# Evaluation of Energy Gradients and Infrared Vibrational Spectra through Molecules-in-Molecules Fragment-Based Approach

K. V. Jovan Jose and Krishnan Raghavachari*

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

S Supporting Information

ABSTRACT: Molecules-in-Molecules (MIM) is a general hybrid fragment-based extrapolation approach for calculating accurate total energies of large molecules, similar in spirit to the popular ONIOM methodology. In this work, the MIM model is extended for the precise evaluation of the energy gradients and infrared (IR) vibrational spectra of large molecules. The overlapping subsystems in this work are constructed from nonoverlapping fragments using a number-based scheme, and the dangling bonds are saturated with link-hydrogen atoms. Independent fragment calculations are performed to evaluate the energies and its gradients. Subsequently, the link-atom energy gradient components are projected back onto the corresponding host and supporting atoms, through the Jacobian projection method, as in the ONIOM approach. After geometry optimization, the Jacobian link-atom projection method is also employed for the precise evaluation of the force constants and dipole derivatives of the full molecule. The performance of the MIM model is benchmarked on 25 small-to-large peptides, with inevitable weak long-range intramolecular interactions. Upon accounting these long-range interactions through a second layer, at an inexpensive low-level of theory (MIM2), the energy accuracy improve by 80%, compared to MIM with one layer (MIM1). The MIM2 IR frequencies and intensities have an ~75% improvement, compared to the corresponding values at the MIM1 level of theory. A similar improvement is also observed for anion, cation, and radical systems constructed from the neutral benchmark molecules. The accuracy and performance of the benchmark systems validate the MIM model for exploring the vibrational infrared spectra of large molecules.
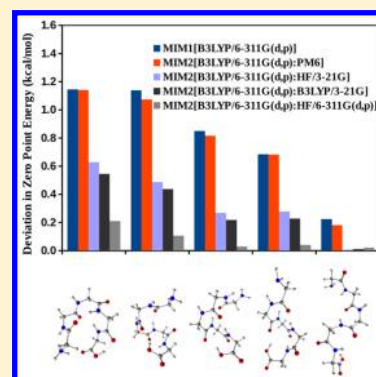
1. INTRODUCTION

The evaluation of equilibrium geometries and vibrational spectra of large molecules requires an accurate and efficient method to calculate first (gradient) and second derivatives (Hessian) of the total energy, with respect to the nuclear displacements. A precise gradient and Hessian evaluation is also necessary for an accurate characterization of the potential energy surface (PES) and searching for transition states in chemical reactions. Moreover, the calculation of the precise analytical Hessian and dipole derivatives is one of the bottlenecks for vibrational normal-mode analysis and accurate thermochemistry for large molecules.

Significant progress in the study of large molecules in the past decade has come from the development of fragment-based methods for the fast evaluation of the total energies. The basic idea involved in these methods is to generate a set of subsystems based on the nearsightedness idea of Kohn[1] and assemble their subsystem energies in a systematic way up to construct the total energy of the large molecule. Gordon et al.[2] have provided a recent review of different fragment-based methods. Although there are many such methods for evaluating the energies, only a few fragment-based methods have implemented the corresponding analytic first- and second-derivative expressions.

An early attempt toward the fragment-based gradient and vibrational spectral evaluation has been done through the integrated multicenter molecular orbital method of Sakai et al.[3,4] Similar applications have been carried out by other

fragmentation methods, such as the systematic fragmentation method of Collins et al.[5−7] or the molecular tailoring approach of Gadre et al.,[8−13] focusing mostly on nonbonded systems such as molecular clusters. In the simplest fragment-based implementations using a single layer, the fragment Hessian evaluations are independent of the chemical environment. In more-rigorous implementations, the chemical environment can be taken into account through point-charge embedding in the fragment gradient and Hessian evaluations. For example, the generalized energy-based fragmentation (GEBF) method has been extended to evaluating vibrational spectra and other ground-state properties.[14,15] Recently, Nakata et al. have derived the analytical expressions for energy derivatives in the fragment molecular orbital method (FMO) for evaluating the vibrational spectra and other related properties.[16−18]

In recent years, for improving the accuracy and efficiency of calculations on large molecules, hybrid methods have become popular in quantum chemistry.[19,20] As defined by Morokuma, *hybrid methods* are techniques that combine two or more computational methods in one calculation and allow one to study the chemistry of large systems accurately. Some of the popular hybrid fragment-based techniques, which combine different computational methods are (a) the multilevel fragment-based

approach (MFBA) of Salahub,[21] (b) combination of ONIOM with MM methods of Morokuma et al.,[22] (c) the multilayer formulation of the fragment molecular orbital method (FMO) of Fedorov et al.,[23] (d) Truhlar's electrostatic embedded molecular tailoring (EE-MTA) method,[24] (e) hybrid many-body interaction method (HMBI) of Beran,[25] (f) the combination of the divide and conquer (DC) with molecular fractionation with conjugated caps (MFCC) of Merz and co-workers,[26] (g) combination of molecular orbital (FMO) method interfaced with effective fragment potentials (EFP) of Gordon et al.,[27] (h) combination of systematic fragmentation method (SFM) and the effective fragment potential (EFP) of Collins et al.,[28] and (i) extended ONIOM XO method of Xu et al.[29] Although many of the above-discussed methods account for the chemical environment at some level of theory, the analytic expression for gradient and Hessian are mathematically complicated and may be laborious to implement.

We have recently proposed a new hybrid extrapolation method called Molecules-in-Molecules (MIM).[30] It uses a multilayer partitioning technique with multiple levels of theory using a generalized hybrid energy expression, similar in spirit to ONIOM. The steps involved in MIM are

(1) selection of nonoverlapping fragments by cutting appropriate covalent bonds in the large molecule,

(2) formation of overlapping primary subsystems by allowing the fragments to interact with neighboring fragments based on a well-defined scheme (distance-based or connectivity-based or number-based, *vide infra*),

(3) formation of derivative subsystems to take into account the overcounting of the regions in the primary subsystems, and

(4) summation of the energies of the primary and derivative subsystems (after link-atom termination of the broken covalent bonds and taking into account the appropriate signs of the energy terms).

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.

The performance of the MIM model has previously been validated for accurate energy evaluations of large chemical systems such as DNA fragments or peptides. Since the long-range interactions are taken into account at a low level of theory, the fragment size at the high level of theory can be fairly small. Using two (or more) layers in the MIM method makes the latter method suitable for efficient evaluations of accurate higher-energy derivatives. Hence, in the present work, we have extended this method for evaluating atomic forces for performing geometry optimizations and Hessian and dipole-derivative matrices for evaluating the infrared (IR) normal modes and intensities.

The following sections of the current paper is organized as follows: Section 2 describes the method involved in evaluating MIM energies and its higher derivatives. Section 3 presents MIM benchmark analysis of the geometrical parameters and infrared spectra on 25 peptide systems containing significant intramolecular interactions. Finally, section 4 provides a summary and our conclusions.

2. METHODOLOGY

This section discusses the steps involved in the implementation of the MIM energy gradients and subsequent geometry optimization and evaluation of infrared (IR) vibrational spectra.

All the actual and MIM calculations are performed through the Gaussian 09 program suite.[31] MIM fragmentation schemes for energy and IR evaluations are implemented through an external Perl script interface.

(1) The inputs for carrying out the MIM algorithm are the full molecular geometry, (either in Cartesian coordinates or the Z-matrix format) and the corresponding (nonoverlapping) fragments. In the original MIM paper,[30] the fragments were determined automatically by cutting all "single bonds". In the current paper that deals exclusively with peptides of different sizes, the initial fragments are generated by cutting only the backbone C−C bonds as commonly done for peptides in many other fragment-based methods.

(2) With this information, the external script constructs all the overlapping primary subsystems, where the interactions between fragments are included via a number-based scheme (*vide infra*). The derivative subsystems that result from the overcounting due to the overlapping primary subsystems are then determined (*vide infra*). The dangling bonds of the primary and derivative subsystems are terminated with link-hydrogen atoms. This link-hydrogen atom is placed along the vector connecting the supporting and host atoms, at a bond distance determined by a scaling factor.[32] Here, the scale factor, g, and the link-atom Cartesian coordinates are related through a linear equation,

$$R_2 = R_1 + g(R_3 - R_1) \tag{1}$$

Here, R_2 is the Cartesian component of the link-atom, and R_1 and R_3 are that of the supporting and host atoms, respectively. The values of g are taken from the standard implementation of the link atoms in the ONIOM method. This scaling factor is dependent on the nature of the bond being cut, while constructing a model system from the real system. The scale factor g is defined as the ratio of sum of the covalent radii of atoms 1 (support) and 2 (link), to that of atoms 1 (support) and 3 (host), respectively.[32]

(3) MIM properties can be evaluated at a single fragmentation layer (MIM1) or with multiple (two or more, MIM2, MIM3, etc.) layers. In this paper, we will focus on the two-layer MIM2 model. For extrapolating MIM2 properties to a high level of theory, all the subsystem energies, atomic forces, and Hessians are evaluated at high and low levels of theory. The definition of the fragments, subsystems, and the resulting layers can be very flexible, as outlined in the original MIM paper.[30] We have used a simpler, more specific, two-layer ONIOM-type notation in this paper. The ONIOM hybrid extrapolated total energy is constructed from the independent real-low (E_{rl}), model-low (E_{ml}), and model-high (E_{mh}) energies, through the general expression

$$E_{total} = E_{rl} - E_{ml} + E_{mh} \tag{2}$$

The above ONIOM extrapolation expression holds for large molecules with only one model system (i.e., active site). However, in MIM, the full system has been fragmented to yield overlapping subsystems, and each of them can be considered as a separate model system. In addition, the overcounting of the overlapping regions of the primary subsystems (at model-high and model-low levels of theory) are corrected through the formation of derivative subsystems. Thus, we can represent MIM2 as a generalized ONIOM, where E_{ml} and E_{mh} have been modified to represent the sum over all the subsystems as follows:

$$E_{ml/mh} = \sum_i E_{l/h}^i - \sum_{ij} E_{l/h}^{i \cap j} + \dots + (-1)^{n-1} \sum_n E_{l/h}^{i \cap j \cap k \dots \cap n} \tag{3}$$

Here, $E_{l/h}^i$ represents the energy of the ith primary subsystem (at low or high level of theory) while the other terms involving the overlapping terms represent the derivative subsystems. Thus, unlike in the original ONIOM, the redefined "model-high" and "model-low" levels in MIM extend through the whole molecule via summation over the subsystems. If the MIM is restricted to a single layer (MIM1), E_{mh} yields the total energy.

(4) For the precise atomic force evaluation, link-atom force components are projected back onto the supporting and host atoms through a general Jacobian projection method.[32]

$$
\begin{aligned}
F_a &= \frac{\partial E_{total}}{\partial X_a} \\
&= \frac{\partial E_{rl}}{\partial X_a} - \frac{\partial E_{ml}}{\partial X_a} J(R_2; R_1, R_3) + \frac{\partial E_{mh}}{\partial X_a} J(R_2; R_1, R_3)
\end{aligned}
\tag{4}
$$

Here, the index a corresponds to the Cartesian components of the X, Y, or Z directions. The elements of the 3×3 Jacobi matrix (J) for projecting the link-atom Cartesian component onto corresponding supporting-atom component is defined as the partial derivative of the Cartesian component of the former with respect to the latter:

$$
\frac{\partial r_{2,a}}{\partial r_{1,b}} = (1 - g)\delta_{a,b}
\tag{5}
$$

Here, δ is the Kronecker delta, and a and b are the Cartesian components.

In a similar manner, the Jacobi matrix element for projecting the link-atom forces onto the host atom is given as the analogous partial derivative, with respect to the host-atom:

$$
\frac{\partial r_{2,a}}{\partial r_{3,b}} = g\delta_{a,b}
\tag{6}
$$

Thus, the force vector elements of the host- and support-atoms are modified from the link-atom Jacobi projection method. This approach is similar to that used in the ONIOM method,[32] and also in the fragmentation method of Řezáč and Salahub.[21]

(5) With the same physical picture, the mass-weighted Hessian of the link atom is projected back onto the supporting and host atoms through the general expression[32]

$$
H_{ab} = \frac{\partial^2 E_{total}}{\partial X_a \partial X_b} = \frac{\partial^2 E_{rl}}{\partial X_a \partial X_b} - J^T \frac{\partial^2 E_{ml}}{\partial X_a \partial X_b} J + J^T \frac{\partial^2 E_{mh}}{\partial X_a \partial X_b} J
\tag{7}
$$

The Hessian ($\mathbf{H}$) is a real and symmetric matrix, and the H_{ab} elements, which are modified from the link-atom Jacobi-projection, are the blocks involving link-atom indices. For example, for a fragment with only one link atom, on projecting link-atom contributions onto the host and support atoms, the blocks of Hessian elements with link-host and link-support indices are modified. The same procedure is used in the standard ONIOM method.[32]

(6) These atomic forces, and Hessians, are fed back to the Gaussian 09 program to perform the geometry optimization.[31] Steps (1)−(4) (and (5), if needed) are followed to construct the extrapolated energy, gradients (and Hessians if needed), for each optimization cycle. The geometry optimization cycle is carried out until convergence.

(7) At this final optimized geometry, the Hessian and the dipole derivative matrices are evaluated for constructing IR spectra. IR intensity is directly proportional to the square of the

dipole derivate, and this matrix is also constructed from the general expression

$$
\mu_a = \frac{\partial \mu_{total}}{\partial X_a} = \frac{\partial \mu_{rl}}{\partial X_a} - \frac{\partial \mu_{ml}}{\partial X_a} J + \frac{\partial \mu_{mh}}{\partial X_a} J
\tag{8}
$$

The dipole derivative of the fragment link-atom contributions are projected back onto the corresponding supporting and host atoms through a general procedure,[32] similar to that in eq 4. Since the dipole derivative is the first derivative in Cartesian coordinates like the force expression, only the elements with the host and supporting atoms are modified.

(8) The mass-weighted Hessian is diagonalized to evaluate the vibrational normal-mode frequencies, and the dipole derivatives are used to derive the corresponding IR intensities.

The above implementation of the algorithm is general, and it will work with any arbitrary fragmentation scheme at the primary level. In this work, it deals with peptide systems; the initial nonoverlapping fragments are constructed by breaking the backbone C−C bonds in the peptide. The latter non-overlapping fragments are combined by employing a number-based criterion, based on the connectivity information for constructing the overlapping primary subsystems. These primary fragments are constructed manually in this work due to the simplicity of the linear peptide systems, though an automated procedure applicable to more-complex chemical systems is also being optimized. Depending on the size of the parent peptide systems (*vide infra*), we use dipeptide, tripeptide, or tetrapeptide units as primary subsystems. The overlapping primary subsystems are then used to obtain the derivative subsystems. The derivative subsystems are also simple for the peptide systems considered and involve only binary overlaps between fragments. Finally, the above-mentioned MIM procedure for performing the geometry optimization and evaluating vibrational spectra of a molecule is independent of the levels of theory and basis sets. While the current description involves the use of hydrogen link atoms to saturate the broken covalent bonds, the same method can be applied with minor modifications for molecular clusters, where no link atoms are involved while constructing fragments.

In the present work, we have employed B3LYP/6-311G(d,p) and MP2/6-311G(d,p) as high levels of theory, for assessing the performance of MIM1 and MIM2 IR spectra. For MIM2, we have employed a second layer, at PM6, HF/3-21G, B3LYP/3-21G, and HF/6-311G(d,p) levels of theory for correcting the molecular properties. These results are compared with the actual calculations performed at the B3LYP/6-311G(d,p) and MP2/6-311G(d,p) levels of theory. The entries in the bracket following "MIM" give the level(s) of theory and basis set(s) employed for the calculations.

3. RESULTS AND DISCUSSIONS

This section discusses the quantitative performance of our method for geometry optimization, and the evaluation of Hessian and dipole-derivative matrices for IR vibrational spectra. The benchmark system chosen for this work are 25 covalently bonded peptide systems with unavoidable intramolecular interactions and are depicted in Figures 1, 2, and 6 (provided later in this work). The structure numbers are the same in all figures and tables throughout the text. To assess the accuracy and applicability of the MIM method, the benchmark systems are divided into two subsets: (A) a variety of 21 small and medium peptides containing 3−8 amino acids, including some
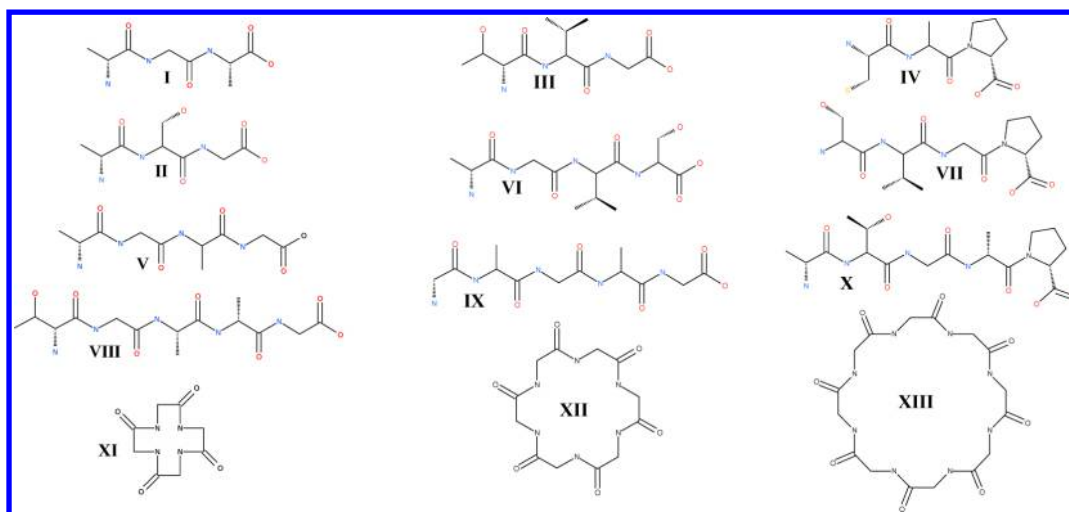
Figure 1. Test systems employed for the assessment of MIM. Structures **I–IV** are amino acid trimers, **V–VII** are amino acid tetramers, **VIII–X** are amino acid pentamers, and **XI–XIII** are cyclic amino acids. The optimized geometries at Actual[B3LYP/6-311G(d,p)] are furnished in the Supporting Information. The structure numbers are same in all the tables and text. Refer to the tables and text for more details.
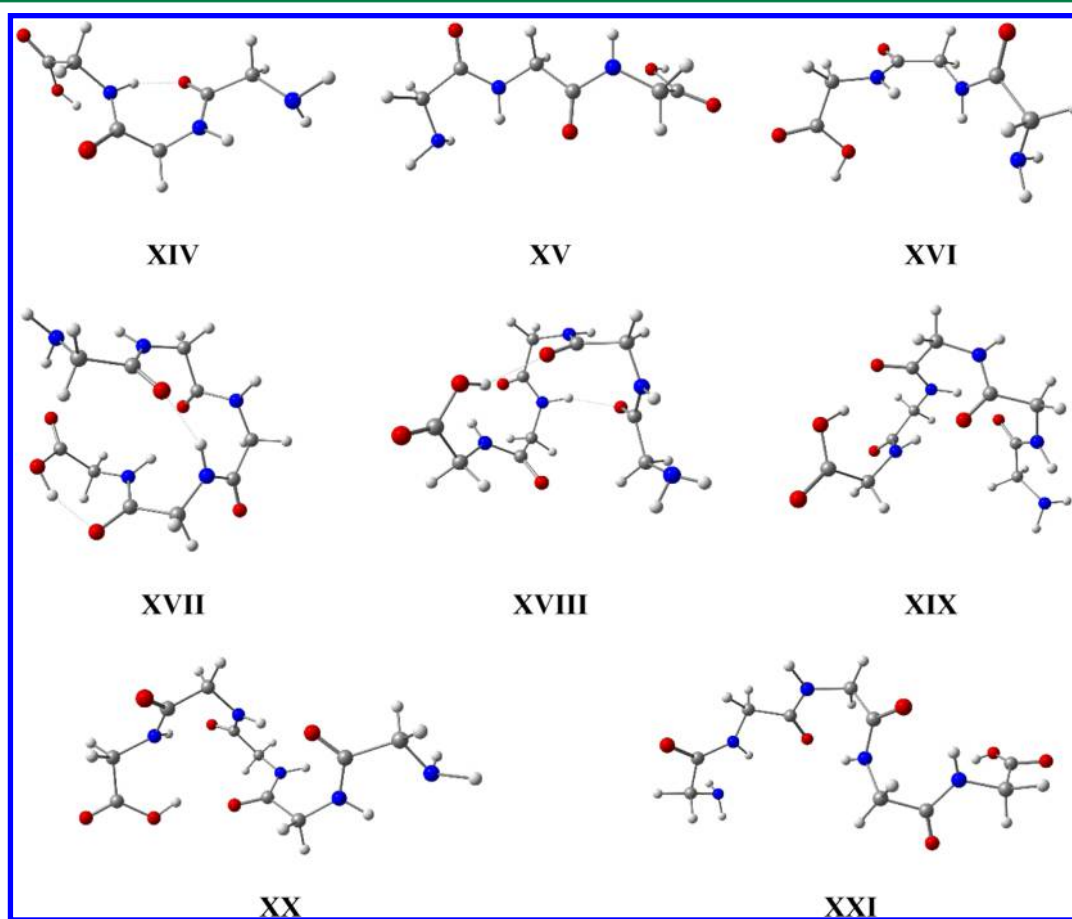

Figure 2. Conformers of (Glycine)$_3$ (**XIV–XVI**), and (Glycine)$_5$ (**XVII–XXI**) optimized at MIM2[B3LYP/6-311G(d,p):HF/6-311G(d,p)]. The corresponding energies are reported in Table 1. Refer to the tables and text for more details.

open as well as cyclic conformational isomers, and (B) α- and β-conformational isomers of large glycine polypeptides (10-mer and 15-mer).

3.1. Test Set A. The initial 21 benchmark molecules chosen are small to medium peptides containing 3–8 amino acids and are depicted in Figures 1 and 2. Ten (10) of these (labelled as **I** in Figure 1) are small peptides containing 3, 4, or 5 amino

acids. A variety of amino acids has been included in forming these peptides. The remaining structures are all peptides composed of glycine. Structures **XI**, **XII**, and **XIII** (see Figure 1) are cyclic peptides composed of 4, 6, and 8 glycines. We have also included conformational isomers of (Glycine)$_3$ and (Glycine)$_5$ (structures **XIV–XXI** in Figure 2) containing prominent intramolecular hydrogen bonding interactions.

Geometry optimizations were performed on the test systems (as described in the Methodology section) using dipeptides as primary subsystems at three levels of theory: (1) full B3LYP (or MP2)/6-311G(d,p), (2) MIM1 at B3LYP (or MP2)/6-311G(d,p), and (3) MIM2 with a combination of B3LYP (or MP2)/6-311G(d,p):HF/6-311G(d,p) levels of theory. While all 21 systems were optimized for B3LYP/6-311G(d,p) as the high level, a subset of systems were considered using MP2/6-311G(d,p) as the high level of theory. The use of HF/6-311G(d,p) as the second layer of theory is more appropriate in the latter case. The use of lower levels of theory may be more appropriate for B3LYP/6-311G(d,p) (*vide infra*).

The optimized geometrical parameters of the test sets at MIM2[B3LYP/6-311G(d,p):HF/6-311G(d,p)] are compared with the actual ones in Figure S1 in the Supporting Information. To evaluate the accuracy of the optimized geometrical parameters, we have compared the MIM2 bond lengths (*x*-axis) of these 21 benchmark systems with the actual optimized values (*y*-axis) in Figure S1(A) in the Supporting Information. A corresponding comparison for all the optimized angles and dihedral angles is shown in Figure S1(B) in the Supporting Information. The fitted lines to these sets of data points have a correlation coefficient of ∼1.0, suggesting that the geometrical parameters of large chemical systems can be reproduced very accurately. A corresponding comparison of geometrical parameters evaluated at MIM2[MP2/6-311G-(d,p):HF/6-311G(d,p)] and MP2/6-311G(d,p) levels of theory, for a subset of molecules, also yields similar accuracy.

The use of a lower level of theory for the second layer was also explored. For a subset of molecules, we have presented a similar comparison in Figures 3A−F, for the MIM1[B3LYP/6-311G(d,p)], MIM2[B3LYP/6-311G(d,p):HF/3-21G], and MIM2[B3LYP/6-311G(d,p):HF/6-311G(d,p)] levels of theory. The bond length correlation coefficient is ∼0.99 in Figure 3 for all combinations (Figures 3A, 3C, and 3E). However, the corresponding angular correlation coefficients show an improvement from 0.987 (intercept 1.560), to 0.997 (intercept 0.054), to 0.999 (*y*-intercept 0.000), respectively, for the three models. In Figure 3, the corresponding root-mean-square (rms) deviations, compared to the actual optimized geometrical parameters, are 2.5×10^{-3} Å (Figure 3A), 8.800° at MIM1[B3LYP/6-311G(d,p)] (Figure 3B), 2.1×10^{-3} Å (Figure 3C), 3.782° at MIM2[B3LYP/6-311G(d,p):HF/3-21G] (Figure 3D), 0.1×10^{-3} Å (Figure 3E), and 0.001° at MIM2[B3LYP/6-311G-(d,p):HF/6-311G(d,p)] (Figure 3F). In Figures 3B, 3D, and 3F, it is clear that, in the MIM1 model, some of the geometrical parameters, particularly the bond angles and dihedral angles, are significantly deficient, because of the missing intramolecular weak interactions. The latter missing long-range interactions are taken into account in the second layer of theory in the MIM2 model.

Table 1 presents the actual energies of the 21 benchmark molecules and a comparison of MIM1 and MIM2 energies, relative to the actual ones at B3LYP/6-311G(d,p) and MP2/6-311G(d,p) levels of theory. The MP2 energies listed in Table 1 were obtained as single points at the optimized B3LYP geometries. The computed energy values show an 85% improvement in accuracy at MIM2, compared to MIM1. The latter improvement in accuracy shows the importance of the second layer of theory for accounting for the intramolecular long-range interactions.

Tables 2 and 3 present the frequencies and intensities of the two most intense normal modes for the 21 test molecules at MIM1 and MIM2 levels, and the mean absolute deviation
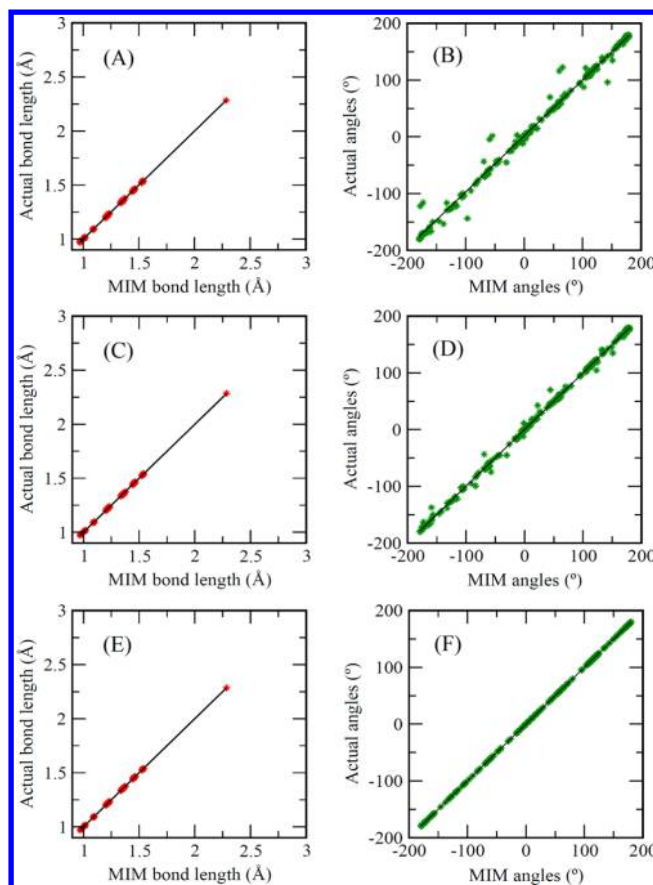

Figure 3. Comparison of the Actual[B3LYP/6-311G(d,p)] bonds lengths (in Å), bond angles (in degrees), and dihedral bond angles (in degrees) with (A, B) MIM1[B3LYP/6-311G(d,p)], (C, D) MIM2[B3LYP/6-311G(d,p):HF/3-21G], and (E, F) MIM2[B3LYP/6-311G(d,p):HF/6-311G(d,p)] levels of theory. The corresponding geometries are depicted in Figure 2, and the energies are listed in Table 1. Refer to the text for more details.

(MAD) values of the spectra including all the $3N − 6$ normal modes at the MIM1 and MIM2 levels, relative to the actual ones at the B3LYP/6-311G(d,p) and MP2/6-311G(d,p) levels of theory. The MP2 energies listed in Table 3 correspond to completely optimized values. It is observed that the accuracy of the full spectrum improves on incorporating a second layer in MIM2. The MAD at MIM2 shows an ∼75% improvement over that observed at MIM1, in both IR normal-mode frequencies and corresponding intensities. These 21 benchmark molecules have an average MAD of 0.51 cm^{-1} in the vibrational frequencies, and 2.32 km/mol in the corresponding vibrational intensities at the MIM2[B3LYP/6-311G(d,p):HF/6-311G(d,p)] level of theory. The corresponding values for a subset of molecules at the MIM2[MP2/6-311G(d,p):HF/6-311G(d,p)] level of theory are 0.86 cm^{-1} and 2.10 km/mol, respectively (see Table 3), similar to the performance observed for B3LYP.

Figures 4A−E depict the corresponding IR spectra of some of the benchmark molecules. The corresponding two highest intensity modes for all of the molecules at the B3LYP/6-311G(d,p) level of theory are reported in Table 2. In all of these amino acid combinations, the listed modes mostly correspond to the C═O and the N−H stretching modes. For comparison, the MIM vibrational spectra are shown, along with the actual ones in Figure S2 in the Supporting Information, between 600 cm^{-1} and 3600 cm^{-1} spectral region, showing

Table 1. Structure Number, System Name, Number of Basis Functions (N_{Basis}), Actual Total Energies (E) at the B3LYP/6-311G(d,p) and MP2/6-311G(d,p) Levels of Theory, along with the Corresponding MIM1 and MIM2 Energy Deviations, Relative to the Actual Energy

No.	system	N_{Basis}	E_{B3LYP}^{Actual} (Hartree)	ΔE_{B3LYP}^{MIM1} (kcal/mol)	ΔE_{B3LYP}^{MIM2} (kcal/mol)	E_{MP2}^{Actual} (Hartree)	ΔE_{MP2}^{MIM1} (kcal/mol)	ΔE_{MP2}^{MIM2} (kcal/mol)
I	AGA	360	−779.31384	0.38	0.22	−777.19860	0.26	0.21
II	ASG	378	−854.54761	−0.23	−0.21	−852.26904	0.01	−0.00
III	TVG	468	−972.52280	1.12	0.81	−969.86289	0.79	0.48
IV	PAC	464	−1294.27510	0.08	−0.04	−1291.25140	0.29	0.41
V	AGAG	450	−987.38502	0.20	0.19	−984.71251	0.50	0.42
VI	AGVS	588	−1219.92060	0.35	0.01	−1216.57760	1.01	0.67
VII	SVGP	636	−1297.35150	0.38	0.33	−1293.77610	0.99	0.94
VIII	TGAAG	618	−1349.34043	0.53	0.28	−1345.69160	1.02	0.77
IX	GAGAG	540	−1195.45492	0.29	0.44	−1192.22520	0.70	0.65
X	ASGAP	666	−1426.77492	0.31	0.41	−1422.89120	0.82	0.72
XI	Cyclic(Gly)$_4$	360	−832.27675	−1.08	−0.80	−830.05593	0.99	0.27
XII	Cyclic(Gly)$_6$	540	−1248.42034	5.13	−1.91	−1245.08540	7.68	0.64
XIII	Cyclic(Gly)$_8$	720	−1664.53620	5.08	0.49	−1660.07940	7.89	3.30
XIV	Con(Gly)$_3$	300	−700.65787	2.65	−0.59	−698.79798	2.81	−0.44
XV	Con(Gly)$_3$	300	−700.65705	−0.48	−0.01	−698.79549	−0.15	0.12
XVI	Con(Gly)$_3$	300	−700.65783	1.30	−0.26	−698.79703	1.54	−0.02
XVII	Con(Gly)$_5$	480	−1116.82181	19.77	−1.40	−1113.85570	22.42	3.04
XVIII	Con(Gly)$_5$	480	−1116.81052	13.40	−1.27	−1113.84290	15.79	1.12
XIX	Con(Gly)$_5$	480	−1116.80757	10.20	−0.95	−1113.84060	11.59	0.43
XX	Con(Gly)$_5$	480	−1116.80438	8.75	−0.43	−1113.83510	9.59	0.41
XXI	Con(Gly)$_5$	480	−1116.79814	0.44	−0.32	−1113.82390	0.98	0.22
	average[a]			3.42	0.50		4.18	0.73

[a]The last row reports the average of all the corresponding values in each column.

Table 2. Two Most Intense Frequencies ($\nu_{B3LYP}^{Highest}$) of the Test Systems (Depicted in Figures 1 and 2) at Actual[B3LYP/6-311G(d,p)] Level, and the Mean Absolute Deviation ($\Delta\nu$) of the Vibrational Modes at the MIM1 and MIM2[B3LYP/6-311G(d,p):HF/6-311G(d,p)] Levels from the Corresponding Actual Ones[a]

No.	ν_{B3LYP} ($I_{B3LYP}^{Highest}$) (cm^{-1} (km/mol))	ν_{B3LYP} ($I_{B3LYP}^{Highest}$) (cm^{-1} (km/mol))	$\Delta\nu_{B3LYP}^{MIM1}$ (cm^{-1})	ΔI_{B3LYP}^{MIM1} (km/mol)	$\Delta\nu_{B3LYP}^{MIM2}$ (cm^{-1})	ΔI_{B3LYP}^{MIM2} (km/mol)
I	1526.84 (579.99)	1735.85 (371.88)	0.53	3.45	0.22	0.80
II	1529.33 (561.69)	1729.68 (310.97)	0.04	0.26	0.01	0.07
III	1518.20 (474.47)	1729.74 (374.89)	0.79	6.37	0.24	1.10
IV	1524.83 (643.16)	1730.27 (329.17)	0.50	3.30	0.12	0.56
V	1522.14 (1052.93)	1719.01 (440.12)	0.48	5.52	0.22	0.93
VI	1516.41 (748.74)	1722.76 (413.54)	0.66	7.25	0.21	1.50
VII	1448.73 (556.86)	1528.05 (522.99)	0.50	8.04	0.17	1.88
VIII	1518.81 (1691.63)	1727.65 (493.15)	0.67	9.09	0.23	1.40
IX	1522.66 (1582.90)	1720.17 (533.98)	0.95	8.73	0.36	1.76
X	1518.86 (1516.07)	1727.74 (581.69)	0.57	22.40	0.21	4.99
XI	1751.90 (734.75)	1560.91 (256.14)	3.82	4.63	1.10	3.78
XII	1749.08 (618.85)	1572.30 (318.18)	2.16	6.27	0.89	2.78
XIII	1764.21 (1103.98)	1546.27 (619.45)	1.31	15.04	0.56	5.53
XIV	1865.55 (360.74)	1785.92 (326.13)	2.28	5.41	0.49	1.27
XV	1537.07 (580.72)	1752.74 (398.97)	1.48	5.55	0.30	1.02
XVI	1834.95 (339.87)	1534.99 (336.23)	1.12	5.00	0.34	0.85
XVII	3254.39 (615.92)	3484.09 (362.84)	5.87	10.28	1.49	6.73
XVIII	3667.39 (464.52)	1747.28 (442.06)	10.46	25.73	1.62	5.00
XIX	1724.22 (407.63)	1840.59 (347.05)	3.69	19.07	0.77	2.42
XX	1762.15 (390.38)	3693.09 (380.57)	3.64	22.52	0.69	2.81
XXI	1543.41 (472.15)	1866.71 (373.13)	1.52	6.02	0.47	1.61
	average[b]		2.05	9.52	0.51	2.32

[a]Refer to text for more details. [b]The last row reports the average of all the corresponding values in each column.

good agreement. The full spectra of all these test systems at the actual, MIM1, and MIM2 levels of theory are reported in the Supporting Information.

Another important test is the ability of MIM to reproduce the relative energies between conformations that have different intramolecular interactions. The most important observation is

Table 3. Structure Number, Actual[MP2/6-311G(d,p)] Energy and Deviation of This Energy from the Corresponding MIM2[MP2/6-311G(d,p):HF/6-311G(d,p)] Energies, the Highest Intense Frequencies of the Test Systems at MIM2 Levels of Theory, and Mean Absolute Deviation (Δ) of All of the $3N - 6$ Normal Mode Frequencies and Intensities from the Actual Values[a]

No.	E_{MP2}^{Actual} (Hartree)	ΔE_{MP2}^{MIM2} (kcal/mol)	$\nu_{MP2}(I_{MP2}^{Highest})$ (cm^{-1}(km/mol))	$\nu_{MP2}(I_{MP2}^{Highest})$ (cm^{-1}(km/mol))	$\Delta\nu_{MP2}^{MIM2}$ (cm^{-1})	ΔI_{MP2}^{MIM2} (km/mol)
I	−777.19939	−0.44	1540.31 (480.73)	1755.92 (273.23)	0.33	1.12
II	−852.27011	−0.50	1544.62 (543.56)	1763.26 (306.23)	0.41	1.01
XI	−830.05701	−1.48	1777.78 (667.99)	1578.86 (220.43)	1.81	4.07
XIV	−698.79883	0.10	1400.00 (324.89)	1800.50 (290.50)	1.24	2.76
XV	−698.79705	−0.47	1549.85 (589.61)	1780.13 (353.60)	0.11	0.29
XVI	−698.79821	−0.12	1541.17 (359.66)	1850.33 (256.76)	0.59	1.81
XVII	−1113.85770	−1.82	3337.45 (633.05)	3547.70 (407.94)	0.37	1.42
XVIII	−1113.84940	−1.36	3702.27 (619.52)	1776.40 (587.68)	0.24	1.56
XIX	−1113.84330	−3.32	1581.81 (324.68)	3556.35 (293.47)	1.46	3.65
XX	−1113.83700	0.48	3762.06 (410.31)	3571.64 (404.15)	0.50	0.86
XXI	−1113.83050	0.75	1873.33 (303.62)	1792.32 (281.17)	2.35	4.57
average[b]		0.99			0.86	2.10

[a]Refer to the text for more details. [b]The last row reports the average of all the corresponding values in each column.
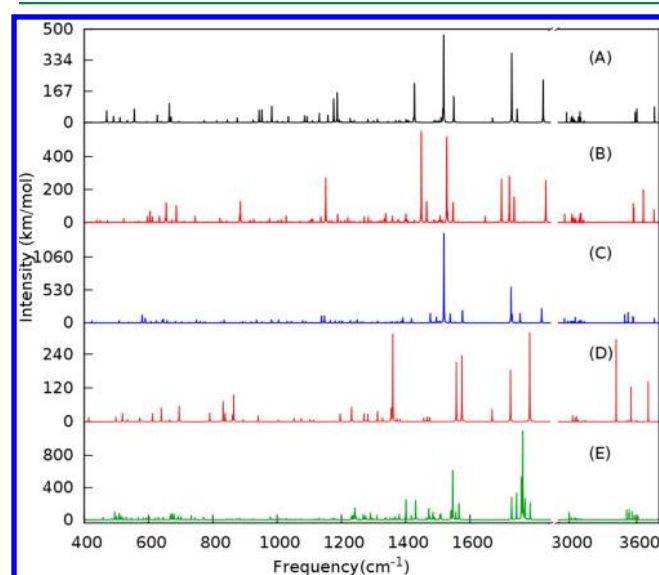

Figure 4. Frequency (in cm^{-1}) versus intensity (in km/mol) infrared spectra of (A) structure **III**, (B) structure **VII**, (C) structure **X**, (D) structure **XIV**, and (E) structure **XIII**, evaluated at MIM2[B3LYP/6-311G(d,p):PM6]. The corresponding geometries are depicted in Figure 1. Refer to the text for more details.

that relative energies between conformers, e.g., between **XVII** and **XXI**, representing five different isomers of (Glycine)$_5$, is not reproduced well at MIM1; however, upon incorporating a second layer at MIM2, the relative ordering is consistent with the actual ones. The MIM2 extrapolated relative energies are ∼85% more accurate than the MIM1 energies.

For benchmarking the effect of combinations of levels of theory and basis, on the accuracy of the frequencies and intensities, a comparative analysis of the performance for five conformational isomers of (Glycine)$_5$ is depicted in Figure 5. The MAD at the MIM1[B3LYP/6-311G(d,p)], MIM2[B3LYP/6-311G(d,p):PM6], MIM2[B3LYP/6-311G(d,p):HF/3-21G], MIM2[B3LYP/6-311G(d,p):B3LYP/3-21G], and MIM2-[B3LYP/6-311G(d,p):HF/6-311G(d,p)] levels of theory are depicted in Figures 5A−D. Figure 5A illustrates the MAD in frequency, using dipeptides as the primary subsystems, and the values obtained using the five different models are 7.24, 5.97,

3.40, 2.90, and 1.30 cm^{-1}, respectively. The corresponding MAD in intensity values (Figure 5B) are 18.92, 13.6, 8.88, 5.78, and 2.34 km/mol, respectively. Similarly, we have evaluated the MIM frequencies and intensities in Figure 5C with the same combinations of levels of theory, but with tripeptides as the primary subsystems. In Figure 5D, the corresponding MAD in frequencies are 2.92, 2.72, 1.40, 1.17, and 0.71 cm^{-1}, respectively, and that of intensity is 8.09, 4.83, 3.24, 2.43, and 1.34 km/mol, respectively. Increasing the size of the primary subsystems from dipeptide to tripeptide results in a 45%−60% improvement in the accuracy of the calculated frequencies and intensities. A comparison of the calculated spectra for the most stable isomer (**XVII**) at MIM1[B3LYP/6-311G(d,p)], MIM2-[B3LYP/6-311G(d,p):HF/3-21G], MIM2[B3LYP/6-311G-(d,p):HF/6-311G(d,p)], and MIM2[B3LYP/6-311G-(d,p):B3LYP/3-21G] levels of theory with the actual spectra are depicted in Figure S3(A−D) in the Supporting Information.

We have performed a similar analysis on the most stable isomer (**XVII**) at the MP2 level of theory in Figures S4 and S5 in the Supporting Information, with dipeptide and tripeptide as the primary fragments. The corresponding MAD in the calculated frequency (intensity) at the MIM1[MP2/6-311G-(d,p)], MIM2[MP2/6-311G(d,p):PM6], MIM2[MP2/6-311G-(d,p):HF/3-21G], MIM2[MP2/6-311G(d,p):B3LYP/3-21G], and MIM2[MP2/6-311G(d,p):HF/6-311G(d,p)] levels of theory from the actual spectra are 5.83 (12.02), 3.83 (9.13), 3.03 (6.43), 2.80 (5.54), and 1.52 (2.97) cm^{-1} (km/mol), respectively, with dipeptide as the primary fragments. The corresponding values with tripeptide as primary fragments are 2.97 (5.52), 1.95 (4.04), 1.52 (2.75), 1.51 (2.53), and 0.75 (0.45) cm^{-1} (km/mol), respectively. Similarly to the B3LYP level of theory, with the size of the primary fragments, there is a 50%−60% improvement in the accuracy of the calculated energies, frequencies, and intensities.

We have also included some charged and radical systems in our benchmark test molecules. The most negative and positive ends in the electrostatic potential topography of the neutral molecules are near the end amino and carboxyl functional groups, respectively. Hence, we have constructed the corresponding cation and anion species from the neutral species by adding and removing a proton, respectively. In addition, the radical species is constructed by removing a hydrogen from the
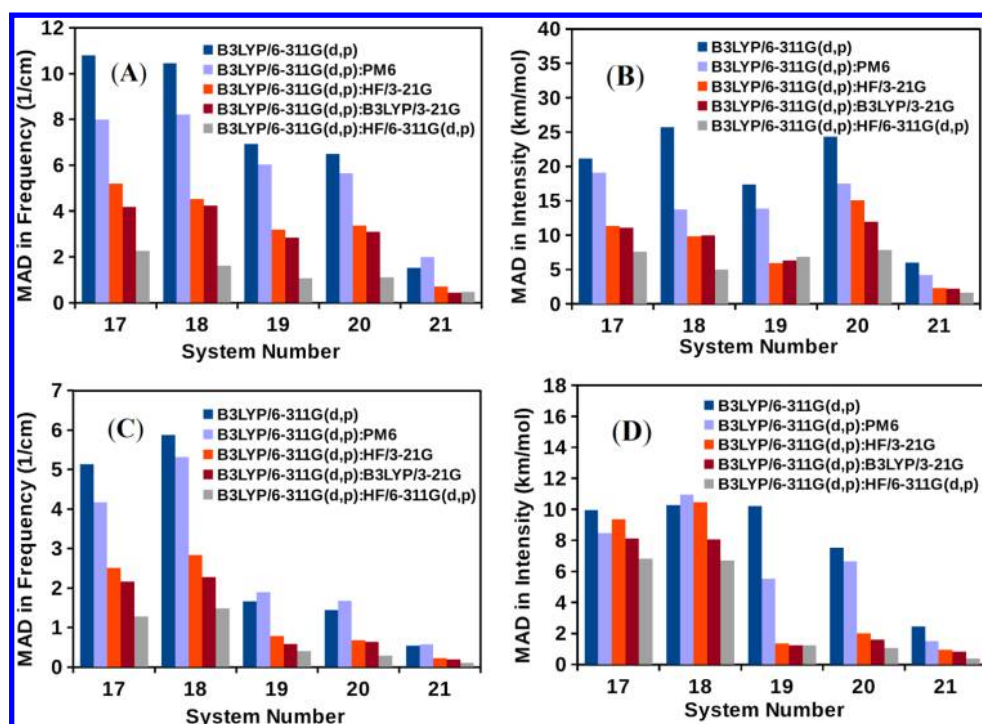
Figure 5. Mean absolute deviation (MAD) for (A, C) all the MIM normal-mode frequencies (cm^{-1}) and (B, D) in the intensities (km/mol) of structures **XVII−XXI**), relative to the actual high calculations, with dipeptides and tripeptides as overlapping primary subsystems. The corresponding MAD at the MIM1[B3LYP/6-311G(d,p)] and MIM2[B3LYP/6-311G(d,p):HF/6-311G(d,p)] levels of theory are reported in Table 2. Refer to the text for more details.

neutral species. We have constructed these anion, cation, and radical species for structures **I**, **VI**, and **IX** reported in Figure 1. A similar number-based fragmentation scheme based on the connectivity information is followed for the charged systems. While constructing the fragments, we assume that the charge is localized within a certain region of the molecule. Hence, in the cation, anion, and radical fragmentation schemes, only one of the fragments is assigned an appropriate charge and multiplicity, and the remaining fragments are assigned as neutral. These geometries are optimized at actual and MIM levels of theory while evaluating the IR spectrum. A comparison of the actual vibrational IR spectra with MIM-based ones is depicted in Figure S6 in the Supporting Information and Figure 6, with one and two layers, respectively. In both illustrations, each spectrum corresponds to neutral forms (spectrum A), anion forms (spectrum B), cation forms (spectrum C), and radical forms (spectrum D) corresponding to structure **IX**. An analysis of the MAD in the energies and frequencies at MIM1 and MIM2 is reported in Table 4. The MAD in the calculated energy at the MIM1[B3LYP/6-311G(d,p)] and MIM2[B3LYP/6-311G(d,p):HF/6-311G(d,p)] levels of theory are 3.39 and 0.90 kcal/mol, respectively. Similarly, the corresponding MAD in frequencies and intensities are 0.92 (6.50) and 0.44 (1.60) cm^{-1} (km/mol), respectively.

3.2. Test Set B. The second benchmark set are the α- and β-conformational isomers of (Glycine)$_n$ polypeptides, with n = 10 and 15. First, we consider the performance of the 2-layer method: MIM2[B3LYP/6-311G(d,p):HF/3-21G]. The overlapping primary subsystems were explored via two models: one composed of tripeptide units or another composed of tetrapeptide units. It is expected that the use of larger primary subsystems (i.e., tetrapeptide) will be more accurate, since some intramolecular interactions are captured directly at the
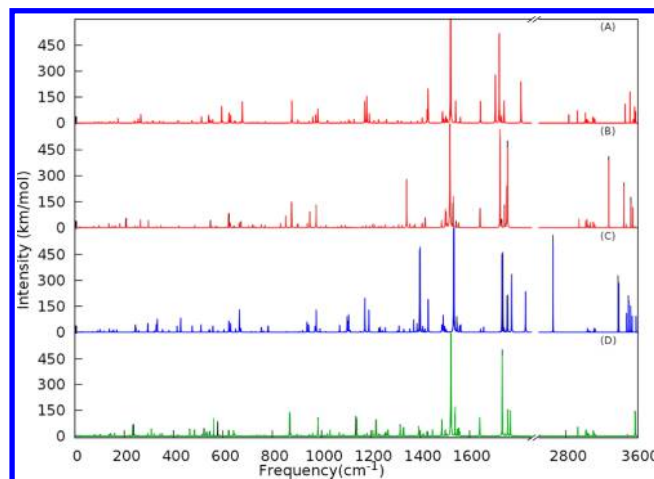

Figure 6. Vibrational IR spectra of the different forms of **IX**: neutral, in red (spectrum A), anion, in brown (spectrum B), cation, in blue (spectrum C), and radical, in green (spectrum D). In all of the vibrational IR spectra, the Actual[B3LYP/6-311G(d,p)] is depicted in black, and the MIM2[B3LYP/6-311G(d,p):HF/6-311G(d,p)] is shown in red, brown, blue, and green, respectively. An analysis of these spectra is reported in Table 4. Here, the primary overlapping fragments are constructed from tripeptide units. Refer to the text for more details.

high level of theory, although all interactions are always included at the low level. The optimized structures of (Glycine)$_n$ (n = 10 and 15) are depicted in Figure 7A−C. Note that the β-isomer of (Glycine)$_{15}$ is also included in the analysis, although it is not shown in Figure 7.

In order to show the difference in the spectral characteristics of open (β) and helical (α) isomers, Hessian matrix elements evaluated through the MIM method are compared with the

Table 4. Structure Number, Actual[B3LYP/6-311G(d,p)] Energy and Deviation of This Energy from the Corresponding MIM1[B3LYP/6-311G(d,p)] and MIM2[B3LYP/6-311G(d,p):HF/6-311G(d,p)] Energies, the Highest Intense Frequencies of the Test Systems at MIM1 and MIM2 Levels of Theory, and the Mean Absolute Deviation (Δ) of All of the $3N - 6$ Normal Mode Frequencies and Intensities from the Actual Values[a]

No.	E^{Actual} (Hartree)	ΔE^{MIM1} (kcal/mol)	ΔE^{MIM2} (kcal/mol)	$\nu(I^{Highest})$ k(cm^{-1} (km/mol))	$\Delta\nu^{MIM1}(I^{MIM1})$ (cm^{-1} (km/mol))	$\Delta\nu^{MIM2}(I^{MIM2})$ (cm^{-1} (km/mol))
I, anion	−778.77439	−0.74	−0.10	1518.50 (725.46)	0.49 (2.76)	0.23 (0.87)
VI, anion	−986.84886	−3.39	−0.74	1520.67 (1553.30)	0.84 (7.55)	0.42 (1.24)
IX, anion	−1194.92033	−4.76	−1.26	1520.21 (2109.35)	0.93 (12.26)	0.43 (1.95)
I, cation	−779.69611	−5.09	−1.37	1414.72 (571.34)	1.43 (4.62)	0.50 (2.01)
VI, cation	−987.77060	−7.31	−1.97	1543.55 (1182.85)	1.42 (12.04)	0.35 (1.50)
IX, cation	−1195.84150	−8.76	−2.41	1536.75 (1669.86)	1.50 (9.44)	0.48 (2.68)
I, radical	−778.63210	−0.22	−0.02	1526.60 (632.83)	0.49 (1.37)	0.48 (0.89)
VI, radical	−986.70241	0.10	−0.05	1523.08 (1003.52)	0.485 (3.28)	0.46 (1.04)
IX, radical	−1194.77211	0.12	−0.18	1524.23 (1669.56)	0.64 (5.21)	0.59 (2.19)
average[b]		3.39	0.90		0.92 (6.50)	0.44 (1.60)

[a]Refer to the text for more details. [b]The last row reports the average of all the corresponding values in each column.
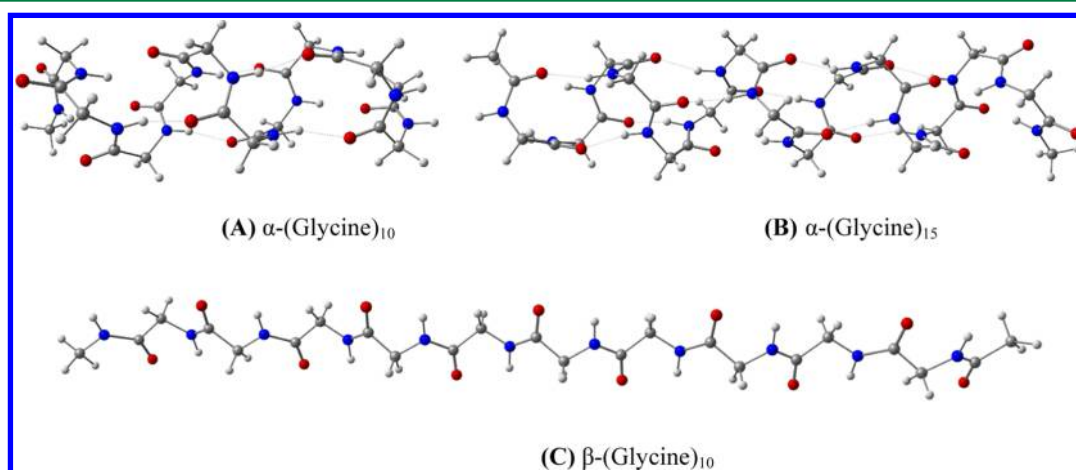

Figure 7. Optimized structures of (A) α-(Glycine)$_{10}$, (B) α-(Glycine)$_{15}$, and (C) β-(Glycine)$_{10}$ at MIM2[B3LYP/6-311G(d,p):PM6]. The β-(Glycine)$_{10}$ structure has also been computed but is not shown. Refer to the text for more details.

actual (i.e., directly evaluated) values. The contour distribution of the force constant matrix elements for α-and β-(Glycine)$_{10}$ evaluated at MIM2[B3LYP/6-311G(d,p):HF/3-21G] are depicted in Figure 8. The Hessian matrix is a real and symmetric matrix, and, hence, in Figures 8A and 8B, the upper triangular matrices correspond to the actual force constant elements and the lower triangular matrices correspond to MIM2 ones. Both α- and β-(Glycine)$_{10}$ conformers have 82 atoms, including the end-capping −CH$_3$ groups. Hence, the dimension of the symmetric Hessian matrix is 264 × 264. Each pixel in Figure 8 corresponds to an element of the full Hessian matrix. The highest valued Hessian matrix elements are the one in red and are distributed along the diagonal of this matrix. As the interatomic distance increases, most of the Hessian elements are close to zero, and these matrix elements are depicted in blue in Figure 8.

Figure 8A depicts the contour distribution of Hessian matrix elements of the α-(Glycine)$_{10}$, and it shows that long-range interactions (leading to small Hessian elements) are important in constructing meaningful force constants and frequencies. MIM2 handles this effectively, since some of the hydrogen bonding interactions are included in the first layer (depending

on the size of the primary subsystems), and the remaining long-range weak interactions are taken into account using the second layer, at a low level of theory. Hence, Figure 8A brings forth the importance of the second layer in capturing the long-range weak intramolecular interactions. An additional important point is that this second layer correction at a low level of theory enables a reduction in the size of the primary overlapping subsystems. If such a second layer is not included, it may be computationally very expensive to have large primary subsystems incorporating all these long-range weak interactions. On the other hand, Figure 8B depicts the contour Hessian distribution of the β-(Glycine)$_{10}$ conformer. For this open conformer, the diagonal matrix elements are the dominant ones, and the off-diagonal elements are small in magnitude, compared to the former ones. In these types of systems, with sufficiently large fragments, one layer of theory may suffice to give accurate structures and properties.

The details of the fragmentation scheme and the number of atoms and basis functions in each fragment for (Glycine)$_{10}$ and (Glycine)$_{15}$ molecules are depicted in Figure S9 in the Supporting Information and Table 5. The sizes of the primary overlapping fragments are the rate-limiting steps in these
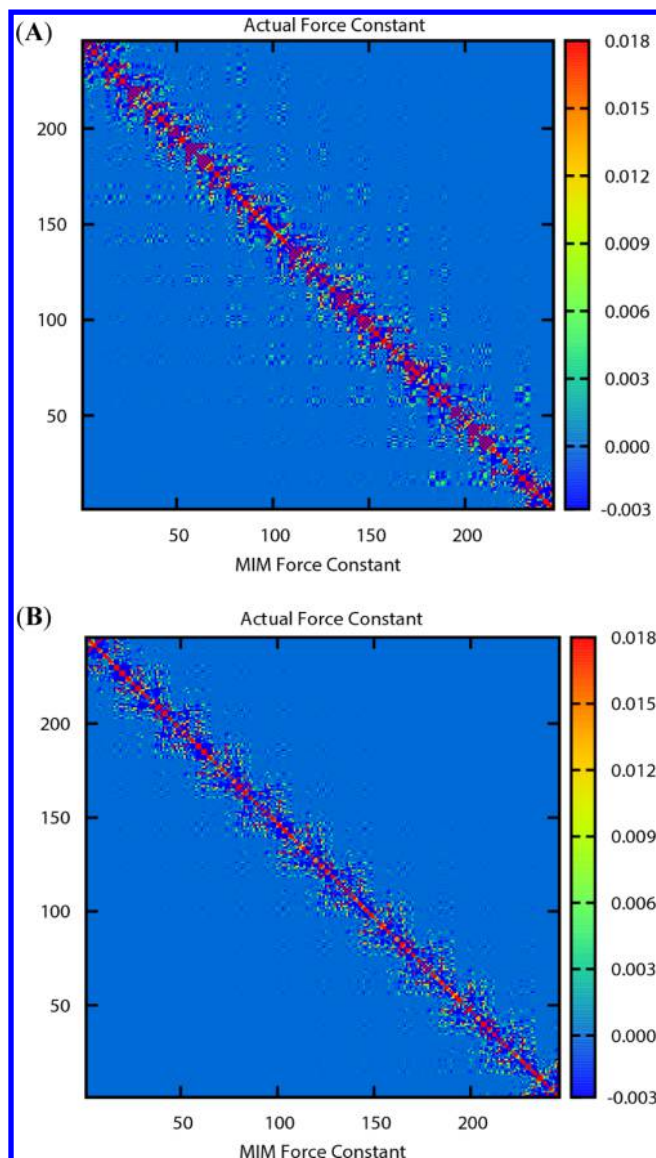
Figure 8. Heat diagram showing the distribution of the force constants of (A) α-(Glycine)$_{10}$ and (B) β-(Glycine)$_{10}$ at the MIM2[B3LYP/6-311G(d,p):HF/3-21G] and Actual[B3LYP/6-311G(d,p)] levels of theory. In panels (A) and (B), the upper triangular matrix corresponds to the actual Hessian and the lower triangular matrix corresponds to the MIM Hessian. The Hessian matrix is 246 × 246 in dimension, and each pixel corresponds to one of the elements. The atomic numbering is according to the connectivity information from one end of the molecule to the other end. Refer to the text for more details.

calculations. In both (Glycine)$_{10}$ and Glycine)$_{15}$ molecules, the number of primary fragments increases linearly with system size, and the size of the fragments is independent of the size of

the system. The distribution of the number of atoms and basis functions in each of the primary fragments are similar. Thus, the asymptotic scaling of MIM1 is linear. In MIM2, each of the fragments is also evaluated with a low level of theory. The calculations on the fragments at the low level of theory are usually very efficient and much faster than those for the high level of theory. In some cases, such as using HF as the low level and MP2 as the high level, they are obtained completely free. However, MIM2 also involves a calculation on the real system (full molecule) at the low level of theory. In most cases, this also tends to be very efficient, but will depend on the high and low methods, as well as the relative sizes of the basis sets used. Overall, the MIM method scales almost linearly with system size in such large molecules.

We have evaluated IR spectra of α- and β-(Glycine)$_n$ isomers (with $n = 10$ and 15), employing MIM2 force constants and dipole derivatives. For further determine the effects of the fragmentation size and combinations of levels of theory on the IR spectra, we have evaluated the spectra with different fragmentation schemes and levels of theory. The IR spectra of α-(Glycine)$_{10}$ isomer with tripeptide and tetrapeptide as the primary subsystems are depicted in Figure S7 in the Supporting Information and Figure 9, respectively. The first entry in Figure S7(A) in the Supporting Information and Figure 9A are the α-(Glycine)$_{10}$ spectra evaluated at an actual high
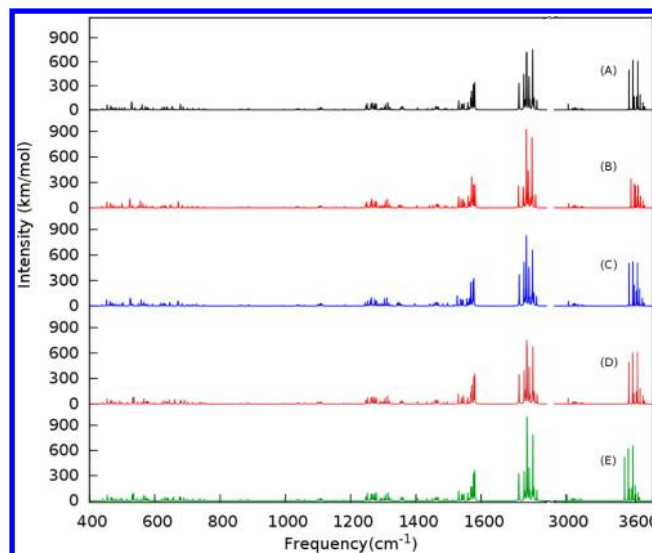

Figure 9. Vibrational IR spectra of α-(Glycine)$_{10}$ at (A) Actual-[B3LYP/6-311G(d,p)] in black, (B) MIM1[B3LYP/6-311G(d,p)] in red, (C) MIM2[B3LYP/6-311G(d,p):PM6] in blue, (D) MIM2-[B3LYP/6-311G(d,p):HF/3-21G] in brown, and (E) MIM2[B3LYP/6-311G(d,p):B3LYP/3-21G] in green, respectively. Here, the primary overlapping fragments are constructed from tetrapeptide units. Refer to the text for more details.

Table 5. System, Number of Basis Functions (N_{Basis}), B3LYP/6-311+G(d,p) Energies, the Corresponding Deviation of MIM2[B3LYP/6-311+G(d,p):PM6] Energies from the Actual Ones, Two Highest Intense Frequencies, and the Corresponding Intensities^a

system	N_{Basis}	E_{B3LYP}^{Actual} (Hartree)	ΔE_{B3LYP}^{MIM2} (kcal/mol)	ν_{B3LYP}^{MIM2} (cm^{-1})	I_{B3LYP}^{MIM2} (km/mol)	ν_{B3LYP}^{MIM2} (cm^{-1})	I_{B3LYP}^{MIM2} (km/mol)
α-(Glycine)$_{10}$	1032	−2329.33676	−3.99	1740.95	1119.83	1758.87	695.83
β-(Glycine)$_{10}$	1032	−2329.31998	−2.29	1522.65	4550.56	1734.34	1310.81
α-(Glycine)$_{15}$	1482	−3369.72589	−2.78	3531.01	2618.81	1738.46	2345.63
β-(Glycine)$_{15}$	1482	−3369.68063	−0.98	1519.91	7172.40	1523.60	1118.34

^{a}Refer to the text for more details.

B3LYP/6-311G(d,p) level of theory. Figures S7(B)−(E) in the Supporting Information and Figures 9B−E are the IR spectra evaluated at the MIM1[B3LYP/6-311G(d,p)], MIM2[B3LYP/6-311G(d,p):PM6], MIM2[B3LYP/6-311G(d,p):HF/3-21G], and MIM2[B3LYP/6-311G(d,p):B3LYP/3-21G] levels of theory, respectively. For a better comparison, in Figure S8 in the Supporting Information, the MIM vibrational spectra of α-(Glycine)$_{10}$ is superimposed on to the actual ones in the most intense spectral region, between 1200 cm^{-1} and 3800 cm^{-1}. The MAD values in the frequency spectra depicted in Figures S7(B)−(E) in the Supporting Information are 7.71, 5.89, 3.62, and 3.37 cm^{-1}, respectively. The MAD values of the corresponding intensities are 28.21, 21.57, 15.64, and 13.64 km/mol, respectively. Similarly, a fragmentation scheme based on a tetrapeptide subsystem (Figures 9B−E) has the following MAD values, relative to the corresponding frequency spectra: 1.98, 2.67, 0.98, and 0.59 cm^{-1}, respectively. The MAD values for the intensity are 16.73, 8.89, 6.91, and 7.85 km/mol, respectively. These values show that there is a 54%−82% improvement in the calculated frequency and a 40%−58% improvement in the corresponding intensities upon increasing the size of the primary subsystems from tripeptide to tetrapeptide. For comparison, the calculated spectra of α- and β-(Glycine)$_n$ (with n = 10 and 15) at the MIM2[B3LYP/6-311G(d,p):PM6] level are shown in Figure S10 in the Supporting Information. The highest intense peaks in α-(Glycine)$_{10}$ and α-(Glycine)$_{15}$ correspond to C=O stretching modes in the amide functional group, while the intense peak in β-(Glycine)$_{10}$ and β-(Glycine)$_{15}$ corresponds to the N−H wagging and C=O stretching modes of amide functional groups.

To further determine the effects of fragment-based methods on the IR spectra, we have evaluated the corresponding zero-point energies (ZPEs) of these conformers at the above-mentioned levels of theory. The errors in ZPE of α-(Glycine)$_{10}$ isomers at (1) MIM1[B3LYP/6-311G(d,p)], (2) MIM2-[B3LYP/6-311G(d,p):PM6], (3) MIM2[B3LYP/6-311G-(d,p):HF/3-21G] and (4) MIM2[B3LYP/6-311G-(d,p):B3LYP/3-21G] levels of theory, when compared to the actual high value at B3LYP/6-311G(d,p), are depicted in Figure 10. The blue bar corresponds to the ZPE error with the
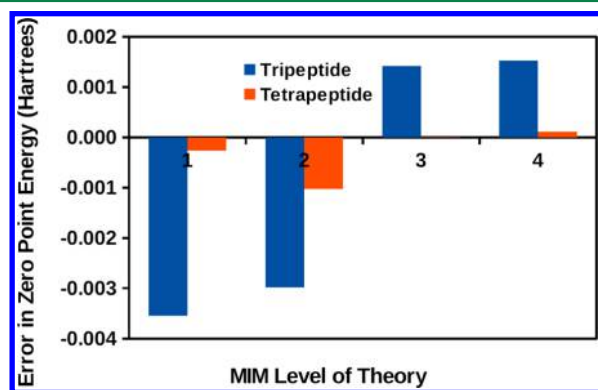

Figure 10. Error in the zero point energy (ZPE) of α-(Glycine)$_{10}$ evaluated at (1) MIM1[B3LYP/6-311G(d,p)], (2) MIM2[B3LYP/6-311G(d,p):PM6], (3) MIM2[B3LYP/6-311G(d,p):HF/3-21G], and (4) MIM2[B3LYP/6-311G(d,p):B3LYP/3-21G] levels of theory, respectively. The blue bars correspond to the ZPE values evaluated with tripeptide primary subsystem, and the red bars correspond to the ZPE values evaluated with tetrapeptide primary subsystem. Refer to the text for more details.

tripeptide as the primary fragment and red bar to that with tetrapeptide as the primary fragment. Upon increasing the size of the primary fragments from tripeptide to tetrapeptide for the α-(Glycine)$_{10}$ isomers, there is a 66%−93% improvement in ZPE accuracy. Some of the residual errors may be alleviated further by using scaled force constants to take into account any known imbalance between the high and low levels of theory. Nevertheless, the errors in the calculated ZPEs are quite small, using the efficient, and inexpensive, 3-21G basis set for the second layer. This improvement is particularly impressive considering that the total ZPE at the high level is more than 350 kcal/mol. This improvement suggests that the MIM model will be a promising tool for predicting the accurate thermochemistry of large molecules.

Overall, IR frequencies at a single level of theory (MIM1) in many cases do not adequately capture the effects of long-range weak interactions. Tables 1, 2, and 3 show ∼86% improvement in energy accuracy and ∼75% improvement in the vibrational spectra at MIM2, compared to MIM1. In addition, from our present work on these test systems, it is very clear that the position of the link atom in the model system affects the energetics and vibrational spectra, and link-atom forces and the Hessian have to be projected back onto the supporting and host atoms correctly to get accurately extrapolated spectra that are consistent with the actual ones. Thus, incorporation of the Jacobian link-atom projection method and the employment of multiple layers of theory in accounting for long-range weak interactions are both paramount to get adequate performance. Overall, the MIM hybrid extrapolation method is a valuable asset for evaluating accurate energies, geometries, and vibrational spectra of large molecules.

4. CONCLUSIONS

In the current work, we have extended the molecules-in-molecules (MIM) fragment-based model for evaluating the structures, energies, atomic forces, and infrared (IR) vibrational spectra of large molecules. The incorporation of the Jacobian link-atom projection method, as well as the employment of multiple layers of theory in accounting for long-range weak interactions, are both imperative factors that contribute to the accuracy of the method. A careful assessment of the calculated vibrational frequencies and IR intensities is carried out for a range of neutral and charged peptides. Employing the MIM2 method, we have optimized the α- and β-(Glycine)$_n$ conformers, with n = 10 and 15, and evaluated the corresponding IR spectra. Finally, our implementation of the algorithm is independent of the combination of levels of theory and fragmentation schemes, and the method is easily parallelizable, without any severe modification to the algorithm.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*kraghava@indiana.edu.

Notes

■ REFERENCES

(1) Kohn, W. Density Functional and Density Matrix Method Scaling Linearly with the Number of Atoms. *Phys. Rev. Lett.* **1996**, *76*, 3168–3171.

(2) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.* **2011**, *112*, 632–672.

(3) Morita, S.; Sakai, S. IMiCMO: A new integrated ab initio multicenter molecular orbitals method for molecular dynamics calculations in solvent cluster systems. *J. Comput. Chem.* **2001**, *22*, 1107–1112.

(4) Sakai, S.; Morita, S. Ab Initio Integrated Multi-Center Molecular Orbitals Method for Large Cluster Systems: Total Energy and Normal Vibration. *J. Phys. Chem. A* **2005**, *109*, 8424–8429.

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

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

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

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

(9) Jovan Jose, K. V.; Gadre, S. R. Ab initio study on $(CO_2)_n$ clusters via electrostatics- and molecular tailoring-based algorithm. *Int. J. Quantum Chem.* **2009**, *109*, 2238–2247.

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

(11) Sahu, N.; Gadre, S. R. Molecular Tailoring Approach: A Route for ab Initio Treatment of Large Clusters. *Acc. Chem. Res.* **2014**, *47*, 2739–2747.

(12) Yeole, S. D.; Sahu, N.; Gadre, S. R. Structures, energetics and vibrational spectra of CO_2 clusters through molecular tailoring and cluster building algorithm. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7718–7723.

(13) Sahu, N.; Gadre, S. R. Accurate vibrational spectra via molecular tailoring approach: A case study of water clusters at MP2 level. *J. Chem. Phys.* **2015**, *142*, 014107.

(14) Li, W.; Li, S.; Jiang, Y. Generalized Energy-Based Fragmentation Approach for Computing the Ground-State Energies and Properties of Large Molecules. *J. Phys. Chem. A* **2007**, *111*, 2193–2199.

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

(16) Nakata, H.; Nagata, T.; Fedorov, D. G.; Yokojima, S.; Kitaura, K.; Nakamura, S. Analytic second derivatives of the energy in the fragment molecular orbital method. *J. Chem. Phys.* **2013**, *138*, 164103.

(17) Nakata, H.; Fedorov, D. G.; Yokojima, S.; Kitaura, K.; Nakamura, S. Simulations of Raman Spectra Using the Fragment Molecular Orbital Method. *J. Chem. Theory Comput.* **2014**, *10*, 3689–3698.

(18) Nakata, H.; Fedorov, D. G.; Yokojima, S.; Kitaura, K.; Nakamura, S. Efficient vibrational analysis for unrestricted Hartree–Fock based on the fragment molecular orbital method. *Chem. Phys. Lett.* **2014**, *603*, 67–74.

(19) Vreven, T.; Morokuma, K. In *Annual Reports in Computational Chemistry*, Vol. 2; David, C. S., Ed.; Elsevier: Amsterdam, 2006; pp 35–51.

(20) Chung, L. W.; Hirao, H.; Li, X.; Morokuma, K. The ONIOM method: Its foundation and applications to metalloenzymes and photobiology. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 327–350.

(21) Řezáč, J.; Salahub, D. R. Multilevel Fragment-Based Approach (MFBA): A Novel Hybrid Computational Method for the Study of Large Molecules. *J. Chem. Theory Comput.* **2009**, *6*, 91–99.

(22) Vreven, T.; Byun, K. S.; Komáromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J. Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM. *J. Chem. Theory Comput.* **2006**, *2*, 815–826.

(23) Fedorov, D. G.; Ishida, T.; Kitaura, K. Multilayer Formulation of the Fragment Molecular Orbital Method (FMO). *J. Phys. Chem. A* **2005**, *109*, 2638–2646.

(24) Isegawa, M.; Wang, B.; Truhlar, D. G. Electrostatically Embedded Molecular Tailoring Approach and Validation for Peptides. *J. Chem. Theory Comput.* **2013**, *9*, 1381–1393.

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

(26) He, X.; Merz, K. M. Divide and Conquer Hartree–Fock Calculations on Proteins. *J. Chem. Theory Comput.* **2010**, *6*, 405–411.

(27) Nagata, T.; Fedorov, D. G.; Sawada, T.; Kitaura, K.; Gordon, M. S. A combined effective fragment potential–fragment molecular orbital method. II. Analytic gradient and application to the geometry optimization of solvated tetraglycine and chignolin. *J. Chem. Phys.* **2011**, *134*, 034110.

(28) Mullin, J. M.; Roskop, L. B.; Pruitt, S. R.; Collins, M. A.; Gordon, M. S. Systematic Fragmentation Method and the Effective Fragment Potential: An Efficient Method for Capturing Molecular Energies. *J. Phys. Chem. A* **2009**, *113*, 10040–10049.

(29) Guo, W.; Wu, A.; Zhang, I. Y.; Xu, X. XO: An extended ONIOM method for accurate and efficient modeling of large systems. *J. Comput. Chem.* **2012**, *33*, 2142–2160.

(30) Mayhall, N. J.; Raghavachari, K. Molecules-in-Molecules: An Extrapolated Fragment-Based Approach for Accurate Calculations on Large Molecules and Materials. *J. Chem. Theory Comput.* **2011**, *7*, 1336–1343.

(31) 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, 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.

(32) Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. *J. Mol. Struct.: THEOCHEM* **1999**, *461–462*, 1–21.