



UNIVERSITY OF LEEDS

This is a repository copy of *ChemDyME: Kinetically Steered, Automated Mechanism Generation through Combined Molecular Dynamics and Master Equation Calculations.*

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/176813/>

Version: Accepted Version

Article:

Shannon, RJ, Martínez-Núñez, E, Shalashilin, DV orcid.org/0000-0001-6104-1277 et al. (1 more author) (2021) *ChemDyME: Kinetically Steered, Automated Mechanism Generation through Combined Molecular Dynamics and Master Equation Calculations*. *Journal of Chemical Theory and Computation*, 17 (8). pp. 4901-4912. ISSN 1549-9618

<https://doi.org/10.1021/acs.jctc.1c00335>

© 2021 American Chemical Society. This is an author produced version of an article, published in *Journal of Chemical Theory and Computation*. Uploaded in accordance with the publisher's self-archiving policy.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

ChemDyME: Kinetically Steered, Automated Mechanism Generation Through Combined Molecular Dynamics and Master Equation Calculations

Robin J. Shannon^{1*}, Emilio Martínez-Núñez², Dmitrii V. Shalashilin³, David R. Glowacki¹

¹School of Chemistry, University of Bristol, Bristol, BS8 1TS, UK

²Department of Physical Chemistry, University of Santiago de Compostela, Spain

³School of Chemistry, University of Leeds, LS2 9JT, UK

Abstract

In many scientific fields, there is an interest in understanding the way in which complex chemical networks evolve. The chemical networks which researchers focus upon, have become increasingly complex and this has motivated the development of automated methods for exploring chemical reactivity or conformational change in a “black-box” manner, harnessing modern computing resources to automate mechanism discovery. In this work we present a new approach to automated mechanism generation implemented which couples molecular dynamics and statistical rate theory to automatically find kinetically important reactions and then solve the time evolution of the species in the evolving network. Key to this ChemDyME approach is the novel concept of “kinetic convergence” whereby the search for new reactions is constrained to those species which are kinetically favorable at the conditions of interest. We demonstrate the capability of the new approach for two systems, a well-studied combustion system, and a multiple oxygen addition system relevant to atmospheric aerosol formation.

1. Introduction

Predicting how a reactive system of molecules evolves over time is important in several scientific fields. In the gas phase, there is a need to unravel the chemical mechanisms at play in a combustion engine¹⁻⁵ or in the earth's atmosphere;⁶ in biological systems there is a need to model complex conformational landscapes;⁷ and in catalysis^{4,8-10} there is a need to screen potential catalytic mechanisms. In all cases an understanding of the important chemical reactions/conformational changes and the rate at which they occur is vital. Given the prevalence of this type of problem, there are a growing number of automated approaches designed to map the network of possible reactive or conformational changes for a given system.^{9,11-21}

What we describe here as automatic network generation of chemical networks, usually in practice involves a number of separate steps. First it is necessary to identify the possible reactions a system can undergo and then it is desirable to identify the transition state or the rate coefficient for these reactions to ascertain the kinetic relevance of each channel. Much work has gone into so called double ended methods²²⁻²⁴ for automatically finding transition state structures, but such methods assume one already knows the reactant and product of a reaction, which is not always the case when exploring a novel reaction mechanism. Several automated mechanism generation approaches then go on to automatically solve the coupled kinetics of the whole system as a final step by employing master equation or related Markov state model calculations.^{4,5,7,14,21,25-27}

In this work we introduce a new automated mechanism generation framework ChemDyME, (**C**hemical **N**etwork **M**apping though combined **D**ynamics and **M**aster **E**quation simulations) designed to automatically map reaction networks and then determine the evolution of the species in the system over time. The key aspect to the current approach is novel concept of “kinetic convergence” whereby the search for possible reactions is guided by of the kinetics network. In ChemDyME, molecular dynamics and master equation calculations are interleaved to ensure that only the kinetically relevant part of the reaction network is sampled. Our approach tracks the timescale associated with each new step we make

in the growing reaction network and the exploration is considered complete once we exceed some specified timescale (we describe this as reaching kinetic convergence). Thus, we aim to confine our search of the full possible reaction network for a system to only include those reactions deemed kinetically important under the conditions of interest. This kinetic convergence approach can be utilized to add new species to the system at the point when a new bimolecular reaction might be kinetically favorable and to our knowledge ChemDyME is the only automated reaction network approach able to add new species and bimolecular reactions to a growing network.

The aim of the paper is to present the details of the ChemDyME approach and to highlight potential applications. To this end sections 2 and 3 are devoted to introducing the reader to our methodology, firstly in the form of a high level, general, summary of the mechanism generation pipeline (Section 2) and then focusing upon more specific implementation and technical details in Section 3. Having laid out the conceptual and technical core of ChemDyME, Section 4 presents some benchmark calculations, where we utilize ChemDyME to study a well understood combustion system (the chemistry of OH + propyne) which has previously been studied by automated methods.²⁸ With these pieces in place, Section 5 demonstrate the solution of atmospheric oxidation system with a reactive bath gas by harnessing the newly developed “kinetic convergence” approach, key to the ChemDyME workflow. Specifically, we turn to the atmospheric reaction of the 1-methyl-hexene with ozone which is an archetypal, peroxy radical forming²⁸ reaction, for investigating the formation of highly oxygenated molecules (HOM),²⁹ key to the chemistry of atmospheric aerosol. Using ChemDyME we explore the network of the 1-methyl-hexene + $n\text{O}_2$ system where the n signifies that additional O_2 molecules are added to the system whenever ChemDyME detects addition of another O_2 is kinetically favorable.

2. High level summary of the three main parts of the ChemDyME pipeline

2.1 Framework overview

Given a starting molecular geometry, ChemDyME is designed to explore the network of reactions that the atomic system may undergo in an automated, “black-box” way. As the network is traversed (via master equation calculations determining transition times between species or nodes), ChemDyME tracks the cumulative time associated with all the transitions which have occurred, and the network exploration continues until some maximum time of interest, specified by the user for a given application, is reached. In this paper we term this “kinetic-convergence”, and the converged network should be representative of the important reactions under some conditions (temperature and pressure) and timescale of interest.

There are two main principles which have guided the development of the ChemDyME and which are emphasized in our chosen methodologies. The first of these principles is simplification: In ChemDyME we wish to focus only upon the subset of possible chemical reactions which are important under the conditions of interest and we have chosen techniques which prioritize exploring reactions in order of kinetic importance rather than exploring reactions quickly. The second principle is generality: ChemDyME is designed to be applicable to a wide range of systems and we have chosen to minimize the use of “chemical-intuition” in the ChemDyME workflow in order to aid flexibility and avoid human bias arising from “perceived-knowledge.” These two principles will be discussed further in sections 2.2 and 2.4.

The overall workflow involved in a ChemDyME calculation may be broadly split into 3 main categories:

- 1) **Molecular Dynamics (MD):** Within ChemDyME, reactive MD simulations are used to determine the possible reactions that the system may undergo from a given starting reactive state where the reactive state could be either a stable chemical species or some pair of bimolecular reactive moieties.

From the MD simulations ChemDyME is interested only in the possible product species which may be formed from the reactive state.

- 2) **Optimisation and Refinement (OR):** Once a set of reactions has been observed through MD, the reactants and products automatically undergo further refinement through geometry optimization calculations and subsequent vibrational frequency and energy calculations. The reaction path is then automatically characterized (see section 2.3) to determine whether there is a defined transition state or saddle point. If present, then the transition state is also subject to automatic refinement.
- 3) **Master Equation (ME):** Once the reactants, products and transition state (if present) have been refined, the structures, energies and vibrational frequencies of these species are then input into a master equation simulation. From the specified starting state these calculations determine which of the possible products is formed and track the transition time associated with the reaction. The kinetically favorable product is then used as a reactive state to seed new MD simulations. Master equation calculations are also used to automatically lump or equilibrate species based upon the relative kinetics of the different reactions in the network (see section 2.4). Once the ChemDyME run is complete, the resulting master equation input may be used to obtain phenomenological rate coefficients for the entire network.

These three different types of calculations create a feedback loop (Figure 1) with MD simulations providing input for OR calculations which then provide the input to the ME calculations, which in turn determine the starting point for the next set of MD calculations.

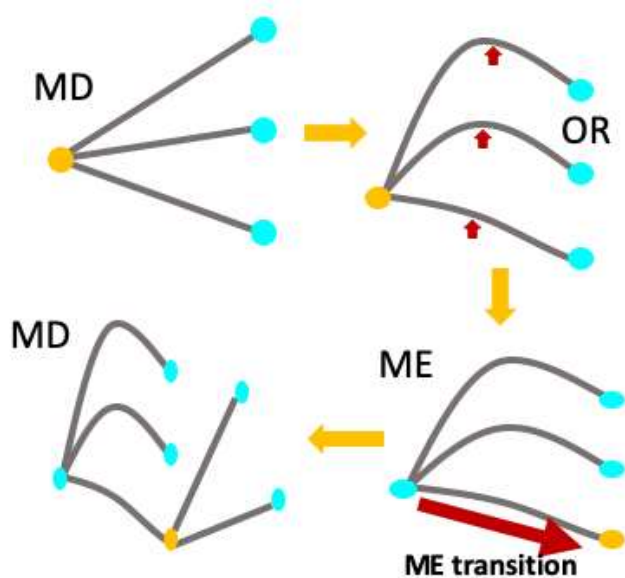


Figure 1: Outline of the steps involved in a ChemDyME calculation. First molecular dynamics (MD) is used to map the potential products which can be formed from a starting reactant or node. Then optimization and refinement (OR) is used to characterize each reaction path. Master equation calculations (ME) can be used to then identify the kinetically favorable product. New MD calculations are performed from this product and the cycle continues.

A flow chart showing the overall ChemDyME process in more detail is shown in Figure 2. To summarize, a ChemDyME calculation starts with a set of cartesian co-ordinates for some reactant state and determines whether the reactant state is a single molecule or a pair of bimolecular fragments. ChemDyME then optimizes the geometry of the reactant state and proceeds to run n_{MD} reactive MD simulations from the optimized reactive state to determine the possible reactions this reactant may undergo. Each MD simulation is tracked until a reactive event and product is identified (see 2.2). These products and reaction paths go through the OR process and each of these reactions is then added to a ME calculation. The ME calculation then returns a product species and the time associated with this transition τ_{trans} . This transition

time is then added to the cumulative system time τ_{tot} and the whole process starts again with new MD simulations initialized from the geometry corresponding to the ME reaction product. If at any point the ME product is a state which has already been used to initialize MD, the ME calculation is repeated until an unvisited (a state which has not been used to initialize MD) product is formed. This continues until τ_{tot} exceeds the specified maximum time of interest τ_{max} , at which point kinetic convergence is considered reached. All reactions in the network are saved in a single ME input file and the ChemDyME process may be repeated multiple times, building upon this input file, to ensure that no channels have been missed. When repeating a previously completed ChemDyME run, τ_{tot} is reset to zero and the ChemDyME process starts with an ME calculation from the initial reactant state. MD simulations are only run if an unvisited product is encountered in the ME simulations.

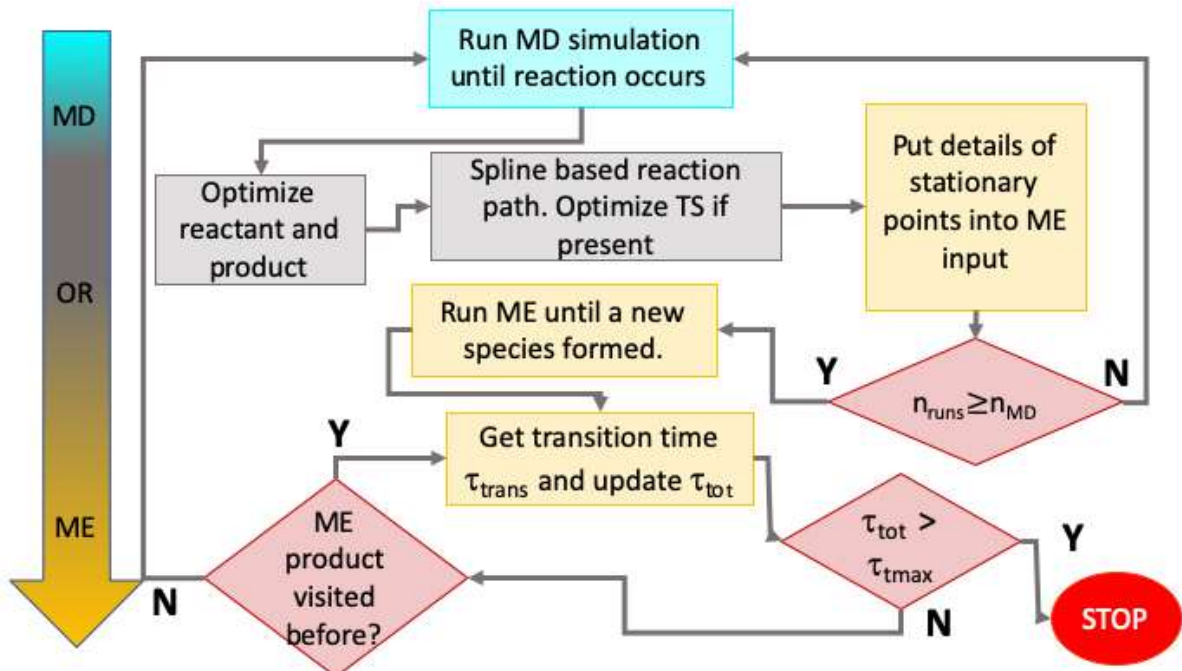


Figure 2: Flow chart showing the parts of a ChemDyME run in more detail. Cyan represents the MD part, grey represents the OR part and orange represents the ME part of the process.

2.2 Molecular dynamics simulations for discovering reactions

The choice of MD simulations to determine the reactions a given state may undergo is central to the principles of simplification and generality which guide the ChemDyME workflow. Apart from limitations associated with the method used to obtain the forces for the MD simulation and the inherent approximations involved in using a classical equation of motion, an MD simulation replicates the “real” (classical) behaviour of a chemical system. As such MD simulations do not explicitly rely upon any assumptions regarding the types of reaction which may occur and are an ideal tool for investigating novel reactions. With an appropriate forcefield and thermostat/barostat, unbiased MD samples reactions which are representative of the real reactions the system may undergo under the same conditions and thus the use of MD helps focus upon the subset of reactions which are important under the conditions of interest.

The disadvantage to using MD simulations is that they can be computationally expensive compared to other methods of searching for chemical reactions. In particular, MD suffers from the commonly encountered “rare- event” problem³⁰⁻³² whereby the particular process or “event” of interest, in this case chemical reaction, is extremely slow, relative to the fundamental timescale of the MD simulations. While rare event acceleration methods are an active area of research, the majority of these methods require some knowledge of the reaction or process one wishes to accelerate³³⁻⁴¹ and are not suitable for accelerating chemical reactions in general. Zheng et al² have successfully used well-tempered metadynamics in conjunction with SPRINT coordinates in order to explore chemical reaction in MD, and recent work from Grimme demonstrated another metadynamics based method for general reaction finding.⁴² In other cases, high temperatures or pressures^{17,25} have been successfully used to alleviate the rare event problem. However, the product channels observed from such simulations may be different

from the products formed at some lower temperature of interest and we recently demonstrated that high temperature MD simulations tend to favor entropically favorable dissociation reactions and may miss the enthalpically favorable but entropically unfavorable reactions which are important at lower temperatures.

6,43

In this automated mechanism generation framework, we utilize the recently developed “boxed molecular dynamics in energy” (BXDE)⁴³ method to accelerate the observation of reaction in MD. This BXDE method was shown to accelerate chemical reaction in MD simulations by several orders of magnitude whilst giving product yields in agreement with unbiased simulations at the same temperature. The BXDE approach is part of the boxed molecular dynamics (BXD),^{44–47} family of methods which steer molecular dynamics simulations into desired regions of configuration space through the use of reflective boundaries. In the BXDE case, the reflective boundaries are potential energy contours, and this causes the MD simulation to be steered towards configurations of high internal energy which are more likely to promote some reactive event.

For the majority of the MD simulations run in the current framework, BXDE is sufficient to promote the chemical reactions we wish to observe. The exception to this, is when we wish to start a ChemDyME run from two separate molecular fragments or moieties which undergo a bimolecular reaction. In this case a standard BXD constraint is automatically implemented along the separation of centers of mass (COM) of the two fragments, r_{com} . This standard BXD procedure has been documented extensively. In this case, the BXD constraint forces the two reacting fragments towards one another until they reach some specified r_{com} value r_{cutoff} . Once r_{com} is less than r_{cutoff} , BXDE is used to help surpass any reaction barriers and the BXD constraint is imposed whenever r_{cutoff} is exceeded, thus keeping the two reacting fragments in close proximity. In this case r_{com} is decreased particularly efficiently by employing a BXD inversion on any MD step at which r_{com} increases. The approach used to accelerate bimolecular reaction is shown schematically in Figure 3.

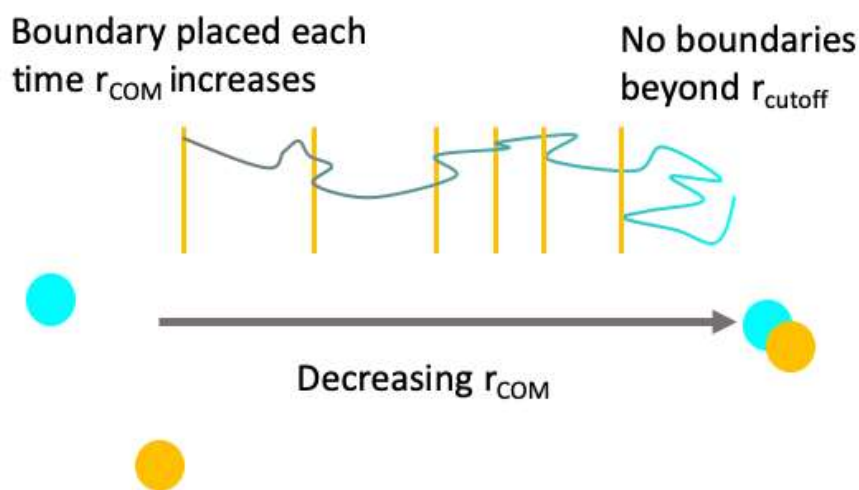


Figure 3: BXD approach for associating two separate fragments. The two circles represent the centers of mass of two fictitious fragments and r_{COM} is the distance between them. As the trajectory progresses, a new reflective boundary is placed whenever r_{COM} would increase until some value r_{cutoff} is reached.

For the proposed black-box framework for mechanism generation to be effective, it is necessary to define a criterion which automatically determines when an MD simulation has caused a chemical reaction to occur. In the current work a chemical reaction is defined as a change in the bonding structure of the molecule, i.e at least one chemical bond is formed or broken, however in the future we hope to add a criterion to optionally track conformational changes also. The reaction criterion used here is taken from the TSSCDS algorithm of⁴⁸ This criterion works as follows: two matrices are defined, d and d^{REF} . The first matrix has elements equal to the distance between atoms i and j in the system; the second matrix has elements, consisting of pre-defined ideal bond distances between atoms i and j . The ideal bond distances used here are shown in Table 1.

ij	d_{ij}^{REF} / Angstrom
CC	1.6
CH	1.2
CO	1.6
OO	1.6
OH	1.2
HH	0.8

Table 1: Ideal bond distances which make up the elements of the matrix

Using these two matrices, we form a connectivity matrix with elements C , where:

$$C_{ij} = \begin{cases} 1 & \text{if } d_{ij} < d_{ij}^{REF} \\ 0 & \text{otherwise} \end{cases} \quad \text{Eq. 1}$$

For the starting structure, this matrix identifies whether two atoms are bonded ($C_{ij} = 1$) or non-bonded ($C_{ij} = 0$). At each time step, the bonding structure (given by d and d^{REF}) of the system is compared with the reactant bonding structure (given by C) to monitor for reaction. Specifically, a reaction is then considered to occur if for an atom i :

$$\max(\delta_{in}) > \min(\delta_{ik}); \delta_{ij} = \frac{d_{ij}}{d_{ij}^{REF}} \quad \text{Eq. 2}$$

Here index n runs over atoms bonded to i (C_{ij} matrix elements equal to one) and index k runs over atoms which do not have a bond to i (C_{ij} matrix elements equal to zero). In ChemDyME the criteria must be met consistently for a specified number of MD steps to ensure that a reactive event has occurred and that

there is no recrossing back to reactants. Assuming the reaction criteria is met consistently, the final molecular configuration is used as a product guess for the OR phase of the ChemDyME workflow.

All BXDE simulations in ChemDyME place the reflective boundaries adaptively as described in our previous work.⁴³ Typically, the MD simulations are thermostated using a Langevin thermostat and the MD simulations are performed at a specified temperature T_{MD} . It should be noted that this temperature is a compromise between the efficiency of reaction sampling (reactions are observed in fewer MD steps at high T_{MD}) versus the number of product channels to be considered (higher T_{MD} simulations will produce more product channels which might not be relevant to low T chemistry). This temperature T_{MD} need not be identical to the temperature of interest T_{ME} which is used to define the conditions for the master equation calculation (see section 2.4).

2.3 Optimization and refinement of the stationary points of discovered reaction paths

Once a reaction has been found in MD, it is necessary to find and fully characterize the stationary points involved, to provide input for the master equation part of the framework. This section of the workflow has two distinct parts. First it is necessary to optimize the molecular structures of the correct stationary points. This is trivial in the case of the product(s) of the reaction, but it is much harder to identify (if present) a first order saddle point corresponding to a transition state between reactant and product. Once the stationary points have been optimized, the second part involves a series of refinement steps, to supply all the information necessary to include the reaction in a master equation simulation.

We will consider the optimization step first and Figure 4 presents a flow chart summarizing the key steps. As stated above, finding a stable minimum for the product is usually trivial. The geometry corresponding to the final frame of the reactive trajectory is taken and a geometry optimization is performed. A canonical SMILES²⁵ string is then obtained for the product and if this string indicates there

is more than one molecular fragment then the refinement steps are performed for each product. In the case of a saddle point, getting a good initial guess geometry for the transition state optimization is vital. An initial guess geometry is taken from the geometry at which the TSSCDS reaction criterion is met. To refine this geometry further we partially minimize the structure by freezing all the bonds which change over the course of the reaction and performing a constrained minimization. This partial relaxation is designed to remove excess energy in modes which are unrelated to the reactive event. Finally, a transition state optimization is performed on this partially relaxed structure. This is equivalent to the approach employed by Martínez-Núñez.²² An intrinsic reaction coordinate (IRC) calculation is then performed to check whether the optimized saddle point (if the TS search was successful)§ connects the expected reactants and products. If the correct transition is not found, this procedure is optionally repeated for several other geometries from snapshots along the reactive MD trajectory.

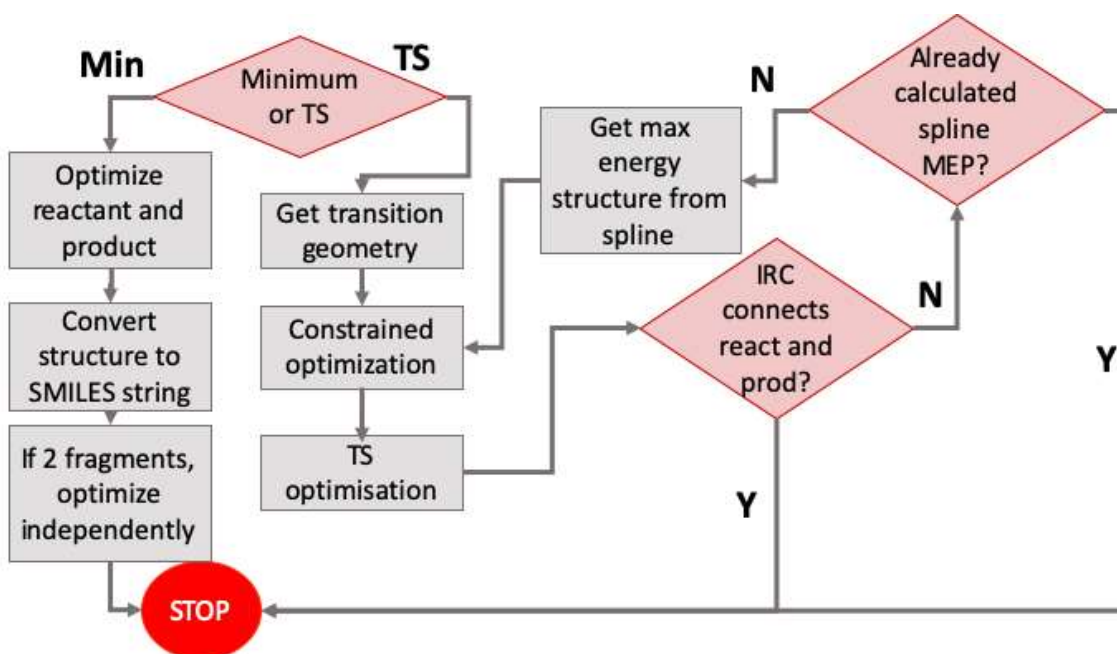


Figure 4: Flowchart outline the steps in the optimization and refinement (OR) part of the ChemDyME process.

If none of these transition state searches are successful, we attempt a different approach. Using the spline-based reaction path method⁴⁹ in the Scine ReaDuct^{50,51} software we optimize the minimum energy path between reactant and product and then attempt another TS search at the geometry corresponding to the maximum in the reaction path. There are of course cases where there is no defined barrier to the reaction and in the particular case where the reaction is either an association (two reactants one product) or a dissociation (one reactant two products), the reaction can be input in the master equation as a barrierless process using the inverse Laplace transform (ILT) method employed in MESMER.⁵¹ If however the reaction is an isomerization (one reactant, one product) and no TS can be identified, then this reaction is flagged for further (manual) investigation by ChemDyME and not included in the automated workflow.

Once an optimized structure (either minimum or saddle point) has been found, the refinement procedure begins. A hierarchy of levels of theory are used for this. A defined “initial optimization level” of theory is initially used for the optimizations already described and then an optional “refinement level” of theory is used to optimize the structure at a higher level of theory. Vibrational frequencies are then obtained using this “refinement level.” Finally, a single point energy calculation is performed at a “Single Point Level” of theory at the optimized ““Refinement Level”” geometry. The total energy of the species is then stored as the sum of zero-point vibrational energy and the Single Point Level electronic energy.

2.4 Master equation calculations for tracking the kinetic evolution of the network and guiding the search

The MD and OR parts of the workflow provide all the necessary information for performing statistical rate theory calculations, and chemical master equation simulations have become a standard tool for determining the kinetic behavior and rate coefficients of a complex chemical network. The

methodological details of the chemical master equation have been described in detail in numerous publications⁵²⁻⁵⁶ and there exist multiple open-source codes^{51,57,58} designed to solve the chemical master equation for a given network of interconnected intermediates, (wells in potential energy space).

In practice the energy space for each species in the ME network is discretized into energy grains. Microcanonical transition state theory and an appropriate energy transfer model are then used to calculate grain to grain rate coefficients. In ChemDyME we make extensive use of the recently developed Boxed Molecular Kinetics (BXK) algorithm.⁵⁹ This BXK methodology has significant conceptual overlap with the BXDE methodology employed for the MD simulations and was recently shown to accelerate Kinetic Monte Carlo, (KMC) ME simulations by several orders of magnitude. Briefly, the KMC approach tracks a stochastic trajectory of the population of the system from one grain to another. All transition probabilities are based upon the microcanonical grain to grain rate coefficients for energy transfer and chemical reaction. The BXK method simply places a boundary in energy space preventing lower energy grains being accessed and helping to alleviate the same rare event problem as encountered in the MD simulations.

A schematic of the BXK KMC approach used here is shown in Figure 5. Usually in a KMC simulation, one would need to run many stochastic trajectories, potentially starting from a range of initial grains, in order to converge the time evolution of the different species. In this work we take a different approach and each ChemDyME run can be viewed as a single KMC trajectory which is paused after each reactive step and for which new reactions are added to the system mid trajectory via the MD and OR procedures. Figure 5 shows such a KMC reactive step. This figure displays a simple two well system and depicts a stochastic trajectory involving a single energy transfer transition followed by reaction from well R to well P. The orange shaded areas signify the inaccessible grains due to the BXK procedure. The total time associated with this particular reactive step τ_{trans} is equal to the sum of the times for each step leading up to and including the reaction, and ChemDyME stores τ_{trans} , and the product energy grain. Before continuing the

stochastic trajectory, ChemDyME examines whether product P has been encountered before and if not, an MD/OR cycle is performed to add additional reaction to the master equation system. This ensures that the MD and OR procedures are only performed for the most kinetically favorable species under the conditions of interest. The KMC simulation finally terminates when τ_{tot} exceeds τ_{max} , constituting one full KMC trajectory. Additional KMC trajectories may be run to ensure no reaction channels are missed and, in each case, MD/OR calculations are only performed if a previously unencountered product is formed in a particular KMC step.

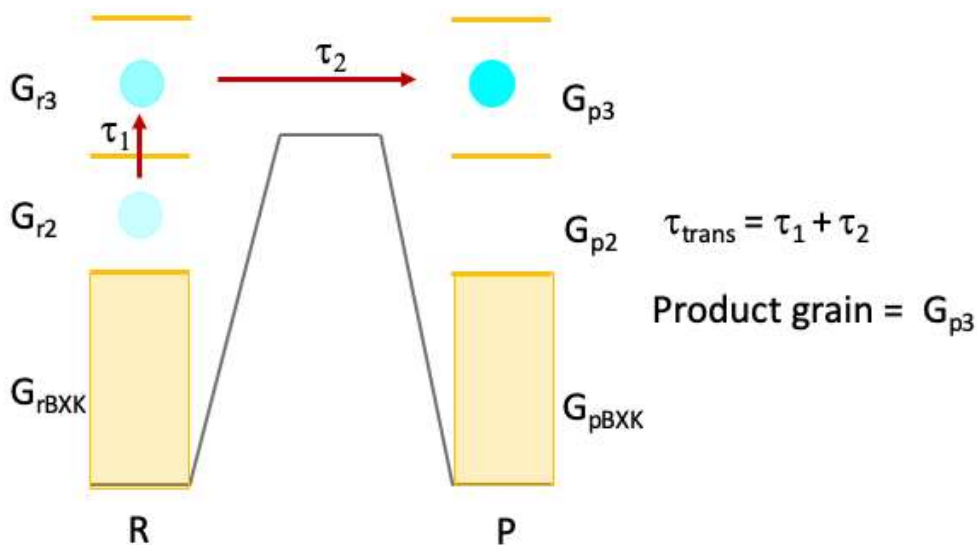


Figure 5: Stochastic master equation example. Here we have a system of two isomers with two active grains each and an excluded volume of rovibrational energies as per the BXK methodology. In this case the stochastic trajectory involves a single energy transfer step with associated time τ_1 and a reactive step with time τ_2 . ChemDyME then terminates the master equation and stores the product grain G_{p3} and the transition time $\tau_{\text{trans}} = \tau_1 + \tau_2$.

3. Technical details of the ChemDyME framework

The framework described in Section 2 has been implemented in the object-orientated Python package ChemDyME and utilises the Atomic Simulation Environment (ASE) package.⁵⁹

The ChemDyME framework may be obtained from: <https://github.com/RobinShannon/ChemDyME> and is distributed under the LGPL licence. ASE contains an “atom” class which stores the cartesian coordinates and various chemical properties for a particular chemical configuration. The “atom” class can be readily interfaced with a variety of existing electronic structure theory codes via the ASE “calculator” classes and allows access to the built in ASE optimisation methods when required.

3.1 External Code interfaces

The ChemDyME source code primarily takes form of a wrapper. The various chemical species in the chemical network are stored as ASE “atom” types and “calculators” are used to run external electronics structure theory calculations which are then gathered and analyzed automatically by ChemDyME. ChemDyME also interfaces with the MESMER⁵¹ code for all master equation calculations. The interaction between ChemDyME and external codes may be broken down into the three parts of the mechanism generation framework as follows:

- 1) MD: All MD calculations are carried out using a bespoke MD integrator class within ChemDyME. The starting structure for the MD is turned into an ASE “atoms” object and at each MD timestep forces are obtained by an external code and read back into ChemDyME. BXD classes in ChemDyME check whether a velocity inversion is required at each timestep and reset positions / update velocities accordingly.

2) By preference optimizations (both minimizations and saddle point searches) and frequency calculations are carried out by external codes, but minimizations and frequency calculations can be performed using built in ASE routines as a backup. If the open source, scine readuct⁴⁹ package is found then the spline-based reaction path ²²methods are utilized as part of the transition state optimization strategy although built in nudged elastic band methods in ASE are an alternate option for minimization the reaction path. Table 2 below lists the currently available codes which may be interfaced with ChemDyME and the capabilities of the interface.

Code	Capabilities	Appropriate ChemDyME level
SCINE (Sparrow and ReaDuct) ^{49,60}	Energies, Forces, Optimizations, Spline, IRC	Trajectory Level Initial Optimization Level Refinement Level Single Point Level
DFTB+ ⁶¹	Energies, Forces	Trajectory Level Single Point Level
XTB ⁶²	Energies, Forces	Trajectory Level Single Point Level
Gaussian ⁶³	Energies, Forces, Optimizations	Refinement Level Single Point Level
Molpro ^{64,65}	Energies, Forces	Single Point Level
NWChem ⁶⁶	Energies, Forces	Single Point Level

Table 2: External codes which may be used to generate forces or energies in a ChemDyME run

3) ME: At the start of every ChemDyME run a template xml input file is created for MESMER. As each reaction is found and characterized, ChemDyME writes the properties for reactant, transition state and product to the input xml and defines the reaction between them. When required ChemDyME updates the

starting species in the xml file and runs the MESMER executable. MESMER runs until a reactive transition occurs and ChemDyME then reads in the time associated with the transition along with the identity and energy gain of the product.

In addition to MESMER and electronic structure theory codes, ChemDyME also utilizes the PyBel interface to the Open Babel ⁶⁷ software to conveniently convert from one chemical format to another. This simplifies bookkeeping. As the Chemical network grows, ChemDyME needs to track each unique species which has been formed, in a convenient way. This is done through converting each new optimized product geometry into a canonical SMILES string using Open Babel. These SMILES strings are used to label each species in the reaction network and are invariant to the exchange of like atoms. These SMILES labels allow ChemDyME to keep track of which reactions have already been discovered and avoid wasted computational effort re-characterizing reactions which have already been identified. Pybel SMILES strings also indicate whether a given set of cartesian coordinates corresponds to multiple molecular fragments and these strings are used by ChemDyME to determine whether a particular reaction is dissociative. PyBel is additionally used to convert optimized cartesian coordinates into Chemical Mark-up Language (.cml) format to aid the creation of the MESMER xml input.

3.2 Dynamically changing the number of the atoms in the system

A ChemDyME run does not need to maintain a static number of atoms. When a dissociation occurs ChemDyME is able to continue exploring the reactive chemistry of one of the fragments formed (by default the fragment with the largest number of atoms). As will be discussed in Section 5, ChemDyME is also capable of adding additional bimolecular channels where it is deemed kinetically appropriate.

To add additional species the ChemDyME input defines a number of SMILES strings corresponding to molecules which might be considered part of the reactive bath for the system of interest and for each molecule the user also specifies a characteristic timescale τ_{bi} associated with estimated rate constant for reaction between the system and the bath molecule. After the MD and OR cycles have been completed for a particular node or species in the reaction network, ChemDyME runs n (default = 10) ME simulations and stores the slowest transition time $\tau_{slowest}$ from these. If $\tau_{slowest} > \tau_{bi}$ for a given bath molecule then it is considered that bimolecular reaction with this species may be competitive and an additional MD OR cycle is performed looking for reactions between the two species.

For each run in the bimolecular MD cycle, initial cartesian coordinates are created such that the COM's of the two moieties (reactive intermediate and bath) are separated by x Angstrom (default value is 8). To generate these coordinates randomly the coordinates of the first fragment are translated to the origin of the coordinate system. We then generate a random unit vector \hat{r} and $x\hat{r}$ thus defines a random point upon a sphere of radius x about the origin. Finally, we translate the COM of fragment 2 to $x\hat{r}$ to give the starting geometry for the bimolecular MD run. This is shown in Figure 6. The MD run then uses standard BXD on the separation of COM's (as described in 2.3) to gradually bring the two moieties into proximity with one another.

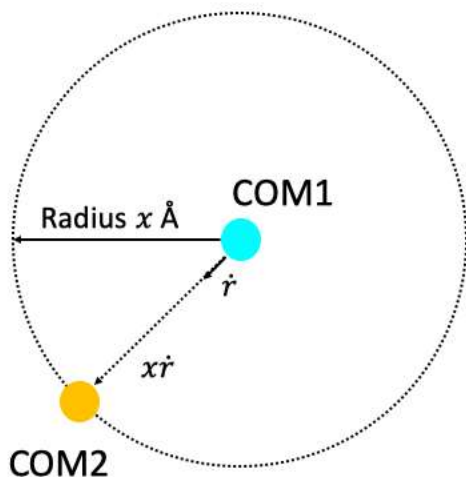


Figure 6: Procedure for generate coordinates for two separate moieties with centers of mass COM1 and COM2 such that they are randomly seperated by x Angstrom.

4. Benchmarking ChemDyME against the well-studied C₃H₅O atomic system starting from OH + propyne

Propyne and allene are key intermediates in combustion chemistry, originating in both propane flames and through the decomposition of larger alkanes.⁶⁸ The reaction of both species with the OH radical can lead to the resonantly stabilized propargyl radical which is of great interest with regards to soot formation mechanisms.^{1,69-76} As a result, the C₃H₅O atomic system, particularly initialized from the OH and propyne reaction has been the subject of a great deal of research, both experimentally and theoretically.^{1,76}

Theoretically there have been two particularly comprehensive studies studying the OH + propyne mechanism.^{1,76} Fitzpatrick⁷⁶ performed a manual search for transition states, intermediates and decomposition products formed from the OH + propyne reaction and subsequently Zador and Miller used an automated approach (Kinbot)¹ to thoroughly investigate the C₃H₅O system starting from both OH +

propyne and OH + allene. This automated study focused upon the lowest energy channels and highlighted a small number of channels that were missed by the Fitzpatrick study.

For the purpose of this paper, the OH + propyne system provides an unusually well studied system and thus represents an ideal benchmark for evaluating ChemDyME's use as an automated mechanism generation framework. To this end we have performed ChemDyME runs, starting with separate OH and propyne moieties. Input and output files for these simulations may be found in the examples folder of the ChemDyME github repository <https://github.com/RobinShannon/ChemDyME>. For the kinetic convergence we have chosen to run ChemDyME with a T_{ME} of 1000K and pressure of 5 bar representative of combustion conditions and a maximum time for kinetic convergence τ_{tot} of 1s. With regards to the MD portion of the ChemDyME workflow, we chose a T_{MD} of 1000K also to promote the discovery of lower energy pathways in the MD reaction finding step. A comparison of the various products formed from different T_{MD} can be found in the supporting information section S1.

The levels of theory used in the ChemDyME run are tabulated in Table 3. PM6 calculations are performed using the SCINE Sparrow code,⁶⁰ dftb2 calculations are performed using the DFTB+ code,⁶¹ all DFT calculations are performed using the Gaussian 09 suite of programs⁶³ and the CCSD(T) calculations are performed using Molpro^{64,65}. The key requirements of the trajectory level of theory are that it is fast (normally necessitates a semi-empirical method) and that it qualitatively orders the possible reaction channels correctly in terms of their kinetic importance. In section S1 of the supporting information we show results from some initial tests where we compared MD runs from two different levels of theory. From these tests we find that with low T_{MD} , the dftb2 method as implemented in DFTB+ is more likely to find the lowest energy pathways when compared to PM6 in Scine. For the "low-level" optimizations in the OR part of the work flow the PM6 method is used, followed by "high-level" optimizations at the B3LYP/3-21G level of theory. Finally, single point energies are calculated at the CCSD(T)-f12/cc-PVDZ level. This hierarchy of levels chosen prioritizes rapid exploration of the reaction network and is of significantly

lower accuracy than the calculations employed to generate the Kinbot surface. However, as we shall show, these levels of theory are perfectly sufficient for ChemDyME to find the important chemistry in the OH + propyne system.

Trajectory Level	DFTB2
Initial Optimization Level	PM6
Refinement Level	B3LYP/3-21G
Single Point Level	CCSD(T)-f12/cc-PVDZ

Table 3: Levels of theory used for the propyne + OH, ChemDyME calculations

The network resulting from the ChemDyME run is shown in Figure 7 and those reactions which were also identified as important by Kinbot are colored in orange (the bold reactions from Figure 1 of the Kinbot work¹) or red (other reactions discovered by Kinbot but not considered kinetically important) and reactions discovered by ChemDyME but not present in the Kinbot network are shown in cyan. This ChemDyME run, tracks the bold reactions found by Kinbot, rapidly forming the wells W4 and W10 before eventually (after many redissociation events back to OH + propyne) forming CH₃ via well skipping reactions across the two intermediates W11 and W1. The x axis marks the logarithmic kinetic time at which each species is visited by the stochastic master equation and species or nodes marked at the inf mark were not visited within the ChemDyME run. This nicely illustrates the principal of kinetic convergence: by tracking the kinetic evolution of the growing network, it is possible to cease the exploration once all the kinetically viable species have been visited, simplifying the overall reaction network.

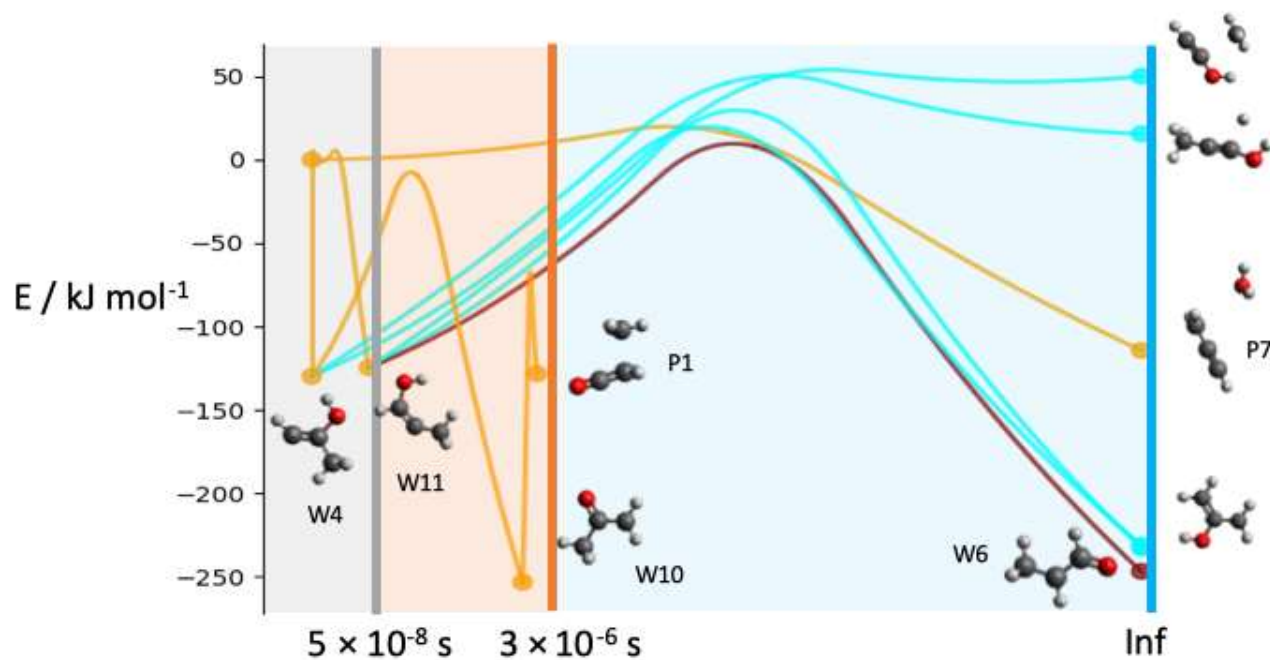


Figure 7: ChemDyME network for the OH + propyne system. The x axis representing the reaction coordinate additionally reflects the time at which each species is first visited in the stochastic master equation and is spaced logarithmically to better view the full range of timescales. The lines delineate and background shading delineate important timescales and the inf label denotes that this species was not visited observed in the stochastic master equation trajectory. Text labels on the different species give the nomenclature for that species from the Kinbot study¹ where available.

Figure 7 shows that ChemDyME prioritizes the bold reactions discovered by Kinbot and inspection of the network reveals that only two of these bold reactions originating from OH + propyne are missed by the above ChemDyME network. Thus far we have used the bold reactions from the Kinbot study as a metric for those which are kinetically important under combustion combustions but upon closer inspection these

bold reactions may be pruned further and the bold two channels “missed” in the ChemDyME network appear to be uncompetitive.

The bold reactions missing from the ChemDyME network are alternative isomerization product channels from intermediate W10. The MD simulations using DFTB2 do not find these product channels, even when running separate calculation with T_{MD} set to 4000K, however both these channels are higher in energy than the dissociation reaction to form P1 are and entropically “tighter” rendering them uncompetitive with the dissociation channel. Indeed, the Kinbot paper, identifies that “The acetyl radical (W10) decomposes exclusively to ketene + CH_3 ”¹. Ignoring the initial initial van der Waals complex found by kinbot, which is discussed below, we argue that ChemDyME does identify all the kinetically important channels from the Kinbot network.

As mentioned above, the other feature missing from the ChemDyME network is the initial van der Waals complex formed between the OH and propyne moieties and the subsequent transition states from this Van der Waals complex. The difficulty of treating van der Waals complexes is a known limitation of ChemDyME currently and this will be addressed in the future. The full master equation xml file resulting from these ChemDyME simulations is in section S3 of the supporting information.

Table 4 compares the energies of the key stationary points in the ChemDyME and Kinbot networks. Here we can see, that despite the much lower level of theory used in the current work, the agreement in energies for the stable species is very good. There are larger discrepancies between the energies of the saddle point structure with the ChemDyME value underestimating the more accurate Kinbot values, although it is noted that the order of the barrier heights is correct in the ChemDyME case. The ChemDyME results could be easily refined through re-optimizing the stationary points with a larger basis, but for the purpose of this work (testing whether ChemDyME finds the important chemistry) the current level of theory is sufficient.

	Current work energy / kJ mol ⁻¹	Kinbot energy / kJ mol ⁻¹
TS PVDW->W4	-	3.51
TS PVDW->W11	-	5.59
TS P->P7	-9.95	8.36
P7	-114.32	-114.20
W4	-124.77	-121.39
W11	-129.94	-130.29
TS W11->W10	-9.37	6.02
W10	-253.32	-250.42
TS W10 ->P1	-70.26	-82.18
P1	-128.31	-127.70
TS W10->W9*	-45.05	-29.09
W9*	-231.13	-227.52
TS W10->W8*	-32.94	-20.69
W8*	-256.46	-251.76

Table 4: Comparison of zero point corrected stationary point energies between the current work (CCSD(T)-f12/cc-PVDZ//B3LYP/3-21G) and the Kinbot study (CCSD(T)-F12b/cc-PVQZ-F12//M06-2X/6-311++G(d,p) . The species marked with asterisks were not found in the initial ChemDyME run, but were found from running separate BXDE trajectories using PM6 rather than dftb2 for forces.

There are a few channels ChemDyME discovers, which are not present in the Kinbot network. Mostly these appear (Figure 7) to have barriers in excess of the Kinbot energy cutoff and thus would have been discounted in the Kinbot study. There are other channels observed for which saddle points couldn't be found and hence they are not shown in Figure 6. The initial adduct W4 is found dynamically to dissociate to P7 with the OH moiety pulling off another hydrogen as it dissociates and W4 and W11 are observed to isomerize directly to one another with the OH moiety dissociating and then rapidly transferring carbons. Further studies are needed to ascertain whether these channels really are operable, and it is unlikely they will significantly affect the wider kinetics of this system; however, this does highlight the potential of a dynamical method to observe roaming type channels.⁷⁷ Recent work has shown that these dynamical

types of reaction, play a larger role in combustion chemistry than previously assumed⁷⁸ and the use of MD in ChemDyME means that such channels may be observed, even when there is no defined transition state for the reaction. Videos showing MD trajectories for these dynamical processes are given in the supplementary information.

5. Exploring the isobutyl + nO₂ surface to for a highly oxygenated molecule (HOM)

The second system we have studied with ChemDyME in the current work, is that of 1-methyl-hexene (1MHE) + ozone (+nO₂)^{79–81} a prototypical oxidation system for the formation of highly oxygenated organic molecules (HOM)^{82–86} in the earth's atmosphere. Such oxidation systems are a key source of atmospheric aerosol.

At the lower temperatures of the earth's atmosphere, bimolecular association reactions are extremely prevalent, and these, typically barrierless, processes are challenging to include in automated approaches. Accurate rate coefficients for a barrierless reaction, usually require variational transition state theory calculations to be performed⁸⁷ and such calculations are complicated by the multireference character of the potential energy surface for associations involving radicals or bimolecular oxygen.

The BXD based sampling of associations used in ChemDyME can in principle, be used to generate rate coefficients. This is something we intend to add to future versions of ChemDyME, however, we recognize that the semiempirical methods used here for the dynamics cannot be relied upon for accurate rate estimations. Instead, we take the pragmatic approach of utilizing the inverse Laplace transform method.⁵⁰ This relies upon the fact that rate coefficients for barrierless associations are usually relatively independent of temperature and have high pressure limiting values in a relatively small window, when

compared to processes which have a defined saddle point.⁸⁸ For the purposes of exploring kinetically feasible reaction channels, we simply wish to know for a given species, whether bimolecular association with some bath gas molecule in the system, is competitive with the fastest of its unimolecular loss process. Since the oxygen concentration in the earth's atmosphere is around 1×10^{18} and the rate coefficients for bimolecular association with oxygen are typically on the order of $1 \times 10^{-12} \text{ cm}^3 \text{ s}^{-1} \text{ molecule}^{-1}$,^{88,89} a typical reaction rate with the bath τ_{O_2} was set to $1 \times 10^6 \text{ s}^{-1}$. This means that any species in the network which exhibited a ME transition time $\tau_{\text{trans}} > \tau_{\text{O}_2}$ was considered a candidate for O_2 addition and BXD association trajectories were performed.

In what follows we discuss results of a ChemDyME run initialized with separated 1-methyl-hexene and ozone moieties. For this run, the “Trajectory Level” and the “Initial Optimization Level” are identical to those used OH + propyne simulations (dftb2 in DFTB+ and PM6 in SCINE Sparrow). The “Refinement Level” is also PM6 however given the open shell singlet structure of the initial part of the network, Gaussian is used for these calculations to ensure the unrestricted, broken symmetry solution to the wavefunction.⁹⁰ CCSD(T) calculations would be extremely time consuming for a system of this potential size so for this run the “Refinement Level” is chosen to be UB3LYP/6-31+G** in Gaussian and again the broken symmetry solution was enforced with the guess=mix keyword for all closed shell species. The master equation temperature and pressure T_{ME} , P_{ME} are set to 298K and 1 atm respectively, representative of the earth's lower atmosphere and t_{max} is set to 1000s to represent the longer timescales at play in the earth's atmosphere. T_{md} is set to 1000 K as a compromise between promoting low energy pathways and speed of reaction finding. The system is then considered to be in a bath of O_2 as described above and whenever a particular species or node exhibits a lifetime of $> 1 \times 10^6 \text{ s}^{-1}$ from a KMC ME run, another oxygen molecule is introduced to the system and potential association pathways are sampled. The input and output files may be found in the examples folder of the ChemDyME repository at <https://github.com/RobinShannon/ChemDyME>

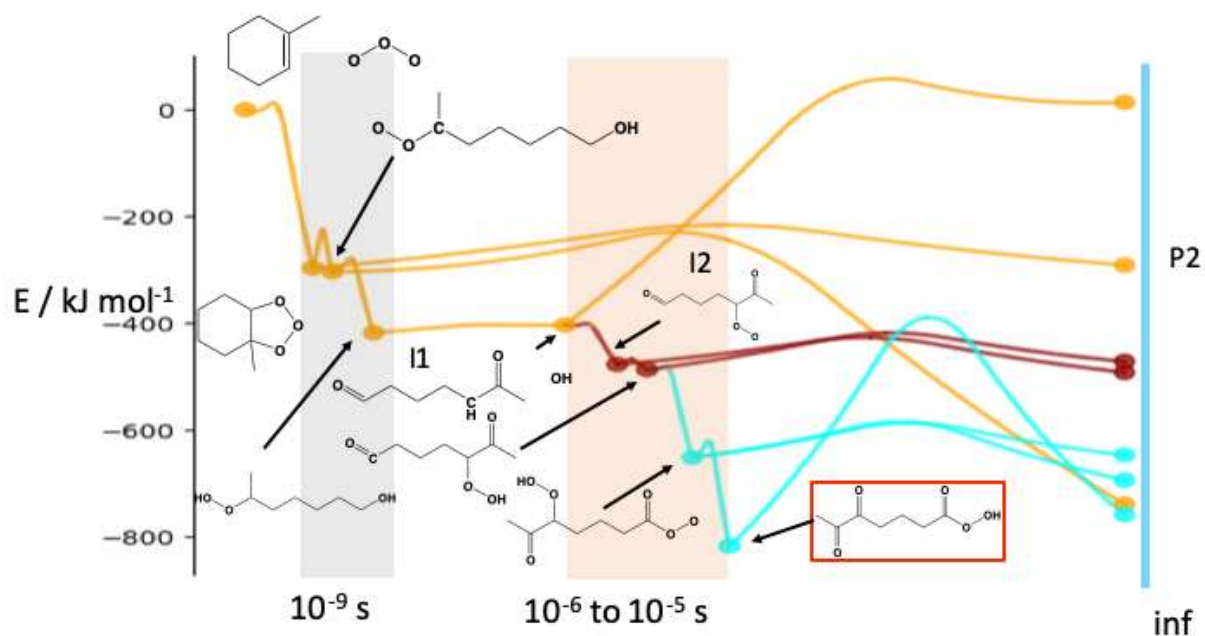


Figure 8: ChemDyME network for the ozone + 1MHE system. The x axis representing the reaction coordinate additionally reflects the time at which each species is first visited in the stochastic master equation however these have been modified slightly to better separate the different nodes. The background shading shows the approximate timescale for different clusters of nodes and the inf label denotes that this species was not visited/observed in the stochastic master equation trajectory. The different colored edges denote the number of O_2 moieties added to the system, with orange = 0, red = 1 and cyan = 2. Labels and the red box highlights the ultimate product species. Key intermediates are labelled. A full figure including all skeletal structures is found in section S2 of the supporting information.

Figure 8 shows the full network from this ChemDyME simulation with each additional O_2 addition colored differently. The ChemDyME run follows much of the expected chemistry. Ozone readily associates with 1MHE forming a 5 membered ring containing the three oxygen atoms. One of the two O-O bonds then

cleaves to form a Criegee intermediate species and subsequent hydrogen transfer leads to OH formation. O₂ can then bind to the lone pair of the resulting radical co-product to form a peroxy radical. Then we observe the standard peroxy radical cycle of hydrogen transfer to a hydroperoxy followed by either dissociation or O₂ addition to the radical center forming a new peroxy radical species.

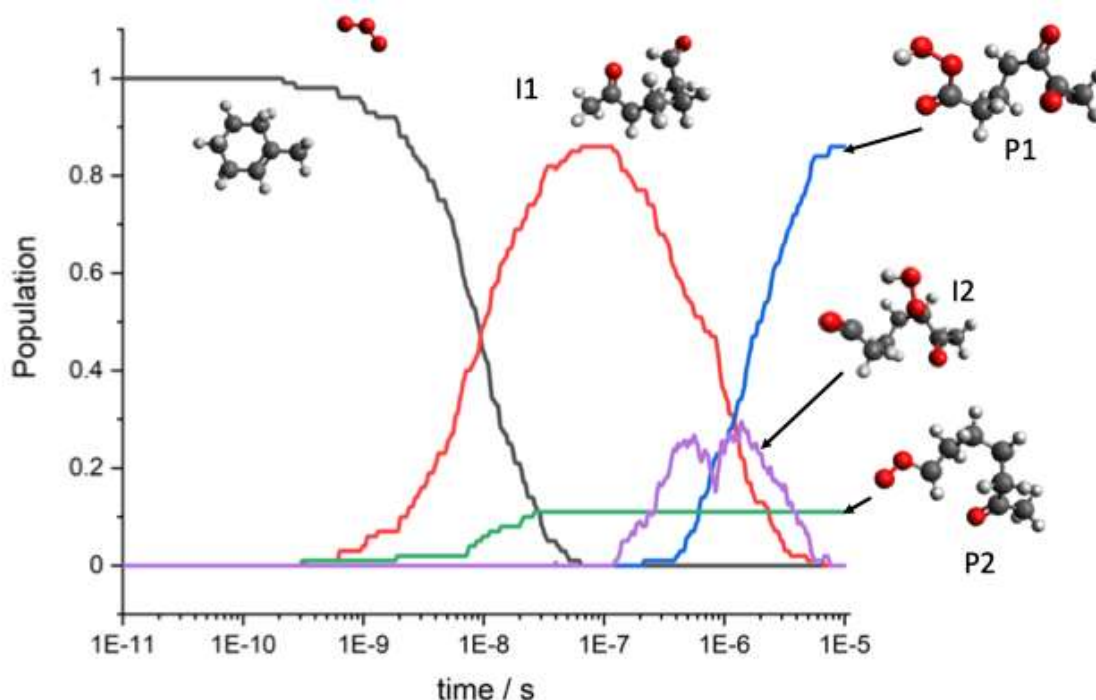


Figure 9: Evolution of key species populations from stochastic master equation simulations. These master equation simulations consist of 100 stochastic trials and a full input file is given in the supporting information

In this single ChemDyME run, after an OH elimination and two O₂ additions the run terminates with a second OH elimination forming the species outlined in red in Figure 8 (P1). Interestingly at the PM6 level the transition state for this reaction appears to be concerted with both H transfer and OH loss occurring simultaneously. A video showing the IRC corresponding to this saddle point is given in the online supporting information.

From the MESMER input file resulting from this single ChemDyME run, it is possible to run a full master equation simulation to track the time evolution of the species in the system. Master equations simulations involving multiple bimolecular additions have rarely been performed^{89,91-95} although there is now an established method for adding bimolecular reaction to activated species in a master equation in such a way as to satisfy detailed balance.^{92,96} Usually, a matrix-based approach would be the preferred method for solving the master equation in MESMER, but given size of the system in terms of species and energy span, we instead use the stochastic BXK algorithm. Figure 9 shows the time evolution of key species in the system from MESMER calculations consisting of 100 stochastic trials. These trials cover a short timescale of up to 10^{-5} s, however considering the channels currently found from the ChemDyME run, this timescale covers the important chemistry of the mechanism in Figure 8. The full MESMER xml file for these master equation simulations is given in section S4 of the supporting information.

Figure 9 demonstrates that two intermediates (I1 and I2) form substantial populations before the majority of the population funnels to the final product P1. A minor channel corresponding to 11% of the total population forms product P2 (Highlighted in green in Figure 7) from an alternate cleavage path of the ozone + 1MHE adduct. Additional ChemDyME runs explore this channel further and in particular find the dominant pathways in Figure 10, however this part of the surface has not been explored in any more detail in the current work as quick gaussian calculations by hand identify that the ground electronic state crosses to triplet upon O₂ addition and the current ChemDyME simulation is set to treat species as either singlet or doublet depending upon the electron number.

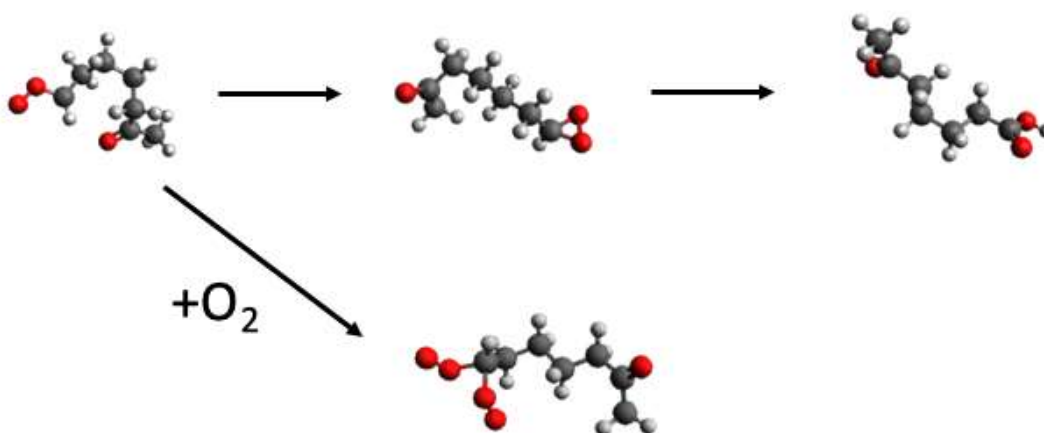


Figure 10: Schematic of key species in an alternate Criegee intermediate pathway found from additional ChemDyME runs.

While the current network captures much of the expected chemistry, more MD trials (n_{MD}), and higher levels of theory would be required for a comprehensive investigation of the 1MHE network, especially given the complexity of the electronic structure. However, within the scope of the current work, these results demonstrate the novel ability of our ChemDyME approach to map the kinetic evolution of a complex system through both dissociation and association processes. Despite the low level of theory used, the ChemDyME simulation captures the chemistry typical of HOM formation and atmospheric oxidation and despite a combinatorial explosion of possible reactions (given the size of the system and the prospect of multiple O_2 additions) the kinetic convergence criterion helps to identify the important chemistry at the level of theory used.

6. Conclusions

In this work we demonstrate the capabilities of the ChemDyME approach for exploring the kinetics of reactive mechanism in an automated manner. Through the combined use of molecular dynamics and master equation simulations, ChemDyME narrows the search for chemical reactions, prioritizing those which are kinetically the most important under the conditions of interest. Through benchmark calculations against an OH + propyne combustion network which has already been pruned to neglect high energy pathways, we demonstrate that ChemDyME prunes this network further to find the key reactions.

By utilizing the concept of “kinetic convergence” ChemDyME is able to add additional molecular moieties to the system, when it is deemed kinetically favorable. We demonstrate this with a prototypical ozonolysis / oxidation system and show that ChemDyME is capable of tracking the network through both the loss of an OH radical and through multiple subsequent O₂ additions to form an example highly oxygenated molecule (HOM). The ChemDyME output produces all the input required to run a master equation on the system, and we present the time evolution of an activated oxidation system of unprecedented size giving the timescale of HOM formation.

This work demonstrates the strengths of the ChemDyME approach and of BXDE in general. Previous work using BXDE has shown its utility in accelerating reactive MD whilst preserving some sense of the important low energy chemistry. By harnessing both BXDE MD and the energy grained master equation, this work shows that ChemDyME can be used to heavily simplify the reaction of network of interest, prioritizing the discovery of the small subset of reactions which are kinetically important at the conditions of interest. ChemDyME also has the benefit of being a dynamical approach which can give a wider picture of the chemical reactivity of a system than gained from knowing the key transition states alone. Finally, the dynamical nature of ChemDyME also makes the approach highly generalizable compared to heuristic based approaches.

Naturally there are also weaknesses to the ChemDyME approach. Currently ChemDyME cannot treat Van der Waals complexes and the reaction finding algorithm does not always identify when multiple reactions have occurred. Advances in both areas have been in the AutoMeKin software and these improvements are planned for future releases of ChemDyME.

7. Acknowledgements

This work was undertaken on ARC3, part of the High Performance Computing facilities at the University of Leeds, UK and using the computational facilities of the Advanced Computing Research Centre, University of Bristol - <http://www.bris.ac.uk/acrc/>. RJS and DRG acknowledge support of this work through the “CHAMPS” EPSRC program grant EP/P021123/1. D.R.G. acknowledges funding from the Royal Society as a University Research Fellow

8. References

1. Zádor, J. & Miller, J. A. Adventures on the C₃H₅O potential energy surface: OH + propyne, OH + allene and related reactions. *Proceedings of the Combustion Institute* **35**, 181–188 (2015).
2. Zheng, S. & Pfaendtner, J. Car-Parrinello molecular dynamics + metadynamics study of high-temperature methanol oxidation reactions using generic collective variables. *Journal of Physical Chemistry C* **118**, 10764–10770 (2014).

3. Kreitz, B., Wehinger, G. D., Goldsmith, C. F. & Turek, T. Development of a Microkinetic Model for the CO₂ Methanation with an Automated Reaction Mechanism Generator. in *Computer Aided Chemical Engineering* vol. 48 529–534 (Elsevier B.V., 2020).
4. Goldsmith, C. F. & West, R. H. Automatic Generation of Microkinetic Mechanisms for Heterogeneous Catalysis. *Journal of Physical Chemistry C* **121**, 9970–9981 (2017).
5. Broadbelt, L. J., Stark, S. M. & Klein, M. T. Computer Generated Pyrolysis Modeling: On-the-Fly Generation of Species, Reactions, and Rates. *Industrial and Engineering Chemistry Research* **33**, 790–799 (1994).
6. Jara-Toro, R. A., Pino, G. A., Glowacki, D. R., Shannon, R. J. & Martínez-Núñez, E. Enhancing Automated Reaction Discovery with Boxed Molecular Dynamics in Energy Space. *ChemSystemsChem* **2**, (2020).
7. Bowman, G. R., Beauchamp, K. A., Boxer, G. & Pande, V. S. Progress and challenges in the automated construction of Markov state models for full protein systems. *Journal of Chemical Physics* **131**, (2009).
8. Varela, J. A., Vázquez, S. A. & Martínez-Núñez, E. An automated method to find reaction mechanisms and solve the kinetics in organometallic catalysis. *Chemical Science* **8**, 3843–3851 (2017).
9. Robertson, C. & Habershon, S. Fast screening of homogeneous catalysis mechanisms using graph-driven searches and approximate quantum chemistry †. *Cite this: Catal. Sci. Technol* **9**, 6357–6369 (2019).
10. Maeda, S. & Morokuma, K. Toward predicting full catalytic cycle using automatic reaction path search method: A case study on HCo(CO)₃-catalyzed hydroformylation. *Journal of Chemical Theory and Computation* **8**, 380–385 (2012).

11. Simm, G. N., Vaucher, A. C. & Reiher, M. Exploration of Reaction Pathways and Chemical Transformation Networks. *Journal of Physical Chemistry A* **123**, 385–399 (2019).
12. Gao, C. W., Allen, J. W., Green, W. H. & West, R. H. Reaction Mechanism Generator: Automatic construction of chemical kinetic mechanisms. *Computer Physics Communications* **203**, 212–225 (2016).
13. Rodríguez, A. *et al.* tsscds2018: A code for automated discovery of chemical reaction mechanisms and solving the kinetics. *Journal of Computational Chemistry* **39**, 1922–1930 (2018).
14. van de Vijver, R. & Zádor, J. KinBot: Automated stationary point search on potential energy surfaces. *Computer Physics Communications* **248**, 106947 (2020).
15. Broadbelt, L. J. & Pfaendtner, J. Lexicography of kinetic modeling of complex reaction networks. *AIChE Journal* **51**, 2112–2121 (2005).
16. Maeda, S., Ohno, K. & Morokuma, K. This journal is c the Owner Societies. *Phys. Chem. Chem. Phys* **15**, 3683 (2013).
17. Wang, L.-P. *et al.* Discovering chemistry with an *ab initio* nanoreactor. *NATURE CHEMISTRY* | **6**, (2014).
18. Unsleber, J. P. & Reiher, M. The Exploration of Chemical Reaction Networks. *Annual Review of Physical Chemistry* **71**, 121–142 (2020).
19. Dewyer, A. L., Argüelles, A. J. & Zimmerman, P. M. Methods for exploring reaction space in molecular systems. *Wiley Interdisciplinary Reviews: Computational Molecular Science* vol. 8 (2018).
20. Simm, G. N., Vaucher, A. C. & Reiher, M. Exploration of Reaction Pathways and Chemical Transformation Networks. *Journal of Physical Chemistry A* **123**, 385–399 (2019).

21. Danilack, A. D., Klippenstein, S. J., Georgievskii, Y. & Goldsmith, C. F. Low-temperature oxidation of diethyl ether: Reactions of hot radicals across coupled potential energy surfaces. *Proceedings of the Combustion Institute* (2020) doi:10.1016/j.proci.2020.07.111.
22. Vaucher, A. C. & Reiher, M. Minimum Energy Paths and Transition States by Curve Optimization. (2018) doi:10.1021/acs.jctc.8b00169.
23. Henkelman, G., Uberuaga, B. P. & Jónsson, H. Climbing image nudged elastic band method for finding saddle points and minimum energy paths. *Journal of Chemical Physics* **113**, 9901–9904 (2000).
24. Zimmerman, P. Reliable transition state searches integrated with the growing string method. *Journal of Chemical Theory and Computation* **9**, 3043–3050 (2013).
25. Martínez-Núñez, E. An automated method to find transition states using chemical dynamics simulations. *Journal of Computational Chemistry* **36**, 222–234 (2015).
26. Husic, B. E. & Pande, V. S. Markov State Models: From an Art to a Science. *Journal of the American Chemical Society* vol. 140 2386–2396 (2018).
27. Gao, C. W., Allen, J. W., Green, W. H. & West, R. H. Reaction Mechanism Generator: Automatic construction of chemical kinetic mechanisms. *Computer Physics Communications* **203**, 212–225 (2016).
28. Rissanen, M. P. *et al.* The formation of highly oxidized multifunctional products in the ozonolysis of cyclohexene. *Journal of the American Chemical Society* **136**, 15596–15606 (2014).
29. Bianchi, F. *et al.* Highly Oxygenated Organic Molecules (HOM) from Gas-Phase Autoxidation Involving Peroxy Radicals: A Key Contributor to Atmospheric Aerosol. *Chemical Reviews* vol. 119 3472–3509 (2019).

30. Warmflash, A., Bhimalapuram, P. & Dinner, A. R. Umbrella sampling for nonequilibrium processes. *Journal of Chemical Physics* **127**, 154112 (2007).
31. Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Physical Review Letters* **100**, 020603 (2008).
32. Bolhuis, P. G., Chandler, D., Dellago, C. & Geissler, P. L. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annual Review of Physical Chemistry* **53**, 291–318 (2002).
33. Faradjian, A. K. & Elber, R. Computing time scales from reaction coordinates by milestoning. *Journal of Chemical Physics* **120**, 10880–10889 (2004).
34. Vanden-Eijnden, E. & Venturoli, M. Markovian milestoning with Voronoi tessellations. *Journal of Chemical Physics* **130**, 194101 (2009).
35. Passerone, D. & Parrinello, M. Action-derived molecular dynamics in the study of rare events. *Physical Review Letters* **87**, 108302/1-108302/4 (2001).
36. Allen, R. J., Frenkel, D. & ten Wolde, P. R. Simulating rare events in equilibrium or nonequilibrium stochastic systems. *Journal of Chemical Physics* **124**, (2006).
37. Allen, R. J., Warren, P. B. & ten Wolde, P. R. Sampling rare switching events in biochemical networks. *Physical Review Letters* **94**, (2005).
38. van Erp, T. S., Moroni, D. & Bolhuis, P. G. A novel path sampling method for the calculation of rate constants. *Journal of Chemical Physics* **118**, 7762–7774 (2003).
39. Pártay, L. B., Bartók, A. P. & Csányi, G. Efficient sampling of atomic configurational spaces. *Journal of Physical Chemistry B* **114**, 10502–10512 (2010).

40. Warmflash, A., Bhimalapuram, P. & Dinner, A. R. Umbrella sampling for nonequilibrium processes. *Journal of Chemical Physics* **127**, (2007).
41. Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Physical Review Letters* **100**, (2008).
42. Grimme, S. Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. *Journal of Chemical Theory and Computation* **15**, 2847–2862 (2019).
43. 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. *Journal of Chemical Theory and Computation* **14**, 4541–4552 (2018).
44. Martínez-Núñez, E. & Shalashilin, D. v. Acceleration of classical mechanics by phase space constraints. *Journal of Chemical Theory and Computation* **2**, 912–919 (2006).
45. Booth, J. *et al.* Recent applications of boxed molecular dynamics: A simple multiscale technique for atomistic simulations. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* vol. 372 (2014).
46. Glowacki, D. R., Paci, E. & Shalashilin, D. v. Boxed molecular dynamics: Decorrelation time scales and the kinetic master equation. *Journal of Chemical Theory and Computation* **7**, 1244–1252 (2011).
47. O'Connor, M., Paci, E., McIntosh-Smith, S. & Glowacki, D. R. Adaptive free energy sampling in multidimensional collective variable space using boxed molecular dynamics. *Faraday Discussions* **195**, 395–419 (2016).

48. Weininger, D., Weininger, A. & Weininger, J. L. SMILES. 2. Algorithm for Generation of Unique SMILES Notation. *Journal of Chemical Information and Computer Sciences* **29**, 97–101 (1989).
49. Brunken, C. *et al.* qcscine/readuct: Release 2.0.0. (2020) doi:10.5281/ZENODO.3768539.
50. Robertson, S. H., Pilling, M. J., Baulch, D. L. & Green, N. J. B. Fitting of pressure-dependent kinetic rate data by master equation/inverse Laplace transform analysis. *Journal of Physical Chemistry* **99**, 13452–13460 (1995).
51. Glowacki, D. R., Liang, C. H., Morley, C., Pilling, M. J. & Robertson, S. H. MESMER: An open-source master equation solver for Multi-Energy well reactions. *Journal of Physical Chemistry A* **116**, 9545–9560 (2012).
52. Pilling, M. J. & Robertson, S. H. Master Equation Models for Chemical Reactions of Importance in Combustion. *Annual Review of Physical Chemistry* **54**, 245–275 (2003).
53. Gillespie, D. T. *A rigorous derivation of the chemical master equation*. *Physica A* vol. 188 (1992).
54. Vereecken, L., Huyberechts, G. & Peeters, J. Stochastic simulation of chemically activated unimolecular reactions. *Journal of Chemical Physics* **106**, 6564–6573 (1997).
55. Klippenstein, S. J. & Miller, J. A. From the time-dependent, multiple-well master equation to phenomenological rate coefficients. *Journal of Physical Chemistry A* **106**, 9267–9277 (2002).
56. Robertson, S. H. Foundations of the master equation. in vol. 43 299–361 (Elsevier, 2019).
57. Barker, J. R. Multiple-Well, multiple-path unimolecular reaction systems. I. MultiWell computer program suite. *International Journal of Chemical Kinetics* **33**, 232–245 (2001).

58. Georgievskii, Y., Miller, J. A., Burke, M. P. & Klippenstein, S. J. Reformulation and solution of the master equation for multiple-well chemical reactions. *Journal of Physical Chemistry A* **117**, 12146–12154 (2013).
59. Hjorth Larsen, A. *et al.* The atomic simulation environment - A Python library for working with atoms. *Journal of Physics Condensed Matter* vol. 29 273002 (2017).
60. Bosia, F., Husch, T., Vaucher, A. C. & Reiher, M. qcscine/sparrow: Release 2.0.1. (2020) doi:10.5281/ZENODO.3907313.
61. Hourahine, B. *et al.* DFTB+, a software package for efficient approximate density functional theory based atomistic simulations. *Journal of Chemical Physics* **152**, 124101 (2020).
62. Bannwarth, C., Ehlert, S. & Grimme, S. GFN2-xTB - An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *Journal of Chemical Theory and Computation* **15**, 1652–1671 (2019).
63. Frisch, M. J. *et al.* Gaussian~09 Revision D.01.
64. Werner, H.-J., Knowles, P. J., Knizia, G., Manby, F. R. & Schütz, M. Molpro: a general-purpose quantum chemistry program package. *WIREs Comput Mol Sci* **2**, 242–253 (2012).
65. Werner, H.-J. *et al.* MOLPRO, version 2019.2, a package of ab initio programs. (2019).
66. Aprà, E. *et al.* NWChem: Past, present, and future. *Journal of Chemical Physics* **152**, 184102 (2020).
67. Refaey, K. M. A. & Chupka, W. A. Photoionization of the Lower Aliphatic Alcohols with Mass Analysis. *The Journal of Chemical Physics* **48**, 5205–5219 (1968).

68. Miller, J. A. & Klippenstein, S. J. The recombination of propargyl radicals: Solving the master equation. *Journal of Physical Chemistry A* **105**, 7254–7266 (2001).
69. Bradley, J. N., Hack, W., Hoyermann, K. & Wagner, H. G. Kinetics of the reaction of hydroxyl radicals with ethylene and with C₃ hydrocarbons. *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases* **69**, 1889–1898 (1973).
70. Hatakeyama, S., Washida, N. & Akimoto, H. Rate constants and mechanisms for the reaction of OH (OD) radicals with acetylene, propyne, and 2-butyne in air at 297 ± 2 K. *Journal of Physical Chemistry* **90**, 173–178 (1986).
71. Taylor, S. E., Goddard, A., Blitz, M. A., Cleary, P. A. & Heard, D. E. Pulsed Laval nozzle study of the kinetics of OH with unsaturated hydrocarbons at very low temperatures. *Physical Chemistry Chemical Physics* **10**, 422–437 (2008).
72. Boodaghians, R. B., Hall, I. W., Toby, F. S. & Wayne, R. P. Absolute determinations of the kinetics and temperature dependences of the reactions of OH with a series of alkynes. *Journal of the Chemical Society, Faraday Transactions 2: Molecular and Chemical Physics* **83**, 2073–2080 (1987).
73. Atkinson, R. & Aschmann, S. M. Kinetics of the gas phase reaction of Cl atoms with a series of organics at 296 ± 2 K and atmospheric pressure. *International Journal of Chemical Kinetics* **17**, 33–41 (1985).
74. Kwon, H. C. *et al.* Crossed beam investigations of the reaction dynamics of O(³P) with allyl radical, C₃H₅. *Journal of Chemical Physics* **116**, 2675–2679 (2002).
75. Park, J. H., Lee, H. & Choi, J. H. A theoretical study of the reaction of O(³P) with an allyl radical C₃H₅. *Journal of Chemical Physics* **119**, 8966–8978 (2003).

76. Fitzpatrick, B. Theoretical Study of Isomerization and Dissociation Transition States of C₃H₅O Radical Isomers: Ab Initio Characterization of the Critical Points and Statistical Transition-State Theory Modeling of the Dynamics. *J. Phys. Chem. A* **115**, 1701–1712 (2011).
77. Bowman, J. M. & Shepler, B. C. Roaming Radicals. *Annual Review of Physical Chemistry* **62**, 531–553 (2011).
78. West, R. H. & Goldsmith, C. F. The impact of roaming radicals on the combustion properties of transportation fuels. *Combustion and Flame* **194**, 387–395 (2018).
79. Mentel, T. F. *et al.* Formation of highly oxidized multifunctional compounds: autoxidation of peroxy radicals formed in the ozonolysis of alkenes—deduced from structure-product relationships. *Atmos. Chem. Phys* **15**, 6745–6765 (2015).
80. Rissanen, M. P. *et al.* Effects of chemical complexity on the autoxidation mechanisms of endocyclic alkene ozonolysis products: From methylcyclohexenes toward understanding α -pinene. *Journal of Physical Chemistry A* **119**, 4633–4650 (2015).
81. Rissanen, M. P. *et al.* The Formation of Highly Oxidized Multifunctional Products in the Ozonolysis of Cyclohexene. (2014) doi:10.1021/ja507146s.
82. Crouse, J. D., Nielsen, L. B., Jørgensen, S., Kjaergaard, H. G. & Wennberg, P. O. Autoxidation of Organic Compounds in the Atmosphere. (2013) doi:10.1021/jz4019207.
83. Hallquist, M. *et al.* The formation, properties and impact of secondary organic aerosol: Current and emerging issues. *Atmospheric Chemistry and Physics* **9**, 5155–5236 (2009).
84. Donahue, N. M., Robinson, A. L. & Pandis, S. N. Atmospheric organic particulate matter: From smoke to secondary organic aerosol. *Atmospheric Environment* **43**, 94–106 (2009).

85. Jimenez, J. L. *et al.* Evolution of organic aerosols in the atmosphere. *Science* **326**, 1525–1529 (2009).
86. Bianchi, F. *et al.* Highly Oxygenated Organic Molecules (HOM) from Gas-Phase Autoxidation Involving Peroxy Radicals: A Key Contributor to Atmospheric Aerosol. (2019) doi:10.1021/acs.chemrev.8b00395.
87. Georgievskii, Y. & Klippenstein, S. J. Transition state theory for multichannel addition reactions: Multifaceted dividing surfaces. *Journal of Physical Chemistry A* **107**, 9776–9781 (2003).
88. Zádor, J., Taatjes, C. A. & Fernandes, R. X. Kinetics of elementary reactions in low-temperature autoignition chemistry. *Progress in Energy and Combustion Science* vol. 37 371–421 (2011).
89. Glowacki, D. R. *et al.* Interception of excited vibrational quantum states by O₂ in atmospheric association reactions. *Science* **337**, 1066–1069 (2012).
90. Vereecken, L., Harder, H. & Novelli, A. The reactions of Criegee intermediates with alkenes, ozone, and carbonyl oxides. *Physical Chemistry Chemical Physics* **16**, 4039–4049 (2014).
91. Goldsmith, C. F., Burke, M. P., Georgievskii, Y. & Klippenstein, S. J. Effect of non-thermal product energy distributions on ketohydroperoxide decomposition kinetics. *Proceedings of the Combustion Institute* **35**, 283–290 (2015).
92. Burke, M. P., Goldsmith, C. F., Georgievskii, Y. & Klippenstein, S. J. Towards a quantitative understanding of the role of non-Boltzmann reactant distributions in low temperature oxidation. *Proceedings of the Combustion Institute* **35**, 205–213 (2015).
93. Maranzana, A., Barker, J. R. & Tonachini, G. Master equation simulations of competing unimolecular and bimolecular reactions: Application to OH production in the reaction of acetyl radical with O₂. *Physical Chemistry Chemical Physics* **9**, 4129–4141 (2007).

94. Pfeifle, M. & Olzmann, M. Consecutive chemical activation steps in the OH-initiated atmospheric degradation of isoprene: An analysis with coupled master equations. *International Journal of Chemical Kinetics* **46**, 231–244 (2014).
95. Shannon, R. J., Robertson, S. H., Blitz, M. A. & Seakins, P. W. Bimolecular reactions of activated species: An analysis of problematic HC(O)C(O) chemistry. *Chemical Physics Letters* **661**, 58–64 (2016).
96. Green, N. J. B. & Robertson, S. H. General master equation formulation of a reversible dissociation/association reaction. *Chemical Physics Letters* **605–606**, 44–46 (2014).

Supporting information

S1. Comparison of BXDE trajectory products at different temperatures and with different methods.

Figures S1 and S2 show the product species from reactive BXDE trajectories initialized from intermediate W4 in the propyne + OH mechanism. Figure S1 compares products from DFTB2 (red writing) and PM6 (black writing) trajectories (from SCINE Sparrow and DFTB+ respectively) with a Langevin thermostat set to 1000K whilst Figure S2 compare the same methods but with the thermostat set to 4000K. The species circled in red (intermediate W10) is the reaction channel with the lowest barrier and it can be seen that whilst DFTB2 finds this product channel at both temperatures, PM6 trajectories fail to observe it at all. PM6 trajectories at low temperatures also miss another hydrogen transfer channel, instead favoring dissociative product channels. When comparing different temperatures, it is found that 4000 K MD simulations find a much larger array of exotic dissociation channels, but are less likely to observe the low

barriered but entropically tight, hydrogen transfer reactions which dominate the kinetics of the system under combustion conditions.

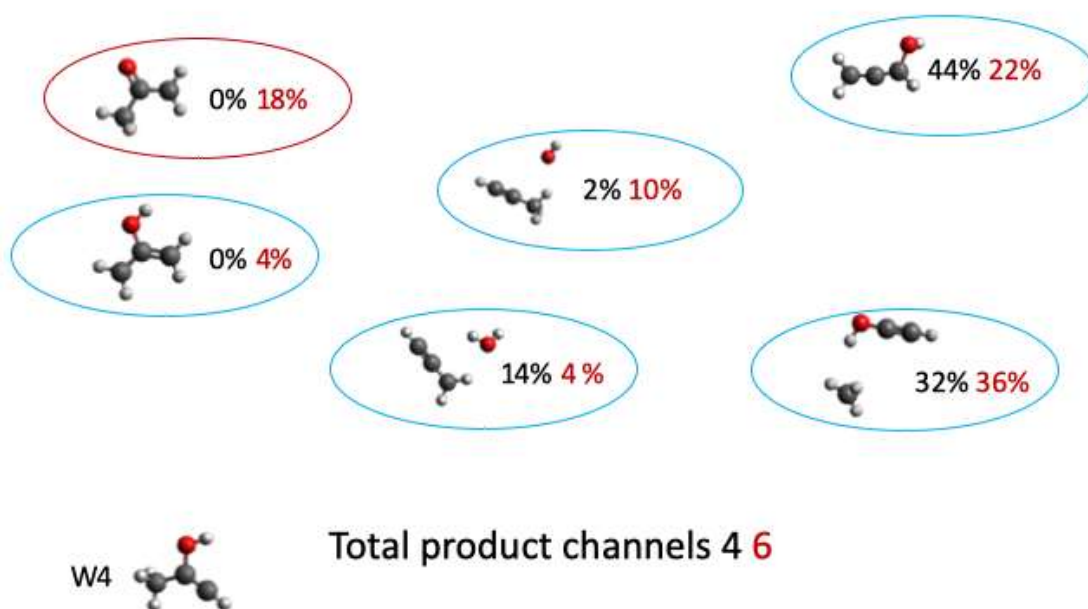


Figure S1: Products from 50 reactive BXDE trajectories starting from species W4 with a Langevin thermostat set to 1000K. Red writing indicates DFTB2 simulations and black writing indicates PM6. The red circled product corresponds to the lowest energy barrier.

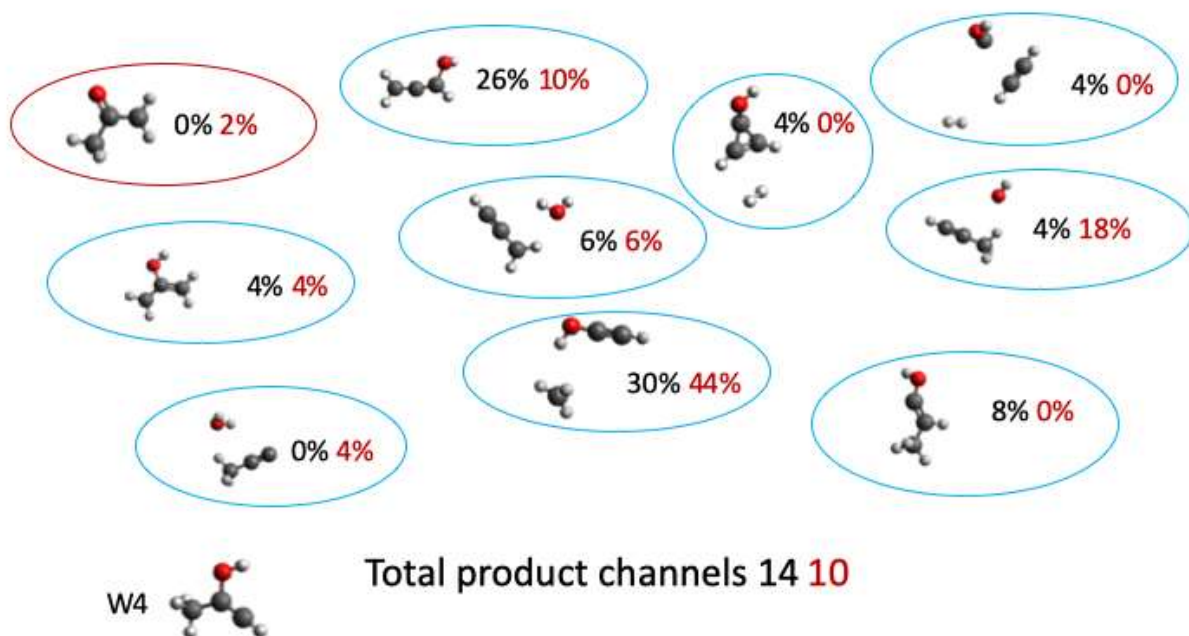


Figure S2: Products from 50 reactive BXDE trajectories starting from species W4 with a Langevin thermostat set to 4000K. Red writing indicates DFTB2 simulations and black writing indicates PM6. The red circled product corresponds to the lowest energy barrier.

S2. More detailed mechanism the 1MHE + ozone network

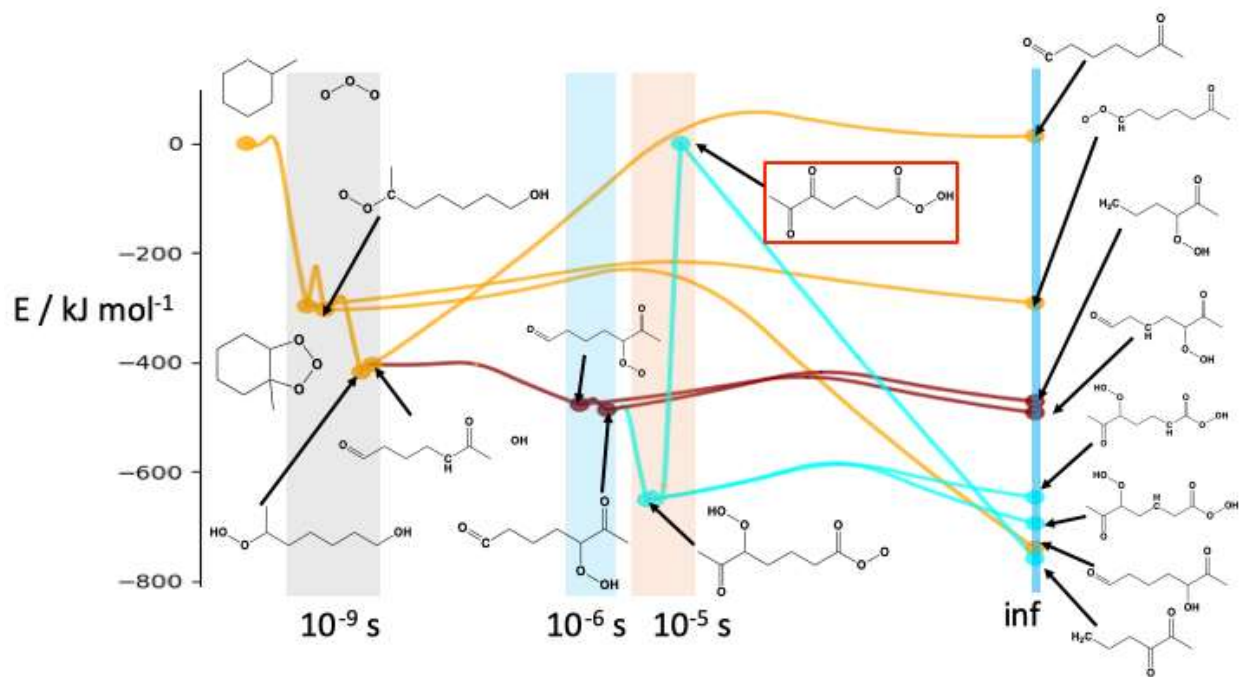


Figure S2: ChemDyME network for the ozone + 1MHE system. The invisible y axis represents the approximate time at which each species is first visited in the stochastic master equation however these have been modified slightly to better separate the different nodes. The background shading shows the approximate timescale for different clusters of nodes and the inf label denotes that this species was not visited/observed in the stochastic master equation trajectory. The different colored edges denote the number of O_2 moieties added to the system, with orange = 0, red = 1 and cyan = 2. Labels and the red box highlights the ultimate product species. Key intermediates are labelled.

S3. OH + propyne MESMER input

S4 1MHE + Ozone MESMER input

