JCTC Journal of Chemical Theory and Computation

Spatially Homogeneous QM/MM for Systems of Interacting Molecules with on-the-Fly ab Initio Force-Field Parametrization

Ali Sebetci[†,‡] and Gregory J. O. Beran*[,†]

Department of Chemistry, University of California, Riverside, California 92521

Received October 15, 2009

Abstract: Quantum and classical mechanics are combined in a hybrid many-body interaction model to enable the computationally affordable study of systems containing many interacting molecules. This model treats intramolecular and pairwise intermolecular interactions quantum mechanically, while many-body electrostatic induction effects are approximated using a polarizable force field. In this paper, we demonstrate that parametrizing the force field with distributed multipoles and atom-centered polarizabilities obtained on-the-fly from ab initio quantum mechanical monomer calculations makes the model very accurate and eliminates nearly all empiricism. Test calculations on water, formamide, hydrogen fluoride, and glycine−water clusters, all of which exhibit strong many-body interactions, are presented. The performance of the hybrid model is competitive with related point-charge embedding models.

1. Introduction

Molecular clusters, liquids, and solids display complex and interesting behaviors, many of which are not well understood. Ab initio quantum chemistry would be an extremely helpful tool for modeling such systems, but its high computational expense and the difficulty in treating intermolecular interactions accurately limits its applicability at present. Reliable, inexpensive approximations to a full quantum mechanical treatment of such systems are greatly needed.

One such approach lowers the computational cost in systems of interacting molecules by using a truncated many-body interaction (MBI) expansion to describe the total system energy in terms of interacting molecules/fragments. This idea is quite old, and many groups have studied it in recent years. For example, symmetry-adapted perturbation theory techniques,[1,2] the fragment molecular orbital method,[3−5] the effective fragment potential,[6,7] electrostatically embedded many-body expansions,[8−11] molecular mechanics for clusters,[12] accurate polarizable force fields,[13−18] and other studies[19−26] all utilize MBI expansions. Notably, the importance of including three-body and higher ("many-body") effects is a recurring theme in most of these works.

Many-body effects include contributions from exchange−repulsion, electrostatic induction (also called polarization), and dispersion forces. However, for systems of interacting polar molecules, induction typically dominates the many-body effects. Induction effects are particularly important when strong electrostatic interactions or hydrogen-bond cooperativity[27] are present.

One of us recently investigated[28] a hybrid quantum mechanics/molecular mechanics (QM/MM) model for describing systems of interacting molecules. This hybrid many-body interaction (HMBI) model partitions the system based on the order of intermolecular interactions in the many-body interaction series: the most important intermolecular interactions are modeled quantum mechanically, while smaller, computationally expensive contributions are approximated classically. Specifically, the one- and two-body interactions are computed quantum mechanically, while the three-body and higher interactions are approximated with an inexpensive classical polarizable force field that describes electrostatic induction. Many-body dispersion and exchange-repulsion effects are completely neglected in this model. Similar partitioning is used very successfully in constructing spec-

* To whom correspondence should be addressed: E-mail: gregory.beran@ucr.edu.

[†] University of California, Riverside.

[‡] Current address: Computer Engineering Department, Zirve University, 27260 Gaziantep, Turkey. E-mail: alisebetci@zirve.edu.tr.

10.1021/ct900545v
Published on Web 12/14/2009

troscopically accurate polarizable force fields/potential energy surface models, such as the TTM-xF models for water.[13,14] The key difference here is that we perform all calculations on-the-fly instead of fitting to a fixed classical functional form.

Unfortunately, the incorporation of a force field into the model raises the question of force-field parametrization. Previously,[28] we tested the Amoeba[29] polarizable force field for approximating the many-body induction effects. While Amoeba works reasonably well, it can exhibit systematic errors in predicting the many-body induction contributions (as compared to quantum mechanical results). For example, Amoeba systematically overestimated the many-body induction energies in a set of water clusters.[28] In addition, like most standard force fields, Amoeba has only been parametrized for a relatively small number of molecules, and its existing parameters are not necessarily transferable to new systems.

In this paper, we demonstrate that force-field parameters needed for the HMBI model can be obtained on-the-fly in an ab initio fashion, virtually eliminating the need for empirical parameters in the model. At each given molecular geometry of a system, the one- and two-body interactions are computed quantum mechanically, and a force field is parametrized to reproduce the many-body induction effects for that specific geometry.

In general, ab initio force-field (AIFF) parametrization is a difficult problem that many research groups have studied.[6,7,15,30−38] However, HMBI simplifies the task in two ways. First, the HMBI model only uses the many-body electrostatic induction effects from the force field. Intramolecular, two-body dispersion, two-body exchange-repulsion, and two-body electrostatic/induction effects are all treated quantum mechanically. Many-body dispersion and exchange−repulsion are neglected (at least in the model's current form). Thus, HMBI requires only an AIFF model for self-consistent polarization. Second, while many-body effects are frequently too large to neglect, their energetic contribution remains relatively small. In water clusters, which have significant many-body effects, these effects account for only about 15% of the total interaction energy[39] (though in small clusters these effects can be much larger[40,41]). In other words, HMBI requires a fairly simple force field, and the relatively small size of its contribution reduces the impact of inadequacies in the force field. As explained in section 2.2, we use existing techniques to determine the actual force-field parameters.

Polarizable force fields come in many forms.[15,42−44] The ones considered here are based on atom-centered distributed multipole expansions and atomic polarizabilities, for which we use ab initio parametrization methods developed in other groups.[34,45−47] In section 2, we discuss the HMBI model and the techniques used for constructing the force fields from first principles. After explaining the computational details in section 3, we demonstrate the method on a series of test clusters exhibiting large many-body effects in section 4.

2. Theory

2.1. The HMBI Model. The particular hybrid many-body interaction (HMBI) expansion approach for modeling systems of interacting molecules used here has been presented previously,[28] so we only briefly review the formalism. The total energy of a system of interacting molecules can be viewed in terms of the energies of individual molecules, their two-body (or pairwise) interactions, and the three-body and higher ("many-body") intermolecular interaction corrections, according to the many-body interaction expansion:

$$E_{\text{total}} = \sum_i E_i + \sum_{ij} \Delta^2 E_{ij} + \sum_{ijk} \Delta^3 E_{ijk} + \ldots \quad (1)$$

where E_i is the energy of the ith molecule, $\Delta^2 E_{ij}$ is the pairwise-interaction energy between two molecules i and j ($\Delta^2 E_{ij} = E_{ij} - E_i - E_j$, where E_{ij} is the total dimer energy), $\Delta^3 E_{ijk}$ is the three-body interaction correction between molecules i, j, and k ($\Delta^3 E_{ijk} = E_{ijk} - \Delta^2 E_{ij} - \Delta^2 E_{ik} - \Delta^2 E_{jk} - E_i - E_j - E_k$, where E_{ijk} is the total trimer energy), etc.

Though they typically contribute less than the one- and two-body terms, the many-body interactions are often non-negligible. In systems containing polar molecules, these many-body terms are typically dominated by induction, which can be approximated relatively easily.[8,26,28] In particular, a classical polarizable force field can approximate these terms at very low computational cost. With this approximation and some straightforward rearrangement, we obtain:[28]

$$E_{\text{total}}^{\text{HMBI}} = E_{\text{total}}^{\text{MM}} + \sum_i (E_i^{\text{QM}} - E_i^{\text{MM}}) + \sum_{ij} (\Delta^2 E_{ij}^{\text{QM}} - \Delta^2 E_{ij}^{\text{MM}}) \quad (2)$$

We emphasize that this HMBI model differs from conventional QM/MM models in that it partitions the system based on the type of interaction rather than by using a spatial criterion. HMBI's spatially homogeneous treatment of the entire system makes it ideal for treating molecular condensed-phase systems in which the important chemistry arises through intermolecular interactions dispersed over a large spatial area.

In addition HMBI sums many-body terms through all orders. Subsequent higher-order terms in the MBI expansion often alternate in sign. Truncating the series at a given order can introduce surprisingly large errors due to the absence of cancelation from higher-order terms.[28] By summing through all orders, HMBI avoids this pitfall.

2.2. Ab Initio Force Field (AIFF) Parameterization. The accuracy of an HMBI model depends critically on the polarizable force field used to approximate the many-body induction effects. The Amoeba force field tested previously performed moderately well, but evidence suggests it could be improved.[28] In this article, we demonstrate that an AIFF for many-body induction, in which the parameters are obtained directly from quantum mechanical monomer calculations, significantly improves results while simultaneously reducing the need for user-intervention during the force-field parametrization.

Specifically, we use a force field based on an atom-centered distributed-multipole representation[48−50] of the electron density and atom-centered local polarizabilities. To compute the polarizabilities, the molecular static polarizability is calculated using Kohn−Sham linear-response theory. The molecular polarizability is then distributed and localized to individual atoms according to the Williams−Stone−Misquitta procedure.[34,46,47] No intramolecular or two-body intermolecular force-field terms are required, because these are treated at the QM level.

The use of such procedures to parametrize a force field is not unique to our work. These particular procedures have been used to parametrize successful force fields for predicting the structures of molecular crystals,[51] for example. More generally, distributed multipoles (and sometimes polarizabilities) are widely used in force fields today (see, for example, refs 6, 7, 15, 34−38). However, these properties are rarely recomputed at every new geometry or point of interest on a potential energy surface to obtain a geometry-specific force field for many-body induction. In a purely classical model, doing such calculations would computationally overwhelm the relatively small effort required to evaluate the force field. Compared to the cost of evaluating the two-body quantum mechanical interactions in eq 2, however, the evaluation of these force-field parameters is reasonable. This means that the many-body induction force-field parameters can be recomputed as needed in HMBI. The repeated QM calculations (both one- and two-body energies and parametrizing the AIFF) obviously make HMBI much much more expensive than a simple force-field evaluation, but it is also widely applicable without requiring substantial hand reparameterization from the user.

Several important technical issues arise in constructing the force field. First, the distributed multipoles and local polarizabilities can be computed to different ranks. Previous studies have found that up to rank 4 (hexadecapole) multipole moments on non-hydrogen atoms and rank 1 moments (dipole) on hydrogen atoms, along with up to rank 2 (quadrupole) polarizabilities, perform well for general-purpose force fields,[34,47] so we use these same ranks here. In other words, the polarization model includes up to induced quadrupole effects.

Second, approximating electrostatic interactions with multipole interactions is only rigorously valid at long ranges. Therefore, it is necessary to introduce a damping function to attenuate the induction energy at short-range and to avoid the "polarization catastrophe." Numerous approaches to damp these interactions exist.[52−57] We apply the commonly used damping function proposed by Tang and Toennies,[58]

$$f_n(R) = 1 - \sum_{k=0}^{n} \left(\frac{(\beta R)^k}{k!} \right) e^{-\beta R} \tag{3}$$

to damp a R^{-n} term in the multipolar interaction. This damping function requires a (usually empirical) damping factor β. Misquitta and co-workers have analyzed these issues in some detail,[47] and they proposed a simple model for predicting the damping factor β single-molecule-containing system based on the ionization potential I,

$$\beta_{\text{pred}} = 2(2I)^{1/2} \tag{4}$$

They also proposed a related expression for systems containing multiple species. We will test this damping factor in our systems, and we will also explore treating this damping factor as an empirical parameter. Though the latter option is contrary to the spirit of the AIFF, we find it is unfortunately necessary in most of the cases examined here.

Third, one would ideally compute the multipoles and polarizabilities separately for each molecule in the cluster (and at each step in a molecular dynamics simulation or geometry optimization). However, the calculation of the polarizabilities in particular can be time-consuming. In practice, the polarizabilities of a water molecule, for example, will need to be computed repeatedly at slightly varying geometries. The polarizabilities do not change substantially with moderate changes in the molecular geometry, suggesting that computational savings might be obtained by approximating the polarizabilities by using the values obtained at similar geometries. In contrast, we find the multipole moments to be more sensitive to geometry. Computation of the multipole moments, however, requires much less computational effort than the polarizabilities. The effect of approximating these quantities with their equilibrium geometry values will be investigated below.

Fourth, here we only consider systems in which each molecule is completely contained within a given "monomer" (as in a set of small, interacting molecules). Monomer boundaries never cross a covalent bond. If instead the monomers were composed of molecular fragments, the "many-body" effects will include intramolecular interactions. These interactions are much stronger and harder to approximate with simple classical electrostatics/induction than intermolecular many-body effects. On the other hand, methods such as the fragment molecular orbital method[4,5] have been adapted to such cases with good results, so such partitioning merits future investigation.

3. Computational Details

The HMBI model requires the specification of both the quantum mechanical method and the polarizable force field. Unless otherwise specified, we use resolution-of-the-identity second-order Møller−Plesset (RI-MP2)[59−61] theory with the Dunning aug-cc-pVTZ basis set[62] and its corresponding auxiliary set[63] in the frozen core approximation for the QM calculations in eq 2. The aug-cc-pVTZ basis provides a reasonable balance between basis-set completeness and computational affordability. Dual-basis Hartree−Fock (HF)/MP2 algorithms accelerate the RI-MP2 calculations further.[64−66] These approximations make the benchmark MP2 calculations on the full clusters considered here affordable while introducing insignificant errors in relative energies. MP2 cannot describe all intermolecular interactions,[67−69] but more accurate methods would make the benchmark calculations unfeasible. MP2 serves as a compromise between accuracy and computational affordability here. Integral thresholds and HF energy convergence criteria were set to 10^{-14} and 10^{-8} a.u., respectively, to minimize numerical noise issues in

computing the many-body energies. These calculations were performed using a development version of Q-Chem, version 3.1.[70]

We compare six different approximations for the many-body (MM) terms in eq 2. In the first approximation, we neglect the many-body terms, which corresponds to setting all MM terms in eq 2 to zero. This approximation is labeled as "no MB" in the tables and figures below, and it serves as a measure of the importance of many-body effects. The second approximation uses supermolecular HF calculations to approximate the many-body induction effects quantum mechanically. This is equivalent to models explored by several other groups.[23,25,26,71] HF can describe many-body induction very accurately, and it serves as a target for the other five methods. Unfortunately, it is very expensive to perform the supermolecular HF calculation on the entire system. Using fast modern MP2 algorithms, the HF step is often the computational bottleneck in a large MP2 calculation.[72] The models that approximate many-body induction classically are much less expensive. In the third approximation, many-body induction is treated using the standard Amoeba polarizable force field,[29] as implemented in the Tinker[73] software package, just as was done in ref 28.

In the fourth approximation, we construct and use the AIFF described in section 2.2. The CamCasp[74] software package, which invokes the Dalton[75] quantum chemistry package, is used to construct the AIFF by predicting the distributed multipoles and polarizabilities from quantum mechanical calculations on each monomer in the system. Multipole moments are computed to rank 4 (hexadecapole) on heavy atoms and rank 1 (dipole) on hydrogen atoms.

The polarizabilities are computed to rank 2 (rank 1 on hydrogen atoms) using the Coupled Kohn−Sham (CKS) propagator. These ranks have been shown to reproduce DFT-based symmetry-adapted perturbation theory induction energies well.[47] Unless otherwise specified, the underlying Dalton calculations use the aug-cc-pVTZ basis and the hybrid PBE0[76] exchange and correlation functional. Our procedure uses two calculations per monomer: one to determine the RI-MP2 energy and another to find the force-field parameters with DFT. This is done to allow the use of pre-existing CamCasp software routines when determining the force-field parameters. The cost of performing an extra single-point energy calculation for each monomer is negligible compared to the other parts of the calculations. Ionization potentials of 0.4638 (water),[77] 0.3720 (formamide),[78] 0.5896 (hydrogen fluoride),[79] 0.3595 (glycine neutral), and 0.3042 au (glycine zwitterion) are employed in an asymptotic correction[80,81] for the PBE0 calculations. The two glycine vertical ionization potentials were calculated at the B3LYP/6-311++G(3df,2pd) level by taking the difference in energies between each molecule and its singly ionized cation. The Orient[82] software package is used to calculate self-consistent induction energies for the clusters based on the CamCasp multipole moments and polarizabilities.

Two different damping factors were considered for truncating short-range multipolar interactions for most systems. First, we used the damping factor predicted by eq 4 and the ionization potentials listed above. Second, we treated the damping factor as an adjustable parameter and loosely optimized it to obtain both a small mean error and a narrow error distribution across the different clusters in each test set. Having a small mean error is required for accurate binding energies, while having a narrow range of errors is critical for accurate relative energies.

The fifth model examined is the electrostatically embedded pairwise-additivity (EE-PA) model of Dahlke and Truhlar.[8] Like the "no MB" case, EE-PA neglects many-body terms; rather, it uses an alternative approach for capturing many-body induction effects. The monomers and dimers are polarized by embedding them in a field of B3LYP/6-31G* Mulliken point charges centered at the nuclear positions of all other monomers. A more accurate variation on EE-PA also includes three-body effects, but that model is substantially more expensive than either EE-PA or the HMBI model discussed here, so we do not consider it. See ref 8 for further details on EE-PA.

The sixth and final model we consider is the self-consistent charge (or charge + dipole) embedded binary interaction scheme of Kamiya and co-workers.[11] In accord with ref 11, we refer to these methods as "binary+ESP" and "binary+ESP-dipole", respectively. Binary+ESP is essentially the same as EE-PA, except that the embedding charges are determined self-consistently by fitting to the electrostatic potential. Binary+ESP-dipole also includes some explicit dipole fitting. Both methods include counterpoise (CP) corrections[83] for basis set superposition error (BSSE). See the original work for more details. The self-consistent treatment of point charges helps capture induction effects through all orders. Because we do not have our own implementation of this method, we present binary+ESP-type results only for two test systems (water hexamers and glycine−water clusters) for which published data is already available.

These approximations are compared to the full RI-MP2 results on water clusters (Test Sets A−D), formamide clusters (Test Sets E and F), hydrogen fluoride clusters (Test Set G), and glycine−water clusters (Test Set H). All of these clusters exhibit strong many-body effects and therefore represent challenging test cases for HMBI models. To facilitate comparisons between the various systems, we report size-intensive energies obtained by dividing the resulting energies by the number of monomers.

Structures for the test clusters in Test Sets B, C, F, and G are provided as Supporting Information. The geometries used in Test Sets A, D, E, and H come from the literature, as discussed below. Tables containing all raw data used to generate the statistical results presented below are also available as Supporting Information.

4. Results and Discussion

4.1. Water Clusters. Water clusters have been investigated extensively in the last two decades because of both their general interest and strong many-body effects,[8,10,11,19,23,40,71,84−91] which means that they provide a challenging test case for approximate MBI expansion methods. We first consider the ability of the approximate methods to reproduce the binding energies of a series of

Table 1. Mean Signed Errors, Standard Deviations, and Total Error Ranges (in kJ/mol per molecule) for Various HMBI Many-Body Approximations Relative to RI-MP2[a]

	errors due to approximate many-body treatment					
	HF[b]	AIFF/β_{pred}[b]	AIFF/β_{opt}[b]	Amoeba[b]	EE-PA[c]	no MB[d]
			Test Set A: 16 $(H_2O)_n$ Clusters, $n = 5-20$			
mean	−0.03	−1.53	−0.04	−0.86	0.11	5.47
std dev	0.05	0.23	0.10	0.17	0.15	0.62
range	0.19	0.73	0.34	0.57	0.53	2.09
			Test Set B: 50 $(H_2O)_8$ Configurations			
mean	−0.02	−1.17	−0.01	−0.65	0.04	4.32
std dev	0.07	0.38	0.17	0.21	0.08	0.61
range	0.35	1.99	0.82	0.90	0.41	2.93
			Test Set C: 36 $(H_2O)_{10}$ Rotation PES Points			
mean	−0.06	−0.66	−0.24	−0.67	−0.43	1.54
std dev	0.09	0.58	0.12	0.54	0.13	1.91
range	0.27	1.86	0.48	1.78	0.59	6.55
			Test Set D: 8 $(H_2O)_6$ Configurations (see Table 2)			
			Test Set E: 51 $(HCONH_2)_8$ Configurations			
mean	0.04	−0.30	−0.04	1.20	0.32	3.06
std dev	0.10	0.19	0.17	0.32	0.18	0.83
range	0.43	0.97	0.71	1.50	0.79	4.00
			Test Set F: 50 Additional $(HCONH_2)_8$ Configurations			
mean	0.00	−0.37	−0.07	1.02	0.34	3.15
std dev	0.07	0.16	0.12	0.25	0.12	0.69
range	0.37	0.67	0.52	1.18	0.51	2.54
			Test Set G: 8 Cyclic $(HF)_n$ Clusters, $n = 3-10$			
mean	0.16	0.69	[e]	[f]	1.46	13.42
std dev	0.02	0.10	[e]	[f]	0.33	3.90
range	0.08	0.33	[e]	[f]	0.93	11.18
			Test Set H: Glycine·$(HO_2)_7$ Clusters (see Table 4)			

[a] For the AIFF, two different damping factors are considered. β_{pred} is the factor computed from eq 4 and β_{opt} is the value that approximately minimizes the mean error and range of errors. The actual β values are given in the text. [b] Technique used to approximate many-body terms in HMBI. [c] The EE-PA method from ref 8. [d] Many-body effects are neglected entirely in eq 2. [e] $\beta_{opt} = \beta_{pred}$, so the results are identical. [f] The Amoeba force field has not been parametrized for the HF molecule.

small- and medium-sized water clusters. Then we investigate how well these methods can reproduce potential energy surface (PES) energetics away from the minimum energy structures. Finally, we examine the ability of these methods to discriminate between a series of low-lying water hexamer isomers.

4.1.1. Performance on Optimized Water Clusters: Test Set A. We begin by computing the binding energies of 16 clusters containing 5 to 20 water molecules (Test Set A) with RI-MP2 and the various approximate many-body models, using optimized structures obtained from the Cambridge Cluster Database.[92,93] These particular clusters have been examined in similar studies previously[8,28] and therefore provide a useful comparison.

Summary statistics comparing HMBI and various treatments of many-body induction interactions with benchmark full-cluster RI-MP2 energies are presented in Table 1. Figure 1 shows the error distributions for each many-body induction approximation as obtained by combining the results from this test set and those from Test Sets B and C. As can be seen in the table, simply neglecting the many-body terms introduces both a large mean error (with error defined as $E_{approx} - E_{MP2}$) and a wide range of errors (defined as the difference between the two most extreme errors in the distribution). In contrast, the HF approximation for many-body induction is extremely accurate, giving a narrow error distribution peaked near zero. Its mean signed error is −0.03 kJ/mol per monomer, and the range of errors is 0.19 kJ/mol per monomer. As noted previously,[28] approximating the HMBI many-body induction terms with Amoeba is much better than simply neglecting them, but Amoeba significantly overestimates the many-body contribution in these water clusters, with a mean error of −0.86 kJ/mol per monomer

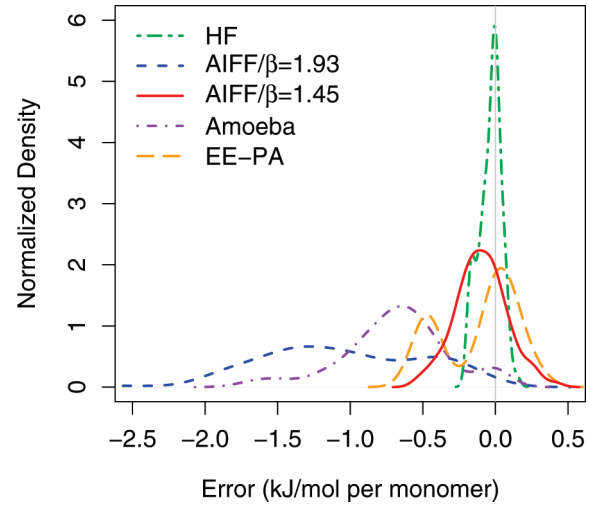

Figure 1. Error distributions for approximate many-body HMBI treatments relative to RI-MP2/aug-cc-pVTZ in 102 water clusters (Test Sets A−C combined).

and a range three times wider than that of the HMBI/HF many-body approximation.

Next we consider the HMBI/AIFF results using two different damping factors. The first, $\beta_{pred} = 1.93$, is obtained using eq 4 and 0.4638 au for the ionization potential of water. For these water clusters, this damping factor appears to be a poor choice for the HMBI model, and the HMBI/AIFF/β_{pred} errors (mean error of −1.53 kJ/mol per monomer) are notably worse than those of the Amoeba.

If we loosely optimize β to minimize the mean error and error range for HMBI/AIFF, we find $\beta_{opt} = 1.45$. With this damping factor, the AIFF substantially out-performs the Amoeba force field for describing many-body induction. Though they are not as accurate as the HMBI/HF many-body results, the HMBI/AIFF/β_{opt} results exhibit a mean error and range of only −0.04 and 0.34 kJ/mol per monomer, respectively, and the cost of HMBI/AIFF is orders of magnitude cheaper than HMBI/HF for systems containing many monomers. The EE-PA point-charge embedding method[8] gives a larger mean error and range of errors (−0.11 and 0.53 kJ/mol per monomer, respectively).

One might be concerned that fitting β to these optimized water structures could bias the many-body force field toward these particular structures. However, this β value also performs well for the other water test sets examined below, which were not used in fitting β_{opt}. Therefore, $\beta_{opt} = 1.45$ appears to be a "universal" value for damping water−water interactions in this AIFF.

4.1.2. Performance across the Potential Energy Surface: Test Sets B and C. To examine the performance of the model away from equilibrium, we consider two different test sets. The first, Test Set B, consists of 50 $(H_2O)_8$ cluster geometries sampled at uniform intervals from a classical molecular dynamics simulation, which is representative of the potential energy surface sampling that occurs in typical condensed-phase studies. The second, Test Set C, consists of a series of points along a one-dimensional potential energy surface (PES) coordinate that disrupts the hydrogen bonding networks. Test Set C demands that the approximate method is able to describe widely varying intermolecular interactions accurately. As mentioned above, we continue to use $\beta_{opt} = 1.45$ for these test sets. Statistical results for both test sets are summarized in Table 1. See also Figure 1.

Optimized water clusters (e.g., Test Set A) often achieve enhanced stability through strong hydrogen bond cooperativity effects. In contrast, clusters sampled along a dynamics trajectory (Test Set B) will frequently exhibit a weaker and less homogeneous distribution of many-body effects. On the basis of the errors in the "no MB" column of Table 1, we observe that the mean many-body effects are smaller for Test Set B (−4.3 kJ/mol per monomer) than for Set A (−5.5 kJ/mol per monomer), and the distribution of many-body effects spans a broader range in Set B (2.9 vs 2.1 kJ/mol per monomer), as expected. The ability to reproduce this broad range of interactions is just as important as reproducing equilibrium structures.

The HMBI/AIFF/β_{opt} mean error for Test Set B is similar to the one for Test Set A. The range of errors is quite a bit larger in this set, but a significant increase in the range of errors is observed for the other HMBI models as well. This observation suggests that the varied many-body effects are indeed harder to approximate in these nonequilibrium structures. Nevertheless, the HMBI/AIFF model still provides good accuracy. EE-PA performs particularly well for this test set, exhibiting a range of errors similar to HMBI/HF and much smaller than HMBI/AIFF.

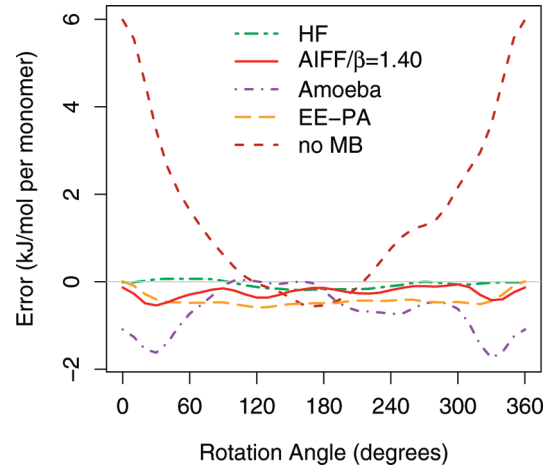

Figure 2. Errors (in kJ/mol per monomer) in binding energies for $(H_2O)_{10}$ relative to RI-MP2/aug-cc-pVTZ as each water molecule is rotated in place, disrupting the hydrogen bonding network (Test Set C).

Next we examine the performance of these models along a challenging coordinate of the potential energy surface, Test Set C. Starting with the $(H_2O)_{10}$ structure from Test Set A, we generate a PES slice by simultaneously rotating each water molecule in place about the global coordinate x-axis passing through its center of mass (the geometries are available as Supporting Information). Energies were evaluated at 10° intervals in the range 0−360°. This rotation completely disrupts the hydrogen bonding network and the corresponding hydrogen-bond cooperativity effects. In fact, the rotation reduces the many-body contribution from −6 kJ/mol per monomer at 0° to +0.5 kJ/mol per monomer (anticooperative effects) at 180°, as reflected by the "no MB" line in Figure 2. The binding energy of this cluster ranges from roughly −40 kJ/mol per monomer (at 0°) to +10 kJ/mol per monomer (at 180°).

As always, HF approximates the many-body induction effects well, and the mean error and standard deviation along the PES slice are quite small. Both the HMBI/AIFF/β_{opt} and EE-PA many-body induction approximations perform well, too. While HMBI/AIFF is slightly more faithful to the RI-MP2 PES than EE-PA across much of the surface (as exhibited by its smaller mean error), the maximum HMBI/AIFF errors are similar to those of EE-PA (as exhibited by their similar standard deviations and error ranges). In either case, the errors are small relative to the overall binding energies. In addition, this system further demonstrates the superiority of the AIFF over an off-the-shelf force field such as Amoeba for our purpose. Amoeba has particular difficulties describing the wide range of interactions present along this PES slice. Together, the results from Test Sets B and C demonstrate that HMBI/AIFF performs well across the PES

Table 2. Errors (in kJ/mol per monomer) in Predicted Counterpoise-Corrected MP2/aug-cc-pVDZ(Cartesian functions) Binding Energies for Low-Lying $(H_2O)_6$ Isomers (Test Set D)[a]

	binding energy	errors due to approximate many-body treatment				
isomer	MP2[b]	aug-cc-pVDZ AIFF[c]	Sadlej AIFF[c]	binary+ESP[d]	binary+ESP-dipole[d]	EE-PA
Book2	−28.88	0.54	0.32	−0.58	−0.69	0.80
C8	−29.01	0.67	0.33	−0.52	−0.64	0.83
C4	−27.53	0.68	0.42	−0.71	−0.79	0.73
C6	−28.59	0.77	0.45	−0.59	−0.70	0.92
Prism3	−28.07	0.55	0.35	−0.70	−0.85	0.79
Cage	−29.09	0.64	0.47	−0.58	−0.71	0.85
Chair	−28.71	0.98	0.43	−0.36	−0.50	1.02
C	−27.90	1.00	0.46	−0.40	−0.51	0.93
mean error		0.73	0.40	−0.55	−0.67	0.86
mean rel error[e]		0.08	−0.07	0.03	0.04	0.01
range		0.46	0.15	0.36	0.36	0.29

[a] Aug-cc-pVDZ and Sadlej in the column headings refer to the basis sets used to compute the AIFF parameters only. [b] Valiron-Mayer[97] counterpoise-corrected result, from ref 11. [c] Using $\beta_{opt} = 1.45$, as determined in section 4.1. [d] From ref 11. [e] Errors in energies relative to the Cage isomer.

and that the optimized damping factor β_{opt} works for water−water interactions well-away from equilibrium structures.

4.1.3. Ability To Differentiate between Low-Lying Isomers of $(H_2O)_6$: Test Set D. Finally, we examine a set of low-lying water hexamer isomers (Test Set D) that were used to benchmark the self-consistent charge-embedding binary interaction method of Kamiya and co-workers.[11] The geometries for these clusters originate from ref 94. The small relative energy differences in these clusters make their accurate prediction difficult. For these particular cluster geometries and MP2 energy calculations, the Cage isomer is the most stable. However, many possible isomers exist even within a single structural motif (such as a prism), and these isomers can differ substantially in energy. In contrast to the results for these particular isomers, recent accurate calculations predict that a Prism isomer, rather than a Cage isomer, is the most stable.[95,96] Nevertheless, our purpose is simply to compare the degree to which the various approximate many-body interaction models can reproduce the MP2 energies for these specific isomers, rather than to try to identify the globally optimal water hexamer structure.

To match the results from ref 11, we use canonical MP2/aug-cc-pVDZ (with Cartesian basis functions and no dual-basis approximation), and we found it necessary to correlate all electrons. Because ref 11 focuses on counterpoise-corrected results, we also apply a standard BSSE CP correction when evaluating each two-body HMBI interaction here. Table 2 presents the errors in binding energies predicted by HMBI, EE-PA, and two binary-interaction methods from ref 11.

Initially we computed the AIFF parameters in the aug-cc-pVDZ basis. However, this basis is too small to predict reliable polarizabilities, and the HMBI/AIFF model with aug-cc-pVDZ AIFF parameters performs much worse than if the compact Sadlej triple-ζ basis set[98,99] is used for parametrizing the AIFF. The Sadlej basis set predicts these polarizabilities with nearly aug-cc-pVTZ quality using roughly half the basis functions. The deficiencies in the aug-cc-pVDZ polarizabilities are not overcome by reoptimizing the damping factor β (rather, we continue to use the same $\beta_{opt} = 1.45$ parameter for all water clusters). Using the aug-cc-pVDZ basis for the MP2 one- and two-body calculations and the Sadlej basis for predicting the AIFF parameters, we obtain a mean error of 0.40 and an error range of 0.15 kJ/mol per monomer relative to the CP-corrected MP2 binding energies. Both EE-PA and the binary-interaction models predict notably larger mean errors and error ranges than the AIFF(Sadlej) model, though they perform similar to or better than the AIFF(aug-cc-pVDZ) many-body induction model. Unfortunately, using a triple-ζ basis to compute the monomer polarizabilities substantially increases the cost of the double-ζ quality HMBI calculation, particularly in such small clusters. On the other hand, triple-ζ or larger basis sets are frequently important when describing intermolecular interaction without excessive BSSE.

If we consider relative energies instead of binding energies, the mean errors decrease substantially for all models. The HMBI/AIFF mean relative error is slightly larger than those for EE-PA and the binary interaction models, but it still performs very well. However, while several of the methods incorrectly order one of more isomers relative to the MP2 results, HMBI/AIFF(Sadlej) alone fails to predict the Cage structure as the most stable. It places the C8 isomer 0.06 kJ/mol per monomer *below* the Cage isomer. While the ordering is incorrect, the energy gap between them is very small and well within there margin of errors for the method. Furthermore, the ordering is corrected if the one- and two-body energies are computed in the aug-cc-pVTZ basis. Furthermore, the AIFF correctly predicts the prism as being more stable than the cage in the two water hexamer geometries considered in Test Set A, in agreement with both our own RI-MP2 calculations and highly accurate calculations.[95,96]

Overall, HMBI with the AIFF represents a good approximation for both single point energies of equilibrium structures (Test Sets A and D) and across the potential energy surface (Test Sets B and C) for small- and medium-sized water clusters. The combined results from Test Sets A−C in Figure 1 demonstrate the importance of using the proper damping factor ($\beta_{opt} = 1.45$ here), and the improvement of the AIFF over the Amoeba force field for describing many-body induction is clear. EE-PA performs similarly to the HMBI/AIFF model here, though it does exhibit slightly broader error distributions. Next, we evaluate the perfor-

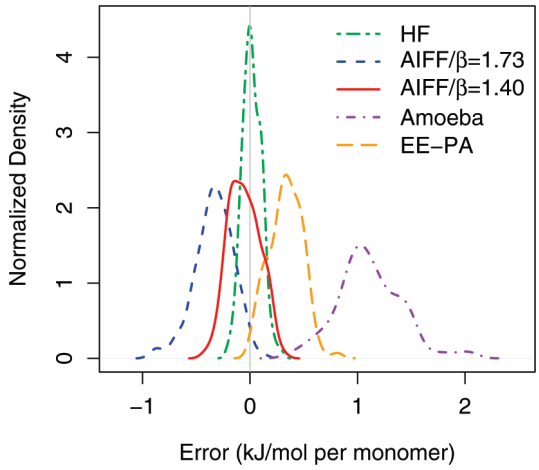

Figure 3. Error distributions for approximate many-body HMBI treatments relative to RI-MP2/aug-cc-pVTZ for 101 formamide clusters (Test Sets E and F combined).

mance of these models on formamide clusters, which also exhibit strong many-body effects.

4.2. $(HCONH_2)_8$ Clusters: Test Sets E and F. Formamide clusters are also noted for their strong many-body effects, and they have been studied by many groups.[100−107] We revisit the 51 formamide octamer geometries (Test Set E) considered in ref 28 with the AIFF. Once again, the two different damping factors are considered. Equation 4 predicts $\beta_{pred} = 1.73$, while optimization gives $\beta_{opt} = 1.40$. The error distributions relative to RI-MP2 for both Test Sets E and F are plotted in Figure 3, and statistical summaries are presented in Table 1.

As before, the AIFF/β_{opt} substantially out-performs Amoeba for approximating the RI-MP2 many-body induction. It also provides a significant improvement over neglecting many-body induction entirely. The mean HMBI/AIFF/β_{opt} error is only −0.04 kJ/mol per monomer (compared to the HMBI/HF error of 0.04 kJ/mol per monomer and an HMBI/Amoeba error of 1.20 kJ/mol per monomer). The HMBI/AIFF/β_{opt} error range of 0.71 kJ/mol per monomer is similar to what was observed for water/Test Set B. Again, this error distribution is broader than that of HMBI/HF, but the AIFF still does a remarkable job of approximating many-body induction effects. The range of the EE-PA error distribution (0.79 kJ/mol per monomer) is slightly broader than for HMBI/AIFF/β_{opt}, and the relatively large EE-PA mean error (0.32 kJ/mol per monomer) reflects a somewhat systematic underestimation of many-body effects.

We test the optimized β_{opt} value for transferability on 50 new $(HCONH_2)_8$ configurations (Test Set F). As listed in Table 1, the AIFF performs very well on this set too. The many-body effects are more uniform in size across the configurations sampled in the second set (compare the error ranges in the "no MB" columns for both Test Sets E and F), and the error ranges for the approximate models are commensurately smaller. Again, HMBI/AIFF/β_{opt} outperforms EE-PA by more than 0.3 kJ/mol per monomer in the mean error, while the error ranges are fairly similar. Overall, HMBI/AIFF/β_{opt} reproduces the total binding energies of these $(HCONH_2)_8$ clusters with a mean absolute percent error of 0.66%, compared to 1.70% for EE-PA and 5.59% for HMBI/Amoeba. The vastly more expensive HMBI/HF approximation reduces the mean absolute percent error in binding energies to 0.35%.

Table 3. Binding Energies and Errors Arising from the Approximate Treatment of Many-Body Interactions (in kJ/mol per monomer) As Compared To Full RI-MP2 Binding Energies for a Series of Cyclic Hydrogen Fluoride Clusters, $(HF)_n$ (Test Set G)

	binding energy	errors due to approximate many-body treatment			
n	RI-MP2[a]	HF[a]	AIFF/β_{pred}[a]	EE-PA[b]	no MB[c]
3	−21.67	0.12	0.81	0.73	4.64
4	−30.24	0.18	0.71	1.19	10.90
5	−33.20	0.14	0.48	1.52	14.24
6	−33.96	0.16	0.64	1.65	15.17
7	−34.17	0.16	0.68	1.65	15.39
8	−34.34	0.16	0.71	1.64	15.63
9	−34.38	0.15	0.70	1.64	15.60
10	−34.66	0.20	0.82	1.65	15.82

[a] Technique used to approximate many-body terms in HMBI. [b] The EE-PA method from ref 8. [c] Many-body effects are neglected entirely in eq 2.

We also investigate the use of the compact Sadlej triple-ζ basis set for computing the AIFF parameters for formamide clusters. As mentioned above, water polarizabilities were poorly reproduced in the aug-cc-pVDZ basis, but they can be predicted accurately with the Sadlej basis set, which is about half the size of the aug-cc-pVTZ basis set. The HMBI/AIFF/β_{opt} mean error, standard deviation, and range of errors for Test Set E with Sadlej-based AIFF are −0.05, 0.16, and 0.70 kJ/mol per monomer, which are essentially identical to the aug-cc-pVTZ results. Predicted many-body contributions for individual cluster geometries differ by only a few hundredths of a kJ/mol per monomer between the two basis sets. As recognized previously,[47] the Sadlej basis set provides a useful compromise between basis set size and accuracy in parametrizing the AIFF.

4.3. Cyclic $(HF)_n$ ($n = 3-10$) Clusters: Test Set G. Next, we consider a series of increasingly large cyclic hydrogen fluoride clusters (Test Set G). Hydrogen fluoride clusters are noted for their extraordinarily strong many-body effects, making them an extremely challenging test case. Many experimental[108−110] and theoretical[111−121] studies on HF clusters have been performed to understand their cluster and condensed-phase behavior.

Table 1 (statistical summary) and Table 3 (detailed results) demonstrate the performance of HMBI on these cyclic hydrogen fluoride clusters. The Amoeba force field has not been parametrized for the HF molecule, so we do not report HMBI/Amoeba results for these clusters. The large many-body hydrogen-bond cooperativity effects are clearly evident in Table 3, and we observe (as has been reported previously[111]) that they saturate at almost 16 kJ/mol per monomer for clusters of eight monomers or more.

As always, Hartree−Fock theory reproduces the many-body induction effects very well, though the mean error of 0.16 kJ/mol per monomer is significantly larger than for any of the other systems examined here. HMBI/AIFF/$\beta_{pred} = 2.17$ also performs rather well, but the mean error (0.69 kJ/mol per monomer) is again much larger than for formamide or water. On the other hand, the width of the error distribution

Table 4. Predicted Energy Difference ($\Delta E = E_{neut} - E_{zwit}$ in kJ/mol) between Neutral and Zwitterionic Glycine$-(H_2O)_7$ Clusters (Test Set H) Using Frozen-Core MP2/aug-cc-pVDZ(Cartesian functions) with and without Counterpoise Correction

	full	HMBI approximate many-body treatment					
isomer	MP2[a]	HF	AIFF, single β[b]	AIFF, three βs[c]	binary+ESP[d]	EE-PA	no MB
ΔE	29.0	28.8	32.8	32.4	n/a	35.4	47.1
CP-ΔE	15.5	15.0	19.0	18.6	20.5	19.1	33.3

[a] Valiron-Mayer[97] counterpoise-corrected result, from ref 11. [b] AIFF determined with Sadlej basis, using $\beta_{opt} = 1.5$. [c] AIFF determined with Sadlej basis, using $\beta_{H_2O-H_2O} = 1.45$, $\beta_{neut-H_2O} = 1.51$, and $\beta_{zwit-H_2O} = 1.44$. [d] From ref 11.

is only 0.33 kJ/mol per monomer. Thus, the AIFF systematically underestimates the many-body effects by about 5% in the larger clusters, which amounts to an average 2.3% error in the total binding energies. Unlike the water and formamide test cases above, no improvement is found by optimizing β. Increasing β decreases the mean error further but at the expense of increasing the error range. For this particular test set, eq 4 provides a useful prediction for β.

Although the HMBI/AIFF errors are more pronounced in these hydrogen fluoride clusters than in the water or formamide clusters examined above, they are substantially smaller than the EE-PA errors. The mean EE-PA error is nearly 1.5 kJ/mol per monomer (nearly 4.5% in the total binding energy), more than twice that of HMBI/AIFF. Likewise, the EE-PA error range is almost three times as wide as that of HMBI/AIFF.

Reference 11 presents results on a similar set of cyclic HF clusters containing 4−9 molecules. They find that binary+ESP gives a mean error and error range of 1.04 and 0.33 kJ/mol per monomer, respectively, while the binary+ESP-dipole model gives a mean error and error range of 0.32 and 0.08 kJ/mol per monomer. These numbers cannot be compared directly to our own because they use slightly different clusters/geometries, smaller basis sets, and include CP corrections. Nevertheless, a loose comparison suggests that HMBI/AIFF is probably performing with accuracy intermediate between binary+ESP and binary+ESP-dipole here.

4.4. Glycine in Water: Test Set H. As a final test, Table 4 presents the energy difference, $\Delta E = E_{neutral} - E_{zwitterion}$, between the neutral and zwitterionic forms of glycine microsolvated with seven water molecules (Test Set H)[10,11] using CP-corrected, frozen-core canonical MP2/aug-cc-pVDZ(Cartesian). These conditions were chosen to enable direct comparison with ref 11. As observed previously, BSSE effects are substantial in this system and basis set, and CP-correction reduces the energy gap by almost a factor of 2.[11] The microsolvated zwitterion is more stable than its neutral counterpart before and after the CP correction. Unsurprisingly, HMBI/HF faithfully reproduces the energy gap to within 0.5 kJ/mol with or without the CP correction. Note that we did not counterpoise-correct the HF many-body contribution. Valiron-Mayer CP corrections for many-body effects become prohibitively expensive for large clusters and cannot be used in practice for most systems. BSSE effects in the many-body terms tend to be less significant than in the two-body terms.[21]

The presence of two different species (or even three if the neutral and zwitterionic forms of glycine are considered distinct) raises the question of how best to treat the damping in the AIFF. Various approaches for handling damping in mixed systems have been proposed and tested in the literature (see, for example, refs 47, 55, 56). Here, we compare two different damping models on a training set of MP2 results from 18 glycine$-(H_2O)_2$ clusters (nine with neutral and nine with zwitterionic glycine), using the Sadlej basis set for the AIFF. The first model uses a single, system-wide empirical damping factor, just as in the examples described previously. We find that $\beta = 1.5$ is approximately optimal here, which is similar to the value obtained for pure water.

The second model uses three separate damping factors: one to describe water−water interactions (which we set to $\beta = 1.45$, as before) and additional parameters to describe water−neutral glycine and water−zwitterionic glycine interactions (which we optimized). We find $\beta_{neutral-H_2O} = 1.51$ and $\beta_{zwitterion-H_2O} = 1.44$.

Both damping models reproduce the counterpoise-corrected binding energy gap to within 3−3.5 kJ/mol. The three-parameter model performs 0.4 kJ/mol better, albeit at the undesirable expense of additional parameters. Once again, the AIFF parametrized in the aug-cc-pVDZ basis set performs slightly worse, predicting an energy gap that is 1.5 kJ/mol larger than the Sadlej AIFF gap. In this difficult system, counterpoise-corrected HMBI/AIFF, EE-PA, and binary+ESP all produce the correct qualitative trend, but they each predict an energy gap that is several kJ/mol to large. In contrast, the HMBI/HF predicted gap differs from the full MP2 one by only 0.5 kJ/mol. This discrepancy may stem from limitations in describing higher-order multipole induction effects in the simple classical models.

4.5. Approximating the AIFF Parameters. Finally, we briefly examine the possibility that the AIFF parameters can be approximated using the distributed multipoles and polarizabilities computed at the equilibrium geometry of a species. This would eliminate the need to compute these properties repeatedly for each monomer structure.

For the water clusters in Test Set A, using the multipoles and polarizabilities for water determined at its equilibrium geometry moderately degrades the results. It produces a mean error, standard deviation, and a range of errors of 0.18, 0.09, and 0.36, respectively. The geometries of individual water molecules do not differ too much in these clusters, so this result is unsurprising.

Formamide molecules, on the other hand, exhibit greater flexibility in our test sets. As can be seen for Test Set E in Table 5 and Figure 4, approximating the moments/polarizabilities with their equilibrium geometry values severely degrades the quality of the HMBI/AIFF approximation. The mean errors increase from −0.04 to 0.12 kJ/mol per monomer, while the range of errors more than doubles, from

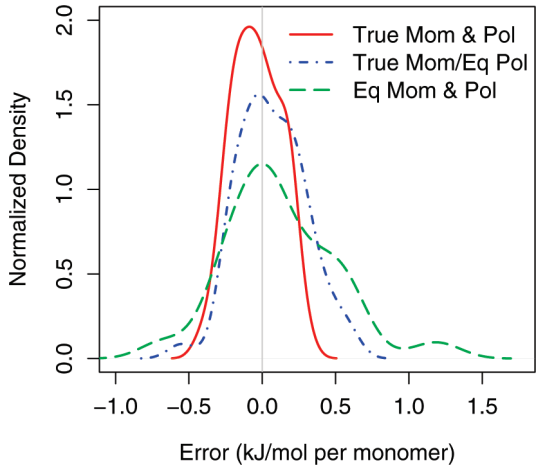

Figure 4. Distribution of errors relative to RI-MP2/aug-cc-pVTZ results as a function of the multipole moments and polarizabilities used in the AIFF/β_{opt} force field for the binding energies in 51 formamide octamer clusters (Test Set E).

Table 5. The Effect of Replacing the True Distributed Multipoles and Polarizabilities in the AIFF/β_{opt} with Those Computed at the Equilibrium Structure of an Isolated Formamide Molecule Using 51 $(HCONH_2)_8$ Configurations (Test Set E)[a]

	true moments and polarizabilities	true moments and equilib. polarizabilities	equilib. moments and polarizabilities
mean	−0.04	0.06	0.12
std dev	0.17	0.23	0.38
range	0.71	1.12	1.97

[a] Errors are given in kJ/mol per monomer.

0.71 to 1.97 kJ/mol per monomer. In contrast, using the equilibrium polarizabilities with the true distributed multipole moments is a much better approximation, with the mean error and range increasing to only 0.06 and 1.12 kJ/mol per monomer, respectively.

Physically, the polarizabilities are essentially an atomic property, even if the specific localization procedure is geometry dependent. So one expects that they will not vary too much with moderate conformational changes in the molecule. On the other hand, the density will likely be more sensitive to the details of the geometry and orbital overlap between atoms. This approximation is also computationally beneficial: calculating polarizabilities by solving the coupled Kohn−Sham equations requires substantially more effort than fitting distributed multipoles to the electron density. Of course, in systems containing many monomers, the QM calculation of the two-body effects will still be the computationally dominant step. In any case, these results make it clear that one should compute both the polarizabilities and multipoles for each molecule if very accurate results are needed. Similar results have been observed in developing more general force fields for flexible molecules.[122]

5. Conclusions

HMBI approximations represent interesting QM/MM models for studying large systems of interacting molecules at very low computational cost. It combines a quantum mechanical treatment of monomers and their pairwise interactions with a classical polarizable force-field approximation for many-body induction effects. Key advantages of HMBI include its quantum/classical partitioning based on the type of intermolecular interaction, its approximate inclusion of all orders in the many-body interaction expansion, and its ability to reproduce full quantum mechanical calculations accurately at much lower computational cost. The biggest weakness of an HMBI-type model has been its need for a parametrized polarizable force field. In this paper, we largely eliminate this weakness and demonstrate that very good HMBI results can be obtained when the force field is constructed on-the-fly from first-principles quantum mechanical calculations on the monomers.

We tested the model on water, formamide, hydrogen fluoride, and mixed glycine−water clusters, all of which exhibit strong many-body effects and test it strenuously. In each case, HMBI with the ab initio polarizable force field performed well. We also demonstrated that one can tolerably approximate the atomic polarizabilities by computing them at a single (or perhaps a few) representative geometries.

The performance of the HMBI/AIFF model demonstrated here is competitive with the related EE-PA and self-consistently embedded binary interaction methods. Using a supermolecular HF calculation to approximate the many-body effects[23,25,71] is significantly more accurate than any of the other approximations considered here, but it becomes cost-prohibitive for larger systems. The HMBI/AIFF formulation exhibits advantages and disadvantages compared to these other methods. The expense of computing distributed polarizabilities is probably similar to the cost of determining self-consistent embedding charges for the binary interaction method, for example. Both of those methods are more expensive than EE-PA, which uses simple fixed B3LYP/6-31G* point charges. While the cost of computing the distributed polarizabilities to evaluate the force field is nontrivial, there are only a linear number of monomers, versus a locally quadratic number of two-body interactions (assuming local truncations are applied, though that is not done here). Therefore, the cost of evaluating the QM two-body interactions will dominate in larger systems.

Implementing software routines to parametrize the AIFF requires more effort than does the implementation of the other methods. Nuclear gradients of the AIFF require derivatives of the multipole moments and polarizabilities, which are also more complicated to implement. On the other hand, the use of point charges in EE-PA and the self-consistent binary interaction method introduces additional computational complexity. The point charges cause the gradient of each one-body and two-body term to depend on all three N_{atoms} coordinates of the entire system, instead of depending only on the coordinates of the atoms of a particular monomer or dimer. The implementation and efficiency of HMBI/AIFF nuclear gradients will be addressed in a future study. All of the methods examined here are very promising, and further investigation is needed to understand their strengths and weaknesses better.

The errors in HMBI can be attributed primarily to (1) the limitations of the classical multipolar description of induction effects and (2) the neglect of many-body dispersion and

exchange−repulsion effects. Describing these latter effects accurately and inexpensively remains a challenge. The largest outstanding difficulty in the former is the need to dampen short-range multipolar interactions. In our tests, we found it necessary to treat this damping factor as an empirical adjustable parameter. On the other hand, optimizing this parameter is relatively straightforward compared to the more general problem of force-field parametrization, and it can in principle be done by running several small cluster benchmark calculations to optimize the damping factor(s) before proceeding with more elaborate studies, just as we did in the glycine−water cluster example.

The procedure used here to compute the distributed polarizabilities can be extended readily to monomers containing a few dozen atoms. Furthermore, the computational cost of evaluating the multitude of quantum mechanical two-body interaction energies will typically dominate over the cost of constructing the force field, at least for systems containing many interacting molecules. Given its low cost and good accuracy, HMBI holds much promise for studying a wide range of interesting condensed-phase chemical systems, and we plan to report on such applications in the near future.

Acknowledgment. The authors thank Profs. Anthony Stone and Anthony Misquitta for sharing their CamCasp and Orient software packages and for providing assistance in using them. They also thank Prof. So Hirata for sharing the geometries that were used in Test Sets D and H. This research was supported by start-up funds and a Regents Faculty Fellowship (Beran) from the University of California at Riverside.

Supporting Information Available: Tables containing the RI-MP2 binding energies and many-body contributions for each method. Geometries for Test Sets B, C, F, and G. This material is available free of charge via the Internet at http://pubs.acs.org.

References

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

(2) Podeszwa, R.; Szalewicz, K. *J. Chem. Phys.* **2007**, *126*, 194101.

(3) Kitaura, K.; Sawai, T.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *312*, 319–324.

(4) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701–706.

(5) Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. A* **2007**, *111*, 6904–6914.

(6) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293–307.

(7) Gordon, M. S.; Slipchenko, L. V.; Li, H.; Jensen, J. H. *Annu. Rep. Comput. Chem.* **2007**, *3*, 177.

(8) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 46–53.

(9) Hirata, S. *J. Chem. Phys.* **2008**, *129*, 204104.

(10) Hirata, S.; Valiev, M.; Dupuis, M.; Xantheas, S. S.; Sugiki, S.; Sekino, H. *Mol. Phys.* **2005**, *103*, 2255–2265.

(11) Kamiya, M.; Hirata, S.; Valiev, M. *J. Chem. Phys.* **2008**, *128*, 074103.

(12) Dykstra, C. E. *J. Am. Chem. Soc.* **1989**, *111*, 6168–6174.

(13) Ganourgakis, G. S.; Xantheas, S. S. *J. Phys. Chem. A* **2006**, *110*, 4100–4106.

(14) Ganourgakis, G. S.; Xantheas, S. S. *J. Chem. Phys.* **2008**, *128*, 074506.

(15) Cisneros, G. A.; Darden, T. A.; Gresh, N.; Reinhardt, P.; Parisel, O.; Pilme, J.; Piquemal, J.-P. Design of next generation polarizable force fields from ab initio computations: Beyond point charges. In *Multi-scale Quantum Models for Biocatalysis: Modern Techniques and Applications*; Springer-Verlag: London, 2009; pp 137−172.

(16) Bukowski, R.; Szalewicz, K.; Groenenboom, G. C.; van der Avoird, A. *Science* **2007**, *315*, 1249.

(17) Bukowski, R.; Szalewicz, K.; Groenenboom, G. C.; van der Avoird, A. *J. Chem. Phys.* **2008**, *128*, 094314.

(18) Wang, Y.; Shepler, B. C.; Braams, B. J.; Bowman, J. M. *J. Chem. Phys.* **2009**, *131*, 054511.

(19) Xantheas, S. S. *J. Chem. Phys.* **1994**, *100*, 7523–7534.

(20) Tauer, T. P.; Sherrill, C. D. *J. Phys. Chem. A* **2005**, *109*, 10475–10478.

(21) Christie, R. A.; Jordan, K. D. *Struct. Bonding (Berlin)* **2005**, *116*, 27–41.

(22) Pedulla, J. M.; Kim, K.; Jordan, K. D. *Chem. Phys. Lett.* **2006**, *291*, 78–84.

(23) Tschumper, G. S. *Chem. Phys. Lett.* **2006**, *427*, 185–191.

(24) Elsohly, A. M.; Shaw, C. L.; Guice, M. E.; Smith, B. D.; Tschumper, G. S. *Mol. Phys.* **2007**, *105*, 2777–2782.

(25) Paulus, B.; Rosciszewski, K.; Gaston, N.; Schwerdtfeger, P.; Stoll, H. *Phys. Rev. B* **2004**, *70*, 165106.

(26) Hermann, A.; Schwerdtfeger, P. *Phys. Rev. Lett.* **2008**, *101*, 183005.

(27) Steiner, T. *Angew. Chem., Int. Ed.* **2002**, *41*, 48–76.

(28) Beran, G. J. O. *J. Chem. Phys.* **2009**, *130*, 164115.

(29) Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933–5947.

(30) Tabacchi, G.; Mundy, C. J.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **2002**, *117*, 1416–1433.

(31) Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A. *J. Chem. Phys.* **2004**, *120*, 10896–10913.

(32) Masia, M. *J. Chem. Phys.* **2008**, *128*, 184107.

(33) Bukowski, R.; Szalewicz, K.; Groenenboom, G. C.; van der Avoird, A. *Science* **2007**, *315*, 1249–1252.

(34) Stone, A. J.; Misquitta, A. J. *Int. Rev. Phys. Chem.* **2007**, *26*, 193–222.

(35) Popelier, P. L. A.; Rafat, M. *Chem. Phys. Lett.* **2003**, *376*, 148–153.

(36) Rafat, M.; Popelier, P. L. A. *J. Chem. Phys.* **2005**, *123*, 204103.

(37) Darley, M. G.; Handley, C. M.; Popelier, P. L. A. *J. Chem. Theory Comput.* **2008**, *4*, 1435–1448.

(38) Plattner, N.; Bandi, T.; Doll, J. D.; Freeman, D. L.; Meuwly, M. *Mol. Phys.* **2008**, *106*, 1675–1684.

(39) Mas, E. M.; Bukowski, R.; Szalewicz, K. *J. Chem. Phys.* **2003**, *118*, 4404–4413.

(40) Hodges, M. P.; Stone, A. J.; Xantheas, S. S. *J. Phys. Chem. A* **1997**, *101*, 9163–9168.

(41) Milet, A.; Moszynski, R.; Wormer, P. E. S.; van der Avoird, A. *J. Phys. Chem. A* **1999**, *103*, 6811–6819.

(42) Halgren, T. A.; Damm, W. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236–242.

(43) Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27–85.

(44) Warshel, A.; Kato, M.; Pisliakov, A. V. *J. Chem. Theory Comput.* **2007**, *3*, 2034–2045.

(45) Stone, A. J. *The Theory of Intermolecular Forces*; Clarendon Press: Oxford, 2002.

(46) Misquitta, A. J.; Stone, A. J. *J. Chem. Theory Comput.* **2008**, *4*, 7–18.

(47) Misquitta, A. J.; Stone, A. J.; Price, S. L. *J. Chem. Theory Comput.* **2008**, *4*, 19–32.

(48) Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233–239.

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

(50) Stone, A. J. *J. Chem. Theory Comput.* **2005**, *1*, 1128–1132.

(51) Misquitta, A. J.; Welch, G. W. A.; Stone, A. J.; Price, S. L. *Chem. Phys. Lett.* **2008**, *456*, 105–109.

(52) Thole, B. T. *Chem. Phys.* **1981**, *59*, 341–350.

(53) Miller, K. J. *J. Am. Chem. Soc.* **1990**, *112*, 8543–8551.

(54) LeSueur, C. R.; Stone, A. J. *Mol. Phys.* **1993**, *78*, 1267–1291.

(55) Piquemal, J.-P.; Perera, L.; Cisneros, G. A.; Ren, P.; Pedersen, L. G.; Darden, T. A. *J. Chem. Phys.* **2006**, *125*, 054511.

(56) Masia, M.; Probst, M.; Rey, R. *Chem. Phys. Lett.* **2006**, *420*, 267–270.

(57) Slipchenko, L. V.; Gordon, M. S. *Mol. Phys.* **2009**, *107*, 999–1016.

(58) Tang, K. T.; Toennies, J. P. *J. Chem. Phys.* **1984**, *80*, 3726–3741.

(59) Dunlap, B. I. *J. Chem. Phys.* **1983**, *78*, 3140–3142.

(60) Feyereisen, M. W.; Fitzgerald, G.; Komornicki, A. *Chem. Phys. Lett.* **1993**, *208*, 359–363.

(61) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283–289.

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

(63) Weigend, F.; Köhn, A.; Hättig, C. *J. Chem. Phys.* **2002**, *116*, 3175–3183.

(64) Liang, W. Z.; Head-Gordon, M. *J. Phys. Chem. A* **2004**, *108*, 3206–3210.

(65) Steele, R. P.; Distasio, R. A., Jr.; Shao, Y.; Kong, J.; Head-Gordon, M. *J. Chem. Phys.* **2006**, *125*, 074108.

(66) Steele, R. P.; Distasio, R. A., Jr.; Head-Gordon, M. *J. Chem. Theory Comput.* **2009**, *5*, 1560–1572.

(67) Chalasinski, G.; Szczesniak, M. M.; Kendall, R. A. *J. Chem. Phys.* **1994**, *101*, 8860–8869.

(68) Jurecka, P.; Sponer, J.; Cerný, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.

(69) Podeszwa, R. *J. Phys. Chem. A* **2008**, *112*, 8884–8885.

(70) Shao, Y.; et al. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172–3191.

(71) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 1342–1348.

(72) Distasio, R. A., Jr.; Jung, Y.; Head-Gordon, M. *J. Chem. Theory Comput.* **2005**, *1*, 862–876.

(73) Ponder, J. W. TINKER v4.2, 2004, http://dasher.wustl.edu/tinker/, accessed Jan 23, 2008.

(74) Misquitta A. J. Stone, A. J. CamCasp v5.2 (2008), http://www-stone.ch.cam.ac.uk/programs.html#CamCASP, accessed Oct 16, 2008.

(75) DALTON, a molecular electronic structure program, Release 2.0 (2005), http://www.kjemi.uio.no/software/dalton/dalton.html. Accessed October 16, 2008.

(76) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158.

(77) Joshipura, K. N.; Gangopadhyay, S.; Limbachiya, C. G.; Vinodkumar, M. *J. Phys.: Conf. Ser.* **2007**, *80*, 012008.

(78) Faubel, M.; Steiner, B.; Toennies, J. P. *J. Electron Spectrosc. Relat. Phenom.* **1998**, *95*, 159–169.

(79) Bruna, P. J.; Grein, F. *J. Chem. Phys. A* **2006**, *110*, 4906–4917.

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

(81) Tozer, D. J. *J. Chem. Phys.* **2000**, *112*, 3507–3515.

(82) Dullweber, A. Engkvist, O. Fraschini, E. Hodges, M. P. Meredith, A. Popelier, P. L. A. Wales, D. J. Stone, A. J. Orient v4.6 (2006) http://www-stone.ch.cam.ac.uk/programs.html#Orient, accessed Oct 16, 2008.

(83) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–566.

(84) Laasonen, K.; Parrinello, M.; Car, R.; Lee, C.; Vanderbilt, D. *Chem. Phys. Lett.* **1993**, *207*, 208–213.

(85) Kim, J.; Majumdar, D.; Lee, H. M.; Kim, K. S. *J. Chem. Phys.* **1999**, *110*, 9128–9134.

(86) Kozmutza, C.; Kryachko, E. S.; Tfirst, E. *J. Mol. Struct. (THEOCHEM)* **2000**, *501*, 435–444.

(87) Day, P.; Pachter, R.; Gordon, M. S.; Merrill, G. N. *J. Chem. Phys.* **2000**, *112*, 2063–2073.

(88) Upadhyay, D. M.; Shukla, M. K.; Mishra, P. C. *Int. J. Quantum Chem.* **2001**, *81*, 90–104.

(89) Bulusu, S.; Yoo, S.; Apra, E.; Xantheas, S.; Zeng, X. C. *J. Phys. Chem. A* **2006**, *110*, 11781–11784.

(90) Olson, R. M.; Bentz, J. L.; Kemdall, R. A.; Schmidt, M. W.; Gordon, M. S. *J. Chem. Theory Comput.* **2007**, *3*, 1312–1328.

(91) Maheshwary, S.; Patel, N.; Sathyamurthy, N.; Kulkarni, A. D.; Gadre, S. R. *J. Phys. Chem. A* **2001**, *105*, 10525–10537.

(92) Maheshwary, S.; Patel, N.; Sathyamurthy, N.; Kulkarni, A. D.; Gadre, S. R. *J. Phys. Chem. A* **2001**, *105*, 10525–10537.

(93) Geometries obtained from the Cambridge Cluster Database. Wales, D. J.; Doye, J. P. K.; Dullweber, A.; Hodges, M. P.; Naumkin F. Y.; Calvo, F.; Hernandez-Rojas, J.; Middleton, T. F. http://www-wales.ch.cam.ac.uk/CCD.html, accessed Aug 21, 2008.

(94) Ohno, K.; Okimura, M.; Akai, N.; Katsumoto, Y. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3005.

(95) Dahlke, E. E.; Olson, R. M.; Leverentz, H. R.; Truhlar, D. G. *J. Phys. Chem. A* **2008**, *112*, 3976–3984.

(96) Bates, D. M.; Tschumper, G. S. *J. Phys. Chem. A* **2009**, *113*, 3555–3559.

(97) Valiron, P.; Mayer, I. *Chem. Phys. Lett.* **1997**, *275*, 46–55.

(98) Sadlej, A. J. *Collect. Czech. Chem. Commun.* **1988**, *53*, 1995–2016.

(99) Sadlej, A. J. *Theor. Chim. Acta* **1991**, *79*, 123–140.

(100) Suhai, S. *J. Chem. Phys.* **1995**, *103*, 7030–7039.

(101) Bende, A.; Vibok, A.; Halasz, G. J.; Suhai, S. *Int. J. Quantum Chem.* **2001**, *84*, 617–622.

(102) Fogarasi, G.; Szalay, P. G. *J. Phys. Chem. A* **1997**, *101*, 1400–1408.

(103) Bellissent-Funel, M.-C.; Nasr, S.; Bosio, L. *J. Chem. Phys.* **1997**, *106*, 7913–7919.

(104) Nasr, S.; Bosio, L. *J. Chem. Phys.* **1998**, *108*, 2297–2301.

(105) Cabaleiro-Lago, E. M.; Rios, M. A. *J. Chem. Phys.* **1999**, *110*, 6782–6791.

(106) Tsuchida, E. *J. Chem. Phys.* **2004**, *121*, 4740–4746.

(107) Mardyukov, A.; Sanchez-Garcia, E.; Rodziewicz, P.; Doltsinis, N. L.; Sander, W. *J. Phys. Chem. A* **2007**, *111*, 10552–10561.

(108) Pine, A. S.; Lafferty, W. L. *J. Chem. Phys.* **1983**, *78*, 2154–2162.

(109) Andrews, L.; Bondybev, V. E.; English, J. H. *J. Chem. Phys.* **1984**, *81*, 3452–3457.

(110) Micheal, D. W.; Lisy, J. M. *J. Chem. Phys.* **1986**, *85*, 2528–2538.

(111) Guedes, R. C.; do Couto, P. C.; Costa Cabral, B. J. *J. Chem. Phys.* **2003**, *118*, 1272–1281.

(112) Hendricks, S. B.; Wulf, O.; Hilbert, G. E.; Liddell, U. *J. Am. Chem. Soc.* **1936**, *58*, 1991–1996.

(113) Quack, M.; Schmitt, U.; Suhm, M. A. *Chem. Phys. Lett.* **1993**, *208*, 446–452.

(114) del Bene, J. E.; Person, W. B.; Szczepaniak, K. *J. Phys. Chem.* **1995**, *99*, 10705–10707.

(115) Hirata, S.; Iwata, S. *J. Phys. Chem. A* **1998**, *102*, 8426–8436.

(116) Hodges, M. P.; Stone, A. J.; Lago, E. C. *J. Phys. Chem. A* **1998**, *102*, 2455–2465.

(117) Halkier, A.; Klopper, W.; Helgaker, T.; Jørgensen, P.; Taylor, P. R. *J. Chem. Phys.* **1999**, *111*, 9157–9167.

(118) Rankin, K. N.; Boyd, R. J. *J. Comput. Chem.* **2001**, *22*, 1590–1597.

(119) Rincon, L.; Almeida, R.; Garcia-Aldea, D.; y Riega, H. D. *J. Chem. Phys.* **2001**, *114*, 5552–5561.

(120) Raynaud, C.; Maron, L.; Jolibois, F.; Daudey, J.-P.; Esteves, P. M.; Ramirez-Solis, A. *Chem. Phys. Lett.* **2005**, *414*, 161–165.

(121) Aviles, M. W.; Gray, P. T.; Curotto, E. *J. Chem. Phys.* **2006**, *124*, 174305.

(122) Murdachaew, G.; Szalewicz, K. *Faraday Discuss.* **2001**, *118*, 121–142.

CT900545V