JCTC Journal of Chemical Theory and Computation

Accurate Induction Energies for Small Organic Molecules. 2. Development and Testing of Distributed Polarizability Models against SAPT(DFT) Energies

Alston J. Misquitta,[†,‡] Anthony J. Stone,*[,†] and Sarah L. Price[‡]

University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, U.K., and University College London, 20 Gordon Street, London WC1H 0AJ, U.K.

Received May 3, 2007

Abstract: In part 1 of this two-part investigation we set out the theoretical basis for constructing accurate models of the induction energy of clusters of moderately sized organic molecules. In this paper we use these techniques to develop a variety of accurate distributed polarizability models for a set of representative molecules that include formamide, *N*-methyl propanamide, benzene, and 3-azabicyclo[3.3.1]nonane-2,4-dione. We have also explored damping, penetration, and basis set effects. In particular, we have provided a way to treat the damping of the induction expansion. Different approximations to the induction energy are evaluated against accurate SAPT(DFT) energies, and we demonstrate the accuracy of our induction models on the formamide−water dimer.

I. Introduction

In this paper, which is the second part in a two-part investigation[1] of the induction energy, we report methods for developing models of the induction energy that are suitable for applications involving organic molecules. In the first part of our study, which we will refer to as part 1, we set out the theoretical framework for the calculation of the nonexpanded and expanded induction energies in a way that is suitable for clusters of organic molecules. We now use that framework to develop the distributed polarizability models needed to model the induction energy in the condensed phase and demonstrate how the refinement procedure described in part 1 can be used to obtain polarizability models that combine accuracy and computational simplicity, making them ideal for real-world applications, in particular those in the field of organic crystal structure prediction.

The dominant part of the induction contribution to the total interaction energy of a cluster of molecules arises from the two-body interactions. In part 1 we argued that the SAPT(DFT) two-body interaction energy should be defined as

$$U = E^{(1)}_{\mathrm{elst}}(\mathrm{KS}) + E^{(1)}_{\mathrm{exch}}(\mathrm{KS}) + E^{(2)}_{\mathrm{ind,tot}} + E^{(2)}_{\mathrm{disp}} + E^{(2)}_{\mathrm{exch-disp}} + U^{(3-\infty)} \quad (1)$$

where $E^{(1)}_{\mathrm{elst}}$ (KS) and $E^{(1)}_{\mathrm{exch}}$ (KS) are the first-order electrostatic and exchange energies, respectively, $E^{(2)}_{\mathrm{disp}}$ and $E^{(2)}_{\mathrm{exch-disp}}$ are the second-order dispersion and exchange-dispersion energies, respectively, and $E^{(2)}_{\mathrm{ind,tot}}$ and $U^{(3-\infty)}$ were defined in part 1. The induction contribution to the two-body interaction energy arises at second and higher orders in the interaction operator. The most computationally efficient and accurate method for the calculation of the two-body induction energy at second order is based on the recently developed symmetry-adapted perturbation theory based on density functional theory, called SAPT(DFT)[2−4] or DFT-SAPT.[5] Higher-order induction energies form the bulk of the higher-order effects for polar molecules and can be very important for hydrogen-bonded dimers, where they can contribute as much as 17% of the two-body energy. In part 1 we described how induction contributions to the two-body energy that arise from terms of third and higher order in the interaction operator can be approximated within SAPT(KS).[3,6] Since a SAPT(DFT) calculation of the interaction energy uses energies computed within SAPT(KS), we are now able to calculate accurate nonexpanded induction energies within one theoretical framework. Additionally, due to the modest

* Corresponding author phone: +44 1223 336375; fax: +44 1223 336362; e-mail: ajs1@cam.ac.uk.
† University Chemical Laboratory.
‡ University College London.

10.1021/ct700105f CCC: $40.75
Published on Web 11/10/2007

computational resources needed for such a calculation, we are able to apply these methods to dimers of small organic molecules.

What remains then is for us to model the induction energy in a way that is suitable for calculations on clusters of polar molecules. In such clusters, many-body induction energies can be almost as important as the two-body induction energies. Both can be described using the damped classical polarizable model,[7] and in fact higher-order contributions to the many-body energy can also be obtained from this model. The latter, which can form as much as 12% of the total energy of the cluster, are obtained from the classical model by iterating the fields and the responses to the fields to self-consistency (see section IV.2 in part 1).

We need accurate molecular polarizabilities for the classical polarizable model. For all but the smallest molecules, they are needed in a distributed form. Recently two of us have developed a method of obtaining distributed polarizabilities based on a partitioning of the molecular transition densities using a constrained density-fitting technique that is both accurate and applicable to large molecules.[8] In ref 8 we observed that the accuracy of the model deteriorates if it is simplified to include only local polarizability terms, that is to omit terms involving pairs of sites. This deterioration can be overcome by refining the polarizability model by the method of Williams and Stone.[9] This two-step procedure has an advantage over the Williams and Stone method alone, which can lead to unphysical nonpositive-definite terms. In the combined procedure, such terms are completely removed from the low-ranking polarizabilities and greatly reduced in the higher-ranking terms. In a recent review[7] some preliminary results of this Williams−Stone−Misquitta (WSM) procedure were presented, and in part 1 we gave its theoretical basis in some detail. Here we describe the numerical details of the procedure and present extensions of the method that can be used to calculate local polarizability models up to rank 2. Rank 1, that is, dipole−dipole, polarizabilities may be sufficient for calculations of moderate accuracy, but the higher-rank dipole−quadrupole and quadrupole−quadrupole terms are needed to achieve higher accuracy. On the other hand, for applications that are so computationally demanding as to require a simpler model, the WSM procedure can provide the best accuracy subject to such constraints.

At short intermolecular separations the classical polarizable model will be in error, and the difference between the nonexpanded induction energies calculated using SAPT(DFT) and the energies from the classical model will need to be accounted for. Additionally, at short separations, the classical polarizable model, which is based on a multipole expansion, can result in divergent energies. This problem arises quite often in the condensed phase. We have analyzed the problems associated with short intermolecular separations in some detail and have tried to provide solutions to many of them.

This paper is organized as follows: In section II we describe a powerful graphical technique for displaying the model induction energies. In section III we discuss the features of basis sets that are needed to obtain accurate molecular polarizabilities and also SAPT(DFT) induction energies.

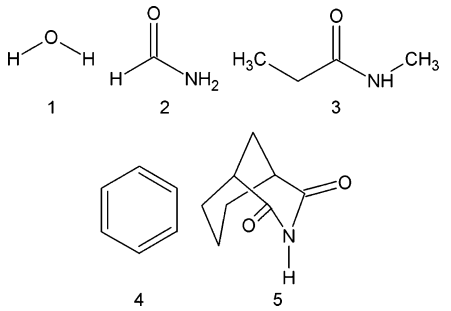

Figure 1. Molecules used to test distributed polarizability models: (1) water, (2) formamide, (3) *N*-methyl propanamide (N-MPA), (4) benzene, and (5) 3-azabicyclo[3.3.1]nonane-2,4-dione(BOQQUT).

The methods that we describe here for distributed polarizabilities can be used to obtain models of varying complexity and accuracy. In section IV we assess these models for the molecules shown in Figure 1, which have been chosen to provide a range of sizes and different types of charge distribution. Benzene tests the modeling of the polarizability of conjugated systems, in contrast to the saturated hydrocarbon functional groups. The other molecules give a range of hydrogen bond donor and acceptor strengths: the imide has a plastic phase[10] indicating that the hydrogen-bonding in its ordered polymorphs is readily disrupted.

In section V we discuss effects of penetration, truncation, and damping and propose a way to determine the damping coefficient based on molecular ionization energies. In section VI we assess approximations for calculating the induction energies. Finally, in section VII we conclude with a summary of the main results of this paper.

I.1. A Note on Notation. The notation we have used for the induction energies defined within SAPT(DFT) is somewhat nonstandard. We have described it in some detail in section III of part 1, but the key ideas are summarized here for convenience.

At order n, there are two components to the induction: the induction as defined through the polarization expansion,[11,12] termed $E^{(n)}_{\mathrm{ind,pol}}$, and the exchange component of the induction, termed $E^{(n)}_{\mathrm{ind,exch}}$. In part 1 we defined the nth order induction energy as the sum of these contributions, i.e., $E^{(n)}_{\mathrm{ind,tot}} = E^{(n)}_{\mathrm{ind,pol}} + E^{(n)}_{\mathrm{ind,exch}}$. The reasons for this definition, rather than the more conventional identification of $E^{(n)}_{\mathrm{ind,pol}}$ as the nth order induction, have been outlined in section III in part 1. In brief, our choice has been made because Coulomb singularities in the interaction operator mean that neither $E^{(n)}_{\mathrm{ind,pol}}$ nor $E^{(n)}_{\mathrm{ind,exch}}$ are meaningful on their own,[13,14] and the observation that the expanded induction energy, termed $E^{(n)}_{\mathrm{ind,d-class}}$, agrees best with $E^{(n)}_{\mathrm{ind,tot}}$ as defined here. Numerical evidence will be provided below.

II. Displaying the Energies

A powerful way to understand the various models that will be presented below is by mapping energies onto a suitable surface around the molecule in question. Such a mapping

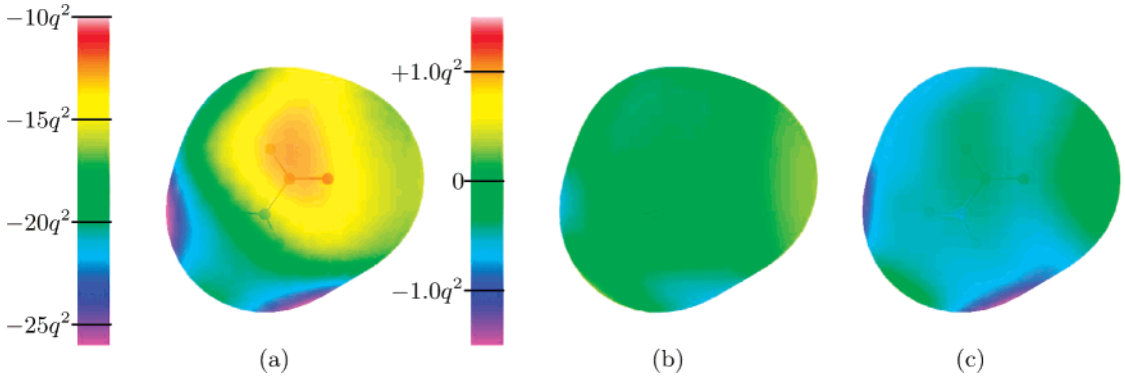

Figure 2. Induction energy map and difference maps (kJ mol^{-1}) arising from a charge q atomic units on the vdW $\times$ 2 surface of formamide obtained using a rank 4 nonlocal polarizability description. In (a) is displayed the induction energy map obtained using distributed polarizabilities calculated with the Sadlej basis set. Also shown are maps comparing the induction energies obtained using distributed polarizabilities calculated with the (b) aug-cc-pVTZ and (c) aug-cc-pVQZ bases with the Sadlej-basis energies.

can be most easily done if the energy probe has spherical symmetry, and a convenient probe for the induction energy is a point charge. The maps in this paper show the induction energies that result from a charge qe on the chosen surface. The SAPT(DFT) expression for the induction energy of a molecule in the field of a point charge is given by eq 17 in part 1. The induction energy of a molecule in the field of a point charge depends quadratically on the magnitude of the charge, and an appropriate value of q needs to be used in interpreting the energy scales in these maps. Setting $q = 1$ gives the response to a unit charge, but this is larger than typical local charges in a molecule, which are not expected to exceed 0.5e.

The surface around the molecule is constructed as follows. If the required distance from atom a is R_a^0 (e.g., twice the van der Waals radius for the vdW $\times$ 2 surface), then the surface is defined by $R_a - R_a^0 = 0$, or equivalently by $\exp[-\xi(R_a - R_a^0)] = 1$, where ξ is an arbitrary constant. We define the surface for the whole molecule by $\sum_a \exp[-\xi(R_a - R_a^0)] = 1$. The effect is as if we shrank an elastic membrane onto the union of vdW $\times$ 2 atomic surfaces, more or less tightly depending on the value of ξ; the intersections between the vdW surfaces of neighboring atoms are smoothed out. A value of $\xi = 2$ has been used for the maps shown here. The SAPT(DFT) maps were generated using the CamCASP program,[15] and the maps using distributed polarizabilities were generated with the ORIENT program, version 4.6.[16]

The van der Waals radii prescribed by Bondi[17] have been used for all but the hydrogen atoms that can form hydrogen bonds, which have their radii set to zero in order to reflect better the small interatomic distances associated with hydrogen bonds. The surface is then approximately the surface of contact for neighboring non-hydrogen atoms.

III. Choice of Basis

We will discuss the basis set requirements for obtaining accurate molecular properties and accurate SAPT(DFT) energies separately, because the requirements for large and small intermolecular separations are generally quite different. At short range orbital overlap effects become important and we generally need to supplement the basis sets used for SAPT(DFT) calculations with additional functions, while at long range the multipole expansion can be used and all we need are basis-saturated molecular properties.

III.1. Basis Set Requirements for Molecular Polarizabilities and Multipole Moments. The large size of even the smaller organic molecules makes the use of basis sets higher than triple-ζ quality hard to use on a routine basis. In fact, sufficiently accurate molecular polarizabilities are obtained with the Sadlej basis set[18,19] *if* used within the linear response Kohn−Sham DFT framework described in section IV of part 1. As an example, consider the formamide molecule. The induction map using distributed rank 4 nonlocal polarizabilities obtained with the Sadlej basis is displayed in Figure 2, together with the corresponding results for the aug-cc-pVTZ and aug-cc-pVQZ basis sets, displayed as *difference* maps against the Sadlej basis results. As would be expected, the differences are very small for the aug-cc-pVTZ basis, indicating that this basis is roughly equivalent to the Sadlej, at least for calculations of the induction energy. The differences are somewhat larger for the aug-cc-pVQZ basis, but, even in this case, the largest difference is around 1.5q^2 kJ mol^{-1} which is an order of magnitude less than the actual energies. For a more realistic charge of 0.5 units, the maximum difference would be about 0.4 kJ mol^{-1}.

We conclude that the Sadlej basis sets provide a very good compromise between size and accuracy for our purposes. They have been optimized for molecular properties but are about half the size of the equivalent aug-cc-pVTZ Dunning basis sets, thus significantly raising the limit on the size of molecules that can be used in such calculations.

As explained in section IV of part 1, we use density-fitting techniques in our calculations of the nonexpanded induction energy and distributed polarizabilities. As yet, there is no auxiliary density-fitting basis optimized for the Sadlej basis, so in view of the similarities between the Sadlej and aug-cc-pVTZ bases we have used the aug-cc-pVTZ auxiliary basis instead. We have confirmed that this is a good choice by carrying out extensive tests of molecular properties and interaction energies.

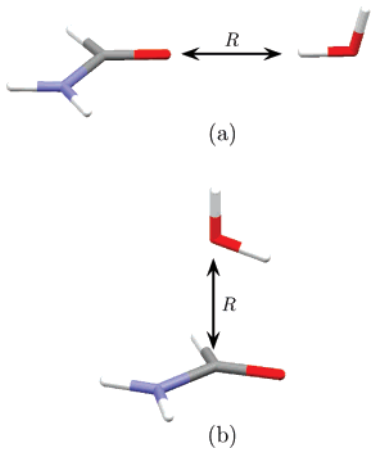

Figure 3. Formamide water dimer geometries used in this paper. Geometry (a) emphasizes the hydrogen oxygen contacts and geometry (b) emphasizes the contacts between the heavy atoms. The radial minima in the interaction energy for geometries (a) and (b) are at approximately 4 and 5.5 bohr, respectively.

III.2. Basis Set Requirements for the SAPT(DFT) Induction Energies. Monomer basis sets of triple-ζ quality are generally suitable for accurate SAPT(DFT) induction energies if the intermolecular separation is large and overlap effects are negligible. Basis sets with functions located only on the atomic sites of the monomer are said to be of the 'monomer centered' or MC type. For intermediate and short molecular separations where overlap effects are significant, basis-converged SAPT(DFT) energies are obtained only with basis sets supplemented with functions located on the nuclei of the interacting partners.[20] These are the so-called 'far-bond' functions which typically comprise just the *s* and *p* symmetry functions of the interacting monomer. To saturate the dispersion energy, a further small set of functions is needed in the region between the interacting molecules—the so-called 'mid-bond' functions. The resulting basis set is said to be of the MC$^+$ type, where the '+' sign indicates the presence of the additional basis functions. Although the mid-bond functions have a negligible effect on the induction energies, we will include them in all SAPT(DFT) calculations, for consistency with later work.

The effect of the far-bond functions on the polarization expression for the induction and exchange-induction energies *individually* is rather dramatic.[20] At the minimum-energy intermolecular separation, these energies can increase in magnitude by 2 orders of magnitude upon inclusion of the far-bond functions; but $E^{(n)}_{\mathrm{ind,pol}}$ is significantly quenched by $E^{(n)}_{\mathrm{ind,exch}}$, at every order *n*.[21] Consequently, as discussed in sections III and IV.1 of part 1, it is more useful to consider $E^{(n)}_{\mathrm{ind,tot}}$, the sum of these two energies at each order.

In Figure 4 we display $E^{(2)}_{\mathrm{ind,tot}}$ and $E^{(3)}_{\mathrm{ind,tot}}$ calculated using the Sadlej basis in MC and MC$^+$ basis types for the formamide water dimer. The MC$^+$ energies are uniformly larger in magnitude (more negative) than the MC results. Near the minimum-energy separations, the difference between the MC$^+$ and MC results for geometry (a) is about 1 kJ mol^{-1} for $E^{(2)}_{\mathrm{ind,tot}}$ and 0.7 kJ mol^{-1} for $E^{(3)}_{\mathrm{ind,tot}}$. For geometry (b) the corresponding differences are about 0.5 and

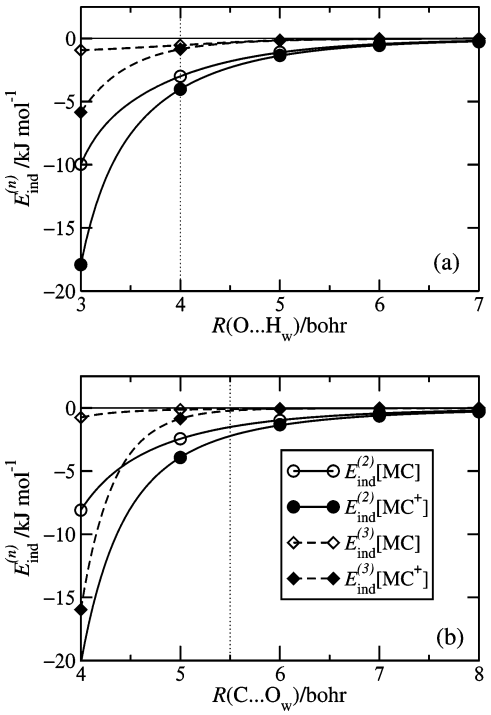

Figure 4. The effect of far-bond functions on the second- and third-order induction energies, respectively, for the formamide−water dimer. See Figure 3 for a description of the dimer geometries.

0.2 kJ mol^{-1}. These changes constitute around 6% of the total interaction energies near the minimum-energy separations for the two configurations. These are not small effects and cannot be ignored in accurate studies. Perhaps more importantly from the point of view of geometry optimizations, the MC$^+$ basis sets result in deeper wells and smaller radial separations, particularly for the hydrogen-bonding geometries.

The above picture remains the same if the aug-cc-pVTZ basis is used in place of the Sadlej basis. The differences in $E^{(2)}_{\mathrm{ind,tot}}$ and $E^{(3)}_{\mathrm{ind,tot}}$ calculated using these two bases are already small (though not negligible) for the MC basis type, and with the MC$^+$ type, the aug-cc-pVTZ and Sadlej bases yield essentially the same energies. We should emphasize here that this does not mean that the *total* interaction energy can be calculated using the Sadlej/MC$^+$ basis. This is because the basis incompleteness error in the second-order dispersion energy may not be negligible in this basis.[22]

IV. Which Model?

With the current version of the CamCASP program, distributed nonlocal polarizabilities can be calculated up to rank 4 using the constrained density-fitting algorithm.[8] If the molecule contains n_s sites and l_{max} is the highest rank included, the nonlocal polarizability model contains $O(n_s^2 l_{\mathrm{max}}^2)$ polarizability components. This can be many thousands for a rank 4 description of a molecule like BOQQUT, which contains 22 atoms. The computational cost of using such a polarizability model is already quite high for a pair of interacting molecules. For a cluster of molecules, a model of such complexity could be impossible to use. However,

Table 1. Maximum and rms Differences between the Model and SAPT(DFT) Second-Order Induction Energies of the Molecule Interacting with a Unit Charge on the vdW × 2 Surface[a]

	formamide		N-MPA		benzene		BOQQUT	
model	max	rms	max.	rms	max	rms	max	rms
NL1	5.40	2.79	7.63	3.95	5.35	3.84	7.35	4.93
NL2	1.48	0.48	1.42	0.57	1.60	0.66	1.92	0.78
NL3	0.57	0.27	0.36	0.17	0.30	0.19	0.54	0.15
NL4	0.56	0.26	0.37	0.17	0.27	0.17	0.57	0.14
L1	3.93	1.90	5.56	2.30	3.34	1.98	5.08	2.35
L2	6.85	1.37	4.66	2.02	3.71	1.74	7.90	1.75
L1,WSM	3.48	1.35	2.32	0.74	2.98	1.58	2.48	0.82
L2,WSM	1.14	0.27	1.29	0.21	0.69	0.25	2.27	0.23
L2/L1,WSM	1.18	0.34	0.99	0.34	0.67	0.42	1.48	0.32

[a] 'NL*n*' denotes a nonlocal model of rank *n*, 'L*n*' denotes a local model of rank *n* obtained using the Le Sueur and Stone[23] localization method, and 'L*n*,WSM' denotes a refined, local model of rank *n* obtained using the WSM procedure. The mixed-rank descriptions are denoted by 'L2/L1'. This means we have used a rank 2 local description on the heavy atoms and a rank 1 local description on the hydrogen atoms. All differences are reported in kJ mol^{-1}.

dramatic simplifications in the polarizability model can be made without significant losses in accuracy, using the localization techniques described in section IV.3 of part 1.

In Table 1 we report the maximum and rms errors made by several polarizability models in reproducing the induction energies of a molecule with a unit point charge placed on the vdW × 2 surface (as described in section II). Notice that the exchange part is zero in this case, because the point charge carries no electrons, so we are studying the ability of distributed polarizability models to reproduce $E_{\mathrm{ind,pol}}$. The exchange part is not negligible in general, but it is short-range in form and cannot in any case be described by a classical polarizability model.

The rank 1 nonlocal models are clearly inadequate, with maximum and rms errors in the range 3−8 kJ mol^{-1} or, for a more realistic charge of 0.5*e*, between 1 and 2 kJ mol^{-1}. These errors are dramatically reduced in the rank 2 nonlocal models and are still better for the rank 3 models for which the errors are only a few tenths of a kJ mol^{-1}. The rank 4 nonlocal models offer negligible improvements over the rank 3 models.

Curiously, at rank 1, the Le Sueur and Stone[23] localization technique results in local models that are *more accurate* than the nonlocal models they were constructed from. At present we do not understand why this is so. It is surprising because, as has been mentioned earlier, the Le Sueur and Stone localization procedure uses truncated multipole expansions to transform the nonlocal terms away, a procedure which is expected to cause a deterioration in the convergence properties of the model by increasing its sphere of divergence.

At rank 2 we see that the Le Sueur and Stone procedure does indeed lead to a deterioration compared with the nonlocal models. While the rms errors are slightly improved over the rank 1 local models, with the exception of *N*-methyl propanamide, the maximum errors are much larger, being in the range 4−8 kJ mol^{-1}, or, for a charge of 0.5*e*, 1−2 kJ mol^{-1}. These errors are probably too large for most applications.

We can obtain more accurate local models by refining the results of the Le Sueur and Stone localization procedure, as described in section IV.3 of part 1. For this refinement, which is the last stage in the WSM procedure, we need to choose the coefficients $g_{kk'}$, in eq 36 in part 1, which determine the weight given to the 'anchors' of the polarizability values. We have found that the following values lead to accurate models while minimizing the occurrence of unphysical terms:

$$g_{kk'} = 0 \text{ if } k \neq k'$$

$$g_{kk} = \begin{cases} 10^{-5} & \text{if } k \in \{10, 10c, 10s\} \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

The dipole−dipole polarizabilities from the constrained density-fitting and Le Sueur and Stone procedure are usually quite accurate, so they are used as anchor values and given nonzero weight, while the higher ranking polarizabilities can be quite poor and are given zero weight in the WSM procedure. The accuracy of the polarizability descriptions obtained from the constrained-DF method varies considerably with molecule size, so a system-dependent weight might yield better results. While the above choice of the weights has been found to be appropriate for all the molecules studied in this work, it is possible that there will be exceptions which will require another choice. In general, the errors made by the models in reproducing the point-to-point polarizabilities or the induction energy with a point charge (see below) should be monitored. As a rule of thumb, the maximum and rms errors in the point-to-point polarizabilities as percentages of the range[8] should be less than 6% and 0.2%, respectively, for a rank 1 model and less than 2% and 0.05% for a rank 2 model.

From Table 1 we see that the WSM rank 1 local models are already rather good, with maximum and rms errors of around 3 and 1 kJ mol^{-1}, respectively, for unit probe charge. At rank 2, the WSM models are good for all molecules. In all cases, the rms errors are only a few tenths of a kJ mol^{-1}. Maximum errors are somewhat larger at around 1−2 kJ mol^{-1}. These are all for unit charge and would be smaller by a factor of 4 or so for a more realistic charge.

In Figure 5 we show difference maps of the induction energy of the formamide molecule in the field of a unit point charge, computed using the nonlocal models, local models, and WSM local models, respectively. The large errors in the nonlocal rank 1 model, particularly near the oxygen and the polar hydrogen atoms, are quite clearly displayed. These errors are reduced in the rank 1 local model and are still smaller in the WSM rank 1 local model. The largest residual errors always seem to occur in the same regions, perhaps indicating the need for higher ranking polarizabilities on some sites than on others. At rank 2, the nonlocal model is in almost perfect agreement with SAPT(DFT), but the local model obtained using the Le Sueur and Stone localization method exhibits rather large deficiencies near the polar hydrogens. These are removed in the WSM rank 2 local model which is comparable in accuracy to the rank 2 nonlocal model.

By and large, the behavior of the polarizability models for *N*-methyl propanamide, benzene, and 3-azabicyclo[3.3.1]-

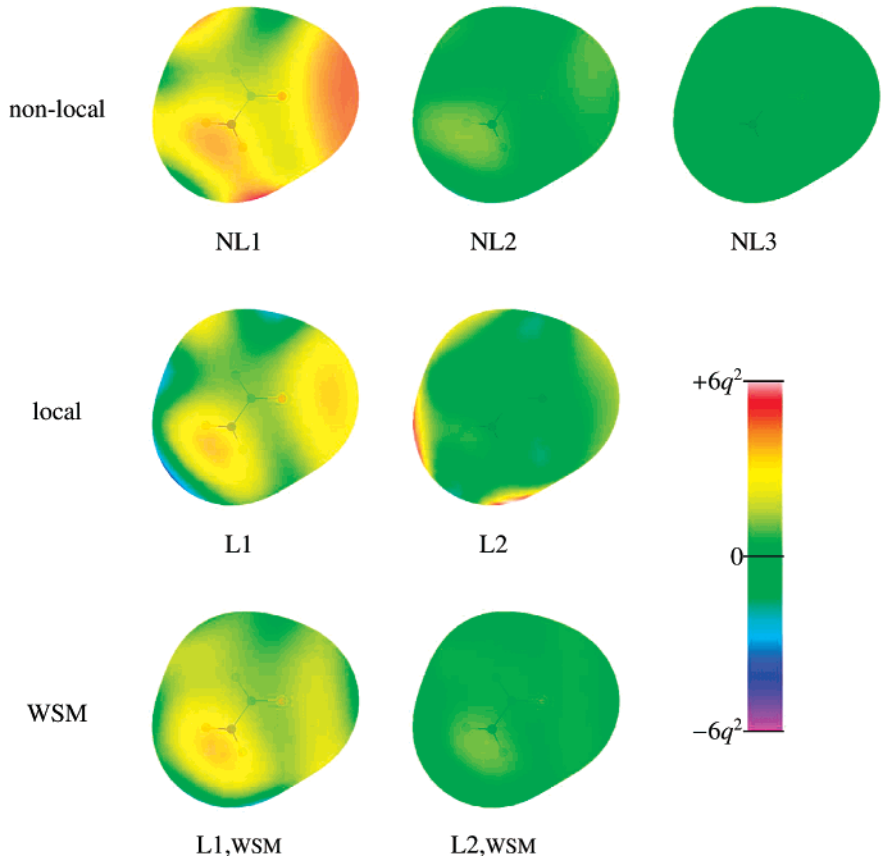

Figure 5. Difference maps of the induction energy (kJ mol^{-1}) arising from a charge q atomic units on the vdW $\times$ 2 surface of formamide using distributed nonlocal description of ranks 1, 2 and 3, distributed local description of ranks 1 and 2 obtained from the nonlocal models using the Le Sueur and Stone localization technique,[23] and WSM distributed local descriptions of ranks 1 and 2. The differences are taken against SAPT(DFT) second-order induction energies obtained using a molecular description with the Sadlej/MC basis set.

nonane-2,4-dione (BOQQUT) are similar to those for formamide, so we will comment only briefly on the models for these systems. The SAPT(DFT) induction energy maps and difference maps for the WSM rank 1 and rank 2 local models for these molecules are displayed in Figure 6. It is clear that the WSM rank 1 local models consistently underestimate the induction energy arising from a point charge, while the WSM rank 2 models offer consistently higher accuracy. This underestimation is notably severe for the benzene molecule, because the π-orbitals need to be described by rank 2 polarizability terms.

More accurate models are possible, but only if we are willing to accept the presence of unphysical terms. Even in the present models, some of the quadrupole−quadrupole polarizabilities violate the requirement of positive-definiteness, but, in the absence of any constraints, even the dipole−dipole terms are not always positive-definite.

The distributed polarizability calculations on BOQQUT brought to light a potential problem with the WSM procedure as currently implemented. BOQQUT is a fairly large molecule and possesses a plane of symmetry. We first calculated refined polarizabilities for this molecule using point-to-point polarizabilities calculated on a grid of 1000 points. The resulting rank 2 local model significantly broke the symmetry of the molecule, because the grid of points used for calculating the point-to-point polarizabilities was chosen at random and did not respect the symmetry of the molecule. The asymmetry could have been avoided simply by ensuring that the parameters of the polarizability model were correctly symmetrized, and we normally do this, but the use of unsymmetrized parameters allowed us to assess the quality of the grid, by determining how large the grid needed to be before the asymmetry was numerically negligible. For BOQQUT, a grid of 2000 points, or over 2 million point-to-point polarizabilities, proved to be adequate. The calculation of the point-to-point polarizabilities needed less than 2 h of CPU time on a single processor. This is due to the computational efficiency of the density-fitting-based algorithm[8] that has been implemented in the CamCASP program[15] and used for the point-to-point polarizabilities.

V. Penetration, Truncation Errors, and Damping

The multipole expansion provides us with a computationally efficient means of calculating the induction energy. However the resulting energies will be in error for a number of reasons, which must be addressed if we are to ensure accurate interaction energies.

(1) The multipole series are expansions in inverse powers of the intersite distance R_{ab}, so they diverge when the sites coincide. The multipoles and polarizabilities that appear in the damped classical polarizable model (eqs 21 and 22 in

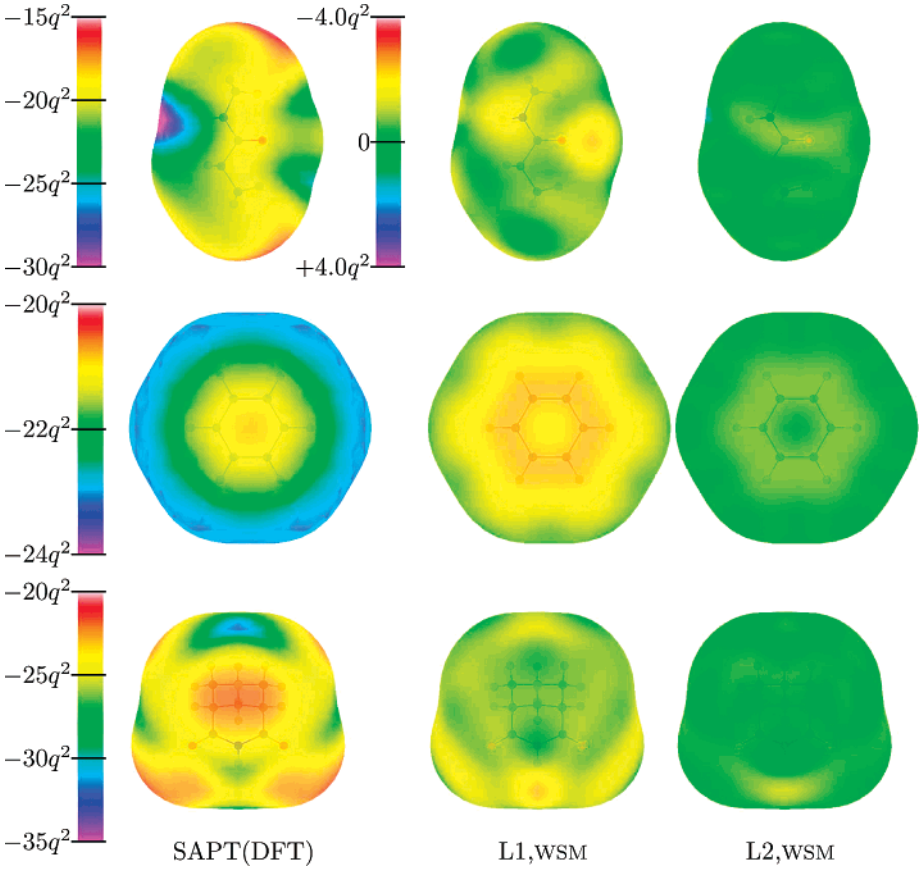

Figure 6. Induction energy maps and difference maps (kJ mol^{-1}) with the Sadlej basis arising from a charge q atomic units on the vdW $\times$ 2 surfaces of the *N*-methyl propanamide (top row), benzene (middle), and BOQQUT (bottom) molecules. The induction energy maps have been obtained using SAPT(DFT). The difference maps for the WSM rank 1 and rank 2 local polarizability models have been taken against the corresponding SAPT(DFT) second-order induction energies obtained using molecular descriptions with the Sadlej/MC basis set. The scale used for the difference maps is the same for all three molecules and is shown for *N*-methyl propanamide only.

part 1) treat the electron charge distribution as if it were concentrated at the local origin. That is, the finite extent of the charge distribution is neglected and orbital penetration effects are absent. The exchange part of the induction energy is also absent from the multipole expansion. These features result in a 'penetration error'. The true damping function—if one could be found—would account for this error.

For the second-order two-body induction energy, the divergence occurs at small enough intersite distances that it can be ignored, except for Monte Carlo simulations where a trial step is taken without regard to energy. Nevertheless, when iterations are included in the damped classical polarizable model (part 1, section IV.2) the role of the damping becomes more important. This can be seen from eq 22 in part 1, for at the mth iteration, the expression involves the product of $m + 1$ damping functions. The number of iterations needed for convergence increases as the intersite distance decreases, so the cumulative effect of the damping increases.

(2) Additionally, in practice, the multipole expansion must be truncated at a low rank. This introduces a 'truncation error'.

These sources of error need to be accounted for in accurate calculations of the interaction energies. This can be done using comparisons with the nonexpanded energies. While this comparison is quite straightforward for the electrostatic energy,[7] it is much less straightforward for the induction energy, because it is not immediately clear which SAPT(DFT) energies we should be using as a reference.

It has generally been assumed that the expanded induction energies approximate the induction energy from the polarization approximation. After all, the former is derived from the latter by using the multipole-expanded form of the interaction operator.[24] For example, in the absence of iterations, the expanded induction energies, $E^{(2)}_{\mathrm{ind,d-class}}$, have been expected to approximate $E^{(2)}_{\mathrm{ind,pol}}$. From the discussion in section IV.1 of part 1 and section III, it should come as no surprise that this is generally not the case. $E^{(2)}_{\mathrm{ind,d-class}}$ does indeed approximate $E^{(2)}_{\mathrm{ind,pol}}$ rather well if medium-sized monomer basis sets are used.[7] This is because the spurious tunneling effects discussed in section IV.1 of part 1 and, consequently, the exchange-induction energies are small for such basis sets. However, when large monomer basis sets are used, and especially with the MC$^+$ basis type, the spurious tunneling effects are quite large and result in large (negative) values for $E_{\mathrm{ind,pol}}$. In such a case, $E_{\mathrm{ind,exch}}$ is also large (and positive) and quenches $E_{\mathrm{ind,pol}}$ quite substantially. These tunneling effects are completely absent from the expanded form of the induction energy based on eq 21 in

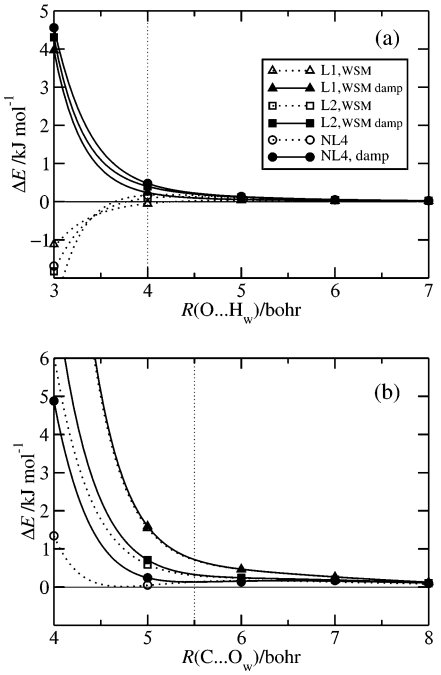

Figure 7. Errors in the expanded induction energy for the formamide−water dimer. The energy difference plotted on the *y*-axis is defined as $\Delta E = E^{(2)}_{\text{ind,d-class}} - E^{(2)}_{\text{ind,tot}}$, so a positive ΔE means that the expanded energies are not attractive enough. The SAPT(DFT) $E^{(2)}_{\text{ind,tot}}$ energies have been calculated with a Sadlej/MC$^+$ basis. Results are displayed for two relative orientations of the formamide and water molecules as described in Figure 3. The dotted vertical lines mark the approximate minimum-total-energy separations. Notice that the damped and undamped rank 1 curves for geometry (b) are almost identical. All local polarizability models have been obtained using the WSM procedure.

part 1. At present we cannot tell whether it is more appropriate to compare $E^{(2)}_{\text{ind,d-class}}$ with $E^{(2)}_{\text{ind,pol}}$, calculated in the monomer basis set, or with $E^{(2)}_{\text{ind,tot}}$, that is, the sum of $E^{(2)}_{\text{ind,pol}}$ and $E^{(2)}_{\text{ind,exch}}$, calculated with the MC$^+$ basis set. The former will include some spurious effects; in the latter they will be larger but partly cancelled out. We are currently investigating this question, but for the present work we use the latter choice in the comparisons of the expanded models below.

We now illustrate the points discussed above with the noniterated induction energies, i.e., $E^{(2)}_{\text{ind,d-class}}$, calculated for the formamide−water dimer using the Sadlej/MC$^+$ basis set. Figure 7 shows the energy difference $\Delta E = E^{(2)}_{\text{ind,d-class}} - E^{(2)}_{\text{ind,tot}}$. The two dimer geometries used, though somewhat artifical, serve the purpose of representing the two main types of intermolecular interactions: geometry (a) is a hydrogen-bond-like interaction with the hydrogen on water making close contact with the oxygen on formamide, and geometry (b) is a nonpolar interaction with the oxygen on water in close contact with the carbon on formamide. First we focus on the results without damping. The rank 4 nonlocal model provides an excellent description for both geometries, with energy differences less than 0.2 kJ mol^{-1} for physically relevant intersite separations. The rank 3 nonlocal model (not shown) gives very similar results to the rank 4 model, which suggests that increasing the rank above 4 will not improve the description significantly. Therefore the residual difference must be due to orbital penetration (a positive energy difference) and the divergence of the multipole expansion at short-range (a negative energy difference). Being of opposite signs, these two effects partially cancel. For geometry (a), the relatively small charge density on the hydrogen atom and the small intersite distance means that the divergence of the multipole expansion is the dominant effect and the rank 4 nonlocal model needs net damping; but for geometry (b), the opposite is the case, and the rank 4 nonlocal model needs a net enhancement at short intersite separations.

The rank 1 and 2 local models behave in a similar manner. However, while all three models result in similar energies for geometry (a), the rank 2 model is substantially better than rank 1 for geometry (b). This probably indicates the importance of higher ranking terms in the polarizability description of the heavy atoms.

We now turn to the issue of damping. As has been mentioned above, the damping function has to account for penetration effects and eliminate the divergence of the multipole expansion that occurs at small intersite distances. These two effects are of opposite signs and very likely depend differently on the intersite separation, as is the case for geometry (b) in the formamide water example. In practical applications the damping function must also compensate for truncation effects, which can complicate the picture quite substantially, as we have seen from the above example. It is unrealistic to expect to find a universal damping function capable of accounting for all these effects. The damping function will have to depend on the sites involved and will probably have to be larger than unity at intermediate distances, where enhancement is needed to account for truncation and penetration effects, and less than unity at short distances, to cancel out the short-range divergence of the multipole expansion. Such a function has been proposed for the dispersion energy,[25] but to the best of our knowledge, not for the induction energy.

We believe that the best we can do at present is to damp out the short-range divergence of the multipole expansion. This is especially necessary in calculations of the induction energy of clusters of polar molecules. Intersite distances can be quite small in such clusters, because cooperative induction effects can be quite large and compensate for the unfavorable exchange energies associated with small intersite distances. In such cases, an undamped multipole expansion will result in nonsensical induction energies, particularly when iterations are included in the evaluation of eqs 21 and 22 in part 1. A convenient choice for the damping functions are those due to Tang and Toennies[26] which have had the greatest success in the generation of high-accuracy potentials for small dimers. This will leave a residual error arising from truncation and penetration effects which can, in principle, be accounted for using an overlap model.[7]

We are then led to the problem of determining the damping factor to be used in the damping functions. It is often assumed that the damping factor can be determined by a comparison of the expanded and nonexpanded energies.

Given the observations of the above paragraphs, such a comparison is fruitless as it would suggest an antidamping. An alternative procedure has been to use the exponential parameter in the Born−Mayer term from a fit to the exchange energies as the damping parameter. This could, in principle, be done on a site−site basis to obtain damping parameters that depend on the pair of sites involved. The argument for this approach is that the exchange−repulsion and damping are both consequences of the wave function overlap. Perhaps a simpler method of determining the damping factor β, though one that depends only on the interacting molecules and not on the individual interacting sites, is based on the considerations presented in Chapter 6 of ref 24 and goes as follows: The potential due to the electronic density of a hydrogen-like atom with wave function $\psi(r) = \sqrt{\alpha^3/\pi}\, e^{-\alpha r}$ is

$$V(r) = -\frac{1}{r} + e^{-2\alpha r}\left(\alpha + \frac{1}{r}\right) \tag{3}$$

The first term in this equation is the multipole expansion of a spherical electron cloud and the second is the penetration correction. This potential can be rewritten as

$$V(r) = -\frac{1}{r} f_1(2\alpha r), \quad \text{where } f_1(2\alpha r) = 1 - e^{-2\alpha r}(1 + \alpha r) \tag{4}$$

so that f_1 is the damping function that correctly incorporates the penetration effect. This result can be plausibly, though not rigorously, generalized to an arbitrary wave function by using the asymptotic form of the wave function[27]

$$\psi(r) \rightarrow e^{-\sqrt{2I}r} \tag{5}$$

where I is the vertical ionization potential. We now see that the damping factor should be

$$\beta = 2\sqrt{2I} \tag{6}$$

where everything is in atomic units, so β is in bohr^{-1} if I is in Hartree. For most organic molecules, this results in an atom−atom damping function with a damping constant between 1.9 and 1.7. For mixed dimers, we suggest using the damping factor $\beta = \sqrt{2I_A} + \sqrt{2I_B}$. This choice is plausible as it is the coefficient of R in the exponential factor of the density-overlap function.

There is reason to believe this is a good choice for the damping factor, but there may be cases for which the issue of damping will have to be re-examined. Numerical evidence from calculations of the induction energy of organic crystals[28] indicates that above damping factor is not only appropriate but also essential when the rank 2 models needed for high accuracies are used. Welch et al.[28] have observed that without damping the small intersite distances present in some organic crystals can lead to very large and unphysical crystal induction energies. However, using the above damping factor leads to a rational progression of the crystal induction energy with rank of the polarizability description. As would be expected, damping has the largest effect on the higher ranking models.

However, there will be molecules for which a single damping factor may be too simplistic. Our definition of β involves the ionization potentials of the molecules. For a large molecule, with very different functional groups, it may be necessary to use a different ionization potential for each of the functional groups. This would then lead to a damping factor that depended on the pair of interacting groups. We have not yet investigated such a possibility.

In the formamide−water example, using ionization potentials of 0.375 au and 0.464 au for formamide and water, respectively, we obtain $\beta = 1.83$. From Figure 7 we see that after damping all models exhibit a positive energy difference which is a combination of penetration effects and truncation errors. From the modeling point of view, their uniformity makes these differences simpler to handle than those made by the undamped models. For example, as suggested in ref 7, the overlap model[29] could be used to model the residual energy differences. We are currently working on a methodology to make this possible on a routine basis.

VI. Approximations

The many issues discussed in the preceding sections are important in high-accuracy calculations of the interaction energy. However, for many applications, it may be sufficient to achieve a moderate accuracy. Indeed, for large systems, some of the prescriptions given above may not even be feasible due to lack of computational resources. Here we discuss the approximations most useful in calculations on large systems and test their accuracy.

VI.1. Calculating $E^{(2)}_{\mathrm{ind,tot}}$ and $E^{(3)}_{\mathrm{ind,tot}}$ Using a Monomer Basis. In section III we have argued that the 'far-bond' functions that are part of the MC$^+$ type of basis should be used in order to obtain basis-saturated induction energies. Since the far-bond functions are placed at the locations of the nuclei of the interacting partner, this means that the MC$^+$ basis set depends on the dimer geometry. This in turn means that the calculation of the Hessians (eqs 7 and 8 of part 1), which is the most computationally expensive step in the evaluation of the induction energy, needs to be repeated for each dimer geometry. This can be computationally prohibitive and would be avoided if the monomer basis (MC) were to be used. Furthermore, there is good evidence to suggest that the MC$^+$ basis type introduces significant errors in the first-order energies[30] for which the MC type should be used.

The first-order energies are reasonably well converged with the Sadlej/MC basis set. The problems lie with the second-order energies. Both the induction and dispersion energies are insufficiently converged in this basis. As we saw in section III, the so-called mid-bond basis functions that are included in the MC$^+$ basis type are needed to obtain basis-converged dispersion energies. Like the far-bond functions, the mid-bond basis functions also make the basis set dependent on the dimer geometry. Neglecting the mid-bond functions will introduce significant errors in the dispersion energy which will complicate the discussion of the induction energy calculations. Therefore, for the purposes of the present paper, we will avoid discussing the basis convergence issues associated with the dispersion energy by using dispersion

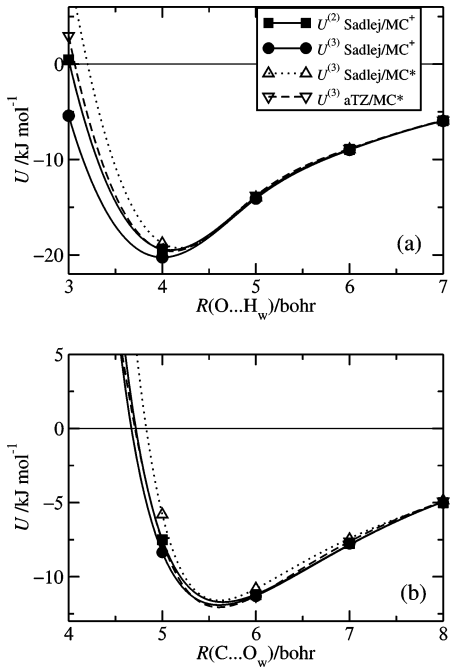

Figure 8. Total interaction energies for the formamide−water dimer using different basis sets. $U^{(2)}$and $U^{(3)}$ are the total SAPT(DFT) interaction energies up to second and third order, respectively. Calculations have been performed using the MC$^+$ and MC basis types. The latter basis has been labeled as 'MC*' because, for reasons explained in the text, the dispersion and exchange-dispersion energies have been calculated using the MC$^+$ basis. The $U^{(2)}$ and $U^{(3)}$ potential curves with the MC* basis are very similar, so, for clarity, only the latter are displayed. Results are displayed for two relative orientations of the formamide and water molecules as described in Figure 3. For geometry (b), the $U^{(3)}$ aTZ/MC* curve is almost obscured by the $U^{(2)}$ Sadlej/MC$^+$ and $U^{(3)}$ Sadlej/MC$^+$ curves.

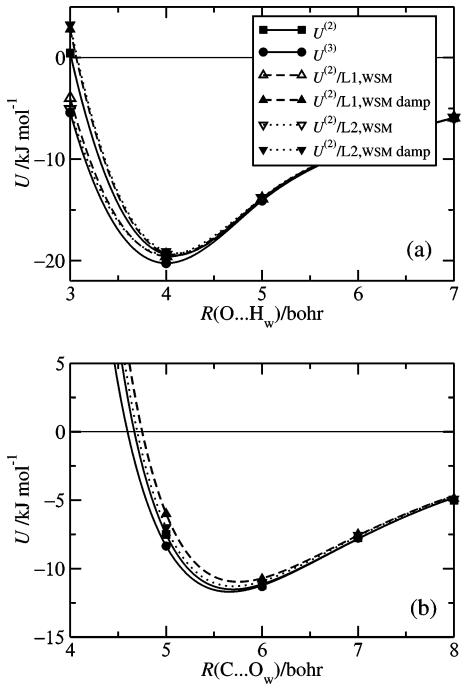

Figure 9. Total interaction energies for the formamide−water dimer. $U^{(2)}$ and $U^{(3)}$ are the SAPT(DFT) interaction energies at second and third order. Approximations to $U^{(2)}$ are labeled $U^{(2)}$/Ln,WSM,(damp). This notation means that the SAPT(DFT) induction energies have been replaced by the induction energies obtained from a damped classical polarizable model with iteration, using a distributed polarizability model of rank n on both molecules. Results are displayed for two relative orientations of the formamide and water molecules as described in Figure 3. For geometry (a), the $U^{(2)}$/Ln,WSM and $U^{(2)}$/Ln,WSM,damp curves are just above the $U^{(3)}$ and $U^{(2)}$ curves, respectively. For geometry (b), damping has almost no effect on the energies, so the damped and undamped curves are indistinguishable. All local polarizability models have been obtained using the WSM procedure.

energies computed using the Sadlej/MC$^+$ basis. We will denote the basis types for these mixed basis calculations by 'MC*'.

In Figure 8 we display total interaction energies for the formamide−water dimer obtained using different basis sets. For the MC* basis type only the $U^{(3)}$ potentials are shown, since $U^{(2)} \approx U^{(3)}$ for this basis type. This is because $E^{(3)}_{\mathrm{ind,tot}}$ is negligibly small when computed using monomer basis sets, as should be apparent from Figure 4. $U^{(3)}$ from the Sadlej/MC* basis results in potentials that are consistently too shallow, particularly for the hydrogen-bonded geometry (a). A considerable improvement is obtained if the aug-cc-pVTZ/MC* basis is used. For both geometries, this basis gives potential curves in good agreement with the $U^{(2)}$ potential obtained with the Sadlej/MC$^+$ basis set. Therefore the aug-cc-pVTZ/MC* basis is a viable alternative to the dimer-geometry dependent Sadlej/MC$^+$ basis.

It must be borne in mind that the aug-cc-pVTZ/MCbasis is twice the size of the Sadlej/MC basis and larger even than the Sadlej/MC$^+$ basis. Therefore, while the aug-cc-pVTZ/MC basis allows fairly accurate induction energies to be evaluated for all dimer geometries using Hessians calculated just once, it does entail a significant increase in the computational cost of evaluating the Hessians.

VI.2. Neglecting the Higher-Order Two-Body Energies. The calculation of the higher-order energies using eq 25 in part 1 is the most time-consuming part of a SAPT(KS) calculation. Its computational scaling is dominated by the evaluation of $E^{(3)}_{\mathrm{ind,exch}}$[KS] which scales as $O(n_o^2 n_v^3)$, where n_o and n_v are the number of occupied and virtual orbitals, respectively. From the discussion in section IV.1.2 of part 1 we know that the higher-order energies cannot be neglected in systems that exhibit strong hydrogen bonds, such as water and hydrogen fluoride, but their effect is far less in other systems and may even be negligible. In Figure 8 we display total interaction energies for the formamide−water system in the two representative geometries used in earlier discussions. In geometry (a), the hydrogen-bonded geometry, the $U^{(3)}$ potential curve is deeper than $U^{(2)}$. Using the $U^{(2)}$ potential curve would result in a slightly longer formamide−water bond, but the errors are not large and could well be acceptable in calculations of moderate accuracy.

The situation is far better for the non-hydrogen-bonded geometry (b). Here the $U^{(3)}$ potential curve is only slightly deeper than $U^{(2)}$, which could be used with almost no loss in accuracy.

Table 2. Contribution of Third- and Higher-Order Corrections to the Interaction Energy for the Water, Hydrogen Fluoride, Carbon Dioxide, and Benzene Dimers and the $H_2O\cdots H_3N$ and $H_2\cdots CO$ Complexes[a]

	$(H_2O)_2$	$(HF)_2$	$H_2\cdots H_3N$	$(CO_2)_2$	$(C_6H_6)_2$	$H_2\cdots CO$
$E_{ind}^{(2)}$ SAPT(DFT)	−5.164	−6.687	−0.479	−0.866	−0.885	−0.122
$E_{ind,tot}^{(2)}$/L1,WSM,damp	−3.932	−4.268	−0.361	−0.358	−0.599	−0.073
$E_{ind,tot}^{(2)}$/L2,WSM,damp	−4.507	−5.303	−0.357	−0.474	−0.710	−0.080
$E_{ind,tot}^{(2)}$/NL4,damp	−5.051	−5.647	−0.420	−0.619	−0.754	−0.081

[a] The first three dimers are at their equilibrium geometries, and the benzene dimer is in the parallel stacked geometry with a center-of-mass separation of 3.8 Å. $H_2\cdots CO$ is in the linear geometry with C toward H_2 and a center-of-mass separation of 7.8 au, and $H_2O\cdots H_3N$ is in geometry (a) of Figure 5 from ref 31. L1,WSM and L2,WSM denote damped, noniterated, classical induction energies evaluated using local rank-1 and rank-2 polarizabilities refined using the WSM procedure, and NL4 denotes damped energies obtained from nonlocal rank-4 polarizabilities. Molecular properties needed for the damped classical model were obtained using the aug-cc-pVTZ/MC basis. All energies are reported in kJ mol^{-1}.

VI.3. Replacing $E_{ind,tot}^{(2)}$ and $E_{ind,tot}^{(3)}$ by $E_{ind,d-class}$. An approximation that is commonly used is to avoid calculating the nonexpanded induction energies altogether by using the expanded induction instead. From Figure 7 it should be apparent that this approximation would be rather good if the undamped rank 4 nonlocal polarizability description were used on both molecules, but this description is very elaborate and would probably be impossible to use in clusters of molecules. The rank 1 and rank 2 local polarizability descriptions are more practicable but less accurate. The accuracy can be improved slightly by using the iterated form of the polarizability models.

We will use $U^{(2)}$/L*n*,WSM,(damp) to denote the total interaction energy obtained by replacing $E_{ind,tot}^{(2)}$ and $E_{ind,tot}^{(3)}$ with $E_{ind,d-class}$, where *n* denotes the rank of the (possibly damped) local polarizability model. In Figure 9 we display total interaction energies obtained using these approximations. For geometry (a), both $E_{ind,tot}^{(2)}$/L1,WSM,damp and $E_{ind,tot}^{(2)}$/L2,WSM,damp are reasonably good approximations to $U^{(2)}$. Curiously, the undamped approximations are very close to $U^{(3)}$. For geometry (b), all approximations fare well, with the rank 2 models being the more accurate. Damping has a very small effect on the energies here. Therefore, $U^{(2)}$/L2,WSM,damp is a viable approximation to the total interaction energy in both cases.

Table 2 shows results for some other small dimers at equilibrium or near-contact geometries. Here too the results are generally satisfactory.

The computational cost of evaluating the induction energy using the rank 1 or 2 local polarizability models is much lower than the cost of evaluating $E_{ind,pol}^{(2)}$ and $E_{ind,exch}^{(2)}$ within SAPT(DFT). This is true even if the SAPT(DFT) energies are calculated using the MC type of basis. Additionally, we see from Figures 8 and 9 that $U^{(2)}$/L*n*,WSM,(damp) is a better approximation to the interaction energy than calculating $U^{(2)}$ in the Sadlej/MC* basis. This is particularly welcome for applications involving large molecules for which we may be able to calculate molecular properties but not the SAPT(DFT) interaction energies. The relatively good accuracy of the $U^{(2)}$/L*n*,WSM,(damp) approximation also helps explain why ab initio intermolecular potentials that used multipole expansion for the induction energy, such as the ASP water potential,[32] have been so successful.

VI.4. Mixed-Rank Polarizability Descriptions. There will be situations where polarizability models of mixed rank can be used. For example, an accurate description of the polarizability of benzene can be obtained with rank 1 polarizabilities on the hydrogen atoms and rank 2 terms on the carbon atoms.[9] For large, compact molecules, it may be desirable to simplify the polarizability model by reducing the rank of the description of those atoms hidden under the van der Waals spheres of neighboring atoms or even omitting them altogether from the model. Further simplifications can be achieved by enforcing symmetries of functional groups or eliminating small terms in the polarizability model. All of these simplifications can be incorporated in the WSM procedure described in section IV.3 of part 1. This procedure, together with the constrained density-fitting distribution method, gives us the flexibility to choose an appropriate polarizability description while avoiding unphysical terms in the polarizability description as far as possible.

In Table 1 we report the maximum and rms errors in polarizability models using a rank 2 description on the heavy atoms and a rank 1 description on the hydrogens. The rms errors in these mixed-rank polarizability descriptions are comparable to those made by the more complex rank 2 descriptions, while the maximum errors, which tend to occur near the hydrogen atoms, are significantly smaller for the larger molecules studied here. Additionally, there is far less loss of positive-definiteness in the mixed-rank descriptions as it is the quadrupole−quadrupole polarizability terms on the hydrogen atoms that tend to be negative.

In the case of BOQQUT, the mixed-rank description has nearly half as many nonzero polarizability components as the rank 2 description. The mixed-rank description of the formamide molecule results in induction energies of the formamide water dimer that are almost identical to those from the rank 2 local description.

We therefore have good physical and computational reasons for using the mixed-rank polarizability descriptions and strongly recommend use of these models in favor of the more complex and less physical rank 2 models.

VII. Summary

We have provided a theoretical and numerical framework for the accurate calculation of the induction energies of clusters of organic molecules. These are large systems and pose quite different problems from those that have faced the high-accuracy, small-molecule community. These problems can be broadly classified as those concerned with the theoretical details of the induction energies at second and higher order in the interaction operator and those concerned

with numerical details and the development of models suitable for applications.

VII.1. Model Induction Energies. Accurate molecular polarizabilities and multipole moments are needed for modeling of the induction energy of clusters of molecules. For all but the smallest of molecules, these properties need to be distributed. The problem of distributing the multipole moments has already been addressed[33,34] in a satisfactory manner. Here we have proposed and demonstrated an accurate and versatile method of obtaining distributed polarizabilities that is suitable for molecules of as many as 30 atoms or so. Our distribution scheme for the polarizabilities is based on the methods of Williams and Stone[9] and our constrained density-fitting method.[8] By combining the strengths of these two methods, we have obtained a distribution procedure with properties that make it ideal for high-accuracy calculations on systems of large molecules. The main features of this distribution scheme are as follows:

(1) The underlying theory used in the polarizability calculations is coupled Kohn−Sham theory (CKS), also known as Kohn−Sham linear response theory. Molecular properties obtained using CKS theory can exceed coupled-cluster methods in accuracy when used with a modern density functional like PBE0[35] with asymptotic corrections,[36,37] thus ensuring an accurate polarizability description at modest computational cost. In the form used in this paper and implemented in the CamCASP program,[15] the CKS equations are solved with a computational effort that scales as $O(n_o^3 n_v^3)$, but this scaling can be reduced using density-fitting techniques.[5,38,39] With the current implementation of the CKS equations we have been able to compute the properties of the 3-azabicyclo[3.3.1]nonane-2,4-dione (BOQQUT) molecule, containing 22 atoms, in less than a day of CPU time on a single Opteron processor using the Sadlej basis set. With the density-fitted form of these equations we expect to be able to perform calculations on even larger systems.

(2) We have tested localized distributed polarizability models of rank 4 for small molecules and rank 2 or 3 for the larger molecules. In principle, local descriptions up to rank 4 could be generated for all molecules, but rank 2 should be sufficient.

(3) Non-positive-definite polarizability tensors do not occur in the dipole−dipole polarizabilities, and at higher rank they arise mainly for the hydrogen atoms. Consequently, this problem is smallest in the mixed-rank models.

(4) There is no fundamental limit to the size of the basis sets that can be used, though computational limitations will restrict it in practice.

(5) Finally, one of the most powerful features of the distribution scheme proposed here is that it gives us the ability to choose any reasonable polarizability model and yet obtain an accurate, physically correct, local polarizability model.

We have tested this Williams−Stone−Misquitta (WSM) distribution scheme using the formamide and *N*-methyl propanamide molecules and have also generated local polarizability models of ranks 1 and 2 for benzene, BOQQUT, and other molecules,[28] only some of which have been reported in this paper. The rank 2 local description is comparable in accuracy to the much more complex rank 4 nonlocal models. The rank 1 models underestimate the induction energy, particularly around the heavy atoms. Mixed rank models with a rank 2 description on the heavy atoms and rank 1 on the hydrogen atoms offer a good compromise between accuracy and simplicity.

The newly developed visualization techniques recently implemented in the ORIENT program[16] have given us a very powerful means of evaluating the polarizability models. By 3-dimensional visualization of the induction maps or error maps made against accurate SAPT(DFT) induction energies, we were able to make assessments of the shortcomings of these models. The 3-D maps enable us to identify sites at which a particular polarizability description may be deficient. This proves invaluable in designing accurate polarizability models where a mixed rank description may be necessary.

VII.2. Numerical Aspects. One of the main considerations in any ab initio calculation is the type of basis set to be used. It will not usually be possible to use large basis sets in calculations on organic molecules of the size considered in this paper, so we have attempted to determine which basis sets are good enough for calculations of the induction energy. The necessary level of accuracy will depend on the application, but for systems of organic molecules with dimer binding energies of 10−20 kJ mol^{-1}, basis set incompleteness errors of less than a few tenths of a kJ mol^{-1} at the important dimer geometries are probably acceptable.

From numerical tests on a variety of molecules (only one example was reported here), we recommend the Sadlej basis sets[18,19] for calculations of molecular multipole moments and polarizabilities. Auxiliary basis sets tuned for the Sadlej bases are not available, but we have found that the aug-cc-pVTZ auxiliary basis,[40,41] though probably too large, works very well. With these basis sets, very accurate polarizability models can be obtained. Comparisons with the much larger aug-cc-pVQZ basis show that the maximum error made by the Sadlej basis models is about 1.5 kJ mol^{-1} on the vdW × 2 surface in the field of a unit point charge. For a more realistic charge of 0.5 units, this would be an error of only 0.4 kJ mol^{-1}.

To get a similar accuracy, SAPT(DFT) induction energies must be evaluated using the Sadlej/MC$^+$ basis, that is, with the inclusion of basis functions located at the positions of the nuclei of the partner molecule.[20] The Sadlej/MC$^+$ and aug-cc-pVTZ/MC$^+$ bases give almost identical induction energies. However, when used in the MC basis type, that is, without the extra off-atomic functions, both the Sadlej and aug-cc-pVTZ bases yield poor induction energies, the latter being the better choice.

On the issue of damping: in our opinion, it is impossible for the damping functions in current use to recover the penetration energy. All that is possible is the damping out of the short-range divergence in the multipole expansion. Apart from special cases such as Monte Carlo simulations, this is not essential for the dimer energy, as intersite distances in a dimer are never small enough to see the onset of the divergence in the multipole series. However, in the bulk, many-body effects can cause small intersite distances,

particularly in hydrogen-bonded geometries, and consequently damping is needed. We have argued that comparisons with nonexpanded energies are not useful in determining the damping needed. Rather, we propose that the Tang−Toennies functions[26] be used with a damping coefficient of $\beta = \sqrt{2I_A} + \sqrt{2I_B}$, where I_A and I_B are the ionization energies (in au) of the interacting molecules.

The remaining error will be entirely due to penetration effects and the truncation error, the latter arising from the truncation of the multipole (1/*R*) expansion. In accurate work, these errors must be accounted for in some way. As they decay exponentially with distance, like the exchange−repulsion energy, they could be included with it and modeled in a similar way. We are currently looking for a robust way to do this.

Comparisons of the model induction energies and the nonexpanded SAPT(DFT) energies brought to light a very unexpected correspondence. Contrary to supposition, the model induction energy at second order, $E^{(2)}_{\mathrm{ind,d-class}}$, does not recover the expression for the second-order induction energy in the polarization approximation, $E^{(2)}_{\mathrm{ind,pol}}$, but rather the *sum* $E^{(2)}_{\mathrm{ind,pol}} + E^{(2)}_{\mathrm{ind,exch}}$. We have defined this sum to be the induction energy, $E^{(2)}_{\mathrm{ind,tot}}$. The reason for this has to do with spurious tunneling effects present due to Coulomb singularities in the interaction operator.[13,14] These singularities are also responsible for the slow convergence of the induction energy with basis set and order in perturbation theory. It is expected that a SAPT(DFT) formulation with a regularized form of this operator will be free from both these problems. Preliminary investigations suggest that this is indeed the case.

We have tried to describe ways of calculating the induction energy with as few approximations as possible. However, there will always be systems for which approximations will be needed. Consequently, we have proposed and analyzed a number of possible approximations of varying complexity. The most useful of these approximations is one in which we calculate the induction energy of a cluster from the induction models only. We recommend using the damped mixed-rank local polarizability description, that is, with a rank 2 description on the heavy atoms and a rank 1 description on the hydrogen atoms. The success of this type of model also explains why potentials that have used a similar description of the induction energy have been so successful. An example is the ASP water potential.[32]

VIII. Programs

Many of the theoretical methods described in this review are implemented in programs available for download. Some of these, together with their main uses in the present work, are as follows:

(1) SAPT2002:[42] SAPT(KS) energy calculations.

(2) CamCASP 4.5:[15] Molecular properties in total and distributed form and SAPT(DFT) dispersion and induction energies. The CamCASP suite includes the GDMA 2.2 program[34] used for calculating the distributed multipoles needed in the induction energy calculations and the PFIT program used to refine the distributed polarizabilities in the WSMprocedure.

(3) ORIENT 4.6:[16] Localization of the distributed polarizabilities and visualization of the energy maps.

(4) DALTON 2.0:[43] DFT and CKS calculations. A patch[42] is needed to enable DALTON 2.0 to work with SAPT2002 and CamCASP 4.5.

Acknowledgment. We would like to thank Gareth Welch for useful discussions and comments and his help in obtaining the structure of BOQQUT. A.J.M. would like to thank Girton College, Cambridge for a research fellowship. This research was supported by EPSRC grant EP/C539109/1.

References

(1) Misquitta, A. J.; Stone, A. J. Accurate induction energies for small organic molecules: I. Theory. *J. Chem. Theory Comput.* **2007**, *3*, 7−18.

(2) Misquitta, A. J.; Jeziorski, B.; Szalewicz, K. Dispersion energy from density-functional theory description of monomers. *Phys. Rev. Lett.* **2003**, *91*, 33201.

(3) Misquitta, A. J.; Szalewicz, K. Symmetry-adapted perturbation-theory calculations of intermolecular forces employing density-functional description of monomers. *J. Chem. Phys.* **2005**, *122*, 214109.

(4) Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. Intermolecular potentials based on symmetry-adapted perturbation theory with dispersion energies from time-dependent density-functional theory. *J. Chem. Phys.* **2005**, *123*, 214103.

(5) Hesselmann, A.; Jansen, G.; Schutz, M. Density-functional theory-symmetry-adapted intermolecular perturbation theory with density fitting: A new efficient method to study intermolecular interaction energies. *J. Chem. Phys.* **2005**, *122*, 014103.

(6) Misquitta, A. J.; Szalewicz, K. Intermolecular forces from asymptotically corrected density functional description of monomers. *Chem. Phys. Lett.* **2002**, *357*, 301−306.

(7) Stone, A. J.; Misquitta, A. J. Atom−atom potentials from *ab initio* calculations. *Int. Rev. Phys. Chem.* **2007**, *26*, 193−222.

(8) Misquitta, A. J.; Stone, A. J. Distributed polarizabilities obtained using a constrained density-fitting algorithm. *J. Chem. Phys.* **2006**, *124*, 024111.

(9) Williams, G. J.; Stone, A. J. Distributed dispersion: a new approach. *J. Chem. Phys.* **2003**, *119*, 4620−4628.

(10) Hulme, A. T.; Johnston, A.; Florence, A. J.; Fernandes, P.; Shankland, K.; Bedford, C. T.; Welch, G. W. A.; Sadiq, G.; Haynes, D. A.; Motherwell, W. D. S.; Tocher, D. A.; Price, S. L. The search for a predicted hydrogen bonded motif − a multidisciplinary investigation into the polymorphism of 3-azabicyclo[3.3.1]nonane-2,4-dione. *J. Am. Chem. Soc.* **2007**, *129*, 3649−3657.

(11) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation theory approach to intermolecular potential energy surfaces of Van der Waals complexes. *Chem. Rev.* **1994**, *94*, 1887−1930.

(12) Jeziorski, B.; Szalewicz, K. Symmetry-adapted perturbation theory. In *Handbook of Molecular Physics and Quantum Chemistry*; Wilson, S., Ed.; Wiley: 2002; Vol. 8, pp 37−83.

(13) Patkowski, K.; Jeziorski, B.; Szalewicz, K. Symmetry-adapted perturbation theory with regularized coulomb potential. *J. Mol. Struct.* (*THEOCHEM*) **2001**, *547*, 293−307.

(14) Patkowski, K.; Jeziorski, B.; Szalewicz, K. Unified treatment of chemical and van der Waals forces via symmetry-adapted perturbation expansion. *J. Chem. Phys.* **2004**, *120*, 6849−6862.

(15) Misquitta, A. J.; Stone, A. J. *CamCASP: a program for studying intermolecular interactions and for the calculation of molecular properties in distributed form*; University of Cambridge: 2006. Inquiries to A. J. Misquitta, am592@cam.ac.uk.

(16) Stone, A. J.; Dullweber, A.; Engkvist, O.; Fraschini, E.; Hodges, M. P.; Meredith, A. W.; Nutt, D. R.; Popelier, P. L. A.; Wales, D. J. *Orient: a program for studying interactions between molecules, version 4.6*; University of Cambridge: 2006. Inquiries to A. J. Stone, ajs1@cam.ac.uk.

(17) Bondi, A. van der Waals volumes and radii. *J. Phys. Chem.* **1964**, *68*, 441−451.

(18) Sadlej, A. J. Medium-size polarized basis sets for high-level correlated calculations of molecular electric properties. *Collect. Czech Chem. Commun.* **1988**, *53*, 1995−2016.

(19) Sadlej, A. J. Medium-sized polarized basis sets for high-level correlated calculations of molecular electric properties. II. Second-row atoms Si−Cl. *Theor. Chim. Acta* **1991**, *79*, 123−140.

(20) Williams, H. L.; Mas, E. M.; Szalewicz, K.; Jeziorski, B. On the effectiveness of monomer-centered, dimer-centered, and bond-centered basis functions in calculations of intermolecular interaction energies. *J. Chem. Phys.* **1995**, *103*, 7374−7391.

(21) Patkowski, K.; Szalewicz, K.; Jeziorski, B. Third-order interactions in symmetry-adapted perturbation theory. *J. Chem. Phys.* **2006**, *125*, 154107.

(22) Misquitta, A. J.; Stone, A. J. Accurate dispersion energies for organic molecules. **2007**, manuscript in preparation.

(23) Le Sueur, C. R.; Stone, A. J. Localization methods for distributed polarizabilities. *Mol. Phys.* **1994**, *83*, 293−308.

(24) Stone, A. J. *The Theory of Intermolecular Forces*; Clarendon Press: Oxford, 1996.

(25) Hodges, M. P.; Stone, A. J. A new representation of the dispersion interaction. *Mol. Phys.* **2000**, *98*, 275−286.

(26) Tang, K. T.; Toennies, J. P. An improved simple model for the Van der Waals potential based on universal damping functions for the dispersion coefficients. *J. Chem. Phys.* **1984**, *80*, 3726−3741.

(27) Levy, M.; Perdew, J. P.; Sahni, V. Exact differential equation for the density and ionization energy of a many-particle system. *Phys. Rev. A* **1984**, *30*, 2745−2748.

(28) Welch, G. W. A.; Karamertzanis, P. G.; Misquitta, A. J.; Stone, A. J.; Price, S. L. Is the induction energy important for modelling organic crystals. *J. Chem. Theory Comput.* **2007**, submitted for publication.

(29) Kim, Y. S.; Kim, S. K.; Lee, W. D. Dependence of the closed-shell repulsive interaction on the overlap of the electron densities. *Chem. Phys. Lett.* **1981**, *80*, 574−575.

(30) Burcl, R.; Chalasinski, G.; Bukowski, R.; Szczesniak, M. M. On the role of bond functions in interaction energy calculations: Ar···HCl, Ar···H_2O, $(HF)_2$. *J. Chem. Phys.* **1995**, *103*, 1498−1507.

(31) Langlet, J.; Caillet, J.; Bergès, J.; Reinhardt, P. Comparison of two ways to decompose intermolecular interactions for hydrogen-bonded dimer systems. *J. Chem. Phys.* **2003**, *118*, 6157−6166.

(32) Millot, C.; Stone, A. J. Towards an accurate intermolecular potential for water. *Mol. Phys.* **1992**, *77*, 439−462.

(33) Stone, A. J.; Alderton, M. Distributed multipole analysis−methods and applications. *Mol. Phys.* **1985**, *56*, 1047−1064.

(34) Stone, A. J. Distributed multipole analysis: Stability for large basis sets. *J. Chem. Theory Comput.* **2005**, *1*, 1128−1132.

(35) Adamo, C.; Cossi, M.; Scalmani, G.; Barone, V. Accurate static polarizabilities by density functional theory: assessment of the PBE0 model. *Chem. Phys. Lett.* **1999**, *307*, 265−271.

(36) Tozer, D. J.; Handy, N. C. Improving virtual Kohn−Sham orbitals and eigenvalues: Application to excitation energies and static polarizabilities. *J. Chem. Phys.* **1998**, *109*, 10180−10189.

(37) Tozer, D. J. The asymptotic exchange potential in Kohn−Sham theory. *J. Chem. Phys.* **2000**, *112*, 3507−3515.

(38) Bukowski, R.; Podeszwa, R.; Szalewicz, K. Efficient generation of the coupled Kohn−Sham dynamic sysceptibility functions and dispersion energy with density fitting. *Chem. Phys. Lett.* **2005**, *414*, 111−116.

(39) Podeszwa, R.; Bukowski, R.; Szalewicz, K. Density-fitting method in symmetry-adapted perturbation theory based on Kohn−Sham description of monomers. *J. Chem. Theory Comput.* **2006**, *2*, 400−412.

(40) Weigend, F.; Häser, M.; Patzelt, H.; Ahlrichs, R. RI-MP2: optimized auxiliary basis sets and demonstration of efficiency. *Chem. Phys. Lett.* **1998**, *294*, 143−152.

(41) Weigend, F.; Kohn, A.; Hättig, C. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *J. Chem. Phys.* **2002**, *116*, 3175−3183.

(42) Bukowski, R.; Cencek, W.; Jankowski, P.; Jeziorski, B.; Jeziorska, M.; Kucharski, S.; Misquitta, A. J.; Moszynski, R.; Patkowski, K.; Rybak, S.; Szalewicz, K.; Williams, H.; Wormer, P. *SAPT2002: an ab initio program for many-body symmetry-adapted perturbation theory calculations of intermolecular interaction energies*; University of Delaware and University of Warsaw: 2002.

(43) *DALTON, a molecular electronic structure program, release 2.0*; 2005. See: http://www.kjemi.uio.no/software/dalton/dalton.html (accessed 21 June 2007).

CT700105F