Simple Position and Orientation Preconditioning Scheme for Minimum Energy Path Calculations.

Christopher Robertson *Scott Habershon †

January 26, 2021

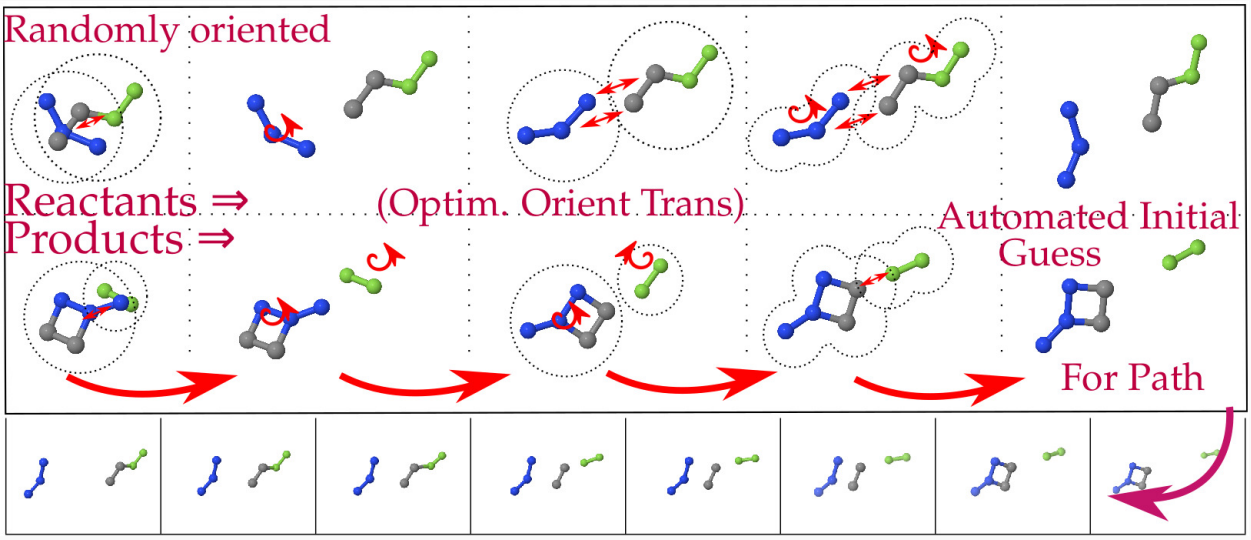
Abstract

Minimum-energy path (MEP) calculations, such as those typified by the nudged elastic band method, require input of reactant and product molecular configurations at initialization. In the case of reactions involving more than one molecule, generating initial reactant and product configurations requires careful consideration of the relative position and orientations of the reactive molecules in order to ensure that the resulting MEP calculation proceeds without converging on an alternative reaction-path, and without requiring excessive numbers of optimization iterations; as such, this initial system set-up is most commonly performed "by hand", with an expert user arranging reactive molecules in space to ensure that the following MEP calculation runs smoothly. In this Article, we introduce a simple preconditioning scheme which replaces this labour-intensive, human-knowledge-based step with an automated deterministic computational scheme. In our approach, initial reactant and product configurations are generated such that steric hindrance between reactive molecules is minimized in the reactant and product configurations, while also simultaneously requiring minimal structural differences between the reactants and products. The method is demonstrated using a benchmark test-set >3400 organic molecular reactions, where comparison of the reactant/product configurations generated using our approach compare very well to initial configurations which were generated on an *ad hoc* basis.

Keywords: Reaction Paths, Nudge Elastic Band, Minimum Energy Paths, Preconditioning Scheme, Reaction Discovery.

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

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



The algorithm starts with reactant and product molecules in any arbitrary orientation and position in space and through several optimization steps, reorients them into a reasonable initial guess for use in double ended, minimum energy path calculations. The top left-most cell shows the reactants (top) and product (bottom) molecules arbitrarily oriented. Successively, each stage represented by the dashed cells performs a different operation on the molecules, resulting in the right-most cell, a reasonable relative initial position and orientation. The resulting reactant and product images can then be used to obtain some approximation to some minimum energy path, as shown in the eight lower cells.

1 INTRODUCTION

There are a number of algorithms in the literature for the optimization of minimum-energy paths (MEPs) between specified reactants and products[1–5], as well as algorithms for providing a reasonable initial guess for these[6–8]. These approaches are enormously powerful tools for providing direct insight into chemical reaction mechanisms, in domains ranging from heterogeneous catalysis to interstellar chemistry, while also affording access to quantitative data which can be directly compared to experimental results; for example, the activation energies determined by MEP calculations can be readily used to estimate the reaction rate constant *via* different approximate methods.[9–11]

However, MEP algorithms can be sensitive to their set-up, with the choice of the initial position and orientation of reactive molecules being particularly important in ensuring smooth convergence to the desired reaction-path. This choice of reactant and product configurations can be problematic for inter-molecular reactions, where the initial relative reactant/product configurations and orientations can both influence which MEP is located using MEP-finding algorithms, but can also influence the efficiency of convergence of MEP calculations.

In many cases, a physically-sensible initial position and orientation of reactive molecules may be known, particular if *ab initio* molecular dynamics simulation (or closely related methods employing reactive potential energy surface [PESs]) are used to sample chemical reactions.[12–15]. However, there are equally situations where prior information about the mechanism, specifically the initial orientation and relative position of reactants, is not available;[16–20] in such cases, the task of preparing the position and orientation of reactants and products can be ambiguous, and in the worst case simply performed "by hand". This "hand-crafted" generation of reactant/product configurations is obviously not ideal, being non-deterministic and reliant on user expert knowledge.

In seeking an alternative to this situation, thereby enabling automated deterministic reactant/product configuration generation, it is useful to note the following "common-sense" rules which are evident in any sensible MEP (described for simplicity as a chain of connected states or images, as in the NEB method):

1. Molecules that will form bonds must lie in each others neighborhood, but not so close that the electronic potential is perturbed.

2. The bond-forming atoms should lie as close as possible (albeit unperturbed), with no other moieties unnecessarily obstructing the direct line of travel between them. In other words, reacting groups in molecules should be facing each other.

3. Any groups of atoms shared in common between the reactant and product structures should be spatially and orientationally correlated along the reaction-path.

4. Molecules that do not share atoms with the reactive molecular species should not occupy the same region of space as the reactive species along the reaction-path.

These simple geometric demands can often uniquely determine or predict the correct positions of reactants and products, even before more detailed analysis of the electronic structure is incorporated. For example, rules (1) and (2) above simply indicate that atoms which are making/breaking bonds in a given reaction must, at some point along the reaction-path, approach each other, and must not be obstructed by other atoms being present. Rule (3) is simply a statement that spectator atoms, which are not directly participating in the reaction, should occupy similar positions and orientations in the reactants and product structures, while rule (4) simply states that molecules which are not directly exchanging fragments, should stay out of each others way. While the "by hand" generation of reactant and product configurations for MEP analysis will invariably rely on these sorts of "common-sense" geometric and chemical ideas, this Article presents an approach which seeks to formalize these rules in a deterministic algorithm.

Of course, it is worth noting that purely geometric assessments can leave ambiguities regarding the orientation of reactants/products that cannot be easily be addressed without evaluating the electronic potential energy. Determining which is the best orientation of "attack" for reactants to adopt is a not trivial problem, and may require substantial computational resources to conclusively determine it. A simple example is the decomposition of formaldehyde into molecular hydrogen and carbon monoxide (see section 3), and whose MEP involves the breaking of the σ_v symmetry of the molecule, something which is not

easily predictable without appealing to its electronic structure. Nevertheless, even for such cases the above geometric rules serve to enable generation of a reasonable starting point for more precise exploration of the orientation of reactants, as we demonstrate below.

The remainder of this Article is organized as follows. Section 2 presents the algorithm, which optimizes a series of simple geometric functions depending on the translational and rotational degrees of freedom (DOF) of the reactive molecules. Section 3 highlights the result of each stage of our algorithm, demonstrating application of our approach to a set of ten reactions from a benchmark set devised by Birkholz and Schlegel[21,22]. Finally, we quantify the degree of reliability of our algorithm by generating initial reactant/product configurations for a large test-set of >3400 organic molecular reactions from the benchmark set devised by Grambow and coworkers.[23] We show that our simple algorithm can position product and reactant molecules in approximately the same place in space as found by using the growing-string method[24–26,30,31] starting from those reactants, for all but a small number ($< 3\%$) of cases.

2 METHODOLOGY

We assume that we have independent geometry-optimized molecular configurations for all of the molecules which are present in the reactant and product configurations of a given reaction. Our aim is to generate total reactant and product configurations, comprising these independent molecules, which represent "good" starting configurations for MEP-finding. The approach taken is to seek the relative positions and orientations which minimizes the steric hindrance between reactant molecules before and after the reaction has taken place, as well as approximately minimizing the overall change in position of all atoms.

We emphasize that our algorithm only acts on the translational and rotational degrees of freedom of the independent reactant and product molecules. In our approach, molecules exert a simple linear attractive or repulsive force so as to minimize steric hindrance; the use of simple parameterized forces means we do not require any computationally-expensive energy evaluations using electronic structure methods. Throughout, we will use the term "image" to describe the collection of molecules constituting either the products or reactants.

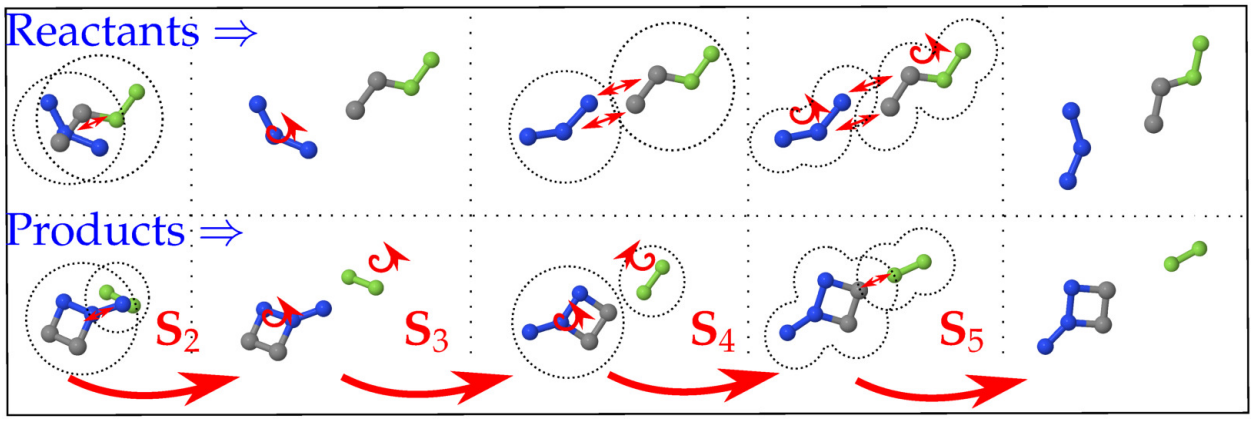


Figure 1: Schematic diagram depicting molecular pre-conditioning algorithm. *From left (Cell 1) to right (Cell 5).* **Cell 1:** The geometric centre of (initially randomly oriented and positioned) reactant molecules are placed near the atoms in their reaction-partner, and the product molecules are placed at the geometric centre of their atoms in the reactant image. The intra-image inter-molecular hard-sphere forces described in stage S_2 are depicted by the red arrows, with approximate hard-sphere radii indicated by dashed circles. **Cell 2:** The results of stage S_2 are shown, with the molecules separated by a hard-sphere-based distance, albeit still randomly oriented. The red curly arrows represent the initial rotation of molecules according to stage S_3. **Cell 3:** The resulting re-oriented molecules from stage S_3 are shown. Further rotational and attractive forces are applied in stage S_4 (represented by the red arrows) to align reactive atoms within the same image (reactants or products), while also seeking to maintain similarity of position and orientation between similar molecules in the reactant and product images. Further hard-sphere forces (as in stage S_2) are also applied in order to maintain appropriate intermolecular distances. **Cell 4:** The optimized positions and orientation of molecules from stage S_4 are shown. Finally, stage S_5 applies the same attractive forces as stage S_4, but changes the repulsive forces from molecular-based hard-sphere forces, to atom-based hard-sphere forces, as depicted by the dashed smaller circles around atoms. This approach better accounts for molecular shape in the final molecular positions. **Cell 5:** The final positions of molecules in reactant and product images are shown; by optimizing the position and orientation of molecular fragments which participate in the reaction, the final reactant and product images should provide sensible (and automated) guesses for further MEP optimization.

Our algorithm will be described in five stages which are summarized below; further mathematical details are given in the appendices. The first three stages, S_1-S_3 (Appendices A.1 to A.3), place reactants and products in space at appropriate initial positions and orientations, as determined by the set of reactive atoms. The following two stages, S_4 and S_5 (Appendices A.4 to A.5), then introduce repulsive and attractive forces in stages that more precisely orient the reactant and product molecules.

A schematic diagram of each stage in our algorithm is shown in Fig. 1. In outline, each stage runs as follows:

- **Stage S_1 (Appendix A.1):** First, in both reactant and product images, we sequentially place the geometric centres of reactive molecules near to their reaction partners.

- **Stage S_2 (Appendix A.2):** Intermolecular "hard-sphere" repulsive forces are used ensures all molecules are sufficiently displaced from each other into a coherent region is space permitted by the original placing of their geometric centres (stage S_1).

- **Stage S_3 (Appendix A.3):** The reactive molecules in the reactant image are oriented such that their reacting-atoms are facing their reaction-partner molecules. In the product image, molecules are reoriented to maximize similarity with the corresponding molecules in the reactant image.

- **Stage S_4 (Appendix A.4):** The translational/rotational DOF of the reactant and product molecules are optimized under the action of two further competing forces: (i) intra-image, inter-molecular forces that attract reactive atoms, and (ii) inter-image, inter-molecular forces which attract and orient molecules that contain common atoms.

- **Stage S_5 (Appendix A.5):** Finally, the translational/rotational DOF of the reactant and product molecules are optimized under the action of refined intra-image, atom-atom hard-sphere forces and inter-image molecular hard-sphere forces to give final positions and orientations of all molecules in both end-point images.

We note that this algorithm involves several geometric operations acting on the reactive molecular species; the equations used are given in the Appendices, as are further implementational details. As will be shown below, the result of this simple geometry-based algorithm

is that reactive molecular species are positioned and oriented in optimal locations to act as initial structures for MEP calculations, such that it is anticipated there are no atomic clashes, or large-scale structural distortions, required on the transit from reactants to products.

3 RESULTS

In this section, we discuss application of our approach to a benchmark set of molecular reactions. First, as an example of our approach in action, we study a single bimolecular reaction in detail (section 3.1); next, we qualitatively demostrate the result on a selection of 10 exemplary reactions (section 3.2); we then report the results of applying our algorithm to a test-set of > 3400 reactions (section 3.3). Comparison to the existing reactant and product structures for this large set of reactions demonstrates that our approach can generate good initial guess structures when compared to those generated either "by hand" or by geometry optimization algorithms.

3.1 Step-by-step example

We first show the action of the key steps of our algorithm with a specific example, namely the reaction of ethyl trifluoroacetate with ammonia; this example is taken from a test-set devised by Zimmerman,[24] It is a reasonably challenging reaction to which to apply our algorithm, having two molecules of approximately similar size in both reactant and products.

The upper cells of Fig. 2 show the positions of molecules in reactant and product images at the end of the different stages described above. We make the following comments:

- $\mathbf{S_1}$: The left-hand cells shows molecules after superimposing their geometric centers according to section A.1. Note that spurious bonds are rendered here due to the proximity of the reactive molecules.

- $\mathbf{S_2}$: The next cell shows the resulting positions of the reactive molecules following optimization under hard-sphere forces (section A.2). Note that molecules are randomly oriented at this stage.

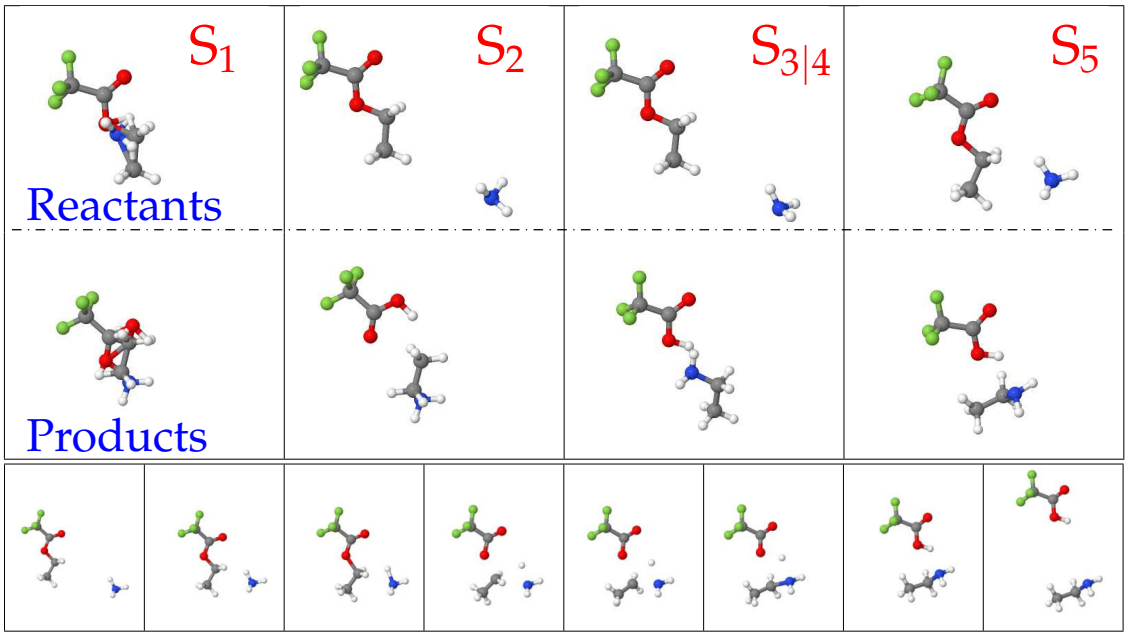


Figure 2: *top cells* Results of different stages (labeled in red) described in the main text for the example of an ethyl transfer reaction between ethyl trifluoroacetate and ammonia. Each cell is described in the text. *bottom cells* An image-dependent pair-potential (IDPP) interpolated reaction-path[6] between resulting reactant and product images, highlighting the suitability of the initial guess provided by this algorithm.

- $\mathbf{S_{3|4}}$: the orientation of molecules and minimization of forces as described in section A.3 and A.4 result in the third top cell. For this example, the minimization step of stage S_4 results in only small adjustments to the original orientation step of section S_3; as a result, the outcome of both steps is shown. Note that reacting atoms are now facing each other (within a given image) and the orientation of corresponding molecules in different images are also very similar.

- $\mathbf{S_5}$: the final atom-based hard-sphere optimization arranges molecules with no close atomic contacts.

The eight frames along the bottom of Fig. 2 shows an image-dependent pair-potential (IDPP) reaction-path[6] obtained from the resulting reactant and products images (noting that two extra images have been added to the two end-point structures to push the moieties further apart). As expected, the resulting IDPP path shows molecules undergoing minimal

changes to their original position and orientation, while also providing a reasonable path for the exchange taking place.

3.2 Further examples

From a set of benchmark reactions provided by Vaucher and Reiher[21], adapted from Birkholz and Schlegel[22], we took the ten intermolecular reactions to demonstrate the results of applying our approach to a variety of different types of reaction classes. The structures for the original datasets were, in turn, obtained from a variety of sources; in some cases, product structures were obtained by geometry optimization of appropriately modified reactant structures using graphical user interface programs,[27] while other examples were obtained from previous benchmark sets, also used for TS optimization algorithms,[28] which were also generally constructed "by hand" from well-known systems and well-informed guesses to their TS structures.

Figure 3 shows the resulting positions of molecules given by our algorithm (the lower set of molecular species in each panel, in gray-scale) while the initial reactant/product positions from the original datasets are shown in color. For our calculations, the position and orientation of all molecules are initially randomized to mimic the conditions for which this algorithm is hoping to be of use; no information from the existing database is used, beyond the identity of the reactive molecules.

It is clear that, in many cases, the initial reactant and product positions given by our approach are nearly indistinguishable from those given in the benchmark datasets. Furthermore, we note that, in the examples where differences are observed (reaction 6-8, and 10), the molecular positions given by our algorithm generally appear more "natural" compared the final position of the atoms than their original placement. As noted in the Introduction, at least for the formaldehyde decomposition (example 6), the formation of H_2 occurs via a distortion which breaks the σ_v symmetry, and the H pointing towards the carbon π cloud likely belies the need for further electronic structure considerations. The suggested position provided by our algorithm is nevertheless a good enough guess such that further NEB calculation can locate the correct MEP.

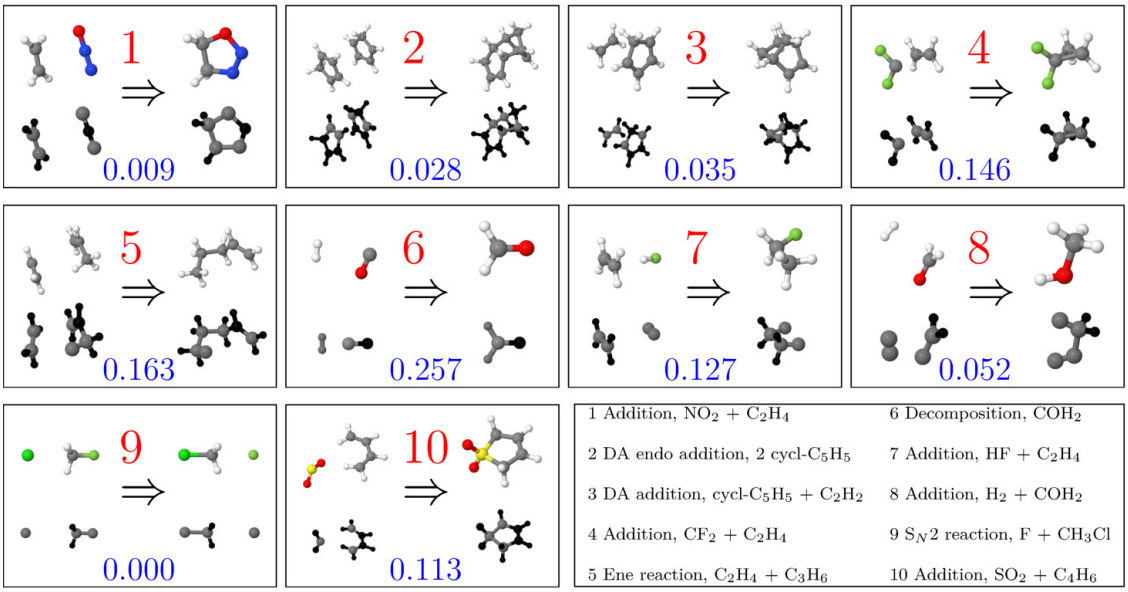


Figure 3: Ten test reactions from Ref.[22]. The initial positions of reactants and product molecules are shown (colored, upper molecular structures in each cell), along with the molecular reactant/product positions given by our algorithm (in gray-scale, lower molecular structures in each panel). In this latter case, gray atoms represent those atoms that form or break bonds, whereas black atoms do not participate in bond changes. Each reaction is indexed in red, and described in the bottom-right cell. In blue text are the calculated d-values of each structure, as described in the text (Eqn. 1).

3.3 Application to large benchmark reaction set

Finally, to obtain a broader quantitative description of the reliability of our algorithm, we have devised a single measure that tries to quantify the difference between the initial reactant/product molecular positions given by our algorithm and those molecular positions provided in the original benchmark datasets. Here. we consider a large database of reactions devised by Grambow and coworkers.[23] This dataset comprises 12,000 reactions of organic molecules derived by taking a subset of the GDB-7 data set. The growing string algorithm[24–26] was then employed on this subset to generate a large set of plausible elementary reactions with at most seven heavy C/N/O heteroatoms, and evenly distribuited molecular sizes therewithin, using an accurate DFT method (ωB97X-D3/def2-TZVP). Of these 12,000 reaction we found 3,435 which were *inter* molecular reactions (as opposed to and intramolecular reaction such as isomerization). For each of these 3,435 reactions, we used our proposed algorithm to generate initial reactant and product structures.

In order to give a quantitative comparison between structures generated by our algorithm and those available from other methods (*e.g.* geometry optimization, or growing-string method), we devise a simple figure-of-merit. The simplest approach to comparing two sets of molecular structures would be to evaluate the norm of the difference vector between the reactant/product images determined by our algorithm and the database. The problem with this simple approach is that the relative positioning of molecules can skew the results. Provided that molecules lie far enough each other that they are not significantly interacting, the interpolated path should be essentially the same for some sensible range of possible intermolecular distances; such trivial difference in intermolecular distances would then appear as a larger norm in the difference vector, despite not actually being meaningfully different.

To obtain a measure for both these features, we directly compare the orientation and position (with respect to the laboratory frame) of the reactant/product images determined by our approach to those provided in the benchmark dataset. It is first necessary to align the reactant+product images from this algorithm with those of the benchmark data. This can be achieved using the Kabsch algorithm, which provides the rotation matrix that minimizes the Frobenius norm of the difference between the position vectors of the reactant+product

super-system provided by our algorithm and those given in the benchmark data. After aligning these reactant/product structures, we calculate the following difference measure, d, where vectors are subscripted either "A" or "B" depending on whether they refer to the structures given by our algorithm or the benchmark data, respectively:

$$
\begin{aligned}
d &= \frac{1}{2}(d_{gom} + d_{ori}) \quad N = N^R + N^e \\
d_{gom} &= N^{-1} \sum_{j=P,R} \sum_{m}^{N^j} \frac{1 - \vec{g}^A_{jm} \cdot \vec{g}^B_{jm}}{2|\vec{g}^A_{jm}||\vec{g}^B_{jm}|} \\
d_{ori} &= N^{-1} \sum_{j=P,R} \sum_{m}^{N^j} \frac{1 - (\vec{r}^A_{jm} - \vec{\mathfrak{g}}^A_{jm}) \cdot (\vec{r}^B_{jm} - \vec{\mathfrak{g}}^B_{jm})}{2|\vec{r}^A_{jm} - \vec{\mathfrak{g}}^A_{jm}||\vec{r}^B_{jm} - \vec{\mathfrak{g}}^B_{jm}|}
\end{aligned}
\tag{1}
$$

The definitions of the geometric variables in Eq. 1 are given in the Appendices. In brief, however, these equations define an average of $\cos(\theta)$, where θ are the angles between

- the geometric centre vector of each molecule m in A with that in B (second equation, d_{gom}).
- the position vector (centered at the origin) between each molecule m in A with that in B (third equation, d_{ori}).

Each separate term in Eq. 1 contributes a value of zero if these vectors are perfectly aligned, or a value of one if they are exactly pointing in opposite directions; a value of $d = 0.5$ is the average expected for completely random orientation and positions of all molecules.

The numbers highlighted in blue in Fig. 3 are the resulting d-values for each test reaction. We note that the d measure captures the correct trend in suitability of the generated reactant and product structures. For example, it is clear that application of our algorithm to examples 1-3 all give reactant and product structures which seem to be very well-aligned with the initial configurations from the database, and show similarly small d-values; on the other hand, as we have already noted, the initial configuration for example 6 generated by our approach is quite different to that from the database, and also demonstrates a corresponding larger value of d.

With this measure of similarity in hand, we then applied our configuration pre-conditioning algorithm to a set of 3,435 multi-molecule reactions taken from the Grambow dataset;[23] Fig.

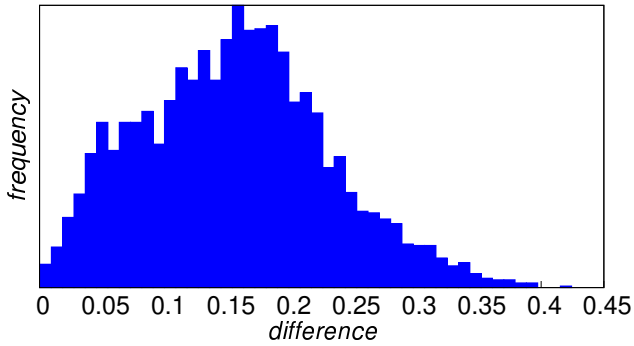


Figure 4: Histogram showing distribution of d-values (given by Eq. 1) between the benchmark dataset and our algorithm. A value of $d = 0$ corresponds to a perfect match to the benchmark data, while $d = 0.5$ corresponds to a random orientation of molecular species relative to the benchmark dataset.

4 illustrates the histogram of d values calculated for this large test-set. We find that the large majority ($\approx 88\%$) of d-values lie below 0.25, suggesting that our algorithm generates initial configurations which are similar to those seen in Fig 3 of subsection 3.2. Encouragingly, these results demonstrate that our algorithm is obtaining initial reactant/product structure which are comparable to those obtained by alternative methods, such as "by hand", by conformation searching, or by geometry optimization. We find that just 3% of reactions considered have d values greater than 0.3; in such cases, the position/orientation of one of the molecules is typically different enough from the benchmark data such that it is unlikely that one can expect the same resulting MEP. However, in the vast majority of cases, the generated initial reactant/product structures appear to be of high quality, and at least comparable to those obtained by other methods (although emphasizing that our approach does not require any electronic structure calculations).

4 DISCUSSION AND CONCLUSIONS

We have presented and tested a relatively-straightforward and free-standing pre-conditioning algorithm for generating initial reactant/product molecular configurations in preparation for geometry optimization and MEP-finding calculations. In brief, this algorithm uses both hard-sphere forces and rigid-molecule rotations in attempting to optimally position reaction

partners such that the reactive atoms are unobstructed, while also seeking to maintain proximity to the original input molecular configurations of the separate reactive molecules. Our algorithm does not require any potential energy evaluations (for example, using *ab initio* methods). In detailed tests to 10 benchmark reactions, as well as in quantitative comparison to a large set of $>$ 3400 benchmark reactions, we find that the initial reactant/product configurations generally appear suitable for further analysis in MEP-finding.

Of course, the algorithm is not fool-proof; as we have noted, around 3% of reactions in the large dataset exhibited relatively significant differences compared to the existing initial configurations. Given the simplicity of our algorithm, it seems likely that additional orientational penalty functions which introduce further "chemical knowledge" will help address these. Nevertheless, our approach generally presents as a suitable starting point for a more systematic re-orientation of the reacting molecules that incorporate better physically motivated non-covalent interactions, while also being consistent with the many available reaction-path interpolation schemes.[6–8] Both datasets tested in Section 3 are restricted to relatively small organic molecules, but it is worth noting that we have obtained good results with larger systems comprising of a few tens of atoms, larger than the bicyclic molecules such as test 2 of Section 3.2. Our algorithm has also been used already to help generate a reaction network used in a study the carbon monoxide oxidation of small platinum nanoparticles[34]. In that study we found the algorithm was also able to cope with tri-molecular reactions, for which no examples where available in the large dataset tested in section 3.

Most relevant to our work, we anticipate that this algorithm may be of particular utility in the context of automated, high-throughput schemes to attempt *de novo* design of potential molecular catalysts[17,19,20,25,29,30,32–36]. As well as preconditioning the position and orientation of molecules, deciding whether the product and reactant molecules require conformational intermediates and providing reasonable guesses (prior to PES evaluations) of these would be a natural next step to implement, and for which work is currently underway.

4.1 ACKNOWLEDGMENTS

The authors gratefully acknowledge funding for this project from the Engineering and Physical Sciences Research Council (EP/R020477/1). Computational resources were provided by

the Scientific Computing Research Technology Platform at the University of Warwick. Data from Figs. 3 and 4 can be found at http://wrap.warwick.ac.uk/.....

A Numerical Details

The definitions of vectors and sets which will be used throughout our algorithm are presented here:

- Variables of the image corresponding to reactant molecules will be labeled by R and the image corresponding to the products by P, although it should be noted that for the purposes of this algorithm the R and P labels are interchangeable.
- The number of molecules in image $R|P$ are given by $N^{R|P}$ respectively.
- Molecules will be referenced by the image label and molecule index, *e.g.* Rn labels molecule n for reactant image R.
- The set of atom indices belonging to molecule Rn will be labeled as $M^{Rn} = \{M^{Rn}\}$.
- The identity of an atom in the reactants image is preserved across to the product image. Namely $\bigcup_m^{N^R} M^{Rm} \equiv \bigcup_m^{N^P} M^{Pm}$.
- The positions of atoms in molecule Rn will be represented by $\vec{r}_{Rn}$, a vector of $3|M^{Rn}|$ dimensions. A superscript $a|b$ will denote the atomic index of the coordinates, $\vec{r}^{\,a}_{Rn}$.
- The geometric centre (average position) of molecule Rn will be labeled by

$$\vec{g}_{Rn} = \frac{1}{|M^{Rn}|} \sum_{a \in M^{Rn}} \vec{r}^{\,a}_{Rn}$$

  .
- The vector of $|M^{Rn}|$ repeating $\vec{g}_{Rn}$ is given by $\vec{\mathfrak{g}}_{Rn}$; this is a vector of length $3|M^{Rn}|$.
- The sub-set of atom indices from molecule Rn which are in molecule Pm after reaction is labeled C^{Rnm}, where the second index (*i.e.* n) specifies the molecule in the image

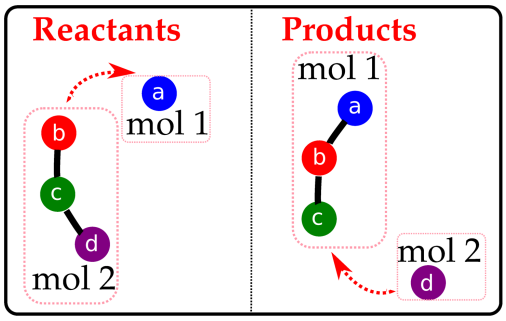


Figure 5: Diagram aiding the definition of sets. **A Sets:** $A^{R12} = a, A^{R21} = b, A^{P12} = c, A^{P21} = d$. **B Sets:** $B^{R11} \equiv B^{P11} = a, B^{R22} \equiv B^{P22} = b$. **C Sets:** $C^{R21} \equiv C^{P12} = a, b$.

structure (first index *i.e.* R) to which these atoms belong (similarly for C^{Pxy}). Note that $C^{Rxy} \equiv C^{Pyx} \equiv M^{Rx} \cap M^{Py}$.

- The sub-set of atom indices of C^{Rnm} for those atoms which are actually involved in bond breaking/forming is given by B^{Rnm}, where the second index (*i.e.* n) specifies the molecule to which these atoms belong. Note that $B^{Rxy} \equiv B^{Pyx}$.

- The sub-set of atom indices belonging to molecule Rn which *will form bonds* to atoms in molecule Rm (bonded in the product image P) is given by A^{Rnm}, where the second index (*i.e.* n) specifies the molecule to which these atoms belong (similarly defined A^{Pxy}).

- The geometric centre of the set of atom $G^{Rn} = \bigcup_m^{N^R} A^{Rnm}$, namely all atoms in molecule Rn which will form bonds, is labeled

$$\vec{\gamma}_{Rn} = \frac{1}{|G^{Rn}|} \sum_{a \in G^{Rn}} \vec{r}^{\,a}_{Rn}$$

.

Figure 5 shows a simple diagram, the caption presenting all sets of type A B and C for it.

A.1 S_1 Initial positioning of reactant and product molecules

First, we position molecules in the reactant and product images at approximate initial locations by shifting geometric centres. For image R, we initially place each molecule at the

origin: $\vec{g}_{Rm} = 0$. This is followed by sequentially re-positioning $\vec{g}_{Rm}$ at the vector $\vec{\alpha}_{Rm}$, given by

$$\vec{\alpha}_{Rm} = \frac{1}{N^R} \sum_{n \neq m}^{N^R} \frac{1}{|A^{Rnm}|} \sum_{a \in A^{Rnm}} \vec{r}^{\,a}_{Rn}$$

, which is the geometric centre of atoms in image R that molecule Rm will participate in bond changes.

For the product structure P, we shift the positions of molecules to $\vec{g}_{Pm} = \frac{1}{2}(\vec{\beta}_{Pm} + \vec{\sigma}_{Pm})$ where

$$\vec{\beta}_{Pm} = \frac{1}{N^R} \sum_{n}^{N^R} \frac{1}{|B^{Rnm}|} \sum_{a \in B^{Rnm}} \vec{r}^{\,a}_{Rn}$$

is the geometric centre of atoms in the reactant configuration R with which molecule Pm will form bonds, and $\vec{\sigma}_{Pm}$ is given by

$$\vec{\sigma}_{Pm} = \frac{1}{N^R} \sum_{n}^{N^R} \frac{1}{|C^{Rnm}|} \sum_{a \in C^{Rnm}} \vec{r}^{\,a}_{Rn}$$

, which is the geometric centre of atoms in reactant configuration R that belong to molecule Pm after reaction.

A.2 S_2 Intra-image inter-molecular hard-sphere force

The positioning described above can leave molecules overlapping each other. To separate them and remove steric clashes, we use a molecular hard-sphere force $\vec{f}^{\,1}_{Rn}$ (affecting only the translational degrees-of-freedom) to any molecule Rn that is too close to other molecules $Rm, m \neq n$, where:

$$\begin{aligned}
\vec{f}^{\,S_2.1}_{Rn} &= \sum_{a \in M^{Rn}} \hat{a} \left[\vec{f}^{\,a}_{Rn} \right] \\
\vec{f}^{\,a}_{Rn} &= \sum_{m \neq n}^{N^R} \phi_{RmRn} \cdot H \cdot (\vec{g}_{Rm} - \vec{g}_{Rn}) / |\vec{g}_{Rm} - \vec{g}_{Rn}| \\
\phi_{RmRn} &= \frac{\kappa_{S_2.1}}{N} \{ |\vec{g}_{Rm} - \vec{g}_{Rn}| - (h^{Rm} + h^{Rn}) \} \\
H &= \begin{cases} 1 & |\vec{g}_{Rm} - \vec{g}_{Rn}| < (h^{Rm} + h^{Rn}) \\ 0 & |\vec{g}_{Rm} - \vec{g}_{Rn}| > (h^{Rm} + h^{Rn}) \end{cases} \qquad N = 3|M^{Rn}| \sum_{m \neq n}^{N^R} H
\end{aligned} \tag{2}$$

Here, $\vec{f}_{Rn}^{S_{2}.1}$ is the $3|M^{Rn}|$ total force vector acting on the geometric centre of molecule Rn and $\vec{f}_{Rn}^{a}$ is the hard-sphere force on atom a in M^{Rn}. Geometry optimization under the hard-sphere forces is simple (and fast) to perform numerically; we also note that the same optimization process is applied to all molecules in both $R|P$ images.

The norm of $\vec{f}_{Rn}^{S_{2}.1}$ is independent from the number of atoms involved and proportional to the force-constant $\kappa_{S_{2}.1}$ by construction. The parameters $h^{Rm|Rn}$ are estimates of the molecular radius of $Rm|Rn$ respectively; this estimate is calculated as the mean distance of each atom with respect to the geometric centre, plus two standard deviations away from this mean.

The outcome of optimization under $\vec{f}_{Rn}^{S_{2}.1}$ (and similarly $\vec{f}_{Pm}^{S_{2}.1}$) is that all molecules in the reactant and product images are displaced to remove overlapping atoms.

A.3 S_3 Initial orientation of molecules

The previous two steps ensure molecules are not overlapping and are placed at appropriate distances to avoid steric clashes. The next step in our algorithm focuses on the orientations of reactive molecules.

For image R only, we rotate each molecule Rn using the matrix that that acts on the vector $\vec{\gamma}_{Rn} - \vec{g}_{Rn}$ to minimize the difference with $\vec{\alpha}_{Rn} - \vec{g}_{Rn}$. This will orient molecule Rn such that any atoms involved in reaction are directed towards their reaction partners in other molecules.

For image P only, we then rotate each molecule Pm (let the 3×3 rotation matrix be $\mathbf{R}^{Pm}$) to minimize the difference between $\vec{r^{0}}_{Pm} = \vec{r}_{Pm} - \vec{\mathfrak{g}}_{Pm}$ and the following average vector $\vec{\mu}_{Pm}$:

$$\begin{aligned} \vec{\mu}_{Pm} &= \sum_{n}^{N^{R}} \frac{|C^{Rnm}|^{2}}{N} \cdot \vec{r}^{\,0}_{Pmn} \qquad N = \sum_{n}^{N^{R}} |C^{Rnm}|^{2} \\ \vec{r}^{\,0}_{Pmn} &= \sum_{a \in M^{Pm}} \hat{a} \left[\mathbf{R}^{PmRn} (\vec{r^{a}}_{Pm} - \vec{g}_{Pm}) \right] \end{aligned} \tag{3}$$

The second equation above defines $\vec{r}^{\,0}_{Pmn}$ as the rotated vector of $\vec{r^{0}}_{Pm}$. The rotation matrix is labeled $\mathbf{R}^{PmRn}$ and minimizes the difference between $\vec{r}^{\,0}_{Pmn}$ and $\vec{r^{0}}_{Rn}$ for the subset of atoms $C^{Rnm} \equiv C^{Pmn} \equiv M^{Rn} \cap M^{Pm}$. In other words, $\vec{\mu}_{Pm}$ is a weighted sum of vectors $\vec{r}^{\,0}_{Pmn}$, where

each is $\vec{r}^{\,0}_{Pm}$ rotated in such a way as to maximize its orientation with the fragment of atoms shared with molecule $\vec{r^0}_{Rn}$.

The rotation matrix $\mathbf{R}^{Pm}$ acts on all atoms in M^{Pm} and does not affect the internal DOF. We shall now describe how to calculate $\mathbf{R}^{PmRn}$ and $\mathbf{R}^{Pm}$. Any rotation matrix which acts on a vector, in these cases $\vec{r}^{\,0}_{Pm}$, to minimize the geometric difference with another vector (*i.e.* $\vec{r^0}_{Rn}$ or $\vec{\mu}_{Pm}$) can be obtained as follows. First, move all vectors to the origin and shape them as $3 \times N$ matrices: $\mathbf{r^0}_{Pm}, \mathbf{r^0}_{Rn}$, where $N = |M^{Pm}|, |M^{Rn}|$; The rotation matrix which maximizes the trace of the overlap matrix $(\mathbf{r^0}_{Pm})^T \mathbf{r^0}_{Rn}$ is then given by the Kabsh procedure.[37]

A.4 S_4 Alignment of reactive atoms

We next apply translational and torsion forces to each molecule to place atoms near those atoms in other molecules with which they will ultimately form bonds.

Here, for each molecule Rn, we calculate two average Euclidean vectors, namely a translational vector $\vec{v}^{\,t}_{Rn}$ and and a rotational $\vec{v}^{\,r}_{Rn}$, defined as:

$$\begin{aligned}
\vec{v}^{\,t}_{Rn} =& N^{-1} \sum_{m \neq n}^{N^R} \sum_{a \in A^{Rnm}} \sum_{b \in A^{Rmn}} \\
& |(\vec{r}^{\,b}_{Rm} - \vec{r}^{\,a}_{Rn}) \cdot (\vec{r}^{\,a}_{Rn} - \vec{g}_{Rn})| \frac{(\vec{r}^{\,b}_{Rm} - \vec{r}^{\,a}_{Rn})}{|\vec{r}^{\,b}_{Rm} - \vec{r}^{\,a}_{Rn}|} \\
\vec{v}^{\,r}_{Rn} =& N^{-1} \sum_{m \neq n}^{N^R} \sum_{a \in A^{Rnm}} \sum_{b \in A^{Rmn}} (\vec{r}^{\,b}_{Rm} - \vec{r}^{\,a}_{Rn}) \times (\vec{r}^{\,a}_{Rn} - \vec{g}_{Rn}) \\
N =& 3|M^{Rn}| \sum_{m \neq n}^{N^R} |A^{Rmn}||A^{Rnm}|.
\end{aligned} \tag{4}$$

Here, $\times$ denotes cross-product and the superscripts ${}^a|^b$ denote atom indices. Each term in the first equation is a vector in the direction from atom a to b, with length equal to the dot product between (i) the position vector of atom a from the geometric centre of its molecule and, (ii) the distance vector between atoms a to atom b. Similarly the second equation takes the cross product between the same two vectors, quantifying their orthogonality, and contributes to the torsional force $\vec{v}^{\,r}_{Rn}$, acting upon molecule Rn. N is the total number of elements in the sums. These average vectors are then used to define translational and

rotational forces for the entire molecule, $\vec{f}_{Rn}^{S_4.1}$:

$$\begin{aligned}\vec{f}_{Rn}^{S_4.1} &= \sum_{a \in M^{Rn}} \hat{a}\left[\vec{f}_{Rn}^{a}\right] \\ \vec{f}_{Rn}^{a} &= \kappa_{S_4.1} \cdot \left\{-\vec{v}_{Rn}^{r} \times (\vec{r}_{Rn}^{a} - \vec{g}_{Rn}) + \vec{v}_{Rn}^{t}\right\}\end{aligned} \tag{5}$$

Ideally any given atom in the reactant configuration R should be close to itself in product configuration P, but without affecting the "integrity" of molecules. Each molecule in the reactant image R will contain atoms that belong to molecules in the product image P. Consequently it is important that reactant molecules are placed near their correlated product molecules. Furthermore, it is also important that the orientations are comparable; if these criteria are not satisfied, reactive fragments will have to undergo (potentially) unnecessary rotation and translation during MEP optimization. To enforce these conditions, we take a similar approach as Eq. 5 and calculate two average Euclidean vectors, namely the translational vector $\vec{w}_{Rn}^{t}$ and rotational vector $\vec{w}_{Rn}^{r}$ defined as:

$$\begin{aligned}\vec{w}_{Rn}^{t} =& N^{-1} \sum_{m}^{N^P} \sum_{a \in C^{Rnm}} \sum_{b \in C^{Pmn}} \\ & \phi \cdot |(\vec{r}_{Pm}^{b} - \vec{r}_{Rn}^{a}) \cdot (\vec{r}_{Rn}^{a} - \vec{g}_{Rn})| \frac{(\vec{r}_{Pm}^{b} - \vec{r}_{Rn}^{a})}{|\vec{r}_{Pm}^{b} - \vec{r}_{Rn}^{a}|} \\ \vec{w}_{Rn}^{r} =& N^{-1} \sum_{m}^{N^P} \sum_{a \in C^{Rnm}} \sum_{b \in C^{Pmn}} \\ & \phi \cdot (\vec{r}_{Pm}^{b} - \vec{r}_{Rn}^{a}) \times (\vec{r}_{Rn}^{a} - \vec{g}_{Rn}) \\ N =& 3|M^{Rn}| \sum_{m}^{N^P} |C^{Rnm}||C^{Pmn}|, \qquad \phi = \begin{cases} 1 & a \in B^{Rnm} \\ \frac{1}{2} & a \notin B^{Rnm} \end{cases}\end{aligned} \tag{6}$$

Note that we have weighted atoms that are going to form bonds more strongly (*i.e.* through the ϕ factor) because it is important these atoms are ultimately close in space. Again, these two vectors are used to define a forces, in analogy as Eqn. 5, but using a force parameter $\kappa_{S_4.2}$ The forces presented in this subsection, as well as the hard-sphere forces from the previous section A.1, are then numerically minimized.

A.5 S_5 Refinement of atomic positions using further hard-sphere forces

It was useful to apply the hard-sphere repulsive molecular forces in sections A.1 and A.4 because it allowed free rotation of molecules into an optimized orientation before attempting to bring them into closer proximity. In this final stage, we remove the hard-sphere forces operating within the same image (R or P; section A.1) and re-apply these forces to those molecules Rn and Pm such that $C^{Rnm} = \emptyset$ (in other words, molecules which do not share or exchange atoms). This ensures that those molecules which are not exchanging atoms are not situated in the same region in space. The force constant associated with this new force is labeled $\kappa_{S_5.1}$.

Furthermore, in this final stage, repulsive forces are applied within the same image (R or P) which better account for the geometry of the molecules. In the same spirit as the forces introduced in section A.4, these impart both a torque and a translational force determined from the average of atom-atom hard sphere forces. These inter-molecular repulsive forces can be interpreted as hard sphere interactions centered on each atom of each molecule. The atom-atom hard-sphere radius dependent thresholds, c_{ab}, are proportional to the average of the typical covalent radius for this atom pair ab.

The two average translation and rotation vectors, $\vec{z}^{\,t}_{Rn}$ and $\vec{z}^{\,r}_{Rn}$, for molecule Rn are:

$$
\begin{aligned}
\vec{z}^{\,t}_{Rn} =& N^{-1} \sum_{m \neq n}^{N^R} \sum_{a \in M^{Rn}} \sum_{b \in M^{Rm}} \\
& H \cdot |(\vec{y}^{\,ab}_{Rnm}) \cdot (\vec{r}^{\,a}_{Rn} - \vec{g}_{Rn})| \frac{\vec{y}^{\,ab}_{Rnm}}{|\vec{y}^{\,ab}_{Rnm}|} \\
\vec{z}^{\,r}_{Rn} =& N^{-1} \sum_{m \neq n}^{N^R} \sum_{a \in M^{Rn}} \sum_{b \in M^{Rm}} \\
& H \cdot (\vec{y}^{\,ab}_{Rnm}) \times (\vec{r}^{\,a}_{Rn} - \vec{g}_{Rn}) \\
\vec{y}^{\,ab}_{Rnm} =& c_{ab} \cdot \frac{\vec{r}^{\,b}_{Rm} - \vec{r}^{\,a}_{Rn}}{|\vec{r}^{\,b}_{Rm} - \vec{r}^{\,a}_{Rn}|} - (\vec{r}^{\,b}_{Rm} - \vec{r}^{\,a}_{Rn}) \\
N =& |M^{Rn}|^2 \sum_{m \neq n}^{N^R} \sum_{a \in M^{Rn}} \sum_{b \in M^{Rm}} H \\
H =& \begin{cases} 1 & |\vec{r}^{\,b}_{Rm} - \vec{r}^{\,a}_{Rn}| < \phi \cdot c_{ab} \\ 0 & |\vec{r}^{\,b}_{Rm} - \vec{r}^{\,a}_{Rn}| > \phi \cdot c_{ab} \end{cases}, \qquad \phi = \begin{cases} \frac{3}{2} & a \in A^{Rnm} \\ 2 & a \notin A^{Rnm} \end{cases}
\end{aligned}
\tag{7}
$$

These repulsive forces are "switched on" when the intermolecular atoms approach each other

Table 1: Values of force constants κ used at the different stages of our algorithm.

stages:	S_2	S_4	S_5
$\kappa_{S_2.1}$	1.0	50.0	0
$\kappa_{S_4.1}$	0	1.0	1.0
$\kappa_{S_4.2}$	0	1.0	3.0
$\kappa_{S_5.1}$	0	0	50.0
$\kappa_{S_5.2}$	0	0	2.0

below the threshold distance c_{ab}. We note that the factor ϕ modifies the hard-sphere radius between atoms depending upon whether the atoms ab are bonding or not.

As above, these two vectors define the direction along which the molecules experience a torque and translational force, as in Eqn. 5 but with a force-constant $\kappa_{S_5.2}$. Unlike the other forces, we note that the normalization constant N here depends on $(M^{Rn})^2$, making $|\vec{f}_{Rn}^{S_5.2}|$ dependent on the size of the molecule (and not simply proportional to $\kappa_{S_5.2}$); larger molecules will thus experience a smaller repulsive force against smaller ones (while the smaller molecules experience a reciprocally stronger force). These two new forces (associated with force-constants $\kappa_{S_5.1}$ and $\kappa_{S_5.2}$), together with those of the previous two subsections, are minimized numerically.

A.6 Other considerations

Because the evaluation of these forces have insignificant computational expense, we use simple steepest descents for all optimizations. The values for the parameters of the different forces at the different stages are shown in Table 1. The relatively high values for $\kappa_{S_2.1}$ and $\kappa_{S_5.1}$ reflect their role as "hard-sphere" forces, compared to the other force constants which are meant to only partially exert their influence. These values give reasonable results for all examples tested to date.

Though the torque formulas in the preceding subsections do not by themselves change the internal degrees-of-freedom of the system for sufficiently small displacements, their linear combination can change the internal degrees-of-freedom. To ensure that the sum of forces at the different optimization stages remains within the space of rotations and translations of the

molecules, a projection operator was applied to the final force at every minimization step. This is an orthogonal projection operator $\mathbf{P} = \mathbf{V}(\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T$ where $\mathbf{V}$ is a $(6 \cdot N^R) \times 3N$ matrix with each of the $6 \cdot N^R$ columns containing a vector representing translation and rotation vectors for each molecule (N being the total number of atoms in images R—P). For systems with a high-degree of group symmetry it is occasionally found the above equations "get stuck" on a saddle point. It was found that a simple application of a small random force on the first few hundred iterations was sufficient to break the symmetry of the forces and find an adequate minimum guess. Note that the repeated construction of this projection matrix on every optimization step turns the linear complexity of the algorithm into a cubic one. However, it should be noted there are ways circumventing the use of these matrices, if application to systems with more degrees-of-freedom was desired. The resulting reactant R and product P images are finally used to perform an IDPP interpolation[6] with a generous number of intermediate images. The centre-of-mass of each image on the resulting path are then shifted to zero and the norm of the overlap matrix of image j and $j-1$ is maximized by rotating image j using the Kabsch algorithm (using mass-scaled coordinates). All images then have the same centre of mass and inertial frame. Finally to ensure that the molecules at the reaction end-point are sufficiently far apart from each other, we linearly translate each molecule Rm in the direction $-\vec{\alpha}_{Rm}$ (similarly for P) by some small distance (~ 2 Å).

References

1. G. Mills H. Jónsson and K. W. Jacobsen. In *Class. Quan. Dyn. Cond. Sim.*, pages 385–404. World Scientific, 1998.

2. G. Henkelman, B.P. Uberuaga, and H. Jónsson. *J. Chem. Phys.*, 113(22):9901–9904, 2000.

3. M. N. Groves E. L. Kolsbjerg and B. Hammer. *J. Chem. Phys.*, 145(9):094107, 2016.

4. V. Asgeirsson A. Vehtari O. P. Koistinen, F. B. Dagbjartsdottir and H. Jonsson. *J. Chem. Phys.*, 147(15), 2017.

5. P. M. Zimmerman. *J. Chem. Theo. Comp.*, 9(7):3043–3050, 2013.

6. K. Stokbro S. Smidstrup, A. Pedersen and H. Jónsson. *J. Chem. Phys.*, 140(21):214106, 2014.

7. X. L. Zhu, K. C. Thompson, and T. J. Martinez. *J. Chem. Phys.*, 150(16), 2019.

8. R. McGibbon V. S. Pande L. Wang, A. Titov and T. J. Martinez. *J. Chem. Theo. Comp.*, 12(2):638–649, 2016.

9. F. Montalenti A. F. Voter and T. C. Germann. *Ann. Rev. Mat. Res.*, 32:321–346, 2002.

10. E. Pollak and P. Talkner. *Chaos*, 15(2), 2005.

11. A. F. Voter. *Rad. Eff. Sol.*, 235:1–23, 2007.

12. R. McGibbon F. Liu V. S. Pande L. Wang, A. Titov and T. J. Martinez. *Nature Chem.*, 6(12):1044–1048, 2014.

13. S. A. Vazquez G. L. Barnes J. P. Stewart A. Rodriguez, R. Rodriguez-Fernandez and E. Martinez-Nunez. *J. Comp. Chem.*, 39(23):1922–1930, 2018.

14. S. Habershon. *The J. Chem. Phys.*, 143(9):094106, 2015.

15. V. Dufour-Decieux C.A. Sing-Long R. Freitas E. Chen, Q. Yang and E. J. Reed. *J. Phys. Chem. A*, 123(9):1874–1881, 2019.

16. M. A. Watson-T. F. Hughes D. Rinaldo S. Ehrlich T. B. Steinbrecher S. Vaitheeswaran D. M. Philipp M. D. Halls L. D. Jacobson, A. D. Bochevarov and R. A. Friesner. *J. Chem. Theo. Comp.*, 13(11):5780–5797, 2017.

17. Y. V. Suleimanov and W. H. Green. *J. Chem. Theo. Comp.*, 11(9):4248–4259, 2015.

18. I. Ismail, C. Curtis C. Robertson H. B. V. A. Stuttaford-Fowler, A. Ochan, and S. Habershon. *The J. Phys. Chem. A*, 123(15):3407–3417, 2019.

19. Z. Kim Y. Kim, J. W. Kim and W. Y. Kim. *Chem. Sci.*, 9(4):825–835, 2018.

20. J. W. Kim K. Lee and W. Y. Kim. *Chem. Sys. Chem.*, 2:e1900057–67, 2020.

21. A. C. Vaucher and M. Reiher. *J. Chem. Theo. Comp.*, 14(6):3091–3099, 2018.

22. A. B. Birkholz and H. B. Schlegel. *J. Comp. Chem.*, 36(15):1157–1166, 2015.

23. L. Pattanaik C. A. Grambow and W. H. Green. *Scientific Data*, 7(1), 2020.

24. P. M. Zimmerman. *J. Chem. Theo. Comp.*, 9(7):3043–3050, 2013.

25. P. M. Zimmerman. *Mol. Sim.*, 41(1-3):43–54, 2015.

26. P. M. Zimmerman. *J. Comp. Chem.*, 36(9):601–611, 2015.

27. C. Y. Peng, P. Y. Ayala, H. B. Schlegel, and M. J. Frisch. *J. Comp. Chem.*, 17(1):49–56, 1996.

28. J. Baker and F. R. Chan. *J. Comp. Chem.*, 17(7):888–904, 1996.

29. C. Poree and F. Schoenebeck. *Acc. Chem. Res.*, 50(3):605–608, 2017.

30. A. J. Arguelles A. L. Dewyer and P. M. Zimmerman. *Wiley Interd. Rev.-Comp. Mol. Sci.*, 8(2), 2018.

31. A. L. Dewyer and P. M. Zimmerman. *Organic & Biomolecular Chemistry*, 15(3):501–504, 2017.

32. K. Y. Baek K. Lee J. W. Kim, Y. Kim and W. Y. Kim. *J. Phys. Chem. A*, 123(22): 4796–4805, 2019.

33. J. P. Unsleber and M. Reiher. *Ann. Rev. Phys. Chem., Vol 71*, 71:121–142, 2020.

34. I. Ismail C. Robertson and S. Habershon. *Chem. Sys. Chem.*, 2(4):e1900047, 2020.

35. C. Robertson and S. Habershon. *Catal. Sci. Technol.*, 9:6357–6369, 2019.

36. S. Habershon. *J. Chem. Theory Comput.*, 12(4):1786–1798, 2016.

37. W. Kabsch. *Acta. Crystallog. Sec. A*, 32(SEP1):922–923, 1976.