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

Söderhjelm, Pär; Ryde, Ulf

Published in:
Journal of physical chemistry. A

DOI:
10.1021/jp8073514

2009

Document Version:
Peer reviewed version (aka post-print)

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, P.O.B. 124, SE-22100 Lund, Sweden

Correspondence to Pär Söderhjelm, E-mail: par.soderhjelm@teokem.lu.se,
Tel: +46 46 2224502, Fax: +46 46 2224543

August 17, 2008

Abstract

A new method for accurately estimating interaction energies involving a large molecule 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 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 dispersion and repulsion is assumed. 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 mainly originates from the omission of the three-body contribution to the coupling of polarization and exchange repulsion.

Keywords: interaction energy, QM/MM, fragmentation, polarization, ligand binding

1 Introduction

In many applications of theoretical methods in chemistry, one is interested in changes in 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 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, which are treated more or less independently. Such *fragmentation* approaches to QM calculations have a long history [1, 2, 3, 4, 5], but they have experienced a strong revival in the last 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 in 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 non-overlapping 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, based on the approximate atom-wise additivity of interaction energies observed by Claverie [7], methods for handling the overlaps by adding and subtracting fragment energies have been proposed, e.g. by Prampolini [8] and by Zhang and Zhang [9, 10, 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, 13, 14]. There are also related methods limited to non-covalently bonded clusters, in which the fragmentation is trivial [15, 16, 17, 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 has 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 pair-wise interaction energies. For calculating the total energy by MFCC, embedding using unit charges, Mulliken charges, ESP charges, or natural population analysis (NPA) charges have been used [14, 23].

In principle, all 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), e.g. 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). In addition, PA can be replaced by the three-body (3B) approximation to include another term in the expansion. 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 non-covalent 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 non-classical terms, stemming from the quantum-mechanical nature of the interaction.

Such physical decomposition of the interaction energy is the foundation of molecular mechanics force fields, in which each term is estimated by a separate expression. Normally, the parameters in the classical terms (e.g. atomic charges and polarizabilities) as well as those in the non-classical 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 non-classical parameters are predefined, whereas the classical parameters are obtained for each considered system, typically by performing QM calculations and analyzing the obtained electron density [26]. Although very appealing in theory, the accuracy of this approach is limited by the transferability of the non-classical parameters, which are normally fitted to reproduce supermolecular energies. Fitting of the exchange-repulsion is a difficult problem, which has been adressed in many studies [27, 26, 28, 29, 30]. The transferability is improved if the overlap of the wavefunctions (or densities) is explicitly taken into account [31, 32, 33]. By this approach, good results may be obtained without fitted parameters [34] or with a small number of element-independent parameters [35]. The dispersion

energy has also been subject to many studies (e.g. [36, 37]) 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 non-classical terms by supermolecular QM calculations, although for smaller subsystems. The advantage of this is twofold: 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 non-classical 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 captures 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 [38]. 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

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

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

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

$$E_{AB}^{PMISP} = E_{AB}^{ele} + E_{AB}^{ind} + E_{AB}^{rest} \tag{2}$$

where E_{ele} and E_{ind} are the electrostatic and induction energies, respectively, when both monomers are treated classically, using multipoles and polarizabilities, and E_{rest} is a rest 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 as charge transfer).

Unlike in standard molecular-mechanics force fields, E_{rest} 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 rest term from each fragment separately, assuming pairwise additivity. In this work, we employ the MFCC fragmentation procedure [9], which is a rigorous 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_{rest} by

$$E_{AB}^{rest} = \sum_{i=1}^{n} c_i \left(E_{A_iB}^{sup} - E_{A_iB}^{ele} - E_{A_iB}^{ind} \right) \tag{3}$$

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 dipole polarizabilities, located at each nuclear position and each covalent bond midpoint. These *properties* are computed at any QM level by the LoProp method [39, 40]. The multipole expansion is truncated after quadrupoles. The electrostatic interaction includes all possible terms formed by the multipoles, i.e. up to and including quadrupole–quadrupole interactions.

In the same spirit as for the calculation of the rest 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 [41]) and estimate the properties of A as

$$P_k^A = \sum_{i=1}^{n} c_i P_k^{A_i} \tag{4}$$

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 cancelled between caps and concaps; the remaining part is moved to the corresponding real atom. Alternative methods for assembling multipoles have been discussed previously [42].

The polarizabilities in the LoProp approach are derived using homogenous electric fields. Thus, when they are used, there should be no explicit coupling between polarizabilities belonging to the same molecule. In fact, each polarizability is strictly correct only under the assumption that the rest of the molecule is responding to the same electric field. Thus, there is an implicit coupling that is correct as long as the field is homogenous but only approximative when the field is inhomogenous. The approximation was recently found to be accurate for small and medium-sized molecules [43], and the present study will provide a test for larger molecules.

When Eq. 4 is used for estimating properties for A, several subtle issues arise. First, it is obvious that the polarizabilities in one fragment should respond to the field from static multipoles in other fragments. For two centers belonging to overlapping fragments, special care is needed. We use the following rule: Each static

multipole contributes to the field at all other centers, except those centers that have been in the same fragment in at least one LoProp calculation. This convention can be rigorously motivated. 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 fragments. The case with $n_k = 1$ is trivial: The polarizability α_k should not respond to multipoles in the same fragment, because these have already influenced the static multipole at k. 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 α_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. [44].

When considering intramolecular coupling of the polarization, i.e. the effect of an induced dipole on other polarizabilities in the same molecule, a similar argument can be applied. Clearly, the polarization response in one part of a large molecule should affect the response in distant fragments, because in that case all implicit coupling between the polarizabilities is absent. However, as soon as the two centers are in the same fragment, the implicit coupling is present (although in an approximate way because of the inhomogeniety of the electric field). Therefore, we have chosen to use the same rule as for the static field, i.e. that each induced dipole contributes to the field at all other centers, except those that have been in the same fragment in at least one LoProp calculation. The importance of the intramolecular coupling will also be specifically tested.

Another issue when applying fragmentation for calculating the properties is the definitions of electrostatic and induction energy, respectively. The most natural definition of electrostatic energy is the static interaction between the monomers, each having been internally "pre-polarized" 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. However, another possible definition uses the functional form of the potential and simply distinguish between on one hand the multipole–multipole terms and on the other hand the multipole–polarizability and polarizability–polarizability terms. To avoid confusion with the natural definition, we denote the latter quantities multipole energy (E^{mult}) and polarization energy (E^{pol}). In the trivial case of having only one fragment representing monomer A, the two definitions are of course equivalent so that $E^{ele} = E^{mult}$ and $E^{ind} = E^{pol}$.

The general case with several fragments is illustrated in Fig. 2, where five relevant "states" are depicted, using a two-fragment description of monomer A. The dashed line in states 1 and 4 indicates that the interaction is turned off. Polarization of a fragment is indicated with an arrow pointing towards the "source" of the polarizing field. Thus, in state 2, each A_i fragment is polarized only by the other A_i fragment, whereas in state 3, it is also polarized by B. The total classical interaction energy is the change from state 1 to 3. The natural definition of E^{ele} and E^{ind} corresponds to the path in the upper row. To clarify the other definition, we introduce a non-polarized version of monomer A, denoted O. We now recognize that

$$E^{mult}_{AB} = E^{ele}_{OB} \tag{5}$$

and

$$E^{pol}_{AB} = E^{ind}_{OB} - E^{ind}_{O} \tag{6}$$

Also, it is evident from Fig. 2 that

$$E_{AB}^{ele} + E_{AB}^{ind} = E_{AB}^{mult} + E_{AB}^{pol} \tag{7}$$

although the equality does not hold termwise (except in the trivial case). Thus, as long as the PMISP result is concerned, it does not matter which pair of quantities we use in Eq. 2. However, to be able to compare individual terms for several fragmentation schemes, it is essential to use the natural definition. To obtain the natural quantities one may calculate $E_{O(B)}^{ind}$, i.e. the internal induction energy of monomer A in the presence of (but not polarized by) monomer B, and exploit the energy conservation around the cycles $1-2-5-4$ and $2-3-5$, as can be seen from Fig. 2.

There is another useful interpretation of the PMISP energy. Inserting Eq. 3 into Eq. 2, changing definition according to Eq. 7, and reordering the terms, gives

$$E_{AB}^{PMISP} = \sum_{i=1}^{n} c_i E_{A_iB}^{sup} + \left(E_{AB}^{pol} - \sum_{i=1}^{n} c_i E_{A_iB}^{pol} \right) \tag{8}$$

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 cancellation only occurs if the same set of fragments is used for the calculation of the rest term and the properties. In Eq. 8, we easily recognize the first term as the pairwise additive (PA) interaction energy and the second term as the many-body contribution to the polarization energy. Thus, one may regard the PMISP energy as an improvement of the PA energy by approximating the many-body energy (which is absent in PA) by the many-body energy from a polarizable multipole description.

2.2 Systems

For testing the method, we use a model of the avidin protein interacting with the seven ligands shown in Fig. 3. This system (with the full protein) has previously been subject to several studies [45, 46, 47, 48]. 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. The exact simulation protocol has been described elsewhere [48]. 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 remaing 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 4. The residue numbering refers to PDB structure 1AVD [49] The whole model belongs to the same subunit (labeled B in the original structure), except for Trp-110, which belongs to a neighbouring subunit. The model includes all atoms within 4 Å of the ligand (using the coordinates of the main structure), but also a minimal set of additional atoms that are necessary to complete chemically reasonable groups such as aromatic rings. Bonds that are cut in the process are capped with hydrogen atoms. Because many bonds are broken, the model is not a single covalently bound molecule, but rather a collection of 15 separate molecules of different sizes, labeled by $A_1, A_2, ..., A_{15}$. All of these 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.

To be able to test the additivity among covalently and non-covalently bound fragments separately, the cutting of the model into fragments 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 4, and each fragment is capped with $-COCH_3$ and $-NHCH_3$ groups at the N and C termini, respectively.

These schemes apply both to the computation of properties for monomer A and the computation of the rest term, although scheme *a* makes no sense for the rest 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 rest 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 to use the same cutting scheme for the rest term as for the properties (for example, it is a prerequisite for Eq. 8), we also investigate other possibilities.

2.3 Computational details

The multipoles and polarizabilities were obtained by the LoProp method [39] as implemented in MOLCAS [50, 51, 52]. 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 [39]) 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 [53] 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 [54, 55] was applied in combination with the local exchange (LK) algorithm [56]. Based on previous analysis of the accuracy of the CD approximation [56, 57, 58], 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 $\sim$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. 8 and is thus obtained as a part of the PMISP procedure. For obtaining FMO and EE-PA energies, we used the FMO procedure in the GAMESS program package [59] 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 distance.

3 Results

3.1 Demonstration of the PMISP method

The accuracy of the PMISP method and various further approximations was tested by performing calculations of the interaction energy between a model of avidin (216 atoms, denoted A, see Fig. 4) and biotin-like ligands ($12 - 41$ atoms, denoted B, see Fig. 3), in various geometries, as explained in the method section. 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 for exchange-repulsion and dispersion, the lower cost of the HF property calculations allows us to investigate the influence of the applied fragmentation scheme.

To help the reader to understand the PMISP procedure, we report some numerical values for the main structure. The quantities depicted in Fig. 2 are listed in Table 1 for each of the three cutting schemes, using HF properties. The difference between the naturally defined electrostatic energy (E^{ele}) and the "direct" multipole–multipole term (E^{mult}) is 18–24 kJ/mol when A is split in the property calculation (this difference is exactly cancelled 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 demonstrates the importance of comparing different methodologies in the natural viewpoint and not term-wise according to the functional form. This is a general conclusion that applies whenever one compares interaction energies computed with polarizable force fields: The direct difference of polarization terms depends on details in the intramolecular polarization description and is in general not directly comparable with another force field. To emphasize this point, the corresponding values for the Amber 2002 polarizable force field 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 energy almost vanishes (in some cases, it actually becomes positive), but this only tells us that the rather arbitrarily defined internal polarization of separated monomers is equally favorable as the polarization of the dimer.

Let us now consider the non-classical term, E^{rest}. All terms contributing to E^{rest} for the main structure are reported in Table 2. Both HF and MP2 results are given and the fragment numbering is the same as in Fig. 4. It can be seen that the contributions vary substantially among the various fragments. This reflects the difference in size as well as the difference in overlap. With a few small-magnitude exceptions, the rest term at the HF level is always positive, as can be expected because it contains 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 rest term varies, because of the balance between exchange-repulsion and dispersion. For each fragment interaction, the difference between the rest terms at the MP2 and HF levels is a good estimate of the dispersion energy. The HF results for A_5 suggests that the rest term calculated for the whole molecule (cutting scheme b) and the rest term calculated by summing over fragments $A_{5a} - A_{5ef}$ (cutting scheme c) are the same within 1 kJ/mol.

3.2 Accuracy of the PMISP method

To obtain statistics regarding the accuracy of PMISP, we use ten different geometries of the BTN1 interaction (geometry set), and seven different ligands (Fig. 3), each in a different geometry (ligand set). In all these calculations, the avidin model remains identical, but its geometry changes significantly.

The results at both HF and MP2 level are given for the geometry set in Table 3 and for the ligand set in Table 4. As previously discussed, each of the three terms E^{ele}, E^{ind}, and E^{rest} can be computed using several cutting schemes. Thus, several estimates of each term are reported in Tables 3–4, although not for MP2 because of computational limitations. The results confirm that the three cutting schemes give almost identical electrostatic energy. The most reliable value should be with properties obtained using scheme a, i.e. in a single QM calculation. Using these results as a reference, the mean absolute errors in E_{ele} are $\sim$1 kJ/mol for the use of either scheme b or c. 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 the method section. 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 higher error than fragmenting across e.g. a hydrogen bond. The explanation for this is probably that the quantum-mechanical error of confining each wavefunction to one fragment is to a large extent eliminated by the rigorous MFCC procedure, so that the remaining error comes from imperfect description of the intermolecular 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 still small, both results deviate significantly from scheme a with mean absolute errors of 7–14 kJ/mol. At first sight, one might assume that this discrepancy is due to the error introduced by fragmentation of monomer A. However, there is another effect that also contributes to the difference. The intramolecular coupling of polarizabilities within a fragment is only treated implicitly, and the error of this approximation usually increases with the size of the molecule [43]. 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 multipole–polarizability interaction within monomer A, on the other hand, is kept because that part is exactly modeled (by QM) in scheme a. With this treatment of intramolecular polarization, the mean absolute difference between schemes a and b goes down to 1.5 and 0.8 kJ/mol for the geometry and ligand set, respectively, and the corresponding differences between schemes a and c are 1.6 and 1.0 kJ/mol. Thus, when intramolecular polarization is treated consistently, the influence of the fragmentation of monomer A on the induction energy is negligible for all practical purposes.

The rest term is also similar for cutting schemes b and c, with a mean absolute difference of 3.1 kJ/mol for the geometry set and 1.6 kJ/mol for the ligand set. This indicates that the MFCC fragmentation procedure succeeds in cancelling the non-classical contributions from capping atoms. Interestingly, if the supermolecular energies are combined directly (as in the PA approach), the differences are about twice as high (5.2 and 3.7 kJ/mol, respectively), showing that the additivity assumption employed in the fragmentation is a better approximation for the rest term than for the supermolecular energy

The total PMISP energy is an approximation to the supermolecular interaction energy (E_{ref}). Therefore, we report the error ΔE^{PMISP} in Tables 3 and 4, and we also consider the mean absolute error (MAE). 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 E^{rest}, 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 cancellation.

If we fix scheme c for the rest term and vary the scheme for the classical terms, we see that the MAE becomes 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, which is clearly unacceptable. 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. Thus, we may conclude that the neglect of explicit intramolecular polarizability coupling in large molecules (as in scheme a) is a more severe approximation than the actual fragmentation procedure.

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 cancellation occurs when using the same cutting scheme for all terms. We cannot directly conclude which of the schemes bb and cc is the best one; the more efficient approach (cc) is therefore preferred.

The MAEs for the ligand set follow the same trends. It is not directly comparable to the MAE for the geometry set, because the magnitude of the error roughly follows the magnitude of the interaction energy itself, which in turn depends on the charge of the ligand (-1 for ligands BTN1–BTN3 but 0 for ligands BTN4–BTN7).

Interestingly, the errors do not increase at the MP2 level, despite that the dispersion could make an additional contribution to the non-additivity. This suggests that the main mechanisms responsible for the error are present already at the HF level. To show this in a quantitative way, one only has to note the similarity between the errors at HF and MP2 level, with a mean absolute difference of only 1 kJ/mol for the geometry set and 3 kJ/mol

for the ligand set (with a maximum of 6 kJ/mol for BTN4). The difference does not seem to depend on the charge of the ligand or the magnitude of the total interaction energy. One could assume that it rather depends on the magnitude of the dispersion energy, but our test set is not diverse enough to investigate this assumption.

In fact, the observed similarity between the PMISP errors at the HF and MP2 levels immediately suggests a MP2–HF hybrid method for estimating MP2 interaction energies, namely by the expression

$$E^{PMISP\text{-}CE} = E^{HF}_{ref} - E^{PMISP,HF} + E^{PMISP,MP2} \tag{9}$$

In analogy with earlier notation [24], we call this approach PMISP-CE (where CE stands for correlation energy), and it is also reported in Tables 3 and 4. This appears to be an excellent approximation, useful in cases for which the HF energy can be obtained in a supermolecular calculation but not the MP2 energy. This method is similar in spirit to the PA-CE method [24], which simply adds the sum of the correlation energy for each pair to the HF energy of the complex. In fact, apart from our extension to cutting covalent bonds, the two energies differ only by a term describing the change of the many-body polarization when correlation is included, i.e.

$$E^{PMISP\text{-}CE} = E^{PA\text{-}CE} + \Delta E^{pol,mb}_{HF\rightarrow MP2} \tag{10}$$

As can be seen in Tables 3 and 4, the neglect of this term does not affect the result significantly. An advantage of the PA-CE method is that no property calculations are needed.

Considering the potential usefulness of the PA-CE and PMISP-CE methods, we have investigated whether it is possible to compute the HF energies with a smaller basis set than that used for the MP2 calculations of fragment interactions. In this case, the assumed pairwise additive quantity is not the pure correlation energy but also contains a basis set correction. In practice, one would be interested in approximating a large and diffuse basis set with a smaller one (e.g. 6-31G*) which could be used for the whole complex. However, to be able to test the method, we use the MP2/6-31G*–HF/3-21G combination. The results, given in Tables 3 and 4, are in general in good agreement with the non-approximated PA-CE results, but the maximal difference is 11 kJ/mol, indicating that this approximation should be used with some care.

3.3 Analyzing 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 present already 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 which effects are important.

For all geometries involving charged ligands, the attraction is overestimated (i.e. the interaction energy is too negative) by the PMISP method. Thus, the rest term of each fragment interaction is underestimated compared to the rest term for the whole complex. This error can be considered from two different viewpoints. If one assumes that the "true" repulsion term is perfectly pairwise additive, there must be an error in the classical

induction term. Typically, the classical calculations tends to give a slight over-polarization because of the neglect of Pauli exclusion effects [43]. The sign of the PMISP error would then indicate that this over-polarization is more prominent in the AB interaction than in the A_iB interactions.

On the other hand, if one instead assumes that the induction term is correct, the results suggest that the repulsion is not perfectly additive and is in fact larger for the whole complex than for the individual pairs. This means that the surrounding molecules influence the repulsion of a given dimer, either by polarizing the dimer wavefunction in a way that increases the repulsion (e.g. by increasing the orbital overlap), or by restricting (by the Pauli exclusion effect) the way in which the dimer density may relax.

The two views are of course inseparable, because the discrimination between induction and repulsion is only possible at a qualitative level. Nevertheless, to simplify the discussion we adopt the second view and test the effect of polarization from surrounding molecules on the individual $E^{rest}_{A_iB}$ terms. This defines a new method, denoted embedded PMISP, in which Eq. 3 is replaced by

$$E^{rest,EMB}_{AB} = \sum_{i=1}^{n} \left(E^{sup}_{A_iB}(\bar{A}_i) - E^{ele}_{A_iB}(\bar{A}_i) - E^{ind}_{A_iB}(\bar{A}_i) \right) \tag{11}$$

where a non-covalent 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 terms, it is important that $\bar{A}_i$ is "pre-polarized" without including internal energy terms.

Two series of test calculations are made, one using the polarizable multipole representation of the fragments for constructing $\bar{A}_i$ (denoted EMB-PMISP), and one in which the polarizabilities of $\bar{A}_i$ are removed (denoted EMB(NP)-PMISP). The result is shown in Tables 3 and 4. For all the negatively charged ligands, the embedding gives a significant improvement, indicating that most of the non-additivity in the repulsion is caused by polarization from surrounding molecules, which can be modeled classically. However, the results for the neutral ligands are less convincing; the embedding shifts the energy away from the reference. One possible reason for this could be that there is an error cancellation in the original PMISP energy, which is lost when the rest term is improved. Another reason could be that the surrounding molecules are modeled only classically, so the effect of the Pauli exlusion principle is not taken into account. Such procedure can be expected to overestimate the effect of embedding for certain fragments and could lead to a net effect in the wrong direction when the errors are small. The effect of embedding is overestimated also for the negatively charged ligands.

By comparing the EMB-PMISP and EMB(NP)-PMISP results, one can see 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 embedding contribution from each fragment rest term is listed in Table 5. 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 cancellation 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 Tables 3 and 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, i.e. not only the rest term. Thus, no classical calculations are needed, but instead calculations have to be done at both the embedded monomer and 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 [60] of the wavefunction), whereas in the FMO2 method, the exact electric field from the charge density is used (except for long-range approximations 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 also given in Tables 3 and 4 together with the uncorrected reference energies. All (counterpoise-corrected) methods are compared in Fig. 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 favour 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. In fact, for the standard method (i.e. not counterpoise-corrected), the EE-PA method gives lower MAEs.

A plausible explanation for these observations is that by including the electrostatic effects from the surrounding (whether exactly or approximately) without including the effects of the Pauli exclusion principle, one introduces an inconsistency that is more severe than the actual approximation of the potential [43]. This has been noted before and is one of the reasons why FMO does not work for diffuse basis sets [61]. Clearly, such inconsistency is also introduced in the embedded PMISP method, but in that case, it enters only in a correction to the rest energy, i.e. 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 [43]. It should also be noted that it seems to be more difficult to reproduce the uncorrected reference results than the counterpoise-corrected ones. This indicates that the uncorrected results contain artificial (numerical) effects that cannot be modeled by a physically sound method such as PMISP or FMO2.

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

Owing to the excellent performance of FMO3, it is evident that the magnitude of each trimer correction gives useful information on where the major errors in the FMO2 method come from. Analogously, trimer corrections to PMISP would give the corresponding information about the PMISP error. Therefore, we define the three-body PMISP method as

$$E^{PMISP3} = E^{PMISP} + \sum_{i<j}^{N} \left(\left[E^{sup}_{[A_i A_j]B} - E^{sup}_{A_i B} - E^{sup}_{A_j B} \right] - \left[E^{pol}_{[A_i A_j]B} - E^{pol}_{A_i B} - E^{pol}_{A_j B} \right] \right) \tag{12}$$

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 polarization energy. Note that the three-body multipole energy vanishes because of the additivity of multipole interactions (cf. Eq. 8). The largest trimer contributions to Eq. 12 for the main structure are listed in Table 6. 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 cancellation of its interactions with A_3 (-5), A_9 (+4), A_8 (+4), and others. Moreover, a geometrical analysis of the results (see Table 6) 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 6. 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 cancellation. However, the fact that this cancellation occurs in all considered systems suggests that it is in fact advantageous to have a more random error (as in PMISP) 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 neighbouring pairs give any significant contribution. This indicates that the neglect of Pauli effects (which are known to be very short-range) is the only major approximation in FMO2, whereas the error from assuming perfect pairwise additivity of the rest term (as in PMISP) can be slightly more long-range.

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 straight-forward 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, in this case, inclusion of polarization improves the performance of the potential by a factor 3-5, using a consistent treatment of the remaining terms. Because the PMISP model can be considered as a perfectly fitted polarizable potential and the PA model similarly as the perfectly fitted pair potential, this is the best accuracy one can ever expect from a polarizable and non-polarizable force field, respectively, for this type of problem. 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 the MFCC fragmentation procedure is negligible. Thus, the only way to systematically obtain more accurate results is by improving the polarization model, 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 higher 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.

The error does not increase significantly when going from the HF to the MP2 level. This indicates that the MP2 dispersion is nearly pairwise additive, as assumed by the method. Of course, some 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 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.

Acknowledgments

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

References

[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. Letters* **1999**, *313*, 701.

[7] Claverie, P. In Pullmann, B., ed., *Intermolecular Interactions: From Diatomics to Biopolymers*. Wiley, New York, 1978 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. Letters* **2006**, *427*, 185.

[19] Fedorov, D. G.; Kitaura, K. *Chem. Phys. Letters* **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. Letters* **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] Engkvist, O.; Åstrand, P.-O.; Karlström, G. *Chem. Rev.* **2000**, *100*, 4087.

[27] Brdarski, S.; Karlström, G. *J. Phys. Chem. A* **1998**, *102*, 8182.

[28] Chen, W.; Gordon, M. S. *J. Chem. Phys.* **1996**, *105*, 11081.

[29] Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933.

[30] Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. A* **2004**, *108*, 621.

[31] Wallqvist, A.; Karlström, G. *Chem. Scripta* **1989**, *29A*, 131.

[32] Gresh, N.; Piquemal, J.-P.; Krauss, M. *J. Comput. Chem.* **2005**, *26*, 1113.

[33] Wheatley, R. J.; Price, S. L. *Mol. Phys.* **1990**, *71*, 1381.

[34] Jensen, J.; Gordon, M. S. *Mol. Phys.* **1996**, *89*, 1313.

[35] Söderhjelm, P.; Karlström, G.; Ryde, U. *J. Chem. Phys.* **2006**, *124*, 244101.

[36] Adamovic, I.; Gordon, M. S. *Mol. Phys.* **2005**, *103*, 379.

[37] Hodges, M. P.; Stone, A. J. *Mol. Phys.* **2000**, *98*, 275.

[38] 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.

[39] Gagliardi, L.; Lindh, R.; Karlström, G. *J. Chem. Phys.* **2004**, *121*, 4494.

[40] Söderhjelm, P.; Krogh, J. W.; Karlström, G.; Ryde, U.; Lindh, R. *J. Comput. Chem.* **2007**, *28*, 1083.

[41] Gao, A. M.; Zhang, D. W.; Zhang, J. Z. H.; Zhang, Y. *Chem. Phys. Letters* **2004**, *394*, 293.

[42] Molina, P. A.; Li, H.; Jensen, J. H. *J. Comput. Chem.* **2003**, *24*, 1971.

[43] Söderhjelm, P.; Öhrn, A.; Ryde, U.; Karlström, G. *J. Chem. Phys.* **2008**, *128*, 014102.

[44] Ren, P.; Ponder, J. W. *J. Comput. Chem.* **2002**, *23*, 1497.

[45] Miyamoto, S.; Kollman, P. A. *Proteins: Struct. Funct. Genet.* **1993**, *16*, 226.

[46] Wang, J.; Dixon, R.; Kollman, P. A. *Proteins: Struct. Funct. Genet.* **1999**, *34*, 69.

[47] Kuhn, B.; Kollman, P. A. *J. Med. Chem.* **2000**, *43*, 3786.

[48] Weis, A.; Katebzadeh, K.; Söderhjelm, P.; Nilsson, I.; Ryde, U. *J. Med. Chem.* **2006**, *49*, 6596.

[49] Pugliese, L.; Coda, A.; Malcovati, M.; Bolognesi, M. *J. Mol. Biol.* **1993**, *231*, 698.

[50] 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. *Computational Materials Science* **2003**, *28*, 222.

[51] Veryazov, V.; Widmark, P.-O.; Serrano-Andrés, L.; Lindh, R.; Roos, B. O. *Int. J. Quantum Chem.* **2004**, *100*, 626.

[52] *MOLCAS 7*, University of Lund, Sweden **2007**. see http://www.teokem.lu.se/molcas.

[53] 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, UK **2006**. see http://www.molpro.net.

[54] Beebe, N. H. F.; Linderberg, J. *Int. J. Quantum Chem.* **1977**, *7*, 683.

[55] Koch, H.; A. Sánchez de Merás; Pedersen, T. B. *J. Chem. Phys.* **2003**, *118*, 9481.

[56] Aquilante, F.; Pedersen, T. B.; Lindh, R. *J. Chem. Phys.* **2007**, *126*, 194106.

[57] Aquilante, F.; Pedersen, T. B. *Chem. Phys. Letters* **2007**, *449*, 354.

[58] Aquilante, F.; Malmqvist, P.-Å.; Pedersen, T. B.; Ghosh, A.; Roos, B. O. *J. Chem. Theory. Comput.* **2008**, *4*, 694.

[59] 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.

[60] Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.

[61] Fedorov, D. G. (internal communication).

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. Various other quantities discussed in the method section (e.g. Fig. 2) are also given. For cutting scheme a, molecule O equals A. Energies are in kJ/mol.

	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_{O}^{ind}		-27.1	-32.3	-200.7*
$E_{O(B)}^{ind}$		-9.0	-8.7	-104.2*
E_{OB}^{ind}		-163.4	-162.1	-202.9*

* In the Amber polarization model, any molecule is "pre-polarized", i.e. even molecule B. Defining U to be the non-polarized version of B, the values given are $E_{O}^{ind} + E_{U}^{ind}$, $E_{O(U)}^{ind} + E_{U(O)}^{ind}$, and E_{OU}^{ind}, respectively.

Table 2: Classical energies ($E^{ele}_{XB} + E^{ind}_{XB}$) and rest terms ($E^{sup}_{XB} - E^{ele}_{XB} - E^{ind}_{XB}$) listed for each fragment X for the main structure, at the HF and MP2 level, respectively. Energies are in kJ/mol.

		HF level		**MP2 level**	
X	c_i	**Classical**	**Rest**	**Classical**	**Rest**
A_1	1	-18.7	0.8	-17.3	-1.9
A_2	1	-0.3	-0.1	-0.4	-1.5
A_3	1	-42.2	27.0	-37.3	18.6
A_4	1	-63.9	38.1	-56.6	24.0
A_5	1	-230.9	65.8 , 65.2*		11.0*
A_{5a}	1	-20.8	21.1	-18.6	6.9
A_{5b}	1	-14.4	-0.4	-12.6	-3.4
A_{5c}	1	-22.4	12.1	-21.0	-3.4
A_{5d}	1	-123.8	31.1	-115.0	7.7
A_{5e}	1	-167.7	33.2	-154.6	12.7
A_{5f}	1	-92.5	4.4	-83.3	-6.3
A_{5ab}	-1	-7.3	-0.2	-5.8	-1.8
A_{5bc}	-1	-6.6	-0.2	-6.4	-1.8
A_{5cd}	-1	-15.3	4.8	-13.4	-4.7
A_{5de}	-1	-98.8	28.4	-93.2	15.6
A_{5ef}	-1	-77.8	3.5	-70.6	-3.9
A_6	1	-1.3	26.4	-0.6	-6.2
A_7	1	-1.1	6.7	-1.3	-1.7
A_8	1	-123.2	72.4	-114.4	59.8
A_9	1	-58.3	5.7	-54.4	-1.9
A_{10}	1	-8.4	27.8	-8.2	16.1
A_{11}	1	-3.6	3.2	-3.5	-3.3
A_{12}	1	-0.4	8.5	0.9	-7.2
A_{13}	1	-3.5	3.3	-3.5	-5.4
A_{14}	1	-22.2	7.3	-17.7	1.2
A_{15}	1	-9.7	19.8	-8.0	-0.9

*Summed result using cutting scheme *c*.

Table 3: Results for the geometry set. The letters in brackets indicate the cutting scheme used for each term. For ΔE^{PMISP}, the two letters indicate the cutting scheme used for the classical (E^{ele}, E^{ind}) and E^{rest} terms, respectively. The mean absolute error (MAE) compared to E_{ref} is also reported. Energies are in kJ/mol.

	Geometry										**MAE**
	1	2	3	4	5	6	7	8	9	10	
HF level:											
E_{ref}	-252.2	-292.1	-264.6	-260.2	-266.1	-271.4	-244.9	-277.9	-304.1	-230.6	
$E^{ele}(a)$	-419.4	-414.4	-407.6	-442.4	-397.0	-400.4	-374.2	-415.2	-436.8	-335.8	
$E^{ele}(b)$	-420.5	-414.5	-409.6	-444.4	-396.5	-398.6	-373.7	-417.0	-438.4	-338.6	1.4*
$E^{ele}(c)$	-419.1	-412.3	-407.9	-442.5	-395.3	-397.0	-372.8	-414.9	-435.4	-337.4	1.2*
$E^{ind}(a)$	-164.6	-173.2	-170.7	-190.4	-162.6	-161.9	-152.9	-172.4	-187.9	-127.6	
$E^{ind}(b)$	-154.5	-160.0	-157.4	-178.6	-151.7	-148.7	-140.4	-160.2	-175.8	-119.9	11.7†
$E^{ind}(c)$	-153.4	-156.4	-155.1	-176.7	-149.3	-147.5	-138.2	-157.5	-172.2	-117.4	14.0†
$E^{rest}(b)$	312.7	271.3	294.8	347.8	273.4	265.0	260.0	285.0	296.0	219.5	
$E^{rest}(c)$	312.1	266.1	291.7	346.2	271.1	263.2	256.5	281.0	290.3	216.3	
$\Delta E^{PMISP}(ac)$	-19.6	-29.4	-21.9	-26.4	-22.4	-27.5	-25.6	-28.8	-30.3	-16.5	24.9
$\Delta E^{PMISP}(bc)$	-10.6	-16.4	-10.6	-16.6	-11.0	-12.6	-12.7	-18.3	-19.9	-11.7	14.0
$\Delta E^{PMISP}(cc)$	-8.1	-10.6	-6.7	-12.8	-7.4	-9.8	-9.5	-13.5	-13.2	-8.0	10.0
$\Delta E^{PMISP}(bb)$	-10.0	-11.1	-7.6	-15.0	-8.6	-10.8	-9.2	-14.3	-14.2	-8.5	10.9
$\Delta E^{EMB\text{-}PMISP}(bb)$	0.2	1.6	4.6	4.2	4.1	2.3	4.9	-0.1	1.2	-3.2	2.6
$\Delta E^{EMB(NP)\text{-}PMISP}(bb)$	0.1	0.5	1.9	0.8	2.3	0.9	3.0	-2.8	-2.4	-3.1	1.8
$\Delta E^{PA}(b)$	-22.8	-31.9	-26.5	-36.5	-26.6	-36.3	-20.6	-37.3	-27.3	-17.6	28.3
$\Delta E^{FMO2}(b)$	9.7	16.2	13.2	17.9	11.9	12.6	14.7	14.2	13.0	12.4	13.6
$\Delta E^{EE\text{-}PA}(b)$	12.0	19.2	15.1	18.2	13.8	16.8	16.0	16.4	15.3	13.5	15.6
HF level (no CP):											
E_{ref}	-358.9	-401.9	-369.5	-368.8	-368.0	-377.7	-347.8	-386.2	-412.9	-329.8	
$\Delta E^{PA}(b)$	-47.3	-62.3	-53.5	-68.6	-51.2	-68.5	-50.0	-68.3	-55.4	-42.9	56.8
$\Delta E^{FMO2}(b)$	-14.9	-14.2	-13.7	-14.2	-12.7	-19.7	-14.8	-16.7	-15.0	-12.9	14.9
$\Delta E^{FMO3}(b)$	0.4	0.5	0.4	0.3	0.4	1.2	0.5	0.3	0.6	0.0	0.5
$\Delta E^{EE\text{-}PA}(b)$	-12.6	-11.3	-11.9	-13.9	-10.8	-15.4	-13.5	-14.6	-12.7	-11.7	12.8
MP2 level:											
E_{ref}	-411.9	-446.1	-419.8	-419.6	-418.7	-430.0	-395.4	-436.7	-452.1	-386.8	
$E^{ele}(c)$	-372.6	-368.0	-364.3	-392.5	-354.0	-354.4	-329.4	-370.7	-386.8	-301.8	
$E^{ind}(c)$	-146.6	-149.2	-148.6	-168.4	-142.0	-140.9	-131.9	-150.2	-163.9	-112.2	
$E^{rest}(c)$	100.6	61.7	87.0	129.2	71.6	56.8	57.9	69.8	85.0	21.7	
$\Delta E^{PMISP}(cc)$	-6.7	-9.4	-6.1	-12.1	-5.7	-8.4	-8.0	-14.4	-13.7	-5.6	9.0
$\Delta E^{PMISP\text{-}CE}(cc)$	1.5	1.2	0.6	0.7	1.8	1.4	1.5	-0.9	-0.5	2.4	1.2
$\Delta E^{PA\text{-}CE}(c)$	2.8	3.0	2.3	2.8	2.6	3.1	3.1	0.3	0.9	2.3	2.3
$\Delta E^{PA\text{-}CE[3\text{-}21G]}(c)$	-0.2	5.6	3.5	2.1	3.2	2.0	0.8	3.3	3.8	-0.4	2.5

*Relative to $E^{ele}(a)$. †Relative to $E^{ind}(a)$.

Table 4: Results for the ligand set. The letter in brackets indicates the cutting scheme used for that particular term. For ΔE^{PMISP}, the two letters indicate the cutting scheme used for the classical (E^{ele}, E^{ind}) and E^{rest} terms, respectively. The mean absolute error (MAE) compared to E_{ref} is also reported. Energies are in kJ/mol.

	Ligand							**MAE**
	BTN1	BTN2	BTN3	BTN4	BTN5	BTN6	BTN7	
HF level:								
E_{ref}	-252.2	-269.0	-227.0	-31.6	-3.4	-47.4	-63.2	
$E^{ele}(a)$	-419.4	-427.3	-401.8	-185.1	-107.1	-119.4	-125.3	
$E^{ele}(b)$	-420.5	-427.8	-401.2	-185.3	-106.9	-119.0	-124.6	0.5*
$E^{ele}(c)$	-419.1	-425.4	-399.1	-185.3	-107.0	-119.4	-125.1	0.8*
$E^{ind}(a)$	-164.6	-190.0	-181.3	-55.8	-50.6	-40.7	-37.1	
$E^{ind}(b)$	-154.5	-176.7	-165.8	-54.1	-46.3	-39.0	-35.5	6.9†
$E^{ind}(c)$	-153.4	-173.2	-162.7	-54.2	-46.3	-39.0	-35.6	8.0†
$E^{rest}(b)$	312.7	322.4	330.3	209.3	154.7	110.1	96.1	
$E^{rest}(c)$	312.1	318.3	324.7	209.1	154.8	110.1	95.8	
$\Delta E^{PMISP}(ac)$	-19.6	-29.9	-31.5	-0.2	0.5	-2.7	-3.3	12.5
$\Delta E^{PMISP}(bc)$	-10.6	-17.2	-15.4	1.4	5.1	-0.5	-1.1	7.3
$\Delta E^{PMISP}(cc)$	-8.1	-11.3	-10.2	1.3	4.9	-1.0	-1.6	5.5
$\Delta E^{PMISP}(bb)$	-10.0	-13.1	-9.7	1.6	5.0	-0.5	-0.9	5.8
$\Delta E^{EMB\text{-}PMISP}(bb)$	0.2	5.9	6.5	2.9	9.0	5.0	5.3	5.0
$\Delta E^{EMB(NP)\text{-}PMISP}(bb)$	0.1	1.6	5.1	2.5	8.8	3.9	4.9	3.8
$\Delta E^{PA}(b)$	-22.8	-27.2	-42.0	9.9	18.2	8.6	7.1	19.4
$\Delta E^{FMO2}(b)$	9.7	8.9	0.0	7.9	4.9	6.0	7.6	6.4
$\Delta E^{EE\text{-}PA}(b)$	12.0	11.4	3.0	9.1	5.1	7.4	8.0	8.0
HF level (no CP):								
E_{ref}	-358.9	-377.8	-334.8	-101.9	-56.4	-98.1	-100.5	
$\Delta E^{PA}(b)$	-47.3	-54.5	-54.5	-7.0	4.2	-4.4	-5.6	25.4
$\Delta E^{FMO2}(b)$	-14.9	-18.5	-12.5	-9.0	-9.1	-7.1	-5.0	10.8
$\Delta E^{FMO3}(b)$	0.4	1.4	4.4	0.6	1.0	0.3	0.1	1.2
$\Delta E^{EE\text{-}PA}(b)$	-12.6	-16.0	-9.5	-7.8	-8.8	-5.6	-4.6	9.3
MP2 level:								
E_{ref}	-411.9	-432.7	-372.4	-183.8	-124.6	-148.2	-119.0	
$E^{ele}(c)$	-372.6	-381.9	-355.2	-155.0	-91.9	-103.3	-112.4	
$E^{ind}(c)$	-146.6	-164.6	-156.0	-48.2	-42.4	-36.7	-33.6	
$E^{rest}(c)$	100.6	104.8	126.2	26.6	18.8	-6.0	28.5	
$\Delta E^{PMISP}(cc)$	-6.7	-9.0	-12.6	7.2	9.1	2.4	1.5	6.9
$\Delta E^{PMISP\text{-}CE}(cc)$	1.5	2.3	-2.5	6.0	4.1	3.4	3.1	3.3
$\Delta E^{PA\text{-}CE}(c)$	2.8	3.4	-0.3	6.9	2.7	4.9	3.4	3.5
$\Delta E^{PA\text{-}CE[3\text{-}21G]}(c)$	-0.2	9.2	10.8	1.7	2.6	-0.1	0.4	3.6

*Relative to $E^{ele}(a)$. †Relative to $E^{ind}(a)$.

Table 5: Contributions to the energy difference between the EMB-PMISP and PMISP methods from each fragment for the main structure (MS), as well as the mean absolute contributions for the geometry set (Gset) and ligand set (Lset). The contributions from polarizabilities (i.e. the difference between EMB-PMISP and EMB(NP)-PMISP) are given within brackets. Energies are in kJ/mol.

Fragment	**MS**		**Gset**		**Lset**	
A_1	1.1	(0.2)	1.7	(0.5)	1.9	(0.4)
A_2	0.0	(0.0)	0.1	(0.1)	0.1	(0.1)
A_3	2.7	(0.0)	3.4	(0.4)	3.3	(0.3)
A_4	3.3	(0.7)	3.3	(0.7)	3.4	(0.7)
A_5	1.6	(-0.5)	1.3	(0.5)	0.9	(0.5)
A_6	-2.2	(-0.6)	1.3	(0.5)	1.0	(0.2)
A_7	-0.6	(-0.3)	0.4	(0.2)	0.3	(0.1)
A_8	2.1	(0.4)	3.2	(0.9)	1.1	(0.3)
A_9	2.9	(0.2)	3.1	(0.6)	1.2	(0.2)
A_{10}	0.7	(0.2)	0.6	(0.2)	0.3	(0.2)
A_{11}	0.1	(0.0)	0.3	(0.1)	0.1	(0.0)
A_{12}	-0.5	(0.1)	0.6	(0.1)	0.3	(0.1)
A_{13}	0.5	(0.2)	0.5	(0.3)	0.3	(0.1)
A_{14}	-0.5	(-0.4)	0.4	(0.3)	0.7	(0.3)
A_{15}	-1.2	(-0.1)	0.9	(0.1)	0.7	(0.1)
Total	10.2	(0.1)	12.8	(1.9)	8.8	(1.2)

Table 6: Trimer results for the main structure. Only the fragment pairs X–Y contributing more than 0.5 kJ/mol to Eq. 12 are explicitly 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.

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

Figure captions

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 $c_i = -1$.

Figure 2: Schematic picture of the relations between various energy terms described in the text for an example case where monomer A consists of two fragments.

Figure 3: 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.

Figure 4: Two-dimensional cartoon of the avidin model interacting with biotin (BTN1). It should give 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} to 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 5: Correlation between the supermolecular (HF) interaction energies and the energies estimated by various methods. Note that the axes are broken to accomodate the results of both the charged and neutral ligands. The line represents perfect correlation.

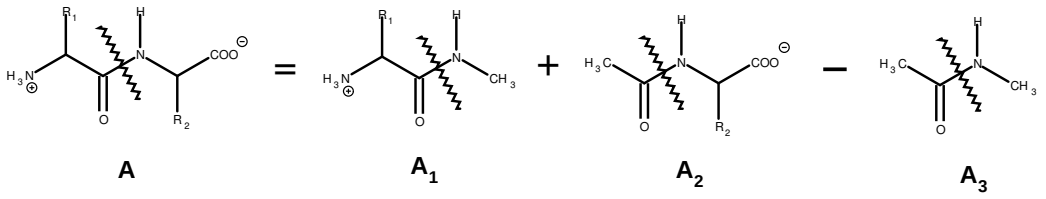

Figure 1: Example illustrating the MFCC procedure for cutting molecule A across the peptide bond and capping with –$COCH_3$ and –$NHCH_3$ groups. The result is two capped fragments A_1 and A_2 with $c_i = 1$, as well as a concap fragment A_3 with $c_i = -1$.

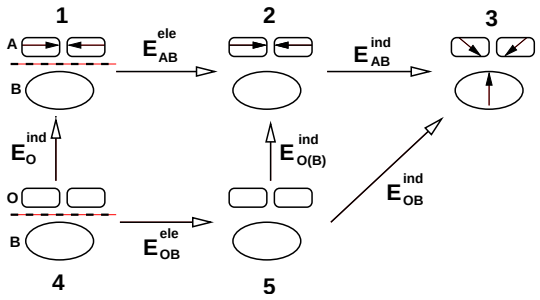

Figure 2: Schematic picture of the relations between various energy terms described in the text for an example case where monomer A consists of two fragments.

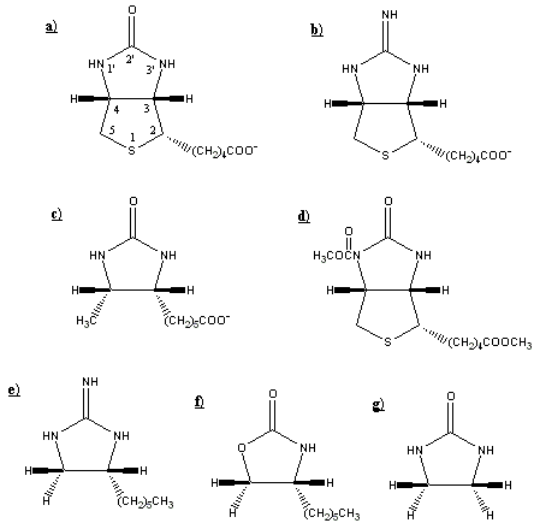

Figure 3: The seven ligands to avidin used in this study: a) BTN1 (biotin), b) BTN2, c) BTN3, d) BTN4, e) BTN5, f) BTN6, g) BTN7. The first three have a molecular charge of -1, whereas the other ligands are neutral.

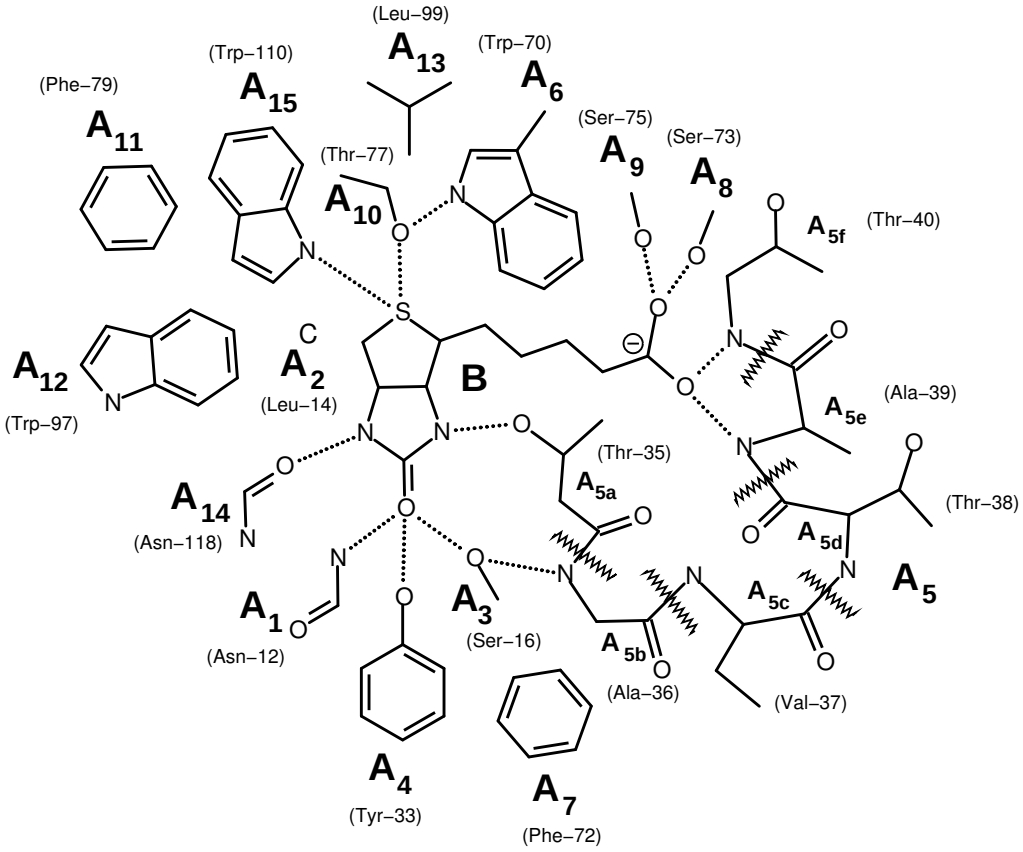

Figure 4: Two-dimensional cartoon of the avidin model interacting with biotin (BTN1). It should give a guidance to the location of each fragment; in reality, the fragments surround the ligand completely. For clarity, hydrogen atoms are omitted. The most prominent hydrogen bonds are indicated by dotted 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} to 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.

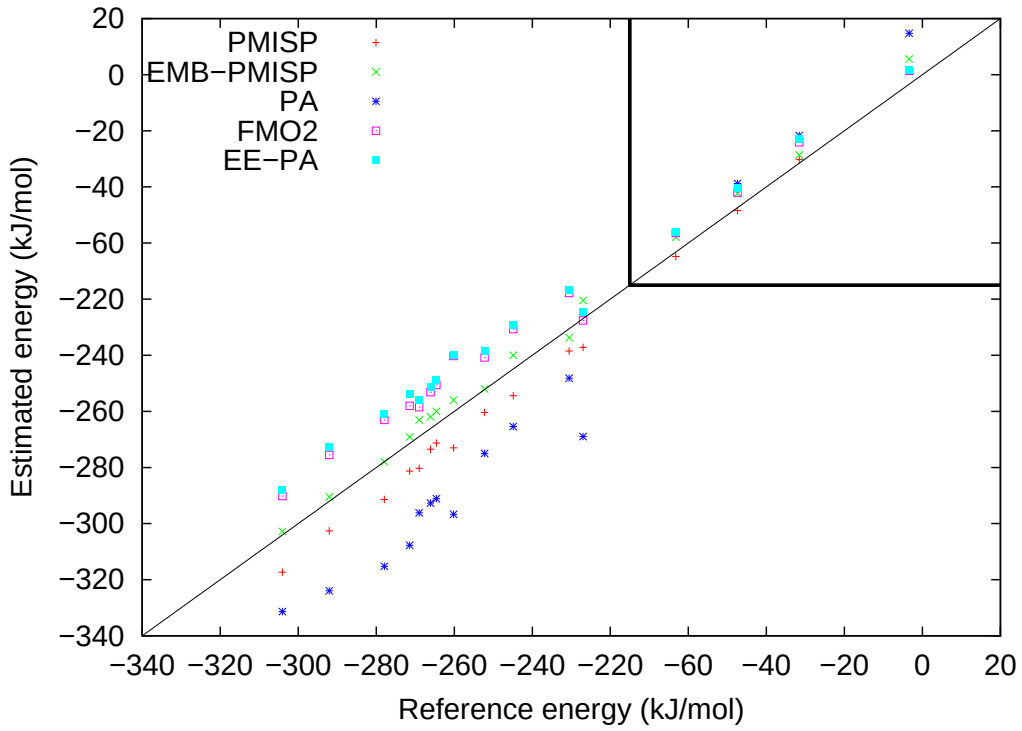

Figure 5: Correlation between the supermolecular (HF) interaction energies and the energies estimated by various methods. Note that the axes are broken to accomodate the results of both the charged and neutral ligands. The line represents perfect correlation.