"Electrostatically Embedded Many-Body Expansion for Large Systems," E. D. Speetzen, H. R. Leverentz, H Lin, and Donald G. Truhlar, in Accurate Condensed-Phase Electronic Structure Theory, edited by F. R. Manby (CRC Press, Boca Raton, FL, 2011), pages 105-127.

chapter five

Electrostatically embedded many-body expansion for large systems

Erin Dahlke Speetzen, Hannah R. Leverentz, Hai Lin, and Donald G. Truhlar

Contents

5.1 *Introduction*

The need to carry out accurate atomistic simulations of condensed-phase systems is a major challenge to computational chemists. For large systems (often tens or hundreds of thousands of atoms) quantum chemistry is usually too expensive to be practical, and more approximate methods such as molecular mechanics force fields have been utilized instead. Molecular mechanics force fields involve classical mechanical force constants and classical electrostatics, but they do not invoke quantum mechanical wave functions and usually do not involve polarizability. Thus they have the advantage of being easy to implement and efficient to compute, but there

so these methods cannot be used to examine chemical reactions. Furthermore, while these methods may give good agreement with experiments for physical properties against which they have been parameterized, they often do not do well when tested on systems or properties outside of the parameterization set.

With increases in computer power, classical potentials have been replaced in many cases by the use of density functional theory (DFT) [1,2], which in some cases [3,4] is able to achieve accuracy comparable to or better than highly correlated levels of wave function theory (WFT), [5–7], such as coupled cluster theory with single and double excitations [8] (CCSD), or coupled cluster theory with single and double excitations and quasiperturbative connected triple excitations [9] (CCSD[T]), while maintaining much more favorable scaling with respect to system size. However, despite these advances the desire to be able to apply correlated levels of wave function theory (such as second-, third-, or fourth-order Møller–Plesset perturbation theory [10–13] or coupled cluster theory) to extended systems remains a goal for many computation researchers.

One of the largest problems standing in the way of this goal is the scaling of such methods with respect to increasing system size. The most basic WFT method, Hartree–Fock theory, has a computational effort that scales as N^4 where, throughout the whole chapter, N is the number of atoms in the system [5]. (Hybrid DFT also scales as N^4.) Hartree–Fock methods have been applied to very large systems; however, even though Hartree–Fock theory fully accounts for the exchange energy of the electrons of the same spin, the lack of dynamical correlation energy means that it does not provide sufficient accuracy for most applications of chemical interest. In order to obtain the accuracy needed, one should use correlated levels of electronic structure theory, but conventional correlated WFT calculations suffer from high-order scaling of the computational cost with respect to increasing system size and are generally impractical for very large systems. For example [5], MP2 scales as N^5, MP3, CISD, MP4SQD, CCSD, and QCISD scale as N^6, and MP4, CCSD(T), and QCISD(T) scale as N^7. The high-order scaling of these methods, as originally developed, has made them too expensive to be practical for use on extended systems without introducing strategies for coping with large size.

One way to overcome the expensive scaling is to introduce localized orbitals and computationally screen out excitations that involve widely spatially separated orbitals [14] (see also Chapters 1, 2, and 3). Another way to circumvent these problems is the introduction of fragment-based methods. Fragment-based methods are built on the assumption that one can divide the system into a set of monomers (which may be molecular fragments, single molecules, or collections of molecules) and express the total energy as a combination of the energies of the monomers, dimers, trimers, etc. These methods have the benefit of reducing a single very

expensive calculation to a large number of small calculations. However, the simplest of these methods, the many-body method, does not give good quantitative results until one includes third-order terms (for moderate accuracy) or higher-order terms (for high accuracy), and the cost to obtain the desired accuracy may be prohibitive. In order to increase the accuracy of these calculations at a low order (i.e., including only dimers or trimers) several groups have tried modifying these methods to include the effects of the electrostatic potential of the other fragments in the system, and a variety of fragment-based approaches have been proposed in the literature [15–27]. (See also Chapters 3, 6, and 7). The emphasis here is on our own methods for carrying out these types of calculations: first, the electrostatically embedded many-body (EE-MB) method [26], which is based on the many-body method but is simpler to implement than other fragment-based methods such as the fragment molecular orbital (FMO) method or the electrostatic-field-adapted molecular-fragment-with-conjugated-caps (EFA-MFCC) method, and second, the electrostatically embedded many-body expansion of the correlation energy (EE-MB-CE) [27].

To apply fragment methods to general systems, one must cut covalent bonds to make fragments, and then the dangling bonds must be capped by link atoms or more complicated strategies [15,16,18–23]. This is an important research goal, but the present article is focused mainly on systems where the division into fragments does not require cutting bonds. This includes molecular clusters, molecular liquids, and molecular solids.

5.2 *Many-body methods*

Consider a system of M interacting units, to be called monomers. These monomers may be covalently connected or noncovalently connected; however, as stated at the end of the introduction, in the present chapter we will focus mainly on monomers that are not covalently connected, and the equations apply only to this case. The total energy of this system, E, can be written, without approximation, as a series of m-body potentials

$$E = V_1 + V_2 + V_3 + \cdots + V_M \tag{5.1}$$

where

$$V_1 = \sum_{i}^{M} E_i \tag{5.2}$$

$$V_2 = \sum_{i<j}^{M} (E_{ij} - E_i - E_j) \tag{5.3}$$

$$V_3 = \sum_{i<j<k}^{M} [(E_{ijk} - E_i - E_j - E_k) - (E_{ij} - E_i - E_j) - (E_{ik} - E_i - E_k) - (E_{jk} - E_j - E_k)] \tag{5.4}$$

and higher-order terms are defined analogously. In these equations E_i represents the energy of each of the M monomers in the system, E_{ij} the energy of each of the $M(M-1)/2!$ dimers in the system, E_{ijk} the energy of each of the $M(M-1)(M-2)/3!$ trimers in the system, and so forth.

In the limit to which this expansion is carried, the M^{th} order, the energy returned is the total energy of the system. In order to invoke the many-body approximation, one makes the choice to truncate this series instead at the m^{th} order (where $m < M$). If one chooses to truncate Equation (5.1) after the second term ($m = 2$), the total energy of the system is approximated as

$$E_{PA} = \sum_{i<j}^{M} E_{ij} - (M-2)\sum_{i}^{M} E_i \tag{5.5}$$

where E_{PA} denotes the pairwise additive energy, and E_{ij} and E_i retain the same meanings as above. If one chooses to truncate after the third term ($m = 3$) the total energy of the system is approximated as

$$E_{3B} = \sum_{i<j<k}^{M} E_{ijk} - (M-3)\sum_{i<j}^{M} E_{ij} + \frac{(M-2)(M-3)}{2}\sum_{i}^{M} E_i \tag{5.6}$$

where E_{3B} denotes the three-body approximation to the total energy, and E_{ijk}, E_{ij}, and E_i retain the same meanings as above.

The benefit of making such an approximation is that for a large system it reduces a single large and expensive calculation to a large number of small and affordable calculations. Moreover, since each of these small calculations is independent of the others, each one may be run on a different processor, leading to a calculation that can easily be parallelized on a large number of processors. The accuracy of the energy depends on how many m-body terms are retained. While the pairwise additive method may provide qualitatively correct results, if one hopes to obtain quantitative accuracy, one must include higher-order terms, particularly if the monomers are known to have large intermolecular interactions, such those that occur among water molecules [28].

One obvious way to avoid the problem mentioned above would be to continue to include higher-order terms until a satisfactory level of accuracy is obtained. However, for large systems the number of such higher-order terms rises quickly. Table 5.1 shows the number of clusters (dimers through hexamers) that one would need to calculate for a system with 64 monomers. As seen in this table, the number of clusters one must calculate rises rapidly with increasing cluster size (roughly as $M^m/m!$). Moreover, the cost of each of these calculations rises, so that not only must you calculate a much larger number of these higher-order terms, but as the size of the cluster increases, so does the cost to calculate its energy.

Table 5.1 **Number of Clusters to Calculate for Cluster Sizes $m = 1 - 6$ for $M = 64$**

Cluster size (m)	Number of clusters
Monomer (1)	64
Dimer (2)	2016
Trimer (3)	41 664
Tetramer (4)	635 376
Pentamer (5)	7 624 512
Hexamer (6)	74 974 368

In order to consider the cost associated with carrying out a many-body expansion it is useful to compare the amount of time it would take to calculate the energy of the system with and without invoking the many-body approximation. The calculation of the energy without using the many-body expansion (i.e., a calculation of the energy of a supersystem containing all M monomers in the system at a given level of theory and with the given basis set) will be referred to as a *conventional calculation*. The calculation of energy of the same system using the pairwise additive approximation will be denoted PA, while the calculation using the three-body approximation will be denoted 3B. Table 5.2 shows a comparison of the theoretical timings for the calculation of the energies computed in the conventional manner and with the pairwise additive and three-body approximations, for methods that scale as aN^5, bN^6, and cN^7, a, b, and c are unknown prefactors specific to each level of electronic structure theory and basis set. One can see that even on a single processor the many-body methods are far more cost effective than conventional calculations. The pairwise additive calculations are between 5 and 7 orders of magnitude faster than a conventional calculation, while the three-body calculations are between 2 and 5 orders of magnitude faster. In fact, for the methods shown in Table 5.2, one may go up to the 5th order in the many-body expansion before the cost of the many-body expansion surpasses the cost of the conventional calculation.

Table 5.2 **Comparison of Hypothetical Timings for Full Calculations and Many-Body Calculations for a System Containing 64 Monomers**

Scaling	Conventional	PA	3B
aN^5	$1.1 \times 10^9 a$	$6.5 \times 10^4 a$	$1.0 \times 10^7 a$
bN^6	$6.9 \times 10^{10} b$	$1.3 \times 10^5 b$	$3.1 \times 10^7 b$
cN^7	$4.4 \times 10^{12} c$	$2.5 \times 10^5 c$	$9.1 \times 10^7 c$

5.3 Electrostatically embedded many-body methods

5.3.1 EE-MB

The electrostatically embedded many-body (EE-MB) methods have been presented previously [26], and a summary of the methods is given here. The EE-MB methods are an extension of the basic many-body idea. The key difference is that in the basic many-body methods, the energy of each monomer, dimer, or trimer is calculated in vacuum; in the EE-MB methods each monomer, dimer, or trimer is calculated in a field of point charges centered on the nuclear positions of the missing monomers. For example, in the electrostatically embedded pairwise additive (EE-PA) method, the energy of the system is written as

$$E_{EE-PA} = \sum_{i<j}^{M} E'_{ij} - (M-2)\sum_{i}^{M} E'_i \tag{5.7}$$

where E'_i are the energies of each monomer embedded in a field of nuclear-centered point charges representing the other $M-1$ monomers and E'_{ij} are the energies of each dimer embedded in a field of nuclear-centered point charges representing the other $M-2$ monomers. Similarly, in the electrostatically embedded three-body method, the energy of the system is written as

$$E_{EE-3B} = \sum_{i<j<k}^{M} E'_{ijk} - (M-3)\sum_{i<j}^{M} E'_{ij} - \frac{(M-2)(M-3)}{2}\sum_{i}^{M} E'_i \tag{5.8}$$

where E'_i and E'_{ij} have the same meaning as above and E'_{ijk} are the energies of each trimer embedded in a field of nuclear-centered point charges representing the other $M-3$ monomers.

Just as with the unembedded many-body expansions, if one does not truncate the EE-MB series the energy obtained is the total energy of the system. The presence of the point charges in the EE-MB method is meant to increase the rate of convergence of Equation (5.1) by incorporating the higher-order many-body effects in an average way. Work on clusters containing pure water [26–28] as well as mixed clusters containing ammonia and water [29], and water, ammonia, sulfuric acid, and ammonium and bisulfate ions [30] have shown the accuracy of the method to be relatively independent of the choice, within reason, of partial atomic charges (and if the expansion is not truncated the result is completely independent of the type of charges used), although in general it has been found that larger charges produce slightly better results. For simplicity, calculations have been carried out on gas-phase monomers (i.e., for water, a water molecule with an O-H bond length of 0.9572 Å and an H-O-H bond angle of 104.52°)

to determine the partial atomic charges for that species. One may use a variety of charge models including Mulliken, ChelpG, or CM4 charges, and those charges in turn are used to represent the missing monomers in the EE-MB calculations.

Recently, the sensitivity of the accuracy of the EE-MB methods has been examined with respect to a wide variety of charge models, including models in which the charges depend on the geometry of the system being studied [30]. While it was found that the use of these geometry-dependent charges did improve the accuracy of the methods very slightly, the small gain in accuracy is not worth the additional cost of calculating the geometry-dependent charges, and so it is recommended to use charges obtained from gas-phase monomers with the method and charge representation of your choice.

The process for calculating the EE-MB energy can then be summarized as follows:

1. Calculate the embedding charges for each monomer type in the system using gas-phase molecules and the electronic structure method and charge model of your choice.
2. For each m-body cluster in the expansion (for EE-PA $m = 1$ and 2; for EE-3B $m = 1$, 2, and 3) embed the cluster in a field of nuclear-centered point charges (as determined in step one) representing the other $M - m$ monomers and calculate the energy of each embedded cluster with the electronic structure method and basis set of your choice.
3. Calculate the total energy of the system using either Equation (5.7) (EE-PA) or Equation (5.8) (EE-3B).

5.3.2 EE-MB-CE

In practice, the EE-MB methods described in the previous section can be used in conjunction with any level of electronic structure theory including both WFT and DFT. However, an additional approximation can be made when using correlated levels of electronic structure theory such as MP2, CCSD, or CCSD(T). When a post-Hartree–Fock correlated level of theory is used the total energy can be written as

$$E_X = E_{HF} + \Delta E_{corr,X} \tag{5.9}$$

where E_X is the electronic energy of correlated method X (where X = MP2, CCSD, CCSD(T), etc.), E_{HF} is the Hartree–Fock energy of the system, and $\Delta E_{corr,X}$ is the correlation energy for method X. Because the total energy is a linear combination of the correlation energy and the Hartree–Fock energy, and since each term, V_m, in the many-body expansion is a

linear combination of energies for the 1- to m-body clusters, we can use Equation (5.9) to write V_m as

$$V_m = V_{m,\mathrm{HF}} + \Delta V_{m,\mathrm{corr}}. \tag{5.10}$$

A consequence of Equation (5.10) is that the energy of the system can be written as the sum of two different many-body expansions: one for the expansion of the Hartree–Fock energy and one for the expansion of the correlation energy, that is,

$$V = (V_{1,\mathrm{HF}} + V_{2,\mathrm{HF}} + V_{3,\mathrm{HF}} + \cdots) + (\Delta V_{1,\mathrm{corr}} + \Delta V_{2,\mathrm{corr}} + \Delta V_{3,\mathrm{corr}} \cdots). \tag{5.11}$$

It is not unreasonable to treat the expansion of the Hartree–Fock energy and the expansion of the correlation energy differently. For chemical systems without a significant amount of static correlation the Hartree–Fock energy often dominates even the changes in the electronic energy of the system, and therefore one expects that changes in Hartree–Fock energy will often be larger than changes in correlation energy for the process of interest, for example, for a binding event. In such cases, much of the error associated with the truncation of the EE-MB expansion is due to the truncation of the Hartree–Fock energy and not the correlation energy. This is not to imply that the error from the truncation of the correlation energy is not important, only that this error is likely to be considerably smaller, at a given order of expansion, than the error from truncating the Hartree–Fock energy. In such cases, one might truncate the expansion of the correlation energy at a lower level than the expansion of the Hartree–Fock energy.

The motivation for EE-MB-CE is strictly practical, based on the effect that such a procedure has on the cost of the calculation. Since Hartree–Fock theory has more favorable scaling than post-Hartree–Fock methods, when considering higher-order terms (such as three-body or four-body terms) one can afford to calculate larger clusters with Hartree–Fock theory than would be practical with post-Hartree–Fock methods. For moderate-sized systems (up to a few hundred atoms) one can, if desired, calculate the Hartree–Fock energy of the entire system (i.e., carry the many-body expansion to M^{th} order) and carry out a many-body expansion of only the correlation energy [27].

5.4 *Performance*

The accuracies of the EE-MB and EE-MB-CE methods were examined in [27] for a series of pure water clusters, ranging in size from 5 to 20 water molecules and taken from the Cambridge Cluster Database [31], at the MP2/jul-cc-pVTZ level of theory (where jul- denotes semidiffuse, so

Table 5.3 Comparison of Mean Errors (kcal/mol) at the MP2/jul-cc-pVTZ Level of Theory for Pure Water Clusters $(H_2O)_n$ with $n = 5 - 20$

	MSE	MUE	RMSE
PA	15.95	15.95	17.55
3B	0.55	0.56	0.71
EE-PA	0.80	0.80	0.84
EE-3B	−0.34	0.35	0.51
PA-CE	0.22	0.22	0.24
3B-CE	−0.05	0.17	0.24
EE-PA-CE	−0.10	0.10	0.11
EE-3B-CE	−0.23	0.23	0.34

the jul-cc-pVTZ basis set uses the aug-cc-pVTZ basis set [32,33] on oxygen and the cc-pVTZ basis set [34] on hydrogen). Eight different many-body methods were used: PA, 3B, EE-PA, EE-3B, PA-CE, 3B-CE, EE-PA-CE, and EE-3B-CE; each method has been described in Section 5.2 or 5.3. The energy calculated with each of these methods was compared to the energy of a conventional calculation on the same cluster at the same level of theory. The mean errors for each of the many-body methods, calculated over the complete data set, are shown in Table 5.3. The key results can be summarized as follows:

1. The inclusion of point charges dramatically improves the accuracy of the pairwise additive approximation. A comparison of the results of the PA method to the EE-PA method shows that the mean unsigned error is reduced from 15.95 kcal/mol to 0.80 kcal/mol. The average binding energy for these clusters was 105.48 kcal/mol, so an error of 0.80 kcal/mol corresponds to 0.8% of the average binding energy.
2. Even without inclusion of the embedding charges, the inclusion of the full Hartree–Fock energy with a pairwise additive treatment of the correlation energy (i.e., comparing PA to PA-CE) reduces the error by a factor of 78. The inclusion of embedding charges (i.e., comparing EE-PA to EE-PA-CE) shows an improvement of a factor of 8 in these errors. This gives errors that are as low as 0.1% of the average binding energy of these clusters by only including two body terms.

In an effort to make these calculations even more affordable (a more detailed discussion of the costs of these methods is in Section 5.6) [27] we also considered the effect of several different cutoffs on the calculation of the correlation energy. Correlation is typically short-ranged in comparison to Hartree–Fock theory, so for monomers separated by large distances it is

Table 5.4 Comparison of Mean Errors (kcal/mol) for PA-CE and EE-PA-CE Methods with a Cutoff of 6 Å and with No Cutoff

	6 Å			No Cutoff		
	MSE	MUE	RMSE	MSE	MUE	RMSE
PA-CE	0.38	0.38	0.46	0.22	0.22	0.23
EE-PA-CE	0.02	0.07	0.09	0.09	0.09	0.10

likely that the correlation would go to zero. A comparison of the mean errors for the PA-CE and EE-PA-CE methods with no cutoff and a cutoff of 6 Å is shown in Table 5.4. These results show that with a cutoff of 6 Å, one retains the accuracy of no cutoff for the EE-PA-CE method. The implications of this result on the cost of these calculations will be considered in the next section.

To assess the accuracy of the EE-MB and EE-MB-CE methods for other levels of correlated electronic structure theory, the relative energies of a series of low-lying water hexamers were used. The water hexamers were chosen for two reasons:

1. Due to the cost of the post-MP2 methods, which scale as N^6 and N^7, we were limited to clusters on the order of 5 heavy atoms.
2. There are 5 different water hexamers that lie within 3 kcal/mol of each other, making this system a good test of the capabilities and accuracies of these methods.

The following many-body methods were used: PA, PA-CE, EE-PA, EE-PA-CE, 3B, 3B-CE, EE-3B, and EE-3B-CE with the following levels of wavefunction theory: HF, MP2, MP3, MP4D, MP4DQ, MP4SDQ, MP4, CCSD, and CCSD(T) with the jul-cc-pVTZ basis set, that is, the aug-cc-pVTZ basis set on oxygen and the cc-pVTZ basis set on hydrogen. In the following discussion of errors, we are always referring to the error relative to a full calculation at the same level of theory. We found that the error of the many-body method was largely independent of the level of wave function theory used. Table 5.5 presents the average mean unsigned error (relative to the conventional calculation at each level of theory) and standard deviation over all nine levels of electronic structure theory for each of the many body methods. Examining this table shows that all of these methods have a very low standard deviation, indicating that the errors associated with the

Table 5.5 Average Mean Unsigned Errors (kcal/mol) and Standard Deviations for Many-Body Methods and Conventional Calculations at the Same Level of Theory for Water Hexamers

	PA	PA-CE	EE-PA	EE-PA-CE	3B	3B-CE	EE-3B	EE-3B-CE
Ave. MUE	11.77	0.10	1.00	0.14	1.24	0.16	0.12	0.03
Std. Dev.	0.06	0.01	0.05	0.05	0.03	0.03	0.01	0.01

many-body methods are consistent over all levels of electronic structure theory tested. This indicates that conclusions drawn about the accuracies of the many-body method at more affordable levels of theory (e.g., MP2) may be extrapolated to more expensive levels of theory (e.g., CCSD(T)).

If one is interested in trying to extrapolate results from one level of electronic structure theory to another, it bears mentioning that we have found that the results obtained by using the EE-MB methods with density functional theory have accuracies similar to those obtained when MP2 is used. While the focus of this review is on methods that are able to accurately calculate correlation energies for condensed-phase systems, the lower scaling of DFT (which scales as N^3 for local functionals, i.e., functionals without Hartree–Fock exchange, and N^4 for hybrid functionals, i.e., functionals with a nonzero percentage of Hartree–Fock exchange, compared to MP2, which scales as N^5) allows one to test the EE-MB expansions against larger and more complex clusters than would be possible using correlated levels of wave function theory. For example, using DFT, the EE-PA and EE-3B methods have been applied to systems containing 64 water molecules [35]; this would be out of reach even for MP2, which has the lowest scaling of any of the post-Hartree–Fock methods. Additionally, we have applied DFT to the study of mixed clusters containing ammonia and water [29] and also clusters containing water, ammonia, sulfuric acid, and ammonium and bisulfate ions [30]. In the case of the ammonia–water clusters we studied the relative energies of a series of five tetramers ($NH_3(H_2O)_3$), five pentamers ($NH_3(H_2O)_4$), and two hexamers ($NH_3(H_2O)_5$) and found that on average, the MP2 results had a slightly lower mean unsigned error over the data set than the DFT methods tested. We also found that the average difference between the mean unsigned error for the DFT results and the MP2 results never exceeded 0.1 kcal/mol per monomer at either the EE-PA level or the EE-3B level of theory. We have also applied the EE-PA and EE-3B methods to the study of water–ammonia–sulfuric acid clusters. In this study we compared the accuracy of the EE-PA and EE-3B methods over eight different clusters (containing either one or two sulfuric acid molecules, one ammonia molecule, and one to six water molecules) with three different basis sets and five different charge models. We found that the relative errors of the EE-PA methods were, in most cases, approximately 5% of the average binding energy of the clusters, and the relative errors of the EE-3B methods were less than 1% of the average binding energy of the clusters. When the relative absolute error was averaged over all eight configurations, three basis sets, and five charge models we found that the EE-PA methods had an average relative absolute error of 3.0% and the EE-3B method had an average relative absolute error of 0.3%. This study is of particular importance because it represents the first test of the EE-MB methods in systems in which ions and charge-transfer complexes were present, which may be important for condensed-phase studies.

5.5 Cost

One of the benefits of the EE-MB methods discussed in Section 5.3 is that they provide an increased accuracy over the traditional many-body methods discussed in Section 5.2, without increasing the costs shown in Table 5.5. While it is possible that the addition of the point charges to the system may affect the convergence of the self-consistent-field(SCF) iterations and therefore affect the cost of the calculation, in practice this change in cost is negligible, and so for a good approximation the cost of the EE-MB calculations is the same as the cost of a traditional many-body calculation. In the limit of large system size, the EE-PA method scales as M^2 and the EE-3B method scales as M^3 where M is the number of monomers in the system. This feature shows the promise of the EE-MB methods for use in conjunction with MP2 or CCSD(T), which have much less favorable scaling.

In considering the cost of the EE-MB-CE method, one must consider that these calculations may be done in multiple ways, depending on the order to which the expansion of the Hartree–Fock energy is carried out. For any size cluster in which the correlation energy is being calculated by post-Hartree–Fock methods (which, except for DFT, are the most common methods at present for calculating correlation energy), the Hartree–Fock energy must be calculated first and so in that respect the Hartree–Fock energy is obtained at no extra cost. The only additional cost above that of the EE-MB methods is the cost to calculate the Hartree–Fock energies for higher-order terms than are considered for the correlation energy. The extreme case would be a calculation in which the expansion of the Hartree–Fock energy is carried out to Mth order (i.e., a conventional Hartree–Fock calculation is carried out on the system). If one considers the cost to carry out an EE-PA-CE calculation at the MP2 level of theory for a system with 64 monomers, one will see that the cost to calculate the correlation energy is two orders of magnitude larger than the cost to calculate the Hartree–Fock energy of the system. Therefore, it is safe to assume that the cost of these calculations can be estimated to a good extent by the cost of the EE-PA and EE-3B calculations at the same level of theory.

The cost of the EE-MB and EE-MB-CE can be further reduced by the use of a cutoff distance in the dimer and trimer calculations. As discussed in Section 5.4, a cutoff can be used to reduce the number of dimer calculations that must be considered. In the case presented previously, approximately 44% of all dimers were able to be excluded, which would reduce the cost of the calculation by nearly the same percentage. While the precise value of the cutoff used would likely differ for each system, the time required to find this cutoff would be short in comparison to the time saved over the course of the simulation. While no study of the effect of a cutoff in the trimer calculations has been completed yet, it is likely that one could implement a similar cutoff to further reduce the cost of the calculation.

The last consideration in the cost of the EE-MB methods is the use of parallelization. Since the calculation of each monomer, dimer, and in the case of the EE-3B method, trimer, is completely independent of all other calculations, each could, in theory, be run on a different processor. The ability to spread these calculations out evenly over a large number of processors makes these calculations attractive.

5.6 *Use in simulations*

5.6.1 *Routes for extending EE-MB to the bulk*

In order to treat liquids or solids, it is necessary to eliminate surface effects by a choice of boundary conditions, such as stochastic boundary conditions, extended-wall boundary conditions, or periodic boundary conditions [36–38]. With periodic boundary conditions, each monomer interacts with all the monomers and embedding charges in all the replicated unit cells, including its own image, until convergence or until a cutoff distance is reached. For long-range forces, the sum over these interactions is only conditionally convergent, and so care must be used to obtain a physical result; a cutoff can cause artifacts. The long-range forces are more important for some properties than for others [39]; for example, short-range structure and dynamics may depend primarily on short-range forces. Various methods such as reaction fields, multipole summation, or Ewald summation can be used to sum the infinite series of interactions or approximate their effect [36–40]. For liquids the periodicity of the images is artificial, and so some workers consider it more appropriate to employ the nearest image convention in which a given monomer interacts only with the nearest from among another given monomer or its various images [36]; however, it has been recommended that while this might be acceptable in Monte Carlo simulations, it should never be used in molecular dynamics simulations [38]. A common choice for a cutoff distance, for a cubic unit cell, is less than or equal to half the width of the cell. When this is applied to a many-body treatment, one should therefore also screen the oligomer contributions and include only those where no two monomers are separated by more than the cutoff distance. This eliminates ambiguities in the three-body contributions as well as the two-body ones [41].

The application of periodic boundary conditions to EE-MB energies can be carried out in much the same way as it is carried out in simulations involving quantum mechanical/molecular mechanical (QM/MM) calculations. In this respect, the largest difference between EE-MB and QM/MM calculations is that care must be taken in an EE-MB calculation in choosing the correct dimer and trimer combinations to include in the energy calculation. When long-range interactions (such as Coulomb or dipole–dipole interactions) are not cut off, one might require an adaptive scheme [42]

to switch between the original dimer or trimer and one involving images when the distance of given monomer to an interacting monomer becomes larger than the distance between one of the monomers and the image of the other.

Ewald sums are commonly used with molecular mechanics potentials, but applying them to EE-MB is more similar to using them with combined QM/MM potentials. The total energy for a QM/MM calculation can be written as

$$E(\text{QM/MM}) = E(\text{MM}) + E(\text{QM}) + V(\text{QM/MM}) \tag{5.12}$$

where E(MM) is the energy of the molecular mechanics region, E(QM) is the energy of the quantum mechanical region, and V(QM/MM) is the energy due to the interaction between the QM and MM regions. Calculations on the QM region are usually carried out with a background charge distribution of molecular mechanics charges. This is called electronic embedding and is similar to the type of embedding done in the EE-MB methods; and therefore each embedded monomer, dimer, or trimer calculation can be seen as a simplified QM/MM calculation where the embedding charges are the molecular mechanics region. A difference between a QM/MM calculation and an EE-MB calculation is that an EE-MB calculation does not include the interactions of the point charges with one another. Thus one needs V(QM/MM) but not E(MM). There are a number of examples in the literature in which periodic boundary conditions have been applied to QM/MM calculations, but almost all of these [43–46] are for semiempirical QM methods. Recently an algorithm suitable for ab initio QM methods has been presented [47].

Another way to sum the infinite series of interactions in periodic ab initio QM calculations is the fast multipole method (FMM), which can handle non-cubic unit cells [48,49].

Another way to account for the effect of a bulk solvent on a simulated active site is provided by a class of solvent boundary potentials, which have also been applied to QM/MM calculations and which could be reformulated for EE-MB calculations [50–52].

Motivated in part by the hybrid QM:QM scheme of Sauer and coworkers [53,54], we propose another approach to extending the many-body expansion to solids. Their scheme involves a periodic DFT calculation with a local functional on the extended system and a correction to higher-level WFT based on a series of larger and larger mechanically embedded clusters; this avoids evaluating Hartree–Fock exchange on the periodic extended system, which is very expensive. To gain this advantage in the context of the present methods, we propose a method with some similarity to the EE-MB-CE method. We call this new suggestion DFT:EE-MB-HL where HL denotes higher level. In DFT:EE-MB-HL, one would carry out a periodic

DFT calculation with a local functional for the extended system and augment this with an EE-PA or EE-3B calculation of the difference (for dimers or for dimers and trimers) between a higher-level calculation (e.g., a hybrid DFT calculation or a coupled cluster calculation) and a DFT calculation with the local functional. This approach allows one to take advantage of the many efficient periodic codes for local functionals and the expected fast convergence of an EE-MB expansion of the higher-level correction. Note that for general solids this may require the extension mentioned at the end of Section 5.1.

5.6.2 *Monte Carlo simulations*

The use of many-body-based methods in Monte Carlo simulations has been discussed previously by Christie and Jordan [55] and in this section we will summarize their findings for unembedded many-body methods and extend the discussion to EE-MB methods. (See also the description of such methods in Chapter 7.) A key issue in Monte Carlo calculations on clusters, liquids, and amorphous solids, is that many Monte Carlo moves involve a change in only one monomer. The present discussion is limited to that kind of move.

For a many-body method that does not involve electrostatic embedding (e.g., PA, 3B, PA-CE, 3B-CE), when one makes a Monte Carlo move in which only one molecule is displaced, the only interaction terms that must be recalculated are those involving the displaced molecule. Therefore, instead of recalculating the energies of all $M(M-1)/2$ dimers, one only needs to recalculate the energy of $M-1$ dimers. Similarly, instead of recalculating all $M(M-1)(M-2)/6$ trimers, one needs only to recalculate the energy of $(M-1)(M-2)/2$ trimers.

Without using many-body expansions, the cost of each Monte Carlo move is largely determined by the scaling with respect to increased system size of the electronic structure method used. As mentioned previously, for large systems MP2 scales as N^5, while more expensive methods such as MP4 and CCSD(T) scale as N^7. For small systems, the scaling is less severe, but in this section all discussion is carried out for systems large enough that asymptotic scaling applies. While the asymptotic limit is never actually reached in real work, the use of this limit facilitates a general discussion that illustrates the key ideas.

Table 5.6 shows a series of hypothetical timings, for the calculation of a many-body (MB) Monte Carlo move for a system with 64 monomers, for methods with costs (computational efforts) that scale as aN^5, bN^6, and cN^7, where a, b, and c are unknown prefactors for each level of electronic structure theory (the prefactor also depends on the basis set). The cost to do a pairwise additive (PA) Monte Carlo move is 6 to 9 orders of magnitude smaller than calculating the energy in the conventional manner, while the

Table 5.6 Comparison of Hypothetical Timings for a Monte Carlo Move on a System of 64 Monomers Using the PA or 3B Methods

Scaling	Conventional	PA move	3B move
aN^5	$1.1 \times 10^9 a$	$2.0 \times 10^3 a$	$4.8 \times 10^5 a$
bN^6	$6.9 \times 10^{10} b$	$4.0 \times 10^3 b$	$1.4 \times 10^6 b$
cN^7	$4.4 \times 10^{12} c$	$8.1 \times 10^3 c$	$4.4 \times 10^6 c$

cost to do a three-body (3B) calculation is 4 to 6 orders of magnitude smaller than calculating the energy in the conventional manner. In general, for a scaling proportional to N^n, the ratio of the cost of a 3B move to a full calculation is $0.5(3^n)N^{2-n}$. Furthermore, since each of the calculations in the MB move is independent of the rest, the MB calculations may be highly parallelized and so if one has access to a large number of processors the wall clock time for the calculation may be decreased significantly. In addition, as mentioned in Sections 5.4 and 5.5, for large systems, a cutoff can be used to decrease the number of dimers and trimers calculated, eventually making the method scale linearly with the number of processors.

For a many-body method that *does* involve electrostatic embedding (e.g., EE-PA, EE-3B, etc.), individual m-mer energies depend not only on the coordinates of the nuclei belonging to the m-mer itself but also on the coordinates of the embedding charges. Thus, when any atom in the system is moved, the energy of every possible embedded m-mer is at least slightly affected and must in principle be recalculated in order to obtain the true EE-MB energy of the system.

However, one of the advantages of the EE-MB method is that it affords ample opportunities to explore and implement additional cost-saving approximations, especially when the EE-MB energies are used during Monte Carlo simulations, which do not require analytic gradients and therefore do not require a perfectly continuous potential energy surface as long as the "jumps" in the surface are small. We already discussed the use of spatial cutoffs to reduce the number of dimer calculations by screening out dimers or trimers where one monomer is separated by a distance greater than the cutoff. The dimers and trimers that must be recalculated because the embedding charges have moved offer additional possibilities for screening or approximation. For example, prior to performing an ab initio calculation of the energy of an embedded m-mer at some given step in a Monte Carlo simulation, one could estimate by first-order perturbation theory how much the movement of the embedding charge changes the m-mers energy. If the change is smaller than some threshold, one can accept the perturbation theory estimate rather than recalculating the energy of the m-mer in the new configuration of embedding charges. In this way, one could use an energy

cutoff to screen out relatively expensive monomer, dimer, and trimer calculations that can be adequately treated by perturbation theory. Multilevel strategies could also be implemented to save time during Monte Carlo simulations. For example, one could perform all monomer and dimer calculations at a high level of theory and then perform all trimer calculations at some lower level of theory to obtain an estimate of the EE-3B energy. Alternatively, one could perform all m-mer calculations at the same level of theory but only calculate the trimer energies at certain reference steps during the Monte Carlo simulations; for intermediate steps, the EE-3B energy could be approximated by adding the difference between the EE-PA energy at the current step and that at a recent reference step to the EE-3B energy at the recent reference step. Or the change in trimers due to movement of faraway embedding charges could be calculated with a smaller basis set than is used for the strongly coupled trimers. (The EE-MB-CE method could also be considered to be a multilevel strategy.) A myriad of possibilities suggest themselves. Thus, the use of screening techniques, perturbation theory, and/or multilevel strategies can yield tremendous savings at many steps during the course of a Monte Carlo simulation without losing a significant amount of the accuracy gained by the inclusion of embedding charges.

5.6.3 *Molecular dynamics*

In a molecular dynamics simulation, the molecules are allowed to evolve following Newton's laws of motion. This requires calculating the forces acting on the molecules. To do this efficiently, one seeks an algorithm for analytical gradients of the potential energy surface. Since the gradient is a linear operator, applying it to the EE-PA and EE-3B energies (Equations [5.8] and [5.7]) gives [35]

$$\nabla E_{EE-PA} = \sum_{i<j}^{M} \nabla E'_{ij} - (M-2)\sum_{i}^{M} \nabla E'_{i} \tag{5.13}$$

and

$$\nabla E_{EE-3B} = \sum_{i<j<k}^{M} \nabla E'_{ijk} - (M-3)\sum_{i<j}^{M} \nabla E'_{ij} + \frac{(M-2)(M-3)}{2}\sum_{i}^{M} \nabla E'_{i}. \tag{5.14}$$

These equations were first presented in [35], but we later discovered a bug in the program. This bug has been corrected in MBPAC-2009 [56]. As a result, analytic gradients are available for any method that has analytic gradients for the individual m-mers provided that the computer program allows for the use of fractionally charged point charges to be used

as psuedonuclei (sometimes called *ghost atoms* or *sparkles*) in the gradient calculation. It is important to note that all of the terms on the right-hand side of Equations (5.12) and (5.13) contribute to all of the components of the gradients. For example, $\nabla E_i'$, $\nabla E_{ij}'$, and $\nabla E_{ijk}'$ all contribute to all the gradient components of all the other monomers in the system and not just to monomers i, j, and k. A key simplification in the EE-MB methods is that the magnitudes of the point charges are fixed and therefore do not need to be updated during the course of the simulation.

As a numerical example, we consider hydrogen fluoride tetramer. The geometry used for these calculations is given in Table 5.7 and the calculations were carried out using the MBPAC-2009 program [56]. Table 5.7 shows the energy of the tetramer relative to four monomers from full calculations and from calculations by EE-PA and EE-3B at three levels of WFT. The errors in binding energies are 0.60–1.6% at the EE-PA level and reduce to 0.15–0.27% at the EE-3B level. Table 5.8 shows the unsigned errors in the gradient magnitudes as well as mean and maximum unsigned errors in the gradient components for EE-PA and EE-3B at this geometry. The gradient magnitude is 0.44–0.52 hartrees per bohr, and the mean absolute value of the 24 Cartesian components of the gradient is 0.071–0.082 hartrees per bohr, depending on the level of the quantum mechanical theory. It can be seen that, in comparison with the gradient magnitudes or gradient components, the errors are generally 3 orders of magnitude smaller in EE-PA and 4 orders of magnitude smaller in EE-3B calculations.

Table 5.7 Energies (kcal/mol) of Hydrogen Fluoride Tetramer Relative to Four Monomers from Conventional Calculations and from Many-Body Methods at the Same Level of Theory[a]

	Conventional	EE-PA	EE-3B
HF/MIDI!	29.36	29.19	29.32
MP2/cc-pVTZ	17.22	17.49	17.18
CCSD/cc-pVTZ	16.87	17.09	16.84

[a] The Cartesian coordinates (in Å) of the tetramer are

F	1.346092	1.346092	0.000000
F	−1.346092	1.346092	0.000000
F	−1.346092	−1.346092	0.000000
F	1.346092	−1.346092	0.000000
H	0.631976	1.763006	0.000000
H	−1.763006	0.631976	0.000000
H	−0.631976	−1.763006	0.000000
H	1.763006	−0.631976	0.000000

The Cartesian coordinates (in Å) of the monomer are

F	1.346092	1.346092	0.000000
H	0.631976	1.763006	0.000000

Table 5.8 **Unsigned Errors in Gradient Magnitudes (hartree/bohr), Mean Unsigned Errors, and Maximum Unsigned Errors in Gradient Components (hartree/bohr) by Many-Body Methods with Respect to Conventional Calculations for Hydrogen Fluoride Tetramer. (The Geometries Are Specified in Table 5.7.)**

	Gradient magnitudes	
	EE-PA	EE-3B
HF/MIDI!	2.7×10^{-3}	2.0×10^{-4}
MP2/cc-pVTZ	5.8×10^{-4}	4.0×10^{-5}
CCSD/cc-pVTZ	7.1×10^{-4}	3.0×10^{-5}

	Gradient components			
	Mean		Maximum	
	EE-PA	EE-3B	EE-PA	EE-3B
HF/MIDI!	2.0×10^{-4}	2.5×10^{-5}	5.4×10^{-4}	5.5×10^{-5}
MP2/cc-pVTZ	5.3×10^{-5}	8.6×10^{-6}	1.5×10^{-4}	2.3×10^{-5}
CCSD/cc-pVTZ	5.3×10^{-5}	6.8×10^{-6}	1.4×10^{-4}	1.9×10^{-5}

One can also write a set of equations similar to Equations (5.13) and (5.14) in which the Hessian has been applied to the EE-MB energies. As a result, the EE-MB methods will also have analytic Hessians for any correlated level of theory that has analytic Hessians. This is important for studying vibrations and phonons.

In contrast to Monte Carlo simulations, the energy of the entire system must be recalculated after each time step of a molecular dynamics calculation, so even with the use of the MB methods the cost of these calculations will be quite high for a large system. However, the scaling with system size is reduced to N^2 for EE-PA and N^3 for EE-3B even without screening, and the method offers numerous possibilities for screening out dimers and trimers to reduce the scaling eventually to linear. Since conventional hybrid DFT scales as N^4, an EE-3B CCSD(T) calculation becomes less expensive than a conventional hybrid DFT calculation for large N.

5.7 Conclusions

The ability to accurately calculate correlation energies for condensed-phase systems remains an important goal of computational scientists. The electrostatically embedded many-body (EE-MB) method and the electrostatically embedded many-body expansion of the correlation energy (EE-MB-CE) method provide promising new routes to achieve this goal. The high accuracy of these methods coupled with their simple implementation, low cost, and easy parallelization make them attractive options for

use on condensed-phase systems. While the EE-MB and EE-MB-CE methods have not been tested on true condensed-phase systems yet, work on large molecular clusters shows that they are able to reproduce the binding energies of these clusters to within 1% of the true energy obtained by a conventional calculation on the system, and this review discusses how the methods can be used in conjunction with available computational methodology to make them applicable to both molecular dynamics and Monte Carlo simulations of condensed-phase systems. The implementation of the EE-MB and EE-MB-CE methods in existing simulation packages and the application of these methods to interesting condensed-phase problems provide interesting and exciting possibilities for further research.

Acknowledgments

This material is based upon work supported by the National Science Foundation under grant no. CHE07-04974. H. Lin was a long-term visitor of the 2008–2009 Mathematics and Chemistry program in the Institute of Mathematics and Its Applications at the University of Minnesota.

References

[1] P. Hohenberg and W. Kohn, Inhomogeneous electron gas. *Phys. Rev.* 136, B864 (1964).

[2] W. Kohn and L. J. Sham, Self-consistent equations including exchange and correlation effects. *Phys. Rev.* 140, A1133 (1965).

[3] Y. Zhao and D. G. Truhlar, Density functionals with broad applicability in chemistry. *Acc. Chem. Res.* 41, 157 (2008).

[4] J. Zheng, Y. Zhao, and D. G. Truhlar, The DBH24/08 database and its use to assess electronic structure model chemistries for chemical reaction barrier heights. *J. Chem. Theory Comput.* 5, 808 (2009).

[5] K. Raghavachari and J. B. Anderson, Electron correlation effects in molecules. *J. Phys. Chem.* 100, 12960 (1996).

[6] M. Head-Gordon, Quantum chemistry and molecular processes. *J. Phys. Chem.* 100, 13213 (1996).

[7] J. A. Pople, Nobel lecture: Quantum chemical models. *Rev. Mod. Phys.* 71, 1267 (1999).

[8] J. Čížek, On the use of the cluster expansion and the technique of diagrams in calculations of correlation effects in atoms and molecules. *Adv. Chem. Phys.* 14, 35 (1969).

[9] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* 157, 479 (1989).

[10] R. Krishnan and J. A. Pople, Approximate fourth-order perturbation theory of the electron correlation energy. *Int. J. Quant. Chem.* 14, 91 (1978).

[11] R. Krishnan, M. J. Frisch, and J. A. Pople, Contribution of triple substitutions to the electron correlation energy in fourth order perturbation theory. *J. Chem. Phys.* 72, 4244 (1980).

[12] M. Frisch, R. Krishnan, and J. Pople, A systematic study of the effect of triple substitutions on the electron correlation energy of small molecules. *Chem. Phys. Lett.* 75, 66 (1980).

[13] G. F. Adams, G. D. Bent, and R. J. Bartlett, Calculation of potential energy surfaces for HCO and HNO using many-body methods, in *Potential energy surfaces and dynamics calculations*, edited by D. G. Truhlar, pages 133–167 (Plenum, New York, 1981).

[14] S. Saebø and P. Pulay, A low-scaling method for second order Møller–Plesset calculations. *J. Chem. Phys.* 115, 3975 (2001).

[15] K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayashi, Fragment molecular orbital method: an approximate computational method for large molecules. *Chem. Phys. Lett.* 313, 701 (1999).

[16] V. Deev and M. A. Collins, Approximate *ab initio* energies by systematic molecular fragmentation. *J. Chem. Phys.* 122, 154102/1 (2005).

[17] S. Hirata, M. Valiev, M. Dupuis, S. S. Xantheas, S. Sugiki, and H. Sekino, *Mol. Phys.* 103, 2255 (2005).

[18] X. H. Chen and J. Z. H. Zhang, Molecular fractionation with conjugated caps density matrix with pairwise interaction correction for protein energy calculation. *J. Chem. Phys.* 125, 44903/1 (2006).

[19] N. Jiang, J. Ma, and Y. Jiang, Electrostatic field-adapted molecular fractionation with conjugated caps for energy calculations of charged biomolecules. *J. Chem. Phys.* 124, 114112/1 (2006).

[20] D. M. Fedorov and K. Kitaura, Theoretical development of the fragment molecular orbital (FMO) method. in *Modern methods for theoretical physical chemistry of biopolymers*, edited by E. B. Starikov, J. P. Lewis, and S. Tanaka, pages 3–38 (Elsevier, Amsterdam, 2006).

[21] W. Li, S. Li, and Y. Jiang, Generalized energy-based fragmentation approach for computing the ground-state energies and properties of large molecules. *J. Phys. Chem. A* 111, 2193 (2007).

[22] D. G. Fedorov and K. Kitaura, Extending the power of quantum chemistry to large systems with the fragment molecular orbital method. *J. Phys. Chem. A* 111, 6904 (2007).

[23] W. Xie, L. Song, D. G. Truhlar, and J. Gao, The variational explicit polarization potential and analytical first derivative of energy: Towards a next generation force field. *J. Chem. Phys.* 128, 234108/1 (2008).

[24] S. Hirata, Fast electron-correlation methods for molecular crystals: An application to the α, β_1, and β_2 modifications of solid formic acid. *J. Chem. Phys.* 129, 204104/1 (2008).

[25] E. E. Dahlke and D. G. Truhlar, Assessment of the pairwise additive approximation and evaluation of many-body terms for water clusters. *J. Phys. Chem. B* 110, 10595 (2006).

[26] E. E. Dahlke and D. G. Truhlar, Electrostatically embedded many-body expansion for large systems, with applications to water clusters. *J. Chem. Theory Comput.* 3, 46 (2007).

[27] E. E. Dahlke and D. G. Truhlar, 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.* 3, 1342 (2007).

[28] M. J. Elrod and R. J. Saykally, Many-body effects in intermolecular forces. *Chem. Rev.* 94, 1975 (1994).

[29] A. Sorkin, E. E. Dahlke, and D. G. Truhlar, Application of the electrostatically embedded many-body expansion to microsolvation of ammonia in water clusters. *J. Chem. Theory Comp.* 4, 683 (2008).

[30] H. R. Leverentz and D. G. Truhlar, Electrostatically embedded many-body approximation for systems of water, ammonia, and sulfuric acid and the dependence of its performance on embedding charges. *J. Chem. Theory Comp.* 5, 1573 (2009).

[31] D. J. Wales, J. P. K. Doye, A. Dullweber, M. P. Hodges, F. Y. Naumkin, F. Calvo, J. Hernández-Rojas, and T. F. Middleton, The Cambridge cluster database, http://www-wales.ch.cam.ac.uk/CCD.html, 2006, accessed March 8, 2006.

[32] R. A. Kendall, T. H. Dunning Jr., and R. J. Harrison, Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* 96, 6796 (1992).

[33] D. E. Woon and T. H. Dunning Jr., Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* 98, 1358 (1993).

[34] T. H. Dunning Jr., Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* 90, 1007 (1989).

[35] E. E. Dahlke and D. G. Truhlar, Electrostatically embedded many-body expansion for simulations. *J. Chem. Theory Comput.* 4, 1 (2008).

[36] W. F. van Gunsteren, Molecular dynamics and stochastic dynamics: A primer, in *Computer simulation of biomolecular systems*, edited by W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson, volume 2, pages 3–36 (ESCOM, Leiden, 1993).

[37] D. C. Rapaport, *The art of molecular dynamics simulation* (Cambridge University Press, Cambridge, 1995).

[38] D. Frenkel and B. Smit, *Understanding molecular simulation* (Academic Press, San Diego, 2002).

[39] H. Hu and W. Yang, Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods. *Annu. Rev. of Phys. Chem.* 59, 573 (2008).

[40] H. M. Senn and W. Thiel, QM/MM methods for biological systems, in *Atomistic approaches in modern biology: From quantum chemistry to molecular simulations*, Topics in Current Chemistry, Vol. 268, edited by M. Reiher, pp. 173–290 (Springer, Berlin, 2007).

[41] P. Attard, Simulation results for a fluid with the Axilrod-Teller triple dipole potential. *Phys. Rev. A* 45, 5649 (1992).

[42] A. Heyden, H. Lin, and D. G. Truhlar, Adaptive partitioning in combined quantum mechanical and molecular mechanical calculations of potential energy functions for multiscale simulations, *J. Phys. Chem. B* 111, 2231 (2007).

[43] J. Gao and C. Alhambra, A hybrid semiempirical quantum mechanical and lattice-sum method for electrostatic interactions in fluid simulations. *J. Chem. Phys.* 107, 1212 (1997).

[44] D. Riccardi, P. Schaefer, and Q. Cui, pKa calculations in solution and proteins with QM/MM free energy perturbation simulations: A quantitative test of QM/MM protocols. *J. Phys. Chem. B* 109, 17715 (2005).

[45] K. Nam, J. Gao, and D. M. York, An efficient linear-scaling ewald method for long-range electrostatic interactions in combined QM/MM calculations. *J. Chem. Theory Comp.* 1, 2 (2005).

[46] R. C. Walker, M. F. Crowley, and D. A. Case, The implementation of a fast and accurate QM/MM potential method in Amber. *J. Comp. Chem.* 29, 1019 (2008).

[47] T. Laino, F. Mohamed, A. Laio, and M. Parrinello, An efficient linear-scaling electrostatic coupling for treating periodic boundary conditions in QM/MM simulations. *J. Chem. Theory Comp.* 2, 1370 (2006).

[48] K. N. Kudin and G. Scuseria, A fast multipole algorithm for the efficient treatment of the Coulomb problem in electronic structure calculations of periodic systems with Gaussian orbitals. *Chem. Phys. Lett.* 289, 611 (1998).

[49] K. N. Kudin and G. E. Scuseria, Analytic stress tensor with the periodic fast multipole method. *Phys. Rev. B* 61, 5141 (2000).

[50] W. Im, S. Berněche, and B. Roux, Generalized solvent boundary potential for computer simulations. *J. Chem. Phys.* 114, 2924 (2001).

[51] B. A. Gregersen and D. M. York, Variational electrostatic projection (VEP) methods for efficient modeling of the macromolecular electrostatic and solvation environment in activated dynamics simulations. *J. Phys. Chem. B* 109, 536 (2005).

[52] P. Schaefer, D. Riccardi, and Q. Cui, Reliable treatment of electrostatics in combined QM/MM simulation of macromolecules. *J. Chem. Phys.* 123, 014905/1 (2005).

[53] C. Tuma and H. Sauer, A hybrid MP2/planewave-DFT scheme for large chemical systems: proton jumps in zeolites. *Chem. Phys. Lett.* 387, 388 (2004).

[54] S. Svelle, C. Tuma, X. Rozanska, T. Kerber, and J. Sauer, Quantum chemical modeling of zeolite-catalyzed methylation reactions: toward chemical accuracy for barriers. *J. Amer. Chem. Soc.* 131, 816 (2009).

[55] R. A. Christie and K. D. Jordan, N-body decomposition approach to the calculation of interaction energies of water clusters, in *Intermolecular forces and clusters II*, Structure and Bonding series, series edited by D. M. P. Mingos, Volume 116, pages 27–41 (Springer, Berlin, 2005).

[56] E. Dahlke, H. Lin, H. Leverentz, and D. G. Truhlar, MBPAC-2009, 2009, University of Minnesota, Minneapolis, URL: http://comp.chem.umn.edu/mbpac/.