THE JOURNAL OF PHYSICAL CHEMISTRY A
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCA

Article

Efficient Acceleration of Reaction Discovery in the *Ab Initio* Nanoreactor: Phenyl Radical Oxidation Chemistry

Published as part of The Journal of Physical Chemistry virtual special issue "Marsha I. Lester Festschrift".

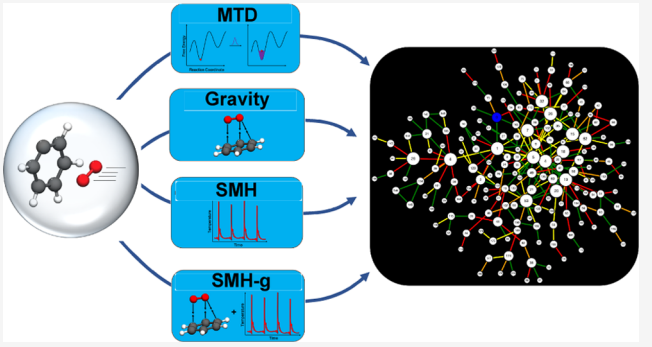
Alexander M. Chang, Jan Meisner, Rui Xu, and Todd J. Martínez*

Cite This: *J. Phys. Chem. A* 2023, 127, 9580−9589

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Over the years, many computational strategies have been employed to elucidate reaction networks. One of these methods is accelerated molecular dynamics, which can circumvent the expense required in dynamics to find all reactants and products (local minima) and transition states (first-order saddle points) on a potential energy surface (PES) by using fictitious forces that promote reaction events. The *ab initio* nanoreactor uses these accelerating forces to study large chemical reaction networks from first-principles quantum mechanics. In the initial nanoreactor studies, this acceleration was done through a piston periodic compression potential, which pushes molecules together to induce entropically unfavorable bimolecular reactions. However, the piston is not effective for discovering intramolecular and dissociative reactions, such as those integral to the decomposition channels of phenyl radical oxidation. In fact, the choice of accelerating forces dictates not only the rate of reaction discovery but also the types of reactions discovered; thus, it is critical to understand the biases and efficacies of these forces. In this study, we examine forces using metadynamics, attractive potentials, and local thermostats for accelerating reaction discovery. For each force, we construct a separate phenyl radical combustion reaction network using solely that force in discovery trajectories. We elucidate the enthalpic and entropic trends of each accelerating force and highlight their efficiency in reaction discovery. Comparing the nanoreactor-constructed reaction networks with literature renditions of the phenyl radical combustion PES shows that a combination of accelerating forces is best suited for reaction discovery.

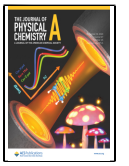


1. INTRODUCTION

The study of a complex chemical reaction system requires an understanding of its potential energy surface (PES), which governs the stability of all reactants, products, and transition states for elementary reactions. From this PES, reaction rates for chemical pathways can be obtained to understand which reaction pathways are important and how the system evolves over time under various initial conditions. However, complete exploration of a PES, even for smaller systems, is challenging due to its high dimensionality and complexity. The number of possible intermediates and elementary reactions grows combinatorially with system size, making accurate assessments of the relevant reaction rates for an entire chemical system difficult to attain.

To address this challenge, various approaches have been employed over the years to explore potential energy landscapes and unravel reaction networks. One approach is to generate reaction mechanisms through various heuristics and often reaction family templates.[1−6] Other methods involve exploiting Monte Carlo-like techniques,[7,8] deep neural networks,[9] PES curvature information,[10−12] and path-optimization methods.[13,14] Another technique uses molecular dynamics (MD) to explore the PES by sampling configuration space according to a chosen thermodynamic ensemble. Using MD for reaction discovery offers several advantages: independence from bias and heuristics generalizes its applicability, allowing the observation of complex reaction mechanisms while preferentially avoiding reactions with infeasibly high energy barriers. However, MD is more computationally expensive than other approaches, especially at the *ab initio* level, and may be unable to find important rare events swiftly. One workaround is accelerated molecular dynamics, which introduces a bias potential (force) that promotes enhanced sampling of the PES during an MD trajectory, hastening reaction discovery. There have been various implementations of accelerated MD techniques over the years, such as metadynamics,[15,16] hyper-

Received: August 14, 2023
Revised: October 2, 2023
Published: November 7, 2023



dynamics,[17,18] boxed MD,[19,20] and high-energy chemical dynamics simulations.[21] The recently developed *ab initio* nanoreactor[22] uses accelerating forces that hasten reaction discovery in MD in order to study large chemical reaction networks from first-principles and construct microkinetic models.[23,24] In the nanoreactor, the discovery phase is performed on a proxy PES (one with a cheaper level of theory) in order to expedite discovery of all reactions needed for a microkinetic model. Transition states and energy barriers are found in the refinement phase, where the reactions discovered on the proxy surface are refined at a higher level of theory and used in the ensuing microkinetic analyses. Because the nanoreactor separates its reaction discovery and path refinement stages, it can drive chemical systems far from equilibrium during MD to accelerate reaction discovery while still obtaining accurate energetic barriers and reaction rates in the later refinement phases. Thus, the process of calculating thermodynamic quantities, such as free energy, does not depend on the choice of accelerating forces.

Nevertheless, the inherent bias introduced by the choice of reaction-accelerating forces can greatly affect which reactions are discovered. Therefore, it is critical to not only understand these biases but also maximize the range of biases that can be employed. In the first use of the nanoreactor framework, reaction acceleration was performed using a piston compression potential.[22] This same piston force was later used to construct kinetic models for nitromethane decomposition[23] and methane pyrolysis.[24] In these instances, a virtual periodic piston force pushed molecules toward the center of the nanoreactor in order to induce collisions. This inherently promotes reaction discovery of associative reactions (A + B → C) that are often entropically unfavorable while (at least somewhat) suppressing the discovery of dissociative reactions (A → B + C). Other techniques, such as the root-mean-squared-deviation (RMSD)-based metadynamics[16] recently used by our nonadiabatic nanoreactor to study nonadiabatic decay mechanisms of benzene,[25] are geared toward conformational and intramolecular rearrangements. It is also essential to gauge the relative efficiencies of these discovery methods as well as their biases. Although both the Urey−Miller and nitromethane nanoreactor studies succeeded in discovering many species and reactions using the piston, this approach consumed a large amount of computational time. Indeed, the nitromethane studies required 5 ns of MD simulations using ReaxFF.[26] A quantitative understanding of the reaction discovery rates of various forces and their different biases will inform us how to better explore PESs.

To compare discovery methods and improve the nanoreactor's ability to find intramolecular and dissociative reactions, we chose phenyl radical oxidation as our model system. Peroxy radicals are ubiquitous in combustion and have thus been of much interest.[27−29] The pathways to condensed-phase soot formation involve gas-phase reactions between aromatic free radicals (e.g., phenyl C_6H_5, naphthalenyl $C_{10}H_7$) and unsaturated hydrocarbons (e.g., acetylene C_2H_2) that are often in competition with oxidation reactions. As the smallest aromatic free radical, the phenyl radical (C_6H_5) plays an integral role in the formation of soot, and thus, its reaction pathways should be fully elucidated. There have been several previous studies on phenyl radical combustion that detail its oxidation channels. A computational study performed by Mebel et al. described pathways and mechanisms involving nearly 40 molecules on the potential energy surface of C_6H_5 + O_2.[30] The initial addition forms a phenylperoxy radical intermediate that goes through various decomposition pathways to form smaller products such as cyclopentadienyl radical, carbon dioxide, carbon monoxide, and pyranyl radical. Kaiser et al. experimentally revealed these decomposition channels, which also proceed through the initial phenylperoxy intermediate to arrive at primary products such as acetylene, ketene, acrolein, and others.[31]

In this article, we introduce new tools for accelerating reaction discovery and showcase their efficacy in exploring the potential energy surface of phenyl radical combustion. We compare these tools and the previous piston in a systematic manner to examine their reaction discovery rates along with enthalpic and entropic trends. Furthermore, we compare the resulting reaction networks to the existing literature network of Mebel et al.[30]

2. METHODS

Creation of the phenyl radical oxidation reaction map requires iterative processes, and each process progresses through three stages: discovery, refinement, and network creation, as shown schematically in Figure 1. In the discovery phase, *ab initio* molecular dynamics (AIMD) with reaction-accelerating forces is performed on initial species to discover reactions and other molecules. In the refinement stage, the endpoints of reactions are minimized, followed by interpolation between the endpoints to generate a guess path for refinement, which aims to obtain a minimum energy path and transition state. These reactions and their calculated reaction barriers are combined to form a reaction network. In order to fully explore the PES, the next iterative process begins with the discovery phase starting in different regions of the PES. Thus, our progressing reaction network informs the subsequent iterations of the discovery phase. In the next three subsections, we will describe the three stages in detail.

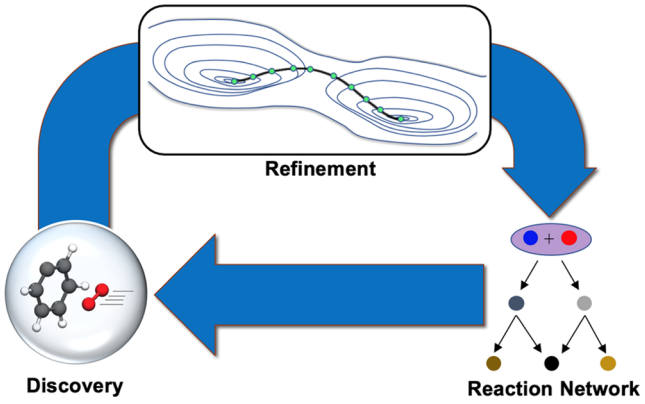


Figure 1. Stages for creating a reaction network. In the discovery phase, reactions are sampled by accelerated *ab initio* molecular dynamics. These reaction paths are then optimized toward their minimum energy paths in the refinement stage. Subsequently, a reaction network of the refined reactions is produced and used to inform the next discovery phase.

2.1. Discovery. All discovery runs begin with the molecules packed in a sphere of radius 5 Bohr using the Packmol program.[32] These molecules are initialized from user-provided SMILES strings and converted into three-dimensional (3D) structures through RDKit.[33] For all discovery simulations (except those with the piston, which requires longer time-

https://doi.org/10.1021/acs.jpca.3c05484
J. Phys. Chem. A 2023, 127, 9580−9589

scales), AIMD proceeds for 2 ps using velocity Verlet integration with a time step of 1 fs. The initial velocities of all atoms were selected from a 1000 K Boltzmann distribution. In order to quench the energy created by the accelerating forces, a Langevin thermostat was employed at 1000 K with a friction coefficient of 41 ps^{-1}.[34] All discovery runs were performed using the B3LYP[35] functional with a 3-21G basis set.[36] TeraChem[37−39] was used for all electronic structure calculations. All calculations were performed using unrestricted self-consistent field (SCF) methods. Level-shifting[40] was used to improve convergence for open-shell states. Reactions were detected in the AIMD trajectories through the changes in the connectivity graph over time using hidden Markov models.[22,41] A pair of atoms is defined as connected if their bond order is greater than 0.1 and the atoms are within 1.2 covalent radii of each other. Reaction events are demarcated as 100 fs windows of time centered around the time frame in which the bonding topology changes. Events with identical atomic indices that occur over overlapping time windows are treated as a single reaction event. Each different bonding topology is then given a state label, to which a hidden Markov model is fit by maximum likelihood. The Viterbi algorithm[42] is then applied to the Markov chain to compute the most probable sequence of states, filtering out transient changes in the bonding topology.[41]

Five accelerating techniques were employed during the AIMD discovery phase, including (1) a piston compression potential, (2) RMSD-based metadynamics (MTD), (3) single-molecule heating (SMH), (4) gravity, and (5) a combined single-molecule heating with gravitational force (SMH-g). A spherical boundary potential was used for all discovery trajectories and is detailed in Section **S6** in the SI. The five accelerating forces used in discovery trajectories are described below.

2.1.1. Piston. A time-dependent piston compression potential $V(r,t)$, which was used in the original nanoreactor study,[22] is applied on all atoms throughout the discovery simulation.

$$V(r, t) = f(t)U(r, r_1, k_1) + (1 - f(t))U(r, r_2, k_2) \quad (1)$$

The compression potential applied on an atom is dependent on the atom's distance from the origin, r, and oscillates over time t, dictated by the functions $U(r, r_2, k_2)$ and $f(t)$, respectively

$$U(r, r_0, k) = \frac{mk}{2}(r - r_0)^2\theta(r - r_0) \quad (2)$$

$$f(t) = \theta\left(\left\lfloor \frac{t}{T} \right\rfloor - \frac{t}{T} + \frac{\tau}{T}\right) \quad (3)$$

Here, θ is the Heaviside step function and $\lfloor\ \rfloor$ represents the floor function. Parameters similar to those used previously were employed, with the radii being reduced fivefold to adjust for the smaller system (k_1 =1.0 kcal mol^{-1} $Å^{-2}$ amu^{-1}, r_1 = 2.8 Å, k_2 = 0.5 kcal mol^{-1} $Å^{-2}$ amu^{-1}, r_2 = 1.6 Å, τ = 1.5 ps, T = 2.0 ps). The mass dependence of the potential function ensures that all atoms experience identical acceleration at the same distance from the origin.

2.1.2. RMSD-Based Metadynamics (MTD). An RMSD biasing potential E_{Bias} was applied during AIMD to accelerate the exploration of reaction space. This biasing potential, introduced by Grimme and co-workers,[16] is given by the sum of Gaussian functions dictated by a root-mean-square deviation (RMSD) metric as shown in eqs 4 and 5.

$$E_{Bias} = \sum_{i=1}^{N} k\ \exp(-\alpha\Delta_i^2) \quad (4)$$

Here, the biasing potential is a sum of N Gaussians where N is the maximum number of reference structures. Each Gaussian has the same strength parameter k and Gaussian width α. The collective variable Δ_i is defined as

$$\Delta_i = \sqrt{\frac{1}{N_{atoms}} \sum_{j=1}^{N_{atoms}} (r_j - R_j^i)^2} \quad (5)$$

where r_j is a component of the Cartesian space position vector of the actual molecule and R_j^i is the corresponding component in the ith reference structure with the same atom numbering as the actual molecule. The quaternion algorithm as described by Coutsias et al.[43] was employed for proper alignment of structures in order to minimize the RMSD between the two sets of vectors. A new reference structure was recorded every 50 fs. The maximum number of reference structures N used for calculation of the bias energy was set to 40. The strength parameter k was the same for each Gaussian and was set at 0.008 hartree/atom. Each Gaussian width α was set to 2.7726 $Bohr^{-2}$ to give a full width at half-maximum (FWHM) of 1 Bohr.

2.1.3. Single-Molecule Heating (SMH). This force raises the internal temperature of a randomly chosen molecule by scaling up its momenta for a duration of time τ = 50 fs and then scaling the momenta back down for a cooling period. The procedure then repeats with another randomly chosen molecule. In the discovery AIMD runs, the momenta were multiplied by a factor of 4.5 at a time t and then divided by a factor of 4.5 at time $t + \tau$, beginning a cooling period of 350 fs, after which the momentum upscaling occurs again.

2.1.4. Gravity. A time-independent pairwise attractive "gravitational" force, F_g, is applied to all atoms.

$$F_g = G \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j\neq i}}^{N} \frac{M_i M_j}{x_{ij}^2} S(x_{ij}) \quad (6)$$

$$S(x_{ij}) = \frac{1}{1 + \exp\left(\frac{c - x_{ij}}{w}\right)} \quad (7)$$

Here, G is a strength constant, for which 0.001 atomic units was used in our system (42 orders of magnitude larger than the true gravitational constant). The attractive force is also dependent on the atomic mass of a given atom i, M_i, and the pairwise distance between two atoms x_{ij}. The sigmoid function $S(x_{ij})$ rescales the forces so that the atoms do not get too close to each other. The center c and width w for this sigmoid function are 5 and 4 au, respectively.

2.1.5. Single-Molecule Heating with Gravity (SMH-g). Single-molecule heating and gravity forces were also used simultaneously. The same parameters as those in Sections 2.1.3 and 2.1.4 were used.

2.2. Refinement. The discovered reaction paths are then refined at the same level of theory (B3LYP/3-21G) using a reaction path and transition state optimization tool called pyGSM, a Python implementation of the growing string method (GSM).[44−46] Usually, the refinement phase occurs at a

https://doi.org/10.1021/acs.jpca.3c05484
J. Phys. Chem. A 2023, 127, 9580−9589

higher level of theory than the discovery phase. However, since our aim here is to study the efficiency of the accelerating forces used in discovery, we use the same level of theory for discovery and refinement. First, the endpoints of reaction paths are optimized to find the nearest local minimum. The bonding information from these refined reactant and product geometries is used to determine the SMILES strings for the endpoints. Geodesic interpolation[47] is used to provide an approximate minimum energy path (MEP) used as an initial guess that is then further refined by pyGSM. Reactions consisting of more than one elementary step are split into separate reactions that are then further refined. This path-splitting procedure is detailed in **Section S7** of the SI.

Spin values used for calculating the multiplicity of individual molecules were obtained by summing the Mulliken spin densities[48] of a given molecule fragment and rounding the total to the nearest half-integer. Because of the use of unrestricted DFT methods, the spin-contamination of all relaxed species was assessed. This was quantified by finding the difference between the contaminated spin value (s) calculated by the unrestricted DFT method and the exact value s assuming a pure spin state for each molecule. A histogram of this spin-contamination difference is shown in **Figure S1** in the Supporting Information (SI).

2.3. Reaction Network Generation. Each of the forces described in Section 2.1 was individually used to progress through the phenyl oxidation reaction network. The initialized molecules for each discovery trajectory were chosen with a breadth-first reaction network bootstrapping method (Figure 2) through the reaction graph. The first initial reactants

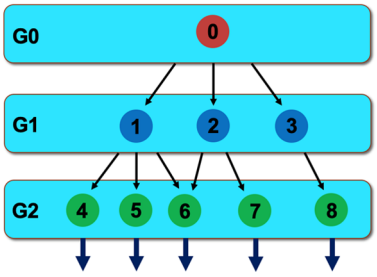


Figure 2. Hypothetical reaction network, where G*i* stands for the *i*th generation of the discovery phase and numbers represent an individual species. The network starts from Node **0** as the first initial species. Ten discovery simulations are run on Node **0** to produce Nodes **1**, **2**, and **3**. Then, each product in G1 (**1**, **2**, **3**) is used as a sole initial species for a set of ten discovery simulations to produce **4**, **5**, **6**, **7**, and **8** in G2, and so on. In this example, a total of 90 discovery simulations would be run (ten discovery simulations for nine different initial species).

(labeled as Generation 0, or G0, in Figure 2) were one phenyl radical molecule (C_6H_5) and one oxygen molecule (O_2). A 2 ps discovery simulation was carried out ten times with these initial molecules for forces described in Sections 2.1.2−2.1.5 (the piston required longer timescales for reaction discovery; thus, we used a single 20 ps trajectory for this force). Following reaction detection and path refinement, each of the direct products of phenyl radical and oxygen molecule (G1 in Figure 2) were separately used as an initial species for a set of ten discovery runs. After these reactions were refined, the process was repeated one last time, with the direct products in G2 each serving as an initial molecule(s) for ten discovery runs. Initial starting molecules were used only if they could be discovered from phenyl and O_2 without having to overcome a reaction barrier greater than 100 kcal/mol. Due to the bootstrapping method, the reaction networks varied in their sets of initial species and number of initial species. These initial species are detailed in **Section S2** of the SI.

Although multiple starting molecules can be used as initial species in the same trajectory, our study constrained the initial species choice to a single molecule (outside of the initial phenyl radical and O_2) in order to examine decomposition channels. Additionally, since it was intended for discovering intramolecular and decomposition reactions, SMH struggled with discovering the initial formation of the phenylperoxy radical; therefore, we used phenylperoxy radical as its starting initial species, skipping the initial round of discovery using phenyl radical and O_2. Nevertheless, it was able to discover the initial formation reaction in the backward direction.

After refinement, a reaction network for phenyl radical oxidation was generated with all of the discovered species and their refined reactions represented as nodes and edges, respectively. Separate reaction networks were formed for each of the five reaction-accelerating forces and are included in **Section S5** of the SI.

3. RESULTS AND DISCUSSION

A total of 188 unique species and 233 unique reactions were discovered among all five reaction networks generated. We defined a "unique" species/reaction by counting a species/reaction discovered multiple times as one species/reaction (whereas "total" species/reactions incorporate the double-counting of multiple discoveries). A Venn diagram detailing how many unique reactions were discovered by each reaction force (excluding the piston, which only discovered one reaction, the initial phenyl radical oxidation addition) or sets of reaction forces is shown in Figure 3. It is clear that SMH-g discovered the most reactions and gravity the least. A more intriguing observation is the lack of overlap in reactions discovered between the four forces. Only 27 of the 233 reactions were discovered by multiple forces, and only three reactions were seen in all four reaction networks. The limiting overlap between forces was the intersection of SMH and

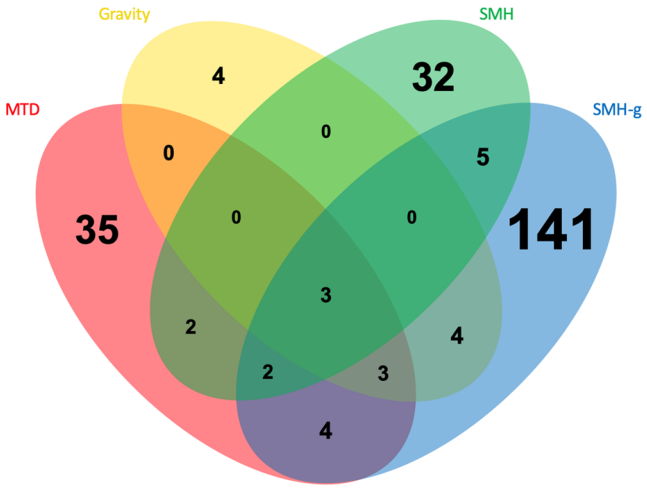


Figure 3. Venn diagram displaying overlaps of discovered unique reactions in the four accelerating force reaction networks. Only three reactions were found in all networks.

https://doi.org/10.1021/acs.jpca.3c05484
J. Phys. Chem. A 2023, 127, 9580−9589

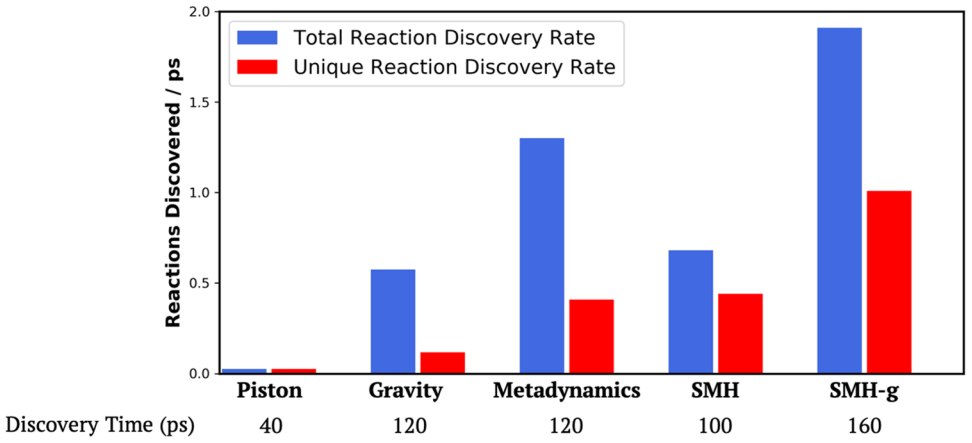


Figure 4. Efficiency of accelerating forces represented by the reaction discovery rate. The discovery rate is calculated using the number of total reactions and unique reactions in each force's reaction network. As the discovery time varies for each force due to the bootstrapping procedure, the number of reactions is divided by the total number of picoseconds of discovery trajectories used by each force. Total reactions account for the same unique reactions being discovered multiple times.

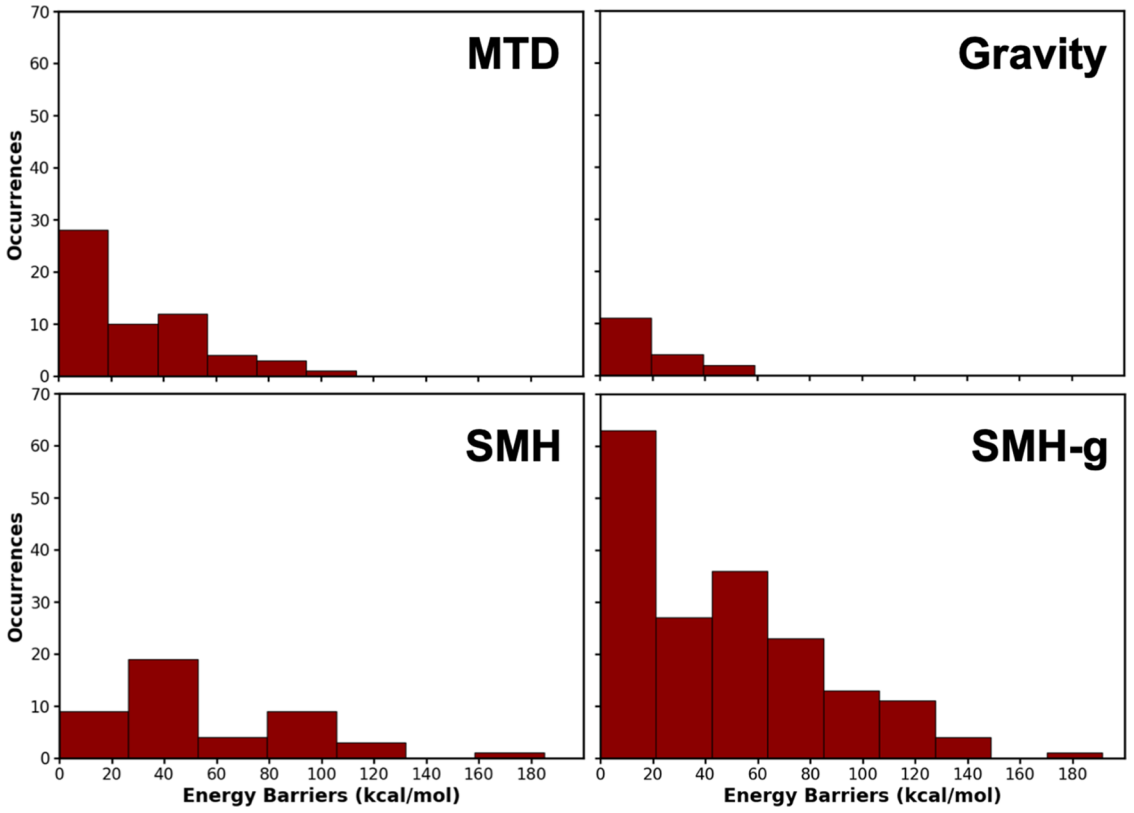


Figure 5. Histograms of reaction activation energy barriers of *unique* reactions for each force's reaction network, including (a) MTD (upper left), (b) gravity (upper right), (c) SMH (lower left), and (d) SMH-g (lower right). Reactions shown are discovered in the forward direction (*reactants* → *transition state* → *products*) during a discovery trajectory. The histogram bin size was determined using the Freedman−Diaconis rule.[51]

gravity, which only had the three aforementioned reactions in common. As SMH is a force intended to drive dissociation through local heating and gravity is a force designed for pulling atoms together, it may not be surprising that the resulting reaction networks have minimal overlap.

Since some forces had more discovery time than others because of the varying number of initial species resulting from the bootstrapping method (Figure 2), the rates of reaction discovery displayed in Figure 4 serve as a better representation of discovery efficiency. SMH-g was the most efficient in discovering reactions, more than doubling the unique reaction discovery rates of its closest competitors. Excluding the piston, which struggles to find the intramolecular and dissociative reactions integral to a unimolecular system, gravity had the lowest rate of discovery of unique reactions; however, as displayed in Figure 5, the gravity force was the best at discovering reactions solely within a lower-energy barrier threshold. All forces seem to follow a general decay in reaction occurrences as energy barriers increase after the ∼40−60 kcal/mol range. Most of the reactions found by gravity and MTD (RMSD metadynamics) have energy barriers of less than 60 kcal/mol (and for gravity, all of the discovered reactions have

https://doi.org/10.1021/acs.jpca.3c05484
J. Phys. Chem. A 2023, 127, 9580−9589

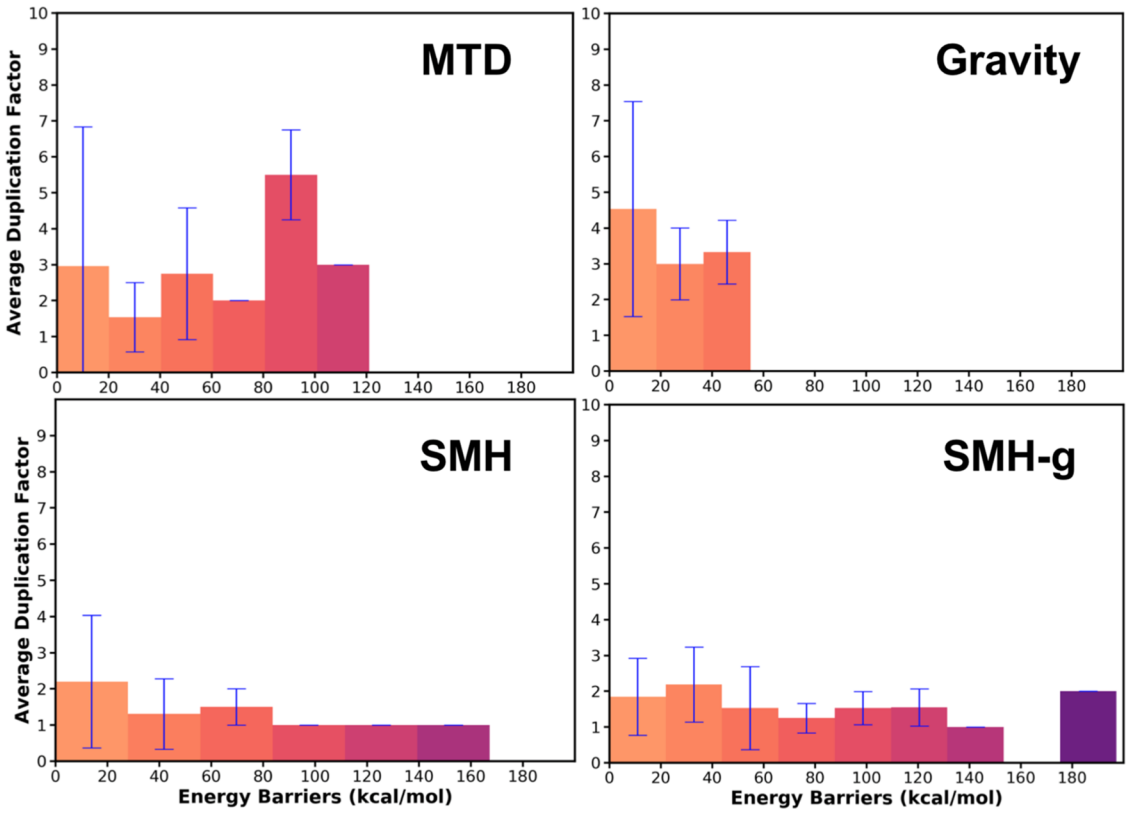


Figure 6. Average duplication factor histogram of unique reactions for the network generated by accelerating forces, including (a) MTD (upper left), (b) gravity (upper right), (c) SMH (lower left), and (d) SMH-g (lower right). The average duplication factor is calculated as the average number of times each unique reaction in a given bin was discovered per 100 ps of discovery time. Error bars represent one standard deviation away from the mean. The error bar for the leftmost bin in the MTD histogram goes below zero, showing the high variance in the duplication factor for that bin. Similar to Figure 5, the bin size was determined using the Freedman−Diaconis rule.

energy barriers of less than 46 kcal/mol). This suggests that while gravity does accelerate chemical reactivity, it may not be able to find rare events without increasing (even further) the gravitational strength parameter. This conclusion could change with longer discovery trajectories, especially for RMSD metadynamics. The longer an AIMD trajectory with metadynamics proceeds, the more Gaussians fill up the free energy well, increasing the frequency of observing high-energy barrier reactions, which would shift the distribution seen in Figure 5a. SMH and SMH-g, on the other hand, found a greater proportion of higher energy barrier reactions. Over a quarter of unique discovered reactions in the SMH network had reaction barriers exceeding 80 kcal/mol, which underscores both its ability to find rare events but also its potential inefficiency by finding many reactions that may be highly improbable. SMH-g found a higher proportion of lower-energy barrier reactions than SMH, again highlighting gravity's tendency to discover reactions with lower enthalpic barriers even when in combination with another force.

The disparity between total and unique reaction discovery rates should also be noted. We can observe in Figure 4 that while MTD and SMH discovered similar numbers of unique reactions, MTD discovered many more total reactions, implying that MTD discovered the same reactions more often than SMH did. We see that MTD and gravity both have higher duplication factors in Figure 6, indicating that they found the same reactions multiple times. The wide error bars for the MTD and gravity duplication factors reflect that most of the lower-energy barrier reactions were found only once, but a few were found many times. This may seem counterintuitive for MTD, where one would expect the Gaussian biasing potentials to steer discovery away from finding the same reaction repeatedly; however, these biasing potentials are not currently preserved between discovery trajectories, allowing MTD to find the same reaction in many separate trajectories. SMH and SMH-g produce duplication rates lower than that of MTD and gravity. Although SMH finds less than half of the total reactions that MTD found, it counteracts this through a lack of redundancy in discovery, helping it explore the PES more effectively.

On the given simulation timescales, the piston is unable to discover reactions past the formation of the unimolecular phenylperoxy radical because it does not easily induce intramolecular and dissociative reactions. In order to understand the capabilities of the other forces to find these intramolecular and dissociative reactions, entropies were computed for each reactant and product species using the rigid rotor/harmonic approximation (more detailed procedures can be found in Section S3).[49] These trends are compared in Figure 7, where the changes in reaction entropy for all discovered reactions in each force's reaction network is depicted. We find that approximately 50% of unique reactions in SMH were entropically favorable reactions ($\Delta\Delta S > 5$ kcal/mol), a consequence of local heating inducing fragmentations. On the other hand, SMH was unable to discover important entropically unfavorable reactions such as the initial oxidative addition to the phenyl radical in the forward direction. It can be hypothesized that before enough time passes for the

https://doi.org/10.1021/acs.jpca.3c05484
J. Phys. Chem. A 2023, 127, 9580−9589

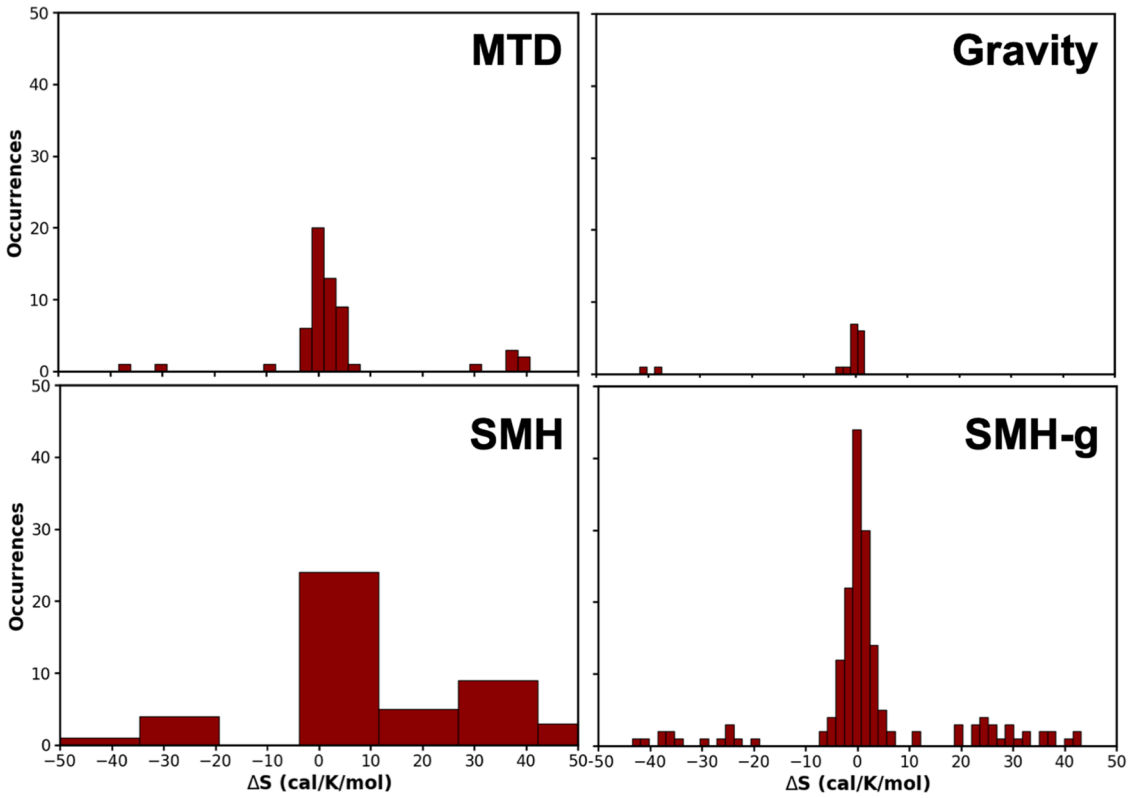


Figure 7. Reaction entropy change for each unique reaction in the network generated by accelerating forces, including (a) MTD (upper left), (b) gravity (upper right), (c) SMH (lower left), and (d) SMH-g (lower right). The procedure for obtaining entropies is detailed in **Section S3** of the SI. Reactions shown are discovered in the forward direction (*reactants* → *products*) during a discovery trajectory. The histogram bin size was determined using the Freedman−Diaconis rule.[49]

entropic barrier of the addition reaction to be overcome, local heating occurs and induces a different reaction. This underscores a limitation in SMH in systems with multiple molecules: the frequency of local heating may preclude discovery of some associative reaction pathways. However, combining gravity with SMH allows the discovery of this initial oxidative addition as gravity allows the initial addition reaction to be discovered before local heating occurs. We also observe that under 20% of unique reactions in SMH-g were entropically favorable, significantly less than the 50% for SMH on its own. The addition of gravity, an attractive force keeping the molecule together, prevents the molecule from completely fragmenting and instead promotes isomerization.

Taking a superset of our reaction networks, we see in Figure 8 that we discovered all but eight species from the Mebel literature network.[30] Only three reactions were found by all four nanoreactor accelerating forces: **0** → **1**, **1** → **5**, and **5** → **19**. While all three species in the latter two reactions are found in the literature, the two elementary reactions themselves are not. Observing the graph, one can see that SMH-g was able to discover the majority of the displayed reactions, with multiple pathways being discovered solely by SMH-g. Its ability to uncover much of the existing literature on potential energy surface suggests that SMH-g can find a plethora of reactions including many of the most probable ones. However, SMH-g is not all-encompassing. There are tens of unique reactions discovered by other forces that SMH-g did not find, including reaction **5** → **154** and the decomposition pathway **6** → **146** shown in Figure 8. Perhaps SMH-g would find these reactions given more discovery time, but reaction **5** → **154** was discovered ten times in the gravity reaction network, and **6** → **134** was discovered three times in the MTD reaction network. This indicates that other forces may be able to discover certain important reactions at a frequency higher than SMH-g. Thus, for complete exploration of a potential energy surface, the use of all forces is recommended.

A resulting question from this statement is whether to use all of the forces in a trajectory *simultaneously*, in separate trajectories, or perhaps a combination of both options. Using Figure 3, we can compare the SMH-g reaction network to the union of the SMH and gravity reaction networks. We find that the intersection of SMH-g with the union of SMH and gravity consists of only 17 reactions and that there are nearly 40 reactions in the union of the SMH and gravity networks that are not observed in SMH-g and more than 150 reactions vice versa. Thus, we see that there is value in the simultaneous application of SMH and gravity, as well as in their independent application, and that both options should be used for reaction discovery.

In order to make quantitative statements about the importance of the discovered reactions in a reaction network, one needs to perform microkinetic analyses (at a higher level of theory than the proxy discovery PES) along with further continuation of the bootstrapping method (as detailed in Figure 2), a task that we leave for future work. Nevertheless, Figure 8 demonstrates that even in limited discovery time, decomposition channels seen in the literature such as those for cyclopentadienyl and pyranyl radicals can be discovered. Absent from Figure 8 and the 2005 Mebel reaction network[30] is the decomposition pathway to cyclopentadienone and HCO,

https://doi.org/10.1021/acs.jpca.3c05484
J. Phys. Chem. A 2023, 127, 9580−9589

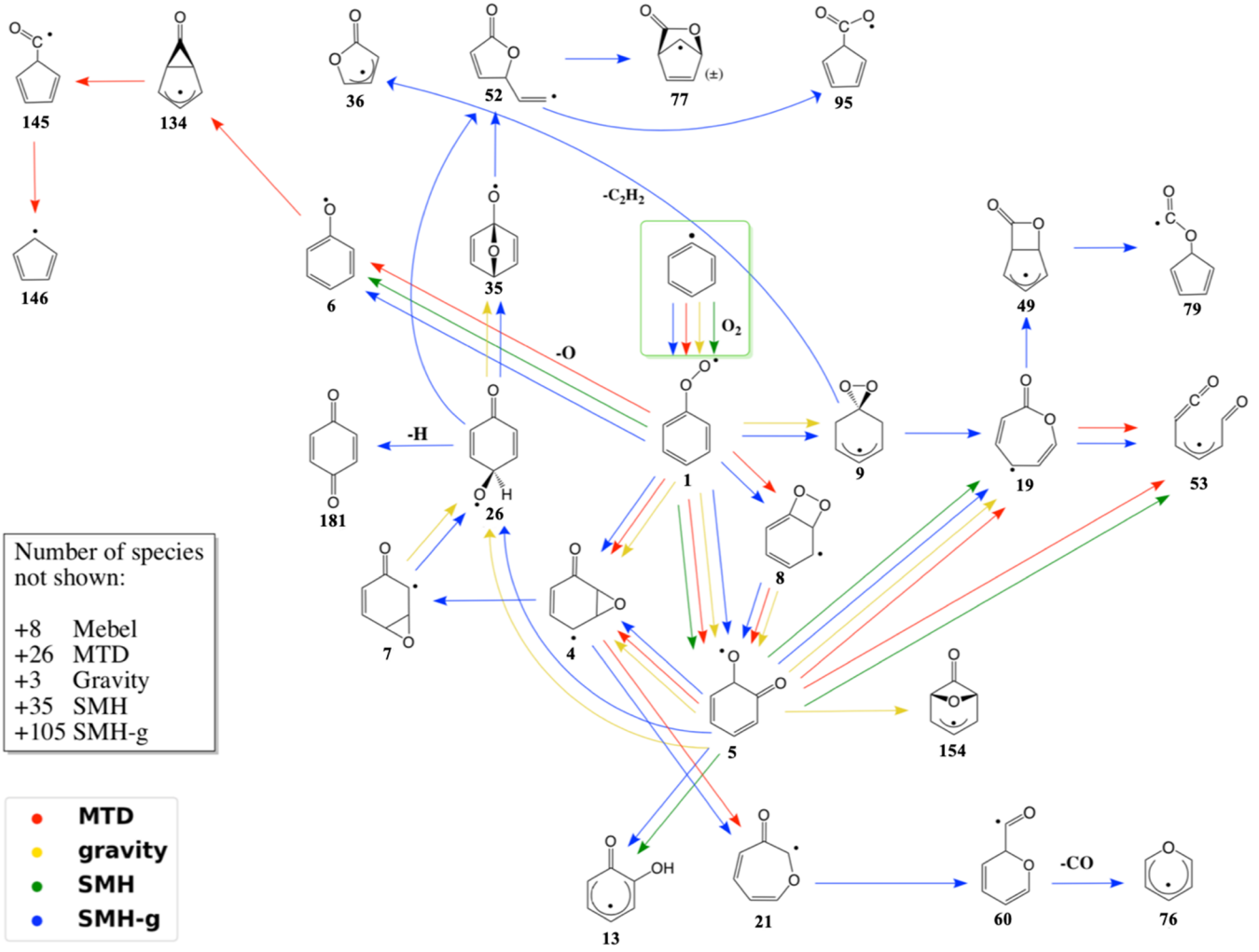


Figure 8. Superset of the reactions and species found by the nanoreactor and the literature reactions and species shown in Mebel et al.[30] Red, yellow, green, and blue arrows signify that the reaction was found by MTD, gravity, SMH, and SMH-g, respectively. The green box signifies the starting phenyl radical oxidation reaction. Only reactions between species found in the literature reaction network are shown.

which was later found to be the predominant channel at lower temperatures both experimentally[31] and theoretically.[50] While cyclopentadienone is not observed in our reaction networks, its direct precursor C_5H_4OHCO (molecule **20**, see Section S1) is discovered in the SMH-g network. Further continuation of the bootstrapping method would help to elucidate this decomposition channel as well as many others.

4. CONCLUSIONS

In this study, we examined the reaction discovery rates and trends of various reaction-accelerating forces in the *ab initio* nanoreactor. All forces studied proved to be superior to the previous nanoreactor piston force for finding decomposition pathways in the phenyl radical combustion system. Among all of the forces, single-molecule heating in combination with gravity (SMH-g) has the highest reaction discovery efficiency and can discover both rare events and lower-energy barrier reactions. RMSD-based metadynamics and gravity are not as efficient at discovering reactions but do tend to find reactions with lower energy barriers (and thus likely more important to the reaction network). Single-molecule heating on its own primarily finds entropically favorable reactions; however, in combination with gravity, the proportion of entropically favorable discoveries is considerably reduced. Furthermore, some forces such as metadynamics find a large number of total reactions compared to unique reactions, suggesting a propensity for finding the same reactions repeatedly, while forces such as single-molecule heating rarely find the same reaction more than once.

When compared to a literature reaction network, we see that the majority of reactions and species in the literature were found by the nanoreactor. The discovery of these literature reactions was not dominated by one particular force, highlighting the importance of using all forces at one's disposal. This is further illustrated by the lack of overlap in the Venn diagram of sets of reactions obtained by each accelerating force. In future work, we aim to address several points related to comparison with the literature reaction network. There were eight species in the literature network that were not seen in any of our reaction networks. This raises the question of whether one simply needs more or longer discovery trajectories or whether, instead, one needs to include a different accelerating force. Since our forces are so far unable to find these reactions, this highlights a need to develop different accelerating techniques that can. Another future direction we aim to take is the microkinetic modeling of our nanoreactor reaction networks, as well as the literature reaction network. A comparison between the nanoreactor kinetic

https://doi.org/10.1021/acs.jpca.3c05484
J. Phys. Chem. A 2023, 127, 9580−9589

models and the literature kinetic model will help establish whether we have discovered all important reactions in the reaction network.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpca.3c05484.

2-D structures and SMILES strings of all discovered molecules; list of initial molecules used in discovery trajectories; procedure for thermodynamic calculations; table of raw data of energies, enthalpies, entropies, and Gibbs free energies; Cytoscape diagrams of all reaction networks; spherical boundary conditions (PDF)

Procedure for reaction splitting in path refinement; and example input files used for each phase of nanoreactor (ZIP)

Special Issue Paper

Published as part of *The Journal of Physical Chemistry* virtual special issue "Marsha I. Lester Festschrift".

■ AUTHOR INFORMATION

Corresponding Author

Todd J. Martínez − *Department of Chemistry and The PULSE Institute, Stanford University, Stanford, California 94305, United States; SLAC National Accelerator Laboratory, Menlo Park, California 94025, United States;* orcid.org/0000-0002-4798-8947; Email: toddjmartinez@gmail.com

Authors

Alexander M. Chang − *Department of Chemistry and The PULSE Institute, Stanford University, Stanford, California 94305, United States; SLAC National Accelerator Laboratory, Menlo Park, California 94025, United States;* orcid.org/0000-0003-1347-7169

Jan Meisner − *Department of Chemistry, Heinrich-Heine-Universität Düsseldorf, Düsseldorf 40225, Germany;* orcid.org/0000-0002-1301-2612

Rui Xu − *Department of Chemistry and The PULSE Institute, Stanford University, Stanford, California 94305, United States; SLAC National Accelerator Laboratory, Menlo Park, California 94025, United States;* orcid.org/0000-0003-3264-9050

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpca.3c05484

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research (N00014-21-1-2151 and N00014-18-1-2659). AMC acknowledges support from the National Science Foundation Graduate Research Fellowship and the Stanford Graduate Fellowship.

■ REFERENCES

(1) Gao, C. W.; Allen, J. W.; Green, W. H.; West, R. H. Reaction Mechanism Generator: Automatic construction of chemical kinetic mechanisms. *Comput. Phys. Commun.* **2016**, *203*, 212−225.

(2) Susnow, R. G.; Dean, A. M.; Green, W. H.; Peczak, P.; Broadbelt, L. J. Rate-based construction of kinetic models for complex systems. *J. Phys. Chem. A* **1997**, *101*, 3731−3740.

(3) Rappoport, D.; Galvin, C. J.; Zubarev, D. Y.; Aspuru-Guzik, A. Complex Chemical Reaction Networks from Heuristics-Aided Quantum Chemistry. *J. Chem. Theory Comput.* **2014**, *10*, 897−907.

(4) Broadbelt, L. J.; Stark, S. M.; Klein, M. T. Computer Generated Pyrolysis Modeling: On-the-Fly Generation of Species, Reactions, and Rates. *Ind. Eng. Chem. Res.* **1994**, *33*, 790−799.

(5) Bergeler, M.; Simm, G. N.; Proppe, J.; Reiher, M. Heuristics-Guided Exploration of Reaction Mechanisms. *J. Chem. Theory Comput.* **2015**, *11*, 5712−5722.

(6) Liu, M.; Grinberg Dana, A.; Johnson, M. S.; Goldman, M. J.; Jocher, A.; Payne, A. M.; Grambow, C. A.; Han, K.; Yee, N. W.; Mazeau, E. J.; et al. Reaction Mechanism Generator v3.0: Advances in Automatic Mechanism Generation. *J. Chem. Inf. Model.* **2021**, *61*, 2686−2696.

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

(8) Habershon, S. Sampling reactive pathways with random walks in chemical space: Applications to molecular dissociation and catalysis. *J. Chem. Phys.* **2015**, *143*, No. 094106.

(9) Segler, M. H. S.; Preuss, M.; Waller, M. P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **2018**, *555*, 604−610.

(10) 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.

(11) Maeda, S.; Harabuchi, Y.; Takagi, M.; Saita, K.; Suzuki, K.; Ichino, T.; Sumiya, Y.; Sugiyama, K.; Ono, Y. Implementation and performance of the artificial force induced reaction method in the GRRM17 program. *J. Comput. Chem.* **2018**, *39*, 233−251.

(12) Maeda, S.; Harabuchi, Y.; Hayashi, H.; Mita, T. Toward Ab Initio Reaction Discovery Using the Artificial Force Induced Reaction Method. *Annu. Rev. Phys. Chem.* **2023**, *74*, 287−311.

(13) Zimmerman, P. M. Automated discovery of chemically reasonable elementary reaction steps. *J. Comput. Chem.* **2013**, *34*, 1385−1392.

(14) Dewyer, A. L.; Zimmerman, P. M. Finding reaction mechanisms, intuitive or otherwise. *Org. Biomol. Chem.* **2017**, *15*, 501−504.

(15) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.

(16) Grimme, S. Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. *J. Chem. Theory Comput.* **2019**, *15*, 2847−2862.

(17) Voter, A. F. Hyperdynamics: Accelerated Molecular Dynamics of Infrequent Events. *Phys. Rev. Lett.* **1997**, *78*, 3908−3911.

(18) Hirai, H. Practical hyperdynamics method for systems with large changes in potential energy. *J. Chem. Phys.* **2014**, *141*, No. 234109.

(19) Glowacki, D. R.; Paci, E.; Shalashilin, D. V. Boxed Molecular Dynamics: A Simple and General Technique for Accelerating Rare Event Kinetics and Mapping Free Energy in Large Molecular Systems. *J. Phys. Chem. B* **2009**, *113*, 16603−16611.

(20) Shannon, R. J.; Amabilino, S.; O'Connor, M.; Shalishilin, D. V.; Glowacki, D. R. Adaptively Accelerating Reactive Molecular Dynamics Using Boxed Molecular Dynamics in Energy Space. *J. Chem. Theory Comput.* **2018**, *14*, 4541−4552.

(21) Martínez-Núñez, E. An automated method to find transition states using chemical dynamics simulations. *J. Comput. Chem.* **2015**, *36*, 222−234.

(22) Wang, L.-P.; Titov, A.; McGibbon, R.; Liu, F.; Pande, V. S.; Martínez, T. J. Discovering chemistry with an ab initio nanoreactor. *Nat. Chem.* **2014**, *6*, 1044−1048.

(23) Ford, J.; Seritan, S.; Zhu, X.; Sakano, M. N.; Islam, M. M.; Strachan, A.; Martínez, T. J. Nitromethane Decomposition via Automated Reaction Discovery and an Ab Initio Corrected Kinetic Model. *J. Phys. Chem. A* **2021**, *125*, 1447−1460.
(24) Xu, R.; Meisner, J.; Chang, A. M.; Thompson, K. C.; Martínez, T. J. First Principles Reaction Discovery: From the Schrodinger Equation to Experimental Prediction for Methane Pyrolysis. *Chem. Sci.* **2023**, *14*, 7447−7464.
(25) Pieri, E.; Lahana, D.; Chang, A. M.; Aldaz, C. R.; Thompson, K. C.; Martínez, T. J. The non-adiabatic nanoreactor: towards the automated discovery of photochemistry. *Chem. Sci.* **2021**, *12*, 7294−7307.
(26) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. ReaxFF: A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A* **2001**, *105*, 9396−9409.
(27) Hansen, A. S.; Bhagde, T.; Qian, Y.; Cavazos, A.; Huchmala, R. M.; Boyer, M. A.; Gavin-Hanner, C. F.; Klippenstein, S. J.; McCoy, A. B.; Lester, M. I. Infrared spectroscopic signature of a hydroperoxyalkyl radical (QOOH). *J. Chem. Phys.* **2022**, *156*, No. 014301.
(28) Bhagde, T.; Hansen, A. S.; Chen, S.; Walsh, P. J.; Klippenstein, S. J.; Lester, M. I. Energy-resolved and time-dependent unimolecular dissociation of hydroperoxyalkyl radicals (QOOH). *Faraday Discuss.* **2022**, *238*, 575−588.
(29) Hansen, A. S.; Bhagde, T.; Moore, K. B.; Moberg, D. R.; Jasper, A. W.; Georgievskii, Y.; Vansco, M. F.; Klippenstein, S. J.; Lester, M. I. Watching a hydroperoxyalkyl radical (QOOH) dissociate. *Science* **2021**, *373*, 679−682.
(30) Tokmakov, I. V.; Kim, G.-S.; Kislov, V. V.; Mebel, A. M.; Lin, M. C. The Reaction of Phenyl Radical with Molecular Oxygen: A G2M Study of the Potential Energy Surface. *J. Phys. Chem. A* **2005**, *109*, 6114−6127.
(31) Parker, D. S. N.; Kaiser, R. I.; Troy, T. P.; Kostko, O.; Ahmed, M.; Mebel, A. M. Toward the Oxidation of the Phenyl Radical and Prevention of PAH Formation in Combustion Systems. *J. Phys. Chem. A* **2015**, *119*, 7145−7154.
(32) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157−2164.
(33) Landrum, G. *RDKit: Open-Source Cheminformatics Software*, 2016.
(34) Bussi, G.; Parrinello, M. Accurate sampling using Langevin dynamics. *Phys. Rev. E* **2007**, *75*, No. 056707.
(35) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(36) Binkley, J. S.; Pople, J. A.; Hehre, W. J. Self-consistent molecular orbital methods. 21. Small split-valence basis sets for first-row elements. *J. Am. Chem. Soc.* **1980**, *102*, 939−947.
(37) Seritan, S.; Bannwarth, C.; Fales, B. S.; Hohenstein, E. G.; Isborn, C. M.; Kokkila-Schumacher, S. I.; Li, X.; Liu, F.; Luehr, N.; Snyder, J. W., Jr; et al. TeraChem: A graphical processing unit-accelerated electronic structure package for large-scale ab initio molecular dynamics. *WIRES: Comp. Mol. Sci.* **2021**, *11*, No. e1494.
(38) Ufimtsev, I. S.; Martínez, T. J. Quantum chemistry on graphical processing units. 1. Strategies for two-electron integral evaluation. *J. Chem. Theory Comput.* **2008**, *4*, 222−231.
(39) Ufimtsev, I. S.; Martínez, T. J. Quantum chemistry on graphical processing units. 2. Direct self-consistent-field implementation. *J. Chem. Theory Comput.* **2009**, *5*, 1004−1015.
(40) Saunders, V. R.; Hillier, I. A "Level−Shifting" method for converging closed shell Hartree−Fock wave functions. *Int. J. Quantum Chem.* **1973**, *7*, 699−705.
(41) Wang, L.-P.; McGibbon, R. T.; Pande, V. S.; Martinez, T. J. Automated Discovery and Refinement of Reactive Molecular Dynamics Pathways. *J. Chem. Theory Comput.* **2016**, *12*, 638−649.
(42) Viterbi, A. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. Info. Theor.* **1967**, *13*, 260−269.
(43) Coutsias, E. A.; Seok, C.; Dill, K. A. Using quaternions to calculate RMSD. *J. Comput. Chem.* **2004**, *25*, 1849−1857.
(44) Zimmerman, P. Reliable Transition State Searches Integrated with the Growing String Method. *J. Chem. Theory Comput.* **2013**, *9*, 3043−3050.
(45) Zimmerman, P. M. Growing string method with interpolation and optimization in internal coordinates: Method and examples. *J. Chem. Phys.* **2013**, *138*, No. 184102.
(46) Zimmerman, P. M. Single-ended transition state finding with the growing string method. *J. Comput. Chem.* **2015**, *36*, 601−611.
(47) Zhu, X.; Thompson, K. C.; Martínez, T. J. Geodesic interpolation for reaction pathways. *J. Chem. Phys.* **2019**, *150*, No. 164103.
(48) Mulliken, R. S. Electronic population analysis on LCAO−MO molecular wave functions. I. *J. Chem. Phys.* **1955**, *23*, 1833−1840.
(49) Mcquarrie, D. A. *Statistical Mechanics*; Harper & Row: New York, 1976; pp 81−138.
(50) Morozov, A. N.; Medvedkov, I. A.; Azyazov, V. N.; Mebel, A. M. Theoretical Study of the Phenoxy Radical Recombination with the $O(^3P)$ Atom, Phenyl plus Molecular Oxygen Revisited. *J. Phys. Chem. A* **2021**, *125*, 3965−3977.
(51) Freedman, D.; Diaconis, P. On the histogram as a density estimator: L_2 theory. *Z. Wahrschein. und verwan. Gebiete* **1981**, *57*, 453−476.

https://doi.org/10.1021/acs.jpca.3c05484
J. Phys. Chem. A 2023, 127, 9580−9589