Advanced Review

Methods for exploring reaction space in molecular systems

Amanda L. Dewyer,[†] Alonso J. Argüelles[†] and Paul M. Zimmerman *

The area of reaction mechanism discovery simulation has taken considerable strides in recent years. Novel methods that make hypotheses for elementary steps and complementary means for reaction path and transition state (TS) optimization are lowering the amount of chemical intuition and user effort required to explore reaction networks. The resulting networks lead from reactants to reactive intermediates and products, and are becoming closer representations of physical mechanisms involved in experiments. This review describes several of these approaches, which are categorized based on their overarching TS finding strategies. Future advances are discussed that may revolutionize the ability of simulation to fully predict not just the reaction mechanism but reaction outcomes. © 2017 Wiley Periodicals, Inc.

How to cite this article:
WIREs Comput Mol Sci 2018, 8:e1354. doi: 10.1002/wcms.1354

HARVESTING REACTION PATHWAYS USING COMPUTATION

Deep mechanistic insight into a chemical reaction can be gained from atomistic studies of elementary steps, where one intermediate is connected to another by a path containing a single transition state (TS). The three-dimensional (3D) structural information provided by reaction paths not only provides a basic framework for conceptualizing the reaction, but also estimates the rates, thermodynamics, and selectivity that quantify key reaction parameters. In a highly successful model of a reaction, all key elementary steps composing the mechanism are known and accounted for. Building such a model is by no means trivial, even for relatively well-studied reactions where many of the elementary steps are already known. For emerging reactions with little mechanistic precedent, the task requires large amounts of effort—both computational and human time—and no current strategy provides any guarantee of success.

*Correspondence to: paulzim@umich.edu

Department of Chemistry, University of Michigan, Ann Arbor, MI, USA

[†]These authors contributed equally.

Conflict of interest: The authors have declared no conflicts of interest for this article.

This article outlines recent contributions to the area of computational mechanism discovery, where sequences and networks of individual elementary steps are created and analyzed. The powerful methods of this area are designed to locate intermediates, reaction paths, and TSs with as little guidance from the researcher as possible. Ultimately, it is hoped that further advances will allow full reaction mechanisms to be revealed purely from computation, even for challenging chemical reactions with large numbers of intermediates, TSs, and products densely populating the reaction landscape. Although reaching this goal by gathering up all vital reaction pathways is not yet within sight, major advances in recent years have started moving toward it.

First principles, or *ab initio*, simulations of reactant molecules and catalysts provide an important, central viewpoint on how computation can approach reactivity with little or no guidance from experiment.[1–3] *Ab initio* simulations provide potential energy surfaces (PESs) and, from these, free energy surfaces,[4,5] which approximate the true reactive landscapes. All potential elementary steps may be located by exploring these landscapes, at least in principle. The high dimensionality (approximately 3 N, where N is the number of atoms in the system) of the surfaces, however, means that exhaustive exploration is usually impossible. Some form of search strategy, often coming from a low-

dimensional re-envisioning of reaction paths, must be employed to locate intermediates and the TSs that connect them. This situation leads us to the two key challenges discussed herein: (1) *Locating reaction paths for single, posited elementary steps, and their associated TSs* and (2) *Identifying hypothetical reaction pathways quickly, automatically, and with good coverage of the chemically relevant elementary steps.*

Locating a pathway for an envisioned reaction often employs local TS finders to optimize saddle points along the path.[6–8] The TS structures can be guessed using intuition or approximated using various interpolation tools before optimization.[9–23] Once a TS is found, intrinsic reaction coordinate (IRC)[24,25] computations provide the reaction path connecting the TS to its neighboring intermediates. Largely, this method has relied on chemical intuition to designate the TSs that are of interest, as well as the intermediates that have to be found.[26,27] Therefore, this strategy has been shown to be largely useful for accomplishing *Goal 1*, but at the same time is not conducive for solving *Goal 2*.

To accomplish *Goal 2*, excellent methods for achieving *Goal 1* must be available nonetheless. Conceptually, elementary step exploration on a PES can occur via three general approaches (Figure 1). TS finding followed by IRC, double-ended path optimization starting from two endpoints, and single-ended searching by moving along a reaction coordinate. While all three method classes lead to the same information about one elementary step, strategies for exploring many such steps on a PES must carefully account for the strengths and weaknesses of each approach. A strong strategy for PES exploration must synergistically combine its reaction path finder with its reaction hypothesis generator to achieve maximal success.

Reaction mechanism discovery is an immense challenge, and the methods discussed in this review will show a variety of tools that have approached the problem. A number of interesting areas must be left out of this review to maintain focus: standard optimization methods[28,29] computer aided synthesis,[30] quantum and molecular mechanical (QM/MM) methods,[31–35] molecular dynamics (MD) simulations that do not look for TSs,[36–39] haptic quantum chemistry,[40,41] and machine learning tools[42–44] will not be described herein. Our focus instead is on PES exploration tools for molecular systems, and closely related atomistic methods that approximate the mechanistic information that would have come from *ab initio* simulations. Furthermore, the accuracy of kinetics emerging for all of these methods depends ultimately on the capabilities of the underlying electronic structure methodologies,[2,45–48] but this area is also outside the scope of this article.

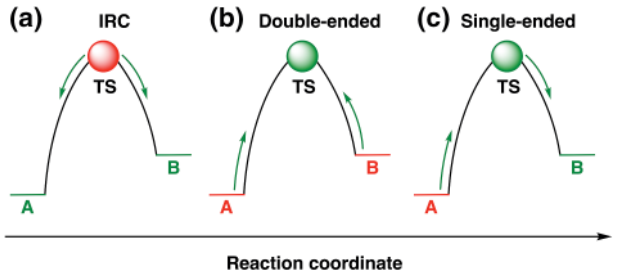

FIGURE 1 | Single elementary step characterization algorithms. (a) IRC, (b) double-ended, (c) single-ended.

FOUR CONCEPTS FOR DISCOVERING ELEMENTARY REACTION NETWORKS

The four categories shown in Figure 2 will be used to organize our discussion on reaction exploration. These methods start from a designated set of reactants and catalysts, but otherwise are supposed to operate with as little input from a researcher as possible. The categories consist of: *Concept 1*. Encoded elementary step types from databases or chemical heuristics are used to describe reaction pathways, with the steps taken depending on the input reactants. In this category, activation energies are frequently approximated, but sometimes computed based on TS optimization. *Concept 2*. Approximate TSs are generated, followed by local TS optimization and IRC computations. For example, a well-known tool in this area pushes two molecules together with artificial forces to generate the approximate TS. *Concept 3*. Putative elementary steps are generated and corresponding intermediates formed, then double-ended methods are used to refine a reaction path and TS. *Concept 4*. Hypothetical reaction coordinates are generated, and single-ended methods perform reaction path searches along these coordinates. All of these methods are closely related to one another in spirit, but differences in the details can be paramount to their degree of success. All generate approximate reaction paths, estimate energetic barriers for reaction, and can be used to sequentially assemble networks of elementary steps. These methods, however, can break down at one point or another, and none achieve a perfectly ideal balance of high accuracy, comprehensive PES searching, and low computational cost. Despite the apparent promise of user-interaction-free exploration, every method

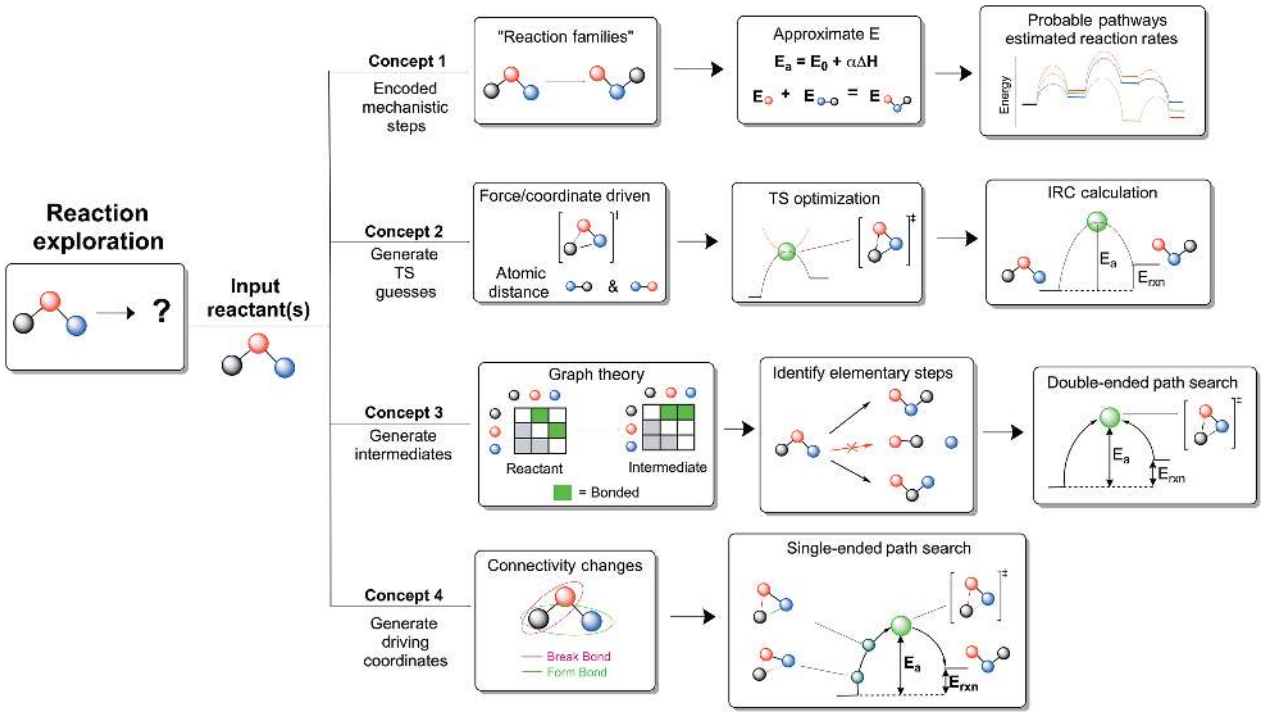

FIGURE 2 | Four categories of automated reaction path exploration methods.

still requires at least some input from chemical intuition.

Concept 1: Reaction Discovery Using Encoded Transformations

One of the earliest attempts at computerized reaction mechanism generation was pioneered by Broadbelt and coworkers in the 1990s, in a method called NetGen.[49] The strategy used graph theory to represent reaction networks, reducing the 3D problem of structural rearrangements of chemistry into a simplified, two-dimensional (2D) view of reactions as modification of the edges (bonds) connecting vertices (atoms) of a graph (see Figure 3). This concept can be implemented in matrix form using adjacency matrices or bond and electron (BE) matrices, allowing chemical transformations to easily be described by matrix operations.[50,51] These transformations correspond to postulated reaction steps, or specifically encoded rules that dictate what may or may not occur at a given substrate.

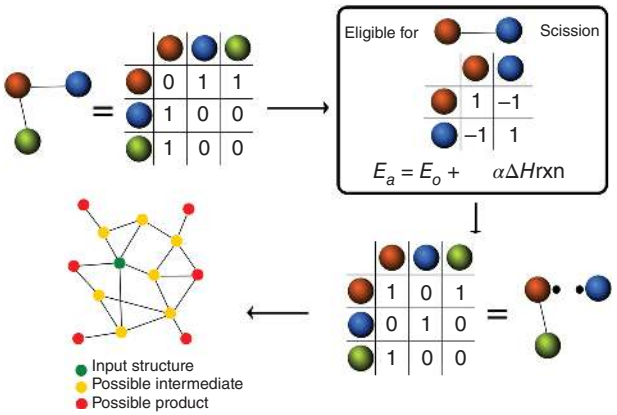

FIGURE 3 | Graph-based reaction network generation using specifically encoded transformations.

Broadbelt's original work estimated kinetic parameters using the concept of reactive families, where each family obeyed uniform principles of reactivity. These rates were calculated using the Evans–Polanyi principle[52] (Eq. (1)):

$$E_a = E_0 + \alpha\Delta H_{rxn} \quad (1)$$

Where E_0 is a reference activation energy based on a known reaction, α is a constant particular to the reaction family, and ΔH_{rxn} is the enthalpy of reaction, which is obtained using Benson group additivity.[53,54] In the Benson approach, a species is divided into predefined groups, and overall thermodynamic properties such as enthalpy of formation (ΔH_f), and therefore the activation energy, can be estimated as a sum of the individual values of its groups (e.g., ΔH_f for propane is a sum of enthalpies for two methyl and one methylene group). In cases where group

additivity was unable to give accurate results, *ab initio* computations provided the necessary thermodynamic information.

The Netgen algorithm begins by applying transformations to the reactant molecule graph, and the newly made molecules are added to a list of species involved in the global mechanism. Iterating on this procedure can lead to a large number of possible reactions, products, and intermediates.[55] Because only a fraction of these species are mechanistically relevant, the reaction network can be trimmed using a rate-based termination algorithm.[56] Early studies examined pyrolysis of tetradecane[57] and even the Fischer–Tropsch reaction catalyzed by Ni and Co surfaces.[58] This latter study demonstrated the flexibility of graphical approaches, but also the challenges for applications to new chemistries: new thermodynamic and kinetic parameters were required, as the gas-phase models were inapplicable to species on the surface.

The Green group subsequently created the Reaction Mechanism Generator (RMG) software to provide insights into complex mechanisms, where manual searches would prove extraordinarily tedious.[59,60] Although their overall strategy is represented by Figure 3, significant steps were taken beyond the original approach by Broadbelt. The graph representation was expanded to include new atom types (e.g., a carbonyl carbon when a $R_2C{=}O$ was present), more efficient graph isomerism algorithms to identify reaction families, and improved approximations for group additivity and rate predictions. To achieve the latter, a hierarchical organization identifies general functional groups, then refines these based on more specific chemical moieties. Importantly, specific kinetic parameters were estimated using quantum chemical optimization of TSs, especially for cases where good rate estimates were otherwise unavailable.[61]

RMG has been used to extensively map pyrolysis reaction mechanisms in an automated way. The mechanism of *n*-butanol pyrolysis, where previous reports found 1446 reactions, was found to be composed of 263 intermediates and 3381 elementary steps.[62] In a related scenario, RMG created a global reaction map for the low-temperature, pressure-dependent pyrolysis of methane, with roughly 100 species and 1000 reversible reactions.[59] In this study, they were able to account for an unexplained phenomenon of autocatalysis at low methane conversion via nonintuitive reactions. The Green and West groups have shown that RMG can be applied in a variety of complex systems with good success,[63–67] and is adaptable enough to study even silicon hydride chemistry.[64] While RMG is clearly a powerful concept, it is ultimately limited by the availability of good reaction data available to its libraries, as RMG itself provides no automated approach for TS finding (but see *Concepts 2 and 3* for recent advances).

In a strategy closely related to RMG, reaction mechanism construction can be achieved by using principles inspired by organic chemistry, rather than specifically encoded elementary step types. In work by Aspuru-Guzik and coworkers,[68] a set of heuristic transformation rules were developed to represent the electron flow of polar reactions (i.e., their work is inspired by the 'arrow pushing' concept). For example in Figure 4, cleavage of a single H—Cl bond generates a reactive charged species. The H^+ and Cl^- sequentially are added to an alkene via ionic intermediates until a neutral product is formed. The kinetics of a reaction pathway are evaluated using a heuristic based on the energy of the highest intermediate (in accord with the Hammond postulate).

Aspuru-Guzik and coworkers heuristic mechanism generator has been used to study the formose reaction, a base-catalyzed reaction that converts formaldehyde into a complex mixture of sugars.[69,70] While it is encouraging that such a complex reaction could be explored and so many intermediates found (hundreds), the predictions included reactions considered infeasible by organic chemists. For instance, 3- and 4-membered rings appear as key intermediates, despite being highly strained structures.

The three *Concept 1* strategies outlined in Table 1 represent a core class of tools for generating reaction mechanisms using knowledge-based approaches. In these methods, TSs are typically implicit, and not computed at an *ab initio* level of theory. This paradigm is not precisely true of RMG, where select TSs are often computed to augment the

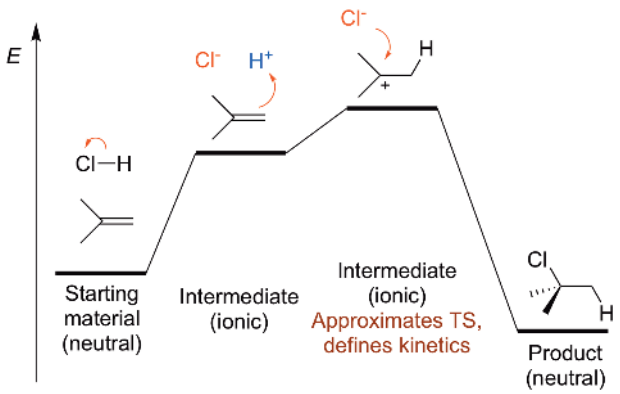

FIGURE 4 | Example of an alkene hydrohalogenation pathway by the Aspuru-Guzik method.

TABLE 1 | Comparison of Automatic, Knowledge-Based Mechanism Generators

Method	Representation of Reactants	Rules	Rates and Thermodynamics
Netgen[49]	Bond-electron matrices	General reaction family templates	Evans–Polanyi relationships, group additivity, quantum chemistry
Reaction Mechanism Generator (RMG)[59,60]	Atomically detailed adjacency lists	Hierarchically organized, more specific reaction family templates	Same as above, plus some transition states (TSs) from quantum chemistry
Heuristics-Aided Quantum Chemistry[68]	SMILES	Heuristic rules for electron flow in polar organic reactions ('arrow pushing')	Heuristic kinetic parameters based on Hammond postulate

kinetic parameter libraries (see *Concepts 2 and 3*). Therefore, activation barriers and rates must be estimated for most species, due to experimental data not being available for the hundreds or thousands of different elementary steps in a particular reaction network. At a basic physical level, different reaction mechanisms can differ by orders of magnitude in their relative rates, and one mechanism can dominate over others due to just a few kcal/mol differences in activation barriers. This well-known, but major challenge has motivated the development of *Concepts 2 to 4*, which all consider the 3D, atomistic description of TSs to be a key factor for uncovering reaction mechanisms.

Concept 2: Reaction Paths by Focus on TSs

Our focus now returns to the PES and methods that explore this surface to determine low-energy reaction mechanisms. In principle, a complete, high-quality PES from an appropriate quantum chemical method is sufficiently detailed to provide all information required for a complete mechanism delineation. In practice, the massive dimensionality and huge breadth of PESs means that only highly local, low-dimensional regions can be reasonably characterized. Up until recent years, the great majority of quantum chemical studies of reaction paths have restricted themselves to regions selected via chemical intuition. Typically, guess TS structures are formed and optimized in manual, stepwise procedures (e.g., the often-used eigenvector following[71–73] and related algorithms[74–87]) that provide atomistic detail, but also serve to verify the intuition from which they came. This strategy has been highly useful in analyzing rate-limiting steps and assessing steric and electronic preferences of reaction, despite the human burdens in which they entail.[88] This section will describe four methods for reaction path searches that center on the TS as the most important aspect of reaction network generation and provide streamlined, more systematic means for TS finding.

The first *Concept 2* strategy was developed by Maeda and coworkers to identify reaction paths using an important insight into PES curvature. Specifically, when traveling along a PES from one intermediate to another, the PES curves downward compared to the harmonic potential centered in the original minimum. This effect was exploited by Maeda and coworkers in a strategy called anharmonic downward distortion following (ADDF),[89–92] shown in Figure 5. By following the ADDs in contours of increasing energy (i.e., further displaced from the minima), direct paths to TSs are found. A major advantage of ADDF is that the paths generated closely match the IRC path, so ADDF paths are not only search paths, in the sense of optimization, but

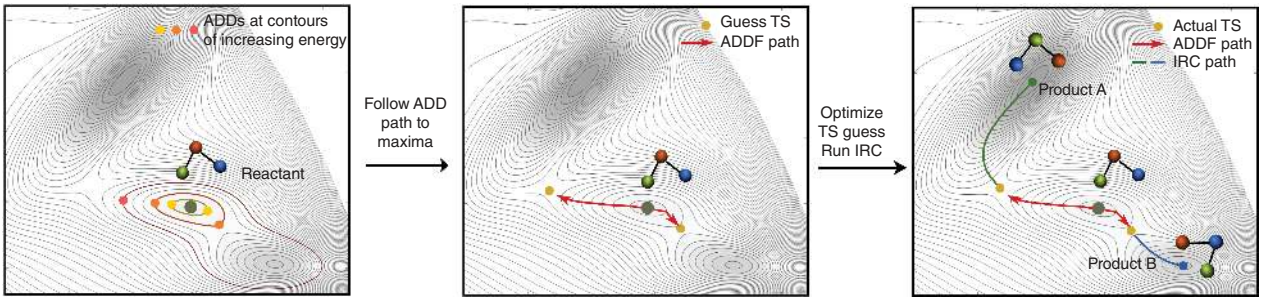

FIGURE 5 | The anharmonic downward distortion following (ADDF) method.

also are reaction pathways that represent the true reaction coordinate. Iterative invocation of ADDF leads to a reaction network, including essentially all relevant pathways. Although the current method scales poorly with system size,[90] it is especially useful for unimolecular reactions, and has been applied successfully to form reaction maps for formaldehyde, propyne, and formic acid reactions.

Limitations in cost and scope provided impetus for Maeda and coworkers to design the artificial force-induced reaction (AFIR) method,[93–96] shown in Figure 6, which is particularly useful for bimolecular reactions. AFIR overcomes intermolecular activation barriers by applying an artificial external force to push the reactants together. When optimizing under applied force, the unbiased maximum energy point (i.e., when forces are removed) along the path toward the product is close in geometry to the true TS and can easily be refined to the actual saddle point. Sampling over many random initial orientations of the two reactants allows a variety of bimolecular reactions to be found.

AFIR has been successfully used, for example, in Claisen rearrangements, Biginelli reactions,[97] and cobalt-catalyzed hydroformylations.[98] These examples include cases where variants of AFIR could approach intramolecular reactions by defining fragments within a molecule, showing that the method is not strictly limited to bimolecular reactions. To accomplish these studies without massive computational cost, human input was required in the form of selecting which pairs of molecules/catalyst react as well as intramolecular fragment selection (i.e., active atom selection). In other words, some input guiding the reactants toward particular types of reaction paths was required (though similar limitations exist for other methods in this review).

Another method that uses energetic bias to push reactions forward was developed by Martínez-Núñez and coworkers.[99–102] Their method, termed transition state search using chemical dynamics simulations (TSSCDS), diagrammed in Figure 7, relies on running high-energy dynamics to induce reactions to occur. To initiate the dynamics, vibrational modes are populated with multiple quanta of energy, providing an effective stimulus to surmount activation barriers. For large systems, vibrational mode selection will inevitably be incomplete and require manual guidance due to the large number of mode combinations that may be populated. These dynamics are run at a semiempirical level of theory to minimize costs, yet still qualitatively capture bond-breaking and bond-forming events. After performing the dynamics simulations, an algorithm identifies atomic connectivity changes throughout each trajectory. When a reaction is found, a TS guess is formed by examining points around the transition region. The TS guesses are optimized at the PM7 level and then at a higher level using density functional theory (DFT)[103] methods, followed by an IRC computation. The process is repeated until a satisfactory set of reaction paths is obtained.

The original TSSCDS procedure was tested in intramolecular reactions involving formaldehyde, formic acid, and vinyl cyanide.[100] In these cases, not only did TSSCDS locate TSs in agreement with other established methods in terms of the lowest energy TSs and reaction paths, but also several other unprecedented TSs were found. Advances in this method have allowed global reaction mapping and treatment of fragments within molecules to allow larger systems to be studied.[102] These advances have allowed examination of pyrolysis of propenal and cobalt catalysis for hydroformylation of ethylene.[101]

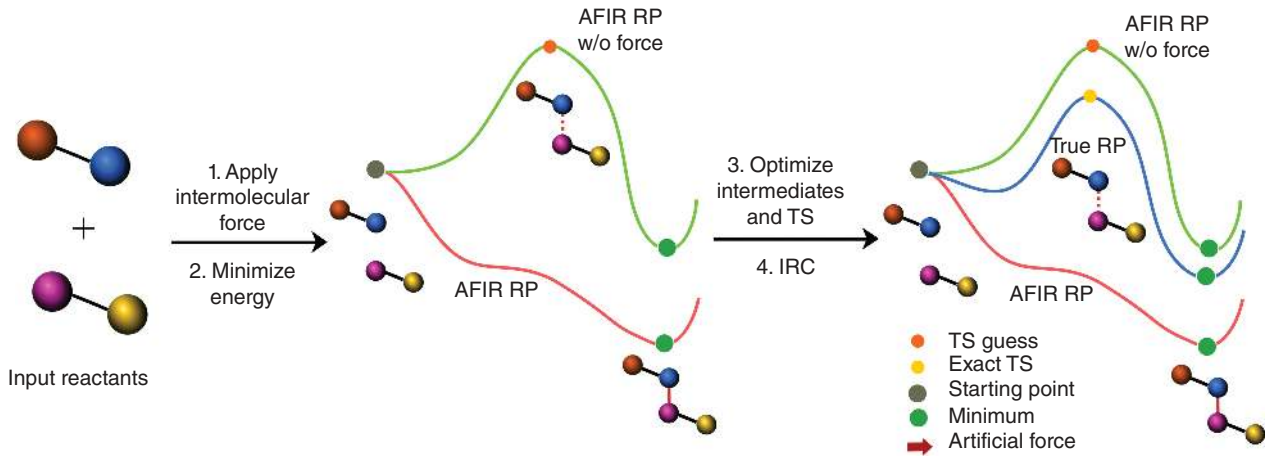

FIGURE 6 | The artificial force-induced reaction (AFIR) method.

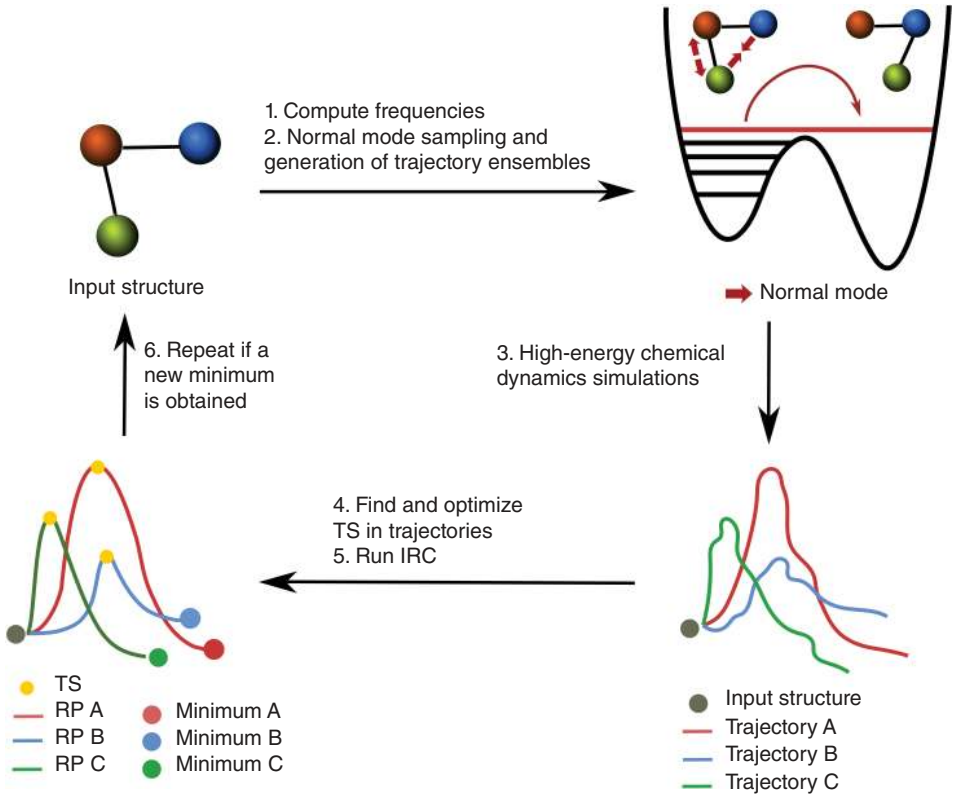

FIGURE 7 | The transition state search using chemical dynamics simulations (TSSCDS) method.

The *Concept 2* examples conclude with developments by West designed to obtain useful kinetic information for particular reaction types[104,105] (Figure 8). This procedure generates good TS structural guesses by using group contribution considerations to predict interatomic distances in the reactive site of the TS, saving a huge amount of labor when similar reactions are examined across many substrates. To form these guesses, the interatomic distances of the reactive atoms are modified based on a library containing typical reactions of the same class. The optimized TSs, as typical for *Concept 2* strategies, are subjected to IRC calculations to confirm the product identity.

West and coworkers applied this method to study H-abstraction reactions from a diisopropyl ketone (DIPK) combustion model. Based on a collection of 1393 H-abstraction reactions,[106] an initial training set of 44 unique reactions failed to provide good TS guesses, but once the training set was expanded to 827 reactions, 65% of the desired TSs could be found. West's approach is a good example of leveraging existing data to create additional useful data,[107] and we believe this concept should be used more frequently in the area of TS optimization and reaction mechanism exploration.

The four *Concept 2* tools are substantially more heterogeneous than those of the *Concept 1* section.

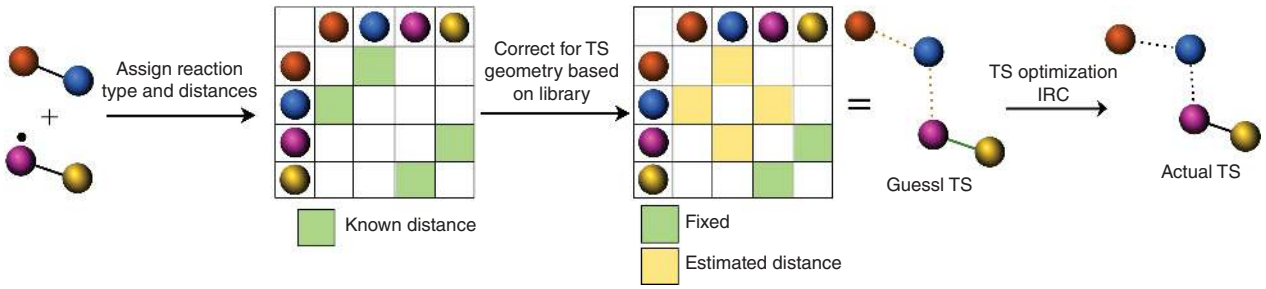

FIGURE 8 | Group additivity strategy for transition state geometry estimation and refinement.

Generally, all four provide means to uncover TSs, and therefore reaction paths. ADDF, AFIR, and TSSCDS focus on finding pathways that are not necessarily known *a priori*, while the group additivity TS-constructor focuses on particular types of elementary steps. These methods, however, are conceptually different from the *Concept 3* approaches, which seek to first locate plausible intermediates, and *then* go back and locate TSs and reaction paths.

Concept 3: Connecting Paths by Stringing Together Intermediates

Double-ended methods, such as the string method[108–110] and nudged elastic band,[20,23,111,112] are powerful tools for refining single-elementary-step reaction pathways where the two endpoints are known. When these are combined with algorithms that generate plausible intermediates starting from given reactant structures, reactions mechanisms can be found by connecting stable intermediates in the hypothetical network. In *Concept 3* techniques, reaction network generation can be successful when the intermediate generator and the double-ended method are each efficient, accurate, and reliable. This section details particular implementations of this two-piece strategy, with more emphasis given to the intermediate-generating tools.

One of the first methods developed in this area was the ZStruct method of the Zimmerman research group.[113,114] This approach utilizes connectivity graphs to determine potential intermediates that could form when connections are made or broken (Figure 9). Graphs with modified connections represent potential intermediates, for which 3D structures are created and then optimized using semiempirical methods. The low energy structures are optimized using DFT, and the stable, relatively low-energy intermediates are subjected to a double-ended reaction path search using the growing string method (GSM).[115–118] GSM has proven to be a powerful method in this regard,[119–123] as it generates a reaction pathway and exact TS structure using only two structures as input. Repeated invocation of ZStruct and GSM on newly generated intermediates provides reaction networks of increasing complexity. The original implementation of ZStruct worked best for intramolecular reactions due to the requirement that reactants be prealigned, which limited the usefulness of the method for systematic reaction exploration. Furthermore, ZStruct did not guarantee that pairs of intermediates were connected by a single elementary step, causing double-ended GSM to struggle in obtaining a single representative TS for a multistep pathway.

Despite its limitations, ZStruct has been used to investigate the mechanisms of several types of main group and transition metal reactions. For instance, this strategy uncovered an unexpected, off-cycle trap that was preventing a Ni-based C-H functionalization catalyst from turning over, which lead to the design of a new catalyst that avoided this trap.[119] A handful of (mostly catalytic) reactions have also been investigated using the original ZStruct, showing it to be a competent tool for reaction exploration.[119,124–126]

In a similar vein as ZStruct, Green and coworkers developed a graph-based approach to find plausible reaction pathways,[127] also described in Figure 9. Two major differences are of note here. First, the freezing string method,[128] which rapidly creates a TS guess structure, was used in tandem with the Berny TS finder[22,129] to refine the TS guesses to exact TSs. From the optimized TS, an IRC calculation identifies the intermediates connected by the TS, just like *Concept 2* algorithms. Second, Green's method sampled not only changes in 2D structure, but performed conformational sampling of each graph-generated intermediate. This ensured that realistic sampling of configurations was included and avoided ZStruct's dependence on initial reactant alignment.

Applications of Green's approach have focused on combustion and atmospheric chemistry, and proved able to uncover complex, unexpected reaction pathways. For the six unimolecular organic reactions explored using the FSM-Berny method, 44 unique TSs were identified, only 6 of which were present in the RMG database. The TS search success rate was low, however, with TSs found for less than 10% of identified products.

Another method for generation reaction pathways was developed by Habershon, which also uses connectivity graphs to describe possible intermediates. Reaction pathways are examined by MD over a Hamiltonian that can be updated over time to sample different connectivity graphs.[130,131] The graphs are switched using a Monte Carlo procedure at a user-defined update probability, and the energetic cost of each switch is estimated. When changes are accepted due to being low enough in energy, the previously sampled reaction paths to existing intermediates are saved. Sampling on the updated Hamiltonian allows analysis of the next reaction pathway to the new intermediate (Figure 10). After dynamics sampling, the unique reaction pathways are subjected to the nudged elastic band method,[20,23,111,112] which further refines the paths.

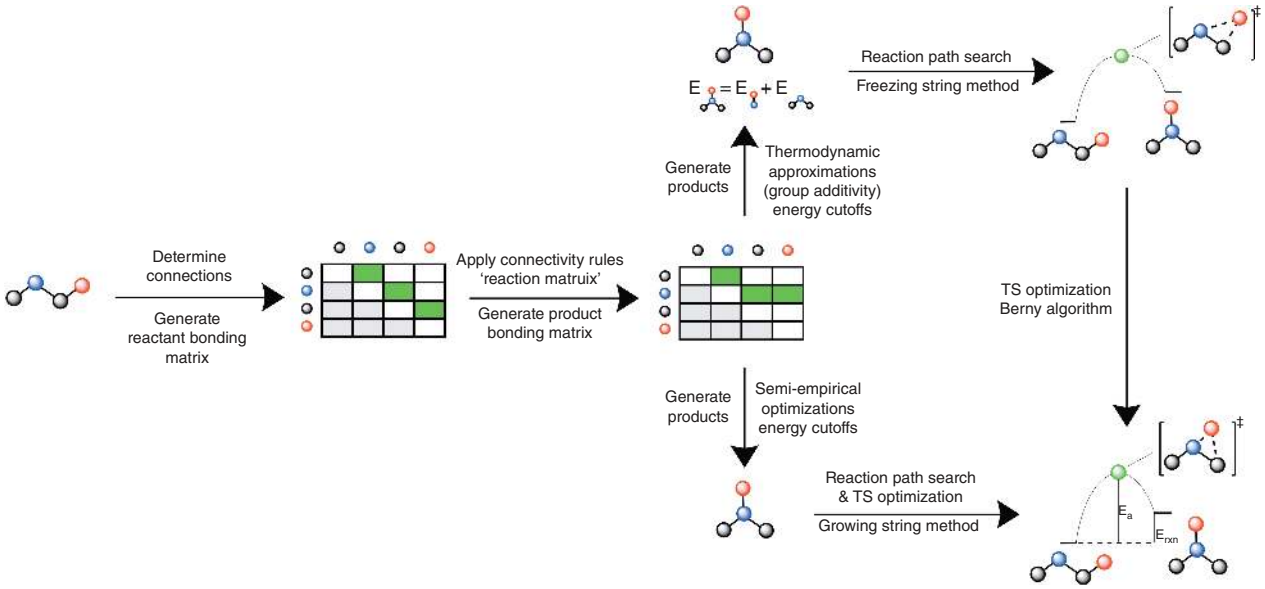

FIGURE 9 | Graphical methods used by Zimmerman (bottom pathway) and Green (top pathway) to generate plausible intermediates, followed by reaction path finding.

When tested on formaldehyde decomposition, the major pathways leading to H dissociation as well as H_2 formation were found, as expected, but some known high-energy intermediates were not found during the sampling procedure. In the Cocatalyzed hydroformylation of ethylene, the main reaction pathways for catalysis were identified (Heck–Breslow mechanism) using a DFT tight binding[132] PES, followed by optimization at an *ab initio* level of theory. Overall, the Habershon approach provides an advantage in more extensive reaction path sampling compared to the Zimmerman and Green methods, at the additional cost of requiring dynamics simulations.

An alternate means for reaction path generation created by Reiher and coworkers utilizes the concept of reactive sites,[133] (Figure 11) instead of reactive atoms connected in graphs, as was the case with Zimmerman, Green, and Habershon. Analogous to the use of arrow pushing or specifically encoded chemical reaction libraries, Reiher's method utilizes chemical descriptors tailored to the target chemical system to identify which atoms will most likely interact with one another during bond forming or breaking elementary steps. This heuristic approach generates high-energy structures by overlaying reactive sites, followed by optimization to identify real intermediates. Interpolation between structures provides TS guesses for subsequent TS optimization. In principle, the selection of heuristics on a system-specific basis means this method can work for organic and organometallic transformations as long as a good chemical descriptors can be defined.

Reiher's heuristics-guided approach was used to examine dinitrogen fixation by molybdenum via the Chatt–Shrock catalytic cycle.[133] Their method identified the traditional Chatt–Schrock cycle and multiple alternative pathways that could operate in tandem, as well as hints of pathways that could lead to catalyst deactivation. Overall, these benchmarks suggested this method—when initiated with quality

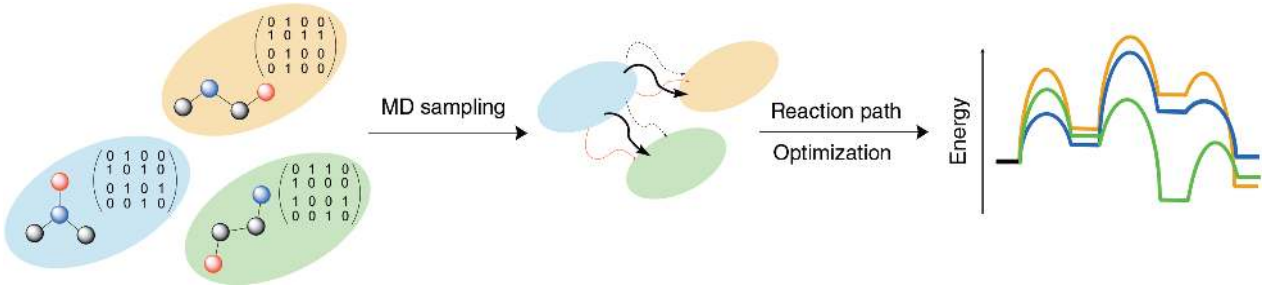

FIGURE 10 | Graph-based, dynamical reaction path sampling method.

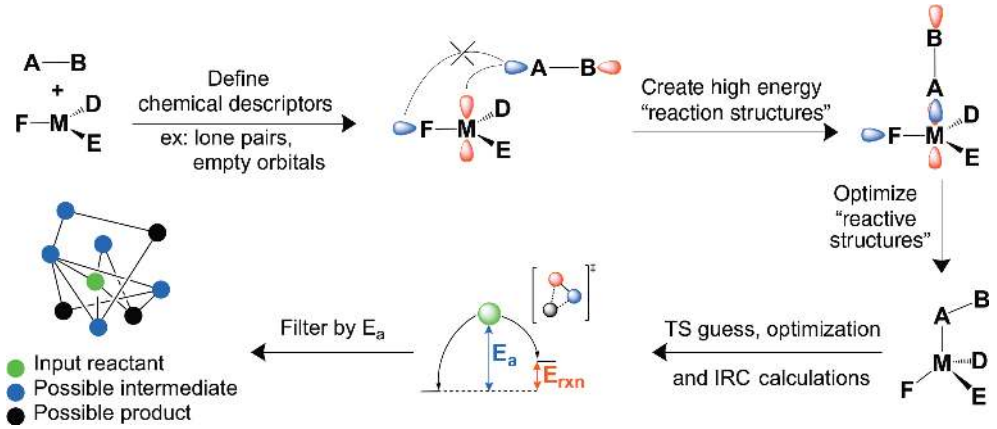

FIGURE 11 | Flow chart for the reaction path discovery method developed by Reiher and coworkers.

chemical intuition—is a powerful tool for examining complex reactive networks.

The methods for reaction path discovery discussed up to this point have proven most useful in exploring chemical systems involving uni- or bimolecular reactions. When multiple reactants are involved, or solvent may participate in the reaction, such methods are less useful. To address this challenge, Martínez and coworkers developed an *ab initio* MD tool for use in complex reaction environments[134] (Figure 12). In their Nanoreactor tool, high temperature and pressure dynamics simulations are run for long timescales through the assistance of fast, graphics processor unit-accelerated quantum chemistry. The chemical transformations that occur during the dynamics serve as input for subsequent double-ended reaction pathway optimization, and in the course of this process, intermediates, TSs, and their associated energetics are compiled.[135,136]

The Nanoreactor method has been applied to acetylene polymerization and Urey–Miller chemistry.[134,137] For acetylene polymerization, the nanoreactor was able to grow polymer chains consisting of more than 70 atoms over the time scale of the simulation (560 ps), showing the power of this tool in finding multistep, multicomponent reaction paths. The exploration of Urey–Miller chemistry generated more than 700 unique reactions during 1296 ps of simulation, including pathways leading to glycine and the formation of other amino acids. These examples show the nanoreactor has significant strengths in unbiased reaction network exploration, but like other MD-based methods, it comes with substantial costs.

Generating putative elementary steps by creating the intermediates first (i.e., *Concept 3*) has certain advantages over the direct TS searches of *Concept 2* tools. For example, intermediates can be screened for thermodynamics, and discarded if they are too high in energy. Double-ended reaction path finders tend to be more reliable than local TS optimizers as well. Double-ended methods, however, tend to struggle to find reaction paths when the two starting intermediates are connected by multiple elementary steps, and there is no guarantee the intermediate-generation tools will strictly identify single- versus multistep pathways. When this problem occurs, individual elementary steps are lost, or additional reaction path optimizations are needed to refine each step, a tedious correction at best. To combat these issues, single-ended reaction finding tools under *Concept 4* have been developed.

Concept 4: Using Reaction Coordinates to Find Reaction Paths

Using quantum chemistry to construct reaction pathways for even a single elementary step is a computationally costly, multistep procedure. In most *Concepts 2 and 3* strategies, TSs and reaction path

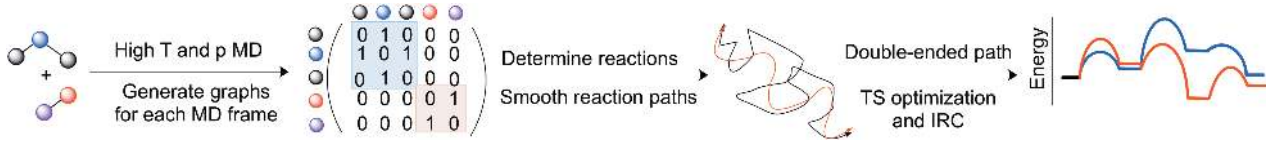

FIGURE 12 | Reaction pathway discovery using the nanoreactor.

approximations are formed first, and then TSs are optimized and IRC computations complete the reaction pathway. Because many of these steps may be costly or fail, constructing full reaction networks becomes hindered by difficulties in optimizing individual reaction paths. Streamlined approaches for finding reaction paths directly from mechanistic hypotheses therefore have considerable advantages over prior elementary step analysis strategies. For instance, the Zimmerman group has developed methods that systematically generate and follow qualitative reaction coordinates to construct a reaction path, find the TS, and locate an intermediate in a single computation. In related work, Li and coworkers developed a method that utilizes bond forming and breaking processes to drive reaction exploration. Both of these methods are designed to find single elementary steps, which reduces the cost of reaction exploration and addresses the challenges apparent in *Concept 3* reaction path optimizations.

The Zimmerman group's most recent reaction exploration tool, ZStruct2, combinatorially samples driving coordinates (DC), which are bond-addition or bond-breaking vectors that describe elementary reactions.[138] These reactive coordinates are designed for use with the single-ended GSM,[117] which generates reaction paths, TSs, and intermediate structures for single elementary steps that are consistent with the CD (Figure 13). ZStruct2 handles intramolecular and bimolecular reactions by aligning reactants in a way consistent with the CD, giving it an important advantage over the original ZStruct method. Variations on the CD and alignment rules, shown in Figure 14, allow for more complete sampling of reactions involving transition metals, which are particularly sensitive to 3D geometric structure.

The incorporation of knowledge of the transition metal center geometry, as well as aligning reactants, has allowed ZStruct2 to usefully inform the study of transition metal-catalyzed reactions. Recent reports from Zimmerman and coworkers have used ZStruct2 to explore Pd-catalyzed C-H arylation of piperidine,[139] $FeCl_3$-catalyzed carbonyl-olefin metathesis,[140,141] and Ni-catalyzed thiazole polymerization,[142] among others.[143,144] ZStruct2 was able to identify all of major steps of the catalytic cycle for piperidine arylation, including the roles that the multiple supporting reagents play in driving forward the reaction. In the study of thiazole polymerization, ZStruct identified an unexpected route for chain termination that was preventing controlled growth. Ultimately, this insight was used to design a new catalyst that showed improved polymerization activity.[142]

A related method developed by Li and coworkers combines MD and coordinate driving (CD) to search for reaction paths, as shown in Figure 15.[145] One advantage of this method is that it can sample conformational isomerization, which is achieved through MD simulations. To generate chemically interesting reactive pathways (e.g., those involving bonds breaking or forming) the interatomic distances between each pair of reactive atoms are set as reaction coordinates, and constrained optimization is used to approximate the reaction path. The lowest and highest energy structures from along the constrained optimization are then fully optimized without constraints to give exact intermediates and TSs.

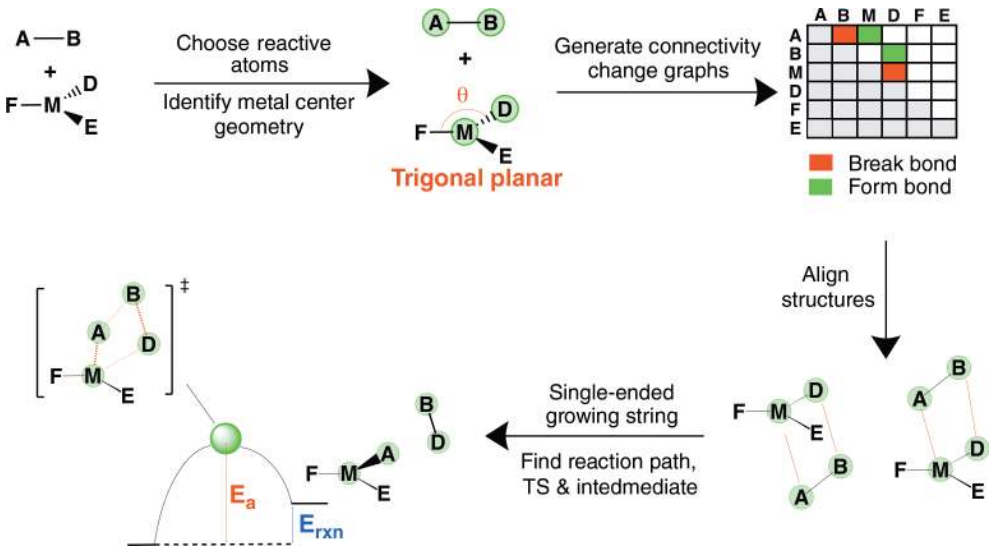

FIGURE 13 | ZStruct2 graphical driving coordinate generation and reaction path searches.

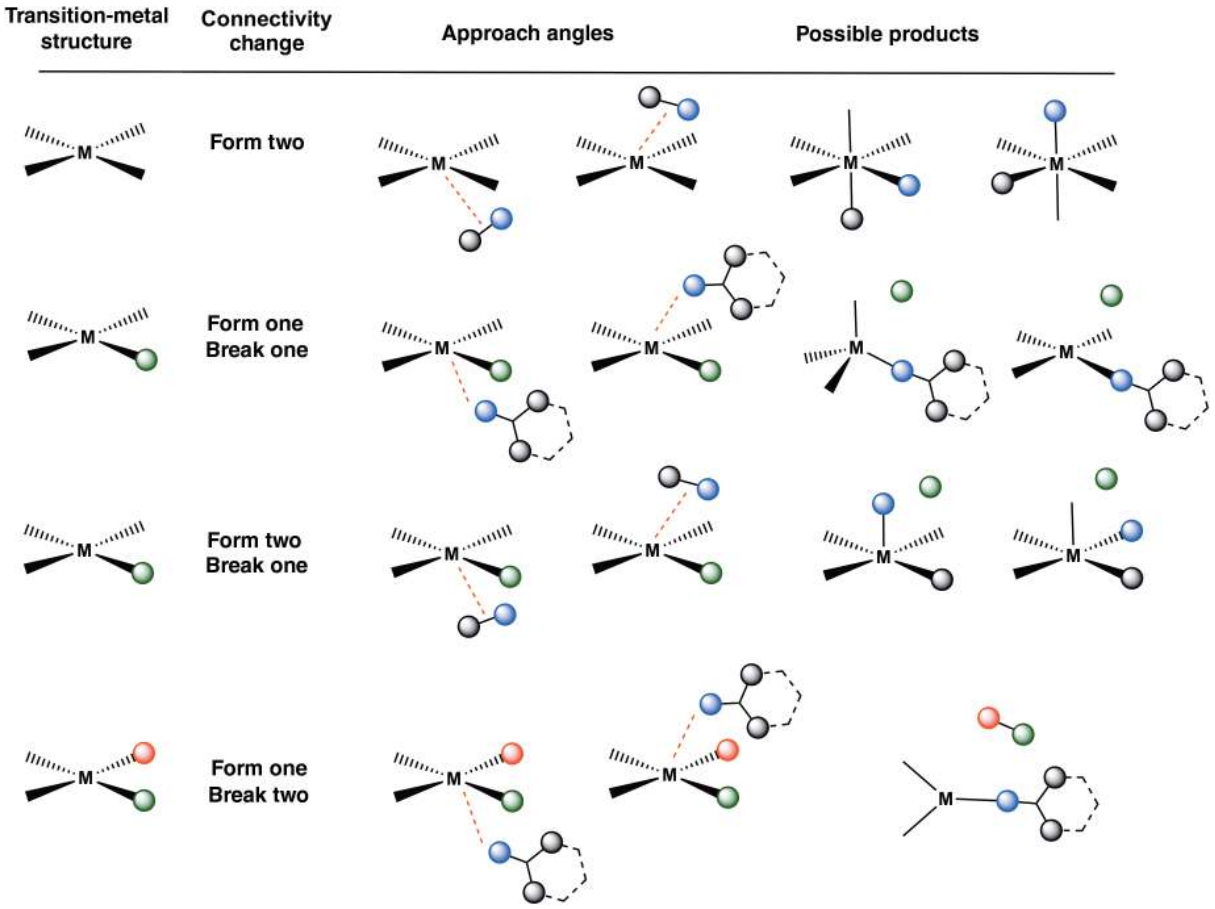

FIGURE 14 | Example of ZStruct2 reaction types and alignment of reactants at a square planar transition metal center.

The MD/CD method was tested on four common organic reactions, three of which were intramolecular, and one intermolecular. Previously identified reaction pathways found with other reaction exploration methods were confirmed, and the variety of chemical systems successfully tested indicates that the method is capable of exploring systems ranging in size from 10 to 50 atoms. The authors note,

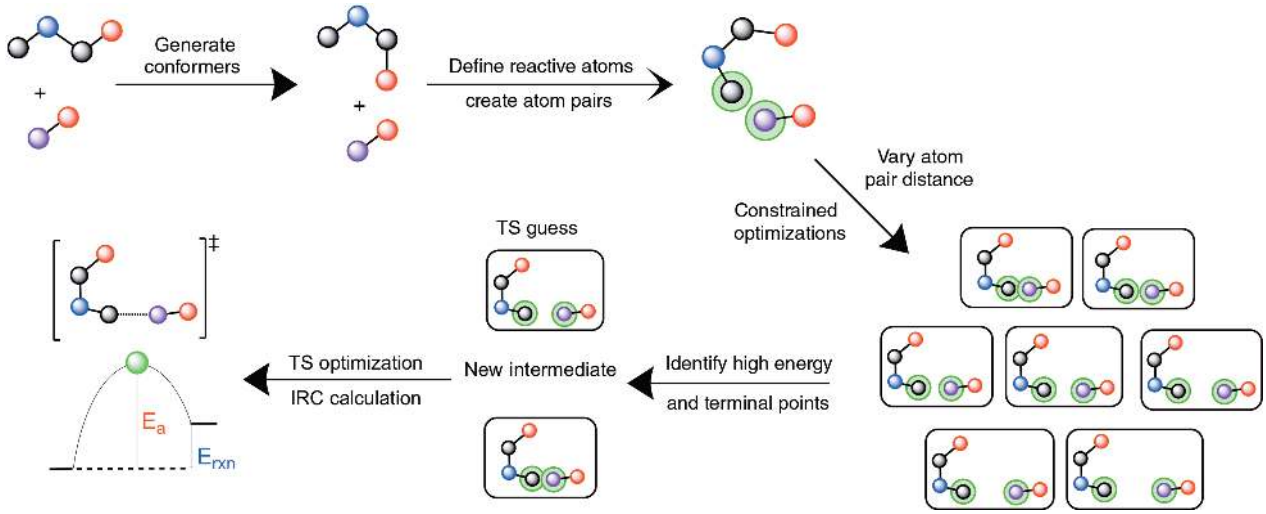

FIGURE 15 | Flow chart for molecular dynamics/coordinate driving method of Li and coworkers.[145]

however, that their MD/CD method is best utilized for organic molecules, as the semiempirical simulations cannot easily handle transition metal geometries.[145] As in Habershon's work (*Concept 3*), this choice is necessary because the use of MD largely restricts the underlying level of theory to be semiempirical, as also observed in the TSSCDS algorithm (*Concept 2*).

CONCLUSIONS

Comparison of Reaction Mechanism Exploration Tools

Concepts 1 to 4 are representative of contemporary and emerging strategies for reaction mechanism exploration. Two broad generalizations may be made here: (1) Graphical methods are highly prevalent, and the concept of the graph is implicit even when not specifically invoked and (2) Successful TS optimization is paramount for success in mechanism discovery. The foundations of graphical methods were laid in *Concept 1*, see Table 1 for a summary of those tools. *Concepts 2 to 4* are specifically relevant to TS optimization, and these are summarized in Table 2 and now discussed further.

A number of factors differentiate the individual methods of Table 2. Fundamentally, cost is a severe limitation because all quantum chemical methods require nontrivial computational resources. These costs can be difficult to estimate, however, as many source articles do not clearly state the computational resources used. For methods utilizing MD, however, semiempirical methods almost always must be employed, unless accelerated *ab initio* methods are available (such as GPU-accelerated DFT). Whenever low levels of theory are involved, the descriptions of reactive events will suffer. In our opinion, having a comprehensive reaction mechanism from a PES with qualitatively incorrect reactive landscape is most useful for benchmarking novel techniques, but not for examining real systems of contemporary chemical interest.

Under the constraint of available computational power, *Concepts 2 to 4* techniques reach different degrees of completeness in reaction path exploration. The ADDF method is likely the most comprehensive in this regard, but is also fundamentally limited by rapid increases in cost with growing system size. On the other side of (in)completeness, the heuristics-based approach of Reiher uses chemical intuition to greatly limit the number of reaction pathways that are searched. Reiher's method still locates a great number of pathways, however, showing how difficult it is to strike a balance between exploring full reactive space and maintaining a tractable computational burden.

In between these limits, a variety of tools are available from *Concepts 2 to 4*. These tools are semi-guided, in that they require some prior labeling of reactive atoms and designation of graph-allowed rearrangements. These choices restrict the reactive space, but it can be unclear in each application whether these restrictions are truly limiting. If new knowledge is gained about the reaction mechanism, any of these methods can claim success to the degree that this knowledge is useful. In practice, we believe that the scientific impact of the discovered knowledge measures the reaction exploration tool, but further conversations in the development community about benchmark systems for reaction discovery approaches will also be useful. We hope that continued development of these tools widens the scope of their application, with examination of emerging and unexplained reactions as an important focus.

These considerations all rely on tools which may or may not consistently converge reaction pathways and TSs. As noted in the beginning of this article, effective exploration of reaction paths (*Goal 2*) requires robustness in TS finding (*Goal 1*). While methods like the single-ended GSM and West's TS estimator have provided some advances in this regard, no method yet developed provides failsafe algorithms. A failed reaction path optimization may mean the path does not exist, is highly unfavorable, or simply that the optimizer was stuck and did not converge. This is troubling, as automated approaches would disregard any failed path, even if it were the actual major reaction pathway. Much work remains to be done to achieve greater fidelity in these simulations.

A Wish-List for Reaction Discovery Simulations

As progress in reaction mechanism exploration continues, we will move closer and closer to uncovering the full details of chemical reactions, with less and less guidance from user input and chemical intuition. The combinatorial complexity of chemistry means that simulation will never be useful for locating *all* elementary steps and TSs on high-dimensional surfaces. On the other hand, only the lowest energy, physically relevant reaction paths need to be found to usefully inform mechanistic understanding. This latter task may well be achievable with advances in simulation. Here, we provide a wish-list of goals that may be vital to accomplish this task.

TABLE 2 | Comparison of Recent Methods for Reaction Path Finding and Transition State Optimization

Method (Concept #)	Input Required	Method of Change	TS Finding Strategy	Intermediate Generation Strategy
ADDF[89,90] (2)	Reactant	Anharmonic downward distortion (ADD)	Anharmonic mode following then TS optimization	IRC
AFIR[93–96] (2)	Reactant(s)	Artificial external force	TS optimization along force-biased pathway	IRC
TSSCDS[100–102] (2)	Reactant(s)	High energy dynamics	Optimize TS from where bond change occurs	Dynamics and IRC
West[104] (2)	Reactant(s), library of TS geometries	Interatomic distances	Reactive atom constraints followed by TS optimization	IRC
ZStruct[113,114] (3)	Reactant, reactive atoms	Graph rules	Double-ended reaction path optimization	Graph rules
Green[127] (3)	Reactant, reactive atoms	Graph rules	Freezing string then local TS optimization	Graph rules
Habershon[130] (3)	Initial reactants and intermediates	Reaction/ graphical Hamiltonian	Double-ended reaction path optimization	Graph rules
Reiher[133] (3)	Reactants, reactive sites	Reactive sites (Heuristics)	Interpolation then local TS optimization	Heuristic rules and IRC
Nanoreactor[134] (3)	Reactants	High p, T dynamics	Double-ended reaction path optimization	MD trajectories
ZStruct2[138] (4)	Reactant(s), reactive Atoms	Graph rules	Single-ended growing string	Single-ended growing string
MD/CD[145] (4)	Reactant(s)	Distance between reactive atoms	Interpolation then local TS optimization	Trajectories, interatomic distances

ADDF, anharmonic downward distortion following; AFIR, artificial force-induced reaction; CD, driving coordinates; MD, molecular dynamics; IRC, intrinsic reaction coordinate; TS, transition state; TSSCDS, transition state search using chemical dynamics simulations.

- Delineation of all possible intermediates that are within a given number of elementary steps from currently known structures, including interactions between many reactive components and accessible conformational isomers.
- Accurate, but low-cost ranking of possible elementary transformations to determine those that are most kinetically viable, prior to quantum chemical optimizations.
- More generalized treatment of transition metals, surfaces, solvents, and nanoporous materials.
- Reaction path and TS optimizers that are fail-safe and require no user input.
- Streamlined procedures for setting up calculations, managing ongoing research, and pushing results to open-access databases.
- Visualization tools, based on kinetic analysis, that show the reaction network, rate-limiting steps, and potential channels for reactions outside of the known network.
- Procedures to generate experimentally testable hypotheses based on the simulated reaction mechanism.

These goals are within reach of imagination, yet far from trivial. Fortunately, many of these tasks can be accomplished individually, and later interfaced with other advances. The field of reaction exploration simulation may have far to travel, but the ultimate promise of fully *ab initio* reaction mechanism prediction is surely a goal worth continued effort for years to come.

ACKNOWLEDGMENT

The authors wish to thank Dow Chemical Company and National Science Foundation (NSF) (CHE-1551994) for continued support.

REFERENCES

1. Bell AT, Head-Gordon M. Quantum mechanical modeling of catalytic processes. *Annu Rev Chem Biomol Eng* 2011, 2:453–477 https://doi.org/10.1146/annurev-chembioeng-061010-114108.
2. Sperger T, Sanhueza IA, Kalvet I, Schoenebeck F, Computational studies of synthetically relevant homogeneous organometallic catalysis involving Ni, Pd, Ir, and Rh: an overview of commonly employed DFT methods and mechanistic insights. *Chem Rev* 115, 9532 (2015). DOI: https://doi.org/10.1021/acs.chemrev.5b00163.
3. Jackson K, Jaffar SK, Paton RS. Computational organic chemistry. *Annu Rev Prog Chem Sect B: Org Chem* 2013, 109:235 https://doi.org/10.1039/C3OC90007J.
4. Wales DJ, Bogdan TV. Potential energy and free energy landscapes. *J Phys Chem B* 2006, 110:20765–20776 https://doi.org/1021/jp0680544.
5. Wales DJ. *Energy Landscapes*. Cambridge: Cambridge University Press; 2003.
6. Schlegel HB. Geometry optimization on potential energy surfaces. In: Yarkony DR, ed. *Modern Electronic Structure Theory*. Singapore: World Scientific Publishing; 1995, 459–500.
7. Schlegel HB. Geometry optimization. In: PVR S, Allinger NL, Kollman PA, Clark T, Schaefer HF III, Gasteiger J, Schreiner PR, eds. *Encyclopedia of Computational Chemistry*, vol. 2. Chichester: John Wiley & Sons; 1998, 1136–1142.
8. Jensen F. Transition structure optimization techniques. In: PVR S, Allinger NL, Kollman PA, Clark T, Schaefer HF III, Gasteiger J, Schreiner PR, eds. *Encyclopedia of Computational Chemistry*, vol. 5. Chichester: John Wiley & Sons; 1998, 3114–3123.
9. Halgren TA, Lipscomb WN. The synchronous-transit method for determining reaction pathways and locating molecular transition states. *Chem Phys Lett* 1977, 49:225–232 https://doi.org/10.1016/0009-2614(77)80574-5.
10. Müller L, Brown LD. Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. *Theor Chim Acta* 1979, 53:75–93 https://doi.org/10.1007/BF00547608.
11. Dewar MJS, Healy EF, Stewart JJP. Location of transition states in reaction mechanisms. *J Chem Soc Faraday Trans* 1984, 80:227–233 https://doi.org/10.1039/F29848000227.
12. Elber R, Karplus M. A method for determining reaction paths in large molecules: application to myoglobin. *Chem Phys Lett* 1987, 139:375–380 https://doi.org/10.1016/0009-2614(87)80576-6.
13. Ionova IV, Carter EA. Ridge method for finding saddle points on potential energy surfaces. *J Chem Phys* 1993, 98:6377–6386 https://doi.org/10.1063/1.465100.
14. Cerjan CJ, Miller WH. On finding transition states. *J Chem Phys* 1981, 75:2800–2806 https://doi.org/10.1063/1.442352.
15. Banerjee A, Adams N, Simons J, Shepard R. Search for stationary points on surfaces. *J Phys Chem* 1985, 89:52–57 https://doi.org/10.1021/j100247a015.
16. Baker J. An algorithm for the location of transition states. *J Comput Chem* 1986, 7:385–395 https://doi.org/10.1002/jcc.540070402.
17. Henkelman G, Jonsson H. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *J Chem Phys* 1999, 111:7010–7022 https://doi.org/10.1063/1.480097.
18. Bolhuis P, Changler D, Dellago C, Geissler PL. Transition path sampling: throwing ropes over Rough Mountain passes, in the dark. *Annu Rev Phys Chem* 2002, 53:291–318 https://doi.org/10.1146/annurev.physchem.52.082301.113146.
19. Dellago C, Bolhuis PG, Geissler P. Transition path sampling. *Adv Chem Phys* 2002, 123:1–84 https://doi.org/10.1002/0471231509.ch1.
20. Henkelman G, Honsson 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 https://doi.org/10.1063/1.1323224.
21. Fischer S, Karplus M. Conjugate peak refinement: an algorithm for finding reaction paths and accurate transition states in systems with many degrees of freedom. *Chem Phys Lett* 1992, 194:252–261 https://doi.org/10.1016/0009-2614(92)85543-J.
22. Ayala PY, Schlegel HB. A combined method for determining reaction paths, minima, and transition state geometries. *J Chem Phys* 1997, 107:375–384 https://doi.org/10.1063/1.474398.
23. Maragakis P, Andreev SA, Brumer Y, Reichman DR, Kaxiras E. Adaptive nudged elastic band approach for transition state calculation. *J Chem Phys* 2002, 117:4651–4658 https://doi.org/10.1063/1.1495401.
24. Fukui K. The path of chemical reactions—the IRC approach. *Acc Chem Res* 1981, 14:363–368 https://doi.org/10.1021/ar00072a001.
25. Maeda S, Harabuchi Y, Ono Y, Taketsugu T, Morokuma K. Intrinsic reaction coordinate: calculation, bifurcation, and automated search. *Int J Quantum Chem* 2015, 115:258–269 https://doi.org/10.1002/qua.24757.
26. Grzybowski BA, Bishop KJM, Kowalczyk B, Wilmer CE. The 'wired' universe of organic chemistry. *Nat Chem* 2009, 1:31–36 https://doi.org/10.1038/nchem.136.

27. Graulich N, Hopf H, Schreiner PR. Heuristic thinking makes a chemist smart. *Chem Soc Rev* 2010, 39:1503–1512 https://doi.org/10.1039/B911536F.
28. Schlegel BH. Exploring potential energy surfaces for chemical reactions: an overview of some practical methods. *J Comput Chem* 2002, 24:1514–1527 https://doi.org/10.1002/jcc.10231.
29. Schlegel BH. Geometry optimization. *Wiley Interdiscip Rev Comput Mol Sci* 2011, 1:790–809 https://doi.org/10.1002/wcms.34.
30. Todd MH. Computer-aided organic synthesis. *Chem Soc Rev* 2005, 34:247–266 https://doi.org/10.1039/B104620A.
31. Mulholland AJ. Modeling enzyme reaction mechanisms, specificity and catalysis. *Drug Disc Today* 2005, 10:1393–1402 https://doi.org/10.1016/S1359-6446(05)03611-1.
32. Zhang Y, Liu H, Yang W. Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined *ab initio* QM/MM potential energy surface. *J Chem Phys* 2000, 112:3483–3492 https://doi.org/10.1063/1.480503.
33. Acevedo O, Jorgensen WL. Advances in quantum and molecular mechanical (QM/MM) simulations for organic and enzymatic reactions. *Acc Chem Res* 2010, 142:142–151 https://doi.org/10.1021/ar900171c.
34. Marti S, Moliner V. Improving the QM/MM description of chemical processes: a dual level strategy to explore the potential energy surface in very large systems. *J Chem Theory Comput* 2005, 1:1008–1016 https://doi.org/10.1021/ct0501396.
35. Sousa SF, Ribeiro AJM, Neves RPP, Bras NF, Cerqueira NMFSA, Fernandes PA, Ramos MJ. Application of quantum mechanics/molecular mechanics methods in the study of enzymatic reaction mechanisms. *Wiley Interdiscip Rev Comput Mol Sci* 2017, 7:e1281. https://doi.org/10.1002/wcms.1281.
36. Ensing B, De Vivo M, Liu Z, Moore P, Klein ML. Metadynamics as a tool for exploring free energy landscapes of chemical reactions. *Acc Chem Res* 2006, 39:73–81 https://doi.org/10.1021/ar040198i.
37. Dontgen M, Przybylski-Freund M-D, Kroger LC, Kopp WA, Ismail AE, 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 https://doi.org/10.1021/acs.jctc.5b00201.
38. Ianuzzi M, Laio A, Parrinello M. Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics. *Phys Rev Lett* 2003, 90:328302 https://doi.org/10.1103/PhysRevLett.90.238302.
39. Laio A, Parrinello M. Escaping free-energy minima. *Proc Nat Acad Sci* 2002, 99:12562–12566 https://doi.org/10.1073/pnas.202247399.
40. Haag MP, Marti KH, Reiher M. Generation of potential energy surfaces in high dimensions and their haptic exploration. *Chem Phys Chem* 2011, 12:3204–3213 https://doi.org/10.1002/cphc.201100539.
41. Marti KH, Reiher M. Haptic quantum chemistry. *J Comput Chem* 2009, 30:2010–2020 https://doi.org/10.1002/jcc.21201.
42. Kayala MA, Azencott CA, Chen JH, Baldi P. Learning to predict chemical reactions. *J Chem Inf Model* 2011, 51:2209–2222 https://doi.org/10.1021/ci200207y.
43. Chen JH, Baldi P. No electron left behind: a rule-based expert system to predict chemical reactions and reaction mechanisms. *J Chem Inf Model* 2009, 49:2034–2043 https://doi.org/10.1021/ci900157k.
44. Kayala MA, Baldi P. Reaction predictor: prediction of complex chemical reactions at the mechanistic level using machine learning. *J Chem Inf Model* 2012, 52:2526–2540 https://doi.org/10.1021/ci3003039.
45. Schwabe T, Grimme S. Theoretical thermodynamics for large molecules: walking the thin line between accuracy and computational cost. *Acc Chem Res* 2008, 41:569–579 https://doi.org/10.1021/ar700208h.
46. Mardirossian N, Head-Gordon M. ωB97X-V: a 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy. *Phys Chem Chem Phys* 2014, 16:9904–9924 https://doi.org/10.1039/C3CP54374A.
47. Yang KR, Jalan A, Green WH, Truhlar DG. Which *ab initio* wave function methods are adequate for quantitative calculations of the energies of biradicals? The performance of coupled-cluster and multi-reference methods along a single-bond dissociation coordinate. *J Chem Theory Comput* 2013, 9:418–431 https://doi.org/10.1021/ct3009528.
48. Zimmerman PM. Incremental full configuration interaction. *J Chem Phys* 2017, 146:104102 https://doi.org/10.1063/1.4977727.
49. Broadbelt LJ, Stark SM, Klein MT. Computer generated pyrolysis modeling: on-the-fly generation of species, reactions, and rates. *Ind Eng Chem Res* 1994, 33:790–799 https://doi.org/10.1021/ie00028a003.
50. Broadbelt LJ, Stark SM, Klein MT. Computer generated reaction modelling: decomposition and encoding algorithms for determining species uniqueness. *Comput Chem Eng* 1996, 20:113–129 https://doi.org/10.1016/0098-1354(94)00009-D.
51. Broadbelt LJ, Pfaendtner J. Lexicography of kinetic modeling of complex reaction networks. *AIChE J*

2005, 51:2112–2121 https://doi.org/10.1002/aic.10599.

52. Evans MG, Polanyi M. Inertia and driving force of chemical reactions. *Trans Faraday Soc* 1938, 34:11 https://doi.org/10.1039/TF9383400011.
53. Benson SW, Buss JH. Additivity rules for the estimation of molecular properties. Thermodynamic properties. *J Chem Phys* 1958, 29:546–572 https://doi.org/10.1063/1.1744539.
54. Cohen N, Benson SW. Estimation of heats of formation of organic compounds by additivity methods. *Chem Rev* 1993, 93:2419–2438 https://doi.org/10.1021/cr00023a005.
55. Broadbelt LJ, Stark SM, Klein MT. Termination of computer-generated reaction mechanisms: species rank-based convergence criterion. *Ind Eng Chem Res* 1995, 34:2566–2573 https://doi.org/101021/ie00047a003.
56. Susnow RG, Dean AM, Green WH, Peczak P, Broadbelt LJ. Rate-based construction of kinetic models for complex systems. *J Phys Chem A* 1997, 101:3731–3740 https://doi.org/10.1021/jp9637690.
57. De Witt MJ, Dooling DJ, Broadbelt LJ. Computer generation of reaction mechanisms using quantitative rate information: application to long-chain hydrocarbon pyrolysis. *Ind Eng Chem Res* 2000, 39:2228–2237.
58. Klinke DJ III, Broadbelt LJ. Construction of a mechanistic model of Fischer–Tropsch synthesis on Ni(1 1 1) and co(0 0 0 1) surfaces. *Chem Eng Sci* 1999, 54:3379–3389 https://doi.org/10.1016/S0009-2509(98)00386-8.
59. Matheu DM, Dean AM, Grenda JM, Green WH. Mechanism generation with integrated pressure dependence: a new model for methane pyrolysis. *J Phys Chem A* 2003, 107:8552–8565 https://doi.org/10.1021/jp0345957.
60. Gao CW, Allen JW, Green WH, West RH. Reaction mechanism generator: automatic construction of chemical kinetic mechanisms. *Comput Phys Commu* 2016, 203:212–225 https://doi.org/10.1016/j.cpc.2016.02.013.
61. Magoon GR, Green WH. 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 https://doi.org/10.1016/j.compchemeng.2012.11.009.
62. Harper MR, Van Geem KM, Pyl SP, Marin GB, Green WH. Comprehensive reaction mechanism for n-butanol pyrolysis and combustion. *Combust Flame* 2011, 158:16–41 https://doi.org/10.1016/J.combustflame.2010.06.002.
63. Hansen N, Merchant SS, Harper MR, Green WH. The predictive capability of an automatically generated combustion chemistry mechanism: chemical structures of premixed iso-butanol flames. *Combust Flame* 2013, 160:2343–2351 https://doi.org/10.1016/j.combustflame.2013.05.013.
64. Slakman BL, Simka H, Reddy H, West RH. Extending reaction mechanism generator to silicon hydride chemistry. *Ind Eng Chem Res* 2016, 55:12507–12515 https://doi.org/10.1021/acs.iecr.6b02402.
65. Petway SV, Ismail H, Green WH, Estupiñán EG, Jusinski LE, Taatjes CA. Measurements and automated mechanism generation modeling of OH production in photolytically initiated oxidation of the neopentyl radical. *J Phys Chem A* 2007, 111:3891–3900 https://doi.org/10.1021/jp0668549.
66. Van Geem KM, Reyniers MF, Marin GB, Song J, Green WH, Matheu DM. Automatic reaction network generation using RMG for steam cracking of n-hexane. *AIChE J* 2006, 52:718–730 https://doi.org/10.1002/aic.10655.
67. Seyedzadeh Khanshan F, West RH. Developing detailed kinetic models of syngas production from bio-oil gasification using reaction mechanism generator (RMG). *Fuel* 2016, 163:25–33 https://doi.org/10.1016/j.fuel.2015.09.031.
68. Rappoport D, Galvin CJ, Zubarev DY, Aspuru-Guzik A. Complex chemical reaction networks from heuristics-aided quantum chemistry. *J Chem Theory Comput* 2014, 10:897–907 https://doi.org/10.1021/ct401004r.
69. Mizuno T, Weiss AH. Synthesis and utilization of formose sugars. *Adv Carbohyd Chem Biochem* 1974, 29:173–227 https://doi.org/10.1016/S0065-2318(08)60250-4.
70. Appayee C, Breslow R. Deuterium studies reveal a new mechanism for the Formose reaction involving hydride shifts. *J Am Chem Soc* 2014, 136:3720–3723 https://doi.org/10.1021/ja410886c.
71. Davis HL, Wales DJ, Berry RS. Exploring potential energy surfaces with transition state calculations. *J Chem Phys* 1990, 9:4308–4319 https://doi.org/10.1063/1.457790.
72. Tsai CJ, Jordan KD. Use of an eigenmode method to locate the stationary points on the potential energy surfaces of selected argon and water clusters. *J Phys Chem* 1993, 97:11227–11237 https://doi.org/10.1021/100145a019.
73. Doye JPK, Wales DJ. Surveying a potential energy surface by eigenvector-following. *Z Phys D* 1997, 40:194–197.
74. Basilevsky MV, Shamov AG. 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 https://doi.org/10.1016/0301-0104(81)80170-X.

75. Basilevsky MV. The topography of potential energy surfaces. *Chem Phys* 1982, 67:337–346 https://doi.org/10.1016.03001-0104(82)85194-X.
76. Hoffman DK, Nord RS, Ruedenberg K. Gradient extremals. *Theor Chim Acta* 1986, 69:265–279 https://doi.org/10.1007/BF00527704.
77. Jorgensen P, Jorgen H, Jensen A, Helgaker T. A gradient extremal walking algorithm. *Theor Chim Acta* 1988, 73:55–65 https://doi.org/10.1007/BF00526650.
78. Quapp W. Gradient extremals and valley floor bifurcations on potential energy surfaces. *Theor Chim Acta* 1989, 75:447–460 https://doi.org/10.1007/BF00527676.
79. Schlegel HB. Following gradient extremal paths. *Theor Chim Acta* 1992, 83:15–20 https://doi.org/10.1007/BF01113240.
80. Sun J, Ruedenberg K. Gradient extremals and steepest descent lines on potential energy surfaces. *J Chem Phys* 1993, 98:9707–9714 https://doi.org/10.1063/1.464394.
81. Bondensgard K, Jensen F. Gradient extremal bifurcation and turning points: an application to the H_2CO potential energy surface. *J Chem Phys* 1996, 104:8025–8031 https://doi.org/10.1063/1.471495.
82. 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 https://doi.org/10.1002/(SICI)1096-987X(19980715)19:9<1087::AID-JCC9>3.0.CO;2-M.
83. Quapp W, Hirsch M, Heidrich D. Bifurcation of reaction pathways: the set of valley ridge inflection points of a simple three-dimensional potential energy surface. *Theor Chem Acc* 1998, 100:285–299 https://doi.org/10.1007/s0021400503.
84. Bofill JM, Anglada JM. Finding transition states using reduced potential-energy surfaces. *Theor Chem Acc* 2001, 105:463–472 https://doi.org/10.1007/s0021400002.
85. Crehuet R, Bofill JM, Anglada JM. A new look at the reduced-gradient-following path. *Theor Chem Acc* 2002, 107:130–139 https://doi.org/10.1007/s00214-001-0306.
86. Dallos M, Lischka H, Ventura Do Monte E, Hirsch M, Quapp W. Determination of energy minima and saddle points using multireference configuration interaction methods in combination with reduced gradient following: the S0 surface of H_2CO and the T1 and T2 surfaces of acetylene. *J Comput Chem* 2002, 23:576–583 https://doi.org/10.1002/jcc.10054.
87. Hirsch M, Quapp W. Reaction channels of the potential energy surface: application of newton trajectories. *J Mol Struct Theochem* 2004, 683:1–13 https://doi.org/10.1016/j.theochem.2004.06.009.
88. Balcells D, Clot E, Eisenstein O, Nova A, Perrin L. Deciphering selectivity in organic reactions: a multifaceted problem. *Acc Chem Res* 2016, 49:1070–1078 https://doi.org/10.1021/acs.accounts.6b00099.
89. Ohno K, Maeda S. A scaled hypersphere search method for the topography of reaction pathways on the potential energy surface. *Chem Phys Lett* 2004, 384:277–282 https://doi.org/10.1021/jp0513162.
90. Maeda S, Ohno K. Global mapping of equilibrium and transition structures on potential energy surfaces by the scaled hypersphere search method: applications to *ab initio* surfaces of formaldehyde and propyne molecules. *J Phys Chem A* 2005, 109:5742–5753 https://doi.org/10.1021/jp0513162.
91. Ohno K, Maeda S. Global reaction route mapping on potential energy surfaces of formaldehyde, formic acid, and their metal-substituted analogues. *J Phys Chem A* 2006, 110:8933–8941 https://doi.org/10.1021/jp061149I.
92. Maeda S, Ohno K, Morokuma K. Systematic exploration of the mechanism of chemical reactions: the global reaction route mapping (GRRM) strategy using the ADDF and AFIR methods. *Phys Chem Chem Phys* 2013, 15:3683–3701 https://doi.org/10.1039/C3CP44063J.
93. Maeda S, Morokuma K. Communications: a systematic method for locating transition structures of A +B→X type reactions. *J Chem Phys* 2010, 132:241102 https://doi.org/10.1063.1.3457903.
94. Maeda S, Morokuma K. Finding reaction pathways of type A + B → X: toward systematic prediction of reaction mechanisms. *J Chem Theory Comput* 2011, 7:2335–2345 https://doi.org/10.1021/ct200290m.
95. 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 https://doi.org/10.1002/jcc.23481.
96. 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 https://doi.org/10.1002/tcr.201600043.
97. Puripat M, Ramozzi R, Hatanaka M, Parasuk W, Parasuk V, Morokuma K. The Biginelli reaction is a urea-catalyzed Organocatalytic multicomponent reaction. *J Org Chem* 2015, 80:6959–6967 https://doi.org/10.1021/acs.joc.5b00407.
98. Hebrard F, Kalck P. Cobalt-catalyzed Hydroformylation of alkenes: generation and recycling of the carbonyl species, and catalytic cycle. *Chem Rev* 2009, 109:4272–4282 https://doi.org/10.1021/cr8002533.
99. Vázquez SA, Martínez-Núñez E. HCN elimination from vinyl cyanide: product energy partitioning, the role of hydrogen–deuterium exchange reactions and a

new pathway. *Phys Chem Chem Phys* 2015, 17:6948–6955 https://doi.org/10.1039/C4CP05626D.

100. Martínez-Núñez E. An automated method to find transition states using chemical dynamics simulations. *J Comput Chem* 2015, 36:222–234 https://doi.org/10.1002/jcc.23790.
101. Martínez-Núñez E. An automated transition state search using classical trajectories initialized at multiple minima. *Phys Chem Chem Phys* 2015, 17:14912–14921 https://doi.org/10.1039/C5CP02175H.
102. Varela JA, Vázquez SA, Martínez-Núñez E. An automated method to find reaction mechanisms and solve the kinetics in organometallic catalysis. *Chem Sci* 2017, 8:3843–3851 https://doi.org/10.1039/C7SC00549K.
103. Cohen AJ, Mori-Sánchez P, Weitao Y. Challenges for density functional theory. *Chem Rev* 2012, 112:289–320 https://doi.org/10.1021/cr200107z.
104. Bhoorasingh PL, West RH. Transition state geometry prediction using molecular group contributions. *Phys Chem Chem Phys* 2015, 17:32173–32182 https://doi.org/10.1039/C5CP04706D.
105. Bhoorasingh PL, Slakman BL, Khanshan FS, Cain JY, West RH. Automated transition state theory calculations for high-throughput kinetics. *J Phys Chem A* 2017, 121:6896–6904 https://doi.org/10.1021/acs.jpca.7b07361.
106. Allen JW, Scheer AM, Gao CW, Merchant SS, Vasu SS, Welz O, Savee JD, Osborn DL, Lee C, Vranckx S, et al. A coordinated investigation of the combustion chemistry of diisopropyl ketone, a prototype for biofuels produced by endophytic fungi. *Combust Flame* 2014, 161:711–724 https://doi.org/10.1016/j.combustflame.2013.10.019.
107. Doney AC, Rooks BJ, Lu T, Wheeler SE. Design of organocatalysts for asymmetric propargylations through computational screening. *ACS Catal* 2016, 6:7948–7955 https://doi.org/10.1021/acscatal.6b02366.
108. E W RW, Vanden-Eijnden E. String method for the study of rare events. *Phys Rev B* 2002, 66:052301 https://doi.org/10.1103/PhysRevB.66.052301.
109. E W RW, Vanden-Eijnden E. Finite temperature string method for the study of rare events. *J Phys Chem B* 2005, 109:6688–6693, 6693 https://doi.org/10.1021/jp0455430.
110. Peters B, Heyden A, Bell AT, Chakraborty A. A growing string method for determining transition states: comparison to the nudged elastic band and other string methods. *J Chem Phys* 2004, 120:7877–7886 https://doi.org/10.1063/1.1691018.
111. Henkelman G, Uberuaga BP, Jonsson H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J Chem Phys* 2000, 113:9901–9904 https://doi.org/10.1063/1.1329672.
112. Trygubenko SA, Wales DJ. A doubly nudged elastic band method for finding transition states. *J Chem Phys* 2004, 120:2082–2094 https://doi.org/10.1063/1.1636455.
113. Zimmerman PM. Automated discovery of chemically reasonable elementary reaction steps. *J Comput Chem* 2013, 34:1385–1392 https://doi.org/10.1002/jcc.23271.
114. Zimmerman PM. Navigating molecular space for reaction mechanisms: an efficient, automated procedure. *Mol Simul* 2014, 41:43–54 https://doi.org/10.1080/08927022.2014.894999.
115. Zimmerman PM. Growing string method with interpolation and optimization in internal coordinates: methods and examples. *J Chem Phys* 2013, 138:184102 https://doi.org/10.1063/1.4804162.
116. Zimmerman PM. Reliable transition state searches integrated with the growing string method. *J Chem Theory Comput* 2013, 9:3043–3050 https://doi.org/10.1021/ct400319w.
117. Zimmerman PM. Single-ended transition state finding with the growing string method. *J Comput Chem* 2015, 36:601–611 https://doi.org/10.1002/jcc.23833.
118. Jafari M, Zimmerman PM. Reliable and efficient reaction path and transition state finding for surface reactions with the growing string method. *J Comput Chem* 2017, 38:645–658 https://doi.org/10.1002/jcc.24720.
119. Nett AJ, Wanxiang Z, Zimmerman PM, 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 https://doi.org/10.1021/jacs.5b04548.
120. Sun Z, Winschel GA, Zimmerman PM, Nagorny P. Enantioselective synthesis of piperidines through the formation of chiral mixed phosphoric acid acetals: experimental and theoretical studies. *Angew Chem Int Edt* 2014, 53:11194–11198 https://doi.org/10.1002/anie.201405128.
121. Khomutnyk YY, Argüelles AJ, Winschel GA, Sun Z, Zimmerman PM, 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 https://doi.org/10.1021/jacs.5b12528.
122. Tay J-H, Argüelles AJ, DeMars MD, Zimmerman PM, Sherman DH, Nagorny P. Regiodivergent Glycosylations of 6-deoxy-erythronolide B and oleandomycin-derived macrolactones enabled by chiral acid catalysis. *J Am Chem Soc* 2017, 139:8570–8578 https://doi.org/10.1021/jacs.7b03198.

123. Bhunya S, Zimmerman PM, Paul A. Unraveling the crucial role of metal-free catalysis in borazine and polyborazylene formation in transition-metal-catalyzed ammonia-borane dehydrogentation. *ACS Catal* 2015, 5:3478–3493 https://doi.org/10.1021/cs502129m.

124. Li M, Pendleton IM, Nett AJ, Zimmerman PM. Mechanism for forming B, C, N, O rings from NH_3BH_3 and CO_2 via reaction discovery computations. *J Phys Chem A* 2016, 120:1135–1144 https://doi.org/10.1021/acs.jpca.5b11156.

125. Pendleton IM, Perez-Temprano MH, Sanford MS, Zimmerman PM. Experimental and computational assessment of reactivity and mechanism in C(sp^3)-N bond-forming reductive elimination from palladium(IV). *J Am Chem Soc* 2016, 138:6049–6060 https://doi.org/10.1021/jacs.6b02714.

126. Zhao Y, Nett AJ, McNeil AJ, Zimmerman PM. Computational mechanism for initiation and growth of poly(3-hexylthiophene) using palladium N-heterocyclic carbene precatalysts. *Macromolecules* 2016, 49:7632–7641 https://doi.org/10.1021/acs.macromol.6b01648.

127. Suleimanov YV, Green WH. Automated discovery of elementary chemical reaction steps using freezing string and Berny optimization methods. *J Chem Theor Comput* 2015, 11:4248–4259 https://doi.org/10.1021/acs.jctc.5b00407.

128. Behn A, Zimmerman PM, Bell AT, Head-Gordon M. Efficient exploration of reaction paths via a freezing string method. *J Chem Phys* 2011, 135:224108 https://doi.org/10.1063/1.3664901.

129. Peng C, Schlegel BH. Combining synchronous transit and quasi-newton methods to find transition states. *Israeli J Chem* 1993, 33:449–454 https://doi.org/10.1002/ijch.199300051.

130. Haberson S. Sampling reactive pathways with random walks in chemical space: applications to molecular dissociation and catalysis. *J Chem Phys* 2015, 143:094106 https://doi.org/10.1063/1.4929992.

131. Haberson S. Automated prediction of catalytic mechanism and rate law using graph-based reaction path sampling. *J Chem Theory Comput* 2016, 12:1786–1798 https://doi.org/10.1021/acs.jctc.6b00005.

132. Seifert G, Joswig JO. Density-functional tightbinding—an approximate density-functional theory method. *Wiley Interdiscip Rev Comput Mol Sci* 2012, 2:456–465 https://doi.org/10.1002/wcms.1094.

133. Bergler M, Simm GN, Proppe J, Reiher M. Heuristics-guided exploration of reaction mechanisms. *J Chem Theor Chem* 2015, 11:5712–5722 https://doi.org/10.1021/acs.jctc.5b00866.

134. Wang LP, Titov A, McGibbon R, Liu F, Pande VS, Martinez TJ. Discovering chemistry with an *ab initio* nanoreactor. *Nat Chem* 2014, 6:1044–1048 https://doi.org/10.1038/nchem.2099.

135. Wang LP, McGibbon RT, Pande VS, Martinez TJ. Automated discovery and refinement of reactive molecular dynamics pathways. *J Chem Theory Comput* 2016, 12:638–649 https://doi.org/10.1021/acs.jctc.5b00830.

136. Martinez TJ. *Ab initio* reactive computer aided molecular design. *Acc Chem Res* 2017, 50:652–656 https://doi.org/10.1021/acs.accounts.7b00010.

137. Goldman N. Accelerated reaction simulations a virtual squeeze on chemistry. *Nat Chem* 2014, 6:1033–1034 https://doi.org/10.1038/nchem.2118.

138. Dewyer AL, Zimmerman PM. Finding reaction mechanisms, intuitive or otherwise. *Org Biomol Chem* 2017, 15:501–504 https://doi.org/10.1039/c6ob02183b.

139. Dewyer AL, Zimmerman PM. Simulated mechanism for palladium-catalyzed, directed γ-arylation of piperidine. *ACS Catal* 2017, 7:5466–5477 https://doi.org/10.1021/acscatal7b01390.

140. Ludwig JR, Zimmerman PM, Gianino JB, Schindler CS. Iron(III)-catalyzed carbonyl olefin metathesis. *Nature* 2016, 533:374–379 https://doi.org/10.1038/nature17432.

141. Ludwig JR, Phan S, McAtee CC, Zimmerman PM, Devery JJ III, Schindler CS. Mechanistic investigations of the iron(III)-catalyzed carbonyl-olefin metathesis. *J Am Chem Soc* 2017, 139:10832–10842 https://doi.org/10.1021/jcas.7b05641.

142. Smith ML, Leone AK, Zimmerman PM, McNeil AJ. Impact of preferential π-binding in catalyst-transfer polycondensation of thiazole derivatives. *ACS Macro Lett* 2016, 5:1411–1415 https://doi.org/10.1021/acsmacrolett.6b00886.

143. Ellington BR, Paul B, Vitek AK, Das D, Zimmerman PM, Marsh ENG. A novel iron-dependent oxidative deformylation reaction providing insight into hydrocarbon biosynthesis in nature. *ACS Catal* 2016, 6:3293–3300 https://doi.org/10.1021/acscatal.6b00592.

144. Hale LVA, Malakar T, Tseng KNT, Zimmerman PM, Szymczak NK. The mechanism of acceptorless double dehydrogenation by N,N,N-amine ruthenium(II) hydrides: a combined experimental and computational study. *ACS Catal* 2016, 6:4799–4813 https://doi.org/10.1021/asccatal.6b01465.

145. 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 https://doi.org/10.1021/acs.jpca.6b12195.

