

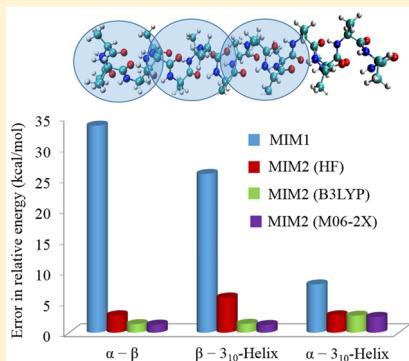
Analysis of Different Fragmentation Strategies on a Variety of Large Peptides: Implementation of a Low Level of Theory in Fragment-Based Methods Can Be a Crucial Factor

Arjun Saha and Krishnan Raghavachari*

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

Supporting Information

ABSTRACT: We have investigated the performance of two classes of fragmentation methods developed in our group (Molecules-in-Molecules (MIM) and Many-Overlapping-Body (MOB) expansion), to reproduce the unfragmented MP2 energies on a test set composed of 10 small to large biomolecules. They have also been assessed to recover the relative energies of different motifs of the acetyl(ala)₁₈NH₂ system. Performance of different bond-cutting environments and the use of Hartree–Fock and different density functionals (as a low level of theory) in conjunction with the fragmentation strategies have been analyzed. Our investigation shows that while a low level of theory (for recovering long-range interactions) may not be necessary for small peptides, it provides a very effective strategy to accurately reproduce the total and relative energies of larger peptides such as the different motifs of the acetyl(ala)₁₈NH₂ system. Employing M06-2X as the low level of theory, the calculated mean total energy deviation (maximum deviation) in the total MP2 energies for the 10 molecules in the test set at MIM($d=3.5\text{\AA}$), MIM($\eta=9$), and MOB($d=5\text{\AA}$) are 1.16 (2.31), 0.72 (1.87), and 0.43 (2.02) kcal/mol, respectively. The excellent performance suggests that such fragment-based methods should be of general use for the computation of accurate energies of large biomolecular systems.



1. INTRODUCTION

Modern quantum mechanical methods have demonstrated their reliability in describing the chemistry of small molecular systems with high accuracy from first principles. Methods based on density functional theory (DFT) allow theoretical chemists to evaluate the properties of fairly large molecules with varying levels of success that depend on the exchange-correlation functional being employed and on the nature of the chemical system being investigated. However, accurate post-Hartree–Fock *ab initio* methods such as CCSD(T) that yield more robust results on a broad range of challenging applications scale steeply as a function of the size of the system. For example, popular correlated *ab initio* methods such as MP2, CCSD, and CCSD(T), scale as N^5 , N^6 , and N^7 , respectively, where N represents the size of the system. Brute force application of such methods for large molecules is still a formidable challenge.

Morokuma et al.^{1–4} have provided a very useful general hybrid energy method called “ONIOM” (our own N -layer integrated molecular orbital molecular mechanics) to compute extrapolated energies for large molecular systems. More generally, “fragment-based methods” have been developed by several research groups and constitute some of the most powerful quantum chemical tools to predict the chemistry of large molecular systems.⁵ The goal in these methods is to make an accurate calculation on a large system computationally tractable by performing independent electronic structure calculations on small portions of the system at a time and

assembling their energies to predict an accurate extrapolated energy for the whole system. Starting from the FMO (fragment molecular orbital) method by Kitaura and co-workers,^{6–9} a broad range of fragment-based methods has been proposed including the following: cardinality-guided molecular tailoring approach (CG-MTA) by Gadre et al.,^{10–17} generalized energy based fragmentation (GEBF) approach by Li et al.,^{18–23} molecular fractionation with conjugate caps (MFCC) approach by Zhang et al.,^{24–29} systematic molecular fragmentation (SMF) method by Collins et al.,^{30–33} combined fragmentation method (CFM) proposed by Bettens and Lee,^{34–36} kernel energy method (KEM) by Huang et al.,^{37,38} electrostatically-embedded many-body (EE-MB) approach by Truhlar et al.,^{39–43} multilevel fragment-based approach (MFBA) approach by Rezac and Salahub,⁴⁴ hybrid many-body interaction (HMBI) method by Beran et al.,^{45–47} multicentered QM:QM method (MC QM:QM) developed by Hopkins and Tschumper,^{48–50} molecules-in-molecules (MIM),⁵¹ many-overlapping-body expansion (MOB),⁵² and dimers of dimers (DOD) developed by Raghavachari et al.,⁵³ and generalized many-body expansion (GMBE) by Richard and Herbert.⁵⁴ All of these methods share the ultimate goal of trying to achieve linear scaling but differ in their strategies for generating the fragments and for describing the local environment in each independent electronic structure calculation.

Received: November 23, 2014

In our group we have been exploring different fragment-based methodologies. Recently we have proposed two useful fragment-based methods termed as “MIM (Molecules-in-Molecules)⁵¹ and “MOB” (Many-Overlapping-Body)⁵² expansion. In the present work, we have applied both of these fragment-based approaches to carefully chosen challenging biomolecular test systems. The selection of the test systems has been carried out with several major objectives. First, we have selected a range of small to large peptides to analyze if their energies using fragment-based methods can be reproduced on an even footing. The ability to describe small to large systems evenly may be important for processes such as growth of peptides. Second, we have considered open as well as two different helical isomers of a large peptide to determine their relative energies. This may be important to describe processes such as the folding of peptides where the shape as well as intramolecular interactions change between the different conformations. Third, since weak interactions are prevalent in biomolecules, we have selected a correlated MP2 level of theory as our target method and carried out some large benchmark calculations. Simple models such as Hartree–Fock (HF) or standard DFT such as B3LYP are highly deficient in describing the relative energies between such isomers (*vide infra*). Finally, we have included some polymers in addition to peptides to have a larger variety of molecules.

The fundamental concepts and some relevant technical details of both of these methods (MIM and MOB) will be briefly described in section 2. Section 3 contains a description of our test set of biomolecules, and a discussion of our results where we perform a careful assessment and analysis of the performance of our methods on these biomolecular systems. Our conclusions are summarized in section 4.

2. METHODS

A. Molecules-in-Molecules. MIM is a fragment-based method⁵¹ that employs a multilayer partitioning technique with multiple levels of theory using a generalized hybrid energy expression, similar in spirit to the ONIOM¹ methodology. The working principles in MIM can be described in four steps: (1) initial fragmentation, (2) primary subsystem formation, (3) derivative subsystem formation, and (4) energy summation. In step 1, the initial “fragments” are generated in this work via a general procedure that breaks each “single” bond between non-hydrogen atoms. For a polyalanine peptide system, for example, this involves cutting the C–C and the N–C backbone bonds as well as the C–C side-chain bonds. In the current project, we also explored the performance of MIM by cutting the peptide bonds (C–N, which may have a partial double bond character) in addition to all of the single bonds. Each of the fragments from step 1 initiate the formation of a primary subsystem in step 2. In this work, formation of primary subsystems resulting from the local interactions between nearby fragments is done via two prescribed fragmentation parameters: distance-based cutoff (r) or number-based cutoff (η). In step 3, derivative subsystems are obtained via the inclusion–exclusion principle to avoid overcounting of the overlapping regions of the primary subsystems. The truncated bonds in the primary and derivative subsystems are then terminated with link hydrogen atoms. Finally in step 4, independent energy calculations are performed on all of the subsystems, and their energies are summed up carefully (taking their signs into account) to predict the total energy of the system. Most importantly, the use of multiple layers with different levels of theory makes it an

efficient extrapolation method compared to many fragment-based methods that use a single layer. More details are available from the original publication.⁵¹

The notation that we follow in this work is as follows. Execution of MIM with a single level of theory is termed as MIM(1) (or MIM1) whereas execution with two different levels of theory (high and low) is termed as MIM(2) (or MIM2). MIM2 essentially indicates a scheme with two fragmentation parameters (e.g., $r < R$). The subsystems generated with the smaller parameter (r) are treated with a higher level of theory and the ones with the larger parameter (R) at a lower level of theory. Thus, the long-range interactions are accounted for at the lower level of theory. If $R = \infty$, the entire molecule can be treated at the lower level of theory. The energy of MIM2 can be written, in general, as

$$E^{\text{MIM2}} = E_{\text{low}}^R - E_{\text{low}}^r + E_{\text{high}}^r \quad (1)$$

Each of the terms is the energy of the total system obtained from the summation of individual energy components for the subsystems at the defined fragmentation parameter. Similar extrapolations can be done up to any arbitrary number of theories depending on the complexity of the system and desired level of accuracy. If only a single level of theory is employed,

$$E^{\text{MIM1}} = E_{\text{high}}^r \quad (2)$$

B. Many-Overlapping-Body Expansion. Many-body expansion is a well-known and widely used method. The energy of a system (containing N fragments) in any many-body expanded fragment-based method is given as

$$E = \sum_i E_i + \sum_{i < j} \Delta E'_{ij} + \sum_{i < j < k} \Delta E'_{ijk} + \dots \quad (3)$$

Here, E is the energy of the supersystem. E_i denotes the energy of fragment i , $\Delta E'_{ij}$ is the two-body interaction term between i and j , and so on. The prime (') in the notation denotes that fragments i and j are assumed not to overlap. In most of the existing fragment-based methods, this expansion is usually truncated at two-body interactions for practical reasons.

Recently we have proposed a method where the many-body energy expression has been generalized to consider overlapping fragments (Many-Overlapping-Body (MOB)).⁵² First, the overlapping fragments and their overlap-canceling derivative subsystems are collectively redefined as “monomers”. The inclusion of two-body interactions between overlapping monomers was suggested as a way of effectively including the critical many-body interactions between fragments to gain accuracy at a tractable cost. The many-body energy expression with overlapping monomers can be given in the generalized form

$$E = \sum_p c_p E_p + \sum_{p < q} c_p c_q \Delta E_{pq} + \sum_{p < q < r} c_p c_q c_r \Delta E_{pqr} + \dots \quad (4)$$

where

$$\Delta E_{pq} = E_{pq} - (E_p + E_q - E_{p \cap q}) \quad (5)$$

c_p and c_q are the generalized monomer coefficients. Unlike MIM, formation of the subsystems in MOB is carried out following the bond connectivity of the system.⁵² This strategy of subsystem formation has been adopted following the connectivity-based hierarchy (CBH) defined by Ramabhadran and Raghavachari,^{55,56} and is related to prior work by Deev and

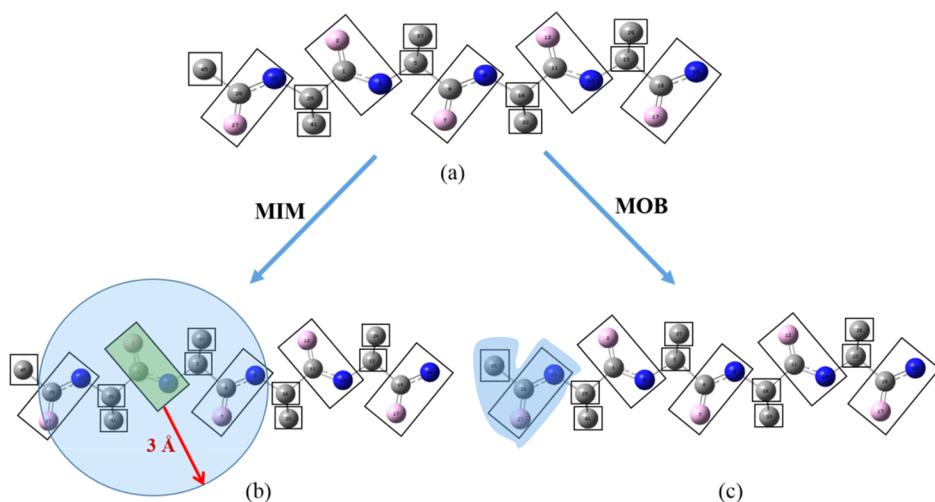


Figure 1. Illustration of MIM and MOB fragmentation schemes on alanine tetrapeptide: (a) fragments generated in both MIM and MOB; (b) primary subsystem generated in distance-based MIM schemes; (c) primary subsystem generated in MOB schemes with monomer degree of 1. Hydrogens are not shown for simplicity.

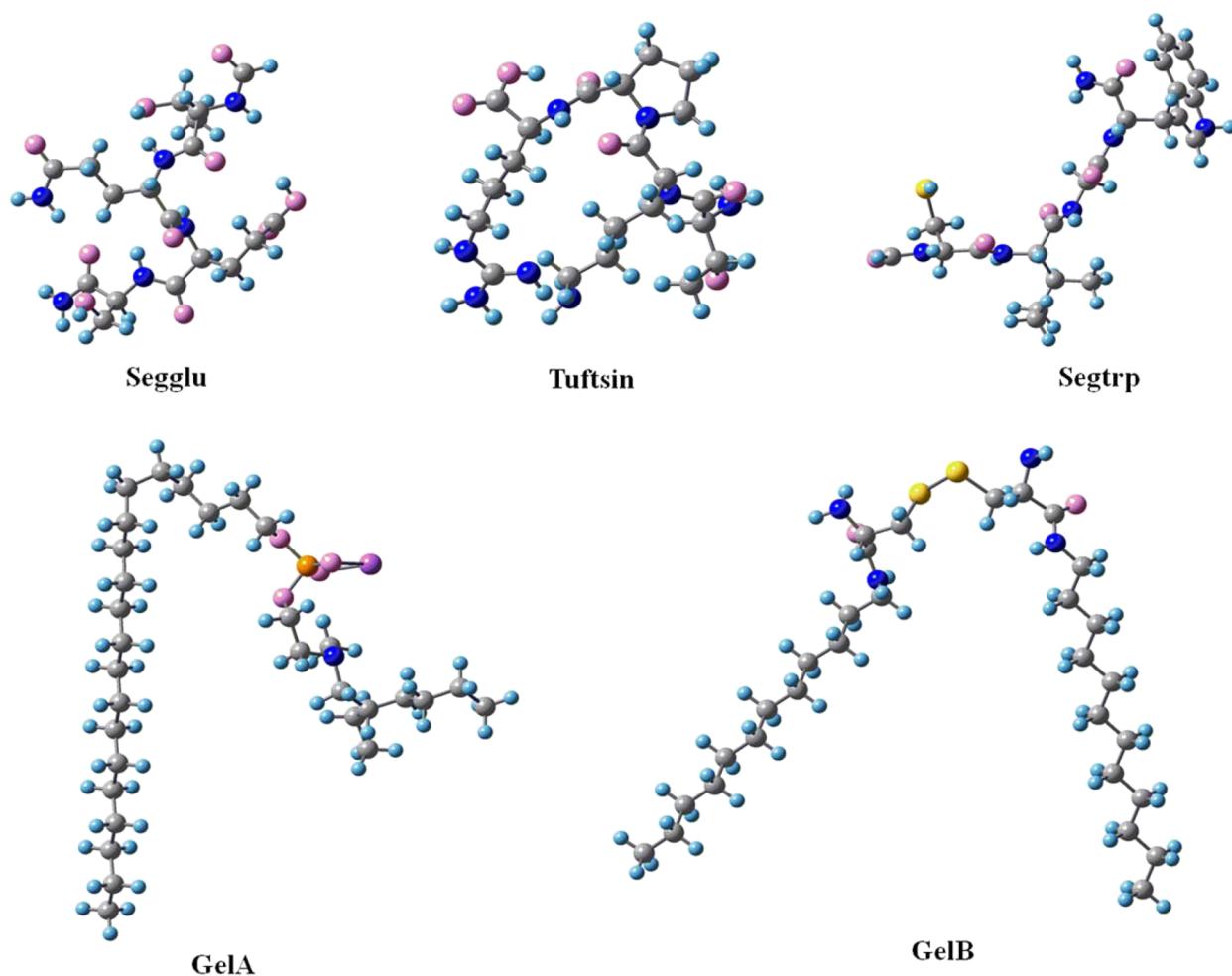


Figure 2. Five small systems studied in this work. Segglu, Tuftsin, and Segtrp are small peptides. GelA and GelB are polymer systems.

Collins³¹ and by Bettens and Lee.⁵⁷ As mentioned in our original MOB report,⁵² successful execution of MOB essentially needs two parameters denoted as $\text{MOB}\alpha,\beta$ (where α denotes the order of MOB expansion and β denotes the monomer degree). In this work, we have consistently used the MOB2.3

model that uses two-body interactions using a monomer degree of 3 (each heavy atom bond has the same environment as in the parent system). In addition, we have also considered the effectiveness of using two MIM-type layers (using high and low levels of theory) within the existing MOB methodology. As in

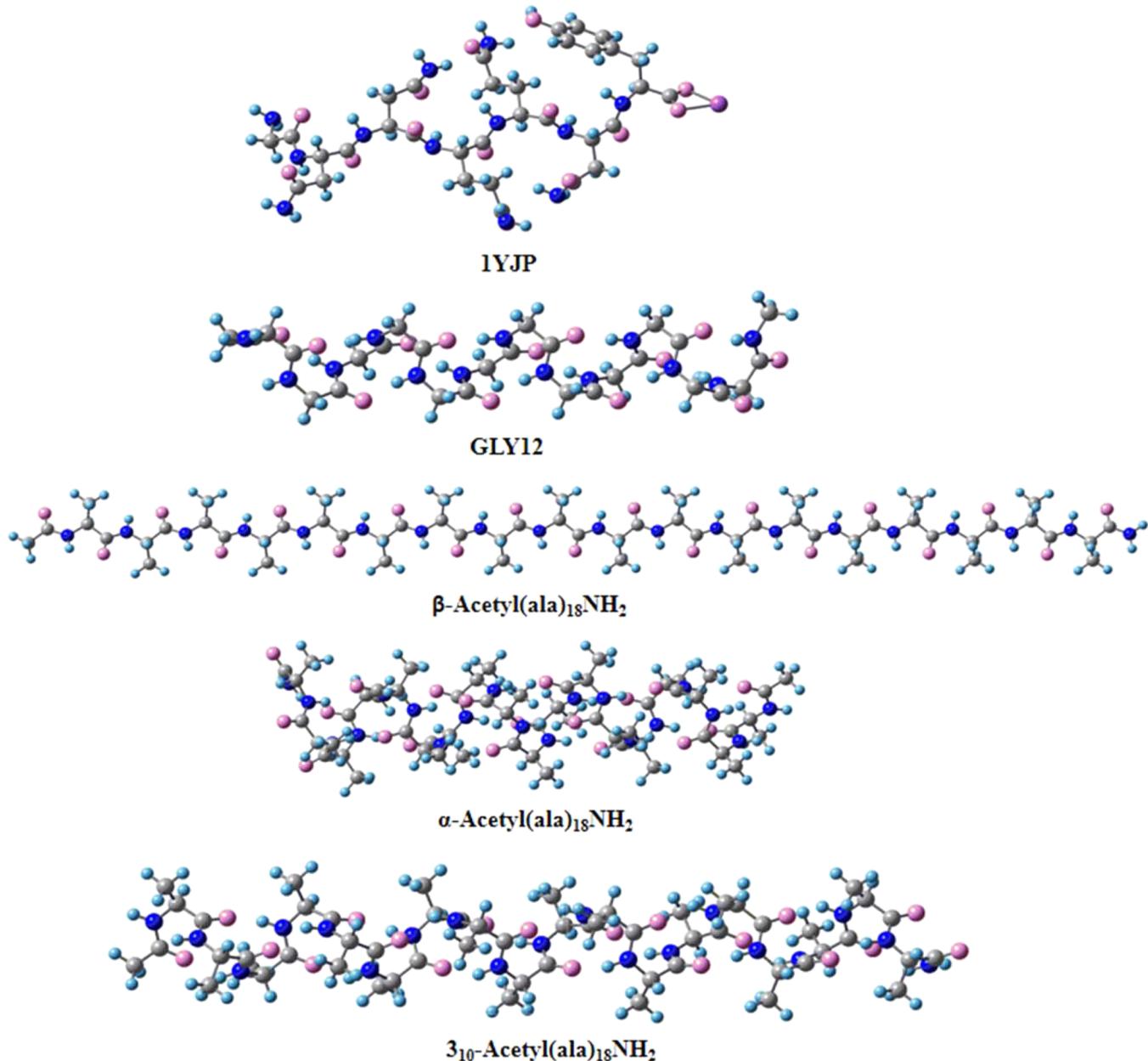


Figure 3. Five large systems studied in this work. Three different motifs (α , β , and 3_{10} -helix) of acetyl(ala)₁₈NH₂ represent the largest system considered.

MIM, we have followed a similar notation to represent them. For example, execution of MOB with a single layer of theory is termed as MOB2.3(1) whereas the same with two layers of theory is termed as MOB2.3(2) (*vide infra*). For simplicity, they will be denoted as MOB(1) and MOB(2).

To better illustrate the fragmentation procedures, we will walk the reader through both MIM and MOB fragmentation schemes in a sample system. Figure 1 illustrates the fragmentation schemes on alanine tetrapeptide. Figure 1a shows that all of the single bonds between heavy atoms are excised to generate fragments. This is valid for any MIM or MOB fragmentation schemes where only single bonds are cut. Figure 1b shows one of the sample primary subsystems generated in a distance-based MIM scheme. If the green fragment is considered as the central fragment unit, then all of the fragments within 3 Å from this central fragment (shown

with a circle) constitute a primary subsystem. Similarly Figure 1c shows a MOB scheme with monomer degree of 1, where each bond initializes a primary subsystem. With the increase in monomer degree, the size of the subsystems will increase. Once these primary subsystems are generated in both MIM and MOB schemes, derivative subsystems are formed to cancel overlapping regions. In MOB, all of the primary and derivative subsystems are together called “monomers”, and they will interact with each other via a two-body interaction within a distance-based cutoff. The primary and derivative subsystems (for MIM) or the monomers and dimers (for MOB) are then fed into the inclusion–exclusion-based energy equation (for MIM) or the many-body-based equation (for MOB) to predict the full energy or any other desired property of the system.

Table 1. Total Energy Deviations (kcal/mol) of All Systems between Unfragmented and Fragment-Based MP2/6-31+G(d,p) Results Using a Single Layer (MIM1 or MOB(1))^a

system	MIM($d=3.0\text{\AA}$)		MIM($d=3.5\text{\AA}$)		MIM($\eta=9$)		MOB($d=3.0\text{\AA}$)		MOB($d=5.0\text{\AA}$)	
	br	no_br	br	no_br	br	no_br	br	no_br	br	no_br
Segglu	-2.26	-2.00	-2.99	-1.29	-1.02	-1.09	-0.20	-0.50	-1.02	-0.67
Tuftsin	-2.69	-2.77	-2.67	-2.91	-1.21	-0.94	-1.23	-0.66	0.51	-0.26
Segtrp	-2.00	-1.70	-0.88	-1.62	-1.81	-0.58	-0.76	-0.16	0.02	-0.09
GelA	NA	-1.47	NA	-1.25	NA	-0.63	NA	-0.60	NA	0.32
GelB	NA	-2.29	NA	-1.56	NA	-0.28	NA	-0.41	NA	-0.29
1YJP	-8.10	-4.38	-5.63	-4.86	-3.65	-1.17	0.63	-0.81	-1.23	0.56
Gly12	-24.7	-12.03	-18.1	-11.4	-17.0	-6.45	-15.4	-17.4	-12.4	-14.8
β	-10.7	-8.27	-10.7	-8.27	-11.1	-2.42	-6.48	-3.12	-3.50	-2.73
α	-52.5	-22.8	-49.1	-22.3	-64.4	-20.2	-45.7	-44.7	-33.5	-36.3
3_{10} -helix	-41.9	-27.3	-29.2	-25.4	-52.9	-27.4	-26.3	-33.7	-20.3	-28.5

^aMIM($d=3.0\text{\AA}$) and MIM($d=3.5\text{\AA}$) denote the distance-based MIM methods with a distance cutoff of 3.0 and 3.5 Å, respectively. MIM($\eta=9$) denotes the number-based MIM method with nine fragments in each primary subsystem. MOB($d=3.0\text{\AA}$) and MOB($d=5.0\text{\AA}$) denote many-overlapping-body method (truncated at two-body level) with an interaction sphere of 3 and 5 Å, respectively. “br” denotes the situation where the peptide bond is broken during initial fragment generation while “no_br” denotes otherwise.

3. RESULTS AND DISCUSSION

To provide an assessment of the performance of our fragment-based methods, we have assembled a test set consisting of 10 small to large biomolecules. Among these, the most challenging systems are three isomers of a large polypeptide containing 18 alanine residues, acetyl(ala)₁₈NH₂.⁵⁸ Optimized structures of the three isomers were taken from ref 58. The three unique motifs of this system have a wide range of binding energies due to the different numbers of intramolecular H-bonds. Since these isomers have a large number of such nonbonded interactions, they are known to be challenging systems to test the capabilities of fragment-based methods. The three isomers are denoted as α , β , and 3_{10} -helix. β is a linear extended structure while α and 3_{10} represent two different helical conformations. Both α and 3_{10} -helices exhibit significant cooperative effects. Distortion from the extended β structure to either form of the helical conformations causes some steric strain. The increase in strain is compensated for by the intramolecular H-bonds formed in the helical structures. The strain energy per H-bond is greater for the 3_{10} -helix than the α -helix. The two helical structures differ in NCCN(ψ) and CNCC(φ) angles. The NCCN(ψ) and CNCC(φ) angles for the α -helix are -46° and -58° , respectively, whereas they are -22° and -60° for the 3_{10} -helix. In addition to the three motifs of acetyl(ala)₁₈NH₂, we have also considered five other small peptides and two polymer systems for the validation of our fragmentation methods. These systems have been chosen from ref 59. All of the test systems are shown in Figures 2 and 3. All the calculations were performed with the Gaussian 09 program suite.⁶⁰

The energies of all of the test systems are evaluated at the MP2/6-31+G(d,p) level, directly as well as with fragment-based methods. The direct calculation on the 18-alanine system is quite large, involving 2261 basis functions. If a two-layer model is considered, HF, B3LYP, and M06-2X methods with the same 6-31+G(d,p) basis are used for the second layer. This is particularly efficient for HF/6-31+G(d,p) since it is a subset of the MP2/6-31+G(d,p) calculation for each subsystem.

Table 1 shows the total energy deviations (kcal/mol; difference between the total energy calculated by fragment-based methods and the full unfragmented reference MP2 calculation) for all of the test set systems. A negative sign in Table 1 denotes that the total energy computed by fragment-

based methods is too low. All of the results in Table 1 represent MIM1 and MOB(1) with a single level of theory (MP2/6-31+G(d,p)). Both distance-based and number-based MIM schemes have been used for the initial subsystem generation. Distance-based cutoff parameters with $d = 2.0$, 2.5, 3.0, and 3.5 Å were explored initially in generating the primary subsystems from each fragment. Larger distance parameters were not directly considered since the subsystem sizes were quite large even at $d = 3.5$ Å (*vide infra*). However, the use of a second layer of theory is considered later to take all of the remaining long-range interactions into account. Table 1 shows the performance with two different distance parameters ($d = 3.0$ and 3.5 Å). For number-based MIM schemes, $\eta = 6, 7, 8$, and 9 were investigated, and an optimal parameter ($\eta = 9$) is chosen to keep the subsystem size reasonable (composed of nine fragments). We have also applied the many-body-based methods (MOB) truncated at the two-body level. To recover the long-range interactions via the two-body term, two different radius parameters have been tested (3.0 and 5 Å). In addition, for each of these methods, the performance is shown for two different bond-cutting environments. Since most of our test set systems are polypeptides, we mainly focus on two different covalent bond breaking scenarios. “br” denotes the methods where all of the peptide (C–N) bonds are broken (in addition to C–C and N–C) during initial fragmentation while “no_br” denotes that the peptide bonds are not broken. Since peptide bonds may have partial double bond characters and we are replacing the broken bonds with hydrogen atoms (within all of our current fragment-based methods), it is interesting to investigate the performance of fragmentation strategies with different bond-breaking situations.

We must mention here that “MP2/6-31+G(d,p)” theoretical model was chosen as a reasonably representative and sufficiently high target level of theory. It does include electron-correlation effects and a polarized basis set including diffuse functions. As mentioned earlier, for alanine-18 systems, the unfragmented MP2/6-31+G(d,p) calculation is quite large and computationally intensive (189 atoms and 2261 basis functions). It is possible, and perhaps likely, that it suffers from basis set superposition errors and the relative energies of the isomers might change if we use larger basis sets. Nevertheless, we use this as the target total energy, and we use the same basis set for all of the fragment-based calculations. This may be akin

Table 2. Total Energy Deviations (kcal/mol) of All Systems between Unfragmented and Fragment-Based MP2/6-31+G(d,p) Results^a

system	HF		B3LYP		M06-2X	
	br	no_br	br	no_br	br	no_br
(a) MIM(2) Model Using a Distance-Based Cutoff ($d = 3 \text{ \AA}$)						
Segglu	-2.26	-1.38	-2.42	-1.41	-2.28	-1.32
Tuftsin	-2.21	-1.86	-2.61	-2.04	-2.36	-1.80
Segtrp	-3.05	-1.65	-2.53	-1.67	-2.70	-1.66
GelA	NA	-1.37	NA	-1.36	NA	-1.20
GelB	NA	-1.51	NA	-1.71	NA	-1.55
1YJP	-5.29	-2.17	-5.62	-2.76	-5.59	-2.44
Gly12	2.44	2.24	-1.49	0.51	-1.44	0.37
β	-8.97	-2.91	-4.60	-1.87	-6.08	-2.19
α	4.47	2.79	-0.02	-0.18	-2.59	-0.14
3 ₁₀ -helix	3.61	3.84	-2.74	-0.22	-3.14	-0.67
(b) MIM(2) Model Using a Distance-Based Cutoff ($d = 3.5 \text{ \AA}$)						
Segglu	-1.38	-0.87	-1.47	-0.83	-1.44	-0.84
Tuftsin	-1.90	-1.37	-2.08	-1.49	-1.95	-1.37
Segtrp	-1.78	-0.61	-1.39	-0.71	-1.51	-0.72
GelA	NA	-1.19	NA	-1.18	NA	-1.03
GelB	NA	-1.07	NA	-1.16	NA	-1.07
1YJP	-3.36	-2.09	-3.80	-2.44	-3.57	-2.31
Gly12	3.47	2.46	0.45	0.80	0.30	0.63
β	-8.97	-2.91	-4.60	-1.87	-6.08	-2.19
α	5.50	3.70	-0.25	0.81	-0.95	0.54
3 ₁₀ -helix	4.61	5.31	1.35	1.56	0.48	0.95
(c) MIM(2) Model Using a Number-Based Cutoff ($\eta = 9$)						
Segglu	-0.97	-0.46	-0.96	-0.36	-0.93	-0.38
Tuftsin	-1.50	-0.66	-1.69	-0.66	-1.60	-0.65
Segtrp	-1.02	-0.13	-0.84	-0.17	-0.96	-0.18
GelA	NA	-0.59	NA	-0.61	NA	-0.57
GelB	NA	-0.36	NA	-0.40	NA	-0.37
1YJP	-2.56	-1.04	-2.54	-1.07	-2.48	-1.01
Gly12	2.30	1.37	-0.46	0.49	-0.39	0.37
β	-3.06	-1.1	-0.44	-0.18	-1.46	-0.46
α	0.50	3.36	-4.01	1.27	-4.87	-1.87
3 ₁₀ -helix	2.21	3.16	-3.38	-0.85	-4.01	-1.30
(d) MOB(2) Scheme with a Distance Cutoff of 3 Å To Include the Two-Body Terms						
Segglu	-0.34	-0.46	-0.15	-0.32	-0.24	-0.31
Tuftsin	-0.50	-0.37	-0.54	-0.38	-0.51	-0.31
Segtrp	-0.38	-0.17	-0.58	-0.16	-0.47	-0.15
GelA	NA	-0.59	NA	-0.60	NA	-0.52
GelB	NA	-0.26	NA	-0.32	NA	-0.31
1YJP	0.89	-0.14	1.12	-0.33	0.17	-0.18
Gly12	1.54	-2.03	0.17	0.19	-0.09	0.01
β	-1.71	-1.59	-0.59	-0.70	-0.98	-0.89
α	-14.81	-1.75	-14.87	-5.09	-3.16	-3.45
3 ₁₀ -helix	3.29	4.93	-2.79	0.46	-1.66	0.08
(e) MOB(2) Scheme with a Distance Cutoff of 5 Å To Include the Two-Body Terms						
Segglu	-0.46	-0.22	-0.41	-0.19	-0.35	-0.15
Tuftsin	-0.18	-0.20	-0.00	-0.13	-0.02	-0.14
Segtrp	-3.05	-0.13	-2.53	-0.12	-2.70	-0.12
GelA	NA	0.13	NA	0.19	NA	0.19
GelB	NA	-0.21	NA	-0.21	NA	-0.21
1YJP	-0.86	-0.17	-0.99	-0.22	-0.65	-0.10
Gly12	1.00	1.57	0.26	0.32	-0.02	0.12
β	-1.43	-1.40	-0.53	-0.5	-0.70	-0.70
α	-1.20	1.40	-1.68	-1.87	-2.94	-2.02
3 ₁₀ -helix	2.41	4.28	-2.43	0.91	-1.23	0.57

^aPerformance of HF and two different DFT methods as a low level theory has been shown. “br” denotes the situation where the peptide bond has been broken during initial fragment generation while “no_br” denotes otherwise.

Table 3. Calculated Relative Energies (kcal/mol) Using Direct Unfragmented Calculations with the 6-31+G(d,p) Basis Set

system	relative energy (kcal/mol)			
	MP2	HF	B3LYP	M06-2X
$\alpha\text{-}\beta$	-124.3	-24.9	-20.3	-81.9
$\beta\text{-}3_{10}\text{-helix}$	-77.6	-171.6	-188.7	-108.7
$\alpha\text{-}3_{10}\text{-helix}$	-201.9	-196.4	-209.0	-190.6

to using a full configuration interaction (CI) reference calculation with a modest basis set to validate the performance of approximate electron-correlation methods. Thus, the unfragmented MP2/6-31+G(d,p) energy is our target “experimental value” that we try to approach using appropriate fragment-based methods. There are numerous other papers in fragment-based QM literature where a similar approximate theoretical model has been used for the assessment of new fragment-based methods. In our current project we hope to identify the features that are needed to perform effectively at this level, with the hope of carrying out fragment-based applications for more challenging systems in the future.

Table 1 (using only a single layer of theory) shows that for all of the small peptides (for system 1 to system 3 and GelA and GelB) the total energy deviations are within 3 kcal/mol regardless of the methods applied. Close inspection on the results (Table 1) also reveals that, for these small peptides, breaking the peptide bonds does not impact the performance significantly. However, it can be seen later that the peptide bond breaking does deteriorate the performance for larger systems and two-layer models. If we compare performance of different methods, MOB-based methods seem to have smaller

deviations than MIM-based methods. Total energy deviations with MOB(1) in Table 1 are all less than 1 kcal/mol for all small peptides (except for two cases) whereas those deviations in MIM-based methods (both distance-based and number-based) can go up to 3 kcal/mol.

The performance of the methods for larger systems in Table 1 is significantly worse. Now the deviations range from a small value of 0.56 kcal/mol to as high as 64.4 kcal/mol. In particular, for the two helical motifs of acetyl(ala)₁₈NH₂ system (last two rows of Table 1), it is clear that regardless of the methods the deviations are more than 20 kcal/mol. It is obvious that interfragment interactions are not included adequately and that substantially larger fragments will be needed within such a single-layer model to reproduce the unfragmented MP2 energies. In addition, there is very uneven performance in the prediction of the energy difference between the open β -motif and the helical motifs. This observation suggests that other strategies have to be considered for possible improvement in the prediction of unfragmented MP2 results.

We implemented a second layer of theory in both of our fragment-based methods (MIM and MOB). This idea helps to recover the long-range fragment–fragment interactions involved in these systems which is otherwise difficult to recover in a single layer. Table 2 shows the total energy deviations (kcal/mol) for all of the systems by using a second low level of theory (*vide infra*). It demonstrates that, on going from a single layer (Table 1) to two layers (Table 2), the deviation improves remarkably regardless of the system. This clearly suggests that adding a cheaper level of theory as a second layer in fragment-based methods (if computationally feasible) can be a very effective strategy relative to the use of larger fragments or more clever fragmentation schemes.

Table 4. Calculated Relative Energies (Columns 3–6, kcal/mol) and Relative Energy Deviations (Columns 7–10, kcal/mol) for 18-Alanine Peptide Systems in Fragment-Based MP2/6-31+G(d,p) Calculations^a

(a) MIM Model Using a Number-Based Cutoff ($\eta = 9$)									
system	full MP2	MIM1	MIM2			relative energy deviation (kcal/mol)			MIM2
			HF	B3LYP	M06-2X	MIM1	HF	B3LYP	
$\alpha\text{-}\beta$	-124.3	-106.5	-128.8	-125.5	-122.9	-17.8	4.5	1.2	-1.4
$\beta\text{-}3_{10}\text{-helix}$	-77.6	-102.7	-73.4	-78.3	-78.5	25.0	-4.3	0.7	0.9
$\alpha\text{-}3_{10}\text{-helix}$	-201.9	-209.2	-202.2	-204.1	-201.4	7.3	0.3	2.0	-0.6

(b) MIM Model Using a Distance-Based Cutoff ($d = 3 \text{ \AA}$)									
system	full MP2	MIM1	MIM2			relative energy deviation (kcal/mol)			MIM2
			HF	B3LYP	M06-2X	MIM1	HF	B3LYP	
$\alpha\text{-}\beta$	-124.3	-109.7	-130.0	-126.0	-126.3	-14.6	5.71	1.7	2.0
$\beta\text{-}3_{10}\text{-helix}$	-77.6	-96.7	-70.9	-76.0	-76.1	19.1	-6.77	-1.6	-1.5
$\alpha\text{-}3_{10}\text{-helix}$	-201.9	-206.4	-200.9	-202.0	-202.5	4.5	-1.01	0.1	0.6

(c) MOB Scheme Used for All of the Systems with a Distance Cutoff of 5 Å To Include the Two-Body Terms ^b									
system	full MP2	MOB1	MOB2			relative energy deviation (kcal/mol)			MOB2
			HF	B3LYP	M06-2X	MOB1	HF	B3LYP	
$\alpha\text{-}\beta$	-124.3	-90.7	-127.1	-122.9	-123.0	-33.6	2.8	-1.4	-1.3
$\beta\text{-}3_{10}\text{-helix}$	-77.6	-103.4	-71.9	-76.2	-76.4	25.8	-5.7	-1.4	-1.3
$\alpha\text{-}3_{10}\text{-helix}$	-201.9	-194.1	-199.1	-199.2	-199.4	-7.8	-2.8	-2.8	-2.6

^aThe unfragmented reference MP2 results are shown in column 2. Performance of single-layer (MIM1) models and two-layer (MIM2) models with HF and two different DFT methods as a low level correction have been shown. Results are shown for situations where peptide bonds are not broken during initial fragment generation. ^bMOB(1) and MOB(2) denote single and double layers of theory, respectively.

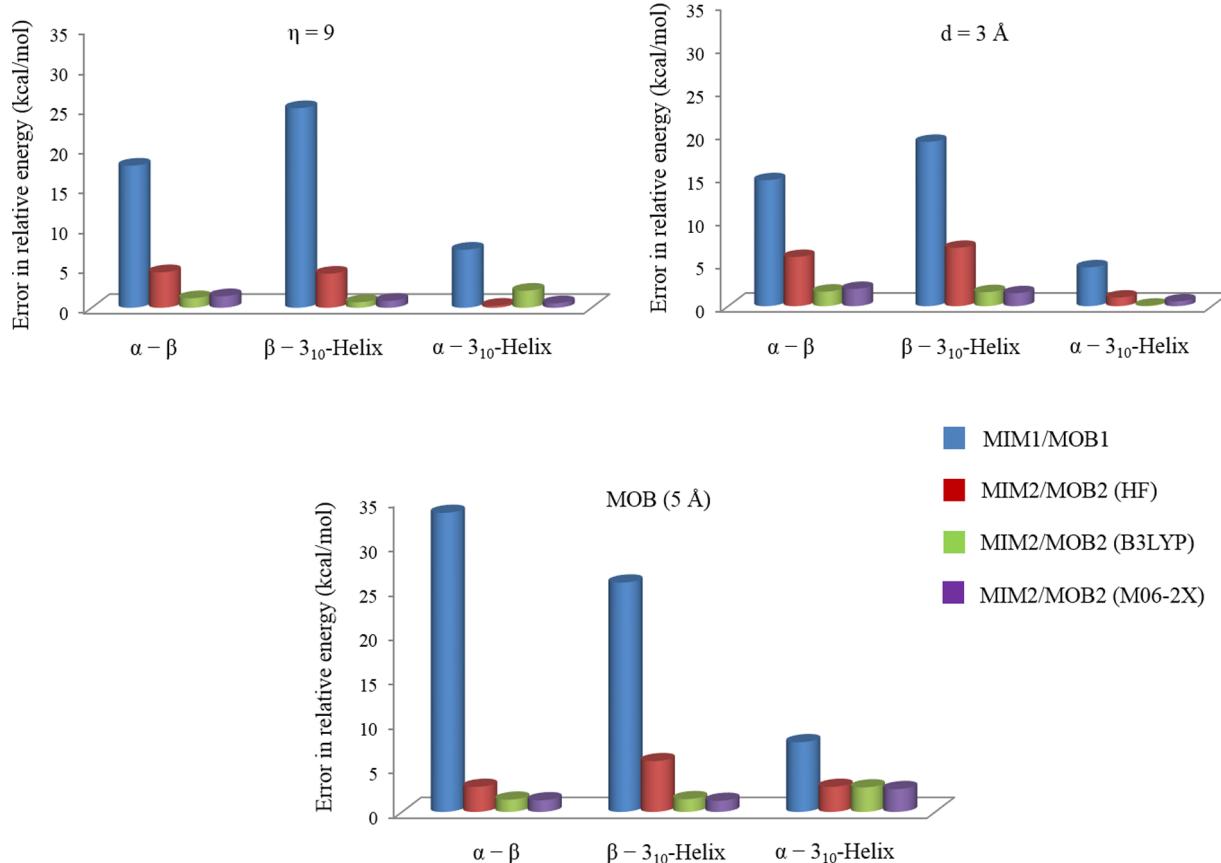


Figure 4. Graphical illustration of “second-layer effect” in relative energy deviations for three motifs of acetyl(ala)₁₈NH₂. MIM1 and MOB1 represent one-layer methods, and MIM2 and MOB2 refer to two-layer methods.

Since we use a low level of theory on the whole system to successfully capture all of the long-range interactions, the performance of both HF and DFT methods were tested rigorously for a clearer analysis of their effectiveness in capturing such effects. Three different low levels of theory have been used here: HF/6-31+G(d,p), B3LYP/6-31+G(d,p), and M06-2X/6-31+G(d,p). Table 2 shows the total energy deviations with the different schemes tested. Parts (a) and (b) of Table 2 show the performance of the distance-based MIM scheme with a cutoff distance of $r = 3 \text{ \AA}$ and $r = 3.5 \text{ \AA}$, respectively. Performance of the number-based MIM scheme with $\eta = 9$ is shown in Part (c). The MOB calculations have been performed with a distance cutoff (radius of interaction sphere) of 3 Å (Part (d)) and 5 Å (Part (e)).

There are some straightforward observations from Table 2. Similar to our previous observation with a single layer in Table 1, different bond-breaking situations do not really impact the predicted energies as much for the *small peptides*. At first glance, this appears to support the concept of breaking a partially double bonded peptide bond and using hydrogen as a link atom. However, such insensitivity to bond-cutting environments is not retained in other systems (larger peptides) as we will show in further discussion. It is also important to note that the inclusion of a second layer is not really needed for small peptides. For these systems (where long-range interactions do not play a significant role), straightforward fragmentation of the systems (with overlapping units) and summation of the fragmented units is sufficient to provide accurate results.

But one has to be very careful about reaching similar general conclusions with acetyl(ala)₁₈NH₂ systems. Breaking the

peptide bonds while forming the initial fragments is no longer validated for the larger peptides. In the single-layer results (Table 1) as well as in the case of two-layer results (Table 2), deviations of fragmentation methods where we retain the peptide bonds are significantly better than the ones where peptide bonds are broken. This clearly demonstrates that breaking partially double bonded peptide bonds indeed influences the performance of fragmentation methods and is highly system dependent. One can also notice that the performance for the larger peptides depends on the nature of the low level of theory regardless of the fragmentation methods. Table 2 shows that for large peptides, while HF as a low level performs with reasonable accuracy, B3LYP and M06-2X as a low level outperform these results with high accuracy. This observation is general for all of the fragment-based methods applied to all of the large systems considered in this report. Overall, M06-2X appears to be the best among the three methods considered, though B3LYP performs almost as well.

We can also make some preliminary observations analyzing the performance of different fragment-based schemes. The compactness of these three isomers can be described as $\alpha > 3_{10}\text{-helix} > \beta$. Obviously, α and 3_{10}-helix motifs represent the most challenging problems among the three isomers. Hence, we will compare the performance of different methods on these larger peptides based on the results on α and 3_{10}-helix forms of the peptide. Since these forms contain a large number of intramolecular hydrogen bonds due to their helical structure and compactness, a (second) lower level of theory appears to be a crucial factor to capture all of the long-range interactions. From Table 2, we observe that the performance of all of the

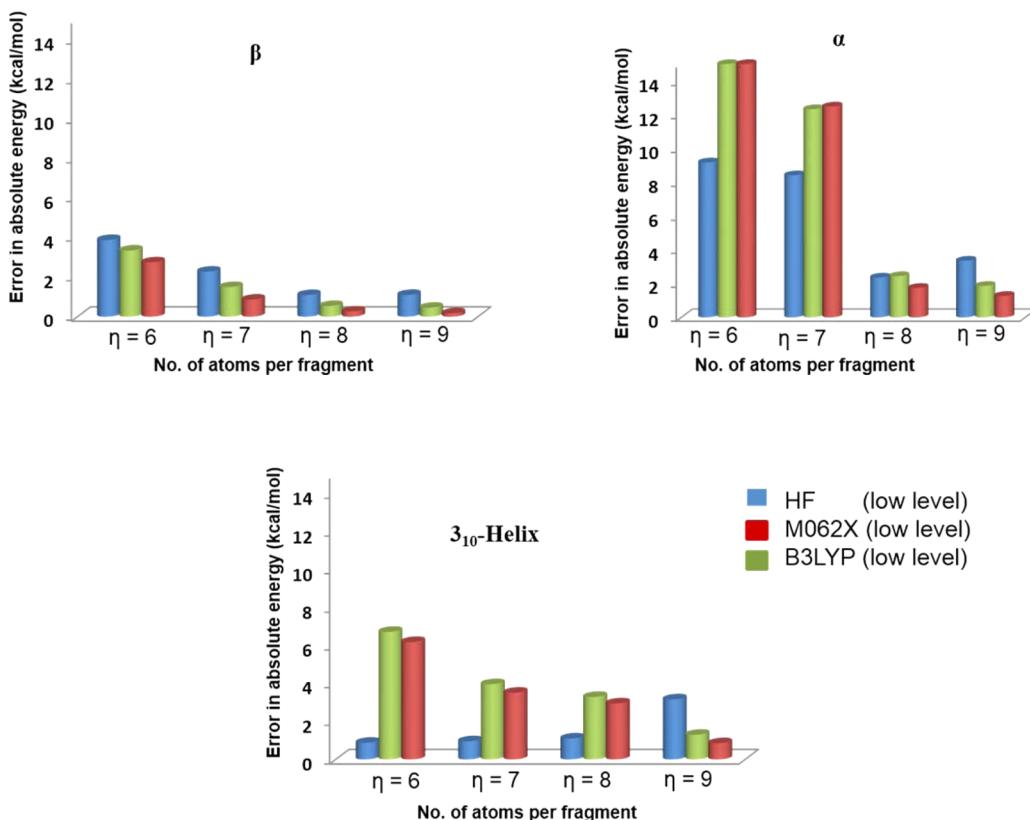


Figure 5. Convergence of the total energy deviations of three motifs of acetyl(ala)₁₈NH₂ with increasing number of fragments per subsystem in the number-based MIM scheme. The performance of different methods as a low level of theory is shown in different color codes.

fragmentation schemes on the helix forms of the peptide noticeably improve when B3LYP and M06-2X functionals are used as a cheaper level of theory.

Since quantum chemistry is mostly concerned with energy differences, it is always important to calculate relative energies of the different isomers. The systematic deviations in any fragmentation procedure may lead to a high discrepancy in the total energy whereas those deviations may get canceled while calculating relative energies. Before we consider the performance of one-layer or two-layer fragment-based models, it is important to understand the intrinsic behavior of the different levels of theory used in this work. Table 3 shows the relative energies with *unfragmented* MP2, HF, B3LYP, and M06-2X calculations. While the energy ordering of the three isomers ($\alpha < \beta < 3_{10}\text{-helix}$) is maintained with all levels of theory, there are large differences between our target MP2 values and those obtained by the other methods. In particular, HF and B3LYP show differences that exceed 100 kcal/mol relative to MP2 in the relative energies! While M06-2X is closer to MP2, the energy difference between α and β isomers differs by more than 40 kcal/mol relative to the target MP2 value. It is important to notice here that HF, B3LYP, and even M06-2X cannot reproduce the MP2 relative energies between linear and helical structures ($\alpha\text{-}\beta$ and $\beta\text{-}3_{10}\text{-helix}$), the most challenging ones considering the completely different structural conformations. However, these methods can reproduce the MP2 relative energies between two different helical motifs ($\alpha\text{-}3_{10}\text{-helix}$) with smaller deviations (10 kcal/mol or less).

Table 4 lists the relative energies of the isomers as well as relative energy deviations (with respect to unfragmented MP2 result) with number-based MIM scheme, distance-based MIM

scheme, and the MOB scheme. The performance of a single level of theory (MIM1 or MOB(1)) as well as the use of a background low layer (MIM2 (or MOB(2))) are assessed. The same trend is observed as in the total energy deviations. The single-layer models (MIM1 and MOB(1)) have substantial deviations in the relative energies ranging from 5 to 35 kcal/mol. Thus, the deviations in the total energies are not systematic enough to yield accurate relative energies within a single-layer model. However, relative energy deviations improve dramatically on going from one layer to two layers. Table 4 shows that B3LYP and M06-2X functionals as a low level of theory can reproduce the unfragmented MP2 relative energy profiles most accurately. Relative energy deviations with B3LYP and M06-2X functionals in MIM2 (both with $\eta = 9$ and $d = 3$ Å) or MOB(2) schemes are always within 3 kcal/mol. If HF is used as the low level of theory, the deviations range as high as 7 kcal/mol. The MIM2 models perform best with deviations from unfragmented MP2 that do not exceed 2 kcal/mol. One can easily notice the novelty of our fragment-based methods here. Irrespective of the structural conformations, both MIM2 and MOB(2) can reproduce the MP2 relative energies with high accuracy. The key point is that treating the local regions with MP2 and the longer range effects with DFT is sufficient to yield accurate results, despite the fact that the DFT methods are themselves not accurate enough to make reliable predictions of the relative energies of such peptides. Figure 4 illustrates the dramatic improvements in the quality of the computed results on going from one layer to two layers.

To put these results in context, we should mention other contemporary fragment-based methods and compare with their performances. Truhlar et al. developed EE-MTA⁶¹ method and

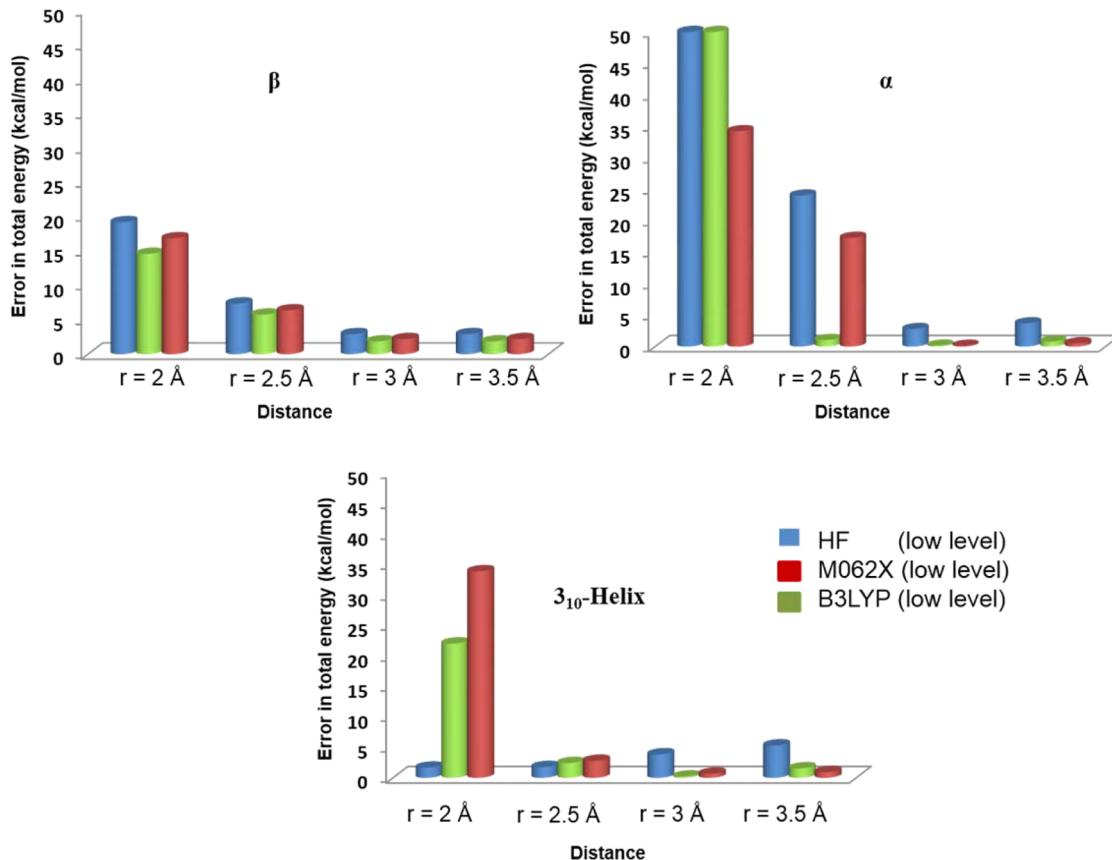


Figure 6. Convergence of the total energy deviations of three motifs of acetyl(ala)₁₈NH₂ with increasing distance parameter in the distance-based MIM scheme. Performance of the different methods as a low level of theory is shown in different color codes.

Table 5. Number of RMS Basis Functions in Fragments Using Different Fragmentation Strategies^a

(a) Size Dependence of MIM with Increasing Number-Based Parameter η				
system	$\eta = 6$	$\eta = 7$	$\eta = 8$	
β	237 (283)	281 (312)	309 (374)	353 (428)
α	218 (355)	262 (432)	276 (465)	260 (494)
3 ₁₀ -helix	247 (374)	303 (403)	364 (465)	401 (494)
(b) Size Dependence of MIM with Increasing Distance-Based Parameter r				
system	$d = 2 \text{ \AA}$	$d = 2.5 \text{ \AA}$	$d = 3 \text{ \AA}$	
β	136 (192)	227 (283)	258 (312)	
α	160 (274)	226 (475)	424 (734)	
3 ₁₀ -helix	171 (274)	427 (494)	458 (523)	
(c) Number of Basis Functions in MOB Scheme ^b				
system	MOB23			
β	264 (403)			
α	268 (403)			
3 ₁₀ -helix	269 (403)			

^aValues in parentheses denote basis functions in the largest fragment. Results are shown for situations where peptide bonds are not broken during initial fragment generation. ^bResults are shown with a cutoff distance of 5 Å to include all of the two-body terms inside a sphere with radius 5 Å.

applied it to the relative energies of alanine-20 systems that are comparable in size to our peptide. They reported relative energy deviations (with M06-2X/6-31G* theoretical model) which range from 7.63 to 2.64 kcal/mol (Table 5 in ref 61) depending on the type of charges used in the embedding. In general, a maximum deviation of 2 kcal/mol for these types of

challenging biochemical systems should be regarded as highly accurate in such fragment-based quantum chemical calculations.^{51,61}

Since our goal is to perform an accurate calculation on a large molecule within a given amount of time, we have carried out a rigorous analysis to find out the optimal procedure (i.e., computationally most efficient) to calculate the energies. We investigate the convergence of total energy deviations for three motifs of acetyl(ala)₁₈NH₂ with increasing distance parameter in distance-based MIM scheme and with increasing number in number-based MIM scheme. The convergence pattern in the calculated total energies of the different methods are shown in Figures 5 and 6. It is interesting to observe that the convergence pattern of total energies in the case of the linear structure (β motif of polypeptide) is very smooth with increasing cutoff values in the number-based scheme (Figure 5a) as well as in the distance-based scheme (Figure 6a). In contrast, smooth convergence is not observed in the case of helical structures in either of these MIM schemes: Rather a slight increase in cutoff values results in large changes in the total energy deviations. This effect is more pronounced in the case of distance-based MIM scheme than in the number-based scheme. Rationalization for these observations can be given in terms of the growth of fragment sizes in different fragmentation strategies. If the system has a linear backbone of the heavy atoms, the distance- and number-based MIM schemes will give rise to systematic growth of fragments. But if the structure is highly compact (such as in helical conformations, α and 3₁₀-helix), these schemes may suffer from unbalanced growth of fragments. This is also evident from Table 5. The number of

basis functions of the largest fragment in distance-based (3 \AA cutoff) MIM scheme is 734 whereas it is as low as 494 and 403 in the case of number-based MIM and MOB schemes, respectively. The overall fragment sizes seem to be larger in the distance-based MIM scheme (root-mean-square (RMS) number of basis functions in the distance-based scheme is 424, whereas it is 260 and 268 in the case of number-based MIM and MOB schemes, respectively). However, in general, more numbers of fragment calculations have to be carried out by MOB compared to MIM to reach the same level of accuracy. We conclude that while MIM generates relatively larger fragments to include the important weak interactions, similar performance is achieved in MOB by executing more two-body interactions of distant but smaller fragments.

The performance of MIM and MOB for the test set of 10 molecules can now be summarized. The unfragmented MP2 total energies of the test set can be reproduced with impressive accuracy with fragment-based methods. The inclusion of a low level of theory to capture the long-range interactions is critical to the success of our models. B3LYP and M06-2X perform much better than HF as the second level of theory, particularly for the larger molecules in the test set. Using M06-2X as the low level of theory, the mean deviation in the calculated total energies for the five different two-layer methods MIM($d=3.0\text{\AA}$), MIM($d=3.5\text{\AA}$), MIM($\eta=9$), MOB($d=3.0\text{\AA}$), and MOB($d=5.0\text{\AA}$) are 1.33, 1.16, 0.72, 0.62, and 0.43 kcal/mol, respectively. The corresponding maximum deviations are 2.44, 2.31, 1.87, 3.45, and 2.02 kcal/mol, respectively. The performance is clearly impressive considering the size of the biomolecular systems.

4. CONCLUSIONS

We have carried out a successful application of two of our fragment-based methods “MIM” and “MOB” on large biomolecular systems. Using two layers of theory, we demonstrate that both MIM and MOB can predict the unfragmented MP2 energies of a range of biomolecules including three isomers of 18-alanine residues with high accuracy. In addition, all of these methods can successfully reproduce the relative energy ordering of the isomers from unfragmented calculations. Rigorous calibration of the low level of theory was carried out choosing HF and two different density functionals. Our study indicates that the performance of DFT (as a low level of theory to capture the long range of interaction) is significantly better than HF, particularly for the larger molecules. Employing M06-2X as the low level of theory, the calculated mean total energy deviation (maximum deviation) in the total MP2 energies for the 10 molecules in the test set at MIM($d=3.5\text{\AA}$), MIM($\eta=9$), and MOB($d=5\text{\AA}$) are 1.16 (2.31), 0.72 (1.87), and 0.43 (2.02) kcal/mol, respectively. This impressive performance suggests that such fragment-based methods are highly promising to evaluate accurate energies of large biomolecular systems.

■ ASSOCIATED CONTENT

Supporting Information

Table listing Cartesian coordinates of all systems investigated in this work and figure showing graphical representation of MIM and MOBE. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: kraghava@indiana.edu.

Funding

This work was supported by funding from NSF Grant CHE-1266154 at Indiana University.

Notes

The authors declare no competing financial interest.

■ REFERENCES

- (1) Humbel, S.; Sieber, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *105*, 1959.
- (2) Morokuma, K. *J. Comput. Chem.* **2000**, *21*, 1419.
- (3) Karadakov, P. B.; Morokuma, K. *Chem. Phys. Lett.* **2000**, *317*, 589.
- (4) Svensson, M.; Humbel, S.; Froese, R.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357.
- (5) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. *Chem. Rev.* **2012**, *112*, 632.
- (6) Kitaura, K. *Chem. Phys. Lett.* **1999**, *313*, 701.
- (7) Pruitt, S. R.; Bertoni, C.; Brorsen, K. R.; Gordon, M. S. *Acc. Chem. Res.* **2014**, *47*, 2786.
- (8) Fedorov, D. G.; Kitaura, K. *Chem. Phys. Lett.* **2014**, *597*, 99.
- (9) Ishida, T.; Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. B* **2006**, *110*, 1457.
- (10) Gadre, S. R.; Shirasat, R. N.; Limaye, A. C. *J. Phys. Chem.* **1994**, *98*, 9165.
- (11) Rahalkar, A. P.; Ganesh, V.; Gadre, S. R. *J. Chem. Phys.* **2008**, *129*, 234101.
- (12) Kavathekar, R.; Khire, S.; Ganesh, V.; Rahalkar, A. P.; Gadre, S. R. *J. Comput. Chem.* **2009**, *30*, 11.
- (13) Sahu, N.; Gadre, S. R. *Acc. Chem. Res.* **2014**, *47*, 2739.
- (14) Rusinska-Roszak, D.; Sowinski, G. *J. Chem. Inf. Model.* **2014**, *1963*.
- (15) Yeole, S. D.; Gadre, S. R. *J. Chem. Phys.* **2011**, *134*, No. 084111.
- (16) Rahalkar, A. P.; Katouda, M.; Gadre, S. R.; Nagase, S. *J. Comput. Chem.* **2010**, *31*, 2405.
- (17) Ganesh, V.; Dongare, R. K.; Balanarayanan, P.; Gadre, S. R. *J. Chem. Phys.* **2006**, *125*, 104109.
- (18) Hua, W. J.; Fang, T.; Li, W.; Yu, J. G.; Li, S. H. *J. Phys. Chem. A* **2008**, *112*, 10864.
- (19) Li, W.; Li, S.; Jiang, Y. *J. Phys. Chem. A* **2007**, *111*, 2193.
- (20) Li, S.; Li, W.; Ma, J. *Acc. Chem. Res.* **2014**, *47*, 2712.
- (21) Wang, K.; Li, W.; Li, S. *J. Chem. Theory Comput.* **2014**, *10*, 1546.
- (22) Hua, S.; Li, W.; Li, S. *ChemPhysChem* **2013**, *14*, 108.
- (23) Hua, S.; Hua, W.; Li, S. *J. Phys. Chem. A* **2010**, *114*, 8126.
- (24) Zhang, D. W.; Zhang, J. Z. H. *J. Chem. Phys.* **2003**, *119*, 3599.
- (25) He, X.; Zhu, T.; Wang, X.; Liu, J.; Zhang, J. Z. H. *Acc. Chem. Res.* **2014**, *47*, 2748.
- (26) Gao, A. M.; Zhang, D. W.; Zhang, J. Z. H.; Zhang, Y. *Chem. Phys. Lett.* **2004**, *394*, 293.
- (27) He, X.; Zhang, J. Z. H. *J. Chem. Phys.* **2005**, *122*, No. 031103.
- (28) Chen, X. H.; Zhang, D. W.; Zhang, J. Z. H. *J. Chem. Phys.* **2004**, *120*, 839.
- (29) Xiao, H.; Zhang, Z. H. *J. Chem. Phys.* **2006**, *124*, 184703.
- (30) Collins, M. A.; Cvitkovic, M. W.; Bettens, P. A. *Acc. Chem. Res.* **2014**, *47*, 2776.
- (31) Collins, M. A.; Deev, V. A. *J. Chem. Phys.* **2006**, *125*, 104104.
- (32) Collins, M. A. *J. Chem. Phys.* **2007**, *127*, No. 024104.
- (33) Collins, M. A. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7744.
- (34) Le, H.; Tan, H.; Ouyang, J.; Bettens, P. A. *J. Chem. Theory Comput.* **2012**, *8*, 469.
- (35) Frankcombe, T. J.; Collins, M. A. *Phys. Chem. Chem. Phys.* **2011**, *13*, 8379.
- (36) Tan, H. J.; Bettens, R. P. A. *Phys. Chem. Chem. Phys.* **2013**, *15*, 7541.
- (37) Huang, L.; Massa, L.; Karle, J. *Int. J. Quantum Chem.* **2005**, *103*, 808.

- (38) Huang, L.; Massa, L.; Karle, J. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 12690.
- (39) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 1342.
- (40) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 46.
- (41) Dahlke, E. E.; Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 33.
- (42) Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009**, *5*, 1573.
- (43) Isegawa, M.; Wang, B.; Truhlar, D. G. *J. Chem. Theory Comput.* **2013**, *9*, 1381.
- (44) Rezac, J.; Salahub, D. R. *J. Chem. Theory Comput.* **2010**, *6*, 91.
- (45) Wen, S.; Nanda, K.; Huang, Y.; Beran, G. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7578.
- (46) Nanda, K.; Beran, G. *J. O. J. Chem. Phys.* **2012**, *137*, 174106.
- (47) Beran, G. J. O.; Nanda, K. *J. Phys. Chem. Lett.* **2010**, *1*, 3480.
- (48) Hopkins, B. W.; Tschumper, G. S. *J. Comput. Chem.* **2003**, *24*, 1563.
- (49) Hopkins, B. W.; Tschumper, G. S. *Chem. Phys. Lett.* **2005**, *407*, 362.
- (50) Tschumper, G. S. *Chem. Phys. Lett.* **2006**, *427*, 185.
- (51) Mayhall, N. J.; Raghavachari, K. *J. Chem. Theory Comput.* **2011**, *7*, 1336.
- (52) Mayhall, N. J.; Raghavachari, K. *J. Chem. Theory Comput.* **2012**, *8*, 2669.
- (53) Saha, A.; Raghavachari, K. *J. Chem. Theory Comput.* **2014**, *10*, 58.
- (54) Richard, M. R.; Herbert, J. M. *J. Chem. Phys.* **2012**, *137*, No. 064113.
- (55) Ramabhadran, R. O.; Raghavachari, K. *J. Chem. Theory Comput.* **2011**, *7*, 2094.
- (56) Ramabhadran, R. O.; Raghavachari, K. *Acc. Chem. Res.* **2014**, *47*, 3596.
- (57) Bettens, R. P. A.; Lee, A. M. *J. Phys. Chem. A* **2006**, *110*, 8777.
- (58) Robert, W.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2004**, *126*, 14198.
- (59) Le, H.; Lee, M. A.; Bettens, P. A. *R. J. Phys. Chem. A* **2008**, *113*, 10527.
- (60) Frisch, M. J.; Trucks, G. W.; Schlegel, H. G.; 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, M. J.; 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, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision D.01*; Gaussian, Inc.: Wallingford, CT, 2009.
- (61) Isegawa, M.; Wang, B.; Truhlar, D. G. *J. Chem. Theory Comput.* **2013**, *9*, 1381.