JCTC Journal of Chemical Theory and Computation

Article

Cite This: J. Chem. Theory Comput. 2020, 16, 475−487

pubs.acs.org/JCTC

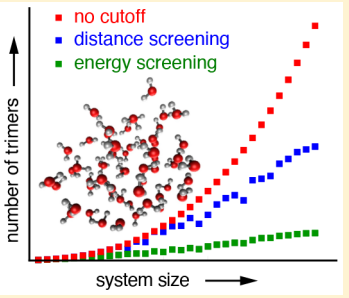
Energy-Screened Many-Body Expansion: A Practical Yet Accurate Fragmentation Method for Quantum Chemistry

Kuan-Yu Liu† and John M. Herbert*

Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, United States

Supporting Information

ABSTRACT: We introduce an implementation of the truncated many-body expansion, MBE(n), in which the n-body corrections are screened using the effective fragment potential force field, and only those that exceed a specified energy threshold are computed at a quantum-mechanical level of theory. This energy-screened MBE(n) approach is tested at the n = 3 level for a sequence of water clusters, $(H_2O)_{N=6-34}$. A threshold of 0.25 kJ/mol eliminates more than 80% of the subsystem electronic structure calculations and is even more efficacious in that respect than is distance-based screening. Even so, the energy-screened MBE(3) method is faithful to a full-system quantum chemistry calculation to within 1−2 kJ/mol/monomer, even in good quality basis sets such as aug-cc-pVTZ. These errors can be reduced by means of a two-layer approach that involves a Hartree−Fock calculation for the entire cluster. Such a correction proves to be necessary in order to obtain accurate relative energies for conformational isomers of $(H_2O)_{20}$, but the cost of a full-system Hartree−Fock calculation remains smaller than the cost of three-body subsystem calculations at correlated levels of theory. At the level of second-order Møller−Plesset perturbation theory (MP2), a screened MBE(3) calculation plus a full-system Hartree−Fock calculation is less expensive than a full-system MP2 calculation starting at N = 12 water molecules. This is true even if all MBE(3) subsystem calculations are performed on a single 40-core compute node, i.e., without significant parallelization. Energy-screened MBE(n) thus provides a fragment-based method that is accurate, stable in large basis sets, and low in cost, even when the latter is measured in aggregate computer time.



1. INTRODUCTION

Fragmentation methods[1−6] are an increasingly popular means to circumvent the steeply nonlinear computational scaling of *ab initio* quantum chemistry, by decomposing a large (super)system into tractable subsystems, then approximating the energy or other properties of the supersystem using electronic structure calculations performed on the subsystems. A formal scaling of $O(N^p)$, where N measures the supersystem size and p is characteristic of a particular quantum-chemical model, is thereby reduced to $O(n^p)$, where n measures the size of the largest fragment necessary to obtain accurate results. If the fragment size (n) remains fixed as the system size (N) increases, then the scaling bottleneck can be defeated by means of distributed computing.

The reduced scaling, however, incurs a prefactor proportional to the number of subsystems, and in practice the total computer time (summed across all subsystem calculations) is sometimes considerably larger than the cost of the supersystem calculation that fragmentation aims to avoid.[1,7−9] As system size grows, the fragment-based approach must eventually become the more economical one, and the crossover point is reached more quickly for correlated wave function methods ($p \geq 4$) as compared to density-functional calculations ($p \leq 3$). Furthermore, there are inherent advantages to the subsystem-based approach in terms of parallelizability and checkpointing, both of which are relatively straightforward in the context of fragment-based methods. Despite these advantages, and notwithstanding the "forced-march towards exascale computing" that characterizes some areas of computational science at the moment, we feel that it is a mistake to focus strictly on time-to-completion or "wall time". An inefficient-but-scalable calculation still occupies processors that are then unavailable for other work, which is especially relevant in the context of shared computing facilities. Such a calculation also consumes power, whereas modern processor designs often use very little power when idle.

Most fragment-based quantum chemistry methods are based at some level on the idea of a many-body expansion (MBE) or else generalizations thereof.[1,10−12] Only the traditional MBE is considered here:

$$E = \sum_{I=1}^{N} E_I + \sum_{I=1}^{N} \sum_{J>I} \Delta E_{IJ} + \sum_{I=1}^{N} \sum_{J>I} \sum_{K>J} \Delta E_{IJK} + \cdots \quad (1)$$

The energy of a supersystem composed of N fragments is thus decomposed into monomer energies E_I along with two-body corrections

$$\Delta E_{IJ} = E_{IJ} - E_I - E_J \quad (2)$$

three-body corrections

Received: November 2, 2019
Published: November 25, 2019

DOI: 10.1021/acs.jctc.9b01095
J. Chem. Theory Comput. 2020, 16, 475−487

$$\Delta E_{IJK} = E_{IJK} - \Delta E_{IJ} - \Delta E_{IK} - \Delta E_{JK} - E_I - E_J - E_K \quad (3)$$

etc. The idea is to truncate eq 1 at relatively low n-body interactions, and we will use the notation MBE(n) to refer to such a truncated expansion. This affords a method whose cost contains a multiplicative prefactor of

$$\binom{N}{n} = \frac{N(N-1)\cdots(N-n+1)}{n!} \quad (4)$$

that represents the number of distinct n-body subsystems. Note that the numerator of eq 4 grows like N^n with system size, for a fixed level of approximation, so the formal scaling of the MBE(n) approach is $O(N^n)$ with system size, in the absence of any attempt to screen or otherwise discard n-body subsystems.

Calculations for small water clusters indicate that electron correlation effects are nearly pairwise-additive;[13−15] nevertheless, four- and five-body terms contribute 1−2 kcal/mol to the total interaction energies of clusters ranging from $(H_2O)_6$ to $(H_2O)_{16}$.[16−18] In larger water clusters, our own work suggests that four-body terms are necessary to obtain (or even approach) chemical accuracy.[7,19,20] This largely remains true even when electrostatic embedding is employed in an effort to capture higher-order polarization effects.[7]

Use of a four-body expansion is often prohibitively expensive in practice, however. For calculations at the level of second-order Møller−Plesset perturbation theory (MP2, for which p = 5), the MBE(4) approximation applied to $(H_2O)_{20}$ is actually more expensive than the MP2 calculation on the full system, when the cost is measured in aggregate computer time.[7] The large number of four-body terms also engenders serious loss-of-precision problems,[19,21] necessitating the use of very tight thresholds.[7,21] For these reasons, fragmentation is not the proverbial "free lunch" that it is sometimes made out to be; it is more like a long-term investment strategy, with significant up-front costs but a potentially sizable payoff as $N \to \infty$.

Distance-based thresholding represents an obvious way to reduce the combinatorial prefactor associated with MBE(n) and has been used with some success in water clusters.[20,22,23] For systems where the fragments are larger and more diverse than H_2O, however, the efficiency and/or the accuracy of distance-based screening is likely to suffer. Fortunately, the essentially classical nature of higher-order n-body interactions means that classical polarization formulas can be used to predict the three- and four-body interaction energies with reasonable fidelity.[24] The success of the "hybrid many-body interaction" (HMBI) scheme developed by Beran and co-workers[25−31] is further testament to the validity of classical approximations for higher-order induction. Within HMBI, one- and two-body interactions are computed from electronic structure theory but higher-order terms are obtained from polarizable force fields.

The success of these classical polarization approaches suggests that energy-based screening, using an appropriate classical model, may be a useful alternative to distance-based screening. Even for water clusters, we have previously demonstrated that the energy-based approach is more efficient than distance-based screening, without loss of accuracy.[20] These were proof-of-concept results, in the sense that we precomputed all of the many-body interactions at a quantum-mechanical (QM) level of theory and then reverse-engineered an energy-screening threshold to preserve the accuracy of the MBE(n) approximation.

Here, we present a practical implementation of energy-based screening for MBE(n) using the effective fragment potential (EFP).[32−34] The EFP method is attractive in this capacity because it is a classical force field and can therefore be evaluated at essentially zero cost (in comparison to QM calculations), yet it is derived in an automated way from QM calculations. As such, EFP can be parametrized for the sort of large, complex systems for which one might want to apply quantum chemistry, and the energy-screened MBE(n) approach need not be limited to systems for which force fields have already been developed. Calculations on non-covalent dimers suggest that the accuracy of EFP is similar to that of MP2,[35] which proves to be adequate for screening purposes, although the calculations reported here also make it clear that EFP is no substitute for MP2.

2. THEORY

2.1. Many-Body Expansion. The MBE in eq 1 will be truncated at three-body terms in the present work. Non-redundant formulas for the two- and three-body energy corrections can be found in ref 21. In our experience, the cost differential between MBE(3) and MBE(4) is rather catastrophic, both in terms of growth in the number of subsystems and in terms of issues with finite precision, which further increases the cost by necessitating the use of tight thresholds.[7,19,21] For that reason, we aim to see whether a composite three-body approach, which combines MBE(3) with EFP, can be made both accurate and tractable.

Importantly, the subsystem calculations used to obtain the dimer energies (E_{IJ} in eq 2) and trimer energies (E_{IJK} in eq 3) each include a full self-consistent field (SCF) calculation. This should be contrasted with the approach taken in the fragment molecular orbital (FMO) method,[2,36−39] which is probably the most widely used fragment-based quantum chemistry method, due to its implementation in the GAMESS code.[36,39] The FMO approach with n-body interactions (FMOn) uses the energy expression in eq 1, but unlike the method that we call MBE(n), in FMOn only the one-body wave functions are iterated to self-consistency. These iterations are performed in the presence of a classical electrostatic embedding field that represents the effects of the other fragments. Because the self-consistent iterations lack interfragment exchange interactions, however, the FMO approximation affords large errors or even convergence failure when it is combined with large basis sets, especially those that contain diffuse functions.[40,41] This severely limits the applicability of FMO in the context of correlated wave function methods. Even in density functional theory (DFT), basis sets such as aug-cc-pVTZ are recommended for some modern functionals including new-developed B97-based functionals,[42−44] which are among the most accurate DFT methods presently available for a variety of properties.[45] Self-consistency also means that analytic gradients of MBE(n) are much more straightforward, even when self-consistent charge embedding is used,[46] as compared to the gradient for FMOn.[47]

2.2. Energy-Based Screening. We adopt the parameter-free version of EFP (sometimes called "EFP2"),[34] whose energy expression is

$$E^{\mathrm{EFP}} = E^{\mathrm{elst}} + E^{\mathrm{pol}} + E^{\mathrm{disp}} + E^{\mathrm{exrep}} \quad (5)$$

DOI: 10.1021/acs.jctc.9b01095
J. Chem. Theory Comput. 2020, 16, 475−487

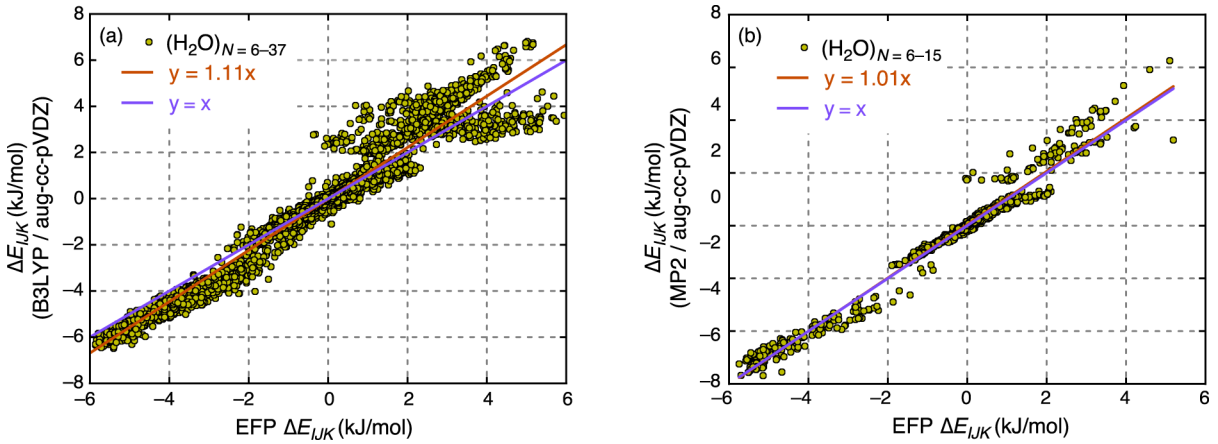


Figure 1. Correlation plot for three-body corrections ΔE_{IJK} in water clusters, computed at (a) the B3LYP/aDZ level of theory and (b) the MP2/aDZ level of theory, comparing against EFP results in either case. The MP2 data set is smaller (using clusters only up to $N = 15$, versus $N = 37$ for B3LYP) because evaluation of the full set of three-body corrections is considerably more expensive than the supersystem calculation at this level of theory. The orange line in either plot represents a linear fit to the data.

Here, E^{elst} is the classical electrostatic (permanent multipole) energy and E^{pol} is the dipole polarization energy. The dispersion energy E^{disp} is truncated at leading order and includes only the R^{-6} term. The exchange-repulsion energy E^{exrep} is based on the Heitler–London formula.[32] The charge-transfer contribution is omitted in this work. EFP calculations are performed using an interface between the Q-Chem program[48] and the LibEFP library.[49] Note that EFP is designed for rigid monomers whose geometries are fixed at the geometry used to parametrize the model. LibEFP uses two vectors to align the target molecule to this reference molecule. This works perfectly well when the monomers are H_2O, as in the present work, but in preliminary testing for other systems we found that this procedure could lead to large deviations for small but nonplanar monomers such as NH_3. In the interest of later extending our scheme to more general systems with minimal modification, we therefore use an optimal rotational matrix in order to minimize the deviation between reference and target molecules.[50]

To reduce the number of QM subsystems, we precompute three-body interactions at the EFP level before performing any QM calculations. Three-body EFP corrections are given by

$$\Delta E_{IJK}^{EFP} = E_{IJK}^{EFP} - E_{IJ}^{EFP} - E_{IK}^{EFP} - E_{JK}^{EFP} \quad (6)$$

Note that $\Delta E_{IJ}^{EFP} = E_{IJ}^{EFP}$ because EFP is an intermolecular (rigid-monomer) force field; that is, its monomer energies are set to zero. Furthermore, the only many-body effects captured by ΔE_{IJK}^{EFP} are induction effects, since EFP is pairwise-additive in all other energy components. However, many-body contributions to other energy components are generally quite small,[1] and the three- and four-body energies computed using classical many-body induction formulas agree quite well with the three- and four-body corrections (ΔE_{IJK} and ΔE_{IJKL}) obtained from QM calculations.[24] Here, a QM calculation is performed on dimer IJ only if $|\Delta E_{IJ}^{EFP}|$ exceeds a specified energy threshold and on trimer IJK only if $|\Delta E_{IJK}^{EFP}|$ exceeds the threshold.

Figure 1 plots three-body corrections ΔE_{IJK} for a large set of water clusters, computed at two different QM levels of theory (B3LYP/aDZ and MP2/aDZ) and compared to EFP estimates ΔE_{IJK}^{EFP}. (Here and elsewhere, "aDZ" is an abbreviation for the aug-cc-pVDZ basis set.) The correlation with EFP is excellent at either level of QM theory, and the data cluster around the $y = x$ line in similar ways for B3LYP and MP2, although we present more data at the B3LYP level, for reasons discussed below. These similarities are not altogether surprising given that the performance of EFP for noncovalent interactions is generally similar to that of MP2,[35] and the performance of B3LYP for water clusters is also roughly similar to that of MP2.[51−53] The good correspondence between QM and classical values of ΔE_{IJK} justifies the use of EFP to screen the QM values at either level of theory.

In these tests, we have used a rigid-monomer parametrization of EFP, neglecting any changes that would be obtained if we were to reparameterize this force field for each individual monomer geometry. Various methods to obtain a flexible EFP, including full reparameterization, were tested in ref 54 for cases involving strong hydrogen bonding (e.g., formic acid dimer) where the monomer geometries do change upon complexation. These changes are likely to be quite minor for the neat water clusters examined herein. For the challenging problem of relative conformational energies (Section 3.3), we use a set of $(H_2O)_{20}$ isomers with strictly rigid monomer geometries so that the issue of changes in EFP due to changes in the monomer geometry does not complicate our evaluation.

Note that the $(H_2O)_N$ cluster data set that is used for the MP2 calculations (Figure 1b) ranges up to $N = 15$, whereas for B3LYP we include data up to $N = 37$ (Figure 1a). For the larger clusters in the data set, a complete MBE(3) calculation at the MP2/aDZ level of theory requires significantly greater computer time than the calculation it aims to approximate, namely, MP2/aDZ applied to the full cluster. This aspect of MBE(n) is seldom discussed, although we have raised this issue before.[1,7] This observation underscores the fact that automated screening methods are mandatory in order to turn MBE(n) into a usable (and useful) method.

In view of this, it is worth noting that the combination of EFP with fragment-based quantum chemistry has been attempted previously, in the form of the "effective fragment molecular orbital" (EFMO) method.[55−60] EFMO was originally introduced as a distance-based screening procedure for FMO2, wherein the dimer energy E_{IJ} is replaced by its EFP approximation when fragments I and J are sufficiently well-

DOI: 10.1021/acs.jctc.9b01095
J. Chem. Theory Comput. 2020, 16, 475−487

separated. (In the EFMO approach, new EFP parameters are generated for each individual monomer geometry.) The EFMO approximation is problematic, however, in part because FMO2 is not an accurate method in the first place. It is also unclear that EFMO is genuinely a systematic or robust approximation to FMO2. Both of these concerns are borne out, e.g., from the water cluster data presented in ref 58. In recent work, the accuracy of EFMO has been discussed as an entity unto itself.[61] In absolute terms, however, its accuracy remains rather poor.[1]

2.3. ONIOM-Type Formalism. To recover higher-order n-body interaction energies ($n \geq 4$), which are dominated by classical induction, we adopt an ONIOM-type formalism[62] that involves some type of calculation on the entire supersystem. The idea to combine ONIOM with a fragment-based method was originally introduced by Tschumper and co-workers,[63−71] and later by Raghavachari and co-workers.[72−81] The two-layer "molecules-in-molecules" (MIM2) approach,[72,73] equivalent to Tschumper's "n-body:many-body" technique,[70] can be expressed in ONIOM-style notation as

$$E_{\text{high}}(\text{super}) \approx E_{\text{high}}(\text{MBE}) - E_{\text{low}}(\text{MBE}) + E_{\text{low}}(\text{super}) \tag{7}$$

Here, "high" and "low" indicate different levels of theory, with a lower-level calculation performed on the entire supersystem [E_{low}(super)]. The relationship between eq 7 and the original ONIOM method can be understood schematically using Figure 2. Alternatively, by rewriting eq 7 as

$$E_{\text{high}}(\text{super}) \approx E_{\text{high}}(\text{MBE}) + \delta^{\text{frag}} \tag{8}$$

it becomes clear that

$$\delta^{\text{frag}} = E_{\text{low}}(\text{super}) - E_{\text{low}}(\text{MBE}) \tag{9}$$

is a correction for errors introduced by fragmentation.

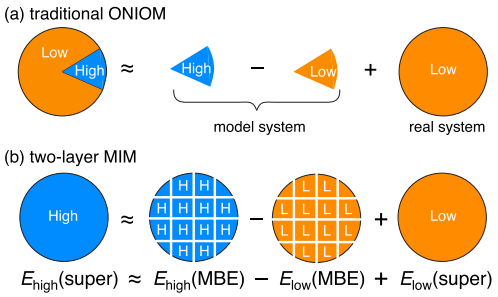


Figure 2. Schematic depictions of the (a) ONIOM and (b) MIM2 approaches. In the traditional ONIOM method,[62] a high level of theory (illustrated in blue) is applied to a small portion of a large system, often called the "model system", while a low level of theory (in orange) is applied to the entire (or "real") system. The low-level theory is applied also to the model system and subtracted out, to avoid double counting. The molecules-in-molecules (MIM) approach[72] is an adaptation in which the high level of theory is applied to the entire system by means of a fragmentation approximation; see eq 7.

The method introduced in this work combines eq 7 with an energy-screened MBE(3) approximation. The high-level E_{high}(MBE) term is first evaluated using EFP to screen an MBE(3) calculation. Energetically important subsystems (as determined using EFP) are replaced by QM calculations, but for below-threshold subsystems we simply retain the EFP energy, which then cancels out in the difference E_{high}(MBE) − E_{low}(MBE). Ideally, the supersystem correction E_{low}(super) would be applied using EFP for the low level of theory, as this calculation is essentially free in comparison to the cost of any QM calculations. However, we find that this approach does not always preserve good fidelity with respect to a high-level QM calculation on the entire supersystem. To improve the accuracy, we also examine a composite approach in which E_{high}(MBE) is evaluated using the EFP-screened MBE(3) approximation but the low-level supersystem calculation uses either Hartree−Fock theory or else a semiempirical method. Although this destroys the linear-scaling nature of the energy-screened MBE(3) approximation, it may yet represent an affordable approximation when the subsystem MBE(3) calculations are performed at a post-Hartree−Fock level of theory. This has been the approach pursued by the Raghavachari group,[73,79−81] where the supersystem calculation is either semiempirical or else a small-basis Hartree−Fock calculation. Similar ideas underlie many-body expansions for the correlation energy (only),[82] sometimes called the "method of increments".[83]

3. RESULTS AND DISCUSSION

3.1. Computational Details. In the first part of this work, we examine how energy-based cutoffs affect the accuracy of the MBE(3) approximation, for a sequence of water clusters $(H_2O)_N$ with N = 6−34.[84] There is one structure at each cluster size, representing the putative global minimum on the TIP4P potential energy surface and taken from ref 85. (This is the same sequence of clusters that we have used in previous examinations of the MBE,[7,19−21] and these clusters have rigid H_2O monomer geometries.) In the second part of this work, we examine relative energies of four different structural motifs of $(H_2O)_{20}$, taken originally from ref 86 and used also in our previous work on MBE(n) methods.[7,20,87]

For many of the convergence tests in Section 3.2, we use the affordable B3LYP/aDZ level of theory. Our goal is to use low-cost calculations to establish an appropriate value of the three-body energy-screening threshold (τ_{3B}), in the hope that if τ_{3B} is chosen conservatively then this value may prove to be transferrable to more accurate MBE(n) calculations at higher levels of theory. Note that the cost of unscreened MBE(n) calculations for medium-sized clusters is typically much larger than the cost of full-system quantum chemistry calculations, even for an $O(N^5)$ method such as MP2,[1,7] assuming that the supersystem calculations are performed using an efficient quantum chemistry code. (Inefficient or nonscalable implementations skew the comparison in favor of the fragment-based approach.[1]) The high cost of unscreened, supersystem MBE(n) calculations means that we cannot possibly reparameterize τ_{3B} at every distinct level of theory and basis set that we might hope to use in practical applications.

The ultimate proof that the threshold τ_{3B} is transferrable from low- to high-level MBE(3) calculations rests in the performance of those higher-level calculations. To that end, we consider (in Section 3.3) the very challenging problem of predicting relative energies of $(H_2O)_{20}$ isomers, for which we use MP2 calculations. We will also test the ωB97X-V functional[42] for the same application, as this functional performs very well for noncovalent interactions.[45]

All calculations use a single water monomer per fragment and were performed with a locally modified version of Q-

DOI: 10.1021/acs.jctc.9b01095
J. Chem. Theory Comput. 2020, 16, 475−487

Chem.[48] The SCF convergence threshold was set to $\tau_{SCF} = 10^{-7}$ a.u. and the integral screening threshold to $\tau_{ints} = 10^{-14}$ a.u.. These are "tight" convergence thresholds, as defined in previous work.[7] Both thresholds, but especially τ_{ints}, are tighter than typical default settings in many electronic structure programs (including Q-Chem), but the use of looser thresholds leads to precision problems in large clusters.[21] In contrast, the DFT quadrature grid is found to make little difference,[21] so the modest SG-1 grid[88] is used for all calculations.

3.2. Validity of the Energy Cutoff Scheme. We first assess the accuracy of the energy-based screening approximation by comparing screened versus unscreened MBE(3) results. We have previously observed that errors in fragment-based approximations are size-extensive,[7,19] and as such it is common to report errors on a per-monomer basis. Ouyang and Betterns[89] have suggested that the target accuracy for such methods should be $0.1 \times (3/2)k_B \times (298\ \mathrm{K}) = 0.4$ kJ/mol per monomer, which represents 10% of the available thermal energy per fragment at room temperature. The idea is that if the fragmentation approximation can be made faithful to the underlying quantum-chemical model to within this level of accuracy, then fragmentation could conceivably be used in an *ab initio* molecular dynamics simulation without concern that the fragmentation approximation skews the results. This turns out to be a rather stringent criterion, which is likely not met by most fragment-based methods in the literature.[1]

Figure 3 reports error statistics for the MBE(3) method as a function of the three-body screening threshold, τ_{3B}, aggregating

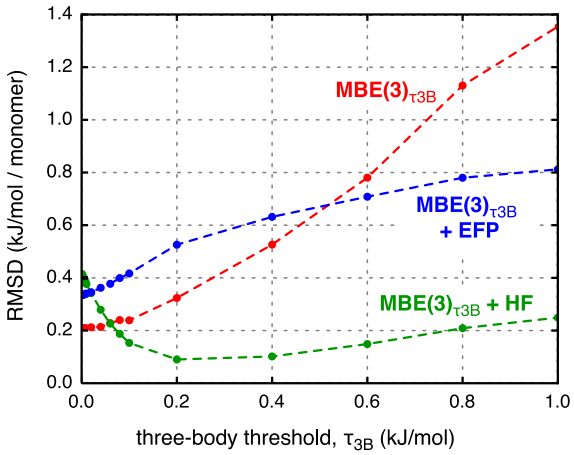


Figure 3. Errors (measured with respective to a supersystem calculation) in the energy-screened three-body approximation, $\mathrm{MBE}(3)_{\tau 3B}$, as applied to a sequence of water clusters $(H_2O)_{N=6-34}$. The error is expressed as the root-mean-square deviation (RMSD) for the entire data set. QM calculations were performed at the B3LYP/aDZ level, and the cutoff τ_{3B} was applied only to the three-body corrections, ΔE_{IJK}, retaining all of the two-body corrections. In the $\mathrm{MBE}(3)_{\tau 3B}$ + EFP and $\mathrm{MBE}(3)_{\tau 3B}$ + HF schemes, higher-order induction is recovered using an ONIOM-style formalism that requires a supersystem calculation at either the EFP level or else the HF/aDZ level.

statistics for the entire cluster sequence $(H_2O)_{N=6-34}$. We will use the notation $\mathrm{MBE}(3)_{\tau 3B}$ to indicate the energy-screened MBE(3) method, in which three-body corrections are simply neglected if $|\Delta E_{IJK}^{\mathrm{EFP}}| < \tau_{3B}$. (In the calculations reported in Figure 3, all two-body corrections are retained at the QM level.) These calculations establish a baseline, namely, that a threshold $\tau_{3B} \approx 0.3$ kJ/mol is required in order to meet the aforementioned dynamic accuracy criterion. This is very similar to the threshold of 0.25 kJ/mol that was suggested in previous work by Ouyang and Bettens,[24] albeit based on proof-of-concept calculations in which all of the QM interaction energies were computed *a priori* and then estimated based on classical polarization formulas. In what follows, we will adopt the slightly more conservative choice of $\tau_{3B} = 0.25$ kJ/mol from ref 24 for the purpose of screening the three-body interactions.

Figure 4 shows how the errors in the screened approximation $\mathrm{MBE}(3)_{\tau 3B}$ differ from errors in a complete

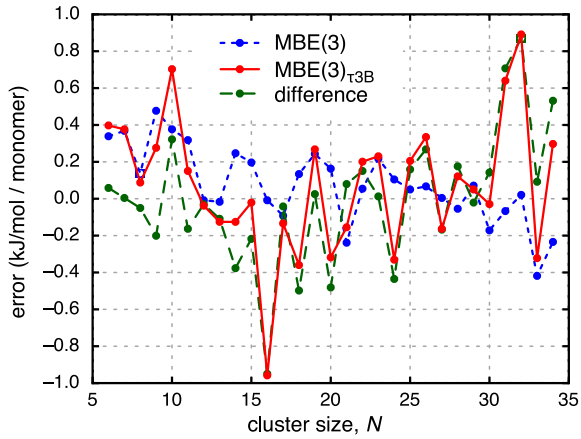


Figure 4. Signed errors per monomer for clusters $(H_2O)_{N=6-34}$, considering both a complete MBE(3) approximation (including all terms) and its three-body-screened variant, $\mathrm{MBE}(3)_{\tau 3B}$ (neglecting terms for which $|\Delta E_{IJK}^{\mathrm{EFP}}| < \tau_{3B}$, with $\tau_{3B} = 0.25$ kJ/mol/monomer). All calculations were performed at the B3LYP/aDZ level of theory, and the error is defined with respect to a supersystem calculation at that level of theory. The data in green represent the difference between the $\mathrm{MBE}(3)_{\tau 3B}$ error and the MBE(3) error.

MBE(3) approximation that includes all of the trimers. Both errors are judged with respect to a supersystem calculation at the same level of theory, which is B3LYP/aDZ in the present case, and are plotted as a function of cluster size N. For most of the individual $(H_2O)_N$ clusters, the absolute error introduced by energy screening is $\lesssim$0.4 kJ/mol/monomer, consistent with the aggregate error statistics in Figure 3 for the selected threshold of $\tau_{3B} = 0.25$ kJ/mol. However, several of the largest individual errors occur in the largest clusters. To understand whether this represents a random fluctuation or else the start of an upward trend, we performed calculations on two larger clusters, $(H_2O)_{79}$ and $(H_2O)_{140}$, obtaining $\mathrm{MBE}(3)_{\tau 3B}$ errors of 0.50 and −0.25 kJ/mol/monomer, respectively. These are of the same magnitude as what is observed for clusters with $N \leq 34$, suggesting that any hint of an upward trend in Figure 4 may be illusory.

Given that we are using EFP calculations to screen the n-body energy corrections, there is no reason to simply drop the three-body interactions when they fall below the threshold, as we have done in the $\mathrm{MBE}(3)_{\tau 3B}$ approach that has been described thus far. Instead, we can include them at the EFP level by means of the two-layer ONIOM-style scheme that is shown schematically in Figure 2b. Results from this composite approach are labeled $\mathrm{MBE}(3)_{\tau 3B}$ + EFP in Figure 3. When $\tau_{3B} \leq 0.5$ kJ/mol, this approach is actually slightly *less* accurate as compared to simply neglecting all of the small interactions. We regard this as a minor artifact of error cancellation; for more

DOI: 10.1021/acs.jctc.9b01095
J. Chem. Theory Comput. 2020, 16, 475−487

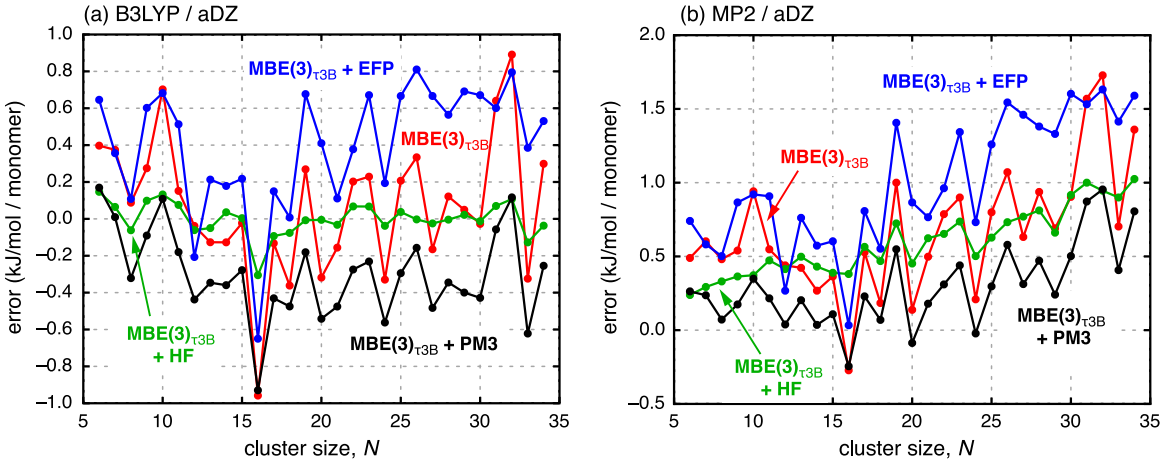


Figure 5. Signed errors per monomer for energy-screened MBE(3) approximations computed at the (a) B3LYP/aDZ or (b) MP2/aDZ level of theory. In either case, the energy cutoff is $\tau_{2B} = \tau_{3B} = 0.25$ kJ/mol and error is defined with respect to a supersystem at the specified level of quantum chemistry. Higher-order induction is recovered via two-layer methods labeled MBE(3)$_{\tau 3B}$ + X, where X = EFP, PM3, or HF/aDZ. These composite methods require a supersystem calculation at level X.

aggressive thresholds (larger values of τ_{3B}), the composite scheme is indeed more accurate than simply neglecting below-threshold values.

More accurate still is a two-layer method where higher-order induction is incorporated by means of a two-layer approach involving a supersystem calculation at the HF/aDZ level. This approach is labeled "MBE(3)$_{\tau 3B}$ + HF" in Figure 3, and its accuracy is within the dynamic accuracy criterion of 0.4 kJ/mol/monomer suggested by Ouyang and Bettens,[89] for values of τ_{3B} ranging all the way up to at least $\tau_{3B} = 1.0$ kJ/mol. The minimum error is obtained close to the value $\tau_{3B} = 0.25$ kJ/mol that was singled out above.

In Figure S1, we fix $\tau_{3B} = 0.25$ kJ/mol and plot the error in MBE(3)$_{\tau 3B}$ calculations as a function of an energy threshold τ_{2B} that is used to screen the two-body corrections. Denoting this approach as MBE(3)$_{\tau 2B,\tau 3B}$, we note that MBE(3)$_{\tau 2B,\tau 3B}$ + HF/aDZ affords better results than either MBE(3)$_{\tau 2B,\tau 3B}$ alone or MBE(3)$_{\tau 2B,\tau 3B}$ + EFP, which is unsurprising in view of the results for three-body screening. With $\tau_{2B} = 0.25$ kJ/mol, however, each of these MBE(3)$_{\tau 2B,\tau 3B}$-based approximations affords errors <0.5 kJ/mol/monomer. This suggests that some two-body screening is feasible without serious loss of accuracy, although this does not do much to reduce the cost of the systems examined here. As such, all other calculations reported in this work use three-body screening only.

Admittedly, the MBE(3)$_{\tau 3B}$ + HF/aDZ results that exhibit the best performance must be regarded as proof-of-concept calculations only, because the MBE(3) calculations reported in Figure 3 are performed at the B3LYP/aDZ level of theory and thus the HF/aDZ calculation that is applied as a supersystem correction is not significantly less expensive than the method we are aiming to approximate and which serves as the benchmark. That said, this example demonstrates that two different levels of theory can successfully be matched in a two-layer composite scheme that facilitates energy-based screening of the MBE(3) portion of the calculation. It also facilitates low-cost testing to establish a value for the τ_{3B} parameter. In what follows, we will consider more expensive levels of theory for the MBE(3) part of the calculation, including MP2 calculations. In such cases, MBE(3)$_{\tau 3B}$ + HF/aDZ proves to be extremely cost-effective in comparison to a supersystem calculation at the MP2 level, and in fact the supersystem HF/aDZ calculation amounts to only a small fraction of the cost of the fragment-based MP2 calculation. This will be demonstrated in Section 3.4, where timing data are presented. First, it is necessary to establish the accuracy of MBE(3)$_{\tau 3B}$ and related methods.

Figure 5 shows the signed errors for MBE(3)$_{\tau 3B}$-based approximations for the entire data set of water clusters, with $\tau_{3B} = 0.25$ kJ/mol as established above. These calculations are reported both at the B3LYP/aDZ level (Figure 5a) and also at the MP2/aDZ level (Figure 5b). At the B3LYP level, the MBE(3)$_{\tau 3B}$ approach alone (sans supersystem correction) exhibits a few data points that lie outside of the dynamic accuracy criterion. The two-layer MBE(3)$_{\tau 3B}$ + EFP approach is unable to rectify this situation, and in fact makes the errors slightly larger in most cases, but MBE(3)$_{\tau 3B}$ + HF/aDZ reduces all of the errors below the target accuracy. In an attempt to reduce the cost of the supersystem correction, and inspired by the success of Raghavachari and co-workers using semiempirical calculations for the supersystem correction in MIM2,[79,80] we also tested the semiempirical PM3 method.[90,91] We find that the MBE(3)$_{\tau 3B}$ + PM3 approach tends to exaggerate the polarization correction, pushing the errors in the opposite direction as compared to MBE(3)$_{\tau 3B}$ + EFP. (It should be noted that Raghavachari and co-workers have used PM6-D3 in their MIM2 calculations,[79,80] but this has not yet been implemented in Q-Chem.) In the end, however, all of the data in Figure 5a are accurate to within ±1.0 kJ/mol/monomer, including those for the MBE(3)$_{\tau 3B}$ and MBE(3)$_{\tau 3B}$ + EFP approaches that include no supersystem correction at all or else an EFP correction with negligible additional cost.

Errors are larger when the underlying quantum-chemical model is MP2/aDZ (Figure 5b), although the underlying trends are the same as those observed for B3LYP calculations. Here, MBE(3)$_{\tau 3B}$ + PM3 is slightly more accurate than MBE(3)$_{\tau 3B}$+ HF/aDZ, but neither method consistently achieves the stringent 0.4 kJ/mol/monomer accuracy criterion. Most errors are <1.0 kJ/mol/monomer, which cannot be said of the MBE(3)$_{\tau 3B}$ + EFP approach. We tried tightening the threshold τ_{3B} in an attempt to reduce the errors at the MP2/aDZ level, but to no effect. The size-dependent trends in the errors that are observed in Figure 5b using the threshold $\tau_{3B} = 0.25$ kJ/mol are nearly identical to those documented in Figure

DOI: 10.1021/acs.jctc.9b01095
J. Chem. Theory Comput. 2020, 16, 475−487

S2 using τ_{3B} = 0.20 kJ/mol, the latter of which results in more subsystem calculations performed at the target (MP2) level of theory. The level of accuracy that is observed in Figure 5b may simply be the intrinsic limit of the method. It is possible that this situation could be rectified by extending the method to MBE(4), but we leave that as a topic for future work.

3.3. Relative Energies. We next examine the challenging problem of predicting relative energies of different cluster isomers, using a set of 80 isomers of $(H_2O)_{20}$. (These are taken originally from ref 86 but are provided in the Supporting Information to the current work.) This data set includes 20 isomers from each of the four structural motifs (Figure 6) that

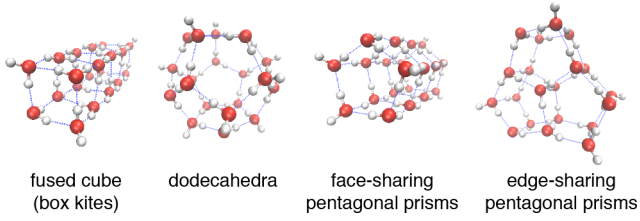


Figure 6. Examples of the four families of $(H_2O)_{20}$ isomers.

are found in water clusters in this size regime.[92] The same data set has been used in our previous work,[7,20,87] and benchmark energies have been computed at the MP2/aDZ and ωB97X-V/aTZ levels of theory, where "aTZ" means aug-cc-pVTZ. Regarding basis sets, we note that one-dimensional dissociation potentials computed at the MP2/aDZ level for water dimer mimic those computed at the CCSD(T)/aQZ level,[93] due to error cancellation. For the DFT calculations, we use the aTZ basis set as this is required for good results using modern B97-based functionals.[42] In order to use large, flexible basis sets such as aTZ, it appears to be necessary to iterate the *n*-body subsystem calculations to self-consistency,[1] which is done here but is notably *not* done in FMO calculations.[37] The absence of interfragment Pauli repulsion interactions renders FMO calculations unstable in large, flexible basis sets.[1,40,41]

Errors with respect to supersystem benchmarks that are reported below are defined as

$$\text{error} = E_{\text{rel}}^{\text{fragment}} - E_{\text{rel}}^{\text{supersys}} \tag{10}$$

These errors are plotted in Figure 7 for calculations at the MP2/aDZ level, using a variety of MBE(3)$_{\tau 3B}$-based approximations, and in Figure S3 for calculations at the ωB97X-V/aTZ level. Unlike the calculations in Section 3.2, where errors were plotted in per-monomer terms, here all of the clusters are the same size so we deal in total energy errors. A target accuracy might therefore be the "chemical accuracy" standard of ∼4 kJ/mol.

Results in Figure 7a correspond to the energy-screened MBE(3)$_{\tau 3B}$ method without any supersystem correction, which proves to be insufficient to reproduce MP2/aDZ relative energies with any real fidelity. An accuracy of ±4 kJ/mol can

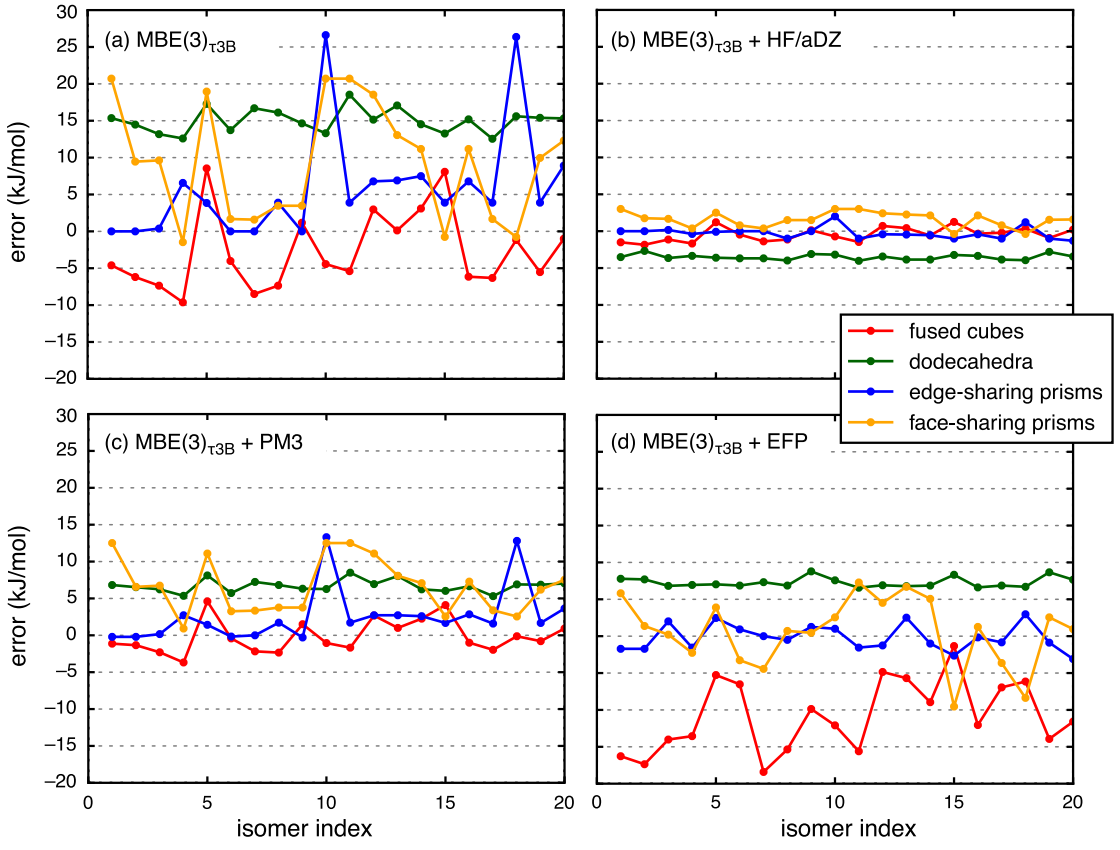


Figure 7. Signed errors (as compared to a supersystem MP2/aDZ calculation) in relative energies for isomers of $(H_2O)_{20}$, using MBE(3)$_{\tau 3B}$-based approximations at the MP2/aDZ level, with τ_{3B} = 0.25 kJ/mol. These approximations include (a) energy screening only, with no supersystem calculation; (b) MBE(3)$_{\tau 3B}$ + HF/aDZ, where a supersystem Hartree−Fock calculation is used to capture higher-order induction; (c) MBE(3)$_{\tau 3B}$ + PM3, where the supersystem correction is semiempirical; and (d) MBE(3)$_{\tau 3B}$ + EFP, where the supersystem calculation can be evaluated at essentially no cost.

DOI: 10.1021/acs.jctc.9b01095
J. Chem. Theory Comput. 2020, 16, 475−487

be achieved by application of an ONIOM-style correction at the HF/aDZ level, as shown in Figure 7b. Roughly the same behavior is observed for ωB97X-V/aTZ calculations, where errors in relative energies computed using the MBE(3)$_{\tau 3B}$ + HF/aDZ approximation all lie between −7 kJ/mol and +3 kJ/mol (Figure S3b). Together, these calculations establish that the MBE(3)$_{\tau 3B}$ + HF/aDZ approach is faithful to the underlying quantum-chemical model for relative energies, to *nearly* (but not quite) a "chemical accuracy" standard of ±4 kJ/mol.

Returning to the MP2/aDZ results, the accuracy is degraded a bit when PM3 is substituted for HF/aDZ (Figure 7c), with a small, systematic error for the dodecahedral structures relative to the others and a few noticeable deviations for the face-sharing pentagonal prisms. In an attempt to bypass the need for a supersystem calculation at the HF/aDZ level of theory, we attempted to substitute HF/6-31G* instead, and also the semiempirical HF-3c method,[94] which requires only a minimal-basis Hartree−Fock calculation. Neither of these methods, when used in conjunction with fragmentation at the MBE(3)$_{\tau 3B}$ level, afforded results that were accurate enough to warrant consideration here.

The MBE(3)$_{\tau 3B}$ + EFP approach is worth considering because the EFP correction is essentially free, but again the results (Figure 7d) represent only a modest improvement upon the underlying MBE(3)$_{\tau 3B}$ approximation, and are considerably worse than results obtained using a supersystem HF/aDZ or PM3 correction. In addition, MBE(3)$_{\tau 3B}$ + EFP exhibits a sizable, systematic error in its prediction of the energies of fused-cube structures with respect to those of the other three structural motifs. Errors for the fused-cube structures are relatively large and negative, indicating that the EFP induction correction overstabilizes these isomers relative to the others. Based on these results, it seems that accurate results for relative isomer energies are achievable with MBE(3)$_{\tau 3B}$ only if it is combined with a low-level quantum chemistry calculation performed on the supersystem. While EFP can be used for screening, its accuracy appears to be insufficient to replace Hartree−Fock theory as the supersystem correction.

We noted in Section 2.2 that MBE(3)$_{\tau 3B}$ + EFP bears some similarity to the EFMO method,[55−60] although the latter uses a two-body QM approximation (namely, FMO2) and distance-based rather than energy-based screening of the two-body subsystem calculations. Nevertheless, it is interesting to see how EFMO performs in comparison to MBE(3)$_{\tau 3B}$-based approaches (especially MBE(3)$_{\tau 3B}$ + EFP) for this same set of cluster isomers. These results are shown in Figure 8. Overall, and in absolute terms, the EFMO approach is no more accurate than MBE(3)$_{\tau 3B}$ + EFP and arguably less so, with errors that range from −10 kJ/mol up to +20 kJ/mol. For MBE(3)$_{\tau 3B}$ + EFP, most of the errors are negative, ranging up to −20 kJ/mol. These EFMO calculations use a dimensionless cutoff parameter R_{cut} = 2.0 to switch between short-range FMO2 and longer-range EFP, based on the dimensionless intermolecular distance[55]

$$R_{IJ} = \min_{i\in I, j\in J}\left(\frac{\|\mathbf{r}_i - \mathbf{r}_j\|}{\|\mathbf{R}_i^{vdW} - \mathbf{R}_j^{vdW}\|}\right) \quad (11)$$

QM calculations are retained only if $R_{IJ} < R_{cut}$, and the value R_{cut} = 2.0 that we use represents the most conservative choice from the systematic tests that were reported in ref 58.

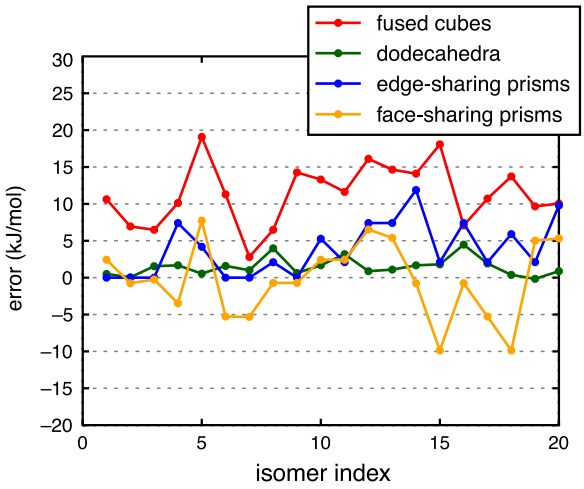


Figure 8. Signed errors in relative energies for isomers of $(H_2O)_{20}$, computed using the EFMO method at the MP2/aDZ level. (Error is defined relative to a supersystem MP2/aDZ calculation.) The color scheme and energy scale are the same as in Figure 7, which compares MBE(3)$_{\tau 3B}$-based approximations for the same data set. EFMO calculations used the default value R_{cut} = 2.0 for the dimensionless cutoff used in the distance screening.[55]

One interesting aspect of the EFMO data is that what appears to be a systematic error in EFP energies for the fused-cube isomers manifests again, as this particular class of isomers exhibits much larger errors as compared to the others, similar to what was observed in the case of MBE(3)$_{\tau 3B}$ + EFP (Figure 7d) but opposite in sign in the case of EFMO. Noting that positive errors correspond to understabilization by the fragment-based approximation (according to eq 10), we observe that EFMO understabilizes the fused-cube structures whereas the MBE(3)$_{\tau 3B}$ + EFP approach overstabilizes them, despite the fact that the ONIOM-style induction correction is applied using the same force field in both calculations. The EFMO calculation lacks three-body QM terms that are present in MBE(3), which may make the results more erratic, but we suspect that the primary explanation for this difference with respect to MBE(3)$_{\tau 3B}$ + EFP lies in the rather aggressive distance-based thresholding that is applied to the two-body corrections ΔE_{IJ} in the EFMO approach. For stable, low-energy cluster geometries, the two-body terms are almost always attractive, whereas three-body corrections ΔE_{IJK} are a mix of attractive and repulsive contributions, as seen in Figure 1. Comparing the MBE(3)$_{\tau 3B}$ results in Figure 7a, which contain no higher-order induction, to the accurate MBE(3)$_{\tau 3B}$ + HF results in Figure 7b, where apparently the higher-order effects are well described, one may determine that the higher-order polarization effects that are needed to correct MBE(3)$_{\tau 3B}$ are net attractive for all of the isomers except the fused-cube structures, where they are net repulsive. According to the EFMO results, underestimation of the two-body interactions does not cancel overestimation (by EFP) of higher-order interactions, so the overall effect is still underestimation of the relative energies.

3.4. Computational Cost. For MBE(3)$_{\tau 3B}$ calculations it seems reasonable to expect that the number of "significant" three-body terms (i.e., those with $|\Delta E_{IJK}^{EFP}| > \tau_{3B}$) will grow linearly with system size. Although we have no formal proof of this fact, it appears to be borne out in practice, so let us assume for the sake of discussion that this is indeed the case. Then

DOI: 10.1021/acs.jctc.9b01095
J. Chem. Theory Comput. 2020, 16, 475−487

energy-based screening of the three-body interactions reduces the formal scaling of MBE(3) to $O(N^2)$ with system size, since there remains a quadratic number of two-body terms, but of course that number can also be reduced to $O(N)$ by means of energy screening.

The efficacy of both energy- and distance-based cutoffs is examined in Figure 9, by plotting the number of trimers that must be computed at the QM level for the cluster data set $(H_2O)_{N=6-37}$. These subsystem counts are based on the energy-screening threshold τ_{3B} = 0.25 kJ/mol that was determined in Section 3.2 and the distance cutoff of 7 Å that was previously determined for water clusters.[20] Distance-based screening results in a provably $O(N)$ algorithm because the size of the QM domains is fixed and cannot grow with respect to supersystem size. That said, for finite clusters the number of trimers grows a bit erratically in the distance-based approach, with a noticeable jump between N = 29 and N = 30, for example (see Figure 9). This particular anomaly results from a structural transition that occurs in this size regime,[95] such that the N = 29 structure has fewer rings than either the N = 28 or N = 30 structures.[84] Changes in the number of rings also explain the dip in the number of trimers that is observed between N = 23 and N = 24.[84] These structural changes have the effect of modifying the distribution of intermolecular distances, which changes the efficacy of the distance-based screening protocol.

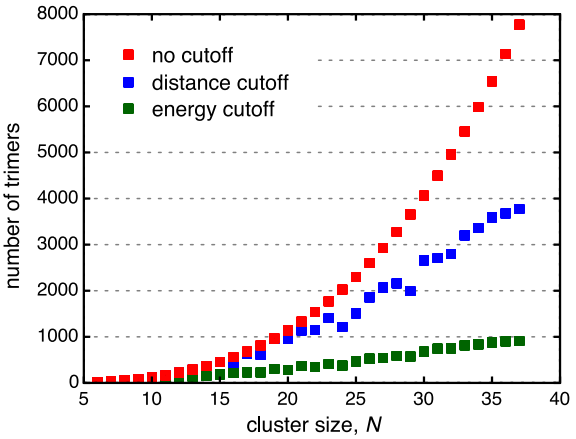


Figure 9. Total number of three-body subsystems that must be computed at the QM level for water clusters $(H_2O)_N$. The "no cutoff" data represent a complete MBE(3) calculation with $O(N^3)$ trimers. Distance-based screening uses a cutoff of 7 Å (as determined in ref 20), and energy-based screening represents the MBE(3)$_{\tau 3B}$ approach with a threshold τ_{3B} = 0.25 kJ/mol.

In contrast, the energy-based screening approach interpolates smoothly through these structural changes with no sudden changes in its efficacy. Furthermore, the number of significant trimers appears to increase only as $O(N)$, with a significantly smaller prefactor as compared to distance-based screening. That the energy-based screening procedure is more efficient even in water clusters is notable, because these systems (where the fragments are very small and identical in size) likely represent a best-case scenario for the efficacy of distance-based thresholds. We expect that the advantages of energy-based screening will become even more pronounced in more complex environments.

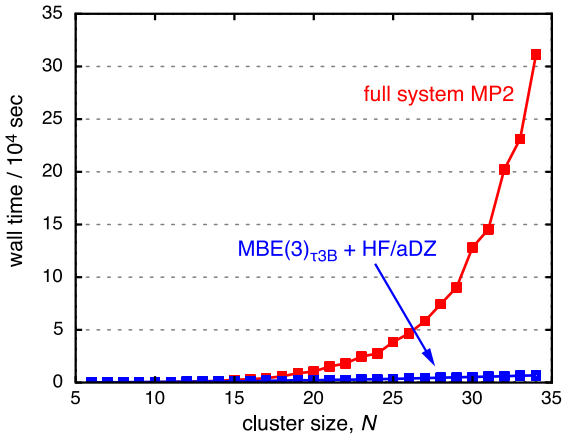


Figure 10. Wall time for MP2/aDZ calculations on clusters $(H_2O)_{N=6-34}$. All calculations were performed on a single 40-core compute node, including all of the MP2 subsystem calculations required for MBE(3)$_{\tau 3B}$ (using τ_{3B} = 0.25 kJ/mol). Timings for the MBE(3)$_{\tau 3B}$ + HF/aDZ approach include the cost of the supersystem Hartree−Fock calculation, which is formally $O(N^3)$ but which cannot be distinguished, on this scale, from the $O(N)$ timings for the MBE(3)$_{\tau 3B}$ calculations.

Figure 10 compares timing data for MP2/aDZ calculations across the sequence of clusters $(H_2O)_{N=6-34}$, comparing the $O(N^5)$ cost of a traditional MP2 calculation to the cost of our best approximation to it, MBE(3)$_{\tau 3B}$ + HF/aDZ. These data represent wall times for calculations performed on a single 40-core node. This means that we have not exploited the trivial parallelizability of the subsystem calculations to report a very low wall time while hiding the true cost of the method behind the number of processors that was used. Instead, the data in Figure 10 reflect the full cost of the method, and the hardware required to realize these timings is quite modest.

According to the data in Figure 9, the number of energetically significant trimers appears to increase linearly with system size, and thus the cost of an MBE(3)$_{\tau 3B}$ calculation is expected to grow as $O(N)$. (Technically, to achieve linear scaling we would need to screen the dimers as well, which is not done in these calculations; however, the dimers represent so little of the total cost that this has no discernible impact on the scaling for the system sizes examined here.) Furthermore, in the timing data for MBE(3)$_{\tau 3B}$ + HF/aDZ that are given in Figure 10, we have not attempted to separate the cost of the supersystem HF/aDZ calculation because the difference between this and the MBE(3)$_{\tau 3B}$ timing is not discernible on the scale of Figure 10. That fact alone is significant: the supersystem Hartree−Fock calculation is not a significant bottleneck in these calculations. This behavior cannot persist as $N \rightarrow \infty$; the supersystem Hartree−Fock cost scales as $O(N^3)$ in principle and as $O(N^{2.x})$ in most practical applications, but with an efficient implementation the prefactor is low. For the system sizes considered in this work, there is thus no reason *not* to use the supersystem HF/aDZ correction, which significantly improves the accuracy while having very little effect on the cost. For the MP2/aDZ calculations whose timings are shown in Figure 10, the crossover point at which the MBE(3)$_{\tau 3B}$ + HF calculation becomes cheaper than the supersystem MP2 calculation occurs at N = 12. For calculations at the much cheaper B3LYP/aDZ level, the

DOI: 10.1021/acs.jctc.9b01095
J. Chem. Theory Comput. 2020, 16, 475−487

crossover point occurs at N = 45. (Timing data for B3LYP calculations are plotted in Figure S4, out to N = 140.)

4. CONCLUSIONS

Calculations in water clusters demonstrate that energy-based screening using the automatically generated EFP force field can be used to significantly reduce the cost of MBE(n) calculations, while still preserving reasonable fidelity with respect to a full-system calculation. An energy cutoff threshold τ_{3B} = 0.25 kJ/mol, used in conjunction with a three-body expansion, maintains accuracy at the level of 1−2 kJ/mol/monomer with respect to the supersystem calculation performed at the same level of theory as the subsystem calculations. We have called this approach MBE(3)$_{\tau 3B}$.

Calculations on a large set of $(H_2O)_{20}$ conformational isomers reveal that MBE(3)$_{\tau 3B}$ alone is insufficient to predict relative energies, at either the MP2/aDZ or ωB97X-V/aTZ level of theory, to within a "chemical accuracy" target of ±4 kJ/mol. However, a two-layer approach that includes an ONIOM-style correction (requiring a supersystem calculation at the HF/aDZ level) does approach this level of accuracy. Although the supersystem HF/aDZ calculation formally introduces a $O(N^3)$ bottleneck, for calculations at the MP2 level this full-system Hartree−Fock calculation amounts to only a small fraction of the subsystem (monomer, dimer, and trimer) MP2 cost. For $(H_2O)_N$ at the MP2/aDZ level, the cost of the composite MBE(3)$_{\tau 3B}$ + HF/aDZ approach is cheaper than a full-system MP2/aDZ calculation already at N = 12. Energy screening of the subsystem calculations can reduce the cost of the fragment-based MBE(3)$_{\tau 3B}$ part of the calculation to $O(N)$ with system size, with an even smaller prefactor as compared to that incurred using distance-based thresholding.

There are many fragment-based quantum chemistry approaches with documented parallel scalability, but fewer whose accuracy is genuinely faithful to the underlying quantum-chemical model.[1,19] This is especially true when it comes to their use with large basis sets appropriate for correlated wave function calculations. In this preliminary report of the energy-screened MBE(n) approach, we have not yet (quite) achieved chemical accuracy for energies, but there is reason to think that this could be improved. We have certainly not exhausted the possibilities for adding an ONIOM-style supersystem correction, and the addition of four-body terms (which we had previously deemed to be borderline-intractable[7,19]) seems feasible in conjunction with energy-based screening. A generalized MBE,[1,10−12] based on dimers of overlapping fragments, exhibits outstanding accuracy for systems ranging from noncovalent clusters[7,12] to proteins.[8] The number of subsystem calculations required for this generalized approach is relatively large, and thus the cost is relatively high,[7,8] but energy-based screening is likely to reduce this cost considerably. These extensions are currently being explored in our group.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.9b01095.

Additional data on the performance of the method (PDF)

Geometries for the data set of $(H_2O)_{20}$ isomers (TXT)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: herbert@chemistry.ohio-state.edu.

ORCID

Kuan-Yu Liu: 0000-0002-2619-3301

John M. Herbert: 0000-0002-1663-2278

Present Address

†(K.-Y.L.) Q-Chem Inc., 6601 Owens Dr., Suite 105, Pleasanton, CA.

Notes

The authors declare the following competing financial interest(s): J.M.H. serves on the Board of Directors of Q-Chem Inc.

■ ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences under Award No. DE-SC0008550. Calculations were performed at the Ohio Supercomputer Center under project no. PAA-0003.[96] All MBE(3)-based calculations were performed with a locally modified version of Q-Chem.[48] EFMO calculations were performed with the GAMESS program.[97]

■ REFERENCES

(1) Herbert, J. M. Fantasy versus reality in fragment-based quantum chemistry. *J. Chem. Phys.* **2019**, *151*, 170901.

(2) Fedorov, D. G., Kitaura, K., Eds. *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*; CRC Press: Boca Raton, 2009.

(3) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation methods: A route to accurate calculations on large systems. *Chem. Rev.* **2012**, *112*, 632−672.

(4) Collins, M. A.; Bettens, R. P. Energy-based molecular fragmentation methods. *Chem. Rev.* **2015**, *115*, 5607−5642.

(5) Raghavachari, K.; Saha, A. Accurate composite and fragment-based quantum chemical methods for large molecules. *Chem. Rev.* **2015**, *115*, 5643−5677.

(6) Gordon, M. S., Ed. *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems*; John Wiley & Sons: Hoboken, NJ, 2017.

(7) Lao, K. U.; Liu, K.-Y.; Richard, R. M.; Herbert, J. M. Understanding the many-body expansion for large systems. II. Accuracy considerations. *J. Chem. Phys.* **2016**, *144*, 164105.

(8) Liu, J.; Herbert, J. M. Pair−pair approximation to the generalized many-body expansion: An efficient and accurate alternative to the four-body expansion, with applications to *ab initio* protein energetics. *J. Chem. Theory Comput.* **2016**, *12*, 572−584.

(9) Liu, J.; He, X. Accurate prediction of energetic properties of ionic liquid clusters using a fragment-based quantum mechanical method. *Phys. Chem. Chem. Phys.* **2017**, *19*, 20657−20666.

(10) Richard, R. M.; Herbert, J. M. A generalized many-body expansion and a unified view of fragment-based methods in electronic structure theory. *J. Chem. Phys.* **2012**, *137*, 064113.

(11) Richard, R. M.; Herbert, J. M. The many-body expansion with overlapping fragments: Analysis of two approaches. *J. Chem. Theory Comput.* **2013**, *9*, 1408−1416.

(12) Jacobson, L. D.; Richard, R. M.; Lao, K. U.; Herbert, J. M. Efficient monomer-based quantum chemistry methods for molecular and ionic clusters. *Annu. Rep. Comput. Chem.* **2013**, *9*, 25−56.

(13) Gillan, M. J.; Alfé, D.; Bygrave, P. J.; Taylor, C. R.; Manby, F. R. Energy benchmarks for water clusters and ice structures from an embedded many-body expansion. *J. Chem. Phys.* **2013**, *139*, 114101.

(14) Heßelmann, A. Correlation effects and many-body interactions in water clusters. *Beilstein J. Org. Chem.* **2018**, *14*, 979−991.

DOI: 10.1021/acs.jctc.9b01095
J. Chem. Theory Comput. 2020, 16, 475−487

(15) Khire, S. S.; Gadre, S. R. Pragmatic many-body approach for economic MP2 energy estimation of molecular clusters. *J. Phys. Chem. A* **2019**, *123*, 5005−5011.
(16) Christie, R. A.; Jordan, K. D. *n*-body decomposition approach to the calculation of interaction energies of water clusters. In *Intermolecular Forces and Clusters II*, Vol. *116*; Wales, D., Christie, R. A., Eds.; Springer-Verlag: Heidelberg, 2005.
(17) Cui, J.; Liu, H.; Jordan, K. D. Theoretical characterization of the $(H_2O)_{21}$ cluster: Application of an *n*-body decomposition procedure. *J. Phys. Chem. B* **2006**, *110*, 18872−18878.
(18) Wang, F.-F.; Deible, M. J.; Jordan, K. D. Benchmark study of the interaction energy for an $(H_2O)_{16}$ cluster: Quantum Monte Carlo and complete basis set limit MP2 results. *J. Phys. Chem. A* **2013**, *117*, 7606−7611.
(19) Richard, R. M.; Lao, K. U.; Herbert, J. M. Aiming for benchmark accuracy with the many-body expansion. *Acc. Chem. Res.* **2014**, *47*, 2828−2836.
(20) Liu, K.-Y.; Herbert, J. M. Understanding the many-body expansion for large systems. III. Critical role of four-body terms, counterpoise corrections, and cutoffs. *J. Chem. Phys.* **2017**, *147*, 161729.
(21) Richard, R. M.; Lao, K. U.; Herbert, J. M. Understanding the many-body expansion for large systems. I. Precision considerations. *J. Chem. Phys.* **2014**, *141*, 014108.
(22) Dahlke, E. E.; Truhlar, D. G. Electrostatically embedded many-body expansion for large systems, with applications to water clusters. *J. Chem. Theory Comput.* **2007**, *3*, 46−53.
(23) Liu, J.; Qi, L.; Zhang, J. Z. H.; He, X. Fragment quantum mechanical method for large-sized ion-water clusters. *J. Chem. Theory Comput.* **2017**, *13*, 2021−2034.
(24) Ouyang, J. F.; Bettens, R. P. A. When are many-body effects significant? *J. Chem. Theory Comput.* **2016**, *12*, 5860−5867.
(25) Beran, G. J. O. Approximating quantum many-body intermolecular interactions in molecular clusters using classical polarizable force fields. *J. Chem. Phys.* **2009**, *130*, 164115.
(26) Beran, G. J. O.; Nanda, K. Predicting organic crystal lattice energies with chemical accuracy. *J. Phys. Chem. Lett.* **2010**, *1*, 3480−3487.
(27) Wen, S.; Beran, G. J. O. Accurate molecular crystal lattice energies from a fragment QM/MM approach with on-the-fly ab initio force field parametrization. *J. Chem. Theory Comput.* **2011**, *7*, 3733−3742.
(28) Nanda, K. D.; Beran, G. J. O. Prediction of organic molecular crystal geometries from MP2-level fragment quantum mechanical/molecular mechanical calculations. *J. Chem. Phys.* **2012**, *137*, 174106.
(29) Wen, S.; Nanda, K.; Huang, Y.; Beran, G. J. O. Practical quantum mechanics-based fragment methods for predicting molecular crystal properties. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7578−7590.
(30) Hartman, J. D.; Beran, G. J. O. Fragment-based electronic structure approach for computing nuclear magnetic resonance chemical shifts in molecular crystals. *J. Chem. Theory Comput.* **2014**, *10*, 4862−4872.
(31) Beran, G. J. O.; Hartman, J. D.; Heit, Y. N. Predicting molecular crystal properties from first principles: Finite-temperature thermochemistry to NMR crystallography. *Acc. Chem. Res.* **2016**, *49*, 2501−2508.
(32) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. The effective fragment potential method: A QM-based MM approach to modeling environmental effects in chemistry. *J. Phys. Chem. A* **2001**, *105*, 293−307.
(33) Ghosh, D.; Kosenkov, D.; Vanovschi, V.; Williams, C. F.; Herbert, J. M.; Gordon, M. S.; Schmidt, M. W.; Slipchenko, L. V.; Krylov, A. I. Noncovalent interactions in extended systems described by the effective fragment potential method: Theory and application to nucleobase oligomers. *J. Phys. Chem. A* **2010**, *114*, 12739−12754.
(34) Slipchenko, L. V.; Gurunathan, P. K. Effective fragment potential method: Past, present, and future. In *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems*; Gordon, M. S., Ed.; Wiley: Hoboken, 2017; Chapter 6, pp 183−208.
(35) Flick, J. C.; Kosenkov, D.; Hohenstein, E. G.; Sherrill, C. D.; Slipchenko, L. V. Accurate prediction of noncovalent interaction energies with the effective fragment potential method: Comparison of energy components to symmetry-adapted perturbation theory for the S22 test set. *J. Chem. Theory Comput.* **2012**, *8*, 2835−2843 Erratum.: *ibid.* **2014**, *10*, 4759−4760.
(36) Fedorov, D. G.; Kitaura, K. Theoretical background of the fragment molecular orbital (FMO) method and its implementation in GAMESS. In *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*; Fedorov, D. G., Kitaura, K., Eds.; CRC Press: Boca Rotan, FL, 2009; Chapter 2, pp 5−36.
(37) Nagata, T.; Fedorov, D. G.; Kitaura, K. Mathematical formulation of the fragment molecular orbital method. In *Linear-Scaling Techniques in Computational Chemistry and Physics*, Vol. *13*; Zaleśny, R., Papadopoulos, M. G., Mezey, P. G., Leszczynski, J., Eds.; Springer: New York, 2011; Chapter 2, pp 17−64.
(38) Fedorov, D. G.; Nagata, T.; Kitaura, K. Exploring chemistry with the fragment molecular orbital method. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7562−7577.
(39) Fedorov, D. G. The fragment molecular orbital method: Theoretical development, implementation in GAMESS, and applications. *WIREs Comput. Mol. Sci.* **2017**, *7*, e1322.
(40) Fedorov, D. G.; Slipchenko, L. V.; Kitaura, K. Systematic study of the embedding potential description in the fragment molecular orbital method. *J. Phys. Chem. A* **2010**, *114*, 8742−8753.
(41) Fedorov, D. G.; Kitaura, K. Use of an auxiliary basis set to describe the polarization in the fragment molecular orbital method. *Chem. Phys. Lett.* **2014**, *597*, 99−105.
(42) Mardirossian, N.; Head-Gordon, M. ωB97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy. *Phys. Chem. Chem. Phys.* **2014**, *16*, 9904−9924.
(43) Mardirossian, N.; Head-Gordon, M. Mapping the genome of meta-generalized gradient approximation density functionals: The search for B97M-V. *J. Chem. Phys.* **2015**, *142*, 074111.
(44) Mardirossian, N.; Head-Gordon, M. ωB97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *J. Chem. Phys.* **2016**, *144*, 214110.
(45) Mardirossian, N.; Head-Gordon, M. Thirty years of density functional theory in computational chemistry: An overview and extensive assessment of 200 density functionals. *Mol. Phys.* **2017**, *115*, 2315−2372.
(46) Liu, J.; Rana, B.; Liu, K.-Y.; Herbert, J. M. Variational formulation of the generalized many-body expansion with self-consistent embedding charges: Simple and correct analytic energy gradient for fragment-based *ab initio* molecular dynamics. *J. Phys. Chem. Lett.* **2019**, *10*, 3877−3886.
(47) Nagata, T.; Brorsen, K.; Fedorov, D. G.; Kitaura, K.; Gordon, M. S. Fully analytic energy gradient in the fragment molecular orbital method. *J. Chem. Phys.* **2011**, *134*, 124115.
(48) Shao, Y.; et al. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **2015**, *113*, 184−215.
(49) Kaliman, I. A.; Slipchenko, L. V. LIBEFP: A new parallel implementation of the effective fragment potential method as a portable software library. *J. Comput. Chem.* **2013**, *34*, 2284−2292.
(50) Liu, P.; Agrafiotis, D. K.; Theobald, D. L. Fast determination of the optimal rotational matrix for macromolecular superpositions. *J. Comput. Chem.* **2009**, *31*, 1561−1563.
(51) Santra, B.; Michaelides, A.; Scheffler, M. On the accuracy of density-functional theory exchange-correlation functionals for H bonds in small water clusters: Benchmarks approaching the complete basis set limit. *J. Chem. Phys.* **2007**, *127*, 184104.
(52) Bryantsev, V. S.; Diallo, M. S.; van Duin, A. C. T.; Goddard, W. A., III Evaluation of B3LYP, X3LYP, and M06-class density functionals for predicting the binding energies of neutral, protonated, and deprotonated water clusters. *J. Chem. Theory Comput.* **2009**, *5*, 1016−1026.

DOI: 10.1021/acs.jctc.9b01095
J. Chem. Theory Comput. 2020, 16, 475−487

(53) Gillan, M. J.; Alfè, D.; Michaelides, A. Perspective: How good is DFT for water? *J. Chem. Phys.* **2016**, *144*, 130901.
(54) Dubinets, N.; Slipchenko, L. V. Effective fragment potential method for H-bonding: How to obtain parameters for nonrigid fragments. *J. Phys. Chem. A* **2017**, *121*, 5301−5312.
(55) Steinmann, C.; Fedorov, D. G.; Jensen, J. H. Effective fragment molecular orbital method: A merger of the effective fragment potential and fragment molecular orbital methods. *J. Phys. Chem. A* **2010**, *114*, 8705−8712.
(56) Nagata, T.; Fedorov, D. G.; Sawada, T.; Kitaura, K.; Gordon, M. S. A combined effective fragment potential−fragment molecular orbital method. II. Analytic gradient and application to geometry optimization of solvated tetraglycine chignolin. *J. Chem. Phys.* **2011**, *134*, 034110.
(57) Steinmann, C.; Fedorov, D. G.; Jensen, J. H. The effective fragment molecular orbital method for fragments connected by covalent bonds. *PLoS One* **2012**, *7*, e41117.
(58) Pruitt, S. R.; Steinmann, C.; Jensen, J. H.; Gordon, M. S. Fully integrated effective fragment molecular orbital method. *J. Chem. Theory Comput.* **2013**, *9*, 2235−2249.
(59) Bertoni, C.; Gordon, M. S. Analytic gradients for the effective fragment molecular orbital method. *J. Chem. Theory Comput.* **2016**, *12*, 4743−4767.
(60) Steinmann, C.; Jensen, J. H. Effective fragment molecular orbital method. In *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems*; Gordon, M. S., Ed.; Wiley: Hoboken, 2017; Chapter 5, pp 165−182.
(61) Pruitt, S. R.; Bertoni, C.; Brorsen, K. R.; Gordon, M. S. Efficient and accurate fragmentation methods. *Acc. Chem. Res.* **2014**, *47*, 2786−2794.
(62) Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z. F.; Liu, F. Y.; Li, H. B.; Ding, L. N.; Morokuma, K. The ONIOM method and its applications. *Chem. Rev.* **2015**, *115*, 5678−5796.
(63) Hopkins, B. W.; Tschumper, G. S. A multicentered approach to integrated QM/QM calculations. Applications to multiply hydrogen bonded systems. *J. Comput. Chem.* **2003**, *24*, 1563−1568.
(64) Hopkins, B. W.; Tschumper, G. S. Multicentered QM/QM methods for overlapping model systems. *Mol. Phys.* **2005**, *103*, 309−315.
(65) Hopkins, B. W.; Tschumper, G. S. Integrated quantum mechanical approaches for extended π systems: Multicentered QM/QM studies of the cyanogen and diacetylene trimers. *Chem. Phys. Lett.* **2005**, *407*, 362−367.
(66) Tschumper, G. S. Multicentered integrated QM:QM methods for weakly bound clusters: An efficient and accurate 2-body:many-body treatment of hydrogen bonding and van der Waals interactions. *Chem. Phys. Lett.* **2006**, *427*, 185−191.
(67) Elsohly, A. M.; Shaw, C. L.; Guice, M. E.; Smith, B. D.; Tschumper, G. S. Analytic gradients for the multicentered QM:QM method for weakly bound clusters: Efficient and accurate 2-body:many-body geometry optimizations. *Mol. Phys.* **2007**, *105*, 2777−2782.
(68) Bates, D. M.; Smith, J. R.; Janowski, T.; Tschumper, G. S. Development of a 3-body:many-body integrated fragmentation method for weakly bound clusters and application to water clusters $(H_2O)_{n=3-10,16,17}$. *J. Chem. Phys.* **2011**, *135*, 044123.
(69) Bates, D. M.; Smith, J. R.; Tschumper, G. S. Efficient and accurate methods for the geometry optimization of water clusters: Application of analytic gradients for the two-body:many-body QM:QM fragmentation method to $(H_2O)_n$, n = 3−10. *J. Chem. Theory Comput.* **2011**, *7*, 2753−2760.
(70) Howard, J. C.; Tschumper, G. S. *N*-body:many-body QM:QM vibrational frequencies: Application to small hydrogen-bonded clusters. *J. Chem. Phys.* **2013**, *139*, 184113.
(71) Sexton, T. M.; Tschumper, G. S. 2-Body:many-body QM:QM study of structures, energetics, and vibrational frequencies for microhydrated halide ions. *Mol. Phys.* **2019**, *117*, 1413−1420.
(72) Mayhall, N. J.; Raghavachari, K. Molecules-in-molecules: An extrapolated fragment-based approach for accurate calculations on large molecules and materials. *J. Chem. Theory Comput.* **2011**, *7*, 1336−1343.
(73) Saha, A.; Raghavachari, K. Analysis of different fragmentation strategies on a variety of large peptides: Implementation of a low level of theory in fragment-based methods can be a crucial factor. *J. Chem. Theory Comput.* **2015**, *11*, 2012−2023.
(74) Jose, K. V. J.; Raghavachari, K. Evaluation of energy gradients and infrared vibrational spectra through molecules-in-molecules fragment-based approach. *J. Chem. Theory Comput.* **2015**, *11*, 950−961.
(75) Jose, K. V. J.; Beckett, D.; Raghavachari, K. Vibrational circular dichroism spectra for large molecules through molecules-in-molecules fragment-based approach. *J. Chem. Theory Comput.* **2015**, *11*, 4238−4247.
(76) Jose, K. V. J.; Raghavachari, K. Molecules-in-molecules fragment-based method for the calculation of chiroptical spectra of large molecules: Vibrational circular dichroism and Raman optical spectra of alanine polypeptides. *Chirality* **2016**, *28*, 755−768.
(77) Jose, K. V. J.; Raghavachari, K. Fragment-based approach for the evaluation of NMR chemical shifts for large biomolecules incorporating the effects of the solvent environment. *J. Chem. Theory Comput.* **2017**, *13*, 1147−1158.
(78) Jose, K. V. J.; Raghavachari, K. Molecules-in-molecules fragment-based method for the accurate evaluation of vibrational and chiroptical spectra for large molecules. In *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems*; Gordon, M. S., Ed.; Wiley: 2017; Chapter 4, pp 141−164.
(79) Thapa, B.; Beckett, D.; Erickson, J.; Raghavachari, K. Theoretical study of protein−ligand interactions using the molecules-in-molecules fragmentation-based method. *J. Chem. Theory Comput.* **2018**, *14*, 5143−5155.
(80) Thapa, B.; Beckett, D.; Jose, K. V. J.; Raghavachari, K. Assessment of fragmentation strategies for large proteins using the multilayer molecules-in-molecules approach. *J. Chem. Theory Comput.* **2018**, *14*, 1383−1394.
(81) Debnath, S.; Sengupta, A.; Jose, K. V. J.; Raghavachari, K. Fragment-based approaches for supramolecular interaction energies: Applications to foldamers and their complexes with anions. *J. Chem. Theory Comput.* **2018**, *14*, 6226−6239.
(82) Dahlke, E. E.; Truhlar, D. G. Electrostatically embedded many-body correlation energy, with applications to the calculation of accurate second-order Møller−Plesset perturbation theory energies for large water clusters. *J. Chem. Theory Comput.* **2007**, *3*, 1342−1348.
(83) Usvyat, D.; Maschio, L.; Schütz, M. Periodic and fragment models based on the local correlation approach. *WIREs Comput. Mol. Sci.* **2018**, *8*, e1357.
(84) Kazachenko, S.; Thakkar, A. J. Improved minima-hopping. TIP4P water clusters, $(H_2O)_n$ with $n \leq 37$. *Chem. Phys. Lett.* **2009**, *476*, 120−124.
(85) Kazachenko, S.; Thakkar, A. J. Water nanodroplets: Predictions of five model potentials. *J. Chem. Phys.* **2013**, *138*, 194302.
(86) Wales, D. J.; Hodges, M. P. Global minima of water clusters $(H_2O)_n$, $n \leq 21$, described by an empirical potential. *Chem. Phys. Lett.* **1998**, *286*, 65−72.
(87) Richard, R. M.; Lao, K. U.; Herbert, J. M. Approaching the complete-basis limit with a truncated many-body expansion. *J. Chem. Phys.* **2013**, *139*, 224102.
(88) Gill, P. M. W.; Johnson, B. G.; Pople, J. A. A standard grid for density-functional calculations. *Chem. Phys. Lett.* **1993**, *209*, 506−512.
(89) Ouyang, J. F.; Bettens, R. P. A. Many-body basis set superposition effect. *J. Chem. Theory Comput.* **2015**, *11*, 5132−5143.
(90) Stewart, J. J. P. Optimization of parameters for semiempirical methods. I. Method. *J. Comput. Chem.* **1989**, *10*, 209−220.
(91) Stewart, J. J. P. Optimization of parameters for semiempirical methods. II. Applications. *J. Comput. Chem.* **1989**, *10*, 221−264.
(92) Yoo, S.; Xantheas, S. S. Structures, energetics, and spectroscopic fingerprints of water clusters n = 2−24. In *Handbook*

DOI: 10.1021/acs.jctc.9b01095
J. Chem. Theory Comput. 2020, 16, 475−487

of Computational Chemistry; Leszczynski, J., Ed.; Springer Science + Business Media: 2012; Chapter 21, pp 761−792.
(93) Willow, S. Y.; Salim, M. A.; Kim, K. S.; Hirata, S. *Ab initio* molecular dynamics of liquid water using embedded-fragment second-order many-body perturbation theory towards its accurate property prediction. *Sci. Rep.* **2015**, *5*, 14358.
(94) Sure, R.; Grimme, S. Corrected small basis set Hartree-Fock method for large systems. *J. Comput. Chem.* **2013**, *34*, 1672−1685.
(95) Kabrede, H. Using vibrational modes in the search for global minima of atomic and molecular clusters. *Chem. Phys. Lett.* **2006**, *430*, 336−339.
(96) Ohio Supercomputer Center, http://osc.edu/ark:/19495/f5s1ph73.
(97) Gordon, M. S.; Schmidt, M. W. Advances in electronic structure theory: GAMESS a decade later. In *Theory and Applications of Computational Chemistry*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier: Amsterdam, 2005; Chapter 41, pp 1167−1189.

DOI: 10.1021/acs.jctc.9b01095
J. Chem. Theory Comput. 2020, 16, 475−487