Automated Discovery of Elementary Chemical Reaction Steps Using Freezing String and Berny Optimization Methods

Yury V. Suleimanov*,[†,‡] and William H. Green[‡]

[†]Computation-based Science and Technology Research Center, Cyprus Institute, 20 Kavafi Street, Nicosia 2121, Cyprus
[‡]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

S *Supporting Information*

ABSTRACT: We present a simple protocol which allows fully automated discovery of elementary chemical reaction steps using in cooperation double- and single-ended transition-state optimization algorithms—the freezing string and Berny optimization methods, respectively. To demonstrate the utility of the proposed approach, the reactivity of several single-molecule systems of combustion and atmospheric chemistry importance is investigated. The proposed algorithm allowed us to detect without any human intervention not only "known" reaction pathways, manually detected in the previous studies, but also new, previously "unknown", reaction pathways which involve significant atom rearrangements. We believe that applying such a systematic approach to elementary reaction path finding will greatly accelerate the discovery of new chemistry and will lead to more accurate computer simulations of various chemical processes.
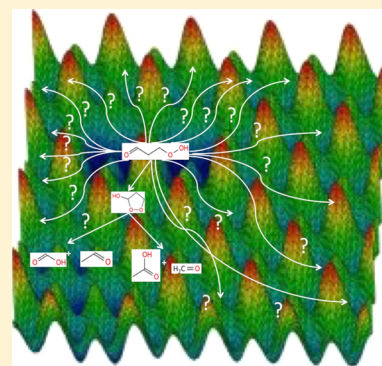
INTRODUCTION

Currently, new chemical understanding is driven mainly by the experimental discovery of new compounds and their reactivity. This experimental work is expensive and time consuming, but the payoff for discovering new chemical reactions is enormous in terms of both fundamental understanding and practical engineering. Examples of important new reactions discovered in recent decades include many advances that followed from the discoveries of olefin metathesis[1] and click chemistry.[2] Theoretical, algorithmic, and computational advances in quantum chemistry over the last two decades have brought it to the stage where it is widely applied for qualitative interpretation of such new experimental findings and increasingly used to construct detailed quantitative descriptions of chemical reaction networks.

The main parameters in chemical kinetic models—equilibrium constants and rate coefficients—are increasingly derived from quantum mechanical calculations on stationary points (local minima and first-order saddle points) along the potential energy surface (PES). Local minima, i.e., reactants, intermediates, and products, are generally easy to characterize due to simple bonding and because the negative of the gradient along the PES always points downhill. Several robust algorithms were developed for identifying local minima.[3] Saddle points, which correspond to transition states, remain a challenge for systematic characterization due to nonstandard variations in reactive centers and because a transition structure optimization must step uphill in one direction and downhill in all other orthogonal directions. At present, most saddle points are found based on human intuition, derived primarily from experience on analogous reactions; the human provides a very good initial guess at the saddle point geometry, and then

various algorithms (so-called single-ended) are used to refine the initial guess geometry to find the true saddle point.[4−8] This has typically been applied to cases where humans expect a reaction to connect the reactant(s) and product(s), and so it also rarely discovers anything surprising. This conventional approach turns out to be quite good for finding "expected reaction" saddle points but rarely discover anything that is unexpected. As an alternative to human input, the initial guess structure for the saddle point can be generated by methods (so-called double-ended searches), which reconstruct the reaction path between known reactant and product sides.[9−14]

Most large chemical kinetic models are constructed in ways which do not allow any unexpected reactions to be included; this is true whether the model is constructed by hand or using automatic mechanism generators.[15] Incorporation of even a single new reaction in chemical kinetic models can significantly alter these models since it almost certainly will be the first member of a new large family of related reactions. For instance, several unexpected low-barrier reactions for JP-10 pyrolysis[16] and for decomposition of the ketohydroperoxides,[17] which are critical intermediates in alkane autoxidation and ignition chemistry, were recently discovered in our group. Including the new unexpected reactions and analogous new reactions in the kinetic models significantly improved the reliability of the kinetic model predictions.[18] All this experience suggests that if we search for them systematically, we can expect to find many "unexpected" reactions, which are omitted from existing chemistry models,

and that incorporating these new reactions will significantly improve the reliability of the model predictions.

Various methods for automated reaction pathway search, which do not rely on any initial guess for saddle points, have been proposed. Some time ago, Irikura and Johnson invented contour-following methods and used them to discover some organic reactions using semi-empirical methods.[19] Several researchers have been pursuing methods based on reactive molecular dynamics,[20−31] most commonly using force fields to reduce the computational expense. However, it is questionable whether parametrized methods such as semi-empirical calculations or force fields will be reliable in regions of the PES where unexpected reactions occur. Car−Parrinello or conventional density functional theory (DFT) calculations can be used instead, but this makes it quite expensive to compute even a few nanoseconds of time evolution using ordinary MD, so other researchers have explored accelerating the search procedure using metadynamics.[32,33] However, even the best methods based on molecular dynamics involve a large number of calculations at geometries far from transition state, so they may not be very efficient at discovering new chemistry. To increase the efficiency of the reaction−discovery process, Maeda, Morokuma, and co-workers proposed anharmonic downward distortion following (ADDF) method (for A → X type isomerization and A → X + Y type dissociation pathways) and artificial force-induced reaction (AFIR) method (for A + B → X(+Y) type associative pathways).[34−36] These methods start from a given reactant structure and focus on the sampling of possible reaction pathways by forcing the reaction to occur toward different products. For similar purposes, an unbiased PES searching method via the stochastic surface walking method, where the second derivative information is taken into account, was proposed by Liu and co-workers.[37,38] AFIR has been shown to be very successful in establishing the reaction network of some organic molecules[37,39] and demonstrated great potential for reactivity prediction. Both AFIR and ADDF follow special paths which tend to pass near saddle points; configurations in such regions can be further optimized to corresponding true saddle points without artificial forces/distortions facilitating accurate rate calculations of the newly discovered reactions.

Zimmerman and co-workers have been working on alternative approaches based on bond-forming rules, which first detect product channels of potential interest and then decide on which to apply saddle point searches. (Usually, these are screened for thermodynamic stability using standard DFT methods.)[40,41] Then, reaction path-finding methods based on double-ended string methods provide the actual energy profile which connects the initial structure to the new products and detects the actual activation barriers. In recent years, the development of string methods has focused on the growing string method (GSM). Advances in the GSM have allowed it to simultaneously locate reaction paths and the corresponding transition states without construction of exact Hessians.[14,42,43] In addition, single-ended modification of the original GSM has been proposed which, similar to AFIR and ADDF, starts from a single structure and drives it to saddle point region(s).[44] The freezing string method (FSM) was proposed as an alternative double-ended algorithm that attempts to avoid the costs associated with the iterative nature of the GSM.[13] It has been shown that the FSM is able to generate a guess for the saddle point structure with significantly lower cost in practical applications.[13,42] However,

the FSM does not reproduce the reaction path exactly, and convergence to the correct saddle point is not guaranteed. Therefore, its practical implementation in automated reaction pathways search represents an important question, which if successful could motivate its further development and more active application. We note that in addition to the GSM and FSM, a variety of double-ended methods have been proposed, including several varieties of nudged elastic band methods,[9,10,12] although to our knowledge these have been used almost exclusively to search for TSs for expected reactions with some human assistance, rather than to automatically search for unexpected reactions.

In the present work, we propose a systematic approach for discovery of new chemical reactions with minimum human effort, which is based on a computational strategy similar to the one developed by Zimmerman and co-workers. The method advances recent work on saddle point optimization algorithms by making use of both single- and double-ended methods in a fully automated procedure. Choosing from a great variety of possible methods,[4−14] we employ a combination of the Berny optimization algorithm as implemented and improved by Schlegel[6−8] (a reliable and widely used quasi-Newton single-ended method) and the freezing string method[13] (double-ended method with low computational cost) as both methods represent certain desirable computational features and were readily available to us at the time we started this project. We also make use of the Benson's group additivity approach[45] for initial thermochemistry calculations instead of computationally more demanding DFT calculations and use tools taken from automated reaction mechanism generation to enumerate all the possible product channels. We find that this combination works fairly well for several single-molecule systems of combustion and atmospheric chemistry importance. We were able to detect not only all "known" reaction pathways, manually detected in the previous studies, but also new, previously "unknown", reaction pathways, which involve significant atom rearrangements. The present work is complementary to recent developments in computational chemistry,[15,46] which demonstrate a tendency of switching from a traditional secondary role of theory in interpreting experimental findings in chemistry to its use as a powerful predictive tool for analysis of chemical reactivity, discovery of chemical reaction pathways, and generation of reaction mechanisms prior to any experiments.

■ THEORY

Matrix Representation of Species and Reactions. There are different ways to describe mathematically compounds presented in a system.[47−54] In the present algorithm, we chose the Bond Electron (BE) matrices[47,48] as one of the most vivid and convenient representations. A BE matrix $\mathbf{R}$ for a compound with n atoms $r_1, r_2..r_n$ is a square matrix of dimension n. The ith row of this matrix corresponds to electron configuration of atom r_i. The R_{ij} element represents the number of covalent bonds between atoms r_i and r_j, and matrix $\mathbf{R}$ is symmetric, i.e. $R_{ij} = R_{ji}$. The diagonal elements R_{ii} represent the number of free valence electrons which are not involved in bonds. The sum over the entries of the ith row or column of a BE matrix is the number of valence electrons which belong to the atom r_i.

The total matrix of the initial reactive system $\mathbf{R}_{\text{sum}}$ is depicted by a block matrix, each block representing one

compound. A chemical reaction is the conversion of the system by the redistribution of valence electrons. It is represented by transformation of the initial matrix $\mathbf{R}_{sum}$ by the addition of the reaction matrix $\mathbf{A}$ to yield a product total matrix $\mathbf{P}_{sum}$ with altered connectivity. The reaction matrix $\mathbf{A}$ is also symmetric, and its elements represent the appearance (positive value) or the disappearance (negative value) of the localized valence electrons. This matrix should conserve the total number of electrons in the system, i.e., $\sum_{i,j} a_{ij} = 0$. Usually in elementary reactions $|a_{ij}| \leq 2$ since the change by two implies a significant chemical process, for instance, formation or rupture of a double bond in a single elementary step.

After applying the reaction matrix $\mathbf{A}$, the product matrix $\mathbf{P}$ can then be rearranged to form a block structure to represent the product molecules. If the product matrix satisfies the mathematical and chemical constraints, the reaction pathway is considered to be feasible and is added to the list of channels for further thermochemistry calculations.

Reducing the Size of the Search Space. *Thermochemistry Calculations.* Even for a small hydrocarbon system, the number of reaction pathways which can be generated using the BE matrix representation described above can be massive, increasing exponentially with the number of carbon atoms. The thermodynamic reaction control principle can be implemented to narrow the search space; highly endothermic reactions are likely to be too slow to be important. The relative stability of the products can be estimated using their thermochemical properties, and reaction channels with less stable products can be eliminated. However, on-the-fly quantum chemical calculations of thermochemical properties for hundreds (or even thousands) species using, e.g., DFT methods, can also be computationally very demanding. One efficient alternative way to do so is to use group contribution methods that estimate the thermochemistry of a molecule based on the submolecular fragments present in the molecule. The Benson group additivity framework is such an example of a group contribution method that has proven to provide accurate estimates of the ideal gas thermochemistry for a large range of molecules. Benson's group additivity approach[45] divides a molecule into functional groups, and the contribution of each functional group to the overall thermochemistry is included. The group additivity enthalpy estimate is typically within a few kcal/mol[55] of the truth, sufficient to reliably separate very endothermic reactions from the rest.

Generating 3D Geometries. While DFT methods can, in principle, be implemented to generate and optimize the starting and final 3D geometries, the large number of reaction channels can prohibit such calculations. A much faster route is to use a molecular mechanics force field. Though, in principle, any type of reasonable force field can be used, we choose Merck Molecular force fields (MMFF94) developed by Merck Research Laboratories,[56−60] which were specifically optimized for a wide range of organic chemistry calculations (from very small molecules to proteins). The method includes parameters from high-quality quantum calculations (rather than experimental data) for a wide range of atom types including those presented in the examples studied in this work.

Locating Transition States. As mentioned in the Introduction, methods for a transition-state structure search can be classified as single-ended and double-ended.[4,5] The single-ended methods refine an initial guess of the transition state to the exact answer by exploring the PES using local gradient and usually the second derivative information. The Berny saddle point optimization algorithm is one of these methods, which is widely used in *ab initio* calculations primarily due its reliability.[6−8] Berny geometry optimization starts with an initial guess for the second derivative matrix (Hessian), which is determined using connectivity derived from atomic radii and a simple valence force field.[7,8] The approximate Hessian matrix is improved at each point using the first derivatives and energies computed along the optimization pathway. The main input for such algorithm is a guess structure that should be located very close to the transition state in order to succeed. At present, initial guess geometries are usually found based on human intuition, and automated generation of good initial guess transition-state structures is the main point of the procedure proposed here.

The double-ended methods start from the reactants and products and work from both sides to find the transition structure and the reaction path. These include the nudged elastic band method,[9,10] string method,[11,12] GSM,[14] and FSM.[13] All these algorithms generate a sequence of configurations located between the reactant and products geometries. In the present algorithm, we choose the FSM as it does not require Hessian calculations and provides one of the most inexpensive algorithms for determining a path with a specified number of intermediate structures (nodes) connecting the known reactant(s) and product(s). The FSM starts from both ends of the path and adds nodes irreversibly until the two ends join. The FSM partially optimizes intermediate structures orthogonally to the reaction path and then freezes them before a new pair of points along the reaction path is added. The number of gradients required to locate a guess saddle point using the FSM is rather small.

Since the reactant and product geometries, corresponding to initial and final minima, are known from the matrix representation, energy cutoff and force field optimization described above, automatic path-finding tools, such as the freezing string method, can characterize a reaction coordinate joining the end-points. The highest point on the pathway becomes a good initial guess for subsequent refinement of the transition structure using the single-ended method such as Berny algorithm.

Summary of Methodology and Computational Details. All the methods just described can be combined in a systematic and automated procedure for identifying reaction pathways. This is summarized in the flowchart presented in Figure 1. Only initial reactant(s) geometry is required at the start. It is converted to the reactant matrix $\mathbf{R}$. Different reaction matrices $\mathbf{A}$ are then generated, which break up to N bonds and make up to M bonds, where N and M are defined by the user. The number of possibilities rises combinatorially with N and M. In the benchmark examples below, we constrain $1 < N < 4$ and $1 < M < 4$. We constrain $N, M < 4$ to reduce computational demands and $N, M > 1$ since most of those simple reactions are well understood. To reduce the computational effort, we restricted the products to those which can be drawn with a Lewis structure with a fixed number of valence electrons for each element —4 for carbon, 2 plus 2 lone pairs for oxygen, 1 for hydrogen, and 3 plus a lone pair for nitrogen.

In the present work, we restrict all the elements R_{ij} and P_{ij} in the reactant and product matrices to be integers. This can potentially lead to issues for products with resonant
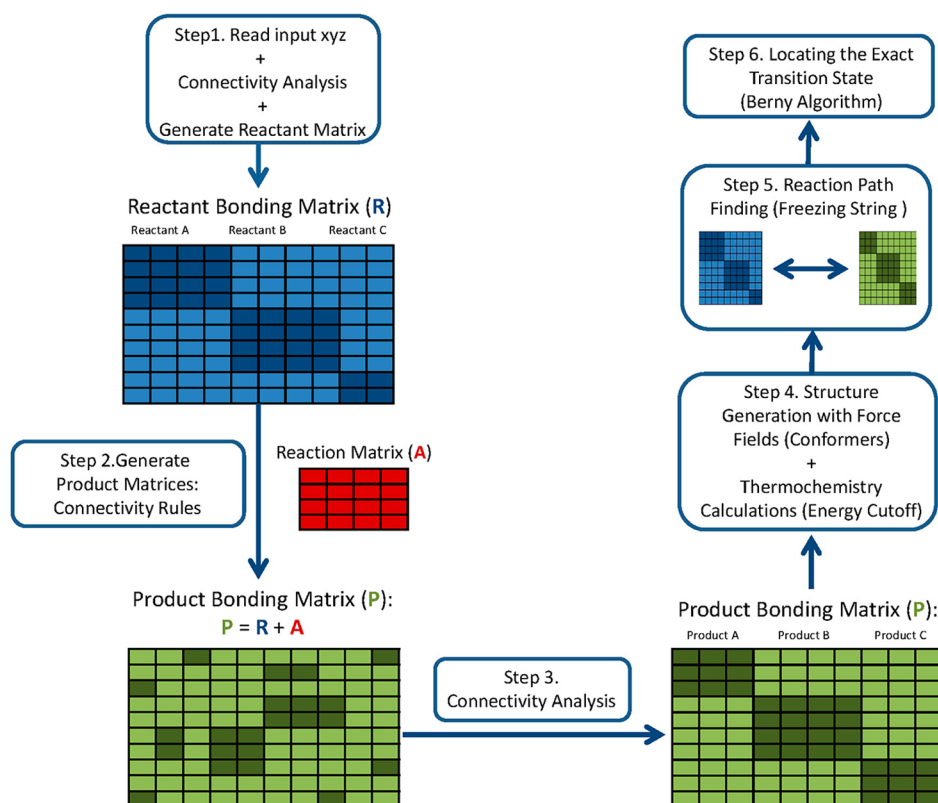
Figure 1. Flowchart for automated discovery of elementary chemical reaction steps using molecular mechanics and quantum chemistry calculations. See the Theory section for details of each step.

structures. We also demand a number of unpaired electrons on an atom not exceed 1. This is certainly ruling out some known species, e.g., biradicals such as methylene and $O(^3P)$, but we can always relax this restriction and include them in future studies. Since BE matrices are invariant with certain permutations, several BE matrices can correspond to chemically equivalent compounds. However, in order not to omit important reaction pathways, all combinations are taken into account initially. We remove any duplicates before the molecular mechanics and quantum chemistry calculations are run.

The product matrices **P** are obtained and converted to block structure by simple linear algebra operations. Thermochemistry calculations are used to narrow the search space; during this step "known" and unstable products are detected. In the present study, a reaction is considered to be too endothermic to be interesting if the standard enthalpy of reaction (denoted as ΔH_r^0 in Tables 2−5) is higher than 20 kcal/mol. These calculations are performed using the Reaction Mechanism Generator (RMG) software package developed in our group.[15]

The next step is an automated saddle point search for "unknown" reaction pathways which lead to stable products. For quantum chemistry calculations, geometries of reactant(s) and product(s) are required. They are generated by converting corresponding blocks of the **P** matrices to SMILES strings (standard format for molecule mechanics calculations) and quick force field optimization using academic version of the MarvinBeans package from ChemAxon.[61] Special care is taken of the conversion between different molecule representation formats, eliminating duplicate channels. In the

present study, we employ the MMFF94 force field to determine the geometry of the products.

The BE matrix representation is identical for all conformers. However, the distinct 3d structure of conformers strongly affects the properties and the reactions of molecules. In order to take into account conformational effects, we used the MarvinBean's Conformer Plugin to generate a series of stable 3d structures (up to three in the present study), i.e., conformers for each compound. If more than one product molecule is formed in a reaction channel, their structures were first aligned to maximum coincidence with the reactant structure in non-mass weighted Cartesian coordinates, and then the length of the vector between the centers of mass of the products was increased in order to avoid overlaps between atoms in different product molecules.

For the determination of all the transition states, we used the FSM implemented in the Q-Chem software package with default spacing parameters (20 nodes, 6 perpendicular gradients per node).[62] In order not to repeat computations on duplicate structures, all input structures are compared prior to initiating the FSM calculations, and identical input files were eliminated. However, chemically equivalent channels with distinct 3d geometries were not excluded from the procedure in order not to omit conformers or atom arrangements which could potentially lead to low-energy reaction pathways. For channels where the FSM found an apparent single barrier along the reaction path (without any restrictions to the energy barrier height), these calculations were followed by a transition-state search using the Berny optimization algorithm implemented in the Gaussian 03 software package[63] starting from the FSM geometry. For each detected saddle point, we also perform intrinsic reaction path

(IRC) calculations which go downhill from the transition state to see which two minima it is connected to and to verify whether this saddle point corresponds to the reaction from R to P.

Geometry optimization and reaction path analysis for the examples herein were performed using M062X/6-311++G* level of theory, while single point energies at the stationary points were obtained using CCSD(T)/6-311++G* level of theory. Thermochemistry of the detected reactions was computed using CBS-QB3 level of theory. Note that the proposed procedure is not dependent on the use of any particular software package or level of theory. Our choice of such parameters as quantum chemistry methods, basis set, force fields, FSM parameters, number of conformers, and method for orientating product structures should be considered as reasonable first attempts which can be improved in the future as the community gains more experience.

Each set of quantum chemistry calculations—structure optimization, energy calculation and FSM calculation—is performed independently in parallel to exploit the full potential of supercomputers. All the computations are automated to reduce time-to-completion and also to reduce bias related to manual intervention to help the computer find a saddle point, which tend to find only "human-expected" known reactions. In order to demonstrate how such a fully automated procedure can accelerate the discovery of new chemistry, we use it to study several systems of combustion, oxidation, and atmospheric chemistry importance, which span a reasonably wide space of chemical reactivity.

■ RESULTS

For our first case study, we have chosen the chemistry of γ-ketohydroperoxide (HOOCH$_2$CH$_2$CHO). Ketohydroperoxides are the main source of radicals during the first stage of low-temperature ignition and are also formed in liquid phase oxidation. Combustion parameters, such as ignition delay, are very sensitive to the details of their chemistry.[17] Most previous combustion models assumed ketohydroperoxides do only one reaction: forming radicals by breaking the weak O—O bond. However, new chemical reactions of ketohydroperoxides which transform them into cyclic peroxides and then acids and carbonyl molecules have been recently characterized[17] in our group (so-called "Korcek reaction"),[64] suggesting that the chemistry of ketohydroperoxides is more complicated. To verify the generality of the proposed algorithm, we also studied other structures—1,2-dioxolan-3-ol (intermediate cyclic peroxide in the Korcek mechanism),[17] ethylene ozonide (an intermediate in reactions between the Criegee intermediate and carbonyl compounds important in atmospheric ozonolysis models),[65] and their nitrogen counterparts N-(hydroperoxymethyl) formamide, 1,2,4-dioxazolidin-3-ol, and 1,2,4-dioxazolidine—model systems for oxidation chemistry of nitrogen-containing oligomers, important components of lubricants and fuel. We also analyzed the chemistry of penta-1,4-diene—a model system for an important functional group in polyunsaturated lipids.

For each initial structure, Table 1 shows the number of product channels generated using the BE representation. The results of thermochemistry analysis using the Benson group additivity approach of all product channels are also given in Table 1. Since the double-ended FSM calculations represent the major part of all computational expenses (~90% for the present examples), the details of these calculations, such as

Table 1. Number of Elementary Steps Investigated for Different Initial Reactants

initial reactant (SMILES)	total number of product channels identified	number of low energy channels		total number of gradient calls in double-ended FSM	number of successful (normal) terminations of FSM calculations (%)	number of automatically identified transition states (vs existing RMG database)[c]
		ΔH_0^0 (298 K) $\leq$ 0 kcal/mol	ΔH_f^0 (298 K) $\leq$ 20 kcal/mol			
γ-ketohydroperoxide (OOCCC=O)	77[a] 415[b]	39[a] 204[b]	53[a] 276[b]	210380	51	6 (RMG database:1)
N-(hydroperoxymethyl) formamide (OOCNC=O)	62[a] 293[b]	22[a] 84[b]	32[a] 123[b]	52565	48	8 (RMG database: 0)
ethylene ozonide (O1COOC1)	42[a] 82[b]	23[a] 59[b]	35[a] 80[b]	410871	71	3 (RMG database: 0)
1,2,4-dioxazolidine (N1COOC1)	56[a] 190[b]	35[a] 137[b]	49[a] 159[b]	361602	33	4 (RMG database: 0)
1,2-dioxolan-3-ol (OC1OOCC1)	90[a] 533[b]	40[a] 233[b]	64[a] 344[b]	265279	26	4 (RMG database: 2)
1,2,4-dioxazolidin-3-ol (OC1OOCN1)	74[a] 378[b]	43[a] 191[b]	59[a] 275[b]	70718	26	11 (RMG database: 0)
penta-1,4-diene (C=CCC=C)	96[a] 652[b]	45[a] 197[b]	64[a] 341[b]	181304	35	8 (RMG database: 3)

[a]Number of breaking bonds = 2; Number of forming bonds = 2. [b]Number of breaking bonds = 3; Number of forming bonds = 3. [c]From http://rmg.mit.edu.
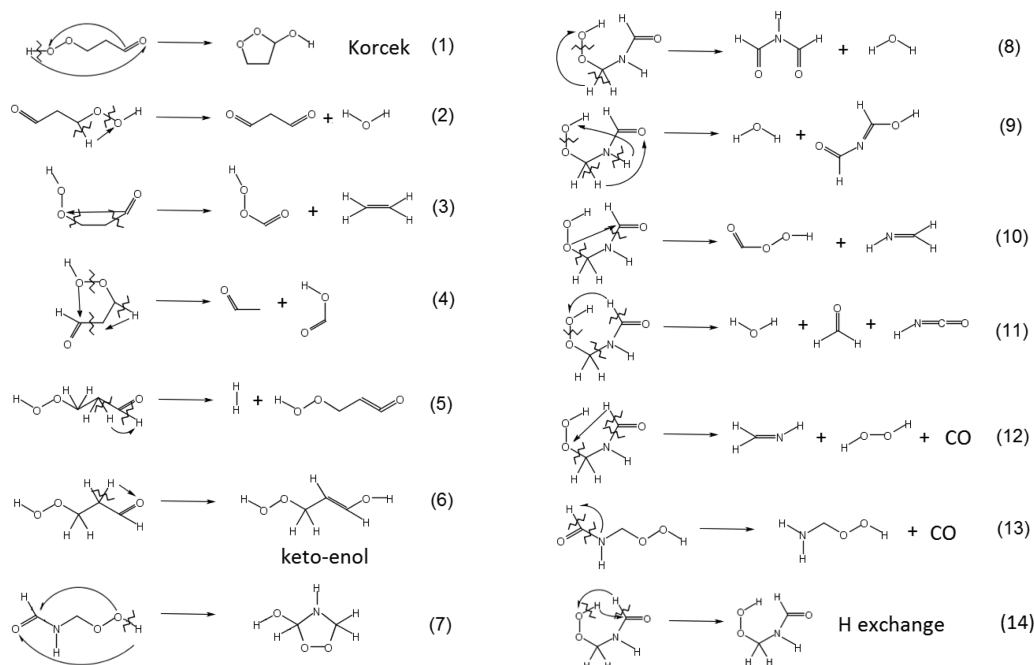
Figure 2. Automatically identified unimolecular reactions of γ-ketohydroperoxide (OOCCC=O) (1–6) and N-(hydroperoxymethyl) formamide (OOCNC=O) (7–14). Reactions (3), (4), and (9)–(12) are unusual and unexpected.
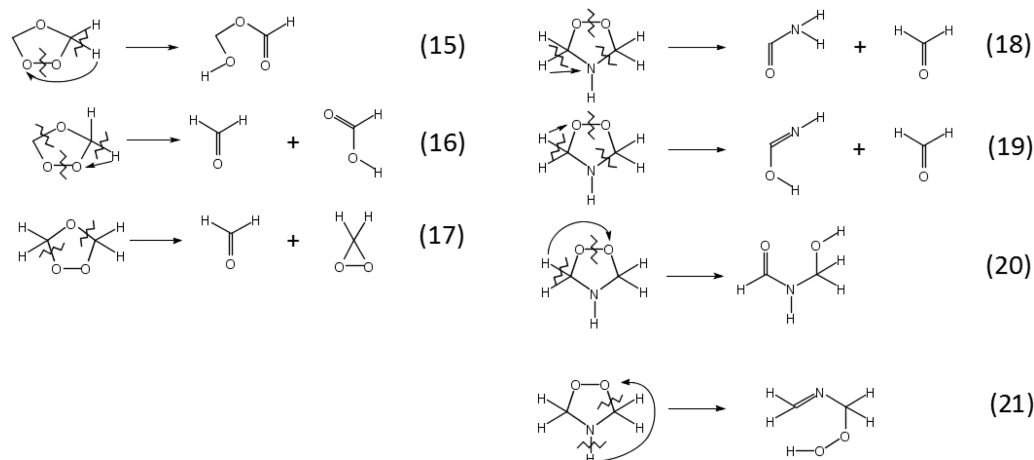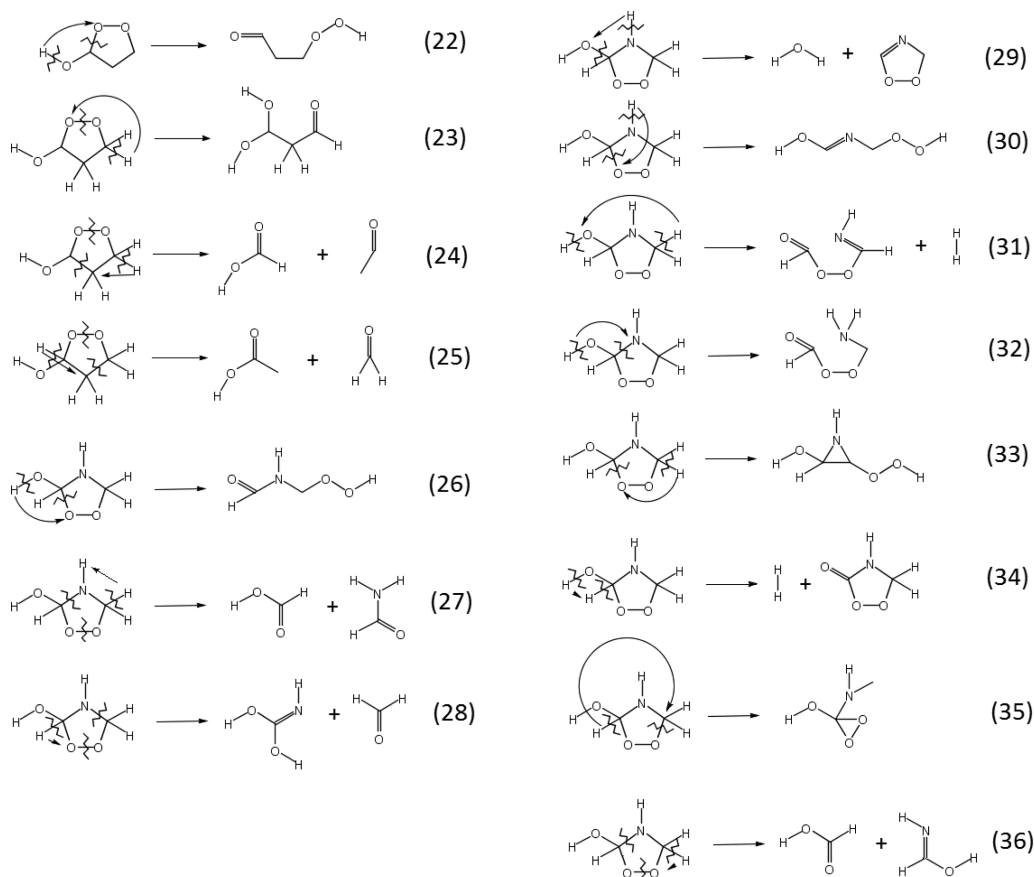

Figure 3. Automatically identified unimolecular reactions of ethylene secondary ozonide (O1COOC1) (15–17) and 1,2,4-dioxazolidine (N1COOC1) (18–21). Reactions (17) and (21) are particularly unusual and unexpected.

the total number of gradient calls in this procedure for all "thermodynamically preferable" channels and the success rate for successful execution (normal termination), are also included in Table 1. Note that these numbers are for all possible channels including chemically equivalent channels with distinct 3d geometries (such different conformers, up to three in the present work). We note that typical single successful FMS calculation required ∼100 gradient calls for all examples studied here. All reaction pathways identified are summarized in Figures 2–5 and Tables 2–5. The Cartesian coordinates of all the automatically detected saddle points are summarized in the Supporting Information.

For the first case study, γ-ketohydroperoxide (Table 1) shows that almost 500 product channels, which break and make two or three bonds, are possible from this small molecule. (Note that this includes distinct reactions which form the same products. Without repeated channels, this

number reduces to 109.). Half of the channels are exothermic, and almost 70% of the channels have $\Delta H_{rxn} < 20$ kcal/mol. At the same time, only one (!) product channel (1,2-dioxolan-3-ol) is reached by a "well-known" reaction type available in our current RMG database—the first step of the Korcek mechanism forming an intermediate cyclic peroxide.[17] As shown in Figure 2 (Reaction (1)), our simple algorithm was able to locate efficiently transition states for this "known" reaction, although it involves a significant rearrangement of atoms including cycle formation and hydrogen transfer between two oxygen atoms. Figure 2 also shows that the automatic search found transition states for five additional reaction pathways. Note that one of these channels leads directly to the products of the Korcek mechanism—Reaction (4) on Figure 2. However, Table 2 shows that the energy barrier (denoted as $\Delta E^{\#}$) for this channel is higher by almost 20 kcal, and therefore, rather than going directly over this

Figure 4. Automatically identified unimolecular reactions of 1,2-dioxolan-3-ol (OC1OOCC1) (22−25) and 1,2,4-dioxazolidin-3-ol (OC1OOCN1) (26−36). Reactions (30)−(33) and (35) are particularly unusual and unexpected.
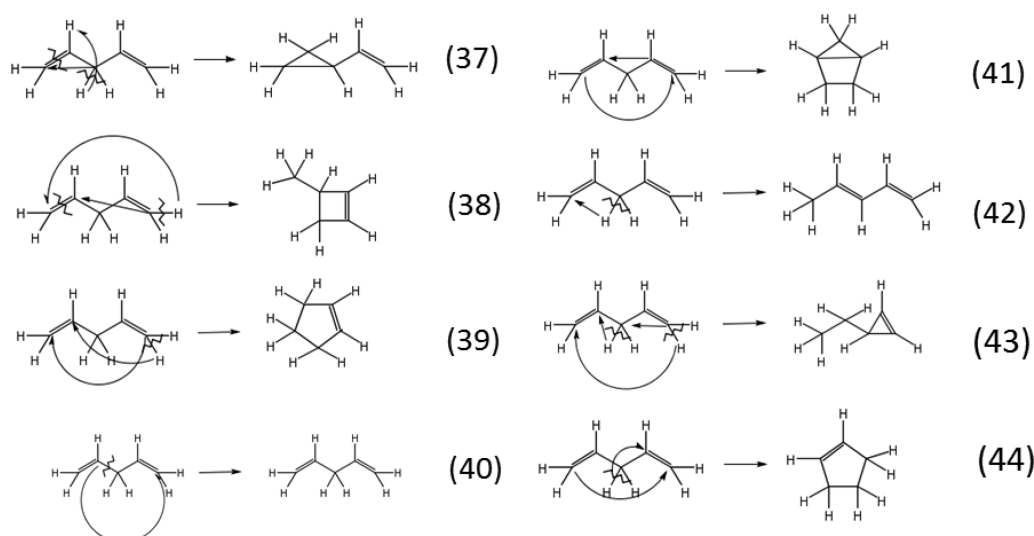
Figure 5. Automatically identified unimolecular reactions of 1,4-pentadiene (C=CCC=C) (37−44). Reactions (38)−(40) and (43) are particularly unusual and unexpected.

high mountain pass, sequential reactions via an intermediate valley (the cyclic peroxide) are expected to be the fastest reaction path. This channel was located while searching for a reaction path to a different product (see comments in Table 2). This example illustrates that our saddle point search procedure cannot guarantee convergence to the desired transition state. As a result of its noniterative nature, the

FSM cannot guarantee conservation of the exact IRC profile, and therefore, further optimization of the guess structure (the highest energy node) with Berny optimization can lead to structures which do not correspond to the initial reaction path. We also occasionally converged to saddle points which correspond to different initial reactants. More importantly, the majority of our reaction path searches do not converge to a

Table 2. Automatically Identified Unimolecular Reactions of γ-Ketohydroperoxide and N-(Hydroperoxymethyl) Formamide*

	Product(s)	$\Delta E^{\#a}$	ΔH_r^0(298 K)^b	Comments
1	(structure)	34.9	-10.4	Korcek reaction, see Ref. 17
2	(structure) + H_2O	60.2	-25.6	
3	(structure) + $H_2C=CH_2$	61.8	4.8	
4	(structure) + (structure)	52.9	-65.1	Intended product: (structure)
5	(structure) + H_2	84.1	26.1	Intended product: (structure) + H_2
6	(structure)	67.9	7.3	Keto-enol
7	(structure)	54.6	9.9	Korcek-type reaction
8	(structure) + H_2O	54.0	-37.2	
9	(structure) + H_2O	45.4	-16.9	Intended product: (structure)
10	(structure) + $HN=CH_2$	57.8	18.1	Intended product: (structure)
11	H_2O + $H_2C=O$ + $HN=C=O$	54.8	-16.1	Intended product: (structure) + H_2O
12	$H_2C=NH$ + H_2O_2 + CO	69.8	62.9	
13	(structure) + CO	83.3	12.6	Intended product: (structure)
14	(structure)	91.6	0.0	H exchange

In several cases, the computer was searching for a saddle point leading to intended products but converged to a transition state for a different product channel. Mechanistic details are given in Figure 2. ^{a}In kcal/mol; geometry optimization from Gaussian 03 using M062X level of theory, energy from Gaussian 03 using CCSD(T) level of theory. Basis set is 6-311++G in both calculations. ^{b}In kcal/mol; thermochemistry from Gaussian 03 using CBS-QB3 level of theory.
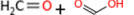
saddle point. This typically occurred because the energy profiles from the FSM calculations for these paths exhibited several energy barriers—not the single elementary-step transition state we were looking for or chemically infeasible pathways with very high energy gradients—and therefore, the Berny-type TS search was not performed. In these cases, it is possible that no single saddle point exists that connects reactant and product(s), or just that with our choices of parameters and options, the FSM is not robust enough to find one. Table 1 also shows that the rate of successful (normal) termination of the FSM step usually does not exceed 50% because of frequent premature stops due to high-energy gradients for infeasible channels. This rate is lower for initial cyclic structures that impose additional constrains on the possible product channels not taken into account by our general matrix representation. We note that usually 80−90% of FSM gradient calculations contribute to the normal execution runs (i.e., premature failure, if it happens, takes place at the beginning of the FSM calculations before we have wasted much computer time). However, we observed systematic failure of our procedure for two cyclic structures, ethylene-ozonide (O1COOC1) and 1,2,4-dioxazolidine (N1COOC1), when after a large number of gradient calculations the FSM calculation eventually stopped. In these unsuccessful (and wasteful) calculations, the reaction pathways represent slow opening of the cycle with further

chemically infeasible conversion. We also note that such failures during the FSM calculation happen more often when three bonds are broken/formed than when two bonds are broken/formed. For example, for penta-1,4-diene (C=CCC=C), there are ∼5 times more channels with three breaking/forming bonds channels than with two breaking/forming bonds but only ∼1.4 more channels were computed successfully (normal termination of the FSM). Clearly more robust methods are needed.

Upon examination of the reactivity of 1,2-dioxolan-3-ol—our second case study—we confirmed that fragmentation of the cyclic peroxide leads to two possible pairs of acid and aldehyde products (Reactions (24) and (25) in Figure 4) previously characterized in ref 17. We were also able to locate the reverse pathway which leads to the initial molecule (Reaction (22) in Figure 4) confirming that the proposed procedure is rather sustainable. The fact that the fully automated search reproduces the results of the previous manual studies of the Korcek mechanism—and also discovers a half-dozen additional transition states to products not found by that manual search—is encouraging.

Another interesting example is the chemistry of the Criegee intermediate with formaldehyde (third case study). These two species quickly recombine forming the more stable secondary ozonide intermediate (ethylene ozonide)[65] whose reactivity was investigated in the present work. All three reaction pathways identified for ethylene ozonide are presented in Figure 3 (Reactions (15)−(17)) and Table 3. In accordance

Table 3. Automatically Identified Unimolecular Reactions of Ethylene Secondary Ozonide (O1COOC1) and 1,2,4-Dioxazolidine (N1COOC1)*

	Product(s)	$\Delta E^{\#a}$	ΔH_r^0(298 K)^b	Comments
15	(structure)	42.4	-73.1	See Ref. 65
16	$H_2C=O$ + (structure)	46.3	-58.4	See Ref. 65
17	$H_2C=O$ + (structure)	45.1	27.6	Intended product: (structure)
18	(structure) + $H_2C=O$	47.4	-60.8	
19	(structure) + $H_2C=O$	48.1	-49.3	
20	(structure)	45.8	-72.2	
21	(structure)	66.6	12.4	

*For mechanistic details see Figure 3. ^{a}In kcal/mol; geometry optimization from Gaussian 03 using M062X level of theory, energy from Gaussian 03 using CCSD(T) level of theory. Basis set is 6-311++G*in both calculations. ^{b}In kcal/mol; thermochemistry from Gaussian 03 using CBS-QB3 level of theory.

with the previous manual search,[65] we find that two low energy pathways for ethylene ozonide decomposition are formation of hydroxylmethyl formate by breaking the weak O−O bond (the lowest energy saddle point) and formation of formaldehyde and formic acid. In addition to the reactions found in ref 65, we also find that ethylene ozonide can decompose to formaldehyde and oxirane with similar barrier height.

We also repeated the calculations for similar structures with nitrogen-substituted analogues: N-(hydroperoxymethyl) formamide, 1,2,4-dioxazolidin-3-ol, and 1,2,4-dioxazolidine. Table

1 shows that the chemistry of these N-containing compounds is even less known—no unimolecular channels with two and three bonds breaking/forming were found in the RMG database. However, we were able to identify and characterize several reaction pathways using our automated procedure. The results are summarized in Tables 2–4 and Figures 2–4. They

Table 4. Automatically Identified Unimolecular Reactions of 1,2-Dioxolan-3-ol (OC1OOCC1) and 1,2,4-Dioxazolidin-3-ol (OC1OOCN1)*

	Product(s)	$\Delta E^{\#a}$	$\Delta H_r^0(298\ K)^b$	Comments
22		45.1	10.4	See Ref. 17
23		77.0	-52.2	
24		38.2	-54.7	See Ref. 17
25		41.3	-53.2	See Ref. 17
26		17.6	-10.4	
27		27.1	-79.8	
28		29.0	-46.9	
29		38.6	2.8	
30		40.2	-1.5	
31		52.2	12.1	
32		54.9	-9.6	
33		64.1	19.5	
34		68.7	-2.3	
35		89.0	0.2	
36		33.8	-64.2	

For mechanistic details see Figure 4. [a]In kcal/mol; geometry optimization from Gaussian 03 using M062X level of theory, energy from Gaussian 03 using CCSD(T) level of theory. Basis set is 6-311++G in both calculations. [b]In kcal/mol; thermochemistry from Gaussian 03 using CBS-QB3 level of theory.

show that the chemistry for these molecules is similar to the first three structures. For example, Reaction (7) is similar to the first step of the Korcek mechanism (Reaction (1)). However, replacing one CH₂ group with an NH group significantly enriches the chemistry. For example, Reaction (9) involves two H-transfers (transition state shown in Figure 2) and Reactions (11) and (12) form three products from one reactant.

Table 2 shows that in the case of N-(hydroperoxymethyl) formamide, half of the identified saddle points do not correspond to the intended product(s) generated by the BE matrix representation. Due to an initial open-chain structure, all these cyclization reactions require significant rearrangement of several atoms, and apparently, the FSM has trouble with these cases. Similarly, the algorithm failed to find several cyclization transition states of the γ-ketohydroperoxide. At the same time, all saddle points detected for more constrained cyclic reactants 1,2,4-dioxazolidine and 1,2,4-dioxazolidin-3-ol are for the intended products (Tables 3 and 4).

As a final example, the chemical reactivity of 1,4-pentadiene is investigated (Table 5 and Figure 5). Due to many possible

Table 5. Automatically Identified Unimolecular Reactions of Penta-1,4-diene (C=CCC=C)*

	Product(s)	$\Delta E^{\#a}$	$\Delta H_r^0(298\ K)^b$	Comments
37		63.0	5.2	
38		95.0	5.6	
39		104.3	-17.4	Unexpected H-shift
40		86.6	0.0	Framework rearrangement
41		66.2	11.3	
42		68.4	-7.1	Intended product: and
43		101.8	29.5	Intended product:
44		61.6	-17.4	1,2 H-shift

For mechanistic details see Figure 5. [a]In kcal/mol; geometry optimization from Gaussian 03 using M062X level of theory, energy from Gaussian 03 using CCSD(T) level of theory. Basis set is 6-311++G in both calculations. [b]In kcal/mol; thermochemistry from Gaussian 03 using CBS-QB3 level of theory.

single and double hydrogen atom transfers, many product channels are possible in this system. Table 1 shows that total number of channels exceeds seven hundred. However, out of this number, only eight saddle points were located, and three of these correspond to known reaction types in the RMG database. All structures were connected to the initial molecule by a large barrier (higher than 60 kcal/mol). Nevertheless, the ability of our algorithm not only to reproduce the existing "known" reaction pathways but also to identify five new reaction pathways for this simple molecule is very promising.

■ DISCUSSION AND CONCLUSIONS

The algorithm proposed here for finding elementary reaction paths possesses several important features: (1) It requires no human intervention. (2) It does not require any information on the chemical reactivity of the reactant(s). In order to detect saddle points, it generates a series of product(s) and tries to find pathways that connect the product(s) to the initial structure; this would be tedious to do without automation. It is much more computationally efficient (i.e., requires fewer energy/gradient calculations) than methods based on *ab initio* molecular dynamics since it rapidly finds and focuses on the structures near the saddle point connecting the reactant with designated products. Though the present calculations demonstrated that the procedure requires further optimization, comparison of both computational expenses and success rate (number of detected transition states) for the present examples with other automated methods for reaction pathway searches, especially those based on similar bond-forming rules, would be desirable in the future.

The present single-molecule examples show that our automated algorithm is able to detect not only previously detected reaction pathways but also several new, previously unknown, types of reactions. Our constraint that allows only two or three bonds to be formed and broken in the same

elementary step worked well for the present examples, keeping the number of possible product channels manageable (<1000 possible reaction paths per reactant) but still allowing us to discover a lot of new chemistry. The present results also show that detecting thermodynamically feasible products does not guarantee kinetic feasibility. The degree of success in locating unexpected kinetically favorable elementary steps is rather low—a significant number of thermodynamically stable structures either are connected to the initial structure only through multiple elementary steps or via chemically infeasible routes. As indicated above, clearly, there is some need to improve the procedure and reduce the search space especially when three chemical bonds are involved or initial structure is cyclic. It is also important to improve, accordingly, the underlying algorithms in the FSM calculations so as to avoid unnecessary gradient calls since the reaction path-finding step is the computationally limiting step.

In the future, some of these routes which are the most obviously unlikely, such as breaking the C=O bond, might be eliminated in order to focus resources on discovering transition states for more likely reactions. Another possibility is to freeze certain bonds and angles that are expected to be unreactive in order to reduce the number of conformers and matrices dimensionality. However, such limitations based on "chemical intuition" should be imposed with caution as they may prohibit detecting unexpected important reaction pathways! The present calculations with minimal constraints on the chemical reactivity demonstrate that automated identification of transition states is indeed a tricky problem. Table 1 shows no connection between number of located transition states and number of total gradient calls (though the latter is obviously related to the complexity of the initial structure). The lowest number of total gradient calls was observed for the initial reactants for which we found the highest number of transition states (OOCNC=O, 8 "unexpected" transition states; OC1OOCN1, 11 "unexpected" transition states), while the lowest number of transition states was determined for the initial reactant with the highest number of total gradient calls (O1COOC1, 3 "unexpected" transition states).

The success of the exact search for the saddle point heavily relies on the ability to generate very accurate corresponding guess structures from a double-ended method, such as the freezing string method (FSM) employed in the present study. The FSM can be considered as one of the fastest double-ended methods. However, the FSM is not guaranteed to converge to the reaction path of interest and therefore convergence to the correct saddle point cannot be rigorously guaranteed.[13] For each detected transition state, we performed intrinsic reaction coordinate (IRC) calculations to recover both reactant and product sides. In several cases, we found that the Berny optimization of the highest node along the reaction path generated by the FSM leads to transition states, which do not connect the specified reactant and product states. Therefore, a good initial guess of the transition state remains one of the major challenges in the use of such algorithms for complex systems.[43] We also note that the Berny optimization procedure and IRC calculations may sometimes fail, although we have not observed this in the present work, suggesting that the FSM (sometimes) generated good guess starting points for TS optimization and that the Berny and IRC methods implemented in Gaussian are quite robust.

As an alternative to the FSM, its predecessor, the growing string method (GSM), an iterative algorithm that connects the reactant and the product via the IRC, can be implemented. As indicated above, recent advances in the GSM have allowed it to locate the exact saddle point for a given reaction path.[43,44] A procedure for automatic reaction path finding which is based on the GSM has been recently proposed.[40] However, while formally very attractive, the GSM, in its original formulation, is more computationally expensive compared to the cost of the FSM calculation, and in this work, the the FSM calculations are the limiting computational step in the algorithm. Moreover, even the GSM can fail to locate a saddle point for a single elementary step as no double-ended string method is perfectly reliable. An important point of the present testing is that, in practical applications, the FSM is able to discover several important transition states recently discovered by manual searches.[17,65] However, future improvement of reliability as well as reducing computational costs of double-ended methods is obviously required.

In the future, we plan to extend the present algorithm to bimolecular chemical reactions and to couple the automated approach for finding the "unexpected" reactions with existing automated mechanism-generation software packages for "expected" reactions, such as the Reaction Mechanism Generator (RMG).[15] The combination would efficiently generate much more comprehensive and reliable chemical kinetic models than is possible with existing model-construction techniques.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00407.

Saddle point structures for the reactions (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: ysuleyma@mit.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) Schrodi, Y.; Pederson, R. L. Evolution and Applications of Second-Generation Ruthenium Olefin Metathesis Catalysts. *Aldrichimica Acta* **2007**, *40*, 45−52.

(2) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem., Int. Ed.* **2001**, *40*, 2004−2021.

(3) Magoon, G. R.; Green, W. H. Design and Implementation of a Next-generation Software Interface for On-the-fly Quantum and Force Field Calculations in Automated Reaction Mechanism Generation. *Comput. Chem. Eng.* **2013**, *52*, 35−45.

(4) Wales, D. J. Energy Landscapes: Calculating Pathways and Rates. *Int. Rev. Phys. Chem.* **2006**, *25*, 237−282.

(5) Wales, D. J. *Energy Landscapes*; Cambridge University Press: Cambridge, 2003.

(6) Schlegel, H. B. Optimization of Equilibrium Geometries and Transition Structures. *J. Comput. Chem.* **1982**, *3*, 214−218.

(7) 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.

(8) Schlegel, H. B. Estimating the Hessian for Gradient-type Geometry Optimizations. *Theor. Chem. Acc.* **1984**, *66*, 333−340.

(9) Henkelman, G.; Uberuaga, B. P.; Jonsson, H. A Climbing Image Nudged Elastic Band Method for Finding Saddle Points and Minimum Energy Paths. *J. Chem. Phys.* **2000**, *113*, 9901−9904.

(10) Henkelman, G.; Jonsson, H. Improved Tangent Estimate in the Nudged Elastic Band Method for Finding Minimum Energy Paths and Saddle Points. *J. Chem. Phys.* **2000**, *113*, 9978−9985.

(11) E, W. N.; Ren, W. Q.; Vanden-Eijnden, E. String Method for the Study of Rare Events. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2002**, *66*, 052301.

(12) Sheppard, D.; Terrell, R.; Henkelman, G. Optimization Methods for Finding Minimum Energy Paths. *J. Chem. Phys.* **2008**, *128*, 134106.

(13) 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.

(14) Zimmerman, P. M. Growing String Method with Interpolation and Optimization in Internal Coordinates: Method and Examples. *J. Chem. Phys.* **2013**, *138*, 184102.

(15) Green, W. H.; Allen, J. W.; Buesser, B. A.; Ashcraft, R. W.; Beran, G. J.; Class, C. A.; Gao, C.; Goldsmith, C. F.; Harper, M. R.; Jalan, A.; Keceli, M.; Magoon, G. R.; Matheu, D. M.; Merchant, S. S.; Mo, J. D.; Petway, S.; Raman, S.; Sharma, S.; Song, J.; Suleimanov, Y. V.; Van Geem, K. M.; Vandeputte, A. G.; Wen, J.; West, R. H.; Wong, A.; Wong, H.-W.; Yelvington, P. E.; Yee, N.; Yu, J. RMG − Reaction Mechanism Generator, 2014. https://github.com/GreenGroup/RMG-Py (accessed July 2015).

(16) Magoon, G.; Aguilera-Iparraguirre, J.; Green, W. H.; Lutz, J. J.; Piecuch, P.; Wong, H.-W.; Oluwole, O. O. Detailed Chemical Kinetic Modeling of JP-10 (exo-tetrahydrodicyclopentadiene) High Temperature Oxidation: Exploring the Role of Biradical Species in Initial Decomposition Steps. *Int. J. Chem. Kinet.* **2012**, *44*, 179−193.

(17) Jalan, A.; Alecu, I. M.; Meana-Pañeda, R.; Aguilera-Iparraguirre, J.; Yang, K. R.; Merchant, S. S.; Truhlar, D. G.; Green, W. H. New Pathways for Formation of Acids and Carbonyl Products in Low-temperature Oxidation: The Korcek Decomposition of γ- ketohydroperoxides. *J. Am. Chem. Soc.* **2013**, *135*, 11100−11114.

(18) Vandewiele, N. M.; Magoon, G. R.; Van Geem, K. M.; Reyniers, M.-F.; Green, W. H.; Marin, G. B. Kinetic Modeling of Jet Propellant-10 Pyrolysis. *Energy Fuels* **2015**, *29*, 413−427.

(19) Irikura, K.; Johnson, R. Predicting Unexpected Chemical Reactions by Isopotential Searching. *J. Phys. Chem. A* **2000**, *104*, 2191−2194.

(20) Van Duin, A. C. T.; Zeiri, Y.; Dubnikova, F.; Kosloff, R.; Goddard, W. A., III. Atomistic-Scale Simulations of the Initial Chemical Events in the Thermal Initiation of Triacetonetriperoxide. *J. Am. Chem. Soc.* **2005**, *127*, 11053−11062.

(21) Döntgen, M.; Przybylski-Freund, M.-D.; Kröger, L. C.; Kopp, W. A.; Ismail, A. E.; Leonhard, K. Automated Discovery of Reaction Pathways, Rate Constants, and Transition States Using Reactive Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2015**, *11*, 2517−2524.

(22) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A., III. ReaxFF: A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A* **2001**, *105*, 9396−9409.

(23) Yue, D.-C.; Ma, T.-B.; Hu, Y.-Z.; Yeon, J.; van Duin, A. C. T.; Wang, H.; Luo, J. Tribochemistry of Phosphoric Acid Sheared

between Quartz Surfaces: A Reactive Molecular Dynamics Study. *J. Phys. Chem. C* **2013**, *117*, 25604−25614.

(24) Abolfath, R. M.; van Duin, A. C. T.; Brabec, T. Reactive Molecular Dynamics Study on the First Steps of DNA Damage by Free Hydroxyl Radicals. *J. Phys. Chem. A* **2011**, *115*, 11045−11049.

(25) Chenoweth, K.; van Duin, A. C. T.; Dasgupta, S.; Goddard, W. A., III. Initiation Mechanisms and Kinetics of Pyrolysis and Combustion of JP-10 Hydrocarbon Jet Fuel. *J. Phys. Chem. A* **2009**, *113*, 1740−1746.

(26) Lümmen, N. ReaxFF-molecular Dynamics Simulations of Non-oxidative and Non-catalyzed Thermal Decomposition of Methane at High Temperatures. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7883−7893.

(27) Liu, L.; Bai, C.; Sun, H.; Goddard, W. A., III. Mechanism and Kinetics for the Initial Steps of Pyrolysis and Combustion of 1,6-Dicyclopropane-2,4-hexyne from ReaxFF Reactive Dynamics. *J. Phys. Chem. A* **2011**, *115*, 4941−4950.

(28) Ding, J.; Zhang, L.; Zhang, Y.; Han, K.-L. A Reactive Molecular Dynamics Study of n-Heptane Pyrolysis at High Temperature. *J. Phys. Chem. A* **2013**, *117*, 3266−3278.

(29) Qi, T.; Bauschlicher, C. W., Jr.; Lawson, J. W.; Desai, T. G.; Reed, E. J. Comparison of ReaxFF, DFTB, and DFT for Phenolic Pyrolysis. 1. Molecular Dynamics Simulations. *J. Phys. Chem. A* **2013**, *117*, 11115−11125.

(30) Cheng, X.-M.; Wang, Q.-D.; Li, J.-Q.; Wang, J.-B.; Li, X.-Y. ReaxFF Molecular Dynamics Simulations of Oxidation of Toluene at High Temperatures. *J. Phys. Chem. A* **2012**, *116*, 9811−9818.

(31) He, Z.; Li, X.-B.; Liu, L.-M.; Zhu, W. The intrinsic mechanism of methane oxidation under explosion condition: A combined ReaxFF and DFT study. *Fuel* **2014**, *124*, 85−90.

(32) Zheng, S.; Pfaendtner, J. Car−Parrinello Molecular Dynamics + Metadynamics Study of High- Temperature Methanol Oxidation Reactions Using Generic Collective Variables. *J. Phys. Chem. C* **2014**, *118*, 10764−10770.

(33) Polino, D.; Parrinello, M. *Combustion Chemistry via Metadynamics: Benzyl Decomposition Revisited*; Talk presented at the 9th International Conference on Chemical Kinetics, Ghent, Belgium, July 1, 2015.

(34) Luo, Y.; Maeda, S.; Ohno, K. Automated Exploration of Stable Isomers of H+(H₂O)n (n = 5−7) via Ab Initio Calculations: An Application of the Anharmonic Downward Distortion Following Algorithm. *J. Comput. Chem.* **2009**, *30*, 952−961.

(35) Maeda, S.; Ohno, K.; Morokuma, K. Automated Global Mapping of Minimum Energy Points on Seams of Crossing by the Anharmonic Downward Distortion Following Method: A Case Study on H₂CO. *J. Phys. Chem. A* **2009**, *113*, 1704−1710.

(36) Maeda, S.; Ohno, K.; Morokuma, K. Systematic Exploration of the Mechanism of Chemical Reactions: the Global Reaction Route Mapping (GRRM) Strategy by the ADDF and AFIR Methods. *Phys. Chem. Chem. Phys.* **2013**, *15*, 3683−3701.

(37) Shang, C.; Liu, Z.-P. Stochastic Surface Walking Method for Structure Prediction and Pathway Searching. *J. Chem. Theory Comput.* **2013**, *9*, 1838−1845.

(38) Zhang, X.-J.; Liu, Z.-P. Reaction Sampling and Reactivity Prediction Using the Stochastic Surface Walking Method. *Phys. Chem. Chem. Phys.* **2015**, *17*, 2757−2769.

(39) Hatanaka, M.; Maeda, S.; Morokuma, K. Sampling of Transition States for Predicting Diastereoselectivity Using Automated Search Method -Aqueous Lanthanide-Catalyzed Mukaiyama Aldol Reaction. *J. Chem. Theory Comput.* **2013**, *9*, 2882−2886.

(40) Zimmerman, P. M. Automated Discovery of Chemically Reasonable Elementary Reaction Steps. *J. Comput. Chem.* **2013**, *34*, 1385−1392.

(41) Zimmerman, P. M. Navigating Molecular Space for Reaction Mechanisms: An Efficient, Automated Procedure. *Mol. Simul.* **2015**, *41*, 43−54.

(42) 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.

(43) Zimmerman, P. M. Reliable Transition State Searches Integrated with the Growing String Method. *J. Chem. Theory Comput.* **2013**, *9*, 3043−3050.

(44) Zimmerman, P. M. Single-ended Transition State Finding with the Growing String Method. *J. Comput. Chem.* **2015**, *36*, 601−611.

(45) Benson, S. W. *Thermochemical Kinetics: Methods for the Estimation of Thermochemical Data and Rate Parameters*, 2nd ed.; Wiley: New York, 1976.

(46) Wang, L.-P.; Titov, A.; McGibbon, R.; Liu, F.; Pande, V. S.; Martinez, T. J. Discovering Chemistry with an Ab Initio Nanoreactor. *Nat. Chem.* **2014**, *6*, 1044−1048.

(47) Tarjan, R. E. Graph Algorithms in Chemical Computation. In *Algorithms for Chemical Computations*; Christoffersen, R. E., Ed.; ACS Symposium Series 46; American Chemical Society: Washington, DC, 1977.

(48) Ugi, I.; Bauer, J.; Brandt, J.; Friedrich, J.; Gasteiger, J.; Jochum, C.; Schubert, W. New Applications of Computers in Chemistry. *Angew. Chem., Int. Ed. Engl.* **1979**, *18*, 111−123.

(49) Yoneda, Y. A Computer Program Package for the Analysis, Creation, and Estimation of Generalized Reactions − GRACE I. Generation of Elementary Reaction Network in Radical Reactions − A/GRACE (I). *Bull. Chem. Soc. Jpn.* **1979**, *52*, 8−14.

(50) Blurock, E. S. Reaction: System for Modeling Chemical Reactions. *J. Chem. Inf. Model.* **1995**, *35*, 607−616.

(51) Quann, R. J.; Jaffe, S. B. Structure-Oriented Lumping: Describing the Chemistry of Complex Hydrocarbon Mixtures. *Ind. Eng. Chem. Res.* **1992**, *31*, 2483−2497.

(52) Broadbelt, L. J.; Stark, S. M.; Klein, M. T. Computer Generated Reaction Modelling Decomposition and Encoding Algorithms for Determining Species Uniqueness. *Comput. Chem. Eng.* **1996**, *20*, 113−129.

(53) Warth, V.; Battin-Leclerc, F.; Fournet, R.; Glaude, P. A.; Come, G. M.; Scacchi, G. Computer Based Generation of Reaction Mechanisms for Gas-Phase Oxidation. *Comput. Chem.* **2000**, *24*, 541−560.

(54) Rangarajan, S.; Bhan, A.; Daoutidis, P. Rule-Based generation of Thermochemical Routes to Biomass Conversion. *Ind. Eng. Chem. Res.* **2010**, *49*, 10459−10470.

(55) Sumathi, R.; Carstensen, H.-H.; Green, W. H. Reaction Rate Prediction via Group Additivity, Part 2: H-Abstraction from Alkenes, Alkynes, Alcohols, Aldehydes, and Acids by H Atoms. *J. Phys. Chem. A* **2001**, *105*, 8969−8984.

(56) Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.

(57) Halgren, T. A. Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **1996**, *17*, 520−552.

(58) Halgren, T. A. Merck molecular force field. III. Molecular Geometries and Vibrational Frequencies for MMFF94. *J. Comput. Chem.* **1996**, *17*, 553−586.

(59) Halgren, T. A.; Nachbar, R. B. Merck Molecular Force Field. IV. Conformational Energies and Geometries for MMFF94. *J. Comput. Chem.* **1996**, *17*, 587−615.

(60) Halgren, T. A. Merck molecular force field. V. Extension of MMFF94 Using Experimental Data, Additional Computational Data, and Empirical Rules. *J. Comput. Chem.* **1996**, *17*, 616−641.

(61) Marvin was used for drawing, displaying, and characterizing chemical structures, substructures, and reactions, Marvin 6.2.0, 2014, ChemAxon. http://www.chemaxon.com (accessed July 2015).

(62) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T.B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kuś, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; Woodcock, H. L., III; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C.-M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.;

Diedenhofen, M.; DiStasio, R. A., Jr.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W. D.; Harbach, P. H. P.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T.-C.; Ji, H.; Kaduk, B.; Khistyaev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S.-P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Neuscamman, E.; Oana, C. M.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y.-C.; Thom, A. J. W.; Tsuchimochi, T.; Vanovschi, V.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Yang, J.; Yeganeh, S.; Yost, S. R.; You, Z.-Q.; Zhang, I. Y.; Zhang, X.; Zhao, Y.; Brooks, B. R.; Chan, G. K. L.; Chipman, D. M.; Cramer, C. J.; Goddard, W. A., III; Gordon, M. S.; Hehre, W. J.; Klamt, A.; Schaefer, H. F., III; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xu, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J.-D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C.-P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. A.; Slipchenko, L. V.; Subotnik, J. E.; Van Voorhis, T.; Herbert, J. M.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. Advances in Molecular Quantum Chemistry Contained in the Q-Chem 4 Program Package. *Mol. Phys.* **2015**, *113*, 184−215.

(63) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A.; *Gaussian 03, Revision C.02*, Gaussian, Inc., Wallingford CT, 2004.

(64) Jensen, R. K.; Korcek, S.; Mahoney, L. R.; Zinbo, M. Liquid-Phase autoxidation of organic compounds at elevated temperatures. 1. The stirred flow reactor technique and analysis of primary products from n-hexadecane autoxidation at 120−180° C. *J. Am. Chem. Soc.* **1979**, *101*, 7574−7584.

(65) Jalan, A.; Allen, J. W.; Green, W. H. Chemically Activated Formation of Organic Acids in Reactions of the Criegee Intermediate with Aldehydes and Ketones. *Phys. Chem. Chem. Phys.* **2013**, *15*, 16841−16852.