# How Accurate Can a Force Field Become? A Polarizable Multipole Model Combined with Fragment-wise Quantum-Mechanical Calculations

**Pär Söderhjelm* and Ulf Ryde**

*Department of Theoretical Chemistry, Lund University, Chemical Center, POB 124, SE-22100 Lund, Sweden*

*Received: August 17, 2008; Revised Manuscript Received: October 31, 2008*

A new method to accurately estimate the interaction energy between a large molecule and a smaller ligand is presented. The method approximates the electrostatic and induction contributions classically by multipole and polarizability expansions, but uses explicit quantum-mechanical fragment calculations for the remaining (nonclassical) contributions, mainly dispersion and exchange repulsion. Thus, it represents a limit of how accurate a force field can ever become for interaction energies if pairwise additivity of the nonclassical term is assumed (e.g., all general-purpose force fields). The accuracy is tested by considering protein−ligand model systems for which the true MP2/6-31G* interaction energies can be computed. The method is shown to be more accurate than related fragmentation approaches. The remaining error (2−5 and ∼10 kJ/mol for neutral and charged ligands, respectively) can be decreased by including the polarizing effect from surrounding fragments in the quantum-mechanical calculations.

## 1. Introduction

In many applications of theoretical methods in chemistry, one is interested in changes in the potential energy. Although there are many empirical potentials developed for specific types of systems, only quantum-mechanical (QM) methods are generally applicable. By these methods, one can in principle attain any accuracy by using a sufficiently large basis set and including electron correlation in a rigorous way. In practice, however, the applicability of the most accurate methods is limited by computational resources. Therefore, there is a need for general potential energy methods with low and predictable error as compared to the exact QM treatment, but still applicable to large molecular systems.

One solution to this problem is to decompose the system into smaller subsystems, fragments that are treated more or less independently. Such fragmentation approaches to QM calculations have a long history,[1−5] but they have experienced a strong revival in the past decade. Most such methods estimate the desired property (e.g., total energy, interaction energy, or electron density) by the formally exact expansion into monomer contributions, two-body contributions, three-body contributions, etc., up to *n*-body contributions, where *n* is the number of fragments. For the method to give a significant computational advantage over the supermolecular calculation of the whole system, the series has to be truncated already after the two- or three-body term. Besides the choice of how to truncate the series, the various methods differ mainly in the treatment of fragmentation across covalent bonds, the selection of subsystems, and the use of embedding to capture some of the many-body effects that are lost in the truncation.

For the fragmentation, several schemes have been developed. The fragment molecular orbital (FMO) method[6] defines a set of nonoverlapping fragments and assigns a number of electrons to each fragment, using local molecular orbitals. A simpler approach is to assign a number of nuclei to each fragment and handle empty valencies by capping with hydrogen atoms or other functional groups. By necessity, the fragments will then overlap. However, on the basis of the approximate atomwise additivity of interaction energies,[7] methods for handling the overlaps by adding and subtracting fragment energies have been proposed.[8−11] The latter model, called molecular fractionation with conjugate caps (MFCC), has later been used with various capping groups, and more sophisticated and automatic procedures for the fractionation have been developed.[12−14] There are also related methods limited to noncovalently bonded clusters, in which the fragmentation is trivial.[15−18]

The selection of subsystems (i.e., the specific supermolecular calculations performed) depends on the quantity to be computed. The FMO method, which exists as a two-body (FMO2) or a three-body expansion (FMO3),[19] computes the total energy and thus requires calculation of all fragment pairs. The MFCC method, on the other hand, is primarily designed for interaction energies with fixed monomer geometries, and thus only fragment pairs belonging to different monomers are computed. The method has also been adapted to calculation of total energies.[20] More elaborate ways of selecting subsystems have been used for clusters,[16] as well as for covalently bonded fragments.[14]

For the embedding, there have also been many proposals. In FMO, each monomer and dimer experiences the exact electric potential from all other fragments in the system. This requires the computation of two-electron repulsion integrals between the fragments, but it has been shown that outside a certain distance, the potential may be approximated by Mulliken charges without loss of accuracy.[21] A simpler alternative to FMO is the binary interaction method[15,22] (or ternary interaction, in analogy to FMO3), in which the potential is approximated by the potential from fragment-centered dipoles or atom-centered electrostatic potential (ESP) charges, giving special attention to the basis-set superposition error (BSSE). In the original MFCC method, there is no embedding; the total interaction energy is simply a sum of pairwise interaction energies. For calculating the total energy by MFCC, embedding using unit charges, Mulliken charges, ESP charges, or natural population analysis (NPA) charges has been used.[14,23]

* Corresponding author. Phone: +46 46 2224502. Fax: +46 46 2224543. E-mail: par.soderhjelm@teokem.lu.se.

10.1021/jp8073514 CCC: $40.75 
Published on Web 12/18/2008

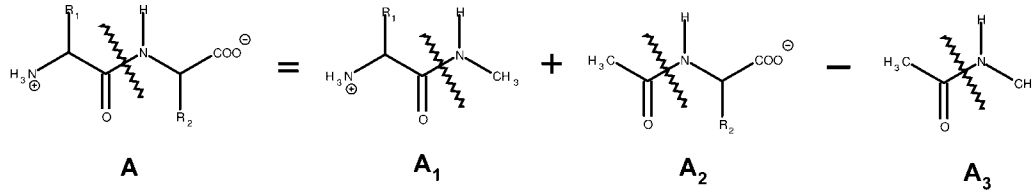


**Figure 1.** Example illustrating the MFCC procedure for cutting molecule A across the peptide bond and capping with $-COCH_3$ and $-NHCH_3$ groups. The result is two capped fragments $A_1$ and $A_2$ with $c_i = 1$, as well as a concap fragment $A_3$ with $c_i = -1$.

In principle, all of these methods are applicable to any level of theory, although most work has been done using Hartree−Fock theory (HF), density functional theory (DFT), or second-order Møller−Plesset perturbation theory (MP2). A fragmentation method may also be combined with a full-system calculation at a lower level of theory (hybrid approach), for example, the MP2−HF[24] or CCSD(T)−MP2[18] combinations.

A consistent series of methods was recently tested by Truhlar et al.[17,24] for water clusters. These include the pairwise additive (PA) approximation, which can be seen as a particular case of MFCC with no covalent bonds between fragments; the electrostatically embedded pairwise additive (EE-PA) approximation, which is similar to FMO2 except for the use of an approximate potential (generated by Mulliken charges) at all distances; and the MP2−HF hybrid versions of these methods (denoted by a CE extension for correlation energy). We will in the following adopt this notation and reserve the term MFCC for the actual fragmentation procedure.

An alternative to the fragmentation approach is to decompose the energy into terms with different physical meaning. If we limit ourselves to noncovalent interactions, the most common description defines four such terms: electrostatic energy, induction energy, dispersion energy, and repulsion energy.[25] The first two terms are usually called classical terms, whereas the latter two are nonclassical terms, stemming from the quantum-mechanical nature of the interaction. The decomposition is ambiguous, and several schemes have been devised.

Apart from permitting the direct evaluation of each term separately,[26,27] such physical decomposition of the interaction energy is the foundation of molecular mechanics force fields, in which each term is estimated by a relatively simple expression. Normally, the parameters in the classical terms (e.g., atomic charges and polarizabilities), as well as those in the nonclassical terms (e.g., Lennard-Jones parameters), are part of the force field itself. They are usually obtained by a major parametrization involving systems of the type for which the force field is designed.

There are also several examples of methodologies in which only the nonclassical parameters are predefined, whereas the classical parameters are obtained for each considered system, typically by performing QM calculations and analyzing the obtained electron density.[28] Although very appealing in theory, the accuracy of this approach is limited by the transferability of the nonclassical parameters, which are normally fitted to reproduce supermolecular energies. Fitting of the exchange repulsion is a difficult problem, which has been addressed in many studies.[28−32] The transferability is improved if the overlap of the wave functions (or densities) is explicitly taken into account.[33−35] By this approach, good results may be obtained without fitted parameters[36] or with a small number of element-independent parameters.[37] The dispersion energy has also been subject to many studies[38,39] with the aim to avoid parameter fitting as much as possible. Any fitting to supermolecular QM calculations or experimental data will include model errors in the classical terms (e.g., in the multipole approximation). This will necessarily introduce unphysical effects into the parameters and restrict their transferability.

To avoid the transferability problem and address the accuracy of the actual physical decomposition, we will go one step further in this study and estimate the nonclassical terms by supermolecular QM calculations, although for smaller subsystems. The advantage of this is 2-fold: No fitted parameters are needed, and the accuracy is expected to be improved. In fact, the pair-potentials are by definition exact (within the given QM methodology applied), so the only approximation in the method is the assumption of pairwise additivity of the nonclassical term. We call the method polarizable multipole interaction with supermolecular pairs (PMISP) to highlight that it is based on a classical (polarizable multipole) interaction model, but enhanced with supermolecular dimer energies.

At the same time, the PMISP method is directly related to the fragmentation methods. It uses a two-body expansion and the MFCC procedure for handling covalently bonded fragments. The use of a polarizable multipole description of the whole system replaces the need of embedding the QM calculations in an electrostatic field; both approaches capture the most important many-body effects. PMISP may in fact be seen as a hybrid fragmentation method, using a polarizable multipole description as the lower level of theory and any QM method as the higher level.

## 2. Methods

**2.1. The PMISP Method.** We consider the interaction between a large molecule A (typically a model of a protein binding site) and a small molecule B, in vacuum. These two molecules will be denoted monomers. The geometries of the isolated monomers are kept fixed as the dimer is formed, a common approximation in ligand-binding calculations.[40] For a given (necessarily size-extensive) method and basis set, the QM (supermolecular) interaction energy between A and B, corrected for basis set superposition error (BSSE) by the counterpoise procedure, is defined by

$$E_{\mathrm{ref}} = E_{\mathrm{AB}}^{\mathrm{sup}} = E_{\mathrm{A+B}} - E_{\mathrm{A+(B)}} - E_{\mathrm{B+(A)}} \quad (1)$$

where $E_{\mathrm{X+(Y)}}$ denotes the energy of monomer X in the dimer basis set. Note that, throughout this study, we adopt the notation $E_{\mathrm{XY}}$ to mean the interaction energy between X and Y, that is, not the total energy.

In the PMISP method, $E_{\mathrm{ref}}$ is estimated by the following expression:

$$E_{\mathrm{AB}}^{\mathrm{PMISP}} = E_{\mathrm{AB}}^{\mathrm{ele}} + E_{\mathrm{AB}}^{\mathrm{ind}} + E_{\mathrm{AB}}^{\mathrm{nc}} \quad (2)$$

where $E^{\mathrm{ele}}$ and $E^{\mathrm{ind}}$ are the electrostatic and induction energies, respectively, when both monomers are treated classically, using multipoles and polarizabilities, and $E^{\mathrm{nc}}$ is a nonclassical term, mainly containing the dispersion and exchange-repulsion energies, but also corrections to the classical terms (e.g., charge

penetration), as well as various coupling terms and corrections for the artificial division of the system into monomers (usually denoted charge transfer).

Unlike in standard molecular-mechanics force fields, $E^{nc}$ is not estimated by an expression with fitted parameters. Instead, it is obtained by splitting A into fragments $A_1$, $A_2$, ..., $A_n$ and evaluating the contribution to the nonclassical term from each fragment separately, assuming pairwise additivity. In this work, we employ the MFCC fragmentation procedure,[9] which is a well-defined and general way to treat fragmentation over covalent bonds. In this method, the fragments are capped with chemically suitable functional groups. Moreover, the capping groups on each side of a bond that is cut are joined to form a concap fragment. An example is shown in Figure 1. The key feature of this procedure is that by adding the sets of atoms in the normal fragments and subtracting the sets of atoms in the concap fragments, one recovers the molecule A, and this additivity is expected to hold approximately for certain properties and energies. Thus, we define $E^{nc}$ by

$$E_{AB}^{nc} = \sum_{i=1}^{n} c_i (E_{A_iB}^{sup} - E_{A_iB}^{ele} - E_{A_iB}^{ind}) \quad (3)$$

where $c_i$ is equal to 1 for a normal fragment and $-1$ for a concap fragment.

The electrostatic and induction energies are calculated by representing each molecule as a collection of multipoles and anisotropic dipole polarizabilities, located at each nuclear position and each covalent bond midpoint. These properties are computed at any QM level by the LoProp method.[41] The multipole expansion is truncated after quadrupoles. The electrostatic interaction includes all possible terms formed by the multipoles, that is, up to and including quadrupole−quadrupole interactions. The accuracy of the LoProp method has been tested before.[42] Including octupoles in the multipole expansion changes the total energies in this study by less than 1 kJ/mol.

In the same spirit as for the calculation of the nonclassical term, the properties of monomer A may be computed fragment-wise to reduce the computational time. To do this, we again apply the MFCC procedure (in terms of electron densities)[43] and estimate the properties of A as

$$P_k^A = \sum_{i=1}^{n} c_i P_k^{A_i} \quad (4)$$

where $P_k^{A_i}$ is a multipole moment or polarizability located at center $k$ (which may be either a nucleus or a bond midpoint), obtained by a QM calculation of fragment $A_i$ (or zero if center $k$ is not within $A_i$). The properties of capping hydrogen atoms are almost perfectly canceled between caps and concaps; the remaining part is moved to the corresponding real atom. Alternative methods for assembling multipoles have been discussed previously.[44]

**2.2. Intramolecular Polarization.** The use of a polarizability model is particularly approximate for treating interactions within a covalently bonded molecule (intramolecular polarization). To avoid polarization catastrophes and obtain a physical behavior of the model, a polarizable force field normally includes a rule for which interactions are neglected or damped (typically between close-lying centers). If one aims at a high accuracy, special care is required when devising this rule. Thus, in our method, the following rule is applied: Polarization is included between all centers, except those pairs of centers that have been in the same fragment in at least one LoProp calculation.

To motivate this rule, we distinguish between polarization due to static multipoles (static polarization) and polarization due to other induced dipoles (polarizability coupling). For the static polarization, the rule can be rigorously justified. First, we note that no matter how the fragments are chosen, each center $k$ belongs to an odd number $n_k$ of fragments, or specifically to one more normal fragment than concap fragment. The case with $n_k = 1$ is trivial: The polarizability $\alpha_k$ should not respond to multipoles in the same fragment, because these have already influenced the static multipole at $k$ through the self-consistent quantum-chemical treatment. For $n_k = 3$, the multipole at $k$ has, through the sum over three fragments, been effectively influenced by the union of the two fragments (with the concap fragment removing double-counting), and thus $\alpha_k$ should not respond to multipoles in any of these fragments. The same argument can be used for higher values of $n_k$, thus giving the simple rule stated above, which can be automatically applied in the calculations. A similar discussion in the context of flexible molecules can be found in ref 45.

For the polarizability coupling, the justification is more qualitative. In principle, a given polarizability $\alpha_k$ should not respond to induced dipoles in the same fragment, because the magnitude of $\alpha_k$ is derived assuming that the whole fragment is polarized simultaneously. One may consider it as an "implicit coupling" manifested through the numerical values of the distributed polarizabilities. Including the coupling explicitly (i.e., letting $\alpha_k$ respond to the field from the other induced dipoles when using the model) would mean that the total polarizability of the fragment would be higher than that obtained in the quantum-chemical calculation. For $n_k > 1$, an argument similar to that above can be applied, resulting in the same rule as for the static polarization. However, the polarizabilities in the LoProp approach are derived using homogeneous electric fields. Therefore, the implicit coupling is strictly correct only if the field is homogeneous (so that the other polarizabilities respond to the same field), but approximate when the field is inhomogeneous. This approximation was recently found to be accurate for small and medium-sized molecules.[46] The present study will provide a test for larger molecules.

**2.3. Systems.** For testing the method, we use a model of the avidin protein interacting with the seven ligands (biotin analogues) shown in Figure 2. This system (with the full protein) has previously been subject to several studies.[47−50] For each ligand, we obtain geometries from snapshots of a simulation of the protein−ligand complex in explicit water, using the Amber 1994 force field.[51] The exact simulation protocol has been described elsewhere.[50] To draw statistically valid conclusions, the first 10 snapshots (separated by 10 ps simulation time) are used for BTN1. This set of 10 geometries will be called the geometry set. Some calculations are performed only for the first snapshot of BTN1. This will be called the main structure. For the remaining ligands, only the first snapshot is used. The first snapshot of each of the seven ligands will be called the ligand set.

The model of the avidin active site (denoted as monomer A) consists of 216 atoms and is shown schematically in Figure 3. The residue numbering refers to PDB structure 1AVD,[52] and the whole model belongs to the same subunit (labeled B in the original structure), except for Trp-110, which belongs to a neighboring subunit. Note that the model is not a single covalently bonded molecule, but rather a collection of 15 separate molecules of different sizes, labeled by $A_1$, $A_2$, ..., $A_{15}$. This is a consequence of the way that the model was constructed, by including all atoms within 4 Å of the ligand (using the

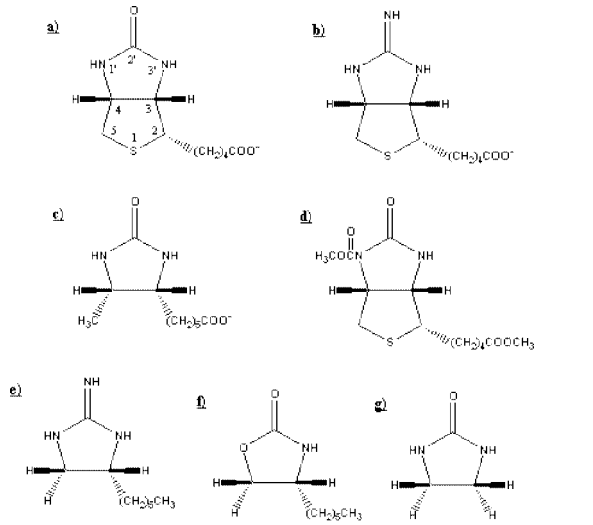

**Figure 2.** The seven ligands to avidin used in this study: (a) BTN1 (biotin), (b) BTN2, (c) BTN3, (d) BTN4, (e) BTN5, (f) BTN6, (g) BTN7. The first three have a molecular charge of −1, whereas the other ligands are neutral when bound to the protein.

coordinates of the main structure) and a minimal set of additional atoms necessary to complete chemically reasonable groups such as aromatic rings. Bonds that were cut during this process were capped with hydrogen atoms. As can be seen in Figure 3, all molecules are rather small, except for $A_5$, which is a chain of six amino acids. This irregular distribution of molecule sizes is unintentional, but happens to be advantageous for some of the tests performed.

The fragmentation of monomer A is done in three different ways (cutting schemes): (a) The whole model A is treated as one fragment. (b) Each of the 15 molecules in A is treated as one fragment. (c) The largest molecule $A_5$ is further divided into 6 capped fragments and 5 concap fragments.

The cuts are done through the peptide bonds, as indicated in Figure 3, and each fragment is capped with $-COCH_3$ and $-NHCH_3$ groups at the N and C termini, respectively.

Note that the word fragmentation is used in a wide sense for splitting the whole monomer A into smaller pieces, no matter whether the pieces are covalently bonded or not (covalent bonds are only cut in scheme c). The cutting schemes apply to both the computation of properties and the nonclassical term, although scheme a makes no sense for the nonclassical term (it would require the quantity $E_{AB}^{sup}$ that we are trying to approximate). In practice, scheme c is the only computationally feasible option for calculating the nonclassical term at a reasonably high level of theory. For a full protein, it corresponds to letting each amino acid residue constitute one fragment. Although there are advantages of using the same cutting scheme for the nonclassical term as for the properties, we also investigate other possibilities.

**2.4. Separating Electrostatic and Induction Energies.** When comparing electrostatic and induction contributions to interaction energies between various polarizable force fields (or cutting schemes), a technical issue arises concerning the definition of each term. The most natural definition of electrostatic energy is the static interaction between the monomers, each having been internally prepolarized before the interaction starts. The induction energy is then the energy change (always negative) caused by the polarization of A by B, and vice versa. This definition is illustrated in the upper row of Figure 4, where the dashed line in state 1 indicates that the interaction is turned off, the random arrows inside the monomers in states 1 and 2 indicate that the monomers are internally prepolarized, and the bigger arrows in state 3 indicate that the whole dimer is self-consistently polarized.

However, a more direct definition simply distinguishes between the interaction involving only multipoles (i.e., obtained by turning off the polarizabilities) and the rest of the energy (caused by the polarizabilities). These are the quantities obtained directly from a standard molecular-mechanics program, and we denote them multipole energy ($E^{mult}$) and polarizability energy ($E^{pol}$). The definition is illustrated in the lower part of Figure 4, where the absence of arrows in states 4 and 5 indicates that the monomers are not prepolarized. If we make the logical definition

$$E_{AB}^{pol} = E_{total}^{pol} - E_{A}^{pol} - E_{B}^{pol} \tag{5}$$

then it is evident from Figure 4 that

$$E_{AB}^{ele} + E_{AB}^{ind} = E_{AB}^{mult} + E_{AB}^{pol} \tag{6}$$

although, in general, the equality does not hold termwise. Thus, as long as the total result is concerned, it does not matter which pair of quantities we use. However, to be able to compare individual terms from various methods, it is essential to use the natural definition. In practice, one can obtain $E_{AB}^{ele}$ by performing a separate calculation of each monomer, saving the induced dipoles, and treating these as static dipoles in the subsequent dimer calculation, taking care not to double-count the internal polarization.

In Table 1, we report numerical values for the quantities depicted in Figure 4 for the main structure using each of the three cutting schemes and HF properties. The difference between the natural electrostatic energy ($E^{ele}$) and the direct multipole energy ($E^{mult}$) is 18−24 kJ/mol when A is split in the property calculation (this difference is exactly canceled by the corresponding $E^{ind} - E^{pol}$ difference). As can be seen, the $E^{mult}$ energy differs significantly between the various cutting schemes, whereas the $E^{ele}$ energy is almost constant.

This discussion is quite general and applies for interaction energies computed with any polarizable force field. To emphasize this point, the corresponding values for the Amber 2002 polarizable force field[53] are also given in Table 1. Interestingly, the direct use of the $E^{mult}$ and $E^{pol}$ terms, as obtained from the AMBER program, gives the impression that the polarization is negligible for this interaction (for some geometries, $E^{pol}$ actually becomes positive). However, the polarizability energy depends on details in the description of intramolecular polarization and has no physical significance (for example, an internally induced dipole that is subjected to an oppositely directed field from the other monomer gives a positive contribution to the polarizability energy). The true induction energy ($E^{ind}$) is attractive and substantial.

Having introduced the direct terms, we note that there is another useful interpretation of the PMISP energy. Inserting eq 3 into eq 2, applying eq 6, and reordering the terms gives

$$E_{AB}^{PMISP} = \sum_{i=1}^{n} c_i E_{A_iB}^{sup} + \left( E_{AB}^{pol} - \sum_{i=1}^{n} c_i E_{A_iB}^{pol} \right) \tag{7}$$

where we have used eq 4 together with the inherent pairwise additivity of the multipole interactions to cancel the $E^{mult}$ terms. Note that the cancelation only occurs if the same cutting scheme is used for the calculation of the nonclassical term and the properties. In eq 7, we recognize the first term as the pairwise additive (PA) interaction energy and the second term as the many-body contribution to the polarizability energy. Thus, one may regard the PMISP energy as an improvement of the

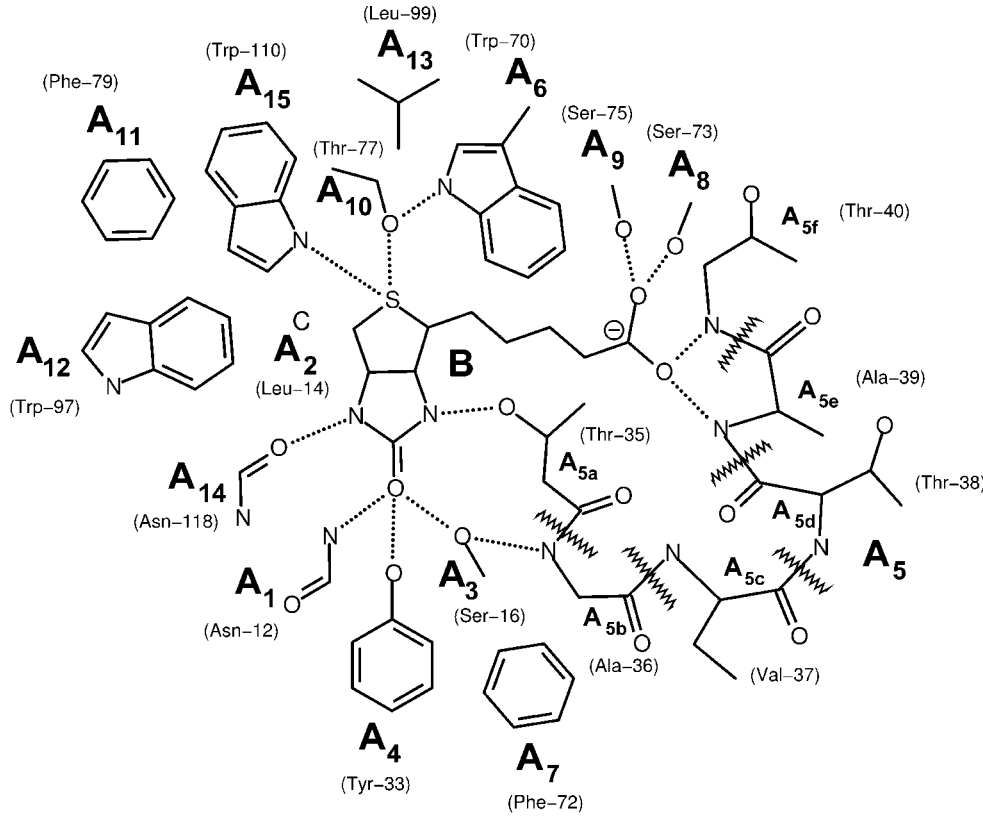


**Figure 3.** Two-dimensional cartoon of the avidin model interacting with biotin (BTN1). It gives a guidance to the location of each fragment; in reality, the fragments surround the ligand completely. For clarity, hydrogen atoms are omitted. The most prominent hydrogen bonds are indicated by dotted lines. The fragments of the model (in cutting scheme b) are labeled from $A_1$ to $A_{15}$, and the additional fragmentation of $A_5$ into fragments $A_{5a}-A_{5g}$ (in cutting scheme c) is also indicated. The avidin residue from which each fragment is derived is shown in brackets. The biotin molecule is labeled B.

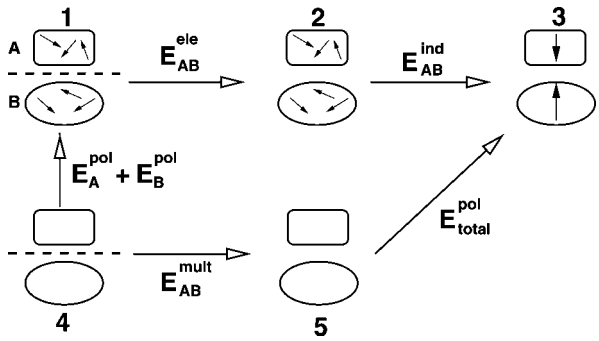


**Figure 4.** Schematic picture of the relations between classical contributions to the interaction energy.

**TABLE 1: Electrostatic and Induction Energies for the Model Complex in the Main Structure, Calculated by PMISP with HF Properties Obtained by Cutting Schemes a, b, and c, as well as by the Amber 2002 Force Field**[a]

| | PMISP | | | Amber |
|---|---|---|---|---|
| | a | b | c | ff02 |
| $E_{AB}^{ele}$ | −419.4 | −420.5 | −419.1 | −403.4 |
| $E_{AB}^{ind}$ | −164.6 | −154.5 | −153.4 | −98.7 |
| $E_{AB}^{mult}$ | −419.4 | −438.6 | −442.7 | −499.9 |
| $E_{AB}^{pol}$ | −164.6 | −136.3 | −129.8 | −2.2 |
| $E_{A}^{pol}$ | 0.0 | −27.1 | −32.3 | −169.9 |
| $E_{B}^{pol}$ | 0.0 | 0.0 | 0.0 | −30.8 |
| $E_{total}^{pol}$ | −164.6 | −163.4 | −162.1 | −202.9 |

[a] Various other quantities (cf., Figure 4) are also given. Energies are in kJ/mol.

PA energy by approximating the many-body energy (which is absent in PA) by the many-body energy from a polarizable multipole description.

**2.5. Computational Details.** The multipoles and polarizabilities were obtained by the LoProp method[41] as implemented in MOLCAS.[54−56] The default settings were used, except for the largest clusters (with 216 atoms), for which we found that the constant α in the penalty function for converting the charge flow to polarizabilities (eq 17 in ref 41) had to be reduced from 7.1 to 2.0 to avoid numerical instability. For MP2, properties were obtained by using the linear-response charge density. Calculating this density is similar in effort to a gradient evaluation, and thus takes significantly more time than an MP2 energy evaluation. For these calculations, the MOLPRO program[57] was used to generate the density needed by LoProp.

The supermolecular calculations were also performed with MOLCAS. The Cholesky decomposition (CD) approximation to the two-electron integrals[58,59] was applied in combination with the local exchange (LK) algorithm.[60] On the basis of a previous analysis of the accuracy of the CD approximation,[60−62] a decomposition threshold of $10^{-4}$ was used in all calculations. The largest effect of the CD approximation was seen in the calculations of $E_{ref}$, in which the error in the interaction energies was up to 2 kJ/mol.

The comparison with other methods was done at the HF level using cutting scheme b. The PA energy is the first term in eq 7 and is thus obtained as a part of the PMISP procedure. To obtain the FMO and EE-PA energies, we used the FMO procedure in the GAMESS program package[63] to calculate the total energy of the AB dimer and monomer A, respectively, whereas the energy of monomer B was obtained by a standard HF calculation. For FMO2, we used the very rigorous settings RESPAP

= 0, RESPPC = 0, and RESDIM = 4.0, and for FMO3, we additionally used RITRIM = 2.0. Tests performed for the main structure showed that the more approximate settings suggested in ref 19 (RESPPC = 2.0 and RESDIM = 2.0) could be used without influencing the result with more than 0.1 kJ/mol. However, applying the previously proposed RESPAP = 1.0 approximation changed the result by several kJ/mol. For the EE-PA energy, we used RESPAP = 0, RESPPC = −1, and RESDIM = 2.0. Thus, the only difference between FMO2 and EE-PA was the use of the exact electrostatic potential in the former and the potential from Mulliken charges in the latter, independently of the distance.

## 3. Results

**3.1. Demonstration of the PMISP Method.** The PMISP method was tested by performing calculations of the interaction energy between a model of avidin (216 atoms, denoted A, see Figure 3) and biotin-like ligands (12−41 atoms, denoted B, see Figure 2), in various geometries. To enable a comparison with the exact supermolecular results, the 6-31G* basis set was used. Using this basis set, the supermolecular interaction energies for the main structure are −252 kJ/mol at the HF level and −412 kJ/mol at the MP2 level. The BSSE is substantial, 105 and 214 kJ/mol, respectively, at the HF and MP2 levels. This indicates that the supermolecular results are very far from the basis set limit, but they still provide a reference for testing approximations within the same basis set.

The difference between the MP2 and HF reference energies is mainly due to dispersion, which is entirely missing in the HF result, but included in the MP2 result. In fact, the dispersion energy (at the MP2 level of theory) is even larger than the energy difference suggests, because the electron correlation also affects the electrostatic and induction energies, in this particular system reducing the attraction by about 50 kJ/mol. For this reason, it is advantageous to use the same level of theory in the supermolecular calculations as in the calculation of properties. Although we are mainly interested in the MP2 results, a separate set of calculations is performed at the HF level. In addition to enabling separate tests of the PMISP approximation at the HF level and for the correlation contributions, the lower cost of the HF property calculations allows us to investigate the influence of the applied fragmentation scheme.

For the main structure, the classical energy is ∼−570 kJ/mol at the HF level (see Table 1) and ∼−520 kJ/mol at the MP2 level (Table S1 in the Supporting Information). The large attraction is partly canceled by the repulsive nonclassical term ($E^{nc}$), being ∼310 and ∼100 kJ/mol at the HF and MP2 levels, respectively, resulting in total PMISP energies within 10 kJ/mol of the reference energies. The contribution from each fragment to the nonclassical term for the main structure is reported in Table 2. It can be seen that the contributions vary substantially among the various fragments. This reflects the difference in size as well as in overlap. With a few small-magnitude exceptions, the fragment contributions at the HF level are always positive, as can be expected because they contain mainly exchange repulsion. The presence of negative terms is a reminder of the approximations involved in the polarizable multipole description of the interactions (e.g., neglect of charge-penetration effects). At the MP2 level, the sign of each contribution varies, because of the balance between exchange repulsion and dispersion. For each fragment interaction, the difference between the nonclassical terms at the MP2 and HF levels is a good estimate of the dispersion energy.

**3.2. Accuracy of the PMISP Method.** To obtain statistics regarding the accuracy of PMISP, we use 10 different geometries of the BTN1 interaction (geometry set), and seven different ligands (Figure 2), each in a different geometry (ligand set). In all of these calculations, the atoms in the avidin model remain the same, but its geometry changes significantly. Detailed results for all terms, all considered cutting schemes, and all geometries are given as Supporting Information (Tables S1 and S2).

**TABLE 2: Contributions to the Nonclassical Term from Each Fragment for the Main Structure Using the PMISP Method (with Cutting Scheme b) at the MP2 and HF Levels, and the Embedded PMISP Method at the HF Level (EMB), as well as the Mean Absolute Contributions to the Difference between the Embedded and Standard PMISP Methods for the Geometry Set (Gset) and Ligand Set (Lset) with the Contributions from Polarizabilities within Brackets[a]**

| fragment | MP2 | HF | EMB | Gset | | Lset | |
|---|---|---|---|---|---|---|---|
| $A_1$ | −1.9 | 0.8 | 2.0 | 1.7 | (0.5) | 1.9 | (0.4) |
| $A_2$ | −1.6 | −0.1 | −0.2 | 0.1 | (0.1) | 0.1 | (0.1) |
| $A_3$ | 18.6 | 27.0 | 29.7 | 3.4 | (0.4) | 3.3 | (0.3) |
| $A_4$ | 24.0 | 38.1 | 41.5 | 3.3 | (0.7) | 3.4 | (0.7) |
| $A_5$ | 11.0[b] | 65.9 | 67.5 | 1.3 | (0.5) | 0.9 | (0.5) |
| $A_6$ | −6.2 | 26.4 | 24.2 | 1.3 | (0.5) | 1.0 | (0.2) |
| $A_7$ | −1.7 | 6.7 | 6.2 | 0.4 | (0.2) | 0.3 | (0.1) |
| $A_8$ | 59.8 | 72.4 | 74.5 | 3.2 | (0.9) | 1.1 | (0.3) |
| $A_9$ | −1.9 | 5.7 | 8.5 | 3.1 | (0.6) | 1.2 | (0.2) |
| $A_{10}$ | 16.2 | 27.8 | 28.4 | 0.6 | (0.2) | 0.3 | (0.2) |
| $A_{11}$ | −3.3 | 3.2 | 3.3 | 0.3 | (0.1) | 0.1 | (0.0) |
| $A_{12}$ | −7.2 | 8.5 | 8.0 | 0.6 | (0.1) | 0.3 | (0.1) |
| $A_{13}$ | −5.4 | 3.3 | 3.8 | 0.5 | (0.3) | 0.3 | (0.1) |
| $A_{14}$ | 1.2 | 7.3 | 6.8 | 0.4 | (0.3) | 0.7 | (0.3) |
| $A_{15}$ | −0.9 | 19.9 | 18.7 | 0.9 | (0.1) | 0.7 | (0.1) |
| total | 100.6 | 312.9 | 322.9 | 12.8 | (1.9) | 8.8 | (1.2) |

[a] Energies are in kJ/mol. [b] Summed result using cutting scheme c.

**TABLE 3: Mean Absolute Difference for the Geometry Set (Gset) and Ligand Set (Lset) between Classical Energies Calculated from Properties Obtained Fragment-wise (Scheme b or c) and for the Whole Monomer A (Scheme a)[a]**

| | scheme b | | scheme c | |
|---|---|---|---|---|
| | Gset | Lset | Gset | Lset |
| $E_{AB}^{ele}$ | 1.4 | 0.5 | 1.2 | 0.8 |
| $E_{AB}^{ind}$ | 11.7 | 6.9 | 14.0 | 8.0 |
| $E_{AB}^{ind,b}$ | 1.5 | 0.8 | 1.6 | 1.0 |

[a] Energies are in kJ/mol. [b] Using the same polarizability coupling as in scheme a.

Let us first consider the influence of the fragmentation on the computed properties and thereby on the classical energies. Using the results with scheme a (i.e., obtaining the properties in a single QM calculation) as the reference, the mean absolute differences in electrostatic and induction energies for schemes b and c are given in Table 3 for the geometry and ligand sets. The results confirm that the three cutting schemes give almost identical electrostatic energy, with mean absolute differences of ∼1 kJ/mol. This is a remarkably accuracy, considering that the b and c calculations involve 15 and 25 fragments, respectively, and that errors from each cut could accumulate. Moreover, the electrostatic energy in schemes b and c depends not only on the multipoles, but also on the polarizabilities and the definition of excluded multipoles, as discussed in section 2.4. The fact that the errors are not higher with scheme c than with scheme b indicates that cutting a covalent bond does not introduce a larger error than fragmenting across, for example, a hydrogen bond. The explanation for this is probably that the quantum-mechanical error of confining each wave function to one fragment is to a large extent eliminated by the MFCC

**TABLE 4: Error As Compared To the Reference Energy for the Main Structure, as well as the Mean Signed Error (MSE), Standard Deviation (SD), and Mean Absolute Error (MAE) for the Geometry and Ligand Sets[a]**

| | | geometry set | | | ligand set | | |
|---|---|---|---|---|---|---|---|
| method | main | MSE | SD | MAE | MSE | SD | MAE |
| | | | HF Level | | | | |
| PMISP (ac) | −19.6 | −24.9 | ±4.6 | (24.9) | −12.4 | ±14.2 | (12.5) |
| PMISP (bc) | −10.6 | −14.0 | ±3.4 | (14.0) | −5.5 | ±8.8 | (7.3) |
| PMISP (cc) | −8.1 | −10.0 | ±2.5 | (10.0) | −3.7 | ±6.2 | (5.5) |
| PMISP (bb) | −10.0 | −10.9 | ±2.7 | (10.9) | −4.0 | ±6.9 | (5.8) |
| EMB-PMISP (bb) | 0.2 | 2.0 | ±2.6 | (2.6) | 5.0 | ±2.8 | (5.0) |
| PA (b) | −22.8 | −28.3 | ±7.0 | (28.3) | −6.9 | ±23.2 | (19.4) |
| FMO2 (b) | 9.7 | 13.6 | ±2.3 | (13.6) | 6.4 | ±3.3 | (6.4) |
| EE-PA (b) | 12.0 | 15.6 | ±2.2 | (15.6) | 8.0 | ±3.2 | (8.0) |
| | | | MP2 Level | | | | |
| PMISP (cc) | −6.7 | −9.0 | ±3.3 | (9.0) | −1.2 | ±8.4 | (6.9) |
| PMISP-CE (cc) | 1.5 | 1.0 | ±1.0 | (1.2) | 2.6 | ±2.6 | (3.3) |

[a] Results for several methods are given. The letters in brackets indicate the cutting scheme used for the classical ($E^{ele}$, $E^{ind}$) and nonclassical ($E^{nc}$) terms, respectively (if no such decomposition is used, only one letter is given). Energies are in kJ/mol.

procedure, so that the remaining error comes from imperfect description of the polarization, which is prominent in hydrogen bonds.

The situation is slightly different for the induction energy. Although the difference between schemes b and c is rather small, both results deviate significantly from scheme a with mean absolute differences of 7−14 kJ/mol. However, this discrepancy is not necessarily an error introduced by fragmentation of monomer A, but the reason could also be that the intramolecular polarizability coupling within a fragment is only treated implicitly, and the error of this approximation is expected to be largest for scheme a. The magnitude of this effect can be tested by using properties derived by scheme b or c, but ignoring the polarizability coupling within monomer A, as in scheme a. The static polarization within monomer A, on the other hand, is kept because that part is exactly modeled (by QM) in scheme a. As can be seen in Table 3, with this treatment of intramolecular polarization, the differences in induction energies are similar to those in electrostatic energies, that is, negligible for all practical purposes.

The total PMISP energy is an approximation to the supermolecular interaction energy ($E_{ref}$). The error of this approximation for the main structure, as well as the mean signed errors, mean absolute errors, and standard deviations for the geometry and ligand sets, are reported in Table 4. Because of the choice of cutting schemes for the individual terms, several combinations are computed, labeled by two letters, indicating the cutting scheme used for the computation of the classical terms ($E^{ele}$ and $E^{ind}$) and nonclassical term ($E^{nc}$), respectively. For computational reasons, the cc combination is most interesting, but within HF theory, we also test the use of larger fragments for the property calculations (ac and bc combinations), as well as the more consistent bb combination to investigate possible error cancelation.

If we fix scheme c for the nonclassical term and vary the scheme for the classical terms, we see that the MAEs become lower as the fragments are made smaller. Using the full monomer A to compute the properties (ac) gives a MAE of 25 kJ/mol for the geometry set. Using the separate molecules (bc) reduces the MAE to 14 kJ/mol, and cutting $A_5$ into fragments (cc) reduces the MAE even more to 10 kJ/mol. As we have already showed that the discrepancy in classical energy between the cutting schemes primarily reflects the differing treatment of intramolecular polarizability coupling, these results indicate that it is a severe approximation to rely on the implicit coupling in large molecules (as in scheme a).

The bb combination gives approximately the same MAE as the cc combination (i.e., lower than the bc combination). This is not a coincidence. A closer examination of the energies shows that the mean absolute difference between the two estimates is only 1 kJ/mol (maximum difference 2 kJ/mol), which is less than the differences in the individual terms. Thus, a systematic error cancelation occurs when using the same cutting scheme for all terms (note that this cancelation also includes the treatment of polarizability coupling). We cannot directly conclude which of the schemes bb and cc is the best one, but for practical reasons we prefer the more efficient approach (cc).

The errors are similar at the MP2 level. This might be expected, because the dispersion energy included in MP2 is pairwise additive (assuming that the molecular orbitals are localized on the fragments). However, it indicates that other types of correlation effects do not make significant contributions to the nonadditivity. This immediately suggests a MP2−HF hybrid method for estimating MP2 interaction energies, by the expression

$$E^{\text{PMISP-CE}} = E_{\text{ref}}^{\text{HF}} - E^{\text{PMISP,HF}} + E^{\text{PMISP,MP2}} \tag{8}$$

In analogy with earlier notation,[24] we call this approach PMISP-CE (where CE stands for correlation energy), and its errors are also reported in Table 4. With mean absolute errors of only 1−3 kJ/mol (with a maximum of 6 kJ/mol for BTN4), it appears to be an excellent approximation, useful in cases where the HF energy (but not the MP2 energy) can be obtained in a supermolecular calculation. The method differs from the previously proposed PA-CE method[24] by a term describing the change of the many-body polarization when correlation is included. As this term has an average magnitude of only 1 kJ/mol, the simpler PA-CE method is preferred, because no property calculations are needed.

**3.3. Analysis of the Error.** Having shown that the error of PMISP (i.e., the part of the many-body effects that cannot be captured by our classical model) is rather independent of the fragmentation scheme and already present at the HF level, it remains to discuss the origin of this effect. To this end, we restrict ourselves to the HF case using cutting scheme b, and follow three paths. First, we try to improve the PMISP method by understanding what physical effects are missing. Second, we compare our result with the corresponding result using other methods, which employ different sets of approximations. Third, we apply a brute force solution to the problem, which also gives useful insights into what effects are important.

The negative mean signed errors in Table 4 reflect that, for the negatively charged ligands, the attraction is systematically overestimated (i.e., the interaction energy is too negative) by the PMISP method. By our definition, this means that the sum of fragment-wise nonclassical terms is lower than the nonclassical term of the whole complex. Such nonadditivity of the nonclassical term can have several physical origins. First, the exchange-repulsion energy can be nonadditive.[64,65] Second, the surrounding molecules can influence the exchange repulsion of a given dimer by polarizing the dimer wave function in a way that increases the repulsion; that is, the coupling between exchange and induction can be nonadditive. Third, the polarization model can be inadequate, so that the nonclassical term does not represent the quantity that we would expect to be additive. Typically, the classical calculations tend to give a slight overpolarization,[46] and the sign of the PMISP error would then

indicate that this overpolarization is more prominent in the whole dimer than in the fragment interactions.

Although these effects can in principle be quantified by symmetry-adapted perturbation theory,[64,65] we take a more pragmatic approach. To test whether the effect of polarization from surrounding molecules on the individual $E^{\mathrm{nc}}_{\mathrm{A}_i\mathrm{B}}$ terms is significant, we define a new method, denoted embedded PMISP (EMB-PMISP), in which eq 3 is replaced by

$$E^{\mathrm{nc,EMB}}_{\mathrm{AB}} = \sum_{i=1}^{n} \left( E^{\mathrm{sup}}_{\mathrm{A}_i\mathrm{B}}(\bar{A}_i) - E^{\mathrm{ele}}_{\mathrm{A}_i\mathrm{B}}(\bar{A}_i) - E^{\mathrm{ind}}_{\mathrm{A}_i\mathrm{B}}(\bar{A}_i) \right) \qquad (9)$$

In this expression, a noncovalent cutting scheme (i.e., all $c_i = 1$) is assumed, and $(\bar{A}_i)$ denotes that the calculation is done in the presence of all fragments $A_j$ satisfying $j \neq i$. To avoid double-counting of energy contributions, it is important that $\bar{A}_i$ is "pre-polarized" without including internal energy terms.

The results for the EMB-PMISP method are shown in Table 4. For all of the charged ligands, the embedding gives a significant improvement. This suggests that the polarization model and its neglected coupling to exchange repulsion are the most probable sources of error and, more importantly, gives a practical method to reduce the problem. However, the effect of embedding is slightly overestimated, and for the neutral ligands the embedding actually shifts the energy away from the reference. Although more statistics is needed to exclude the possibility of a lost error cancelation, this indicates that, in addition, the effect of the Pauli exclusion principle from the surrounding molecules needs to be taken into account. Such approach would avoid possible overpolarization in the embedded calculations and also capture nonadditive effects in the exchange repulsion.

By repeating the embedded calculations without polarizabilities, we note that most of the effect of embedding (in average ~9 kJ/mol) comes from the multipoles, whereas only a small part (in average ~2 kJ/mol) comes from the polarizabilities. The average embedding contributions from each fragment are listed in Table 2. The largest contributions (~3 kJ/mol) come from $A_3$, $A_4$, $A_8$, and $A_9$, although the latter two give large contributions only for the ligands with a carboxylate group. The total effect of embedding is positive for all considered structures, but the sign of each contribution varies. Thus, in cases where cancelation is less prominent, the effect of embedding may be larger.

To continue the analysis, we also give the results using some other fragmentation methods in Table 4. The pairwise additive (PA) method, in which the supermolecular interaction energies are simply summed, gives significantly worse results (MAEs 28 and 19 kJ/mol for the geometry and ligand sets, respectively). This shows that inclusion of many-body effects is important. We tested two other methods that include many-body contributions: the EE-PA[17] and FMO2[6] methods. Both use embedding to model the surrounding molecules, but in contrast to EMB-PMISP, the embedding is used to calculate all interaction terms, that is, not only the nonclassical term. Thus, no classical calculations are needed, but instead calculations have to be done at both the embedded monomer and the embedded dimer level. At the monomer level, an iterative procedure is performed to ensure a self-consistent treatment of polarization (in the original EE-PA method, this iterative scheme is omitted).

The difference between the EE-PA and FMO2 methods is the type of embedding employed. In the EE-PA method, each surrounding molecule is modeled by a point-charge representation (obtained through a Mulliken analysis[66] of the wave function), whereas in the FMO2 method, the exact electric field

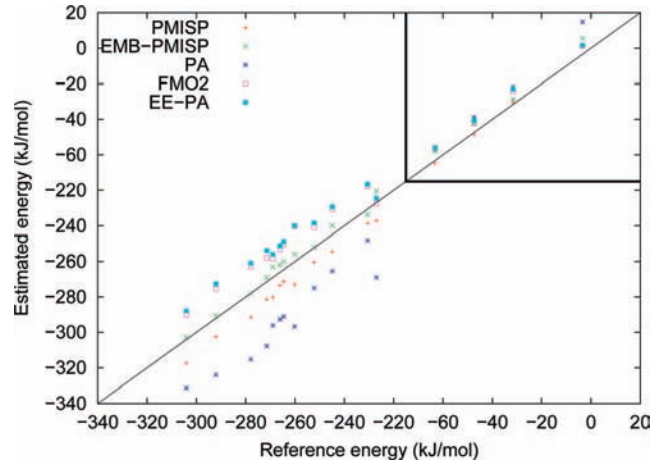


**Figure 5.** Correlation between the supermolecular (HF) interaction energies and the energies estimated by various methods. Note that the axes are broken to accommodate the results of both the charged and the neutral ligands. The line represents perfect correlation.

from the charge density is used (except for approximations in the long-range interactions that were shown not to influence the result). A drawback with the current implementation of the EE-PA and FMO2 methods is that there is no method to correct for BSSE. Therefore, to enable comparison with the PMISP results, the results are shifted by the sum of the counterpoise corrections for each dimer calculation. This procedure is exact for the PA method, but only approximate for the EE-PA and FMO2 methods. Therefore, the uncorrected results are given in the Supporting Information (Tables S1 and S2), together with the uncorrected reference energies. For the FMO3 method, only uncorrected energies are computed. All (counterpoise-corrected) methods are compared in Figure 5.

Regardless of whether the approximate BSSE-correction is applied or not, the results with the EE-PA and FMO2 methods are worse than those obtained by the PMISP method. This is surprising, considering the presumably better treatment of the coupling between induction and repulsion in the former two. Note that, unlike the PMISP method, the accuracy of EE-PA and FMO necessarily increases with the size of the fragments; thus, the restriction to cutting scheme b does not favor the PMISP method. It is also interesting that the FMO2 method is not significantly better than the EE-PA method, despite that the latter uses a rather crude approximation to the electric potential from the surrounding molecules.

A plausible explanation of these observations is that including the electrostatic effects from the surrounding (either exactly or approximately) without including the effects of the Pauli exclusion principle may introduce an inconsistency that is more severe than the actual approximation of the potential.[46] This has been noted before and is one of the reasons why FMO does not work for diffuse basis sets.[67] Clearly, such inconsistency is also introduced in the embedded PMISP method, but in that case, it enters only in a correction to the nonclassical energy, that is, not in the computation of the main part of the many-body effect. The main part is computed using a polarizability model, whose neglect of the local inhomogeneity of the electric field usually cancels the lack of Pauli effects.[46]

The three-body FMO method (FMO3) is expected to reduce the inconsistency problem, because it corrects the embedded polarization by performing supermolecular calculations (which include Pauli effects) for all trimers. The trimer correction reduces the error dramatically so that essentially exact results are obtained (MAEs 1 kJ/mol for both the geometry and the

**TABLE 5: Trimer Results for the Main Structure[a]**

| X−Y | dist (Å) | $E^{sup3}$ | $E^{pol3}$ | $E^{sup3} - E^{pol3}$ | FMO3 |
|---|---|---|---|---|---|
| $A_3-A_5$ | 1.9 | −8.6 | −3.5 | −5.1 | 2.0 |
| $A_8-A_9$ | 2.3 | 9.1 | 4.7 | 4.5 | 1.3 |
| $A_5-A_9$ | 4.8 | 11.2 | 6.9 | 4.3 | 0.3 |
| $A_3-A_4$ | 2.4 | 4.0 | −0.3 | 4.3 | 1.8 |
| $A_5-A_8$ | 2.6 | 16.2 | 12.6 | 3.6 | 0.6 |
| $A_6-A_9$ | 2.5 | −5.3 | −3.1 | −2.3 | 0.3 |
| $A_1-A_3$ | 2.0 | 3.4 | 1.1 | 2.3 | 0.9 |
| $A_1-A_4$ | 2.7 | 1.8 | 0.5 | 1.2 | 0.6 |
| $A_6-A_{10}$ | 1.8 | −2.0 | −1.0 | −1.0 | 1.9 |
| $A_3-A_{14}$ | 5.0 | −2.5 | −1.7 | −0.8 | −0.1 |
| $A_5-A_{15}$ | 2.5 | −0.5 | 0.3 | −0.8 | 0.5 |
| $A_{12}-A_{14}$ | 2.2 | 1.6 | 0.8 | 0.8 | 0.5 |
| $A_4-A_{14}$ | 4.5 | −3.1 | −2.4 | −0.7 | −0.1 |
| $A_5-A_6$ | 2.5 | −1.9 | −1.3 | −0.6 | 0.4 |
| $A_5-A_7$ | 2.5 | 1.8 | 2.4 | −0.6 | 0.2 |
| $A_{10}-A_{11}$ | 2.1 | 1.3 | 0.8 | 0.5 | 0.9 |
| sum | | 25.4 | | 9.9 | 15.3 |
| error | | 2.7 | | −0.2 | 0.4 |

[a] Only fragment pairs X−Y that contribute more than 0.5 kJ/mol to eq 10 are listed, but the sum contains all contributions. Dist is the closest distance between any atom in X and any atom in Y. The error is given relative to the reference energy (not counterpoise-corrected in the FMO case) after addition of the corresponding two-body quantity (PA, PMISP, and FMO2, respectively). Energies are in kJ/mol.

ligand sets). However, the computational cost is several times higher; in fact, the FMO3 calculation takes much longer time than the exact supermolecular calculation (although this will of course not be true with a larger basis set).

Because of the excellent performance of FMO3, it is evident that the trimer correction terms indicate where the error in the FMO2 method arises. Likewise, trimer corrections to PMISP would give the corresponding information about the PMISP error. Therefore, we define the three-body PMISP method as

$$E^{\mathrm{PMISP3}} = E^{\mathrm{PMISP}} + \sum_{i<j}^{n} \left( \left[ E^{\mathrm{sup}}_{[\mathrm{A}_i\mathrm{A}_j]\mathrm{B}} - E^{\mathrm{sup}}_{\mathrm{A}_i\mathrm{B}} - E^{\mathrm{sup}}_{\mathrm{A}_j\mathrm{B}} \right] - \left[ E^{\mathrm{pol}}_{[\mathrm{A}_i\mathrm{A}_j]\mathrm{B}} - E^{\mathrm{pol}}_{\mathrm{A}_i\mathrm{B}} - E^{\mathrm{pol}}_{\mathrm{A}_j\mathrm{B}} \right] \right) \quad (10)$$

where the expression in the first bracket (denoted $E^{sup3}$) is the three-body contribution to the supermolecular interaction energy between the $A_iA_j$ pair and B, and the expression in the second bracket (denoted $E^{pol3}$) is the corresponding three-body contribution to the polarizability energy. Note that the three-body multipole energy vanishes because of the additivity of multipole interactions (cf., eq 7). The largest trimer contributions to eq 10 for the main structure are listed in Table 5. With only three exceptions, the $E^{pol3}$ term for a given trimer has the same sign as the corresponding $E^{sup3}$ term but is smaller in magnitude. Thus, the many-body classical polarization, through which PMISP approximates the total many-body effects, contains qualitatively the correct effect, but systematically underestimates it for each trimer. This observation provides a perfect test case for future improvement of the description of polarization (e.g., by including explicit coupling to the repulsion).

The difference between $E^{sup3}$ and $E^{pol3}$, which is the contribution to PMISP3 for a given trimer, is at most 5 kJ/mol. The detailed results give the same picture as those using the embedded PMISP method: The fragments that contributed most to the embedded correction are in general those present in the most important trimers. However, the trimer results give more detailed information. For example, it can be seen that the rather small embedded correction for $A_5$ is a result of cancelation of its interactions with $A_3$ (−5), $A_9$ (+4), $A_8$ (+4), and others. Moreover, a geometrical analysis of the results (see Table 5) shows that all fragment pairs that give large contributions (>1 kJ/mol) are directly interacting except for the $A_5-A_9$ pair, which is linked by the carboxylate group of B.

For comparison, the corresponding contributions to FMO3 are also reported in Table 5. Interestingly, they seem to be completely uncorrelated to the PMISP3 contributions. In contrast to the latter, the FMO3 contributions are almost consistently positive. Therefore, they add up to a larger sum than the PMISP3 contributions, despite that the individual contributions are in general smaller in magnitude. Thus, the lower error for PMISP than for FMO2 is partially caused by error cancelation. However, the fact that this cancelation occurs in all considered systems suggests that it is in fact advantageous to have a more random error (as in PMISP) as compared to a systematic error (as in FMO2). It is also interesting to note that the FMO3 contributions are even more strongly related to the distance: only neighboring pairs give any significant contribution. This indicates that the neglect of Pauli effects (which are of course very short-ranged) is the dominant approximation in FMO2.

Finally, it can be noted that, just as the FMO3 method, the PMISP3 method gives essentially the exact result. Thus, it may be a useful method when one needs very high accuracy and can afford the increased computational cost. Moreover, because of the geometrical dependence, it may be possible to select the important trimers a priori. The extension to cutting covalent bonds is straightforward with the MFCC approach. The simple three-body expansion (i.e., PA extended with three-body terms) gives an error of 3 kJ/mol, indicating that higher-order many-body effects are rather small for this system.

## 4. Conclusions

We have developed and tested a computationally efficient method (PMISP) to estimate the quantum-mechanical interaction energy between a large and a small molecule in vacuum. The method, which is based on a polarizable multipole description supplemented by a set of supermolecular calculations, can be used for benchmarking simpler potentials, but also applied directly in the calculation of interaction energies.

Tests on model complexes with ∼250 atoms using HF theory show that, for charged ligands, the error of the PMISP model is ∼10 kJ/mol, whereas the corresponding error of the pairwise additive (PA) model is ∼30 kJ/mol. For the neutral ligands, the corresponding errors are significantly lower, but show the same trend, being ∼2 and ∼10 kJ/mol, respectively. Thus, inclusion of polarization improves the performance of the potential by a factor 3−5, using a consistent treatment of the remaining terms. The reason why such large effects of polarization are seldom reported in other published tests of force fields is probably that these errors are hidden behind the large parametrization errors.

The error from calculating multipoles and polarizabilities by the MFCC fragmentation procedure is negligible (within 1 kJ/mol). Thus, the only way to systematically obtain more accurate results is by improving the modeling of many-body effects, in particular by including coupling between polarization and repulsion. However, our test indicates that two of the previously proposed methods that include this coupling, FMO and EE-PA, do not give better results; in fact, they give slightly larger errors than PMISP to a higher computational cost. Although

our attempt to include embedding in the PMISP method does indeed decrease the error for charged ligands, this is still an area where more development is needed.

When going from the HF to the MP2 level, the PMISP error increases slightly for the neutral ligands (to ∼5 kJ/mol) but not for the charged ligands. This indicates that the correlation effects on the nonclassical energy are nearly pairwise additive. Of course, many-body contributions to the dispersion are not captured by the MP2 method, so these may still be significant. The dispersion part of interaction energies is known to converge very slowly with basis set. Therefore, to obtain quantitative results, a large basis set including diffuse functions must be used; the modest 6-31G* basis set was used in this study only to enable an exact reference calculation.

The obtained errors have important consequences for force-field development in general. Assuming that the nonadditivity of intermolecular interactions is only modeled by a classical polarization term (as in all current general-purpose force fields that consider the nonadditivity at all), the PMISP method can be considered as a perfectly fitted polarizable force field. Thus, without introducing new physical principles for modeling nonadditivity,[68] this is the best accuracy one can ever expect from a polarizable force field for this type of problem. Similarly, the PA model uses exact pair potentials, so its accuracy provides a corresponding limit for nonpolarizable force fields.

The PMISP method does not put any restriction on the size of the large molecule, and the size of the small molecule is only limited by the applied QM method. Thus, the method can in principle be used for a full protein−ligand complex. However, the method, as presented here, wastes computational power by treating residues far from the ligand in the same way as the nearest ones. In a future publication, we will show how the method can be adapted for this type of application.

**Acknowledgment.** We thank F. Aquilante for helpful comments and suggestions, LUNARC at Lund University for generously providing computational resources, and AstraZeneca and the Swedish Science Research Council for financial support.

**Supporting Information Available:** Detailed results for the geometry and ligand sets in Tables S1 and S2, respectively. This material is available free of charge via the Internet at http://pubs.acs.org.

## References and Notes

(1) Newton, M. D.; Boer, F. P.; Lipscomb, W. N. *J. Am. Chem. Soc.* **1966**, *88*, 2353.
(2) Niessen, W. V. *Theor. Chim. Acta* **1973**, *31*, 111.
(3) Degand, P.; Leroy, G.; Peeters, D. *Theor. Chim. Acta* **1973**, *30*, 243.
(4) Yang, W. *Phys. Rev. Lett.* **1991**, *66*, 1438.
(5) Mezey, P. G. *J. Med. Chem.* **1995**, *18*, 141.
(6) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701.
(7) Claverie, P., Pullmann, B., Eds. *Intermolecular Interactions: From Diatomics to Biopolymers*; Wiley: New York, 1978; p 69.
(8) Amovilli, C.; Cacelli, I.; Campanile, S.; Prampolini, G. *J. Chem. Phys.* **2002**, *117*, 3003.
(9) Zhang, D. W.; Zhang, J. Z. H. *J. Chem. Phys.* **2003**, *119*, 3599.
(10) Zhang, D. W.; Xiang, Y.; Zhang, J. Z. H. *J. Phys. Chem. B* **2003**, *107*, 12039.
(11) Zhang, D. W.; Xiang, Y.; Gao, M.; Zhang, J. Z. H. *J. Chem. Phys.* **2004**, *120*, 1145.
(12) Deev, V. A.; Collins, M. A. *J. Chem. Phys.* **2005**, *122*, 154102.
(13) Bettens, R. P. A.; Lee, A. M. *J. Phys. Chem. A* **2006**, *110*, 8777.
(14) Li, W.; Li, S.; Jiang, Y. *J. Phys. Chem. A* **2007**, *111*, 2193.
(15) Hirata, S.; Valiev, M.; Dupuis, M.; Xantheas, S.; Sugiki, S. *Mol. Phys.* **2005**, *103*, 2255.
(16) Morita, S.; Sakai, S. *J. Comput. Chem.* **2001**, *22*, 1107.
(17) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 46.
(18) Tschumper, G. S. *Chem. Phys. Lett.* **2006**, *427*, 185.
(19) Fedorov, D. G.; Kitaura, K. *Chem. Phys. Lett.* **2006**, *433*, 182.
(20) Li, S.; Li, W.; Fang, T. *J. Am. Chem. Soc.* **2005**, *127*, 7215.
(21) Nakano, T.; Kaminuma, T.; Sato, T.; Fukuzawa, K.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2002**, *351*, 475.
(22) Kamiya, M.; Hirata, S.; Valiev, M. G. *J. Chem. Phys.* **2008**, *128*, 074103.
(23) Jiang, N.; Ma, J.; Jiang, Y. *J. Chem. Phys.* **2006**, *124*, 114112.
(24) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 1342.
(25) Buckingham, A. D. *Adv. Chem. Phys.* **1967**, *12*, 107.
(26) Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887.
(27) Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. *J. Chem. Phys.* **2005**, *123*, 214103.
(28) Engkvist, O.; Åstrand, P.-O.; Karlström, G. *Chem. Rev.* **2000**, *100*, 4087.
(29) Brdarski, S.; Karlström, G. *J. Phys. Chem. A* **1998**, *102*, 8182.
(30) Chen, W.; Gordon, M. S. *J. Chem. Phys.* **1996**, *105*, 11081.
(31) Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933.
(32) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. A* **2004**, *108*, 621.
(33) Wallqvist, A.; Karlström, G. *Chem. Scr.* **1989**, *29A*, 131.
(34) Gresh, N.; Piquemal, J.-P.; Krauss, M. *J. Comput. Chem.* **2005**, *26*, 1113.
(35) Wheatley, R. J.; Price, S. L. *Mol. Phys.* **1990**, *71*, 1381.
(36) Jensen, J.; Gordon, M. S. *Mol. Phys.* **1996**, *89*, 1313.
(37) Söderhjelm, P.; Karlström, G.; Ryde, U. *J. Chem. Phys.* **2006**, *124*, 244101.
(38) Adamovic, I.; Gordon, M. S. *Mol. Phys.* **2005**, *103*, 379.
(39) Hodges, M. P.; Stone, A. J. *Mol. Phys.* **2000**, *98*, 275.
(40) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. *Acc. Chem. Res.* **2000**, *33*, 889.
(41) Gagliardi, L.; Lindh, R.; Karlström, G. *J. Chem. Phys.* **2004**, *121*, 4494.
(42) Söderhjelm, P.; Krogh, J. W.; Karlström, G.; Ryde, U.; Lindh, R. *J. Comput. Chem.* **2007**, *28*, 1083.
(43) Gao, A. M.; Zhang, D. W.; Zhang, J. Z. H.; Zhang, Y. *Chem. Phys. Lett.* **2004**, *394*, 293.
(44) Molina, P. A.; Li, H.; Jensen, J. H. *J. Comput. Chem.* **2003**, *24*, 1971.
(45) Ren, P.; Ponder, J. W. *J. Comput. Chem.* **2002**, *23*, 1497.
(46) Söderhjelm, P.; Öhrn, A.; Ryde, U.; Karlström, G. *J. Chem. Phys.* **2008**, *128*, 014102.
(47) Miyamoto, S.; Kollman, P. A. *Proteins: Struct., Funct., Genet.* **1993**, *16*, 226.
(48) Wang, J.; Dixon, R.; Kollman, P. A. *Proteins: Struct., Funct., Genet.* **1999**, *34*, 69.
(49) Kuhn, B.; Kollman, P. A. *J. Med. Chem.* **2000**, *43*, 3786.
(50) Weis, A.; Katebzadeh, K.; Söderhjelm, P.; Nilsson, I.; Ryde, U. *J. Med. Chem.* **2006**, *49*, 6596.
(51) Cornell, W.; Cieplak, P.; Bayly, C.; Gould, I.; Merz, K. M., Jr.; Ferguson, D.; Spellmeyer, D.; Fox, T.; Caldwell, J.; Kollman, P. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(52) Pugliese, L.; Coda, A.; Malcovati, M.; Bolognesi, M. *J. Mol. Biol.* **1993**, *231*, 698.
(53) Cieplak, P.; Caldwell, J.; Kollman, P. A. *J. Comput. Chem.* **2001**, *22*, 1048.
(54) Karlström, G.; Lindh, R.; Malmqvist, P.-Å.; Roos, B. O.; Ryde, U.; Veryazov, V.; Widmark, P.-O.; Cossi, M.; Schimmelpfennig, B.; Neogrady, P.; Seijo, L. *Comput. Mater. Sci.* **2003**, *28*, 222.
(55) Veryazov, V.; Widmark, P.-O.; Serrano-Andrés, L.; Lindh, R.; Roos, B. O. *Int. J. Quantum Chem.* **2004**, *100*, 626.
(56) *MOLCAS 7*; University of Lund: Sweden, 2007; see http://www.teokem.lu.se/molcas.
(57) Werner, H.-J.; Knowles, P. J.; Lindh, R.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Rauhut, G.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Lloyd, A. W.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T. *MOLPRO, version 2006.1, a package of ab initio programs*; Cardiff: U.K., 2006; see http://www.molpro.net.

(58) Beebe, N. H. F.; Linderberg, J. *Int. J. Quantum Chem.* **1977**, *7*, 683.
(59) Koch, H.; Sánchez de Merás, A.; Pedersen, T. B. *J. Chem. Phys.* **2003**, *118*, 9481.
(60) Aquilante, F.; Pedersen, T. B.; Lindh, R. *J. Chem. Phys.* **2007**, *126*, 194106.
(61) Aquilante, F.; Pedersen, T. B. *Chem. Phys. Lett.* **2007**, *449*, 354.
(62) Aquilante, F.; Malmqvist, P.-Å.; Pedersen, T. B.; Ghosh, A.; Roos, B. O. *J. Chem. Theory Comput.* **2008**, *4*, 694.
(63) 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.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.
(64) Moszynski, R.; Wormer, P. E. S.; Jeziorski, B.; van der Avoird, A. *J. Chem. Phys.* **1995**, *103*, 8058.
(65) Lotrich, V. F.; Szalewicz, K. *J. Chem. Phys.* **1997**, *106*, 9668.
(66) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.
(67) Fedorov, D. G., personal communication.
(68) Mas, E. M.; Bukowski, R.; Szalewicz, K. *J. Chem. Phys.* **2003**, *118*, 4386.

JP8073514