

Molecules-in-Molecules: An Extrapolated Fragment-Based Approach for Accurate Calculations on Large Molecules and Materials

Nicholas J. Mayhall and Krishnan Raghavachari*

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

ABSTRACT: We present a new extrapolated fragment-based approach, termed molecules-in-molecules (MIM), for accurate energy calculations on large molecules. In this method, we use a multilevel partitioning approach coupled with electronic structure studies at multiple levels of theory to provide a hierarchical strategy for systematically improving the computed results. In particular, we use a generalized hybrid energy expression, similar in spirit to that in the popular ONIOM methodology, that can be combined easily with any fragmentation procedure. In the current work, we explore a MIM scheme which first partitions a molecule into nonoverlapping fragments and then recombines the interacting fragments to form overlapping subsystems. By including all interactions with a cheaper level of theory, the MIM approach is shown to significantly reduce the errors arising from a single level fragmentation procedure. We report the implementation of energies and gradients and the initial assessment of the MIM method using both biological and materials systems as test cases.

I. INTRODUCTION

Although electronic structure theory can now routinely yield chemically accurate results for small molecules,^{1–17} the steep computational scaling of the most accurate methods precludes their direct application to large molecular systems. For sufficiently local chemical processes, hybrid-energy methods^{18–20} provide a strategy for using accurate (and expensive) computational methods on only those parts of the molecule that comprise the chemically active region. Morokuma and co-workers have developed a particularly useful hybrid energy method called ONIOM (our own N-layer integrated molecular orbital molecular mechanics),^{21–27} in which the hybrid energy for a system divided into two regions (I and II) is given as

$$E_{\text{ONIOM}} = E(\text{I} + \text{II})_{\text{Low-Level}} + E(\text{I})_{\text{High-Level}} - E(\text{I})_{\text{Low-Level}} \quad (1)$$

Although hybrid methods such as QM/MM (quantum mechanics/molecular mechanics) and ONIOM have been incredibly useful in a variety of situations, the necessity to identify a chemically active region prevents them from being truly black-box methods. Moreover, it is clearly important to develop theoretical methods that call for a uniformly accurate treatment of the entire molecule without any bias.

To develop a strategy for performing high level of theory computations on large systems, one must also overcome the $O(N^5)$ – $O(N^8)$ computational scaling of the most accurate electronic structure methods. In this direction, several researchers have focused their efforts on developing linear scaling methods which exploit the local nature of a molecule's electronic structure in a single calculation.^{28–52} In the alternative fragmentation-based approach, significant scaling reductions can be achieved by dividing a molecule into smaller fragments, performing electronic structure calculations on each independently and then summing up the results appropriately. An additional, and possibly equally important, benefit of the

fragmentation approach is the inherent ease in constructing a massively parallel implementation.

One of the earliest fragment-based techniques developed is the fragment molecular orbital method (FMO) of Kitaura and co-workers, in which the fragment subsystems are embedded in a Coulomb bath.^{53–59} Gadre and co-workers have developed the cardinality-guided molecular tailoring approach (CG-MTA) for energies, gradients, and properties.^{60–66} Other fragmentation methods include the molecular fractionation with conjugate caps (MFCC) method of Zhang and co-workers,^{67,68} the systematic molecular fragmentation (SMF) method of Collins and co-workers,^{69–73} which was later coupled with the effective fragment potential (EFP) of Gordon and co-workers,^{74–76} Truhlar et al.'s electrostatically embedded many-body (EE-MB) method,^{77–82} the generalized energy-based fragmentation (GEBF) method of Li and co-workers,^{83–85} the kernel energy method (KEM) of Karle and co-workers,^{86,87} the hybrid many-body interaction (HMBI) model of Beran and Nanda,^{88,89} and the multilevel fragment-based approach (MFBA) of Rezač and Salahub.⁹⁰

The various existing fragment-based methods use a variety of different approaches to implement and to perform the computations. However, it is convenient to discuss them in terms of three main components common to most fragmentation methods: (1) partitioning, (2) subsystem formation, and (3) energy summation.

A. Partitioning. First, the system (large molecule) is partitioned into defined subunits or fragments, most often by cleaving single bonds between adjacent non-hydrogen atoms. This would, as an example, divide hexane into six fragments: $2(-\text{CH}_3)$ and $4(-\text{CH}_2-)$. Less general but possibly more useful fragmentation schemes may be employed if one is well versed in the chemical composition of the target system. For example, as the

Received: January 17, 2011

Published: April 05, 2011

MFCC and MFBA methods have been developed for modeling polypeptide systems, the fragmentation scheme in these two methods involves only the cleavage of specific bonds in the peptide backbone. Since the fragments typically have dangling bonds, direct electronic structure calculations on the fragments are not possible without further manipulation in most cases. The fragments are analogous to “atoms” and make up the fundamental building blocks of the system.

B. Subsystem Formation. After partitioning the molecule into fragments, the interacting fragments are combined to form subsystems that can be used in individual electronic structure calculations. The subsystems must be constructed in a way which provides the appropriate balance between accuracy and computational efficiency. A key step in subsystem formation involves the capping of truncated bonds to form well-defined units (i.e., small molecules) on which direct electronic structure calculations can be performed. Most methods (i.e., MFCC, MFBA, GEBF) use hydrogen atoms to cap the dangling bonds, whereas the FMO methods use a more sophisticated potential to satisfy valencies. For the formation of subsystems involving a systematic inclusion of interfragment interactions, many methods (FMO, EE-MB, MFCC, MFBA, SMF, KEM, HMBI) employ the many-body expansion described by Xantheas.⁹¹ This provides a simple way to systematically improve one’s results, albeit at greatly increasing computational cost for high expansion orders. Linear scaling can be recovered by including only the n -body interactions which fall within a specified interfragment distance. Typically, however, most such calculations include only two-body (and three-body, if possible) interactions.

In an alternative but equally powerful approach, one may form overlapping subsystems comprised of a central fragment surrounded by proximal fragments (either in connectivity or spatial distance). The overlap denotes regions where the interfragment interactions are overcounted and must be appropriately subtracted out in accordance with the Inclusion–Exclusion Principle:

$$|A_1 \cdots \cup A_n| = \sum_i |A_i| - \sum_{i < j} |A_i \cap A_j| + \sum_{i < j < k} |A_i \cap A_j \cap A_k| \dots + (-1)^{n-1} |A_i \cap \dots \cap A_n| \quad (2)$$

This is the approach taken by the GEBF and CG-MTA methods.

C. Energy Summation. The final part involves a summation of the energies from the individual subsystems (small molecules) to yield the total energy for the entire system (large molecule). The overall energy expression is clearly dependent on the manner in which the individual subsystems are formed. Methods which use a many-body expansion are then summed according to the form of the expansion. The energy of methods which use overlapping subsystems must be assembled more carefully, as the energy for each subsystem must be summed along with its appropriate coefficient. This will be discussed in more detail in the Methodology section.

In this contribution, we report the development and initial assessment of a new hybrid energy method which we have termed molecules-in-molecules (MIM). In this method, we use multiple levels of theory to extrapolate the fragment-based energy to obtain better convergence of the total energy of the large molecule. Our method is similar in spirit to the popular

ONIOM methodology (*vide infra*) and allows one to couple very accurate electronic structure methods, performed at a modest level of fragmentation (small subsystems), with cheaper methods performed at more aggressive levels of fragmentation (larger subsystems). We use features from many of the existing approaches and generalize other features.

II. METHODOLOGY

Most previous treatments consider a single level of partitioning to create the fragments and treat the resulting subsystems (interacting fragments including capping) with one or more levels of theory. The central point of the MIM method is that it is a multilevel fragment/subsystem approach, in which cheaper *ab initio* or semiempirical methods can be used to describe longer range interactions. In the most straightforward approach, after the initial partitioning, subsystems of different sizes can be generated by using different cutoff distances to describe the interactions between fragments. Alternatively, different partitioning schemes can be used to create fragments of different sizes. As an example, in a three-level MIM scheme (MIM3), the entire molecule can be treated at a low level of theory (e.g., HF), medium subsystems can be treated at a medium level of theory (e.g., MP2), and small subsystems can be treated at a high level of theory (e.g., CCSD). For a biomolecule containing hundreds of atoms, in a two-level MIM scheme (MIM2), B3LYP can be used to treat the subsystems while the entire molecule can be treated with PM6 to provide a correction for long-range interactions. This is very similar in spirit to the popular ONIOM approach by Morokuma, though the subsystems are not centered on the active site (as in ONIOM) but are individually centered throughout the large molecule (in MIM). Thus, the energy expressions can also be written in a manner similar to that in the ONIOM approach. However, unlike in ONIOM we are approximating the high level of theory on the whole molecule. Therefore, the error associated with the fragmentation method is

$$\text{Error} = E_{\text{High}}^r - E_{\text{High}} \quad (3)$$

In our method, we extrapolate the fragmentation energy by approximating the error of the fragmentation method with a more efficient level of theory as shown here:

$$E_{\text{High}} = E_{\text{High}}^r - (E_{\text{High}}^r - E_{\text{High}}) \approx E_{\text{High}}^r - (E_{\text{Low}}^r - E_{\text{Low}}) \quad (4)$$

Use of the approximation shown in eq 4 sets up a general hierarchy for an arbitrary number of extrapolations.

$$E^{\text{MIM1}} = E_{\text{High}}^r \quad (5)$$

$$E^{\text{MIM2}} = E_{\text{High}}^r - (E_{\text{Low}}^r - E_{\text{Low}}^\infty) \quad (6)$$

$$E^{\text{MIM3}} = E_{\text{High}}^r - (E_{\text{Med}}^r - E_{\text{Med}}^r) - (E_{\text{Low}}^r - E_{\text{Low}}^\infty) \quad (7)$$

where the superscript r indicates the accuracy threshold parameter used in that fragmentation level (i.e., distance/number cutoff or the order of a many-body expansion), and the Low, Med, and High refer to the low level, medium level, and high level of theory, respectively. It should be noted, however, that E_{High}^r and E_{Low}^r each represent a composite energy for the entire system assembled from a number of subsystem calculations at the fragmentation level

denoted by the parameter r . Since the energy of the entire system (large molecule) is assembled from the individual energies of the subsystems (small molecules), we label our method as “molecules in molecules” (MIM).

In eq 7, the MIM3 expression now has two defined parameters, both r and r' , where $r > r'$. The method can also be generalized to more than three levels (multilevel MIM) in more complex systems in an analogous manner. Although we have used the full system calculation in the above equations, E_{Low}^∞ we do not necessarily have to extrapolate to the full calculation, which may be intractable for very large systems. We can alternatively decide to extrapolate only to a more accurate method of fragmentation, such as in

$$E_{MIM2} = E_{High}^r - (E_{Low}^r - E_{Low}^{r'}) \quad (8)$$

where $r' \gg r$. However, in this paper, we will only be considering the case in which $r' = \infty$.

The energy expressions given in eqs 5–7 are generic, and independent of the actual procedure used for generating the subsystems. Therefore, it may be used with any fragmentation procedure such as the GEBF or the FMO methods, allowing one to couple various fragmentation methods together. Furthermore, as the MIM approach uses an ONIOM framework, it exhibits several attractive features: (1) When the two (or three) levels of theory become identical, the exact energy (E_{High}^∞) is recovered. (2) When the distance parameter r becomes sufficiently large, the exact energy (E_{High}^∞) is recovered. (3) As all individual subcalculations are performed on well-defined molecular systems, any electronic structure theory method may be used, which allows this approach to take advantage of the most recent advances in semiempirical methods or electron correlation methods. (4) Finally, an efficient parallel implementation is easy since the individual subsystem calculations can be carried out on different processors. The basic methodology of the MIM method is displayed in Figure 1 with a comparison to the ONIOM methodology.

In the initial implementation of the MIM method reported in this paper, we use a fragmentation scheme similar to the GEBF method.⁸⁵ In the GEBF approach, all important interactions must be included in the subsystem calculations. However, since we are using a multilevel approach, we should be able to relax our conditions for assembling subsystems to permit smaller calculations, while picking up the additional long-range interactions with the cheaper methods.

For a MIM calculation to be performed, each level of calculation is carried out in the following manner:

- 1 **Defining Fragments.** To form the initial fragments, a connectivity table is processed, and all single bonds between heavy atoms are cleaved. As noted earlier, fragments are the most fundamental units considered and are never broken up in the subcalculations. Therefore, a manual modification of the connectivity table provides one with a simple mechanism for controlling the composition of the subsystems generated by the automated program. Customized fragment lists may also be used for generating the subsystems in the next step.
- 2 **Primary Subsystem Formation.** Subsystems which are centered on a particular fragment with nearby fragments appended are called *primary subsystems*.¹⁰⁸ Primary subsystems may be assembled either by appending all fragments which are

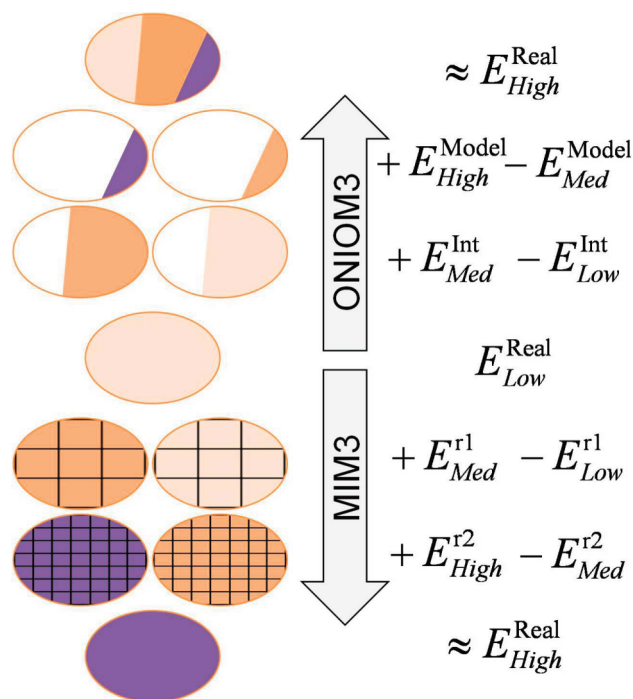


Figure 1. Schematic representation of the MIM methodology. The faint orange areas represent regions computed with the low level of theory. The strong orange areas represent regions computed with the medium level of theory. The purple areas represent regions computed with the high level of theory.

within some distance cutoff r (distance-based cutoff) or by appending a given number of fragments η which are closest to the central fragment (number-based cutoff).⁸⁵ We assemble the primary subsystems according to the following pseudocode:

- 1 Define r , for distance-based cutoff, or η , for number-based cutoff.
- 2 For each fragment (f)
 - a Initiate a new primary subsystem (p).
 - b If distance-based cutoff:
 - Append all fragments to the primary subsystem that are less than distance r away from the central fragment (f)
 - c If number-based cutoff:
 - Append $\eta - 1$ fragments to p which are nearest to the central fragment (f). p now contains η fragments.
 - In case of systems with high symmetry, include any additional fragments if they are at the same distance as the farthest included fragment.
 - d Are there any atoms (a) not in p which are also covalently connected to two or more atoms in p ?
 - If yes:
 - Append a to p .
 - Go to step d again.
- 3 Reduce list of primary subsystems to only unique primary subsystems, which are not subsets of other primary subsystems. Here, step d is done to ensure that the same center is not replaced twice by link atoms, as would be the case when only five atoms of a six-membered ring are part of a primary subsystem.

3 *Derivative Subsystem Formation.* A consequence of using overlapping primary subsystems is that much of the molecule is overcounted. The primary subsystems overlap with one another, and we must therefore construct subsystems which cancel the overlapping regions of the primary subsystems. These are called *derivative subsystems*⁸⁵ and are generated in accordance with eq 2. This can become a time-consuming process for large values of r , though our algorithm as implemented above is sufficient to get quite accurate results for reasonably large molecules (*vide infra*). In the future, we plan to investigate other methods of subsystem determination such as the algorithm put forth by Bettens and Lee.⁹²

While we have used a GEBF-like fragmentation procedure in this initial MIM implementation, we have major differences. The first key difference between the current fragmentation scheme and the GEBF method is that we do not use the “extension rules” employed in constructing the GEBF primary subsystems, as our aim is to use a cheaper electronic structure method for the longer range interactions. This permits us to get accurate results employing significantly smaller high-level subsystem calculations (*vide infra*). Since we are currently not including any high-level electronic embedding in our calculations, our energy gradients are exact.⁸⁵ However, we are working on implementing electronic embedding in a formalism similar to our earlier ONIOM-EE work,^{93–95} in which all of the small subsystem calculations are performed in the presence of the charge distribution obtained in the E_{Low}^{∞} level in a MIM calculation. Electronic embedding done this way will remove the necessity to iteratively obtain charge distributions.

In essence, the MIM method provides a general hierarchy for coupling different electronic structure methods to achieve accurate energies and properties for large molecules. This is, to the best of our knowledge, the first *fully general, extrapolated ONIOM-like methodology for a multilevel fragmentation energy approach* (*vide infra*). The method allows transparent coupling of many different methods including semiempirical schemes, easy implementation of analytical gradients for the exploration of potential energy surfaces or dynamics, and efficient parallelization across many platforms including massively parallel architectures. An additional advantage is that our energy extrapolation procedure is generic and is easily coupled with many possible fragmentation schemes. We have already implemented schemes based on bond-space cutoff, real space (distance-based) cutoff, and number-based cutoff treatments. Moreover, the seamless inclusion of link atoms (when needed) makes it possible to treat bonded systems (such as peptides) or nonbonded systems (such as water clusters) on an equal footing. Another major advantage of our methodology is that multilevel MIM calculations do *include all interactions*, albeit, at the low-level of theory. Finally, unlike most previous schemes, our use of the ONIOM-like extrapolation scheme results in our method having the formal property that it yields exact energies when the different levels of theory become identical.

Of the many previous fragment-based approaches developed by other groups, several share some of the advantageous details featured here. The ability to obtain system–system interactions by using overlapping subsystems has been demonstrated previously by Li et al. (GEBF^{83–85}) and Gadre et al. (CG-MTA^{60–66}). However, both of these methods have thus far used only a single level of fragmentation that attempts to incorporate all significant

interactions into the fragmented energy calculations, e.g., via electrostatic embedding of the subsystems. As an alternative, the MIM approach uses a lower level of theory to compute the energy of the full molecule (or simply a less fragmented system) which does include all interactions. Several previous methods (MC-QM:QM,^{96–100} HMBI,^{88,89} EE-MB-CE,⁷⁸ MFBA,⁹⁰) make use of multiple levels of theory, as we do in this work. Typically, many-body energy expansions are used in these approaches, where the lower levels of theory are employed for evaluations of the higher order many-body terms. However, the MIM method is defined in a general and flexible manner, and thus it can couple both overlapping subsystem approaches and many-body expansion approaches (though the latter has not yet been fully implemented in our code). The more recent XO method¹⁰¹ not only uses overlapping subsystems for describing fragment-fragment interactions but also couples multiple levels of theory, as we do in our current approach. However, the XO method has not yet included a systematic and programmable algorithm for the fragmentation/subsystem assembly procedure as we have done in this contribution. Overall, while MIM is clearly related to many previous fragment-based methods, its broad definition and generality make it an accurate and applicable approach for investigating a wide range of problems for both bonded and nonbonded systems.

In our implementation, Gaussian 09 is used for both the electronic structure calculations and the geometry optimizer.¹⁰²

III. ASSESSMENT OF MIM

A. Absolute Energies—DNA. For the MIM method to provide accurate results, the low level of theory must be capable of recovering the long-range interactions lost in the high-level subsystem calculations. In this section, we couple a semiempirical method (PM6) with an ab initio method (HF/6-31G) to compute the MIM energy of a large molecular system to test the energy convergence with increasing r both with and without the PM6 level. This allows us to gauge the ability to capture long-range effects with only a low-level of theory. As a test molecule, we have chosen the DNA poly(dA·dT) decamer taken from the nucleic acid database,¹⁰³ a system considered previously with the GEBF method.⁸⁵ However, as we are currently not including any solvation effects, the presence of 18 excess electrons may make the results difficult to interpret. We therefore cap each of the phosphate groups with a sodium ion to neutralize the highly anionic systems as we have done in a previous study.¹⁰⁴ Although electron correlation effects are known to be important in the description of π – π interactions, we have used the HF method to permit comparison to previously reported results on a similar system.⁸⁵ The modified system is shown in Figure 2a.

Figure 2b illustrates the deviations in absolute energy between the unfragmented HF/6-31G results and two fragmented schemes: HF/6-31G(r) (black columns) and HF/6-31G:PM6- r) (green columns). The benefits of including the PM6 correction for long-range interactions are immediately seen. The PM6 correction greatly reduces the deviations at all levels of fragmentation.

B. Relative Energies—2NPV. While reproducing the absolute energies of the full unfragmented method is clearly a sufficient condition for accuracy, it is not necessary, as quantum chemistry is largely concerned with energy differences. Therefore, while systematic errors in a fragmentation method may prevent rapid

convergence of the absolute energies, these errors may cancel when computing relative energies.

Here, we assess the ability of the MIM method to reproduce the relative conformer energies of a cyclic lipopeptide surfactin (PDB entry 2NPV).¹⁰⁵ We have chosen 2NPV 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 3a). This molecule has a well-defined, saddle-shaped polypeptide backbone with a long alkyl chain extension. In this example, we use the first 10 conformers given in the PDB file.

Using both distance-based cutoffs (r) and number-based cutoffs (η) for the primary subsystem assembly, we calculated the energy of the 10 conformers with both 1-level [MP2/6-31G*(r or η)] and 2-level [MP2/6-31G*:B3LYP/6-31G*(r or η)] fragmented methods and compared the results to the full, unfragmented MP2/6-31G* (1262 basis functions) relative

energies. This allows us to examine the effect of the low-level correction on relative energies. For the distance-based cutoff, we use an r value of 3.0 Å,¹⁰⁹ and for the number-based cutoff, we use a value of $\eta = 9$. As can be seen in Figure 3c, the low-level (B3LYP/6-31G*) correction greatly improves the results by closely reproducing the unfragmented MP2/6-31G* relative energies. In fact, for all but conformer 4, there is essentially no discernible difference between the MP2:B3LYP($\eta = 9$) results (dashed line) and the MP2 results (solid line), despite significant deviations for the single level fragmented MP2($\eta = 9$) results (dotted line).

The results, summarized in Table 1, show that while the low-level correction does indeed decrease both the RMS and maximum absolute errors, the relative energies are improved more dramatically. A careful analysis of the distance- and number-based cutoff results illustrates the differences in using the different procedures of subsystem assembly. We can make a preliminary observation from this study that the $\eta = 9$ results appear to be more accurate and have smaller individual MP2 calculations while the $r = 3.0$ results require fewer subsystem calculations. These results are clearly impressive and underscore the extent of error cancellation in the fragmented methods. While the errors in absolute energies are still rather large (RMS = 3.27: max = 3.9 kcal/mol for MP2:B3LYP($r = 3.0$) and RMS = 2.44: max = 2.94 kcal/mol for MP2:B3LYP($\eta = 9$)), the errors in relative energies are much smaller; both the RMS and maximum errors (for both r and η) are less than 1 kcal/mol. Our best results are obtained for the number-based cutoff method where the RMS deviation in relative energies falls from 1.18 kcal/mol to an impressive 0.25 kcal/mol on going from MIM1 to MIM2.

C. Geometry Optimizations—Si(100)-2×1. The theoretical study of semiconductor surface chemistry has benefited greatly from the development of efficient and accurate cluster models.¹⁰⁶ However, the computational demands for studying processes such as a chemical line growth across a surface are greatly increased due to the large cluster models required to capture the lateral movement of the adsorbed molecules. Here, we demonstrate the ability to use the MIM method to optimize the geometry of a large silicon cluster with an adsorbed allylic mercaptan molecule.¹⁰⁷ The structure is shown in Figure 4.

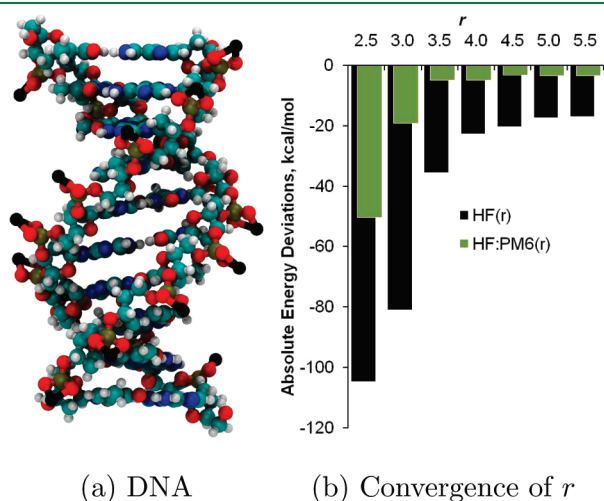


Figure 2. DNA: (a) DNA decamer used in this study. Na⁺ ions are shown as black centers. (b) Convergence of absolute energy with r . Deviations between unfragmented and fragmented calculations reported as HF(r)–HF or HF:PM6(r)–HF.

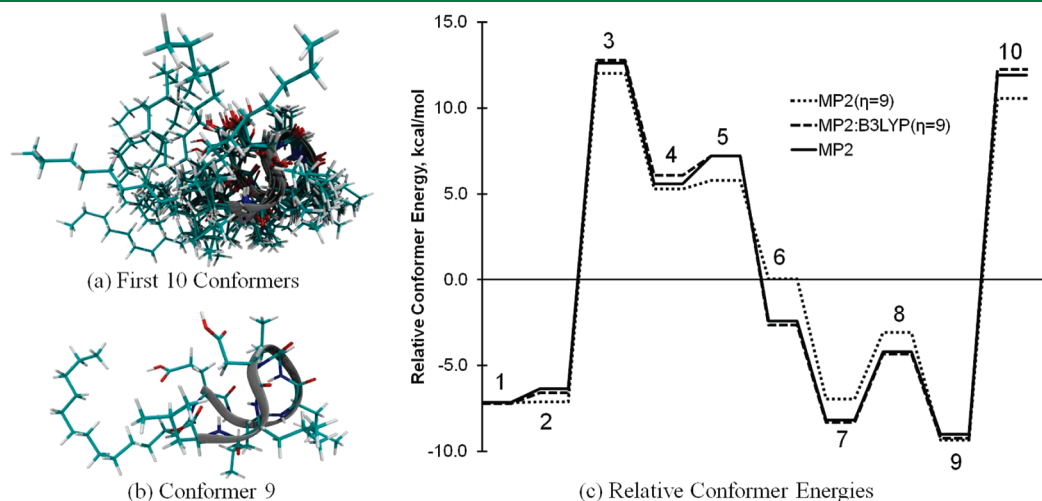


Figure 3. 2NPV: (a) Superposition of the first 10 conformers. (b) Structure of conformer 9 given in the PDB file. (c) Relative energies for the first 10 conformers of 2NPV centered at zero. Dotted line: MP2($\eta = 9$) results. Dashed line: MP2:B3LYP($\eta = 9$) results. Solid line: unfragmented MP2 results. All calculations use the 6-31G* basis set. Relative energies for each method are shown centered at zero.

Table 1. 2NPV—Differences between the Unfragmented MP2 Results and the Fragmented Results^a

RMS(MAX)	MIM1		MIM2	
	MP2($r = 3.0$)	MP2($\eta = 9$)	MP2:B3LYP($r = 3.0$)	MP2:B3LYP($\eta = 9$)
relative, kcal/mol	1.31(2.51)	1.18(2.44)	0.48(0.90)	0.25(0.51)
absolute, kcal/mol	4.69(7.02)	4.02(6.29)	3.27(3.91)	2.44(2.94)

^a Both distance-based cutoffs [MP2($r = 3.0$) and MP2:B3LYP($r = 3.0$)] and number-based cutoffs [MP2($\eta = 9$) and MP2:B3LYP($\eta = 9$)] were used. Energy results in kcal/mol. Values reported are RMS, with MAX values in parentheses. The 6-31G* basis set used in all calculations.

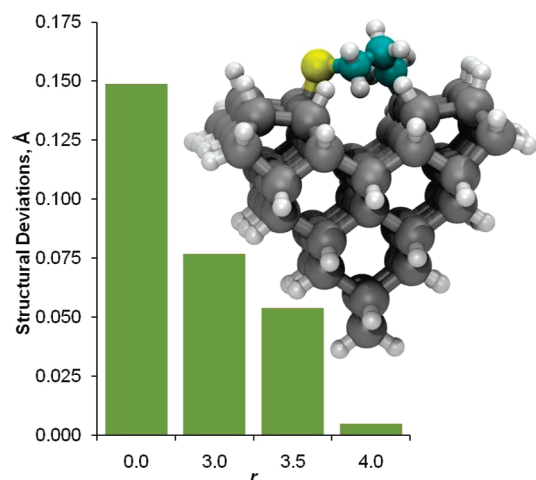


Figure 4. Si(100)—RMS deviations between the B3LYP/6-31G*:AM1(r) optimized geometries and the unfragmented B3LYP/6-31G* optimized geometries. Units in Å.

As the MIM gradient expressions look just like the energy expressions, we can easily obtain MIM gradients from gradient evaluations of the individual subcalculations. Link atom force projections are performed as outlined for the ONIOM methodology. In Figure 4, we show the convergence of the optimized geometries with increasing r . The columns represent RMS deviations (in Å) between the B3LYP/6-31G*:AM1(r) optimized geometry and the unfragmented B3LYP/6-31G* optimized geometry. Using a value of $r = 3.0$, the RMS is nearly halved with respect to the pure AM1 results ($r = 0.0$). Further improvements to the geometry can be made by increasing r , with $r = 4.0$ giving nearly exact results.

IV. CONCLUSIONS

In this paper, we present a new hybrid energy fragmentation method called molecules-in-molecules (MIM), which treats individual fragmentation calculations as levels in an ONIOM-like hybrid energy framework.

As an initial assessment of the approach, we have studied the convergence of the total energy for a large DNA molecule. By comparing to the single-level fragmentation approach, we observe significant improvements to the total energies upon inclusion of a low-level correction.

By calculating the relative conformer energies of a surfactin molecule, we report improvements to the reproduction of MP2 relative energies, with an RMS deviation from the unfragmented MP2 calculations of only 0.25 kcal/mol.

We have also demonstrated the use of the MIM method for the optimization of molecular geometries for a silicon cluster

model system, and our results indicate that accurate molecular geometries can be obtained using the MIM methodology.

Overall, our work suggests that our multilevel MIM approach can be employed for a wide range of large systems with high accuracy.

Finally, a cautionary note on the applicability of MIM is warranted. Central to our method is the assumption that the electronic characteristics are relatively local or near-sighted. Therefore, classes of molecular systems, such as metallic or highly conjugated systems, which are characterized by long-range electron delocalization are not expected to be treatable with the MIM methodology. Further, as we are currently using only hydrogen atoms for capping truncated bonds, our current implementation is only expected to be successful for those systems which are dominated by covalent bonds. Fragmentation of dative or ionic bonds is not recommended. However, we are currently investigating alternative bond-capping approaches (such as the use of pseudoatoms) for various bonding scenarios.

AUTHOR INFORMATION

Corresponding Author

*E-mail: kraghava@indiana.edu.

ACKNOWLEDGMENT

This work was supported by a National Science Foundation grant, CHE-0911454, at Indiana University. The authors would like to thank Professor Balaji Raghavachari for insightful discussions facilitating the development of this work.

REFERENCES

- (1) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2007**, *126*, 084108–084119.
- (2) DeYonker, N. J.; Cundari, T. R.; Wilson, A. K. *J. Chem. Phys.* **2006**, *124*, 114104.
- (3) Petersson, G. A.; Bennett, A.; Tensfeldt, T. G.; Allaham, M. A.; Shirley, W. A.; Mantzaris, J. J. *J. Chem. Phys.* **1988**, *89*, 2193–2218.
- (4) Ochterski, J. W.; Petersson, G. A.; Montgomery, J. A. *J. Chem. Phys.* **1996**, *104*, 2598–2619.
- (5) Karton, A.; Rabinovich, E.; Martin, J. M.; Ruscic, B. *J. Chem. Phys.* **2006**, *125*, 144108.
- (6) Tajti, A.; Szalay, P.; Csaszar, A.; Kallay, M.; Gauss, J.; Valeev, E.; Flowers, B.; Vazquez, J.; Stanton, J. F. *J. Chem. Phys.* **2004**, *121*, 11599.
- (7) Bomble, Y. J.; Vazquez, J.; Kallay, M.; Michauk, C.; Szalay, P. G.; Csaszar, A. G.; Gauss, J.; Stanton, J. F. *J. Chem. Phys.* **2006**, *125*, 064108.
- (8) Harding, M. E.; Vazquez, J.; Ruscic, B.; Wilson, A. K.; Gauss, J.; Stanton, J. F. *J. Chem. Phys.* **2008**, *128*, 114111.
- (9) Kowalski, K.; Piecuch, P. *J. Chem. Phys.* **2000**, *113*, 18–35.
- (10) Piecuch, P.; Kucharski, S. A.; Kowalski, K.; Musial, M. *Comput. Phys. Commun.* **2002**, *149*, 71–96.
- (11) Martin, J. M. L.; de Oliveira, G. *J. Chem. Phys.* **1999**, *111*, 1843–1856.

- (12) Boese, A. D.; Oren, M.; Atasoylu, O.; Martin, J. M. L.; Kallay, M.; Gauss, J. *J. Chem. Phys.* **2004**, *120*, 4129–4141.
- (13) Karton, A.; Taylor, P. R.; Martin, J. M. L. *J. Chem. Phys.* **2007**, *127*, 064104.
- (14) Karton, A.; Parthiban, S.; Martin, J. M. L. *J. Phys. Chem. A* **2009**, *113*, 4802–4816.
- (15) Karton, A.; Martin, J. M. L. *J. Chem. Phys.* **2010**, *133*, 144102.
- (16) Feller, D.; Dixon, D. A. *J. Phys. Chem. A* **2000**, *104*, 3048–3056.
- (17) Barnes, E. C.; Petersson, G. A.; Montgomery, J. A.; Frisch, M. J.; Martin, J. M. L. *J. Chem. Theory Comput.* **2009**, *5*, 2687–2693.
- (18) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700–733.
- (19) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170–1179.
- (20) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718–730.
- (21) Humbel, S.; Sieber, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *105*, 1959–1967.
- (22) Svensson, M.; Humbel, S.; Froese, R.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357–19363.
- (23) Karadakov, P. B.; Morokuma, K. *Chem. Phys. Lett.* **2000**, *317*, 589–596.
- (24) Vreven, T.; Morokuma, K. *J. Comput. Chem.* **2000**, *21*, 1419–1432.
- (25) Vreven, T.; Mennucci, B.; da Silva, C.; Morokuma, K.; Tomasi, J. *J. Chem. Phys.* **2001**, *115*, 62–72.
- (26) Vreven, T.; Morokuma, K. *Theor. Chem. Acc.* **2003**, *109*, 125–132.
- (27) Rega, N.; Iyengar, S.; Voth, G.; Schlegel, H.; Vreven, T.; Frisch, M. J. *J. Phys. Chem. B* **2004**, *108*, 4210–4220.
- (28) He, X.; Merz, K. M. *J. Chem. Theory Comput.* **2010**, *6*, 405–411.
- (29) Kobayashi, M.; Nakai, H. *J. Chem. Phys.* **2009**, *131*, 114108.
- (30) Kobayashi, M.; Imamura, Y.; Nakai, H. *J. Chem. Phys.* **2007**, *127*, 074103.
- (31) Monard, G.; Bernal-Uruchurtu, M. I.; van der Vaart, A.; Merz, K. M.; Ruiz-Lopez, M. F. *J. Phys. Chem. A* **2005**, *109*, 3425–3432.
- (32) der Vaart, A. V.; Gogonea, V.; Dixon, S. L.; Merz, K. M. *J. Comput. Chem.* **2000**, *21*, 1494–1504.
- (33) Yang, W. T. *Phys. Rev. Lett.* **1991**, *66*, 1438–1441.
- (34) Makowski, M.; Korchowiec, J.; Gu, F. L.; Aoki, Y. *J. Comput. Chem.* **2010**, *31*, 1733–1740.
- (35) Imamura, A.; Aoki, Y.; Maekawa, K. *J. Chem. Phys.* **1991**, *95*, 5419.
- (36) Saebo, S.; Pulay, P. *J. Chem. Phys.* **1987**, *86*, 914–922.
- (37) Maslen, P.; Head-Gordon, M. *Chem. Phys. Lett.* **1998**, *283*, 102–108.
- (38) Lee, M. S.; Maslen, P. E.; Head-Gordon, M. *J. Chem. Phys.* **2000**, *112*, 3592–3601.
- (39) Subotnik, J. E.; Head-Gordon, M. *J. Chem. Phys.* **2005**, *123*, 064108.
- (40) Pisani, C.; Maschio, L.; Casassa, S.; Halo, M.; Schtz, M.; Usvyat, D. *J. Comput. Chem.* **2008**, *29*, 2113.
- (41) Hampel, C.; Werner, H.-J. *J. Chem. Phys.* **1996**, *104*, 6286–6297.
- (42) Adler, T. B.; Werner, H.-J. *J. Chem. Phys.* **2009**, *130*, 241101.
- (43) Flocke, N.; Bartlett, R. J. *J. Chem. Phys.* **2004**, *121*, 10935–10944.
- (44) Stratmann, R. E.; Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys. Lett.* **1996**, *257*, 213–223.
- (45) Strain, M. C.; Scuseria, G. E.; Frisch, M. J. *Science* **1996**, *271*, 51–53.
- (46) Scuseria, G. E.; Ayala, P. Y. *J. Chem. Phys.* **1999**, *111*, 8330–8343.
- (47) Scuseria, G. E. *J. Phys. Chem. A* **1999**, *103*, 4782–4790.
- (48) Millam, J. M.; Scuseria, G. E. *J. Chem. Phys.* **1997**, *106*, 5569–5577.
- (49) Kudin, K. N.; Scuseria, G. E. *Phys. Rev. B* **2000**, *61*, 16440–16453.
- (50) Daniels, A. D.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 1321–1328.
- (51) Burant, J. C.; Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys.* **1996**, *105*, 8969–8972.
- (52) Ayala, P. Y.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 3660–3671.
- (53) Fujimoto, H.; Koga, N.; Fukui, K. *J. Am. Chem. Soc.* **1981**, *103*, 7452–7457.
- (54) Fedorov, D. G.; Ishida, T.; Kitaura, K. *J. Phys. Chem. A* **2005**, *109*, 2638–2646.
- (55) Fedorov, D. G.; Ishida, T.; Uebayasi, M.; Kitaura, K. *J. Phys. Chem. A* **2007**, *111*, 2722–2732.
- (56) Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. A* **2007**, *111*, 6904–6914.
- (57) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701–706.
- (58) Fedorov, D. G.; Slipchenko, L. V.; Kitaura, K. *J. Phys. Chem. A* **2010**, *114*, 8742–8753.
- (59) Pruitt, S. R.; Fedorov, D. G.; Kitaura, K.; Gordon, M. S. *J. Chem. Theory Comput.* **2010**, *6*, 1–5.
- (60) Gadre, S. R.; Shirsat, R. N.; Limaye, A. C. *J. Phys. Chem.* **1994**, *98*, 9165–9169.
- (61) Gadre, S. R.; Ganesh, V. *J. Theor. Comput. Chem.* **2006**, *5*, 835–855.
- (62) Ganesh, V.; Dongare, R. K.; Balanarayan, P.; Gadre, S. R. *J. Chem. Phys.* **2006**, *125*, 104109.
- (63) Kavathekar, R.; Khire, S.; Ganesh, V.; Rahalkar, A. P.; Gadre, S. R. *J. Comput. Chem.* **2009**, *30*, 1167–1173.
- (64) Rahalkar, A. P.; Ganesh, V.; Gadre, S. R. *J. Chem. Phys.* **2008**, *129*, 234101.
- (65) Rahalkar, A. P.; Katouda, M.; Gadre, S. R.; Nagase, S. *J. Comput. Chem.* **2010**, *31*, 2405–2418.
- (66) Yeole, S. D.; Gadre, S. R. *J. Chem. Phys.* **2010**, *132*, 094102.
- (67) Mei, Y.; Ji, C. G.; Zhang, J. Z. H. *J. Chem. Phys.* **2006**, *125*, 094906.
- (68) Zhang, D. W.; Zhang, J. Z. H. *J. Chem. Phys.* **2003**, *119*, 3599–3605.
- (69) Addicoat, M. A.; Collins, M. A. *J. Chem. Phys.* **2009**, *131*, 104103.
- (70) Collins, M. A.; Deev, V. A. *J. Chem. Phys.* **2006**, *125*, 104104.
- (71) Collins, M. A. *J. Chem. Phys.* **2007**, *127*, 024104.
- (72) Deev, V.; Collins, M. A. *J. Chem. Phys.* **2005**, *122*, 154102.
- (73) Netzloff, H. M.; Collins, M. A. *J. Chem. Phys.* **2007**, *127*, 134113.
- (74) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293–307.
- (75) Mullin, J. M.; Roskop, L. B.; Pruitt, S. R.; Collins, M. A.; Gordon, M. S. *J. Phys. Chem. A* **2009**, *113*, 10040–10049.
- (76) Slipchenko, L. V.; Gordon, M. S. *J. Comput. Chem.* **2007**, *28*, 276–291.
- (77) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 46–53.
- (78) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 1342–1348.
- (79) Dahlke, E. E.; Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 33–41.
- (80) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 1–6.
- (81) Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009**, *5*, 1573–1584.
- (82) Sorkin, A.; Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 683–688.
- (83) Li, W.; Li, S.; Jiang, Y. *J. Phys. Chem. A* **2007**, *111*, 2193–2199.
- (84) Hua, W. J.; Fang, T.; Li, W.; Yu, J. G.; Li, S. H. *J. Phys. Chem. A* **2008**, *112*, 10864–10872.
- (85) Hua, S. G.; Hua, W. J.; Li, S. H. *J. Phys. Chem. A* **2010**, *114*, 8126–8134.
- (86) Huang, L.; Massa, L.; Karle, J. *Proc. Natl. Acad. Sci.* **2006**, *103*, 1233.
- (87) Huang, L.; Massa, L.; Karle, J. *Int. J. Quantum Chem.* **2005**, *103*, 808.
- (88) Beran, G. J. O. *J. Chem. Phys.* **2009**, *130*, 164115.
- (89) Beran, G. J. O.; Nanda, K. *J. Phys. Chem. Lett.* **2010**, *1*, 3480.
- (90) Řezáč, J.; Salahub, D. R. *J. Chem. Theory Comput.* **2010**, *6*, 91–99.
- (91) Xantheas, S. S. *J. Chem. Phys.* **1994**, *100*, 7523.
- (92) Bettens, R. P. A.; Lee, A. M. *J. Phys. Chem. A* **2006**, *110*, 8777–8785.
- (93) Hratchian, H. P.; Parandekar, P. V.; Raghavachari, K.; Frisch, M. J.; Vreven, T. *J. Chem. Phys.* **2008**, *128*, 034107.
- (94) Mayhall, N. J.; Raghavachari, K.; Hratchian, H. P. *J. Chem. Phys.* **2010**, *132*, 114107.
- (95) Parandekar, P. V.; Hratchian, H. P.; Raghavachari, K. *J. Chem. Phys.* **2008**, *129*, 145101.
- (96) Elsohly, A. M.; Shaw, C. L.; Guice, M. E.; Smith, B. D.; Tschumper, G. S. *Mol. Phys.* **2007**, *105*, 2777–2782.

- (97) Hopkins, B. W.; Tschumper, G. S. *Chem. Phys. Lett.* **2005**, 407, 362–367.
- (98) Hopkins, B. W.; Tschumper, G. S. *Mol. Phys.* **2005**, 103, 309–315.
- (99) Hopkins, B. W.; Tschumper, G. S. *J. Comput. Chem.* **2003**, 24, 1563–1568.
- (100) Tschumper, G. S. *Chem. Phys. Lett.* **2006**, 427, 185–191.
- (101) Guo, W.; Wua, A.; Xu, X. *Chem. Phys. Lett.* **2010**, 498, 203–208.
- (102) 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, A.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision A.1; Gaussian Inc.: Wallingford, CT, 2009.
- (103) <http://ndbserver.rutgers.edu> (accessed: 12/2010).
- (104) Herbert, H. E.; Halls, M. D.; Hratchian, H. P.; Raghavachari, K. *J. Phys. Chem. B* **2006**, 110, 3336.
- (105) Tsan, P.; Volpon, L.; Besson, F.; Lancelin, J.-M. *J. Am. Chem. Soc.* **2007**, 129, 1968–1977.
- (106) Raghavachari, K.; Halls, M. D. *Mol. Phys.* **2004**, 102, 381–393.
- (107) Ferguson, G. A.; Than, C. T.-L.; Raghavachari, K. *J. Phys. Chem. Lett.* **2010**, 1, 679–685.
- (108) In the GEBF work, these subsystems are referred to as primitives.⁸⁵
- (109) To obtain this value of r , we first performed a series of calculations with a much cheaper B3LYP:PM6(r) model to calibrate the convergence of r . For $r = 30$ Å, both the RMS and the MAX relative conformer energy deviations between the fragmented B3LYP:PM6(3.0) and the unfragmented B3LYP methods were less than 1 kcal/mol (RMS = 0.46 kcal/mol; MAX = 0.92 kcal/mol).
- (110) The largest number of calculations required for a MIM(r) energy was 311, whereas for a MIM(η) energy, it was 507. The largest MP2 subcalculation for MIM(r) had 580 basis functions, whereas MIM(η) had only 402 basis functions. A more detailed comparison of the different types of cutoffs (r or η) will be the focus of future work.