THE JOURNAL OF PHYSICAL CHEMISTRY Letters

Letter

Cite This: J. Phys. Chem. Lett. 2019, 10, 3877−3886

pubs.acs.org/JPCL

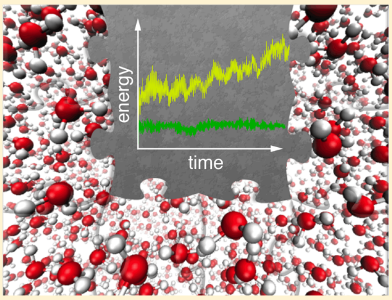
Variational Formulation of the Generalized Many-Body Expansion with Self-Consistent Charge Embedding: Simple and Correct Analytic Energy Gradient for Fragment-Based *ab Initio* Molecular Dynamics

Jie Liu,[†] Bhaskar Rana, Kuan-Yu Liu, and John M. Herbert*

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

Supporting Information

ABSTRACT: The many-body expansion (MBE) and its extension to overlapping fragments, the generalized (G)MBE, constitute the theoretical basis for most fragment-based approaches for large-scale quantum chemistry. We reformulate the GMBE for use with embedding charges determined self-consistently from the fragment wave functions, in a manner that preserves the variational nature of the underlying self-consistent field method. As a result, the analytic gradient retains the simple "sum of fragment gradients" form that is often assumed in practice, sometimes incorrectly. This obviates (without approximation) the need to solve coupled-perturbed equations, and we demonstrate stable, fragment-based *ab initio* molecular dynamics simulations using this technique. Energy conservation fails when charge-response contributions to the Fock matrix are neglected, even while geometry optimizations and vibrational frequency calculations may yet be accurate. Stable simulations can be recovered by means of straightforward modifications introduced here, providing a general paradigm for fragment-based *ab initio* molecular dynamics.



Fragment methods[1−3] provide a practical and intuitive means to reduce the computational scaling of *ab initio* quantum chemistry. The common idea amongst these methods is decomposition of a large (super)system into smaller subsystems. Numerous subsystem electronic structure calculations, which are trivially parallelizable, are then used to approximate the total energy or other properties of the supersystem. An appealing application of these methods is *ab initio* molecular dynamics (AIMD) of liquid water, where fragment-based approaches have recently been used to perform simulations at correlated wave function levels of theory.[4−7]

In an attempt to improve accuracy by capturing many-body polarization, fragment methods often embed the subsystem calculations in a classical electrostatic representation of the entire supersystem, using atomic point charges derived from the fragment wave functions.[8−14] In principle, these embedding charges should be iteratively updated in order to reflect changes in the fragment wave functions;[8−11] however, self-consistency significantly complicates the formulation of analytic energy gradients by introducing response terms that describe how the embedding charges respond to changes in the wave functions.[9,10] In practice these response terms are often neglected,[8,11,13,15−18] meaning that the putative fragment-based analytic gradient is formally incorrect insofar as it is not the derivative of the corresponding fragment-based energy expression. It has been argued that these terms are numerically small,[13] and their neglect has been justified *a posteriori* by noting that fragment-based geometry optimizations and harmonic frequency calculations accurately approximate supersystem results.[13,15−17,19−21]

There do exist fragment-based methods for which correct analytic gradients have been reported with self-consistent electrostatic embedding.[9,10,22−25] However, these methods are formulated in such a way that the embedding compromises the variational nature of the underlying self-consistent field (SCF) method, and therefore a correct analytic gradient requires solution of coupled-perturbed equations, which is ordinarily not required for SCF first derivatives.[26] A notable example is the fragment molecular orbital (FMO) method,[27−31] the most widely used fragment-based approach. First proposed in 1999,[32] with an approximate analytic gradient reported soon thereafter,[33] various pieces of the gradient were developed subsequently,[34−38] until finally in 2011 the exact analytic gradient was reported.[22−24] Notably, inclusion of the response terms is found to improve energy conservation in AIMD simulations.[39] That said, the long delay in reporting the exact analytic gradient for FMO is a testament to its complexity and motivates our attempt to simply the formalism. In the present work, we introduce an alternative to FMO that satisfies a variational principle and therefore obviates the need to solve coupled-perturbed equations, without sacrificing a rigorously correct energy gradient.

Received: April 29, 2019
Accepted: June 6, 2019
Published: June 6, 2019

DOI: 10.1021/acs.jpclett.9b01214
J. Phys. Chem. Lett. 2019, 10, 3877−3886

Most fragment-based approaches, including FMO, rely at some level upon a many-body expansion (MBE) of the supersystem energy:[40,41]

$$E = \sum_{A} E_A + \sum_{A<B} \Delta E_{AB} + \sum_{A<B<C} \Delta E_{ABC} + \cdots \tag{1}$$

The first term is the sum of fragment energies and the subsequent terms represent n-body corrections, e.g., for $n = 2$:

$$\Delta E_{AB} = E_{AB} - E_A - E_B \tag{2}$$

Truncation of eq 1 at n-body terms forms the basis of the FMOn method,[42,43] but other methods fall within this framework as well.[40] These include Stoll's "method of increments"[44] as well as electrostatically embedded (EE)-MBE methods proposed independently by Hirata et al.,[45,46] by Truhlar and co-workers,[47,48] and by Beran et al.[49−52]

Many other fragment-based methods can be understood within the framework of a generalized (G)MBE developed by our group.[14,40,53,54] The GMBE extends eq 1 to the case where the supersystem is tessellated into overlapping fragments A,B,C, ... while also introducing new fragments $A \cap B$, $A \cap B \cap C$, ... in order to avoid double counting. At the one-body ($n = 1$) level, this leads to an energy expression

$$E = \sum_{A} E_A - \sum_{A<B} E_{A\cap B} + \sum_{A<B<C} E_{A\cap B\cap C} - \cdots \tag{3}$$

We call this approach GMBE(1),[40,54] but a variety of other methods can be understood as applications of eq 3. These include Li's "generalized energy-based fragmentation" (GEBF) method;[8,15−18,21] Gadre's "molecular tailoring approach" (MTA);[55−60] Zhang's "molecular fractionation with conjugated caps" (MFCC) approach;[61−63] Raghavachari's "molecules-in-molecules" (MIM) method;[64−67] and the "systematic molecular fragmentation" (SMF) method developed by Collins and co-workers.[13,19,68] The GEBF, MTA, and MFCC methods use precisely the energy formula in eq 3,[40,64] whereas SMF and MIM include a subset of the intersections.[40,53] These interconnections do not seem to be widely appreciated.

In this work, we compute subsystem energies in the presence of atomic point charges representing the rest of the supersystem. If these charges are fixed *a priori*, then the gradient $\mathrm{d}E/\mathrm{d}x$ obtained via term-by-term differentiation of eq 1 or 3 is a simple combination of subsystem gradients, but the situation is more complicated if the charges are updated on-the-fly using the fragment wave functions. Often these embedding charges are simply inserted as "external" charges, in the spirit of a mixed quantum mechanics/molecular mechanics (QM/MM) approach, without modifying the underlying SCF procedure. The result is that the SCF procedure no longer satisfies a variational property.[9,22,40,46] This fact has not always been recognized,[8,11,15−18] leading to incorrect gradients. Note that we *can* express the exact gradient (of eq 3, for example) in a sum-over-fragments form,

$$\frac{\mathrm{d}E}{\mathrm{d}x} = \sum_{A} \frac{\mathrm{d}E_A}{\mathrm{d}x} - \sum_{A<B} \frac{\mathrm{d}E_{A\cap B}}{\mathrm{d}x} + \cdots \tag{4}$$

but quantities such as $\mathrm{d}E_A/\mathrm{d}x$ must be interpreted as *total* derivatives of E_A with respect to perturbation of x. For embedding charges that depend on the fragment wave functions, $\mathrm{d}E_A/\mathrm{d}x$ is *not* equivalent to the result of a straightforward gradient calculation on subsystem A.

Assume henceforth that all calculations are performed at an SCF level of theory. The energy of fragment K can be written in the form

$$E^K_{\mathrm{SCF}} = E^K_0 + E^K_{\mathrm{emb}} \tag{5}$$

where $E^K_0[\mathbf{P}^K]$ denotes the SCF energy functional and E^K_{emb} is the classical embedding energy. Denoting the Fock matrix for isolated fragment K as $\mathbf{f}^K = \delta E^K_0/\delta \mathbf{P}^K$, the Fock matrix $\mathbf{F}^K = \delta E_{\mathrm{SCF}}/\delta \mathbf{P}^K$ for charge-embedded fragment K is[69]

$$F^K_{\mu\nu} = f^K_{\mu\nu} - \frac{1}{2}\sum_{j\notin K} q_j(\mathbf{I}_j)_{\mu\nu} + \sum_{k\in K} \frac{\partial E_{\mathrm{emb}}}{\partial q_k}\frac{\partial q_k}{\partial P^K_{\mu\nu}} \tag{6}$$

where q_k is a point charge located at $\mathbf{R}_k$ and

$$(\mathbf{I}_k)_{\mu\nu} = \int \frac{g_\mu(\mathbf{r})g_\nu(\mathbf{r})}{\|\mathbf{r} - \mathbf{R}_k\|}\,\mathrm{d}\mathbf{r} \tag{7}$$

represents the electrostatic potential generated by the basis-function pair $g_\mu(\mathbf{r})\,g_\nu(\mathbf{r})$. The *total* derivative of E^K_{SCF} with respect to x is[70]

$$\frac{\mathrm{d}E^K_{\mathrm{SCF}}}{\mathrm{d}x} = \frac{\partial E^K_0}{\partial x} + \sum_{\mu\nu} F^K_{\mu\nu}\frac{\partial P^K_{\mu\nu}}{\partial x} + \frac{\mathrm{d}E_{\mathrm{emb}}}{\mathrm{d}x} \tag{8}$$

The first term $(\partial E^K_0/\partial x)$ is the Hellmann−Feynman contribution to the SCF gradient, which involves derivatives of one- and two-electron integrals. The second term can be rewritten as $-\mathbf{W}^K(\partial \mathbf{S}^K/\partial x)$,[26] where $\mathbf{W}^K = \mathbf{P}^K\mathbf{F}^K\mathbf{P}^K$ is the energy-weighted density matrix and $\mathbf{S}^K$ is the overlap matrix. Together, these two terms constitute the usual SCF energy gradient.

The final term in eq 8 is

$$\frac{\mathrm{d}E_{\mathrm{emb}}}{\mathrm{d}x} = \frac{\partial E_{\mathrm{emb}}}{\partial x} + \sum_{j\notin K} \frac{\partial E_{\mathrm{emb}}}{\partial q_j}\frac{\partial q_j}{\partial x} \tag{9}$$

The second term in this equation involves the response of the embedding charges to perturbation of x, and it is these terms that are neglected if the derivatives in eq 4 are naïvely replaced by a gradient calculation for each subsystem. Collins[13] has argued that the magnitude of the neglected response terms is small and thus they are negligible. Numerical comparisons using the FMO analytic gradient suggest that the total contribution arising from the response terms is $\sim 10^{-4}$ au,[71] which is indeed comparable to a typical gradient-based stopping criterion for geometry optimizations. However, neglect of the response terms is thought to be the origin of the failure to conserve energy that is observed in some AIMD simulations using fragment-based methods.[39,72−74]

Whatever the magnitude of the response terms, we argue that neglecting them is philosophically apart from neglect of terms in the (G)MBE and should not be considered as just another aspect of the fragment-based approximation. Whatever approximations or assumptions may have gone into the design and formulation of a fragment-based energy expression (based on eq 1 or 3, for example), our view is that once this *ansatz* for the total energy has been established, one should insist that the analytic energy gradient be the exact derivative of this energy expression.

That said, calculation of the response terms in the FMO method does complicate its gradient, requiring solution of fragment-based coupled-perturbed equations.[9,10,22−25] We avoid this complexity, without sacrificing correctness of the gradient, by formulating a variational version of EE-(G)MBE, whose exact analytic gradient is formally simple and computationally facile yet where embedding charges are nevertheless computed self-consistently from the fragment wave functions. The key to doing so is to use the explicit polarization ("XPol") method[75,76] to derive Fock matrices for the fragments based on application of the variational theorem to an embedded SCF energy expression. In fact, this is precisely the mathematics that leads from the fragment SCF energy expression in eq 5 to the Fock matrix defined in eq 6.[69]

Our new approach can be cast in a general form, applicable to any EE-(G)MBE method, by recognizing that all such methods express the total energy as a sum of subsystem energy calculations,

$$E = \sum_{K} p_K E_{\mathrm{SCF}}[\rho_K] \tag{10}$$

with certain coefficients p_K. For example, the conventional n-body expansion based on disjoint fragments, which we call MBE(n), corresponds to an energy expression[41]

$$\begin{aligned} E^{\mathrm{MBE}(n)} = {} & \sum_{m=0}^{n-1} (-1)^m \binom{N-n-1+m}{m} \\ & \times \sum_{K=1}^{\binom{N}{n-m}} E_K^{\mathrm{MBE}(n-m)} \end{aligned} \tag{11}$$

In this case, p_K is a combinatorial factor that multiplies $E_K^{\mathrm{MBE}(n-m)}$ and accounts for the fact that this particular ($n - m$)-body subsystem appears numerous times through order n in eq 1.

To extend this to the GMBE, we need to consider the set-theoretical inclusion/exclusion principle, which states that the cardinality of a union

$$S = \bigcup_{n=1}^{N} S_n \tag{12}$$

of sets $\{S_n\}$, which may or may not be disjoint, can be expressed in terms of cardinalities of their mutual intersections:

$$\begin{aligned} |S| &= \sum_{i=1}^{N} |S_i| - \sum_{i<j}^{N} |S_i \cap S_j| + \sum_{i<j<k}^{N} |S_i \cap S_j \cap S_k| - \cdots \\ &= \sum_{n=1}^{N} (-1)^{n+1} \sum_{i_1<i_2<\cdots<i_n} |S_{i_1} \cap S_{i_2} \cap \cdots \cap S_{i_n}| \end{aligned} \tag{13}$$

This principle has been used as heuristic motivation for the MTA method,[55,56] but in fact it underlies the GEBF, SMF, and MIM approaches as well,[40,64] and can be used to derive eq 3 based on a partition of the supersystem's Hamiltonian.[40,53,54]

The total density of the supersystem can be expressed as a linear combination of subsystem densities, since the latter form something analogous to a "cover", in the topological sense:[14]

$$\rho(\mathbf{r}) \approx \sum_{n=1}^{N} (-1)^{n+1} \sum_{i_1<i_2<\cdots<i_n} \rho_{i_1\cap i_2\cap\cdots\cap i_n}(\mathbf{r}) \tag{14}$$

This is simply a more general version of the approximation used for the FMO density,[29]

$$\rho_{\mathrm{FMO2}}(\mathbf{r}) = \sum_{I}^{N} \rho_I(\mathbf{r}) + \sum_{I>J}^{N} [\rho_{IJ}(\mathbf{r}) - \rho_I(\mathbf{r}) - \rho_J(\mathbf{r})] \tag{15}$$

Similarly, we take eq 14 to define the GMBE approximation to the supersystem's density, and for convenience we replace the subsystem index $i_1 \cap i_2 \cap \cdots \cap i_n$ with a single index K, à la eq 10:

$$\rho_{\mathrm{GMBE}}(\mathbf{r}) = \sum_{K} p_K \rho_K(\mathbf{r}) \tag{16}$$

Coefficients $p_K = \pm 1$ depend on the number of primitive fragment intersections needed to generate subsystem K. The total embedding charge on atom k, as deduced from eq 14, is[14]

$$\begin{aligned} q_k^{\mathrm{GMBE}} &= \sum_{n=1}^{N} (-1)^{n+1} \sum_{i_1<i_2<\cdots<i_n} q_k \delta_{k\in i_1\cap i_2\cap\cdots\cap i_n} \\ &= \sum_{K} p_K \delta_{k\in K} q_k \end{aligned} \tag{17}$$

Here, q_k is the charge on atom k in a particular subsystem, and the δ-functions are equal to 1 or 0 depending on whether k belongs to the indicated subsystem or not.

In principle we could use the fragment densities to generate a Coulomb embedding potential, as is sometimes done in FMO calculations,[29] but in the present work we pursue a point-charge embedding instead. This is done also in GEBF,[8] MFCC,[11] SMF,[13] and MTA[12] calculations. (In practice, FMO calculations often use a hierarchical embedding consisting of density embedding at short range and point-charge embedding at longer range. Inclusion of exchange is not recommended.[29]) For point charges, the embedding contribution to fragment K's energy is

$$E_{\mathrm{emb}}^{K} = \frac{1}{2} \sum_{j\notin K} \left(\sum_{k\in K} \frac{q_j Z_k}{\|\mathbf{R}_j - \mathbf{R}_k\|} - \frac{1}{2} \int \frac{q_j \rho_K(\mathbf{r})}{\|\mathbf{r} - \mathbf{R}_j\|}\, d\mathbf{r} \right) \tag{18}$$

The original XPol method used Mulliken charges for the $\{q_j\}$,[75] although other definitions are possible including Löwdin charges,[77] Hirshfeld charges,[78] and charges derived from the electrostatic potential.[69,70,79]

Individual subsystem energies in eq 10 are evaluated according to eq 5. By requiring that the embedded SCF energy E_{SCF}^{K} be stationary with respect to variation of ρ_K, one obtains an SCF equation for fragment K:[69,75,77]

$$\mathbf{F}^K \mathbf{C}^K = \mathbf{S}^K \mathbf{C}^K \boldsymbol{\varepsilon}^K \tag{19}$$

The fragment Fock matrices are given by eq 6, the form of which guarantees that the SCF procedure is variational.[69,75,77]

As an aside, we note that the fragment Fock matrices used in FMO have the form[29]

$$F_{\mu\nu}^{\mathrm{QM/MM},K} = f_{\mu\nu}^{K} - \sum_{j\notin K} q_j (\mathbf{I}_j)_{\mu\nu} \tag{20}$$

We call this a "QM/MM-style" Fock matrix because it is identical to that used when external point charges are introduced into an SCF calculation. As compared to the XPol Fock matrix in eq 6, the QM/MM-style Fock matrix omits charge-response terms of the form

$$(\boldsymbol{\Lambda}_k)_{\mu\nu} = \frac{\partial q_k}{\partial P^K_{\mu\nu}} \tag{21}$$

along with a factor of 1/2 in the charge−density interaction. The FMO energy expression is not variational and therefore its analytic gradient requires solution of coupled-perturbed equations.[22−24]

Returning to the variational Fock matrix $\mathbf{F}^K$, note that the derivatives $\partial E_{\text{emb}}/\partial q_k$ in eq 6 are easy to evaluate using eq 18. The remaining term in $F^K_{\mu\nu}$ involves the charge-response derivatives defined in eq 21. These can only be evaluated once we have specified how the embedding charges will be obtained from the fragment densities. In the case of Mulliken charges,

$$q_k = Z_K - \sum_{\mu\in K}\sum_{\nu} S_{\mu\nu}P_{\mu\nu} \tag{22}$$

the result is[77]

$$(\boldsymbol{\Lambda}_k)_{\mu\nu} = -S_{\mu\nu}\delta_{\mu\in K} \tag{23}$$

where $\delta_{\mu\in K} = 1$ or 0 depending on whether the basis function g_μ is located on fragment K or not. Formulas for Löwdin and for Hirshfeld embedding charges are also simple,[77,78] whereas those for "CHELPG" charges[80] (obtained from the molecular electrostatic potential, evaluated on a grid) are significantly more complicated.[69,70,79]

Operationally, our variational formulation of EE-(G)MBE requires a self-consistent XPol calculation for each subsystem K, in order to determine the embedding charges. The flowchart in Figure 1 outlines the method. A similar idea was used by Gao and Wang in the context of energy decomposition analysis using the MBE,[81] though no simulations were performed; it has not previously been suggested for GMBE-based methods that exploit overlapping fragments. The self-consistent XPol iterations can be performed very efficiently because an excellent initial guess is always available based on a superposition of one-body wave functions, therefore once the gas-phase fragment wave functions have been computed, the subsequent fragment SCF procedures usually converge in 2 or 3 cycles.

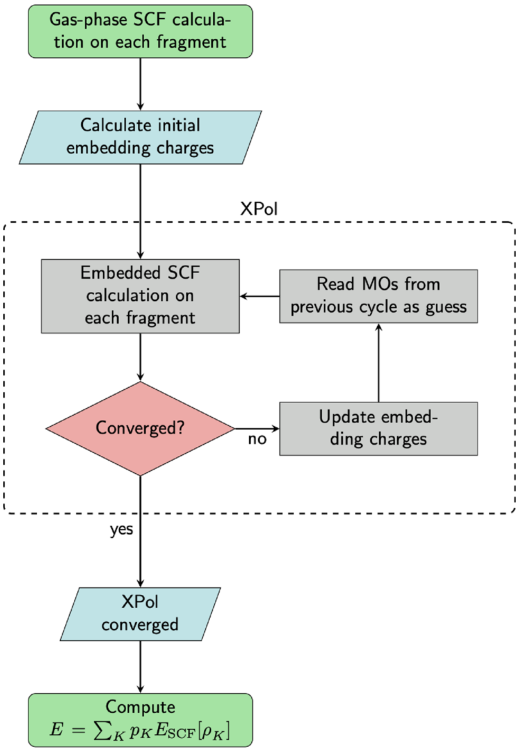


Figure 1. Flowchart for the variational EE-GMBE approach.

The derivative of the EE-(G)MBE energy expression (eq 10) can be evaluated term-by-term, i.e., subsystem-by-subsystem. For each subsystem, the two contributions to the total derivative dE/dx are

$$\frac{\mathrm{d}E_0}{\mathrm{d}x} = \frac{\partial E_0}{\partial x} + \sum_{\mu\nu}\frac{\partial E_0}{\partial P_{\mu\nu}}\frac{\partial P_{\mu\nu}}{\partial x} \tag{24}$$

and

$$\frac{\mathrm{d}E_{\text{emb}}}{\mathrm{d}x} = \frac{\partial E_{\text{emb}}}{\partial x} + \sum_{\mu\nu}\frac{\partial E_{\text{emb}}}{\partial P_{\mu\nu}}\frac{\partial P_{\mu\nu}}{\partial x} \tag{25}$$

The partial derivatives $\partial E_0/\partial x$ and $\partial E_{\text{emb}}/\partial x$ indicate derivatives of any *explicit* dependence on x. These are the Hellmann−Feynman contributions to the gradient, containing derivatives of the one- and two-electron integrals. The gradient of the total energy for fragment K is then given by eq 8, whose first two terms are equivalent to a traditional SCF gradient for subsystem K except that the Fock matrix is the modified one in eq 6. To this must be added the gradient of the embedding potential (eq 9), the key aspect of which is an explicit expression $\partial q_k/\partial x$ for how the embedding charges change as the nuclei are perturbed. The same derivatives are needed to perform Ewald summation in QM/MM simulations,[70,79] and appropriate expressions can be found in ref 70.

The variational EE-(G)MBE method and its analytic gradient have been implemented in a modified version of Q-Chem.[82] Tests versus finite-difference calculations confirm the validity of the gradient; see Tables S1 and S2. The method is applicable to any approach that combines a linear combination of subsystem SCF calculations to obtain the total energy (eq 10). This includes methods based on the traditional MBE but also those based on the inclusion/exclusion principle such as MTA, MFCC, SMF, and "level 1" of the MIM approach, each of which uses a variant of the energy formula in eq 3.[40] We have previously called this approach "GMBE(1)" to emphasize this generality along with the fact that this approach can be extended to GMBE(n).[14,40,54] In fact, GMBE(2) proves to be essentially exact for many noncovalent clusters,[54,83] and a pared-down approximation to GMBE(2) affords good results for proteins.[14] Only GMBE(1) is considered here, in conjunction with Mulliken embedding charges.

Our general philosophical approach to fragment-based quantum chemistry is to understand the mechanics of these methods first, before making pragmatic approximations. This systematic approach has led us to question the efficacy of low-order MBE(n) methods,[41,83−85] at least in the absence of some kind of supersystem calculation to capture high-order induction. (The latter approach has been used with success by Beran et al.[50,52] and by Raghavachari and co-workers.[65−67]) GMBE(n) represents a possible alternative. This approach generates far more subsystem calculations as compared to the traditional MBE, and we choose to begin conservatively with an exhaustive fragmentation scheme, before introducing screening approximations[85,86] that promise to make the calculations more practical.

Test systems include the clusters $(H_2O)_{32}$ and $F^-(H_2O)_{31}$. As in previous work,[40,53,54] these are first divided into "groups" corresponding to a single H_2O or F^-. We then loop through all groups, constructing fragments out of any groups that lie within a specified distance cutoff (ξ) from one another. Redundant fragments are eliminated, and then intersections of nonredundant fragments are added as additional fragments, per eq 3. This manner of constructing fragments has been used in the GEBF approach as well,[8,87−89] and for $(H_2O)_{32}$ a cutoff of ξ = 3.0 Å results in 169 fragments (including intersections) that contain 15−22 water molecules each. A single subsystem calculation thus constitutes a sizable portion of the full cluster, but this fact alone does not guarantee energy conservation in AIMD simulations when charge-response contributions to the gradient are neglected.

Before examining AIMD simulations, we first validate the EE-GMBE(1) approximation to the supersystem density, eq 14. Figure 2 shows an isocontour plot of the difference between the supersystem density for $(H_2O)_{10}$ and the GMBE approximation to it ($\rho - \rho_{\text{GMBE}}$), using an isocontour value that is typically used for evaluating molecular size.[90] Differences between the two densities are extremely small, as confirmed (and more easily quantified) by comparing Mulliken charges obtained from either density. The mean absolute difference (MAD) between the supersystem and GMBE-based Mulliken charges is 0.004 electrons, and the maximum difference is 0.0105 electrons; see Table S3. Results are just as good for a $F^-(H_2O)_{10}$ cluster (Table S4), with a MAD of just 0.002 electrons and a maximum deviation of 0.077 electrons. While various studies using the GEBF, MTA, MFCC, SMF, and MIM1 methods have evaluated the accuracy of what is effectively the EE-GMBE(1) approximation for the energy, this example suggests that the overlapping-fragment approach also affords an accurate approximation to the total supersystem density. This in turn suggests that EE-GMBE(1) might be useful for properties other than energy, such as polarizabilities,[8,17] vibrational intensities,[15] NMR chemical shifts,[65,91,92] and other response properties.[66,93−95]

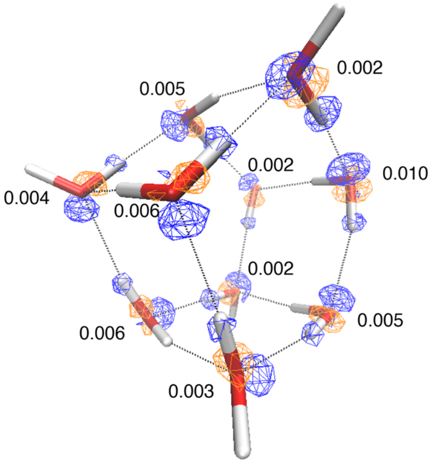


Figure 2. Difference density ($\rho - \rho_{\text{GMBE}}$) for $(H_2O)_{10}$ computed at the B3LYP/6-31G* level of theory and plotted with a contour value of 0.001 au. Also shown are the differences in Mulliken charges on the oxygen atoms, comparing those computed from ρ to those computed using ρ_{GMBE} (eq 16). The cluster was fragmented using ξ = 3.0 Å, resulting in 12 fragments with 6 water molecules per fragment.

The focus of this work is the GMBE analytic gradient. To evaluate the importance of possessing a proper gradient, we face the peculiar task of deciding what is the *right* way to implement the *wrong* gradient. The only reasonable answer, in our view, is to test variants that resemble other methods that have been reported in the literature. This suggests several options. For example, we could omit the charge-response contributions to the Fock matrix in eq 6, using instead a Fock matrix

$$\tilde{F}^K_{\mu\nu} = f^K_{\mu\nu} - \frac{1}{2}\sum_{j\notin K} q_j(\mathbf{I}_j)_{\mu\nu} \tag{26}$$

This resembles the QM/MM-style Fock matrix in eq 20 but retains the factor of 1/2 in the charge−density interaction that arises naturally (from the variational principle) in the XPol Fock matrix. We will refer to the modification where $\tilde{\mathbf{F}}^K$ replaces $\mathbf{F}^K$ as the "modified Fock matrix" version of EE-GMBE(1). In addition, we might also omit the response terms from the gradient of the embedding potential in eq 25, taking $dE_{\text{emb}}/dx \approx \partial E_{\text{emb}}/\partial x$. We call this second degradation of the formalism the "modified Fock, modified gradient" version of EE-GMBE(1). Both modifications invalidate the variational nature of the method.

To facilitate rapid comparison of variational and nonvariational strategies, the remaining calculations in this work use the semiempirical "HF-3c" method,[96] which we have implemented in Q-Chem for this purpose. HF-3c combines a minimal-basis Hartree−Fock calculation with simple, parametrized corrections for dispersion, basis-set superposition error, and basis-set incompleteness. Despite its simplicity and low cost, this method nevertheless affords reasonable molecular geometries[96] and noncovalent interaction energies[78,96,97] for large molecules and supramolecular complexes.

Water clusters are challenging test cases for fragment-based geometry optimizations due to the existence of numerous local minima that may differ only in the orientation of one or more hydrogen-bonded O−H moieties, so that small differences in the gradient between two different approaches might be enough to lead the optimization down a rather different path. This is *not* the case, however, when our rather conservative fragmentation strategy is applied to $(H_2O)_{32}$. Figure 3 plots total energy versus optimization cycle for both supersystem and fragment-based geometry optimizations of this cluster. Three independent geometry optimizations are reported in Figure 3 but in no case do the total energies stray further than ~0.3 mE_h from one another. In this particular case, charge-response terms in EE-GMBE(1) are apparently unimportant.

We explore this further in Table 1, which presents the results of geometry optimizations and vibrational frequency calculations on both $(H_2O)_{32}$ and $F^-(H_2O)_{31}$, comparing supersystem and EE-GMBE(1) results, including both variational and nonvariational versions of the latter. The final, optimized geometries differ by $\lesssim$2 mE_h in energy despite having been optimized by different methods with (slightly) different gradients. Vibrational frequencies for the optimized structures are almost identical across methods, with mean deviations of only 1−4 cm^{-1} and maximum deviations of ~50 cm^{-1} in the case of $F^-(H_2O)_{31}$. Frequencies were obtained by finite difference of analytic gradients because we have not implemented analytic Hessians for the XPol method. Previous work suggests that the finite-difference error is likely <1 cm^{-1}.[98]

DOI: 10.1021/acs.jpclett.9b01214
J. Phys. Chem. Lett. 2019, 10, 3877−3886

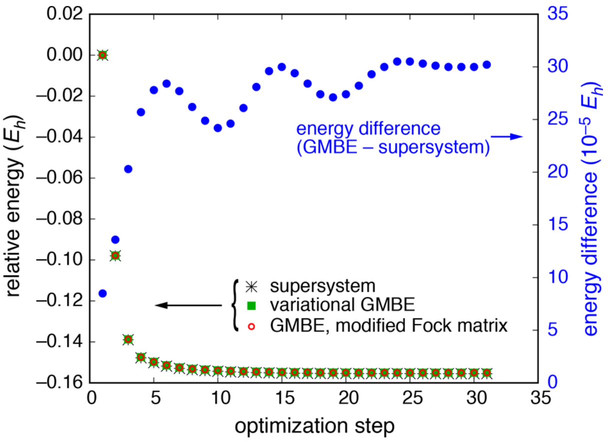


Figure 3. Geometry optimizations for $(H_2O)_{32}$ at the HF-3c level using the variational EE-GMBE(1) method and a nonvariational analogue thereof, in comparison to a traditional supersystem optimization. The axis on the left shows the total energy for each of the three methods, with a supersystem calculation at the initial geometry taken to define zero. Blue symbols, which should be read from the scale on the right, indicate the difference between the supersystem energy and the variational EE-GMBE(1) energy at geometries obtained from the supersystem optimization.

The importance of correct analytic gradients is unclear from these examples, but the situation is dramatically different for AIMD simulations. We examine simulations for the same systems considered above, noting that $(H_2O)_{32}$ was also used to evaluate energy conservation in FMO-based simulations, both with and without response terms in the gradient.[39] Figure 4 depicts energy fluctuations obtained from microcanonical AIMD simulations whose initial velocities are consistent with $T = 150$ K. (All trajectories were propagated using the velocity Verlet algorithm with a time step of 1.0 fs and an SCF convergence criterion of 10^{-7} E_h, which has been shown to provide good energy conservation.[99]) The variational EE-GMBE(1) method exhibits root-mean-square (RMS) energy fluctuations of ≈1 mE_h for $(H_2O)_{32}$, with no sign of energy drift. The two nonvariational approaches, on the other hand, exhibit slightly larger fluctuations and, more importantly, a clear energy drift over just a few picoseconds of simulation time. The problem is especially severe for the "modified Fock" variant that uses $\tilde{\mathbf{F}}^K$ in eq 26 but leaves intact the gradient dE_{emb}/dx in eq 9. For this method, the energy drift for $F^-(H_2O)_{31}$ amounts to almost 30 mE_h in just the first picosecond of dynamics! The version of EE-GMBE(1) that neglects charge-response contributions to both the Fock matrix and the gradient exhibits smaller (albeit still noticeable) energy drift, suggesting that the "modified Fock" scheme may represent something of an unbalanced approximation.

Of the methods considered in Figure 4, the "modified Fock, modified gradient" version of EE-GMBE(1) is arguably closest to what has most often been used in fragment-based methods other than FMO, including GEBF, SMF, and MFCC. However, the factor of 1/2 that appears in the charge−density term in $\mathbf{F}^K$ (eq 6), and which we have retained in the modified Fock matrix $\tilde{\mathbf{F}}^K$ of eq 26, is difficult to infer in the absence of a variational principle. Most electrostatic embedding schemes use the full electrostatic potential instead, which in the absence of charge-response terms corresponds to the QM/MM-style Fock matrix of eq 20. We have therefore tested additional modifications to the EE-GMBE(1) approach in which $\mathbf{F}^{\text{QM/MM},K}$ is used for the Fock matrix. In these "QM/MM-style" variants, we also adjust the electrostatic embedding term in the gradient, multiplying it by a factor of 2 as compared to what is used in the variational EE-GMBE(1) method. Note that this is *not* simply a fragment-based QM/MM calculation because the embedding charges change in each SCF cycle.

The performance of these QM/MM-style variants is examined for $(H_2O)_{32}$ in Figure 5. Modification of the Fock matrix alone leads to even larger energy drift as compared to the previous simulations, but energy conservation is greatly improved upon modifying the gradient as well. In order to obtain this level of energy conservation, however, it is necessary to incorporate an *ad hoc* factor of 2 into the charge-embedding gradient. This is not ordinarily done in fragment-based approaches as there is no apparent reason to do so, in the absence of an understanding of the formalism derived herein.

We have proposed a variational formulation of the generalized MBE that uses electrostatic embedding charges that are iterated to self-consistency using the "XPol" SCF procedure.[75] This ensures that the fragment-based total energy expression satisfies a variational principle, which in turn greatly simplifies the formulation of analytic energy gradients. The

Table 1. Comparison of Harmonic Frequencies Computed at Optimized Geometries by EE-GMBE(1) Methods[a]

	minimized energy		frequency errors (cm^{-1})		
method	absolute[b] (E_h)	relative[c] (mE_h)	MAD	std. dev.	max
$\mathbf{(H_2O)_{32}}$					
variational GMBE	−2417.00571859	0.463	1.9	2.5	21.1
GMBE (using $\tilde{\mathbf{F}}$)	−2417.00567628	0.505	2.1	2.3	16.3
GMBE ($\tilde{\mathbf{F}}$ and modified gradient)	−2417.00568608	0.496	1.8	2.2	19.8
GMBE (using $\mathbf{F}^{\text{QM/MM}}$)	−2417.00553605	0.646	1.8	2.4	20.8
GMBE ($\mathbf{F}^{\text{QM/MM}}$ and modified gradient)	−2417.00556923	0.612	1.0	1.2	11.0
supersystem	−2417.00618173	0.000	–	–	–
$\mathbf{F^-(H_2O)_{32}}$					
variational GMBE	−2440.27297135	2.290	3.3	5.9	53.4
GMBE (using $\tilde{\mathbf{F}}$)	−2440.27292684	2.246	4.1	6.1	48.1
supersystem	−2440.27068107	0.000	–	–	–

[a] HF-3c level, using $\xi = 3.0$ Å for $(H_2O)_{32}$ and $\xi = 4.0$ Å for $F^-(H_2O)_{31}$. [b] Computed using the method indicated in the left-most column, at geometries obtained from the same method. [c] Final energy obtained from each method, with the supersystem result taken to define zero.

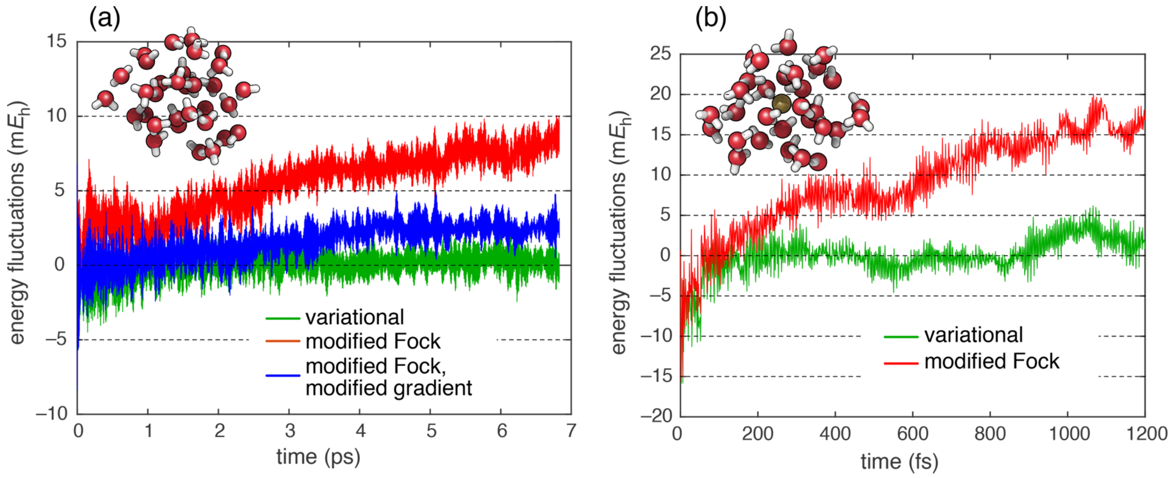


Figure 4. Energy fluctuations for EE-GMBE(1) AIMD simulations of (a) $(H_2O)_{32}$ and (b) $F^-(H_2O)_{31}$ at the HF-3c level. Simulations were performed under microcanonical conditions with initial velocities sampled from a Maxwell−Boltzmann distribution at T = 150 K. The "modified Fock" scheme uses $\tilde{\mathbf{F}}^K$ (eq 26), and the "modified gradient" scheme neglects charge-response terms in eq 9. The zero of energy is chosen to be the average energy from the variational GMBE simulation.

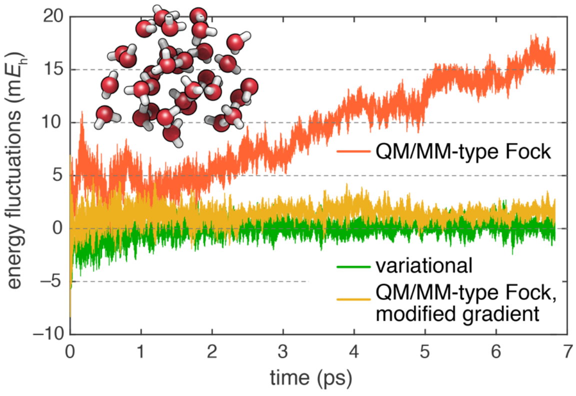


Figure 5. Energy fluctuations for EE-GMBE(1) AIMD simulations of $(H_2O)_{32}$ at the HF-3c level using the QM/MM-style Fock matrix of eq 20. Simulations were performed under microcanonical conditions with initial velocities sampled from a Maxwell−Boltzmann distribution at T = 150 K. The zero of energy is chosen to be the average energy from the variational GMBE simulation.

gradient reported here is rigorously correct despite the absence of any coupled-perturbed equations to compute orbital-response terms. By means of this procedure, a fragment approximation to the total density is variationally optimized in order to obtain the fragment-based approximation to the total energy. Notably, the GMBE(1) energy expression is similar (or in some cases equivalent) to energy expressions used in many contemporary fragment-based quantum chemistry procedures, including those based on overlapping fragments. These include the GEBF,[18] SMF,[68] MIM1,[64] MFCC,[61,63] and MTA[59] methods. In almost all cases where self-consistent embedding charges have been used in these methods, the requisite charge-response terms have been omitted from the gradient. Putative analytic gradients are therefore not equal to derivatives of the fragment-based energy expression.

Comparing results of our variational EE-GMBE(1) approach to several nonvariational alternatives, we find that charge-response terms make little difference when it comes to geometry optimizations and vibrational frequency calculations. This is consistent with the fact that the omitted terms are comparable in magnitude to the gradient-based stopping criteria used for geometry optimizations in quantum chemistry programs. It also explains how previous nonvariational approaches are able to obtain good agreement with supersystem calculations for forces[88] and vibrational frequencies,[13,15,17,19−21] despite the use of incorrect gradients. This favorable performance, however, is insufficient to guarantee that these methods are appropriate for use in AIMD simulations, as has sometimes been assumed.[88,100] Indeed, failure of energy conservation in EE-GEBF simulations has been documented previously,[74] and the present work provides compelling evidence that this failure is a direct result of using a gradient that is not the derivative of the energy. Small errors that may be negligible in the search for a local minimum evidently accumulate in a simulation.

Energy conservation is restored by means of the straightforward modifications described herein, which add only a small amount of additional overhead in order to achieve mutual self-consistency of the fragment wave functions and embedding charges. In return, one obtains a variational principle (assuming that the underlying electronic structure method is variational) and a rigorously correct energy gradient, without the need to solve fragment-based coupled-perturbed equations. These modifications can be used in conjunction with a variety of fragmentation schemes, encompassing all of the fragment-based methods mentioned above, so that the present formalism can be considered a general paradigm for fragment-based AIMD simulations.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpclett.9b01214.

> Mulliken charges for $(H_2O)_{10}$ and $F^-(H_2O)_{10}$ and finite-difference results to verify our implementation of the analytic gradient (PDF)

■ AUTHOR INFORMATION

Corresponding Author

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

ORCID

Bhaskar Rana: 0000-0002-8751-7314

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

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

Present Address

†Hefei National Laboratory of Physical Science at the Microscale, University of Science and Technology of China, Anhui 230026, China.

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.[101]

■ REFERENCES

(1) 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.

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

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

(4) 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.

(5) Hirata, S.; Gilliard, K.; He, X.; Keçeli, M.; Li, J.; Salim, M. A.; Sode, O.; Yagi, K. *Ab initio* ice, dry ice, and liquid water. In *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems*; Gordon, M. S., Ed.; Wiley, 2017; Chapter 9, pp 245−296.

(6) Liu, J.; He, X.; Zhang, J. Z. H. Structure of liquid water—A dynamical mixture of tetrahedral and 'ring-and-chain' like structures. *Phys. Chem. Chem. Phys.* **2017**, *19*, 11931−11936.

(7) Liu, J.; He, X.; Zhang, J. Z. H.; Qi, L.-W. Hydrogen-bond structure dynamics in bulk water: Insights from *ab initio* simulations with coupled cluster theory. *Chem. Sci.* **2018**, *9*, 2065−2073.

(8) Li, W.; Li, S.; Jiang, Y. Generalized energy-based fragmentation approach for computing the ground-state energies and properties of large molecules. *J. Phys. Chem. A* **2007**, *111*, 2193−2199.

(9) Hratchian, H. P.; Parandekar, P. V.; Raghavachari, K.; Frisch, M. J.; Vreven, T. QM:QM electronic embedding using Mulliken atomic charges: Energies and analytic gradients in an ONIOM framework. *J. Chem. Phys.* **2008**, *128*, 034107.

(10) Parandekar, P. V.; Hratchian, H. P.; Raghavachari, K. Applications and assessment of QM:QM electronic embedding using generalized asymmetric Mulliken atomic charges. *J. Chem. Phys.* **2008**, *129*, 145101.

(11) Wang, X.; Liu, J.; Zhang, J. Z. H.; He, X. Electrostatically embedded generalized molecular fractionation with conjugate caps method for full quantum mechanical calculation of protein energy. *J. Phys. Chem. A* **2013**, *117*, 7149−7161.

(12) Isegawa, M.; Wang, B.; Truhlar, D. G. Electrostatically embedded molecular tailoring approach and validation for peptides. *J. Chem. Theory Comput.* **2013**, *9*, 1381−1393.

(13) Collins, M. A. Molecular forces, geometries, and frequencies by systematic molecular fragmentation including embedding charges. *J. Chem. Phys.* **2014**, *141*, 094108.

(14) 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.

(15) Hua, W.; Fang, T.; Li, W.; Yu, J.-G.; Li, S. Geometry optimizations and vibrational spectra of large molecules from a generalized energy-based fragmentation approach. *J. Phys. Chem. A* **2008**, *112*, 10864−10872.

(16) Hua, S.; Hua, W.; Li, S. An efficient implementation of the generalized energy-based fragmentation approach for general large molecules. *J. Phys. Chem. A* **2010**, *114*, 8126−8134.

(17) Li, W.; Hua, W.; Fang, T.; Li, S. The energy-based fragmentation approach for ab initio calculations of large systems. In *Computational Methods for Large Systems: Electronic Structure Approaches for Biotechnology and Nanotechnology*; Reimers, J. R., Ed.; Wiley: Hoboken, NJ, 2011; Chapter 7, pp 227−258.

(18) Li, S.; Li, W.; Ma, J. Generalized energy-based fragmentation approach and its applications to macromolecules and molecular aggregates. *Acc. Chem. Res.* **2014**, *47*, 2712−2720.

(19) Deev, V.; Collins, M. A. Approximate *ab initio* energies by systematic molecular fragmentation. *J. Chem. Phys.* **2005**, *122*, 154102.

(20) Yang, Z.; Hua, S.; Hua, W.; Li, S. Low-lying structures and stabilities of large water clusters: Investigation based on the combination of the AMOEBA potential and generalized energy-based fragmentation approach. *J. Phys. Chem. A* **2010**, *114*, 9253−9261.

(21) Fang, T.; Li, Y.; Li, S. Generalized energy-based fragmentation approach for modeling condensed phase systems. *WIREs Comput. Mol. Sci.* **2017**, *7*, e1297.

(22) 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.

(23) Nagata, T.; Fedorov, D. G.; Ishimura, K.; Kitaura, K. Analytic energy gradient for second-order Møller-Plesset perturbation theory based on the fragment molecular orbital method. *J. Chem. Phys.* **2011**, *135*, 044110.

(24) Nagata, T.; Fedorov, D. G.; Kitaura, K. Analytic gradient for the embedding potential with approximations in the fragment molecular orbital method. *Chem. Phys. Lett.* **2012**, *544*, 87−93.

(25) Brorsen, K. R.; Zahariev, F.; Nakata, H.; Fedorov, D. G.; Gordon, M. S. Analytic gradient for density functional theory based on the fragment molecular orbital method. *J. Chem. Theory Comput.* **2014**, *10*, 5297−5307.

(26) Pople, J. A.; Krishnan, R.; Schlegel, H. B.; Binkley, J. S. Derivative studies in Hartree−Fock and Møller−Plesset theories. *Int. J. Quantum Chem. Symp.* **1979**, *16*, 225−241.

(27) Fedorov, D. G.; Kitaura, K. Extending the power of quantum chemistry to large systems with the fragment molecular orbital method. *J. Phys. Chem. A* **2007**, *111*, 6904−6914.

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

(29) Nagata, T.; Fedorov, D. G.; Kitaura, K. Mathematical formulation of the fragment molecular orbital method. In *Linear-Scaling Techniques in Computational Chemistry and Physics*; Zalesńy, R., Papadopoulos, M. G., Mezey, P. G., Leszczynski, J., Eds.; Springer: New York, 2011; Vol. *13*, Chapter 2, pp 17−64.

(30) Fedorov, D. G.; Nagata, T.; Kitaura, K. Exploring chemistry with the fragment molecular orbital method. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7562−7577.

(31) Fedorov, D. G. The fragment molecular orbital method: Theoretical development, implementation in GAMESS, and applications. *WIREs Comput. Mol. Sci.* **2017**, *7*, e1322.

(32) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment molecular orbital method: An approximate computational method for large molecules. *Chem. Phys. Lett.* **1999**, *313*, 701−706.

(33) Kitaura, K.; Sugiki, S.-I.; Nakano, T.; Komeiji, Y.; Uebayasi, M. Fragment molecular orbital method: Analytical energy gradients. *Chem. Phys. Lett.* **2001**, *336*, 163−170.

(34) Fedorov, D. G.; Kitaura, K. Second-order Møller-Plesset perturbation theory based upon the fragment molecular orbital method. *J. Chem. Phys.* **2004**, *121*, 2483−2490.

(35) Fedorov, D. G.; Ishida, T.; Uebayasi, M.; Kitaura, K. The fragment molecular orbital method for geometry optimizations of polypeptides and proteins. *J. Phys. Chem. A* **2007**, *111*, 2722−2732.

(36) Nagata, T.; Fedorov, D. G.; Kitaura, K. Derivatives of the approximated electrostatic potentials in the fragment molecular orbital method. *Chem. Phys. Lett.* **2009**, *475*, 124−131.

(37) Nagata, T.; Fedorov, D. G.; Kitaura, K. Importance of the hybrid orbital operator derivative term for the energy gradient in the fragment molecular orbital method. *Chem. Phys. Lett.* **2010**, *492*, 302−308.

(38) Mochizuki, Y.; Nakano, T.; Komeiji, Y.; Yamashita, K.; Okiyama, Y.; Yoshikawa, H.; Yamataka, H. Fragment molecular orbital-based molecular dynamics (FMO-MD) method with MP2 gradient. *Chem. Phys. Lett.* **2011**, *504*, 95−99.

(39) Brorsen, K. R.; Minezawa, N.; Xu, F.; Windus, T. L.; Gordon, M. S. Fragment molecular orbital molecular dynamics with the fully analytic energy gradient. *J. Chem. Theory Comput.* **2012**, *8*, 5008−5012.

(40) 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.

(41) 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.

(42) Fedorov, D. G.; Kitaura, K. The three-body fragment molecular orbital method for accurate calculations of large systems. *Chem. Phys. Lett.* **2006**, *433*, 182−187.

(43) Nakano, T.; Mochizuki, Y.; Yamashita, K.; Watanabe, C.; Fukuzawa, K.; Segawa, K.; Okiyama, Y.; Tsukamoto, T.; Tanaka, S. Development of the four-body corrected fragment molecular orbital (FMO4) method. *Chem. Phys. Lett.* **2012**, *523*, 128−133.

(44) Stoll, H. Correlation energy of diamond. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1992**, *46*, 6700−6704.

(45) Hirata, S.; Valiev, M.; Dupuis, M.; Xantheas, S. S.; Sugiki, S.; Sekino, H. Fast electron correlation methods for molecular clusters in the ground and excited states. *Mol. Phys.* **2005**, *103*, 2255−2265.

(46) Hirata, S. 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.* **2008**, *129*, 204104.

(47) 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.

(48) Speetzen, E. D.; Leverentz, H. R.; Lin, H.; Truhlar, D. G. Electrostatically embedded many-body expansion for large systems. In *Accurate Condensed-Phase Electronic Structure Theory*; Manby, F. R., Ed.; CRC Press: Boca Rotan, FL, 2011; Chapter 5, pp 105−127.

(49) Beran, G. J. O. Approximating quantum many-body intermolecular interactions in molecular clusters using classical polarizable force fields. *J. Chem. Phys.* **2009**, *130*, 164115.

(50) Beran, G. J. O.; Nanda, K. Predicting organic crystal lattice energies with chemical accuracy. *J. Phys. Chem. Lett.* **2010**, *1*, 3480−3487.

(51) 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.

(52) 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.

(53) 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.

(54) 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.

(55) Ganesh, V.; Dongare, R. K.; Balanarayan, P.; Gadre, S. R. Molecular tailoring approach for geometry optimization of large molecules: Energy evaluation and parallelization strategies. *J. Chem. Phys.* **2006**, *125*, 104109.

(56) Rahalkar, A. P.; Ganesh, V.; Gadre, S. R. Enabling *ab initio* Hessian and frequency calculations of large molecules. *J. Chem. Phys.* **2008**, *129*, 234101.

(57) Gadre, S. R.; Jovan Jose, K. V.; Rahalkar, A. P. Molecular tailoring approach for exploring structures, energetics and properties of clusters. *J. Chem. Sci.* **2010**, *122*, 47−56.

(58) Rahalkar, A. P.; Gadre, S. R. Tailoring approach for obtaining molecular orbitals of large systems. *J. Chem. Sci.* **2012**, *124*, 149−158.

(59) Sahu, N.; Gadre, S. R. Molecular tailoring approach: A route for *ab initio* treatment of large clusters. *Acc. Chem. Res.* **2014**, *47*, 2739−2747.

(60) Khire, S. S.; Bartolotti, L. J.; Gadre, S. R. Harmonizing accuracy and efficiency: A pragmatic approach to fragmentation of large molecules. *J. Chem. Phys.* **2018**, *149*, 064112.

(61) He, X.; Zhu, T.; Wang, X. W.; Liu, J. F.; Zhang, J. Z. H. Fragment quantum mechanical calculation of proteins and its applications. *Acc. Chem. Res.* **2014**, *47*, 2748−2757.

(62) 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.

(63) Liu, J.; Zhu, T.; He, X.; Zhang, J. Z. H. MFCC-based fragmentation methods for biomolecules. In *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems*; Gordon, M. S., Ed.; Wiley, 2017; Chapter 11, pp 323−348.

(64) 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.

(65) 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.

(66) 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.

(67) 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.

(68) Collins, M. A.; Cvitkovic, M. W.; Bettens, R. P. A. The combined fragmentation and systematic molecular fragmentation methods. *Acc. Chem. Res.* **2014**, *47*, 2776−2785.

(69) Herbert, J. M.; Jacobson, L. D.; Un Lao, K.; Rohrdanz, M. A. Rapid computation of intermolecular interactions in molecular and ionic clusters: Self-consistent polarization plus symmetry-adapted perturbation theory. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7679−7699.

(70) Holden, Z. C.; Rana, B.; Herbert, J. M. Analytic energy gradients for the QM/MM-Ewald method using atomic charges derived from the electrostatic potential: Theory, implementation, and application to *ab initio* molecular dynamics of the aqueous electron. *J. Chem. Phys.* **2019**, *150*, 144115.

(71) Nakata, H.; Schmidt, M. W.; Fedorov, D. G.; Kitaura, K.; Nakamura, S.; Gordon, M. S. Efficient molecular dynamics simulations of multiple radical center systems based on the fragment molecular orbital method. *J. Phys. Chem. A* **2014**, *118*, 9762−9771.

(72) Komeiji, Y.; Nakano, T.; Fukuzawa, K.; Ueno, Y.; Inadomi, Y.; Nemoto, T.; Uebayasi, M.; Fedorov, D. G.; Kitaura, K. Fragment molecular orbital method: Application to molecular dynamics simulation, 'ab initio FMO-MD'. *Chem. Phys. Lett.* **2003**, *372*, 342−347.

(73) Komeiji, Y.; Mochizuki, Y.; Nakano, T.; Fedorov, D. G. Fragment molecular orbital-based molecular dynamics (FMO-MD), a

quantum simulation tool for large molecular systems. *J. Mol. Struct.: THEOCHEM* **2009**, *898*, 2−7.

(74) Zhang, L.; Li, W.; Fang, T.; Li, S. Ab initio molecular dynamics with intramolecular noncovalent interactions for unsolvated polypeptides. *Theor. Chem. Acc.* **2016**, *135*, 34.

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

(76) Gao, J.; Truhlar, D. G.; Wang, Y.; Mazack, M. J. M.; Löffler, P.; Provorse, M. R.; Rehak, P. Explicit polarization: A quantum mechanical framework for developing next generation force fields. *Acc. Chem. Res.* **2014**, *47*, 2837−2845.

(77) Jacobson, L. D.; Herbert, J. M. An efficient, fragment-based electronic structure method for molecular systems: Self-consistent polarization with perturbative two-body exchange and dispersion. *J. Chem. Phys.* **2011**, *134*, 094118.

(78) Liu, K.-Y.; Carter-Fenk, K.; Herbert, J. M. Self-consistent charge embedding at very low cost, with application to symmetry-adapted perturbation theory. **2019**, submitted.

(79) Holden, Z. C.; Richard, R. M.; Herbert, J. M. Periodic boundary conditions for QM/MM calculations: Ewald summation for extended Gaussian basis sets. *J. Chem. Phys.* **2013**, *139*, 244108. Holden, Z. C.; et al. Erratum. *J. Chem. Phys.* **2015**, *142*, 059901.

(80) Francl, M. M.; Chirlian, L. E. The pluses and minuses of mapping atomic charges to electrostatic potentials. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 2000; Vol. *14*, Chapter 1, pp 1−32.

(81) Gao, J.; Wang, Y. Communication: Variational many-body expansion: Accounting for exchange repulsion, charge delocalization, and dispersion in the fragment-based explicit polarization method. *J. Chem. Phys.* **2012**, *136*, 071101.

(82) Shao, Y.; et al. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **2015**, *113*, 184−215.

(83) 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.

(84) 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.

(85) 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.

(86) Ouyang, J. F.; Bettens, R. P. A. When are many-body effects significant? *J. Chem. Theory Comput.* **2016**, *12*, 5860−5867.

(87) Yuan, D.; Shen, X.; Li, W.; Li, S. Are fragment-based quantum chemistry methods applicable to medium-sized water clusters? *Phys. Chem. Chem. Phys.* **2016**, *18*, 16491−16498.

(88) Zhang, L.; Li, W.; Fang, T.; Li, S. Accurate relative energies and binding energies of large ice−liquid water clusters and periodic structures. *J. Phys. Chem. A* **2017**, *121*, 4030−4038.

(89) Yuan, D.; Li, Y.; Ni, Z.; Pulay, P.; Li, W.; Li, S. Benchmark relative energies for large water clusters with the generalized energy-based fragmentation method. *J. Chem. Theory Comput.* **2017**, *13*, 2696−2704.

(90) Zhan, C.-G.; Chipman, D. M. Cavity size in reaction field theory. *J. Chem. Phys.* **1998**, *109*, 10543−10558.

(91) Lee, A. M.; Bettens, R. P. A. First principles NMR calculations by fragmentation. *J. Phys. Chem. A* **2007**, *111*, 5111−5115.

(92) Kobayashi, R.; Amos, R. D.; Reid, D. M.; Collins, M. A. Application of the systematic molecular fragmentation by annihilation method to *ab initio* NMR chemical shift calculations. *J. Phys. Chem. A* **2018**, *122*, 9135−9141.

(93) Jose, K. V. J.; Raghavachari, K. Molecules-in-molecules fragment-based method for the evaluation of Raman spectra of large molecules. *Mol. Phys.* **2015**, *113*, 3057−3066.

(94) 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.

(95) 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.

(96) Sure, R.; Grimme, S. Corrected small basis set Hartree-Fock method for large systems. *J. Comput. Chem.* **2013**, *34*, 1672−1685.

(97) Lao, K. U.; Herbert, J. M. Atomic orbital implementation of extended symmetry-adapted perturbation theory (XSAPT) and benchmark calculations for large supramolecular complexes. *J. Chem. Theory Comput.* **2018**, *14*, 2955−2978.

(98) Liu, K.-Y.; Liu, J.; Herbert, J. M. Accuracy of finite-difference harmonic frequencies in density functional theory. *J. Comput. Chem.* **2017**, *38*, 1678−1684.

(99) Herbert, J. M.; Head-Gordon, M. Accelerated, energy-conserving Born-Oppenheimer molecular dynamics *via* Fock matrix extrapolation. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3269−3275.

(100) Collins, M. A. Can systematic molecular fragmentation be applied to direct ab initio molecular dynamics? *J. Phys. Chem. A* **2016**, *120*, 9281−9291.

(101) *Ohio Supercomputer Center*. http://osc.edu/ark:/19495/f5s1ph73 (2019).