Fast and automatic estimation of transition state structures using tight binding quantum chemical calculations (#50973)

First submission

Guidance from your Editor

Please submit by **3 Aug 2020** for the benefit of the authors (and your $200 publishing discount) .

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Raw data check
Review the raw data.

Image check
Check that figures and images have not been inappropriately manipulated.

Privacy reminder: If uploading an annotated PDF, remove identifiable information to remain anonymous.

Files

Download and review all files from the materials page.

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- All underlying data have been provided; they are robust, statistically sound, & controlled.
- Speculation is welcome, but should be identified as such.
- Conclusions are well stated, linked to original research question & limited to supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

Fast and automatic estimation of transition state structures using tight binding quantum chemical calculations

Maria H Rasmussen [1], **Jan H Jensen** [Corresp. 1]

[1] Department of Chemistry, University of Copenhagen, Copenhagen, Denmark

Corresponding Author: Jan H Jensen
Email address: jhjensen@chem.ku.dk

We present a method for the automatic determination of transition states (TSs) that is based on Grimme's RMSD-PP semiempirical tight binding reaction path method (*J. Chem. Theory Comput*. 2019, 15, 2847-2862), where the maximum energy structure along the path serves as an initial guess for DFT TS searches. The method is tested on 100 elementary reactions and located a total of 89 TSs correctly. Of the 11 remaining reactions, nine are shown not to be elementary reaction after all and for one of the two true failures the problem is shown to be the semiempirical tight binding model itself. Furthermore, we show that the RMSD-PP barrier is a good approximation for the corresponding DFT barrier for reactions with DFT barrier heights up to about 30 kcal/mol. Thus, RMSD-PP barrier heights, which can be estimated at the cost of a single energy minimisation, can be used to quickly identify reactions with low barriers, although it will also produce some false positives.

Fast and automatic estimation of transition state structures using tight binding quantum chemical calculations

Maria H. Rasmussen[1] **and Jan H. Jensen**[1]

[1]**Department of Chemistry, University of Copenhagen, Copenhagen, Denmark**

Corresponding author:
Jan H. Jensen[1]

Email address: jhjensen@chem.ku.dk, Twitter @janhjensen

ABSTRACT

We present a method for the automatic determination of transition states (TSs) that is based on Grimme's RMSD-PP semiempirical tight binding reaction path method (*J. Chem. Theory Comput.* 2019, 15, 2847-2862), where the maximum energy structure along the path serves as an initial guess for DFT TS searches. The method is tested on 100 elementary reactions and located a total of 89 TSs correctly. Of the 11 remaining reactions, nine are shown not to be elementary reaction after all and for one of the two true failures the problem is shown to be the semiempirical tight binding model itself. Furthermore, we show that the RMSD-PP barrier is a good approximation for the corresponding DFT barrier for reactions with DFT barrier heights up to about 30 kcal/mol. Thus, RMSD-PP barrier heights, which can be estimated at the cost of a single energy minimisation, can be used to quickly identify reactions with low barriers, although it will also produce some false positives.

INTRODUCTION

The computational determination of chemical reaction networks[1;2;3;4;5;6] requires that the estimation of barrier heights and/or location of transition states (TSs) be automated. Many methods for automated barrier height estimation and TS location have been proposed.[7;8;9;10;11;12;13;2] However, the computational demand of these methods are significantly higher than for locating minima.

Recently, Grimme[14]. presented a method (RMSD-PP) for the rapid estimation of reaction paths based on a semiempirical tight-binding model (GFN2-xTB[15;16]). The predicted path can be used in a barrier estimate and the maximum energy structure as a TS guess in more expensive methods. Here, the performance on both are tested. This method is attractive to use when screening large amounts of reactions, as it is not much more expensive than a geometry optimization and the GFN2-xTB method has been parameterised for the entire periodic table up to Z = 86. However, for it to be practically useful it needs to work in an automated framework.

The paper is organized as follows. First, the automated procedure for locating transition states is presented. Then, the method is applied to 100 elementary reactions suggested by Zimmerman[13;17] and lastly we conclude on the results.

METHOD

The idea behind the RMSD-PP method is to add a Gaussian biasing potential "pushing" the molecule away from the reactant structure and a Gaussian biasing potential "pulling" the molecule towards the product structure. A geometry optimization should (provided that good parameters for sizes and widths of the biasing potentials are used) take the reactant structure to the product structure along the minimum energy path. The path is further refined by 2-4 optimization steps without any bias at every point on the

path.

Figure 1 shows a flowchart of the automated procedure for locating transition states (TSs). The reactant and product structures with same atomic ordering are required as input. The procedure starts with an RMSD-PP path search run with respective k_{pull} and k_{push} values of -0.02 and 0.01 E_h and an α of 0.6 $1/a_0$ (parameter set 1, Table S1). In addition to this run, two additional runs are performed where the k_{pull} and k_{push} values are multiplied by 1.5 and 2.25. The number of optimization steps used to refine the GFN2-xTB path is fixed to 3 in all runs. A run is deemed successful if the root mean square deviation (RMSD) of the end structure compared to the product structure is less than 0.3 Bohr and the reaction path with the smallest absolute values of k_{pull} and k_{push} are selected. If the reaction does not complete, the setup for the path search is changed: the last structure of the run is saved and used as product structure in the next run while the product structure is used as reactant structure (trial 2, parameter set 1, Table S1). The same procedure is then repeated for trials 3-4 and 5 (Table S1) until completion is achieved. If all five attempts fail, then the entire procedure is repeated with an electronic temperature of 6000 K (increased from 300 K). If the reaction again fails to complete then the method is deemed to fail for the reaction, although we did not observe this for the reactions considered in this paper. We also test a slightly different parameter set (parameter set 2, Table S1), where k_{push} is lowered to 0.008 E_h for the first try.

Once the reaction has completed and the path found, the maximum energy structure along the path is extracted along with the two neighbouring structures. A linear interpolation (10 points from maximum energy structure to both neighbours) is performed and the interpolated structures are subjected to single point energy calculations using both Density Functional Theory (DFT) and GFN2-xTB. All DFT calculations are performed with the Gaussian 16 program[18]. The maximum GFN2-xTB energy along the interpolated path is used to estimate the GFN2-xTB barrier (orange part of the flow chart, Figure 1). The maximum energy structure based on DFT calculations is used as initial guess for the TS structure in a DFT TS search [*opt=(calcall, ts, noeigen)*]. Whether the correct TS is found is evaluated based on an intrinsic reaction coordinate (IRC) path search in both forward and reverse direction from the found TS. From the endpoint structures of the IRC, the adjacency matrices are extracted. The adjacency matrix for an N atom system is an $N \times N$ matrix with 1 on the off-diagonal elements linking atoms that are bonded and 0 if the atoms are not bonded. The structures are converted from coordinates to adjacency matrix using *xyz2mol*.[19] The assignment of bond/no bond is done using the xyz2mol program based on a simple extended Hückel theory (EHT) calculation and the Mulliken overlap population between each pair of atoms as implemented in `RDKit`[20]. The adjacency matrices for the endpoints of the IRC are compared with the adjacency matrices for the intended reactant and product structures to determine if a TS for the intended reaction is found. If the adjacency matrices of the IRC endpoint structures do not match those of the input reactant and product structures it may be due to the IRC not haven completed as the IRC calculations often crash before converging to reactant/product structures. Thus, the endpoints of the IRC are geometry optimized, and these structures checked by the same procedure. If either sets of structures (based on adjacency matrices) match the input structures, the TS for the given reaction is concluded to have been found and the search procedure terminated.

If the IRC did not result in a path connecting the input reactant and product, a constrained optimization on the TS guess, obtained as the maximum energy structure of the interpolated structures, is performed. The bond constraints are set up automatically by considering the difference in adjacency matrices of input reactant and product structure, resulting in a set of bonds being formed/broken during the reaction. only connectivity changes are considered, meaning that, e.g., going from a double bond to a single bond is not considered bond breaking. The length of the set of bonds are fixed to the values in the guess structure from the interpolation, and the remaining structure relaxed. The new TS guess is taken through the same procedure with TS optimization, IRC and check. If the TS is still not found, the entire procedure is repeated but using an electronic temperature of 6000 K in the RMSD-PP reaction path step.

Dataset

To test the TS localizer protocol, a preexisting data set from the literature is chosen to avoid bias in the choice of reactions studied. The data set used by Zimmerman to test his double-ended GSM, consisting

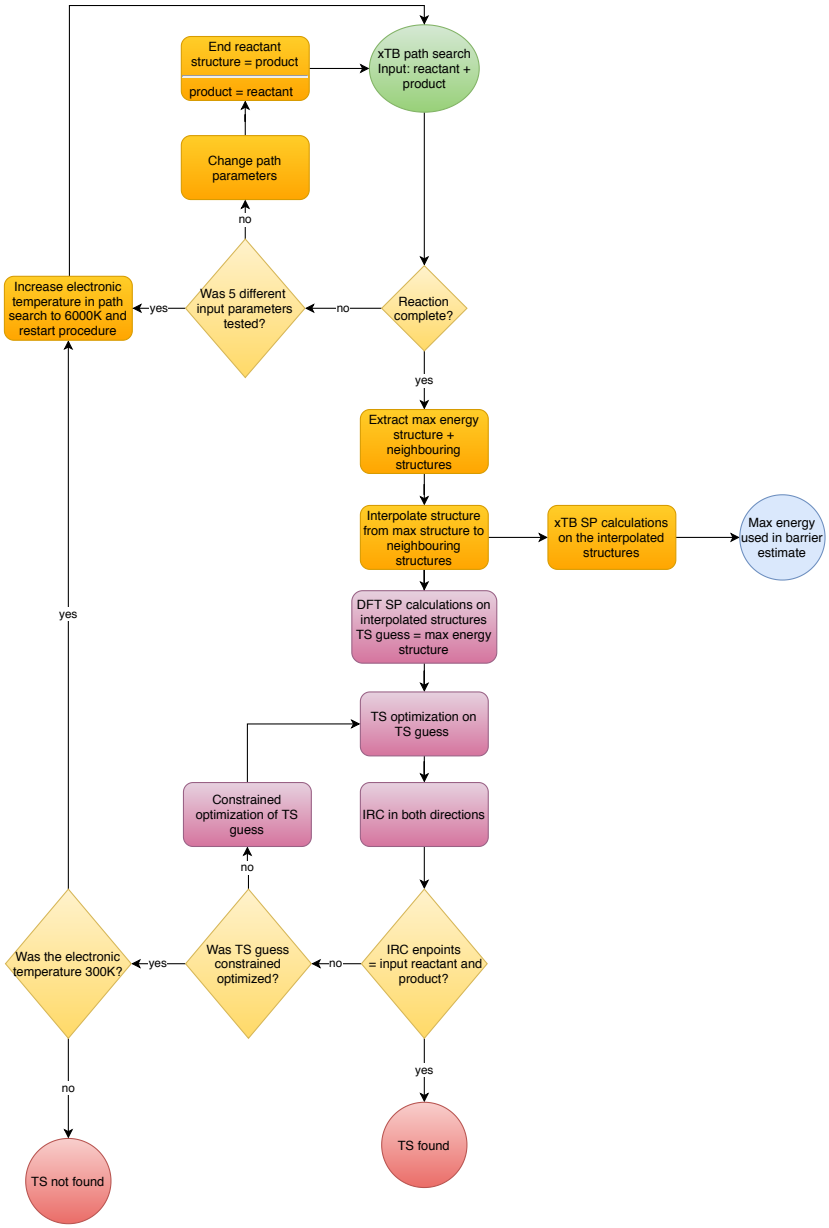


Figure 1. Flowchart describing the automated workflow implemented. Orange steps depend solely on GFN2-xTB calculations, while purple steps rely on DFT calculations

of 105 elementary reactions is used[17;13]. Only reactions of neutral molecules and reactions where bond breaking/formation take place are included (i.e. excluding conformational changes). Thus, the test set consists of 100 elementary reactions including both simple and complicated reactions with between 1 and 6 bond changes (Table S2). To be able to use the TSs located by Zimmerman, the same level of theory for the DFT part of the procedure is used: UB3LYP/6-31G**[21;22;23;24;25].

All reactant and product structures were reoptimised using GFN2-xTB to verify that the structures have corresponding minima on the GFN2-xTB potential energy surface. This is the case for all reactions but reaction 16, as discussed further below. The DFT geometries for the reactant and product are used as input for the procedure described above.

Approximate TS validation procedures

A popular approach in automated TS procedures is to either skip the IRC step and use alternative validation procedures for the TS or first screen the TS with alternative validation procedures before doing the IRC in an effort to save computational time[26;27;13]. Though the TS validation here is based on the IRC path and whether it connects the reactant and product, some of these alternative approaches are also tested. In particular, the TS vetting requirements suggested by Jacobson *et al.*[26] are tested. The three requirements are: 1) There should be exactly 1 imaginary frequency of the Hessian, 2) at least one of the active bonds (bonds being broken or formed during the reaction) should have an intermediate length, and 3) that the eigenvector corresponding to the imaginary frequency should have motion along at least one of the active bond stretching modes. We use the same cutoff values for when a bond length is considered intermediate and when it is considered that the eigenvector has motion along a bond stretching mode as in the original article, that is: A bond length r_{ij} between atom i and j is considered intermediate if

$$1.2 \leq \frac{r_{ij}}{r_i^{cov} + r_j^{cov}} \leq 1.7 \tag{1}$$

where r_i^{cov} is the covalent radius of atom i[20]. The eigenvector corresponding to the imaginary frequency, $\boldsymbol{v}^{TS}$ is considered to move along the stretching mode of bond i, $\boldsymbol{v}_i^{stretch}$ (unit vector), if the absolute value of the scalar projection of $\boldsymbol{v}^{TS}$ on $\boldsymbol{v}_i^{stretch}$ is larger than 0.33:

$$|\boldsymbol{v}_i^{stretch} \cdot \boldsymbol{v}^{TS}| \geq 0.33 \tag{2}$$

RESULTS AND DISCUSSION

Success rate

For each of the 100 reactions, the procedure is run three times with two different but similar parameter strategies for the xTB path calculations (Table S1) for a total of 6 runs. The reason for running three times per parameter set is that the RMSD-PP procedure includes a random ”initial distortion parameter” which can lead to slightly different reactions paths for each run.

Figure 2 shows the distribution of success rates for each of the 6 runs. Run 1-3 are with the same parameters (parameter set 1 in Table S1) and run 4-6 are with the same parameters (parameter set 2 in Table S1). The parameter sets are almost identical, the only difference is that the first run in parameter set 2 is initiated with a smaller push strength. The total number of successes is quite similar within the 6 runs (ranging between 81 and 83 TS located) and the majority of the TSs are located using the guess structure from the RMSD-PP path directly. Combining all TSs located during the 6 runs, a total of 89 TSs are found. For the first parameter set (run 1-3) 85 TSs are located and for the second parameter set (run 4-6) 88 of the TSs are located. It is possible, that exploring a larger part of the parameter space allows localization of the last reactions.

For the reactions not located by the procedure (reactions 6, 10, 11, 16, 20, 35, 54, 68, 84, 90, and 96), the TS structures proposed by Zimmerman was further analysed. However, they were first put through the same IRC validation procedure (with and without reoptimization of the TS). Only two of the remaining 11 reactions (reactions 16 and 84) went to minima corresponding to the proposed reactant and product

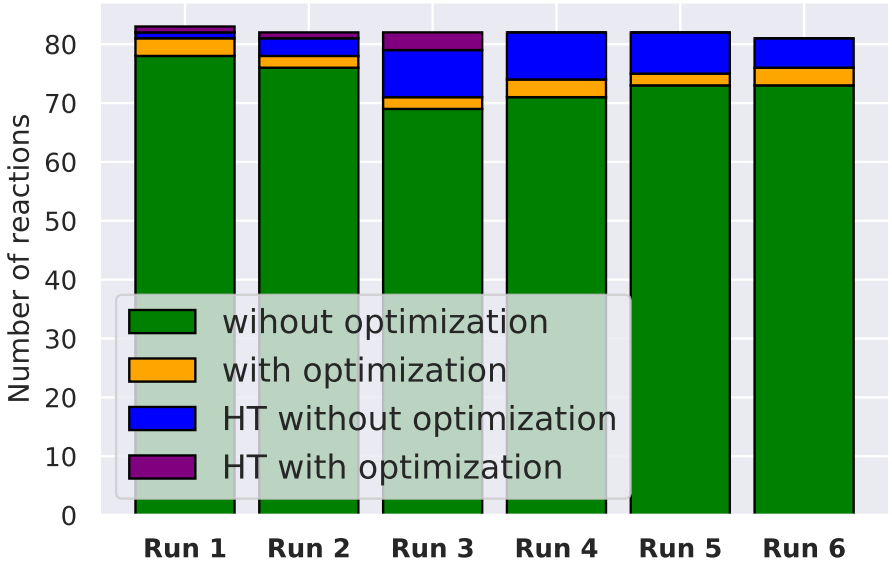


Figure 2. Distribution of the successful TSs localized for each of the 6 runs. The xTB TS guess structure is first used directly (without optimization) in a UB3LYP/6-31G** TS optimization. If that fails to find the TS, a constraint optimization is done and the TS optimization tried again. Finally, for the failed searches, the entire procedure is run at a higher electronic temperature (HT) of 6000 K. Run 1-3 is done with parameter set 1 and Run 4-6 done with parameter set 2 (SI).

structures, while the majority of the 9 reactions found an intermediate minimum structure along the way, indicating that the reaction (at least within this level of theory) is not an elementary reaction. The 9 reactions are not used in the following analysis, where the data set is now reduced to 91 reactions (89 of which the procedure managed to locate a TS for).

The two reactions, for which the TS search was unsuccessful, are shown in Figure 3. The product in reaction 16 was the only structure that reacted when optimized with GFN2-xTB. After optimization the product became NH_3 + BH_3 + NH_2BH_2 and the product is thus not stable on the GFN2-xTB potential energy surface, which can affect the path optimization and thus the TS guess. During the DFT TS optimization the TS guess structure instead goes to the TS of reaction 9 (Table S2), which has a ≈ 8 kcal/mol lower barrier than reaction 16. The other reaction not found, reaction 84, is a pretty simple reaction and it is not clear why the TS of this reaction would be difficult to locate. Instead the TS of the reaction in Figure 4 is found every time. Comparing the found TS with the true TS (Figure 5) shows that the TSs are quite similar. The important difference seems to be the orientation of the methylene group in the middle.

Comparison of xTB barrier estimates and DFT barriers

In this section we test whether the RMSD-PP reaction path can be used to distinguish reactions that have high and low barriers at the DFT level. If so, the RMSD-PP method could be used in the high throughput determination of reaction networks, where one is usually interested in relatively low-energy barriers. The 91 reactions for which a DFT TS is found, can be used to calculate the barrier of the reactions, which can be compared to the very cheap barrier estimates from the GFN2-xTB path. The barrier is calculated as the electronic energy of the TS (or maximum energy along the GFN2-xTB path) minus the electronic energy of the reactant. The reactant structures used were the same in both DFT and GFN2-xTB calculations (from [13]). This can affect the RMSD-PP barriers especially if either reactant or product structures are not stable on the GFN2-xTB surface as this can affect the path. All reactant and product structures were optimized with GFN2-xTB and only the product of reaction 16 changed bonding during optimization.

Figures 6(a), 6(c) and 6(e) show the correlation between the barrier estimated with GFN2-xTB and that calculated by DFT for the first parameter set. For each point is indicated the pull strength (color)

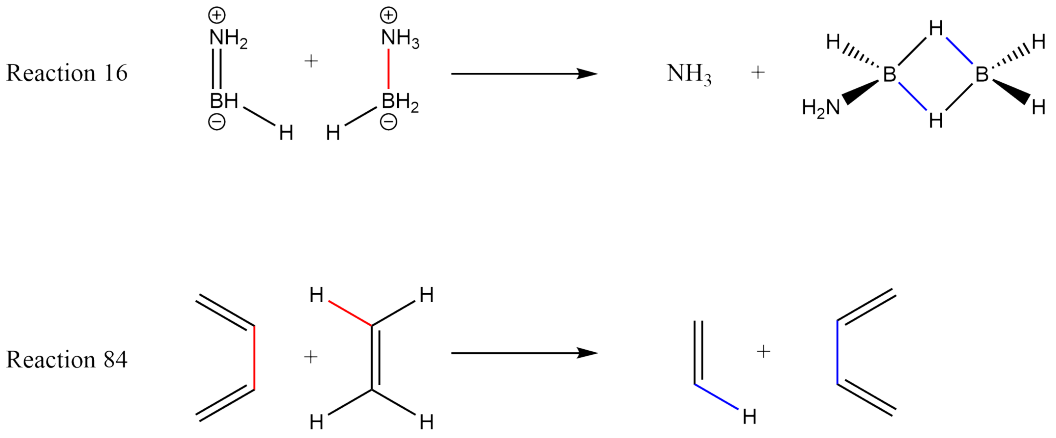


Figure 3. The two reactions not found by the procedure

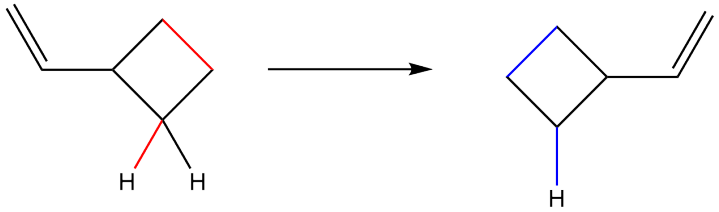


Figure 4. The reaction of the TS located when searching for the TS for reaction 84

and the push strength (size). Reactions, where the search was unsuccessful are labelled with red edges. Similarly, Figures 6(b), 6(d) and 6(f) show the GFN2-xTB barrier estimate vs. DFT barriers for the three runs with parameter set 2. As one would expect, higher pull and push values are needed for higher barriers. The mean absolute error (MAE) is between 14.9 and 19.2 kcal/mol for all six runs, and there is a wide spread of values and several outliers. So, generally speaking, the xTB barrier from the RMSD-PP reaction path is a poor estimate of DFT barrier heights. However, in many reaction network studies the goal is to identify reactions that proceed at measurable rates at room temperature, which translates into barrier heights of no more than 30 kcal/mol. The correlation between xTB and DFT is considerably better for these reactions. Reactions where the xTB barrier is less than 40 kcal/mol, includes all seven reactions with DFT barriers less than 30 kcal/mol, in addition to 14-20 false positives (14, 17, 20, 14, 17, and 15 for runs 1-6) where the DFT barrier is higher than 30 kcal/mol. If one excludes points where the absolute pull values are higher than 0.03 then the number of false positives drops to 11-14 (12, 11, 14, 11, 12 and 11 for runs 1-6).

TS validation procedure

Here we test the performance of the validation procedures described in the Methods section: 1) exactly 1 imaginary frequency of the Hessian, 2) at least one of the active bonds (bonds being broken or formed during the reaction) has an intermediate length, and 3) the eigenvector corresponding to the imaginary frequency has motion along at least one of the active bond stretching modes. The tests are applied to both the correct (83) and incorrect (8) TSs located during run 1. For the incorrect TSs, the first TS found (without constrained optimization) is used in the analysis. The outcome of the individual tests along with the combination of all three tests is shown in Figure 8(a) for the found transition states of run 1 and in Figure 8(b) for the failed transition states of run 1.

An effective validation procedure should discard as many wrong TSs as possible while not removing true transition states. The requirement, that the found transition state should have exactly 1 imaginary frequency is fulfilled for all 83 found TSs, but is also fulfilled for all but 1 (TS optimization failed) of the wrong transition states. Though the requirement can be applied without fear of throwing away true TSs, it is not very effective in filtering out wrong TSs. The requirement, that the TS structure should have at

(a) (b)

Figure 5. Comparison of the TS located ((a)) searching for the TS of reaction 84 ((b)). (a) taken from the first run, coordinates for (b) from SI of the work of Zimmerman[13]

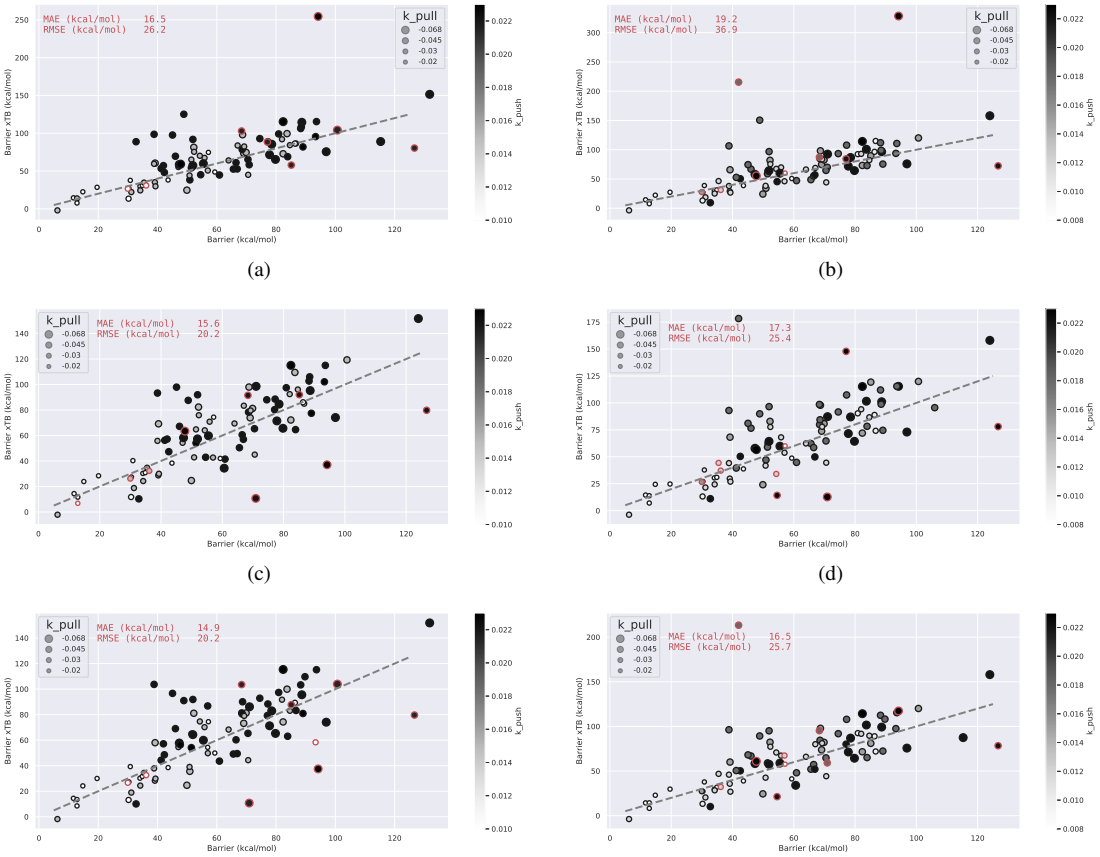


Figure 6. Barrier estimate from xTB compared to DFT (UB3LYP/6-31G**) barriers for the 6 runs shown in Figure 2. k_pull and k_push values are given per atom. For each point is indicated the pull strength (color) and the push strength (size). Reactions, where the search was unsuccessful are labelled with red edges. Figures (a), (d), and (e) correspond to runs 1, 2, and 3, respectively.

least one of the active bonds at an intermediate distance is fulfilled for 77 out of 83 true transition states and not fulfilled for three out of eight wrong transition states. Thus, applying this validation test to the

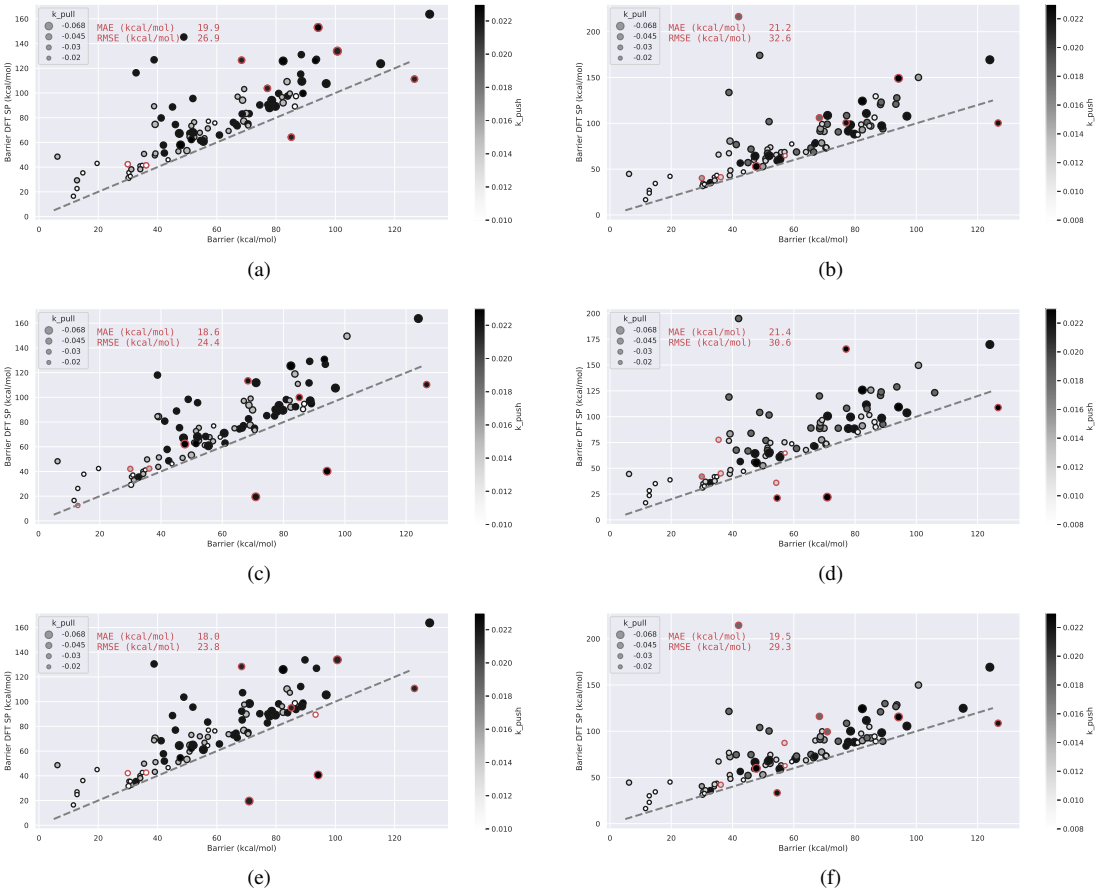


Figure 7. Barrier estimate from DFT single points compared to DFT (UB3LYP/6-31G**) barriers for the 6 runs shown in Figure 2. k_pull and k_push values are given per atom. For each point is indicated the pull strength (color) and the push strength (size). Reactions, where the search was unsuccessful are labelled with red edges. Figures (a), (d), and (e) correspond to runs 1, 2, and 3, respectively.

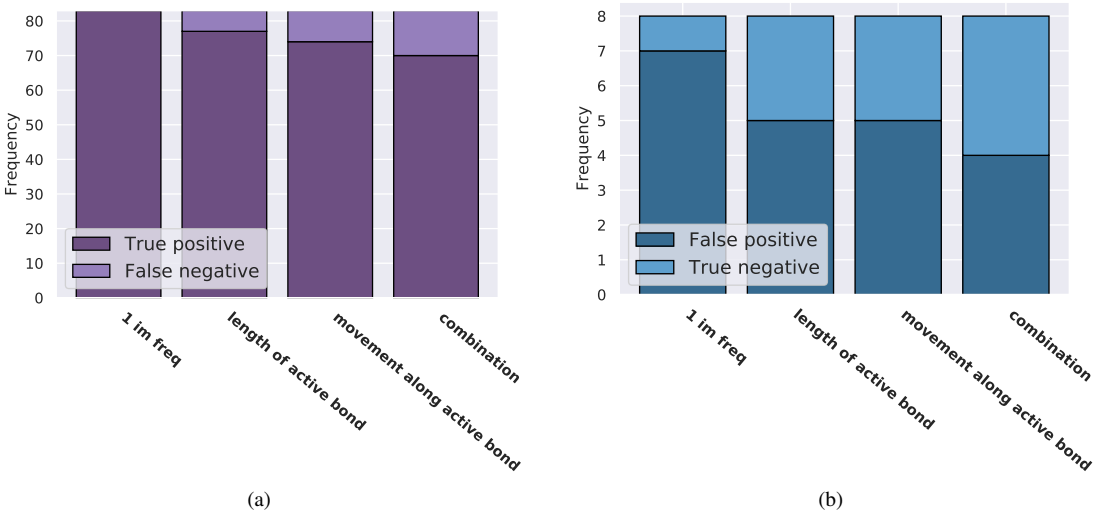


Figure 8. Test of different TS validation methods for the (a) True TSs ; and (b) Wrong TSs of run 1.

transition state structures would have resulted in six correct TSs being filtered out. The last validation test, that the displacement vector of the imaginary frequency should be along at least one of the active bonds given the cutoff value above, is not fulfilled for nine of the transition states confirmed to be true by an IRC. Requiring all three validation tests to be fulfilled would have resulted in 13 of the 83 true transition states to have been filtered out. Four out of eight of the wrong transition states would also have been filtered out, but one needs to be very careful when applying these alternative validation tests, considering whether the saved computational time is worth more than the wrongly rejected transition states.

SUMMARY

We present a method for the automatic determination of transition states (TSs) that is based on Grimme's RMSD-PP method[14] for the rapid estimation of reaction paths using the GFN-xTB semiempirical tight binding models (Figure 1). The RMSD-PP method estimates a reaction path between reactants and products by a geometry optimisation using an energy function augmented by a Gaussian biasing potential that "pushes" and "pull" the structure away from the reactant and towards the product. Our method starts with a series of RMSD-PP calculations with increasingly larger push and pull strengths until reaction completion. The additional structures near the highest point on the reaction path are generated by interpolation and used for DFT single points and the highest energy structure is then used as an initial guess for a TS search. Upon convergence the TS is tested by an IRC calculation and if the TS is found to be incorrect then the initial guess structure is reoptimised with key bond lengths constrained and used as an initial guess for a new TS search. If that fails the *entire* procedure is repeated but with using an electronic temperature of 6000K for the RMSD-PP calculations.

The method is tested on 100 elementary reactions used previously by Zimmerman and co-workers (Table S2).[17;13] For each of the 100 reactions, the procedure is run three times with two different but similar parameter strategies for the xTB path calculations (Table S1) for a total of 6 runs. Combining all TSs located during the six runs, a total of 89 TSs are found. Only two of the remaining 11 reactions (reactions 16 and 84) went to minima corresponding to the proposed reactant and product structures, while the majority of the 9 reactions found an intermediate minimum structure along the way, indicating that the reaction (at least within this level of theory) is not an elementary reaction. Thus our method failed for only two reactions (Figure 3), where the product is not a stable structure on the xTB potential energy surface.

Furthermore, we show that the RMSD-PP barrier is a good approximation for the corresponding DFT barrier for reactions with DFT barrier heights up to about 30 kcal/mol. Thus, RMSD-PP barrier heights, which can be computed at the cost of a single energy minimisation, can be used to quickly identify reactions with low barriers, although it will also produce some false positives.

Finally, we show that various tests of whether the correct TSs have been found, produce several false positives and false negatives and should be used with care.

ACKNOWLEDGMENTS

MHR is supported by a research grant (00022896) from VILLUM FONDEN

REFERENCES

[1] S. Maeda, K. Ohno, K. Morokuma, *Phys. Chem. Chem. Phys.* **2013**, *15*, 3683–3701.
[2] P. L. Bhoorasingh, B. L. Slakman, F. Seyedzadeh Khanshan, J. Y. Cain, R. H. West, *J. Phys. Chem. A* **2017**, *121*, 6896–6904.
[3] Y. Kim, J. W. Kim, Z. Kim, W. Y. Kim, *Chem. Sci.* **2018**, *9*, 825–835.
[4] J. P. Unsleber, M. Reiher, *Annu. Rev. Phys. Chem.* **2020**, *71*, 121–142.
[5] C. Robertson, I. Ismail, S. Habershon, *ChemSystemsChem* **2019**, 2607.
[6] Y. V. Suleimanov, W. H. Green, *J. Chem. Theory Comput.* **2015**, *11*, 4248–4259.
[7] G. Mills, H. Jónsson, *Physical review letters* **1994**, *72*, 1124.

[8] H. Jónsson, G. Mills, G. Schenter, *Surface Science* **1995**, *324*, 305–337.
[9] G. Henkelman, B. P. Uberuaga, H. Jónsson, *The Journal of chemical physics* **2000**, *113*, 9901–9904.
[10] E. Weinan, W. Ren, E. Vanden-Eijnden, *Physical Review B* **2002**, *66*, 052301.
[11] E. Weinan, W. Ren, E. Vanden-Eijnden, *J. Phys. Chem. B* **2005**, *109*, 6688–6693.
[12] B. Peters, A. Heyden, A. T. Bell, A. Chakraborty, *The Journal of chemical physics* **2004**, *120*, 7877–7886.
[13] P. Zimmerman, *J. Chem. Theory Comput.* **2013**, *9*, 3043–3050.
[14] S. Grimme, *J. Chem. Theory Comput.* **2019**, *15*, 2847–2862.
[15] C. Bannwarth, S. Ehlert, S. Grimme, *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671.
[16] S. Grimme, C. Bannwarth, P. Shushkov, *J. Chem. Theory Comput.* **2017**, *13*, 1989–2009.
[17] P. M. Zimmerman, *J. Comput. Chem.* **2013**, *34*, 1385–1392.
[18] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, *Gaussian˜16 Revision A.03*, **2016**, Gaussian Inc. Wallingford CT.
[19] J. H. Jensen, *xyz2mol*, `https://github.com/jensengroup/xyz2mol`.
[20] *RDKit: Open-source cheminformatics*, `http://www.rdkit.org`.
[21] A. D. Becke, *Phys. Rev. A Gen. Phys.* **1988**, *38*, 3098–3100.
[22] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B Condens. Matter* **1988**, *37*, 785–789.
[23] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652.
[24] R. Ditchfield, W. J. Hehre, J. A. Pople, *J. Chem. Phys.* **1971**, *54*, 724–728.
[25] W. J. Hehre, R. Ditchfield, J. A. Pople, *J. Chem. Phys.* **1972**, *56*, 2257–2261.
[26] L. D. Jacobson, A. D. Bochevarov, M. A. Watson, T. F. Hughes, D. Rinaldo, S. Ehrlich, T. B. Steinbrecher, S. Vaitheeswaran, D. M. Philipp, M. D. Halls, R. A. Friesner, *J. Chem. Theory Comput.* **2017**, *13*, 5780–5797.
[27] C. Grambow, L. Pattanaik, W. H. Green.

SUPPORTING INFORMATION

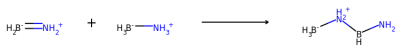
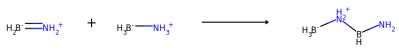
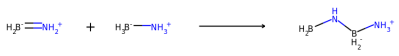
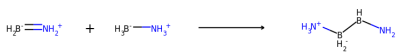
The code and data resulting from this study can be found here https://github.com/jensengroup/RMSD_PP_TS and https://sid.erda.dk/sharelink/EPvv68fOTp, respectively.

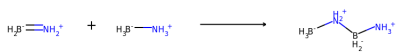
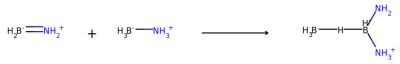
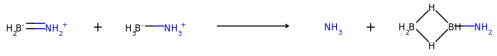
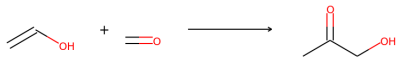
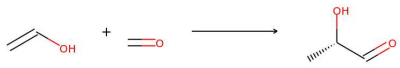
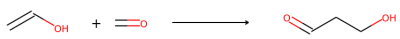
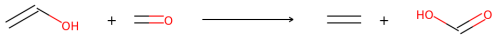
Parameter sets

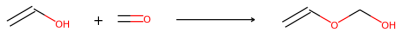
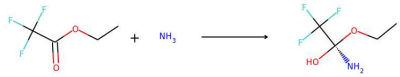
Parameter set 1				
Trial	k_pull (E_h)	k_push (E_h)	α ($1/a_0$)	direction
1	-0.02	0.01	0.6	R⟶P
2	-0.02	0.01	0.3	P⟶R
3	-0.02	0.01	0.3	R⟶P
4	-0.03	0.01	0.6	P⟶R
5	-0.03	0.01	0.6	R⟶P
Parameter set 2				
Trial	k_pull (E_h)	k_push (E_h)	α ($1/a_0$)	direction
1	-0.02	0.008	0.6	R⟶P
2	-0.02	0.01	0.3	P⟶R
3	-0.02	0.01	0.3	R⟶P
4	-0.03	0.01	0.6	P⟶R
5	-0.03	0.01	0.6	R⟶P

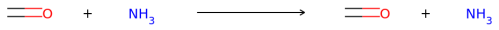
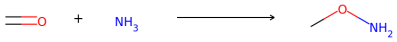
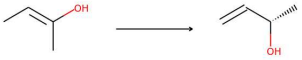
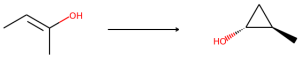
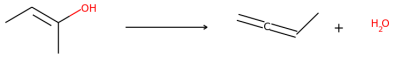
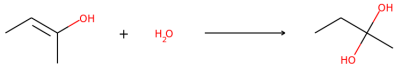
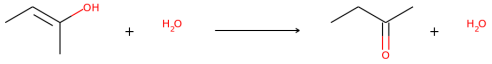
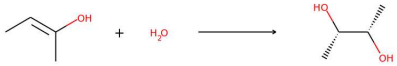
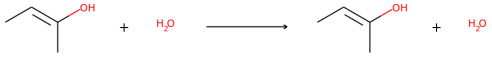
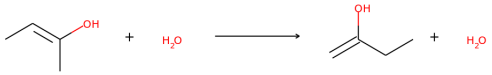
Table S1. The two parameter sets tested. For each trial three runs are done with the k_push and k_pull, 1.5*k_push and 1.5*k_pull, and 2.25*k_push and 2.25*k_pull. The direction is indicated as reactant (R) to product (P) or the other way around

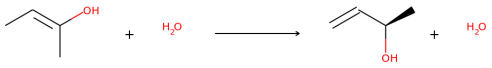
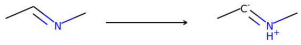
Reactions

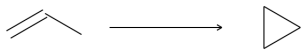
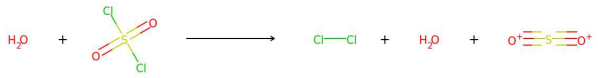
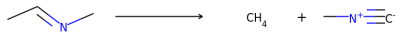
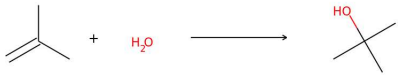
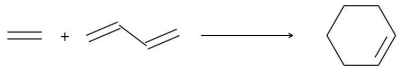
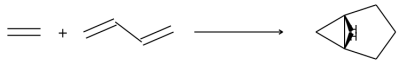
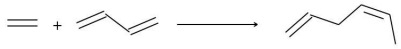
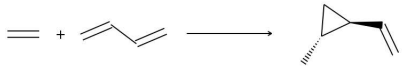
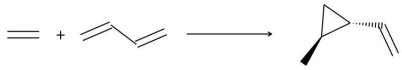
	Reaction	Comments
1		
2		
3		
4		
5		
6		TS could not be confirmed to exist by IRC
7		

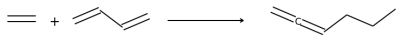
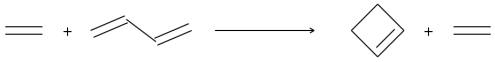
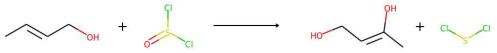
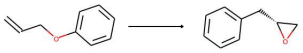
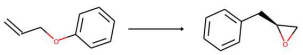
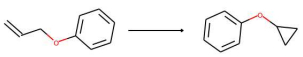
8		
9		
10		TS could not be confirmed to exist by IRC
11		TS could not be confirmed to exist by IRC
12		
13		
14		
15		
16		TS confirmed to exist but was not found by the procedure
17		
18		
19		
20		TS could not be confirmed to exist by IRC
21		
22		

23		
24		
25		
26		
27		
28		
29		
30		
31		
32		
33		
34		
35		TS could not be confirmed to exist by IRC
36		
37		

38		
39		
40		
41		
42		
43		
44		
45		
46		
47		
48		
49		
50		
51		
52		

53		
54		TS could not be confirmed to exist by IRC
55		
56		
57		
58		
59		
60		
61		
62		
63		
64		
65		
66		
67		

68		TS could not be confirmed to exist by IRC
69		
70		
71		
72		
73		
74		
75		
76		
77		
78		
79		
80		
81		
82		

83		
84		TS confirmed to exist but was not found by the procedure
85		
86		
87		
88		
89		
90		TS could not be confirmed to exist by IRC
91		
92		
93		
96		TS could not be confirmed to exist by IRC
97		
98		
99		

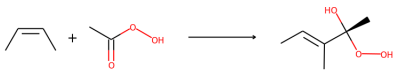
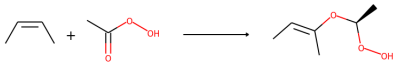
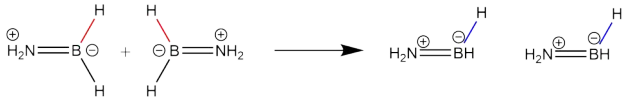
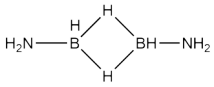
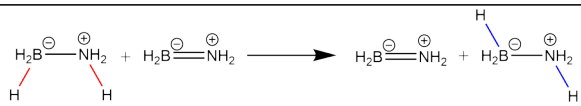
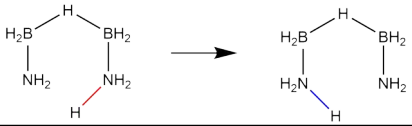
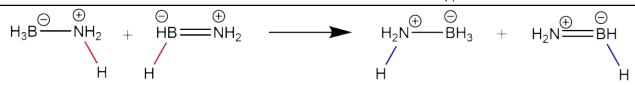
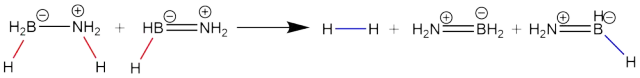
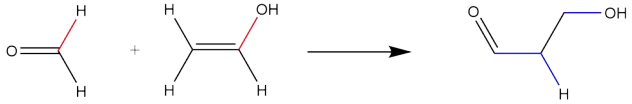
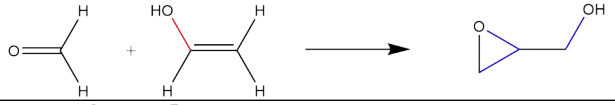
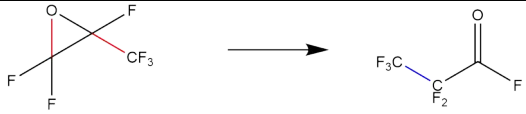
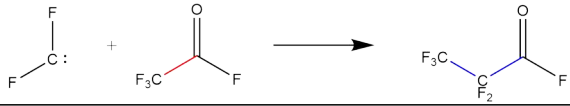
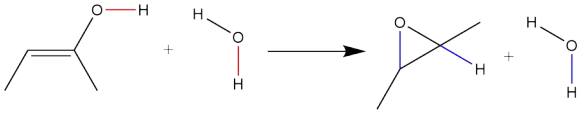
100		
101		
102		

Table S2. Table of the 100 studied reactions

#	Type	Reaction/Structure	Comments
6	intended		A minimum is found at the indicated structure. The TS correspond to a conformational change
6	interme-diate		
10	intended		TS correspond to step in the reaction but find minimum before reaction complete
10	Other reaction		
11	intended		TS correspond to another reaction
11	Other reaction		
20	intended		TS combines true reactant with indicated intermediate product
20	Other reaction		
35	intended		TS combines true product with the indicated (unstable) intermediate
35	Other reaction		
54	intended		TS correspond to another reaction with same product

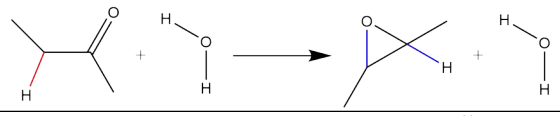
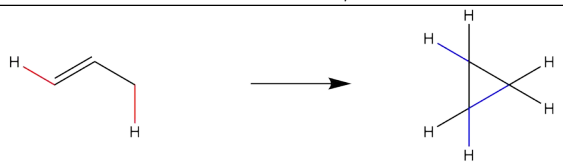
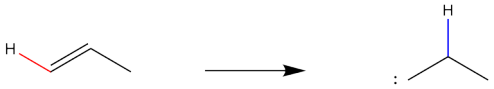
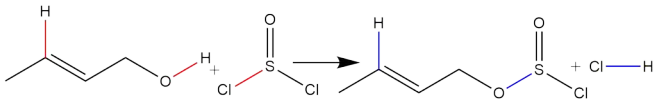
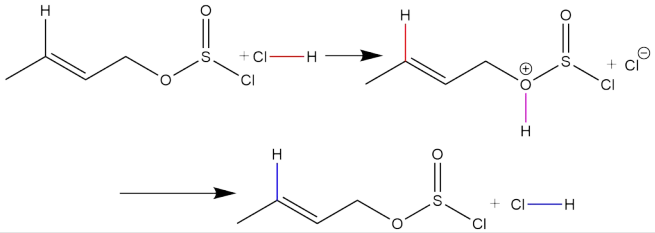
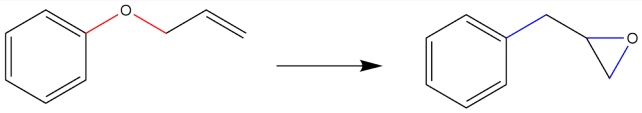
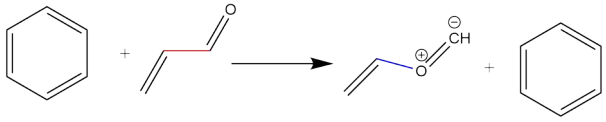
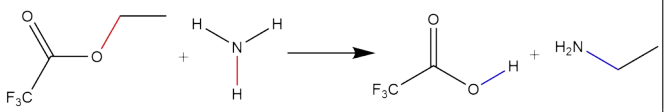
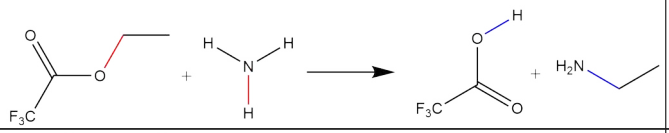
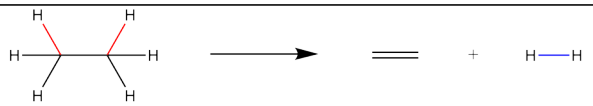
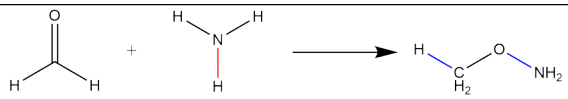
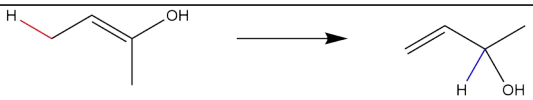
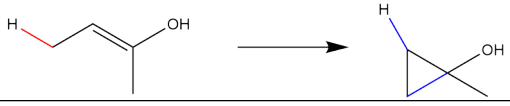
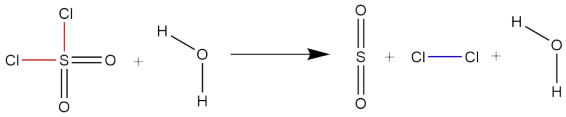
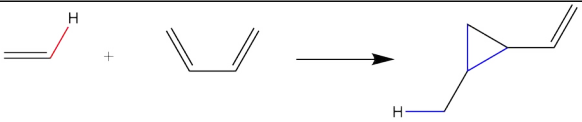
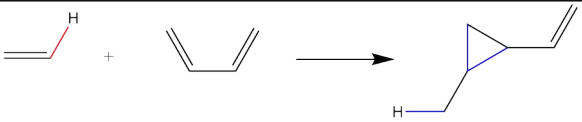
54	Other reaction		
68	intended		The reaction found correspond to first step in intended reaction but find the indicated unstable intermediate along the way
68	Other reaction		
90	intended		No re-optimization: 2 imaginary frequencies IRC gets 'stuck' Re-optimization: TS correspond to indicated reaction (pink indicates that bond is both formed and broken
90	Other reaction		
96	intended		No re-optimization: 2 imaginary frequencies both side of IRC goes to reactant. Re-optimization: TS correspond to another reaction
96	Other reaction		

Table S3. Table of the 9 reactions where a TS could not be confirmed

#	Type	Reaction/Structure	Comments
30	intended		TS correspond to different but similar reaction. The H from ammonia goes to the carbonyl oxygen atom instead
30	Other reaction		
34	intended		No re-optimization: Both sides of IRC go to product Re-optimization: TS corresponds to a reaction
34	Other reaction		
40	intended		TS goes to reactant along both directions of the IRC
44	intended		TS corresponds to a different reaction
44	Other reaction		
70	intended		No Re-optimization: 2 imaginary frequencies (below 150 cm^{-1}). IRC can only be followed in reactant direction. Re-optimization: No imaginary frequencies left
80	intended		The IRC stops almost immediately in both directions while still at the proposed TS structure.
81	intended		Reaction equivalent to 80. No re-optimization: 2 imaginary frequencies, IRC stops almost immediately in reactant direction. Re-optimization: Find reaction with different reactant

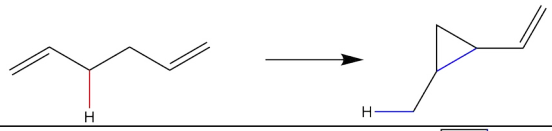
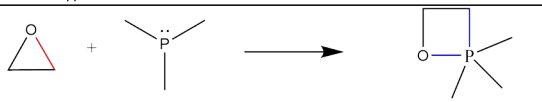
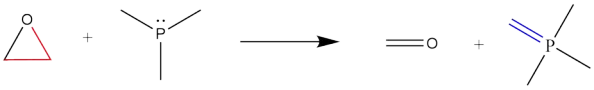
81	Other reaction		
88	intended		TS corresponds to a different reaction
88	Other reaction		

Table S4. 7 Reactions for which TSs were found by the RMSD-PP procedure but where the TSs given in[1] went to different reactants/products during the IRC

Timings

Reaction	N_{trials}	wall-time
75	1	54 s
55	4	3 min 17 s

Table S5. examples of wall-times for the RMSD-PP part of the procedure. The examples are for the first parameter set (Table S1), run1, running on a single CPU. N_{trials} is the number of trials needed before the reaction completed

REFERENCES

[1] P. Zimmerman, *J. Chem. Theory Comput.* **2013**, *9*, 3043–3050.