Many-Overlapping-Body (MOB) Expansion: A Generalized Many Body Expansion for Nondisjoint Monomers in Molecular Fragmentation Calculations of Covalent Molecules

Nicholas J. Mayhall and Krishnan Raghavachari*

Department of Chemistry, Indiana University, Bloomington, Indiana 47405, United States

Ⓢ Supporting Information

ABSTRACT: A common approach to approximating the full electronic energy of a molecular system is to first divide the system into nonoverlapping (disjoint) fragments and then compute the two-body or three-body fragment−fragment interactions using a many-body expansion. In this paper, we demonstrate that, by using a set of fragments which overlap with each other, a many-body expansion converges much faster than using nonoverlapping fragments. A new hierarchical fragmentation scheme is therefore proposed which generalizes the many-body expansion expressions and describes a simple procedure for generating the set of overlapping monomers. This method is referred to as the many-overlapping-body (MOB) expansion and is evaluated with two example systems: four dendritic isomers of $C_{29}H_{60}$ and 10 conformational isomers of a polypeptide molecule. In both examples, the MOB methodology significantly improves the two-body corrected energies.

I. INTRODUCTION

In order to dodge the steep computational scaling of *ab initio* quantum chemical calculations, fragmentation methods aim to break a given molecular system up into computationally efficient units and piece them back together to approximate the unfragmented energy. In our previous work,[1] we demonstrated how fragmentation calculations can be systematically improved by using multiple levels of theory to extrapolate to the unfragmented results. This method, termed molecules-in-molecules (MIM), was shown to provide very accurate approximations to the unfragmented electronic energies for a variety of molecular systems.[1]

MIM is a general methodology, applicable to any type of molecular fragmentation method. Of the numerous fragmentation methods which have been presented, two distinct classes emerge that differ in the manner in which the fragments are constructed. We will identify a fragmentation method as either an inclusion−exclusion based method or a many-body based method. While a full review of all current fragmentation methods would exceed the scope of this article, we refer to a recent review paper for a more thorough documentation.[2]

Inclusion−Exclusion Based Methods. Inclusion−exclusion based methods take a top-down approach by first creating *overlapping* subsystems, each containing all necessary interactions (**primary subsystems**). The inclusion−exclusion principle (eq 1) is then used to obtain the remaining subsystems (**derivative subsystems**) required for appropriately canceling out the overcounted atoms.

$$|A_1 \ldots \cup A_n| = \sum_i |A_i| - \sum_{i<j} |A_i \cap A_j|$$
$$+ \sum_{i<j<k} |A_i \cap A_j \cap A_k| \ldots + (-1)^{n-1}$$
$$|A_i \cap \ldots \cap A_n| \tag{1}$$

Examples of inclusion−exclusion based methods include the cardinality-guided molecular tailoring approach (CG-MTA)[3−9] and the generalized energy-based fragmentation (GEBF) method.[10−12] Although eq 1 is not explicitly referred to, both the molecular fractionation with conjugate caps (MFCC) method[13,14] and the systematic molecular fragmentation (SMF) method[15−19] can also be considered inclusion−exclusion based methods, as well as the fragmentation scheme used in the original MIM paper.[1] The accuracy of an inclusion−exclusion based method is primarily determined by the size of the primary subsystems (e.g., the distance cutoff parameter).

Many-Body Based Methods. Many-body based methods take a bottom-up approach, by first obtaining small, *nonoverlapping* (disjoint) subsystems and then combining them according to eq 2, to form larger subsystems that contain the desired interactions.

$$E = \sum_i E_i + \sum_{i>j} \Delta E_{ij}' + \sum_{i>j>k} \Delta E_{ijk}' + \ldots \tag{2}$$

where the interaction terms are obtained using supersystem calculations[49]

$$\Delta E_{ij}' = E_{ij} - (E_i + E_j) \tag{3}$$

$$\Delta E_{ijk}' = E_{ijk} - (\Delta E_{ij}' + \Delta E_{ik}' + \Delta E_{jk}') - (E_i + E_j + E_k) \tag{4}$$

Here, E_i is the energy of subsystem i and E_{ij} is the energy of combined subsystem $i + j$. The prime (′) on the interaction energies indicates that the monomers are not overlapping. Typically this is truncated at two-body or three-body interactions. A second approximation may be used at this

point to avoid calculating those interaction terms which are not spatially proximal.

Examples of many-body based methods are the fragment molecular orbital method (FMO),[20−26] the electrostatically embedded many-body (EE-MB) method,[27−31] the multi-centered integrated QM:QM methods,[32−36] the kernel energy method (KEM),[37,38] the hybrid many-body interaction (HMBI) method,[39,40] and the multilevel fragment-based approach (MFBA).[41] The accuracy of many-body methods is often determined by two parameters: the number of interacting monomers (e.g., two-body vs three-body) and the distance cutoff (if used).

Comparing the Two Classes of Fragmentation. Inclusion−exclusion based methods may be thought of as a full N-body expansion within some distance cutoff, which is determined by the size of each individual subsystem. Clearly, if the distance cutoff is large enough to incorporate the entire molecule, the molecule's energy will become exact. Many-body methods also typically employ a distance cutoff to attain a linear scaling cost. However, contrary to the inclusion−exclusion based methods, truncated many-body methods do not necessarily converge to the exact energy with increasing distance cutoff.

Just as an inclusion−exclusion based method can be thought of as an untruncated many-body method (with a distance cutoff), many-body methods can also be thought of as an inclusion−exclusion based method. For example, if one were to choose primary subsystems to be all pairs of monomers, and then use eq 1 to obtain the derivative subsystems, we would get the same resulting subsystems as in a many-body method truncated at two-body terms.

Regardless of the class of fragmentation used, the resulting subsystems may be embedding in a sea of point charges to mimic the electrostatic environment of the extended molecule. This has been shown to significantly improve the results of several fragmentation methods.[10,27,42,43]

In this paper, we present a new fragmentation scheme in an attempt to combine the most advantageous characteristics of both the inclusion−exclusion based methods and the many-body methods. Recognizing the ability of inclusion−exclusion based methods to reproduce the absolute energy of large molecules, we take this as our starting point and then do a many-body expansion on top of this. In other words, we generalize the many-body energy expressions for use with a set of monomers which overlap with one another.

II. METHODS

Inclusion−exclusion based methods include the important fragment−fragment interactions by forming strongly over-lapping primary subsystems and subsequently subtracting out the overlapping regions in accordance with the inclusion−exclusion principle. One benefit of this approach is that since the primary subsystems are assembled on the basis of some distance criteria, only the most important interactions are included, and the fragmentation naturally exhibits a linear scaling with molecule size. However, for systems with polar or charged groups, long-range interactions are critical for reproducing the molecular energies. In such cases, many-body based methods offer benefits by allowing one to include longer range interactions without significantly increasing the resulting subsystem size. However, by defining the many-body expansion as in eq 2, we are implying that the monomers are disjoint sets of atoms (there are no atoms common to more than one

monomer). It seems reasonable to expect accelerated convergence of the many-body expansion when the self-energies (or monomer energies) form a larger percentage of the overall binding energy, as this would require less of the higher order terms. Therefore, we propose a generalized many-body expansion which is built from a set of *overlapping monomers*. The revised many-body formalism is presented below, followed by a procedure for obtaining the overlapping monomers for covalent systems.

A. Many-Overlapping-Body Expansion. Straightforward application of eq 2 with overlapping monomers would clearly double count the energies of atoms and interactions. Therefore, the use of overlapping monomers (denoted by the indices p, q, r, ...) requires a simple generalization of the many-body expansion expression:

$$E = \sum_p c_p E_p + \sum_{p<q} c_p c_q \Delta E_{pq} + \sum_{p<q<r} c_p c_q c_r \Delta E_{pqr} + \cdots \tag{5}$$

where c_p is the monomer coefficient obtained from inclusion−exclusion principle for the pth subsystem (i.e., $c_i = 1$, $c_{i\cap j} = -1$, $c_{i\cap j\cap k} = 1$, ...). Note, however, that since the same intersection may arise from different overlaps, the coefficients may differ from +1 or −1.

Summation of the one-body energies produces the original inclusion−exclusion based fragmentation energy. The two-body and three-body terms are now defined as

$$\Delta E_{pq} = E_{pq} - (E_p + E_q - E_{p\cap q}) \tag{6}$$

$$\Delta E_{pqr} = E_{pqr} - (\Delta E_{pq} + \Delta E_{pr} + \Delta E_{qr})$$
$$- (E_p + E_q + E_r - E_{p\cap q} - E_{p\cap r} - E_{q\cap r})$$
$$- E_{p\cap q\cap r} \tag{7}$$

where $E_{p\cap q}$ is the energy of the intersection of monomers A_p and A_q, i.e., $E_{p\cap q} = E(A_p \cap A_q)$.

If the Δ terms are expanded and compared to eqs 3 and 4, the generalized interaction expressions can be written in terms of the conventional interaction expressions:

$$\Delta E_{pq} = \Delta E_{pq}' + E_{p\cap q} \tag{8}$$

$$\Delta E_{pqr} = \Delta E_{pqr}' - E_{p\cap q\cap r} \tag{9}$$

From eqs 8 and 9, it is clear that in the absence of monomer overlap, the MOB interaction energies reduce to the conventional (non-overlapping) many-body expansion interaction energies. Use of this simple expression permits a fragment-based calculation in which the individual monomers of a many-body expansion form a set of overlapping subsystems. We call this approach the MOB (many-overlapping-body) expansion.

B. Overlapping Monomer Generation. In order to fully define the MOB method, one must also describe the method for generating the overlapping monomers. Although several approaches could be taken to generate monomers, we have chosen a very simple method which only takes into account the covalent network of the molecule. We define a hierarchy of monomer sets which are based on previous work from our research group concerning thermochemical reaction schemes.[44]

In this approach to developing overlapping monomers, we first visit each atom (or bond) in the molecule and initialize a new primary subsystem. We then append nearby connected

atoms according to the chosen *monomer degree*, ensuring that we do not truncate nonsingle bonds. Monomers of even degree are atom-centric, meaning that the primary subsystems are initialized for each heavy atom center. Monomers of odd degree are bond-centric, meaning that the primary subsystems are initialized for each bond between two heavy atoms.

- For 0th degree monomers, each atom initializes a new primary subsystem, and no further atoms are added. In this case, the list of monomers is the same as the list of atoms.
- For first degree monomers, each bond initializes a primary subsystem, and no further atoms are added.
- For second degree monomers, each atom initializes a new primary subsystem, and then all directly connected atoms are appended.
- For third degree monomers, each bond initializes a new primary subsystem, and then each atom directly connected to that bond is appended.
- For fourth degree monomers, each atom initializes a new primary subsystem, and then each atom that falls within two bonds of the central atom is appended.

This hierarchy can continue further.

After all primary subsystems are formed, the inclusion–exclusion principle (eq 1) is used to determine all of the overlap-canceling derivative subsystems. All of the primary and derivative subsystems are collectively referred to as monomers. This is a subtle but important aspect, because if the many-body expansion is only done with the primary subsystems, the number of interactions does not sum properly. Only when all interactions of all monomers are evaluated (weighted by their respective coefficients as shown in eq 5) do the atom–atom interactions sum appropriately.

To perform a MOB calculation, there are two fields needing definition: (1) order of the MOB expansion and (2) monomer degree. Therefore, throughout this paper, we use the notation MOB$\alpha.\beta$ where α is the order of the MOB expansion and β is the monomer degree. For example, if we used third degree monomers and computed all two-body interactions via MOB, we would refer to this as a MOB2.3 calculation. An illustration of the MOB$\alpha.\beta$ hierarchy is shown in Figure 1.
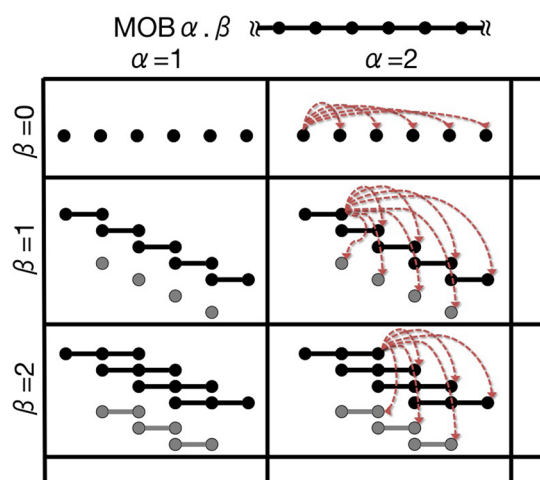

Figure 1. Schematic representation of the MOB methodology. Black monomers represent positive monomers. Gray monomers represent negative monomers. Red dashed lines denote interaction terms.

As with the MIM work,[1] we use hydrogen link atoms to cap each subsystem's severed covalent bonds.[50] Therefore, only covalent single bonds are allowed to be cut during the fragmentation process. When a double bond, for instance, is needed to be cut to satisfy the above criteria, the subsystem increases in size until the nearest single bond is found. The use of hydrogen link-atoms creates a situation for some subsystems which would have link atoms replacing the same center twice, forcing two link atoms into close proximity (imagine using link atoms to cap an ethane subsystem obtained from cyclopropane). When this occurs, we simply add the atom to be replaced into the subsystem. A graphic to help illustrate this point is shown in the Supporting Information.

We note that work has been done to include a two-body energy correction to the MFCC method.[45] This is similar to the MOB method in that it is a many-body based correction to an inclusion–exclusion based method (MFCC). However, in that work, hydrogen caps were used for the two-body interaction terms, whereas larger caps were used for the one-body terms. This is fundamentally different than the approach taken in this paper, since we treat the one-body, two-body, etc. terms all on an equal footing. Our two-body correction runs over all unique pairs of all subsystems, both primary and derivative.

As is common to all fragmentation methods, the MOB methodology provides significant reductions in computational complexity compared to unfragmented calculations. The MOB$(\alpha.\beta)$ method exhibits a computational scaling of CN^α, with C being determined by β and N monomers growing linearly with number of atoms. However, if only local many-body interactions are included, the method asymptotically becomes linear scaling. In addition to the attractive computational scaling, the method is fully parallelizable. Our programs for the generation of the fragments and organization of the electronic structure calculations were written as a series of scripts consisting of *perl* modules. In our implementation, Gaussian 09 is used for both the electronic structure calculations and the geometry optimizer.[46]

III. ASSESSMENT OF THE MOB EXPANSION

In this section, we examine the MOB method's ability to reproduce the unfragmented absolute energies and relative energies. We have selected two example systems which will appraise the MOB's ability to describe short-range dominated systems (alkane dendrimer energies which are largely determined by the description of dispersion[51]) and longer-range systems (polypeptide, PDB: 2NPV with many polar groups). For both assessments, we use only MIM1 calculations (no low level of theory extrapolation) so that performance of the fragmentation method is clearly demonstrated. As with any fragmentation method, the MIM2 methodology may be applied to extrapolate to better results using more efficient levels of theory.

A. Alkane Dendrimers: $C_{29}H_{60}$. The absolute energies of four dendritic isomers are used to assess the ability of overlapping monomers to accelerate the convergence of the MOB expansion. The four conformers are shown in Figure 2, and the Cartesian coordinates are given in the Supporting Information.

In Figure 3a, the deviations from the unfragmented MP2/6-311+G* absolute energies are shown for the MOB2.β ($\beta = 0$–3) fragmentation methods. As previously mentioned, the MOB2.0 method is the standard two-body approach, using nonoverlapping monomers. In this case, the two-body
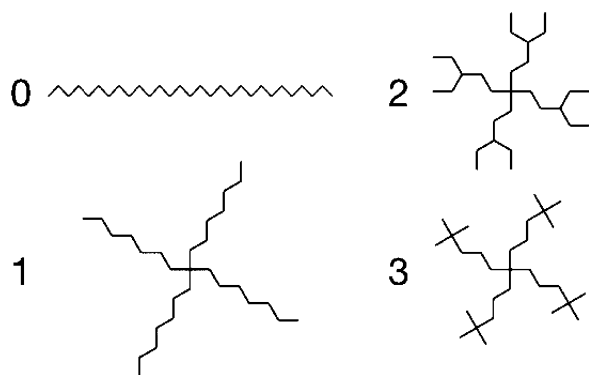
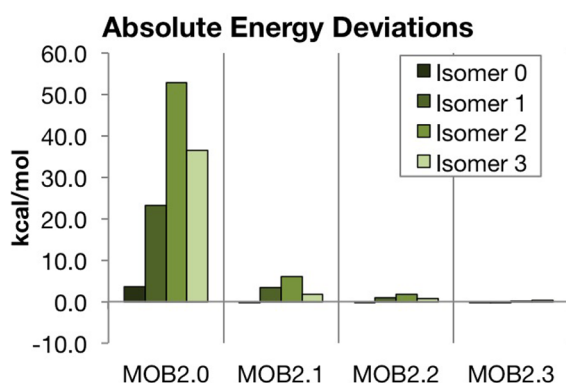**Figure 2.** Dendrimers: All four $C_{29}H_{60}$ isomers for are shown.

Figure 3. Dendrimers: Absolute energy deviations (kcal/mol) from the full, unfragmented MP2/6-311+g* energies. Isomer 0 (dark green). Isomer 1 (medium-dark green). Isomer 2 (medium-light green). Isomer 3 (light green). Stoichiometry of all isomers, $C_{29}H_{60}$.

correction terms are computed between monomers with C_1H_n ($n = 0-3$) stoichiometries. It is clear that the MOB2.0 method does not accurately reproduce the energies of the unfragmented results, with absolute error deviations of up to 52.8 kcal/mol. However, as soon as the monomers overlap, the deviations are significantly reduced as seen with the MOB2.1 results, where the largest deviation is reduced to 6.1 kcal/mol. If the monomers overlap even further, MOB2.2 and MOB2.3, the largest deviations are reduced even further to 1.7 and 0.3 kcal/mol, respectively. The RMS deviations are given in Table 1.

Table 1. Dendrimers: Absolute Energy Deviations for the Four Isomers of $C_{29}H_{60}$ (Energies in kcal/mol)

isomer	MOB2.0	MOB2.1	MOB2.2	MOB2.3
0	3.6	−0.4	−0.2	−0.1
1	23.1	3.3	0.9	−0.1
2	52.8	6.1	1.7	0.0
3	36.4	1.7	0.8	0.3
RMS	34.15	3.57	1.04	0.15
σ	20.85	2.73	0.76	0.17

Although convergence of the absolute energy ensures convergence of the relative energies, the relative energies may in fact converge significantly faster than the absolute energies due to a cancellation of systematic errors present in a quantum chemical approach. In Table 1, the standard deviations (σ) of the deviations from the unfragmented relative energies for the different MOB calculations are reported, revealing the presence

of any systematic errors in our results. Here, the standard deviations are quite large for smaller values of β, demonstrating a strong isomer-dependent error. This is because for all cases, isomer 0 is well described by fragment calculations, while the more heavily branched structures are less accurately described by fragment calculations. However, as the amount of monomer overlap increases, σ decreases. For these methods, MOB2.0, MOB2.1, MOB2.2, and MOB2.3, the standard deviations are 20.85, 2.73, 0.76, and 0.17 kcal/mol, respectively. The larger, overlapping monomers clearly provide a simple route to improving the absolute and relative energies computed in a two-body truncated expansion for alkane systems.

B. Polypeptide Conformers: 2NPV. In the previous example, we demonstrated how the MOB2.β methodology can improve the convergence of the two-body energies with increasing monomer overlap (increasing β) for systems dominated by short-range dispersion interactions. Polar molecules, such as polypeptides, however, are dominated by the long-range interactions of electrostatically charged groups. In this section, we assess the performance of the MOB2.β expansion using a small (163 atoms) polypeptide system PDB: 2NPV.

The 2NPV molecule has been chosen as a test molecule based both on its small size (which allows full, unfragmented calculations to be performed) and on the wide range of conformations sampled by the fatty acid group (as shown in Figure 4 (left)). This molecule has a well-defined, saddle-
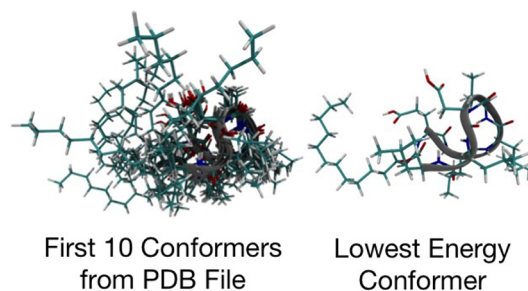

First 10 Conformers from PDB File   Lowest Energy Conformer

Figure 4. 2NPV: The first 10 conformers from PDB file superimposed (left). Lowest energy conformer (right).

shaped polypeptide backbone with a long alkyl chain extension. In this example, we use the first 10 conformers given in the PDB file. Table 2 lists the absolute energy deviations for the 2NPV conformers. As the MOB2.0 calculations do not have any overlap between monomers, both the errors in absolute

Table 2. 2NPV: Absolute Energy Deviations in kcal/mol

conformer	MOB2.0	MOB2.1	MOB2.2	MOB2.3
1	44.23	−12.78	−1.84	−0.36
2	42.80	−12.22	−2.10	−0.52
3	48.86	−14.82	−0.26	−0.04
4	44.50	−11.64	0.80	0.95
5	46.97	−10.63	0.82	0.89
6	45.13	−13.89	−1.45	0.26
7	47.48	−12.53	−0.03	0.80
8	39.80	−13.75	−1.76	0.05
9	47.53	−13.62	−1.73	−0.13
10	51.69	−10.72	1.22	1.11
RMS	46.01	12.73	1.37	0.64
σ	3.35	1.39	1.28	0.59

energies and the standard deviations of the errors are quite large. As soon as the monomers share overlapping regions, the performance is significantly improved. For the four values of β calculated, MOB2.0, MOB2.1, MOB2.2, and MOB2.3, the results have RMS deviations of 46.01, 12.73, 1.37, and 0.64 kcal/mol (Figure 5).
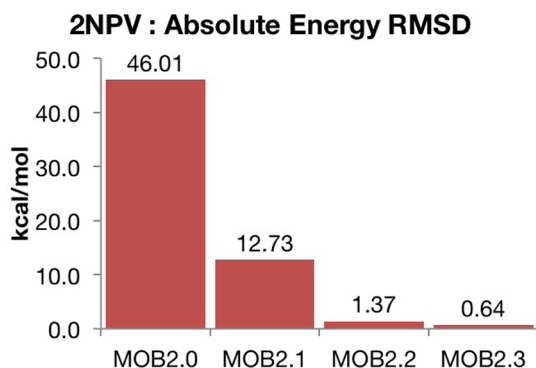

Figure 5. 2NPV: Absolute energy RMS deviations (kcal/mol) from the full, unfragmented MP2/6-31 g* energies for the first 10 conformers in the PDB file.

Compared to the $C_{29}H_{60}$ isomers, the 2NPV conformers have a markedly greater potential for error cancellation in computing relative energies. This is because the 2NPV structures are only conformational isomers, whereas the $C_{29}H_{60}$ structures are constitutional isomers, which have different bonding arrangements. The 2NPV deviations, therefore, have smaller standard deviations. For MOB2.0, MOB2.1, MOB2.2, and MOB2.3, the standard deviations are 3.35, 1.39, 1.28, and 0.59 kcal/mol, respectively.

IV. CONCLUSIONS

In this paper, we have put forth a methodology for accelerating the convergence of the many-body expansion by transforming the set of nonoverlapping monomers, to a set of monomers which overlap. This requires a modification of the many-body expansion expressions to include coefficients in front of the interaction terms and a generalization of the many-body interaction terms.

The two-body MOB method (MOB2.β) has been tested on two different examples with a varying degree of monomer overlap (various values of β). The nonpolar $C_{29}H_{60}$ dendritic isomers were found to converge much faster than the polar 2NPV conformers. However, at the MOB2.2 and MOB2.3 levels, the absolute energies of both example systems were converged to less than 1.5 and 0.7 kcal/mol, respectively. For systems with charged groups, convergence is expected to be slower, although, electronic embedding may be used to accelerate convergence. Various methods of electronic embedding will be investigated in future work.

For both example systems, we find significant performance improvements by using the overlapping monomers. The improvements observed in the accuracy of the MOB energies come at the expense of computing larger subsystems. However, because the monomers are created according to the molecular connectivity, increasing the overlap by one value of β will only increase a given monomer's size by one fragment for linear systems. For nonlinear systems, the monomers will grow faster depending on the extent of fragment coordination. Although the resulting monomers are somewhat larger than the standard

nonoverlapping monomers, we expect the MOB methodology to be more useful than including the numerous higher order corrections (three-body, four-body, etc.) to improve the energy approximations.

Because the previously presented MIM methodology is independent of the fragmentation method chosen, lower levels of theory may be employed to extrapolate the MOB2.β results to increase the accuracy even further. This will be the topic of future work.

As with all fragmentation methods, accurate results can only be expected when the chemical system studied is sufficiently local. Therefore, the approach presented here is not likely to be useful for studying highly delocalized electronic structures such as long conjugated systems like graphene or metal systems. Furthermore, for systems which have high degrees of cooperative interactions, such as the strongly interacting network of hydrogen bonds in α helices, inclusion of only the two-body terms is not anticipated to provide a suitable description of the electronic structure, and the inclusion of three-body or higher terms (larger values of α) may be necessary.

By combining the inclusion−exclusion based methods and the many-body based methods, the MOB fragmentation method is expected to be a useful tool in approximating the electronic energies of large molecules.

■ ASSOCIATED CONTENT

⑤ Supporting Information

Graphical illustration of the inclusion of doubly connected link atoms into subsystem. Cartesian coordinates for the alkane dendrimers from section III.A. This information is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: kraghava@indiana.edu.
Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) 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.

(2) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.* **2011**, DOI: 110826081820043.

(3) Gadre, S. R.; Shirsat, R. N.; Limaye, A. C. Molecular tailoring approach for simulation of electrostatic properties. *J. Phys. Chem.* **1994**, *98*, 9165−9169.

(4) Gadre, S. R.; Ganesh, V. Molecular tailoring approach: Towards PC-based ab initio treatment of large molecules. *J. Theor. Comput. Chem.* **2006**, *5*, 835−855.

(5) 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.

(6) Kavathekar, R.; Khire, S.; Ganesh, V.; Rahalkar, A. P.; Gadre, S. R. WebMTA: A Web-Interface for Ab Initio Geometry Optimization of

Large Molecules Using Molecular Tailoring Approach. *J. Comput. Chem.* **2009**, *30*, 1167−1173.

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

(8) Rahalkar, A. P.; Katouda, M.; Gadre, S. R.; Nagase, S. Molecular Tailoring Approach in Conjunction with MP2 and RI-MP2 Codes: A Comparison with Fragment Molecular Orbital Method. *J. Comput. Chem.* **2010**, *31*, 2405−2418.

(9) Yeole, S. D.; Gadre, S. R. On the applicability of fragmentation methods to conjugated pi systems within density functional framework. *J. Chem. Phys.* **2010**, *132*, 094102.

(10) 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.

(11) Hua, W. J.; Fang, T.; Li, W.; Yu, J. G.; Li, S. H. Geometry Optimizations and Vibrational Spectra of Large Molecules from a Generalized Energy-Based Fragmentation Approach. *J. Phys. Chem. A* **2008**, *112*, 10864−10872.

(12) Hua, S. G.; Hua, W. J.; Li, S. H. An Efficient Implementation of the Generalized Energy-Based Fragmentation Approach for General Large Molecules. *J. Phys. Chem. A* **2010**, *114*, 8126−8134.

(13) Mei, Y.; Ji, C. G.; Zhang, J. Z. H. A new quantum method for electrostatic solvation energy of protein. *J. Chem. Phys.* **2006**, *125*, 094906.

(14) Zhang, D. W.; Zhang, J. Z. H. Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein-molecule interaction energy. *J. Chem. Phys.* **2003**, *119*, 3599−3605.

(15) Addicoat, M. A.; Collins, M. A. Accurate treatment of nonbonded interactions within systematic molecular fragmentation. *J. Chem. Phys.* **2009**, *131*, 104103.

(16) Collins, M. A.; Deev, V. A. Accuracy and efficiency of electronic energies from systematic molecular fragmentation. *J. Chem. Phys.* **2006**, *125*, 104104.

(17) Collins, M. A. Molecular potential energy surfaces constructed from interpolation of systematic fragment surfaces. *J. Chem. Phys.* **2007**, *127*, 024104.

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

(19) Netzloff, H. M.; Collins, M. A. Ab initio energies of nonconducting crystals by systematic fragmentation. *J. Chem. Phys.* **2007**, *127*, 134113.

(20) Fujimoto, H.; Koga, N.; Fukui, K. A coupled fragment molecular-orbital method for interacting systems. *J. Am. Chem. Soc.* **1981**, *103*, 7452−7457.

(21) Fedorov, D. G.; Ishida, T.; Kitaura, K. Multilayer formulation of the fragment molecular orbital method (FMO). *J. Phys. Chem. A* **2005**, *109*, 2638−2646.

(22) 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.

(23) 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.

(24) 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.

(25) Fedorov, D. G.; Slipchenko, L. V.; Kitaura, K. Systematic Study of the Embedding Potential Description in the Fragment Molecular Orbital Method. *J. Phys. Chem. A* **2010**, *114*, 8742−8753.

(26) Pruitt, S. R.; Fedorov, D. G.; Kitaura, K.; Gordon, M. S. Open-Shell Formulation of the Fragment Molecular Orbital Method. *J. Chem. Theory Comput.* **2010**, *6*, 1−5.

(27) 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.

(28) Dahlke, E. E.; Leverentz, H. R.; Truhlar, D. G. Evaluation of the electrostatically embedded many-body expansion and the electrostatically embedded many-body expansion of the correlation energy by application to low-lying water hexamers. *J. Chem. Theory Comput.* **2008**, *4*, 33−41.

(29) Dahlke, E. E.; Truhlar, D. G. Electrostatically embedded many-body expansion for simulations. *J. Chem. Theory Comput.* **2008**, *4*, 1−6.

(30) Leverentz, H. R.; Truhlar, D. G. 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 Comput.* **2009**, *5*, 1573−1584.

(31) Sorkin, A.; Dahlke, E. E.; Truhlar, D. G. Application of the electrostatically embedded many-body expansion to microsolvation of ammonia in water clusters. *J. Chem. Theory Comput.* **2008**, *4*, 683−688.

(32) Elsohly, A. M.; Shaw, C. L.; Guice, M. E.; Smith, B. D.; Tschumper, G. S. Analytic gradients for the multicentred integrated QM: QM method for weakly bound clusters: efficient and accurate 2-body: many-body geometry optimizations. *Mol. Phys.* **2007**, *105*, 2777−2782.

(33) Hopkins, B. W.; Tschumper, G. S. Integrated quantum mechanical approaches for extended pi systems: Multicentered QM/QM studies of the cyanogen and diacetylene trimers. *Chem. Phys. Lett.* **2005**, *407*, 362−367.

(34) Hopkins, B. W.; Tschumper, G. S. Multicentred QM/QM methods for overlapping model systems. *Mol. Phys.* **2005**, *103*, 309−315.

(35) Hopkins, B. W.; Tschumper, G. S. A multicentered approach to integrated QM/QM calculations. Applications to multiply hydrogen bonded systems. *J. Comput. Chem.* **2003**, *24*, 1563−1568.

(36) Tschumper, G. S. Multicentered integrated QM: QM methods for weakly bound clusters: An efficient and accurate 2-body: many-body treatment of hydrogen bonding and van der Waals interactions. *Chem. Phys. Lett.* **2006**, *427*, 185−191.

(37) Huang, L.; Massa, L.; Karle, J. The Kernel Energy Method: Application to a tRNA. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 1233.

(38) Huang, L.; Massa, L.; Karle, J. Kernel energy method illustrated with peptides. *Int. J. Quantum Chem.* **2005**, *103*, 808.

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

(40) Beran, G. J. O.; Nanda, K. Predicting Organic Crystal Lattice Energies with Chemical Accuracy. *J. Phys. Chem. Lett.* **2010**, *1*, 3480.

(41) Řezàč, J.; Salahub, D. R. Multilevel Fragment-Based Approach (MFBA): A Novel Hybrid Computational Method for the Study of Large Molecules. *J. Chem. Theory Comput.* **2010**, *6*, 91−99.

(42) Dahlke, E. E.; Truhlar, D. G. Electrostatically Embedded Many-Body Correlation Energy, with Applications to the Calculation of Accurate Second-Order Moller-Plesset Perturbation Theory Energies for Large Water Clusters. *J. Chem. Theory Comput.* **2007**, *3*, 1342−1348.

(43) Tempkin, J. O. B.; Leverentz, H. R.; Wang, B.; Truhlar, D. G. Screened Electrostatically Embedded Many-Body Method. *J. Phys. Chem. Lett.* **2011**, *2*, 2141.

(44) Ramabhadran, R. O.; Raghavachari, K. Theoretical Thermochemistry for Organic Molecules: Development of the Generalized Connectivity-Based Hierarchy. *J. Chem. Theory Comput.* **2011**, *7*, 2094−2103.

(45) Li, S.; Li, W.; Fang, T. An Efficient Fragment-Based Approach for Predicting the Ground-State Energies and Structures of Large Molecules. *J. Am. Chem. Soc.* **2005**, *127*, 7215.

(46) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.;

Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ã.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision A.1; Gaussian Inc.: Wallingford, CT, 2009.

(47) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes. *Chem. Rev.* **1994**, *94*, 1887.

(48) 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.

(49) Alternative approaches to obtaining the interactions may, in general, be used such as symmetry adapted perturbation theory (SAPT).[47,48] However, we will not consider this at this time.

(50) We are investigating other approaches to capping dangling bonds based on pseudoatoms or group-capping.

(51) The energy gap between isomers 1 and 4 is 41 kcal/mol higher for B3LYP/6-311+G* than for MP2/6-311+G*, reversing the relative orderings.