a), Gly2 (b), Gly3 (c), Gly4 (d), Gly5 (e), and Gly6 (f) dimers calculated at different levels of theory.

The inclusion of diffuse functions has its greatest effect on Gaussian DFT (more pronounced for pure DFT functionals) and MP2 energies; is slightly less for CCSD, QCISD, and MP4SDQ; and is the least for HF and CISD methods.

In general, the importance of diffuse functions for the calculation of intermolecular E_{es} reported here is in accord with the results of previous studies, for example, those using symmetry-adapted perturbation theory[3] for various types of systems.[47−49] Similar conclusions were also drawn from the studies of supermolecular interaction energies in both hydrogen-bonded[50,51] and $\pi−\pi$ interacting[52] systems, molecular electric moments and polarizabilities,[53] and so-called correlated cumulative atomic multipole moments.[54] Our results confirm that augmenting a given basis set is more

important for the calculation of E_{es}, and also E_{int},[52] than adding a shell of valence functions (i.e., aug-cc-pVDZ vs cc-pVTZ).

It is noteworthy that, within the BLYP method, the electrostatic interaction energies obtained with augmented Gaussian triple- and quadruple-ζ basis sets are in an excellent agreement with those from TZP and QZ4P Slater calculations.

2. Method Dependence of E_{es}. The results summarized in Figure 4a−f show both the basis set and method dependences of E_{es} for each of the α-glycine dimers. The Gaussian-type basis sets are listed at the bottom along the x axis, while Slater-type basis sets are listed at the top of the graphs. Values of E_{es} obtained from the DB+EPMM approach are represented by the solid horizontal line.

The overall spread of E_{es} obtained from first principle calculations is remarkably large. For example, it is as large as 74 kJ/mol between pure DFT calculations with cc-pVDZ or 6-31G** basis sets and HF/DZP+ calculations in the Gly4 dimer. Judging from the results of our best calculations, these DFT values are estimated to be ~22% too large and the HF values are ~19% too low. Similarly, the spread of calculated E_{es} values is about 40 kJ/mol in dimers Gly1 and Gly3.

Given the large basis-set dependence, we will analyze the effect of the method on E_{es} within each basis set to arrive at general conclusions. In general, we can distinguish five groups in terms of their E_{es} method dependence: (1) calculations with pure DFT functionals using both Slater and Gaussian functions (BLYP in ADF and BLYP/PBE in G03); (2) hybrid DFT with Gaussian functions (B3LYP in G03); (3) HF; (4) CISD; and (5) MP2, CCSD, QCISD, and MP4SDQ, the last four groups all with Gaussian functions.

For the strongly bonded dimers Gly1, Gly3, and Gly4, all DFT calculations yield less-negative E_{es} values than do the advanced correlated methods, such as CCSD, QCISD, and MP4SDQ. HF, on the other hand, overestimates E_{es} (by 10−20 kJ/mol), as does CISD, but by a smaller amount than HF (~5−10 kJ/mol). The advanced correlated methods, and also MP2, show consistent results with differences of only a few kJ/mol. This is also true for the somewhat more weakly bonded dimer Gly6, except that with aug-cc-pVTZ, and the higher basis set B3LYP method shows an excellent agreement with MP2 results.

For the only repulsive dimer, Gly5, included in this study, the same trend with respect to the method is observed but with the opposite sign, that is, repulsion is largest for HF and more advanced methods.

For the weakly bonded Gly2 dimer, the situation is the opposite of the one described above. HF and CISD calculations underestimate E_{es}, while pure DFT overestimates it. The behavior of the hybrid DFT B3LYP functional in this dimer is similar to that of the MP2, CCSD, QCISD, and MP4SDQ calculations.

Within a given Gaussian basis set approximation, the values of E_{es} either increase or decrease, depending on the spatial orientation of the monomers, in the following order: HF, CISD, (CCSD, QCISD, MP4SDQ), MP2, B3LYP, and pure DFT functionals. As expected, advanced correlated methods, such as CCSD, QCISD, and MP4SDQ, are consistently in good agreement with one another. Electron correlation effects are significant. Large-basis HF calculations yield values that differ from comparable correlated results by factors ranging from 0.95 (Gly1) to 1.19 (Gly5) in reasonable agreement with factors of about 0.94 previously reported for H_2O and HF dimers.[47] MP2 consistently overestimates the magnitude of the electron correlation correction, but never by more than 3 kJ/mol in the six dimers studied here, and these small deviations are removed at the MP4SDQ level of theory. This agrees with early studies of the convergence of the Møller−Plesset perturbation expansion applied to the calculation of electrostatic interaction energies[55] and electron density distributions[56] in simple closed-shell molecules. The CISD method, which suffers from nonsize consistency, recovers only half of the electron correlation correction; that is, CISD E_{es} values are roughly halfway between those of HF and advanced correlated methods. Clearly, CISD is inappropriate for molecules comparable to or larger than glycine.

Hybrid DFT B3LYP calculations often deviate significantly from advanced correlated methods for double-ζ quality basis sets, but the agreement improves for more extended basis sets. The deviations of pure DFT calculations (using either Gaussian or Slater functions) from advanced correlated methods always have the same sign but larger magnitude than those of hybrid B3LYP calculations. Problems with pure DFT functionals have been attributed to their inability to correctly describe long-range correlations,[57,58] which in general can be remedied by incorporation of the special asymptotic correction.[59,60] Hybrid DFT functionals, such as B3LYP, by their very nature, already include a part of correct asymptotics via Hartree−Fock exchange, which improves the overall asymptotic behavior of these functionals. Accordingly, electrostatic energies calculated with pure DFT functionals almost always deviate much more from advanced correlated methods than does hybrid B3LYP. The latter energies are almost always intermediate between those from pure DFT and HF calculations and, in some cases, are even in very good agreement with MP2 results. It is anticipated that, once the asymptotic correction is applied to pure DFT functionals, their performance should improve dramatically and produce electrostatic interaction energies close to those of CCSD.[61]

Several previous studies relate to the method dependence of intermolecular electrostatic interaction energies, either based on the perturbation approach, which adds correlation corrections to the Hartree−Fock E_{es} from perturbation contributions,[47,48,55,62,63] or calculated from relaxed correlated densities.[63] In general, our results, obtained on systems much larger than those studied previously, confirm (a) the importance of intramolecular correlation for the calculation of intermolecular electrostatic interaction energies, (b) almost complete convergence of E_{es} at the MP4SDQ level, and (c) the relative unimportance of higher-order terms included in the CCSD theory.[63] We find that intramolecular correlation included even at the MP2 level yields highly satisfactory electrostatic interaction energies for the type of systems studied here.

3. Effectiveness of the Databank in the E_{es} Calculation. One of the goals of this study is to obtain reliable reference values for E_{es} in the test dimers in order to provide a benchmark of accuracy for E_{es} obtained with the DB+EPMM approach. Two questions have to be addressed: (1) how does the databank approach compare with the B3LYP/6-31G** method on which it is based, and (2) how does it compare with much more advanced correlated methods?

As to the first question, the agreement between electrostatic interaction energies calculated with the DB+EPMM method and B3LYP/6-31G** values is quite good —under 4 kJ/mol (~1 kcal/mol) for five out of six dimers. This good agreement should be viewed in light of the fact that the glycine molecule was not included in the set of molecules used in the construction of the pseudoatom databank. For the Gly3 dimer, the difference is slightly larger—9 kJ/mol.

Intermolecular Electrostatic Interaction Energy

J. Chem. Theory Comput., Vol. 2, No. 1, 2006 **87**

Taking account of the fact that, in constructing the databank, B3LYP/6-31G** Gaussian-type densities were projected onto the Slater-type basis set used in the Hansen−Coppens pseudoatom model, and that the final set of pseudoatom parameters is obtained by averaging over many slightly different chemical environments and atomic conformations, a root-mean-square (RMS) discrepancy of 4 kJ/mol is quite acceptable.

As to the second question, DB+EPMM, like B3LYP/6-31G** itself, always underestimates the attractive electrostatic interaction energy compared to our best ab initio CCSD/aug-cc-pVTZ calculation. The differences can be fairly significant. Thus, for dimers Gly3 and Gly4, the differences between DB+EPMM and CCSD/aug-cc-pVTZ E_{es} values are as large as ∼30 and 20 kJ/mol, respectively. But, for the Gly1, Gly2, and Gly6 dimers, the DB+EPMM approach underestimates E_{es} by only 5−7 kJ/mol (1−2 kcal/mol). For the only repulsive dimer, Gly5, the DB+EPMM energy is in excellent agreement with the CCSD/aug-cc-pVTZ value. The RMS discrepancy between DB+EPMM and CCSD/aug-cc-pVTZ E_{es} values is only 16 kJ/mol for the set of six dimers, essentially due to the less advanced method on which the databank is based. For comparison, the RMS deviation between B3LYP/6-31G** and CCSD/aug-cc-pVTZ energies is 14 kJ/mol, and between the best ADF BLYP/QZ4P calculation and CCSD/aug-cc-pVTZ it is 9 kJ/mol.

4. Dependence of Electron Density Distributions on the Level of Theory. Electrostatic interaction energies described in this paper are, of course, intimately related to the electron density distribution in the monomer of α-glycine. Figure 5a shows the difference between HF/cc-pVTZ and HF/cc-pVDZ electron density distributions plotted in the plane of an oxygen, the carbon atom of the CH$_2$ group, and the nitrogen atom. The extension of the basis set from DZ to TZ significantly affects the spherical component of the electron density near the atom cores, which is expected to be relatively unimportant for E_{es} calculations, and increases the density in the bonding and tail regions of the density distributions, which is expected to be more important. Figure 5b illustrates the effect of including diffuse functions in the cc-pVDZ orbital basis set at the HF level. The results for other methods and basis sets are similar. Surprisingly, the effect is not confined to the tails of the density distributions but is also pronounced near the atoms and in the bonding regions. Most remarkable are the nonspherical features around the atoms. The effect of electron correlation is shown in Figure 5c. As observed in previous studies,[30] correlation builds charge density near the nuclei and decreases it in bonding regions. Contrary to earlier studies, the charge density is actually depleted in a very small region in the immediate vicinity of the oxygen and nitrogen atoms. To ensure that this feature is not an artifact of our calculations, we repeated the formaldehyde calculations previously reported by Wiberg et al.,[30] computing charge densities on a finer grid of points, and found the same feature in that molecule.

In general, the effects of the basis set and method of computation are rather significant and sufficiently complicated to account for the observed changes in the inter-
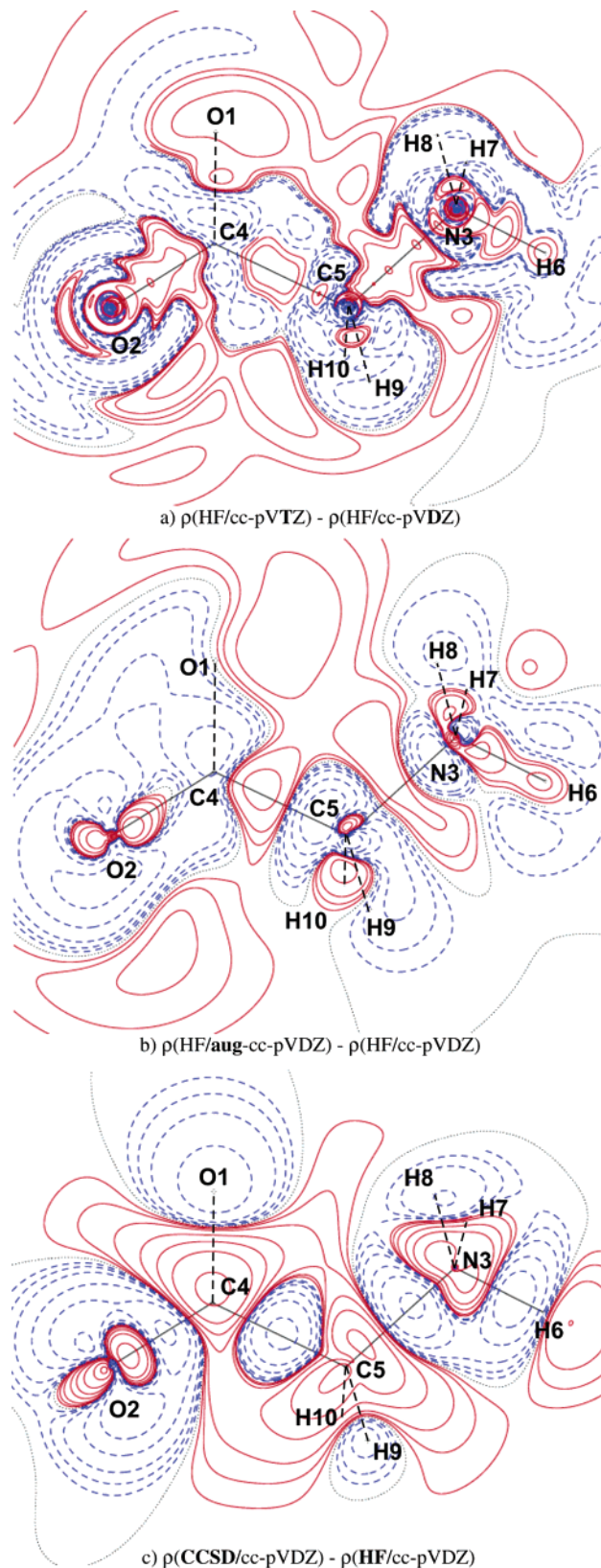

a) ρ(HF/cc-pVTZ) - ρ(HF/cc-pVDZ)

b) ρ(HF/aug-cc-pVDZ) - ρ(HF/cc-pVDZ)

c) ρ(CCSD/cc-pVDZ) - ρ(HF/cc-pVDZ)

Figure 5. Differences in the charge density distribution in the glycine molecule between various levels of theory in the plane of the oxygen, the carbon of the CH$_2$ group, and the nitrogen atom. Positive contours are shown with a solid red line, negative with a dashed blue line, and zero with a dotted black line. Contour levels are $\pm 2 \times 10^{-4}$, $\pm 4 \times 10^{-4}$, $\pm 8 \times 10^{-4}$, $\pm 2 \times 10^{-3}$, $\pm 4 \times 10^{-3}$, $\pm 8 \times 10^{-3}$, $\pm 2 \times 10^{-2}$, and so forth e/au^3.

molecular electrostatic interaction energy and its orientational dependence.

Concluding Remarks

The results of an extensive analysis of the basis set and method dependence of intermolecular electrostatic interaction energies in six dimers of α-glycine show that the most significant effect on E_{es} in Gaussian-type calculations is produced by the inclusion of diffuse functions, even in the case of the fairly extended cc-pVQZ basis set. For augmented Gaussian basis sets, the convergence of E_{es} is nearly complete at the aug-cc-pVDZ level. The basis set dependence in Slater-type calculations of E_{es} is somewhat smaller than that for Gaussians.

The method dependence of the calculated E_{es} is also pronounced. Advanced correlated methods, such as QCISD, CCSD, and MP4SDQ, and also MP2, show very consistent results, usually within a range of $1-2$ kJ/mol. Of these, MP2 is much less computationally demanding and when combined with aug-cc-pVTZ or (if possible) a larger basis set is capable of producing accurate benchmark electrostatic interaction energies. Electrostatic energies obtained with HF and CISD methods deviate considerably from these results, generally overestimating the magnitude of the electrostatic interaction. Pure DFT functionals with both Gaussian and Slater basis functions almost always show large deviations from advanced correlated methods, apparently because of an incorrect long-range behavior of these functionals. Despite their different origins, BLYP and PBE functionals yield very similar E_{es} energies. Electrostatic interaction energies from the hybrid DFT B3LYP functional are in much better agreement with those of advanced correlated methods, especially when augmented TZ- or QZ-type basis sets are used. This is due to the inclusion of Hartree−Fock exchange, which by itself contains correct asymptotics and improves the overall asymptotic behavior of a hybrid functional.

The combination of the pseudoatom databank and EPMM method is well able to reproduce the results of the B3LYP/6-31G** calculations on which it is based (well within 10 kJ/mol, usually within 4 kJ/mol only), confirming the transferability of the pseudoatom densities among the types of molecules considered. However, because electrostatic interaction energies calculated at the B3LYP/6-31G** level of theory deviate (sometimes by $20-30$ kJ/mol) from advanced correlated results, the databank results show analogous discrepancies. This indicates that the databank can be improved by the use of electron densities from advanced correlated methods.

Nevertheless, the combination of the current databank for the evaluation of electrostatic interaction energies in molecular systems with high-quality atom−atom potentials for the description of exchange−repulsion, dispersion, and induction forces should provide total bonding energies at an accuracy similar to or better than those obtained by the standard DFT methods. This approach is now being pursued.[64]

References

(1) Xu, D.; Lin, S. L.; Nussinov, R. *J. Mol. Biol.* **1997**, *265*, 68−84.

(2) Stone, A. J. *The Theory of Intermolecular Forces*; Oxford University Press: Oxford, England, 1996; International Series of Monographs on Chemistry 32.

(3) Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887−1930.

(4) Bickelhaupt, F. M.; Baerends, E. J. *Rev. Comput. Chem.* **2000**, *15*, 1−86.

(5) Volkov, A.; Coppens, P. *J. Comput. Chem.* **2004**, *25*, 921−934.

(6) Buckingham, A. D. *Adv. Chem. Phys.* **1967**, *12*, 107−142.

(7) 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*, 5179−5197.

(8) MacKerell, A. D.; Wiorkiewicz-Kuczera, J.; Karplus. M. *J. Am. Chem. Soc.* **1995**, *117*, 11946−11975.

(9) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490−519.

(10) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.

(11) Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551−8566.

(12) Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233−239.

(13) Stone, A. J.; Alderton, M. *Mol. Phys.* **1985**, *56*, 1047−1064.

(14) Spackman, M. A. *J. Chem. Phys.* **1986**, *85*, 6587−6601.

(15) Volkov, A.; Koritsanszky, T. S.; Coppens, P. *Chem. Phys. Lett.* **2004**, *391*, 170−175.

(16) Hansen, N. K.; Coppens, P. *Acta Crystallogr., Sect. A* **1978**, *34*, 909−921.

(17) Coppens, P. *X-ray Charge Densities and Chemical Bonding*; Oxford University Press: New York, 1997.

(18) Koritsanszky, T.; Volkov, A.; Coppens, P. *Acta Crystallogr., Sect. A* **2002**, *58*, 464−472.

(19) Volkov, V.; Li, X.; Koritsanszky, T.; Coppens, P. *J. Phys. Chem. A* **2004**, *108*, 4283−4300.

(20) Ziegler, T.; Rauk, A. *Theor. Chim. Acta* **1977**, *46*, 1−10.

(21) Morokuma, K. *J. Chem. Phys.* **1971**, *55*, 1236−1244.

(22) 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*, 931−967.

(23) Fonseca Guerra, C.; Snijders, J. G.; te Velde, G.; Baerends, E. *Theor. Chem. Acc.* **1998**, *99*, 391−403.

(24) *ADF*, version 2004.01; SCM, Theoretical Chemistry, Vrije Universiteit: Amsterdam, The Netherlands, 2004. http://www.scm.com.

Intermolecular Electrostatic Interaction Energy

J. Chem. Theory Comput., Vol. 2, No. 1, 2006 **89**

(25) Destro, R.; Roversi, P.; Barzaghi, M.; Marsh, R. E. *J Phys Chem A* **2000**, *104*, 1047−1054.

(26) 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.01; Gaussian, Inc.: Wallingford, CT, 2004.

(27) Handy, N. C.; Schaefer, H. F., III. *J. Chem. Phys.* **1984**, *81*, 5031−5033.

(28) Diercksen, G. H. F.; Roos, B. O.; Sadlej, A. J. *Chem. Phys.* **1981**, *59*, 29−39.

(29) Diercksen, G. H. F.; Sadlej, A. J. *J. Chem. Phys.* **1981**, *75*, 1253−1266.

(30) Wiberg, K. B.; Hadad, C. M.; LePage, T. J.; Breneman, C. M.; Frisch, M. J. *J. Phys. Chem.* **1992**, *96*, 671−679.

(31) Dupuis, M.; Rys, J.; King, H. F. *J. Chem. Phys.* **1976**, *65*, 111−116.

(32) Rys, J.; Dupuis, M.; King, H. F. *J. Comput. Chem.* **1983**, *4*, 154−157.

(33) King, H. F.; Dupuis, M. *J. Comput. Phys.* **1976**, *21*, 144−165.

(34) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213−222.

(35) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257−2261.

(36) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100.

(37) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.

(38) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1997**, *78*, 1396−1396.

(39) Dunning, T. H., Jr. *J. Chem. Phys.* **1970**, *53*, 2823−2833.

(40) Dunning, T. H., Jr.; Hay, P. J. In *Methods of Electronic Structure Theory*; Schaefer, H. F., III, Ed.; Plenum Press: New York, 1977; Vol. 3.

(41) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(42) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007−1023.

(43) Davidson, E. R. *Chem. Phys. Lett.* **1996**, *260*, 514−518.

(44) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1994**, *100*, 2975−2988.

(45) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796−6806.

(46) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347−1363.

(47) Rybak, S.; Jeziorski, B.; Szalewicz, K. *J. Chem. Phys.* **1991**, *95*, 6576−6601.

(48) Williams, H. L.; Mas, E. M.; Szalewicz, K.; Jeziorski, B. *J. Chem. Phys.* **1995**, *103*, 7374−7391.

(49) Szalewicz, K.; Cole, S. J.; Koos, W.; Barlett, R. J. *J. Chem. Phys.* **1988**, *89*, 3662−3673.

(50) Lii, J.-H.; Ma, B.; Allinger, N. L. *J. Comput. Chem.* **1999**, *20,* 1593−1603.

(51) Halkier, A.; Koch, H.; Jørgensen, P.; Christiansen, O.; Beck Nielsen, I. M.; Helgaker, T. *Theor. Chem. Acc.* **1997**, *97*, 150−157.

(52) Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. *J. Am. Chem. Soc.* **2002**, *124*, 10887−10893.

(53) van Duijneveldt-van de Rijdt, J. G. C. M.; van Duijneveldt, F. B. *J. Mol. Struct.* **1982**, *89*, 185−201.

(54) Sokalski, W. A.; Sawaryn, A. *J. Chem. Phys.* **1987**, *87*, 526−534.

(55) Moszynski, R.; Jeziorski, B.; Ratkiewicz, A.; Rybak, S. *J. Chem. Phys.* **1993**, *99*, 8856−8869.

(56) Amos, R. D. *Chem. Phys. Lett.* **1980**, *73*, 602−606.

(57) Misquitta, A. J.; Szalewicz, K. *Chem. Phys. Lett.* **2002**, *357*, 301−306.

(58) Misquitta, A. J.; Szalewicz, K. *J. Chem. Phys.* **2005**, *122*, 214109-1−214109-19.

(59) Tozer, D. J.; Handy, N. C. *J. Chem. Phys.* **1998**, *109*, 10180−10189.

(60) Casida, M. E.; Salahub, D. R. *J. Chem. Phys.* **2000**, *113*, 8918−8935.

(61) Szalewicz, K. Personal communication.

(62) Bukowski, R.; Szalewicz, K.; Chabalowski, C. F. *J. Phys. Chem.* **1999**, *103*, 7322−7340.

(63) Korona, T.; Moszynski, R.; Jeziorski, B. *Mol. Phys.* **2002**, *100*, 1723−1734.

(64) Volkov, A.; Li, X.; Coppens, P.; Szalewicz, K. To be published.

(65) Flükiger, P. F. Development of the molecular graphics package MOLEKEL and its application to selected problems in organic and organometallic chemistry, Thèse No 2561, Département de chimie physique, Université de Genève, Genève, Switzerland, 1992.

(66) Portmann, S.; Lüthi, H. P. *CHIMIA* **2000**, *54*, 766−769.