Automatic potential energy surface exploration by accelerated reactive MD simulations: from pyrolysis to oxidation chemistry

Wassja A. Kopp, Can Huang, Yuqing Zhao, Peiyang Yu, Felix Schmalz, Lukas Krep, and Kai Leonhard*

Institute of Technical Thermodynamics, RWTH Aachen University, 52062 Aachen, Germany

E-mail: kai.leonhard@ltt.rwth-aachen.de

Phone: +49 (0)241 80 98174. Fax: +49 (0)241 80 92255

Abstract

Automatic potential energy surface (PES) exploration is important to better understand reaction mechanisms. Existing automatic PES mapping tools usually rely on predefined knowledge or on computationally expensive on-the-fly quantum-chemical calculations. In this work, we have developed a method for discovering novel reaction pathways and automatically mapping out the PES when only being given a starting species. We have therefore extended the reactive molecular dynamics simulation tool ChemTraYzer2.0 (Chemical Trajectory Analyzer, CTY) for this PES-mapping algorithm. To explore PESs with low-temperature reactions, we applied the acceleration method Collective Variable-driven Hyperdynamics (CVHD). This involved the development of tailored collective variable (CV) templates, which are discussed in this study. The PES mapping algorithm can generate new seed species, automatically start replica simulations for new pathways, and stop the simulation when a reaction has been found, reducing the computational cost of the algorithm. It is validated for known pathways in various pyrolysis and oxidation systems: hydrocarbon isomerization and dissociation (C_4H_7 and C_8H_7 PES), mostly dominant at high temperatures, low-temperature oxidation of n-butane ($C_4H_9O_2$ PES) and cyclohexane ($C_6H_{11}O_2$ PES). As a result, in addition to new pathways showing up in the simulations, common isomerization and dissociation pathways were found very fast: for example, 44 reactions of butenyl radicals including major isomerizations and decompositions within about 30 minutes wall time and low-temperature chemistry such as the internal H-shift of $RO_2 \longrightarrow QO_2H$ within one day wall time. Furthermore, since CTY keeps track of all found geometries of the reactants, transition states, and products, PES mapping facilitates automated higher-level calculations without a priori knowledge. Lastly, we applied PES mapping to the oxidation of the recently proposed bio-hybrid fuel 1,3-dioxane and validated that the tool could be used to discover new reaction pathways of larger molecules that are of practical use.

Introduction

Kinetic modeling of fuel pyrolysis and combustion is crucial for applications like design of new biohybrid fuel or optimization of advanced engines.[1] To study reactions that are essential in combustion mechanisms, such as fuel pyrolysis and low-temperature oxidation (LTO), automated potential energy surface (PES) searching is a building block for further determination of rate constants through quantum chemistry (QC). Hence, a multitude of reaction network discovery tools has been developed to date,[2] each of which come with strengths and weaknesses and a limited success rate in discovering a full reaction network in a given computational time. The following provides a non-exhaustive list of popular tools, which rely more or less on a priori knowledge of reaction classes, templates, and heuristics. The well-known tool Reaction Mechanism Generator (RMG)[3] is based on encoded elementary steps from databases or heuristics and can explain novel phenomena via non-intuitive reactions. Artificial Force-Induced Reaction (AFIR) finding,[4] as well as the Single-ended Growing String Method (SSM),[5] generates approximate transition states (TSs) also based on databases, heuristics, or low-level methods, followed by TS optimization. The advanced AutoMeKin program,[6] successor of the tsscds method,[7–9] employs accelerated molecular dynamics (MD) simulations with semi-empirical density functional theory (DFT) and detects reactions by applying geometrical criteria (adjacency matrix, SPRINT coordinates) to find TS candidates. These procedures can find novel reaction pathways but are relatively computationally demanding for larger systems. Kinetik roBot (KinBot)[10] has proven to be efficient for the PES mapping of several systems but requires predefined reaction templates.[11,12] If tentative elementary steps are known, which requires already-given reactants and products, double-ended methods like the Nudged Elastic Band (NEB),[13–15] Freezing String Method (FSM),[16] and Growing String Method (GSM)[13,17] can be used to optimize the corresponding TSs.

In turn, MD simulations with reactive force fields (FFs) do not require a priori knowledge of possible reactions or products, are very fast, and allow for trajectory-based reaction detection from bond-order analysis. Reactive molecular dynamics (RMD) has also shown the capability to describe combustion reactions requiring much fewer computational resources than ab initio methods.[18] However, the time scale in those simulations still remains a problem when temperature is lowered, especially below 1000 K. To accelerate the simulation process and tackle the time scales required to study complex reaction systems, various methods have been developed that apply some sort of bias in thermal or potential energy or its gradient. Enhanced sampling methods, examples of which are metadynamics,[19] Temperature-Accelerated Dynamics (TAD),[20] and Adaptive Biasing Force (ABF),[21] concentrate on improving the overall sampling of the configuration space and thus address the problem of ergodicity in the MD simulation. However, these methods either lack flexibility, restricting the applicable scope to a small range of problems, which is a common weakness of accelerated MD methods, or, due to their biases, fail to accurately describe some reaction pathways and system properties. The requirements for the generality of a comprehensive PES searching process have led us to the Collective Variable-Driven Hyperdynamics (CVHD) method, which combines the strengths of accelerated MD methods and enhanced sampling methods and was recently developed by K. M. Bal.[22] Here, a small number of collective variables (CVs) is used to distinguish all relevant states of the system, and a successively added biasing potential ΔV fills up energy minima, thereby considerably reducing the waiting time between well-to-well transitions. As a result, an exhaustive exploration of the full dynamics of the fuel combustion process at technically relevant temperatures and time scales can be reached efficiently.[23]

In this work, we have developed an automated PES exploration by using parallel MD simulations with the Chemical Trajectory analYzer (CTY) software package.[24–26] The CTY software package can identify complex reaction mechanisms and derive rate constants from a single MD trajectory, as was demonstrated for high-temperature methane combustion[24] and $H_2 + O_2$ reactions with coupled recalculations via QC.[26] For the present work, we have extended CTY by an algorithm that explores the PES by subsequently starting further simulations for obtained products, facilitated by parallel simulations and acceleration by CVHD.

In the test cases, PESmapping of parallel simulations explored high-temperature chemistry of butenyl radicals C_4H_7 and the acetylene addition to phenyl radical $C_2H_2 + C_6H_5$, while CVHD was adopted for PESmapping of LTO of typical paraffins n-butane ($C_4H_7O_2$ network) and cyclohexane (cycC_6H_{11}OO network). For LTO, which requires CVHD acceleration, a thorough CV study was performed to optimize parameter sets of CVs. These have been integrated as templates into the CTY software code[24] so that CVs can be generated on the fly for MD simulations. The efficiency and performance of the generalized CV parameters was evaluated. To validate their effectiveness and flexibility, the updated CTY was then applied to 1,3-dioxane, a recently proposed bio- and e-fuel candidate.[27–29]

Methods

The automated PES mapping (see Fig. 1 for a workflow diagram) is enabled by interaction of CTY,[24] Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS),[30,31] and Plumed.[32] LAMMPS provides the platform of reactive MD simulations with the CVHD method implemented in version 2.7.0 of Plumed[33] to accelerate simulations, and CTY enables on-the-fly trajectory processing, reaction identifications, and reaction pathways extraction. The explored trajectories and related structures were carefully checked via Visual Molecular Dynamics (VMD),[34] and some TSs were validated by DFT calculations with Gaussian16.[35] CTY2.0 allows for an automated recalculation of stationary points (reactants, products, and transition states) with higher-level QC methods, but this study focuses on the actual mapping-out of PES as a necessary step in complete characterization of reactive systems. Geometries extracted from reactive MD trajectories serve as excellent starting points for higher-level QC optimizations,[26] especially with more sophisticated TS-finding algorithms.[36–38]

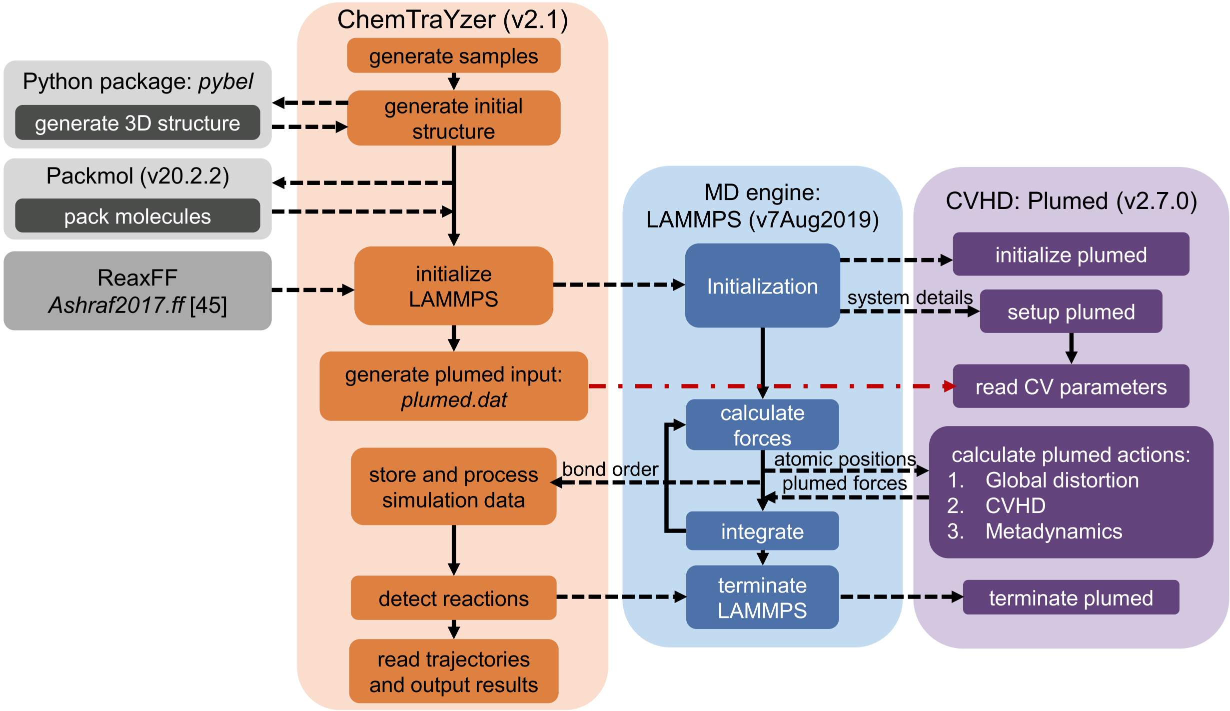


Figure 1: Workflow of PESmapping. Solid lines represent transitions of activities and dashed lines represent interaction of different programs. The red dash-dotted line emphasizes a newly constructed connection in this work.

Simulation details

The reactive force field ReaxFF provides a good compromise between accuracy and computational effort. We employ the parametrization for C/H/O from the work of Ashraf and van Duin,[39] which is an extension of the first ReaxFF combustion force field for hydrocarbons ($> C_3$) developed by Chenoweth et al.[40] in 2008. Therein, QC data of CO/HCO reactions as well as H_2/O_2 chemistry was added. We perform MD simulations with this force field using LAMMPS[30,31] and QEq charge equilibration.[41]

The PES-mapping algorithm only requires choosing an initial molecular structure to be investigated (the seed species). For high-temperature pyrolysis (2500, 3500, and 4500 K) of the radicals, 50 replicas have been used for butenyl (C_4H_7) and 30 replicas for acetylene and phenyl (C_8H_7). The accelerating effect of high temperature here has been enough to observe a sufficient number of reactions, i.e. large parts of the network, in combination with an adequate total simulation time achieved by parallel replica sampling. In contrast, for LTO of butane, cyclohexane, and 1,3-dioxane, all of the simulations have been carried out with CVHD acceleration. For temperatures of 500, 700, and 1000 K, 100 parallel samples each have been simulated. All samples were run in parallel on one core each within a maximum of two hours wall-time. Each sample contains a single molecular structure and the box dimensions are chosen to correspond to a density of $10\,\mathrm{mol\,m^{-3}}$. For each sample, random initial velocity distributions have been assigned to the same initial positions. Each configuration has been simulated using a time step of 0.1 fs and a Nosé-Hoover thermostat[42,43] with a damping constant of 10 fs. The atomic configurations and connectivities have been written to text files every 1 ps and 20 fs, respectively.

For Trajectory analysis of our parallel simulations we have been using CTY. CTY performs on-the-fly analysis to detect reactive, i.e. bond order changing, events. Once a reaction is detected, energies and structures of molecules and TSs are extracted and subsequently optimized in LAMMPS. In this study, for each new well, further simulations have been started with 100 samples per default until no new well was formed (Since the number of atoms remains constant, the number of wells must be finite). For the TSs extracted by CTY,[24] a pre-optimization by a freezing method has been employed, which assigns zero force to the atoms actively involved in the reaction, which also have been detected and stored by CTY.

On-the-fly generation of collective variables for hyperdynamics

The CVHD method[32] is a CV-based algorithm that combines metadynamics with aspects from hyperdynamics by using a history-dependent bias constructed from a sum of repulsive Gaussians. The changes in CV values reflect how far the system has traveled away from the reference state and are thus defined as local distortions X_i. The local distortions are combined into a global distortion X_t using a p-norm to globally describe the distance from the reference states according to

$$X_\mathrm{t} = \left(\sum_i^n X_i^p\right)^{1/p} \tag{1}$$

Here, parameter p controls the weighting of the deviations from the reference states. A greater value of p provokes more localized events dominated by a small number of CVs with large distortions, whereas a smaller value of p leads to more "concerted" reactions influenced by more of the n CVs. Due to the limited description of the configuration space by a single set of CVs, redefinition of the CVs is generally required when the system reaches a new state after a transition, i.e. reaction. Therefore, a criterion of the occurrence of a transition and the following change in the reference state is needed, which is defined here based on the global distortion. In case of possible transitions, the global distortion should take a value equal to or greater than a predefined value X_cut.

In this work, the CVHD method is implemented with Plumed,[32,33] which serves as a plug-in in our MD simulation of reaction systems. While biasing CVs that monotonously increase with distance quite naturally favors bond dissociations, bond formation processes are conceptually more difficult to accelerate using CVHD. A convenient solution to also bias bond formation in addition to bond dissociation is provided by the use of a switching function (SWF),[44] enabled by the CVHD implementation in Plumed.[33] A rational SWF is described in Eq. 2. It maps bond distances r from the interval of $]0, +\infty[$ to a variable s between $]0, 1[$.

$$s(r) = \frac{1 - \left(\frac{r-d_0}{r_0}\right)^n}{1 - \left(\frac{r-d_0}{r_0}\right)^m} \tag{2}$$

Default values amount to $d_0 = 0$, $n = 6$, and $m = 12$. The adjustable parameter r_0 describes the turning point of the switching function and is chosen to be 1.5 Å here, cf. Fig. 2. This value should be reached for typical reaction coordinate values near a TS.

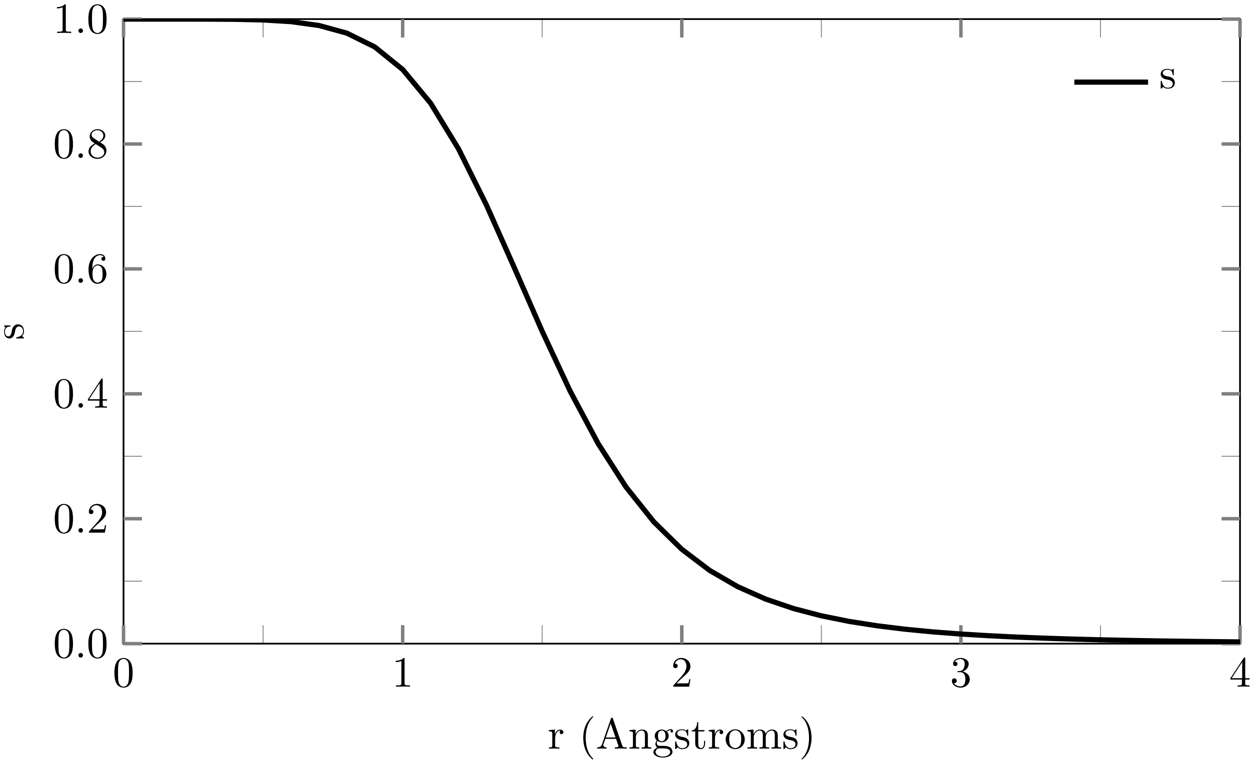


Figure 2: Rational switching function with default parameter values and $r_0 = 1.5\,\text{Å}$.

The local distortions from Eq. 1 are then redefined as $X_i = \|s_i - s_{\text{ref}}\|$. Choosing the reference value s_{ref} to zero corresponds to decreasing bond length and thus accelerating bond formation. In turn, a reference value $s_{\text{ref}} = 1$ leads to increasing bond length and therefore acceleration of bond breaking. In this way, bond formation and bond breaking can be treated consistently under the same Plumed-based framework. At the start of our simulations, upper and lower bounds of the transformed CVs are determined and the reference s_{ref} is automatically assigned to zero for distances less than r_0 with its default value of $1.5\,\text{Å}$ and to unity for distances greater than r_0.

Ultimately, along the CVs, Gaussian bells pile up during the simulation so that the bias can adapt itself to the barriers of the relevant transitions, as is shown in Fig. 3. The height and width of the Gaussian pulses, as well as the deposition rate influence not only the effectiveness, but also the efficiency of the acceleration, which is represented by the parameters HEIGHT in unit kcal/mol, SIGMA, and PACE in number of time steps, respectively.

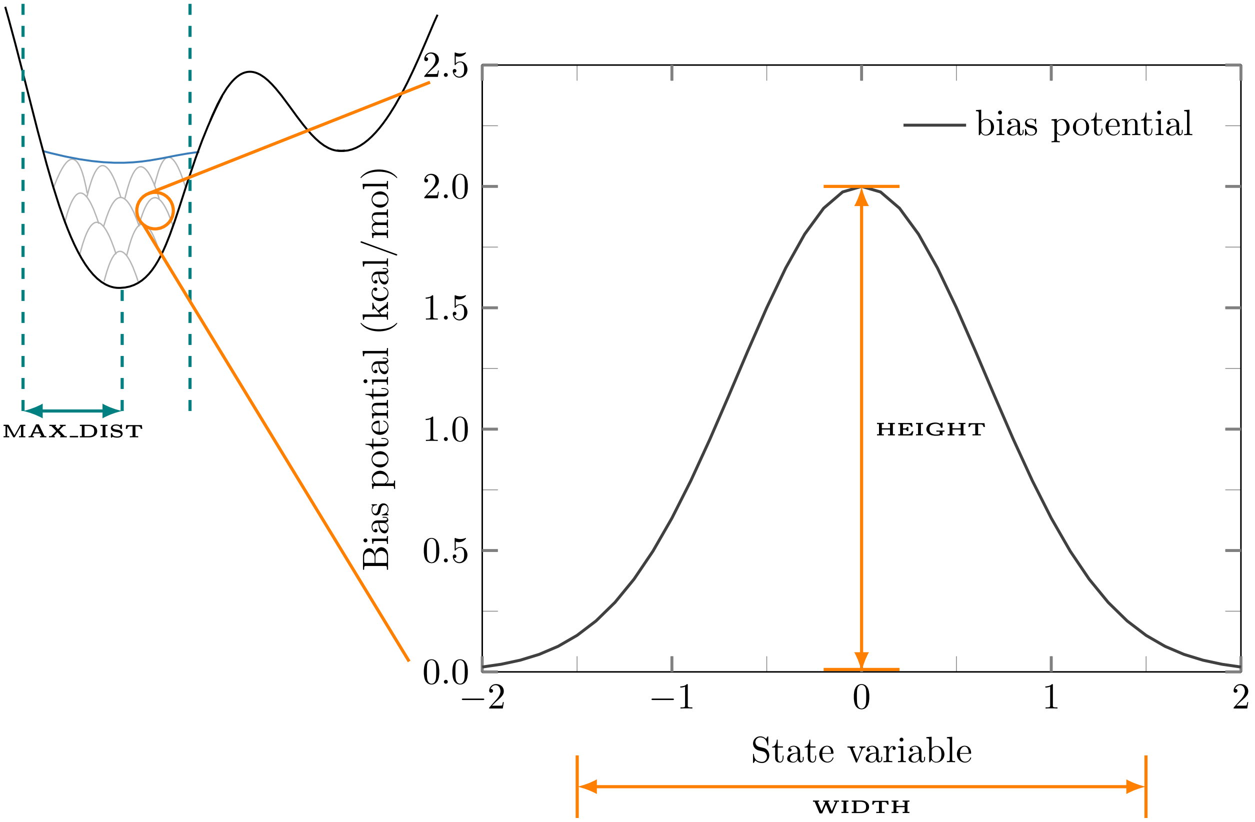


Figure 3: Bias in form of Gaussian pulses deposited on the energy surface

This large space of adjustable parameters poses the challenge of determining effective parameter sets. We here adopt a systematic approach that consists of, first, choosing proper CVs inspired by the LTO reaction mechanisms investigated here, and, second, acceleration parameter testing. Note that the PESmapping approach is independent of the application cases from this study. Optimization of the CVs employed herein only renders the exploration more efficient. For different reaction classes and systems (especially of LTO), we have developed seven templates with optimized CVs and parameter sets detailed in the Results section.

Validation of reaction pathways by quantum chemical calculations

In order to validate species and TSs, DFT optimizations and frequency calculations have been carried out at B3LYP/6-311+G(d,p) level.[45] For some ambiguous TSs, intrinsic reaction coordinate (IRC) calculations have been performed at the same level. Moreover, to increase the accuracy of the energy of some essential transition states on the PES of 1,3-dioxane, additional optimizations at the M06-2X/aug-cc-pVTZ level[46] starting from the B3LYP/6-311+G(d, p) results have been performed. While uncertainties in barrier heights evaluated for the DBH24 database amount to more than $4\,\mathrm{kcal\,mol^{-1}}$ for B3LYP, the M06-2X/aug-cc-pVTZ method yields uncertainties below $1\,\mathrm{kcal\,mol^{-1}}$.[47] This is the second-best method scaling at fourth order in the benchmark by Zheng et al.[47] and therefore well-suited to efficiently provide reliable barrier heights in this study.

Results and discussion

PES for pyrolysis: parallel simulations of butenyl isomerization and decomposition of C_4H_7 and C_8H_7

Reaction pathways on the C_4H_7 PES extracted by CTY from ReaxFF simulations are shown in Fig. 4. In total, 44 reactions of butenyl radicals, including the major isomerization and decompositions that are reported in the literature, were found within only ≈ 30 minutes wall-time. One shortcoming of the ReaxFF parametrization can be seen from e.g. the left and right ends of Fig. 4: the sum of energies of decomposition products is often higher than the TS connecting them. In principle, such submerged barriers may exist, but in these cases this is owed to the ReaxFF parametrization badly reproducing the energy of the system (similar issues may arise for B3LYP barrier optimizations). Nevertheless, the obtained pathways can be validated and the resulting geometries of TSs and products are descriptive starting points for higher-level optimizations. The inset of Fig. 4 shows the comparison of the TS geometry extracted by CTY at the ReaxFF level and the geometry optimized at QCISD/6-311++G(2df,2p) level from the literature.[48] The difference in bond lengths amounts to 10 % to 30 %, but the TS features are well captured and good enough to serve as an initial geometry guess for high-level quantum chemical calculation for both β-scission and isomerization

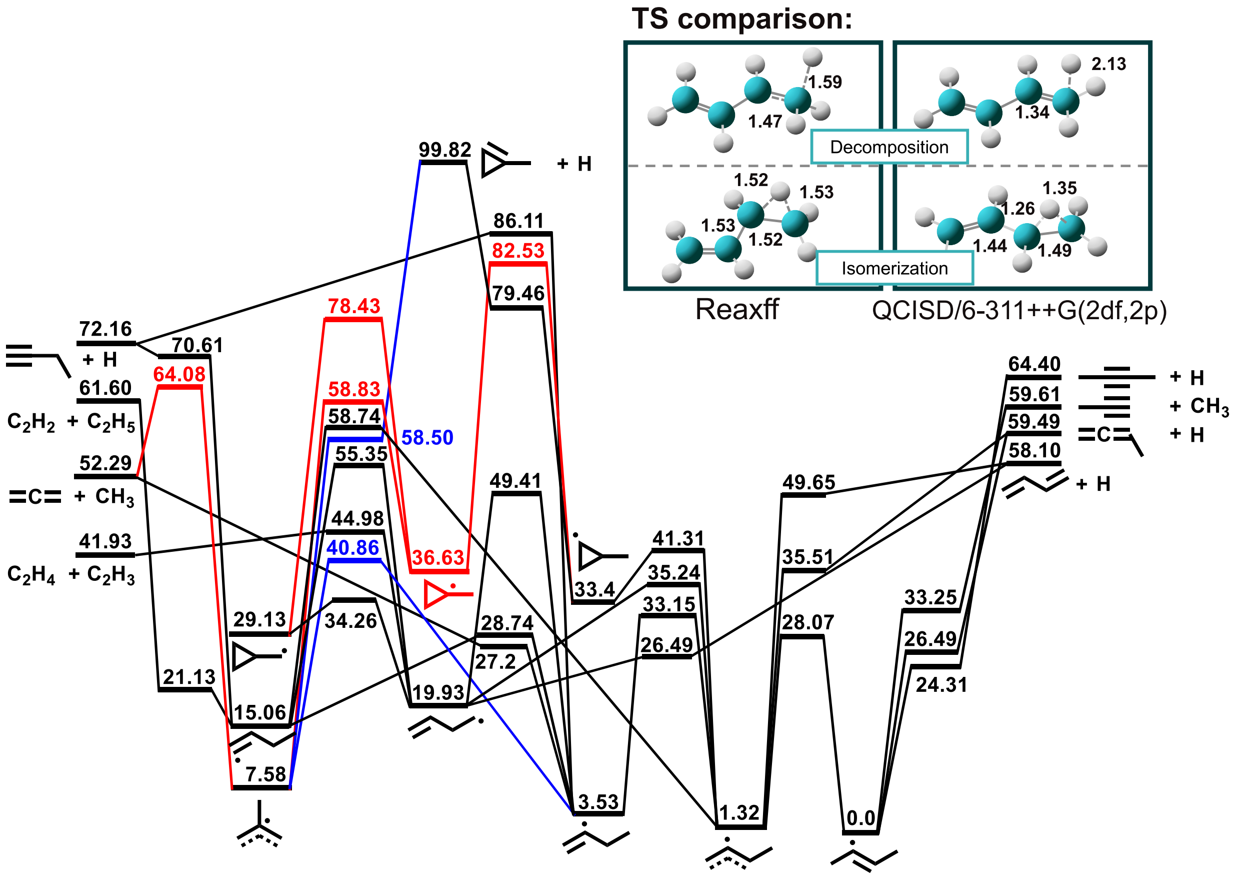


Figure 4: PES of C_4H_7 extracted with CTY from reactive MD simulations at ReaxFF level, Ashraf2017 parametrization.[39] Reactions that are the same as in Huang et al.[48] are denoted in black, newly-observed reactions in blue, other not observed yet possible reactions in red. Energy unit: kcal/mol

reaction of hydrocarbon radicals. Three reactions are marked in red, which are not found by the MD simulation. The reason is mainly that MD simulation naturally favors reactions that have lower barrier heights and we only simulated a limited temperature range and sample numbers. If more high-barrier reactions are desired, acceleration methods can be used which will be introduced below for the LTO systems.

Fig. 5 gives the explored PES of the reaction of acetylene and the phenyl radical, which serves as the first step of the hydrogen-abstraction-acetylene-addition (HACA) mechanism.[49] The most important decomposition and isomerization steps are observed (the same as observed by Mebel et al.[49]). The difference between TSs geometries at the ReaxFF level and B3LYP/6-311G** level from Mebel et al.[49] is very small: both the bond lengths and bond

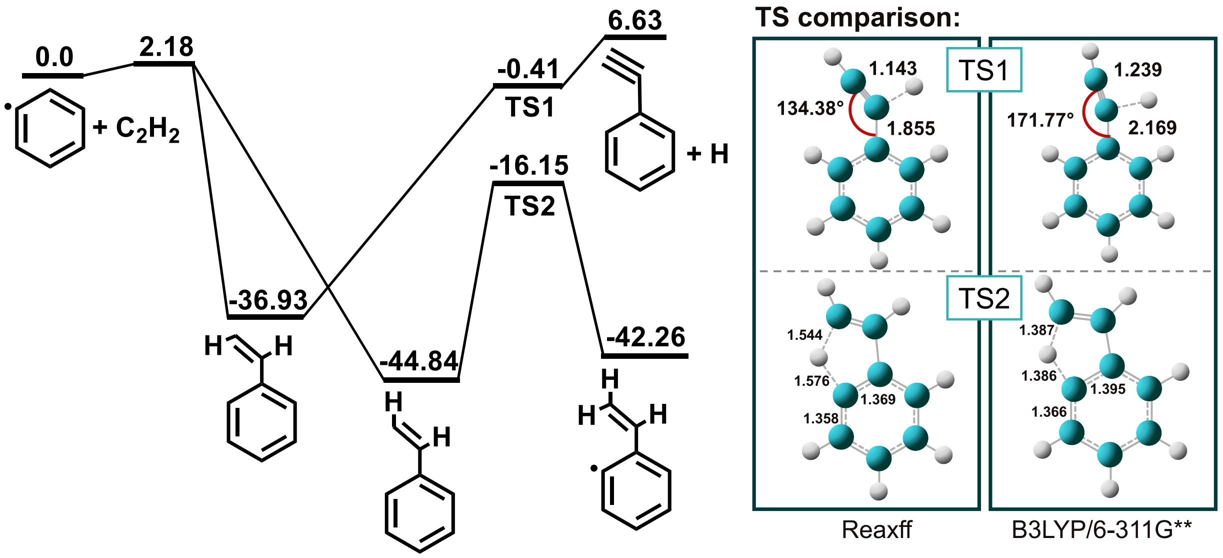


Figure 5: PES of C_8H_7 extracted with CTY from reactive MD simulations at ReaxFF level, Ashraf2017 parametrization.[39]

angles differ less than 15 %.

PES for low-temperature oxidation: CVHD-accelerated parallel simulations for $C_4H_9O_2$ and cyclic $C_6H_{11}O_2$

We observe typical LTO reactions such as RO_2 isomerization to QO_2H or QO_2H decomposition to cyclic ethers when CVHD acceleration is used. Our studies on butyl and cyclohexyl radical oxidation also serve to train the CVs required for the acceleration method. The $C_4H_9O_2$ PES from PESmapping is shown in Fig. 6. The PESmapping algorithm does not only reproduce a large number of important reactions known from literature, cf. Duan et al.,[50] but also leads to some reactions not considered in state-of-the-art mechanisms. The corresponding barriers are mostly larger than the ones of the known reactions and do not produce products of great importance to the butane oxidation mechanism, which may be the reason for them not being part of state-of-the-art butane mechanisms. Rather, the resulting reactions indicate that PESmapping reasonably explores chemical space without prior knowledge about the involved mechanism.

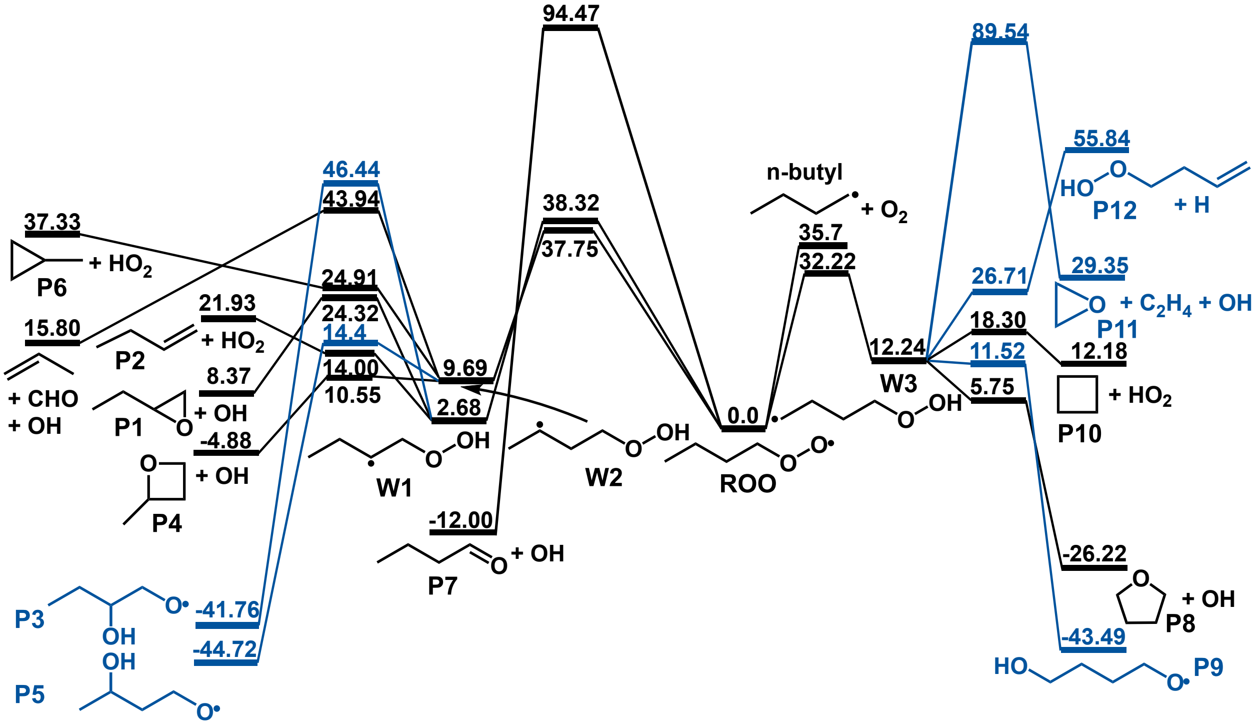


Figure 6: PES of the n-butyl peroxy radical $C_4H_9O_2$ extracted with PESmapping in CTY from reactive MD simulations at ReaxFF level, Ashraf2017 parametrization.[39] Reactions that are known from Duan et al.[50] are depicted in black, newly-observed reactions in blue. The energy unit is kcal/mol.

Fig. 7 presents the comparison between the extracted TS geometry for isomerization of the butyl peroxy radical to QO_2H, and decomposition of QO_2H at ReaxFF level, and the TS geometries optimized from the ReaxFF initial structures at the B3LYP/6-311++G(d,p) level. Here, the average relative deviation in relevant bond lengths ranges from 4.1 % to 19.5 % for TS structures. Once again, the ReaxFF level structures serve as a good initial TS guess.

We further validate PESmapping for a cyclic fuel, cyc-$C_6H_{11}O_2$. Knepp et al.[51] have laid out a detailed mechanism for the LTO of cyclohexane. The key reactions are reproduced by PESmapping, and two more pathways show up in the simulations, cf. Fig. 8.

A phenomenon showing up both in the butane and cyclohexane reaction systems, is the OH-shift isomerization, a pathway also noted by Green et al.[52] in their $QO_2H \longrightarrow HOQO$ isomerization study. In MD simulations, atoms are in constant vibration, especially lighter

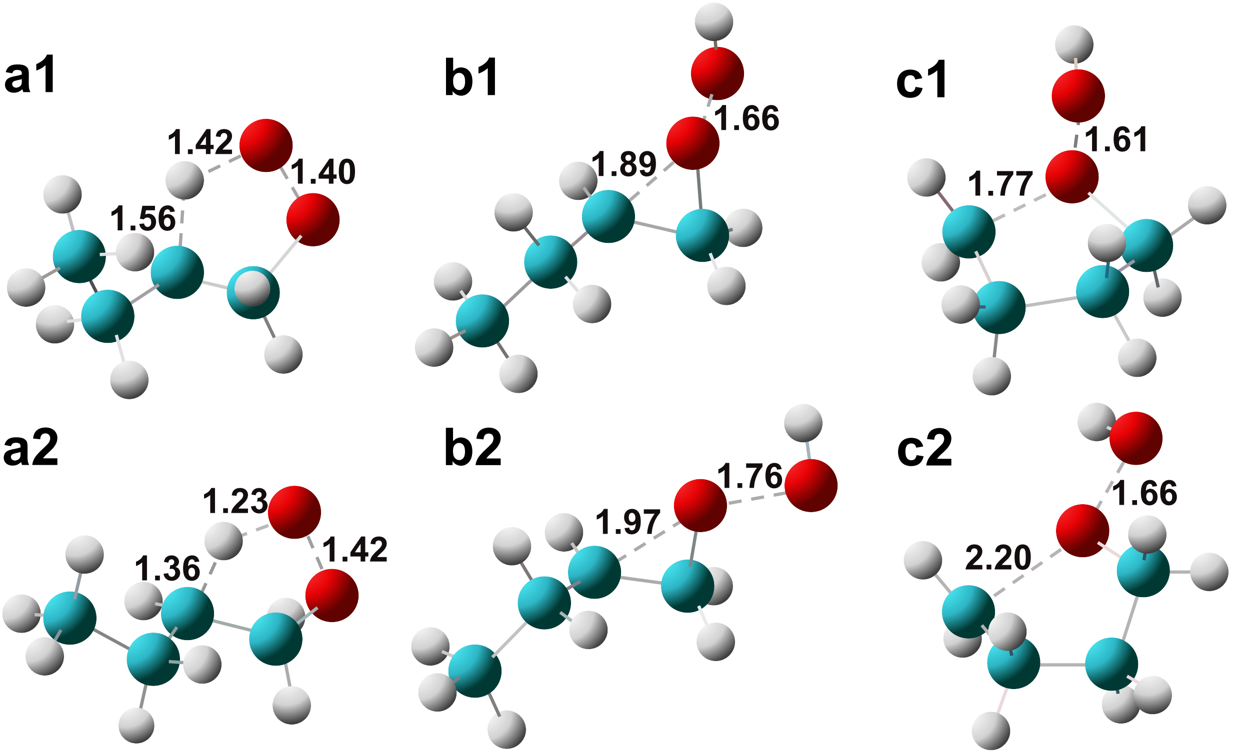


Figure 7: Transition state structure comparison for reaction $RO_2 \rightarrow W1$ (a1, a2), $W1 \rightarrow P1 +$ OH (b1, b2) and $W3 \rightarrow P8 + OH$ (c1, c2) (cf. Fig. 1 in Supplementary Information). a1, b1, c1 were optimized at the ReaxFF level, while b1, b2, b3 were calculated at the B3LYP/6-311++G(d,p) level (which was also used by Duan et al.[50] in 2021). Distances are reported in Angstrom.

atoms like H atom or groups like OH with less inertia. This results in a strong tendency to drift away either as a free radical or to other radical sites. Also due to the low barrier at the ReaxFF level ($< 10\,\mathrm{kcal\,mol^{-1}}$), reactions like H- or OH-shifts were frequently observed. To treat this reaction pathway rigorously, our QC approach at B3LYP/6-311G(d,p) level of theory (and further 6-31+G(d,p) basis to reproduce Knepp et al.[51] values) was adopted to calculate the barrier height for OH-transfer from cyclic C_6O_2H-2 structure. As shown in Fig. 9, the to-be-formed C−O bond in TS b initiated at the ReaxFF geometry a is almost identical to that reported by Knepp et al.[51] in their cyclohexyl oxidation study. The activation energy amounts to $21.8\,\mathrm{kcal\,mol^{-1}}$ according to the same B3LYP/6-31G(d,p) calculation as by Knepp et al.,[51] which decreases to $20.0\,\mathrm{kcal\,mol^{-1}}$ using the 6-311+G(d,p) basis set. In their ReaxFF study on low-temperature pentane oxidation, Krep et al.[53] also

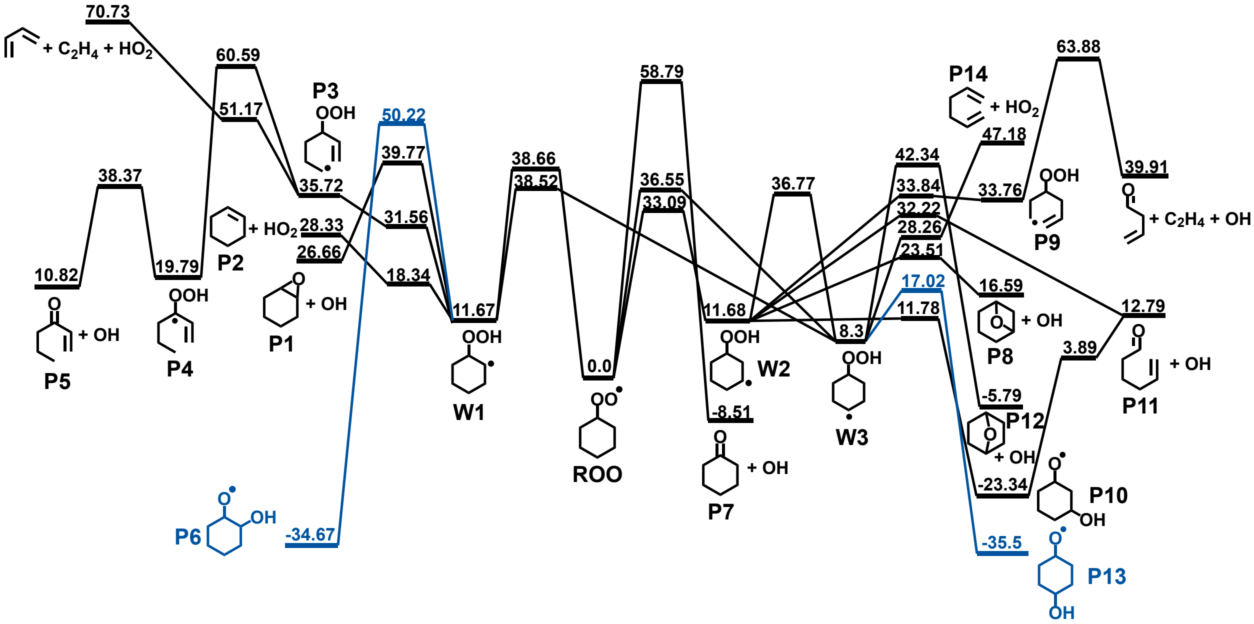


Figure 8: PES of the cyclohexyl peroxy radical cyc-$C_6H_{11}O_2$ extracted with PESmapping in CTY from reactive MD simulations at ReaxFF level, Ashraf2017 parametrization.[39] Reactions that are known from Knepp et al.[51] are depicted in black, newly-observed reactions in blue. The energy unit is kcal/mol.

noted a significant underestimation of methyl group migration (from 2-pentyl to isopentyl) by $45\,\mathrm{kcal\,mol^{-1}}$. Such isomerization barriers seem to be generally underestimated by ReaxFF.

CV efficiency study

Inspired by the various reactions of the LTO of butane and cyclohexane, we have developed various CV templates to efficiently explore the space of (LTO) reactions and studied the influence of CV parameters on effectively found reaction types and CPU times. The corresponding templates are named with four letters: the first two describe the elements of the atoms involved in the respective bond, followed by "bb" if bond breaking is accelerated and "bf" if bond formation is accelerated. Seven templates have been generated (and equally applied during the simulation setup) that make use of CVs with optimized parameter settings given in the SI: (1) chbb without switching function, (2) cobf, (3) cobf, cobf + ccbb, cobf + ccbf, (4) ccbb without switching function, (5) ccbb, (6) ccbf, (7) ohbf. Template 3 contains three cases with the same weight of one third.

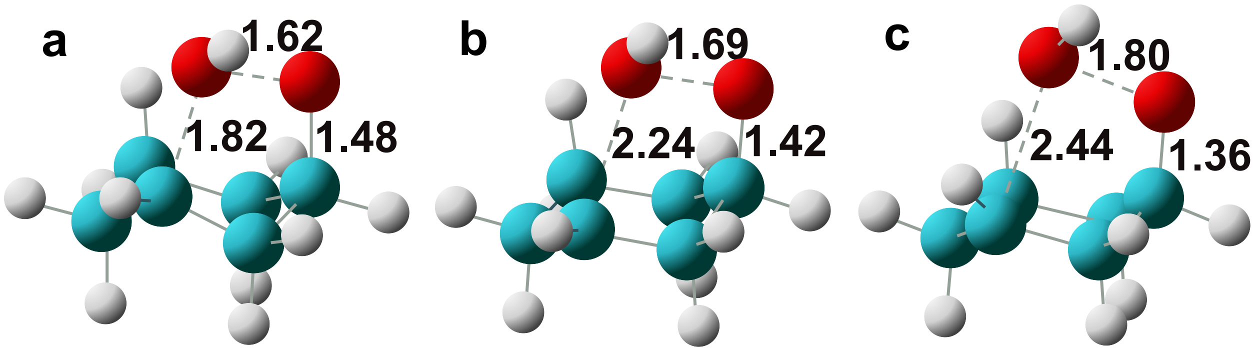


Figure 9: Comparison of TS structure for the OH-isomerization of cyclic C_6O_2H, structure 2 (W2 → P10 in Fig. 2 in Supplementary Information). Indices a, b, and c refer to the TS structure obtained, respectively, at ReaxFF level, B3LYP/6-311+G(d,p) level based on structure a, and at the B3LYP/6-31G(d, p) level used by Knepp et al.[51] in 2007. Distances are given in Angstroms.

To illustrate the major reaction classes addressed by the integrated templates, we use reactions from the cyclohexane 2-peroxy radical system as test case. We use eight representative key reactions from this network to study CV templates, listed in Table 1.

For the key reactions from Table 1, five templates that show an interesting distribution among accelerating these key reactions are presented in Fig. 10.

Different templates have a focus on different reaction classes as can bee seen from Fig. 10. Template 2 (cobf) derived from the cyclohexane system exhibits a good performance in discovering epoxides and OH-transfer reactions. Template 3 can be regarded as the most all-around and robust template owing to its complex composition. Template 4 (ccbb w/o switching function) and template 5 (ccbb) derived from cyclohexane and *n*-butane system, respectively, exert their effects on accelerating C – C bond fission. Both are implemented to improve the diversity of this kind of reactions in paraffin and naphthene systems. Template 6 (ccbf) was also obtained from the *n*-butane system in order to promote the formation of methyl-cyclopropane and cyclobutane. Its capability in discovering the same reaction class in cyclohexane system has been successfully proven by the template tests.

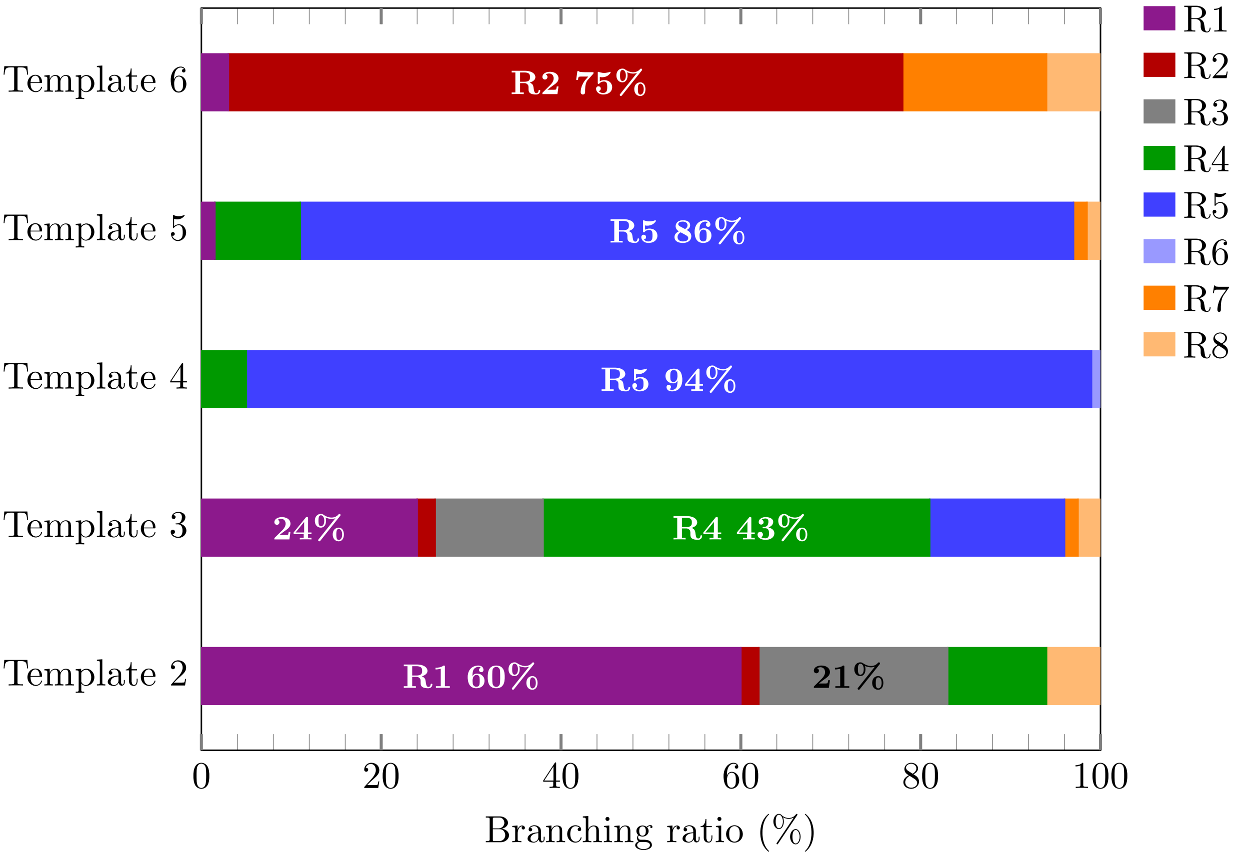
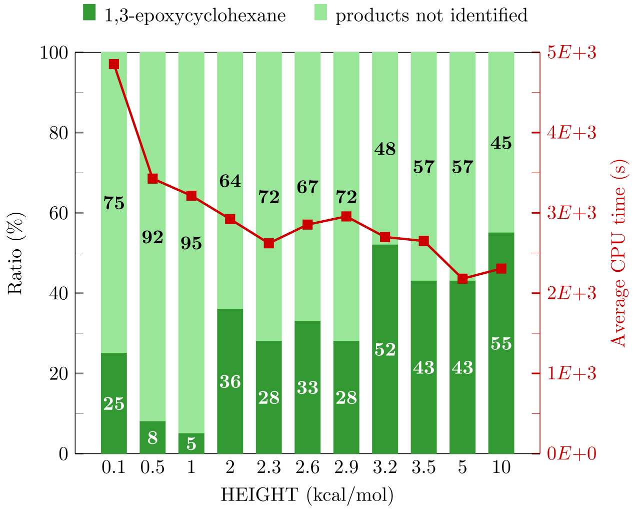


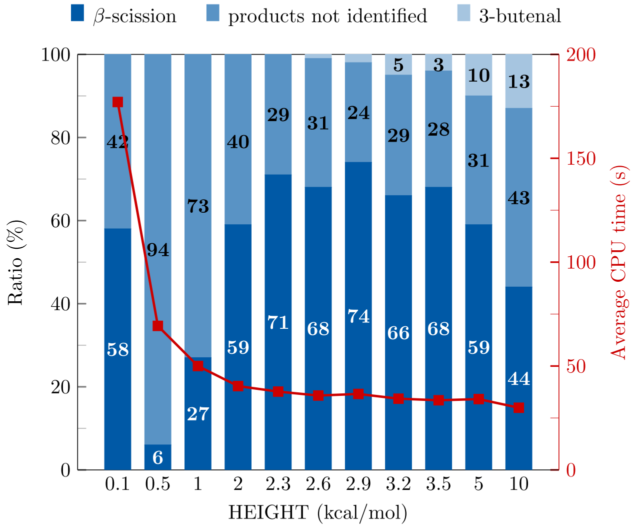
Figure 10: The distribution of key reactions defined in Table 1 in template 2, 3, 4, 5, and 6.

Influence of parameter HEIGHT

Fig. 11 shows the influence of parameter HEIGHT on the key reactions for cyclic C_6O_2H-2. As written above, template 2 will mainly yield epoxides for the cyclohexane case when working properly. We therefore check for how often 1,3-epoxycyclohexane is produced when varying parameter HEIGHT. For HEIGHT larger or equal to $3.2\,\mathrm{kcal\,mol^{-1}}$, the desired product is obtained in more than 40 % of the samples with this reaction event, cf. Fig. 11a. CPU time was saved by nearly 50 % compared to the case of $0.1\,\mathrm{kcal\,mol^{-1}}$. A further increase in HEIGHT continued to lower computational costs without diminishing the recognition rate. Through a comparison regarding the average CPU time, it was shown that the bond-breaking process, cf. template 4 in Fig. 11b, evolves an order of magnitude faster than the bond formation process. The latter one usually spends more time trapped in one well until sufficient energies have been accumulated. To further study the increased proportion of the desirable product 3-butanal, we extend the bias potential range by 5 kcal mol^{-1} and 10 kcal mol^{-1} in our automatically generated CVs, facilitated by the extremely low CPU time needed, cf. Fig. 11.



(a) Template2



(b) Template 4

Figure 11: Influence of parameter HEIGHT on finding key reactions for cyclic $C_6H_{11}O_2$-2, with corresponding product recognition rates, where R1 with template 2 and R5, R6 with template 4 serve as examples. The average CPU time is given by averaging the sum of CPU cost for each parallel simulation over the sample number and is recorded in seconds.

Influence of parameter PACE

Another CV parameter PACE determines the frequency of the deposition of Gaussian pulses in the number of time steps, which has retained a value of 200 during previous simulations. As is depicted in Fig. 12, a relatively high ratio of intended reactions to all reactions has been observed for values of 100, 200, and 300, which reflects the robustness of our CV templates, yet a further increase led to worse performance in identifying the key products as a result of the slower bias behavior. After testing a series of PACE values, the effectiveness of the unified PACE of 200 in the present work has been validated. Nevertheless, regarding computational cost a PACE of 100 or 300 may be a bit more efficient (note that the CPU time axis in Fig. 12 is cut off at 1000) and can be examined in future applications of our templates on similar reaction systems. Altogether, the parameter PACE has slighter influence than the parameter HEIGHT.

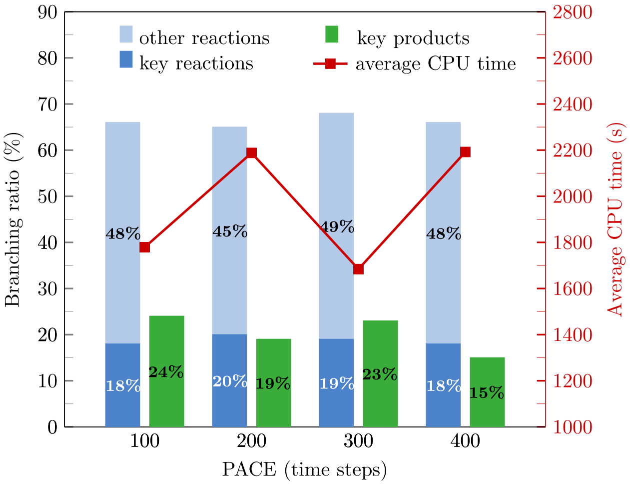


Figure 12: Influence of parameter PACE on the simulation of cyclohexane oxidation (cyc-$C_6H_{11}O_2$-2. The average CPU time is given by averaging the sum of CPU cost for each parallel simulation over sample number and recorded in seconds.

a priori application of PESmapping and CVHD to cyclic $C_4H_7O_4$

PESmapping in combination with proper CVHD templates can predictively reveal chemical mechanisms for fuels without knowledge on existing models. For the recently proposed potential biohybrid fuel 1,3-dioxane, no previous calculations have been available except the very recent contributions involving our research group.[27–29] We therefore have selected 1,3-dioxane oxidation upon designing this study as an interesting system to verify the effectiveness and reliability of our approach. This involved application of on-the-fly generated CVs using the preoptimized and integrated CV templates. As is summarized in Fig. 13, important reaction classes in the first O_2 addition were successfully reproduced within 600 samples for each peroxy radical and the hydroperoxyl radical. The batch time for each parallel simulation was set to a constant value of 2 hours. All three peroxy radicals of 1,3-dioxane followed similar reaction channels, including H-isomerization to form hydroperoxyl radicals, formation of bridged cyclic ethers and dissociation to bimolecular products or trimolecular fragments via β-scission.

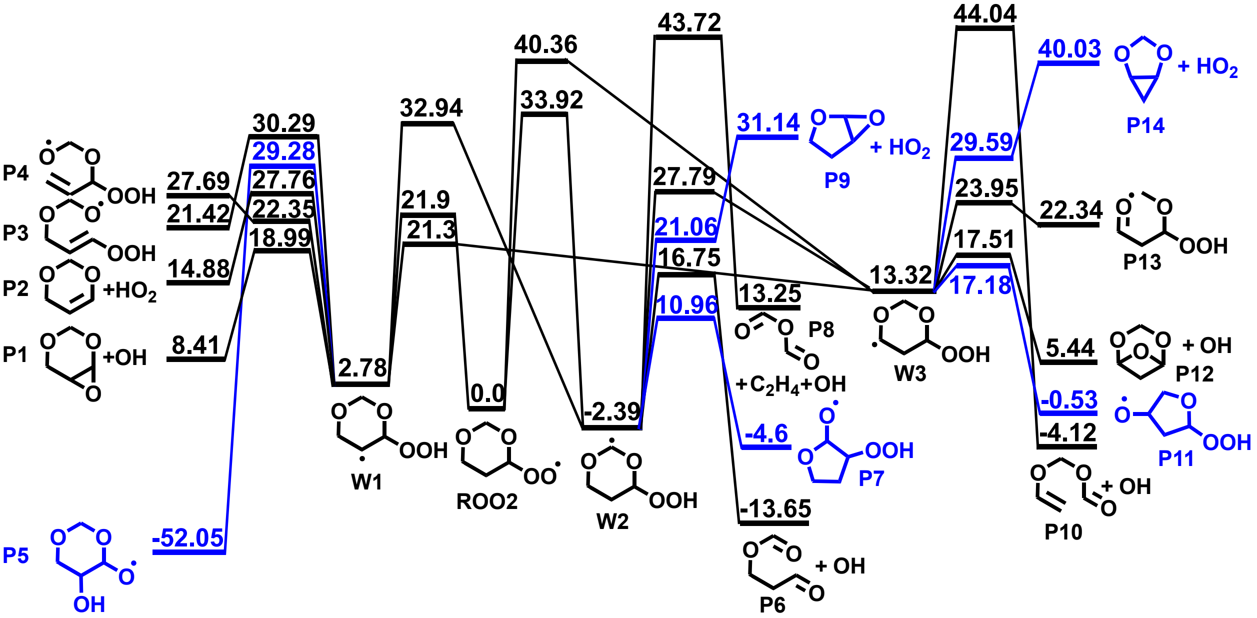


Figure 13: PES of $C_4H_7O_4$-2 obtained from reactive MD simulations at ReaxFF level, Ashraf2017 parametrization.[39] The reactions of the same class as on the PES of cyclic C_6O_2 is denoted in black, newly-observed reactions in blue. The energy unit is kcal/mol. PESs for the other two peroxy radicals are given in the Supplementary Information.

One reaction class of interest is the decomposition into a fused bicyclic compound and HO_2, which is not so important in hydrocarbon cyclic structures due to high strain energy.[50] As shown in Fig. 14, structures at the ReaxFF level identified by ChemTraYzer[24] were good approximations to DFT results, with a maximum deviation of 0.2 Å in the length of the critical forming C−O bond and of 0.4 Å in the lengths of the breaking C−O bond. Based on ReaxFF level results, the TSs were further validated at B3LYP/6-311+G(d,p) and M06-2X/aug-cc-pVTZ levels as shown in Table 2. The barrier heights for three/four/five-membered-ring TSs are all below 20 kcal mol^{-1}and that leading to the three-membered ring species is even below 5 kcal mol^{-1}. After refinement at the M06-2X/aug-cc-pVTZ level, the activation energies slightly increased, but are still below 30 kcal mol^{-1}. Thus these paths could possibly contribute to QO_2H consumption in 1,3-dioxane oxidation.

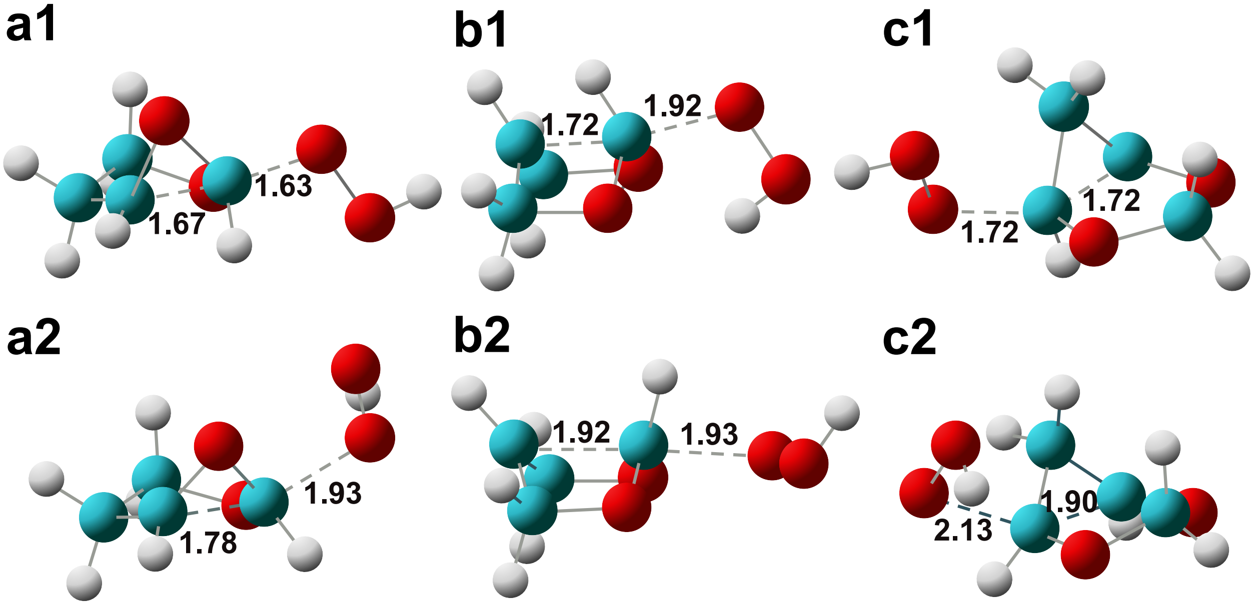


Figure 14: TS structures of reaction W4 → P9 $+HO_2$ (a1, a2), W5 → P24 $+HO_2$ (b1, b2) in Fig. 3 in Supplementary Information and W3 → P14 $+HO_2$ (c1, c2) in fig 13. Index 1 refers to structures optimized at the ReaxFF level, while structures calculated at the B3LYP/6-311+G(d, p) level carry index 2. Distances are reported in Angstroms.

PESmapping together with CV templates has revealed a multitude of oxidation reactions for the three dioxane peroxy radicals. We expect the eight CV templates developed in this study to accelerate oxidation reactions in general. Substantially different reaction systems — e.g. involving substitutions, diradicals, other chemical elements — will need further templates to work with CVHD acceleration. The PESmapping algorithm is independent of the acceleration method employed. At high temperatures, it can without acceleration quickly explore large chemical networks, as shown for butenyl and for the acetylene/phenyl system, which is promising for studying soot precursors and formation pathways.

Conclusions

In this work, we made progress in the automatization of PES mapping incorporating a new CVHD module into the CTY software, thus enabling for an efficient PES searching for hydrocarbons and oxygenated fuels regardless of temperature and their molecular structure. With and without this acceleration, a PESmapping algorithm could automatically explore large networks in pyrolysis and oxidation of fuel molecules. We showed that with parallel simulations and PES mapping, common isomerization and dissociation pathways can be found in wall-time scales of hours for hydrocarbon radicals (C_4H_7, C_8H_7). By deriving effective and flexible CV parameters and developing valuable templates that are highly adaptive to different reaction systems at low temperatures, the reproduction of almost all LTO reactions in butane, cyclohexane, and a newly proposed biohybrid fuel 1,3-dioxane, is achieved. Furthermore, some important reaction pathways have been validated using QC. The corresponding TS structures extracted from ReaxFF simulations have been proved to be excellent initial approximations, which underlines the effectiveness of the well-tuned CVs in accelerating correct dynamics.

A systematic study revealed influence of the bias potential parameters (HEIGHT and PACE) on the consumed CPU time and recognition rate of products. This deepens the understanding of the biasing process and offers guidance for future applications. Discussion of the Gaussian height parameter HEIGHT and additional discussion about the frequency of adding bias, namely PACE, confirms the reasonableness of choice (HEIGHT=$3.2\,\mathrm{kcal\,mol^{-1}}$ to $3.5\,\mathrm{kcal\,mol^{-1}}$ and PACE=200) while PACE may also be chosen to be 100 as well. The modular CV templates can be further expanded for more reaction systems in an elegant manner, such as the automatic exploration of soot formation and soot oxidation, which was previously inaccessible using existing computational chemistry methods.

While the PESmapping algorithm of this study is independent of the employed MD acceleration method, the optimized CVHD templates are expected to generally work for oxidation networks. Our CV template development procedure can be used and extended to further chemical elements and processes.

Table 1: **Key reactions for cyc-$C_6H_{11}O_2$-2 with corresponding products.**

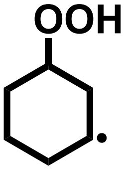
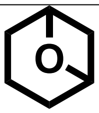
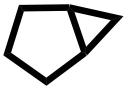
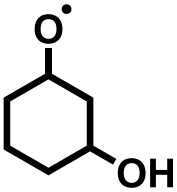
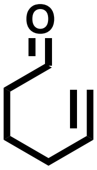
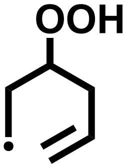
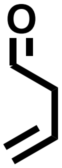
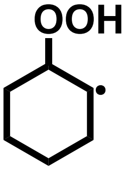
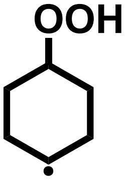
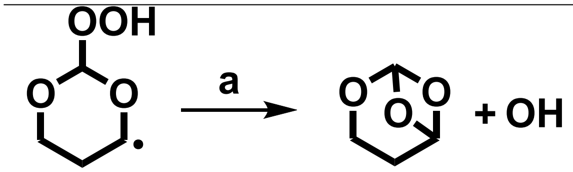
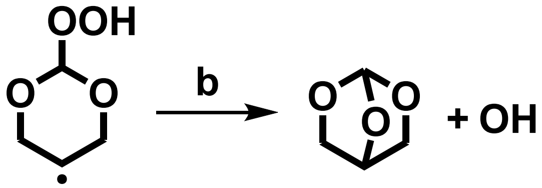
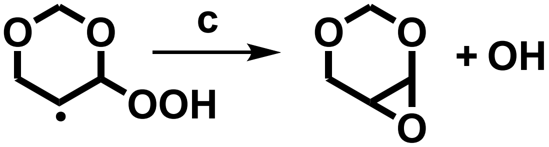
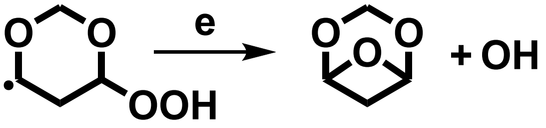
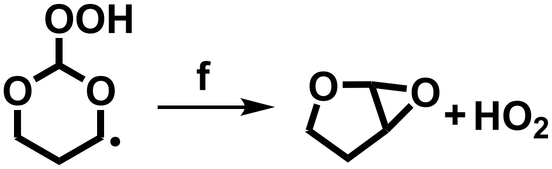
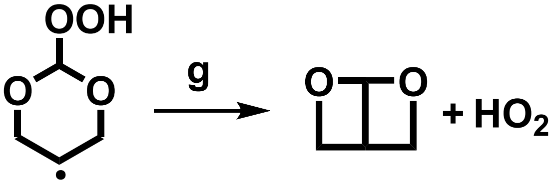
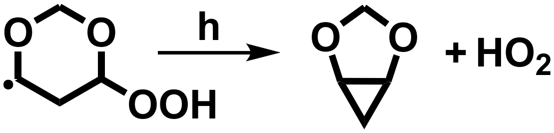
reactant	product(s)	reaction label
	 + OH	R1
	 + HO_2	R2
		R3
	 + OH	R4
		R5
	 + C_2H_4 + OH	R6
		R7
		R8

Table 2: **Activation energies for dioxane QO_2H decomposition reactions verified at DFT level (energies in kcal/mol).**

Reactions	B3LYP/ 6-311+G(d,p)	M06-2X/ aug-cc-pVTZ
	15.84	26.02
	11.16	20.38
	4.50	13.81
	13.13	24.20
	45.95	-
	62.43	-
	42.89	-

Acknowledgement

The authors thank Dr. Kristof Bal for the help with the PLUMED version of CVHD. The authors also thank B.Sc. Iryna Yevlakhovych for graphical data processing. This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – Cluster of Excellence 2186 ”The Fuel Science Center” – ID:390919832. Simulations were performed with computing resources granted by RWTH Aachen University under project rwth0468.

Supporting Information Available

The details of CVs, geometries obtained at the ReaxFF level, and full potential energy surfaces as well as additional figures are provided in SM1-3.

The following files are available free of charge.

- PESmapping.zip: archive containing two python scripts: PESmapping.py with the PES mapping algorithm used in this study and simulation.py with the CTY module containing CVHD templates developed in this study
- SM1-PESmapping.xls: optimized parameter sets for optimized CVs
- SM2-PESmapping.xls: all stationary point structures from this study
- SM3-PESmapping.pdf: PES diagrams for the other two cyclohexyl peroxy radical systems along with a flowchart depicting the on-the-fly generation of CVs

References

(1) Leitner, W.; Klankermayer, J.; Pischinger, S.; Pitsch, H.; Kohse-Höinghaus, K. Advanced Biofuels and Beyond: Chemistry Solutions for Propulsion and Production. *Angewandte Chemie International Edition* **2017**, *56*, 5412–5452.

(2) Dewyer, A. L.; Argüelles, A. J.; Zimmermann, P. M. Methods for Exploring Space in Molecular Systems. *WIREs Comput. Mol. Sci.* **2018**, *8*, e1354.

(3) 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.

(4) Maeda, S.; Harabuchi, Y. Exploring paths of chemical transformations in molecular and periodic systems: An approach utilizing force. *WIREs Comput. Mol. Sci.* **2021**, *11*, e1538.

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

(6) Martínez-Núñez, E.; Barnes, G. L.; Glowacki, D. R.; Kopec, S.; Peláez, D.; Rodríguez, A.; Rodríguez-Fernández, R.; Shannon, R. J.; Stewart, J. J. P.; Tahoces, P. G. et al. AutoMeKin2021: An open-source program for automated reaction discovery. *J. Comput. Chem.* **2021**, *42*, 2036–2048.

(7) Rodríguez, A.; Rodríguez-Fernández, R.; A. Vázquez, S.; L. Barnes, G.; J. P. Stewart, J.; Martínez-Núñez, E. tsscds2018: A code for automated discovery of chemical reaction mechanisms and solving the kinetics. *J. Comput. Chem.* **2018**, *39*, 1922–1930.

(8) Martìnez-Núñez, E. An Automated Transition State Search Using Classical Trajectories Initialized at Multiple Minima. *Phys. Chem. Chem. Phys.* **2015**, *17*, 14912–14921.

(9) Martìnez-Núñez, E. An Automated Method to Find Transition States Using Chemical Dynamics Simulations. *J. Comput. Chem.* **2015**, *36*, 222–234.

(10) Zádor, J.; Najm, H. N. *Automated exploration of the mechanism of elementary reactions*; 2012.

(11) Zádor, J.; Martí, C.; Van de Vijver, R.; Johansen, S. L.; Yang, Y.; Michelsen, H. A.; Najm, H. N. Automated Reaction Kinetics of Gas-Phase Organic Species over Multiwell Potential Energy Surfaces. *J. Phys. Chem. A* **2023**, *127*, 565–588, Description of features implemented in kinbot, of challenges for kinbot, and 5 case studies. Discussion of the coupling to ME, MR effects; symmetry and stereo chemistry; mismatch of stationary points between different levels, viewer, fireworks, non-statistical effects. (KL, 7/2023).

(12) Van de Vijver, R.; Zádor, J. KinBot: Automated stationary point search on potential energy surfaces. *Comput. Phys. Commun.* **2020**, *248*, 106947, Unimolecular reactions for C, H, O and S only; TS search for predefined reaction types; New calculations from the products automatically; conformer searches; frequency and hindered rotor calculations; symmetry numbers; Automated workflow for pressure and temperature dependent rate coefficients; Visualization and production of a graph; 4 Applications; (KL, 3/2020).

(13) Jónsson, H.; Mills, G.; Jacobsen, K. W. Nudged Elastic Band Method for Finding Minimum Energy Paths of Transitions. Classical and Quantum Dynamics in Condensed Phase Simulations. 1998; pp 385–404.

(14) Henkelman, G.; Jónsson, H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.* **2000**, *113*, 9978.

(15) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. A climbing image nugded elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **2000**, *113*, 9901.

(16) Behn, A.; Zimmerman, P. M.; Bell, M., A. T.; Head-Gordon Efficient Exploration of Reaction Paths via a Freezing String Method. *J. Chem. Phys.* **2011**, *135*, 224108.

(17) Zimmerman, P. Reliable Transition State Searches Integrated with the Growing String Method. *J. Chem. Theor. Comput.* **2013**, *9*, 3043–3050.

(18) Senftle, T. P.; Hong, S.; Islam, M. M.; Kylasa, S. B.; Zheng, Y.; Shin, Y. K.; Junkermeier, C.; Engel-Herbert, R.; Janik, M. J.; Aktulga, H. M. et al. The ReaxFF Reactive Force-Field: Development, Applications and Future Directions. *npj Comput. Mat.* **2016**, *2*, 15011.

(19) Laio, A.; Parrinello, M. Escaping free-energy minima. *PNAS* **2002**, *99*, 12562–12566.

(20) Maragliano, L.; Vanden-Eijnden, E. A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations. *Chem. Phys. Lett.* **2006**, *426*, 168–175.

(21) Darve, E.; Pohorille, A. Calculating free energies using average force. *J. Chem. Phys.* **2001**, *115*, 9169–9183.

(22) Bal, K. M.; Neyts, E. C. Merging Metadynamics into Hyperdynamics: Accelerated Molecular Simulations Reaching Time Scales from Microseconds to Seconds. *J. Chem. Theory Comput.* **2015**, *11*, 4545–4554.

(23) Bal, K. M.; Neyts, E. C. Direct observation of realistic-temperature fuel combustion mechanisms in atomistic simulations. *Chem. Sci.* **2016**, *7*, 5280–5286.

(24) Döntgen, M.; Przybylski-Freund, M.-D.; Kröger, L. C.; Kopp, W. A.; Ismail, A. E.; Leonhard, K. Automated discovery of reaction pathways, rate constants and transition states using reactive molecular dynamics simulations. *J. Chem. Theory Comput.* **2015**, *11*, 2517–2524.

(25) Kröger, L. C.; Kopp, W. A.; Döntgen, M.; Leonhard, K. Assessing Statistical Uncertainties of Rare Events in Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2017**, *13*, 3955–3960.

(26) Döntgen, M.; Schmalz, F.; Kopp, W. A.; Kröger, L. C.; Leonhard, K. Automated Chemical Kinetic Modeling via Hybrid Reactive Molecular Dynamics and Quantum Chemistry Simulations. *J. Chem. Inf. Model.* **2018**, *58*, 1343–1355, PMID: 29898359.

(27) Huang, C.; Zhao, Y.; Roy, I. S.; Chen, B.; Hansen, N.; Pitsch, H.; Leonhard, K. Pathway exploration in low-temperature oxidation of a new-generation bio-hybrid fuel 1,3-dioxane. *Proceedings of the Combustion Institute* **2023**, *39*, 385–393.

(28) Hellmuth, M.; Chen, B.; Bariki, C.; Cai, L.; Cameron, F.; Wildenberg, A.; Huang, C.; Faller, S.; Ren, Y.; Beeckmann, J. et al. A Comparative Study on the Combustion Chemistry of Two Bio-hybrid Fuels: 1,3-Dioxane and 1,3-Dioxolane. *J. Phys. Chem. A* **2023**, *127*, 286–299.

(29) Huang, C.; Zhao, Y.; Roy, I. S.; Cai, L.; Pitsch, H.; Leonhard, K. Effect of methyl substituents, ring size, and oxygen on bond dissociation energies and ring-opening kinetics of five- and six-membered cyclic acetals. *Combustion and Flame* **2022**, *242*, 112211.

(30) Plimpton, S. J. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19.

(31) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in 't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D. et al. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **2022**, *271*, 108171.

(32) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A. et al. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications* **2009**, *180*, 1961–1972.

(33) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185*, 604–613.

(34) Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics. *J. Molec. Grraphics* **1996**, *14*, 33–38.

(35) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H. et al. Gaussian˜16 Revision C.01. 2016; Gaussian Inc. Wallingford CT.

(36) Maeda, S.; Harabuchi, Y. On Benchmarking of Automated Methods for Performing Exhaustive Reaction Path Search. *J. Chem. Theory Comput.* **2019**, *15*, 2111–2115.

(37) Krep, L.; Schmalz, F.; Solbach, F.; Komissarov, L.; Nevolianis, T.; Kopp, W. A.; Verstraelen, T.; Leonhard, K. A Reactive Molecular Dynamics Study of Chlorinated Organic Compounds. Part II: A ChemTraYzer Study of Chlorinated Dibenzofuran Formation and Decomposition Processes. *Chem. Phys. Chem.* **2023**, *24*, e20220078.

(38) Papusha, M.; Schmalz, F.; Kopp, W.; Leonhard, K. Reaction Barriers from MD Simulations. *J. Chem. Inf. Model.* **2023**,

(39) Ashraf, C.; van Duin, A. C. T. Extension of the ReaxFF combustion force field toward syngas combustion and innitial oxidation kinetics. *J. Phys. Chem. A* **2017**, *121*, 1051–1068.

(40) Chenoweth, K.; van Duin, A. C. T.; Goddard III, W. A. ReaxFF Reactive Force Field for Molecular Dynamics Simulations of Hydrocarbon Oxidation. *J. Phys. Chem. A* **2008**, *112*, 1040–1053.

(41) Rappe, A. K.; Goddard, W. A. Charge equilibration for molecular dynamics simulations. *J. Phys. Chem.* **1991**, *95*, 3358–3363.

(42) Nosé, S. A Unified Formulation for the Constant Temperature Molecular-Dynamics Methods. *J. Chem. Phys.* **1984**, *81*, 511–519.

(43) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.

(44) Switching function in Plumed documentation. `https://www.plumed.org/docv2.5/user-doc/html/switchingfunction.html`.

(45) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of Electron Density. *Phys. Rev. B* **1988**, *37*, 785–789.

(46) (a) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215–241; (b) Dunning Jr., T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(47) Zheng, J.; Zhao, Y.; Truhlar, D. G. The DBH24/08 Database and Its Use to Assess Electronic Structure Model Chemistries for Chemical Reaction Barrier Heights. *Journal of Chemical Theory and Computation* **2009**, *5*, 808–821, PMID: 26609587.

(48) Huang, C.; Yang, B.; Zhang, F. Pressure-dependent kinetics on the C4H7 potential energy surface and its effect on combustion model predictions. *Combust. Flame* **2017**, *181*, 100–109.

(49) Mebel, A. M.; Georgievskii, Y.; Jasper, A. W.; Klippenstein, S. J. Temperature- and pressure-dependent rate coefficients for the HACA pathways from benzene to naphthalene. *Proc. Combust. Inst.* **2017**, *36*, 919–926.

(50) Duan, J.; Ji, J.; Ye, L.; Zhai, Y.; Zhang, L. A theoretical kinetics study on low-temperature oxidation of n-C4H9 radicals. *Proc. Combust. Inst.* **2021**, *38*, 681–689.

(51) Knepp, A. M.; Meloni, G.; Jusinski, L. E.; Taatjes, C. A.; Cavallotti, C.; Klippenstein, S. J. Theory, Measurements, and Modeling of OH and HO_2 Formation in the Reaction of Cyclohexyl Radicals with O_2. *Phys. Chem. Chem. Phys.* **2007**, *9*, 4315–4331.

(52) Green, W. H.; Wijaya, C. D.; Yelvington, P. E.; Sumathi, R. Predicting Chemical Kinetics with Computational Chemistry: Is QOOH->HOQO Important in Fuel Ignition? *Mol. Phys.* **2004**, *102*, 371–380.

(53) Krep, L.; Kopp, W.; Kröger, L. C.; Döntgen, M.; Leonhard, K. Exploring the Chemistry of Low-Temperature Ignition by Pressure-Accelerated Dynamics. *Chem. Systems. Chem.* **2020**, *2*, e1900043.

TOC Graphic

Some journals require a graphical entry for the Table of Contents. This should be laid out "print ready" so that the sizing of the text is correct.

Inside the `tocentry` environment, the font used is Helvetica 8 pt, as required by *Journal of the American Chemical Society*.

The surrounding frame is 9 cm by 3.5 cm, which is the maximum permitted for *Journal of the American Chemical Society* graphical table of content entries. The box will not resize if the content is too big: instead it will overflow the edge of the box.

This box and the associated title will always be printed on a separate page at the end of the document.