Combination of the exact potential and multipole methods (EP/MM) for evaluation of intermolecular electrostatic interaction energies with pseudoatom representation of molecular electron densities

Anatoliy Volkov a,*, Tibor Koritsanszky b, Philip Coppens a,*

a Department of Chemistry, State University of New York at Buffalo, 747 NS Complex, Buffalo, NY 14260-3000, USA
b Department of Chemistry, Middle Tennessee State University, MTSU Box 0595, 1301 East Main Street, Murfreesboro, TN 37132, USA

Received 18 February 2004; in final form 30 April 2004
Available online 18 May 2004

Abstract

A new method (EP/MM) for calculation of intermolecular electrostatic interaction energies from pseudoatom expansions of molecular densities is presented. It combines numerical evaluation of the exact Coulomb integral for the short-range with the Buckingham-type multipole approximation for the long-range interatomic interactions. In first instance the method is combined with the pseudoatom representation of the atomic densities as derived from theoretical wavefunctions. The resulting electrostatic interaction energies of monomers in molecular dimers are in very good agreement with Morokuma–Ziegler decomposition of triple-zeta DFT energies. The combination of EP/MM with densities from the theoretical databank of aspherical pseudoatoms provides more accurate results than the MMFF94 force-field at a moderate increase in computing time. EP/MM is also applicable to experimental electron densities.

1. Introduction

The calculation of accurate intermolecular interaction energies is of great importance in evaluation of a host of physical properties, including lattice energies of molecular crystals, protein binding and drug-substrate interactions. One of the most significant contributions to the total binding energy is the electrostatic interaction energy ($E_{es}$).

In our previous study of several molecular dimers [1] we investigated the accuracy of calculation of $E_{es}$ with atomic moments obtained from different types of electron density partitioning using a Buckingham-type (multipole, MM) approximation [2]

\[
E_{es}(\text{MM}) = \sum_{a \in A} \sum_{b \in B} T_{ab}\left(\frac{1}{3}q_a \Theta_{ab} - \frac{1}{3}q_b \Theta_{ba} - \mu_{ab} \mu_{ba}\right) \\
+ T_{abc}\left(\frac{1}{15}q_a \Omega_{bc} - \frac{1}{15}q_b \Omega_{ca} - \frac{1}{15}q_c \Omega_{ab}\right) \\
- \frac{1}{3} \mu_{ab} \Theta_{bc} + \frac{1}{3} \mu_{ba} \Theta_{cb} + \cdots,
\]

where the indices $a$ and $b$ run over atoms constituting monomers A and B, respectively; parameters $T_{ab}^{\gamma}$ are the interaction tensors ($\nabla \cdot \nabla \cdot \nabla \cdot \cdots \nabla \cdot r_{ab}^{\gamma}$, with $r_{ab}$ being the internuclear separation) and $q, \mu, \Theta, \Omega$ are the atomic multipole moments. The resulting $E_{es}(\text{MM})$ were compared with those obtained with the Morokuma–Ziegler decomposition scheme [3,4] implemented in the program ADF [5], in which the $E_{es}$ between two unperturbed molecular charge distributions is defined by the exact potential (EP) [6]

* Corresponding authors. Fax: +17-166-456-948 (A. Volkov).
E-mail addresses: volkov@chem.buffalo.edu (A. Volkov), coppens@buffalo.edu (P. Coppens).

0009-2614/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
\[ E_{es}(\text{EP}) = \sum_{\alpha \in A} \sum_{\beta \in B} Z_\alpha Z_\beta \int \rho_\alpha(r_\alpha) V^\text{nuc}_B(r_\alpha) \, dr_\alpha \\
+ \int_B \rho_B(r_B) V^\text{nuc}_B(r_B) \, dr_B \\
+ \int_A \int_B \frac{\rho_A(r_A) \rho_B(r_B)}{|r_A - r_B|} \, dr_A \, dr_B, \tag{2} \]

where \( \rho_\alpha \) and \( \rho_B \) are the molecular electron densities and \( V^\text{nuc}_A \) and \( V^\text{nuc}_B \) are the nuclear potentials of molecules A and B, respectively. The primary focus of the study was on the performance of the atomic moments derived from multipole populations of nucleus-centered pseudoatom expansions of the electron density [7], obtained from: (a) least-squares projection of theoretical structure factors calculated from Kohn–Sham wavefunctions [8] (referred to as model) and (b) from an extended version of a library of transferable pseudoatoms derived by partitioning of theoretical electron densities [9] (referred to as the databank or DB). Though the agreement was quite reasonable, discrepancies as large as 69 kJ/mol remained, even though the databank and model densities compared well with those from the original calculations [9]. The discrepancies were attributed to the limitation of the multipole model (Eq. (1)) when applied to short-range interactions, as the deviations of \( E_{es}(\text{MM}) \) were much larger for the dimers with shortest intermolecular contacts (\( R_{O-H} \approx 1.5-1.8 \) A).

We present here an improved hybrid scheme for the evaluation of electrostatic intermolecular interactions which combines the evaluation of the exact potential with the multipole approximation in one algorithm. The approach is analogous to that used in the QM/MM method [10] for calculation of large molecular systems, in which the most important part of the system is treated at the \textit{ab initio} level while molecular mechanics methods are applied for the rest of the system. The molecular dimers examined in our earlier work [1] are used to test the method.

2. The EP/MM method

The Exact Potential and Multipole Method (EP/MM) numerically evaluates the exact Coulomb integral (Eq. (2)) for pseudoatom–pseudoatom interactions in the inner sphere, while the Buckingham-type multipole approximation Eq. (1), valid for non-overlapping charge distributions, is used for the interactions with pseudoatoms in the outer region.

The \( E_{es}(\text{EP}) \) between the two pseudoatom charge distributions \( \rho_{a \in A} \) and \( \rho_{b \in B} \) is evaluated by numerical (quadrature) integration over both pseudoatoms. The 6D integral in the last term of Eq. (2) can be much simplified when rewritten as [5]

\[ \int_a \int_b \frac{\rho_a(r_a) \rho_b(r_b)}{|r_a - r_b|} \, dr_a \, dr_b = \int_b \rho_b(r_b) V^\text{elec}_a(r_b) \, dr_b = \int_b \rho_b(r_b) V^\text{elec}_b(r_b) \, dr_b, \tag{3} \]

where \( V^\text{elec}_a \) and \( V^\text{elec}_b \) are the electronic potentials of pseudoatoms \( a \in A \) and \( b \in B \). This leads to the same type of integrals as in the second and third terms of Eq. (2), which can be evaluated in a single integration step [5]. Thus, if integrals \( \int_a \rho_a(r_a) V^\text{nuc}_b(r_a) \, dr_a \) and \( \int_b \rho_b(r_b) V^\text{elec}_a(r_b) \, dr_b \) are evaluated over pseudoatom \( a \in A \), then only one integral \( \int_b \rho_b(r_b) V^\text{nuc}_a(r_b) \, dr_b \) remains to be evaluated over pseudoatom \( b \in B \). In practice, however, both integrals in Eq. (3) are evaluated, providing an independent test for the accuracy of the integration – the values of the integrals have to agree within the desired level of accuracy.

3. Implementation and test systems

The numerical integration implemented in the new version of XDPROP [11] uses Gauss-Chebyshev [12] and Lebedev [13] quadratures for the radial and angular parts, respectively. Radial coordinates and weights are remapped using formula (M4) of Treutler and Ahlrichs [14]. A typical integration involves 50 radial and 194 angular quadrature points. In principle, the electronic potential \( V^\text{elec} \) can be evaluated from the pseudoatom model in various ways [15,16]. However, the method described in [15] is rather inaccurate and can only be used outside of the charge distribution, while the second method [16] involves evaluation of fairly complicated integrals. To overcome these disadvantages \( V^\text{elec} \) is obtained by calculating the potential of each Slater function [5] in the pseudoatom expansion. Details will be published elsewhere [17].

The current analysis is restricted to the systems studied previously [1]: six dimers of \( \alpha \)-glycine (Gly), two dimers of \( N \)-acetylglucosamine (AcG) and three dimers of \( L-(+)-\text{lactic acid} \) (Lac). The reference values for \( E_{es} \) were obtained using the Morokuma–Ziegler energy decomposition scheme implemented in ADF, which gives electrostatic interaction energies between monomers which are exact within the approximations of the theoretical calculation. The dimer and monomer calculations were performed at the BLYP/TZP and BLYP/DZP levels of theory [18,19].

The pseudoatom parameters of the model densities were obtained by least-squares fitting to the valence-only structure factors obtained by Fourier transform of B3LYP/6-31G** monomer densities [19–21] calculated with GAUSSIAN 98 (G98) [22]. The construction and application of the databank pseudoatoms have been described previously [8,9]. In both cases the
pseudoatom expansion was truncated at the hexadecapolar level ($l_{\text{max}} \leq 4$) for the non-hydrogen atoms and at the quadrupolar level ($l_{\text{max}} \leq 2$) for hydrogen atoms [9]. Test calculations varying $l_{\text{max}}$ show that on adding the $l = 4$ terms convergence in $E_{\text{es}}$ is reached at a level of 1% or better, in agreement with our earlier results [1].

For comparison, the $E_{\text{es}}$(MM) calculated using the atoms-in-molecules (AIM) [23] moments ($l_{\text{max}} \leq 4$), extracted from Gaussian 98/B3LYP/6-311++G** [24,25] densities, are also included in the current analysis. These moments have been increasingly used in recent years due to enhanced computational power and readily available software. The application of these moments to the calculation of $E_{\text{es}}$ has been studied by Popelier et al. [26,27]. A comparison with $E_{\text{es}}$ calculated with the MMFF94 [28] force-field [29] is also presented.

4. Results and discussion

Because EP/MM requires a user-supplied critical interatomic distance separating the EP and MM regions, we analyzed the differences in $E_{\text{es}}$ between the EP and MM methods, $\Delta E_{\text{es}} = E_{\text{es}}(\text{EP}) - E_{\text{es}}(\text{MM})$, for all intermolecular O–·–H interactions in the test dimers shorter than 7 Å. The results (Fig. 1a) show the adequacy of the MM approach for interactions above 3 Å and a pronounced underestimation of $E_{\text{es}}$(MM) for interactions below that distance, a discrepancy which increases exponentially up to 50 kJ/mol as the inter-atomic O–·–H distance approaches 1.5 Å. The discrepancy of 50 kJ/mol for the O–·–H interaction at 1.53 Å in dimer AcG2 by itself accounts for over 80% of the difference in $E_{\text{es}}$ between Morokuma–Ziegler and MM for both model and databank pseudoatoms. Similar, but not as dramatic curves are observed for other types of interactions, i.e. O–·–N (Fig. 1b), O–·–C, etc. but only because these atoms are positioned at much longer distances than O and H.

As convergence of $E_{\text{es}}$(EP/MM) to $E_{\text{es}}$(EP) was reached at around 4 Å, this critical distance was used in all EP/MM calculations. However, even at 4 Å the gain in speed with EP/MM compared to EP is very significant since only few pseudotom–pseudoatom interactions have to be evaluated by numerical integration. Clearly, the EP/MM method achieves the accuracy of the full EP evaluation at a much smaller computational cost.

Table 1 and Fig. 2 compare the previous results [1] obtained with MM with those obtained with EP/MM. The $E_{\text{es}}$(EP/MM) values are much closer to those obtained with Morokuma–Ziegler partitioning of ADF/BLYP/TZP energies (used as a reference) than $E_{\text{es}}$(MM). For several dimers the improvement is very significant—about 60 KJ/mol for AcG2, 45–50 kJ/mol for Lac1 and 30 kJ/mol for Gly1 and Gly4. The only dimer for which the $E_{\text{es}}$(EP/MM) becomes somewhat more deviant from the reference value is Lac3, but with deviations of only 1–3 kJ/mol. In 10 out of 11 dimers the $E_{\text{es}}$(EP/MM) are within 10 kJ/mol from the reference values. The exception is Gly3, for which the values are off by 20–24 kJ/mol, and the improvement of EP/MM over MM is only 7 kJ/mol. This discrepancy is not due to the electrostatic calculation which is exact, but attributed to the bias introduced by the pseudoatom model in the electron density upon projection via the X-ray structure factors into the Fourier space, as demonstrated previously (see, for example, [30,31]). The electrostatic interaction energy of Gly3 seems to be affected by this bias more than that of other dimers. A direct-space pseudotom partitioning, which is being investigated, may be needed to eliminate such a deviation.
The differences between EP/MM and the TZP reference values (RMS ~8-9 kJ/mol) are comparable in magnitude with those between the TZP and DZP ADF calculations (RMS = 5 kJ/mol). When comparison is made with a broader range of first-principle calculations the differences are generally comparable with or smaller than the spread among first-principle calculations. For example, in the dimer Gly4 the difference in electrostatic interaction energy between ADF BLYP/TZP (~165 kJ/mol) and GAMESS [32] Hartree–Fock/6-311++G** (~211 kJ/mol) calculations is as large as 46 kJ/mol, while the pseudoatom model agrees with the ADF BLYP/TZP electrostatic energy within 4 kJ/mol. In view of the differences between results from first principle calculations, the agreement for EP/MM using the pseudoatom model may be considered quite satisfactory. Note that for dimer Gly6 all methods yield essentially the same $E_{es}$, possibly due to the large value (3.2 Å) of the shortest (O···O) contact. At such a separation the Buckingham approximation is valid.

Table 1
Electrostatic interaction energies and their RMS deviations from Morokuma–Ziegler ADF/BLYP/TZP values for each of the methods examined (all in kJ/mol)

<table>
<thead>
<tr>
<th>Dimer</th>
<th>Morokuma–Ziegler ADF BLYP</th>
<th>MM AIM G98 B3LYP 6-311+G**</th>
<th>Pseudoatoms Model G98 B3LYP 6-31G**</th>
<th>Databank SCDS G98 MP2 6-31G**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly1</td>
<td>−115</td>
<td>−108</td>
<td>−106</td>
<td>−112</td>
</tr>
<tr>
<td>Gly2</td>
<td>−37</td>
<td>−35</td>
<td>−15</td>
<td>−31</td>
</tr>
<tr>
<td>Gly3</td>
<td>−109</td>
<td>−102</td>
<td>−110</td>
<td>−85</td>
</tr>
<tr>
<td>Gly4</td>
<td>−166</td>
<td>−165</td>
<td>−163</td>
<td>−167</td>
</tr>
<tr>
<td>Gly5</td>
<td>+43</td>
<td>+35</td>
<td>+55</td>
<td>+49</td>
</tr>
<tr>
<td>Gly6</td>
<td>−26</td>
<td>−26</td>
<td>−26</td>
<td>−25</td>
</tr>
<tr>
<td>AcG1</td>
<td>−48</td>
<td>−47</td>
<td>−35</td>
<td>−53</td>
</tr>
<tr>
<td>AcG2</td>
<td>−99</td>
<td>−90</td>
<td>−94</td>
<td>−91</td>
</tr>
<tr>
<td>Lac1</td>
<td>−70</td>
<td>−63</td>
<td>−56</td>
<td>−83</td>
</tr>
<tr>
<td>Lac2</td>
<td>−44</td>
<td>−42</td>
<td>−28</td>
<td>−39</td>
</tr>
<tr>
<td>Lac3</td>
<td>−13</td>
<td>−14</td>
<td>−11</td>
<td>−18</td>
</tr>
<tr>
<td>RMS</td>
<td></td>
<td></td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

*4 Å critical internuclear distance.

Fig. 2. Differences between the reference Morokuma–Ziegler electrostatic interaction energies from ADF/BLYP/TZP calculations and those from other methods (kJ/mol).

1 Root-Mean-Square.
One of the main goals of this study was to evaluate the performance of the theoretical pseudoatom databank in the calculation of $E_{es}$. The agreement between the model and databank pseudoatoms obtained with EP/MM is very good – with discrepancies never larger than 5 kJ/mol. Rather unexpectedly the performance of the databank with EP/MM (RMS = 8 kJ/mol) is better in most of the cases (dimer Gly3 being the only exception) than that of the MM with AIM moments (RMS = 11 kJ/mol) obtained at a significantly higher level of theory (B3LYP/6-311++G**) than the databank which is based on B3LYP/6-31G** densities. The EP/MM results also compare well with semi classical density sum (SCDS) energies (only reported for dimers of $\alpha$-glycine) obtained with MP2/6-31G** molecular densities [33,34], even though the latter are based on a higher level of theory. Note that while those two methods used electron densities obtained via a theoretical calculation for a given molecule, the databank used the averaged values of pseudoatoms. This illustrates the remarkable transferability of pseudoatoms and provides yet another proof of the self-consistency of the databank.

When comparing $E_{es}$ obtained with the databank and EP/MM with those calculated with the MMFF94 forcefield, the superiority of the pseudatom approach is evident. The databank performs better than the empirical force field for practically all the dimers, even for Gly3.

In general, our results illustrate the advantage of density-based partitioning schemes over pure multipole methods [35] which cannot make use of the exact Coulomb integral (Eq. (2)) and, thus, are unable to correctly reproduce the short-range electrostatic interaction energies [35].

5. Concluding remarks

The combination of the exact potential and multipole methods (EP/MM) is a highly effective tool for evaluation of electrostatic interaction energies from aspherical pseudoatom expansions of electron densities. The approach combines numerical quadrature evaluation of integrals involving the electron density and potentials for short-range pseudoatom-pseudoatom interactions with the standard Buckingham-type multipole approximation for long-range interactions. For the tested systems the optimal internuclear distance separating the EP and MM regions was found to be 4 Å, at which the $E_{es}$(EP/MM) become practically indistinguishable from the $E_{es}$(EP). Improvements over the MM obtained for test dimers are as large as 60 kJ/mol, which brings the $E_{es}$(EP/MM) values based on pseudoatoms within 10–20 kJ/mol (usually <10 kJ/mol) of energies from the Morokuma–Ziegler decomposition scheme based on ADF/BLYP/TZP calculations. Pseudoatom partitioning of the electron density in direct space is likely to reduce the remaining discrepancies with the Morokuma–Ziegler energy decomposition scheme.

In general, the integration part of EP/MM takes about 1 s on an Athlon XP 2000+ processor for each pseudoatom-pseudoatom interaction using 50 radial and 194 angular quadrature points. This speed will not change for more complex systems because only two pseudoatom expansions are needed for each pairwise interaction. The significant gain in speed in evaluation of $E_{es}$ compared to the EP, combined with much higher accuracy in $E_{es}$ than can be achieved with the MM, strongly supports use of EP/MM in large molecular calculations.

Compared with the SCDS method [33], which is also more computationally intensive, EP/MM does not require a pre-calculated grid – only the analytically defined pseudoatoms, either from aspherical refinement of theoretical structure factors or extracted from the library of aspherical pseudoatoms.

Any electron density that can be represented by a nucleus-centered expansion can be used with EP/MM, including electron densities derived from accurate experimental X-ray diffraction data. Experimental densities may in principle yield more accurate $E_{es}$ than those constructed from the databank since they are based on the measurements of a physically-observable property and in principle include the effect of intermolecular interactions in the crystal. However, bias in the experimental estimates, due to the lack of phase information, approximations made in the thermal motion model, and systematic experimental errors, may significantly affect the accuracy of the results.

Finally we want to point out that EP/MM can also be applied for calculation of intramolecular electrostatic interaction energies, as the electron–nuclear attraction and electron–electron repulsion integrals are accurately determined. In fact, with the addition of the exchange-correlation integral it is possible to obtain the total intramolecular potential energy of a molecule, the electron density of which is represented by a nucleus-centered pseudoatom expansion.

Further improvements of EP/MM may include incorporation of different critical internuclear distances for particular types of interacting pseudoatoms, i.e. for example for O···C, O···N, etc., and possibly more efficient quadrature integration methods. This would further improve the speed of the calculation without sacrificing the accuracy.

Acknowledgements

We thank Xue Li (SUNY at Buffalo, USA) for help in constructing the pseudoatom databank and Dr. Stan van Gisbergen (Vrije Universiteit, Amsterdam) for
valuable suggestions concerning the ADF program and the evaluation of integrals. Support by the National Institutes of Health (GM56829) and the National Science Foundation (CHE0236317) is gratefully acknowledged.

References

[29] SYBYL 6.7, Tripos Inc., 1699 South Hanley Road, St. Louis, Missouri, 63144, USA.