

Transition state-finding strategies for use with the growing string method

Anthony Goodrow,¹ Alexis T. Bell,^{1,a)} and Martin Head-Gordon^{2,a)}¹*Department of Chemical Engineering, University of California, Berkeley, California 94720-1462, USA*²*Department of Chemistry, University of California, Berkeley, California 94720-1462, USA*

(Received 1 March 2009; accepted 29 May 2009; published online 26 June 2009)

Efficient identification of transition states is important for understanding reaction mechanisms. Most transition state search algorithms require long computational times and a good estimate of the transition state structure in order to converge, particularly for complex reaction systems. The growing string method (GSM) [B. Peters *et al.*, *J. Chem. Phys.* **120**, 7877 (2004)] does not require an initial guess of the transition state; however, the calculation is still computationally intensive due to repeated calls to the quantum mechanics code. Recent modifications to the GSM [A. Goodrow *et al.*, *J. Chem. Phys.* **129**, 174109 (2008)] have reduced the total computational time for converging to a transition state by a factor of 2 to 3. In this work, three transition state-finding strategies have been developed to complement the speedup of the modified-GSM: (1) a hybrid strategy, (2) an energy-weighted strategy, and (3) a substring strategy. The hybrid strategy initiates the string calculation at a low level of theory (HF/STO-3G), which is then refined at a higher level of theory (B3LYP/6-31G*). The energy-weighted strategy spaces points along the reaction pathway based on the energy at those points, leading to a higher density of points where the energy is highest and finer resolution of the transition state. The substring strategy is similar to the hybrid strategy, but only a portion of the low-level string is refined using a higher level of theory. These three strategies have been used with the modified-GSM and are compared in three reactions: alanine dipeptide isomerization, H-abstraction in methanol oxidation on VO_x/SiO₂ catalysts, and C–H bond activation in the oxidative carbonylation of toluene to *p*-toluic acid on Rh(CO)₂(TFA)₃ catalysts. In each of these examples, the substring strategy was proved most effective by obtaining a better estimate of the transition state structure and reducing the total computational time by a factor of 2 to 3 compared to the modified-GSM. The applicability of the substring strategy has been extended to three additional examples: cyclopropane rearrangement to propylene, isomerization of methylcyclopropane to four different stereoisomers, and the bimolecular Diels–Alder condensation of 1,3-butadiene and ethylene to cyclohexene. Thus, the substring strategy used in combination with the modified-GSM has been demonstrated to be an efficient transition state-finding strategy for a wide range of types of reactions. © 2009 American Institute of Physics. [DOI: 10.1063/1.3156312]

I. INTRODUCTION

Efficient techniques for identifying transition states are required for the analysis of proposed reaction pathways and calculating activation energies for elementary processes.^{1–6} The need for robust and highly efficient transition state-finding algorithms is particularly important for the analysis of catalyzed reactions, since these often involve very large numbers of atoms in order to provide a reasonable description of the catalytically active site and the reactants. During the last few decades, two broad classes of transition state-finding methods have been developed, referred to as surface walking and interpolation methods.^{6–8} Surface-walking methods, such as that introduced by Cerjan and Miller,⁹ begin with a reactant configuration and move uphill along the potential energy surface in the direction of the largest negative Hessian eigenvalue. Three major difficulties with this class of methods are that the initial configuration dictates which transition state will be found, the analytical Hessian

must be calculated, and the calculations are slow to converge, especially for complex reaction systems. Interpolation methods, such as the nudged-elastic band (NEB),^{10–13} begin with both reactant and product configurations as well as an initial guess of the reaction pathway. Typically the initial pathway is the linear synchronous transit (LST) joining the two configurations. The difficulty with these methods is that electronic structure calculations can fail to converge if the transition state lies far from the LST, since in this case the initial guess of the transition state can lie on high energy portions of the potential energy surface.

The growing string method (GSM), developed by Peters *et al.*,⁷ is an interpolation method that does not require a guess of the initial transition state or minimum energy pathway (MEP). This approach is based on the string method (SM) introduced by E *et al.*¹⁴ and recently has been modified by Burger and Yang.^{15,16} The GSM only requires knowledge on the reactant and product states. However, the GSM, as well as other interpolation methods, requires repeated quantum mechanical (QM) calculations to determine the energy and the energy gradient of intermediate geometries, which make these methods computationally demanding. To address

^{a)}Authors to whom correspondence should be addressed. Electronic addresses: alexbell@berkeley.edu and mhg@bastille.cchem.berkeley.edu.

this issue, Goodrow *et al.*¹⁷ recently presented several modifications that reduce the number of QM gradient evaluations required in the GSM, thereby decreasing the computational time needed to determine the transition state. These modifications consist of (1) replacing Cartesian coordinates with internal coordinates to describe the molecular geometry, (2) replacing the steepest descent method with the conjugate gradient method for the minimization of orthogonal forces at nodes along the reaction pathway, and (3) introduction of a potential energy surface interpolation scheme in order to minimize the number of QM calculations. It is noted that Sheppard *et al.*¹⁸ recently showed that a further slight improvement in the speed of minimizing orthogonal forces during a search for a transition state via the NEB method can be achieved by using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method instead of the conjugate gradient method.

The aim of the present study was to examine additional strategies for reducing the computational requirements of the GSM. Three approaches were explored: (1) a hybrid strategy in which the initial string is grown at a low level of theory and then refined at a higher level, (2) an energy-weighted distribution strategy in which the node density is increased near the transition state rather than being equally spaced along the reaction pathway, and (3) a substring strategy in which only a portion of the string is refined at a higher level of theory. All three strategies have been tested on a series of increasingly complex reaction systems.

II. THEORY

A. Modified-growing string method

Before discussing the transition state-finding strategies to be used with the modified-GSM, it is useful to summarize the method. The modified-GSM consists of four steps: (1) initiation, (2) evolution, (3) reparametrization, and (4) growth.^{7,17}

The initiation step begins with determination of the geometry and energy of the reactant and product states. The list of coordinates and center of mass in the optimized reactant and product states are adjusted to ensure that the geometries are defined consistently. Two initial string segments are then created with each segment consisting of two configurations (referred to as nodes). The string emanating from the reactant state consists of the reactant geometry and a node added along the LST close to the reactant. The string emanating from the product state is constructed in a similar manner. The two string segments are then grown independently up to a point where they are joined together. Details about the growth of the string segments and their ultimate unification into a single string are given below.

In the evolution step, the QM code is used to determine the energy and gradient (forces) at each node along the reaction pathway. The norm of the force orthogonal to the reaction pathway $\|f_{\perp}\|$ on each node is minimized using the conjugate gradient method. In performing these calculations, the required energies and forces are determined either by QM calculations or use of an interpolation scheme. The interpolation scheme estimates the energy of a node using a weighted sum of second-order Taylor series based on nonre-

dundant delocalized internal coordinates in a manner similar to that described by Collins and co-workers.¹⁹⁻²⁴ The interpolation scheme is used when there are at least M previously calculated QM points within a confidence length of d_{tol} . In this work, the values of $M=5$ and $d_{\text{tol}}=1.0d_{\text{node}}$ were used, where d_{node} is the curvilinear distance between adjacent nodes. Previous work¹⁷ has shown that these values of M and d_{tol} successfully balance the frequency of interpolation while still maintaining high accuracy of the interpolated results. The two parts of an evolution step consisting of determining the new geometry and calculating its energy and gradient explicitly from either a QM calculation or by interpolation are repeated until the norm of the orthogonal force on each node falls below a user-specified value $\|f_{\perp}^{\text{tol}}\|$. Each node on the evolving string or string segments is treated in the same manner until $\|f_{\perp}\|$ is minimized for all existing nodes.

Upon completion of the evolution step, the nodes are redistributed along a string of normalized arclength during the reparametrization step. The arclength s is determined by integrating along a cubic spline that passes through all existing nodes in each string segment. In the case where the reactant and product string segments are fully joined, then only a single spline is fitted. The string is normalized such that the reactant and product nodes are located at $s=0$ and $s=1$, respectively. Then, the nodes are redistributed based on a parametrization density $\rho(s)$. In the case of a uniform parametrization density, the nodes are equally spaced (in terms of arclength) along the path between the reactant and product states.

In the growth step, an additional node is added as the string grows inward from the reactant and product ends until the maximum number of nodes N_{max} has been added. The location of each added node is determined by extrapolating the fitted cubic spline. For example, the location of a node added to the string segment emanating from the reactant configuration is determined by extrapolating the spline by a user-specified distance from the end of the reactant string segment. The evolution and reparametrization steps are then repeated until the vacant arclength between the two string segments vanishes. Once the two string segments are joined and there are N_{max} nodes along the string, only the evolution and reparametrization steps are repeated until the string is fully converged. The final convergence criterion is that $\sum_{i=1}^{N_{\text{max}}} \|f_{\perp,i}\|$ is less than a user-specified value $\|F_{\perp}^{\text{tol}}\|$.

B. Hybrid strategy

The hybrid strategy used for the modified-GSM has been described previously¹⁷ and is illustrated in Fig. 1. A string is first grown between the reactant and product states using a low level of theory and minimal basis sets to get the correct topology of the reaction pathway. Then, the low-level string is refined at a higher level of theory using more complete basis sets. For example, a low-level GSM calculation can be performed using HF with STO-3G basis sets and then refined using density functional theory (DFT) with the B3LYP functional and 6-31G* basis sets. The choice of which low level of theory to use is important and should be made with caution because the HF/STO-3G level is not always appropriate.

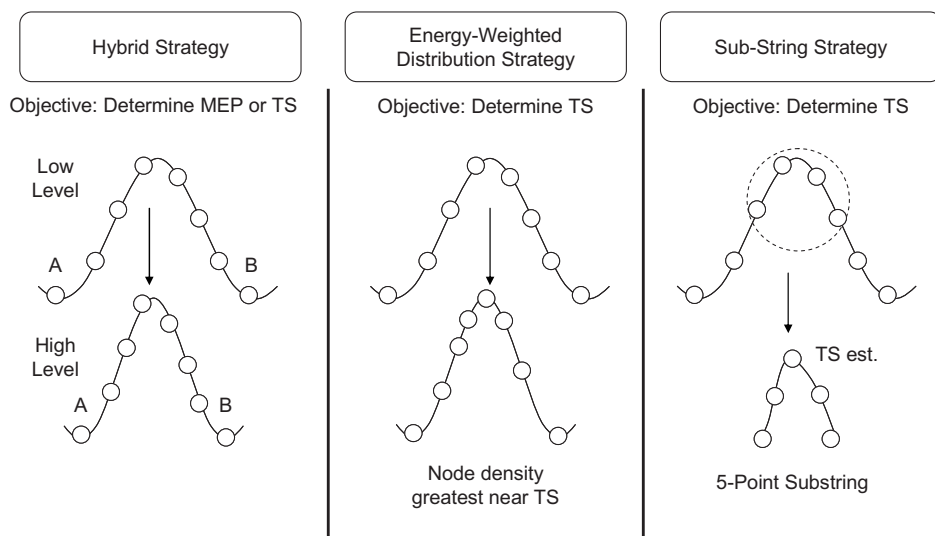


FIG. 1. Three transition state (TS)-finding strategies for use with the modified-GSM. The three strategies are used for determining the TS joining reactant (A) and product (B).

In cases where the reaction topology is incorrectly described by using HF/STO-3G, a low level of theory with higher fidelity can be used such as PBE/6-31G. The transition state estimate obtained from the B3LYP/6-31G* string is then used as a starting geometry for a transition state search calculation at the same level of theory. The hybrid strategy can be used to determine both the MEP and the transition state.

C. Energy-weighted strategy

An energy-weighted strategy for the SM was first suggested by E *et al.*²⁵ It involves distributing the nodes on a string lying between the reactant and product states in such a manner that the node density is highest near where the energy is highest. Recent work by Aguilar-Mogas *et al.*²⁶ showed that a similar strategy can be used to determine the intrinsic reaction coordinate (IRC) path by spacing the nodes equidistant in energy. The advantage of using an energy-weighted strategy is that it gives finer resolution of nodes near the transition state, as shown in Fig. 1; however, this occurs at the expense of a lower density of nodes near the reactant and product states. Consequently, the energy-weighted strategy is best used for identifying the transition state.

During the reparametrization step, the nodes along the string are redistributed according to a prescribed parametrization density $\rho(s)$. In the energy-weighted strategy, the parametrization density is an inverse-weighted Boltzmann distribution, Eq. (1), which weights nodes with a higher energy more heavily than those with a lower energy. In Eq. (1), $E(s)$ is the energy of a node at a given location along the string. E_{\min} is the minimum between the energy of the reactant and product states, and hence is equal to the energy of the reactant for endothermic reactions and the energy of the product for exothermic reactions. E_{\max} is the maximum energy along the string, which changes as the calculation progresses. The final value for E_{\max} is the energy of the transition state estimate

$$\rho(s) = \exp\left(\frac{E(s) - E_{\min}}{E_{\max} - E_{\min}}\right). \quad (1)$$

The local convergence criterion for each node during the evolution of the string is related to the local perpendicular gradient. For the three nodes with the highest parametrization density (i.e., the three nodes closest to the transition state), the convergence criterion is tightened by a factor of 20. On the three nodes with the highest energy, the perpendicular gradient $\|f_{\perp}\|$ must be less than 0.001 hartree/bohr for convergence, and for all other nodes $\|f_{\perp}\|$ must be less than 0.02 hartree/bohr.

A disadvantage of the above energy-weighted distribution is that it will only find the transition state with the highest energy. This can become problematic for reactions that consist of multiple transition states joining the reactant and product states along a single reaction coordinate such as on the Müller-Brown²⁷ potential energy surface. To remedy this situation, when a cubic spline is fit through the existing nodes during the reparametrization step, a node can be placed at all locations where the slope is zero. This step ensures that multiple transition states and minima connecting the reactant and product are found.

D. Substring strategy

In the substring strategy only a portion of the low-level string is refined at a higher level. This strategy is similar to the IRCMAX method developed by Malick *et al.*,²⁸ in which the activation energy at a high level of theory is estimated by using a series of geometries from a low-level IRC calculation. The subset of the nodes is chosen to consist of the node with the highest energy along the string and the geometry of P nodes ($P < N_{\max}/2$) immediately to the left and right of the one with the highest energy. For example, in a five-point substring calculation, the node with the highest energy and the two nodes immediately to the left and right of the highest energy node ($P=2$) are selected, as shown in Fig. 1. The leftmost point becomes the new “reactant” configuration and the rightmost point becomes the new “product” configuration.

There exists the possibility that when the high-level substring calculation is started, the geometry of the node that was once highest in energy is no longer so because the val-

ues of the energy in the low-level string were not accurate. For instance, rather than having a substring that passes through an energy maximum, the substring could now consist of a series of nodes which either increase or decrease monotonically in energy. If the node with the highest energy is either the new reactant or product state upon refining the low-level substring at a higher level of theory, then additional nodes from the low-level string can be added until the maximum in energy exists between the two end points.

The use of a single-node substring was explored. However, in most cases it was not possible to determine the correct transition state by carrying out a high-level transition state search using the transition state estimate from a low-level string calculation. In these cases, either the transition state search calculation failed to converge or it converged to an incorrect transition state. In cases where the single-node substring converged to the correct transition state, the total calculation time was longer than if a multinode substring was used due to the many optimization cycles required to converge the poor transition state estimate. Thus, P must be nonzero to ensure that an accurate transition state estimate is contained within the substring.

The advantage of the substring strategy is that it focuses the modified-GSM calculation on the portion of the string where the transition state is located. This feature is similar to the energy-weighted strategy because both strategies obtain finer resolution of the reaction pathway near the transition state. If multiple transition states exist, the substring strategy can be repeated. For instance, on the Müller–Brown potential energy surface, the substring strategy first determines the transition state with the highest energy. A second calculation can then be performed to focus on the region where the other transition state is located. Therefore, the substring strategy is best used for determining the transition state rather than the MEP.

E. Computations

QCHEM 3.0 (Ref. 29) was used to perform energy and force evaluations. Low-level strings were determined using HF and STO-3G basis sets. All high-level strings and transition state searches were carried out using DFT with the B3LYP functional and 6-31G* basis sets for nontransition metal atoms, whereas transition metal atoms (V and Rh) were treated using the LANL2DZ effective core potential. Single-point calculations were performed on all optimized reactant, product, and transition state geometries using B3LYP/6-311++G** to obtain more accurate values of the overall reaction energy and the activation energy.

The following parameters were specified for each force calculation: the self-consistent field (SCF) energy was considered converged if the difference between successive iterations was less than 10^{-7} hartree, the cutoff for neglecting two-electron integrals (referred to as the thresh value) was set to 10^{-10} , and the default SG-0 exchange-correlation integration grid was used. These three parameters were increased for final single-point calculations: a difference of less than 10^{-8} hartree for the SCF energy, a thresh value of 10^{-11} , and use of the SG-1 integration grid.

The same convergence criteria were used for all systems examined. For the low-level and high-level string calculations, the norm of the perpendicular gradient $\|f_{\perp}\|$ on each leading node was required to be less than 0.04 hartree/bohr before a new node was added and the extrapolated distance for placement of a new node was 0.08. Final convergence of the string was achieved when $\sum_{i=1}^{N_{\max}} \|f_{\perp,i}\|$ was less than 0.04 for the low-level string and 0.02 for the high-level string. In the case of the energy-weighted strategy and the substring strategy, the final convergence criterion for the low-level and high-level strings was that $\|f_{\perp}\|$ be less than 0.001 hartree/bohr for the nodes with the three highest energies. All high-level substring calculations involved five nodes.

The geometry of the node corresponding to the highest energy was used as the initial geometry for a transition state search calculation in QCHEM 3.0. All transition state search calculations were started by first computing the analytic Hessian (force constants), which were then read into a transition state optimization algorithm. All final transition states were verified to have one imaginary frequency.

The transition state-finding strategies are compared based on the number of QM evaluations of the energy, the CPU time required to completely grow the string, the CPU time required for the transition state search calculation, and the total CPU time for determining the transition state. The total CPU time is defined as the time needed to grow the string (number of QM gradients multiplied by the average time per QM gradient calculation) plus the time needed to perform the transition state search calculation. The time required for a transition state search calculation is equal to the time needed to calculate the initial Hessian plus the time needed to perform the transition state optimization.

III. RESULTS AND DISCUSSION

The first three examples used to evaluate the effectiveness of the methods presented here are identical to those reported earlier.¹⁷ They include alanine dipeptide isomerization, hydrogen transfer from a methoxide group associated with a vanadate species supported on silica, and hydrogen transfer occurring during the activation of toluene coordinated to a Rh complex. In each case, the final string contained 11 nodes.

A. Alanine dipeptide isomerization

Alanine dipeptide isomerization is a well-studied example for comparing the performance of transition state-finding methods^{30,31} and has been used to evaluate the performance of the GSM and the modified-GSM.¹⁷ The reaction is easily described using internal coordinates by the rotation of two dihedral angles ϕ (C–N–C _{α} –C) and ψ (N–C _{α} –C–N) from the reactant (C₅) to the product (C_{7AX}),³² as shown in Fig. 2. A single force calculation for the geometries of alanine dipeptide along the reaction pathway took ~ 8 min at B3LYP/6-31G* and only 10–15 s at HF/STO-3G.

Using the hybrid strategy, 324 QM gradient calculations were required to converge the low-level HF/STO-3G string. The final HF/STO-3G string was used to initiate a

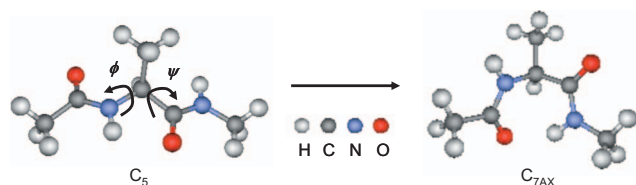


FIG. 2. (Color) Alanine dipeptide isomerization from species C_5 (reactant) to C_{7AX} (product) characterized by the rotation of the two dihedral angles ϕ and ψ .

B3LYP/6-31G* string calculation that required an additional 216 QM gradient calculations. The total time to perform the transition state search calculation, including the time to calculate the initial force constants, was 35 min. The total CPU time to determine the transition state for alanine dipeptide isomerization using the hybrid strategy was 32 h.

Using the energy-weighted strategy, the HF/STO-3G string required 431 QM gradients. This is 107 QM gradients more than was required for the hybrid strategy because of the additional convergence criteria on the three nodes closest to the transition state estimate. However, since a single QM gradient calculation is so rapid using HF/STO-3G, the time to perform these additional calculations was negligible. The B3LYP/6-31G* string required 128 QM gradients, which is 88 fewer gradients than was required for the hybrid strategy. Thus, performing 107 more low-level QM gradient calculations, for which the time per gradient was minimal, led to a more accurate initial string for the high-level refinement and also reduced the number of high-level QM calculations, for which the time is appreciable. The CPU time to determine the optimized transition state was 32 min and the total CPU time for the energy-weighted strategy was 20 h. The large reduction in computational time using the energy-weighted strategy was due to the need to calculate 41% fewer B3LYP/6-31G* gradients relative to what was required for the hybrid strategy.

In implementing the substrating strategy, the low-level HF/STO-3G string required the same number of QM gradients for convergence as that required for the energy-weighted

strategy, 431. A five-point substrating was refined at B3LYP/6-31G* and required 93 QM gradients, which is 123 fewer QM gradients than were required for the hybrid strategy and 35 fewer QM gradients than were required for the energy-weighted strategy at the same level of theory. The large reduction in the number of QM gradients is a consequence of using only those nodes closest to the transition state. The transition state search calculation required 32 min and 14 optimization cycles. The total CPU time using the five-point substrating strategy was 16 h.

A single-node substrating calculation was initiated using the geometry corresponding to the highest energy in the low-level HF/STO-3G string. The transition state search calculation did not converge when performed at B3LYP/6-31G*. However, repeating the low-level string calculation at HF/6-31G* (of higher fidelity than HF/STO-3G), the single-node substrating calculation was able to determine the correct transition state. The low-level HF/6-31G* string required 249 QM gradients, each taking ~ 4 min for a total time to grow the string at 17 h. The transition state search at B3LYP/6-31G* required 62 optimization cycles taking 121 min. The time to perform this transition state search was four times longer than the time required for the five-point substrating strategy because of the poor quality of the initial transition state estimate from the HF/6-31G* string. The transition state determined using the single-node substrating is very similar to that determined from the five-point substrating: the activation energy differs by less than 0.1 kcal/mol and the imaginary frequency differs by 1.5 cm^{-1} . The total CPU time for determining the final transition state using the single-node substrating strategy was 19 h, which is 3 h more than the time required using the five-point substrating strategy.

The number of QM gradients and the total CPU time needed for each transition state-finding strategy are summarized in Table I. These strategies are also compared to the original GSM and the modified-GSM. As can be seen, the total CPU time decreased from 129 h using the original GSM to 15 h using the modified-GSM and five-point substrating strategy; however the CPU time for the transition state opti-

TABLE I. Time required to determine the TS for alanine dipeptide isomerization using different calculational methods and TS-finding strategies. All TS estimates were further optimized at B3LYP/6-31G* to determine the final TS.

Calculation method and strategy	Functional/basis set used to grow string	No. of QM gradients	CPU time/gradient	CPU time to grow string (h) ^a	CPU time for TS search (min)	Total CPU time (h) ^b
GSM ^c	B3LYP/6-31G*	980	~ 8 min	129	35	130
Modified-GSM ^d	B3LYP/6-31G*	339	~ 8 min	45	36	46
Modified-GSM hybrid strategy ^d	HF/STO-3G	324	~ 10 – 15 s	31	35	32
	B3LYP/6-31G*	216	~ 8 min			
Modified-GSM energy-weighted strategy	HF/STO-3G	431	~ 10 – 15 s	19	32	20
	B3LYP/6-31G*	128	~ 8 min			
Modified-GSM substrating strategy	HF/STO-3G	431	~ 10 – 15 s	15	32	16
	B3LYP/6-31G*	93	~ 8 min			

^aCPU time to grow string=number of QM gradients \times CPU time/gradient.

^bTotal CPU time=CPU time to grow string+CPU time to perform TS search.

^cReference 7.

^dReference 17.

TABLE II. Comparison of the final TS geometries for alanine dipeptide isomerization using different calculational methods and TS-finding strategies. See Table I for the functional and basis sets used to grow the string. All TS estimates were refined at B3LYP/6-31G* to determine the final TS. The energies of the optimized reactant, product, and TS structures were determined from single-point calculations at B3LYP/6-311++G** and zero-point corrected.

Calculation method and strategy	ϕ (deg)	ψ (deg)	ν_{img} (cm^{-1})	ΔE^\ddagger (kcal/mol)
GSM ^a	107.3	-139.7	-34.3	7.6
Modified-GSM ^b	105.8	-150.9	-34.9	7.2
Modified-GSM hybrid strategy ^b	99.8	-165.1	-31.5	6.3
Modified-GSM energy-weighted strategy	107.1	-153.2	-34.7	7.8
Modified-GSM substring strategy	118.3	-133.4	-34.7	7.5
Perczel <i>et al.</i> ^c (B3LYP/6-31G*)	114.4	-147.9	...	7.33 (7.20, B3LYP/ 6-311++G**)

^aReference 7.

^bReference 17.

^cReference 32.

mization calculation was reduced by only 3 min. These results suggest that for all calculation methods in Table I, the final transition state estimates from each high-level string are very similar and that the transition state search calculation has a fixed cost of 32–35 min. Thus, in the case of alanine dipeptide isomerization, nearly all of the improvement in total computational time is derived from reductions in the time required to grow the high-level string. Lastly, the final transition states identified in the five cases in Table II are all similar; the imaginary frequencies differ by less than 4 cm^{-1} , the activation energies differ by 1.5 kcal/mol, and each dihedral angle differs by 20°. The geometries of the calculated transition states are also in good agreement with the previous theoretical work of Perczel *et al.*³²

It is important to note that while the variances are large in the two dihedral angles (ϕ and ψ) for the converged transition states using different transition state-finding methods/strategies in Table II, these variances are not uncommon for the reaction of C_5 to $C_{7\text{AX}}$ for two important reasons. First, a contour plot of the potential energy surface as a function of ϕ and ψ (referred to as the Ramachandran surface) reveals that the surface is relatively flat at the transition state joining the reactant C_5 and the product $C_{7\text{AX}}$. This observation is supported by the previous work of Perczel *et al.*,³² which showed that ϕ can vary by 15° and ψ can vary by 30° at this transition state. Second, strict convergence criteria were not used to determine the transition state structures in Table II. If the convergence criteria are increased and a finer integration grid is used, then the variance in ϕ and ψ decreases and the results are more in line with the values determined by Perczel *et al.*,³² particularly for the modified-GSM/hybrid strategy and the modified-GSM/substring strategy. Also, while the time to perform a transition state search using a more strict convergence criteria took longer, the same ordering of strategies shown in Table I is observed with respect to the CPU time for the transition state optimization calculation and the total CPU time required.

B. H-abstraction from methoxide groups associated with isolated vanadate sites supported on silica

The next example is related to the oxidation of methanol on isolated vanadate sites supported on silica, VO_x/SiO_2 . Previous experimental^{33,34} and theoretical^{35,36} works have shown that the rate-limiting step in this process is the transfer of one of the three H atoms of the adsorbed methoxy group to the vanadyl O, as illustrated in Fig. 3. This reaction represents a harder challenge for transition state-finding methods relative to the previous example because of the large number of atoms, including the transition metal V, leading to longer computational times both to grow the string and to perform the transition state search calculation.

Using the hybrid strategy, the low-level HF/STO-3G string converged in 330 QM gradients with each QM gradient calculation taking 4–6 min. The high-level B3LYP/6-31G* string, initiated using the final low-level string, required 151 QM gradients with each gradient taking 80–120 min. The transition state search calculation required 114 h, including the time to compute the initial analytic Hessian. The total CPU time to determine the transition state using the hybrid strategy was 393 h. Without using a transition state-finding strategy, the modified-GSM required 134 h for optimizing the transition state with a total CPU time of 700 h. While only 20 h is saved during the transition state search calculation using the hybrid strategy, almost 300 h is saved during the growth and evolution of the string.

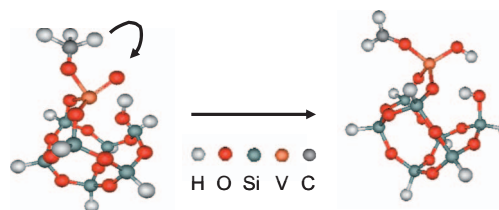


FIG. 3. (Color) H-abstraction in methanol oxidation on VO_x/SiO_2 involving one of the three methoxy H atoms transferring to the vanadyl O atom.

TABLE III. Time required to determine the TS for H-transfer in methanol oxidation on VO_x/SiO₂ using different calculational methods and TS-finding strategies. All TS estimates were further optimized at B3LYP/6-31G*/LANL2DZ to determine the final TS. The LANL2DZ basis set was used for V.

Calculation method and strategy	Functional/basis set used to grow string	No. of QM gradients	CPU time/gradient (min)	CPU time to grow string (h) ^a	CPU time for TS search (h)	Total CPU time (h) ^b
GSM ^c	B3LYP/6-31G*	651	~80–120	1085	140	1225
Modified-GSM ^d	B3LYP/6-31G*	339	~80–120	566	134	700
Modified-GSM hybrid strategy ^d	HF/STO-3G	330	~4–6	279	114	393
	B3LYP/6-31G*	151	~80–120			
Modified-GSM energy-weighted strategy	HF/STO-3G	653	~4–6	259	75	334
	B3LYP/6-31G*	123	~80–120			
Modified-GSM substrating strategy	HF/STO-3G	653	~4–6	180	76	256
	B3LYP/6-31G*	75	~80–120			

^aCPU time to grow string=number of QM gradients×CPU time/gradient.

^bTotal CPU time=CPU time to grow string+CPU time to perform TS search.

^cReference 36.

^dReference 17.

Using the energy-weighted strategy, the low-level HF/STO-3G string required 653 QM gradients and the refined high-level B3LYP/6-31G* string required 123 QM gradients. Despite requiring 323 additional HF/STO-3G gradients than the hybrid strategy, a much better estimate of the transition state geometry was obtained. The transition state estimate converged to the final transition state in 75 h, representing a 34% reduction in time spent on the transition state search calculation alone, as compared with the hybrid strategy. The total CPU time to determine the optimized transition state was 334 h using the energy-weighted strategy.

Using the substrating strategy, the low-level HF/STO-3G string required the same number of QM gradients for convergence as the energy-weighted strategy, 653. A five-point substrating was then refined at B3LYP/6-31G* using nodes three through seven in the final HF/STO-3G string. The B3LYP/6-31G* string required 75 QM gradients, representing 76 fewer QM gradients than using the hybrid strategy. The large reduction in the number of B3LYP/6-31G* QM gradients required compared to that for the hybrid strategy is due only to those nodes closest to the transition state that are refined. In other words, performing more QM gradient calculations by refining all nodes in a high-level string calculation using the hybrid strategy does not lead to a better transition state estimate; rather, the location of the nodes that are refined is what matters. Similar to the energy-weighted strategy, the transition state search calculation required 76 h. The total CPU time using the substrating strategy was only 256 h, a 35% reduction in computational time compared to the hybrid strategy.

The number of QM gradients and the total CPU time needed for each transition state-finding strategy are listed in Table III. These strategies are also compared against the original GSM and the modified-GSM. A large part of the improvement that comes from using the transition state-finding strategies described here is a consequence of the reduced computational time needed to grow the string, similar to the case of alanine dipeptide isomerization. However, there is also a marked improvement in generating a more

accurate transition state estimate because of the reduced computational time for optimizing the transition state, particularly in the case of the energy-weighted and substrating strategies. As can be noted, the total CPU time is decreased from 1225 h using the GSM to 256 h using the modified-GSM together with the five-point substrating strategy. The CPU time for optimizing the transition state is also reduced from 140 to 76 h using these two methods.

The final transition states from each calculation method are identical. Table IV compares the geometry, activation energy, and vibrational frequency corresponding to a hydrogen atom transferring from the adsorbed methoxy group to the vanadyl O for each of the five calculational approaches. A very good agreement in the results is seen for all the methods. For these five methods, the C–H bond length varies by 0.02 Å, the O–H bond length varies by 0.01 Å, the imaginary frequency associated with the transition state differs by 1.6 cm⁻¹, and the activation energy differs by 0.6 kcal/mol.

C. C–H bond activation in oxidative carbonylation of toluene to *p*-toluic acid on Rh(CO)₂(TFA)₃

The third example concerns the oxidative carbonylation of toluene to *p*-toluic acid. The reaction is catalyzed by Rh complexes containing two carbonyl groups and three trifluoroacetate (TFA) ligands.³⁷ Previous work by Zheng and Bell³⁸ showed that the rate-limiting step involves activating the C–H bond in toluene, as seen in Fig. 4. Determining the transition state for this reaction is particularly challenging because there are 41 atoms and, hence, a very large number of degrees of freedom. A major difficulty in finding the correct transition state is associated with the TFA ligands, which undergo rotation during the transition from the reactant to the product configuration. To overcome this challenge, Zheng and Bell³⁸ used the GSM in combination with an iterative partial optimization approach (IPOA). The IPOA involved freezing and unfreezing Cartesian coordinates associated with a user-defined guess of the transition state. While it was

TABLE IV. Comparison of the final TS geometries for H-abstraction in methanol oxidation on VO_x/SiO_2 using different calculational methods and TS-finding strategies. See Table III for the functional and basis sets used to grow the string. All TS estimates were refined at B3LYP/6-31G*/LANL2DZ to determine the final TS. The energies of the optimized reactant, product, and TS structures were determined from single-point calculations at B3LYP/6-311++G**/LANL2DZ and zero-point corrected.

Calculation method and strategy	$R_{\text{C-H}}$ (Å)	$R_{\text{O-H}}$ (Å)	ν_{img} (cm^{-1})	ΔE^\ddagger (kcal/mol)
GSM ^a	1.63	1.11	-870.9	40.3
Modified-GSM ^b	1.63	1.11	-870.9	40.8
Modified-GSM hybrid strategy ^b	1.63	1.11	-871.9	40.8
Modified-GSM energy-weighted strategy	1.64	1.11	-871.4	40.7
Modified-GSM substring strategy	1.62	1.12	-870.3	40.9

^aReference 36.

^bReference 17.

possible to identify the correct transition state using the IPOA, it required extensive user intervention and over 1800 h of computational time.

Goodrow *et al.*¹⁷ showed that neither the GSM nor the modified-GSM is able to converge to a transition state structure.¹⁷ However, these authors did show that the hybrid strategy could be used to identify the transition state successfully. The low-level HF/STO-3G string converged in 738 QM gradients. The string was then refined using 114 QM gradients at the B3LYP/6-31G* level. A single HF/STO-3G QM gradient calculation required 9–10 min, whereas a B3LYP/6-31G* QM gradient calculation required 100–120 min. The transition state search calculation required 228 h with a total CPU time of 554 h to converge to the final transition state. Thus, the hybrid strategy used in combination with the modified-GSM results in a substantial reduction in computational time. For example, performing a modified-GSM calculation using B3LYP/6-31G* in the absence of a transition state-finding strategy required over 1000 QM gradients, after which point the calculation was terminated because it had not converged.

Using the energy-weighted strategy, the HF/STO-3G string required 921 QM gradients, representing 183 more gradients than in the hybrid method. However, only 96 QM gradients were needed to refine the high-level B3LYP/6-31G* string. As in the previous example of methanol oxidation on VO_x/SiO_2 , a much better estimate of the transition state was obtained using the energy-weighted strategy compared to the hybrid strategy. The transition state

search calculation converged in 162 h after 92 SCF cycles and the total CPU time required was 484 h. The energy-weighted strategy and the hybrid strategy both required a very similar time to grow the string (322 h compared to 326 h, respectively). Despite the similar time for growing the string, the energy-weighted strategy was able to generate a much better estimate of the transition state and took 29% less time to converge the transition state estimate to the final transition state.

Using the substring strategy, the HF/STO-3G string required the same number of QM gradients for convergence as the energy-weighted strategy, 921. A five-point substring was then refined at B3LYP/6-31G* using nodes four through eight in the final HF/STO-3G string. The B3LYP/6-31G* string required 84 QM gradient calculations. On a per non-frozen node basis (the reactant and product end points of the string are frozen for which no QM gradients are calculated), more QM calculations are required using the substring strategy than the energy-weighted strategy, which leads to a better estimate of the transition state. Using the substring strategy, the transition state search required 112 h and 65 SCF cycles and a total CPU time of 412 h. The 72 h improvement in total CPU time over the energy-weighted strategy comes from saving 22 h to grow the string and 50 h to optimize the transition state.

It is also necessary to point out the importance of using the increased convergence criteria of $\|f_{\perp}\| < 0.001$ hartree/bohr for the three points with the highest energy along the reaction pathway in both the energy-weighted and substring strategies. If this parameter is reduced to $\|f_{\perp}\| < 0.02$ hartree/bohr, as it is for all other points, then the transition state estimate obtained from the converged B3LYP/6-31G* string does not converge to the correct transition state. Calculation of the initial analytic Hessian for the transition state estimate obtained using the reduced convergence criteria exhibits two negative Hessian eigenvalues that are close in magnitude ($\lambda_1 = -0.00067$, $\lambda_2 = -0.00022$). One of these Hessian eigenvalues ($\lambda_2 = -0.00022$) corresponds to the motion of the H atom in the C–H bond which is being broken. The other eigenvalue ($\lambda_1 = -0.00067$) corresponds to the rotation of a F atom in one of the TFA ligands. During a transition state search calculation in QCHEM 3.0, the eigenvector-following scheme was

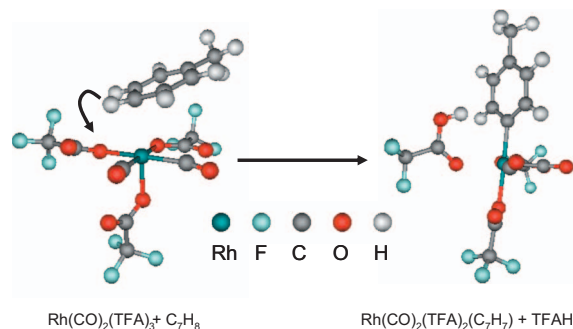


FIG. 4. (Color) C–H bond activation in the oxidative carbonylation of toluene to *p*-toluic acid on a Rh complex coordinated with two CO species and three TFA solvent ligands.

TABLE V. Time required to determine the TS for C–H bond activation in toluene on $\text{Rh}(\text{CO})_2(\text{TFA})_3$ using different calculational methods and TS-finding strategies. All TS estimates were further optimized at B3LYP/6-31G*/LANL2DZ to determine the final TS. The LANL2DZ basis set was used for Rh.

Calculation method and strategy	Functional/basis set used to grow string	No. of QM gradients	CPU time/gradient (min)	CPU time to grow string (h) ^a	CPU time for TS search (h)	Total CPU time (h) ^b
GSM ^c	B3LYP/6-31G*	>1000	~100–120	>1800 did not converge
Modified-GSM ^d	B3LYP/6-31G*	>1000	~100–120	>1800 did not converge
Modified-GSM hybrid strategy ^d	HF/STO-3G	738	~9–10	326	228	554
	B3LYP/6-31G*	114	~100–120			
Modified-GSM energy-weighted strategy	HF/STO-3G	921	~9–10	322	162	484
	B3LYP/6-31G*	96	~100–120			
Modified-GSM substrating strategy	HF/STO-3G	921	~9–10	300	112	412
	B3LYP/6-31G*	84	~100–120			

^aCPU time to grow string=number of QM gradients×CPU time/gradient.

^bTotal CPU time=CPU time to grow string+CPU time to perform TS search.

^cReference 38.

^dReference 17.

used, which maximizes the potential energy along the lowest Hessian mode (the eigenvector corresponding to the most negative eigenvalue) while simultaneously minimizing the potential energy along all other modes. Since λ_1 is more negative than λ_2 , the calculation proceeds to an incorrect transition state. Further optimizing the B3LYP/6-31G* string in the energy-weighted and substrating strategies with $\|f_{\perp}\| < 0.001$ hartree/bohr for the three points with the highest energy results in one negative Hessian eigenvalue in the initial analytic Hessian ($\lambda_1 = -0.00206$) and leads to the correct transition state.

The number of QM gradients and total CPU time needed for each transition state-finding strategy are summarized in Table V. The results obtained using these strategies are also compared to those obtained using the original GSM and the modified-GSM, which were terminated after failing to converge to a transition state after calculation of 1000 QM B3LYP/6-31G* gradients. As can be seen from Table V, the transition state could only be determined using one of the three strategies. Compared to the hybrid strategy, the other two strategies spent little time converging nodes far from the

transition state, especially nodes that involved rotation of the TFA ligands. The energy-weighted and substrating strategies are particularly effective because they generate a very good estimate of the transition state structure by focusing on the highest energy portion of the MEP. For instance, the time required to optimize the transition state decreased by 116 h moving from the hybrid strategy to the substrating strategy, while the time to grow the string decreased only 26 h. Thus, in the case of the oxidative carbonylation of toluene to *p*-toluic acid, the primary improvement of the energy-weighted and substrating strategies comes from obtaining a much better transition state estimate rather than from a reduction in the CPU time needed to grow the string. Lastly, the geometry and activation energy of the final transition state obtained using these three strategies are in good agreement with the transition state determined by Zheng and Bell,³⁸ as can be seen in Table VI. There is some slight disagreement between the activation energies, which vary by 0.9 kcal/mol. If the convergence criteria are increased and a finer integration grid is used, as previously done in the case of alanine dipeptide isomerization, then the variance in the

TABLE VI. Comparison of the final TS geometries for C–H bond activation in toluene on $\text{Rh}(\text{CO})_2(\text{TFA})_3$ using different calculational methods and TS-finding strategies. See Table V for the functional and basis sets used to grow the string. All TS estimates were refined at B3LYP/6-31G*/LANL2DZ to determine the final TS. The energies of the optimized reactant, product, and TS structures were determined from single-point calculations at B3LYP/6-311++G**/LANL2DZ and zero-point corrected.

Calculation method and strategy	$R_{\text{C-H}}$ (Å)	$R_{\text{O-H}}$ (Å)	$R_{\text{Rh-C}}$ (Å)	$R_{\text{Rh-O}}$ (Å)	ΔE^{\ddagger} (kcal/mol)
IPOA with GSM ^a	1.30	1.35	2.20	2.11	9.1
Modified-GSM ^{b,c}
Modified-GSM hybrid strategy ^b	1.28	1.38	2.18	2.11	9.4
Modified-GSM energy-weighted strategy	1.30	1.35	2.21	2.10	9.2
Modified-GSM substrating strategy	1.28	1.38	2.18	2.11	10.0

^aReference 38.

^bReference 17.

^cCalculation did not converge to a transition state.

activation energies is reduced from 0.9 to 0.2 kcal/mol. In addition, the root mean square differences in the transition state geometries for each approach vary by only 0.013 Å.

The three examples discussed above demonstrate that the computational time required for the modified-GSM can be improved significantly by application of the three strategies developed in this study. The hybrid strategy is able to reduce the computational time required for growing the string and the energy-weighted strategy is able to obtain a much better estimate of the transition state, thereby reducing the time required for the transition state search calculation. The substring strategy combines the advantages of both of these methods, and in all three previous examples it resulted in the lowest total CPU time required for convergence to a transition state. It should be noted, however, that all of the previous examples involve only a small number of degrees of freedom to define the reaction coordinate. For instance, alanine dipeptide isomerization involves rotation of the two dihedral angles ϕ and ψ , H-abstraction during methanol oxidation on VO_x/SiO_2 catalysts involves a single H atom that moves from the bound methoxy group to the vanadyl oxygen, and the oxidative carbonylation of toluene to *p*-toluic acid on $\text{Rh}(\text{CO})_2(\text{TFA})_3$ catalysts involves a single H atom that transfers from toluene to a TFA ligand, albeit, in this case, significant rotation of the atoms in the three TFA ligands also occurs.

To test the applicability of the modified-GSM/substring strategy to determine transition states for more complex reactions in which the reaction coordinate is difficult to determine, the following three additional examples were considered: isomerization of cyclopropane to propylene, isomerization of methylcyclopropane to four isomers of butene, and Diels–Alder condensation of 1,3-butadiene and ethylene to form cyclohexene. For each example, the transition state was determined and the activation energy and rate constant were compared with previous experimental and theoretical work. All of the string calculations for these examples were performed using 15 nodes. The series of geometries along the converged MEP and the transition state are provided in the supplementary materials for each example.³⁹

D. Intramolecular rearrangement of cyclopropane to propylene

There are three types of cyclopropane isomerization reactions: geometrical, optical, and structural.⁴⁰ The geometrical and optical isomerization reactions involve breaking a C–C bond in the cyclopropane ring to form trimethylene. The structural isomerization reaction is more complex and involves both cleaving a C–C bond and abstracting a hydrogen atom to form propylene. The geometries of the reactant, transition state, and product for the structural isomerization reaction are shown in Fig. 5 as well as the converged B3LYP/6-31G* string obtained from a modified-GSM calculation.

The results of using the substring strategy with the modified-GSM for the structural isomerization of cyclopropane are shown in Table VII. The calculated enthalpy of

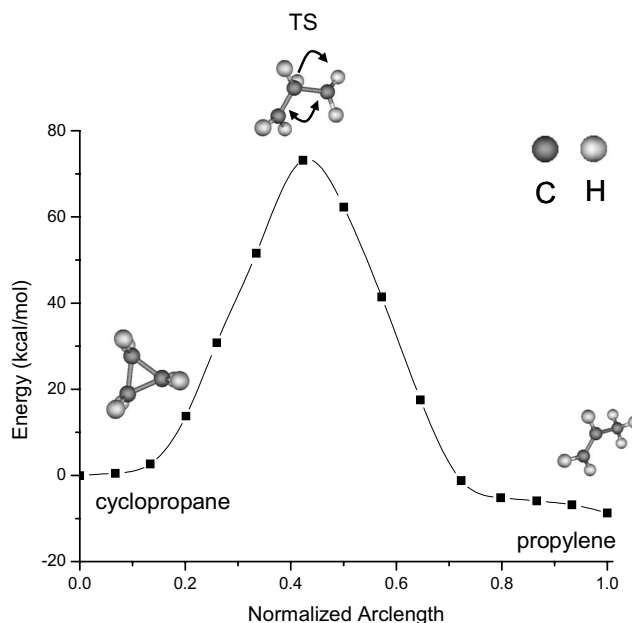


FIG. 5. Intramolecular rearrangement of cyclopropane to propylene. The optimized geometries of reactant, transition state, and product are shown along with the final string.

reaction for the standard state ($T=298$ K, $P=1$ atm), $\Delta H^\circ = -8.09$ kcal/mol, is in very good agreement with the experimental value⁴¹ of -7.96 kcal/mol. The activation energy of the converged B3LYP/6-311++G** transition state is $\Delta E^\ddagger = 63.3$ kcal/mol, which agrees well with the experimental work of Rabinovitch *et al.*⁴² ($\Delta E^\ddagger = 65.5$ kcal/mol) and others^{43–45} (62–67 kcal/mol) as well as with previous theoretical work by Dubnikova and Lifshitz⁴⁰ (63.6–65.1 kcal/mol, depending on the calculational method). Lastly, the pre-exponential factor $k^\circ = 10^{14.5}$ s⁻¹ and the first-order rate constant at 773 K, $k_{773\text{ K}} = 1.43 \times 10^{-4}$ s⁻¹, are also in good agreement with previous experimental⁴² and theoretical⁴⁰ values. Thus, the substring strategy is effective in this case for a reaction that involves concerted ring opening and H transfer.

TABLE VII. Comparison between experimental and theoretical values obtained in the present work for the energetics and kinetics of cyclopropane isomerization to propylene. The reaction enthalpy at standard state (ΔH°), activation energy (ΔE^\ddagger), preexponential factor (k°), and rate constant at 773 K ($k_{773\text{ K}}$) are shown. The energies of the optimized reactant, product, and TS structures were determined from single-point calculations at B3LYP/6-311++G** and zero-point corrected.

Parameter	Experiment	Dubnikova <i>et al.</i> ^a	Present work
ΔH° (kcal/mol)	-7.96^b	...	-8.09
ΔE^\ddagger (kcal/mol)	65.5^c $62\text{--}67^d$	63.58 (uB3LYP) 65.08 (uQCISDT)	63.3
k° (s ⁻¹)	$10^{15.2^c}$	$10^{15.91}$ (uB3LYP)	$10^{14.5}$
$k_{773\text{ K}}$ (s ⁻¹)	4.78×10^{-4} ^c	1.12×10^{-3} (uB3LYP)	1.43×10^{-4}

^aReference 40.

^bReference 41.

^cReference 42.

^dReferences 43–45.

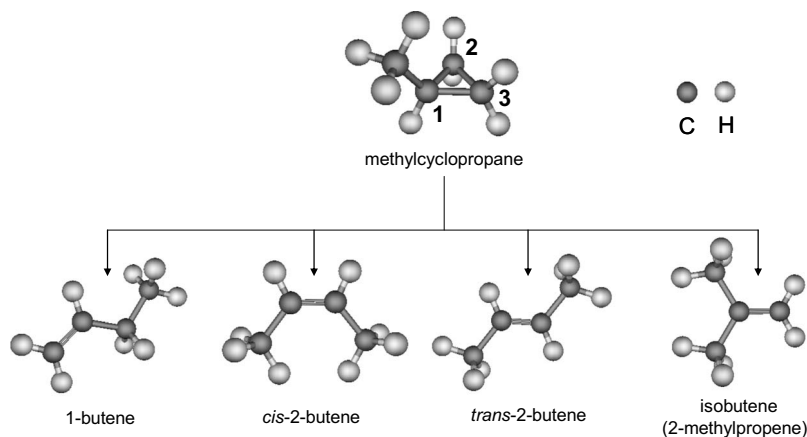


FIG. 6. The rearrangement of methylcyclopropane to form the stereoisomers 1-butene, *cis*-2-butene, *trans*-2-butene, and isobutene. The C atoms in the ring of methylcyclopropane are numbered.

E. Rearrangement of methylcyclopropane to four stereoisomers

The rearrangement of methylcyclopropane is a challenging reaction system because four stereoisomers are formed: 1-butene, *cis*-2-butene, *trans*-2-butene, and isobutene (2-methylpropene), as shown in Fig. 6. The four isomerizations are parallel, first-order reactions,⁴⁶ and proceed along different reaction pathways.

The substrating strategy was used with the modified-GSM to determine the transition state for each isomerization process. The calculated reaction enthalpy (ΔH°), the activation energy (ΔE^\ddagger), and the relative reaction rate at 698 K ($k_{698\text{ K}}$) for all four methylcyclopropane isomerization reactions are compared with previous experimental work^{46–48} in Table VIII. The experimental values of ΔH° are tabulated for liquid-phase species,⁴¹ whereas the theoretically calculated values are for gas-phase species. The standard state heat of formation for methylcyclopropane in the gas phase has not been measured because it is very reactive. Nevertheless, the values of ΔH° obtained from theory and experiment are in very good agreement, as is the trend in ΔH° , with isobutene being the most energetically favored thermodynamic product.

The activation energies for the converged

B3LYP/6-311++G** transition states leading to each of the products differ by less than 4 kcal/mol compared to the experimental values reported by Kalra *et al.*⁴⁸ and Chesick.⁴⁶ The electronic structures of the transition states leading to 1-butene, *cis*-2-butene, and isobutene each have a singlet spin state. A stability analysis on the transition state leading to *trans*-2-butene revealed the triplet spin state to be a more stable electronic structure than the singlet spin state. It is also important to note that the products that are more energetically favorable based on ΔH° do not necessarily have lower activation energies than the products that are less energetically favorable. For instance, while the calculated and experimental values of ΔH° for *trans*-2-butene are lower, and hence more favorable than that for *cis*-2-butene, ΔE^\ddagger for *trans*-2-butene is higher.

The rate constant at 698 K for each isomerization reaction was calculated and then multiplied by the number of different ways in which that particular reaction could occur, referred to as the symmetry factor.⁴⁹ For example, the rearrangement of methylcyclopropane to 1-butene can be achieved in four different ways. If the C₁–C₂ bond is broken, as shown in Fig. 6, then one of two equivalent H atoms on C₃ can be transferred to C₁. Likewise, if the C₁–C₃ bond

TABLE VIII. Comparison between experimental and theoretical values obtained in the present work for ΔH° , ΔE^\ddagger , and the relative reaction rates for the rearrangement of methylcyclopropane for each stereoisomer. The energies of the optimized reactant, product, and TS structures were determined from single-point calculations at B3LYP/6-311++G** and zero-point corrected. All energies are reported in units of kcal/mol.

Product	Experiment $\Delta H^{\circ\text{a,b}}$	Present ΔH°	Experiment ΔE^\ddagger	Present ΔE^\ddagger	Symmetry factor, σ	Experiment relative rate ^c	Present work relative rate
1-butene	-5.38	-6.17	64.5 ± 0.5 ^c	67.3	4	1.0	1.0
<i>cis</i> -2-butene	-7.53	-8.08	63.3 ± 0.3 ^c 63.9 ^d	64.7	4	0.8	0.67
<i>trans</i> -2-butene	-8.37	-9.55	64.9 ± 0.3 ^c 68.1 ^d	68.8	4	0.4	0.24
Isobutene	-9.37	-9.76	66.4 ± 0.2 ^c 66.0 ^d	66.3	2	0.2	0.083

^aReference 41.

^bDue to the instability of methylcyclopropane in the vapor phase, the tabulated values of ΔH° are based on the liquid phase at standard state conditions.

^cReference 48.

^dReference 46.

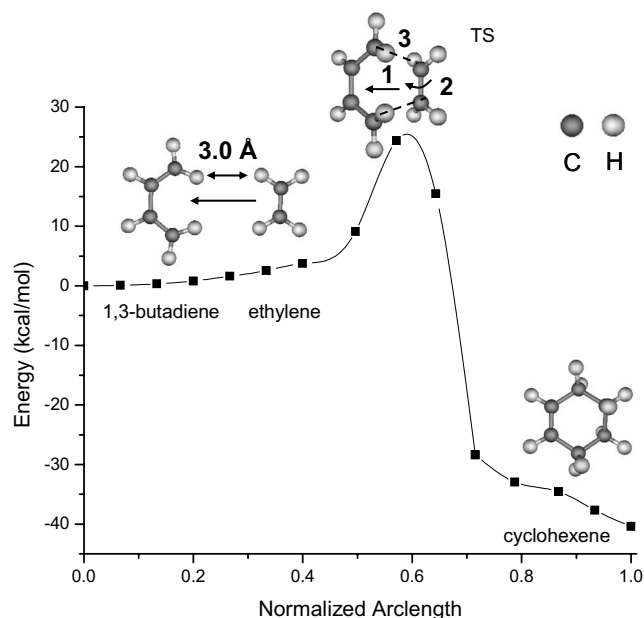


FIG. 7. Bimolecular Diels–Alder reaction of 1,3-butadiene and ethylene to cyclohexene. The optimized geometries of reactant, transition state, and product are shown along with the final string. The two reactant species are located 3.0 Å apart. The transition state consists of the concerted motion of the two reactant species moving toward each other (step 1), the rotation of ethylene (step 2), and the formation of two weak C–C bonds (step 3).

is broken, then one of two equivalent H atoms on C₂ can be transferred to C₃. The symmetry factors for the formation of *cis*-2-butene and *trans*-2-butene are both 4 and were calculated in a manner similar to that of 1-butene. The symmetry factor for the formation of isobutene is two because the C₂–C₃ bond must be broken, followed by the transfer of the H atom on C₁ to either C₂ or C₃. It was also assumed that the activation energy for interconversion of the four different products was higher than the activation barrier for isomerization of methylcyclopropane.

The calculated relative reaction rates at 698 K are in reasonable agreement with experiment,⁴⁸ and the rate of formation of 1-butene is the fastest. In addition, the ratio of *cis*-2-butene to *trans*-2-butene calculated from theory is 2.79, in good agreement with the value of 2.0 determined by Kalra *et al.*⁴⁸ at 698 K and 2.25 determined by Chesick⁴⁶ at 740 K.

F. Bimolecular Diels–Alder reaction

The last reaction examined was the bimolecular Diels–Alder condensation of 1,3-butadiene and ethylene to form cyclohexene. The modified-GSM calculation was initiated by first determining the optimized geometries for 1,3-butadiene, ethylene, and cyclohexene at the B3LYP/6-31G* level of theory. The reactant configuration consisted of the optimized structures of 1,3-butadiene and ethylene with the terminating H atoms on each respective species arranged 3.0 Å apart, as shown in Fig. 7. The converged B3LYP/6-31G* string from a modified-GSM calculation and the transition state is also shown in Fig. 7.

The reaction proceeds in several concerted steps. The separation distance between 1,3-butadiene and ethylene de-

TABLE IX. Comparison between experimental and theoretical values obtained in the present work for the energetics and kinetics of the Diels–Alder reaction of 1,3-butadiene and ethylene to cyclohexene. The reaction enthalpy at standard state (ΔH°), activation energy (ΔE^\ddagger), and rate constant at 823 K ($k_{823\text{ K}}$) are shown. The energies of the optimized reactant, product, and TS structures were determined from single-point calculations at B3LYP/6-311++G** and zero-point corrected.

Parameter	Experiment	VTST ^a	Present work
ΔH° (kcal/mol)	–40.0 ^b	–47.8	–40.1
ΔE^\ddagger (kcal/mol)	25.9 ^c	21.9	21.2
$k_{823\text{ K}}$ (cm ³ mol ^{–1} s ^{–1})	1230–1590 ^c	1734	1169

^aReference 52 (the theoretical rate was matched with the experimental rate by varying the height of the activation barrier for the forward and reverse reactions).

^bReference 41.

^cReference 50.

creases while the two terminating H atoms on 1,3-butadiene are forced out of plane and the ethylene molecule rotates to form two weak C–C bonds. These three steps are illustrated in Fig. 7.

The theoretical results of the reaction energetics and kinetics from using the substring strategy are compared to those from experiment in Table IX. The standard state reaction enthalpy, $\Delta H^\circ = -40.1$ kcal/mol, agrees very well with that determined from tabulated values of the enthalpies of formation for the reactants and products,⁴¹ $\Delta H^\circ = -40.0$ kcal/mol. The activation energy determined from the converged substring transition state, $\Delta E^\ddagger = 21.2$ kcal/mol, is in reasonable agreement with the experimental work by Rowley and Steiner⁵⁰ ($\Delta E^\ddagger = 25.9$ kcal/mol), a theoretical study by Goldstein *et al.*⁵¹ ($\Delta E^\ddagger = 24.8$ kcal/mol), and a theoretical study using variational transition state theory (VTST) by Huang *et al.*⁵² ($\Delta E^\ddagger = 21.9$ kcal/mol). The underestimation of the activation energy is typical when using the B3LYP functional, as done in this work and the work of Goldstein *et al.*⁵¹ The calculated rate constant at 823 K, $k_{823\text{ K}} = 1169$ cm³ mol^{–1} s^{–1}, is also in good agreement with previous experimental⁵⁰ and theoretical work.⁵² Thus, this example illustrates the ability of the substring strategy, when used with the modified-GSM, to determine the reaction pathway and the transition state for bimolecular reactions.

IV. CONCLUSIONS

Three transition state-finding strategies for use with the modified-GSM were developed and applied to a wide range of types of chemical reactions. The hybrid strategy is a valuable way for determining both the MEP and the transition state from a single set (low level/high level) of string calculations. The energy-weighted strategy gives finer resolution of the transition state but can also be used to give finer resolution at areas along the MEP where the energy is changing the most. The substring strategy is effective only if the transition state is desired.

The strategies investigated in the course of this study were able to reduce the CPU time needed for determining a transition state using the modified-GSM in two distinct ways. First, the time needed to grow the string was reduced through

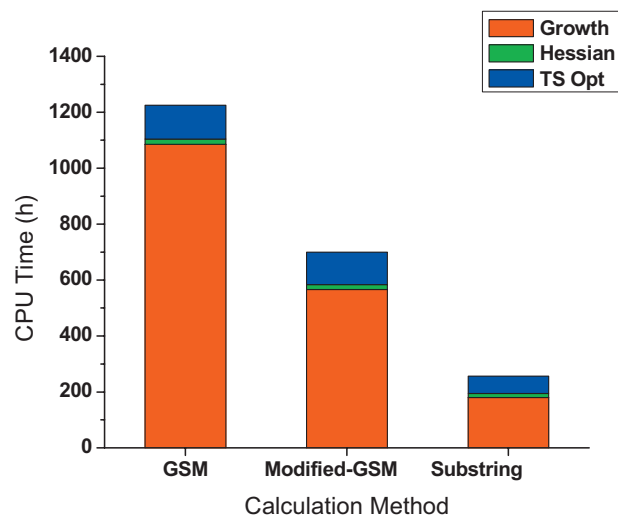


FIG. 8. (Color) Comparison of the distribution of time spent on growth of the string, determination of the Hessian, and optimization of the transition state estimate for the original GSM, the modified-GSM, and the modified-GSM using the substring strategy. The calculated times in hours are shown for H-transfer in methanol oxidation on VO_x/SiO_2 .

the use of low-level QM gradient calculations. Second, the time needed to converge the transition state search calculation was reduced by obtaining a better initial transition state estimate.

Of the three strategies examined, the substring strategy was proved to be the most effective at accurately determining the correct transition state while minimizing the total CPU time. The success of the substring strategy is due to its ability to quickly determine an estimate of the reaction pathway using a low level of theory, followed by refining a portion of the string at a higher level of theory where the transition state is estimated to be located. The strategy is particularly effective in reactions where a large number of atoms are moving, and hence have several very low vibrational modes. A large amount of computational time is reduced because unnecessary time spent in converging nodes along the MEP far from the transition state is omitted. This is most evident in the example of C–H bond activation in toluene on $\text{Rh}(\text{CO})_2(\text{TFA})_3$, where additional time is spent in converging nodes involved in rotation of the TFA ligands in the absence of a transition state-finding strategy. In addition, the substring strategy was applied successfully to reactions in catalysis, stereochemistry, and those involving ring opening and ring formation. The single-node substring (low level: HF/6-31G*, high level: B3LYP/6-31G*) was shown to converge to the correct transition state in the case of alanine dipeptide isomerization but required longer computational time than the five-point substring (low level: HF/STO-3G, high level: B3LYP/6-31G*). This result indicates that a delicate balance exists between accuracy and computational cost for minimizing the time to determine the transition state using the single-node substring strategy.

To summarize the results of the present study, the reduction in the CPU time required for the growth of the string, calculation of the initial Hessian for the transition state estimate, and optimization of the transition state estimate to the final transition state are shown in Fig. 8 for the case of

methanol oxidation on VO_x/SiO_2 . As can be seen, the time required for the growth of the string in the modified-GSM and the modified-GSM/substring strategy has been greatly reduced compared to the original GSM. There is still a fixed cost for performing the transition state search calculation, which involves calculating the initial Hessian and optimizing the transition state estimate. Improvements to the transition state estimate can aid in reducing the time required for the transition state search calculation and represent a productive direction for future work.

ACKNOWLEDGMENTS

This work was supported by the Methane Conversion Cooperative funded by BP.

- ¹D. G. Truhlar, B. C. Garrett, and S. J. Klippenstein, *J. Phys. Chem.* **100**, 12771 (1996).
- ²D. G. Truhlar, *Faraday Discuss.* **110**, 521 (1998).
- ³D. G. Truhlar and K. Morokuma, *Transition State Modeling for Catalysis* (ACS National Conference, Dallas, TX, 1998).
- ⁴F. Jensen, *Introduction to Computational Chemistry* (Wiley, New York, 1999).
- ⁵H. B. Schlegel, *J. Comput. Chem.* **24**, 1514 (2003).
- ⁶A. T. Bell, *Mol. Phys.* **102**, 319 (2004).
- ⁷B. Peters, A. Heyden, A. T. Bell, and A. Chakraborty, *J. Chem. Phys.* **120**, 7877 (2004).
- ⁸A. Heyden, A. T. Bell, and F. J. Keil, *J. Chem. Phys.* **123**, 224101 (2005).
- ⁹C. J. Cerjan and W. H. Miller, *J. Chem. Phys.* **75**, 2800 (1981).
- ¹⁰R. Elber and M. Karplus, *Chem. Phys. Lett.* **139**, 375 (1987).
- ¹¹G. Henkelman, B. P. Uberuaga, and H. Jonsson, *J. Chem. Phys.* **113**, 9901 (2000).
- ¹²G. Henkelman and H. Jonsson, *J. Chem. Phys.* **113**, 9978 (2000).
- ¹³S. A. Trygubenko and D. J. Wales, *J. Chem. Phys.* **120**, 2082 (2004).
- ¹⁴W. N. E, W. Q. Ren, and E. Vanden-Eijnden, *Phys. Rev. B* **66**, 052301 (2002).
- ¹⁵S. K. Burger and W. T. Yang, *J. Chem. Phys.* **124**, 054109 (2006).
- ¹⁶S. K. Burger and W. Yang, *J. Chem. Phys.* **127**, 164107 (2007).
- ¹⁷A. Goodrow, A. T. Bell, and M. Head-Gordon, *J. Chem. Phys.* **129**, 174109 (2008).
- ¹⁸D. Sheppard, R. Terrell, and G. Henkelman, *J. Chem. Phys.* **128**, 134106 (2008).
- ¹⁹J. Ischtwan and M. A. Collins, *J. Chem. Phys.* **100**, 8080 (1994).
- ²⁰K. C. Thompson, M. J. T. Jordan, and M. A. Collins, *J. Chem. Phys.* **108**, 564 (1998).
- ²¹K. C. Thompson, M. J. T. Jordan, and M. A. Collins, *J. Chem. Phys.* **108**, 8302 (1998).
- ²²M. A. Collins, *Theor. Chem. Acc.* **108**, 313 (2002).
- ²³G. E. Moyano and M. A. Collins, *J. Chem. Phys.* **121**, 9769 (2004).
- ²⁴G. E. Moyano and M. A. Collins, *Theor. Chem. Acc.* **113**, 225 (2005).
- ²⁵W. N. E, W. Q. Ren, and E. Vanden-Eijnden, *J. Chem. Phys.* **126**, 164103 (2007).
- ²⁶A. Aguilar-Mogas, X. Giménez, and J. M. Bofill, *J. Chem. Phys.* **128**, 104102 (2008).
- ²⁷K. Müller and L. D. Brown, *Theor. Chim. Acta* **53**, 75 (1979).
- ²⁸D. K. Malick, G. A. Petersson, and J. A. Montgomery, *J. Chem. Phys.* **108**, 5704 (1998).
- ²⁹Y. Shao, L. F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. DiStasio, R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M. Herbert, C. Y. Lin, T. Van Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C. P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, C. D. Sherrill, A. C. Simmonett, J. E. Subotnik, H. L. Woodcock, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer, J. Kong, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon,

- Phys. Chem. Chem. Phys.* **8**, 3172 (2006).
- ³⁰ J. W. Chu, B. L. Trout, and B. R. Brooks, *J. Chem. Phys.* **119**, 12708 (2003).
- ³¹ W. Quapp, *J. Comput. Chem.* **28**, 1834 (2007).
- ³² A. Perczel, O. Farkas, I. Jakli, I. A. Topol, and I. G. Csizmadia, *J. Comput. Chem.* **24**, 1026 (2003).
- ³³ L. J. Burcham, M. Badlani, and I. E. Wachs, *J. Catal.* **203**, 104 (2001).
- ³⁴ J. L. Bronkema and A. T. Bell, *J. Phys. Chem. C* **111**, 420 (2007).
- ³⁵ J. Döbler, M. Pritzsche, and J. Sauer, *J. Am. Chem. Soc.* **127**, 10861 (2005).
- ³⁶ A. Goodrow and A. T. Bell, *J. Phys. Chem. C* **111**, 14753 (2007).
- ³⁷ J. J. Zakzeski and A. T. Bell, *J. Mol. Catal. A: Chem.* **276**, 8 (2007).
- ³⁸ X. B. Zheng and A. T. Bell, *J. Phys. Chem. C* **112**, 2129 (2008).
- ³⁹ See EPAPS supplementary material at <http://dx.doi.org/10.1063/1.3156312> for the series of geometries along the MEP and the structure of the final transition state in XYZ format.
- ⁴⁰ F. Dubnikova and A. Lifshitz, *J. Phys. Chem. A* **102**, 3299 (1998).
- ⁴¹ *CRC Handbook of Chemistry and Physics*, 89th ed., edited by D. R. Lide (CRC/Taylor & Francis, Boca Raton, FL, 2009).
- ⁴² B. S. Rabinovitch, E. W. Schlag, and K. B. Wiberg, *J. Chem. Phys.* **28**, 504 (1958).
- ⁴³ J. Langrish and H. O. Pritchard, *J. Chem. Phys.* **62**, 761 (1958).
- ⁴⁴ E. W. Schlag and B. S. Rabinovitch, *J. Am. Chem. Soc.* **82**, 5996 (1960).
- ⁴⁵ E. V. Waage and B. S. Rabinovitch, *J. Phys. Chem.* **76**, 1695 (1972).
- ⁴⁶ J. P. Chesick, *J. Am. Chem. Soc.* **82**, 3277 (1960).
- ⁴⁷ S. Y. Ho, *J. Chin. Chem. Soc. (Taipei)* **16**, 11 (1969).
- ⁴⁸ B. L. Kalra, J. Y. Cho, and D. K. Lewis, *J. Phys. Chem. A* **103**, 362 (1999).
- ⁴⁹ S. W. Benson, *Thermochemical Kinetics*, 2nd ed. (Wiley, New York, 1976).
- ⁵⁰ D. Rowley and H. Steiner, *Discuss. Faraday Soc.* **10**, 198 (1951).
- ⁵¹ E. Goldstein, B. Beno, and K. N. Houk, *J. Am. Chem. Soc.* **118**, 6036 (1996).
- ⁵² C. H. Huang, L. C. Tsai, and W. P. Hu, *J. Phys. Chem. A* **105**, 9945 (2001).