JCTC Journal of Chemical Theory and Computation

Article

Cite This: J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

pubs.acs.org/JCTC

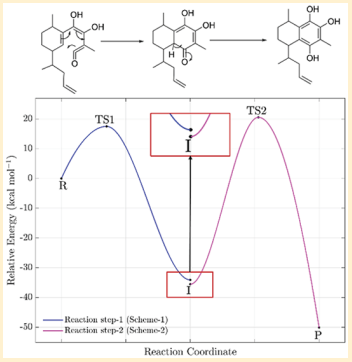
Exploring Reaction Energy Profiles Using the Molecules-in-Molecules Fragmentation-Based Approach

Ankur Kumar Gupta, Bishnu Thapa, and Krishnan Raghavachari*

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

Supporting Information

ABSTRACT: The Molecules-in-Molecules (MIM) fragmentation-based approach has been successfully used in previous studies to obtain the energies, optimized geometries, and spectroscopic properties of large molecular systems. The present work delineates a protocol to study the potential energy profiles for multistep chemical reactions using the MIM methodology. In a complex multistep chemical reaction, the fragmentation scheme needs to be changed as the reacting species transition into a new reaction step, resulting in a discontinuity in the potential energy curve of the reaction. In our approach, the fragmentation scheme for a particular step in a reaction is chosen on the basis of the nature of the bonding changes associated with that step. Thus, the reactant, transition state, and product are treated consistently throughout the reaction step, leading to an accurate energy barrier for that step. The discontinuity now occurs in describing the energies of reaction intermediates at the transition point between two reaction steps that are treated by two different fragmentation schemes. To address this issue, we propose a systematic procedure for obtaining continuous potential energy curves that are least shifted from their initial positions. The corrected MIM potential energy curves are continuous with activation energies preserved. Following this approach, energy profiles of complex reactions involving large molecular species can be obtained at high levels of theory with a reasonable computational cost.



1. INTRODUCTION

Methods based on quantum mechanics can provide an accurate description of chemical systems, but at a high computational cost. The steep computational scaling of popular quantum chemical methods such as MP2, CCSD, or CCSD(T) with system size ($O(N^5)$−$O(N^7)$) often prevents their use in many problems of practical interest involving large molecules (viz., protein-ligand interactions, surface catalysis, etc.). With the advent of hybrid models such as QM/MM[1] and ONIOM,[2] it has become possible to study chemically active regions of a large molecular system (for example, active site of an enzyme) with a high level of theory. Nevertheless, broadly applicable methods, which can treat the whole molecule at the same high level of theory at a reduced computational cost, are desirable. On the basis of the reasonable assumption that chemical interactions are mostly local, fragmentation methods have been developed, which involve independent calculations on subsystems of a molecule at some high level of theory.[3−9] The quantities obtained from such individual calculations are then appropriately summed to obtain the property of the whole system. Because high level calculations are being done only on molecular subsystems, these methods asymptotically scale linearly with system size. The Molecules-in-Molecules (MIM) fragmentation method[10] developed by our group is one such method. MIM's usefulness has been demonstrated for calculating accurate total energies, optimized geometries, and spectroscopic (IR, Raman, VCD, ROA, NMR) properties of a variety of large molecules.[11−16] The MIM methodology has also been used to study a variety of chemical aspects of large molecular systems such as protein−ligand binding energies,[17] supramolecular interactions in foldamers,[18] etc. In this context, the present work extends the use of MIM to study multistep chemical reactions and their corresponding energy profiles.

To obtain a potential energy curve (PEC) of a reaction, the energy is plotted as a function of the corresponding reaction coordinate. In a multistep chemical reaction involving several transition states and intermediates, as the reaction proceeds, the bonding configuration of the molecule often changes, necessitating a change in the fragmentation scheme of the molecule. This change in fragmentation scheme during the course of a reaction results in a discontinuity in the energy profile. Such discontinuities are very prevalent in MD trajectories obtained from QM/MM and ONIOM calculations, especially in cases where the boundary between QM and MM region evolves, for example, solvated ions.[19,20] In such an example, the discontinuity occurs whenever there is an exchange of atoms between the QM and MM regions during the simulation. This change in the level of theory for the exchanged atom appears in the form of sudden shifts in the potential energy and forces, potentially leading to unstable or error-prone MD simulations. Several models have been proposed to solve this problem of discontinuity in the time evolution of such systems in the context of molecular dynamics.[19−30] Popular schemes include the ONIOM-XS

Received: February 14, 2019

DOI: 10.1021/acs.jctc.9b00152
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

method[20] and the Adaptive-Partitioning QM/MM method.[21] Both of these methods involve incorporating a buffer region between the QM and MM spheres, which then allows the use of smoothing functions to remove the discontinuities in the energy and forces when atoms enter or leave the buffer zone.

This Article deals with the calculation of potential energy curves for multistep chemical reactions involving bond rearrangements (bond formation, bond breaking) in each step. We are most interested in accurate calculations of the energies of reaction intermediates and the corresponding activation barriers leading to their formation. Unlike MD trajectories, reaction coordinate diagrams in quantum chemistry are drawn using only the stationary points of the PEC, which means that the fragmentation schemes could be predefined, and not necessarily be determined on the fly. Consequently, the region (or point) of discontinuity in the energy profile could be foreseen, which is generally not the case in dynamics studies. While many different fragmentation approaches exist in the literature, and have been used extensively to study the energy profile of chemical reactions,[31−37] their main focus has been the accurate calculation of the energy barriers so as to decipher the lowest energy pathway of a complex chemical reaction. In the present work, we focus on a problem that has received less attention, the discontinuities in the potential energy curve associated with the change in the fragmentation scheme during the course of a multistep reaction. The discontinuous energy curves obtained from MIM calculations are made continuous by appropriately shifting them relative to each other. While the ideas are demonstrated using DFT methods for relatively small systems for calibration purposes, they should be broadly applicable, and future work will be aimed at applications for systems that are too large to study using accurate correlated electronic structure methods.

2. METHOD

2.1. MIM Methodology. Molecules-in-Molecules (MIM)[10] is a multilayer fragmentation-based method, which can be used to study large molecules uniformly at a high level of theory with asymptotically linear scaling. To capture interfragment and long-range interactions in the full molecule, multiple layers with different fragmentation schemes and levels of theory can be combined in MIM. Briefly, the molecule is first divided into nonoverlapping fragments by cutting nonpolar single bonds (e.g., C−C bond). Each fragment interacts with neighboring fragments according to a well-defined criteria (connectivity-based or distance-based) to yield primary subsystems.[10] In general, the subsystems are overlapping in nature, and overcounting of the overlapping regions is taken into account via the *inclusion−exclusion* principle. The summation of the energies of the subsystems at some high level of theory (after capping the dangling bonds with hydrogen atoms[38]) then yields the energy of the molecule, E_r^{high} (where r is a generic parameter that defines the fragmentation scheme), and is termed as a 1-layer model (MIM1) within the MIM formalism. Thus:

$$E_{\text{MIM1}} = E_r^{\text{high}} \tag{1}$$

The missing long-range interactions in MIM1 can be approximated by using a low level of theory to yield a two-layer model (MIM2), as follows:[10]

$$E_{\text{MIM2}} = E_r^{\text{high}} - (E_r^{\text{low}} - E_\infty^{\text{low}}) \tag{2}$$

Here, E_{low}^{∞} represents a calculation on the full molecule at a low level of theory. Similarly, an n-layer model can be developed with different types of interfragment and long-range interactions.[15]

2.2. MIM for Chemical Reactions. A reaction coordinate diagram (or energy profile) of a chemical reaction is a qualitative representation of the minimum energy path traversed by the reacting species on the potential energy surface (PES). Typically, optimization calculations are done on the structures that represent stationary points on the potential energy curve (PEC), and the energies of the optimized structures are then plotted with respect to an arbitrary reaction coordinate to represent the energy profile of the reaction. In addition, intrinsic reaction coordinate (IRC)[39−41] calculations are often done to confirm the nature of the reacting species associated with the individual transition states.

If the molecules involved in the reaction are large enough, it becomes difficult to study them using conventional quantum chemistry methods. Hence, MIM fragmentation method can be adapted to obtain optimized energies of all of the stable and transient chemical species involved in the reaction. The first step in any MIM calculation is the selection of a fragmentation scheme, based on which its energy will be calculated. While drawing a potential energy curve (PEC) for a chemical reaction, our approach is to keep the fragmentation scheme unchanged during each reaction step, so that the energy barriers could be determined without any errors from having different fragmentation schemes. However, the task of selecting a single fragmentation scheme for the transition state, and the adjoining reactants and products, is nontrivial because the bond order between multiple atom pairs may change as the reaction proceeds. As mentioned earlier, only single nonpolar bonds are allowed to break to form fragments. However, in the case of a chemical reaction, a bond that is single in the reactant may not remain so in the transition state or product. Therefore, for the fragmentation scheme to remain consistent throughout the reaction step, the scheme is chosen on the basis of the structure of the transition state (e.g., from chemical intuition, or, in more complex cases, derived from a computationally efficient low level of theory). The reason behind this choice is that the structure and connectivity of a transition state clearly indicate the bonds that are going through a bond order transition, which helps in selecting a suitable fragmentation scheme for the reaction step. In a given step of a reaction, the change in chemical structure is localized to a small part of the molecular system, which includes changes in bond order, transfer of an atom (or a group), etc. Because the bond orders in that local region are uncertain, that part needs to be confined in a single fragment, whereas the rest of the molecule can be fragmented as usual, that is, by breaking nonpolar single bonds. Most of the reactions of practical interest involve multiple steps having several intermediates and transition states. In such cases, each step's fragmentation scheme needs to be chosen on the basis of the corresponding transition state. However, in such a model, an intermediate (which represents a common point for any two consecutive reaction steps in a coordinate diagram) must be calculated using two different fragmentation schemes, which leads to a discontinuity in the potential energy curve of the

DOI: 10.1021/acs.jctc.9b00152
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

reaction. The present work attempts to find solutions to fix this discontinuity in the energy profile of reactions.

The way to study a chemical reaction using MIM, and the resultant discontinuity that arises in the energy profile, can best be understood with the help of an example. To this end, a simple intramolecular hydrogen transfer reaction (keto−enol tautomerism) was selected for illustration purposes. Figure 1

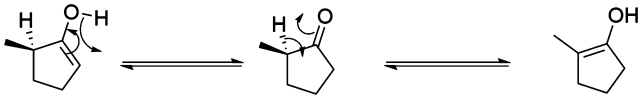


Figure 1. Methyl cyclopentanone−cyclopentenol interconversion through intramolecular hydrogen transfer.

shows an interconversion between two enols (differing by a methyl substitution), via a keto intermediate. It is a two-step reaction involving two transition states and an intermediate. To describe the associated fragmentation schemes, a 3D representation of the reaction is shown in Figure 2.

In the first step of the reaction, the hydrogen of the hydroxyl group transfers to the α-carbon on the right, which results in the formation of a ketone from an enol. The net effect of this reaction is the change in bond order of the carbon−oxygen bond from 1 to 2, whereas that between the right α-carbon and carbonyl-carbon changes from 2 to 1. This region where the change in bond orders takes place (see transition state-1 in Figure 2) is confined in a single fragment, represented by the blue shaded region in Figure 3a. The rest of the molecule can be fragmented as usual, that is, based on whether the bond being broken is a single nonpolar bond or not. Because of the small size of the molecule used in this example, a number-based scheme is used to form subsystems; that is, the primary subsystems are formed by combining two adjacent non-overlapping fragments. This particular fragmentation scheme is then used to optimize the structures associated with the first step of the reaction, reactant (R), transition state-1 (TS1), and intermediate (I) (refer Figure 2) using MIM2 [M06-2X[42]/6-31+G(d,p):M06-2X/6-31G], where the colon separates the levels of theory used in the high and low layers. An IRC calculation is also done to ensure that TS1 connects to the specified reactant and product. In the second step, the hydrogen atom on the left α-carbon transfers to the oxygen atom resulting in a keto to enol isomeric conversion. As described for the first reaction step, the fragmentation scheme for the second step was chosen on the basis of the corresponding transition state (transition state-2 in Figure 2), which is depicted in Figure 3b.

Notice that Intermediate (I) must be computed (optimized) twice, because it is involved in both reaction steps. After obtaining the optimized energies of all of the reacting species using their corresponding fragmentation schemes, a plot of the reaction energy profile is obtained, as shown in Figure 4.

Because the intermediate's energy was computed using two different fragmentation schemes, there are two points corresponding to it on the energy profile, which is a form of discontinuity in the energy curve. The energy difference between the two schemes at the intermediate's geometry is quite small, 0.4 kcal mol^{-1}, not surprising because the two steps differ only by a methyl substitution. To determine the accuracy of the MIM method with respect to the full molecule calculation, the PEC of the reaction obtained using M06-2X/6-31+G(d,p) (which is the MIM2 high level theory) was drawn together with the curves obtained with MIM2 (Figure 5).

A generalized procedure of shifting the curves to remove the discontinuities can be formally translated into a set of mathematical equations for any given network of chemical reactions. For every two-step process (say ith and jth steps) in a reaction network involving a common intermediate (I_k), the following equality needs to hold for the PEC to be continuous:

$$x_i + E_i^{I_k} = x_j + E_j^{I_k} \Rightarrow x_i - x_j = E_j^{I_k} - E_i^{I_k} \quad (3)$$

where x_i and x_j are unknown variables representing the shifts in the ith and jth reaction steps' curves, and $E_i^{I_k}$ and $E_j^{I_k}$ represent the energies of intermediate I_k obtained using the ith and jth steps' fragmentation scheme, respectively. The RHS of eq 3 is nothing but the energy difference (or discontinuity) for the involved intermediate (I_k) between the two fragmentation schemes (i and j). Equation 3 is just a mathematical representation of the condition that the intermediate energies obtained from the conflicting fragmentation schemes need to coincide with each other after shifting their respective curves. To correct all discontinuities in the

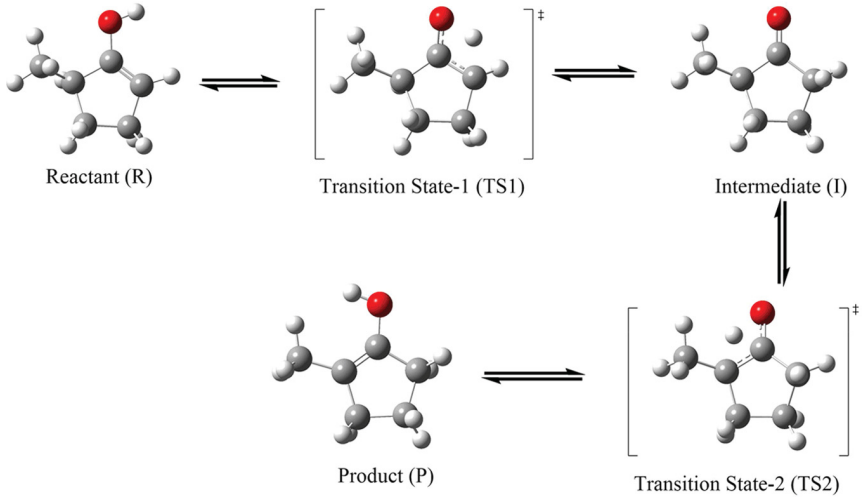


Figure 2. 3D representation of methyl cyclopentanone−cyclopentenol interconversion through intramolecular hydrogen transfer.

DOI: 10.1021/acs.jctc.9b00152
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

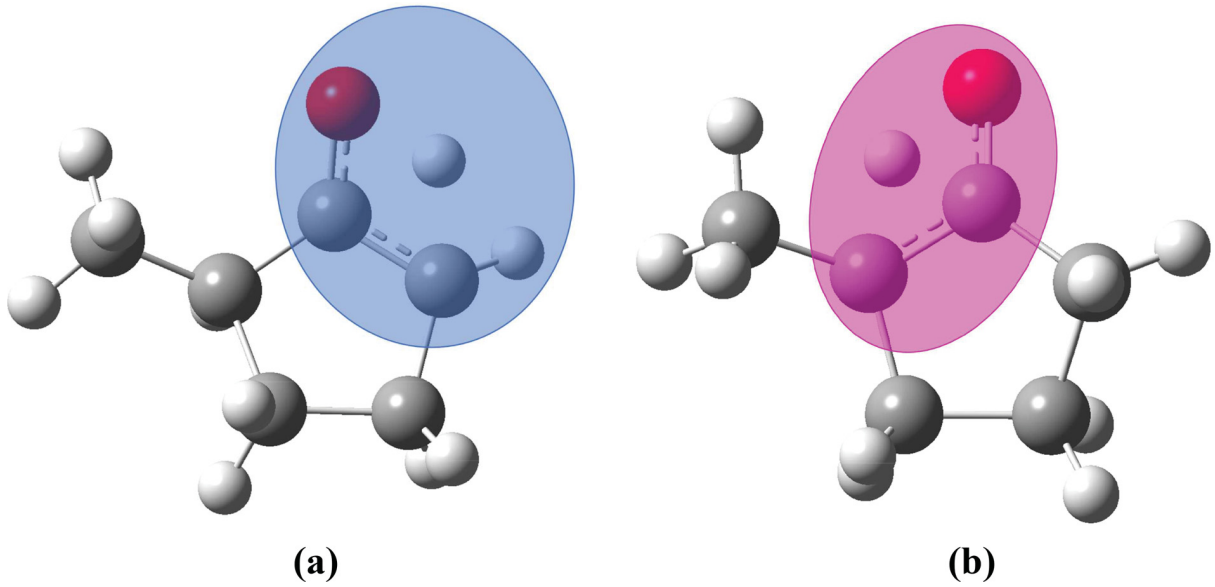


Figure 3. Fragmentation schemes for (a) transition state-1 (TS1) and (b) transition state-2 (TS2).

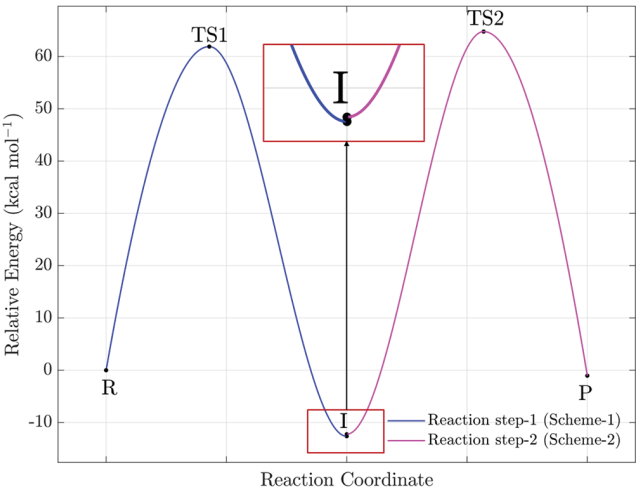


Figure 4. Energy profile of methyl cyclopentanone−cyclopentenol interconversion obtained using MIM2 method.

PES of a given chemical reaction network, eq 3 for every two-step process in the reaction network needs to be solved simultaneously to obtain the set of x_i's (shifts) corresponding to each reaction step, which makes the whole PES continuous. However, because eq 3 represents a system of linear equations, its solution depends on the number of equations relative to the number of unknown variables (x_i's) present in the system, which further depends on the type of reaction network under study.

The keto−enol tautomerism (Figure 2) being a simple two-step linear reaction is a good starting point to demonstrate the working of the mathematical procedure outlined above. Because this reaction has only two steps, and hence one intermediate, only one mathematical condition can be written (using eq 3) to solve for the discontinuity at the intermediate:

$$x_1 - x_2 = E_2^{\mathrm{I}} - E_1^{\mathrm{I}} = 0.4 \text{ kcal mol}^{-1} \tag{4}$$

or in the matrix form as

$$\mathbf{A}\boldsymbol{x} = \boldsymbol{b} \tag{5}$$

where

$$\mathbf{A} = (1 \quad -1),\ \boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},\ \text{and}\ \boldsymbol{b} = E_2^{\mathrm{I}} - E_1^{\mathrm{I}} = 0.4 \tag{6}$$

Because there are two unknown variables (x_1,x_2) and only one equation to solve, eq 4 is underdetermined. In this work, we obtain a physically motivated solution corresponding to the least norm ($\|\boldsymbol{x}\|_2$). The minimum norm ("least shifted") solution can be found by calculating the pseudoinverse[43] (Moore−Penrose inverse) of matrix $\mathbf{A}$ (usually using singular value decomposition (SVD)), and is formally written as follows:

$$\boldsymbol{x} = \mathbf{A}^{+}\boldsymbol{b} \tag{7}$$

where $\mathbf{A}^{+}$ represents the pseudoinverse of $\mathbf{A}$. Solving for x using eq 7 for the given values of $\mathbf{A}$ and $\boldsymbol{b}$ in eq 6 gives $x_1 = 0.2$ and $x_2 = -0.2$, which physically means that the first and second reaction steps' curves need to be shifted by 0.2 kcal mol^{-1} in the upward (positive) and downward (negative) directions, respectively, for the PEC to be continuous. In this case, the two reaction steps' curves need to be shifted equally toward each other to obtain a least shifted continuous PEC for the whole reaction, consistent with the simple intuition for a two-step reaction. The resultant shifted curve is shown in Figure 5b. Moreover, because the whole curve has been shifted by the same amount, this fixes the discontinuity in the energy profile while keeping the relative energies between the involved chemical species intact. As discussed in section 4, the same protocol can be used to obtain a continuous PEC for any linear multistep reaction network. Equation 3 holds for multistep cyclic and branched reaction networks as well, and is discussed in detail in section 4.3.

In addition to energies, the forces would also be discontinuous in the case of a change in the fragmentation scheme while dynamically traversing the PES, potentially leading to faulty MD simulations. In the stationary picture, however, only the optimized energies of the intermediate are needed with respect to the two conflicting fragmentation schemes. The forces on the optimized geometry are zero, only with respect to the fragmentation scheme within which it is

DOI: 10.1021/acs.jctc.9b00152
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

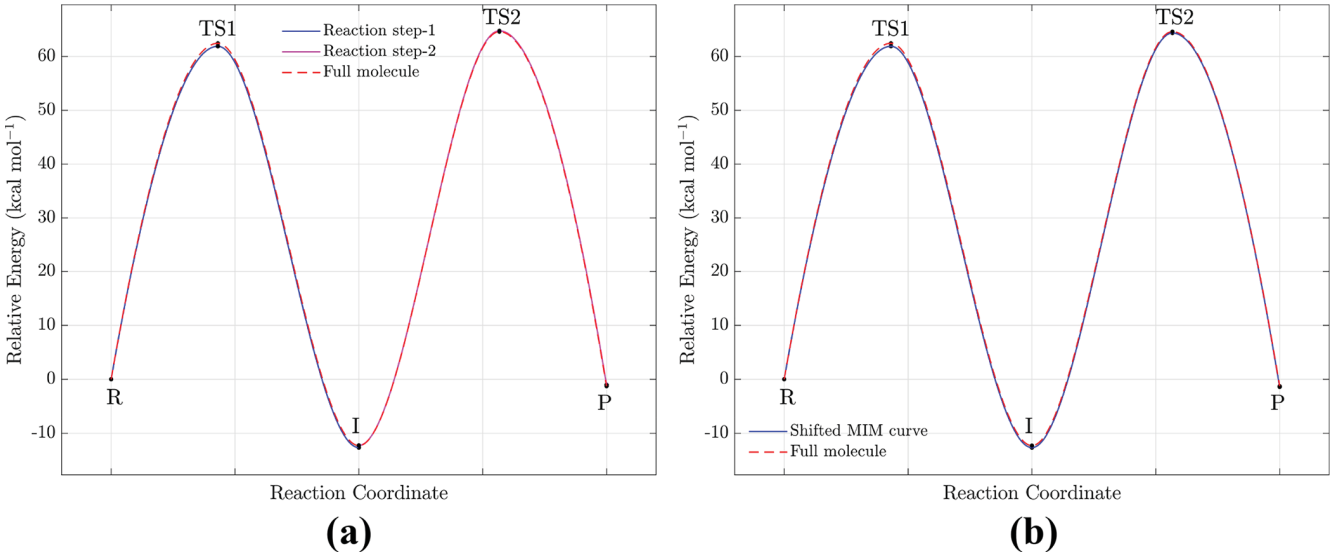


Figure 5. Energy profile of methyl cyclopentanone−cyclopentenol interconversion: (a) as obtained from MIM2 calculation, and (b) obtained after shifting the MIM energy curves.

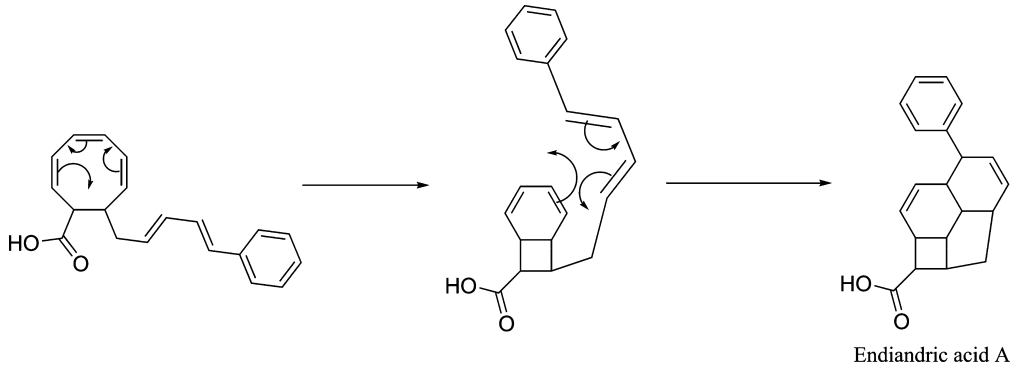


Figure 6. Reaction for the synthesis of endiandric acid A.

optimized. If the fragmentation scheme changes during the optimization procedure itself, then the discontinuous forces may result in wandering behavior of the optimization steps, which in turn may cause the procedure to take a large number of steps to complete (or to not even able to find the minimum). However, unlike MD simulations, fragmentation scheme during an optimization procedure can be kept constant, thus circumventing the need to make the forces continuous.

3. COMPUTATIONAL DETAILS

To study discontinuities in reaction coordinate diagrams, pericyclic reactions were chosen because of their concerted mechanism,[44,45] which makes them ideal to study using the MIM fragmentation method. The energy profiles for all reactions were calculated using MIM2. In particular, all molecular structures associated with the stationary points for each reaction were optimized using MIM2, with M06-2X[42]/6-311++G(3df,2p) and M06-2X/6-31+G being the high and low levels of theory, respectively. For illustrative purposes, the density functional was kept the same at both layers to minimize any error due to the difference in quality of the levels of theory. Formally, there is no restriction in using two different levels of theory although compatibility should be taken into account (as in ONIOM[2]). A number-based scheme was used to form primary subsystems by combining two adjacent fragments for the high layer calculation. The energy curves were also obtained without fragmenting the molecule at the M06-2X/6-311++G(3df,2p) level of theory to verify the accuracy of the MIM2 method. IRC calculations were carried out for every transition state to confirm the structure of its associated reactant and product. All computations were performed using the *Gaussian 16* program package[46] with the MIM method being implemented through an external Perl script.

4. RESULTS AND DISCUSSION

4.1. Synthesis of Endiandric Acid A. The reactions involved in the synthesis of endiandric acid A[47] were investigated using MIM2. Endiandric acid A is formed through a cascade of pericyclic reactions as shown in Figure 6. A 3D representation of this reaction, showing optimized geometries of all of the involved chemical species, is portrayed in Figure 7. In the first step of the reaction, the eight-membered ring electrocyclizes (6π) to yield a six-membered and a four-membered ring. This cyclooctene derivative was confined in a single fragment in the fragmentation scheme of step-1 (Figure 8a). The product obtained in the first step (which is the intermediate for the whole reaction) undergoes an intramolecular Diels−Alder cycloaddition resulting in the requisite

DOI: 10.1021/acs.jctc.9b00152
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

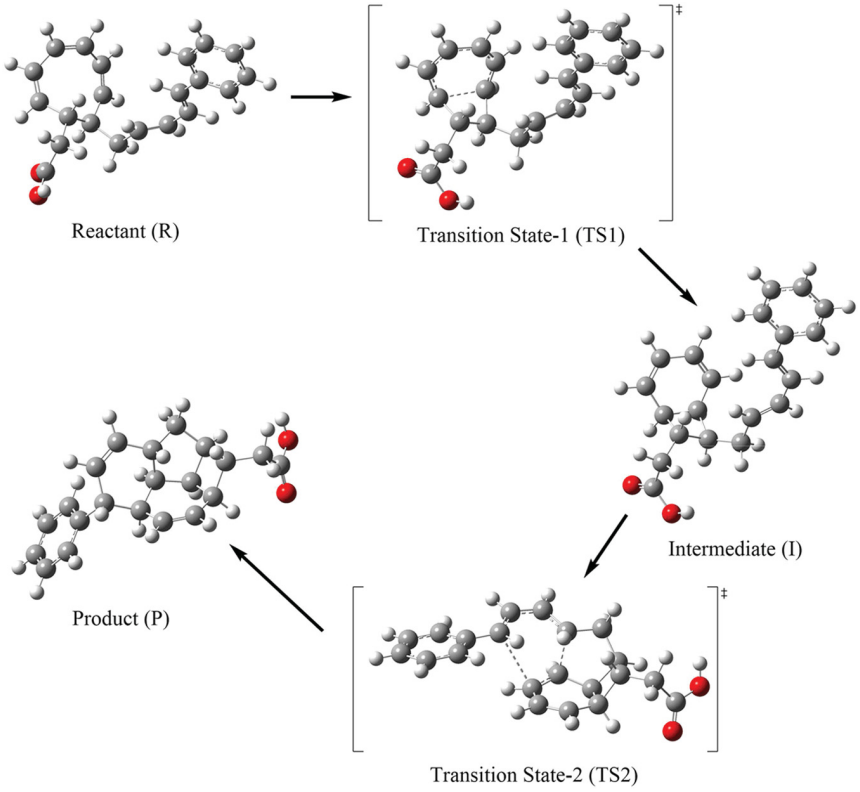


Figure 7. 3D representation of the synthesis of endiandric acid A.

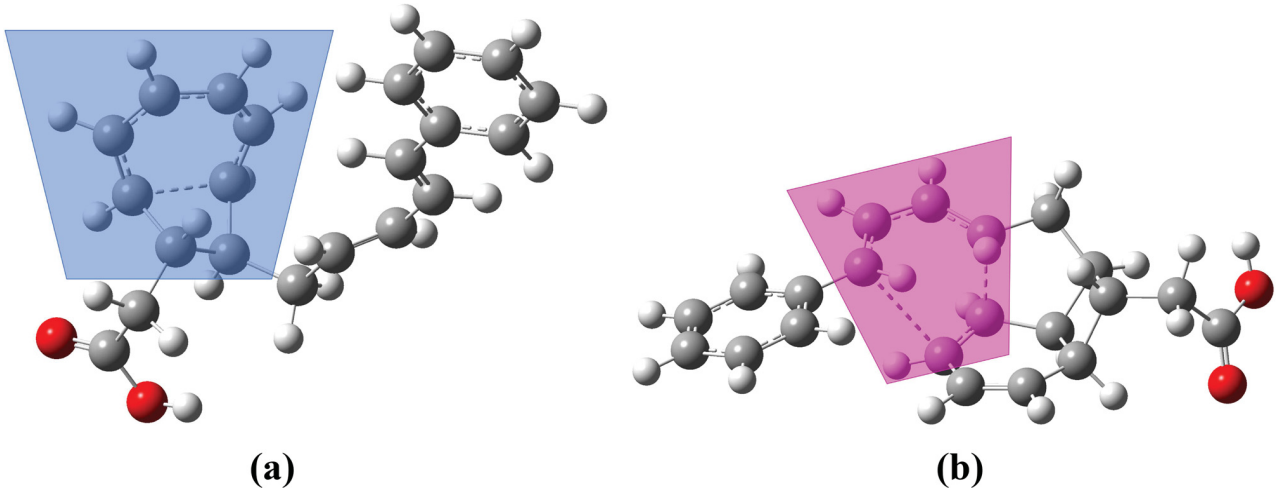


Figure 8. Fragmentation schemes for (a) transition state-1 (TS1) and (b) transition state-2 (TS2).

product, endiandric acid A. The associated diene and dienophile were enclosed in a single fragment (Figure 8b) to obtain the fragmentation scheme for the second step of the reaction.

The energy profile of the individual reaction steps obtained using their respective fragmentation schemes is shown in Figure 9a. The actual PEC for the reaction, obtained without fragmentation, is also depicted in this figure. Although the total energy difference between the MIM curve and the actual curve is around 1 kcal mol^{-1}, the curves run parallel along the reaction coordinate, indicating very small differences in the relative energies (e.g., activation energies) between the two curves, which are the target quantities of interest in any chemical reaction. The energy difference between the two fragmentation schemes at the point of discontinuity is 0.3 kcal mol^{-1}, which is quite small (as compared to the reaction barriers) due to the comparable size of the fragments in the two fragmentation schemes (Figure 8). The curves obtained from the MIM calculations are shifted equally toward each other to obtain a continuous MIM energy curve (Figure 9b). More specifically, the following equation can be written (using eq 3) for this reaction:

$$x_1 - x_2 = E_2^{\mathrm{I}} - E_1^{\mathrm{I}} = -0.3 \text{ kcal mol}^{-1} \quad (8)$$

Following the same procedure as described for the keto−enol reaction in section 2, we obtain $x_1 = -0.15$ and $x_2 = 0.15$,

DOI: 10.1021/acs.jctc.9b00152
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

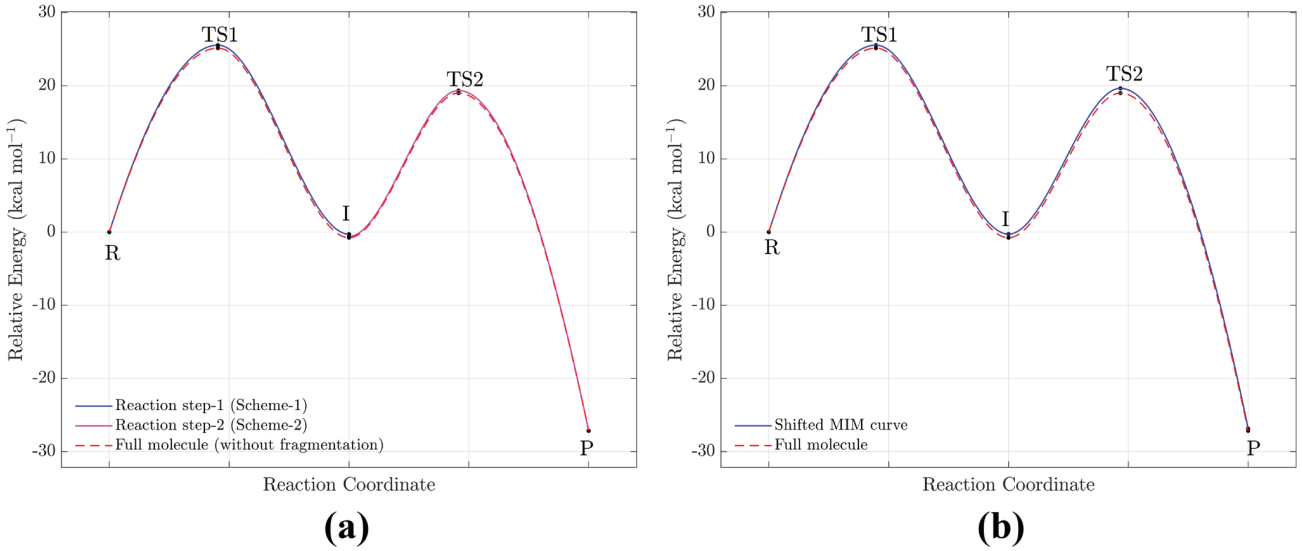


Figure 9. Energy profile for the synthesis of endiandric acid A (a) as obtained from MIM2 calculation and (b) obtained after shifting the MIM energy curves.

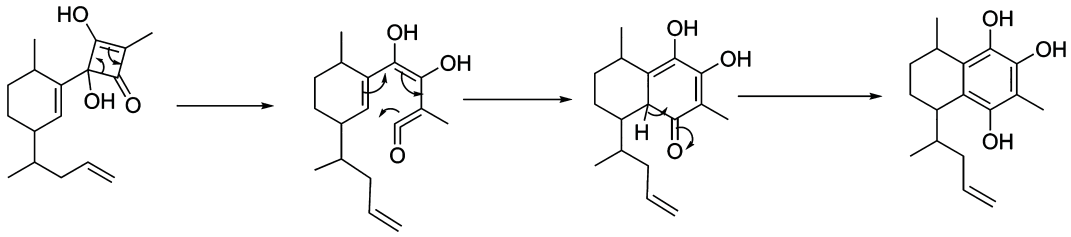


Figure 10. Pericyclic reactions during the synthesis of Colombiasin.

meaning PECs for reaction steps 1 and 2 need to be shifted by 0.15 kcal mol^{-1} in the downward (negative) and upward directions (positive), respectively, for the whole curve to be continuous. This shifting makes the MIM energy profile continuous while preserving the original activation energies in the individual reaction steps.

4.2. Pericyclic Reactions in Colombiasin Synthesis. The synthesis of Colombiasin involves a series of pericyclic reactions.[48] This three-step reaction (Figure 10) was studied using MIM2. The reaction starts with the electrocyclic ring opening of the highly constrained cyclobutene derivative. In the next step, an electrocyclic ring closure occurs to form a very stable six-membered ring, and, finally, the molecule undergoes an intramolecular hydrogen transfer to attain aromaticity in one of its rings. These steps are illustrated in Figure 11, along with all of its optimized reacting species. The fragmentation scheme for each step was chosen on the basis of the corresponding transition state, which shows where the reaction is localized (Figure 12).

The energy profile of the whole reaction as obtained from MIM2 calculations is depicted in Figure 13a. Because the reaction passes through two intermediate configurations, two discontinuities are present in the PEC (0.6 kcal mol^{-1} at intermediate-1 (I1), and 1.5 kcal mol^{-1} at intermediate-2 (I2)), and hence two equations need to be written to solve for the discontinuities in the PEC:

$$x_1 - x_2 = E_2^{I1} - E_1^{I1} = 0.6 \text{ kcal mol}^{-1} \quad (9)$$

$$x_2 - x_3 = E_3^{I2} - E_2^{I2} = -1.5 \text{ kcal mol}^{-1} \quad (10)$$

Using eq 7, the set of x_i's that provide a continuous curve with minimum $\|\boldsymbol{x}\|_2$ can be obtained, $x_1 = -0.1$, $x_2 = -0.7$, and $x_3 = 0.8$. Hence, for the PEC to be continuous, first, second, and third steps' curves need to be shifted by 0.1 kcal mol^{-1} in the downward direction, 0.7 kcal mol^{-1} in the downward direction, and 0.8 kcal mol^{-1} in the upward direction, respectively (Figure 13b). Notice that the discontinuity between the second and third reaction steps (1.5 kcal mol^{-1}) is relatively larger than other cases in this study. This is a consequence of the significant change in fragment size in going from step 2 to step 3 (Figure 12). In the fragmentation scheme for step 2, the whole six-membered ring is confined in a fragment (green shaded region in Figure 12b), whereas in step 3, only a small enol unit (pink shaded region in Figure 12c) is one of the largest fragments. This disparity in fragment sizes in going from one reaction step to another may give rise to large discontinuities in the PEC.

4.3. Multistep Nonlinear (Branched and Cyclic) Reaction Networks. A nonlinear reaction network can be a complex web of chemical species interconnected to one another through multiple mechanistic routes. Unlike a linear reaction network, a nonlinear (cyclic and branched) reaction network may not have a definite starting reactant and end product. Broadly, certain features can be attributed to such reactions, which are common to all nonlinear networks. For instance, a chemical species may be formed through more than one reaction pathway. In such a case, the concerned species is connected to more than two transition states, due to which more than two different fragmentation schemes may come in conflict for a single molecule. Consequently, multiple equal-

DOI: 10.1021/acs.jctc.9b00152
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

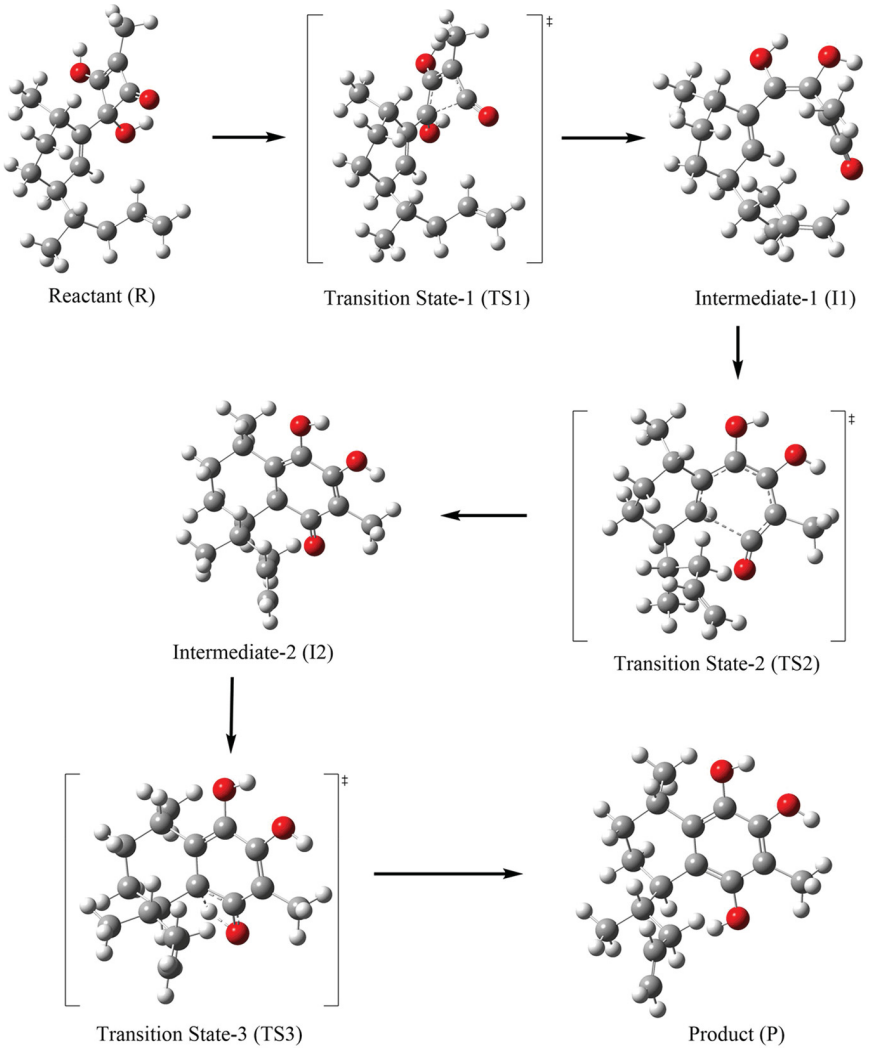


Figure 11. 3D representation of the pericyclic reactions involved in the synthesis of Colombiasin.

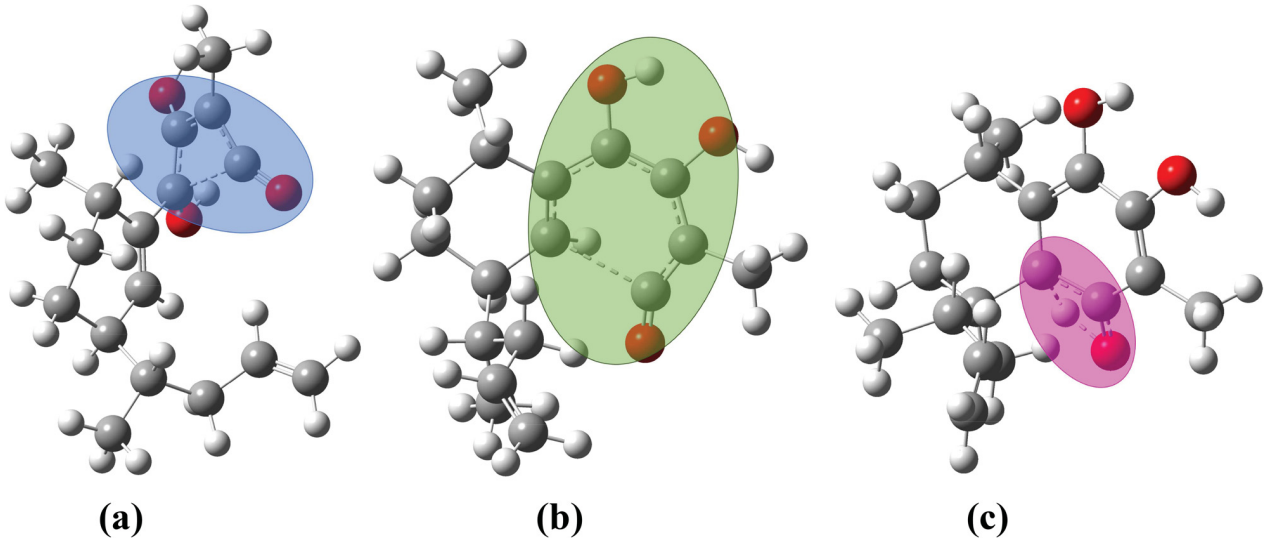


Figure 12. Fragmentation schemes for (a) transition state-1 (TS1), (b) transition state-2 (TS2), and (c) transition state-3 (TS3).

ities (eq 3) need to be satisfied to fix the discontinuity for a single species, in contrast to linear reaction networks where only one conflict (and, hence one eq 3) for one intermediate needs to be resolved between two fragmentation schemes. For example, if a species is connected to three transition states, and hence needs to be computed using three fragmentation

DOI: 10.1021/acs.jctc.9b00152
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

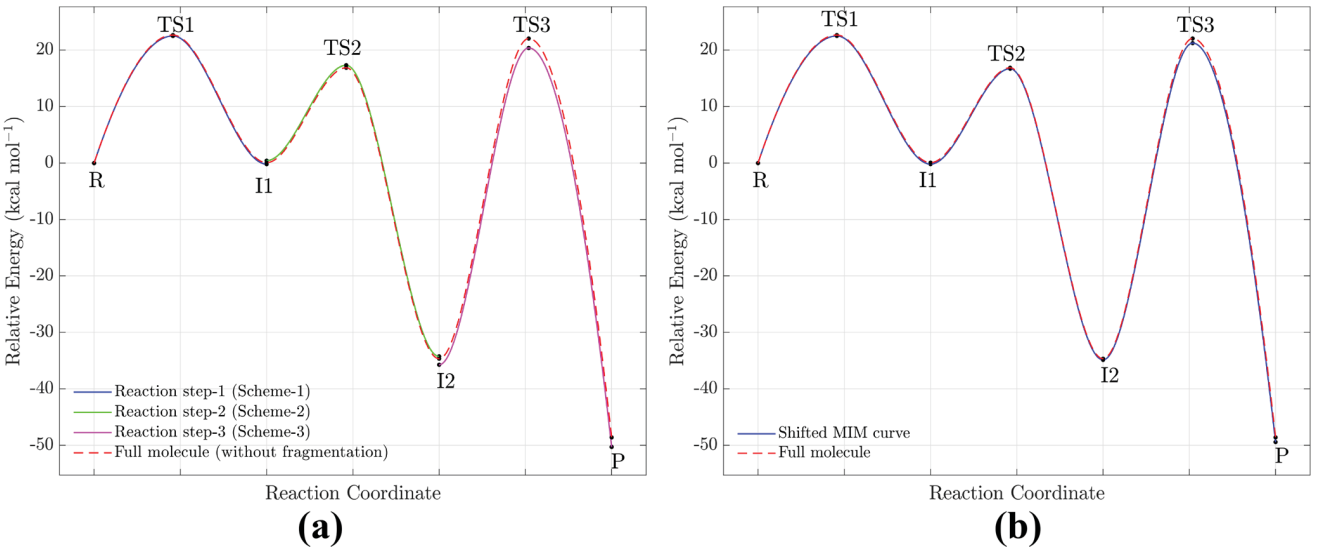


Figure 13. Energy profile for the pericyclic reactions in the synthesis of Colombiasin (a) as obtained from MIM2 calculation and (b) obtained after shifting the MIM energy curves.

schemes (say *a*, *b*, and *c*), then, to solve the resulting discontinuity, three eq 3's (between schemes *a* and *b*, *a* and *c*, and *b* and *c*) need to hold simultaneously for the intermediate of interest. In general, if a chemical species in a reaction network is connected to *n* transition states, and hence needs to be calculated using *n* fragmentation schemes, then the number of equalities (from eq 3) that needs to be satisfied for that species is $\frac{n(n-1)}{2}$. Hence, the number of equations to be solved simultaneously are usually more than the number of unknown variables x_i (amount of shift in a reaction step), resulting in an overdetermined system of equations to solve for all discontinuities in a nonlinear reaction network.

Using these ideas, a protocol to resolve discontinuities in a general hypothetical cyclic reaction (Figure 14) is discussed. In the given cyclic reaction, every chemical species can be obtained from every other species. As such, there is no specific starting or end point of this reaction. Overall, there are six distinct reaction steps in this cyclic reaction, and hence six unknown variables (x_1,x_2,..., x_6) corresponding to shift in each of the reaction steps (which is also equal to the total number of transition states, and fragmentation schemes in the reaction network) need to be determined to make the whole PEC continuous. The given reaction network can also be interpreted as a combination of various two-step processes (B⇌A⇌D, B⇌A⇌C, C⇌A⇌D, etc.) in which each of the four species is acting as an intermediate for three different two-step processes. This means that every species is connected to three different transition states, and hence needs to be calculated using three different fragmentation schemes. Written explicitly (using eq 3), we form the following overdetermined system of equations:

$$
\begin{aligned}
x_1 - x_2 &= E_2^{\mathrm{A}} - E_1^{\mathrm{A}} = \Delta E_{21}^{\mathrm{A}}\\
x_2 - x_6 &= E_6^{\mathrm{A}} - E_2^{\mathrm{A}} = \Delta E_{62}^{\mathrm{A}}\\
x_1 - x_6 &= E_6^{\mathrm{A}} - E_1^{\mathrm{A}} = \Delta E_{61}^{\mathrm{A}}\\
x_2 - x_3 &= E_3^{\mathrm{B}} - E_2^{\mathrm{B}} = \Delta E_{32}^{\mathrm{B}}\\
x_3 - x_5 &= E_5^{\mathrm{B}} - E_3^{\mathrm{B}} = \Delta E_{53}^{\mathrm{B}}\\
x_2 - x_5 &= E_5^{\mathrm{B}} - E_2^{\mathrm{B}} = \Delta E_{52}^{\mathrm{B}}\\
x_3 - x_4 &= E_4^{\mathrm{C}} - E_3^{\mathrm{C}} = \Delta E_{43}^{\mathrm{C}}\\
x_4 - x_6 &= E_6^{\mathrm{C}} - E_4^{\mathrm{C}} = \Delta E_{64}^{\mathrm{C}}\\
x_3 - x_6 &= E_6^{\mathrm{C}} - E_3^{\mathrm{C}} = \Delta E_{63}^{\mathrm{C}}\\
x_1 - x_4 &= E_4^{\mathrm{D}} - E_1^{\mathrm{D}} = \Delta E_{41}^{\mathrm{D}}\\
x_4 - x_5 &= E_5^{\mathrm{D}} - E_4^{\mathrm{D}} = \Delta E_{54}^{\mathrm{D}}\\
x_1 - x_5 &= E_5^{\mathrm{D}} - E_1^{\mathrm{D}} = \Delta E_{51}^{\mathrm{D}}
\end{aligned}
\tag{11}
$$

The above set of equations can be written in the matrix notation as follows:

$$\mathbf{A}\boldsymbol{x} = \boldsymbol{b} \tag{12}$$

where

$$
\mathbf{A} = \begin{pmatrix}
1 & -1 & 0 & 0 & 0 & 0\\
0 & 1 & 0 & 0 & 0 & -1\\
1 & 0 & 0 & 0 & 0 & -1\\
0 & 1 & -1 & 0 & 0 & 0\\
0 & 0 & 1 & 0 & -1 & 0\\
0 & 1 & 0 & 0 & -1 & 0\\
0 & 0 & 1 & -1 & 0 & 0\\
0 & 0 & 0 & 1 & 0 & -1\\
0 & 0 & 1 & 0 & 0 & -1\\
1 & 0 & 0 & -1 & 0 & 0\\
0 & 0 & 0 & 1 & -1 & 0\\
1 & 0 & 0 & 0 & -1 & 0
\end{pmatrix},\ x = \begin{pmatrix} x_1\\ x_2\\ x_3\\ x_4\\ x_5\\ x_6 \end{pmatrix},\ \text{and } b = \begin{pmatrix}
\Delta E_{21}^{\mathrm{A}}\\ \Delta E_{62}^{\mathrm{A}}\\ \Delta E_{61}^{\mathrm{A}}\\ \Delta E_{32}^{\mathrm{B}}\\ \Delta E_{53}^{\mathrm{B}}\\ \Delta E_{52}^{\mathrm{B}}\\ \Delta E_{43}^{\mathrm{C}}\\ \Delta E_{64}^{\mathrm{C}}\\ \Delta E_{63}^{\mathrm{C}}\\ \Delta E_{41}^{\mathrm{D}}\\ \Delta E_{54}^{\mathrm{D}}\\ \Delta E_{51}^{\mathrm{D}}
\end{pmatrix}
\tag{13}
$$

For most cases, an exact solution does not exist for overdetermined systems; therefore, a least-squares solution is usually found by minimizing $\|\boldsymbol{b}-\mathbf{A}\boldsymbol{x}\|_2$, which can be interpreted as an optimum solution to the problem. However,

DOI: 10.1021/acs.jctc.9b00152
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

for the problem at hand (and potentially for other nonlinear reaction networks as well), the matrix **A** is rank deficient, meaning its columns are not linearly independent. In such a case, the least-squares solution with the smallest $\|\boldsymbol{x}\|_2$ can be determined. This solution is unique and is called the minimum norm least-squares solution. The desired solution has the same mathematical form as that for an underdetermined problem (eq 7), that is:

$$\boldsymbol{x} = \mathbf{A}^{+}\boldsymbol{b} \tag{14}$$

where $\mathbf{A}^{+}$ represents the pseudoinverse of **A**. This least-squares solution holds even when matrix **A** has full column rank. Physically, the solution obtained from eq 14 does not make the curves exactly continuous, but only *least discontinuous*.

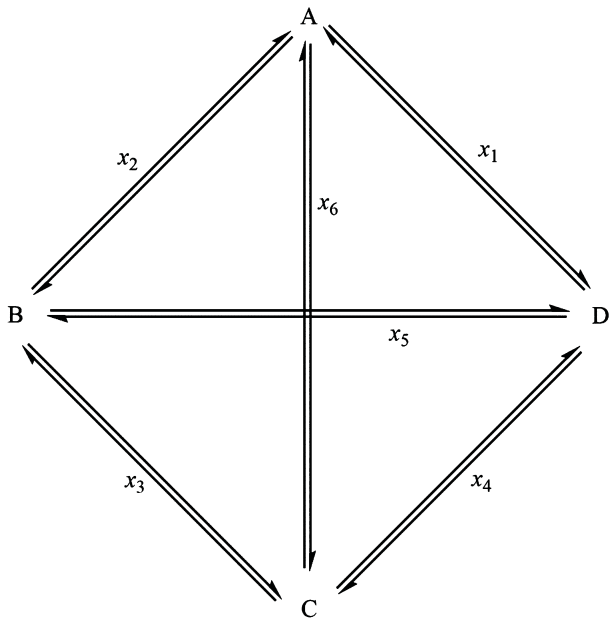


Figure 14. A hypothetical cyclic reaction network.

Although the protocols developed in this work cater to the discontinuities in PECs inherent to the fragmentation methods, the mathematical treatment itself is quite general and independent of the origin of the discontinuity. Most commonly, energetic discontinuities in a PEC seem to arise due to a mismatch in the levels of theory between two consecutive steps of a chemical reaction. For instance, while studying a reaction with a multireference method like CASSCF, different steps of a chemical reaction may require the use of different active spaces, resulting in a qualitatively similar discontinuity in the PEC, which can be easily resolved with the proposed method.

5. CONCLUSIONS

In the present work, the energy landscape of chemical reactions has been studied using MIM fragmentation method. Because of structural changes in the reacting species during the progress of a reaction, the fragmentation scheme of the participating molecule was modified accordingly. To avoid any errors in energy barriers due to differences in the fragmentation scheme, the scheme was kept consistent throughout a step of the reaction. Consequently, the points on PEC corresponding to intermediate geometries became a point for transition between two fragmentation schemes, and hence also a point of discontinuity. The magnitude of discontinuity between the PECs for the conflicting fragmentation schemes was observed to be in the range of 0−2 kcal mol^{-1}, quite small as compared to activation energies. Moreover, the discontinuities do not seem to be dependent on system size, but rather on the difference in fragment sizes in the involved fragmentation schemes. For linear reactions, the discontinuities were removed by shifting the MIM energy curves (using the minimum norm solution) such that the energies of intermediates obtained from the conflicting fragmentation schemes coincide with each other, whereas for the inconsistent case of nonlinear reactions, the discontinuities were minimized using the minimum norm least-squares solution.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.9b00152.

S1, An example involving intramolecular Diels−Alder reaction; S2−S4, Cartesian coordinates and total energies of all of the structures (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: kraghava@indiana.edu.

ORCID

Bishnu Thapa: 0000-0003-3521-1062

Krishnan Raghavachari: 0000-0003-3275-1426

Funding

This work was supported by the National Science Foundation grant CHE-1665427 at Indiana University.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Computations were facilitated through the use of advanced computational, storage, and networking infrastructure provided by the BigRed2 supercomputer system at Indiana University.

■ REFERENCES

(1) Field, M. J.; Bash, P. A.; Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.* **1990**, *11*, 700−733.

(2) Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H.-B.; Ding, L.; Morokuma, K. The ONIOM Method and Its Applications. *Chem. Rev.* **2015**, *115*, 5678−5796.

(3) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.* **2012**, *112*, 632−672.

(4) Gadre, S. R.; Yeole, S. D.; Sahu, N. Quantum Chemical Investigations on Molecular Clusters. *Chem. Rev.* **2014**, *114*, 12132−12173.

(5) Collins, M. A.; Bettens, R. P. A. Energy-Based Molecular Fragmentation Methods. *Chem. Rev.* **2015**, *115*, 5607−5642.

(6) Raghavachari, K.; Saha, A. Accurate Composite and Fragment-Based Quantum Chemical Models for Large Molecules. *Chem. Rev.* **2015**, *115*, 5643−5677.

(7) Tschumper, G. S. Reliable Electronic Structure Computations for Weak Noncovalent Interactions in Clusters. *Reviews in Computational Chemistry* **2009**, 39.

DOI: 10.1021/acs.jctc.9b00152
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(8) Fedorov, D. G. The fragment molecular orbital method: theoretical development, implementation in GAMESS, and applications. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2017**, *7*, No. e1322.
(9) Ratcliff, L. E.; Mohr, S.; Huhs, G.; Deutsch, T.; Masella, M.; Genovese, L. Challenges in large scale quantum mechanical calculations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2017**, *7*, No. e1290.
(10) 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.
(11) Jose, K. V. J.; Beckett, D.; Raghavachari, K. Vibrational Circular Dichroism Spectra for Large Molecules through Molecules-in-Molecules Fragment-Based Approach. *J. Chem. Theory Comput.* **2015**, *11*, 4238−4247.
(12) Jose, K. V. J.; Raghavachari, K. Evaluation of Energy Gradients and Infrared Vibrational Spectra through Molecules-in-Molecules Fragment-Based Approach. *J. Chem. Theory Comput.* **2015**, *11*, 950−961.
(13) Jovan Jose, K. V.; Raghavachari, K. Molecules-in-molecules fragment-based method for the evaluation of Raman spectra of large molecules. *Mol. Phys.* **2015**, *113*, 3057−3066.
(14) Jovan Jose, K. V.; Raghavachari, K. Raman Optical Activity Spectra for Large Molecules through Molecules-in-Molecules Fragment-Based Approach. *J. Chem. Theory Comput.* **2016**, *12*, 585−594.
(15) Thapa, B.; Beckett, D.; Jovan Jose, K. V.; Raghavachari, K. Assessment of Fragmentation Strategies for Large Proteins Using the Multilayer Molecules-in-Molecules Approach. *J. Chem. Theory Comput.* **2018**, *14*, 1383−1394.
(16) Jose, K. V. J.; Raghavachari, K. Fragment-Based Approach for the Evaluation of NMR Chemical Shifts for Large Biomolecules Incorporating the Effects of the Solvent Environment. *J. Chem. Theory Comput.* **2017**, *13*, 1147−1158.
(17) Thapa, B.; Beckett, D.; Erickson, J.; Raghavachari, K. Theoretical Study of Protein−Ligand Interactions Using the Molecules-in-Molecules Fragmentation-Based Method. *J. Chem. Theory Comput.* **2018**, *14*, 5143−5155.
(18) Debnath, S.; Sengupta, A.; Jose, K. V. J.; Raghavachari, K. Fragment-Based Approaches for Supramolecular Interaction Energies: Applications to Foldamers and Their Complexes with Anions. *J. Chem. Theory Comput.* **2018**, *14*, 6226−6239.
(19) Kerdcharoen, T.; Liedl, K. R.; Rode, B. M. A QM/MM simulation method applied to the solution of Li+ in liquid ammonia. *Chem. Phys.* **1996**, *211*, 313−323.
(20) Kerdcharoen, T.; Morokuma, K. ONIOM-XS: an extension of the ONIOM method for molecular simulation in condensed phase. *Chem. Phys. Lett.* **2002**, *355*, 257−262.
(21) Heyden, A.; Lin, H.; Truhlar, D. G. Adaptive Partitioning in Combined Quantum Mechanical and Molecular Mechanical Calculations of Potential Energy Functions for Multiscale Simulations. *J. Phys. Chem. B* **2007**, *111*, 2231−2241.
(22) Bulo, R. E.; Ensing, B.; Sikkema, J.; Visscher, L. Toward a Practical Method for Adaptive QM/MM Simulations. *J. Chem. Theory Comput.* **2009**, *5*, 2212−2221.
(23) Bernstein, N.; Várnai, C.; Solt, I.; Winfield, S. A.; Payne, M. C.; Simon, I.; Fuxreiter, M.; Csányi, G. QM/MM simulation of liquid water with an adaptive quantum region. *Phys. Chem. Chem. Phys.* **2012**, *14*, 646−656.
(24) Takenaka, N.; Kitamura, Y.; Koyano, Y.; Nagaoka, M. An improvement in quantum mechanical description of solute-solvent interactions in condensed systems via the number-adaptive multiscale quantum mechanical/molecular mechanical-molecular dynamics method: Application to zwitterionic glycine in aqueous solution. *J. Chem. Phys.* **2012**, *137*, 024501.
(25) Waller, M. P.; Kumbhar, S.; Yang, J. A Density-Based Adaptive Quantum Mechanical/Molecular Mechanical Method. *ChemPhysChem* **2014**, *15*, 3218−3225.
(26) Watanabe, H. C.; Kubař, T.; Elstner, M. Size-Consistent Multipartitioning QM/MM: A Stable and Efficient Adaptive QM/MM Method. *J. Chem. Theory Comput.* **2014**, *10*, 4242−4252.
(27) Böckmann, M.; Doltsinis, N. L.; Marx, D. Adaptive Switching of Interaction Potentials in the Time Domain: An Extended Lagrangian Approach Tailored to Transmute Force Field to QM/MM Simulations and Back. *J. Chem. Theory Comput.* **2015**, *11*, 2429−2439.
(28) Zheng, M.; Waller, M. P. Adaptive quantum mechanics/molecular mechanics methods. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2016**, *6*, 369−385.
(29) Duster, A. W.; Wang, C.-H.; Garza, C. M.; Miller, D. E.; Lin, H. Adaptive quantum/molecular mechanics: what have we learned, where are we, and where do we go from here? *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2017**, *7*, No. e1310.
(30) Delle Site, L. Some fundamental problems for an energy-conserving adaptive-resolution molecular dynamics scheme. *Phys. Rev. E* **2007**, *76*, 047701.
(31) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels-Alder Reactions and Pt(P(t-Bu)3)2 + H2 Oxidative Addition. *J. Phys. Chem.* **1996**, *100*, 19357−19363.
(32) Re, S.; Morokuma, K. ONIOM Study of Chemical Reactions in Microsolvation Clusters: (H2O)nCH3Cl + OH-(H2O)m (n + m = 1 and 2). *J. Phys. Chem. A* **2001**, *105*, 7185−7197.
(33) Basiuk, V. A. ONIOM Studies of Chemical Reactions on Carbon Nanotube Tips: Effects of the Lower Theoretical Level and Mutual Orientation of the Reactants. *J. Phys. Chem. B* **2003**, *107*, 8890−8897.
(34) Martí, S.; Moliner, V.; Tuñón, I. Improving the QM/MM Description of Chemical Processes: A Dual Level Strategy To Explore the Potential Energy Surface in Very Large Systems. *J. Chem. Theory Comput.* **2005**, *1*, 1008−1016.
(35) Osuna, S.; Morera, J.; Cases, M.; Morokuma, K.; Solà, M. Diels-Alder Reaction between Cyclopentadiene and C60: An Analysis of the Performance of the ONIOM Method for the Study of Chemical Reactivity in Fullerenes and Nanotubes. *J. Phys. Chem. A* **2009**, *113*, 9721−9726.
(36) Steinmann, C.; Fedorov, D. G.; Jensen, J. H. Mapping Enzymatic Catalysis Using the Effective Fragment Molecular Orbital Method: Towards all ab initio Biochemistry. *PLoS One* **2013**, *8*, No. e60602.
(37) Nakata, H.; Fedorov, D. G.; Nagata, T.; Kitaura, K.; Nakamura, S. Simulations of Chemical Reactions with the Frozen Domain Formulation of the Fragment Molecular Orbital Method. *J. Chem. Theory Comput.* **2015**, *11*, 3053−3064.
(38) 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 derivatives1Dedicated to Professor Keiji Morokuma in celebration of his 65th birthday.1. *J. Mol. Struct.: THEOCHEM* **1999**, *461−462*, 1−21.
(39) Fukui, K. The path of chemical reactions - the IRC approach. *Acc. Chem. Res.* **1981**, *14*, 363−368.
(40) Hratchian, H. P.; Schlegel, H. B. Accurate reaction paths using a Hessian based predictor−corrector integrator. *J. Chem. Phys.* **2004**, *120*, 9918−9924.
(41) Hratchian, H. P.; Schlegel, H. B. Using Hessian Updating To Increase the Efficiency of a Hessian Based Predictor-Corrector Reaction Path Following Method. *J. Chem. Theory Comput.* **2005**, *1*, 61−69.
(42) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215−241.
(43) Penrose, R. A generalized inverse for matrices. *Math. Proc. Cambridge Philos. Soc.* **1955**, *51*, 406−413.

DOI: 10.1021/acs.jctc.9b00152
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(44) Houk, K. N.; Li, Y.; Evanseck, J. D. Transition Structures of Hydrocarbon Pericyclic Reactions. *Angew. Chem., Int. Ed. Engl.* **1992**, *31*, 682−708.
(45) Houk, K. N.; Gonzalez, J.; Li, Y. Pericyclic Reaction Transition States: Passions and Punctilios, 1935−1995. *Acc. Chem. Res.* **1995**, *28*, 81−90.
(46) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16*, revision A; Gaussian, Inc.: Wallingford, CT, 2016.
(47) Nicolaou, K. C.; Petasis, N. A.; Zipkin, R. E.; Uenishi, J. The endiandric acid cascade. Electrocyclizations in organic synthesis. I. Stepwise, stereocontrolled total synthesis of endiandric acids A and B. *J. Am. Chem. Soc.* **1982**, *104*, 5555−5557.
(48) Nicolaou, K. C.; Edmonds, D. J.; Bulger, P. G. Cascade Reactions in Total Synthesis. *Angew. Chem., Int. Ed.* **2006**, *45*, 7134−7186.

DOI: 10.1021/acs.jctc.9b00152
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX