Rapid Route-Finding for Bifurcating Organic Reactions

Sanha Lee[†], Jonathan M. Goodman[†,*]

[†]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW

ABSTRACT: A large number of organic reactions feature post transition state bifurcations. Selectivities in such reactions are difficult to analyze because they cannot be determined by comparing the energies of competing transition states. Molecular dynamics approaches can provide answers but are computationally very expensive. We present an algorithm which predicts the major products in bifurcating organic reactions with negligible computational cost. The method requires two transition states, two product geometries and no additional information. The algorithm correctly predicts the major product for about 90% of the organic reactions investigated. For the remaining 10% of the reactions, the algorithm returns a warning indication that the conclusion may be uncertain. The method also reproduces the experimental and the molecular dynamics product ratios within 15% for more than 80% of the reactions. We have successfully applied the method to a trifurcating organic reaction, a carbocation rearrangement and solvent dependent Pummerer-like reactions, demonstrating the power of the algorithm to simplify and to help to understand highly complex reactions.

INTRODUCTION

Selectivity usually arises in a chemical reaction when the reactants can proceed via two or more competing reaction pathways to form two or more sets of products. Under kinetically controlled conditions, the product with the lowest activation barrier will be formed fastest. However, in growing number of reactions, the selectivity cannot be reduced to a simple choice between pathways with different barrier heights.[1–3] Many recently discovered chemical reactions show 'post transition state bifurcation'. The mechanism is characterized by two or more product pathways originating from a single transition state (TS). The potential energy surface (PES) that describes the two sequential TSs with no energy minimum between is characterized by a 'valley-ridge inflection' (VRI) point. VRI is the point where the valley along the minimum energy pathway from the first TS turns to a ridge. The new reaction channel is accessible at the VRI before reaching the second TS and, therefore, the two products do not have distinct barriers to formations, Figure 1.

Analyzing the IRC trajectory from the first transition state can provide some clues that VRI may be present. The IRC trajectory typically contains a flat region. For bifurcating reactions, optimization at regular intervals along the trajectory may converge to a different product to the IRC product. Other clues are discussed elsewhere in the literature.[2,3] Numerous pericyclic reactions proceed on a bifurcating PES,[4–12] like the example shown in Scheme 1 and Figure 2. The VRI points have been located for radical reactions including ene reactions,[13,14] and the reaction between ketyl anion radicals and alkyl halides.[15] In separate studies, Singleton et al. reported post transition state bifurcations for the singlet oxygen ene reaction with an alkene[16,17] and for the enyne-allene cyclizations.[18] The phenomenon is not limited to cyclizations; Yamataka et al. observed bifurcations in Beckmann and Schmidt rearrangements.[19,20] Furthermore, theoretical investigations of nucleophilic substitution reactions have also identified bifurcating PES.[21,22] The observations also include catalyzed processes, such as Rh-carbenoid C-H insertions,[23,24] gold-catalyzed cyclizations,[25–27] asynchronous nitrene insertions[28], and natural product biosynthesis catalyzed by enzymes.[24,29–33] Tantillo et al. have observed complex multiple sequential bifurcations in the carbocation rearrangement mechanism of terpene synthase enzymes.[34]
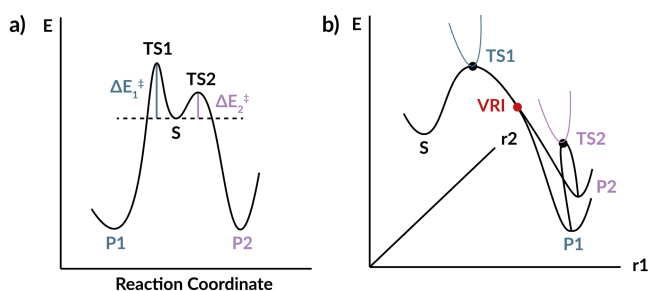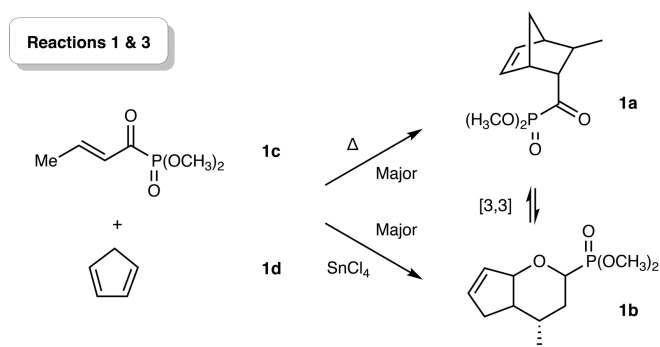

Figure 1. a) The traditional selectivity description based on the relative heights of the activation barriers corresponding to the two product reaction pathways. b) The PES of a bifurcating reaction.

Scheme 1. Selected bifurcating reaction example by Houk et al.[4]

Despite their common occurrence, a reaction with selectivity controlled by a valley-ridge inflection point is hard to analyze. On an asymmetrical bifurcating PES, the product selectivity will not usually
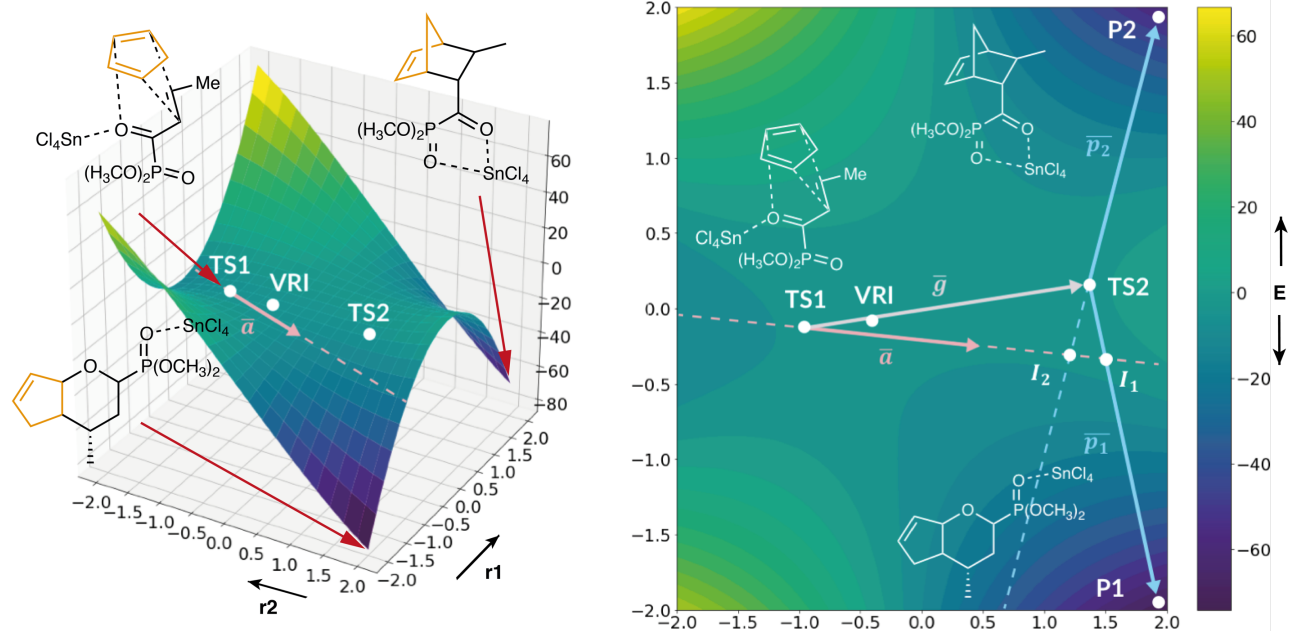
Figure 2. Qualitative representation of the PES for bifurcating reaction from a) – the vectors are drawn on the surfaces to clarify the definitions. $\overline{a}$ is the imaginary eigenvector of TS1, $\overline{g}$ is the vector defining the separation of the two transition states and $\overline{p}_i$ is the vector defining the displacement of product i from TS2.

be 50:50. Simple transition state theory[35] cannot be used to determine the major product because the selectivity is governed by post-transition state dynamic effects. This means that the motions of the nuclei, or how the reaction trajectory passes through the first TS impacts the preferred product formation. Therefore, there are no computationally inexpensive methods for determining the selectivities generalizable to all bifurcating reactions.

Although the methods for locating the VRI point are available[36], the selectivity problem is typically addressed using molecular dynamics (MD) simulations for reaction paths over the first TS and on towards the products.[37] The number of trajectories continuing on one side of the ridge can be compared to the trajectories diverting to the adjacent valley. The average behavior of many trajectories from MD studies can reproduce the experimental findings quite well.[1] However, the significant drawback is that the accurate determination of the major product requires a large number of trajectories to be calculated, each of which has a significant computational cost. Such a procedure is computationally expensive and time-consuming. Further inputs are usually needed, such as scanning the PES around the two TSs. The scans also require many single point calculations which significantly add to the overall computational time.[16]

We were therefore interested in developing a simple and efficient methodology to determine the selectivities in bifurcating organic reactions which would get around the need for an expensive computational method and may provide insights into the reasons or the selectivity. Herein, we present an algorithm that predicts the major product in reactions proceeding via a VRI point using only the two transition state data, the two product geometries and no additional information. The method can also estimate the experimental or the MD simulation product ratios for the majority of the reactions we have investigated.

RESULTS AND DISCUSSION

The Reactions Studied. The reactions were chosen to cover as wide variety of organic reactions discussed in the introduction as possible (see ESI Section 1). A few examples are outlined in Figure 3. The set included pericyclic reactions, nucleophilic substitutions, rearrangements, fragmentations, ene reactions and selected catalysis examples. All the reactions have had their selectivity thoroughly investigated either by experimental validation or by MD trajectory studies. The bifurcating reactions available in the literature were included only if TS geometries were provided, if either experimental or MD data was available and if the system sizes were not too expensive computationally.

Some of the reactions are variants of each other. For instance, Reaction 1 and Reaction 3 differ only by the presence of a Lewis acid, SnCl₄, Scheme 1.[4] The variants are included because the original report showed that the selectivity changes when the experimental conditions are altered. The inclusion of a Lewis acid changes the experimental major product from the Diels-Alder product (**1a**) to the Hetero-Diels-Alder product (**1b**).[38] The algorithm verification would be more rigorous if the change in the major product selectivity could be detected, despite the similar TS geometries in the variant reactions.

Major Product Prediction Algorithm. We reasoned that, although there will be numerous reaction trajectories over the first TS (TS1), the average direction will be associated with the imaginary eigenvector of TS1, which will point backwards towards the starting materials and onwards in the direction of TS2. Analysis of this pathway from TS1 to TS2 might let us work out the identity of the major product and, perhaps, the approximate ratio of the products. To investigate this idea, we wrote a Python program (ValleyRidge.py, see
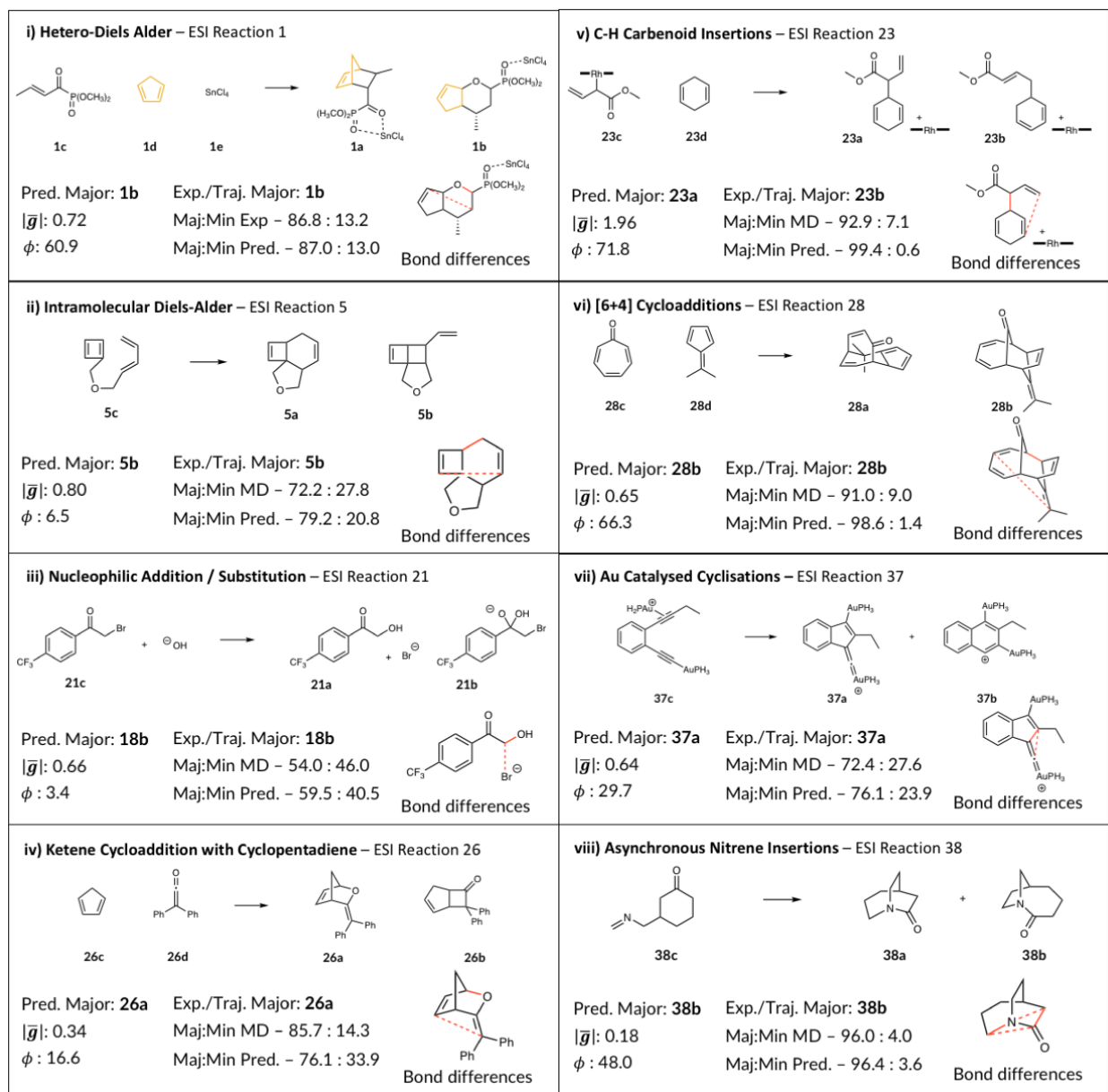
Figure 3. Selected example reactions used to validate the algorithm. The major product and the selectivity ratios predicted by the algorithm is compared to the experimental or the MD major product and the ratios. The separation of the two transition states is given by $|\overline{g}|$ (in Angstroms). ϕ is the angle the vector $\overline{a}$ makes with $\overline{g}$ in degrees.

SI section 6) which takes the geometries and the imaginary eigenvectors of the two transition states (TS1 and TS2) and the two products (P1 and P2) as inputs. We label the separation of the second transition state (TS2) from TS1 as the vector $\overline{g}$ and the displacement vectors from TS2 to product 1 and product 2 of the bifurcation as $\overline{p_1}$ and $\overline{p_2}$ respectively, Figure 4(iv).

The high dimensionality of the system is addressed by identifying key bonds that change as the reaction progresses, using RDKit[39]. In many of the reactions we studied, there is one bond that exists in P1 but not in P2 and one bond that exists in P2 but not in P1. These bonds are selected as the two key variables (see Figure 3 for examples). For some reactions, there are more than two bond differences between the two products. Reactions 31-33 and Reactions 38-41 from Houk *et al.* 2018 and Tantillo *et al.* 2019 respectively, show

such behavior. The two key bonds are then selected by computing all possible permutations of bond difference pairs between the two products and ranking. The ranking is performed by calculating the angle the vector $-\overline{g}$ makes with the vectors $\overline{p_1}$ (θ_1) and with $\overline{p_2}$ (θ_2). If θ_1 and θ_2 (in degrees) are both obtuse, that bond 1 and bond 2 pair gets the priority. The algorithm then computes $|\theta_1 - 90°|$ and $|\theta_2 - 90°|$ and the bond pair that minimizes the sum ($|\theta_1 - 90| + |\theta_2 - 90°|$) receives the highest priority (see ESI section 2.4 for the details). In few cases, such as the fragmentation reactions reported by Yamataka *et al.* 2010 and Tantillo *et al.* 2017, only one bond difference is observed between P1 and P2. We then search for bond differences between P1 and TS1. All possible pair permutations are computed for the P2/P1 and TS1/P1 bond differences. The highest ranked pair is selected as before. These two key
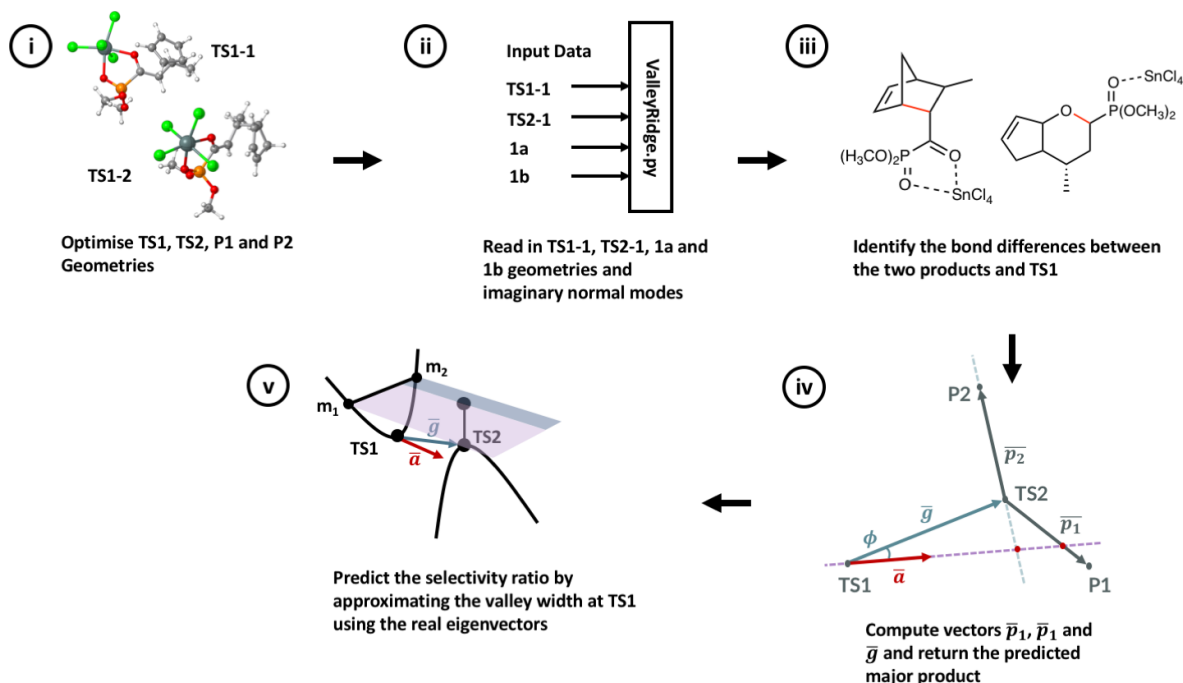
Figure 4. Summary of the selectivity prediction algorithm workflow using Reaction 1 as an example. . i) The user optimises the two transition states involved in the bifurcation. ii) The optimised geometries are read in. iii) The algorithm searches the bond differences between the products. iv) We compute the required vectors in the bond difference space. $\overline{g}$ defines the separation of TS1 and TS2, $\overline{p_1}$ and $\overline{p_2}$ are vectors pointing from TS2 to one of the products, $\overline{a}$ is the imaginary eigenvector of TS1. We then analyze which side of TS2 the imaginary eigenvector of TS1 points. v) The selectivity ratio is predicted by projecting the real vibrational eigenvectors of TS1 to the bond difference space and approximating the maximum valley width for each real normal mode.

bonds are used to define the coordinates for a two-dimensional description of the reaction

The key feature of the algorithm is determining the side of the ridge the imaginary eigenvector at TS1 points towards, using the 2D basis of the two key bonds. Let the vector $\overline{a}$ be the reaction vector: the imaginary eigenvector of TS1 making angle ϕ with vector $\overline{g}$, Figure 4(iv). The vectors $\overline{a}$ and the two $\overline{p}$ vectors in the bond difference space will meet at two positions. The reaction trajectories passing over the first barrier is more likely to fall down the side of the ridge the TS1 valley is facing.[4] In bifurcating reactions, the product vectors $\overline{p_1}$ and $\overline{p_2}$ are expected to point in the opposite directions from TS2. Therefore, the one of the product vectors will intercept vector $\overline{a}$ in the positive direction and the other product vector will intercept the vector $\overline{a}$ in the negative direction. The product vector of the major product will intercept the vector $\overline{a}$ in the positive direction because the imaginary eigenvector points in its direction (see ESI section 2.2).

Occasionally, for a given bond pair, the two product vectors may not have one positive and one negative interception with the vector $\overline{a}$ (see ESI section 2.5 and 2.6). Such a bond pair is unlikely to be a good coordinate choice to represent bifurcation because the product vectors are not in the opposite direction from TS2. The algorithm then uses the bond difference pair which is next highest in the rank until one positive and one negative interception is found.

For the Reaction 3 example in Scheme 1, we found the Diels-Alder product **3a** vector intercepts vector $\overline{a}$ in the positive direction and the Hetero-Diels-Alder product **3b** vector intercepts vector $\overline{a}$ in the negative direction. Therefore, the major product is correctly predicted to be the Diels-Alder product **3a**. In the Reaction 1 example,
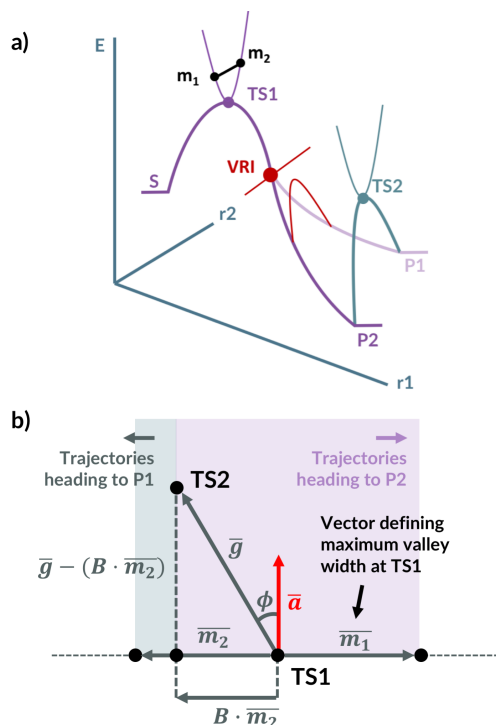

Figure 5. The vectors $\overline{m_1}$ and $\overline{m_2} = -\overline{m_1}$ correspond to the maximum valley width orthogonal to the reaction vector, calculated from the real vibrational eigenvectors of TS1 projected into the bond difference space. The constant B is the scalar multiple of vector $\overline{m_1}$ or $\overline{m_2}$ to move from TS1 to the point of closest approach of TS1 along the $\overline{m_1}$ or $\overline{m_2}$ direction. $\overline{a}$ is the reaction vector of TS1.

the imaginary eigenvector points to the Hetero-Diels-Alder product unlike the uncatalyzed Reaction 3. Hence, the major product is predicted, correctly, to be the Hetero-Diels-Alder product **1b**. The algorithm is summarized in Figure 4.

The Selectivity Ratio Prediction Algorithm. It is possible to predict the ratios for all the reactions by considering the likely distribution of the trajectories from TS1. This distribution may be estimated by using the information in the real normal modes of the TS1. Reaction trajectories that are close to TS1, but do not pass directly through it, may have accessible energies and will travel towards the products at different places along the directions $\overline{p_1}$ and $\overline{p_2}$. For some reactions, the shape of the energy surface around the transition state will constrain these trajectories in a narrow band, but for other reactions, a broader spread of trajectories is possible which will lead to lower selectivity. Working in the two-dimensional space defined by the two key changing bonds, the direction perpendicular to the reaction coordinate defines a valley which we treat with a harmonic approximation. The parameter maximum valley width describes the region of this valley which is within $\frac{3}{2}RT$ of the valley floor (see ESI section 4 for the details), Figure 5. Let the vectors $\overline{m_1}$ and $\overline{m_2}$ define the maximum valley width orthogonal to the reaction vector at TS1. We approximate that the trajectories passing through the P1 side of the point of the closest approach of TS2 (given by $B \cdot \overline{m_2}$) on the maximum valley width vector will arrive at P1 and *vice versa* for P2. The curvature of PES is implicitly taken into account in the present model because the maximum valley width will be dependent on the curvature at TS1. By considering the width of trajectory stream and how much of it favors each product, the selectivity of the reaction can be estimated.

The Algorithm at Work. The accuracy of the selectivity ratio prediction was assessed by comparing the results generated by the algorithm to the experimental and the MD trajectory ratios published in the literature. The major product ratio was found by dividing the number of trajectories that go to the major product by the total number of trajectories going to the major or the minor product, not including the trajectories that have recrossed. The selectivity ratio prediction algorithm was tested against all 41 reactions we modelled.

The predicted selectivity data can be grouped depending whether $|\overline{g}| < 0.5$, $0.5 < |\overline{g}| < 1.0$ or $|\overline{g}| > 1.0$ (in units of Ångstrom in bond difference space) and whether ϕ (the angle vector $\overline{a}$ makes with $\overline{g}$ in degrees) is $\phi < 20$, $20 < \phi < 50$, or $\phi > 50$. The group where $|\overline{g}| < 0.5$ and $20 < \phi < 50$, the algorithm slightly overestimates the major product selectivity (see ESI, section 5). However, a strong correlation is observed between the predicted and the experimental product ratios and the predictions are corrected through linear scaling (see ESI section 5).

The algorithm predicts the product ratios within 8% for 61.0% of the reactions investigated. When the acceptable error range is increased to 15%, the algorithm predicts the ratio correctly for 82.9% of the reactions, Figure 6. For Singleton 2006 (Reactions 9-12) reactions the predictions agree better with the MD trajectory ratios than the experimental ratios. Singleton *et al.* stated that the observed experimental ratio was 100% for Reactions 9,10 and 12 because the products rapidly converted to the thermodynamic product *via* low barrier Cope-rearrangement transition state. Furthermore, despite predicting the incorrect major products for three Yamataka 2011 reactions (Reactions 17-22), the algorithm correctly predicts the major product selectivity to be near 50:50. Thus, the results agree with the borderline selectivity predicted by the MD simulations.
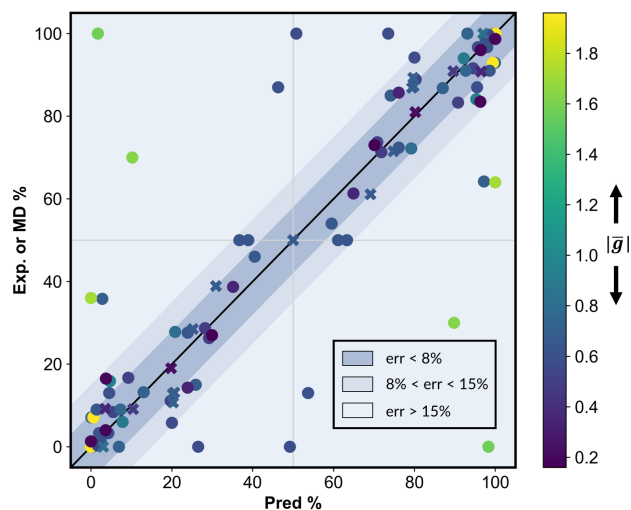

Figure 6. Predicted selectivity ratio from the algorithm plotted against the experimental or the MD trajectory selectivity ratio for all the reactions modelled. The plot is divided in to three regions according to the prediction error less than 8% and less than 15%. The points are color-coded according to the magnitude of $|\overline{g}|$. The circular and the crossed points are the results from the training and the unseen examples respectively.

The algorithm predicts the major product correctly for 87.8% of all the reactions modelled (see ESI, section 2.1). The method also captures the change in the major product selectivity in the similar variant reactions with subtle changes in the reaction conditions. The simplified PES visualization of the Houk 2007 study (Reactions 1 and 3) illustrated earlier shows that the imaginary eigenvector points to the Diels-Alder product side of TS2 for the uncatalyzed reaction and points to the Hetero-Diels-Alder product side for the catalyzed reaction, Figure 7a. Therefore, the experimental major products are correctly predicted by the algorithm. The simplified PES visualization of the Houk 2003 study (Reactions 5 and 7) in Figure 7b also illustrates the power of the algorithm to pick out the subtle changes in the reaction variation. The major product selectivity in these reactions depends on the chain lengths between the cyclobutadiene and the diene components, with the longer chains preferring the (4+2) cycloaddition (in Huisgen notation). The imaginary eigenvector for the shortest 2-oxatrimethylene tether points to the (2+2) cycloaddition side and the longer 2-oxatetramethylene tether points to the (4+2) side, correctly predicting the experimental major product.

The only studies where the algorithm has struggled to predict the correct major product are some of the reactions from Yamataka 2010 and Yamataka 2011. The IRC pathway in Yamataka 2010 is very curved and consequently the orientation of the imaginary eigenvector at TS1 cannot capture the preferred reaction trajectory (see ESI section 2.7). The length of vector $\overline{g}$ corresponding to the TS1 and TS2 separation is larger ($|\overline{g}| > 1$) in Yamataka 2010 compared to other reactions. Therefore, when $|\overline{g}| > 1$ the prediction from the algorithm should be treated with caution because the pathway between TS1 and TS2 have space to be complex. Davis 2011 is the only other reaction with $|\overline{g}| > 1$ and also shows a more complicated reaction pathway despite the correct prediction (see ESI section 2.7). Therefore, the algorithm returns a warning sign to the user when $|\overline{g}| > 1$.
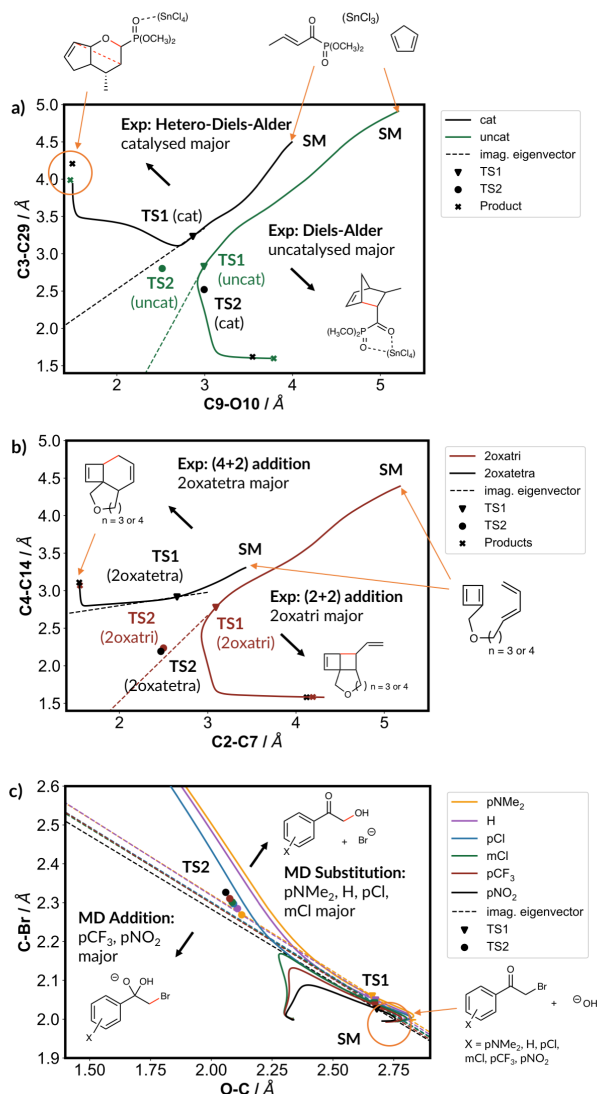
Figure 7. Visualisation of the IRC pathway and the orientation of the imaginary eigenvector at TS1 with respect to the two key bonds selected by the algorithm shown in red. a) Reactions 1 and 3 from Houk 2007 study, b) Reactions 5-8 from Houk 2003 study, c) Reactions 17-22 from Yamataka 2011 study. The IRC pathway is shown with a solid line; the dashed line represents the orientation of the imaginary eigenvector at TS1.

For Yamataka 2011 study, the algorithm correctly predicts the major product when the MD trajectory selectivity is large (pNMe2 and pNO2) but sometimes predicts the wrong product when the selectivities are close to 50% (H, pCl, mCl and pCF3). The imaginary eigenvector for the low selectivity reactions occasionally point to the minor product side of TS2 because the angle it makes with the vector $\bar{g}$ is so small, Figure 7c. Yamataka *et al.* performed the MD simulations with Gaussian 03 (G03) and with PEACH. The G03 and PEACH simulations predicted different major products for the low selectivity reactions pCl and mCl. The difference may be due to different initial sampling and different number of trajectories used in the two studies. The IRC calculation also converged to the minor product side for the mCl reaction. Therefore, experimentally product ratio is likely to be very close to 50:50 for these reactions and the major product predictions are less meaningful. The algorithm returns a warning sign if the major product selectivity is close to 50:50.

An alternative approach to determine the major product for these reactions is to use IRC calculations. This has three significant drawbacks: (i) IRC calculations cannot predict the selectivity ratios; (ii) despite a similar correct prediction percentage, the time required for an IRC calculation is significantly longer than our algorithm; (iii) for Reactions 28-30, predicting the major product via IRC analysis is challenging. The PES is very flat which means that the major product is not easily reached from the first transition state (see ESI section 2.3).

Although our approach offers a quick and computationally inexpensive method for predicting the selectivity for bifurcating reactions, we would like to highlight some advantages of the computationally more expensive MD approach. Tantillo *et al.* found the carbocation rearrangement reaction of pimarenyl cation occurs on a complex network of ambimodal transition states.[34] The MD approach is able to address all competing reactions together whereas our method addresses one bifurcation at a time. Furthermore, branching ratios for some organic reactions are dependent upon the conservation of momentum and reaction trajectories can pass over energetically unfavorable barriers on PES.[40–42] MD simulations predicts the product distributions well for such reactions. Moreover, we have assumed in our model that MD simulations can accurately model the selectivity even though the MD simulations are performed with varying degrees of precision in the literature with different numbers of sampling trajectories and a variety of functionals and the basis sets.
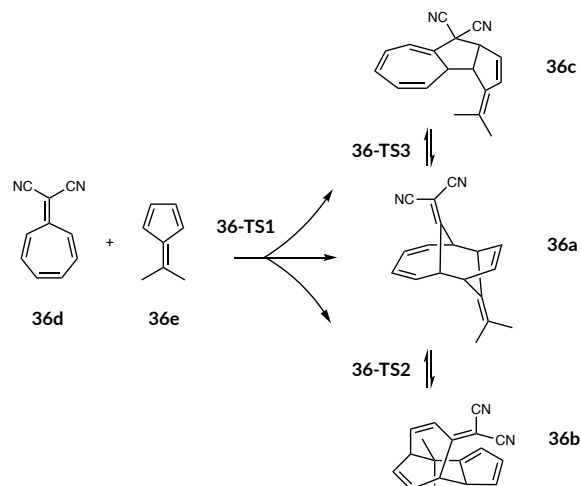

Scheme 2. Trifurcating reaction by Houk *et al.*[44]

Table 1. Percentage selectivity results for the trifurcating reaction from Houk *et al.* [44]

Study	React. No.	Maj.	$|\bar{g}|$	Pred. %	Real %	|err|
36a/36b	36	**36a**	0.59	95.7	96.7	-1.0
36a/36c	36	**36a**	0.59	98.0	96.7	1.3

In 2018, Houk *et al.*[43] proposed a method for predicting selectivity in bifurcating pericyclic reactions using the bond lengths of TS1. Although the prediction method is a very useful approach for pericyclic reactions, we believe our method is more generalizable to wider family of bifurcating reactions and less prone to human error. Houk *et al.*'s model assumes the product with shorter partial bond at TS1 is the major product and this is true for all the pericyclic reactions he reports. For the reactions we have modelled outside the pericyclic
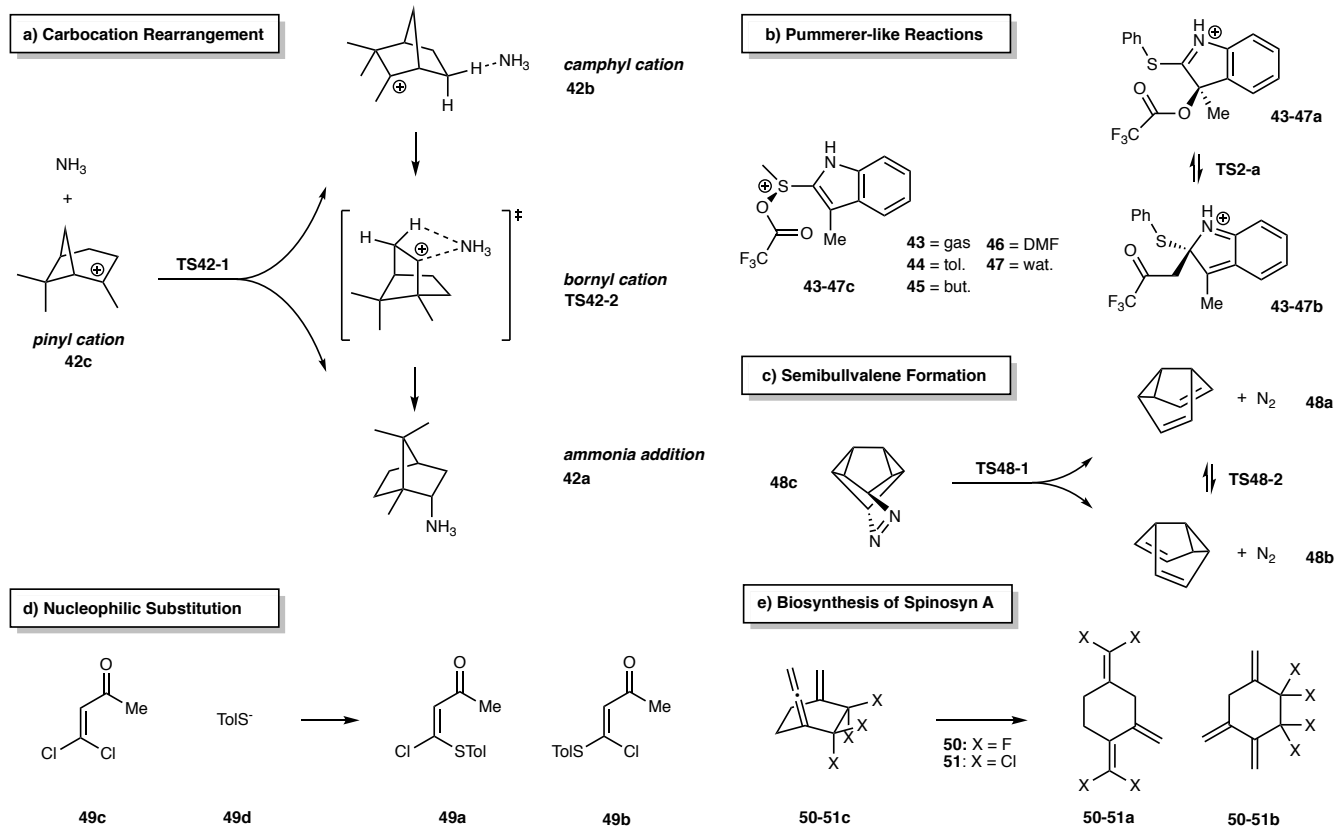
Scheme 3. Reactions not included in the model development but was used to test the ValleyRidge.py algorithm. a) carbocation rearrangement with an elec-tron pair donor. b) Pummerer-like reactions show solvent dependent selectivities. Solvents modelled in the present work include toluene, 1-buta-nol, DMF and water. c) Semibullvalene formation in deazetization reaction taking place on bifurcating PES. d) Unseen nucleophilic substitu-tion re-action. e) Unseen [3,3] sigmatropic rearrangement reaction.

family, we have found the assumption is not always true (see ESI sec-tion 7). Some of the reactions we modelled show more than two bond differences between the two products (e.g. reaction 31-33, re-action 38-31). The ValleyRidge.py automatically selects the two bond differences to be used for selectivity prediction so we always get the consistent difference combination. The Manual bond selec-tions in such reactions could lead to inconsistencies and is harder to apply to large datasets.

Application to Trifurcating Reaction. The recent study by Houk *et al.*[44] reported a tripericyclic transition state for a reaction between 8,8-disubstituted heptafulvenes and 6,6-dimthylfulvene. Three [4+6], [6+4] and [8+2]-cycloadducts all form initially from a single transition state at a 87:3:3 ratio, Scheme 2. We tested the potential application of the algorithm for the Houk *et al.* trifurcating reaction by allocating the three products in pairwise groups. Products 36a and 36b and the transition states 36-TS1 and 36-TS2 was treated as one set of bifurcation and the products 36a and 36c and the transition states 26-TS1 and 36-TS3 as another set. In both groups, the product 36a is correctly identified as the major product. The percentage selectivity is 95.7% and 98.0% for 36a:36b and 36a:36c respectively, showing a good agreement with the MD simulation results, Table 1.

Applications to Examples not Used for Model Development. To rigorously test the performance of our method, we have mod-elled the reactions not included in our algorithm development. Ma-jor and Weitman suggested the biosynthetic mechanism for the for-mation of bornyl diphosphate may involve post transition state

bifurcation.[45] The carbocation rearrangement of pinyl cation pro-ceeds *via* bornyl cation which is a transition state between the cam-phyl cation and the bornyl diphosphate products, Scheme 3. In 2010, Hong and Tantillo modelled the carbocation rearrangement using ammonia as lone pair donors.[46] The IRC calculations showed evidence of bifurcating PES.[46] Although ammonia will have very dif-ferent properties to enzyme bound phosphate, the carbocation reac-tivity could still be examined with respect to a small nucleophile. We have optimized the transition states and the two products found by Hong and Tantillo. ValleyRidge.py predicted the major product to be the addition product with 89.6% selectivity. Despite the simplifi-cation of the diphosphate addition with ammonia, the result agreed well with the MD simulation product distribution by Major and Weitman (1:0.1).

Recently, Tantillo *et al.* observed solvent dependent selectivities on family of Pummerer-like reactions, Scheme 3. The MD [3,3] product selectivity is low (61.1%) when the reaction is modelled in the gas phase. The selectivity initially increases as the dielectric con-stant of the solvent is increases. The increase in the selectivity is eventually flattened out when the solvent is more polar than 1-buta-nol. We have optimized the transition states and the [2,3] and [3,3] products for four selected solvents and for the gas phase. Passing the geometries through the ValleyRidge.py algorithm replicates the MD simulation results well, Figure 8. In contrast to the results from our algorithm, Tantillo *et al.* highlights the Newton program by Carpen-ter *et al.*[47], which was built on a single reaction, and which does not capture the solvent dependent selectivities for Pummerer-like reac-tions. The Newton program predicts the selectivity based on the

relative locations of the products to the imaginary eigenvector of TS1. The selectivity in our method has been tested on a much greater number of bifurcating reactions and depends on more aspects of the PES: the flexibility of the TS1 geometry (approximated by the force constants), directions the molecules can vibrate (approximated by imaginary and the real eigenvectors) and the location of the ridge (approximated by the location of TS2). The visualization of the simplified PES shows the imaginary eigenvector points increasingly further away from the TS2 as the solvent polarity increases (ESI, section 6). Therefore, the results illustrate the inclusion of TS2 geometries and the real eigenvectors are important for the selectivity predictions on bifurcating PES.
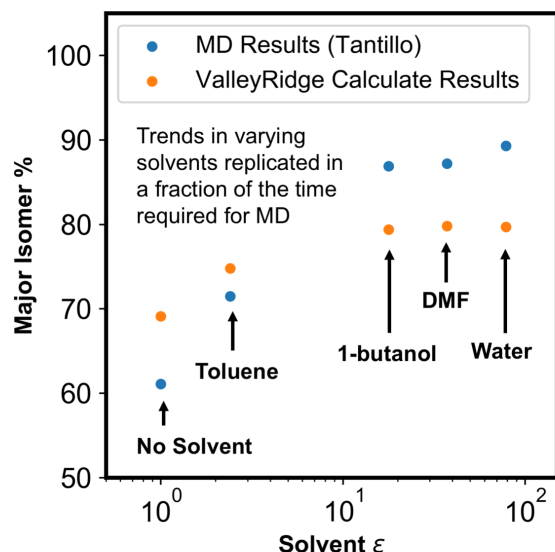

Figure 8. The product selectivity for Pummerer-like reactions for varying solvent polarities.

In 2018, Mandal and Datta[48] investigated the semibullvalene formation in deazetization reactions, Scheme 3. The MD investigation found the ratio of the products **48a** and **48b** to be 1:1 which the ValleyRidge.py successfully predicted. The nucleophilic substitution reaction modelled by Singleton *et al.*[22] has experimental preference for **49a** over **49b** by 81:19. Our algorithm predicted the major product selectivity to be 80.2%, showing a good agreement. The [3,3] sigmatropic rearrangement modelled by López[49] *et al.* shows contrasting selectivity depending on the halogen substituents. ValleyRidge.py predicts 96.5% selectivity for **50b** and 97.0% selectivity for **51a**, which are good agreements with the MD works. Overall, the mean absolute error (MAE) across the 10 unseen examples is 4.7% and the MAE for 46 reactions with $|\overline{g}| < 1$ is 7.8%, supporting the wide applicability of our algorithm to many bifurcating reactions.

CONCLUSIONS

We present an algorithm which can predict the product ratios for bifurcating organic reactions in a fraction of a time required from the traditional molecular dynamics simulation methods. The algorithm has predicted the correct major product for 87.8% of all reactions studied in this work. The algorithm can further predict the product ratios within 8% in 61.0% of the reactions. The proportion of correct ratio prediction increases to 82.9% when the tolerance is increased to 15% error. Our method also captured the changes in the major product in the variant reactions when the reaction conditions are subtly changed. We have shown that the selectivity predictions becomes more challenging when the separation of the two transition states is large or the ratios are near 50:50. Finally, we demonstrate the application of the algorithm to a trifurcating organic reaction, a carbocation rearrangement and solvent dependent Pummerer-like reactions which were not used to develop the algorithm. The algorithm is able to predict the major product and the ratios successfully, illustrating its ability to simplify more complicated reaction examples.

COMPUTATIONAL METHODS

All quantum mechanical calculations were performed using the Gaussian 16 package.[50] The geometry optimizations were performed using the functionals and the basis sets reported by the originally published work. The initial cartesian coordinates for the optimizations were also taken from the reported work. A full list of reactions studied and the method used respectively are provided in the electronic supporting information (ESI), Table S8. The frequency calculations were used to characterize the stationary points and to ensure the structures were optimized to the minimum or the transition state.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

All optimized geometries, computational details and the derivations are available free of charge. (PDF)

ValleyRidge.py python code and the user manual are also available (.zip) and can be downloaded from GitHub github.com/sanha0213/ValleyRidge

AUTHOR INFORMATION

Corresponding Author

*jmg11@cam.ac.uk

Author Contributions

The manuscript was written through contributions of all authors. / All authors have given approval to the final version of the manuscript. / †These authors contributed equally.

REFERENCES

(1) Rehbein, J.; Carpenter, B. K. Do We Fully Understand What Controls Chemical Selectivity? *Phys. Chem. Chem. Phys.* **2011**, 13, 20906–20922. https://doi.org/10.1039/c1cp22565k.

(2) Ess, D. H.; Wheeler, S. E.; Iafe, R. G.; Xu, L.; Çelebi-Ölçüm, N.; Houk, K. N. Bifurcations on Potential Energy Surfaces of Organic Reactions. *Angew. Chemie. Int. Ed.* **2008**, 47, 7592–7601. https://doi.org/10.1002/anie.200800918.

(3) Hare, S. R.; Tantillo, D. J. Post-Transition State Bifurcations Gain Momentum – Current State of the Field., *Pure Appl. Chem.* **2017**, 89, 679–698. https://doi.org/10.1515/pac-2017-0104.

(4) Celei-Olcm, N.; Ess, D. H.; Aviyente, V.; Houk, K. N. Lewis Acid Catalysis Alters the Shapes and Products of Bis-Pericyclic Diels - Alder Transition States. *J. Am. Chem. Soc.* **2007**, 1, 4528–4529.

(5) Yu, P.; Chen, T. Q.; Yang, Z.; He, C. Q.; Patel, A.; Lam, Y. H.; Liu, C. Y.; Houk, K. N. Mechanisms and Origins of Periselectivity of the Ambimodal [6 + 4] Cycloadditions of Tropone to Dimethylfulvene. *J. Am. Chem. Soc.* **2017**, 139, 8251–8258. https://doi.org/10.1021/jacs.7b02966.

(6) Thomas, J. B.; Waas, J. R.; Harmata, M. Singleton, D. A. Control Elements in Dynamically Determined Selectivity on a Bifurcating Surface. *J. Am. Chem. Soc.* **2008**, 130, 14544-14555 https://doi.org/10.1021/ja802577v.

(7) D. A.; Wang, Z.; Hirschi, J. S.; Singleton, D. A. Recrossing and Dynamic Matching Effects on Selectivity in a Diels-Alder Reaction. *Angew. Chem. Int. Ed.* **2009**, 48, 9156–9159. https://doi.org/10.1002/anie.200903293.

(8) Limanto, J.; Khuong, K. S.; Houk, K. N.; Snapper, M. L. Intramolecular Cycloadditions of Cyclobutadiene with Dienes: Experimental and Computational Studies of the Competing (2 + 2) and (4 + 2) Modes of Reaction. *J. Am. Chem. Soc.* **2003**, 125, 16310–16321. https://doi.org/10.1021/ja0380547.

(9) Chen, S.; Yu, P.; Houk, K. N. Ambimodal Dipolar/Diels-Alder Cycloaddition Transition States Involving Proton Transfers. *J. Am. Chem. Soc.* **2018**, 140, 18124–18131. https://doi.org/10.1021/jacs.8b11080.

(10) Yu, P.; Patel, A.; Houk, K. N. Transannular [6 + 4] and Ambimodal Cycloaddition in the Biosynthesis of Heronamide A. *J. Am. Chem. Soc.* **2015**, 137, 13518–13523. https://doi.org/10.1021/jacs.5b06656.

(11) Ussing, B. R.; Hang, C.; Singleton, D. A. Dynamic Effects on the Periselectivity, Rate, Isotope Effects, and Mechanism of Cycloadditions of Ketenes with Cyclopentadiene. *J. Am. Chem. Soc.* **2006**, 128, 7594–7607. https://doi.org/10.1021/ja0606024.

(12) Pham, H. V.; Houk, K. N. Diels-Alder Reactions of Allene with Benzene and Butadiene: Concerted, Stepwise, and Ambimodal Transition States. *J. Org. Chem.* **2014**, 79, 8968–8976. https://doi.org/10.1021/jo502041f.

(13) Sheppard, A. N.; Acevedo, O. Multidimensional Exploration of Valley-Ridge Inflection Points on Potential-Energy Surfaces. *J. Am. Chem. Soc.* **2009**, 131, 2530–2540. https://doi.org/10.1021/ja803879k.

(14) Griesbeck, A. G.; Goldfuss, B.; Leven, M.; De Kiff, A. Comparison of the Singlet Oxygen Ene Reactions of Cyclic versus Acyclic β,γ-Unsaturated Ketones: An Experimental and Computational Study Dedicated to Professor Waldemar Adam on the Occasion of His 75th Birthday. *Tetrahedron Lett.* **2013**, 54, 2938–2941. https://doi.org/10.1016/j.tetlet.2013.03.099.

(15) Bakken, V.; Danovich, D.; Shaik, S.; Schlegel, H. B. A Single Transition State Serves Two Mechanisms : An Ab Initio Classical Trajectory Study of the Electron Transfer and Substitution Mechanisms in Reactions of Ketyl Radical Anions with Alkyl Halides. *J. Am. Chem. Soc.* **2001**, 123, 130–134. https://doi.org/10.1021/ja002799k.

(16) Singleton, D. A.; Hang, C.; Szymanski, M. J.; Meyer, M. P.; Leach, A. G.; Kuwata, K. T.; Chen, J. S.; Greer, A.; Foote, C. S.; Houk, K. N. Mechanism of Ene Reactions of Singlet Oxygen. A Two-Step No-Intermediate Mechanism. *J. Am. Chem. Soc.* **2003**, 125, 1319–1328. https://doi.org/10.1021/ja027225p.

(17) Singleton, D. A.; Hang, C.; Szymanski, M. J.; Greenwald, E. E. A New Form of Kinetic Isotope Effect. Dynamic Effects on Isotopic Selectivity and Regioselectivity. *J. Am. Chem. Soc.* **2003**, 125, 1176–1177. https://doi.org/10.1021/ja027221k.

(18) Bekele, T.; Christian, C. F.; Lipton, M. A.; Singleton, D. A. "Concerted" Transition State, Stepwise Mechanism. Dynamics Effects in C2-C6 Enyne Allene Cyclizations. *J. Am. Chem. Soc.* **2005**, 127, 9216–9223. https://doi.org/10.1021/ja0508673.

(19) Yamamoto, Y.; Hasegawa, H.; Yamataka, H. Dynamic Path Bifurcation in the Beckmann Reaction: Support from Kinetic Analyses. *J. Org. Chem.* **2011**, 76, 4652–4660. https://doi.org/10.1021/jo200728t.

(20) Katori, T.; Itoh, S.; Sato, M.; Yamataka, H. Reaction Pathways and Possible Path Bifurcation for the Schmidt Reaction. *J. Am. Chem. Soc.* **2010**, 132, 3413–3422.

(21) Itoh, S.; Yoshimura, N.; Sato, M.; Yamataka, H. Computational Study on the Reaction Pathway of α -Bromoacetophenones with Hydroxide Ion : Possible Path Bifurcation in the Addition / Substitution Mechanism. *J. Org. Chem.* **2011**, 76, 8294–8299.

(22) Bogle, X. S.; Singleton, D. A. Dynamic Origin of the Stereoselectivity of a Nucleophilic Substitution Reaction. *Org. Lett.* **2012**, 14, 2528–2531. https://doi.org/10.1021/ol300817a.

(23) Hansen, J. H.; Gregg, T. M.; Ovalles, S. R.; Lian, Y.; Autschbach, J.; Davies, H. M. L. On the Mechanism and Selectivity of the Combined C - H Activation / Cope Rearrangement. *J. Am. Chem. Soc.* **2011**, 133, 5076–5085.

(24) Hare, S. R.; Tantillo, D. J. Cryptic Post-Transition State Bifurcations That Reduce the Efficiency of Lactone-Forming Rh-Carbenoid C-H Insertions. *Chem. Sci.* **2017**, 8, 1442–1449. https://doi.org/10.1039/c6sc03745c.

(25) Ye, L.; Wang, Y.; Aue, D. H.; Zhang, L. Experimental and Computational Evidence for Gold Vinylidenes: Generation from Terminal Alkynes via a Bifurcation Pathway and Facile C-H Insertions. *J. Am. Chem. Soc.* **2012**, 134, 31–34. https://doi.org/10.1021/ja2091992.

(26) Zhang, L.; Wang, Y.; Yao, Z. J.; Wang, S.; Yu, Z. X. Kinetic or Dynamic Control on a Bifurcating Potential Energy Surface? An Experimental and DFT Study of Gold-Catalyzed Ring Expansion and Spirocyclization of 2-Propargyl-β-Tetrahydrocarbolines. *J. Am. Chem. Soc.* **2015**, 137, 13290–13300. https://doi.org/10.1021/jacs.5b05971.

(27) Noey, E. L.; Wang, X.; Houk, K. N. Selective Gold(I)-Catalyzed Formation of Tetracyclic Indolines: A Single Transition Structure and Bifurcations Lead to Multiple Products. *J. Org. Chem.* **2011**, 76, 3477–3483. https://doi.org/10.1021/jo200556f.

(28) Campos, R. B.; Tantillo, D. J. Designing Reactions with Post-Transition-State Bifurcations: Asynchronous Nitrene Insertions into C–C σ Bonds. *Chem* **2019**, 5, 227–236. https://doi.org/10.1016/j.chempr.2018.10.019.

(29) Major, D. T.; Weitman, M. Electrostatically Guided Dynamics-the Root of Fidelity in a Promiscuous Terpene Synthase? *J. Am. Chem. Soc.* **2012**, 134, 19454–19462. https://doi.org/10.1021/ja308295p.

(30) Siebert, M. R.; Zhang, J.; Addepalli, S. V.; Tantillo, D. J.; Hase, W. L. The Need for Enzymatic Steering in Abietic Acid Biosynthesis: Gas-Phase Chemical Dynamics Simulations of Carbocation Rearrangements on a Bifurcating Potential Energy Surface. *J. Am. Chem. Soc.* **2011**, 133, 8335–8343. https://doi.org/10.1021/ja201730y.

(31) Hong, Y. J.; Tantillo, D. J. A Potential Energy Surface Bifurcation in Terpene Biosynthesis. *Nat. Chem.* **2009**, 1, 384–389. https://doi.org/10.1038/nchem.287.

(32) Hong, Y. J.; Tantillo, D. J. How Cyclobutanes Are Assembled in Nature-Insights from Quantum Chemistry. *Chem. Soc. Rev.* **2014**, 43, 5042–5050. https://doi.org/10.1039/c3cs60452g.

(33) Ohashi, M.; Liu, F.; Hai, Y.; Chen, M.; Tang, M. C.; Yang, Z.; Sato, M.; Watanabe, K.; Houk, K. N.; Tang, Y. SAM-Dependent Enzyme-Catalysed Pericyclic Reactions in Natural Product Biosynthesis. *Nature* **2017**, 549, 502–506. https://doi.org/10.1038/nature23882.

(34) Hong, Y. J.; Tantillo, D. J. Biosynthetic Consequences of Multiple Sequential Post-Transition-State Bifurcations. *Nat. Chem.* **2014**, 6, 104–111. https://doi.org/10.1038/nchem.1843.

(35) Truhlar, D. G.; Garrett, B. C.; V, B. B.; K-, M. S.; Klippenstein, S. J. Current Status of Transition-State Theory. *J. Phys. Chem.* **1996**, 100, 12771–12800. https://doi.org/10.1021/jp953748q.

(36) Quapp, W.; Hirsch, M.; Heidrich, D. An Approach to Reaction Path Branching Using Valley – Ridge Inflection Points of Potential-Energy Surfaces. *Theor. Chim. Acta* **2004**, 112, 40–51.

https://doi.org/10.1007/s00214-003-0558-8.

(37) Hase, W. L.; Song, K.; Gordon, M. S. Direct Dynamics Simulations. *Comput. Sci. Eng.* **2003**, 5, 36–44.

(38) Hanessian, S.; Compain, P. Lewis Acid Promoted Cyclocondensations of Alpha;-Ketophosphonoenoates with Dienes - From Diels-Alder to Hetero Diels-Alder Reactions. *Tetrahedron* **2002**, 58, 6521–6529. https://doi.org/10.1016/S0040-4020(02)00662-2.

(39) RDKIT: open-source cheminformatics http://www.rdkit.org (accessed February 09, 2020).

(40) Merrer, D. C.; Rablen, P. R. Dichlorocarbene Addition to Cyclopropenes : A Computational Study. *J. Org. Chem.* **2005**, 70, 1630–1635. https://doi.org/10.1021/jo048161z.

(41) Oyola, Y.; Singleton, D. A. Dynamics and the Failure of Transition State Theory in Alkene Hydroboration. *J. Am. Chem. Soc.* **2009**, 131, 3130–3131.

(42) Suhrada, C. P.; Houk, K. N. Potential Surface for the Quadruply Degenerate Rearrangement of Bicyclo[3.1.0]Hex-2-Ene. *J. Am. Chem. Soc.* **2002**, 124, 8796–8797. https://doi.org/10.1021/ja020601l.

(43) Yang, Z.; Dong, X.; Yu, Y.; Yu, P.; Li, Y.; Jamieson, C.; Houk, K. N. Relationships between Product Ratios in Ambimodal Pericyclic Reactions and Bond Lengths in Transition Structures. *J. Am. Chem. Soc.* **2018**, 140, 3061-3067. https://doi.org/10.1021/jacs.7b13562.

(44) Xue, X. S.; Jamieson, C. S.; Garcia-Borràs, M.; Dong, X.; Yang, Z.; Houk, K. N. Ambimodal Trispericyclic Transition State and Dynamic Control of Periselectivity. *J. Am. Chem. Soc.* **2019**, 141, 1217-1221. https://doi.org/10.1021/jacs.8b12674.

(45) Weitman, M.; Major, D. T. Challenges Posed to Bornyl Diphosphate Synthase : Diverging Reaction Mechanisms in Monoterpenes. *J. Am. Chem. Soc.* **2010**, 132, 6349–6360.

(46) Hong, Y. J.; Tantillo, D. J. Quantum Chemical Dissection of the Classic Terpinyl / Pinyl / Bornyl / Camphyl Cation Conundrum — the Role of Pyrophosphate in Manipulating Pathways to Monoterpenes. *Org. Biomol. Chem.* **2010**, 8, 4589–4600.

https://doi.org/10.1039/c0ob00167h.

(47) Peterson, T. H.; Carpenter, B. K. Estimation of Dynamic Effects on Product Ratios by Vectorial Decomposition of a Reaction Coordinate. Application to Thermal Nitrogen Loss from Bicyclic Azo Compounds. *J. Am. Chem. Soc.* **1992**, 114, 766–767. https://doi.org/10.1021/ja00028a057.

(48) Mandal, N.; Datta, A. Dynamical Effects Along the Bifurcation Pathway Control Semibullvalene Formation in Deazetization Reactions. *J. Chem. Phys. B.,* **2018**, 8, 1239-1244. https://doi.org/10.1021/acs.jpcb.7b09533.

(49) Lopez, R. V.; Faza, O. N.; Lopez, C. S. Dynamic E Ff Ects Responsible for High Selectivity in a [3,3] Sigmatropic Rearrangement Featuring a Bispericyclic Transition State. *J. Org. Chem.* **2017**, 82, 4758–4765. https://doi.org/10.1021/acs.joc.7b00425.

(50) Gaussian 16, Revision B.01, 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, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016.

Insert Table of Contents artwork here