Automatic Proposal of Multi-Step Reaction Mechanisms Using a Graph-Driven Search

Idil Ismail, Holly Stuttaford-Fowler, Curtis Ochan-Ashok, Christopher Robertson, and Scott Habershon*

Department of Chemistry and Centre for Scientific Computing, University of Warwick, Coventry, CV4 7AL, United Kingdom

E-mail: S.Habershon@warwick.ac.uk

Abstract

Proposing and testing mechanistic hypotheses stands as one of the key applications of contemporary computational chemistry. In the majority of computational mechanistic analyses, the individual elementary steps leading from reactants to products are proposed by the user, based on learnt chemical knowledge, intuition, or comparison to an existing well-characterized mechanism for a closely-related chemical reaction. However, the pre-requisite of prior chemical knowledge is a barrier to automated (or 'black box') mechanistic generation and assessment, and may simultaneously preclude mechanistic proposals which lie outside the 'standard' chemical reaction set. In this Article, we propose a simple random-walk algorithm that searches for the set of elementary chemical reactions which transform defined reactant structures into target products. Our approach operates exclusively in the space of molecular connectivity matrices, seeking out the set of chemically-sensible bonding changes which link connectivity matrices for input reactant and product structures. We subsequently illustrate how atomic coordinates for each elementary reaction can be generated under the action of a graph-restraining potential, prior to further analysis by quantum chemical calculations. Our approach is successfully demonstrated for carbon monoxide oxidation, the water-gas shift reaction, and n-hexane aromatization, all catalyzed by Pt nanoparticles.

1 Introduction

A key application of contemporary computational chemistry lies in analyzing the elementary steps of complex chemical reaction mechanisms associated with processes such as catalysis,[1–5] polymerization[6] and combustion.[7–9] Given initial reactant and product molecular configurations (*i.e.* atomic coordinates), *ab initio* electronic structure methods can now be straightforwardly used to perform geometry optimization and normal-mode analysis. The resulting optimized molecular geometries and corresponding normal-mode vibrational coordinates can then be used within the standard rigid-rotor/harmonic oscillator model to evaluate the free energies of the reactant and products, making direct connection to experimentally-observable thermodynamic quantities such as reaction free energies.[10,11] Furthermore, reaction-path analysis tools, such as nudged elastic band (NEB[12–14]), the dimer method[15] and the growing-string method[16,17] can be used to seek out minimum-energy paths (MEPs) connecting known reactant and product structures, providing insight into the reaction mechanism of elementary steps. Finally, transition-state (TS) searching algorithms[17–20] can be used to identify TS configurations; free energy evaluations for the TS then enables determination of activation free energies, which can be used within transition-state theory (TST) to approximate reaction rates.[10,11]

While the procedure described above is well-defined for a single elementary chemical reaction step, the challenge of investigating multi-step reaction mechanisms is much greater. Here, to analyze a multi-step chemical reaction, one must first generate a mechanistic proposal, namely the sequence of elementary reaction-steps which comprise the full mechanism, as well as the identities of the participating atoms and molecules. Generating molecular models for each elementary reaction step "by hand" is a tedious task, and leaves open the question of whether the proposed reaction mechanism is truly representative of experimental facts, or whether it is somehow biased by the user's own intuition.

An alternative approach is to develop new computer algorithms which can be used to propose multi-step reaction mechanisms in an automated, black-box manner, without re-

quiring user-guidance; this is the key goal of this Article. The last few years have seen a major growth in methods which aim to construct kinetic networks describing complex chemical systems in an automated manner.[3,17,21–32] A prominent example, particularly in relation to combustion kinetics, is the Reaction Mechanism Generator (RMG) scheme,[21] which uses the idea of molecular connectivity matrices (or graphs) in combination with chemical reaction rules which can be used to build complex reaction networks and simultaneously predict thermodynamic and kinetic parameters for the elementary reaction steps. The RMG scheme has been used to construct a wide variety of kinetic models describing complex combustion and pyrolysis processes,[7–9,21] with the resulting models being used to predict emergent rate laws and product concentrations by direct microkinetic simulations. However, the approach taken in RMG is to start from a "seed" mechanism and iteratively construct a kinetic model using stopping criteria based on predicted reaction rates; this is somewhat different to the aim of the methodology proposed herein, which aims at predicting reaction mechanisms which definitively connect two known reaction end-points. In addition, RMG has not, to date, been used to study catalysis by molecular species such as organometallic complexes or nanoparticles (although recent application to heterogeneous catalysis has been reported[33]); it is exactly these molecular application domains which we are interested in here, as described further below.

In automatically generating complex kinetic networks for molecular catalysis, a number of computational schemes have been suggested recently. For example, a combination of molecular dynamics-based schemes for transition-state searching with graph-based tools for molecular identification has been shown to be effective in modelling catalytic processes.[34] On a similar graph-based theme, our own recent work[35,36] has shown how one can construct an automated scheme for reaction sampling by treating reaction-paths as dynamic objects associated with a classical Hamiltonian incorporating an effective potential which "imprints" a given connectivity matrix onto sampled molecular configurations; by periodically updating the connectivity matrices associated with the reaction end-points using chemical reaction

rules, this scheme can be used to construct reaction networks for subsequent microkinetic modeling, as demonstrated in application to cobalt-catalyzed hydroformylation. Further related graph-based schemes include the ZStruct system proposed by Zimmerman,[30] employing an efficient growing-string method for transition-state searching,[17] or the recent work by Kim and coworkers, employing graph-based heuristics to rapidly populate reaction networks.[32] Other schemes in this domain include the artificial force-induced reaction (AFIR) approach,[27,28] which supplements *ab initio* calculations with an artificial potential energy surface to overcome reaction barriers, and the *ab initio* nanoreactor approach of Martinez and coworkers,[23] which combines *ab initio* molecular dynamics with an artificial piston to push reactive molecules together to accelerate reactions. These and other related approaches have been reviewed recently.[22,37]

However, to the best of our knowledge, none of the schemes described above have been employed to directly identify "double-ended" reaction pathways which connect well-defined reactants and products. In this Article, we show how this challenge can be addressed by transformation into an optimization problem which can be readily approached using established methods such as simulated annealing (SA) or genetic algorithms. Key to our approach is the idea that connectivity (or adjacency) matrices can be used to discretize chemical space into molecular species, such that chemical reactions then correspond to moves between chemical isomers defined by different connectivity matrices.[21,30–32,35,36] Then, given input molecular structures for reactant and product molecules, as well as a library of possible connectivity matrix moves corresponding to generic chemical transformations such as dissociation, insertion and so on, we show that the identification of a many-step reaction mechanism connecting reactants and products can be formulated as the search for a sequence of "graph moves" which transform the connectivity matrix of reactants into the connectivity matrix of the products. Finally, we demonstrate how geometry optimization under a graph-restraining potential energy surface, employed in our previous reaction discovery simulations, can be used to generate Cartesian-space reaction-paths for each individual elementary step in the

predicted sequence of reactions; these reaction-paths can then be subjected to standard thermodynamic (*e.g.* reaction free energies) and kinetic (*e.g.* TST rates) analysis in combination with *ab initio* quantum chemical calculations.

The remainder of this Article is organized as follows. First, in Section 2, we describe our connectivity-matrix-based approach for finding double-ended reaction paths connecting defined reactants and products. Subsequently, in Section 3 we successfully demonstrate our double-ended mechanism search strategy to find reaction paths for oxidation of carbon monoxide, the water-gas shift reaction, and aromatization of n-hexane, all in the presence of a platinum nanoparticle. Finally, in Section 4, we conclude by highlighting some interesting extensions of our approach in the context of molecular catalyst design.

2 Theory

As noted above, we have recently shown how the idea of connectivity matrices (CMs) can be used as the basis of an algorithm to automatically explore chemical reaction networks by constructing a classical Hamiltonian which describes a continuous reaction-path (described by either Fourier coefficients or discrete "images"). In this approach,[35,36] an artificial potential energy surface (PES) is introduced, referred to hereafter as a graph-restraining potential (GRP), to restrain the reaction-path end-points to sample *only* those molecular configurations which correspond to defined CMs; by then introducing changes to the end-point CMs, the GRP then enforces sampling of distinct chemical reactions, allowing automated construction of kinetic networks when combined with *ab initio* quantum chemistry and TST.

In this Article, we show how a similar idea, namely the discretization of chemical-space using CMs, can be employed as the basis of a search algorithm to identify reaction mechanisms connecting user-defined reactants and products. The reactant and product molecular structures are first transformed into CMs based on bonding distance cut-offs, as described below. Changes in chemical bonding correspond to changes in the elements of the CMs; the

search for a chemical reaction mechanism connecting reactants and products can then be viewed as the problem of finding the sequence of "chemically-allowed" CM updates which transforms the reactant CM into the product CM. Finally, using the input molecular structure of the reactants, the sequence of CM updates, and the idea of optimization under the GRP, we show how one can construct a series of reaction-paths in the Cartesian space of atomic coordinates for each elementary step in the full reaction mechanism; further reaction-path analysis methods, such as NEB, can then be used to refine the proposed reaction path, providing a fully-atomistic view of the reaction mechanism. In the following subsections, the ingredients of this approach are presented, before the final algorithm is summarized.

2.1 Connectivity Matrices

Our mechanism proposal algorithm described below operates in the space of CMs. A CM for an n-atom system is an $n \times n$ matrix with entries which are 0 if two atoms are not bonded, and 1 if two atoms are bonded; we note that the *type* of bonding (*e.g.* single, double, *etc.*) is not considered in this definition. The element G_{ij} of the CM $\mathbf{G}$ are then:

$$G_{ij} = \begin{cases} 1 & \text{if } r_{ij} < r_{ij}^{cut}, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Here, r_{ij}^{cut} is a distance cut-off value which indicates whether or not two atoms i and j are bonded. In what follows, we define this cut-off as

$$r_{ij}^{cut} = \gamma(R_i + R_j), \tag{2}$$

where R_i and R_j are approximate covalent radii for the element-types of atoms i and j, and γ is a parameter which allows for some chemical variation in bonding definitions, with a typical value $\gamma = 1.1$. An important point for the remainder of our approach is that it is straightforward to calculate the CM for *any* input molecular structure.

2.2 Chemical reactions as CM updates

The key aim of the approach developed here is to seek out the sequence of elementary chemical reactions which result in transformation of an input reactant structure into an input product structure. Taking the viewpoint that a chemical reaction *must* involve changes in bonding arrangements (this is discussed further below), it is clear that any chemical reaction can itself be defined as a CM operation. This is illustrated in Fig. 1, which shows how molecular hydrogen dissociation from formaldehyde corresponds to a change in the reactant CM to yield a new CM; as a result, this transformation can be viewed as the operation (*e.g.* addition) of a CM update matrix acting on the reactant CM.

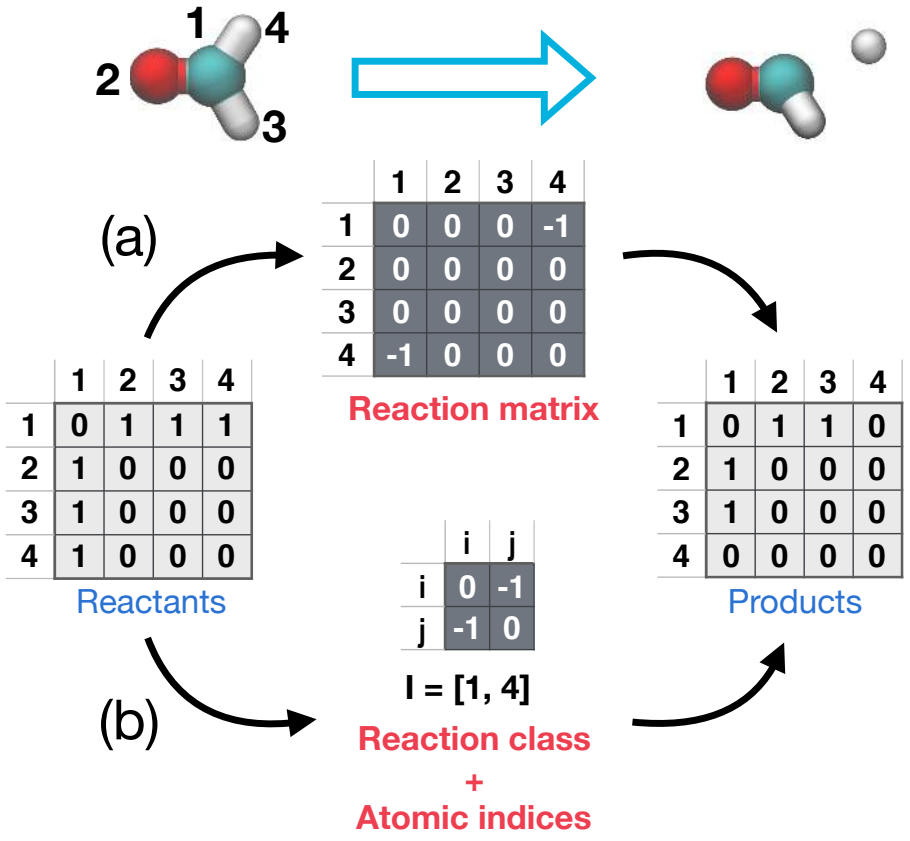


Figure 1: Two methods for defining chemical reactions as connectivity matrix updates, as illustrated for the four-atom formaldehyde system. (a) Definition of an $n \times n$ reaction matrix, describing bond breaking and bond formation, and (b) definition of a reaction class, here acting on two atoms i and j, and atomic indices.

This view of chemical reactions as CM updates is very convenient for developing algorithms for mechanism suggestion, as described below. However, a prerequisite of our approach is that one must provide a library of possible chemical reactions which *might* occur for any given set of atoms and molecules. Unfortunately, as presented in Fig. 1(a), providing a library of $n \times n$ CM update matrices corresponding to possible chemical reactions for a large collection of atom is inconvenient. For example, in the case of formaldehyde illustrated in Fig. 1(a), defining individual reaction matrices would require definition of dissociation reactions for both labelled hydrogen atoms 3 and 4; however, the two corresponding C−H bonds are chemically equivalent. As such, when defining n-atom reaction matrices, the number of possible CM updates grows rapidly as the number of atoms increases because commonalities in chemical reactivity are ignored.

Instead, the approach described below rests on defining a small number of *reaction classes* which operate on small numbers of atoms, rather than defining CM update operations for the full n-atom system, in a similar manner as employed in the RMG scheme.[21] This simplification is illustrated in Fig. 1(b). Here, rather than describing dissociation reactions using a library of n-atom CM update matrices (Fig. 1(a)), an equivalent definition of this set of reactions is to define a 2×2 CM update matrix giving the changes in bonding between *any* two atoms i and j for the reaction-class (*e.g.* dissociation here), as well as identifying the atom indices i and j; this information is equivalent to providing a full $n \times n$ CM update matrix describing dissociation. In other words, by defining a simpler reaction-class CM update matrix, we implicitly reduce the number of CM update matrices which would be required to completely define the set of association/dissociation reactions; using reaction-class CM updates, we simply have a single 2×2 CM update matrix for a dissociation reaction, and the atom indices i and j then identify which atoms participate in this reaction. In what follows, we use $\mathbf{R}^i(\mathbf{I})$ to indicate reaction-class i operating on the set of atomic indices $\mathbf{I}$.

The important point here is that the definition of reaction classes, rather than n-atom reaction matrices, dramatically reduces the number of matrix operations which must be

defined in order to describe the reactivity of a chemical system. Furthermore, reaction-classes can be easily defined using 'chemical common-sense', with standard reactions, such as dissociation, association or elimination, all being readily defined using a small number of reaction-classes.

2.3 Mechanism searching as a global optimization problem

In the above, we have highlighted how:

1. Reactant and product CMs can be readily evaluated using input reactant and product molecular structures;

2. Chemical reactions can be defined as matrix addition or subtraction operations;

3. With a few simple 'chemical common-sense' restraints, one can readily generate a flexible library of possible chemical reaction operations for any given molecular system.

We now show how these ingredients can be used to search for a reaction mechanism (*i.e.* sequence of elementary steps) which connect user-defined reactant and product structures.

We assume that we have input reactant and product molecular structures, which can be converted into CMs $\mathbf{G}^R$ and $\mathbf{G}^P$, respectively. In addition, we assume that we have a library of M chemically-allowed reactions, $\left[\mathbf{R}^1(\mathbf{I}_1), \mathbf{R}^2(\mathbf{I}_2), \ldots, \mathbf{R}^M(\mathbf{I}_M)\right]$, where $\mathbf{I}$ indicates the set of atomic indices which are modified by each reaction.

The goal is then to find a sequence of elementary reactions, as well as participating atomic indices, which connect $\mathbf{G}^R$ and $\mathbf{G}^P$. In other words, we seek the reaction steps such that

$$\mathbf{G}^P = \mathbf{G}^R + \sum_{i=1}^{N_r} \mathbf{R}^{m(i)}(\mathbf{I}_i), \tag{3}$$

where we allow a maximum of N_r elementary steps, $m(i)$ labels the elementary reaction-class occurring at step i, and $\mathbf{I}_i$ labels the set of atomic indices in the n-atom molecular structure which are operated upon by chemical reaction $\mathbf{R}^{m(i)}$.

Now, suppose we have generated a *trial* sequence of reaction steps $\left[\mathbf{R}^{m(1)}(\mathbf{I}_1), \mathbf{R}^{m(2)}(\mathbf{I}_2), \ldots, \mathbf{R}^{m(N_r)}(\mathbf{I}_M)\right]$ and associated atom indices $\mathbf{I}_i$ $(i = 1, 2, \ldots, N_r)$. Using Eq. 3 it is straightforward to calculate the product CM which would be generated upon applying this trial sequence to $\mathbf{G}^R$; we label the resulting product CM as $\tilde{\mathbf{G}}^P$, and note that our goal is to find the reaction sequence and associated atomic indices such that $\tilde{\mathbf{G}}^P = \mathbf{G}^P$.

For any trial sequence, we can define an error function as the simple element-wise difference between the trial product CM and the target product CM, given by:

$$F = \sum_{j>i} \left(\tilde{G}^P_{ij} - G^P_{ij} \right)^2 . \tag{4}$$

Clearly, when $F = 0$, we have identified a sequence of elementary reaction steps (and associated atomic indices) that directly connect the input reactant and product CMs; in other words, a trial path with $F = 0$ represents a possible reaction mechanism to convert $\mathbf{G}^R$ into products $\mathbf{G}^P$.

With the above development, we see that the search for reaction mechanism can be transformed into an optimization problem; we seek the sequence of N_r reactions $\mathbf{R}_i^{m(i)}$ and atomic indices $\mathbf{I}_i$ for which $F = 0$, as illustrated schematically in Fig. 2. Any trial solution can be readily defined as a discrete series of reaction and atomic indices; this discrete optimization problem can, in principle, be addressed using a wide variety of optimization strategies, with examples including genetic algorithms, SA or army ants algorithm. In the examples given below we use SA, with moves at each iteration corresponding to changes in the reaction-class and/or atomic indices of a randomly-selected reaction in the current sequence of N_r steps. After this change is made, the new error function is evaluated using Eq. 4, and the standard Metropolis method is used to accept or reject the move. After a sufficient number of iterations, it is found that a reaction sequence with $F = 0$ can be generated with some reliability. However, it is important to note that, because this optimization takes place solely in the discrete space of CMs, the evaluation of F is very fast, such that large numbers of

SA moves can be generated and tested in a short space of time; for example, a representative calculation with 10^6 Monte Carlo moves takes about one minute on a standard laptop computer.

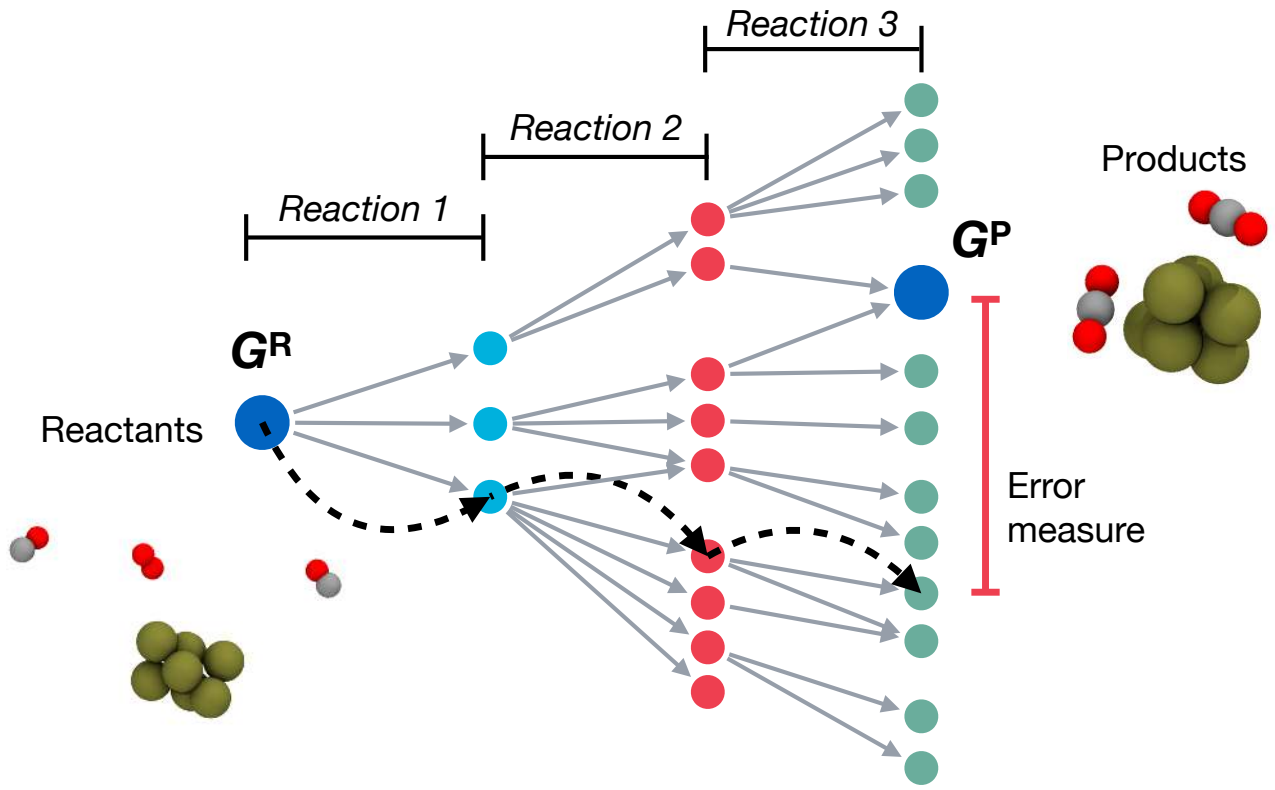


Figure 2: Overview of optimization scheme for multi-step mechanism proposal. Starting with user-defined reactants and products, the end-point connectivity matrices, $\mathbf{G}^{\mathrm{R}}$ and $\mathbf{G}^{\mathrm{P}}$, are evaluated. Next a series of CM updates are proposed leading to generation of a sequence of new connectivity matrices; in this case, we illustrate a path (black dotted line), with the colored circles representing the different possible connectivity matrices which can be generated by matrix updates. After a maximum fixed number of reaction-steps N_r, the resulting connectivity matrix is used to evaluate an error measure F describing distance from the target connectivity matrix; simulated annealing by modifying the intermediate reaction steps is then used to minimize this error measure.

An important factor in the success of this this graph-based optimization procedure, as demonstrated below, is that it can be easily directed to prevent exploration of reaction-path sequences which lead towards non-physical CMs. Specifically, for some trial reaction sequence, as each term is summed in Eq. 3 (*i.e.* after each proposed elementary step) one

can perform checks on the current CM to ensure that standard chemical valence ranges, or similar restraints, are not violated. As an example, one would expect thermodynamically-stable many-atom molecules containing carbon to only exhibit bonding motifs in which the valence of carbon is between one and four. So, if an intermediate CM generated during the summation of Eq. 3 leads to formation of a molecular species with a carbon atom with a valence of six, one can reject this reaction sequence on simple chemical 'common-sense', with little runtime cost to the algorithm. In practice, this can be achieved during SA by giving these constraint-violating reaction-sequences an arbitrarily large error function, such that they are *always* rejected during the Metropolis test. This chemical constraints procedure can also be used to direct the reaction-sequence search towards specific types of chemistry, such as searching only for reaction sequences in which all carbon-carbon double bonds remain intact; this tunable specificity is not greatly exploited in the reactions modeled below, beyond ensuring that reactions take place at a nanocluster surface, but represents an interesting avenue to explore further.

2.4 Molecular structure generation using CMs

The minimization of F described above results in a sequence of elementary reactions and associated atomic indices participating in each reaction, $\left[\mathbf{R}^{m(1)}(\mathbf{I}_1), \mathbf{R}^{m(2)}(\mathbf{I}_2), \ldots, \mathbf{R}^{m(N_r)}(\mathbf{I}_M)\right]$. For further analysis, such as evaluation of thermodynamics and kinetic properties, we then require the atomic coordinates of all of the intermediate molecular structures generated along this sequence of elementary reaction-steps. In other words, we need to generate molecular structures which conform to the bonding pattern encoded in each intermediate CM.

To achieve this, we can use GRPs to *impose* a target CM on a molecular structure. The GRP is a function of both atomic coordinates $\mathbf{r}$ *and* a target CM $\mathbf{G}^T$, and is defined to be zero only when the CM calculated from $\mathbf{r}$ corresponds exactly to $\mathbf{G}^T$. As such, starting from some reasonable initial atomic coordinates and target CM, optimization of the GRP yields a molecular structure which corresponds to the target CM.

The GRP $W(\mathbf{r}, \mathbf{G})$ used here is similar to that used in previous work,[35,36] and has the following functional form:

$$W(\mathbf{r}, \mathbf{G}) = \sum_{j>i} \Bigg[\delta(G_{ij} - 1) \left[H(r_{ij}^{min} - r_{ij}) \sigma_1 (r_{ij}^{min} - r_{ij})^2 + H(r_{ij} - r_{ij}^{max}) \sigma_1 (r_{ij}^{max} - r_{ij})^2 \right] + \delta(G_{ij}) \sigma_2 e^{-r_{ij}^2/(2\sigma_3^2)} \Bigg] + V_{mol}(\mathbf{r}, \mathbf{G}). \tag{5}$$

The summation in Eq. 5 runs over all pairs of atoms. The first term is a harmonic restraining force which acts on pairs of atoms which are *bonded* in order to keep their bond-length between the fixed limits r_{ij}^{min} and r_{ij}^{max}. The "delta" function is defined such that

$$\delta(x) = \begin{cases} 1 & \text{if } x = 0, \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

so the term $\delta(G_{ij} - 1)$ implies that the corresponding term operates on bonded atoms i and j. Furthermore, $H(x)$ is the Heaviside function, defined as:

$$H(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 & \text{if } x > 0. \end{cases} \tag{7}$$

So, $H(r_{ij}^{min} - r_{ij})$ is zero as long as $r_{ij} > r_{ij}^{min}$; in this case, this first harmonic restraint term does not contribute to the GRP. If, on the other hand, $r_{ij} < r_{ij}^{min}$, then $H(r_{ij}^{min} - r_{ij}) = 1$ and the first term does contribute to the potential energy. In particular, a harmonic term of the form $\sigma_1 (r_{ij}^{min} - r_{ij})^2$ is applied to the system, where σ_1 is a user-defined constant. The effect of this potential energy term is to push the atoms i and j to bond lengths such that $r_{ij} > r_{ij}^{min}$. In other words, this term pushes the atoms apart until they are some minimum distance r_{ij}^{min} away from each other. The second harmonic restraint term has a similar effect, but instead of maintaining some *minimum* distance, it makes sure that a *bonded* pair

of atoms always remain closer than some maximum allowed distance r_{ij}^{max}. Together, the influence of the harmonic restraint terms in Eq. 5 is to ensure that a pair of *bonded* atoms *always* have bond-lengths between the pre-defined limits r_{ij}^{min} and r_{ij}^{max}. The final term in the pair-potential of Eq. 5 acts as a repulsive potential between pairs of atoms which are *not* bonded. The $\delta(G_{ij})$ term makes sure that this term only applies to pairs of atoms for which $G_{ij} = 0$. The remainder of this term is a simple Gaussian repulsive potential with a strength parameter σ_2 and a range parameter σ_3.

In addition to the pairwise additive terms in Eq. 5, we also include a *molecular* term which only operates between distinct molecular species. Here, the $V_{mol}(\mathbf{r}, \mathbf{G})$ term has the following form:

$$V_{mol}(\mathbf{r}, \mathbf{G}) = \sum_{J>I} [H(R^{min} - R_{IJ})\sigma_4(R^{min} - R_{IJ})^2], \tag{8}$$

where R_{IJ} is the distance between the centers-of-mass of two molecules I and J, and R^{min} is a user-defined minimum separation distance between any pair of molecules. By comparing to the bonding term in Eq. 5, we see that the molecular term V_{mol} is designed to make sure that distinct molecules are simply "kept apart" from each other.

The GRP defined in Eq. 5 is clearly not unique, but provides a simple pair-potential-like PES which depends on both the target CM and the atomic coordinates. Optimization of atomic coordinates under $W(\mathbf{r}, \mathbf{G}^T)$ yields a molecular structure which obeys the bonding restraints of the target CM; this approach can therefore be used to convert between the CM-space within which our reaction-mechanism search works, into the atomic coordinate-space required for further reaction-path analysis calculations. As a final point, we note that the GRP employed here is a simple pair potential which does not account for the finer details of molecular structure; as demonstrated below, geometry optimization with an empirical force-field or *ab initio* quantum chemistry calculation can subsequently be used to correct molecular structures.

2.5 Algorithm summary

Putting together the different aspects described above, our final reaction-mechanism proposal scheme proceeds as follows:

1. Input molecular structures for reactants ($\mathbf{r}^R$) and products ($\mathbf{r}^P$) are converted into the corresponding CMs $\mathbf{G}^R$ and $\mathbf{G}^P$, respectively;

2. A maximum number of allowed reactions in the reaction sequence, N_r, is selected;

3. The initial reaction sequence is set to simply be a series of N_r 'null' reactions, corresponding to zero CM changes at each step, and the initial error function F is evaluated;

4. A SA simulation is then performed for a maximum of N_{max} iterations. At each iteration, one of the CM updates in the current reaction sequence is updated, and the new reaction-sequence is accepted or rejected based on the Metropolis criterion using F as a target minimization function;

5. Once a reaction-sequence with $F = 0$ is found, the SA calculation is terminated;

6. For the optimized reaction sequence, molecular coordinates for each elementary step are generated by optimization under each successive CM; in each case, the initial molecular configuration for optimization is simply that produced by optimization at the previous step (noting that the input reactant configuration is used as the starting point for the optimization in the first elementary reaction step);

7. Finally, further analysis is performed using the molecular coordinates of each individual elementary step, such as geometry optimization or NEB refinement.

The outcomes of this algorithm are: (i) identification of a set of elementary reactions, as well as the atoms participating in each reaction, which lead from input reactants to input products, (ii) initial molecular coordinates for a series of images along each elementary reaction-path, which can be used in further analysis.

3 Application, Results and Discussion

In this Section, we apply our double-ended reaction path finding method to three different reactions, all occurring in the presence of a model platinum cluster catalyst: (i) oxidation of carbon monoxide, (ii) water-gas shift reaction, and (iii) aromatization of *n*-hexane to benzene. In each case, our goal is to use our reaction-path search algorithm to identify a sequence of elementary reactions connecting defined reactants and products. We do not claim that the reaction sequences generated are the *minimum-energy* sequences, just that they represent 'chemically-sensible' reaction mechanisms leading from reactants to products; further analysis of a large number of such paths would be required to positively identify the globally minimum-energy reaction sequence, as discussed later in Section 4.

3.1 General computational details

All simulations used a custom-written computer code, and were performed on a standard laptop computer. Details of simulation parameters are given in Table 1, notably GRP parameters and approximate cut-off distances. As described below, NEB was used to confirm that the generated reaction sequences for each studied reaction can be converted into atomic coordinate-space; for computational convenience, and because we are only interested in qualitative calculations as an envisaged first step in a hierarchy of increasingly-accurate analysis simulations, we use the ReaxFF[38,39] PES to describe intra- and intermolecular interactions during all geometry optimizations and NEB refinements. Of course, the approach described above is not tied to any particular PES model; for example, we could have equally used density functional theory (DFT) to perform NEB, but the additional computational expense is not justified here.

In all of the reactions considered below, we used a small set of chemical reaction types in the library of possible CM updates for the SA search. In particular, we limited the set of possible chemical reactions to simple association/dissociation reactions, as well as 3-

Table 1: Parameters of the graph-restraining potential, and effective atomic radii for connectivity-matrix calculation.

Parameter	Value / atomic units
σ_1	0.01
σ_2	0.02
σ_3	2.20
σ_4	0.05

Atom type	Effective radius / a_0
C	0.72
O	0.72
H	0.40
Pt	1.46

body insertion/elimination reactions, as shown in Table 2. This reaction library is intended to be somewhat generic; of course, more tailored reaction libraries for any system under consideration could be easily developed too, allowing a large degree of flexibility in the reaction search. However, we note that an important condition is imposed on the reaction library considered here; specifically, all reactions *must* involve one of the platinum catalyst atoms. This is to force the target reaction to take place at the Pt cluster, as might be expected in a 'real-world' nanoparticle catalysis system.

Table 2: Library of reaction classes used in simulated annealing searches. As described in the main text, to ensure that reactions occur at the Pt_7 cluster, one of the reactive atomic species is assigned as Pt in each reaction class.

Reaction	Notes
$A - Pt \rightleftharpoons A + Pt$	Single atom association/dissociation
$A - Pt - B \rightleftharpoons A - B + Pt$	Elimination
$A - Pt + B \rightleftharpoons A - B + Pt$	Atom transfer

At each iteration of the SA calculations, one of two updates were attempted at random. The first update move attempted to change the atomic indices of a randomly selected reaction step in the current reaction-sequence; the second possible update move attempted to change

both the reaction type (drawn from the available reaction library) *and* the associated atomic indices. A typical starting temperature for the SA simulation was 200×10^3 K (here, we are assuming that the function F is given in atomic units; typical changes of $\Delta F = \pm 1$ in atomic units necessitate such high temperatures to ensure that sufficient moves are accepted early in the SA calculation).

Following identification of a reaction-sequence with $F = 0$, atomic coordinates for N_r reaction-sequence end-points; linear interpolation between these structures was then used to generate $n = 10$ images along each elementary reaction step. Optimization of atomic coordinates under action fo the GRP for each successive CM in the reaction-sequence was performed using a simple steepest-descents algorithm until the RMS of the atomic forces was less than $5 \times 10^{-4}\ E_h a_0^{-1}$.

When required, NEB calculations were performed by first optimizing the geometry of each reaction end-point using the ReaxFF PES, then subsequently using the QuickMin algorithm to refine the internal images in the reaction path. NEB refinement continued until the RMS forces on the images was less than $3 \times 10^{-3}\ E_h a_0^{-1}$.

3.2 Oxidation of carbon monoxide

As a first example, we consider oxidation of carbon monoxide[40–42] in the presence of a Pt_7 cluster, namely:

$$2\,CO + O_2 \longrightarrow 2\,CO_2.$$

The initial reactant configuration comprised two carbon monoxide molecules (CO), one oxygen molecule (O_2), and a Pt_7 cluster with D_{5h} symmetry (as determined to be the lowest-energy structure in previous calculations[43]). The target product configuration comprised two carbon dioxide (CO_2) molecules, in addition to the Pt_7 cluster. These input reactant and product configurations were used to generate the target CMs, $\mathbf{G}^R$ and $\mathbf{G}^P$.

Subsequently, we performed a SA optimization to search for a reaction-sequence connect-

ing $\mathbf{G}^R$ and $\mathbf{G}^P$, using the library of reaction-moves shown in Table 2. The maximum allowed length of the reaction sequence was (somewhat arbitrarily) chosen to be $N_r = 12$, although we note that a "null" reaction (resulting in no change in CM) is included in the allowed library of moves, so it is possible that the actual number of active reactions in the sequence N_r represents the maximum number of *active* chemical reactions. SA was performed for a maximum of 10^6 iterations.

Figure 3 shows the results of NEB refinements for each of the $N_r = 12$ elementary steps in a reaction-sequence with $F = 0$. This reaction-sequence was located after around 600×10^3 SA iterations (see inset of Fig. 3). Repeated simulations show that our approach can reliably find a reaction-sequence connecting reactant and product structures within a few hundred thousand iterations, although we note that the determined reaction-sequence is, of course, not always the same. Here, we focus on a representative reaction path, and the challenge of further sampling over multiple reaction-paths is discussed below.

The sequence of molecular structures generated along the indicative CO oxidation reaction pathway are also shown in Fig. 3. The first reaction involves adsorption of O_2 at the Pt_7 surface, followed by dissociation of O_2 into two individual oxygen atoms on the Pt cluster surface; this molecular oxygen dissociation step has the highest potential energy barrier along the entire reaction path (~ 140 kJ mol^{-1}). In subsequent reaction steps, the two separate CO molecules adsorb at Pt sites adjacent to the atomic oxygen, then participate in dissociative reactions to form CO_2, with the barriers to these dissociative steps being of the order 40-90 kJ mol^{-1}. Of course, we anticipate that the actual calculated barriers are less accurate than using, for example, DFT or other *ab initio* methods, but our results suggest that the ReaxFF model provides a qualitatively correct picture of the relative potential energy surface along the reaction path.

The important result in terms of this work is that our double-ended reaction-search method has proven capable of generating candidate reaction mechanism for a complex multi-step reaction, using only minimal "chemical common-sense" as input. To move further, to-

wards mechanistic hypothesis testing, one could repeat this double-ended GDS simulation multiple times to generate multiple candidate mechanisms; NEB refinements for each sequence of reaction paths would then enable identification of the "most likely" mechanism, while reaction-rate calculations *via* TST would enable one to make contact with experimental rate-laws. These sorts of simulations will be computationally-demanding (depending on the exact number of independent reaction sequences generated), but provide a well-defined approach to mechanism testing; this approach will be the subject of future work.

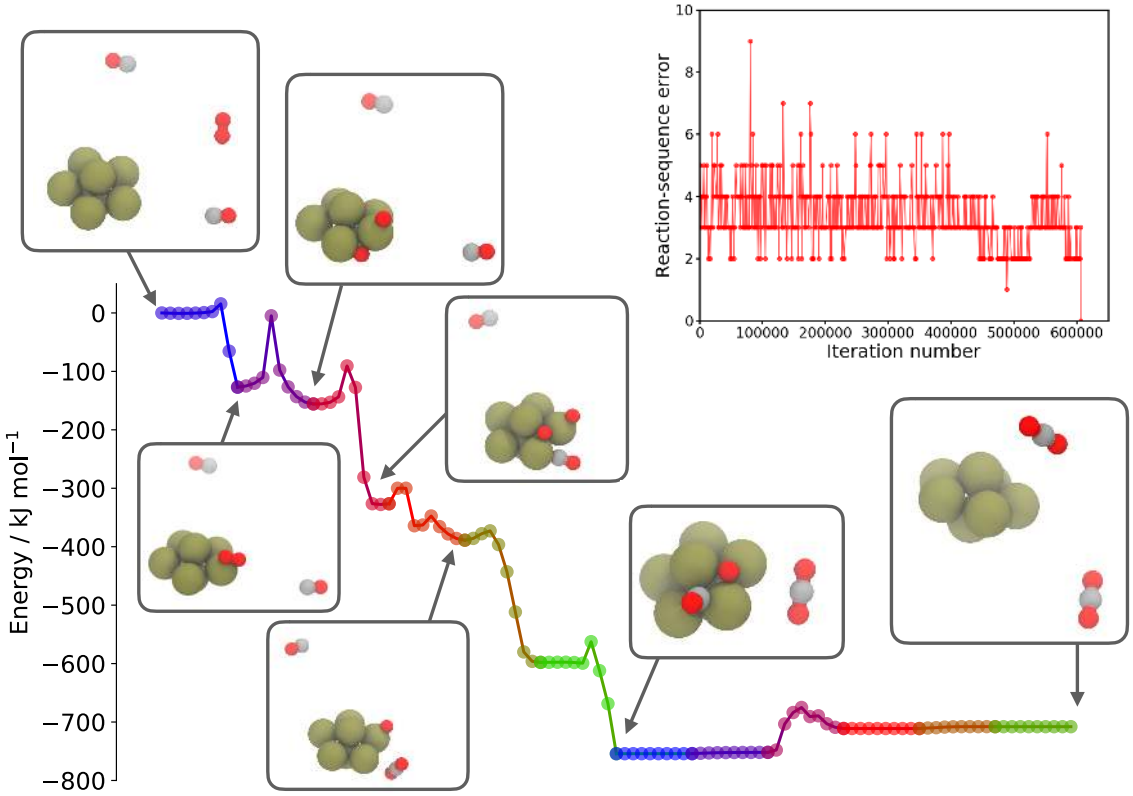


Figure 3: Automatically-proposed reaction mechanism for CO oxidation on Pt_7 nanoparticle. The colored central line shows the calculated potential energy profile for a series of 12 NEB calculations, each connecting end-points generated by simulated annealing combined with structure refinement under the GRP; each different colored segment in the line shows one of the NEB simulations, with the dots showing the potential energies of the NEB images in each refinement. Selected optimized structures along the reaction profile are also shown. The upper-right panel shows the variation in the reaction-sequence error function F during the simulated annealing optimization calculation; note that the error drops to zero after around 600×10^3 iterations.

3.3 Water-gas shift reaction

The second example considered here is the water-gas shift reaction (WGSR), a key route for generating molecular hydrogen:

$$2\,\mathrm{CO} + \mathrm{H_2O} \longrightarrow \mathrm{H_2} + 2\mathrm{CO_2}.$$

The WGSR is typically catalyzed by metal nanoparticles of Pt, Cu or Au, supported on alumna or ceria.[41,44,45] The mechanism of the WGSR has been suggested to proceed *via* either a redox mechanism or an associative mechanisms; however, insight into the atomistic details of the mechanism is complicated by a number of factors, including catalyst size effects, the influence of the support, and the role of the catalyst redox state. We note that the influence of these factors can be investigated using our double-ended reaction-sequence search; however, the aim of this initial investigation is to determine whether our approach can successfully predict possible reaction-mechanisms for complex reaction systems. Ultimately, as noted above, we envisage that full investigation of complex mechanistic questions would require generation of multiple candidate reaction mechanisms, followed by evaluation of each using *ab initio* calculations and reaction rate theory.

Figure 4 shows the results of two independent reaction-sequence calculations for the WGSR; in each calculation, $N_r = 12$. Each calculation was performed with the same set of possible chemical reaction-classes, but using different random-number sequences in each SA simulation; however, both illustrated reaction paths have $F = 0$, and correspond to plausible reaction mechanisms using the allowed set of CM updates. As an aside, the general behavior of our approach seems to be that, as long as N_r is "large enough" to allow sufficient flexibility in the reaction-path, the current SA approach can find a zero-error ($F = 0$) reaction sequence with good reliability; however, exactly how one should choose N_r without prior bias on the mechanism is an area which requires further investigation.

The upper reaction-sequence in Fig. 4 proceeds with initial association of the water

molecule at the Pt_7 surface, followed by dissociation of both hydrogen atoms. This sequential double-dissociation event actually leads to the double-peaked NEB energy profile observed at the second step of the upper reaction-path in Fig. 4. In subsequent steps, the hydrogen atoms diffuse across the Pt_7 cluster surface. Meanwhile, the CO molecule can be seen to bind (*via* the carbon atom) to a Pt atom adjacent to the surface-bound oxygen, followed by rearrangement to form a loosely-bound CO_2 molecule. In the latter reaction steps, the surface-bound CO_2 and H_2 molecules dissociate to form the target products.

In contrast, the lower reaction sequence in Fig. 4 proceeds in a different manner, highlighting the fact that our SA protocol does not just generate a single reaction sequence but is flexible enough to generate multiple candidates; as above, we emphasize that assessing exactly which reaction sequence is kinetically or thermodynamically preferred would require a second assessment step requiring comparison of thermodynamic properties and free energy barriers for multiple candidate pathways. In the second candidate reaction sequence, the broad reaction steps are similar to the first, with dissociation of hydrogen from water followed by recombination to form molecular hydrogen running in parallel with CO adsorption and CO_2 elimination. However, the key difference here seems to be be that the dissociation of hydrogen from water is split into a two-step process, with the first hydrogen atom dissociating to occupy an interstitial site in the Pt_7 cluster, and the second dissociation producing a surface-bound hydrogen atom. After rearrangements of the hydrogen atoms, the remainder of the pathway then broadly follows the upper reaction sequence, obviously leading to the same target products, although it is notable that the barriers to CO_2 dissociation in the reaction

$$CO_{ads} + O_{ads} \longrightarrow CO_2$$

are quite different for the two reactions shown. In the upper reaction, the barrier to this reaction is around 74 kJ mol^{-1}, compared to roughly 250 kJ mol^{-1} for the lower reaction sequence. The difference here seems to lie in the nature of the binding site of the adsorbed oxygen atom. In the upper reaction sequence, the oxygen atom is bound at one of the

edges of Pt_7 cluster, shared by two Pt atoms, whereas the lower reaction sequence finds the oxygen atom bond in a hollow formed by three Pt atoms, leading to stronger binding (at least according to the ReaxFF model used here) and a larger potential energy barrier to dissociation. Most interestingly from our point-of-view is that these sorts of insights emerge directly from automatically-generated reaction-sequences and NEB analysis, without requiring a user-guided search over relevant reactive conformations.

3.4 Aromatization of hexane

The final example application for our approach is the aromatization of n-hexane to benzene,

$$C_6H_{14} \longrightarrow C_6H_6 + 4\,H_2.$$

This reaction has been observed to occur on solid surfaces, clusters and zeolites,[46,47] typically at reaction temperatures over 700 K. The complexity of this reaction, as well as the diversity of catalytic species, suggests an equally complex manifold of possible reaction mechanisms; our intention here is not to fully investigate all possible mechanisms, but simply to demonstrate that our graph-based search approach can successfully determine candidate reaction mechanisms even in the case of complex many-step reactions.

Figure 5 illustrates one of the candidate mechanisms for n-hexane aromatization which was postulated by our graph-based optimization procedure. This $F = 0$ reaction-sequence contains $N_r = 20$ steps, and was found after roughly 600×10^3 SA iterations, as shown in the inset of Fig. 5. Rather than present NEB results for all 20 reaction steps, we only show the selected reaction end-points along the reaction-sequence for clarity.

The illustrated reaction-path proceeds with dissociative adsorption of n-hexane at the Pt_7 surface, cleaving one of the C−H bonds in a terminal CH_3. After further hydrogen dissociation events, an intermediate structure is formed in which the two ends of the C_6 chain form a bridge on the surface of the Pt_7 cluster. Subsequent hydrogen rearrangements on the

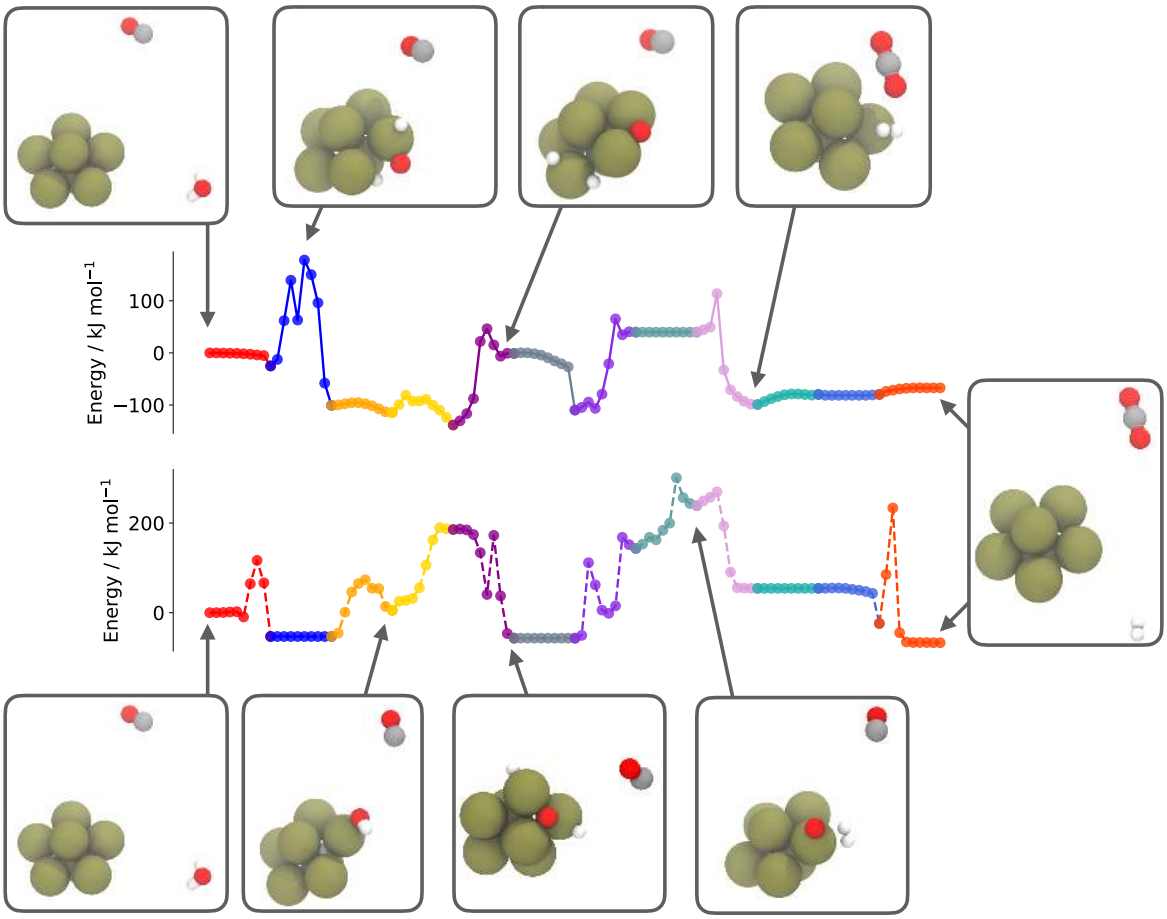


Figure 4: Automatically-proposed reaction mechanisms for the water-gas shift reaction (WGSR) on a Pt_7 nanoparticle. The two colored central lines show the calculated potential energy profiles for a series of 12 NEB calculations, each connecting end-points generated by simulated annealing combined with structure refinement under the GRP; each different colored segment in the line shows one of the NEB simulations, with the dots showing the potential energies of the NEB images in each refinement. Selected optimized structures along both reaction profiles are also shown. The two different sets of potential energy profiles shown arise from two independent runs of our simulated-annealing algorithm for reaction-sequence generation; both reaction sequences have $F = 0$, and so both correspond to plausible mechanisms which can be generated using the allowed reaction set.

Pt_7 cluster, followed by dissociation events, leads to step-wise generation of four molecular hydrogen molecules, with the final step in the process being simultaneous dissociation and completion of the benzene ring. The overall reaction is energetically-favored by around 50 kJ mol^{-1}, although we note that the performance of ReaxFF is not well-characterized for this system; nevertheless, the stability of the geometry optimization process at each stage, as well as the overall characteristics of the optimized structures, suggest that these structures are at least 'chemically sensible' according to the ReaxFF prescription.

We emphasize that the reaction mechanism suggested in Fig. 5 is just one of many possible. Repeated runs of our SA algorithm give different reaction sequences; efficiently sifting through this set of reaction mechanism candidates to search for the MEP is the next anticipated step in the development of our reaction sampling scheme. It seems very likely that an alternative reaction-path would be overall favoured compared to the one shown in Fig. 5, which is found to demonstrate several intermediate steps which are strongly 'uphill' based on the calculated potential energy values.

Compared to CO oxidation and WGSR, *n*-hexane aromatization represents a much more challenging reaction system; the increased number of atoms means that the possible number of reactions available at each intermediate step is very large. However, we find that the SA optimization process proceeds quite straightforwardly; as a rough estimate, we find that $\sim 90\%$ of SA runs (each performed with different random number sequence) are successful when $N_r = 20$. However, this success rate does decrease as N_r is decreased, presumably due to the the random-walk becoming 'trapped' in certain reaction pathways due to lack of flexibility. Although this observation is encouraging, there is obviously some further work required to understand the impact of algorithm parameters such as N_r in order to provide a robust computational strategy.

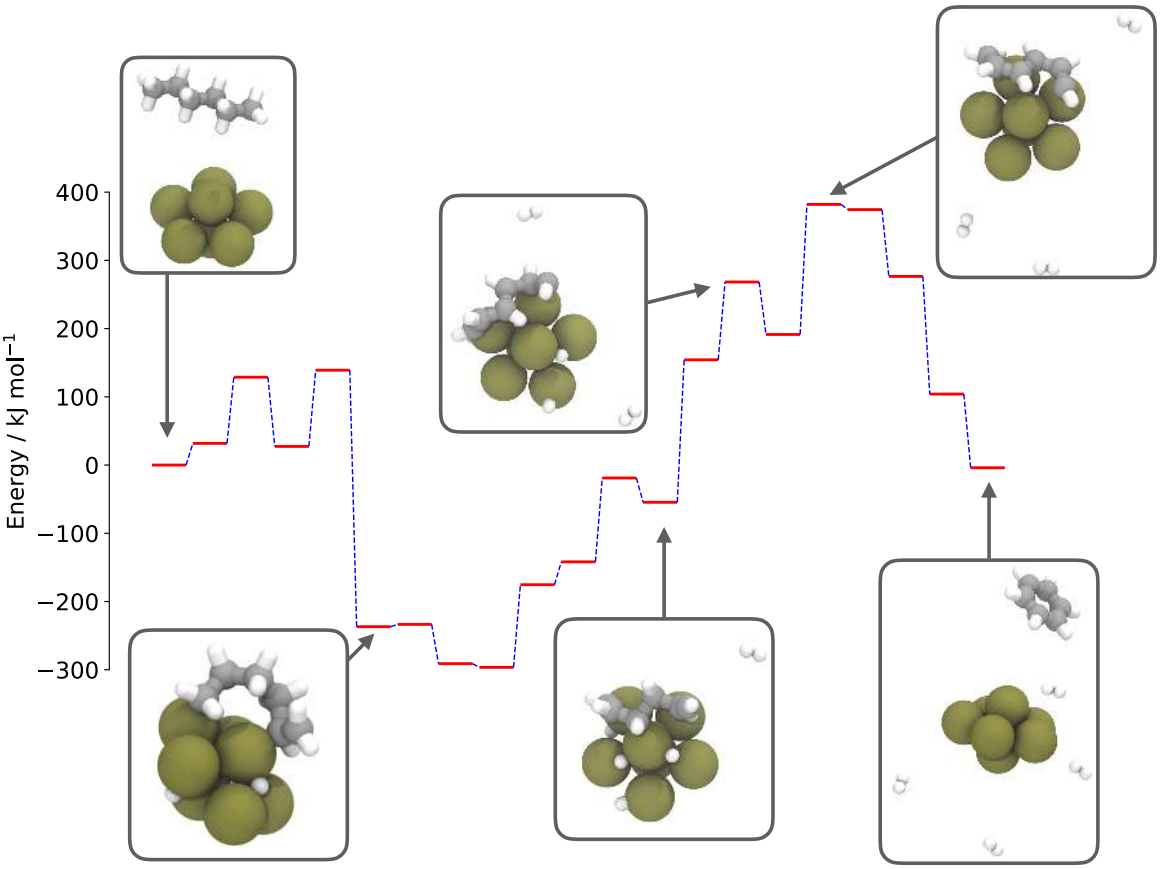


Figure 5: Automatically-proposed reaction mechanism for aromatization of n-hexane on a Pt_7 nanoparticle. The central line shows the energies of the geometry-optimized reactants and products for $N_r = 20$ reactions along a pathways connecting reactants ($C_6H_{14} + Pt_7$) to products ($C_6H_6 + 4\,H_2 + Pt_7$); selected structures along his reaction path are shown.

4 Conclusions

In this Article, we have developed and successfully tested a new automated approach to generating reaction sequences connecting user-defined reactant and product structures. Our approach is based on optimizing an error function which quantifies the difference between the target connectivity matrix of the product structure and the connectivity matrix obtained by applying a series of chemical reactions to the input reactant structure; by updating the sequence of chemical reactions along the path, we turn the search for a reaction sequence connecting reactants and products into a task in optimization. In the present Article, we have used a simple SA method to perform this optimization, but it could equally be performed by other optimization schemes.

A major advantage of our scheme is that is is very fast because it works solely in the space of the connectivity matrices of the reactants, products and intermediate structures. We have also shown how, by optimizing atomic coordinates under the action of a graph-restraining potential energy surface, we can generate atomic coordinates for each reaction in a proposed reaction sequence; this then enables further analysis, such as using NEB refinement (as performed above) or transition-state theory rate calculations.

Our approach towards automatic reaction mechanism proposal has been tested for three different reactions occurring on Pt_7 clusters, namely CO oxidation, the water-gas shift reaction, and *n*-hexane aromatization. In each case, we find that we can generate candidate reaction steps connecting reactants and products in a reliable manner, with a typical optimization simulation taking less than a minute on a standard laptop computer. We have also demonstrated how the emerging reaction paths can be combined with NEB refinement to generate potential energy profiles and atomic mechanisms for further analysis.

Although successful in our original aims of developing a scheme for proposing 'double-ended' reaction mechanisms, there are some clear avenues to improve our approach. For example, a better understanding of how the maximum path-length N_r influences the optimization efficiency would be beneficial. In addition, it is currently unclear how the number

of 'chemically-sensible' reaction sequences connecting defined end-points changes as one increases N_r, changes the constraints on chemical valences, or modifies the identities of reaction classes; again, better insight here might lead to optimization improvements.

Overall, however, this paper represents a potential first step towards novel schemes aimed at, for example, automatic catalyst performance prediction or proposing retrosynthetic paths from complex organic molecules. In particular, there is scope to combine our reaction mechanism proposal scheme with the recently-developed *AARON* code[48] for automated transition-state optimization for catalytic reactions, as well as providing input data for microkinetic modelling schemes which are increasingly finding application in homogeneous catalysis.[35,49] However, perhaps the most exciting opportunity is the goal of developing new computational schemes for direct homogeneous catalyst design and optimization;[50] these aspects, as well as further optimization of our approach, will be discussed in the near-future.

Acknowledgement

The authors gratefully acknowledge the award of funding by the Engineering and Physical Sciences Research Council (EPSRC; EP/R020477/1). We also gratefully acknowledge high-performance computing facilities provided by the Scientific Computing Research Technology Platform at the University of Warwick. Data from Figs. 3, 4 and 5 can be found at http://wrap.warwick.ac.uk/113322.

References

(1) Raugei, S.; DuBois, D. L.; Rousseau, R.; Chen, S.; Ho, M.-H.; Bullock, R. M.; Dupuis, M. Toward molecular catalysts by computer. *Acc. Chem. Res.* **2015**, *48*, 248–255.

(2) Houk, K. N.; Cheong, P. H.-Y. Computational prediction of small-molecule catalysts. *Nature* **2008**, *455*, 309–313.

(3) Nett, A. J.; Zhao, W.; Zimmerman, P. M.; Montgomery, J. Highly active nickel catalysts for C-H functionalization identified through analysis of off-cycle intermediates. *J. Am. Chem. Soc.* **2015**, *137*, 7636–9.

(4) Rush, L. E.; Pringle, P. G.; Harvey, J. N. Computational kinetics of cobalt-catalyzed hydroformylation. *Angew. Chemie Int. Ed.* **2014**, *53*, 8672–8676.

(5) van Leeuwen, P. W. N. M. *Homogeneous catalysis: Understanding the art*; Kluwer Academic Publishers, 2004.

(6) Zhang, X.; Fevre, M.; Jones, G. O.; Waymouth, R. M. Catalysis as an enabling science for sustainable polymers. *Chem. Rev.* **2017**,

(7) Keçeli, M.; Elliott, S. N.; Li, Y.-P.; Johnson, M. S.; Cavallotti, C.; Georgievskii, Y.; Green, W. H.; Pelucchi, M.; Wozniak, J. M.; Jasper, A. W., *et al.* Automated computational thermochemistry for butane oxidation: A prelude to predictive automated combustion kinetics. *Proc. Combust. Inst.* **2019**, *37*, 363 – 371.

(8) Chen, X.; Goldsmith, C. F. A Theoretical and computational analysis of the methylvinyl + O_2 reaction and its effects on propene combustion. *J. Phys. Chem. A* **2017**, *121*, 9173–9184.

(9) Class, C. A.; Liu, M.; Vandeputte, A. G.; Green, W. H. Automatic mechanism gener-

ation for pyrolysis of di-tert-butyl sulfide. *Phys. Chem. Chem. Phys.* **2016**, *18*, 21651–21658.

(10) Laidler, K. J. *Chemical kinetics*, 3rd ed.; Harper Collins: New York, 1987.

(11) Henriksen, N. E.; Hansen, F. Y. *Theories of molecular reaction dynamics: The microscopic foundation of chemical kinetics*; Oxford University Press, 2011.

(12) Mills, G.; Jónsson, H. Quantum and thermal effects in H_2 dissociative adsorption: Evaluation of free energy barriers in multidimensional quantum systems. *Phys. Rev. Lett.* **1994**, *72*, 1124–1127.

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

(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.; Jónsson, H. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *J. Chem. Phys.* **1999**, *111*, 7010–7022.

(16) Peters, B.; Heyden, A.; Bell, A. T.; Chakraborty, A. A growing string method for determining transition states: Comparison to the nudged elastic band and string methods. *J. Chem. Phys.* **2004**, *120*, 7877–7886.

(17) Zimmerman, P. M. Reliable transition state searches integrated with the growing string method. *J. Chem. Theory Comput.* **2013**, *9*, 3043–3050.

(18) Peters, B.; Liang, W.; Bell, A. T.; Chakraborty, A. Biasing a transition state search to locate multiple reaction pathways. *J. Chem. Phys.* **2003**, *118*, 9533–9541.

(19) Pozun, Z. D.; Hansen, K.; Sheppard, D.; Rupp, M.; Müller, K.-R.; Henkelman, G. Optimizing transition states via kernel-based machine learning. *J. Chem. Phys.* **2012**, *136*, 174101.

(20) Govind, N.; Petersen, M.; Fitzgerald, G.; King-Smith, D.; Andzelm, J. A generalized synchronous transit method for transition state location. *Comput. Mat. Sci.* **2003**, *28*, 250 – 258, .

(21) 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.

(22) Simm, G. N.; Vaucher, A. C.; Reiher, M. Exploration of reaction pathways and chemical transformation Networks. *J. Phys. Chem. A* **2019**, *123*, 385–399.

(23) Wang, L.-P.; Titov, A.; McGibbon, R.; Liu, F.; Pande, V. S.; Martínez, T. J. Discovering chemistry with an ab initio nanoreactor. *Nature Chem.* **2014**, *6*, 1044–8.

(24) Maeda, S.; Ohno, K. Global mapping of equilibrium and transition structures on potential energy surfaces by the scaled hypersphere search Method: Applications to ab Initio surfaces of formaldehyde and propyne molecules. *J. Phys. Chem. A* **2005**, *109*, 5742–5753.

(25) Ohno, K.; Maeda, S. Automated exploration of reaction channels. *Phys. Scripta* **2008**, *78*, 058122.

(26) Ohno, K.; Maeda, S. A scaled hypersphere search method for the topography of reaction pathways on the potential energy surface. *Chem. Phys. Lett.* **2004**, *384*, 277–282.

(27) Maeda, S.; Morokuma, K. Finding reaction pathways of type A + B $\rightarrow$ X: Toward systematic prediction of reaction mechanisms. *J. Chem. Theory Comput.* **2011**, *7*, 2335–2345.

(28) Maeda, S.; Morokuma, K. Toward predicting full catalytic cycle using automatic reaction path search method: A case study on $HCo(CO)_3$-catalyzed hydroformylation. *J. Chem. Theory Comput.* **2012**, *8*, 380–385.

(29) Kale, S.; Sode, O.; Weare, J.; Dinner, A. R. Finding chemical reaction paths with a multilevel preconditioning protocol. *J. Chem. Theory Comput.* **2014**, *10*, 5467 – 5475.

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

(31) Kim, Y.; Choi, S.; Kim, W. Y. Efficient basin-hopping sampling of reaction intermediates through molecular fragmentation and graph theory. *J. Chem. Theory Comput.* **2014**, *10*, 2419–2426.

(32) Kim, Y.; Kim, J. W.; Kim, Z.; Kim, W. Y. Efficient prediction of reaction paths through molecular graph and reaction network analysis. *Chem. Sci.* **2018**, *9*, 825–835.

(33) Goldsmith, C. F.; West, R. H. Automatic generation of microkinetic mechanisms for heterogeneous catalysis. *J. Phys. Chem. C* **2017**, *121*, 9970 – 9981.

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

(35) Habershon, S. Automated prediction of catalytic mechanism and rate law using graph-based reaction path sampling. *J. Chem. Theory Comput.* **2016**, *12*, 1786–1798.

(36) Habershon, S. Sampling reactive pathways with random walks in chemical space: Applications to molecular dissociation and catalysis. *J. Chem. Phys.* **2015**, *143*, 094106.

(37) Dewyer, A. L.; Argüelles, A. J.; Zimmerman, P. M. Methods for exploring reaction space in molecular systems. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, e1354.

(38) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. ReaxFF: A reactive force field for hydrocarbons. *J. Phys. Chem. A* **2001**, *105*, 9396–9409.

(39) Gai, L.; Shin, Y. K.; Raju, M.; van Duin, A. C. T.; Raman, S. Atomistic adsorption of oxygen and hydrogen on platinum catalysts by hybrid grand canonical Monte Carlo/reactive molecular dynamics. *J. Phys. Chem. C* **2016**, *120*, 9780–9793.

(40) van Spronsen, M. A.; Frenken, J. W. M.; Groot, I. M. N. Surface science under reaction conditions: CO oxidation on Pt and Pd model catalysts. *Chem. Soc. Rev.* **2017**, *46*, 4347–4374.

(41) Ding, K.; Gulec, A.; Johnson, A. M.; Schweitzer, N. M.; Stucky, G. D.; Marks, L. D.; Stair, P. C. Identification of active sites in CO oxidation and water-gas shift over supported Pt catalysts. *Science* **2015**, *350*, 189–192.

(42) Alavi, A.; Hu, P.; Deutsch, T.; Silvestrelli, P. L.; Hutter, J. CO oxidation on Pt(111): An ab initio density functional theory study. *Phys. Rev. Lett.* **1998**, *80*, 3650–3653.

(43) Doye, J. P. K.; Wales, D. J. Global minima for transition metal clusters described by Sutton-Chen potentials. *New. J. Chem.* **1998**, *22*, 733 – 744.

(44) Liu, P.; Rodriguez, J. A. Water-gas-shift reaction on metal nanoparticles and surfaces. *J. Chem. Phys.* **2007**, *126*, 164705.

(45) Magadzu, T.; Yang, J. H.; Henao, J. D.; Kung, M. C.; Kung, H. H.; Scurrell, M. S. Low-temperature water–gas shift reaction over Au supported on anatase in the presence of copper: EXAFS/XANES analysis of gold–copper ion mixtures on TiO_2. *J. Phys. Chem. C* **2017**, *121*, 8812 – 8823.

(46) Davis, R. J.; Derouane, E. G. A non-porous supported-platinum catalyst for aromatization of n-hexane. *Nature* **1991**, *349*, 313–315.

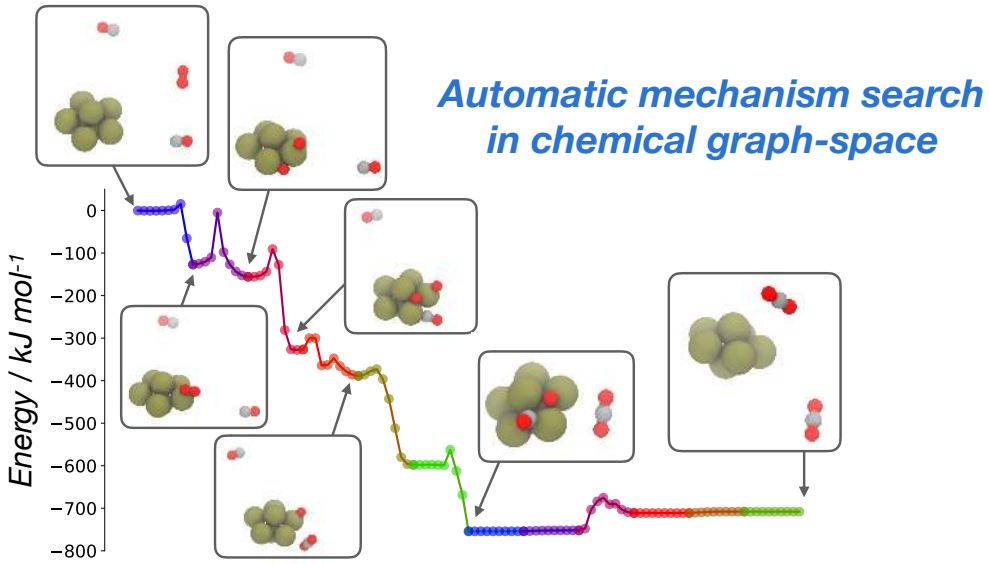
(47) Musselwhite, N.; Na, K.; Sabyrov, K.; Alayoglu, S.; Somorjai, G. A. Mesoporous aluminosilicate catalysts for the selective isomerization of *n*-hexane: The roles of sur-

face acidity and platinum metal. *J. Am. Chem. Soc.* **2015**, *137*, 10231–10237, PMID: 26168190.

(48) Guan, Y.; Ingman, V. M.; Rooks, B. J.; Wheeler, S. E. AARON: An automated reaction optimizer for new catalysts. *J. Chem. Theory Comput.* **2018**, *14*, 5249–5261..

(49) Besora, M.; Maseras, F. Microkinetic modeling in homogeneous catalysis. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, e1372.

(50) Foscato, M.; Venkatraman, V.; Occhipinti, G.; Alsberg, B. K.; Jensen, V. R. Automated building of organometallic complexes from 3D fragments. *J. Chem. Inf. Model.* **2014**, *54*, 1919–1931.


Automatic mechanism search
in chemical graph-space
Energy / kJ mol-1
0
−100
−200
−300
−400
−500
−600
−700
−800

TOC graphic

Biography - Scott Habershon

Scott Habershon is an Associate Professor in Chemistry at the University of Warwick. He studied for a PhD at the University of Birmingham (2001-2004), using genetic algorithms and neural networks to predict molecular crystal structures from powder diffraction data. Subsequently, he undertook post-doctoral research at the California Institute of Technology (2004-2005, with Prof. Ahmed Zewail) and the University of Oxford (2005-2010, with Prof. David Manolopoulos), before being awarded a Leverhulme Trust Early Career Fellowship (University of Bristol, 2010-2012). Since being appointed at Warwick in 2012, his research has focussed on developing simulation tools for predictive chemical dynamics, with particular emphasis on photochemistry, energy transfer and automatic characterization of complex reaction networks.