

Facilitating Transition State Search with Minimal Conformational Sampling Using Reaction Graph

Kyunghoon Lee,[†] Jinwon Lee,[†] Shinyoung Park,[†] and Woo Youn Kim*Cite This: *J. Chem. Theory Comput.* 2025, 21, 2487–2500

Read Online

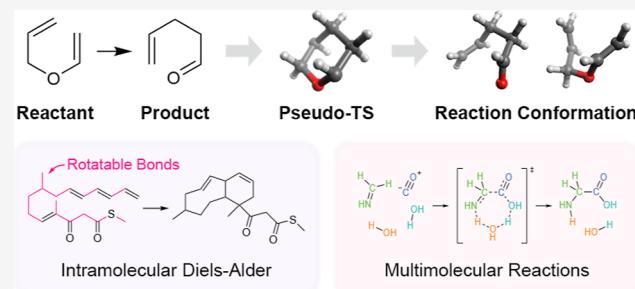
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Elucidating transition states (TSs) is crucial for understanding chemical reactions. The reliability of traditional TS search approaches depends on input conformations that require significant effort to prepare. Previous automated methods for generating input reaction conformations typically involve extensive exploration of a large conformational space. Such exhaustive search can be complicated by the rapid growth of the conformational space, especially for reactions involving many rotatable bonds, multiple reacting molecules, and numerous bond formations and dissociations. To address this problem, we propose a new approach that generates reaction conformations for TS searches with minimal reliance on sampling. This method constructs a pseudo-TS structure based on a reaction graph containing bond formation and dissociation information and modifies it to produce reactant and product conformations. Tested on three different benchmarks, our method consistently generated suitable conformations without necessitating extensive sampling, demonstrating its potential to significantly improve the applicability of automated TS searches. This approach offers a valuable tool for a broad range of applications such as reaction mechanism analysis and network exploration.



1. INTRODUCTION

Transition states (TSs) are crucial in elucidating the mechanisms of chemical reactions. They provide kinetic information about chemical reactions by determining the activation barrier for each reaction step.^{1–7} Despite their importance, experimental identification of TS structures remains an arduous task due to their transient nature. However, thanks to remarkable advances in quantum chemistry, they can be characterized computationally with acceptable accuracy at a reasonable cost.^{8–10} As a result, it has become common practice to elucidate reaction mechanisms at the atomistic level by estimating TS structures with quantum chemical calculations.

Over the past decades, various methods have been developed to find TS structures.^{11–37} Most of these methods rely on the molecular geometries of reactants or products. Representative examples include the nudged elastic band (NEB) method,^{15,38–40} the growing string method (GSM),^{22,41,42} and the artificial force induced reaction (AFIR) method.^{25,43,44} They can be categorized into single-ended or double-ended methods depending on the type of input used. Double-ended methods utilize both the reactant and product geometries to identify TS structures along the reaction pathway connecting the two structures. On the other hand, single-ended methods use only reactant geometry to discover product structures as well as their associated TS structures. These algorithms have been used for various

purposes, including mechanism analysis,^{45–51} reaction network exploration,^{52–58} and reaction database construction.^{59–62}

In practice, these methods require input conformations of reactants and products—known as reaction conformations—that are properly aligned along the reaction coordinate.^{63–65} For double-ended methods, they additionally require the two geometries to have a consistent atom ordering, as interpolation techniques are often employed to find TS structures. While these requirements can be addressed manually, such an approach is often tedious and time-consuming, especially for complex or large reaction systems. Moreover, manual workflows become infeasible for large-scale applications involving hundreds or thousands of reactions.

In this context, substantial efforts have been made to automate the generation of optimal reaction conformations for TS searches.^{51,63,64,66–70} Zimmerman and co-workers developed a tool called ZStruct, which automatically explores reactions based on predefined reaction rules and generates suitable reaction conformations consistent with those rules.^{66–68} Similarly, Young et al. presented autoDE,⁵¹ another

Received: December 11, 2024

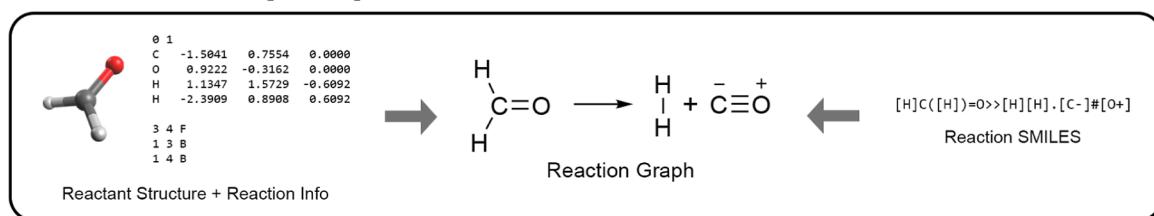
Revised: February 10, 2025

Accepted: February 13, 2025

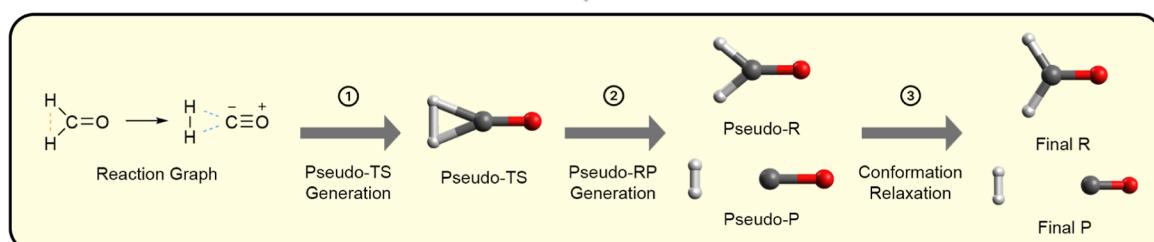
Published: February 25, 2025



Input: Reaction Graph of $\text{H}_2\text{CO} \rightarrow \text{H}_2 + \text{CO}$



This work: Reaction Conformation Generation from Reaction Graph



Output: Reaction Conformation

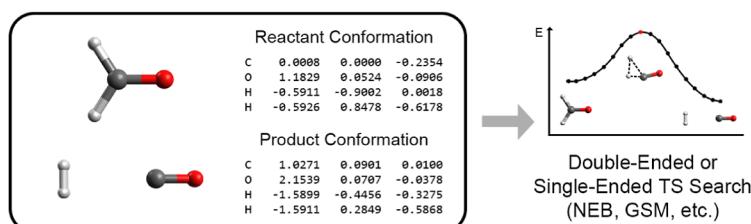


Figure 1. Graphical overview of the reaction conformation generation algorithm presented in this work, using formaldehyde dissociation as an example. Starting with a reaction graph, obtained either from a 3D reactant structure with reaction information or from a reaction SMILES, the algorithm generates a reaction conformation for subsequent use in TS searches. The process begins by constructing a pseudo-TS structure from the reaction graph. This structure is then transformed into approximate reactant/product geometries, referred to as pseudo-RP structures. Finally, these pseudo-RP structures are relaxed to produce the final reaction conformation. The resulting reaction conformation can be used as input for various conventional TS search methods.

automated TS search package tailored for reaction profile calculations that employs an approach akin to ZStruct. For bimolecular reactions, autodE aligns reacting molecules by positioning reacting atoms in close proximity through a customized potential. Robertson and Habershon proposed a scheme for positioning molecules that also minimizes structural differences between reactants and products, making it suitable for double-ended methods.⁶³ However, these methods treat reacting molecules as rigid entities, neglecting their conformational variations. This simplification can lead to failures in reactions involving flexible molecules with many rotatable bonds.⁷⁰

To address this limitation, Geiger et al. developed a precomplex generator (hereafter referred to as “Precomplex Generator”) which considers variations in the torsional angles of rotatable bonds.^{70,71} By adjusting these angles, it samples multiple reactant conformations and identifies structures whose reaction vectors—determined by reaction rules—are optimally aligned using custom scoring functions. Using this workflow, they successfully identified TS structures for several organic reactions involving large, flexible molecules. Similarly, Zhao et al. employed a machine learning-based classifier to find optimal reaction conformations.^{64,72} It first samples a large number of reaction conformations with the CREST algorithm, and then selects the conformations that are most likely to lead to the desired TS structure with the classifier. They integrated

their method into an automated reaction network exploration package, Yet Another Reaction Program (YARP), and used it to explore and characterize diverse types of unimolecular decomposition reactions and a set of bimolecular reactions comprising more than a hundred elementary reactions.^{64,73}

Both methods start by sampling the conformational space of the reactants and/or products. However, the number of possible conformations grows exponentially with the number of rotatable bonds. Furthermore, when multiple molecules are involved, the total conformational space expands rapidly, since the total size is proportional to the product of the individual conformational spaces. This combinatorial explosion makes these methods highly complicated for large reaction systems containing many rotatable bonds or multiple reacting molecules. The motivation behind this work is to investigate whether one can pinpoint a suitable reaction conformation for any desired TS without resorting to an exhaustive search, particularly in a large conformational space.

Herein, we present a robust algorithm for generating reaction conformations appropriate for TS searches. Our algorithm is inspired by the manual preparation process in which experts selectively generate desired conformations by exploiting information about bond formation and dissociation. By automating this intuitive approach, our algorithm enables efficient access to TS structures with minimal conformational sampling. We tested the algorithm on a diverse set of reactions,

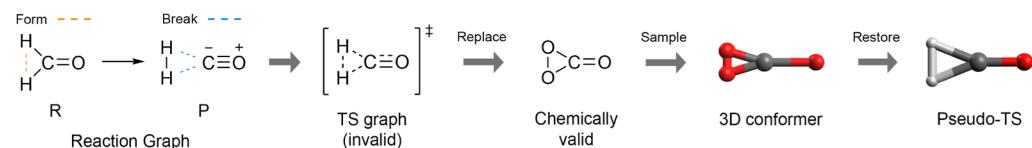
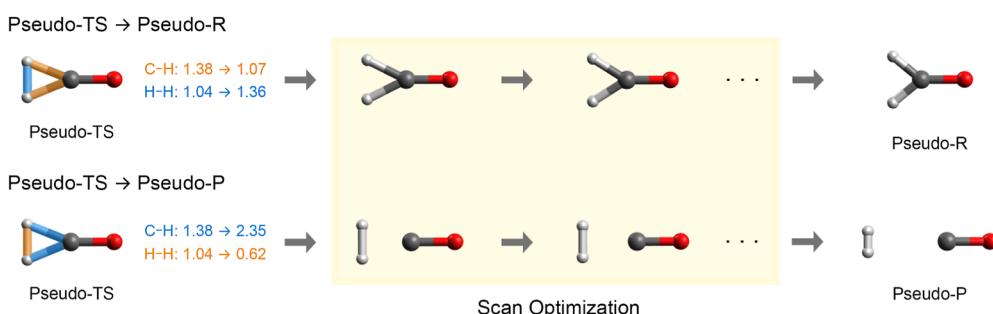
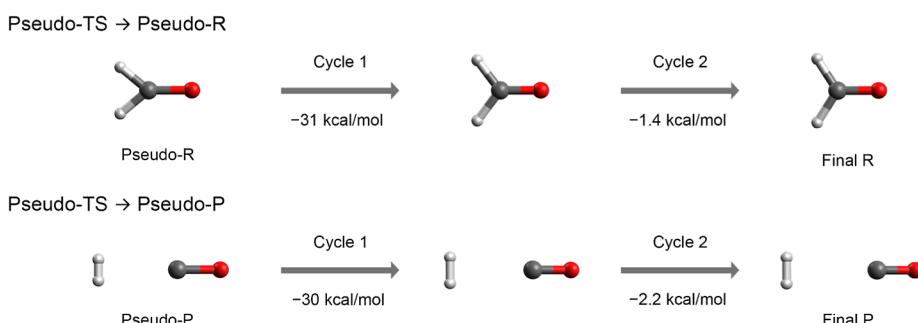
a. Pseudo-TS Generation**b. Pseudo-RP Generation****c. Conformation Relaxation**

Figure 2. Illustration of the three main steps of the proposed algorithm. (a) Pseudo-TS generation: a TS graph is constructed from a reaction graph. This graph is then used to generate a 3D pseudo-TS structure. (b) Pseudo-RP generation: the pseudo-TS structure undergoes constrained scan optimization to generate structures that resemble the reactants (pseudo-R) and the products (pseudo-P). (c) Conformation relaxation: these pseudo-RP structures are further relaxed through several microcycles of geometry optimization to yield the final reaction conformation.

including a collection of organic reactions prepared by Birkholz and Schlegel,⁷⁴ intramolecular Diels–Alder reactions of a ketothioester,⁵⁴ and catalytic reactions from the Urey-Miller experimental system.⁷⁵ These results confirm that our algorithm maintains a consistent level of computational resource usage, irrespective of the rotational degrees of freedom of reactants or products (e.g., the number of rotatable bonds) or the types of reactions (e.g., topologies of bond formations and dissociations). Additionally, the algorithm effectively generates input for both single-ended and double-ended TS search methods, establishing it as a valuable tool for TS search applications.

2. METHODS

Figure 1 illustrates the workflow of our algorithm, using the decomposition of formaldehyde into molecular hydrogen and carbon monoxide as an example. Our algorithm generates a reaction conformation suitable for TS searches of the reaction of interest. It takes a reaction graph as input, consisting of a reactant molecular graph and a product molecular graph, and outputs a reaction conformation for TS search input structures. The reaction graph can also be derived from a 3D reactant structure combined with bond change information or from a

reaction SMILES. Using this reaction graph, our algorithm proceeds with three main steps to produce the final reaction conformation: pseudo-TS generation, pseudo-RP generation, and conformation relaxation. In the first step, the algorithm uses the reaction graph to construct a three-dimensional structure resembling a hypothetical TS, referred to as a pseudo-TS (pseudo-TS generation). While this structure is not an exact TS, it provides a well-aligned geometry with the reacting atoms positioned appropriately along the reaction coordinate. This is the key to our algorithm, which allows for suitable reaction conformation generation with minimal conformational sampling. In the next step, the pseudo-TS structure is transformed into preliminary geometries resembling true reactants and products, referred to as pseudoreactant (pseudo-R) and pseudoproduct (pseudo-P) (pseudo-RP generation). A constrained optimization procedure is applied in this step to ensure that the resulting geometries are well-aligned with the desired atomic connectivities of the reactants and products. The prefix “pseudo” is used because these geometries are similar to the true reactants and products, but not exactly the same, and are often slightly less stable. Finally, the pseudo-R and pseudo-P structures are further optimized to yield the final reaction conformations (conformation relaxa-

tion). These optimized conformations can then be used as input for the TS search methods, such as the NEB method and GSM. Figure 2 details the three main steps, which are described in the following sections.

2.1. Pseudo-TS Generation. A TS graph is first derived from the given reaction graph. This TS graph is constructed by adding new bonds to the molecular graph of the reactants according to the bond formation information contained in the reaction graph. In Figure 2a, the TS graph containing a three-membered ring is formed from the reaction graph, where the orange dotted line represents the bond formation in the reactant. The bond formations in the products, which also correspond to the bond dissociations in the reactants, are represented by blue dotted lines. Since the reacting atoms are connected in the TS graph, the corresponding 3D structure aligns the reacting atoms close together along the reaction coordinate. However, in most cases, a 3D structure cannot be generated from this TS graph because it is usually chemically invalid (e.g., violates the octet rule), making it incompatible with cheminformatics tools such as RDKit. To resolve this problem, the graph is temporarily modified into a chemically valid counterpart that retains its original atomic connectivity. In the figure, the two hydrogen atoms that violate the octet rule by bonding with two other atoms are replaced by oxygen atoms. A conformer is generated from this modified graph, after which the original atoms are restored. In addition, the bond lengths in the resulting structure are adjusted to specific values to more closely resemble an actual TS structure. The target values are based on suggestions from the work of Birkholz and Schlegel.⁷⁴ It is important to note that these adjustments do not need to be precise, nor does the resulting pseudo-TS need to closely resemble the true TS structure. The primary purpose of generating the pseudo-TS is to provide an initial alignment between the reacting molecules that is suitable for the given reaction. Further details on the modification and adjustment process are available in the Supporting Information.

2.2. Pseudo-RP Generation. Constrained scan optimization is applied to the pseudo-TS to generate the corresponding pseudo-R and pseudo-P structures. During the scan optimization, the interatomic distances between the reacting atoms are gradually increased or decreased to the target distances by directly manipulating them, without using external forces. The target distances are determined by the sum of the atomic radii of the two reacting atoms, scaled by specific scale factors denoted as λ_f for bond formation and λ_d for bond dissociation

$$d_{ij} = \begin{cases} \lambda_f(R_i + R_j) & \text{if } (i, j) \in S_f \\ \lambda_d \max_{(k,l) \in S_d} (R_k + R_l) & \text{if } (i, j) \in S_d \end{cases}$$

where d_{ij} denotes the target distance between the i th and j th atoms, and R_i denotes the atomic radius of the i th atom. S_f and S_d denote the sets of bond formations and dissociations, respectively. In general, λ_f values range from 1.0 to 1.2, while λ_d values range from 1.8 to 2.2. Within these ranges, the distance between atoms involved in bond formation generally decreases while the distance between atoms involved in bond dissociation increases.

The pseudo-TS is inherently unstable, making it difficult to transform into the desired reactant and product geometries using simple scan optimization or intrinsic reaction coordinate

(IRC) calculations.⁷⁶ Furthermore, the alignment of reacting atoms in the pseudo-TS must be preserved throughout the transformation. To meet these requirements, we devised a special constrained scan optimization technique. This technique introduces two additional constraints to the standard scan optimization procedure: (1) fixing all interatomic distances between the bonded atoms, and (2) limiting the maximum number of optimization steps and the maximum displacement per step. The first constraint is intended to maintain the atomic connectivities of molecules during optimization. The second constraint suppresses large structural changes, preserving the initial atomic alignment and preventing undesired bond changes. To strike a balance between accuracy and computational efficiency, semiempirical methods such as PM6⁷⁷ and GFN2-xTB⁷⁸ are used by default for this step.

2.3. Conformation Relaxation. Each pseudo-R and pseudo-P undergoes several microcycles of geometry optimization. This step is often essential, as these structures tend to be unstable due to the constraints imposed in the previous step, particularly those that fix interatomic distances. Unlike the previous step, no geometric constraints are applied during optimization. However, limits on the maximum number of optimization steps and the maximum displacement are retained to preserve proper alignments. The relaxation process terminates when the stabilization energy within an optimization cycle falls below a predefined threshold. Based on our experience, an energy threshold between 3 and 10 kcal/mol, comparable to typical conformational energy differences,^{79–81} is appropriate. In Figure 2c, the process was completed within two cycles for both pseudo-R and pseudo-P, as the stabilization energies of the second cycle fell below the 10 kcal/mol threshold. The relaxed structures are closer to the stable reactant and product structures, while maintaining their molecular alignment along the reaction coordinate. These resulting structures can now be used as input for various TS search methods to identify TS structures.

Several options are available for this relaxation step. First, different levels of theory can be used during the optimization cycle. In general, using the same level of theory as in the subsequent TS search is optimal for maximizing the success rate of the TS search; however, this approach can make the generation process slow. Semiempirical methods often strike the best balance between efficiency and reliability. Additionally, preoptimization can be performed using force field methods, such as the Universal force field (UFF)⁸² or the Merck molecular force field (MMFF).⁸³ While this option often enhances the performance of the algorithm, it can occasionally disrupt the alignment of reaction conformations. Users can select a specific option or combine several approaches based on their requirements.

3. RESULTS AND DISCUSSION

We assessed the proposed method by applying it to three benchmark sets. All quantum chemical calculations for the reaction conformation generation were carried out with the GFN2-xTB method as implemented in the ORCA⁸⁴ package, and preoptimization with UFF was used in the conformation relaxation step. The detailed procedures for TS searches are described in the corresponding sections. The identified TS structures were rigorously validated by vibrational analysis and IRC calculations.⁷⁶ The hyperparameters used in each section are summarized in Section S4 of the Supporting Information.

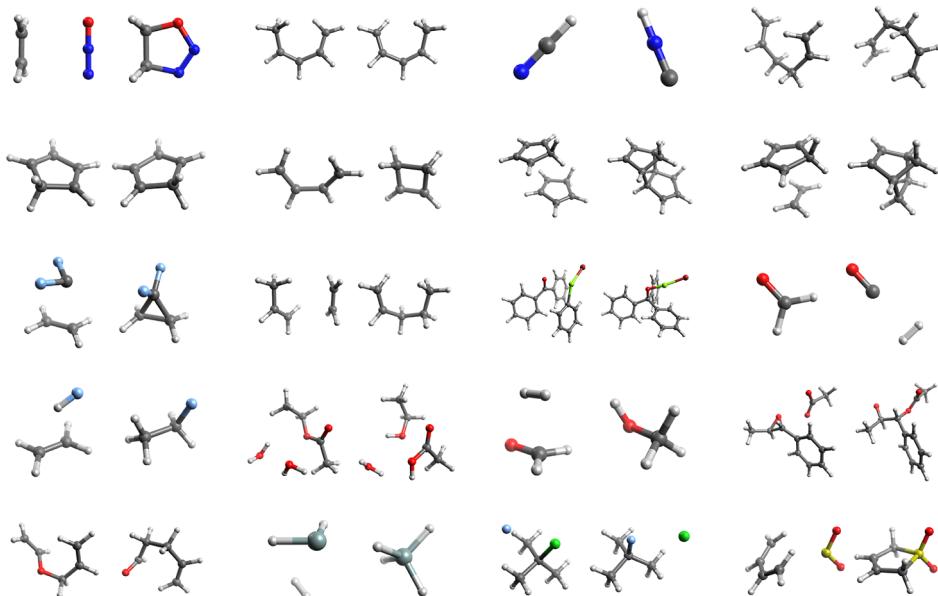


Figure 3. Reaction conformations of the Birkholz and Schlegel set generated by our algorithm.

Table 1. Electronic Energies of TS Structures of the Birkholz and Schlegel Set Obtained Using a Single Reaction Conformation for Each Reaction^a

	E_{ref}	E_{STQN}	E_{NEB}	$E_{\text{SE-GSM}}$
C2N2O	-263.2163	-263.2163	-263.2163	-263.2163
CSHT	-195.2614	-195.2614	-195.2614	-195.2614
HCN	-93.3474	-93.3474	-93.3474	-93.3474
Cope	-234.5720	-234.5720	-234.5720	-234.5720
CPHT	-194.0664	-194.0664	-194.0664	-194.0664
cyc-but	-155.9262	-155.9262	-155.9262	-155.9262
DACP2	-388.1802	-388.1857	-388.1857	-388.1857
DACP + eth	-272.6725	-272.6725	-272.6725	-272.6725
DFCP	-316.2708	-316.2708	-316.2708	-316.2708
ene	-196.4582	-196.4582	-196.4582	-196.4582
Grignard	-3580.1803	-3580.1803	-3580.1803	-3580.1803
H2 + CO	-114.3649	-114.3649	-114.3649	-114.3649
HF + eth	-178.9610	-178.9610	-178.9610	-178.9610
hydro	-460.5188	-460.5188	-460.5188	-460.5188
MeOH	-115.5732	-115.5732	-115.5732	-115.5732
oxirane	-652.6848	-652.6372	-652.6372	-652.6372
oxyCope	-270.4669	-270.4669	-270.4669	-270.4669
silane	-291.7942	-291.7942	-291.7942	-291.7942
SN2	-717.9007	-717.8410	-717.8410	-717.8410
sulfolene	-704.5600	-704.5600	-704.5600	-704.5600

^aAll energies are reported in E_h and calculated at the B3LYP/6-31G(d,p) level of theory. E_{ref} is the electronic energy of the TS structures as reported in ref 74, while E_{STQN} , E_{NEB} , and $E_{\text{SE-GSM}}$ are energies of the TS structures obtained using the STQN, NEB, and SE-GSM methods, respectively.

3.1. Benchmark on the Birkholz and Schlegel Set. The first benchmark set consists of reactions reported by Birkholz and Schlegel⁷⁴ (hereafter referred to as the Birkholz and Schlegel set), which have been widely used to evaluate various TS search methods.^{33,74,85,86} This set includes diverse reaction types such as insertion, addition, elimination, hydrolysis, ring opening/closing, substitution, cycloaddition, and rearrangement. The generated conformations were used as input for both single- and double-ended methods to assess the versatility of our method. Three TS search methods were used for this purpose: the synchronous transit-guided quasi-Newton (STQN) method^{14,16} and the nudged elastic band (NEB)

method^{15,38–40} for the double-ended methods, and the single-ended growing string method (SE-GSM)⁴² for the single-ended method. The STQN method employs a synchronous transit interpolation algorithm to directly search for saddle points near the interpolated reaction path between the given reactant and product structures. Alternatively, the reaction path optimization methods like the NEB method and the SE-GSM find an approximate TS structure by exploring a minimum energy path (MEP) between the reactant and products.^{2,32} As a double-ended method, the NEB method tries to optimize the initial path generated by interpolating between the reactant and product structures toward the MEP.

The reaction path is optimized under a spring force between images that is intended to make the spacing between them uniform.¹⁹ In contrast, SE-GSM is a single-ended method that only uses the reactant structure to search for TS structures. Starting from the reactant, it iteratively adds new images to the reaction path and optimizes the current path.⁴² The direction of the path growth is guided by a specified reaction coordinate, such as bond formation and dissociation rules. For both NEB and SE-GSM, the local maxima along the optimized reaction paths serve as approximate TS structures. These are further refined using other TS optimization methods, such as Berny optimization, to obtain exact TS structures.

In this work, the two-input quadratic synchronous transit (QST2) option, which uses reactant and product structures as inputs, served as the default for the STQN method. When QST2 failed, the three-input quadratic synchronous transit (QST3) option was applied, with the pseudo-TS structure as the third input. SE-GSM and NEB were performed using the pyGSM⁸⁷ and pysisyphus⁸⁸ packages, respectively, followed by Berny optimization⁸⁹ for the TS refinement. All quantum calculations for TS searches were performed with B3LYP/6-31G(d,p) as implemented in Gaussian 09,⁹⁰ in line with the original work.

Our method should be able to generate suitable reaction conformations without having to consider numerous conformations of reactants or products, provided that our algorithm effectively replicates the manual input preparation process. To test this hypothesis, we first evaluated the performance of our method using only a single reaction conformation for each reaction. These conformations were obtained by sampling a single pseudo-TS structure from each TS graph. Figure 3 shows the reaction conformations generated by our algorithm for each reaction in the test set. As shown in the figure, the reactant and product structures are properly aligned along the reaction coordinate. Furthermore, the structural differences between the reactant and product structure pairs appear to be minimal, facilitating TS searches with both double- and single-ended methods. Table 1 lists the electronic energies of the TS structures identified by each TS search method, utilizing these structures as input. The electronic energies of the reference TS structures are also listed for comparison. Notably, all TS structures were successfully obtained by only considering a single reaction conformation per reaction, proving the previous hypothesis. In addition, the electronic energies of most TS structures were close to the reference values, demonstrating the high reliability of our method. The raw output logs are available in ref 91.

To assess the efficiency of our algorithm, we measured the elapsed time for reaction conformation generation and TS search with the three methods. Table 2 summarizes the results. The time required for conformation generation ranged from a few seconds to under two min. This is slower than existing approaches; for example, Precomplex Generator typically generates reaction conformations of common organic reactions within a few seconds. The greater computational cost of our method is primarily attributed to the use of semiempirical methods to transform pseudo-TS structures into reactant and product structures. The intrinsic time complexity of semiempirical methods causes the time cost of our method to significantly increase with the number of atoms. For instance, the HCN reaction, consisting of three atoms, took 8 s, while the Grignard reaction involving 37 atoms needed about two min, as shown in Table 2.

Table 2. Computational Time for Reaction Conformation Generation and Subsequent TS Searches^a

	N_{atom}	t_{gen}	t_{STQN}	t_{NEB}	$t_{\text{SE-GSM}}$
C2N2O	9	15 s	32 s	15 min 0 s	40 min 50 s
CSHT	13	12 s	56 s	27 min 51 s	1 h 24 min 21 s
HCN	3	8 s	12 s	1 min 12 s	19 min 32 s
Cope	16	20 s	1 min 37 s	50 min 19 s	2 h 20 min 12 s
CPHT	11	15 s	35 s	17 min 2 s	1 h 24 min 3 s
cyc-but	10	22 s	25 s	20 min 59 s	1 h 3 min 24 s
DACP2	22	40 s	4 min 42 s	48 min 38 s	4 h 27 min 12 s
DACP + eth	17	34 s	1 min 12 s	26 min 35 s	2 h 3 min 37 s
DFCP	9	16 s	34 s	19 min 4 s	42 min 1 s
ene	15	22 s	1 min 26 s	42 min 21 s	1 h 49 min 46 s
Grignard	37	1 min 49 s	2 h 33 min 0 s	16 h 21 min 30 s	31 h 5 min 30 s
H2 + CO	4	9 s	17 s	7 min 22 s	41 min 40 s
HF + eth	8	10 s	24 s	15 min 7 s	26 min 18 s
hydro	20	39 s	12 min 36 s	1 h 30 min 2 s	4 h 12 min 44 s
MeOH	6	8 s	29 s	11 min 18 s	12 min 29 s
oxirane	27	43 s	27 min 57 s	3 h 57 min 28 s	17 h 15 min 28 s
oxyCope	14	17 s	1 min 33 s	49 min 4 s	1 h 36 min 37 s
silane	5	8 s	17 s	7 min 54 s	22 min 48 s
SN2	15	19 s	3 min 7 s	53 min 29 s	1 h 28 min 27 s
sulfolene	13	31 s	1 min 47 s	51 min 8 s	2 h 17 min 53 s

^a N_{atom} is the number of atoms, t_{gen} is the time taken by our algorithm to generate one reaction conformation. t_{STQN} , t_{NEB} , and $t_{\text{SE-GSM}}$ are the times taken by the STQN method, the NEB method (including TS refinement), and SE-GSM (including TS refinement). All computations were run on a 108-node cluster composed of two Intel(R) Xeon(R) E5-2667 v4 @ 2.40 GHz CPUs (14 effective cores for each CPU) and 128 GB of memory per node. Reaction conformation generation was performed on a single CPU core, whereas TS searches utilized 7-core parallelization.

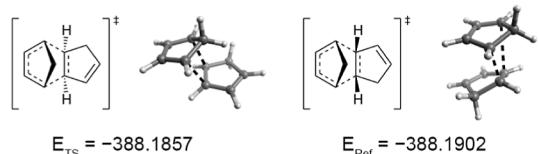
However, the computational cost of our method was often negligible compared to the TS search methods at the DFT level. Specifically, the time required for reaction conformation generation was around one to two orders of magnitude shorter compared to the time taken by the NEB method and the SE-GSM. On the other hand, the difference was less pronounced for the STQN method. This is because STQN directly optimizes a single structure to the TS, while the NEB method and SE-GSM optimize an entire reaction pathway, typically consisting of more than nine images in this work. For reactions involving a small number of atoms, the computation times for STQN and our algorithm were comparable, both requiring only a few seconds. However, as the system size increases (>15 atoms), the computational time for STQN becomes significantly longer than that for reaction conformation generation. It should also be noted that all TS search methods were run on seven CPU cores, while our conformation generation was performed using only a single CPU core. In the case of the Grignard reaction, which showed the largest

difference, the generation time was only about 1.3% of the STQN method and less than 0.2% of NEB or SE-GSM.

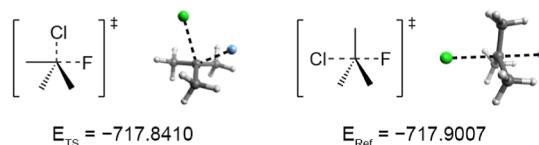
In summary, although our algorithm may not be as fast as other existing approaches in generating reaction conformations, its relative computational cost is negligible compared to the cost required for TS searches. Moreover, it remains efficient even when dealing with complex conformational spaces. This efficiency stems from its ability to generate suitable input structures with minimal conformational sampling. In addition, the algorithm is compatible with both single- and double-ended TS search methods.

Meanwhile, for four reactions in the benchmark set, we obtained TS structures that were less stable than those in the reference. Figure 4 shows three of them: the DACP2, SN2, and

a. DACP2



b. SN2



c. Hydro

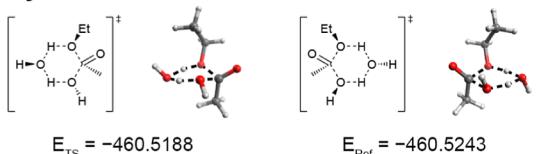


Figure 4. Examples of TS structures identified by our method that are less energetically stable than the reference structures. The TS structures obtained by our method are shown on the left, while the reference TS structures are shown on the right. (a) DACP2 reaction: our method identified an exo-TS that is slightly less stable than the endo-TS reported in the reference. (b) SN2 reaction: our method produced an unusual front-side attack TS, while the reference structure corresponds to the more stable back-side attack TS. (c) Hydro reaction: differences in electronic energy arise from the distinct placement of functional groups in the equatorial versus axial positions around the six-membered ring TS structure.

hydro reactions. The left side shows the TS structures identified by our algorithm, while the right side shows the reference structures. For the DACP2 reaction, our method identified the exo-TS, which is slightly less stable than the endo-TS reported in the reference. In the case of the SN2 reaction, we identified an unusual front-side attack TS, while the reference provided the more stable back-side attack TS, with a difference in energy of over 30 kcal/mol. A similar problem was observed for the oxirane reaction, as shown in Figure S7. For the hydro reaction, the differences in electronic energy resulted from variations in the placement of functional groups in the equatorial versus axial positions around the six-membered ring TS structure. These results suggest that these reactions can have multiple TS conformations.

Conventionally, TS conformational sampling is employed to explore several different TS conformations. This technique

involves the conformational sampling on the identified TS structures and the subsequent TS optimization. However, the success rate of the entire process is often unpredictable since the accuracy of the TS optimization is very sensitive to the initial TS guess.^{12,26,92–94} In this regard, we employed an alternative strategy. Rather than sampling directly from identified TS structures, we sampled diverse reaction conformations for TS searches to obtain different TS structures. This approach is based on the assumption that the TS structure obtained from each reaction conformation will closely resemble the pseudo-TS structure from which the reaction conformation originated. In our algorithm, a variety of pseudo-TS structures is sampled initially. The sampled pseudo-TS structures are then used to generate corresponding reaction conformations. From these reaction conformations, multiple TS structures can be obtained. While this strategy involves additional computational steps, it can have a higher success rate than the conventional TS conformational sampling strategy. This is because sampled pseudo-TS structures do not need to closely resemble true TS structures to yield high-quality reaction conformations for TS searches, as demonstrated previously.

Two sampling strategies were employed. First, we enumerated pseudo-TS stereoisomers by varying the stereochemistry at the reaction centers, as described in Section S2 of the Supporting Information. Second, we applied the CREST method^{78,95} to each enumerated stereoisomer to sample more stable pseudo-TS structures. Constraints on the bonds involved in the reaction were introduced during the sampling process. For the sake of computational efficiency, only the three most stable structures from CREST were considered, assuming that highly unstable pseudo-TS structures will lead to unstable TS structures. We also excluded pseudo-TS structures with energy differences less than 0.5 kcal/mol to avoid finding duplicate TS structures. The resulting reaction conformations were then subjected to TS searches using the STQN method, which had the shortest computational time. In cases where the STQN method failed, the NEB method was used for additional TS search.

Table 3 summarizes the results obtained using multiple reaction conformations. The lowest energy TS structure of each reaction was comparable to or more stable (oxyCope and hydro reaction) than the reference structures. In most cases, only a single pseudo-TS structure was still considered despite using the above strategies, implying that no other viable TS conformations may exist. For these reactions, the electronic energies of the TS structures match the reference values, except for the oxyCope reaction and the hydro reaction, which had even lower electronic energies than the reference. In cases where multiple reaction conformations were considered, diverse TS structures were identified for several reactions, demonstrating the possibility of multiple TS conformations. Notably, the overall success rate for these and other tested reactions were 100%, underscoring the high reliability of using multiple reaction conformations for TS searches.

Figure 5 illustrates how the sampling strategies contributed to the identification of more stable TS structures. For the DACP2 reaction (Figure 5a), a pseudo-TS close to the endo-TS was sampled using stereoisomer enumeration, eventually leading to the TS structure with an energy matching the reference. This enumeration also helped identify the back-side attack TS for the SN2 reaction (Figure 5b). Without stereoisomer enumeration, only the pseudo-TS structure with

Table 3. Summary of TS Searches of the Birkholz and Schlegel Set Using Multiple Reaction Conformations for Each Reaction^a

	E_{ref}	N_{conf}	N_{success}	N_{diff}	E_{lowest}
C2N2O	-263.2163	1	1	1	-263.2163
CSHT	-195.2614	1	1	1	-195.2614
HCN	-93.3474	1	1 ^b	1	-93.3474
Cope	-234.5720	1	1	1	-234.5720
CPHT	-194.0664	1	1	1	-194.0664
cyc-but	-155.9262	1	1	1	-155.9262
DACP2	-388.1902	2	2	2	-388.1902
DACP + eth	-272.6725	1	1	1	-272.6725
DFCP	-316.2708	1	1 ^b	1	-316.2708
ene	-196.4582	1	1	1	-196.4582
Grignard	-3580.1803	3	3 ^b	1	-3580.1803
H2 + CO	-114.3649	1	1	1	-114.3649
HF + eth	-178.9610	1	1	1	-178.9610
hydro	-460.5252	6	6	6	-460.5252
MeOH	-115.5732	1	1	1	-115.5732
oxirane	-652.6848	7	7	4	-652.6848
oxyCope	-270.4669	1	1	1	-270.4734
silane	-291.7942	1	1	1	-291.7942
SN2	-717.9007	3	3	2	-717.9007
sulfolene	-704.5600	1	1	1	-704.5600

^aAll energies are reported in E_h , and the STQN method was used for the TS searches unless noted otherwise. E_{ref} is the electronic energy of the TS structures as reported in ref 74. N_{conf} is the total number of reaction conformations used in TS searches. N_{success} is the number of successful TS searches. N_{diff} is the number of TS structures with distinct energies. E_{lowest} is the electronic energy of the most stable TS structure.

^bThe NEB method was additionally used to locate the TS structure from one of the generated reaction conformations.

an F–C–Cl angle of about 90° was considered, leading to the front-side attack TS structure. With the stereoisomer enumeration, a pseudo-TS structure with an F–C–Cl of 180° was also considered, resulting in the typical back-side attack TS. A similar improvement was observed for the oxirane reaction (Figure S8). These results support our assumption that true TS structures tend to be close to the corresponding pseudo-TS structures. Additional examples of this correspondence can be found in Figures S9 and S10. For the hydro reaction (Figure 5c), both stereoisomer enumeration and CREST played crucial roles in identifying stable TS structures. Starting from a pseudo-TS stereoisomer that matches the stereochemistry of the reference TS structure, CREST generated multiple pseudo-TS structures, yielding several reaction conformations from that initial stereoisomer. TS search from these conformations produced one TS structure that exactly matched the reference, while another TS structure was found to be even more stable than the reference. Interestingly, CREST also improved the TS structure of the oxyCope reaction (Figure 5d). Initially, a boat-form TS consistent with the reference was derived from a boat-form pseudo-TS structure. However, applying CREST yielded a chair-form pseudo-TS, which ultimately led to a chair-form TS that was 4 kcal/mol lower in energy than the reference. Overall, these results demonstrate that our algorithm can reliably identify lower-energy TS structures by considering multiple reaction conformations.

Because applying these sampling strategies can significantly increase the overall computational cost, we offer the following recommendations on when to consider their use. First, stereoisomer enumeration is generally recommended because it can identify lower energy TS structures with an insignificant

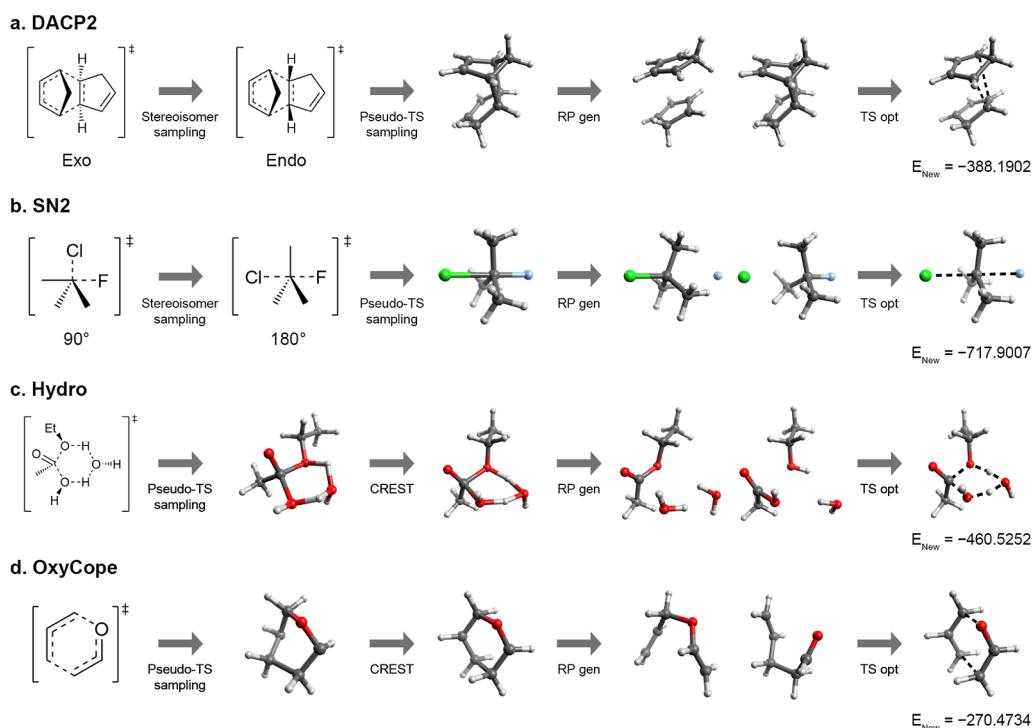


Figure 5. Illustration of improvements in TS structures through stereoisomer enumeration and conformational sampling. (a) DACP2 reaction: stereoisomer enumeration led to the identification of the endo-TS. (b) SN2 reaction: stereoisomer enumeration resulted in the identification of the back-side attack TS. (c) Hydro reaction: stereoisomer enumeration and CREST enabled us to obtain a TS structure with an electronic energy even lower than the reference. (d) oxyCope reaction: CREST allowed the discovery of a TS structure even more stable than the reference.

increase in cost. Second, conformational sampling methods, such as CREST, should be used cautiously due to their high computational expense, especially in reaction systems with large conformational spaces. When these methods are employed, retaining the three most stable conformations for each stereoisomer seems to be sufficient. Typically, enumerating only stereoisomers results in up to 2^n isomers, where n is the number of reacting atoms. If three conformations are considered for each stereoisomer, a maximum of 3×2^n structures is evaluated.

For comparison, we briefly report on the performance of other existing methods on the Birkholz and Schlegel set. Two most recently developed methods were adopted: Precomplex Generator⁷⁰ and the conformational sampling strategy⁶⁴ implemented in YARP, henceforth referred to simply as “YARP”. The structures generated by Precomplex Generator were subjected to SE-GSM, since the method generates reactant conformations (also called “precomplexes”) for single-ended methods. The HCN, CPHT, and H₂ + CO reactions were omitted because these reactions each have only one reactant conformer. Meanwhile, the reaction conformations generated by YARP were subjected to the STQN and NEB methods, which are double-ended methods intended to be used with YARP. The Grignard and silane reactions were omitted for YARP as its ML-based classifier, which filters reaction conformations, does not support elements Mg and Si. All generated reactant/reaction conformations were subjected to TS searches. Detailed procedures for Precomplex Generator and YARP, along with the following TS searches, are described in Section S7 of the Supporting Information.

Table 4 provides a side-by-side comparison of the results. Precomplex Generator failed for six reactions. The CSHT reaction failed because Precomplex Generator did not output any reactant conformations. The other reactions failed due to SE-GSM failures. These SE-GSM failures stem from the inadequate alignment between the reacting atoms, as can be seen in Figure S11. Among the successful cases, while Precomplex Generator obtained a lower energy TS structure than the reference for the oxyCope reaction, it produced higher energy TS structures for the oxirane and DACP2 reactions. Additional TS conformational sampling may be required to obtain lower energy TS structures for those two reactions.

YARP failed for four reactions with STQN and two reactions with NEB. This is likely because the CREST algorithm, used at first to generate many reactant conformations, tends to find minimum energy structures rather than structures suitable for TS searches. This weakness is particularly evident in the hydro reaction, as shown in Figure S12, where both TS search methods failed. The main cause of the failure was the hydrogen bonds between the reactant molecules, which disrupted the reacting atoms from being properly aligned. Like Precomplex Generator, YARP yielded TS structures with energies higher than the reference for some reactions (HCN, DACP2, and oxirane reactions), indicating that additional TS conformational sampling is also required to find more stable TS structures.

3.2. Study on the Intramolecular DA Reactions of a Ketothioester. The second test set consists of intramolecular Diels–Alder (DA) reactions of a ketothioester, originally studied by Yang et al.⁹⁶ In this study, we selected five DA reactions previously investigated by Zhao and co-workers with YARP.^{64,97} Figure 6a illustrates these reactions and their

Table 4. Results of TS Searches of the Birkholz and Schlegel Set Performed with Precomplex Generator and YARP^a

	Precomplex Generator	YARP	
	$E_{\text{SE-GSM}}$	E_{STQN}	E_{NEB}
C2N2O	failed	-263.2163	-263.2163
CSHT	failed	-195.2614	-195.2614
HCN	N/A ^b	-93.2943	failed
Cope	-234.5720	-234.5720	-234.5720
CPHT	N/A ^b	failed	-194.0664
cyc-but	failed	-155.9262	-155.9262
DACP2	-388.1857	-388.1857	-388.1857
DACP + eth	-272.6725	failed	-272.6725
DFCP	-316.2708	-316.2708	-316.2708
ene	-196.4582	-196.4582	-196.4582
Grignard	-3580.1803	N/A ^c	N/A ^c
H ₂ + CO	N/A ^b	failed	-114.3649
HF + eth	-178.9610	-178.9610	-178.9610
hydro	failed	failed	failed
MeOH	-115.5732	-115.5732	-115.5732
oxirane	-652.6454	-652.6791	-652.6791
oxyCope	-270.4734	-270.4734	-270.4669
silane	-291.7942	N/A ^c	N/A ^c
SN2	failed	-717.8410	-717.8410
sulfolene	failed	-704.5600	-704.5600

^aMultiple reactant/reaction conformations were used for TS searches for each reaction, with more details in Section S7 of the Supporting Information. $E_{\text{SE-GSM}}$ is the electronic energy of the most stable TS structure obtained with Precomplex Generator and SE-GSM. E_{STQN} and E_{NEB} are the electronic energies of the most stable TS structure obtained with YARP in conjunction with STQN and NEB, respectively. All energies are reported in E_h . ^bReaction involves reactant with only one conformer. ^cReaction includes elements not supported by YARP.

respective thermodynamic properties in kcal/mol. The values in parentheses are those reported in ref 64. All of these reactions have activation barriers below 40 kcal/mol, indicating their kinetic feasibility. Starting with a single reactant molecule containing many double bonds and rotatable bonds, these DA reactions produce a diverse range of complex products, making the generation task more challenging than the previous test set. Despite these complexities, we successfully generated appropriate reaction conformations for TS searches in less than a minute, as detailed in Supporting Information S10. TS searches were performed with the STQN method at the B3LYP/6-31G level implemented in Gaussian 16,⁹⁸ and thermochemical properties were obtained with vibrational calculations at the B3LYP-D3(BJ)/TZVP level of theory, as conducted in the original studies.

To identify stable TS conformations, we employed the pseudo-TS sampling strategies described in the previous section. For these DA reactions, stereoisomer enumeration was also critical in generating conformers with the correct E/Z stereochemistry, as the R/S stereochemistry of the reaction centers influences the E/Z stereochemistry of the resulting reactant (see Figure S13). For this reason, only reactant conformations where the E/Z stereochemistry of the generated reactant matched that of the original reactant were used for TS searches. In addition, to reduce computational cost, we used only the most stable pseudo-TS structure provided by CREST for each stereoisomer.

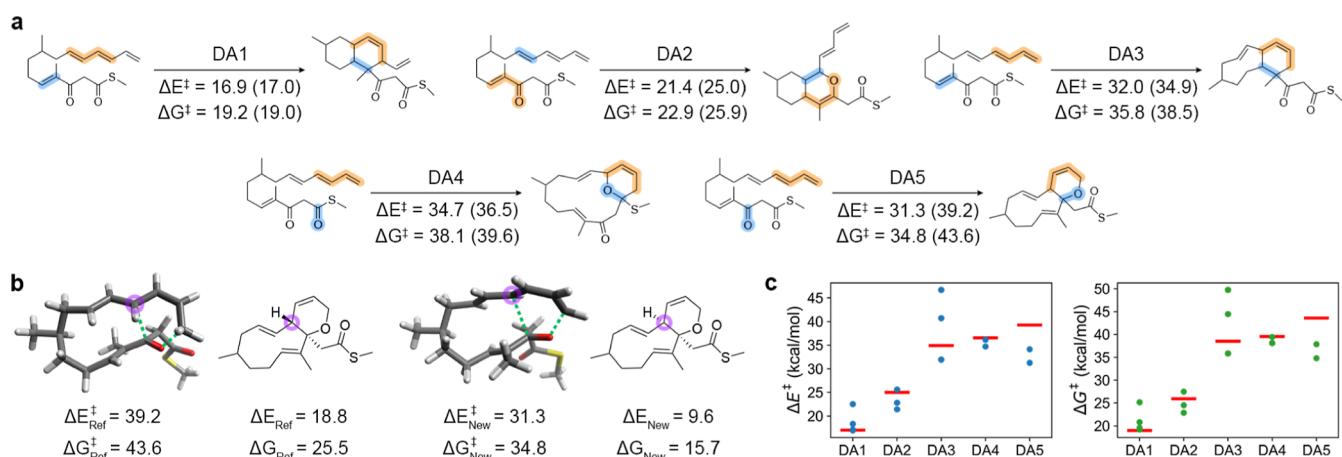


Figure 6. (a) Illustration of the five intramolecular Diels–Alder reactions studied in this work, with reacting dienes highlighted in orange and dienophiles in blue. Activation energies (ΔE^\ddagger) and free energies of activation (ΔG^\ddagger) are given in kcal/mol, with reference values from ref 64 provided in parentheses. (b) Comparison of the TS structures and the corresponding products between the reference and those identified by our method for the DA5 reaction. The energy and free energy values, expressed in kcal/mol, show that our method identified a pathway that is more favorable both kinetically and thermodynamically, leading to a different product stereoisomer with an energy difference of approximately 10 kcal/mol. (c) Distribution of activation energies and free energies of activation of the TS structures identified by our method across all reactions. The red bars indicate the energy and free energy values of the most stable reference TS structure.

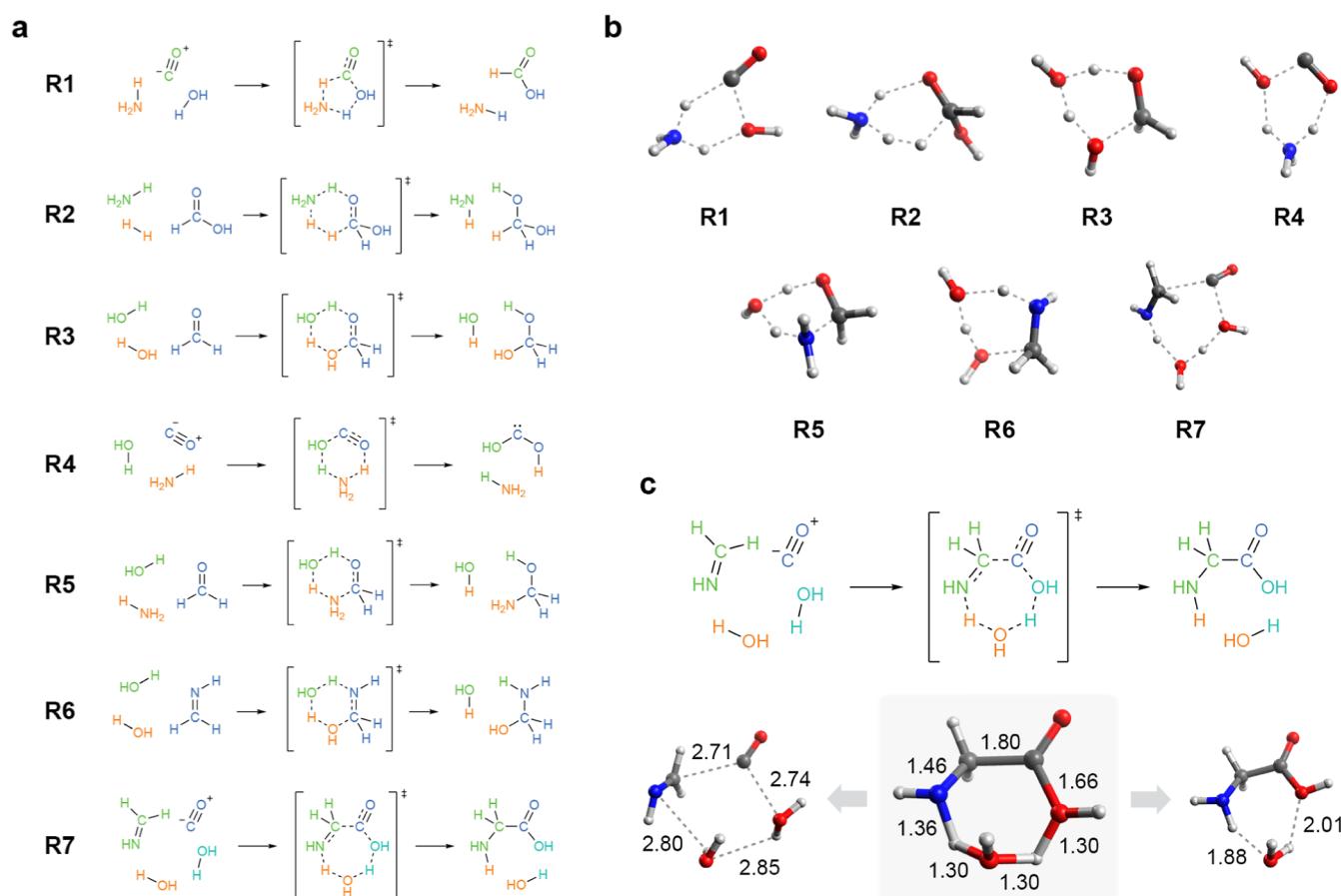


Figure 7. (a) A set of high-molecularity elementary reactions discovered with nanoreactor simulations, which has been selected to test our method. Each involves three or more reactant molecules. (b) TS structures obtained from the reaction conformations generated by our method. (c) Example of R7, demonstrating the method's capability to intelligently generate reaction conformations using a pseudo-TS structure. Distances between atoms are in angstroms (Å). Our use of pseudo-TS allows the alignment of multiple molecules along the reaction coordinate without exhaustive sampling, which is particularly effective in complex reactions such as R7 involving four reactant molecules.

The calculated thermochemical properties of the reactions are summarized in Figure 6a. For all five reactions except DA1, the calculated activation energies (ΔE^\ddagger) and free energies of

activation (ΔG^\ddagger) were lower than their respective reference values. In DA1, the activation barrier remained nearly unchanged. The other four reactions showed notable improve-

ments. Three reactions showed modest decreases of 2–4 kcal/mol, while DAS showed a significant decrease of about 8 kcal/mol; the calculated activation energy was 31.3 kcal/mol, while the reference was 39.2 kcal/mol. *Figure 6b* compares the newly identified DAS TS structure with the reference structure. The new TS structure leads to a product stereoisomer that is about 10 kcal/mol more stable than the reference, suggesting that the new pathway is more favorable both kinetically and thermodynamically. For the remaining reactions (DA2–DA4, see *Figure S14*), both new and reference TS structures yield the same product stereoisomer. The improvements in these reactions were less pronounced than in DAS due to only minor differences in their side chain conformations. *Figure 6c* shows the distribution of the calculated ΔE^\ddagger and ΔG^\ddagger of all TS structures identified by our algorithm. The red bars indicate the lowest energy values reported in the reference for each reaction. The result shows that our method enables us to find various TS conformers of large and complex molecules, including at least one TS conformer with an energy lower than the corresponding reference value.

For Precomplex Generator, it successfully identified several TS structures for reactions DA1, DA2, and DA3 with SE-GSM. However, no TS structure was found for DA4 and DAS. For DA4, two reactant conformations were generated, but SE-GSM failed to identify TS structures for both. In the case of DAS, the method did not output any reactant conformations. These results exemplify the challenge of finding suitable reactant conformations for TS searches when a reaction contains many rotatable bonds. Detailed results can be found in Section S11 of the *Supporting Information*. Note that YARP was not tested on this set, as the reference structures for comparison were generated with YARP.

3.3. Study on Multimolecular Reactions Discovered by Nanoreactor Simulations. The final test set consists of high-molecularity elementary reactions discovered from automated methods for reaction exploration, such as nanoreactors. For instance, nanoreactor simulations produce many possible products triggered by high pressure and temperature conditions or external potentials. Because the simulations provide energetic states along a reaction path connecting two products, they can be used as initial guesses for subsequent TS searches. Diverse systems have been studied with such methods.^{75,99–102} To examine whether our method can be used in such cases, we applied it to a glycine formation reaction network discovered from a nanoreactor simulation.⁷⁵ Seven elementary reactions involving three or more reactant molecules were selected from the network, as shown in *Figure 7a*. Unlike the previous test sets, the reactant molecules in this test are relatively small, but the complexity of the conformational space remains high due to the many possible orientations between three of four molecules in each reactant state, making the generation task difficult. Adding to this difficulty, for each reaction, only one input reaction conformation was generated and evaluated with the STQN method at the B3LYP/6-31G+(d,p) level of theory.

Despite the aforementioned complexities, we successfully identified TS structures for all reactions, as listed in *Figure 7b*. *Figure 7c* illustrates how our method generated reactant and product geometries that are suitably aligned for the R7 reaction. This reaction involves four reactant molecules, which can have numerous molecular arrangements. The figure highlights the critical role of the pseudo-TS in intelligently positioning the molecules along the reaction coordinate

without requiring extensive sampling. Starting from this pseudo-TS structure, our method gradually modifies it to generate a suitable reaction conformation without sampling for multiple conformations. Notably, Precomplex Generator and YARP failed to generate reactant/reaction conformations for about half of the reactions, with the former producing four conformations and the latter three. Among them, Precomplex Generator successfully identified TS structures only for reactions R3 and R6, while YARP succeeded only for R3 using the STQN method. These results underscore the difficulty of generating well-aligned reaction conformations for multimolecular reactions. Detailed results are provided in Section S12 of the *Supporting Information*.

4. CONCLUSIONS AND OUTLOOK

Automating TS searches has become increasingly important for advancing computational studies of chemical reactions. Traditional approaches have relied on the conformational sampling of reactants and products, which, while effective, can become highly complicated for reaction systems with large conformational spaces. To address this challenge, we have developed a robust method that generates reaction conformations suitable for TS searches with minimal reliance on conformational sampling. A unique feature of our method is the construction of a pseudo-TS structure based on the bond formation and dissociation information on a given reaction, which is subsequently transformed into reactant and product geometries. We have rigorously tested the method on three benchmark sets encompassing a diverse set of organic reactions, reactions with extensive conformational spaces, and high-molecularity reactions. Remarkably, our method successfully characterized TS structures across all test cases. In many instances, a single pseudo-TS was sufficient to identify the TS structure with negligible computational cost compared to traditional TS search workflows. We compared the performance of our method with those of two existing methods, Precomplex Generator and YARP. The results show that our method achieved a higher success rate in TS searches and generally produced higher-quality TS structures. This advantage becomes especially pronounced in reactions with expansive conformational spaces or those involving three or more molecules. Moreover, by sampling multiple pseudo-TS structures, we discovered several distinct TS structures, some of which had energies lower than those found in the reference. This demonstrates the high reliability and efficacy of our approach in preparing input structures for TS searches, regardless of reaction complexity.

Despite these advancements, further development is necessary. Currently, the use of semiempirical calculations renders our algorithm slower than some existing methods. To overcome this limitation, faster electronic energy calculation methods, such as machine learning interatomic potentials,^{103–107} can be employed. Another area for future work is to extend our method to support reactions involving transition metals, such as organometallic reactions. The complexity of generating conformers for metal-containing species presents a challenge in creating appropriate pseudo-TS structures. Finally, while our method significantly reduces the need for exhaustive sampling of reaction conformations, exhaustive conformational sampling is still required to identify the most stable TS conformations. Future work in this area may enable our method to pinpoint the most stable TS structures without exhaustive sampling. Our previous work

demonstrated the feasibility of generative AI in efficiently generating appropriate TS conformers directly from the molecular graphs of reactants and products.⁶⁵

Nonetheless, our new method shows substantial promise in improving both the reliability and efficiency of automated TS searches. We believe that our approach will facilitate broader applications, including comprehensive reaction mechanism analysis, extensive reaction network exploration, and the development of elementary reaction databases.

■ ASSOCIATED CONTENT

Data Availability Statement

The code used in this paper is available at <https://github.com/kyunghoonlee777/autoCG>.

■ Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.4c01692>.

Algorithm details for the generation of pseudo-TS structures and stereoisomer enumeration, algorithm hyperparameters and NEB/SE-GSM hyperparameters, additional examples from the Birkholz and Schlegel set, examples demonstrating the importance of stereoisomer enumeration for intramolecular DA reactions, results produced with our method, Precomplex Generator, and YARP ([PDF](#))

■ AUTHOR INFORMATION

Corresponding Author

Woo Youn Kim – Department of Chemistry, KAIST, Daejeon 34141, Republic of Korea;  orcid.org/0000-0001-7152-2111; Email: wooyoun@kaist.ac.kr

Authors

Kyunghoon Lee – Department of Chemistry, KAIST, Daejeon 34141, Republic of Korea;  orcid.org/0000-0002-5454-595X

Jinwon Lee – Department of Chemistry, KAIST, Daejeon 34141, Republic of Korea;  orcid.org/0009-0001-0619-3646

Shinyoung Park – Department of Chemistry, KAIST, Daejeon 34141, Republic of Korea;  orcid.org/0009-0000-3648-5124

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jctc.4c01692>

Author Contributions

[†]K.L., J.L., and S.P. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by Korea Environmental Industry and Technology Institute (grant no. RS202300219144) to K.L., W.Y.K., and Samsung Electronics Co., Ltd (grant no. IO231103-07636-01) to K.L., W.Y.K. We thank Prof. Mu-Hyun Baik of the Center for Catalytic Hydrocarbon Functionalizations, Institute for Basic Science (IBS) and Department of Chemistry at KAIST, for providing the computing resources.

■ REFERENCES

- (1) Bao, J. L.; Meana-Pañeda, R.; Truhlar, D. G. Multi-path variational transition state theory for chiral molecules: the site-dependent kinetics for abstraction of hydrogen from 2-butanol by hydroperoxyl radical, analysis of hydrogen bonding in the transition state, and dramatic temperature dependence of the activation energy. *Chem. Sci.* **2015**, *6*, 5866–5881.
- (2) Dewyer, A. L.; Argüelles, A. J.; Zimmerman, P. M. Methods for exploring reaction space in molecular systems. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, No. e1354.
- (3) Unsleber, J. P.; Reiher, M. The exploration of chemical reaction networks. *Annu. Rev. Phys. Chem.* **2020**, *71*, 121–142.
- (4) Lewis-Atwell, T.; Townsend, P. A.; Grayson, M. N. Machine learning activation energies of chemical reactions. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2022**, *12*, No. e1593.
- (5) Park, S.; Han, H.; Kim, H.; Choi, S. Machine learning applications for chemical reactions. *Chem.-Asian J.* **2022**, *17*, No. e202200203.
- (6) Ismail, I.; Robertson, C.; Habershon, S. Successes and challenges in using machine-learned activation energies in kinetic simulations. *J. Chem. Phys.* **2022**, *157*, 014109.
- (7) Ismail, I.; Chantreau Majerus, R.; Habershon, S. Graph-driven reaction discovery: Progress, challenges, and future opportunities. *J. Phys. Chem. A* **2022**, *126*, 7051–7069.
- (8) Houk, K. N.; Cheong, P. H.-Y. Computational prediction of small-molecule catalysts. *Nature* **2008**, *455*, 309–313.
- (9) Bursch, M.; Mewes, J.-M.; Hansen, A.; Grimme, S. Best-practice DFT protocols for basic molecular computational chemistry. *Angew. Chem., Int. Ed.* **2022**, *61*, No. e202205735.
- (10) Bryenton, K. R.; Adeleke, A. A.; Dale, S. G.; Johnson, E. R. Delocalization error: The greatest outstanding challenge in density-functional theory. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2023**, *13*, No. e1631.
- (11) Halgren, T. A.; Lipscomb, W. N. The synchronous-transit method for determining reaction pathways and locating molecular transition states. *Chem. Phys. Lett.* **1977**, *49*, 225–232.
- (12) Cerjan, C. J.; Miller, W. H. On finding transition states. *J. Chem. Phys.* **1981**, *75*, 2800–2806.
- (13) Basilevsky, M.; Shamov, A. The local definition of the optimum ascent path on a multi-dimensional potential energy surface and its practical application for the location of saddle points. *Chem. Phys.* **1981**, *60*, 347–358.
- (14) Peng, C.; Bernhard Schlegel, H. Combining synchronous transit and quasi-newton methods to find transition states. *Isr. J. Chem.* **1993**, *33*, 449–454.
- (15) Mills, G.; Jónsson, H. Quantum and thermal effects in H₂ dissociative adsorption: Evaluation of free energy barriers in multidimensional quantum systems. *Phys. Rev. Lett.* **1994**, *72*, 1124.
- (16) Peng, C.; Ayala, P. Y.; Schlegel, H. B.; Frisch, M. J. Using redundant internal coordinates to optimize equilibrium geometries and transition states. *J. Comput. Chem.* **1996**, *17*, 49–56.
- (17) Quapp, W.; Hirsch, M.; Imig, O.; Heidrich, D. Searching for saddle points of potential energy surfaces by following a reduced gradient. *J. Comput. Chem.* **1998**, *19*, 1087–1100.
- (18) Henkelman, G.; Jónsson, H. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *J. Chem. Phys.* **1999**, *111*, 7010–7022.
- (19) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **2000**, *113*, 9901–9904.
- (20) Hirsch, M.; Quapp, W. Improved RGF method to find saddle points. *J. Comput. Chem.* **2002**, *23*, 887–894.
- (21) Weinan, E.; Ren, W.; Vanden-Eijnden, E. String method for the study of rare events. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2002**, *66*, 052301.
- (22) Peters, B.; Heyden, A.; Bell, A. T.; Chakraborty, A. A growing string method for determining transition states: Comparison to the nudged elastic band and string methods. *J. Chem. Phys.* **2004**, *120*, 7877–7886.

- (23) Berente, I.; Náray-Szabó, G. Multicoordinate driven method for approximating enzymatic reaction paths: Automatic definition of the reaction coordinate using a subset of chemical coordinates. *J. Phys. Chem. A* **2006**, *110*, 772–778.
- (24) Del Campo, J. M.; Köster, A. M. A hierarchical transition state search algorithm. *J. Chem. Phys.* **2008**, *129*, 024107.
- (25) Maeda, S.; Morokuma, K. Communications: A systematic method for locating transition structures of $A + B \rightarrow X$ type reactions. *J. Chem. Phys.* **2010**, *132*, 241102.
- (26) Schlegel, H. B. Geometry optimization. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 790–809.
- (27) Behn, A.; Zimmerman, P. M.; Bell, A. T.; Head-Gordon, M. Efficient exploration of reaction paths via a freezing string method. *J. Chem. Phys.* **2011**, *135*, 224108.
- (28) Behn, A.; Zimmerman, P. M.; Bell, A. T.; Head-Gordon, M. Incorporating linear synchronous transit interpolation into the growing string method: Algorithm and applications. *J. Chem. Theory Comput.* **2011**, *7*, 4019–4025.
- (29) Mallikarjun Sharada, S.; Zimmerman, P. M.; Bell, A. T.; Head-Gordon, M. Automated transition state searches without evaluating the Hessian. *J. Chem. Theory Comput.* **2012**, *8*, 5166–5174.
- (30) Chaffey-Millar, H.; Nikodem, A.; Matveev, A. V.; Krüger, S.; Rösch, N. Improving upon string methods for transition state discovery. *J. Chem. Theory Comput.* **2012**, *8*, 777–786.
- (31) Zhang, X.-J.; Shang, C.; Liu, Z.-P. Double-ended surface walking method for pathway building and transition state location of complex reactions. *J. Chem. Theory Comput.* **2013**, *9*, 5745–5753.
- (32) Simm, G. N.; Vaucher, A. C.; Reiher, M. Exploration of reaction pathways and chemical transformation networks. *J. Phys. Chem. A* **2019**, *123*, 385–399.
- (33) Vaucher, A. C.; Reiher, M. Minimum energy paths and transition states by curve optimization. *J. Chem. Theory Comput.* **2018**, *14*, 3091–3099.
- (34) Maeda, S.; Harabuchi, Y. On benchmarking of automated methods for performing exhaustive reaction path search. *J. Chem. Theory Comput.* **2019**, *15*, 2111–2115.
- (35) Grimme, S. Exploration of chemical compound, conformer, and reaction space with meta-dynamics simulations based on tight-binding quantum chemical calculations. *J. Chem. Theory Comput.* **2019**, *15*, 2847–2862.
- (36) Shim, J.; Lee, J.; Yu, J. Efficient discovery of multiple minimum action pathways using Gaussian process. *J. Phys. Commun.* **2023**, *7*, 025004.
- (37) Lee, K.; Kim, J. H.; Kim, W. Y. pyMCD: Python package for searching transition states via the multicoordinate driven method. *Comput. Phys. Commun.* **2023**, *291*, 108831.
- (38) Mills, G.; Jónsson, H.; Schenter, G. K. Reversible work transition state theory: application to dissociative adsorption of hydrogen. *Surf. Sci.* **1995**, *324*, 305–337.
- (39) Jónsson, H.; Mills, G.; Jacobsen, K. W. *Classical and Quantum Dynamics in Condensed Phase Simulations*; World Scientific, 1998; pp 385–404.
- (40) Kolsbjergr, E. L.; Groves, M. N.; Hammer, B. An automated nudged elastic band method. *J. Chem. Phys.* **2016**, *145*, 094107.
- (41) Zimmerman, P. M. Growing string method with interpolation and optimization in internal coordinates: Method and examples. *J. Chem. Phys.* **2013**, *138*, 184102.
- (42) Zimmerman, P. M. Single-ended transition state finding with the growing string method. *J. Comput. Chem.* **2015**, *36*, 601–611.
- (43) Maeda, S.; Taketsugu, T.; Morokuma, K. Exploring transition state structures for intramolecular pathways by the artificial force induced reaction method. *J. Comput. Chem.* **2014**, *35*, 166–173.
- (44) Maeda, S.; Harabuchi, Y.; Takagi, M.; Taketsugu, T.; Morokuma, K. Artificial force induced reaction (AFIR) method for exploring quantum chemical potential energy surfaces. *Chem. Rec.* **2016**, *16*, 2232–2248.
- (45) Nett, A. J.; Zhao, W.; Zimmerman, P. M.; Montgomery, J. Highly active nickel catalysts for C–H functionalization identified through analysis of off-cycle intermediates. *J. Am. Chem. Soc.* **2015**, *137*, 7636–7639.
- (46) Khomutnyk, Y. Y.; Argüelles, A. J.; Winschel, G. A.; Sun, Z.; Zimmerman, P. M.; Nagorny, P. Studies of the mechanism and origins of enantioselectivity for the chiral phosphoric acid-catalyzed stereoselective spiroketalization reactions. *J. Am. Chem. Soc.* **2016**, *138*, 444–456.
- (47) Hayashi, H.; Katsuyama, H.; Takano, H.; Harabuchi, Y.; Maeda, S.; Mita, T. In silico reaction screening with difluorocarbene for N-difluoroalkylative dearomatization of pyridines. *Nat. Synth.* **2022**, *1*, 804–814.
- (48) Wu, Y.; He, C.; Zhang, W. Capture-backdonation-recapture mechanism for promoting N₂ reduction by heteronuclear metal-free double-atom catalysts. *J. Am. Chem. Soc.* **2022**, *144*, 9344–9353.
- (49) Wang, Z.; Goddard, W. A., III; Xiao, H. Potential-dependent transition of reaction mechanisms for oxygen evolution on layered double hydroxides. *Nat. Commun.* **2023**, *14*, 4228.
- (50) Li, Y.; Xue, B.; Yang, J.; Jiang, J.; Liu, J.; Zhou, Y.; Zhang, J.; Wu, M.; Yuan, Y.; Zhu, Z.; Wang, Z. J.; et al. Azobenzene as a photoswitchable mechanophore. *Nat. Chem.* **2024**, *16*, 446–455.
- (51) Young, T. A.; Silcock, J. J.; Sterling, A. J.; Duarte, F. autodE: automated calculation of reaction energy profiles—application to organic and organometallic reactions. *Angew. Chem., Int. Ed.* **2021**, *60*, 4266–4274.
- (52) Suleimanov, Y. V.; Green, W. H. Automated discovery of elementary chemical reaction steps using freezing string and Berny optimization methods. *J. Chem. Theory Comput.* **2015**, *11*, 4248–4259.
- (53) Simm, G. N.; Reiher, M. Context-driven exploration of complex chemical reaction networks. *J. Chem. Theory Comput.* **2017**, *13*, 6108–6119.
- (54) Grambow, C. A.; Jamal, A.; Li, Y.-P.; Green, W. H.; Zádor, J.; Suleimanov, Y. V. Unimolecular reaction pathways of a γ -ketohydroperoxide from combined application of automated reaction discovery methods. *J. Am. Chem. Soc.* **2018**, *140*, 1035–1048.
- (55) Zhao, Q.; Savoie, B. M. Simultaneously improving reaction coverage and computational cost in automated reaction prediction tasks. *Nat. Comput. Sci.* **2021**, *1*, 479–490.
- (56) Ramos-Sánchez, P.; Harvey, J. N.; Gámez, J. A. An automated method for graph-based chemical space exploration and transition state finding. *J. Comput. Chem.* **2023**, *44*, 27–42.
- (57) Zhao, Q.; Savoie, B. M. Deep reaction network exploration of glucose pyrolysis. *Proc. Natl. Acad. Sci. U.S.A.* **2023**, *120*, No. e2305884120.
- (58) Zhao, Q.; Garimella, S. S.; Savoie, B. M. Thermally Accessible Prebiotic Pathways for Forming Ribonucleic Acid and Protein Precursors from Aqueous Hydrogen Cyanide. *J. Am. Chem. Soc.* **2023**, *145*, 6135–6143.
- (59) Iron, M. A.; Janes, T. Evaluating transition metal barrier heights with the latest density functional theory exchange–correlation functionals: The MOBH35 benchmark database. *J. Phys. Chem. A* **2019**, *123*, 3761–3781.
- (60) Semidalas, E.; Martin, J. M. The MOBH35 metal–organic barrier heights reconsidered: Performance of local-orbital coupled cluster approaches in different static correlation regimes. *J. Chem. Theory Comput.* **2022**, *18*, 883–898.
- (61) Grambow, C. A.; Pattanaik, L.; Green, W. H. Reactants, products, and transition states of elementary chemical reactions based on quantum chemistry. *Sci. Data* **2020**, *7*, 137.
- (62) Zhao, Q.; Vaddadi, S. M.; Woulfe, M.; Ogunfowora, L. A.; Garimella, S. S.; Isayev, O.; Savoie, B. M. Comprehensive exploration of graphically defined reaction spaces. *Sci. Data* **2023**, *10*, 145.
- (63) Robertson, C.; Habershon, S. Simple position and orientation preconditioning scheme for minimum energy path calculations. *J. Comput. Chem.* **2021**, *42*, 761–770.
- (64) Zhao, Q.; Hsu, H.-H.; Savoie, B. M. Conformational sampling for transition state searches on a computational budget. *J. Chem. Theory Comput.* **2022**, *18*, 3006–3016.

- (65) Kim, S.; Woo, J.; Kim, W. Y. Diffusion-based generative AI for exploring transition states from 2D molecular graphs. *Nat. Commun.* **2024**, *15*, 341.
- (66) Dewyer, A. L.; Zimmerman, P. M. Finding reaction mechanisms, intuitive or otherwise. *Org. Biomol. Chem.* **2017**, *15*, 501–504.
- (67) Zimmerman, P. M. Automated discovery of chemically reasonable elementary reaction steps. *J. Comput. Chem.* **2013**, *34*, 1385–1392.
- (68) Zimmerman, P. M. Navigating molecular space for reaction mechanisms: an efficient, automated procedure. *Mol. Simul.* **2015**, *41*, 43–54.
- (69) Jacobson, L. D.; Bochevarov, A. D.; Watson, M. A.; Hughes, T. F.; Rinaldo, D.; Ehrlich, S.; Steinbrecher, T. B.; Vaitheeswaran, S.; Philipp, D. M.; Halls, M. D.; et al. Automated Transition State Search and Its Application to Diverse Types of Organic Reactions. *J. Chem. Theory Comput.* **2017**, *13*, 5780–5797.
- (70) Geiger, J.; Settels, V.; Deglmann, P.; Schäfer, A.; Bergeler, M. Automated input structure generation for single-ended reaction path optimizations. *J. Comput. Chem.* **2022**, *43*, 1662–1674.
- (71) Precomplex Generator; BASF, 2024. https://github.com/bASF/precomplex_generator (accessed 12 09, 2024).
- (72) YARP; zhaoqy, 2022. <https://github.com/zhaoqy1996/YARP/> (accessed 01 26, 2025).
- (73) Zhao, Q.; Savoie, B. M. Algorithmic explorations of unimolecular and bimolecular reaction spaces. *Angew. Chem., Int. Ed.* **2022**, *61*, No. e202210693.
- (74) Birkholz, A. B.; Schlegel, H. B. Using bonding to guide transition state optimization. *J. Comput. Chem.* **2015**, *36*, 1157–1166.
- (75) Wang, L.-P.; Titov, A.; McGibbon, R.; Liu, F.; Pande, V. S.; Martínez, T. J. Discovering chemistry with an ab initio nanoreactor. *Nat. Chem.* **2014**, *6*, 1044–1048.
- (76) Fukui, K. The path of chemical reactions—the IRC approach. *Acc. Chem. Res.* **1981**, *14*, 363–368.
- (77) Stewart, J. J. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **2007**, *13*, 1173–1213.
- (78) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671.
- (79) Smith, G. D.; Jaffe, R. L. Quantum chemistry study of conformational energies and rotational energy barriers in n-alkanes. *J. Phys. Chem.* **1996**, *100*, 18718–18724.
- (80) Gruzman, D.; Karton, A.; Martin, J. M. Performance of Ab Initio and Density Functional Methods for Conformational Equilibria of C_nH_{2n+2} Alkane Isomers (n = 4–8). *J. Phys. Chem. A* **2009**, *113*, 11974–11983.
- (81) Ehlert, S.; Grimme, S.; Hansen, A. Conformational energy benchmark for longer n-alkane chains. *J. Phys. Chem. A* **2022**, *126*, 3521–3535.
- (82) Rappe, A. K.; Casewit, C. J.; Colwell, K.; Goddard, W. A., III; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.
- (83) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490–519.
- (84) Neese, F. The ORCA program system. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 73–78.
- (85) Ásgeirsson, V.; Birgisson, B. O.; Björnsson, R.; Becker, U.; Neese, F.; Ripplinger, C.; Jónsson, H. Nudged elastic band method for molecular reactions using energy-weighted springs combined with eigenvector following. *J. Chem. Theory Comput.* **2021**, *17*, 4929–4945.
- (86) Liu, Y.; Qi, H.; Lei, M. Improved elastic image pair method for finding transition states. *J. Chem. Theory Comput.* **2023**, *19*, 2410–2417.
- (87) Aldaz, C.; Kammeraad, J. A.; Zimmerman, P. M. Discovery of conical intersection mediated photochemistry with growing string methods. *Phys. Chem. Chem. Phys.* **2018**, *20*, 27394–27405.
- (88) Steinmetzer, J.; Kupfer, S.; Gräfe, S. pysisyphus: Exploring potential energy surfaces in ground and excited states. *Int. J. Quantum Chem.* **2021**, *121*, No. e26390.
- (89) Schlegel, H. B. Optimization of equilibrium geometries and transition structures. *J. Comput. Chem.* **1982**, *3*, 214–218.
- (90) Frisch, M. J.; et al. *Gaussian 09 Revision D.01*; Gaussian Inc.: Wallingford CT, 2013.
- (91) Lee, K.; Lee, J.; Park, S. Output for Facilitating Transition State Search with Minimal Conformational Sampling Using Reaction Graph. **2024**; DOI: [10.6084/m9.figshare.27132603](https://doi.org/10.6084/m9.figshare.27132603).
- (92) Pattanaik, L.; Ingraham, J. B.; Grambow, C. A.; Green, W. H. Generating transition states of isomerization reactions with deep learning. *Phys. Chem. Chem. Phys.* **2020**, *22*, 23618–23626.
- (93) Choi, S. Prediction of transition state structures of gas-phase chemical reactions via machine learning. *Nat. Commun.* **2023**, *14*, 1168.
- (94) Makoś, M. Z.; Verma, N.; Larson, E. C.; Freindorf, M.; Kraka, E. Generative adversarial networks for transition state geometry prediction. *J. Chem. Phys.* **2021**, *155*, 024116.
- (95) Pracht, P.; Bohle, F.; Grimme, S. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* **2020**, *22*, 7169–7192.
- (96) Yang, M.; Zou, J.; Wang, G.; Li, S. Automatic reaction pathway search via combined molecular dynamics and coordinate driving method. *J. Phys. Chem. A* **2017**, *121*, 1351–1361.
- (97) Zhao, Q. YARP reaction database. 2021; <https://figshare.com/articles/dataset/YARP/14766624>.
- (98) Frisch, M. J.; et al. *Gaussian 16. Revision B.01.*; Gaussian Inc: Wallingford CT, 2016.
- (99) Zhang, Y.; Xu, C.; Lan, Z. Automated Exploration of Reaction Networks and Mechanisms Based on Metadynamics Nanoreactor Simulations. *J. Chem. Theory Comput.* **2023**, *19*, 8718–8731.
- (100) Wang, L.-P.; McGibbon, R. T.; Pande, V. S.; Martínez, T. J. Automated discovery and refinement of reactive molecular dynamics pathways. *J. Chem. Theory Comput.* **2016**, *12*, 638–649.
- (101) Lei, T.; Guo, W.; Liu, Q.; Jiao, H.; Cao, D.-B.; Teng, B.; Li, Y.-W.; Liu, X.; Wen, X.-D. Mechanism of graphene formation via detonation synthesis: a DFTB nanoreactor approach. *J. Chem. Theory Comput.* **2019**, *15*, 3654–3665.
- (102) Zhang, S.; Makóś, M. Z.; Jadrich, R. B.; Kraka, E.; Barros, K.; Nebgen, B. T.; Tretiak, S.; Isayev, O.; Lubbers, N.; Messerly, R. A. Exploring the frontiers of condensed-phase chemistry with a general reactive machine learning potential. *Nat. Chem.* **2024**, *16*, 727–734.
- (103) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **2017**, *8*, 3192–3203.
- (104) Gao, X.; Ramezanghorbani, F.; Isayev, O.; Smith, J. S.; Roitberg, A. E. TorchANI: a free and open source PyTorch-based deep learning implementation of the ANI neural network potentials. *J. Chem. Inf. Model.* **2020**, *60*, 3408–3415.
- (105) Devereux, C.; Smith, J. S.; Huddleston, K. K.; Barros, K.; Zubatyuk, R.; Isayev, O.; Roitberg, A. E. Extending the applicability of the ANI deep learning molecular potential to sulfur and halogens. *J. Chem. Theory Comput.* **2020**, *16*, 4192–4202.
- (106) Zheng, P.; Zubatyuk, R.; Wu, W.; Isayev, O.; Dral, P. O. Artificial intelligence-enhanced quantum chemical method with broad applicability. *Nat. Commun.* **2021**, *12*, 7022.
- (107) Chen, Y.; Ou, Y.; Zheng, P.; Huang, Y.; Ge, F.; Dral, P. O. Benchmark of general-purpose machine learning-based quantum mechanical method AIQM1 on reaction barrier heights. *J. Chem. Phys.* **2023**, *158*, 074103.