# Development of Reaction Discovery Tools for Photochemistry and Condensed Phases

by

Cody Aldaz

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in The University of Michigan
2020

Doctoral Committee:

Professor Paul M. Zimmerman, Chair
Professor Eitan Geva
Professor Heather B. Mayes
Professor Corey R. Stephenson

Cody R. Aldaz

craldaz@umich.edu

ORCID iD: 0000-0001-7672-8506

# DEDICATION

To those who have supported me,

even when it was undeserved.

Also, my Mom and Crystal,

they won't read this, but they like the pictures.

# ACKNOWLEDGMENTS

This thesis has been made possible by so many amazing, smart, and caring people. First, I want to thank Diane Marshall and Shannon McCoy-Hayes of the Maximizing Access to Research Careers (MARC) program at the University of New Mexico; at the beginning of my research career I was woefully ignorant of what it meant to be a researcher but they believed in me and made this all possible. Secondly, I want to thank my wonderful advisor, Paul Zimmerman, for giving me the opportunity to do research in his lab. He is truly one-of-a-kind when it comes to giving research help in a timely manner and I have been incredibly fortunate to have that resource. I would also like to thank Professor Roseanne Sension for allowing me to rotate in her lab and for her continued help and patience, without which I would have surely never completed Chapter 5. Similarly, I want to thank Professor Todd Martinez for having me out to Stanford for a summer. Todd and his amazing group of researchers welcomed me like family and opened my eyes to the wider nuances of computation, which ultimately encouraged me to develop my own software the "python growing string method" which is, in my opinion, the greatest accomplishment of my thesis. Todd also provided computational support well beyond my fellowship which I have greatly appreciated. I also want to thank all the Chemistry administrators and staff, Liz Oxford and Katie Foster always helped me out when I needed it. David Braun, our resident computer expert, kept our computer clusters running flawlessly and was always available to crack a joke. Finally, I would like to thank my committee members for spending their time guiding yet another graduate student towards graduation.

My research experience has also been positively impacted by many people in the lab and they deserve recognition. First, I want to thank Josh Kammeraad, Amanda Dewyer, Andy Vitek, Mina Jafari and Alan Chien. They were my first mentors and colleagues at the University of Michigan. They set the bar high for what it meant to be a researcher at Michigan and were always a reliable source of help when I needed it. I cherish them greatly. Secondly, I want to thank my mentors in the Sension lab, Ted Wiley and Nick Miller. They, like Roseanne, were instrumental in Chapter 5 and I am grateful for their help. Next, I want to thank the members of the Martinez lab, they made me feel at home in California and encouraged me to succeed which helped motivate me through tough times.

In particular, I want to thank Rob Parrish, Monica Williams, Stefan Seritan, Alice Walker, and Xiolei Zhu. They are some of smartest and most caring people that I have ever met. Finally, I want to give a shout out to the newer members of the Zimmerman group (with respect to when I joined): Michael Robo, Kevin Skinner, Alan Rask, Matt Hannigan, Timothy "Raz" Jugovic, Khoi Dang, Hanje Jiang, Zeke Punzalan, and anyone else that I missed. Their wit and talent always kept me on my toes. I'm so proud to see their research projects grow and develop and I can't wait to see what they will accomplish in the future.

My stay in Michigan has also been positively affected by my time outside the lab (although it hasn't been as much as I'd like!). I'd like to thank Alan Chien and Andy Vitek for attempting to teach me how to rock climb which went poorly. In turn, I also attempted to teach Alan how to skateboard which went just as bad. My friend Bobby Bonsall was a much better skateboarder and I wish him well on the remainder of his PhD, keep shredding dude! I'd like to thank my workout partners and friends Alan McClean, Alan Rask, and Zeke Punzalan. You all helped me stay fit and inspired a passion for fitness which I hope I can continue to support. Finally, I want to thank my friends Michael Robo and Alan McClean for all the fun times we had playing board games.

Last, but not least, I want to thank my family. My mom deserves so much credit for raising three boys and a daughter by herself. I attribute my strength and dedication in completing this thesis to her. I also want to acknowledge my girlfriend, Crystal. She has filled my time in Michigan with so much love and laughter.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF EQUATIONS

# LIST OF APPENDICES

# ABSTRACT

Photochemistry obeys different rules than ground-state chemistry and by doing so opens avenues for synthesis and materials properties. However, the different rules of photochemistry make understanding the fine details of photochemical reactions difficult. Computational chemistry can provide the details for understanding photochemical reactions, but the field of computational photochemistry is still new, and many techniques developed for ground-state reactions are not directly applicable to photochemical reactions. As a result, many photochemical mechanisms are not understood, and this hinders the rational design and synthesis of new photochemistry.

To address this need, this thesis develops techniques to search for and study photochemical reactions. Chapter 2 and 3 develop methods to calculate photochemical reactions in gas- and condensed-phases via minimum energy reaction paths. First, Chapter 2 develops a method to search the molecular 3N-6 space for photochemical reactions. This space, although vast, is not chaotic and can be efficiently searched using a concept familiar to chemists: breaking and adding bonds and driving angles and torsions. Furthermore, this procedure can be automated to predict new chemistry not previously identified by experiments. Chapter 3 furthers this research by leveraging the concept of molecules to enable the computational study of reactions in large multi-molecular systems like crystals. Specifically, the use of a new coordinate system involving translational and rotational coordinates allows decoupling of the coordinate systems of the individual molecules, which is necessary for the efficient algebra. Importantly, these methods are general, they can be used to study single molecules and crystals, and much in between.

These methods are demonstrated on complex chemical problems including the isomerization pathways of ethylene and stilbene (Chapter 2), the photocycloaddition of butadiene (Chapter 2), the rotation of a crystalline gyroscope (Chapter 3), the bicycle pedal rotation of cis,cis-diphenylbutadiene (Chapter 4), and the mechanism of a reversible photoacid (Chapter 5). These problems have value in understanding the processes of vision, optomechanics, and high-energy materials, and through their

study much needed insight is gained that can be useful for designing new syntheses and materials. Furthermore, the new computational methods open the possibility for many future investigations.

The results of Chapter 2 find a novel roaming-atom and hula-twist isomerization pathway and use automated reaction discovery tools to identify a missing butadiene photoproduct and why the [4+2] cycloaddition is forbidden. The results of Chapter 3 and 4 build on Chapter 2 by including the influence of a steric environment. Chapter 3 demonstrates by application to a molecular gyroscope that extreme long-range correlated motion can be captured with GSM, and Chapter 4 details how the one-bond flip and hula-twist mechanisms are suppressed by the crystal cavity, the nature of the seam space in steric environments, and the features of the bicycle pedal mechanism. For example, the bicycle pedals rotate through the passageway in the adjacent monomers. However, the models do not capture the quantitative activation barriers and more work is needed. Finally, Chapter 5 provides the ultrafast details of how the photoacid isomerizes and ring-closes with experimental and computational evidence. Unfortunately, quantitative calculation of $pK_a$ cannot be provided with the computations employed herein.

In summary, this thesis provides an advancement in the knowledge of photochemical mechanisms that can be used for the development of new syntheses and offers new tools with capacity to study complex photochemical problems.

# **Chapter 1** Introduction

## **1.1** Overview and Importance

Photochemistry is the branch of chemistry concerned with the chemical effects of light. Photochemical reactions proceed differently than temperature-driven reactions by accessing high-energy excited-state intermediates, which have different chemical properties and reactivity than their ground-state counterparts[1–4]. Under light irradiation double-bonds become single-bonds[5,6], aromatics become anti-aromatic[7], and reactions which are ground-state forbidden become allowed[8]. The use of light energy also offers several physical advantages such as the possibility of using sunlight to drive reactions[9,10], and the use of lasers to control the spatial and temporal extent of a reaction (e.g., photo-lithography and photo-switching)[11]. However, despite these benefits, the dark side of photochemistry is that we have developed a less clear picture of photochemical reactions than thermal reactions because they occur on an ultrafast timescale and obey more complex principles. For example, the modern theory of photochemical reactions was not widely accepted until the early 1990s[12] and even the details of "simple" photochemical reactions like double-bond isomerization (which will be discussed in detail in this thesis) are still being discovered. Consequently, photochemistry is an area of chemistry where many reactions are useful but a general lack of understanding limits rational design and synthesis.

The obvious solution, then, is to learn more about this intriguing class of chemistry. Alas, learning more about photochemical mechanisms is easier said than done. Exposing many compounds to light over a long enough time can be a great way to learn about the scope of photochemical reactions, and in fact photochemical "farms" were some of the first examples of photochemistry research[13]. By learning the patterns in the experimental data, some details about the underlying mechanism and its atomic properties can be ascertained. However, without better spatial and temporal resolution, or otherwise cleverly trapping the intermediates[14,15], which is not always possible, the finer

details of photochemical reactions will always remain unknown. For example, meta arene-alkene photocycloaddition is one of the most widely studied photochemical reactions with several hundred examples known[16–18]. It reacts first by [3+2] cycloaddition of the arene and alkene and is followed by cyclopropane ring formation. The direction of the [3+2] cycloaddition is well explained by the polarization of the arene upon interaction with the alkene, but there exists no satisfying explanation of the later stage of the reaction[19,20]. Another example that demonstrates why photochemistry is more than meets the eye is Woodward-Hoffman rules. Woodward-Hoffman's rules describe why certain types of reaction are allowed in the excited-state by considering the symmetries of the molecular orbitals. While this theory is based in quantum mechanics, the actual mechanisms of photoexcited molecules is much more complex. Therefore, the application of simpler trends and models are of only limited practical application and more advanced analyses are required.

Fortunately, there is a better way to gain a closer look at the mechanism of photochemical reactions, and that is through computation. For example, detailed information can be acquired by experiments like transient absorption spectroscopy which can probe the ultrafast (e.g. less than 1 nanosecond) mechanisms of photoexcited reactions, including the nature and dynamics of the excited-states. This information, however, must be interpreted from spectroscopic signatures, and this can be done through first-principles calculations (see Chapter 5). The fine details of reaction mechanisms can also be computed with high precision[21,22] and used to guide experiments. For example, in computational chemistry, the location of the reactant, product and pathway connecting them via a transition state (TS) provide a great amount of detail including the structural details of the reaction like the electronic and steric parameters, and the kinetics (i.e. rate) of the reaction[23].

The true power and promise of computation, however, lies in its predictive capabilities such as automated reaction discovery. Much work in recent years has shown that the location of transition states, which normally requires prior chemical knowledge and tedious manual steps, can be automated to predict reaction intermediates and products which were previously unidentified[24,25]. However, photochemical reactions occur on two or more potential energy surfaces, the excited-state(s) and the ground-state, and typically involve a crossing "seam" between the potential energy surfaces[26,27]. Therefore, the computational approaches which have been successful for studying and predicting ground-state reactions are not directly applicable to photochemical reactions. Consequently, most

computational photochemistry relies on extensive prior chemical knowledge and intuition and this limits their predictive and exploratory capabilities.

Another area where computational chemistry is lacking is the study of reactions in condensed phases. Constraining photochemical reactions in condensed phases (e.g. crystal) is a particularly promising avenue[28,29] because the lifetime of excited molecules is short-lived (~ps) so pre-binding the molecules greatly increases their probability of reacting. Furthermore, photochemistry can react differently under confinement than in the gas-phase. For example, photoisomerization in crystals and proteins can take on new volume-conserving mechanisms which are not present in non-sterically demanding environments; an example of this can be found in Chapter 4 and reference [30]. Restrictions on intramolecular rotations or specific interactions with the environment can also help make the reactions more efficient by directing the photon energy towards specific productive reaction outcomes[31,32]. For example, the process of vision in the eye is initiated by the photoisomerization of retinal in the rhodopsin protein. In the protein environment, isomerization occurs exclusively around a single bond 11-cis→all-trans[33], and is highly photon efficient (>50%). In contrast, in solution the isomerization results in several photoproducts, is much less efficient (~25%), and is about 20 times slower[34]. The computational study of photochemical reactions in condensed phases, however, is more difficult because of the increased system size, and/or increased difficulty in treating the dielectric medium.

This thesis will present new computational techniques for the study of and prediction of photochemical reactions in gas- and condensed-phases and analyze photochemical reactions via a combination of theoretical and experimental evidence. These approaches leverage, where possible, the mathematical structure of chemistry to best solve the chemistry at hand. The first part will deal with the effective study of photochemical reactions, including the prediction of photochemical reactions from first principles[27], the second part will deals with techniques for studying reactions in condensed phases, including crystals, and the last part will deal with a detailed theoretical and experimental investigation of two photoacids in water and aprotic solvent[35,36]. We begin with preliminary details on the theory of photochemistry and computational photochemistry, and then provide an outline for the remaining thesis.

## **1.2** Preliminary Details

Here is provided preliminary details on the key concepts in computational chemistry and ultrafast spectroscopy which will be useful in later chapters.

### **1.2.1** Minimum energy reaction paths

This section will describe minimum energy reaction paths, which is a central computational strategy used in this thesis to study photochemical reactions. To begin, we know that a chemical system can be composed of atoms, ions, and molecules. These particles are very small. To get a feeling for just how small atoms are, astronomers have estimated that there are about 10,000 billion, billion stars in the observable universe ($10^{22}$). In comparison, you would only need about 1 gram of sand (about 1 grain) to have an equal number of atoms.

The very small size of atoms means that only the most advanced instruments can view individual atoms, and that it is impossible to model any macroscopic object, with a galactic number of atoms. The very largest simulations currently available through first-principles calculations is about 1 million atoms, or about the size of the smallest viruses[37]. More commonly, however, fully quantum mechanical calculations of greater than a few hundred atoms can become too computationally expensive to be feasible. Quantum mechanics/molecular mechanics can also be used to reduce the total cost of studying large systems by treating a subset of atoms with simpler empirical forces, but these are still memory intensive calculations[38].

Another factor to consider is that atoms are always moving, vibrating, and rotating. Even at the theoretically coldest temperature, 0 Kelvin, a molecule will vibrate with so-called zero-point energy. So, even if we can view atoms in a chemical system like a solution, it would appear like random vibrations and fluctuations. The picture is also more complex when considering reactions. Chemical reactions must satisfy several criteria to be successful for example, in a bimolecular reaction:

1. the reacting particles must collide with each other
2. the reacting particles must have enough energy to break the old bonds
3. the particles must have the proper orientation

Consequently, because the vibrations of the individual particles occur on the $10^{-15}$ s timescale, and the chance of two particles colliding with enough energy and the proper orientation is much rarer, and far between (e.g. seconds or longer), it would not be easy to observe a reaction take place.

Fortunately, however, atoms and molecules follow straightforward principles of energy minimization which can make studying reactions and molecules easier. Like how a ball will roll down a bowl to the bottom, chemical systems will approach a minimum energy at equilibrium, and the random fluctuations of vibrating molecules can be averaged out. For example, the average positions of an ensemble of identical molecules with different vibrational phases at equilibrium is the minimum energy geometry on the potential energy surface. Similarly, if one were to look at the reaction events of a real system (i.e. the fluctuating ensemble of reactants), only the reactants with the proper orientation will lead to products and their geometries will be distributed normally around the minimum energy path (Figure 1-1).

The highest point along the minimum energy valley is called the transition state, which can be mathematically described as a first-order saddle point. Following from the principle of energy minimization the transition state has the special property that it determines the speed of the reaction: if the transition state energy is large, then the principle of energy minimization tells us that getting to that point is a very rare event, but if the energy is small then reaching this geometry happens much more frequently. The relationship between speed of a reaction i.e. rate and energy of the transition state, at a certain temperature, is described in the important equation

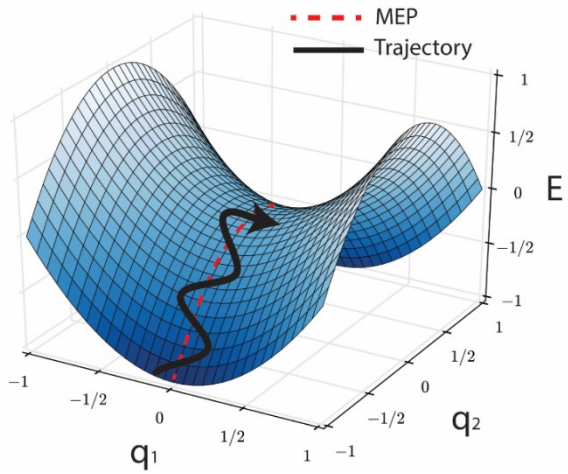



**Figure 1-1** Potential energy surface around the transition state geometry, i.e. saddle point. A real molecule with vibrational energy will tend to sample geometries equally distributed around the minimum energy path as it passes through the transition state.

$$k = \frac{k_B T}{h} e^{-\frac{E_a}{RT}}$$

**Equation 1-1** The Eyring equation

Where $E_a$ is the activation energy, R is the gas constant, $k_B$ is Boltzmann's constant, h is Planck's constant, and T is temperature.

Consequently, finding the transition state geometry is important task for the computational chemist for learning about chemical reactions. For the great majority of chemical reactions, it is not necessary to simulate a realistic looking chemical system with millions of atoms or to have the chemical system obey any time-parameterized equation of motion (e.g. Newton's equation of motion). Instead, we can think about the individual reactants and find the minimum energy pathways that connects reactants to products, because these are representative of the average positions that the real reactions would follow. Solvation effects can be added via implicit solvent models which simulate the dielectric nature of the chosen solvent, and entropic and enthalpic effects can be added by treating the atoms as a quantum rigid-rotor harmonic oscillator[39,40].

The process of calculating a transition state geometry, however, is still difficult[41]. There are 3N-6 degrees of freedom for a reacting system, where N is the number of atoms. Geometry

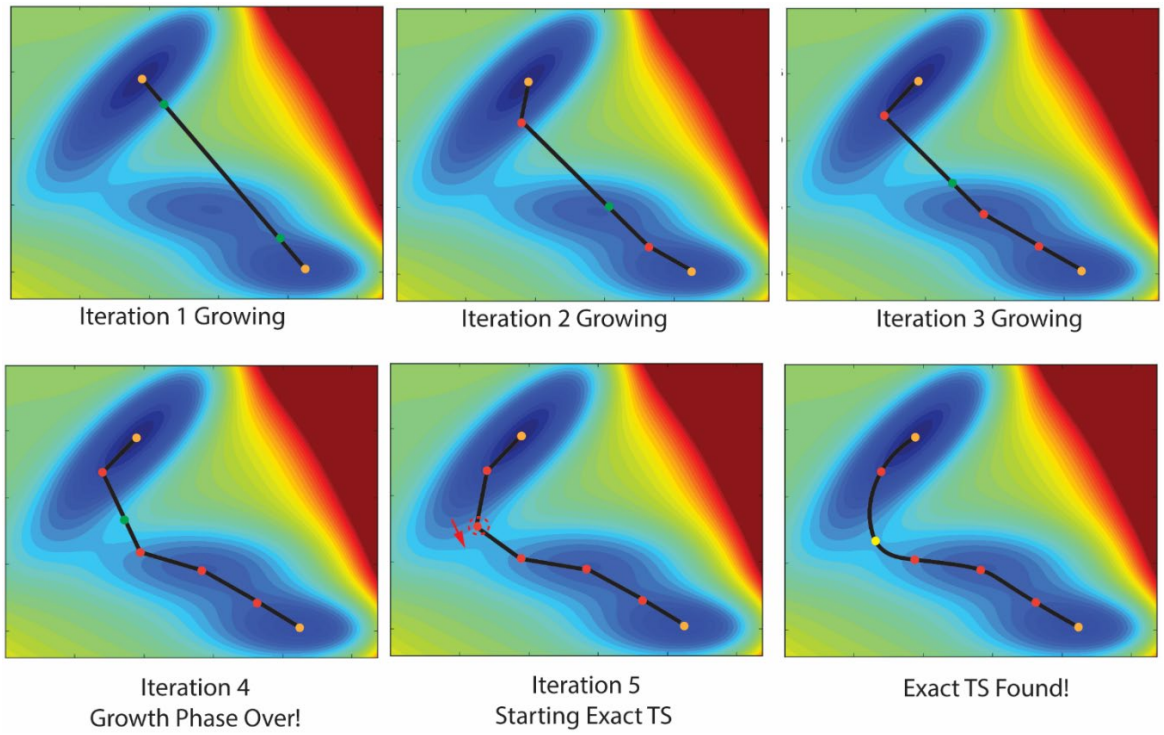



**Figure 1-2** Contour plot illustration of the double-ended growing string method (GSM) on the Muller-Brown potential. GSM iteratively grows the reaction path to avoid placing nodes in high-energy regions of space (e.g. the green region) which can cause the simulation to fail. After finishing growing an exact TS is found. The yellow node is the transition state which is a first-order saddle point connecting the reactant and product wells.

optimization of reactants should therefore proceed in (ideally) no more than 3N-6 steps, one step for each degree of freedom. However, the process of finding a transition state is more difficult than reactant optimization because these algorithms will not optimize to the higher energy transition state given a lower energy (more stable) geometry nearby. Thus, the common strategy to find transition state geometries is to manually move the geometry closer to the transition state before starting the optimization algorithm. Understandably this technique is very error prone, even for experts, because the transition state is unstable in the reaction direction. For these reasons, much effort has gone into the development of transition states algorithms that are more reliable, and do not require manually guessing the transition state geometry[23,42–44]. The approach that this thesis focuses on is called the Growing String Method (GSM)[23,42].

Chapters 2 and 3 will provide more information on GSM so only a brief overview will be given now. First, GSM can operate in either double-ended[23] or single-ended mode[42]. An illustration of the double-ended GSM is given in Figure 1-2. DE-GSM begins by interpolating two geometries along the reaction path tangent towards the transition state, and then optimizing them in all orthogonal degrees of freedom. If either of the frontier nodes become sufficiently low energy in the directions orthogonal to the reaction tangent, as determined by the gradient of the potential energy surface, then new nodes can be grown along the tangent connecting the frontier nodes. See for example, the green nodes in Iteration 2 and 3 of Figure 1-2. This process is repeated until a specified number of nodes are grown, usually 9 or 11[23]. After the string has converged to a predefined threshold an exact transition state search is started and the TS is found. Overall, this process is much preferable to the guess and check method of standard TS optimization. Furthermore, DE-GSM is more efficient in comparison to other reaction-path optimizers because the entire path is not interpolated at once but is done so iteratively which avoids placing the nodes in high-energy regions of space.

In SE-GM the reaction path is grown from the reactant side given a search direction called a driving coordinate, which can consist of any combination of adding and breaking bonds and driving angle and torsion. For example, "ADD ATOM 1 to ATOM 2", if they are not already bonded. This is a highly efficient way of looking for transition states because a product structure is not required, and SE-GSM operates identically to DE-GSM after the string is fully grown. Furthermore, in addition to guessing the driving coordinate based on the geometry and the expected products, the driving coordinates can be generated combinatorially between a set of reactive atoms. Therefore, SE-GSM

can help discover new chemistry and identify missing pathways which would not be obvious to even the expert chemist[45,46].

### **1.2.2** Photochemical Mechanisms

This section will describe the elementary differences between thermal and photochemical reactions, in Chapter 2 the more advanced aspects of photochemical reactions will be described. The basic additional requirement for a photochemical reaction is that light must be absorbed, and the reaction must take place in an excited-state (i.e. thermal heating by the light absorption is not photochemistry). The actual details of a photochemical reaction, however, are more complicated and can proceed in more than one way

1. A photochemical reaction can occur in a similar way as a reaction on the ground-state except that it starts with the absorption of light and ends with the fluorescence of light. A reaction step that proceeds between two potential energy basins on the same electronic state is referred to as an **adiabatic** process[47].
2. A photochemical reaction can proceed by decay through the crossing region between potential energy surfaces known as a conical intersection. A reaction step that proceeds by passage through a conical intersection, either directly or first over a barrier without an intermediate energy well is known as a **non-adiabatic** process[48].
3. A photochemical reaction can proceed by intersystem crossing to a potential energy surface which has different spin-multiplicity. This is like decay through a conical intersection but involves spin-orbit coupling instead of non-adiabatic coupling[49].

An example of the difference between photochemical and thermal reactions is given in Figure 1-3. Interestingly, real crossing points were once believed to be rare, but it is now known that they are common[12].

This thesis deals primarily with photochemical reactions that involve conical intersections, but both adiabatic and non-adiabatic pathways are discussed. The study of intersystem crossing is also possible through the same machinery discussed in greater detail below, except for a slight difference: the seam is 3N-8 dimensional for conical intersections but is 3N-7 dimensional for intersystem crossing. Chapter 2 dives deeper into the details of conical intersections, so the meaning of this difference won't be explained here but should be kept in mind when studying intersystem crossing.

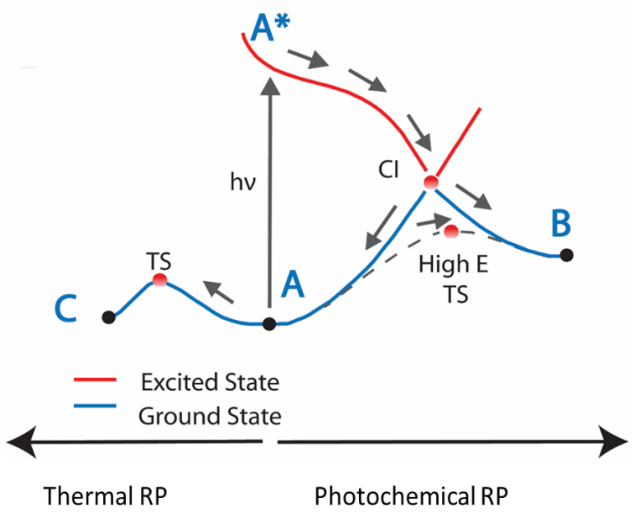



**Figure 1-3** Example difference between thermal and photochemical reaction paths. Thermal reactions involve adiabatic passage over a transition state. Photochemical reactions can involve non-adiabatic passage (which can also include a transition state) through a conical intersection. This figure also illustrates why photochemical reactions are often considered forbidden on the ground-state because the ground-state transition state is often very high energy for these transformations.

### **1.2.3** Electronic Structure Theory

The energy of molecules is obtained from first principles by the Schrodinger equation

$$\hat{H}\Psi(\mathbf{r};\mathbf{R}) = E\Psi(\boldsymbol{r};\boldsymbol{R})$$

**Equation 1-2** The Schrodinger equation

Where $\Psi$ is the wavefunction of the electrons, which is a function of the positions of the electrons, $\boldsymbol{r}$, and parameterized by the positions of the atomic nuclei, $\boldsymbol{R}$, $\hat{H}$ is the electronic Hamiltonian, and $E$ is the electronic energy. The Hamiltonian includes the kinetic energy of the electrons, the kinetic energy of the nuclei, the electron-electron repulsion, the electron-nuclear attraction, and the nuclear-nuclear repulsion. The exact solution of Equation 1-2 is not possible for multi-electron wavefunctions but the exact answer can be approached with sophisticated theories.

The common solutions of Equation 1-2 is the Hartree-Fock theory, and density functional theory[50]. The investigation of conical intersections, however, require electronic structure theories which treat ground- and excited-states on an equal footing, known as multi-reference theories. Single-reference methods like Hartree-Fock and DFT, do not properly capture the cone-shape of a conical intersection[51,52]. In this thesis, we will use the multi-reference, complete-active space self-consistent (CASSCF) theory for all calculations of conical intersections. Unlike HF and DFT, however, CASSCF

is not a blackbox method and requires carefully selecting the size and identity of the orbitals in the active space[53]. For example, for cycloaddition and cyclization reactions the orbitals determined by Woodward Hoffmann rules to be important are a good place to start. Also, the bonding and antibonding orbital of any bond that is breaking or forming should be in the active space. However, the choice of the active space is a non-trivial task and the interested reader should consult the literature[53]. Nevertheless, once chosen, the geometry and reaction path optimization algorithms described in chapters 2 and 3 which obtain CASSCF energies and energy gradients from external packages like MOLPRO[54] and TeraChem[55] take careful precaution to maintain active space consistency.

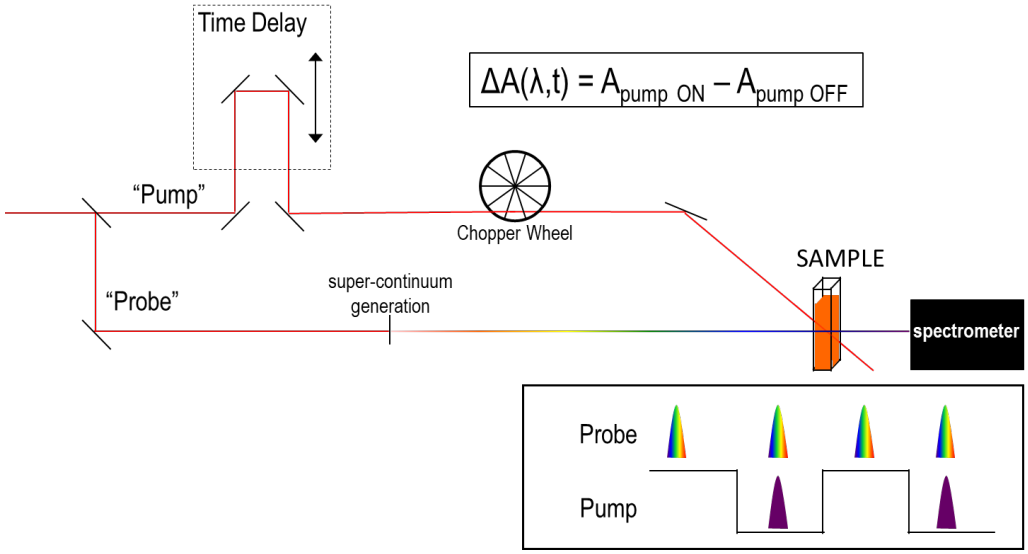



**Figure 1-4** Schematic of the transient absorption spectrometer. Note that the signal reported in transient absorption is the difference with pump on and pump off as function of time, $\boldsymbol{\Delta A(\lambda, t)}$. This allows the excited-state processes to be studied as a function of time without convolution with the ground-state. Courtesy of Dr. Ted Wiley.

### **1.2.4** Transient Absorption Spectroscopy

Transient absorption spectroscopy is an experimental technique to measure the ultrafast details of a light-activated process. This technique can track the evolution of processes that occur on very short timescale that are relevant to photochemistry (e.g. femtosecond to nanoseconds) and can give access to information about the excited-states.

Transient absorption spectroscopy is performed by a series of pump and probe measurements (Figure 1-4). First, the probe pulse which is a "white" light continuum measures the ground-state absorption. Next, the pump and probe, which are delayed by the variable time-delay stage, impinge

on the sample. The pump pulse was previously blocked by the chopper wheel. The pump arrives first and excites some of the molecules to the excited state (but not all!), and therefore when the probe arrives it measures a combination of ground and excited-state absorptions. The probe also causes some stimulated emission from the excited-state. However, because this pump on measurement measures a combination of ground and excited state absorption it is necessary to remove ground-state absorption and this is the purpose of the pump off measurement. By taking the difference of pump on, and pump off measurements the contribution of the ground-state can be removed. Therefore, the four types of signals in transient absorption spectroscopy are:

1. ground-state bleach, a negative signal in the region of the ground-state absorption that arises from a decrease of ground-state population due to photoexcitation
2. excited-state absorption, a positive signal arising from an increase in absorption from the excited-state
3. stimulated emission, a negative signal that stems from the stimulated fluorescence of excited molecules by light. From the detectors point of view this is negative absorption. Typically, this band resides in the same region as fluorescence.
4. photoproduct absorption, a positive signal that will not have accompanying stimulated emission, and will also be long lived

## **1.3** Dissertation Outline

Chapter 1 presents an introduction to the importance and challenges of computational photochemistry including the need to effectively search for conical intersections and study photochemistry in condensed phases. A background on minimum energy reaction paths, photochemical mechanisms and transient absorption spectroscopy is given because later chapters will rely on advanced aspects of these topics, without much review.

Chapter 2 presents algorithms for the study of photochemical reactions including a single-ended method for the location of conical intersections and a growing string method for the location of seams of conical intersections in either single- or double-ended mode. These techniques are applied to the photoisomerization of stilbene and the photocycloaddition of butadiene. In the former, a wider array of isomerization pathways was found, including a conical intersection likely responsible for

isomerization in sterically constraining environments like proteins and crystals. In the latter, the experiments failed to identify one of the photoproducts. The computations identify the most likely missing photoproduct via an automated reaction discovery algorithm.

Chapter 3 presents a linear scaling growing string method for the study of large multi-molecule systems such as crystals and proteins. The new algorithms achieve linear scaling in multi-molecule systems by block-matrix linear algebra and by distributing the reaction tangent, which is used to grow and optimize the minimum energy path, into each block which represents a molecule or fragment. This tangent is a multi-dimensional vector, and acts as a constraint, which prevents motion along that local direction. This results in a stable GSM algorithm for large multi-molecule systems and has the added benefit of being a good proxy for correlation analysis. The method is demonstrated by application to a crystalline gyroscope, which shows extreme long-range correlated motion.

Chapter 4 presents a computational picture of the bicycle pedal isomerization, a photochemical process which occurs exclusively in the crystal phase. Much physical insight is gained into how the steric environment influences isomerization mechanisms including how the one-bond flip and hula-twist mechanisms are suppressed by the crystal cavity, the nature of the seam space in steric environments, and the mechanics of the bicycle pedal motion.

Chapter 5 present a combined theoretical and experimental analysis of two reversible photoacids. The study provides the early time details of the mechanism including how the trans-acid isomerizes to the cis-form, and the timescale for deprotonation, and ring-closing. Also, the reason for why the photoacids are acidic is investigated by computation.

Chapter 6 includes Future Work and Conclusions, in which the findings of the prior chapters are reviewed, and some ideas on extending the tools and capabilities outlined in this thesis are given. Minimum energy reaction paths are a foundational tool for the study of chemical reactions, including photochemistry. The work presented herein opens many new avenues for computational investigations of reaction mechanisms, and much insight into long-standing chemical problems has already been found

## **1.4** References


(1) Barachevsky, V. A. Applied Aspects of Organic Photochemistry. **2003**, *37* (1), 6–16.

(2) Kagan, J. *Organic Photochemistry*; Cambridge Texts in Chemistry and Biochemistry; Cambridge University Press, 1993. https://doi.org/10.1016/B978-0-12-394320-0.50015-6.

(3) Pfoertner, K. H. Photochemistry in Industrial Synthesis. *J. Photochem. Photobiol. A Chem.* **1990**, *51* (1), 81–86. https://doi.org/10.1016/1010-6030(90)87045-D.

(4) Bitterwolf, T. E. Organometallic Photochemistry at the End of Its First Century. **2004**, *689*, 3939–3952. https://doi.org/10.1016/j.jorganchem.2004.06.023.

(5) Gagliardi, L.; Orlandi, G.; Molina, V.; Malmqvist, P.-A.; Roos, B. O. Theoretical Study of the Lowest 1 B U States of Trans -Stilbene. *J. Phys. Chem. A* **2002**, *106* (32), 7355–7361. https://doi.org/10.1021/jp0256138.

(6) Hohlneicher, G.; Dick, B. Experimental Determination of the Low-Lying States of Trans-Stilbene. *J. Photochem.* **1984**, *27* (2), 215–231.

(7) Zimmerman, H. E. The Meta Effect in Organic Photochemistry: Mechanistic and Exploratory Organic Photochemistry. *J. Am. Chem. Soc.* **1995**, *117* (98), 8988–8991. https://doi.org/10.1021/ja00140a014.

(8) Bernardi, F.; De, S.; Olivucci, M.; Robb, M. a. Mechanism of Ground-State-Forbidden Photochemical Pericyclic Reactions: Evidence for Real Conical Intersections. *J. Am. Chem. Soc.* **1990**, *112* (5), 1737–1744. https://doi.org/10.1021/ja00161a013.

(9) Balzani, V.; Clemente-Leon, M.; Credi, A.; Ferrer, B.; Venturi, M.; Flood, A. H.; Stoddart, J. F. Autonomous Artificial Nanomotor Powered by Sunlight. *Proc. Natl. Acad. Sci.* **2006**, *103* (5), 1178–1183. https://doi.org/10.1073/pnas.0509011103.

(10) Oelgemoller, M. Solar Photochemical Synthesis: From the Beginnings of Organic Photochemistry to the Solar Manufacturing of Commodity Chemicals. *Chem. Rev. (Washington, DC, United States)* **2016**, *116* (17), 9664–9682. https://doi.org/10.1021/acs.chemrev.5b00720.

(11) Russew, M. M.; Hecht, S. Photoswitches: From Molecules to Materials. *Adv. Mater.* **2010**, *22* (31), 3348–3360. https://doi.org/10.1002/adma.200904102.

(12) Yarkony, D. R. Nonadiabatic Quantum Chemistry — Past , Present , and Future. *Chem. Rev.* **2012**, *112*, 481–498.

(13) Roth, H. D. The Beginnings of Organic Photochemistry. *Angew. Chemie Int. Ed. English* **1989**, *28* (9), 1193–1207. https://doi.org/10.1002/anie.198911931.

(14) Saltiel, J.; Redwood, C. E. Photochemistry of the 1,4-Diphenyl-1,3-Butadienes in Ethanol. Trapping Conical Intersections. *J. Phys. Chem. A* **2016**, *120* (18), 2832–2840. https://doi.org/10.1021/acs.jpca.6b02330.

(15) Park, K.; West, A.; Raheja, E.; Sellner, B.; Lischka, H.; Windus, T. L.; Hase, W. L. Singlet and Triplet Potential Surfaces for the O2+ C2H4reaction. *J. Chem. Phys.* **2010**, *133* (18). https://doi.org/10.1063/1.3490480.

(16) Chappell, D.; Russell, A. T.; Chappell, D. From a -Cedrene to Crinipellin B and Onward : 25 Years of the Alkene – Arene Meta -Photocycloaddition Reaction in Natural Product Synthesis. **2006**, 4409–4430. https://doi.org/10.1039/B614011B.

(17) Cornelisse, J. The Meta Photocycloaddition of Arenes to Alkenes. *Chem. Rev.* **1993**, *93* (2), 615–669. https://doi.org/10.1021/cr00018a002.

(18) Streit, U.; Bochet, C. G. The Arene–alkene Photocycloaddition. *Beilstein J. Org. Chem.* **2011**, *7*, 525–542. https://doi.org/10.3762/bjoc.7.61.

(19) Wegmann, M.; Bach, T. Influence of the -CH2X Substituent on the Regioselectivity of Intramolecular Meta -Photocycloaddition Reactions. *J. Org. Chem.* **2015**, *80* (3), 2017–2023. https://doi.org/10.1021/jo5028613.
(20) Serrano-Pérez, J. J.; De Vleeschouwer, F.; De Proft, F.; Mendive-Tapia, D.; Bearpark, M. J.; Robb, M. a. How the Conical Intersection Seam Controls Chemical Selectivity in the Photocycloaddition of Ethylene and Benzene. *J. Org. Chem.* **2013**, *78* (5), 1874–1886. https://doi.org/10.1021/jo3017549.
(21) Boggio-pasqua, M. *Computational Mechanistic Photochemistry : The Central Role of Conical Intersections*; Universite Toulouse III <tel-01184241>, 2015.
(22) Schapiro, I.; Melaccio, F.; Laricheva, E. N.; Olivucci, M. Using the Computer to Understand the Chemistry of Conical Intersections. *Photochem. Photobiol. Sci.* **2011**, *10* (6), 867–886. https://doi.org/10.1039/c0pp00290a.
(23) Zimmerman, P. Reliable Transition State Searches Integrated with the Growing String Method. *J. Chem. Theory Comput.* **2013**, *9* (7), 3043–3050. https://doi.org/10.1021/ct400319w.
(24) Zimmerman, P. M. Automated Discovery of Chemically Reasonable Elementary Reaction Steps. *J. Comput. Chem.* **2013**, *34* (16), 1385–1392. https://doi.org/10.1002/jcc.23271.
(25) Dewyer, A. L.; Zimmerman, P. M. Finding Reaction Mechanisms, Intuitive or Otherwise. *Org. Biomol. Chem.* **2017**, 501–504. https://doi.org/10.1039/C6OB02183B.
(26) Coe, J. D.; Ong, M. T.; Levine, B. G.; Martínez, T. J. On the Extent and Connectivity of Conical Intersection Seams and the Effects of Three-State Intersections. *J. Phys. Chem. A* **2008**, *112* (49), 12559–12567. https://doi.org/10.1021/jp806072k.
(27) Aldaz, C.; Kammeraad, J. A.; Zimmerman, P. M. Discovery of Conical Intersection Mediated Photochemistry with Growing String Methods. *Phys. Chem. Chem. Phys.* **2018**, *20* (43), 27394–27405. https://doi.org/10.1039/C8CP04703K.
(28) Ramamurthy, V.; Sivaguru, J. Supramolecular Photochemistry as a Potential Synthetic Tool: Photocycloaddition. *Chem. Rev.* **2016**, *116* (17), 9914–9993. https://doi.org/10.1021/acs.chemrev.6b00040.
(29) Liu, R. S. H.; Yang, L.-Y.; Liu, J. Mechanisms of Photoisomerization of Polyenes in Confined Media. From Organic Glasses to Protein Binding Cavities. *Photochem. Photobiol.* **2006**, 2–10. https://doi.org/10.1562/2006-01-27-RA-786.
(30) Saltiel, J.; Krishna, T. S. R.; Clark, R. J. Photoisomerization of Cis,Cis-1,4-Diphenyl-1,3-Butadiene in the Solid State: The Bicycle-Pedal Mechanism. *J. Phys. Chem. A* **2006**, *110* (5), 1694–1697. https://doi.org/10.1021/jp056700d.
(31) Schapiro, I. The Origin of Bond Selectivity and Excited-State Reactivity in Retinal Analogues. *J. Phys. Chem. A* **2016**, *120* (19), 3353–3365. https://doi.org/10.1021/acs.jpca.6b00701.
(32) Ben-Nun, M.; Molnar, F.; Schulten, K.; Martínez, T. J. The Role of Intersection Topography in Bond Selectivity of Cis-Trans Photoisomerization. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (4), 1769–1773. https://doi.org/10.1073/pnas.032658099.
(33) Tuna, D.; Lefrancois, D.; Dreuw, A.; Wolanski, L.; Andruniow, T.; Gozem, S.; Schapiro, I.; Olivucci, M.; Olivucci, M. Assessment of Approximate Coupled-Cluster and Algebraic-Diagrammatic-Construction Methods for Ground- and Excited-State Reaction Paths and the Conical-Intersection Seam of a Retinal-Chromophore Model. *J. Chem. Theory Comput.* **2015**, *11* (12), 5758–5781.
(34) Cembran, A.; Bernardi, F.; Olivucci, M.; Garavelli, M. Counterion Controlled Photoisomerization of Retinal Chromophore Models: A Computational Investigation. *J. Am.*

*Chem. Soc.* **2004**, *126* (49), 16018–16037. https://doi.org/10.1021/ja048782+.
(35) Abeyrathna, N.; Liao, Y. A Reversible Photoacid Functioning in PBS Buffer under Visible Light. *J. Am. Chem. Soc.* **2015**, *99* (35), 8–11. https://doi.org/10.1021/jacs.5b06218.
(36) Shi, Z.; Peng, P.; Strohecker, D.; Liao, Y. Long-Lived Photoacid Based upon a Photochromic Reaction. *J. Am. Chem. Soc.* **2011**, *133* (37), 14699–14703. https://doi.org/10.1021/ja203851c.
(37) Andermatt, S.; Cha, J.; Schiffmann, F.; VandeVondele, J. Combining Linear-Scaling DFT with Subsystem DFT in Born-Oppenheimer and Ehrenfest Molecular Dynamics Simulations: From Molecules to a Virus in Solution. *J. Chem. Theory Comput.* **2016**, *12* (7), 3214–3227. https://doi.org/10.1021/acs.jctc.6b00398.
(38) Lin, H.; Truhlar, D. G. QM/MM: What Have We Learned, Where Are We, and Where Do We Go from Here? *Theor. Chem. Acc.* **2007**, *117* (2), 185–199. https://doi.org/10.1007/s00214-006-0143-z.
(39) Marenich, A. V; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. **2009**, 6378–6396.
(40) Kamerlin, S. C. L.; Haranczyk, M.; Warshel, A. Are Mixed Explicit/Implicit Solvation Models Reliable for Studying Phosphate Hydrolysis? A Comparative Study of Continuum, Explicit and Mixed Solvation Models. *ChemPhysChem* **2009**, *10* (7), 1125–1134. https://doi.org/10.1002/cphc.200800753.
(41) Schlegel, H. B. Geometry Optimization. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1* (5), 790–809. https://doi.org/10.1002/wcms.34.
(42) Zimmerman, P. M. Single-Ended Transition State Finding with the Growing String Method. *J. Comput. Chem.* **2015**, *36* (9), 601–611. https://doi.org/10.1002/jcc.23833.
(43) Weinan, E.; Ren, W.; Vanden-eijnden, E. Simplified and Improved String Method for Computing the Minimum Energy Paths in Barrier-Crossing Events Simplified and Improved String Method for Computing the Minimum Energy Paths in Barrier-Crossing Events. **2009**, *164103* (2007). https://doi.org/10.1063/1.2720838.
(44) Kundu, S.; Bhattacharjee, S.; Lee, S.-C.; Jain, M. PASTA: Python Algorithms for Searching Transition StAtes. *Comput. Phys. Commun.* **2018**, No. April. https://doi.org/10.1016/j.cpc.2018.06.026.
(45) Nett, A. J.; Montgomery, J.; Zimmerman, P. M. Entrances, Traps, and Rate-Controlling Factors for Nickel-Catalyzed C − H Functionalization. **2017**. https://doi.org/10.1021/acscatal.7b02919.
(46) Pendleton, I. M.; Pérez-Temprano, M. H.; Sanford, M. S.; Zimmerman, P. M. Experimental and Computational Assessment of Reactivity and Mechanism in C(Sp 3 )–N Bond-Forming Reductive Elimination from Palladium(IV). *J. Am. Chem. Soc.* **2016**, *138* (18), 6049–6060. https://doi.org/10.1021/jacs.6b02714.
(47) Turro, N. J.; McVey, J.; Ramamurthy, V.; Lechtken, P. Adiabatic Photoreactions of Organic Molecules. *Angew. Chemie Int. Ed. English* **1979**, *18* (8), 572–586. https://doi.org/10.1002/anie.197905721.
(48) Matsika, S.; Krause, P. Nonadiabatic Events and Conical Intersections. *Annu. Rev. Phys. Chem.* **2011**, *62*, 621–643. https://doi.org/10.1146/annurev-physchem-032210-103450.
(49) Marian, C. M. Spin-Orbit Coupling and Intersystem Crossing in Molecules. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2* (2), 187–203. https://doi.org/10.1002/wcms.83.
(50) Yu, H. S.; Li, S. L.; Truhlar, D. G. Perspective: Kohn-Sham Density Functional Theory

Descending a Staircase. *J. Chem. Phys.* **2016**, *145* (13). https://doi.org/10.1063/1.4963168.

(51) Tuna, D.; Lefrancois, D.; Wolański, Ł.; Gozem, S.; Schapiro, I.; Andruniów, T.; Dreuw, A.; Olivucci, M. Assessment of Approximate Coupled-Cluster and Algebraic-Diagrammatic-Construction Methods for Ground- and Excited-State Reaction Paths and the Conical-Intersection Seam of a Retinal-Chromophore Model. *J. Chem. Theory Comput.* **2015**, *11* (12), 5758–5781. https://doi.org/10.1021/acs.jctc.5b00022.

(52) Blancafort, L.; Ogliaro, F.; Olivucci, M.; Robb, M. A.; Bearpark, M. J.; Sinicropi, A. Computational Investigation of Photochemical Reaction Mechanisms. *Mol. Supramol. Photochem.* **2005**, *13* (Computational Methods in Photochemistry), 31–110.

(53) Veryazov, V.; Malmqvist, P. Å.; Roos, B. O. How to Select Active Space for Multiconfigurational Quantum Chemistry? *Int. J. Quantum Chem.* **2011**, *111* (13), 3329–3338. https://doi.org/10.1002/qua.23068.

(54) Werner, H.-J. J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Sch??tz, M.; Schütz, M. Molpro: A General-Purpose Quantum Chemistry Program Package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2* (2), 242–253. https://doi.org/10.1002/wcms.82.

(55) Snyder, J. W.; Curchod, B. F. E.; Martínez, T. J. GPU-Accelerated State-Averaged Complete Active Space Self-Consistent Field Interfaced with Ab Initio Multiple Spawning Unravels the Photodynamics of Provitamin D $_3$. *J. Phys. Chem. Lett.* **2016**, 2444–2449. https://doi.org/10.1021/acs.jpclett.6b00970.

# **Chapter 2** Discovery of Conical Intersection Mediated Photochemistry with Growing String Methods

**This chapter is largely based upon published work:**

Aldaz, C., Kammeraad, J. A. & Zimmerman, P. M. *Phys. Chem. Chem. Phys.* **20,** 27394–27405 (2018).

## **2.1** Abstract

Conical intersections (CIs) are important features of photochemistry that determine yields and selectivity. Traditional CI optimizers require significant human effort and chemical intuition, which typically restricts searching to only a small region of the CI space. Herein, a systematic approach utilizing the growing string method is introduced to locate multiple CIs. Unintuitive MECI are found using driving coordinates that can be generated using a combinatorial search, and subsequent optimization allows reaction pathways, transition states, products, and seam-space pathways to be located. These capabilities are demonstrated by application to two prototypical photoisomerization reactions and the dimerization of butadiene. In total, many reaction pathways were uncovered, including the elusive stilbene hula-twist mechanism, and a previously unidentified product in butadiene dimerization. Overall, these results suggest that growing string methods provide a predictive strategy for exploring photochemistry.

## **2.2** Introduction

Photochemistries such as double-bond isomerization and electrocyclic reactions often pass through conical intersections (CI) on the pathways leading from initial excitation to the product state.[1–5] In these reactions, CIs act like funnels that enable ultrafast, nonradiative decay to lower-lying electronic states (Figure 2-1). This funneling effect causes the molecule to access nuclear configurations that would otherwise be thermally disallowed, which in turn allows photochemical reactions to reach unique outcomes compared to their ground state counterparts. In a sense, CIs can be envisioned as analogous to ground state transition states (TS), where passing through these regions correspond to important reactive events. Just like TSs, CIs are not single points, but whole regions of

space. These regions are called seams, and different sections of the seam can lead to qualitatively different photo products. When molecules are excited with energetic photons, the product distribution will be determined by the presence and accessibility of these seams.[6,7]

Minima on the seam, called minimum energy conical intersections (MECI), are important structures for describing photochemical reactions[8–18]. Conventional MECI and seam-space optimization algorithms, however, require extensive expert knowledge of photochemistry and CI geometries, which are far different from stable chemical structures. These techniques therefore do not easily permit the discovery of new types of photochemical reactions or simple ways to explore seam spaces. To explore complex photochemical reaction mechanisms and improve computation's ability to discover photoreactivity, new methods are required.

A few unconventional methods have proposed to enable excited-state reaction discovery. For example, this was attempted by Maeda and Morokuma with the anharmonic downward distortion following (ADDF) and the artificial-force-induced reaction (AFIR) methods, which have been used to locate MECI.[19–25] Uncovering complete photochemical reaction mechanisms—including MECI, the seam, excited-state TS, and photoproducts—is still a challenging task, however, leaving room for new methodologies.

Research in our laboratory has made extensive use of the growing string method (GSM) to determine thermal reaction pathways and transition states (TSs) without prior knowledge of the TS structure.[26–28] In GSM, qualitative reactive coordinates such as bond additions/breaking and/or angle changes—called driving coordinates—are used to locate TSs, products, and the reaction paths connecting them. When combined with the reaction discovery tool, ZStruct,[29–31] many TSs, some unexpected, can be found at once in a parallel computation. These features would be equally useful in the context of photochemistry where MECI optimization and seam minimum energy paths are challenging to optimize.

The GSM/ZStruct strategy is herein investigated for the first time for photochemistry, where reaction pathways include CIs and seams as well as TSs and products (Figure 2-1). This method includes the integration of GSM with CI optimizers, penalty function optimization, and CI topography analysis (see Theoretical Details). To show that complex photochemical reaction spaces can be explored, photoisomerizations of ethylene and stilbene, and dimerization of butadiene are examined.

Photoisomerization pathways are of interest for understanding molecular motors, molecular switches and retinal chromophores, but are notoriously challenging to study because the $S_1/S_0$ degeneracy persists along the torsional coordinate, leading to complex reaction mechanisms.[6] Butadiene dimerization was chosen because it can involve several modes of reaction ([2+2], [3+2], [4+2]), which, to our knowledge, has not been completely elucidated.[32,33] Overall, the results will indicate that GSM/ZStruct provides a thorough, robust and highly predictive method for identifying photochemical reaction mechanisms.

### **2.3** Background

The theory of conical intersection mediated photochemistry dates to the early 1930s, but was not widely accepted nor a practical theory until the development of high performance computing and multireference electronic structure theory.[34] Robb et al, pioneered this effort with the development of a general purpose MECI algorithm.[35] Locating MECIs has subsequently been performed using other optimizers such as the composed-step, and double-Newton-Raphson algorithms, which strive to reduce computational cost.[36–40] Although there are several variants of MECI solver, these optimizers all operate by minimizing the energy gap between states and total energy simultaneously and all require an initial starting structure that is geometrically near the MECI to be successful. This article will show that single-ended growing string method (SE-GSM) can be used to generate these structures and optimize to the MECI.

Beyond MECIs, the seam[8,9,11,12,14–17,41] and non-MECI points such as seam TSs (saddle points), minimum energy path CIs,[42] and minimum distance CIs[43] are also vital components of the photochemical reaction landscape. Methods for seam optimization include seam minimum energy paths (e.g from a seam TS),[36] constrained MECI optimization,[40,44,45] and the nudged elastic band (NEB) method which connects two MECIs.[46] GSM will also be shown capable of connecting two MECIs, and the single-ended GSM will additionally be useful for locating seams starting

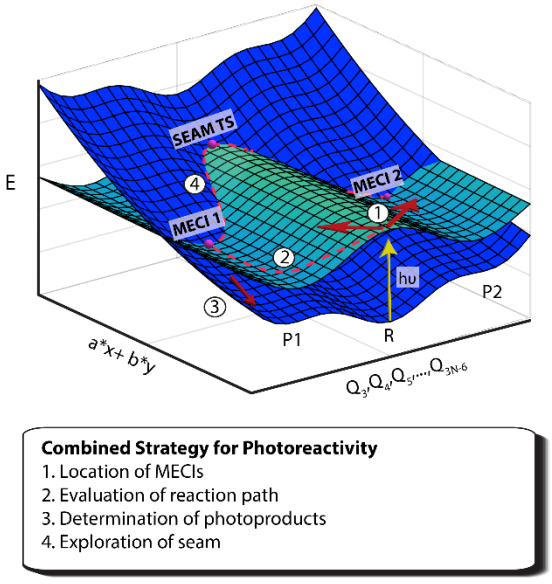



**Figure 2-1** Key points of interest for photochemistry involving crossing between potential energy surfaces.

from just one MECI. Seam mapping with GSM is expected to be advantageous due to its lower cost compared to NEB and standard string methods,[47,48] and its ability to be easily integrated with the SE-GSM method.

MECIs and seams give key transition regions between electronic states, and further analysis can then determine the photoproducts. First, the existence or absence of TS along the reaction path from the Frank Condon region to the CI shows whether the CI is accessible. Given that the CI is accessible, the topology near the CI influences the relaxation directions to the photoproducts.[49–51] One way to find these directions is to plot the energy around the CI within the branching plane (BP), which breaks the degeneracy to first order. The BP topography can be parametrized through the CI pitch, tilt, and asymmetry.[52] Ultimately, dynamical trajectories to and from the CI dictate photoproduct yields,[5,46,53–59] and several approaches have been used to approximate such trajectories.[60,61] Such simulations, however, are highly costly, leaving room for potential energy surface analysis to map out complicated photochemical reaction pathways.

When performing analysis of seams, CIs, and photochemical reaction paths, the results will be sensitive to the choice of electronic structure theory. Therefore, we note that the methods used here are compatible with any electronic structure theory that treats conical intersections, gradients, and derivative couplings. For methods without derivative coupling vectors programmed, these vectors can be approximated using the branching plane updating method,[62] or using a Davidson algorithm, recently developed in our lab, which requires only standard gradients.[63]

### **2.4** Theoretical Methods

The CI space is a central concept for the proposed methodologies, and is defined[64] in terms of the coordinate system

$$\boldsymbol{Q} = (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Q}_3, \boldsymbol{Q}_4, \dots, \boldsymbol{Q}_{3N-6}) \, .$$

**Equation 2-1** The CI space

The first subspace $(\boldsymbol{x}, \boldsymbol{y})$ is called branching plane (BP), where $\boldsymbol{x}$ and $\boldsymbol{y}$ are the difference gradient and derivative coupling vectors, respectively. $\boldsymbol{x}$ and $\boldsymbol{y}$ are orthogonal to all other coordinates, $\boldsymbol{Q}_i$. At a CI, motion along $\boldsymbol{x}$ and $\boldsymbol{y}$ changes the energy gap and produces the characteristic cone shape of a CI between two electronic states. The second subspace $(\boldsymbol{Q}_3 \dots \boldsymbol{Q}_{3N-6})$ is denoted the seam space (SS) and

corresponds to the 3N-8 internal coordinates which do not affect the energy gap to first order. Throughout this text, **bold** font signifies vectors and matrices.

### **2.4.1** Composed-Step Optimizer

MECI optimization reduces the SS gradient and energy gap between two electronic states to zero. Most MECI optimizers achieve this by starting with a good guess structure for the MECI, projecting out the BP contributions from the (Cartesian) gradient, and optimizing in the SS. On the other hand, one can directly construct the CI space in delocalized internal coordinates ($\boldsymbol{U}$) which are non-redundant vectors that fully span the 3N-6 degrees of freedom within a molecular system and have improved geometry optimization properties.[26,65,66]

To form the CI subspaces, $\boldsymbol{U}$ is formed using the standard procedure,[65] and $\boldsymbol{x}$ and $\boldsymbol{y}$ (originally Cartesian vectors) are projected into the basis of internal coordinates $(\boldsymbol{U}_x, \boldsymbol{U}_y)$, but are not orthonormal. Then, the Gram-Schmidt process is applied to the set $\{(\boldsymbol{U}_x, \boldsymbol{U}_y), (\boldsymbol{U}_k; k = 1 \dots 3N - 6)\}$ to produce the orthonormal vectors

$$\boldsymbol{V} = \{(\boldsymbol{U}_x, \boldsymbol{V}_y), (\boldsymbol{V}_k; k = 3{,}4{,}5 \dots 3N - 6)\}$$

**Equation 2-2** The CI space in delocalized internal coordinates

where the first two vectors represent the BP, orthonormalized with respect to $\boldsymbol{U}_x$, and the $\boldsymbol{V}_k$ vectors define the SS.

With these vectors, the composed-step optimizer (Figure 2-2)[36,38] consists of two component, the first

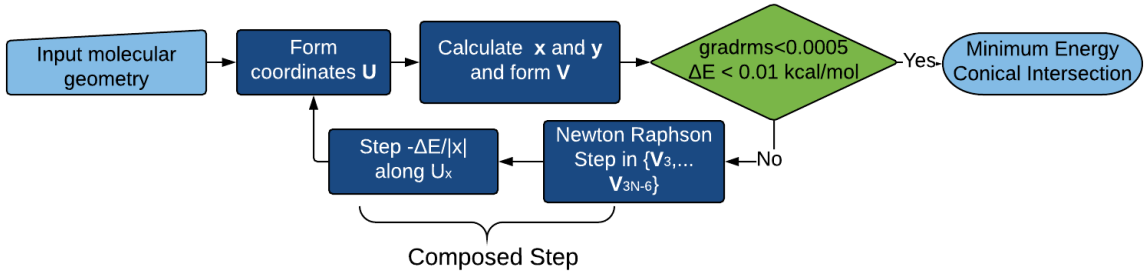



**Figure 2-2** Composed-step optimizer in delocalized internal coordinates. Each cycle of the optimizer takes a composed step until the gradient and energy gap are converged.

$$\Delta \boldsymbol{v}_x = -\frac{\Delta E}{|\boldsymbol{g}_x|}\boldsymbol{U}_x$$

**Equation 2-3** First composed step minimizing energy gap between potential energy surfaces

moves the geometry closer to the CI, where $\Delta E$ is the energy gap between the upper and lower surfaces and $|\boldsymbol{g}_x|$ is the magnitude of the difference gradient. The second component minimizes the total energy by an eigenvector optimization in the seam-space:

$$\Delta \tilde{\boldsymbol{v}}_{SS,i} = -\frac{\tilde{\boldsymbol{g}}_i}{\tilde{\boldsymbol{H}}_{ii}+\lambda}\tilde{\boldsymbol{V}}_{SS,i}$$

**Equation 2-4** Second and final composed step part minimizing energy along the seam

where $\tilde{\boldsymbol{V}}_{SS,i}$ and $\tilde{\boldsymbol{H}}_{ii}$ are the eigenvectors and eigenvalues of the seam-space Hessian, respectively, $\tilde{\boldsymbol{g}}_i$ is the average gradient, and $\lambda$ a shift factor. After back-transforming Equation 2-4 to the CI basis ($\boldsymbol{V}$), the optimization step is composed of Equation 2-3 and Equation 2-4:

$$\Delta \boldsymbol{v} = \Delta \boldsymbol{v}_x + \Delta \boldsymbol{v}_{SS}$$

**Equation 2-5** Composed step

These steps are taken until the MECI is reached.

### **2.4.2** Locating MECI with GSM

GSM develops a reaction path by iteratively adding and optimizing discrete structures, called nodes, along a specified reaction tangent,[26,28] and herein will be shown to be useful for locating MECIs starting far from the seam space. Figure 2-3 details the key steps of the GSM algorithm for finding MECI. The reaction tangent of GSM is defined using the standard delocalized internal coordinates ($\boldsymbol{U}$) as

$$\boldsymbol{U}_c = \sum_{k=1}^{3N-6} (\Delta \boldsymbol{q}^p | \boldsymbol{U}_k)\boldsymbol{U}_k$$

**Equation 2-6** GSM reaction coordinate

where $\Delta \boldsymbol{q}^p$ is the specified change in primitive internal coordinates ($\boldsymbol{q}$) describing the reaction. These vectors, called driving coordinates, allow GSM to explore in a specific direction (e.g. by the arrows of Figure 2-1 or Figure 2-3) using the coordinate system

$$\boldsymbol{V}^* = \{(\boldsymbol{U}_c), (\boldsymbol{V}_k^*; k = 2,3,4 \ldots 3N-6)\}$$

**Equation 2-7** Delocalized internal coordinates with reaction coordinate

where coordinates $\boldsymbol{V}_k^*$ are orthogonal to $\boldsymbol{U}_c$.

To locate a MECI, single-ended GSM moves along $\boldsymbol{U}_c$ and optimizes in all orthogonal directions on the energy surface[43] defined by

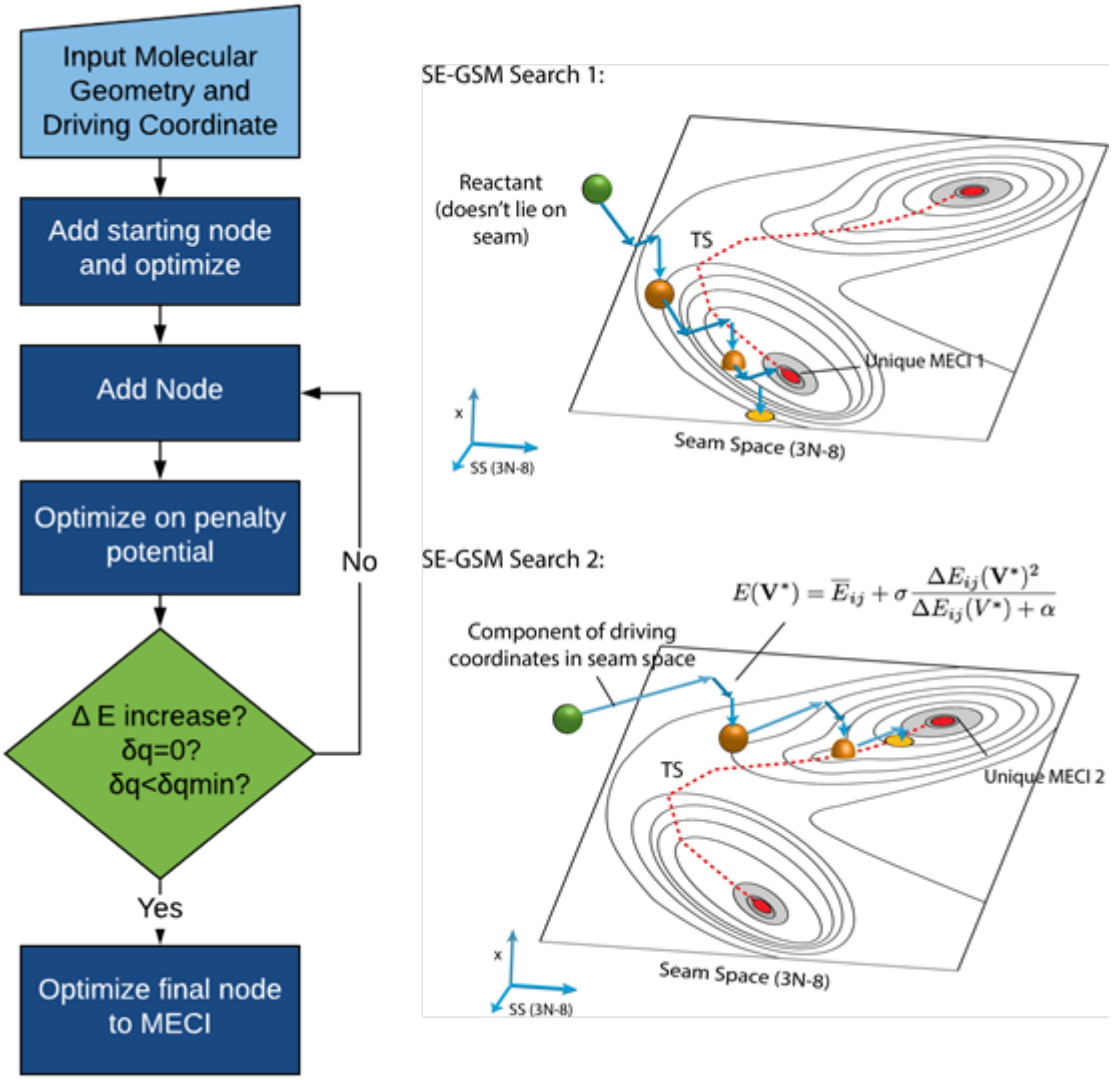



**Figure 2-3** Single-ended growing string method for locating minimum energy conical intersections.

$$E(\boldsymbol{V}^*) = \bar{E}_{ij}(\boldsymbol{V}^*) + \sigma \frac{\Delta E_{ij}(\boldsymbol{V}^*)^2}{\Delta E_{ij}(\boldsymbol{V}^*) + \alpha}$$

**Equation 2-8** Penalty potential energy for finding MECI

where $\bar{E}_{ij}$ is the average energy, $\sigma$ and $\alpha$ are penalty parameters, and $\Delta E_{ij}$ is the energy gap. This strategy therefore permits a simultaneous search and optimization to the nearest conical intersection in the search direction. Once the frontier node of GSM gets close to a CI, the composed step optimizer fully refines the MECI. The computational details section describes the criteria for switching between driving and optimization steps.

### **2.4.3** Seam Space GSM

The seams connecting MECIs will be mapped using GSM as well. This procedure combines the tangent vector (defined using Equation 2-9) with the two BP vectors in a coordinate system

$$\boldsymbol{V}' = \{(\boldsymbol{U}_c), (\boldsymbol{V}_x, \boldsymbol{V}_y), (\boldsymbol{V}_k; k = 4,5,6 \ldots 3N-6)\}.$$

**Equation 2-9** Delocalized internal coordinates with orthogonal reaction coordinate and branching space

As before, the Gram-Schmidt process is applied to produce a nonredundant, orthonormal set, $\boldsymbol{V}'$, with 3N-6 total degrees of freedom. All vectors are orthonormalized with respect to $\boldsymbol{U}_c$ to preserve

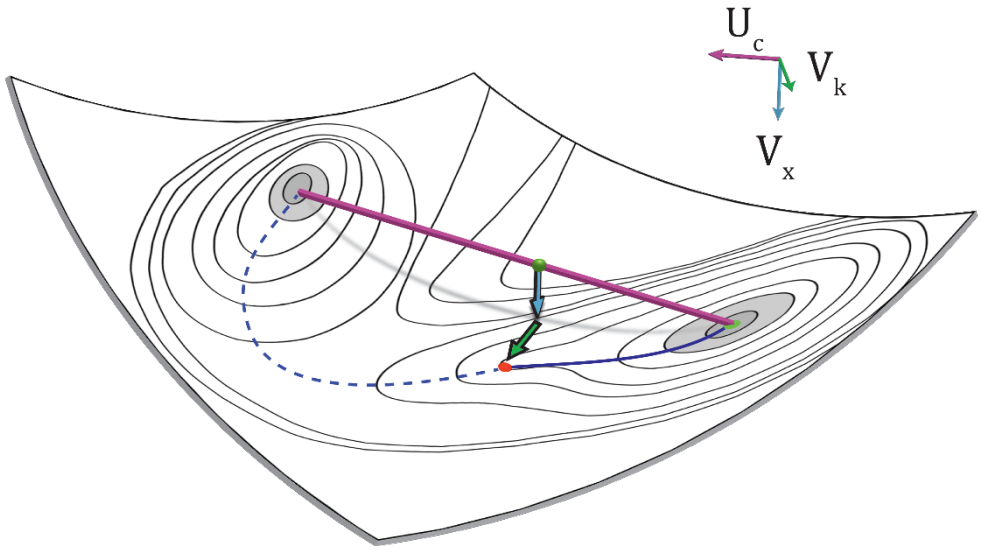



**Figure 2-4** Multidimensional representation that shows the operation of the growing string method seam mapping. Curved contour plot is the seam space (3N-8 dimensional) represented in the full molecular space (3N-6 dimensional). Minima on the contour plot are minimum energy conical intersections. Purple line is the tangent constraint vector ($\mathbf{U}_c$). A composed step along along ($\mathbf{V}_x$) and ($\mathbf{V}_k$; k=4,5,6,…,3N-6) is required to optimize the minimum energy seam path (dashed line).

the direction of $\boldsymbol{U_c}$. This means that $\boldsymbol{U_c}$ will typically overlap with the original BP until the path converges fully to the seam, and in practice this choice results in stable optimization. Node optimization is performed using the composed-step algorithm in the unconstrained part (3N-7 dimensions) of the coordinate space (Figure 2-4).

This coordinate system enables two modes of operation for seam optimization in GSM. First, double-ended GSM is available to connect pairs of MECIs, by growing inward and adding additional nodes until the string connects. Alternatively, a single-ended search starting from one MECI can explore via driving coordinates, permitting a search along a seam for a second MECI. After the full string is formed with either method, full optimization of the seam in the coordinate system of Equation 2-9 is performed, providing a minimum energy path along the seam (Figure 2-4).

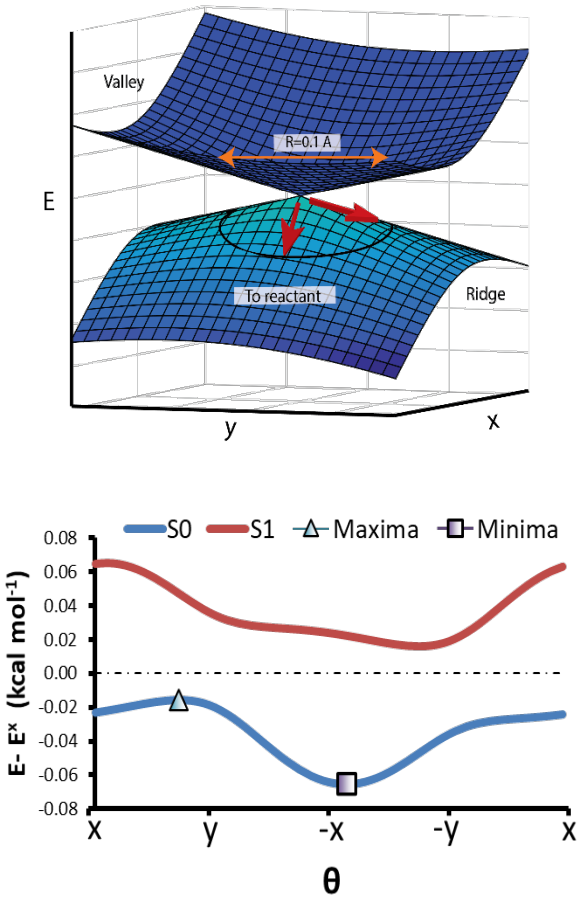



**Figure 2-5** Strategy for optimizing photoproducts from a conical intersection: take a small step along the minimum and maximum directions leading from the CI.

**2.4.4** Determining photoproducts

Photoproduct determination and optimization is initialized by taking small displacements (0.1 Å) in the BP, based on a cross-section analysis defined in polar coordinates as[52]

$$E(r,\theta) = E^{MECI} + \delta_{gh}\left(\sigma\cos(\theta-\theta_s) - \sqrt{1+\Delta_{gh}\cos 2\theta}\right)$$

**Equation 2-10** Energy around CI in branching space

where $\delta_{gh}$ is the pitch, $\theta_s$ is the tilt heading, and $\Delta_{gh}$ is the asymmetry. These parameters can be calculated given $\boldsymbol{x}, \boldsymbol{y}$ and the average energy gradient. Because the BP vectors can be interchanged at a CI through rotation of the electronic states, $\boldsymbol{x}, \boldsymbol{y}$ are chosen such that $\Delta_{gh}>0$ and $\theta_s \in [0, \frac{\pi}{2}]$.

Displacement in the direction of the minima and maxima of Equation 2-10 are chosen as starting pathways to photoproducts, see Figure 2-5. Optimization from the minima is analogous to the minimum energy pathways proceeding downhill from the CI, whereas optimization from the maxima in the cross-section are also investigated because they sometimes lead to minima not in the immediate vicinity of the CI cross-section.[49] Therefore, photoproduct optimization can be initialized given the BP vectors, and the sensitivity of the decay channels to the topography (e.g. minimum versus maximum) will be captured by sampling of both types of pathways.

## **2.5** Computational Details:

GSM methods are implemented in C++, and will be made available online on Github.[67] The Hessian is formed at each step of the optimization, from the Hessian in primitive coordinates

$$\boldsymbol{H} = \boldsymbol{U}^{T\prime}\boldsymbol{H}^{p}\boldsymbol{U}'$$

**Equation 2-11** Hessian in delocalized internal coordinates.

but only the non-constrained seam space coordinates are diagonalized and used in the Newton-Raphson optimization. The Broyden−Fletcher−Goldfarb−Shanno (BFGS) method is used to update a diagonal initial Hessian in primitive coordinates, which is transformed into a delocalized internal coordinate Hessian at each step.[68–71] The MECI is considered converged when the energy gap between the two states is less than 0.01 kcal mol$^{-1}$ and the seam-space root-mean-squared (RMS) average gradient is less than 0.0005 Ha/Å. For seam mapping the calculation terminates when all the nodes are converged to the string within a predefined threshold or the total gradient of the string is below a predefined threshold that is dependent on the system size and number of nodes. Each node is considered converged when the energy gap between the two states is less than 1.0 kcal mol$^{-1}$ and the gradient is converged to 0.0025 Ha/Å. The total gradient convergence criterion was chosen to be $(M-2)*0.0025*\sqrt{3N-6}$, where M is the number of GSM nodes.

Knowing where to stop the SE-GSM algorithm and initiate the composed step optimizer is an important criterion for the successful location of MECI. Since SE-GSM should be generally moving closer to the CI—not further away—string growth terminates when $\Delta E_{ij}$ increases at the frontier node. Alternatively, if the string reaches the "product" specified by $\Delta\mathbf{q}$ (Equation 2-7), this

indicates that no further growth of the string is possible, and the composed step optimizer is turned on immediately.

MECI SE-GSM calculations employed a maximum of 5 iterations per node, and a maximum growth step size (dqmax) of 0.4 Å-radians. Larger dqmax generally decreases the number of nodes in a calculation and therefore decreases the number of total gradients. In the small systems investigated here, however, a small dqmax is beneficial for MECI optimization because the final node is usually closer to the seam. The last node before MECI optimization is optimized for at most 60 steps, or until converged within 0.001 Ha/Å. The penalty function optimizer utilized σ as 1.0, for all nodes except for the final guess node, which is optimized using σ as 3.5. Staggering σ from small to large helps SE-GSM correctly find the desired MECI.

Seam mapping GSM calculations employed 3-5 optimization iterations per node per cycle, and dqmax values from 0.2 to 0.8 Å-radians. The optimization success is sensitive to these parameters and string growth should be monitored. In general, a smaller dqmax should be used for small molecules and short seams to ensure a refined reaction pathway.

The ZStruct method was used to generate driving coordinates consisting of add and break moves between reactive atoms. This is usually subject to the constraint that the driving coordinate can have up to two adds and one break, and the coordination of the atoms doesn't exceed a maximum and minimum coordination number.[30] In this work, however, add connection moves are used exclusively, leading to $N^4$ scaling in the number of generated driving coordinates, where N is the number of reactive atoms.

The methods invoke MOLPRO CASSCF to provide the quantum mechanical energies, gradients and derivative coupling vectors.[72] Active space selection was performed by considering the electrons involved in the excitation and reaction coordinate. Active space choices are listed in the figures below.

## **2.6** Results and Discussion

### **2.6.1** Exploratory via SE-GSM

In many photoreactions the qualitative reaction coordinates are known but not precisely quantified. For example, in the photoisomerization of ethylene the qualitative reaction coordinate is double bond torsion, but the CI also introduces pyramidalization of the carbon. SE-GSM captures this behavior by adding nodes along a driving coordinate (torsion) and optimizing in orthogonal directions (pyramidalization). Therefore SE-GSM identifies and follows directions to MECI that are not necessarily specified prior to the search. Importantly, different MECI within the same reactive system can be found by sampling a variety of driving coordinates. To show the capabilities of these methods, photoisomerization pathways for ethylene and stilbene are investigated as the initial test cases.

### **2.6.2** Ethylene Conical intersections

The well-known twisted-pyramidalized (also called H-bridging[73]) and ethylidene-like MECIs between $S_0$ and $S_1$ are reproduced here using SE-GSM (Figure 2-6).[74] While both MECIs were found by driving a torsional coordinate, the second case also used two driving coordinates, one of which specified H-atom transfer (**Table 2-1**).

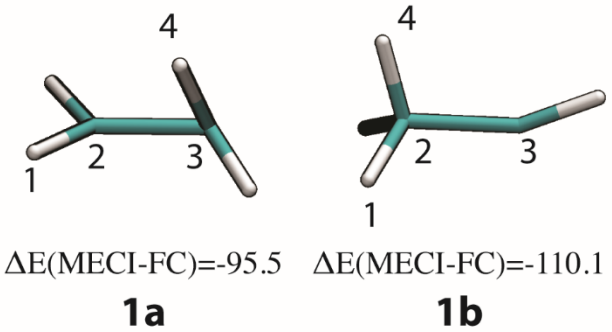



**Figure 2-6** Ethylene minimum energy conical intersections calculated using CAS(4,4)/6-31G* with SE-GSM. Energy is reported in kcal/mol with respect to Franck-Condon point. Paths to the MECI are barrierless.[60]. Numbers in figure are driving coordinate indices (see **Table 2-1**).

Performance metrics for SE-GSM with these two MECI are reported in **Table 2-1**. Driving cycles represent the computational costs of generating the MECI guess structures. Once this initial structure is found, the MECI optimization quickly reaches accurate MECIs because the RMS distance from the guess to the true MECI is small, under 0.07 Å. Compared to a standard MECI optimizer in Molpro,[35] the final geometries agree well (to within 0.01 Å RMS) and the energies to within 0.05 kcal/mol. These examples demonstrate that the SE-GSM process for MECI optimization provides reasonable results for ethylene, at computational costs of less than 100 gradients for each MECI.

**Table 2-1** Ethylene MECI Results from SE-GSM, compared to standard optimizers.

| MECI | Driving Coordinates[a] | Driving Cycles | MECI Cycles | Guess to MECI Distance (RMSD Å)[b] | Calc. MECI vs. Ref. MECI, (RMSD Å)[c] | Calc. E(MECI) – Ref. E(MECI) (kcal/mol)[c] |
|---|---|---|---|---|---|---|
| **1a** | TORSION(1,2,3,4) = 90 | 56 | 10 | 0.0611 | 0.0011 | 0.01 |
| **1b** | TORSION(1,2,3,4)=120, ADD(4,2), BREAK(3,4) | 55 | 39 | 0.0658 | 0.0076 | -0.03 |

[a] TORSION(1,2,3,4)=90, refers to a driving coordinate to push angle 1-2-3-4 towards 90 degrees, starting from its current position. [b] root-mean squared distance in Å from penalty optimized structure to composed-step (CS) optimized MECI. [c] root-mean squared distance in Å from composed-step (CS) optimized MECI to MECI from MOLPRO, see computational details.

Three seam reaction pathways and corresponding TSs were found for the ethylene system using the GSM seam optimization approach, and these are reported in Figure 2-7. Among these three, a seam TS (**2a**) between **1a** and **1b** has been previously reported and is reproduced here.[46] **2a** can be found using 7 or 9 GSM nodes and has energy relative to the FC point of -66.5 kcal/mol. A second, unexpected low energy seam TS was also found between **1a** and **1b,** which occurs along a roaming atom reaction pathway (**2b**).[21,75] Seam TS **2b** was found using 11 or 13 GSM nodes, and has energy relative to the FC point of -84.5 kcal/mol. Two seam TSs were located (using the same string endpoints) due to a difference in step size during GSM growth, with 7 nodes having a larger step size than 11 nodes. Interestingly, GSM can locate the two paths of this bifurcating potential energy surface, a "long" path and a "short" path, by just adjusting the step size during string growth. Lastly, a seam TS **2c** between tilt isomers of **1a** has been reported[46] and is reproduced here. This conical intersection is distinguished by a pyramidalized, but not tilted methylene group. Seam TS **2c** was found using 11 GSM nodes, and has an energy relative to the FC point of -84.1 kcal/mol.

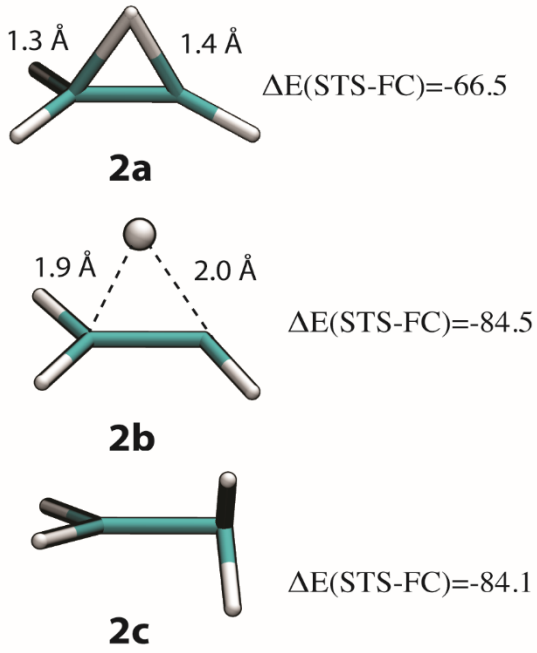



**Figure 2-7** Ethylene seam TS (STS) calculated using CAS(4,4)SCF/6-31G* with DE-GSM. Energies in kcal/mol with respect to the Franck-Condon point.

The tests on the ethylene system demonstrate that GSM has significant exploratory ability, providing not only the CI structures that are previously known, but an additional seam-space TS that has not been reported. The tests that follow will affirm these initial indications.

**2.6.3** Stilbene Conical Intersections

Compared to ethylene, stilbene has additional available photoreaction pathways due to the asymmetry around its central double bond. Stilbene is most stable in the *trans* conformation but can isomerize after absorption of a photon to *cis* via one-bond flip, *cis* via hula-twist, or ring-close to dihydrophenanthrene (DHP). The hula twist mechanism, which preserves the relative position of the phenyl groups has been hypothesized as favored for *cis-trans* isomerization in constrained media, as this mechanism reasonably preserves the molecular volume.[73,76,77]

For stilbene, *cis*-kinked (twisted pyramidalized[78]) and DHP-like MECI have been reported.[79] For related polyene systems, such as butadiene and hexatriene, additional CIs have been found such as *cis*-kinked diene, *trans*-kinked diene and H/vinyl (H-bridging) MECI, as well as seam saddle points hypothesized to be responsible for the hula-twist mechanism.[80] Analogies for all of these structures were found for stilbene using SE-GSM starting from a *trans* geometry. The driving coordinates and structures are shown **Table 2-2** and Figure 2-8, respectively.

In total, five MECI (**3a-3e**) and one seam TS (**3f**) were located using SE-GSM (Figure 2-8). Additional seam TSs for the tilt seam coordinate (between **3a** and **3c**) and hydrogen transfer coordinate (between **3c** and **3d**) were found using DE-GSM (see Supporting Information). **3a** and **3c** are *cis*-kinked MECI analogous to twisted-pyramidalized MECI **1a**, but differ from each other by the tilt of their phenyl ring with respect to the central double bond. **3b** is the trans-kinked MECI related to CIs in polyenes,[80] and was found by a one-bond flip of the central double bond. **3d** is analogous to the ethylidene-like conical intersection **1b** and leads to a reactive carbene intermediate or *cis*- and *trans*-stilbene. **3e** is the DHP-like MECI which can ring close to DHP. Lastly, **3f** is related to the volume conserving hula-twist CI reported by Houk et al for dienes,[80] and was found using SE-GSM by rotation of the phenyl ring in **3c**. **3a** and **3e** agree with structures reported in the literature.[78,79] To our knowledge **3b**-**3d** and **3f** have not been reported for stilbene, but they do resemble conical intersection geometries in dienes.[73,80]

**Table 2-2** Stilbene MECI results from SE-GSM.

| Minimum Energy Conical Intersection | Driving Coordinates | Driving Cycles | MECI cycles | φ(2,3,4,5) | φ(H3,2,3,H4) | φ(3,4,5,6) |
|---|---|---|---|---|---|---|
| **3a** | TORSION(2,3,4,5)=90, TORSION(H3,3,4,H4)=0 | 83 | 25 | 114.3 | 41.5 | -98.1 |
| **3b** | TORSION(2,3,4,5)=90 | 81 | 20 | 93.9 | 140.9 | 36.3 |
| **3c** | TORSION(2,3,4,5)=180, TORSION(H3,3,4 H4)=90 | 77 | 41 | 146.6 | 61.1 | -53.1 |
| **3d** | TORSION(2,3,4,5)=120, ADD(H4,3), BREAK(H4,4) | 95 | 18 | 84.5 | 114.4 | 179.4 |
| **3e** | TORSION(H3,3,4,H4)=0, TORSION(2,3,4,5)= 0, ADD(1,6) | 119 | 11 | 21.4 | 16.7 | 10.7 |
| **3f**[a] | TORSION(3,4,5,6)=50 | - | - | 137.9 | 71.1 | 47.9 |

[a]Seam transition state found from MECI **3c.**

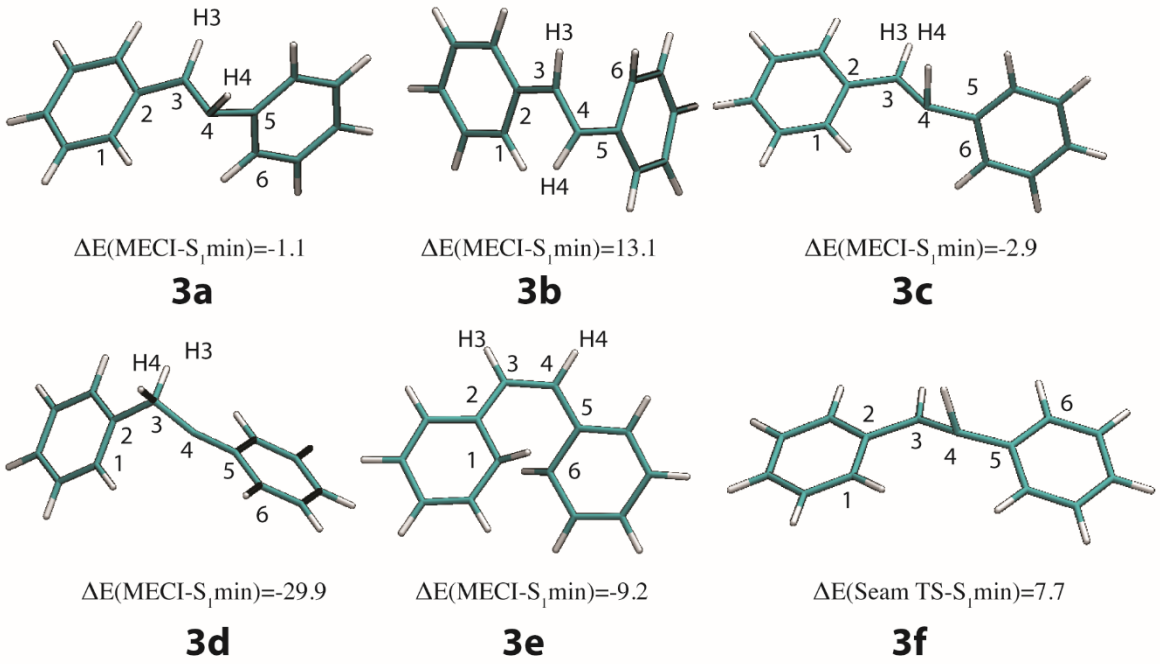



**Figure 2-8** Stilbene conical intersections calculated using CAS(2,2)SCF/6-31G with SE-GSM. Energies in kcal/mol with respect to the $S_1$ minimum. Atom numbers are driving coordinate indices (see **Table 2-2**).

Thus, a wide range of tilt and dihedral angles, and degree of hydrogen bond transfer are present in the MECI and seam TSs and these lead to different reaction possibilities. As can be seen in the CI structures, **3a**-**3c** flip the orientation of the phenyl groups (i.e. atoms 1 and 6 will land on the same side), whereas **3f** preserves the relative orientation of the phenyl groups with respect to each other.

Therefore, **3a-3c** are important for the one-bond flip mechanism, and **3f** is predicted to be important for the hula-twist mechanism.

Importantly, a comparison of the driving coordinates (which are qualitative) to the (quantitative) dihedral angles of the MECIs (Figure 2-8) shows that prior quantitative details of these MECI geometries are unnecessary for their optimization via SE-GSM. Instead, the driving coordinates take the Franck-Condon initial geometry, perturb it in the direction of each of these MECIs, and thus permit facile discovery of all 5 MECIs *from the same starting structure*. Together with the results from ethylene, the stilbene simulations demonstrate that GSM is effective at optimizing MECIs and seam-space reaction pathways. In the next section, a more complex photoreaction will be studied to show the full power of MECI exploration using GSM.

**2.6.4** Butadiene dimerization reaction discovery

Photochemical reactions with changes in covalent bonding are key challenges for simulation. In this space, reaction path discovery is possible with SE-GSM using driving coordinates that are systematically generated to sample many possible changes in bonding. To do this, the ZStruct reaction discovery method[29–31] is used to combinatorially sample search directions based on bond addition and breaking coordinates, where SE-GSM completes the MECI optimizations. ZStruct has seen extensive use in ground-state reactions,[81–87] and is applied here for the first time for photo-induced reactivity.

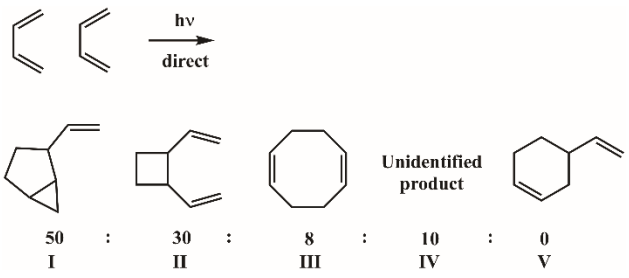


**Figure 2-9** Experimental photoproduct distributions from ref. [42].

As a challenging test case, the dimerization of butadiene is studied[32,33] as a complex photoreaction with many product channels (Figure 2-9). Under direct UV irradiation, butadiene has been observed to form 50% 2-vinylbicyclo[3.1.0]hexane and/or 3-vinylbicyclo[3.1.0]hexane (I), 30% 1,2-divinylcyclobutane (II), and 8% cycloocta-1,5-diene (III). A fourth species accounting for the remaining product (~10%) was unidentified. While 1,3-divinylcyclobutane was not mentioned in the original study, it is possible that this was not separated from 1,2-divinylcyclobutane. Interestingly, this product distribution is markedly different than seen in triplet sensitization and thermal reactions. For example, 4-vinylcyclohex-1-ene (V) is a major product in triplet-sensitized reactions of butadiene and

has also been observed in the thermal dimerization but does not appear in the UV chemistry. This diversity of products provides a great opportunity for reaction discovery simulations to provide details about the photoproduct generation mechanisms and corresponding CI space.

SE-GSM and ZStruct method were applied to a pair of *cis*-butadiene molecules, which were promoted to their lowest energy excited state. ZStruct generated 120 combinations of driving coordinates for the dimerization reaction, considering all carbon, but not hydrogen, as reactive. Of these combinations, 37 produced unique MECI based upon root-mean-squared distances and coordinate connectivity metrics.

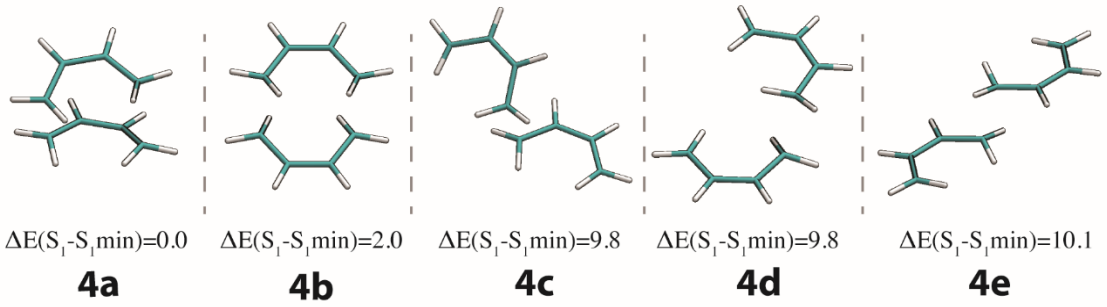



**Figure 2-10** Optimized geometries for $S_1$ minima located with CAS(8,8)SCF/6-31G* from aligned ZStruct structures. Energies in kcal/mol with respect to the global $S_1$ minimum, structure **4a**.

Because existence of a MECI does not guarantee its accessibility, multiple reaction paths to each MECI were considered, starting from the handful of local minima of the excited state surface near the Franck-Condon points. From the initial ZStruct-aligned structures, five excited-state minima were located (Figure 2-10). $S_1$ minima **4a** and **4b** are tetraradicaloid and the lowest in energy. **4b** is related to the so-called pericyclic minimum.[33] **4c**-**4e** minima have one radical from each molecule interacting with the other.

The reaction paths from the lower energy intermediates, **4a** and **4b**, typically result in higher activation energies to reach the MECIs, compared to **4c**-**4e**. A typical case is shown in Figure 2-11. From **4b** to the 5-coordinate MECI (**c-1** in supporting information), the reaction path has an activation barrier of 12.0 kcal/mol, which is in agreement with previous results calculated using CAS(4,4)SCF/4-31G.[33] The pathway, in contrast, is largely downhill from **4c**, with a small barrier of 7.8 kcal/mol. A similar effect is found in other pathways to reach the various MECI (see reaction pathways in SI).

From the accessible MECI, photoproducts were optimized from the minima and maxima of the MECI branching plane cross-sections. In total, 12 energetically accessible constitutional isomers and their diastereomers were identified (Figure 2-12). Additional photoproducts were also identified but are not accessible due to having high energy MECI or large excited state activation barriers with respect to the $S_1$ local minima.

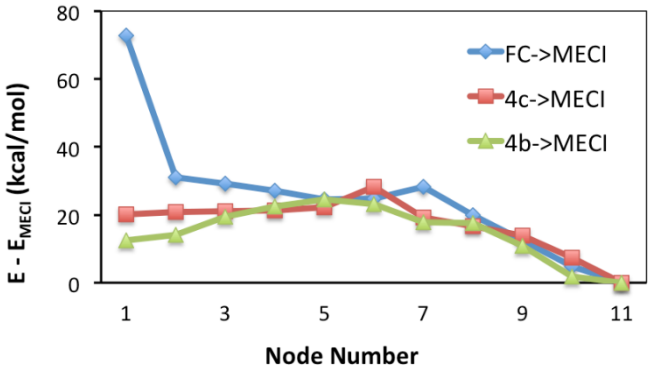



**Figure 2-11** Reaction path analysis reveals that excited-state barrier depends on the binding of the dimers before reaching the MECI. Reaction paths were computed using DE-GSM with CAS (8,8)SCF/6-31G*. The FC point is the aligned, but unoptimized planar monomers.

All experimentally observed photoproducts, (I), (II), and (III) have at least one low energy pathway, as well as additional, higher energy pathways arising from differences in $S_1$ minima, MECI or stereochemistry. Diastereomers of (I) form from a [3+2] cycloaddition, which upon ring-closing form an unstable methylene. The methylene carbon promptly ring-closes to form the resulting three-membered ring (see Figure 2-13). If the methylene and vinyl group are on the opposite sides of the 5-membered ring, then (1S,2R,5R)-(I) or (1S,2S,5R)-(1) forms. When these groups are on same side, (1R,3s,5S)-(I) or (1R,3r,5S)-(I) forms. 4-vinylcyclohex-1-ene (V) also forms from all MECI leading to the diastereomers of (I) and therefore is produced from the same reaction pathways. (V) forms, however, from different initial directions leading away from the MECI funnel. Diastereomers of (II) form from rhomboidal [2+2] cycloaddition MECI which agree with previously reported MECI.[33] The boat and chair conformations of cyclo-1,5-octadiene (III) form from [4+4] cycloaddition MECI close to the initial geometries of **4a** and **4b**. Bicyclo[5.1.0]oct-3-ene forms from the same MECI of (III) along different initial relaxation coordinates. Based on the results in Figure 2-12, the unidentified dimer (IV) from experiment is likely bicyclo[5.1.0]oct-3-ene, which shares the same conical intersection as (III), and has similar yield (~10 %).

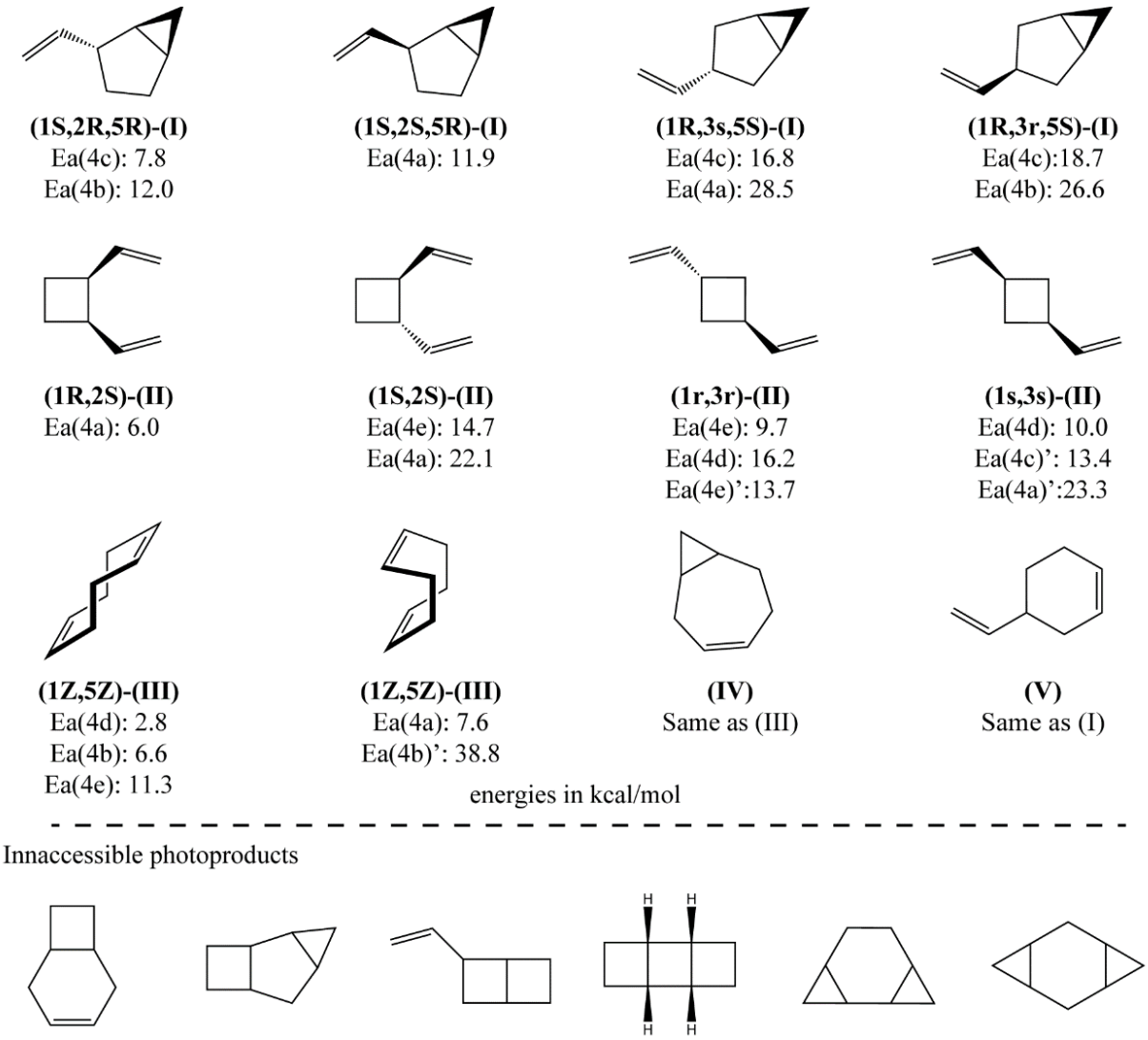



**Figure 2-12** Reaction products discovered using combinatorial reactive hypothesis generator ZStruct and single-ended GSM. Activation energies are in kcal/mol with respect to the labeled local minima. Ea' correspond to different MECI leading to the same product. Inaccessible products due to large reaction pathway barriers and high MECI energies.

These results are largely in agreement with available experimental results, but lead to an important question: why are there energetically accessible 4-vinylcyclohex-1-ene (V) pathways, but this species is not observed in experiment? A possible explanation comes from a reaction path to the MECI that gives (V) and (I)-(1R,3r,5S), which contains a nearby extended seam (Figure 2-13a). The first accessible CI along this seam is a minimum distance conical intersection (MDCI),[43] which would be the first CI approached along this path starting from the $S_1$ minimum.

The photoproducts of this (V)/(I) reaction path are shown in Figure 2-13b. (I)-(1R,3r,5S) forms from the maxima of the MECI cross-section, and (V) forms from the minima. In the latter case, the biradical recouples in a [4+2] arrangement. At the MDCI, which considerably differs in bond length **A** from the MECI, forms (I) or returns to the reactant structure without producing photoproduct. The MDCI can revert to reactants because carbon-carbon bond **A** has not yet been formed and therefore the geometry does not recouple to form (V). Thus, we hypothesize that decay at the MDCI, rather than the MECI, prevents the formation of (V).

In arene-alkene cycloaddition, the seam TS connecting [3+2] cycloaddition MECI has been hypothesized to be responsible for para ([4+2]) selectivity.[10] In butadiene dimerization, however, it is possible that the seam contains high energy points that prevent [4+2] dimerization (Figure 2-13b). Therefore, to provide additional insight**,** a SE-GSM seam calculation was performed (Figure 2-13c) from the MECI in Figure 2-13a using a breaking carbon-carbon bond **A** and a forming carbon-carbon bond **B** as driving coordinates. The SE-GSM seam method was necessary since the [3+2] MECI with carbon-carbon bond **A** was not found in the MECI search using ZStruct, and therefore no DE-GSM seam could be optimized. Notably, unlike the seam in arene-alkene cycloaddition that contains a low lying seam TS and MECI along this coordinate,[10] no such TS or MECI was located, which is consistent with the SE-GSM and ZStruct findings. Thus, the seam landscape is significantly

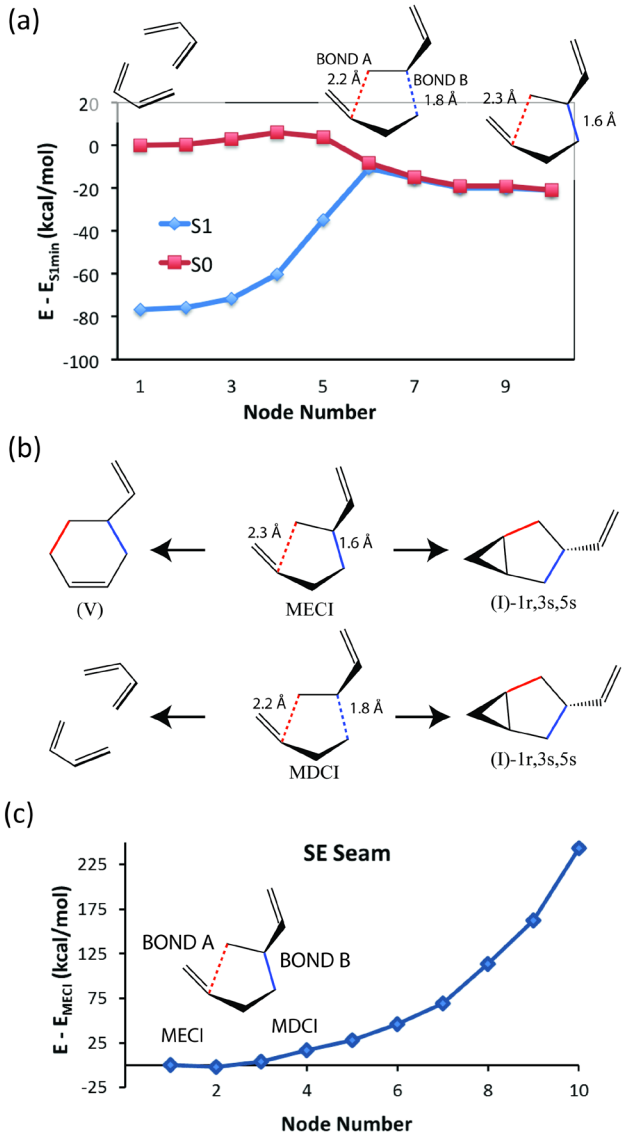



**Figure 2-13** (**a**) Reaction path from **4c** to MECI leading to (I)-1r,3s,5s and (V) products that shows ground and excited-state merging to form a seam and a minimum distance CI (MDCI) (**b**) photoproducts from the minimum and maximum of the CI branching plane cross-section for the MECI and MDCI (**c**) SE-GSM seam using breaking bond A and adding bond B driving coordinate.

different from the arene-alkene cycloaddition and may be the source of the difference in reactivity. Future studies including dynamics will be valuable to better investigate how the extended seam affects photoproduct distributions.

## **2.7** Conclusions

The growing string method is a powerful tool for reaction path optimization, and this tool has now been enabled to search for conical intersections and seam space reaction pathways. These methods provide a useful platform for studying photochemical reactions and have the important distinction of being useful even when little prior knowledge of the photochemical reaction pathways is available.

GSM located new excited state reaction pathways even in the well-understood ethylene and stilbene isomerization reactions. In addition, a combination of GSM with ZStruct was able to systematically explore reaction possibilities for the photodimerization of butadiene. The major, experimental photoproducts were found in the set of computed, energetically accessible products. The unidentified product in the experiment is likely bicyclo[5.1.0]oct-3-ene, since it shares the same conical intersection as cycloocta-1,5-diene and has similar yield. Interestingly, a non-stationary point on the seam, explored using GSM, is found to be important in the formation (or lack thereof) of 4-vinylcyclohex-1-ene. We anticipate that continued applications of the ZStruct/GSM combination will be useful to map CI spaces in emerging, complex photoreactions.

**Acknowledgements**

Support for this project was provided by the National Science Foundation through CHE-1551994.

## **2.8** References

(1) Klessinger, M. Theoretical Models for the Selectivity of Organic Singlet and Triplet Photoreactions. *Pure Appl. Chem.* **1997**, *69* (4), 773–778. https://doi.org/10.1351/pac199769040773.

(2) Robb, M. A.; Bernardi, F.; Olivucci, M. Conical Intersections as a Mechanistic Feature of Organic Photochemistry. *Pure Appl. Chem.* **1995**, *67* (5), 783–789. https://doi.org/10.1351/pac199567050783.

(3) Levine, B. G.; Martínez, T. J. Isomerization Through Conical Intersections. *Annu. Rev. Phys.*

*Chem.* **2007**, *58* (1), 613–634. https://doi.org/10.1146/annurev.physchem.57.032905.104612.

(4) Boggio-pasqua, M. *Computational Mechanistic Photochemistry : The Central Role of Conical Intersections*; Universite Toulouse III <tel-01184241>, 2015.

(5) Kazaryan, A.; Kistemaker, J. C. M.; Schafer, L. V; Browne, W. R.; Feringa, B. L.; Filatov, M. Understanding the Dynamics behind the Photoisomerization of a Light-Driven Fluorene Molecular Rotary Motor. *J. Phys. Chem. A* **2010**, *114* (15), 5058–5067. https://doi.org/10.1021/jp100609m.

(6) Schapiro, I.; Melaccio, F.; Laricheva, E. N.; Olivucci, M. Using the Computer to Understand the Chemistry of Conical Intersections. *Photochem. Photobiol. Sci.* **2011**, *10* (6), 867–886. https://doi.org/10.1039/c0pp00290a.

(7) Blancafort, L. Photochemistry and Photophysics at Extended Seams of Conical Intersection. *ChemPhysChem* **2014**, 3166–3181. https://doi.org/10.1002/cphc.201402359.

(8) Curchod, B. F. E.; Sisto, A.; Martínez, T. J. Ab Initio Multiple Spawning Photochemical Dynamics of DMABN Using GPUs. *J. Phys. Chem. A* **2017**, *121* (1), 265–276. https://doi.org/10.1021/acs.jpca.6b09962.

(9) Paterson, M. J.; Robb, M. a.; Blancafort, L.; DeBellis, A. D. Mechanism of an Exceptional Class of Photostabilizers: A Seam of Conical Intersection Parallel to Excited State Intramolecular Proton Transfer (ESIPT) in o-Hydroxyphenyl-(1,3,5)-Triazine. *J. Phys. Chem. A* **2005**, *109* (33), 7527–7537. https://doi.org/10.1021/jp051108+.

(10) Serrano-Pérez, J. J.; De Vleeschouwer, F.; De Proft, F.; Mendive-Tapia, D.; Bearpark, M. J.; Robb, M. a. How the Conical Intersection Seam Controls Chemical Selectivity in the Photocycloaddition of Ethylene and Benzene. *J. Org. Chem.* **2013**, *78* (5), 1874–1886. https://doi.org/10.1021/jo3017549.

(11) Reguero, M.; Boggio-pasqua, M.; Robb, M. a. Intramolecular Charge Transfer in 4-Aminobenzonitriles Does Not Necessarily Need the Twist. *J. Am. Chem. Soc.* **2005**, *127* (5), 315–318.

(12) Gómez, I.; Reguero, M.; Robb, M. A. Efficient Photochemical Merocyanine-to-Spiropyran Ring Closure Mechanism through an Extended Conical Intersection Seam. A Model CASSCF/CASPT2 Study. *J. Phys. Chem. A* **2006**, *110* (11), 3986–3991. https://doi.org/10.1021/jp056208u.

(13) Kobylecka, M.; Migani, A.; Asturiol, D.; Rak, J.; Blancafort, L. Benign Decay vs. Photolysis in the Photophysics and Photochemistry of 5-Bromouracil. A Computational Study. *J. Phys. Chem. A* **2009**, *113* (19), 5489–5495. https://doi.org/10.1021/jp811330v.

(14) Boggio-Pasqua, M.; Bearpark, M. J.; Hunt, P. A.; Robb, M. A. Dihydroazulene/Vinylheptafulvene Photochromism: A Model for One-Way Photochemistry via a Conical Intersection. *J. Am. Chem. Soc.* **2002**, *124* (7), 1456–1470. https://doi.org/10.1021/ja0161655.

(15) Boggio-Pasqua, M.; Ravaglia, M.; Bearpark, M. J.; Garavelli, M.; Robb, M. a. Can Diarylethene Photochromism Be Explained by a Reaction Path Alone? A CASSCF Study with Model MMVB Dynamics. *J. Phys. Chem. A* **2003**, *107* (50), 11139–11152. https://doi.org/10.1021/jp036862e.

(16) Migani, A.; Blancafort, L.; Robb, M. A.; DeBellis, A. D. An Extended Conical Intersection Seam Associated with a Manifold of Decay Paths: Excited-State Intramolecular Proton Transfer in O-Hydroxybenzaldehyde. *J. Am. Chem. Soc.* **2008**, *130* (22), 6932–6933. https://doi.org/10.1021/ja8013924.

(17) Garavelli, M.; Page, C. S.; Celani, P.; Olivucci, M.; Schmid, W. E.; Trushin, S. A.; Fuss, W. Reaction Path of a Sub-200 Fs Photochemical Electrocyclic Reaction. *J. Phys. Chem. A* **2001**, *105* (Ci), 4458. https://doi.org/10.1021/jp010359p.

(18) Coe, J. D.; Ong, M. T.; Levine, B. G.; Martínez, T. J. On the Extent and Connectivity of Conical Intersection Seams and the Effects of Three-State Intersections. *J. Phys. Chem. A* **2008**, *112* (49), 12559–12567. https://doi.org/10.1021/jp806072k.

(19) Maeda, S.; Ohno, K.; Morokuma, K. Exploring Multiple Potential Energy Surfaces: Photochemistry of Small Carbonyl Compounds. *Adv. Phys. Chem.* **2012**, *2012*, 1–13. https://doi.org/10.1155/2012/268124.

(20) Harabuchi, Y.; Maeda, S.; Taketsugu, T.; Minezawa, N.; Morokuma, K. Automated Search for Minimum Energy Conical Intersection Geometries between the Lowest Two Singlet States S0/S1-MECIs by the Spin-Flip TDDFT Method. *J. Chem. Theory Comput.* **2013**, *9* (9), 4116–4123. https://doi.org/10.1021/ct400512u.

(21) Maeda, S.; Taketsugu, T.; Ohno, K.; Morokuma, K. From Roaming Atoms to Hopping Surfaces: Mapping out Global Reaction Routes in Photochemistry. *J. Am. Chem. Soc.* **2015**, 150223131927005. https://doi.org/10.1021/ja512394y.

(22) Harabuchi, Y.; Taketsugu, T.; Maeda, S. Exploration of Minimum Energy Conical Intersection Structures of Small Polycyclic Aromatic Hydrocarbons: Toward an Understanding of the Size Dependence of Fluorescence Quantum Yields. *Phys. Chem. Chem. Phys.* **2015**, *17* (35), 22561–22565. https://doi.org/10.1039/c5cp02103k.

(23) Maeda, S.; Taketsugu, T.; Morokuma, K. Exploring Pathways of Photoaddition Reactions by Artificial Force Induced Reaction Method: A Case Study on the Paternò–Büchi Reaction. *Z. Phys. Chem.* **2013**, *227* (9–11), 1421–1433. https://doi.org/10.1524/zpch.2013.0401.

(24) Maeda, S.; Saito, R.; Morokuma, K. Finding Minimum Structures on the Seam of Crossing in Reactions of Type A + B → X: Exploration of Nonadiabatic Ignition Pathways of Unsaturated Hydrocarbons. *J. Phys. Chem. Lett.* **2011**, *2* (8), 852–857. https://doi.org/10.1021/jz200262m.

(25) Maeda, S.; Harabuchi, Y.; Taketsugu, T.; Morokuma, K. Systematic Exploration of Minimum Energy Conical Intersection Structures near the Franck–Condon Region. *J. Phys. Chem. A* **2014**, *118* (51), 12050–12058. https://doi.org/10.1021/jp507698m.

(26) Zimmerman, P. M. Growing String Method with Interpolation and Optimization in Internal Coordinates: Method and Examples. *J. Chem. Phys.* **2013**, *138* (18), 1–11. https://doi.org/10.1063/1.4804162.

(27) Zimmerman, P. Reliable Transition State Searches Integrated with the Growing String Method. *J. Chem. Theory Comput.* **2013**, *9* (7), 3043–3050. https://doi.org/10.1021/ct400319w.

(28) Zimmerman, P. M. Single-Ended Transition State Finding with the Growing String Method. *J. Comput. Chem.* **2015**, *36* (9), 601–611. https://doi.org/10.1002/jcc.23833.

(29) Dewyer, A. L.; Zimmerman, P. M. Finding Reaction Mechanisms, Intuitive or Otherwise. *Org. Biomol. Chem.* **2017**, 501–504. https://doi.org/10.1039/C6OB02183B.

(30) Zimmerman, P. M. Automated Discovery of Chemically Reasonable Elementary Reaction Steps. *J. Comput. Chem.* **2013**, *34* (16), 1385–1392. https://doi.org/10.1002/jcc.23271.

(31) Zimmerman, P. M. Navigating Molecular Space for Reaction Mechanisms: An Efficient, Automated Procedure. *Mol. Simul.* **2015**, *41* (1–3), 43–54. https://doi.org/10.1080/08927022.2014.894999.

(32) R. Srinivasan, F. I. S.; Eisenberglk, R. Energy Level of the First Excited Singlet State of 1,3-Butadiene. *J Am Chem Soc* **1965**, *1735* (15), 1964–1965. https://doi.org/10.1021/ja01094a047.

(33) Bearpark, M. J.; Deumal, M.; Robb, M. A.; Vreven, T.; Yamamoto, N.; Olivucci, M.; Bernardi, F. Modeling Photochemical [4 + 4] Cycloadditions: Conical Intersections Located with CASSCF for Butadiene + Butadiene. *J. Am. Chem. Soc.* **1997**, *119* (4), 709–718. https://doi.org/10.1021/ja962576n.

(34) Yarkony, D. R. Nonadiabatic Quantum Chemistry — Past , Present , and Future. *Chem. Rev.* **2012**, *112*, 481–498.

(35) Bearpark, M. J.; Robb, M. a.; Bernhard Schlegel, H. A Direct Method for the Location of the Lowest Energy Point on a Potential Surface Crossing. *Chem. Phys. Lett.* **1994**, *223* (3), 269–274. https://doi.org/10.1016/0009-2614(94)00433-1.

(36) Sicilia, F.; Blancafort, L.; Bearpark, M. J.; Robb, M. a. New Algorithms for Optimizing and Linking Conical Intersection Points. *J. Chem. Theory Comput.* **2008**, *4* (2), 257–266. https://doi.org/10.1021/ct7002435.

(37) Keal, T. W.; Koslowski, A.; Thiel, W. Comparison of Algorithms for Conical Intersection Optimisation Using Semiempirical Methods. *Theor. Chem. Acc.* **2007**, *118* (5–6), 837–844. https://doi.org/10.1007/s00214-007-0331-5.

(38) Ruiz-Barragan, S.; Robb, M. A.; Blancafort, L. Conical Intersection Optimization Based on a Double Newton-Raphson Algorithm Using Composed Steps. *J. Chem. Theory Comput.* **2013**, *9* (3), 1433–1442. https://doi.org/10.1021/ct301059t.

(39) Chachiyo, T.; Rodriguez, J. H. A Direct Method for Locating Minimum-Energy Crossing Points (MECPs) in Spin-Forbidden Transitions and Nonadiabatic Reactions. *J. Chem. Phys.* **2005**, *123* (9), 0–9. https://doi.org/10.1063/1.2007708.

(40) De Vico, L.; Olivucci, M.; Lindh, R. New General Tools for Constrained Geometry Optimizations. *J. Chem. Theory Comput.* **2005**, *1* (5), 1029–1037. https://doi.org/10.1021/ct0500949.

(41) Sicilia, F.; Blancafort, L.; Bearpark, M. J.; Robb, M. A. Quadratic Description of Conical Intersections: Characterization of Critical Points on the Extended Seam. *J. Phys. Chem. A* **2007**, *111* (11), 2182–2192. https://doi.org/10.1021/jp067614w.

(42) Cembran, A.; Bernardi, F.; Olivucci, M.; Garavelli, M. Counterion Controlled Photoisomerization of Retinal Chromophore Models: A Computational Investigation. *J. Am. Chem. Soc.* **2004**, *126* (49), 16018–16037. https://doi.org/10.1021/ja048782+.

(43) Levine, B. G.; Coe, J. D.; Martínez, T. J. Optimizing Conical Intersections without Derivative Coupling Vectors: Application to Multistate Multireference Second-Order Perturbation Theory (MS-CASPT2) †. *J. Phys. Chem. B* **2008**, *112* (2), 405–413. https://doi.org/10.1021/jp0761618.

(44) Passerone, D.; Laino, T. Exploring Conical Intersection Spaces Using Pseudo-Dynamics and Band Optimization: A Novel Strategy. *Comput. Phys. Commun.* **2005**, *169* (1–3), 305–308. https://doi.org/10.1016/j.cpc.2005.03.068.

(45) Laino, T.; Passerone, D. Pseudo-Dynamics and Band Optimizations : Shedding Light into Conical Intersection Seams. **2004**, *389*, 1–6. https://doi.org/10.1016/j.cplett.2004.02.110.

(46) Mori, T.; Martínez, T. J. Exploring the Conical Intersection Seam: The Seam Space Nudged Elastic Band Method. *J. Chem. Theory Comput.* **2013**, *9* (2), 1155–1163. https://doi.org/10.1021/ct300892t.

(47) Jafari, M.; Zimmerman, P. M. Reliable and Efficient Reaction Path and Transition State Finding for Surface Reactions with the Growing String Method. *J. Comput. Chem.* **2017**, 645–658. https://doi.org/10.1002/jcc.24720.

(48) Peters, B.; Heyden, A.; Bell, A. T.; Chakraborty, A. A Growing String Method for Determining

Transition States: Comparison to the Nudged Elastic Band and String Methods. *J. Chem. Phys.* **2004**, *120* (17), 7877–7886. https://doi.org/10.1063/1.1691018.

(49) Garavelli, M.; Celani, P.; Fato, M.; Bearpark, M. J.; Smith, B. R.; Olivucci, M.; Robb, M. a. Relaxation Paths from a Conical Intersection: The Mechanism of Product Formation in the Cyclohexadiene/Hexatriene Photochemical Interconversion. *J. Phys. Chem. A* **1997**, *101* (11), 2023–2032. https://doi.org/10.1021/jp961554k.

(50) Celani, P.; Robb, M. a.; Garavelli, M.; Bernardi, F.; Olivucci, M. Geometry Optimisation on a Hypersphere. Application to Finding Reaction Paths from a Conical Intersection. *Chem. Phys. Lett.* **1995**, *243* (1–2), 1–8. https://doi.org/10.1016/0009-2614(95)00821-K.

(51) Garavelli, M.; Bernardi, F.; Cembran, A.; Castaño, O.; Frutos, L. M.; Merchán, M.; Olivucci, M. Cyclooctatetraene Computational Photo- and Thermal Chemistry: A Reactivity Model for Conjugated Hydrocarbons. *J. Am. Chem. Soc.* **2002**, *124* (46), 13770–13789. https://doi.org/10.1021/ja020741v.

(52) Fdez. Galván, I.; Delcey, M. G.; Pedersen, T. B.; Aquilante, F.; Lindh, R. Analytical State-Average Complete-Active-Space Self-Consistent Field Nonadiabatic Coupling Vectors: Implementation with Density-Fitted Two-Electron Integrals and Application to Conical Intersections. *J. Chem. Theory Comput.* **2016**, *12* (8), 3636–3653. https://doi.org/10.1021/acs.jctc.6b00384.

(53) Garavelli, M.; Bernardi, F.; Cembran, A.; Castaño, O.; Frutos, L. M.; Merchán, M.; Olivucci, M. Cyclooctatetraene Computational Photo- and Thermal Chemistry: A Reactivity Model for Conjugated Hydrocarbons. *J. Am. Chem. Soc.* **2002**, *124* (46), 13770–13789. https://doi.org/10.1021/ja020741v.

(54) Mignolet, B.; Curchod, B. F. E.; Martínez, T. J. Rich Athermal Ground-State Chemistry Triggered by Dynamics through a Conical Intersection. *Angew. Chemie Int. Ed.* **2016**, *55* (48), 14993–14996. https://doi.org/10.1002/anie.201607633.

(55) Althorpe, S.; Althorpe, S. C.; Worth, G. A. *Quantum Dynamics at Conical Intersections*; 2004.

(56) Quenneville, J.; Ben-Nun, M.; Martínez, T. J. Photochemistry from First Principles - Advances and Future Prospects. *J. Photochem. Photobiol. a-Chemistry* **2001**, *144* (2–3), 229–235. https://doi.org/http://dx.doi.org/10.1016/S1010-6030(01)00452-X.

(57) Tomasello, G.; Ogliaro, F.; Bearpark, M. J.; Robb, M. A.; Garavelli, M. Modeling the Photophysics and Photochromic Potential of 1,2-Dihydronaphthalene (DHN): A Combined CASPT2//CASSCF-Topological and MMVB-Dynamical Investigation. *J. Phys. Chem. A* **2008**, *112* (41), 10096–10107. https://doi.org/10.1021/jp802613m.

(58) Malhado, J. P.; Hynes, J. T. Non-Adiabatic Transition Probability Dependence on Conical Intersection Topography. *J. Chem. Phys.* **2016**, *145* (19). https://doi.org/10.1063/1.4967259.

(59) Michael S. Schuurman, A. S. Dynamics at Conical Intersections. *Annu. Rev. Phys. Chem.* **2018**, *69*, 427–450. https://doi.org/https://doi.org/10.1146/annurev-physchem- 052516-050721.

(60) Quenneville, J.; Marti, T. J.; Ben-Nun, M.; Quenneville, J.; Martínez, T. J.; Marti, T. J. Ab Initio Multiple Spawning: Photochemistry from First Principles Quantum Molecular Dynamics. *J. Phys. Chem. A* **2000**, *104* (22), 5161–5175. https://doi.org/10.1021/jp994174i.

(61) Guo, H.; Yarkony, D. R. Accurate Nonadiabatic Dynamics. *Phys. Chem. Chem. Phys.* **2016**, *18*, 26335–26352. https://doi.org/10.1039/C6CP05553B.

(62) Maeda, S.; Ohno, K.; Morokuma, K. Updated Branching Plane for Finding Conical Intersections without Coupling Derivative Vectors. *J. Chem. Theory Comput.* **2010**, *6* (5), 1538–1545. https://doi.org/10.1021/ct1000268.

(63) Kammeraad, J. A.; Zimmerman, P. M. Estimating the Derivative Coupling Vector Using Gradients. *J. Phys. Chem. Lett.* **2016**, 5074–5079. https://doi.org/10.1021/acs.jpclett.6b02501.
(64) Worth, G. a; Cederbaum, L. S. Beyond Born-Oppenheimer: Molecular Dynamics through a Conical Intersection. *Annu. Rev. Phys. Chem.* **2004**, *55*, 127–158. https://doi.org/10.1146/annurev.physchem.55.091602.094335.
(65) Baker, J.; Kessi, A.; Delley, B. The Generation and Use of Delocalized Internal Coordinates in Geometry Optimization. *J. Chem. Phys.* **1996**, *105* (1996), 192–212. https://doi.org/10.1063/1.471864.
(66) Baker, J. Constrained Optimization in Delocalized Internal Coordinates. *J. Comput. Chem.* **1997**, *18* (8), 1079–1095. https://doi.org/10.1002/(SICI)1096-987X(199706)18:8<1079::AID-JCC12>3.0.CO;2-8.
(67) Zimmerman, P. molecularGSM https://github.com/ZimmermanGroup/molecularGSM.
(68) Broyden, C. G. The Convergence of a Class of Double-Rank Minimization Algorithms 1. General Considerations. *IMA J. Appl. Math.* **1970**, *6* (1), 76–90. https://doi.org/10.1093/imamat/6.1.76.
(69) Fletcher, R. A New Approach to Variable Metric Algorithms. *Comput. J.* **1970**, *13* (3), 317–322. https://doi.org/10.1093/comjnl/13.3.317.
(70) Goldfarb, D. A Family of Variable-Metric Methods Derived by Variational Means. *Math. Comput.* **1970**, *24* (109), 23–23. https://doi.org/10.1090/S0025-5718-1970-0258249-6.
(71) Shanno, D. F. Conditioning of Quasi-Newton Methods for Function Minimization. *Math. Comput.* **1970**, *24* (111), 647–647. https://doi.org/10.1090/S0025-5718-1970-0274029-X.
(72) Werner, H.-J. J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Sch??tz, M.; Schütz, M. Molpro: A General-Purpose Quantum Chemistry Program Package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2* (2), 242–253. https://doi.org/10.1002/wcms.82.
(73) Wilsey, S.; Houk, K. N. H–Vinyl Conical Intersections for Dienes: A Mechanism for the Photochemical Hula Twist¶. *Photochem. Photobiol.* **2002**, *76* (6), 616. https://doi.org/10.1562/0031-8655(2002)076<0616:HVCIFD>2.0.CO;2.
(74) Ben-Nun, M.; Martínez, T. J. Photodynamics of Ethylene: Ab Initio Studies of Conical Intersections. *Chem. Phys.* **2000**, *259* (2–3), 237–248. https://doi.org/10.1016/S0301-0104(00)00194-4.
(75) Jordan, M. J. T.; Kable, S. H. Roaming Reaction Pathways Along Excited States. *Science (80-. ).* **2012**, *335* (6072), 1054–1055. https://doi.org/10.1126/science.1218767.
(76) Liu, R. S. H. Photoisomerization by Hula-Twist: A Fundamental Supramolecular Photochemical Reaction. *Acc. Chem. Res.* **2001**, *34* (7), 555–562. https://doi.org/10.1021/ar000165c.
(77) Fuß, W.; Kosmidis, C.; Schmid, W. E.; Trushin, S. A. The Photochemical Cis-Trans Isomerization of Free Stilbene Molecules Follows a Hula-Twist Pathway. *Angew. Chemie - Int. Ed.* **2004**, *43* (32), 4178–4182. https://doi.org/10.1002/anie.200454221.
(78) Quenneville, J.; Martı, T. J. Ab Initio Study of Cis - Trans Photoisomerization in Stilbene and Ethylene. *J. Phys. Chem. A* **2003**, *107*, 829–837.
(79) Ioffe, I. N.; Granovsky, A. A. Photoisomerization of Stilbene: The Detailed XMCQDPT2 Treatment. *J. Chem. Theory Comput.* **2013**, *9* (11), 4973–4990. https://doi.org/10.1021/ct400647w.
(80) Norton, J. E.; Houk, K. N. H/Vinyl Conical Intersections of Hexatrienes Related to the Hula-Twist Photoisomerization. *Mol. Phys.* **2006**, *104* (5–7), 993–1008.

https://doi.org/10.1080/00268970500417606.
(81) Pendleton, I. M.; Pérez-Temprano, M. H.; Sanford, M. S.; Zimmerman, P. M. Experimental and Computational Assessment of Reactivity and Mechanism in C(Sp 3 )–N Bond-Forming Reductive Elimination from Palladium(IV). *J. Am. Chem. Soc.* **2016**, *138* (18), 6049–6060. https://doi.org/10.1021/jacs.6b02714.
(82) Nett, A. J.; Zhao, W.; Zimmerman, P. M.; Montgomery, J. Highly Active Nickel Catalysts for C–H Functionalization Identified through Analysis of Off-Cycle Intermediates. *J. Am. Chem. Soc.* **2015**, 150612094813002. https://doi.org/10.1021/jacs.5b04548.
(83) Nett, A. J.; Montgomery, J.; Zimmerman, P. M. Entrances, Traps, and Rate-Controlling Factors for Nickel-Catalyzed C − H Functionalization. **2017**. https://doi.org/10.1021/acscatal.7b02919.
(84) Dewyer, A. L.; Zimmerman, P. M. Simulated Mechanism for Palladium-Catalyzed, Directed γ - Arylation of Piperidine. **2017**. https://doi.org/10.1021/acscatal.7b01390.
(85) Khomutnyk, Y. Y.; Arg??elles, A. J.; Winschel, G. A.; Sun, Z.; Zimmerman, P. M.; Nagorny, P. Studies of the Mechanism and Origins of Enantioselectivity for the Chiral Phosphoric Acid-Catalyzed Stereoselective Spiroketalization Reactions. *J. Am. Chem. Soc.* **2016**, *138* (1), 444–456. https://doi.org/10.1021/jacs.5b12528.
(86) Tay, J. H.; Argüelles, A. J.; Demars, M. D.; Zimmerman, P. M.; Sherman, D. H.; Nagorny, P. Regiodivergent Glycosylations of 6-Deoxy-Erythronolide B and Oleandomycin-Derived Macrolactones Enabled by Chiral Acid Catalysis. *J. Am. Chem. Soc.* **2017**, *139* (25), 8570–8578. https://doi.org/10.1021/jacs.7b03198.
(87) Smith, M. L.; Leone, A. K.; Zimmerman, P. M.; McNeil, A. J. Impact of Preferential Π-Binding in Catalyst-Transfer Polycondensation of Thiazole Derivatives. *ACS Macro Lett.* **2016**, 1411–1415. https://doi.org/10.1021/acsmacrolett.6b00886.

# **Chapter 3** A Linear Scaling Growing String Method with Correlated Motions

## **3.1** Introduction

The computation of minimum energy paths (MEPs) and transition states (TSs) is an important task for the investigation of reaction mechanisms. The location and characterization of TSs provides information of a reaction including the rate of the elementary steps (i.e. kinetics) and steric and electronic parameters. However, the location of TSs is a non-trivial task because it is a first-order saddle point on the potential energy surface. Therefore, geometry optimization of TSs can easily revert to the more stable equilibrium geometries in the direction along the MEP if given a bad initial starting point.

Consequently, there has been much work on the development of methods for the location of TSs such as the nudged-elastic band (NEB)[1,2], string method (SM)[3] and growing string methods (GSM)[4,5] which endeavor to optimize TSs in a more reliable way. These methods optimize a chain-of-states along the MEP, instead of a single point toward the transition structure, and therefore tend to produce more reliable TS optimization in the direction of the MEP in comparison to the standard guess-and-check optimization approach. Chain-of-states methods, however, can break down in several ways. NEB and SM requires the complete initial path to be present at the onset. However, creating a good initial pathway is a significant challenge for interpolation methods because the MEP is a curved pathway in a highly dimensional space, therefore NEB and SM can cause significant problems if the initial pathway is of poor quality[6,7].

GSM improves upon NEB and the SM[8–10] by growing the path, i.e. string, from the outside-in, such that new nodes are only added after the frontier nodes are sufficiently optimized which avoids placing new nodes into high energy regions of space. The interpolation can also be improved by using

methods such as linear synchronous transit[11], or interpolation in delocalized internal coordinates (DLC), which are combinations of bonds, angles, and torsions[9,12]. The latter is particularly efficient because DLC are a curvilinear coordinate system which means they have improved interpolation and optimization for angular and rotational motion[13].GSM with DLC have been heavily utilized[4,5,9,14] and will be investigated further herein. A drawback of GSM with DLC, however, is that it is not directly usable in environments such as crystals, proteins, metal surfaces and other condensed phases because DLCs become numerically unstable or too computationally intensive. For example, GSM using DLCs will cause error in the study of metal surface chemistry, where the atoms of the surface are packed tightly, because the concept of bonds, angles and torsions is not meaningful for metals. However, changing to mixed delocalized internal (mDLC) coordinate system where the metal atoms are treated with Cartesian coordinates and the adsorbate is treated with internal coordinates, enables GSM to investigate metal-surface chemistry[10]. Changing the coordinate system for geometry optimization is also known to be an effective way to improve geometry optimization performance. For example, DLCs were shown to decrease the number of geometry optimization steps in comparison to Cartesian coordinates because DLCs are less coupled than Cartesian coordinates[13]. There have also been studies that employ primitive internal coordinates such as explicit translations and rotational internal coordinates (TRIC)[15] and lattice parameters[16], which can improve the quality of geometry optimization for different types of chemistry.

Despite the known benefits of modifying coordinate systems which can help extend GSM to more complex systems, only DLCs and mDLC were previously available in GSM[9,10]. Furthermore, these approaches do not address the challenging task of coordinate transformation, which scales as the number of internal coordinates cubed, and becomes prohibitive in many-atom simulations. Thiel et al, developed a hybrid delocalized coordinate system (HDLC)[17] which combines internal coordinates and Cartesian coordinates (this is similar to mDLC except every atom has Cartesian coordinates) and performs linear scaling coordinate transformation, but this has not been applied in GSM until now.

In this letter, a GSM algorithm is presented which employs efficient block-matrix linear algebra to achieve linear scaling in the number of molecules or fragments. This new algorithm is necessary to study large multi-molecule systems like crystals and is highly customizable for different types of reactions (e.g. gas phase, or solvent phase); it can use DLC, HDLC, TRIC, or mDLC. To demonstrate

the new methods, we show how GSM can capture the long-range correlated motions of a crystalline molecular gyroscope.

## **3.2** Theoretical Details

### **3.2.1** GSM Background

To begin, the growing string method (GSM) computes minimum energy paths connecting intermediates and transition state geometries[9,18,19]. Until now, GSM specialized in molecular systems (e.g. less than 300 atoms), and substantial changes were needed to handle larger, condensed phase systems (greater than 1,000 atoms). Therefore, we briefly review the approach before describing the advances in GSM for the condensed phase.

GSM works in the basis of delocalized internal coordinates (DLC)[13]. The coordinate basis ($\boldsymbol{U}$) is modified to isolate a reaction coordinate $U_c$ by applying the Gram-Schmidt procedure to obtain a set of nonredundant, orthogonal vectors that span the possible atomic motions,

$$\boldsymbol{V} = \{U_c; V_k \; k = 2,3, \cdots 3N-6\}$$

**Equation 3-1 Delocalized internal coordinates with reaction coordinate**

where $V_k$ are unrestrained motions. Iterative growth and optimization within this coordinate system is used to find reaction paths and transition states. The original coordinate vectors are represented by $U_k$, the modified coordinate vectors are represented by $V_k$. The coordinate $U_c$ represents the local tangent of the reaction path, and is obtained by projection of a primitive internal coordinate differences $\Delta q^p$ into the space spanned by $\boldsymbol{U}$

$$U_c = \alpha \sum_{k=1}^{3N} \langle \Delta q^p | U_k \rangle U_k$$

**Equation 3-2 GSM reaction coordinate**

Here $\alpha$ is a normalization constant, and $\Delta q^p$ can be computed between two geometries (e.g. reactant and product) or defined from a single structure. The former is called double-ended GSM, which iteratively grows and optimizes the reaction path between two structures. The latter is called single-

ended GSM, which can find reaction paths, transition states and products given a user defined reaction coordinate. For more details see reference [5,9] and [18].

**3.2.2** Delocalized Internal Coordinates

A major difficulty with the above approach, however, is the formation of the DLCs $\boldsymbol{U}$, which is formed by diagonalization of the $\boldsymbol{G}$ matrix, $\boldsymbol{G} = \boldsymbol{B}\boldsymbol{B}^T$, where $B_{ij} = dq_i/dx_j$ is the Wilson B-Matrix. This step scales as the number of internal coordinates cubed and must be performed multiple times because back-transformation to Cartesian coordinates requires iterative inversion of $\boldsymbol{G}$. For these reasons, the formation and utilization of DLC is prohibitive when too many atoms are present. Furthermore, the quality of optimization with DLC declines for intermolecular systems because a fictitious bond must be added to couple the fragments into one supramolecular structure, Figure 3-1. Rotations are not well defined with respect to a single axis of rotation and the Hessian elements of the intermolecular bond are too strong because they are meant to represent a true bond not a fictitious one. The latter problem can be partially alleviated by treating the intermolecular bond with the long-range portion of the Lennard-Jones potential, $r^{-6}$, which can give improvement of intermolecular optimization[20]. Overall, however, while DLCs are great for intramolecular optimization they are not well suited for intermolecular optimization.

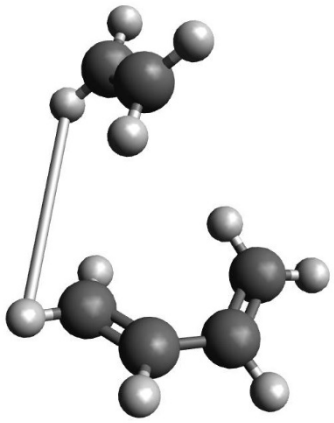


**Figure 3-1** Delocalized internal coordinates require a minimum spanning bond to couple intermolecular fragments.

Another type of coordinate system, known as hybrid delocalized internal coordinates (HDLC), add Cartesian coordinates to the set of internal coordinates and has been developed to reduce the computational cost of coordinate transformation[17]. HDLC negates the need for a minimum spanning bond which couple the fragments because the full 3N degrees of freedom including the rotations and translations are described by Cartesian coordinates. Consequently, linear scaling formation can be achieved if the Wilson B-matrix is organized into blocks corresponding to each non-bonded fragment and only the non-zero elements stored and used in computations such as matrix multiplication, and matrix diagonalization[17].

Other variant of DLC have also been developed including adding explicit translation and rotation internal coordinates (TRIC).[15] TRIC shows improved performance for intermolecular optimization in comparison to DLC, Cartesian coordinates, and HDLC. Furthermore, TRIC do not require a minimum spanning bond and therefore block-matrix techniques can be applied. There are also other variants of DLC which include the explicit dependence of the lattice parameters on the positions of all unit cell atoms, which can be helpful for optimizing crystals, but these internals are not used here[16]. Finally, a subset of atoms can be treated with Cartesian coordinates, and the rest treated with TRIC, HDLC or DLC. We refer to this as mixed DLC (mDLC). Using different

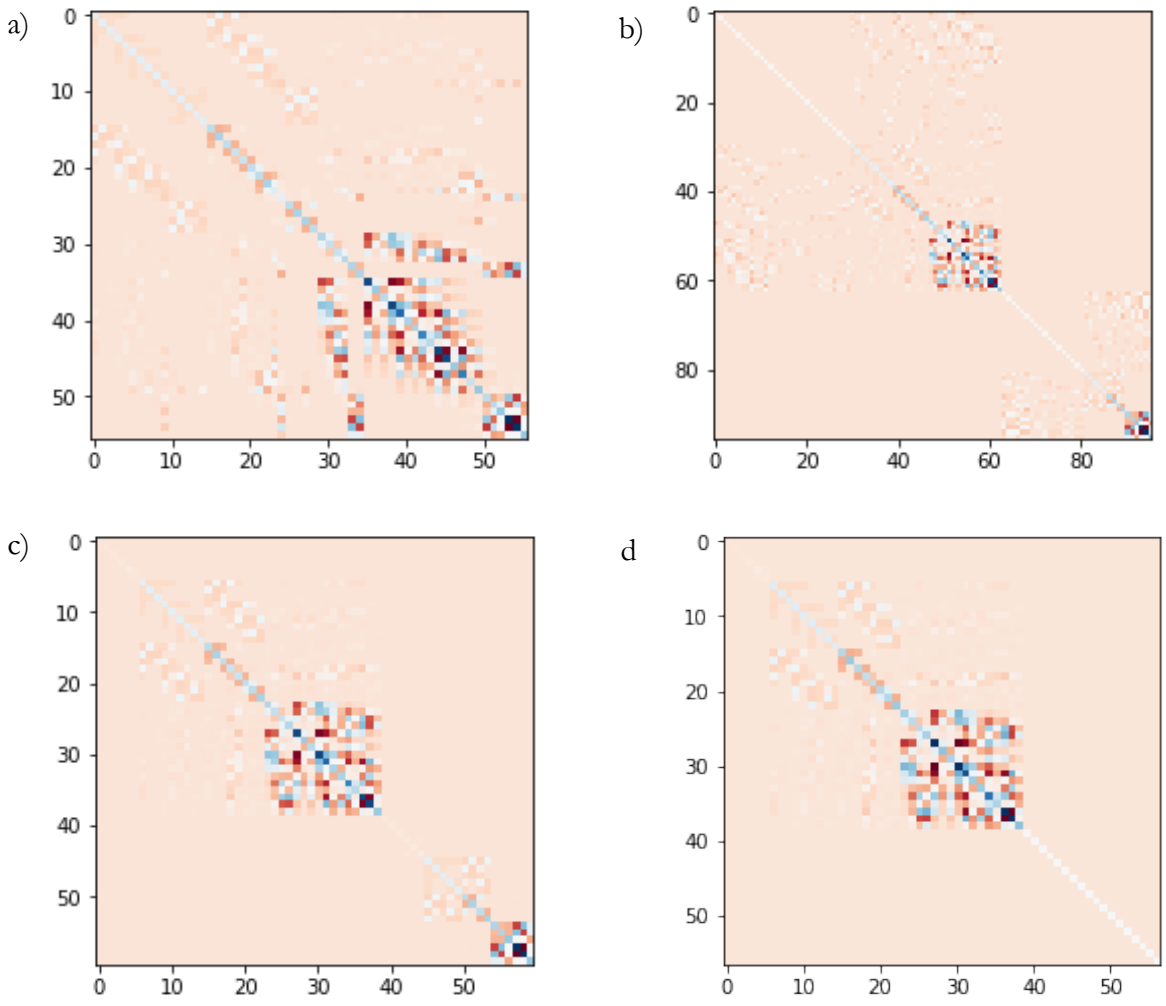



**Figure 3-2** Comparison of G-Matrices, $\mathbf{G} = \mathbf{BB^T}$, where $\mathbf{B_{ij}} = \mathbf{dq_i/dx_j}$ , with different combinations of primitive internal coordinates for butadiene+ethylene a) DLC b) HDLC c) TRIC d) mDLC (TRIC for butadiene and Cartesian coordinates for ethylene).

coordinates for different regions was first done in the context of metal-surface chemistry[10] (vide supra) but can also make the total size of the matrix smaller to diagonalize. To summarize, there are four types of primitive internal coordinate sets studied here:

a) Delocalized internal coordinates (DLC) composed of bonds, angles and torsions
b) Hybrid DLC (HDLC) which include Cartesian coordinates for every atom
c) Translation and rotation internal coordinates (TRIC)
d) Mixed delocalized internal coordinates (mDLC) which treat a subset of atoms with TRIC (or HDLC), and the remainder with Cartesian coordinates

A comparison of the four different G-Matrices are given in Figure 3-2 to demonstrate the differing structure and sparsity. Note that all coordinate systems besides DLC (which is one block), can take advantage of the block-diagonal structure to do efficient block-matrix linear algebra necessary for coordinate transformations.

### **3.2.3** Linear Scaling GSM

Achieving linear scaling in GSM however, requires further consideration due to the inclusion of $U_c$ which represents the reaction coordinate. One possible solution would be to consider the constraint in only the reactive fragment, which, for example, is done in the partial nudged elastic band[21]. However, to include the possibility for correlated motion between the reactive fragment and the remainder of the system, we use a projection scheme which can include the contribution of multiple fragments (Figure 3-3).

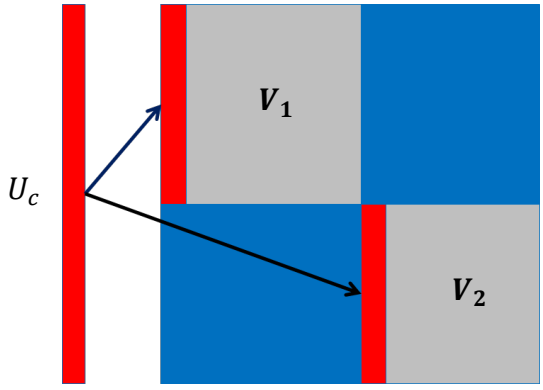



**Figure 3-3** Illustration for distributing the constraint coordinate into the block matrices, for example, butadiene and ethylene.

$U_c$ is split into segments defined by the length of each block, which corresponds to the number of primitive internal coordinates,

$$U_c = |U_{c_1}, U_{c_2}, \dots, U_{c_{3N-1}}, U_{c_{3N}}\rangle = |(U_c)_1, (U_c)_2, \dots, (U_c)_n\rangle \approx \boldsymbol{V}|C\rangle$$

**Equation 3-3** Constraint coordinate in the block-matrix form.

Where n is the number of blocks, and the final approximate equality will become clear soon.

For each block, the magnitude of the constraint (a fraction of one) is saved and stored in a vector with length equal to the number of degrees of freedom (3N)

$$C = \big| \left(c_{1,1}, c_{1,2}, \dots, c_{1,3N_1}\right), \dots, \left(c_{i,1}, \dots\ c_{i,3N_i}\right), \dots, \left(c_{n,1}, \dots\ c_{n,3N_n}\right)\big\rangle$$

$$c_{i,1} = |(U_c)_i|$$

$$c_{i,k} = 0; k = 2,3, \dots 3N_i$$

**Equation 3-4** The C-vector

Where the magnitudes are stored in the first position of each corresponding block, and the remaining positions are zeros. Finally, each $\boldsymbol{U}_i$ are orthogonalized with respect to normalized $\boldsymbol{(U_c)_i}$. Importantly, only block constraints with magnitude greater than a certain threshold are allowed because a vector cannot be orthogonalized with respect to zero. Therefore, a slightly reduced tangent vector (Equation 3-2) is used in practice, and this also has an effect on the extent of the correlations (see below) . Here we have chosen a threshold of $1x10^{-3}$ which has shown stable numerical performance in a wide variety of cases. The resulting orthogonalized blocks are

$$\boldsymbol{V} = \{\boldsymbol{V}_1, \boldsymbol{V}_2, \dots, \boldsymbol{V}_n\}$$

**Equation 3-5** Delocalized internal coordinates in block-matrix form

where each block $\boldsymbol{V}_i$ may or may not contain a constraint vector, depending on value of $c_{1,i}$. Now, multiplication of $\boldsymbol{V}$ and $C$ picks out the constraint coordinate and restores the original individual magnitudes of each constraint component.

In a similar way, the overlap of $C$ (Equation 3-4) and the gradient in the DLC basis provides the magnitude of the gradient along the constraint. Therefore, to optimize the geometry in the non-constrained coordinates the gradient can be formed by projection

$$g_{nc} = g - \langle g|C\rangle C$$

**Equation 3-6** The non-constrained gradient

Likewise, the Hessian in the non-constrained coordinates can be found by projection with $C$

$$\boldsymbol{H}_{nc} = (\boldsymbol{I} - |C\rangle\langle C|)\boldsymbol{H}(\boldsymbol{I} - |C\rangle\langle C|)$$

**Equation 3-7** The non-constrained Hessian

The Hessian is initialized in primitive internal coordinates and updated using the BFGS method in primitive internal coordinates[9]. For large systems (>1000 atoms) the Hessian or primitive Hessian is not formed explicitly but second-order optimization is still achieved using the limited-memory BFGS method[22].

Transition state optimization can be performed with an eigenvector-following algorithm[4], or in the case of large systems a climbing image algorithm which does not require diagonalizing the Hessian[2,9]. The gradient in the direction of the constraint is

$$g_c = \langle g|C\rangle C$$

**Equation 3-8** The constrained gradient

Lastly, in reactions which involves crossings between surfaces, e.g. internal conversion or intersystem crossing, the evaluation of the crossing seams between the potential energy surfaces can provide information on the range of decay pathways to the lower state[18,23,24]. GSM can calculate minimum energy paths in the seam by adding additional constraints which preserve the electronic degeneracy[18]. For example, the coordinates which break the degeneracy of a conical intersection to first order are known as the branching plane (x,y). Thus, to map the seam the vectors **U** are orthogonalized with respect to the constraint coordinate and the branching plane,

$$\boldsymbol{V} = \{U_c, V_x, V_y; V_k \; k = 3{,}4{,}5 \ldots 3N\}$$

**Equation 3-9** Delocalized internal coordinates with orthogonal reaction coordinate and branching space

In the block—matrix formalism, this can be rewritten using

$$C_c = ||(U_c)_1|, 0{,}0 \ldots, |(U_c)_2|, 0{,}0, \ldots, \ldots, |(U_c)_n|, 0{,}0, \ldots\rangle$$

$$C_x = |0, |(V_x)_1|, 0 \ldots, 0, |(V_x)_2|, 0 \ldots, \ldots, 0, |(V_x)_n|, 0 \ldots\rangle$$

$$C_y = |0, 0, \left|(V_y)_1\right| \ldots, 0{,}0, \left|(V_y)_1\right| \ldots, \ldots, 0{,}0, \left|(V_y)_n\right| \ldots\rangle$$

**Equation 3-10** C-vectors for reaction coordinate and branching space

Where the non-zero magnitudes are in the appropriate location within each block, and the corresponding vectors can be recovered as

$$U_c \approx \boldsymbol{V}|C_c\rangle$$

$$V_c \approx \boldsymbol{V}|C_x\rangle$$

$$V_y \approx \boldsymbol{V}|C_y\rangle$$

**Equation 3-11 P**hotochemical coordinates in block-matrix form

## **3.3** Results

First the results present speed comparison of the formation of the different DLC for increasingly larger systems. Then, the capabilities of GSM are demonstrated for a crystal gyroscope, which is known to have correlated motions[25,26].

### **3.3.1** Speed comparison of delocalized internal coordinate systems

In this section we will demonstrate the cost of DLC formation with different primitive internal coordinate sets for cyclohexane crystals. Figure 3-4 presents the computational cost (time) of

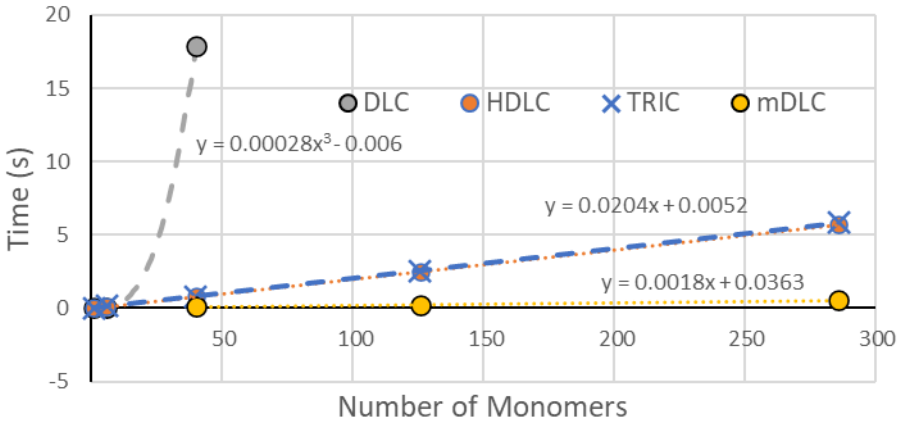



**Figure 3-4** Computational cost for forming coordinate systems of a cyclohexane crystal of increasing size: delocalized internal coordinates (DLC), hybrid delocalized internal coordinates (HDLC), translation and rotation internal coordinates (TRIC), and mixed DLC (mDLC) (TRIC for one cyclohexane monomer and Cartesian coordinates for the rest).

coordinate formation, without constraints, for cyclohexane systems ranging from 1 monomer to a 4x4x4 unit cell which contains 286 monomers.

As expected, the scaling of DLCs is $N^3$ because it diagonalizes the full sparse G-Matrix. However, the scaling for formation of DLC is also worse initially because forming the set of primitive internal coordinates is more computationally intensive than TRIC or HDLC because it adds bonds, angles and torsions between all the fragments. Therefore, even getting to the point of doing interesting computational chemistry (e.g geometry optimization) could take a considerable amount of time. In contrast, for the coordinate systems which do not require a minimum spanning bond it is only necessary to check the connections within the fragment for each fragment, which is linear scaling with respect to the number of fragments.

The scaling for formation of HDLC and TRIC are nearly indistinguishable in the cyclohexane systems. The difference is that HDLC have simpler to evaluate derivatives, but there are more primitive internals for HDLC than TRIC so the scaling can depend slightly on the system. The TRIC coordinate system, however, is recommended owing to the improved intermolecular optimization[15]. In both cases, most of the computation time was spent on the formation of the G-Matrix, not diagonalization. But this depend on the system (and the code implementation). For larger fragments, the computational cost of diagonalization of the G-Matrix blocks will become more dominant.

Finally, further reduction in coordinate transformation can be achieved by treating a subset of the atoms with Cartesian coordinates, and the rest with primitive internal coordinates (mDLC Figure 3-4d). The decrease in computational cost is proportional to the decrease in primitive internal coordinates. However, the choice of the TRIC and Cartesian regions should be carefully selected as the computational cost of coordinate transformation is not the only factor in determining the efficiency of geometry optimization. Different simulation strategies require different amount of correlated motion and therefore different coordinate combinations. For example, in the study of highly correlated motions the use of TRIC is preferable to Cartesian coordinates because of its improved optimization performance[15]. In contrast, in the solvent-perturbed approach[27] all the atoms except the reacting system are frozen to approximate a dynamic system, which has short ranged correlations, so most atoms can be treated with Cartesians to yield the fastest coordinate transformation. Lastly, in the study of crystal nano-clusters with quantum mechanics/molecular

mechanics (QM/MM) there are (usually) no periodic boundary conditions, therefore the outer layer of the nano-cluster can be frozen in order to enforce a crystal packing[28] and can be treated with Cartesian coordinates to reduce computational cost of coordinate transformations.

### **3.3.2** Example GSM for correlated motions

In this section GSM will be applied to the rotation of a molecular gyroscope[25,26] in a crystalline environment. The rotation of the molecular gyroscope in crystals has been previously shown to display long-range correlated motions[25], and is therefore an excellent example to demonstrate how correlated motions are treated in the block-matrix formulism (Equation 3-4).

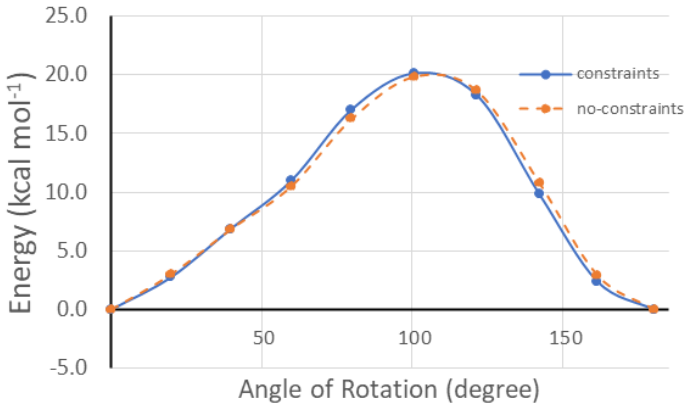



**Figure 3-5** GSM reaction path for rotation of a crystalline molecular gyroscope, with and without constraints

The previous study[25] applied a series of constraints (frozen atoms) and translational and dihedral restraints to maintain crystal packing. When all atoms except the central rotor were frozen the rotational barrier was high (80.4 kcal $mol^{-1}$) owing to the lack of relaxation of the surrounding crystal in response to the rotation. By systematically releasing the constraints, and therefore enabling more lattice relaxation, they obtained rotational barriers of 25.3, 17.5 and 15.5 kcal/mol, which shows that long-range correlated motions are important. Therefore, in this study all atomic coordinates will be treated with TRIC.

In Figure 3-5, the pathway for rotation of a single gyroscope in the crystalline environment is shown, and the structures along the pathway are shown in Figure 3-6. Like the previous investigation[25], the MEP shows a saw-tooth potential owing to the asymmetry with respect to rotation. An exact reproduction of the previous study with GSM, however, is not possible. The previous study only scanned the rotational coordinate (via a dihedral angle) and did not find the exact transition state. Herein, the exact TS is found via the climbing image algorithm[9]. Furthermore, the set of molecular mechanics parameters and constraint set that we have used are slightly different (see Computational Details below). We have also decided to use periodic boundary conditions since this should only affect

the energies and gradients to a minor extent compared to non-periodic molecular mechanics if the constraint set is good. The TS energy we have calculated is 20.1 kcal/mol which is ~5 kcal/mol higher than the previous investigation. Releasing all the constraints within the crystalline environment results in a barrier of 19.8 kcal/mol. The experimental free-energy of rotation has been estimated to be between 12-14 kcal/mol[26,29]. Nevertheless, considering the differences in level of theory, constraints, and numerical procedure the transition state energy is in-line with the previous calculations.

Figure 3-6 shows the structures of the nodes along the reaction path with the structures with high correlated motion shown in licorise representation and with colors representing the magnitude of the correlation. Dark green have lower correlation and red has high correlation. Correlation refers to the response of the environment to the reacting system, via the reaction tangent, Equation 3-4. Specifically, the colors are the magnitudes of the reaction tangent for each molecule/fragment which are stored in the $C$ vector, and are a measure of importance of the motion of that fragment to the total reaction path. The rotor and stator of each gyroscope were treated as individual fragments to view the differences in correlation between rotors and stators.

Sensibly, the fragments with higher correlation to the reaction path changes as the gyroscope rotates. When the gyroscope is in the equilibrium "vertical" position, the motion out of this well disturbs the fragments which are above and below the gyroscope. Likewise as the gyroscope rotates the fragments with largest contribution to the reaction path rotates with the gyroscope. For example, the correlated fragments swap sides going from node 8 to node 1.

The most correlated fragment to the rotational reaction path (besides the rotating rotor) are the benzene solvents which co-crystalize in a tight packing around the gyroscope rotor. Specifically, there are four benzenes which are in the immediate vicinity of the gyroscope rotor; node 5 of Figure 3-6 shows them clearly. The two benzenes that lay on the diagonal edges of the rotor and are represented in pink color (because of their strong correlation) have strong correlation at every angle of rotation. The hydrogen of these benzenes come into close contact with the rotor as the rotor rotates and this causes the benzenes to translate back and forth in a breathing motion. The other two benzenes are angled on the opposite side of the diagonal edge of the rotor so they are less affected except when the rotor is nearly perpendicular to them; see for example, node 6.

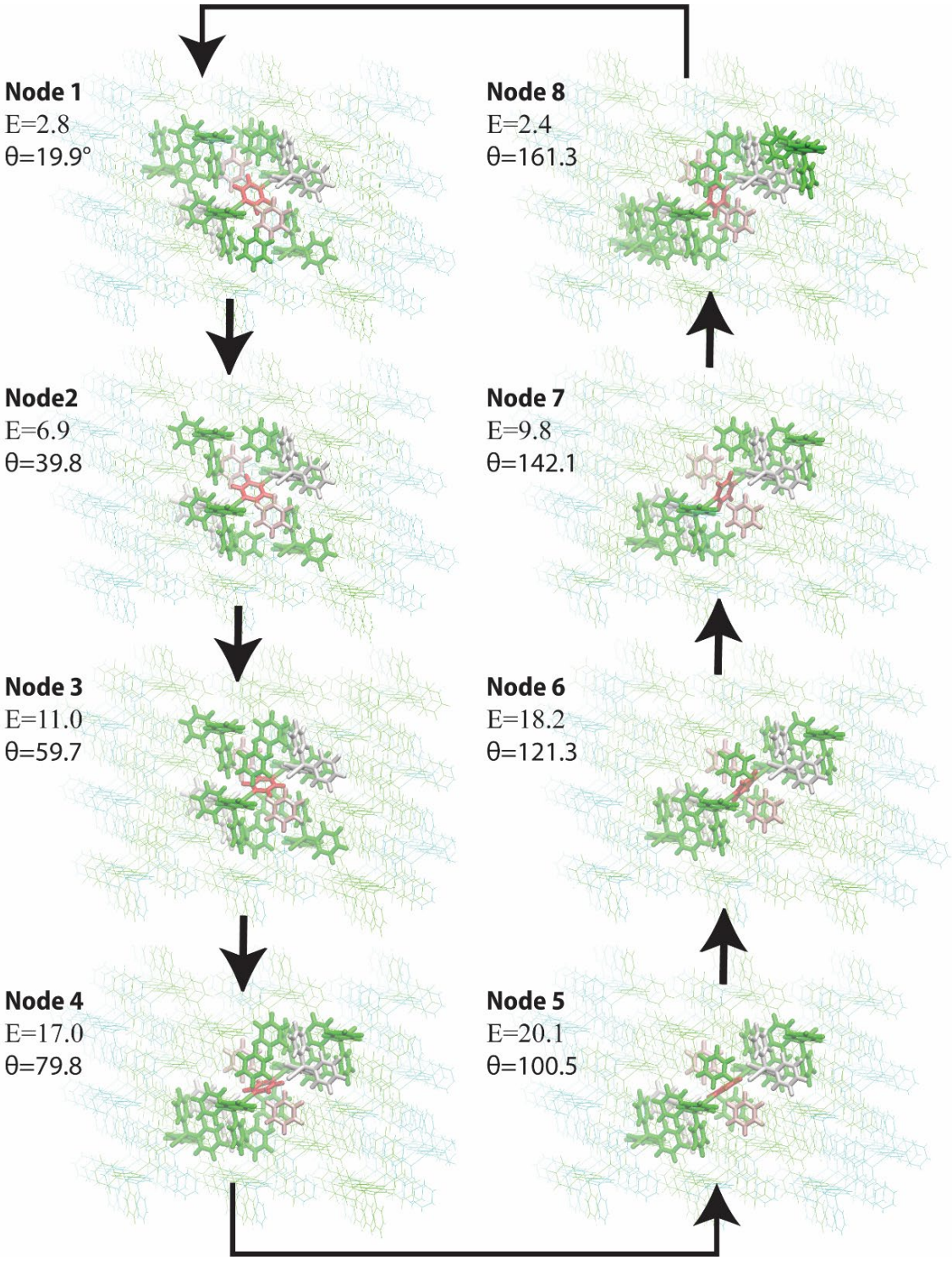



**Figure 3-6** Rotation of a gyroscope 180 degrees embedded in the crystalline environment. The color of the fragment represents the contribution of the fragment to the reaction tangent (Equation 3-4) i.e. correlation. Red licorice representation represents the largest correlation. Light green line representations have magnitudes less than 0.05, cyan lines have zero correlation. Node 0 and node 9 (not shown) have the same correlations as node 1 and node 9 respectively because they use the same tangent vector.

The second strongest correlated fragment is the stators shown in the white color (stronger correlation than dark green) which have phenyls that intercalate between the benzene solvent and rest approximately perpendicular to the rotor at the equilibrium position. These stators experience a strong steric repulsion during the rotation and behave in a breathing motion as the rotor rotates, similar to the two closest benzene. Also note that the rotor attached to those rotors are not significantly correlated, they are light green lines. Looking closely, the primary way that these stators breathe is by rotating which affects surrounding stators but does not affect its own rotor.

Finally, it is relevant to consider the weak correlations with magnitude less than 0.05 shown in light green lines. It is interesting that the motion of the rotor is correlated completely in the vertical direction. Only in the horizontal direction does the correlation die out (cyan lines). The extent of correlation is dependent on the threshold for determining if a fragment's contribution is signifant enough (see Theoretical Details). GSM uses a threshold of $1x10^{-3}$ which has given a stable numerical procedure in a variety of cases, so there is no reason to change this value.

## **3.4** Conclusions

In summary, a general-purpuse, linear scaling growing string method has been presented which can be used to study correlated motions in multi-molecular systems like crystals and proteins, or it can be tailored to study reacting systems embedded in a frozen system, and much in between. The new algorithms achieve linear scaling in multi-molecule systems by block-matrix linear algebra and by distributing the reaction tangent, which is used to grow and optimize the minimum energy path, into each block which represents a molecule or fragment. This tangent is a multi-dimensional vector, and acts as a constraint which prevents motion along that local direction. This results in a stable GSM algorithm for large multi-molecule systems and has the added benefit of being a good proxy for correlation analysis.

### **3.5** Computational Details

All computations were done in the Python3 version of the growing string method (pyGSM) which is available for download, for free, on Github.com[30] and is interfaced with OpenMM for MM calculations[31]. The gyroscope was parameterized with the generalized Amber force-field (GAFF) and use restrained electrostatic potential charges (RESP) derived from quantum mechanics (HF/6-31g*). The atom types and charges are provided in the supporting information.

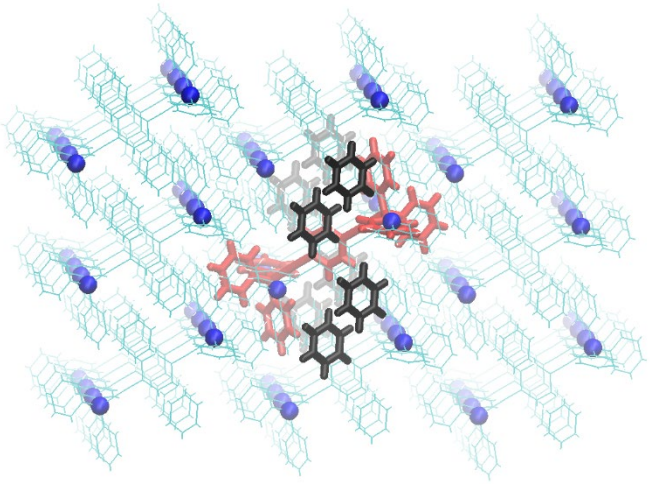


**Figure 3-7** Constraints used for gyroscope.

The transition states are considered optimized when the gradient RMS, max gradient, and reaction tangent are converged to 0.0005 Ha/ Bohr rad, and the difference in energy between successive iterations was less than 0.1 kcal/mol.

In an effort to reproduce the previous investigation, similar constraints were used (Figure 3-7). The red licorice representation is the reacting gyroscope and is free from all constraints and restraints. The benzenes in black licorise representation were kept frozen throughout optimization, this is supposed to preven slippage of the cage benzenes during rotation. The blue spheres were restrained to their initial position by a 100 kcal/mol·Å harmonic restraining potential. Lastly, for all gyroscopes except the reacting system three torsion restraints were applied between the trityl groups (the stators), one for each pair, to restrain them to the initial relative rotation.

### **3.6** References


(1) Henkelman, G.; Jónsson, H. Improved Tangent Estimate in the Nudged Elastic Band Method for Finding Minimum Energy Paths and Saddle Points. *J. Chem. Phys.* **2000**, *113* (22), 9978–9985. https://doi.org/10.1063/1.1323224.

(2) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. Climbing Image Nudged Elastic Band Method for Finding Saddle Points and Minimum Energy Paths. *J. Chem. Phys.* **2000**, *113* (22), 9901–9904. https://doi.org/10.1063/1.1329672.

(3) Weinan, E.; Ren, W.; Vanden-eijnden, E. Simplified and Improved String Method for Computing the Minimum Energy Paths in Barrier-Crossing Events Simplified and Improved String Method for Computing the Minimum Energy Paths in Barrier-Crossing Events. **2009**, *164103* (2007). https://doi.org/10.1063/1.2720838.

(4) Zimmerman, P. Reliable Transition State Searches Integrated with the Growing String Method. *J. Chem. Theory Comput.* **2013**, *9* (7), 3043–3050. https://doi.org/10.1021/ct400319w.

(5) Zimmerman, P. M. Single-Ended Transition State Finding with the Growing String Method. *J. Comput. Chem.* **2015**, *36* (9), 601–611. https://doi.org/10.1002/jcc.23833.

(6) Rong, Z.; Kitchaev, D.; Canepa, P.; Huang, W.; Ceder, G. An Efficient Algorithm for Finding the Minimum Energy Path for Cation Migration in Ionic Materials. *J. Chem. Phys.* **2016**, *145* (7). https://doi.org/10.1063/1.4960790.

(7) Smidstrup, S.; Pedersen, A.; Stokbro, K.; Jónsson, H. Improved Initial Guess for Minimum Energy Path Calculations. *J. Chem. Phys.* **2014**, *140* (21), 1–14. https://doi.org/10.1063/1.4878664.

(8) Peters, B.; Heyden, A.; Bell, A. T.; Chakraborty, A. A Growing String Method for Determining Transition States: Comparison to the Nudged Elastic Band and String Methods. *J. Chem. Phys.* **2004**, *120* (17), 7877–7886. https://doi.org/10.1063/1.1691018.

(9) Zimmerman, P. M. Growing String Method with Interpolation and Optimization in Internal Coordinates: Method and Examples. *J. Chem. Phys.* **2013**, *138* (18), 1–11. https://doi.org/10.1063/1.4804162.

(10) Jafari, M.; Zimmerman, P. M. Reliable and Efficient Reaction Path and Transition State Finding for Surface Reactions with the Growing String Method. *J. Comput. Chem.* **2017**, 645–658. https://doi.org/10.1002/jcc.24720.

(11) Behn, A.; Zimmerman, P. M.; Bell, A. T.; Head-Gordon, M. Incorporating Linear Synchronous Transit Interpolation into the Growing String Method: Algorithm and Applications. *J. Chem. Theory Comput.* **2011**, *7* (12), 4019–4025. https://doi.org/10.1021/ct200654u.

(12) Baker, J. Constrained Optimization in Delocalized Internal Coordinates. *J. Comput. Chem.* **1997**, *18* (8), 1079–1095. https://doi.org/10.1002/(SICI)1096-987X(199706)18:8<1079::AID-JCC12>3.0.CO;2-8.

(13) Baker, J.; Kessi, A.; Delley, B. The Generation and Use of Delocalized Internal Coordinates in Geometry Optimization. *J. Chem. Phys.* **1996**, *105* (1996), 192–212. https://doi.org/10.1063/1.471864.

(14) Grimme, S. Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. *J. Chem. Theory Comput.* **2019**, *15* (5), 2847–2862. https://doi.org/10.1021/acs.jctc.9b00143.

(15) Wang, L. P.; Song, C. Geometry Optimization Made Simple with Translation and Rotation Coordinates. *J. Chem. Phys.* **2016**, *144* (21), 214108. https://doi.org/10.1063/1.4952956.

(16) Bučko, T.; Hafner, J.; Ángyán, J. G. Geometry Optimization of Periodic Systems Using Internal Coordinates. *J. Chem. Phys.* **2005**, *122* (12). https://doi.org/10.1063/1.1864932.

(17) Billeter, S. R.; Turner, J.; Thiel, W.; Zu, C. H.; Turner, A. J.; Thiel, W. Linear Scaling Geometry Optimisation and Transition State Search in Hybrid Delocalised Internal Coordinates. *Phys. Chem. Chem. Phys.* **2000**, *2* (10), 2177–2186. https://doi.org/10.1039/a909486e.

(18) Aldaz, C.; Kammeraad, J. A.; Zimmerman, P. M. Discovery of Conical Intersection Mediated Photochemistry with Growing String Methods. *Phys. Chem. Chem. Phys.* **2018**, *20* (43), 27394–27405. https://doi.org/10.1039/C8CP04703K.

(19) Dewyer, A. L.; Zimmerman, P. M. Finding Reaction Mechanisms, Intuitive or Otherwise. *Org. Biomol. Chem.* **2017**, 501–504. https://doi.org/10.1039/C6OB02183B.
(20) Baker, J.; Pulay, P. Geometry Optimization of Atomic Microclusters Using Inverse-Power Distance Coordinates. *J. Chem. Phys.* **1996**, *105* (24), 11100–11107. https://doi.org/10.1063/1.472911.
(21) Bergonzo, C.; Campbell, A. J.; Walker, R. C.; Simmerling, C. A Partial Nudged Elastic Band Implementation for Use with Large or Explicitly Solvated Systems. *Int. J. Quantum Chem.* **2009**, *109* (15), 3781–3790. https://doi.org/10.1002/qua.22405.
(22) Liu, D. C.; Nocedal, J. On the Limited Memory BFGS Method for Large Scale Optimization. *Math. Program. Ser. B* **1989**, *45* (3), 503–528.
(23) Mori, T.; Martínez, T. J. Exploring the Conical Intersection Seam: The Seam Space Nudged Elastic Band Method. *J. Chem. Theory Comput.* **2013**, *9* (2), 1155–1163. https://doi.org/10.1021/ct300892t.
(24) Paterson, M. J.; Bearpark, M. J.; Robb, M. a.; Blancafort, L. The Curvature of the Conical Intersection Seam: An Approximate Second-Order Analysis. *J. Chem. Phys.* **2004**, *121* (23), 11562–11571. https://doi.org/10.1063/1.1813436.
(25) Jarowski, P. D.; Houk, K. N.; Garcia-Garibay, M. A. Importance of Correlated Motions on the Low Barrier Rotational Potentials of Crystalline Molecular Gyroscopes. *J. Am. Chem. Soc.* **2007**, *129* (11), 3110–3117. https://doi.org/10.1021/ja0637907.
(26) Godinez, C. E.; Zepeda, G.; Garcia-Garibay, M. A. Molecular Compasses and Gyroscopes. II. Synthesis and Characterization of Molecular Rotors with Axially Substituted Bis[2-(9-Triptycyl)Ethynyl]Arenes. *J. Am. Chem. Soc.* **2002**, *124* (17), 4701–4707. https://doi.org/10.1021/ja012550i.
(27) Yang, Z.; Doubleday, C.; Houk, K. N. QM/MM Protocol for Direct Molecular Dynamics of Chemical Reactions in Solution: The Water-Accelerated Diels-Alder Reaction. *J. Chem. Theory Comput.* **2015**, *11* (12), 5606–5612. https://doi.org/10.1021/acs.jctc.5b01029.
(28) Zimmerman, H. E.; Zuraw, M. J. Photochemistry in a Box. Photochemical Reactions of Molecules Entrapped in Crystal Lattices: Mechanistic and Exploratory Organic Photochemistry. *J. Am. Chem. Soc.* **1989**, *111* (20), 7974–7989. https://doi.org/10.1021/ja00202a044.
(29) Garcia-Garibay, M. A. Crystalline Molecular Machines: Encoding Supramolecular Dynamics into Molecular Structure. *Proc. Natl. Acad. Sci.* **2005**, *102* (31), 10771–10776. https://doi.org/10.1073/pnas.0502816102.
(30) Aldaz, C. https://github.com/ZimmermanGroup/pyGSM.
(31) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L. P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**, *13* (7), 1–17. https://doi.org/10.1371/journal.pcbi.1005659.

# **Chapter 4** The Mechanics of the Bicycle Pedal Photoisomerization in Crystalline cis,cis-1,4-diphenyl-1,3-butadiene

## **4.1** Abstract

Direct irradiation of crystalline cis,cis-1,4-diphenyl-1,3-butadiene (cc-DPB) forms trans,trans-1,4-diphenyl-1,3,-butadiene via a concerted two bond isomerization called the bicycle pedal mechanism. However, little is known about photoisomerization pathways in the solid-state and there has been much debate surrounding the interpretation of volume-conserving isomerization mechanisms. The bicycle pedal photoisomerization is investigated using the QM/MM complete active space self-consistent field (CASSCF)/generalized amber force field (GAFF) method. Important details about how the steric environment influences isomerization mechanisms are revealed including how the one-bond flip and hula-twist mechanisms are suppressed by the crystal cavity, the nature of the seam space in steric environments, and the features of the bicycle pedal mechanism. Specifically, in the bicycle pedal the phenyl rings of cc-DPB are locked in place and the intermolecular packing has an open passageway allowing for a smooth rotation of the central diene in a volume-conserving manner.. In contrast, the bicycle pedal rotation in the gas-phase is not a stable pathway, so single-bond rotation mechanisms become operative instead. Lastly, the reaction barriers of the different crystalline conformations within the unit cell of cc-DPB are compared to investigate the possibility for conformation-dependent isomerization. Although some difference in reaction barriers are observed, the difference is most likely not responsible for the experimentally observed periods of fast and slow conversion. However, presently the models do not capture the quantitative activation barriers and more work is needed to better model reactions in crystals.

## **4.2** Introduction:

The ability of light to trigger mechanical changes via photoisomerization is well known in vision and phototaxy, driving efforts to replicate these features in artificial molecular machines and optomechanical devices. This has led to the study of specific isomerization mechanisms such as one-bond flip (OBF, also known as the double-bond twist), volume-conserving hula-twist (HT), and the bicycle pedal (BP) mechanisms[1]. The relative importance of these photoisomerization steps in condensed phases, however, is particularly sensitive to the steric environment. Furthermore, the study of these mechanisms is notoriously difficult because thermal isomerization about carbon-carbon single bonds can mix the photoproducts. As a result, many questions and debate[2–6] surround the interpretation of photoisomerization mechanisms, and this confusion inhibits rational design.

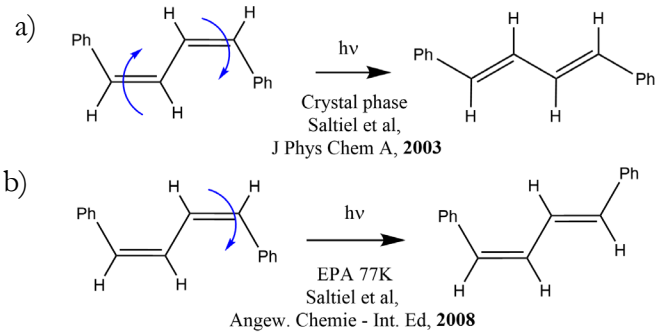



**Figure 4-1** Irradiation of cc-DPB in different steric environments results in different isomerization products a) crystal b) viscous environment.

Herein, the isomerization of crystalline cis,cis-1,4-diphenyl-1,3-butadiene (cc-DPB),[7,8] to trans,trans-1,4-diphenyl-1,3-butadiene (tt-DPB) is investigated (Figure 4-1a). This isomerization may proceed through the concerted BP mechanism, which has been noted in crystalline butadienes[9–11,12], some of which produce macroscopic motion like peeling[12], hexatrienes[13,14], photoactive yellow protein[15], and has been implicated in the visual retinoid cycle[16]. Photoirradiation of crystalline cc-DPB is interesting amongst these because in the solution and glassy phase the BP isomerization is minor or absent. For example, photoirradiation of cc-DPB in ethanol[17] produces ct-DPB via OBF or HT, and a minor amount of tt-DPB through a phenylallylbenzyl intermediate which is not a concerted rotation as portrayed in the BP mechanism[17]. Furthermore, irradiation of cc-DPB in viscous media like glassy phases mostly do not undergo BP isomerization: photoirradiation in the viscous isopentane glass resulted a mixture of single double-bond isomerization and BP isomerization[7], whereas photoirradiation in the harder EPA (ether:isopentane:ethyl alcohol=5:5:2) glass resulted in no BP isomerization(Figure 4-1b)[18]. This is unexpected because isomerization in volume restricting environments is generally expected to proceed via the more volume-conserving mechanism (e.g. HT

vs OBF)[1]. These observations therefore suggest that specific steric and structural interactions must be present in the crystal which inhibit the single double-bond isomerization and enable the BP pathway.

In order to gain more information into this intriguing mechanism, it is useful to perform *ab initio* computational analysis. Previous studies on related compounds[19–25] have demonstrated that quantum chemical simulations can provide important details of photoisomerization mechanisms that would be useful for understanding cc-DPB photochemistry. Specifically, conical intersections (crossings between potential energy surfaces) enable efficient non-radiative decay between states, and therefore control photoproduct formation. Quantum mechanical studies of conical intersections responsible for the BP mechanism, however, are to the best of our knowledge unexplored. Furthermore, the investigation of reaction mechanisms in crystal environments is more difficult than their gas-phase counterparts. H. Zimmerman was the first to report *ab initio* investigations of crystal photochemical reactions by approximating crystal packing with an inert gas-shell model[26]. Later, the same author optimized mini-lattices, also known as a cluster model, using a quantum mechanics/molecular mechanics (QM/MM) ONIOM approach[27,28]. Recent advances in related techniques applied to crystal photochemistry[29] have permitted improved geometry optimization as well as better treatment of short-range[30] and long-range electrostatic effects[31,32] in crystals with quantum mechanics/quantum mechanics' methods (QM/QM') methods.

In this article, a recently developed reaction path optimization method (Chapter 3)[33,34] is applied to analyze photoisomerization of cc-DPB. First, analysis of traditional single double-bond isomerization pathways provides evidence towards why these pathways are forbidden in crystals. Detail of the BP mechanism are then revealed for the first time. The descriptions of these pathways will reveal how the crystal phase affects conical intersections and on how the broader seam topography depends on restrictions on intramolecular rotation. Finally, a hypothesis about the isomerization rate varying on a conformer-specific level is evaluated[8,35,36]. Altogether, this information provides a close look at the effects of environment on photoisomerization mechanisms and explains how qualitative changes in reaction pathways can come with entry of molecules into the crystal phase.

## **4.3** Methods:

### **4.3.1** Cluster model

The cc-DPB crystal structure was retrieved from the Cambridge crystal database (CIF #29021)[7]. A 3x3x3 cluster was prepared and a four-layer QM/MM approach was utilized for all calculations unless otherwise stated (Figure 4-2). The innermost layer is treated with quantum mechanics, the second layer is treated with molecular mechanics and the outermost layer is kept frozen to enforce crystal packing. An extra layer arises here because of the hybrid Cartesian and translation-rotation internal coordinate (TRIC) system (see below)[34]. The choice of layers is as follows. The first layer is the QM residue, which depends on the specific conformation of the unit cell being investigated (see below). The second layer is defined as all monomers with all atoms less than 5 Å from the QM residue. The third layer is defined as monomers with all atoms greater than 5 Å away from the QM residue and less than 10 Å from the QM residue. The fourth layer is all monomers with all atoms greater than 10 Å from the QM residue which is kept frozen. The QM layer is treated with the state-averaged complete active space self-consistent field (SA-CASSCF) method. All layers besides the QM layer are treated with the Amber force field and utilize restrained electrostatic potential charges (RESP) derived from quantum mechanics. The atom types and charges are provided in the **Supporting Information**. The choice of active space in the SA-CASSCF is discussed below because that requires further experimental and theoretical considerations.

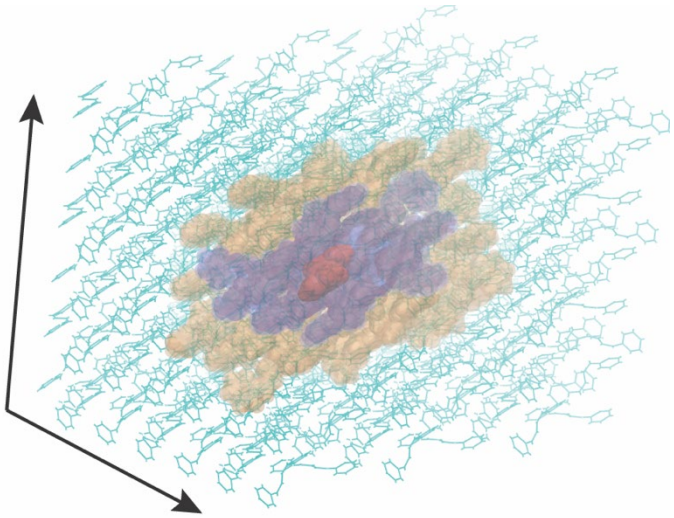


**Figure 4-2.** 3x3x3 cluster model of cc-DPB. The colors represent the four-layer geometry optimization approach. Red VDW: non-frozen QM region treated with TRIC, blue VDW: non-frozen MM region treated with TRIC, transparent VDW orange: non-frozen MM region treated with Cartesians, line cyan: frozen MM region treated with Cartesians. There are 12960 atoms total.

### **4.3.2** Reaction Path Optimization

Reaction path optimization was performed with the growing string method (GSM) as implemented in Python3[37]. The core GSM algorithm has been described separately[38,39], and modifications necessary to treat large systems will be described in detail elsewhere[33]. Therefore, only a brief description is given here. GSM is an efficient reaction path optimization algorithm that can be

operated in either single-ended or double-ended mode. Single-ended mode requires an initial structure and a driving coordinate direction (e.g. adding or breaking bond) and is therefore well suited for reaction discovery and exploration[40]. This mode has also been shown to be useful for finding MECI starting from stable equilibrium geometries[41]. The double-ended mode requires an initial and final structure as input. GSM employs delocalized internal coordinates (DLCs) which are superior to Cartesian coordinates for geometry optimization[42]. In the present implementation of GSM, translation and rotation internal coordinates[34] are used to reduce the computational cost of coordinate transformations. With the inclusion of TRIC, the individual molecules of the cluster are decoupled and therefore block-matrix diagonalization techniques are used to achieve $N_{mol}$ scaling computational costs[33].

Python GSM is available for free on Github[37] and is interfaced with the GPU accelerated TeraChem for QM and QM/MM calculations[43–45], and OpenMM for MM calculations[46]. Initial geometry optimization of the 3x3x3 crystal was performed with OpenMM without the frozen constraints to relax the MM region. MECI optimization was performed using an iterative penalty potential[47]. Gradient RMS, gradient maximum, and energy difference criteria were used to determine optimization convergence. Transition state optimization was performed using a climbing image algorithm[38]. The transition states are considered optimized when the gradient RMS, and reaction tangent are converged to 0.0005 Ha/ Bohr rad, and the difference in energy between successive iterations was less than 0.1 kcal/mol. Seam saddle point were optimized with climbing image optimization using the same criteria except on the penalty potential with $\sigma$ of 10.

**4.3.3** Complete active space self-consistent field

Photoisomerization can proceed through crossings between potential energy surfaces, which requires electronic structure theory that can treat electronic degeneracies[48]. The CASSCF method is therefore a workhorse in the study of photoisomerization and photochemical reactions. Previous computations on 1-phenyl-1,3-butadiene have analyzed the potential energy surface with the SA2-CAS(10,10)SCF/6-31+G* level of theory, which include the 10 carbon p orbitals involved in π-bonding[22]. These computations showed the excited state is more planar due to increased coupling between the phenyl and diene moiety, and a low energy (3 kcal $mol^{-1}$ higher than the $S_1$ minimum) conical intersection was found leading to a spiro-cyclization photoproduct, which agrees with available experimental data. In contrast, the isomerization pathways were found to be much higher at about 25

kcal mol$^{-1}$ above the $S_1$ minimum. The isomerization MECI can be described as *s-transoid* (s-trans with respect to the phenyl diene single bond), and a *central* MECI by analogy with butadiene[19].

Butadiene has also been extensively studied and it has been found that SA-CASSCF with the nonintuitive active space of four electrons in three orbitals closely approximates both second order MS-CAS(4,4)PT2[49] and experiment. This choice of active space destabilizes the $S_1$ surface while leaving the $S_2$ surface largely unchanged, restoring the near-degeneracy of $S_2$ and $S_1$ in the Franck-Condon region. The *transoid* MECI is about 39 kcal mol$^{-1}$ below the Franck Condon point, the $S_1$ minimum of butadiene is a twisted methylene. At the higher SA-3-(4,4)-MSPT2/6-31G** level of theory [21] the *transoid* MECI is about 22 kcal mol$^{-1}$ below the planar 2 $^1A_g$ state, and 41.5 kcal mol$^{-1}$ below the Franck-Condon point.

The full π space of cc-DPB requires a (16,16) active space which is at the high-end of computational capabilities, especially for the relaxed reaction path optimization performed herein. Furthermore, a larger active space is not always better, and this is perhaps evident in 1-phenyl-1,3-butadiene that has a large isomerization barrier at the CAS(10,10) level. Therefore, for an initial investigation it is reasonable to begin with a smaller active space. To our satisfaction, SA3-CAS(4,3)SCF captures spectroscopic characteristics in agreement with SA-3-(4,4,)-MS-PT2 calculations: the $S_1$ is the bright state and, like PB, is described as bichromophoric (see **Supporting Information**). The SA3-CAS(4,4)SCF level, in contrast, predicts $S_2$ as the bright state and the excitation is localized on the butadiene chromophore.

### **4.4** Results and Discussion

The analysis of cc-DPB photoisomerization pathways begins with a brief description of the crystal and gas-phase ground- and excited-state structures. Next, single double-bond isomerization mechanisms are revealed, with details about why these are prohibited in the crystal. Then a comprehensive analysis of BP pathways is provided. Finally, the BP reaction paths for all conformers within the unit crystal are compared to determine if certain conformations isomerize with faster rates than others.

### **4.4.1** Crystal Description

The cc-DPB unit cell contains two conformers of the monomer, for a total of four DPB molecules. One of the conformers has the phenyls in perpendicular planes (**A**,**B**) and the other has phenyls in parallel planes (**C**,**D**), see Figure 4-3. The two conformers of cc-DPB arrange in edge-to-face alternating arrays: the perpendicular conformers clasp together in a row of the array, and the parallel conformers clasp together in the next row of the array. The parallel and perpendicular conformers are also stable geometries in the gas-phase. However, at the CASSCF level of theory and in the gas phase the angle of the diene with respect to the phenyl planes is 90°.

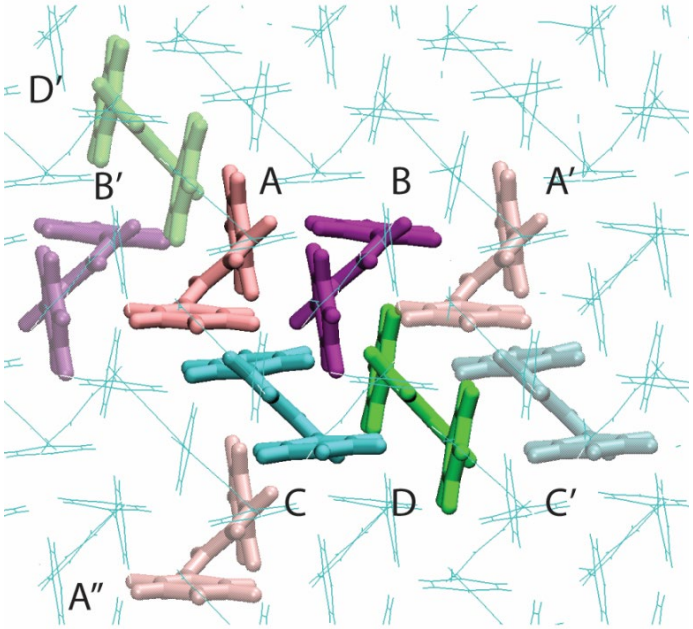



**Figure 4-3.** Unit cell with labels. Cyan: A, Purple: B, Red: C, Green D. The lighter colored monomers are members of adjacent cells, which interact with the central conformers.

**Table 4-1** CASSCF excitation energies in eV, (CASPT2 energies of the single monomer )

| Geometry | $\Delta E_{vert}$ | $\Delta E_{adiab}$ |
|---|---|---|
| **A** isomer | 5.942 | 4.468 (3.794) |
| **B** isomer | 6.000 | 4.600 (3.767) |
| **C** isomer | 5.827 | 4.250 (3.562) |
| **D** isomer | 6.050 | 4.453 (3.771) |
| **Perpendicular** Vacuum | 5.727 | 4.317 |
| **Parallel** Vacuum | 7.513* | 4.175 |

* $C_{2h}$ symmetry

The calculated vertical and adiabatic excitation energies of each of the conformers in the crystal phase and gas-phase are reported in **Table 4-1**. The experimental vertical and adiabatic excitation energy can be estimated from the excitation and emission spectrum, respectively, and are approximately 3.387 eV (366 nm) and 3.024 eV (410 nm)[8]. In contrast, in cyclohexane solution the vertical and adiabatic excitation energies are about 4.27 eV (290 nm), and 3.56 eV (360 nm)[50]. Therefore, the crystal exhibits a strong bathochromic shift with respect to solution[31,32]. CASSCF is known to lack dynamical electron correlation which causes large excitation energies. The inclusion of dynamical correlation with CASPT2 bring the energies down in line with solution values but still lacks the electrostatic effects of the bulk crystal.

**4.4.2** Single Double Bond Rotation

The photoisomerizations of conformer **C** were examined first because this conformer has the lowest energy initial excited state. Two *cis,cis-s-transoid* MECIs were found, where the *s-trans* nomenclature is with respect to the central single bond. These MECI are labeled as *cc-transoid-L* and *cc-transoid-R* because they resemble the rotation of the left and right side of a bicycle pedal, respectively (top row Figure 4-4, and **Table 4-2**). To understand how the crystal steric environment affects these MECI, the gas-phase MECI were also found. The structures are overlaid to show how the crystal affects the geometry and the energies are shown in the color matching the conformer (e.g. dark blue for gas phase). The crystalline *cc-transoid* MECI differ from the gas phase, in that they are higher in energy than their respective $S_1$ minimum by 6.6 and 13.1 kcal $mol^{-1}$ for *cc-transoid-L* and *cc-transoid-R*, respectively[51]. The increase in energy of conical intersections from gas-phase to crystal can be explained by restrictions on intramolecular rotation (RIR)[52,53]. As can be seen in Figure 4-4, the gas-phase cc-*transoid* MECI are able to significantly deviate from planarity ($\alpha$=43° and 51°) whereas the crystalline phenyls require less than half the amount of out-of-plane rotation, consistent with RIR. Rotating out of plane is energetically favorable in the gas phase because it helps the phenyl maintain conjugation with the rotating interior hydrogen (H4 or H5). In contrast, $\alpha$ is prevented from changing due to steric repulsion in the crystal form.

**Table 4-2** Single double-bond isomerization MECI values

| MECI | | S0 | S1 | θ | φ | α* |
|---|---|---|---|---|---|---|
| cc-transoid-L | gas-phase | -1.8 | -1.7 | -83.6 | -130.9 | 42.8 |
| | crystal-phase | 6.6 | 6.7 | -86.6 | -137.7 | 20.6 |
| cc-transoid-R | gas-phase | -1.8 | -1.7 | 85.3 | 131.4 | 50.6 |
| | crystal-phase | 13.1 | 13.1 | 83.6 | 137 | 23.7 |
| s-transoid | gas-phase | -1.4 | -1.3 | -69.6 | -166.1 | 40.3 |
| | crystal-phase | 3.0 | 3.2 | -29.5 | -174.4 | 3.8 |
| s-central | gas-phase | -7.0 | -6.9 | 115.6 | -173.1 | 120.8 |
| | crystal-phase | -0.6 | -0.6 | 38.0 | 166.8 | 29.1 |

* $\alpha$ is defined as the angle between the planes of the phenyl and the $S_1$-minimum.

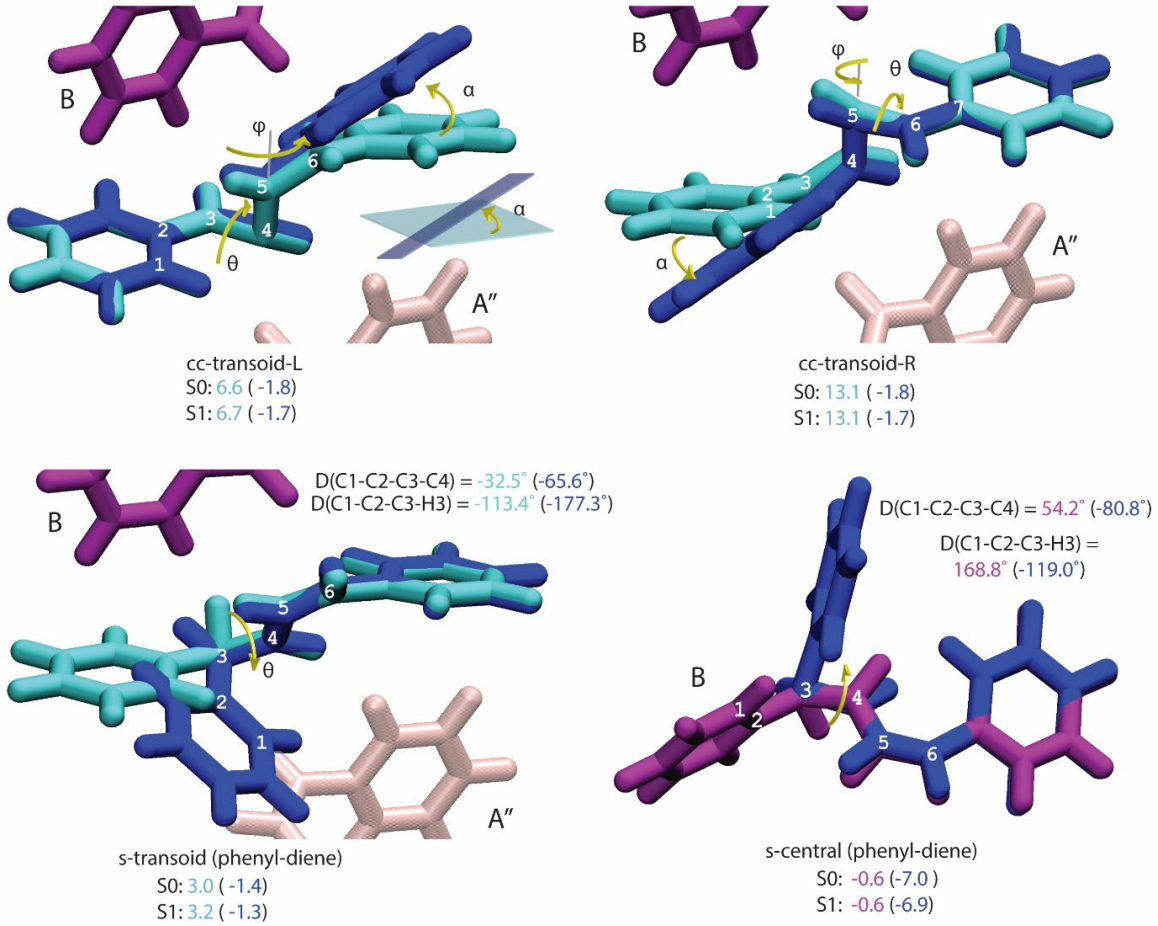



**Figure 4-4.** Geometries of gas-phase (dark blue) and crystalline transoid conical intersections (see **Figure 4-3** for color scheme) involved in single double bond isomerization. The important coordinates are labeled one-bond flip coordinate (θ), single-bond rotation coordinate (φ) and phenyl out-of-plane angle (α). α is defined as the angle between the planes of the phenyls of the MECI and $S_1$ minimum. The energies and values for the different parameters are shown in the color matching the conformer (e.g. dark blue for gas phase). The $S_0$ and $S_1$ energies, in kcal $mol^{-1}$ are relative to the respective $S_1$ minimum geometry.

The MECI geometries imply the existence of specific photoproducts. Herein the gas-phase *cc-transoid* are assumed to lead to OBF and HT-2 (Figure 4-5) because they contain a combination of double-bond rotation (θ) and adjacent single-bond rotation (φ) coordinates (Figure 4-4)[24]. In the OBF mechanism the phenyl completes the ~180° twist about the carbon-carbon double bond. For example, OBF characteristics can be seen in the angle α as the phenyl group flips over C3-C4 in Figure 4-4. In the HT mechanism single-bond rotation about the central double bond allows the CH group (e.g. C4-H4) to complete the twist about the double bond without flipping the phenyl; α decreases as the hydrogen finishes the hula-twist. Interestingly, the perpendicular geometry α begins at ~90°, therefore only half the necessary phenyl rotation is required for OBF, which may facilitate OBF in the gas-phase.

The expected photoproducts in the crystal-phase, however, are much different. The *cc-transoid* MECI distortions imply that completion of the OBF and HT are no longer allowed[24]. OBF is clearly suppressed in the crystal state via restrictions on out-of-plane rotation due to the constrained cavity. Furthermore, although α is small, rotation about the central single bond (φ) is also suppressed in the crystal lattice, and this is evident in the MECI with φ about 7° larger; a smaller φ means greater rotation. This is interesting because it shows that although HT is more volume conserving than OBF it still requires that the phenyls "slide" via the single-bond rotation coordinate φ, which is prohibited in the crystal phase.

Another class of CI important for single double-bond isomerization occur via carbon pyramidalization at the first carbon of the diene chain. Two CI[22] with pyramidalization at the first carbon can be found in the gas-phase for cc-DPB. The two differ in the dihedral phenyl-diene single bond, and can be described as *s-central* and *s-transoid* where the *s-trans* and *s-central* naming refers to the phenyl-diene single bond (Figure 4-4 bottom row and **Table 4-2**)[22]. These CI are assumed to lead to HT-1 and OBF photoproducts (Figure 4-5). The difference in photoproducts is the orientation of the resulting phenyl: in OBF the phenyl ring flips 180° about the phenyl-diene carbon-carbon single bond, whereas HT-1 preserves the relative phenyl position.

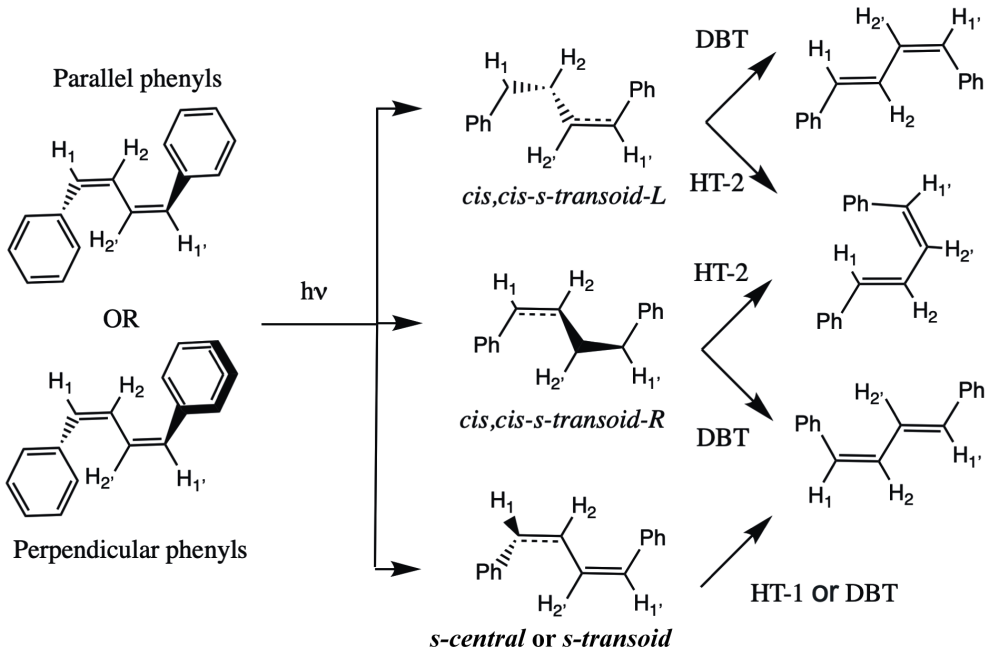



**Figure 4-5.** Single double-bond isomerization mechanisms. Operative only in the gas phase due to steric blockage in the crystal.

Crystal phase planar analogues of *s-transoid* and *s-central* were found starting from the parallel phenyl geometries (**C**) and perpendicular phenyl geometries (**B**), respectively. In comparison to *cc-transoid*, an even greater difference between gas phase and crystal phase exists for this pair of conical intersections. In the crystal phase, the dominant motion rotates the hydrogen out of plane D(C1-C2-C3-H3) = -113.4° and the phenyls remain in plane (e.g. $\alpha$=4°). This is unusual because rotation of hydrogen out of plane has only been found in the gas-phase as a high-energy seam saddle point[23,41,54]. In order for this conical intersection to become responsible for the HT in the condensed phase, the steric surroundings must in some way cause this conical intersection to become lower energy. This has previously been hypothesized[23,41,54], but was not tested until now due to the lack of capacity to simulate the reaction mechanism in the condensed phase, and therefore highlights the importance of explicitly modeling the steric environments: the nature of the potential energy surface topology including the seam space can change dramatically in the condensed phase. Incidentally, SE-GSM was particularly beneficial here because the gas-phase and crystal phase structures were found from the same input reaction coordinates. In contrast, the optimization of these structures via traditional techniques where the gas-phase structure is used as a guess for the reacting system of the crystal would have posed a significant challenge because the gas-phase structures do not fit within the crystal cavity. Regardless, the crystalline *s-transoid* and *s-central* CI analogues are non-productive in the crystal phase as well due to restrictions on side-to-side rotations.

### **4.4.3** Bicycle Pedal Rotation

The CI responsible for bicycle pedal isomerization in the crystal phase is now identified and analyzed. As demonstrated above, the *cc-transoid-L* and *cc-transoid-R* MECI each resemble one half-rotation of a bicycle pedal. Putting these motions together in a concerted manner completes the cycle. As can be seen in Figure 4-6a, the bicycle-pedal CI be described as *trans,trans-s-transoid,* or *tt-transoid* for short. As above, the gas-phase *tt-transoid* is more non-planar than the crystalline form, and this extra flexibility allows gas-phase *tt-transoid* to reach a lower energy than the crystal phase.

Excited-state reaction paths were found connecting the $S_1$-minimum and the *tt-transoid* CI. In the crystal state, the reaction path is a single, smooth elementary step consisting of ~180 degree rotation of the central diene, followed by pyramidalization (Figure 4-6b). In contrast, in the gas-phase the reaction path from the $S_1$-minimum to *tt-transoid* undergoes significant out of plane distortion as the phenyls attempt to maintain conjugation with the central rotating diene (see **Supporting Information)**. No exact transition state geometry could be optimized, probably due to the unstable (high energy) nature of the reaction path. Taken together with the steepness of the potential energy surface towards the *cc-transoid* CI in the gas-phase (Figure 4-4), this indicates that the BP pathway is not operable in the gas phase. The crystal BP mechanism is also much different than the solution-phase cc-DPB→tt-DPB mechanism which proceeds through a cis-phenylallylbenzyl excited state intermediate[17]. The cis-phenylallylbenzyl intermediate for example, would require side-to-side motion of the phenyl, which is not allowed in the crystal. Therefore, the crystal state is unique in its operation: the steric environment opens the bicycle pedal isomerization reaction valley by restricting out-of-plane motion allowing smooth rotation of the diene.

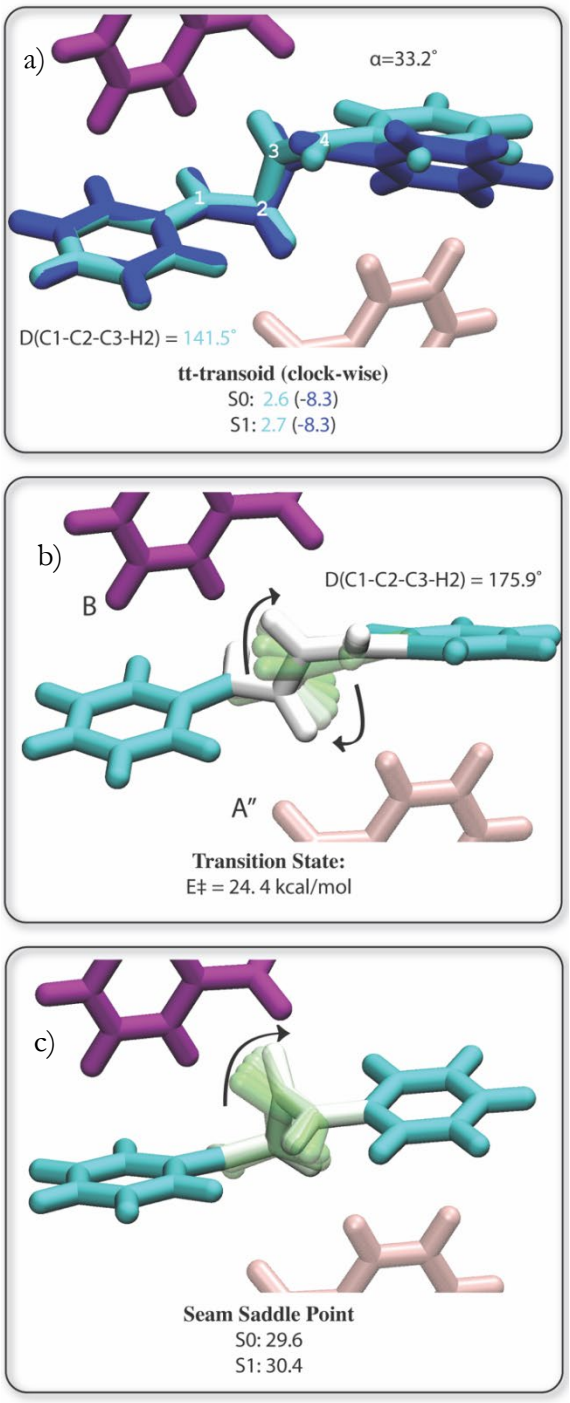



**Figure 4-6.** Relevant geometries for the bicycle pedal isomerization a) comparison of gas and crystal phase tt-transoid MECI, b) excited-state transition state connecting S1 minimum and transoid MECI c) seam saddle point connecting cc-transoid-R and tt-transoid MECI. The $S_0$ and $S_1$ energies, in kcal $mol^{-1}$ are relative to the $S_1$ minimum geometry of C.

Close inspection of the transition state geometry reveals additional important details. First, as can be seen in Figure 4-6b, the pedal rotates neatly through the phenyl-diene gaps of the adjacent **B** and **A"** molecules. The adjacent phenyls prevent out-of-plane rotation by locking the pedals into gear. Thus, the BP isomerization resembles the mechanical operation of real bicycles in more than one way: the concerted rotation, and the mechanical rigidity in being locked into this rotation pathway. Secondly, the transition state geometry has no pyramidalization (D(C1-C2-C3-H2)=175.9°) which indicated to us that the isomerization might proceed directly to tt-DPB on the excited-state (an adiabatic pathway). Indeed, an excited-state reaction pathway was found that connects the $S_1$-minimum and tt-DPB $S_1$-minimum with an identical transition state geometry as Figure 4-6b, although they were computed independently. Therefore, this shows that the transition state bifurcates to lead to the tt-DPB $S_1$-minimum and *tt-transoid* MECI. At the $S_1$-minima, the excited-molecule can decay non-radiatively at the *tt-transoid* which involves pyramidalization, or decay radiatively through fluorescence.

MECI, however, are not enough to fully describe the photochemistry because non-radiative decay can occur at any point along the high-dimensional seam-space. In particular, regions of the seam are often found parallel to the adiabatic reaction coordinate and can contribute via motion orthogonal to the reaction coordinate[55]. Therefore, to explore the potential energy surface more widely for other possible bicycle pedal mechanisms, a minimum energy seam pathway was found connecting *cc-transoid* and *tt-transoid*. This seam runs roughly parallel to the adiabatic pathway and the seam saddle point occurs near the excited-state transition state (Figure 4-6c). Non-radiative decay through CI points before the seam saddle point are considered to return to cc-DPB, whereas all CI points at or after the saddle point lead to tt-DPB. With this seam established a complete picture of the relevant BP potential energy surface is now known and a summary of the important pathways, productive and non-productive, is given in Figure 4-7.

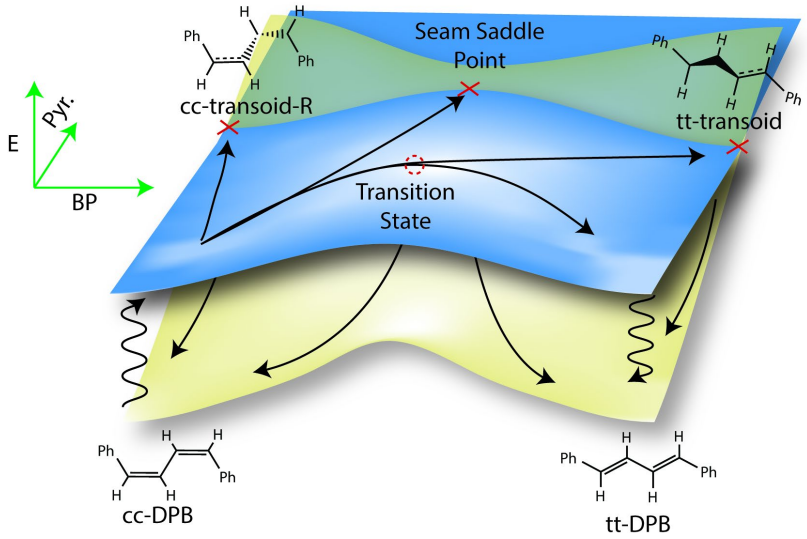



**Figure 4-7.** Summary of major excited-state reaction channels of conformer C, including productive and non-productive routes.

The results of Figure 4-7 allow us to examine the feasibility of the proposed photoisomerization pathways. Starting from photoexcited cc-DPB conformer **C**, the transition state energy leading to the bicycle pedal CI (24.4 kcal $mol^{-1}$) is below the vertical excitation energy (36.3 kcal $mol^{-1}$) suggesting that enough energy is available to overcome the barrier (at least at short timeframes following excitation). However, a significant region of the seam runs roughly parallel to the adiabatic coordinate and can contribute to non-productive decay back to reactants (i.e. an aborted OBF or HT). The productive formation of tt-DPB would only start around the seam-saddle point connecting *cc-transoid* and *tt-transoid* which is much higher in energy. Nevertheless, despite the seemingly poor energetics the mechanism appears the most probable. For example, if the phenyls are held stationary the motion would have to be confined to the diene portion of the molecule. The phenyl-diene gaps also enable unimpeded passage for hydrogen rotation. Therefore, at present the large barrier is considered an artifact of the lack of quantitative energies provided by CASSCF in computing these electronic states, or long-range relaxation of the crystal is necessary which is not captured in the four-layer model. Future studies will test the accuracy of this relatively simple level of theory against higher-accuracy methods.

### **4.4.4** Conformation Comparison

Now that the mechanism for cis-trans photoisomerization in cc-DPB have been outlined, the various conformers within the unit cell are compared to one another. Besides categorization into parallel and perpendicular units, the steric environment surrounding the central double bond of each conformer is different and could result in different isomerization efficiencies. If the conformations have different isomerization efficiencies, this could manifest as periods of fast and slow conversion because certain conformers will isomerize at lower photon fluxes[35]. Isomerization in the clockwise direction versus the counter-clockwise direction can also lead to differences in isomerization efficiencies, depending on conformer.

**Table 4-3.** Intermolecular interactions during the bicycle pedal isomerization for each unit in the clockwise and anti-clockwise direction. See **Figure 4-3** for the point-of-reference used to make this table.

| | | Neighbor | |
|---|---|---|---|
| Unit | Bicycle-Pedal Direction | Top Pedal* | Bottom Pedal* |
| **A** | clockwise | **B** | **B'** |
| | counter-clockwise | **D'** | **B** |
| **B** | clockwise | **A'** | **A** |
| | counter-clockwise | **A** | **D** |
| **C** | clockwise | **B** | **A''** |
| | counter-clockwise | **A''** | **B** |
| **D** | clockwise | **C'** | **C** |
| | counter-clockwise | **C** | **C'** |

*Top and bottom pedal refer to central hydrogen of diene with point-of-view in Figure 4-3.

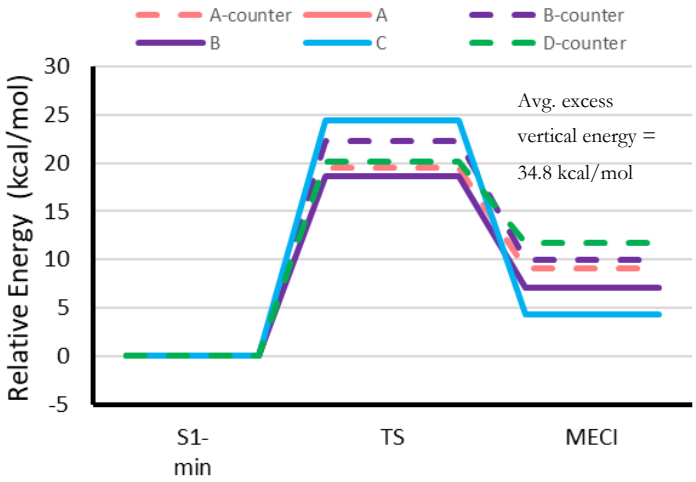



**Figure 4-8.** Comparison of bicycle pedal reaction paths to MECI for all conformers. The average vertical excess energy is over conformers A-D which have different microenvironments and energies.

As we have seen, the primary motion during the BP isomerization is the rotation of the central diene, the hydrogens of which rotate between the phenyl-diene gap of neighboring units. The steric interaction of the hydrogen pedal and the neighboring monomer is therefore of concern. To familiarize oneself with the different steric environments, consider the counter-clockwise rotations of **A** and **B** as an example. The diene of unit **A** and **B** are oriented similarly, however the steric interactions upon isomerization are different. The top pedal (hydrogen) of **A** slips between the gap of the phenyl and diene of **D'** and the bottom pedal

becomes perpendicular with the phenyl of **C** and rests near **B.** On the other hand, the bottom pedal of **B** slips between the phenyl diene gap of **D** and the top pedal sterically clashes with **A**. Similar analysis has been performed for all conformers in both rotation directions, and is present in **Table 4-3**.

With all the expected interactions tabulated, excited-state reaction pathways between the $S_1$-minimum of each conformer and the appropriate *tt-transoid*-CI were found for clockwise and counter-clockwise rotations (Figure 4-8). A range of about 6 kcal/mol is observed between all conformer: conformer **B** has the lowest energy barrier of 18.6 kcal/mol in the clockwise direction, and conformer **C** has the largest barrier of 24.5 kcal/mol. Clockwise BP rotation of conformer **B** causes the top pedal to pass first through the phenyl-diene gap of **A'** and then of **D**. The bottom pedal of conformer **B** skirts past **A** and does not pass directly through a phenyl-diene gap. In contrast when **B** rotates in the counter-clockwise direction, the bottom pedal rotates through **D** and as it approaches **A'** it relaxes before going through the gap. The top pedal of conformation **C** also does not pass through the phenyl diene gap of **B** during clockwise rotation. Overall, however, without quantitative confidence in the underlying model (see above) these differences cannot be rigorously analyzed. For all MECI, and RP diagrams and figures, consult the **Supporting Information**.

Nevertheless, to answer the question of whether the different conformations are responsible for different rates (fast and then slow), the answer is most likely "no". All conformers can rotate without major impediment and the major steric interaction provided by the crystal lattice is to restrict out-of-plane rotations of the phenyl, which is present in all conformers. Furthermore, additional experiments have observed that the related compound cis,trans-1,4-di-o-tolyl-1,3-butadiene (cc-DTB) has no conformational anisotropy[11] yet also displays stepwise isomerization kinetics. As a result, it is reasonable to conclude that conformational differences are not responsible for the observed stepwise reaction rate, but instead the effect is due to more complex changes as the crystal reorganizes.

## **4.6** Conclusions

The present computational investigation of cc-DPB provided substantial insight into the diverse OBF, HT, and BP mechanisms in the gas-phase and crystal states. The out-of-plane angle α and the side-to-side rotation angle φ were found to be useful parameters to explain why OBF and HT were prohibited in the crystal phase. Interestingly, the crystal phase can fundamentally change the seam topology, as seen in how the HT-1 conical intersection in the crystal phase resembles gas-phase seam saddle points. The dramatic influence of the steric environment is also evident in the BP isomerization, where the phenyls are locked in place and the diene must rotate in a smooth volume-conserving manner. The specific way the monomers clasp together is also important for ensuring an unimpeded passageway for the rotating diene. In contrast, cc-DPB is not able to isomerize via the BP in the gas phase because it lacks a rigid structure that is essential for ensuring a smooth rotation, without relaxation into non-BP geometries.

BP isomerization paths for unit-cell conformers revealed that all conformers isomerize in a similar fashion. This indicates that the fast and slow reaction kinetics observed in experiment are not likely to be a result of the different microenvironments in the unit cell, at least in the early stages of isomerization before the crystal reorganizes.

The growing string method was instrumental in locating the reaction pathways, conical intersections, and seam spaces throughout this study. In particular, the qualitative differences in isomerization pathways between the gas phase and solid state were straightforwardly delineated by GSM, allowing these mechanisms to be brought to light without undue reliance on prior chemical intuition. However, quantitative accuracy in describing the bicycle pedal has not yet been achieved. Therefore, in addition to providing relevant details of the BP mechanism we hope that this study can further motivate computational analysis of photochemical reactions in the solid state.

## **4.7** Data availability

MM parameters, molecular orbital diagrams, and reaction path profiles are provided in the Supporting Information.

## **4.8** References


(1) Liu, R. S. H.; Yang, L.-Y.; Liu, J. Mechanisms of Photoisomerization of Polyenes in Confined Media. From Organic Glasses to Protein Binding Cavities. *Photochem. Photobiol.* **2006**, 2–10. https://doi.org/10.1562/2006-01-27-RA-786.

(2) Gerwien, A.; Schildhauer, M.; Thumser, S.; Mayer, P.; Dube, H. Direct Evidence for Hula Twist and Single-Bond Rotation Photoproducts. *Nat. Commun.* **2018**, *9* (1), 1–9. https://doi.org/10.1038/s41467-018-04928-9.

(3) Redwood, C.; Bayda, M.; Saltiel, J. Photoisomerization of Pre- and Provitamin D3 in EPA at 77 K: One-Bond-Twist, Not Hula-Twist. *J. Phys. Chem. Lett.* **2013**, *4* (5), 716–721. https://doi.org/10.1021/jz302108c.

(4) Fuß, W.; Kosmidis, C.; Schmid, W. E.; Trushin, S. A. The Photochemical Cis-Trans Isomerization of Free Stilbene Molecules Follows a Hula-Twist Pathway. *Angew. Chemie - Int. Ed.* **2004**, *43* (32), 4178–4182. https://doi.org/10.1002/anie.200454221.

(5) Liu, R. S. H.; Hammond, G. S. The Case of Medium-Dependent Dual Mechanisms for Photoisomerization: One-Bond-Flip and Hula-Twist. *Proc. Natl. Acad. Sci.* **2000**, *97* (21), 11153–11158. https://doi.org/10.1073/pnas.210323197.

(6) Fuß, W. Previtamin D: Z-E Photoisomerization via a Hula-Twist Conical Intersection. *Phys. Chem. Chem. Phys.* **2019**, *21* (13), 6776–6789. https://doi.org/10.1039/c9cp00500e.

(7) Saltiel, J.; Krishna, T. S. R.; Clark, R. J. Photoisomerization of Cis,Cis-1,4-Diphenyl-1,3-Butadiene in the Solid State: The Bicycle-Pedal Mechanism. *J. Phys. Chem. A* **2006**, *110* (5), 1694–1697. https://doi.org/10.1021/jp056700d.

(8) Saltiel, J.; Krishna, T. S. R.; Laohhasurayotin, S.; Fort, K.; Clark, R. J. Photoisomerization of Cis , Cis - to Trans , Trans -1,4-Diaryl-1,3-Butadienes in the Solid State: The Bicycle-Pedal Mechanism. *J. Phys. Chem. A* **2008**, *112* (2), 199–209. https://doi.org/10.1021/jp077342c.

(9) Furukawa, D.; Kobatake, S.; Matsumoto, A. Direct Observation of Change in the Molecular Structure of Benzyl (Z,Z)-Muconate during Photoisomerization in the Solid State. *Chem. Commun.* **2008**, No. 1, 55–57. https://doi.org/10.1039/b714792a.

(10) Odani, T.; Matsumoto, A.; Sada, K.; Miyata, M. One-Way EZ-Isomerization of Bis(n-Butylammonium) (Z,Z)-Muconate under Photoirradiation in the Crystalline State. *Chem. Commun.* **2001**, *19* (c), 2004–2005. https://doi.org/10.1039/b106155k.

(11) Saltiel, J.; Bremer, M. A.; Laohhasurayotin, S.; Krishna, T. S. R. Photoisomerization Ofcis,Cis-Andcis,Trans-1,4-Di-o-Tolyl-1,3-Butadiene in Glassy Media at 77 K: One-Bond-Twist and Bicycle-Pedal Mechanisms. *Angew. Chemie Int. Ed.* **2008**, *47* (7), 1237–1240. https://doi.org/10.1002/anie.200704465.

(12) Tong, F.; Al-Haidar, M.; Zhu, L.; Al-Kaysi, R. O.; Bardeen, C. J. Photoinduced Peeling of Molecular Crystals. *Chem. Commun.* **2019**, *55* (26), 3709–3712. https://doi.org/10.1039/c8cc10051a.

(13) Sonoda, Y.; Kawanishi, Y.; Tsuzuki, S.; Goto, M. Crystalline-State Z,E-Photoisomerization of a Series of (Z,E,Z)-1,6-Diphenylhexa-1,3,5-Triene 4,4'-Dicarboxylic Acid Dialkyl Esters. Chain Length Effects on the Crystal Structure and Photoreactivity. *J. Org. Chem.* **2005**, *70* (24), 9755–9763. https://doi.org/10.1021/jo051137g.

(14) Saltiel, J.; Redwood, C. E.; Laohhasurayotin, K.; Samudrala, R. Photochemistry of the 1,6-Dideuterio-1,3,5-Hexatrienes in Solution: Efficient Terminal Bond Photoisomerization in One-Bond-Twist and Bicycle Pedal Ways. *J. Phys. Chem. A* **2018**, *122* (43), 8477–8489. https://doi.org/10.1021/acs.jpca.8b08288.

(15) Jung, Y. O.; Lee, J. H.; Kim, J.; Schmidt, M.; Moffat, K.; Šrajer, V.; Ihee, H. Volume-Conserving Trans–cis Isomerization Pathways in Photoactive Yellow Protein Visualized by Picosecond X-Ray Crystallography. *Nat. Chem.* **2013**, *5* (3), 212–220. https://doi.org/10.1038/nchem.1565.
(16) A, W. Bicycle-Pedal Model for First Step in Vision Process. *Nature* **1976**, *260*, 679.
(17) Saltiel, J.; Redwood, C. E. Photochemistry of the 1,4-Diphenyl-1,3-Butadienes in Ethanol. Trapping Conical Intersections. *J. Phys. Chem. A* **2016**, *120* (18), 2832–2840. https://doi.org/10.1021/acs.jpca.6b02330.
(18) Saltiel, J.; Bremer, M. A.; Laohhasurayotin, S.; Krishna, T. S. R. Photoisomerization of Cis,Cis- and Cis,Trans-1,4-Di-o-Tolyl-1,3-Butadiene in Glassy Media at 77 K: One-Bond-Twist and Bicycle-Pedal Mechanisms. *Angew. Chemie - Int. Ed.* **2008**, *47* (7), 1237–1240. https://doi.org/10.1002/anie.200704465.
(19) Olivucci, M.; Ragazos, I. N.; Bernardi, F.; Robb, M. A. A Conical Intersection Mechanism for the Photochemistry of Butadiene. A MC-SCF Study. *J. Am. Chem. Soc.* **1993**, *115* (9), 3710–3721. https://doi.org/10.1021/ja00062a042.
(20) Kuhlman, T. S.; Glover, W. J.; Mori, T.; Moller, K. B.; Martínez, T. J. Between Ethylene and Polyenes - The Non-Adiabatic Dynamics of Cis-Dienes. *Faraday Discuss.* **2012**, *157*, 193–212. https://doi.org/10.1039/c2fd20055d.
(21) Glover, W. J.; Mori, T.; Schuurman, M. S.; Boguslavskiy, A. E.; Schalk, O.; Stolow, A.; Martínez, T. J. Excited State Non-Adiabatic Dynamics of the Smallest Polyene, Trans 1,3-Butadiene. II. Ab Initio Multiple Spawning Simulations. *J. Chem. Phys.* **2018**, *148* (16). https://doi.org/10.1063/1.5018130.
(22) Ogliaro, F.; Wilsey, S.; Bearpark, M. J.; Sardo-Infirri, S. Interpreting the Excited States and Decay Processes of Bichromophoric 1-Phenyl-1,3-Butadiene Using CASSCF Calculations. *Mol. Phys.* **2006**, *104* (5–7), 1017–1032. https://doi.org/10.1080/00268970500418307.
(23) Norton, J. E.; Houk, K. N. H/Vinyl Conical Intersections of Hexatrienes Related to the Hula-Twist Photoisomerization. *Mol. Phys.* **2006**, *104* (5–7), 993–1008. https://doi.org/10.1080/00268970500417606.
(24) Sampedro Ruiz, D.; Cembran, A.; Garavelli, M.; Olivucci, M.; Fuß, W. Structure of the Conical Intersections Driving the Cis–trans Photoisomerization of Conjugated Molecules. *Photochem. Photobiol.* **2002**, *76* (6), 622. https://doi.org/10.1562/0031-8655(2002)076<0622:SOTCID>2.0.CO;2.
(25) Wilsey, S.; Houk, K. N. H-Vinyl Conical Intersections for Dienes: A Mechanism for the Photochemical Hula Twist. *Photochem. Photobiol.* **2002**, *76* (6), 616–621. https://doi.org/10.1562/0031-8655(2002)076<0616:HVCIFD>2.0.CO;2.
(26) Zimmerman, H. E.; Sebek, P.; Zhu, Z. Ab Initio Computations of Reacting Species in Crystal Lattices; Mechanistic and Exploratory Organic Photochemistry [17]. *Journal of the American Chemical Society*. 1998, pp 8549–8550. https://doi.org/10.1021/ja980042b.
(27) Zimmerman, H. E.; Alabugin, I. V; Smolenskaya, V. N. Experimental and Theoretical Host-Guest Photochemistry; Control of Reactivity with Host Variation and Theoretical Treatment with a Stress Shaped Reaction Cavity; Mechanistic and Exploratory Organic Photochemistry. *Tetrahedron* **2000**, *56* (36), 6821–6831. https://doi.org/10.1016/S0040-4020(00)00504-4.
(28) Zimmerman, H. E.; Nesterov, E. E. Development of Experimental and Theoretical Crystal Lattice Organic Photochemistry : The Quantitative Cavity . Mechanistic and Exploratory Organic. **2002**, *35* (2), 77–85. https://doi.org/10.1021/ar000210g.

(29) Ruiz-Barragan, S.; Morokuma, K.; Blancafort, L. Conical Intersection Optimization Using Composed Steps Inside the ONIOM(QM:MM) Scheme: CASSCF:UFF Implementation with Microiterations. *J. Chem. Theory Comput.* **2015**, *11* (4), 1585–1594. https://doi.org/10.1021/acs.jctc.5b00004.
(30) Presti, D.; Labat, F.; Pedone, A.; Frisch, M. J.; Hratchian, H. P.; Ciofini, I.; Cristina Menziani, M.; Adamo, C. Modeling Emission Features of Salicylidene Aniline Molecular Crystals: A QM/QM' Approach. *J. Comput. Chem.* **2016**, *37* (9), 861–870. https://doi.org/10.1002/jcc.24282.
(31) Wilbraham, L.; Adamo, C.; Labat, F.; Ciofini, I. Electrostatic Embedding to Model the Impact of Environment on Photophysical Properties of Molecular Crystals: A Self-Consistent Charge Adjustment Procedure. *J. Chem. Theory Comput.* **2016**, *12* (7), 3316–3324. https://doi.org/10.1021/acs.jctc.6b00263.
(32) Rivera, M.; Dommett, M.; Crespo-Otero, R. ONIOM(QM:QM') Electrostatic Embedding Schemes for Photochemistry in Molecular Crystals. *J. Chem. Theory Comput.* **2019**, *15* (4), 2504–2516. https://doi.org/10.1021/acs.jctc.8b01180.
(33) Aldaz, C.; Wang, L.-P.; Martinez, T. J.; Zimmerman, P. M. A Linear Scaling Growing String Method with Correlated Motions, In Preparation. **2020**.
(34) Wang, L. P.; Song, C. Geometry Optimization Made Simple with Translation and Rotation Coordinates. *J. Chem. Phys.* **2016**, *144* (21), 214108. https://doi.org/10.1063/1.4952956.
(35) Zimmerman, H. E.; Nesterov, E. E. Crystal Lattice Photochemistry Often Proceeds in Discrete Stages. Mechanistic and Exploratory Organic Photochemistry 1 , 2. *Org. Lett.* **2000**, *2* (8), 1169–1171. https://doi.org/10.1021/ol0057838.
(36) Zimmerman, H. E.; Alabugin, I. V; Chen, W.; Zhu, Z.; June, R. V. Dramatic Effects of Crystal Morphology on Solid State Reaction Course; Control by Crystal Disorder; Mechanistic and Exploratory Organic Photochemistry. *J. Am. Chem. Soc.* **1999**, *121* (6), 11930–11931. https://doi.org/10.1021/ja992208u.
(37) Aldaz, C. https://github.com/ZimmermanGroup/pyGSM.
(38) Zimmerman, P. M. Growing String Method with Interpolation and Optimization in Internal Coordinates: Method and Examples. *J. Chem. Phys.* **2013**, *138* (18), 1–11. https://doi.org/10.1063/1.4804162.
(39) Zimmerman, P. M. Single-Ended Transition State Finding with the Growing String Method. *J. Comput. Chem.* **2015**, *36* (9), 601–611. https://doi.org/10.1002/jcc.23833.
(40) Dewyer, A. L.; Zimmerman, P. M. Finding Reaction Mechanisms, Intuitive or Otherwise. *Org. Biomol. Chem.* **2017**, 501–504. https://doi.org/10.1039/C6OB02183B.
(41) Aldaz, C.; Kammeraad, J. A.; Zimmerman, P. M. Discovery of Conical Intersection Mediated Photochemistry with Growing String Methods. *Phys. Chem. Chem. Phys.* **2018**, *20* (43), 27394–27405. https://doi.org/10.1039/C8CP04703K.
(42) Baker, J.; Kessi, A.; Delley, B. The Generation and Use of Delocalized Internal Coordinates in Geometry Optimization. *J. Chem. Phys.* **1996**, *105* (1996), 192–212. https://doi.org/10.1063/1.471864.
(43) Ufimtsev, I. S.; Martinez, T. J. Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *J. Chem. Theory Comput.* **2009**, *5* (10), 2619–2628. https://doi.org/10.1021/ct9003004.
(44) Hohenstein, E. G.; Luehr, N.; Ufimtsev, I. S.; Martínez, T. J. An Atomic Orbital-Based Formulation of the Complete Active Space Self-Consistent Field Method on Graphical Processing Units. *J. Chem. Phys.* **2015**, *142* (22). https://doi.org/10.1063/1.4921956.

(45) Isborn, C. M.; Götz, A. W.; Clark, M. A.; Walker, R. C.; Martínez, T. J. Electronic Absorption Spectra from MM and Ab Initio QM/MM Molecular Dynamics: Environmental Effects on the Absorption Spectrum of Photoactive Yellow Protein. *J. Chem. Theory Comput.* **2012**, *8* (12), 5092–5106. https://doi.org/10.1021/ct3006826.

(46) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L. P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**, *13* (7), 1–17. https://doi.org/10.1371/journal.pcbi.1005659.

(47) Levine, B. G.; Coe, J. D.; Martínez, T. J. Optimizing Conical Intersections without Derivative Coupling Vectors: Application to Multistate Multireference Second-Order Perturbation Theory (MS-CASPT2) †. *J. Phys. Chem. B* **2008**, *112* (2), 405–413. https://doi.org/10.1021/jp0761618.

(48) Levine, B. G.; Martínez, T. J. Isomerization through Conical Intersections. *Annu. Rev. Phys. Chem.* **2007**, *58*, 613–634. https://doi.org/10.1146/annurev.physchem.57.032905.104612.

(49) Levine, B. G.; Martínez, T. J.; Martı, T. J. Ab Initio Multiple Spawning Dynamics of Excited Butadiene: Role of Charge Transfer. *J. Phys. Chem. A* **2009**, *113* (46), 12815–12824. https://doi.org/10.1021/jp907111u.

(50) Saltiel, J.; Krishna, T. R. S.; Laohhasurayotin, K.; Ren, Y.; Phipps, K.; Davis, P. H.; Yee, W. A. Medium Effects on the Direct Cis-Trans Photoisomerization of 1,4-Diphenyl-1,3-Butadiene in Solution. *J. Phys. Chem. A* **2011**, *115* (11), 2120–2129. https://doi.org/10.1021/jp111482m.

(51) The Gas-Phase MECI Are below the S1 Minimum like the MECI for Butadiene, but Unlike the MECI for Phenylbutadiene Which Are above the S1 Minimum.

(52) Wang, B.; Wang, X.; Wang, W.; Liu, F. Exploring the Mechanism of Fluorescence Quenching and Aggregation-Induced Emission of a Phenylethylene Derivative by QM (CASSCF and TDDFT) and ONIOM (QM:MM) Calculations. *J. Phys. Chem. C* **2016**, *120* (38), 21850–21857. https://doi.org/10.1021/acs.jpcc.6b07963.

(53) Peng, X. L.; Ruiz-Barragan, S.; Li, Z. S.; Li, Q. S.; Blancafort, L. Restricted Access to a Conical Intersection to Explain Aggregation Induced Emission in Dimethyl Tetraphenylsilole. *J. Mater. Chem. C* **2016**, *4* (14), 2802–2810. https://doi.org/10.1039/c5tc03322e.

(54) Sumita, M.; Saito, K. Theoretical Study on Hula-Twist Motion of Penta-2,4-Dieniminium on the S1 Surface under Isolated Condition by the Complete Active Space Self-Consistent Field Theory. *Chem. Phys. Lett.* **2006**, *424* (4–6), 374–378. https://doi.org/10.1016/j.cplett.2006.04.093.

(55) Boggio-Pasqua, M.; Ravaglia, M.; Bearpark, M. J.; Garavelli, M.; Robb, M. a. Can Diarylethene Photochromism Be Explained by a Reaction Path Alone? A CASSCF Study with Model MMVB Dynamics. *J. Phys. Chem. A* **2003**, *107* (50), 11139–11152. https://doi.org/10.1021/jp036862e.

# **Chapter 5** Experimental and Theoretical Characterization of Ultrafast Water-Soluble Photochromic Photoacids

This experimental work done in this chapter was performed in collaboration with Drs. Ted Wiley, Nicholas Miller, and Professor Roseanne Sension.

## **5.1** Introduction

The photochromic properties of spiropyran have long been utilized for light activated molecular devices[1,2]. Irradiation of spiropyran (SP) causes reversible ring opening to form merocyanine (MC). Likewise, merocyanine can reversibly ring close to form spiropyran. These two species are chemically and spectroscopically distinct[3,4]. As a result, spiropyran has inspired technology ranging from optical data storage[5,6], photoswitchable coatings and membranes[7–10], and light controlled fuel cells[11], to DNA photoswitches[12,13]. More recently, water-soluble merocyanine/spiropyran (MC/SP) photochromic photoacids have been developed[14–16]. Their utility for studies of proton transfer processes in biological assays has garnered interest[14]. The ability of the MC/SP photoacids to catalyze esterification reactions[16], ring-opening polymerizations[17], and reversible self-assembly[18–20], to drive a molecular shuttle[21], and to modulate the volume of pH sensitive polymers[16] was demonstrated in earlier work.

Two classes of merocyanine/spiropyran (MC/SP) photochromic switches (Figure 5-1) have been developed that experience a large drop in $pK_a$ after photoexcitation[14–16,22,23]. The photochemical mechanism of 2-[(E)-2-(2-hydroxyphenyl)ethenyl]-3,3-dimethyl-1-(3-sulfopropyl)-3H-indol-1-ium hereafter referred to as "phenylhydroxy-MCH" has been proposed to proceed through a trans-to-cis photoisomerization mechanism (Figure 5-1c) following excitation in the visible absorption band[16]. The photochemical mechanism of 2-[(E)-2-(1H-indazol-7-yl)ethenyl]-3-(3-sulfopropyl)-1,3-benzothiazol-3-ium hereafter referred to as "indazole-MCH" has been proposed to proceed through similar trans-to-cis photoisomerization mechanism (Figure 5-1d) with tautomerization preceding loss of a proton and formation of spiropyran[14]. However, these mechanisms have not yet been investigated on an ultrafast timescale or compared with theoretical simulations. As a result, the early sequences in

deprotonation and the formation of spiropyran are unknown, and this may limit the refinement for effectiveness of these photoacids in certain applications.

In this study, transient absorption spectroscopy and quantum mechanical computations are applied to the study of the two photochromic photoacids in buffered water and DMSO solvent. The investigation of these photoacids reveals the early time dynamics of these photoacids including how the trans-acid isomerizes to the cis-form, and the timescales for deprotonation and ring-closing. Furthermore, although not definitive evidence, the TA spectra are best fit by considering a one-bond flip mechanism, which suggests that the hula-twist is not a dominant pathway in solution phase. Lastly, the decrease in $pK_a$ upon isomerization is investigated to benchmark quantum chemical methods for their accuracy. Overall, these results provide a detailed picture of the operation of reversible photoacids which are useful for the refinements and development of photoacid applications.

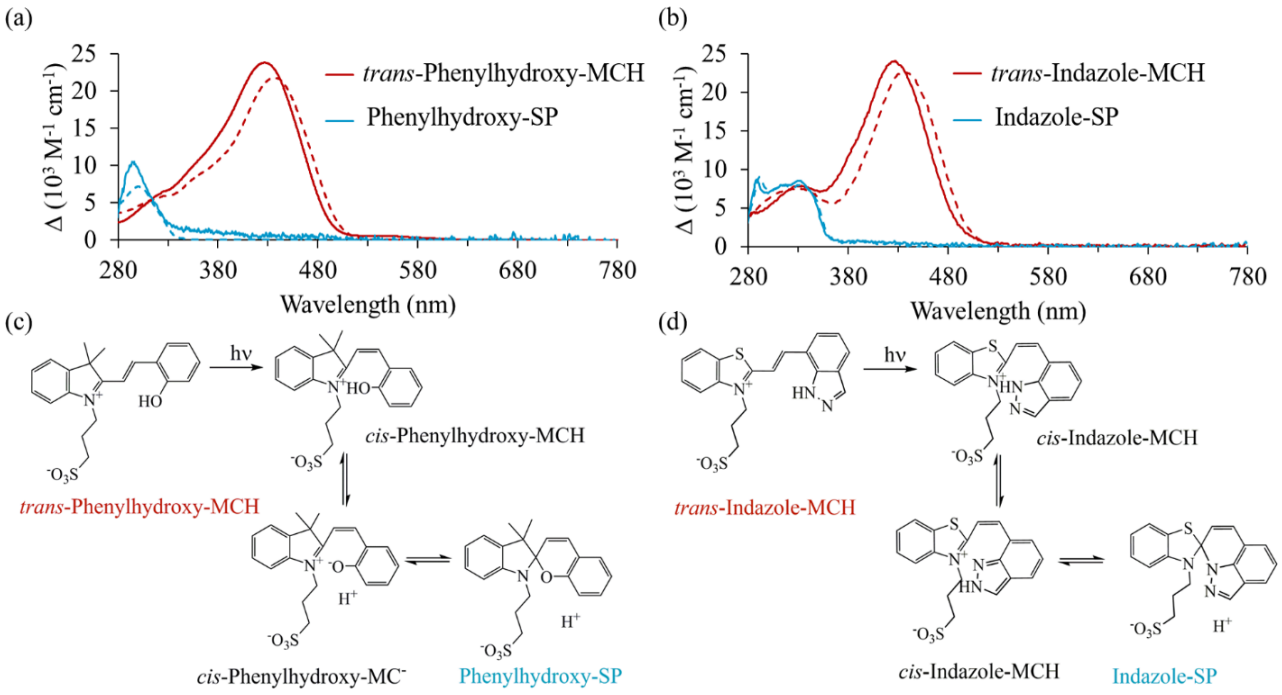



**Figure 5-1** (a) Phenylhydroxy-MCH absorption spectrum in aqueous solution buffered at pH 5.5 (red, solid) and in DMSO (red, dash). Phenylhydroxy SP in aqueous solution buffered at pH 5.5 (blue, solid) and in DMSO (blue, dash). (b) Indazole-MCH absorption spectrum in aqueous solution buffered at pH 7.4 (red, solid) and in DMSO (red, dash). Indazole SP in aqueous solution buffered at pH 7.4 (blue, solid) and in DMSO (blue, dash). (c) Proposed scheme of photoswitching by trans-phenylhydroxy-merocyanine (MCH) to phenylhydroxy-spiropyran (SP).[16] (d) Proposed scheme of photoswitching by trans-indazole-merocyanine (MCH) to indazole-spiropyran (SP).[14] Spectroscopic results in this figure courtesy of Dr. Ted Wiley.

## **5.2** Methods

DFT calculations were performed using the Q-CHEM suite of packages.[24] Geometry optimizations of all e.g. conformations (SP, TTC, TTT, CTC, CTT, CCC, TCT, CCT, and TCC, see below) and their conjugate base were performed with the B3LYP/6-31+G** and wB97xD/6-31+G** level of theory and solvation effects were included with the integral equation formalism polarizable continuum model (IEF-PCM)[25–27]. The propyl sulfonate group of the photoacids was truncated to a methyl group to facilitate geometry optimization. The IEF-PCM calculations used the default Bondi radii. The existence of the stationary points was checked by inspection of the Hessian matrix eigenvalues. All minimum energy geometries have zero imaginary frequencies unless otherwise noted; see supporting information Table D-1 through Table D-5. Enthalpy and entropy components were computed at the same level of theory using the rigid-rotor harmonic oscillator approximation. The entropy component was adjusted by increasing the low vibrational frequencies (<50 $cm^{-1}$) to 50 $cm^{-1}$. Single-point corrections were performed at higher levels of theory including B3LYP/G3LARGE, wB97xD/G3LARGE, DLPNO-CCSD(T)/cc-pvtz[28], and RI-MP2/cc-pvtz. The DLPNO-CCSD(T) calculations were performed with ORCA[29] and use the SMD solvation model if specified. All Gibbs Free energy were calculated at 292.15 K (19 °C). Vertical excitations were calculated at the optimized geometries using TDDFT within the Tamm-Damcoff approximation and the B3LYP/6-3+1G** level of theory[30]. Non-equilibrium solvation of the excited states was taken into account using the linear response (LR) approach[31]. Transition states were calculated using the growing string method[32].

## **5.3** Results

### **5.3.1** Simulations of ground state trans-photoacids.

Six *trans* conformations of the ground state phenylhydroxy-MCH and eight *trans* conformations of indazole-MCH result from rotation of the two functional groups around the single bonds on either side of the central carbon-carbon double bond (Figure 5-2). The conformers are given names to specify the cis/trans relationship between the functional groups and the central double bond and are labeled TTT, CTT, TTC, or CTC. The conformations can exist in two helicities with slightly different energies; M and P for rotation about the two single bonds connecting the central double bond, or in a "flat" rotation, which has less than 10-degree rotation. For example, in Figure 5-2, the trans-acid conformations are helical with respect the single-bond connecting indoline and the central double

bond but are flat with respect to the single-bond connecting phenylhydroxy and the central double bond.

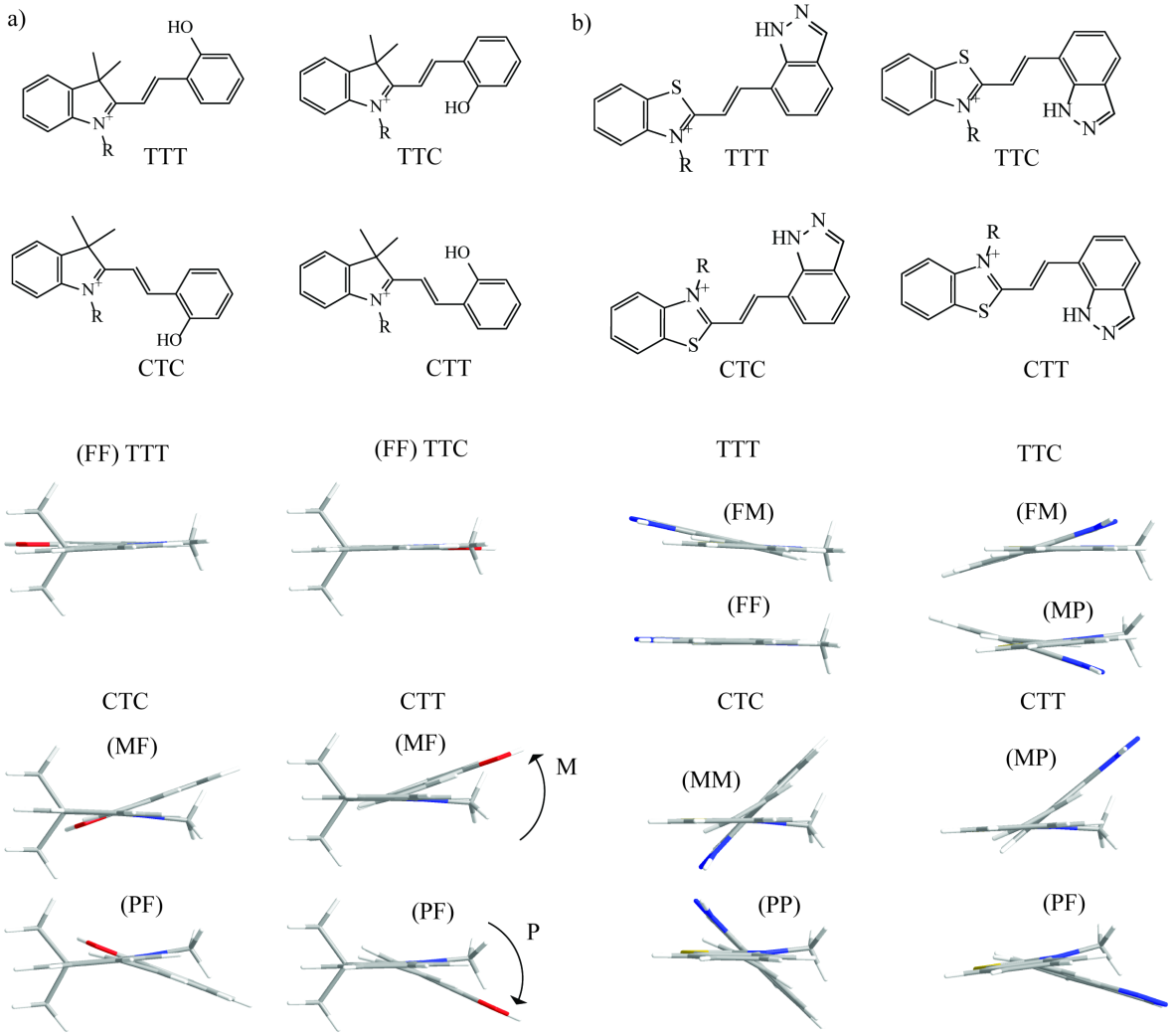



**Figure 5-2** Calculated stable conformations of (a) *trans*-phenylhydroxy-MCH and (b) *trans*-imidazole-MCH. Structures in the lower panels are aligned to look along the indazole ring of *trans*-phenylhydroxy-MCH or the benzothiazole ring of *trans*-imidazole-MCH and highlight the helicity of some conformations. The labels M (counterclockwise) and P (clockwise) refers to the rotation around the single bonds connecting to the central double bond. The label F means the rotation is less than 10 degrees.

The ground state free energies of these conformations were calculated using DFT with a polarized continuum model for water and DMSO (see Figure 5-2). The results are summarized in **Table 5-1** (with more details in the supporting information). The equilibrium solution is predicted to be a complex mixture of conformations, dominated by TTT and TTC. The CTC and CTT conformations

account for as much as 13% of the populations for *trans*-phenylhydroxy-MCH, and less than 5% for trans-indazole-MCH.

**Table 5-1** Relative populations for conformations of trans-phenylhydroxy-MCH and trans-indazole-MCH calculated with wb97xD/6-31++g** and B3lYP/6-31++g** in water and DMSO solvent.

| Conformer | B3LYP (H2O) | B3LYP(DMSO) | wB97xD (H2O) | wB97xD(DMSO) |
|---|---|---|---|---|
| *trans*-phenylhydroxy-MCH | | | | |
| (FF) TTT | 63.7% | 63.3% | 78.2% | 81.9% |
| (FF) TTC | 23.0% | 23.4% | 9.5% | 8.6% |
| (MF) CTT | 3.8% | 3.8% | 1.1% | 1.1% |
| (PF) CTT | 5.8% | 5.8% | 4.8% | 7.0% |
| (PF) CTC | 2.2% | 2.2% | 1.0% | 1.0% |
| (MF) CTC | 1.5% | 1.5% | 5.3% | 0.4% |
| *trans*-indazole-MCH | | | | |
| (FF) TTT | 28.8% | 31.1% | 23.0% | 23.0% |
| (FM) TTT | 31.3% | 29.9% | 39.7% | 39.5% |
| (MP) TTC | 25.0% | 19.7% | 19.8% | 20.0% |
| (FM) TTC | 14.2% | 18.6% | 12.6% | 12.8% |
| (PF) CTT | 0.2% | 0.2% | 1.8% | 1.7% |
| (MF) CTT | 0.2% | 0.2% | 0.9% | 0.9% |
| (PP) CTC | 0.1% | 0.1% | 1.0% | 1.0% |
| (MM) CTC | 0.1% | 0.1% | 1.1% | 1.1% |

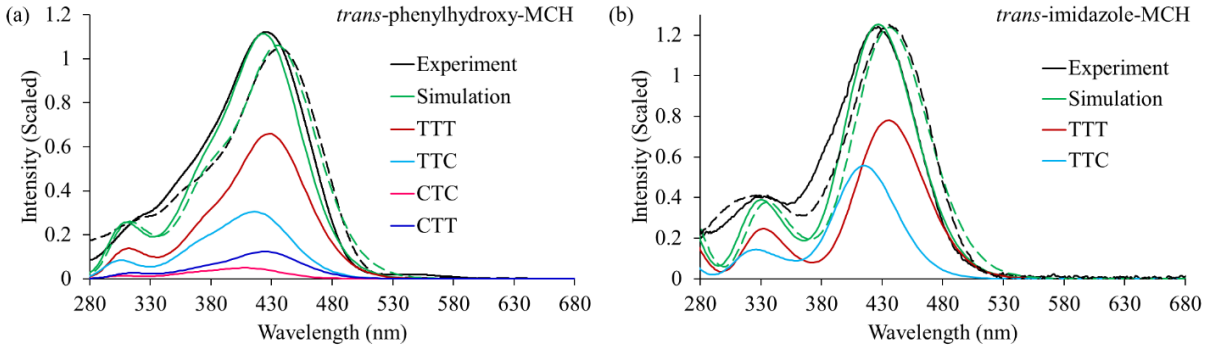



**Figure 5-3** Simulated and experimental absorption spectra the photoacids. (a) Spectra of trans-phenylhydroxy-MCH in pH 5.5 aqueous solution (solid lines) and in anhydrous DMSO (dashed lines). The separate calculated contributions for the conformers are also plotted for water weighted by population. The spectra of the helical conformations of CTC and CTT are essentially identical and the sum is plotted in the figure. (b) Spectra of trans-indazole-MCH in pH 7.4 aqueous solution (solid lines) and in anhydrous DMSO (dashed lines). The spectra have been blue-shifted 500 cm-1 in water and 200 $cm^{-1}$ in DMSO to match the peak of the major transition. The separate calculated contributions for the significant conformers are also plotted for water weighted by population. The spectra of the helical conformations of TTT and TTC are nearly identical and the sum is plotted in the figure. Experimental spectra courtesy of Dr. Ted Wiley.

The UV visible transitions of the stable conformers were calculated using TD-DFT methods (Figure 5-3). The stick spectra were convoluted with a Gaussian lineshape of 0.42 eV full width at half maximum (FWHM) for comparison with experimental data. The simulated spectra of *trans*-phenylhydroxy-MCH and *trans*-indazole-MCH with the different conformations weighted by the

populations are in good agreement with the experimental spectra (Figure 5-3), however in the latter a blue-shift was required to match the observed transition energies.

### **5.3.2** Photoisomerization Pathways

The excited state isomerization of the *trans*-photoacids could follow several different pathways identified as one-bond twist around the central double bond or hula twist around the carbon atom on either side of the central double bond (Figure 5-4). Given an initial distribution of *trans* conformations, these pathways will result in different distributions of *cis*-acids. Single-bond rotation and cyclization reaction paths for the cis-conjugate base isomers were found with the B3LYP/6-31+G* level of theory and are given in Figure 5-4. A single-elementary pathway was found connecting TCT-phenylhydroxy-MC$^-$ and SP-phenylhydroxy, and a single-bond rotation pathway followed by ring-closing pathway was found CCT- to SP-phenylhydroxy. The calculated values for ring cyclization from CCC to SP are in close agreement with values found before[33]. A similar picture exists for indazole-MCH (Figure D-1), but *cis*-indazole-MCH has been hypothesized to tautomerize prior to ring-closing.

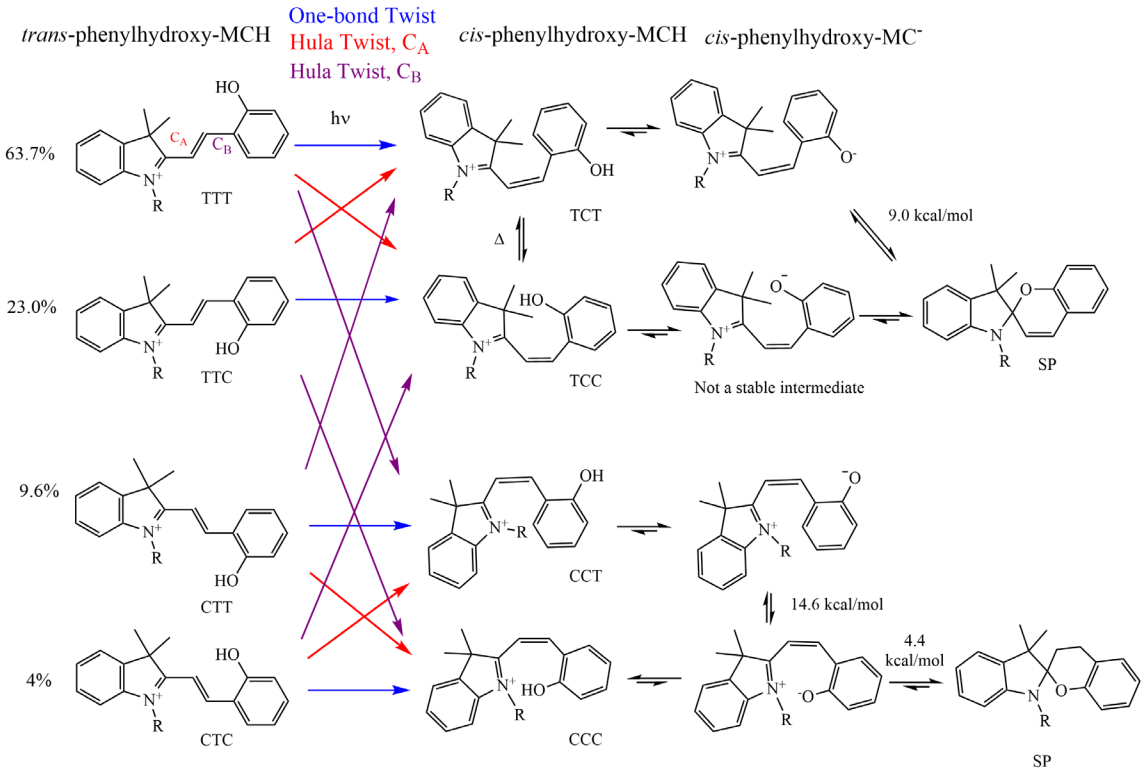



**Figure 5-4** Conformations and reaction pathways for phenylhydroxy-photoacids. R is the sulfonate tail, replaced by $CH_3$. Single-bond rotation and cyclization pathways were calculated using the B3LYP/6-31++G** IEF-PCM (H2O) level of theory. A similar picture exists for indazole-MCH except for the inclusion of the cis-indazole tautomer which is higher energy, see Figure D-1.

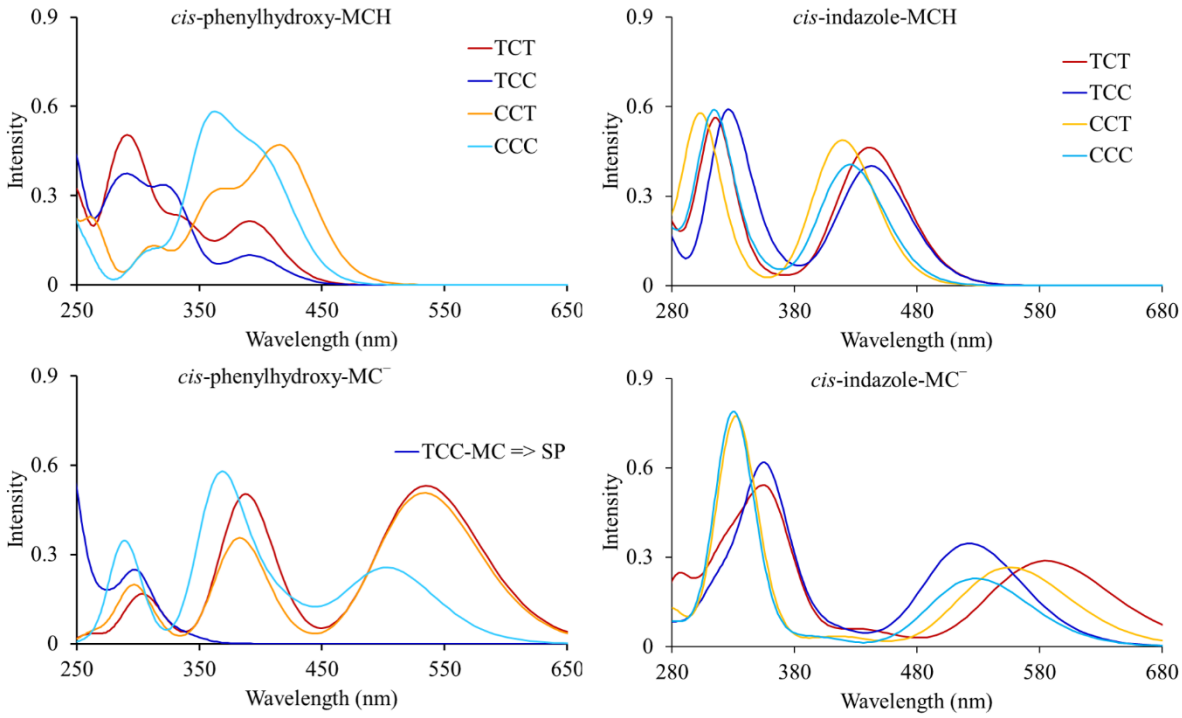



**Figure 5-5** Calculated absorption spectra for the four primary conformations of the cis photoacids. Left: *cis*-phenylhydroxy-MCH and the conjugate base *cis*-phenylhydroxy-MC⁻. The TCC conjugate base is unstable and optimizes to spiropyran. The absolute intensity scale is arbitrary, but identical to the scale used in **Figure 5-3**.

TD-DFT calculations were performed on the possible protonated and deprotonated cis-isomers to compare to the UV-visible spectra (Figure 5-5). The TD-DFT vertical transition energy and oscillator strengths were broadened in the same way as the *trans*-conformers (see above) to obtain the absorption spectra. The cis-isomer transitions are weaker than the corresponding *trans*-conformers which has a peak intensity ca. 1.1 for *trans*-phenylhydroxy-MCH and 1.3 to 1.4 *trans*-indazole-MCH on the same scale (Figure 5-3).

Phenylhydroxy-SP has no visible absorption bands, and the 3.5 ns TA spectra collected in citrate buffer has a prominent positive absorption peak from 475 – 600 nm (Figure D-2). The only structures with significant absorption intensity in the range of 475 – 600 nm are deprotonated *cis* conformations. Thus, the intermediate photoproduct observed following excitation of *trans*-phenylhydroxy-MCH at pH 5.5 is likely the TCT-phenylhydroxy-MC⁻ conformation. The weaker red photoproduct absorption observed following excitation of *trans*-indazole-MCH is also assigned to the *cis* conjugate base (Figure D-6).

## **5.4** Discussion

### **5.4.1** Primary Photolysis Products.

The results of the transient absorption experiments (supporting information) revealed that photoproduct with red-shifted absorption is formed on an ultrafast timescale and this formation is solvent-dependent. Starting with pH 5.5, if the initial population is a mixture of the *trans*-phenylhydroxy-MCH conformations with the calculated weights from the wB97xD level of theory (**Table 5-1**), a good estimate of the photoproduct difference spectrum is obtained for the one-bond flip mechanism followed by deprotonation (Figure 5-6 top). Similar agreement is also obtained with the B3LYP populations and hula-twist at $C_B$ (Figure D-7). However, agreement with hula twist at $C_A$ is not as good because the anticipated percentage of the visible absorbing TCT-MC$^-$ and CCT-MC$^-$ conformations is much smaller. Consequently, the most likely conformer following isomerization is TCT-MCH. The TCC-MCH constitutes only 10-20% of the population, assuming equal isomerization quantum yields for all the conformers.

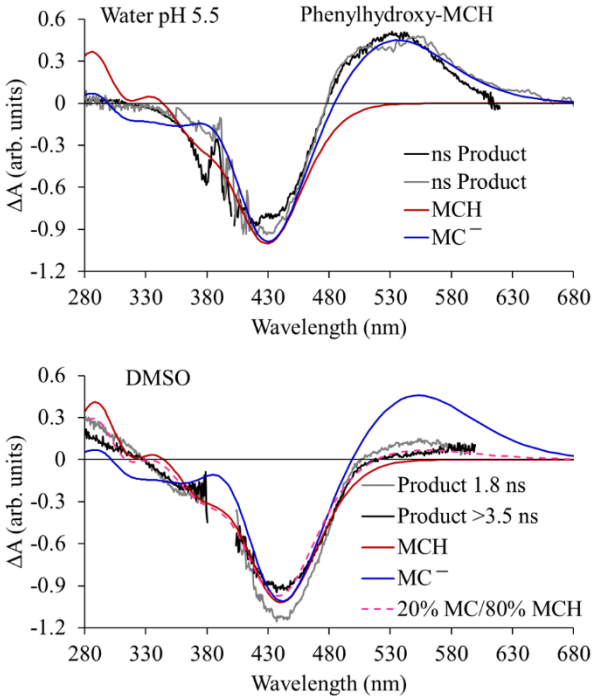



**Figure 5-6** Comparison of the measured (black, gray lines) and simulated product difference spectra for *trans* to *cis* isomerization of the phenylhydroxy-MCH photoacid in both water and DMSO. In water the product spectra come from two independent measurements, in DMSO they represent the two long-lived spectral components. The blue lines represent the simulated difference spectrum for the formation of *cis*-phenylhydroxy-MC$^-$ via a one bond twist mechanism. The red lines represent the simulated difference spectra for the formation of *cis*-phenylhydroxy-MCH via a one bond twist mechanism. The dashed pink line in DMSO assumes 20% deprotonation. Spectroscopic results in this figure courtesy of Dr. Ted Wiley.

TCT-MC- however, is not directly reactive to SP and is meta-stable. The rotational barrier for isomerization of TCT-MC$^-$ → TCC-MC$^-$ is 9.0 kcal/mol or 3100 cm$^{-1}$ (Figure 5-4). Using simple transition state theory with $k_{TST} = \nu e^{-\Delta E/RT}$, a torsional frequency ca. 100 cm$^{-1}$ to 1000 cm$^{-1}$ provides an estimated lifetime of ca. 1 to 10 μs, significantly longer than the 3.4 ns range of the present time-resolved measurements, but still consistent with rapid formation of SP in steady state measurements. The small amount of TCC-MC$^-$ that forms directly is expected to go directly to SP (Figure 5-4) within

3.4 ns, but within spectral range investigated here (260-700 nm) it is not possible to differentiate SP from the cis-conjugate bases owing to the overlapping absorption spectra (Figure 5-5).

A similar analysis can be performed for DMSO (Figure 5-6 bottom). In contrast to pH 5.5, the 3.5 ns TA spectra collected in DMSO has only a weak visible absorption spectrum at 530 nm (see also Figure D-3). Therefore, the dominant product following excitation at 404 nm is a mixture of *cis*-phenylhydroxy-MCH conformers. The small amplitude 1.8 ns dynamics observed in DMSO is assigned to the relaxation of conformers approaching a quasi-equilibrium distribution persisting for much longer than 3.5 ns (Figure D-4). Evidently, the $pK_a$ of *trans*-phenylhydroxy-MCH is higher in DMSO than in water, which is consistent with the behavior of acids in aprotic environments. However, the weak visible absorption suggests that some *cis* conformers, perhaps as high as 20%, are deprotonated (Figure 5-6). Like the analysis for water a better agreement of one-bond flip isomerization than hula-twist at $C_B$ is obtained (Figure D-8).

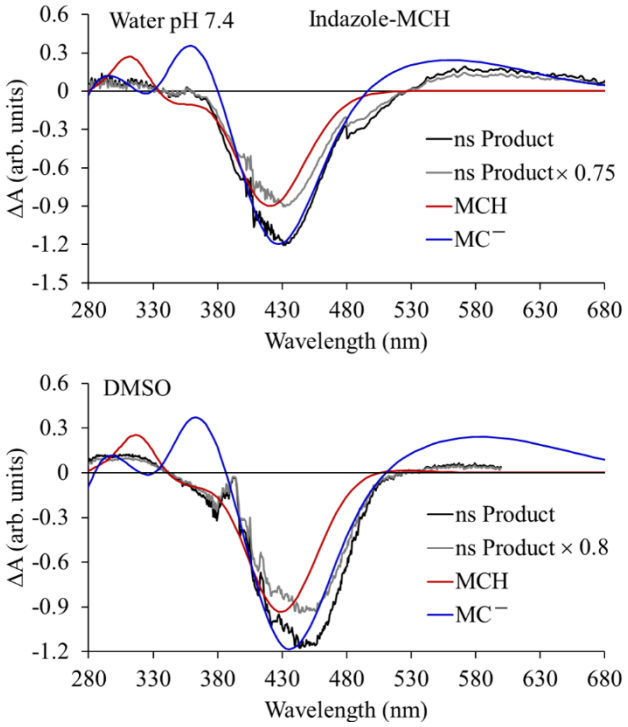



**Figure 5-7** Comparison of the measured (black, gray lines) and simulated product difference spectra for trans to cis isomerization of the indazole-MCH photoacid in both water and DMSO. The blue lines represent the simulated difference spectrum for the formation of cis-indazole-MC– via a one bond twist mechanism. The red lines represent the simulated difference spectra for the formation of cis-phenylhydroxy-MCH via a one bond flip mechanism. Spectroscopic results in this figure courtesy of Dr. Ted Wiley.

Lastly, an analysis can be performed for the indazole-based photoacids (Figure 5-7). At first glance it appears that the situation is different for the indazole-MCH. However, comparison with the simulated difference spectra suggest that the primary photochemical pathways in the buffered water solution are internal conversion to the ground state *trans*- and *cis*-acid combined with rapid formation of the *cis*-conjugate base. The visible product absorption is weaker than for phenylhydroxy-MCH, but the relative intensities of the absorption and bleach are consistent with the calculations. This could also be due to the formation of the cis-indazole tautomer, although the calculated energies make this

unlikely (~0.5 kcal/mol higher in energy). Furthermore, like phenylhydroxy-MCH, the formation of the *cis*-conjugate base is suppressed for indazole-MCH in anhydrous DMSO.

**5.4.2** $pK_a$ calculations

Finally, it is relevant to consider why the photoacids become acidic and whether the quantum chemical calculations can reliably predict the change in acidity. All of the evidence strongly shows that the *cis*-acid is more acidic than the *trans*-acid. In principle the *cis*-acid could ring-close directly to spiropyran with elimination of the phenol proton, and therefore the reactivity of the *cis*-acid would explain the drop in pH upon isomerization. However, the transient absorption spectra best agree with the formation of the *cis*-conjugate base prior to the formation of spiropyran. Furthermore, the results rule out other novel mechanisms like excited-state proton transfer in either the *trans* or *cis* forms; the *trans*-conjugate base does not isomerize when exposed to light (Figure D-5), and the evolution associated difference spectra (EADs, Figure D-4) showed that all the intermediates are excited-state species prior to photoisomerization. Therefore, the evaluation of $pK_a$ is an excellent opportunity to benchmark quantum chemical $pK_a$ calculations.

The difference in $pK_a$ between *cis* and *trans* isomers ($\Delta pK_a$) is calculated by

$$\Delta pK_a = \frac{\Delta\Delta G}{2.303RT}$$

Where

$$\Delta\Delta G = \Delta G_{Cis} - \Delta G_{Trans}$$

And $\Delta G_{Cis}$ and $\Delta G_{Trans}$ are the Gibb's free energy differences between the deprotonated and protonated phenol moiety of the *trans* and *cis* acid, respectively. Before evaluating the performance of the different computational methods, it is important to consider how the photoacid should behave. Previous results have found that charge separation explains the upshift in $pK_a$ upon the isomerization of a model photoactive yellow protein chromophores[34]. The phenoxide and thiomethyl groups of the PYP chromophore experience a repulsive negative-negative charge interaction. Consequently, when isomerizing from trans-to-cis the groups become closer together and this destabilizes the *cis*-

deprotonated form with respect to the cis-protonated form more than the *trans*-deprotonated form with respect to the *trans*-protonated form. In other words, $\Delta G_{Cis}$ becomes greater than $\Delta G_{Trans}$ resulting in a positive $\Delta\Delta G$ and an increase in $pK_a$. In contrast, the photoacids experience an attractive negative-positive interaction in the deprotonated form between the negative phenoxide and positive nitrogen. Consequently, the *cis* conjugate bases should be more stabilized with respect to the *cis*-acids ($\Delta G_{Cis}$ smaller) than the *trans*-conjugate bases are with respect to the *trans* acids ($\Delta G_{Trans}$ larger) and should result in an negative $\Delta\Delta G$, and decrease in $pK_a$.

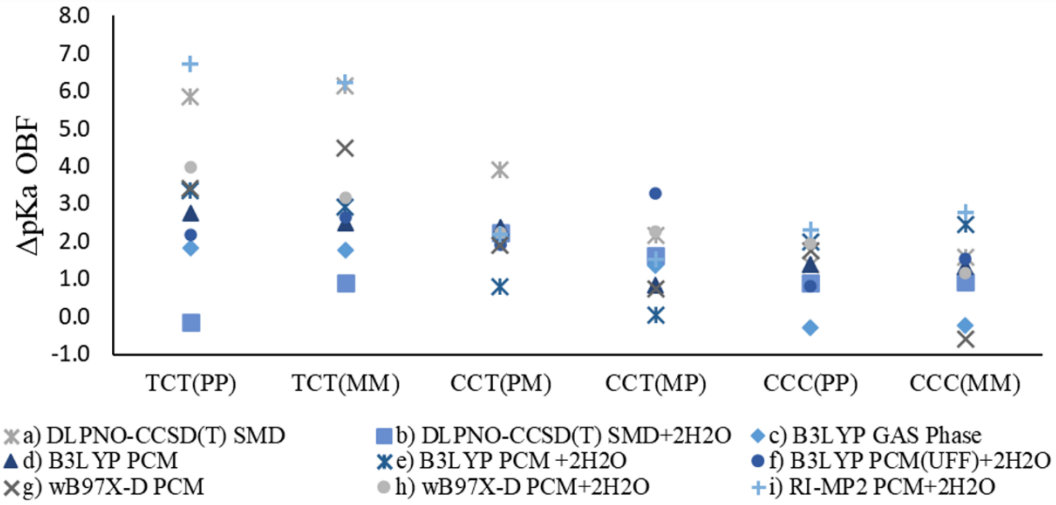



**Figure 5-8** Calculated change in $pK_a$ going from trans- to cis-phenylhydroxy by means of a one-bond flip for the different conformers at varying levels of theory: a) DLPNO-CCSD(T) SMD//B3LYP-PCM/6-31++G**, b) DLPNO-CCSD(T) SMD +2H2O//B3LYP PCM+2H2O/6-31++G**, c) B3LYP/G3LARGE//B3LYP/6-31+G**, d) B3LYP PCM/G3LARGE//B3LYP PCM/6-31++G**, e) B3LYP PCM +2H2O/G3LARGE//B3LYP PCM+2H2O/6-31++G**, f) B3LP PCM(UFF) +2H2O/G3LARGE//B3LYP PCM(UFF)+2H2O/6-31++G**, g) wB97X-D PCM/G3LARGE//wB97X-D PCM/6-31++G**, h) wB97X-D PCM +2H2O/G3LARGE//wB97X-D PCM+2H2O/6-31++G**, i) RI-MP2 SMD+2H2O/cc-pvtz//B3LYP PCM+2H2O/6-31++G**.

In Figure 5-8, $\Delta pK_a$ at varying levels of theory for the cis-conformers of phenylhydroxy-MCH are presented. TCC-MC- goes directly to SP which results in a negative $\Delta pK_a$, but because this does not result in the formation of *cis*-conjugate base it is not considered further. These results show that for all levels of theory $\Delta\Delta G$ is positive or only slightly negative which cannot explain the observed experimental decrease in $pK_a$ upon photoexcitation. While it is generally known that $pK_a$ calculations can be difficult to calculate with quantitative precision, the results are overwhelmingly in the wrong direction. For example, previous computations on the basicity of a azo-stilbene photoswitch showed deficiencies with DFT[35]. Furthermore, calculations with implicit solvent models are known to lack accuracy for specific strong interactions like hydrogen-bonding, and therefore the inclusion of explicit solvent at hydrogen-bonding sites are known to improve the accuracy of $pK_a$ calculations[36,37]. The

specific type of implicit solvent model has also been shown to be important, for example the difference between UFF and Bondi solvent radii was important for the prediction of reduction potentials in water[38]. However, the positive $\Delta pK_a$ result is consistent at all levels of theory including highly correlated methods, the different solvation models and solvents, and with explicit solvent interactions. Note that this is not a result of truncation of the alkyl sulfonate. The alkyl sulfonate can slightly shield the positive charge on the nitrogen on the indoline moiety[16], so if the electrostatic interaction is the dominant force leading to photoacidity the truncation of the sulfonate group is expected to increase the acidity.

The results in Figure 5-8, can therefore be interpreted in one of two ways: (1) the *cis*-acid is not more acidic than the *trans*-acid and some other species must be responsible for the visible absorption within 3.5 ns, or (2) the *cis*-acid is more acidic than the *trans*-acid but the computations lack the necessary accuracy to predict the photoacidity. Given no physical evidence to the contrary, these results must be interpreted in the second way. This, however, does imply that the quantum chemical models lack the necessary precision or interaction which is necessary for properly describing the acidity of phenylhydroxy and indazole based photoacids. Therefore, the results presented here represent a striking example of how $pK_a$ calculations can be difficult to calculate and future research must be given to investigate this apparent discrepancy.

## **5.5** Conclusions

When the *trans*-phenylhydroxy-MCH is dissolved in water, photoexcitation of the molecule produces a species with red-shifted absorption with respect to the *trans*-phenylhydroxy-MCH on a ~10 ps timescale. This product persists for times much longer than 3.5 ns before the formation of phenylhydroxy-SP. The intermediate is assigned to a TCT-, CCC- and CCT-phenylhydroxy-$MC^-$ based on TD-DFT calculations. Any formation of TCC- will result in rapid formation of the ring-closed spiropyran product. Single-bond rotation from TCT- to TCC- occurs on a timescale longer than the 3.5 ns consistent with a rotation and/or cyclization barrier. In DMSO solvent a significantly smaller amount of *cis*-phenylhydroxy-$MC^-$ photoproduct is formed within 50 ps, persisting for >> 3.5 ns, which is consistent with a suppressed acidity in aprotic environments. The picture is similar for indazole-MCH; *cis*-indazole-$MC^-$ is formed an ultrafast timescale.

Lastly, $pK_a$ calculations were performed with 9 different combinations of quantum chemical methods including correlated methods, solvation models, and explicit solvation. These calculations all consistently show that the difference in $pK_a$ is in the wrong direction from expectation. Given no physical evidence to the contrary this suggests that common computational methods for the evaluation of $pK_a$ are not sufficient to describe photoacidity in phenylhydroxy- and indazole-based photoacids and more work is still necessary to properly evaluate these properties.

Overall, the transient absorption results combined with computational results lend insight into the early time details of the operation of merocyanine/spiropyran photoacids, and we hope that the results provided here can provide insight into future computational and experimental research of reversible photoacids.

## **5.6** Acknowledgments

This work was supported by a grant from the National Science Foundation NSF-CHE 1464584 to RJS. CRA is thankful to. Dr. Ted Wiley and Dr. Nicholas Miller who performed the TA spectroscopy, and Professor Roseanne Sension helped with data analysis. Dr. Nawodi Abernathi and Professor Li Yiao provided the samples.

## **5.7** References

(1) Bouas-Laurent, H.; Dürr, H. *Photochromism : Molecules and Systems*; Elsevier, 2003.
(2) Ernsting, N. P.; Arthen-Engeland, T. "Photochemical Ring-Opening Reaction of Indolinospiropyrans Studied by Subpicosecond Transient Absorption" *J. Phys. Chem.* **1991**, *95*, 5502-5509.
(3) Heiligman-Rim, R.; Hirshberg, Y.; Fischer, E. "Photochromism in Some Spiropyrans. Part III. The Extent of Phototransformation" *Journal of the Chemical Society* **1961**, *297*, 156-163.
(4) Zhu, M. Q.; Zhu, L.; Han, J. J.; Wu, W.; Hurst, J. K.; Li, A. D. Q. "Spiropyran-based photochromic polymer nanoparticles with optically switchable luminescence" *J. Am. Chem. Soc.* **2006**, *128*, 4303-4309.
(5) Mahvidi, S.; Takeuchi, S.; Kusumoto, S.; Sato, H.; Nakagawa, T.; Yokoyama, Y. "Gated Photochromic System of Diarylethene with a Photon-Working Key" *Organic Letters* **2016**, *18*, 5042-5045.
(6) London, G.; Chen, K.-Y.; Carroll, G. T.; Feringa, B. L. "Towards Dynamic Control of Wettability by Using Functionalized Altitudinal Molecular Motors on Solid Surfaces" *Chem. - Eur. J.* **2013**, *19*, 10690-10697.

(7) Rosario, R.; Gust, D.; Garcia, A. A.; Hayes, M.; Taraci, J. L.; Clement, T.; Dailey, J. W.; Picraux, S. T. "Lotus effect amplifies light-induced contact angle switching" *J. Phys. Chem. B* **2004**, *108*, 12640-12642.
(8) Rosario, R.; Gust, D.; Hayes, M.; Jahnke, F.; Springer, J.; Garcia, A. A. "Photon-modulated wettability changes on spiropyran-coated surfaces" *Langmuir* **2002**, *18*, 8062-8069.
(9) Johns, V. K.; Patel, P. K.; Hassett, S.; Calvo-Marzal, P.; Qin, Y.; Chumbimuni-Torres, K. Y. "Visible Light Activated Ion Sensing Using a Photoacid Polymer for Calcium Detection" *Anal. Chem.* **2014**, *86*, 6184-6187.
(10) Bao, H.; Li, F. F.; Lei, L. C.; Yang, B.; Li, Z. J. "ON/OFF states of a microbial fuel cell controlled by an optical switching system" *Rsc Advances* **2014**, *4*, 27277-27280.
(11) Andersson, J.; Li, S. M.; Lincoln, P.; Andreasson, J. *Light controlled DNA-binding of spiropyrans*, 2008; Vol. 10.
(12) Beyer, C.; Wagenknecht, H. A. "Synthesis of DNA with spirobenzopyran as an internal covalent modification" *Synlett* **2010**, 1371-1376.
(13) Abeyrathna, N.; Liao, Y. "A Reversible Photoacid Functioning in PBS Buffer under Visible Light" *J. Am. Chem. Soc.* **2015**, *137*, 11282-11284.
(14) Peng, P.; Wang, C. M.; Shi, Z.; Johns, V. K.; Ma, L. Y.; Oyer, J.; Copik, A.; Igarashi, R.; Liao, Y. "Visible-light activatable organic CO-releasing molecules (PhotoCORMs) that simultaneously generate fluorophores" *Organic & Biomolecular Chemistry* **2013**, *11*, 6671-6674.
(15) Shi, Z.; Peng, P.; Strohecker, D.; Liao, Y. "Long-lived photoacid based upon a photochromic reaction" *J. Am. Chem. Soc.* **2011**, *133*, 14699-14703.
(16) Fu, C. K.; Xu, J. T.; Boyer, C. "Photoacid-mediated ring opening polymerization driven by visible light" *Chem. Comm.* **2016**, *52*, 7126-7129.
(17) Kundu, P. K.; Samanta, D.; Leizrowice, R.; Margulis, B.; Zhao, H.; Borner, M.; Udayabhaskararao, T.; Manna, D.; Klajn, R. "Light-controlled self-assembly of non-photoresponsive nanoparticles" *Nat. Chem.* **2015**, *7*, 646-652.
(18) Maity, C.; Hendriksen, W. E.; van Esch, J. H.; Eelkema, R. "Spatial Structuring of a Supramolecular Hydrogel by using a Visible-Light Triggered Catalyst" *Angew. Chem., Int. Ed.* **2015**, *54*, 998-1001.
(19) Samanta, D.; Klajn, R. "Aqueous Light-Controlled Self-Assembly of Nanoparticles" *Adv. Opt. Mater.* **2016**, *4*, 1373-1377.
(20) Johns, V. K.; Peng, P.; DeJesus, J.; Wang, Z. Z.; Liao, Y. "Visible-Light-Responsive Reversible Photoacid Based on a Metastable Carbanion" *Chem. - Eur. J.* **2014**, *20*, 689-692.
(21) Klajn, R. "Spiropyran-based dynamic materials" *Chem. Soc. Rev.* **2014**, *43*, 148-184.
(22) Cances, E.; Mennucci, B.; Tomasi, J. "A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics" *J. Chem. Phys* **1997**, *107*, 3032-3041.
(23) Lange, A. W.; Herbert, J. M. "A smooth, nonsingular, and faithful discretization scheme for polarizable continuum models: The switching/Gaussian approach" *J. Chem. Phys* **2010**, *133*.
(24) Lange, A. W.; Herbert, J. M. "Symmetric versus asymmetric discretization of the integral equations in polarizable continuum solvation models" *Chemical Physics Letters* **2011**, *509*, 77-87.
(25) Hirata, S.; Head-Gordon, M. "Time-dependent density functional theory within the Tamm-Dancoff approximation" *Chemical Physics Letters* **1999**, *314*, 291-299.

(26) Cossi, M.; Barone, V. "Time-dependent density functional theory for molecules in liquid solutions" *J. Chem. Phys* **2001**, *115*, 4708-4717.
(27) Shao, Y. H.; Gan, Z. T.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X. T.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kus, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; Woodcock, H. L.; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C. M.; Chen, Y. Q.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; DiStasio, R. A.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W. D.; Harbach, P. H. P.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T. C.; Ji, H. J.; Kaduk, B.; Khistyaev, K.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S. P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Neuscamman, E.; Oana, C. M.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T. "Advances in molecular quantum chemistry contained in the Q-Chem 4 program package" *Molecular Physics* **2015**, *113*, 184-215.

# **Chapter 6** Future Works and Conclusions

Light irradiation gives access to "forbidden" chemistries that can be applied to synthesize novel compounds including strained rings, and materials with switchable mechanical properties. The development of these technologies requires an in-depth understanding of the reaction mechanisms which can be provided through computational chemistry. However, most computational photochemistry relies on tedious manual steps that are difficult to automate and require prior chemical knowledge or intuition, which limits their predictive and exploratory capabilities. Furthermore, photochemistry is particularly sensitive to the external steric and electrostatic environment. The treatment of these problems are the main areas where this thesis has contributed.

The initial Chapters of this thesis focused on the development of growing string methods for photochemistry and condensed phases, and the latter chapters applies GSM to challenging cases. Chapter 2's investigation developed excited state growing string methods that allow conical intersections to be found from stable equilibrium geometries given chemically intuitive reaction directions that can be easily generated and tested. Furthermore, this protocol can be automated by generating combinations of driving coordinates within a set of reactive atoms and can be used to map the seam of conical intersections. These methods have been tested on photoisomerization and photocycloaddition and in both cases have made important discoveries that would have been difficult to make otherwise. Chapter 3's investigation extends the capabilities of the growing string methods to large multi-molecular systems and was shown to capture the highly correlated motions in a crystalline gyroscope. Chapter 4's investigation applies the techniques developed in Chapter 2 and 3 to the isomerization of crystalline cis,cis-diphenylbutadiene, and the results are complementary to the results of Chapter 2 on stilbene. Finally, chapter 5 provides a close look at the photoisomerization of reversible photoacids with a combined experimental and theoretical analysisHowever, much work is still needed to obtain quantitative accuracy. Starting with Chapter 2, the investigations revealed an interesting seam of conical intersections that affects the [4+2] photocycloaddition (Figure 2-13). Formally, the [4+2] photocycloaddition is forbidden by WoodwardHoffmann rules, but this only

applies to concerted reactions, which it is not. However, further verifying this feature requires going beyond the strategy outlined in Chapter 2. Although a good approximation for many photochemistries, the true decay point in a CI mediated transformation is not the minimum energy conical intersection (MECI), as evident in Figure 2-13. Therefore, what is needed is *ab initio* molecular dynamics (MD) that can appropriately model the dynamics through the extended seam of conical intersections. Nevertheless, if there is a transition state separating the excited-state minimum and the seam of conical intersections then the MD can be started at the transition state in the direction of the conical intersection. This strategy eliminates the slow and uninteresting dynamics of climbing up the transition state valley, and captures a more realistic process of decay through the seam of conical intersections. Furthermore, the MD can be used to estimate the quantum yields which is not possible in the approach outlined in Chapter 2.

In Chapters 3 and 4, GSM was applied to crystals. Crystals are a special class of condensed phases where the concept of a MEP still holds. However, when moving to the liquid phase the minimum energy geometry of the system does not correspond to a physical solution anymore. First, the potential energy surface for the solution phase is incredibly rugged so there are many local minima, e.g. a single hydrogen bond is ~3 kcal/mol. Second, optimizing the geometry of the system removes all the vibrational and translational energy, i.e. it freezes the solution, which is not desirable. Nevertheless, what can be done is to assume that the movement of the reacting system is not correlated with the environment. Therefore, the solution environment can be fixed at a geometry which corresponds to the proper pressure and density, and the reaction path calculated for only the reacting system (for example, see reference 27 of Chapter 3). This method could be performed within the framework of Chapter 3, but it is a difficult job to obtain good statistics because there is a lot of variability in the environment. Furthermore, the contribution of entropy would need to be acounted for somehow. Ideally, a growing string method that could calculate the free-energy directly in the solution phase would be preferred. Some work on this has already been done by Professor Vanden Eijnden on the finite-temperature string methods. Incorporating these techniques within the extensible framework of the Python Growing String Method would therefore be worthwhile.

Chapter 4, has an apparently opposite problem – it is most likely too constrained. In Chapter 4, the QM/MM methods only currently allow for finite molecular models. Consequently, we used a four-layer cluster model to constrain the outer layers to the crystalline geometry. However, the bicycle pedal isomerization may require long-range lattice relaxation, like what was observed for the the

crystalline gyroscope of Chapter 3. Future work should test other kinds of constraints that can maintain the crystal packing but enable the crystal to relax with the motion of the reacting system.

The computational results of Chapter 5 provided excellent interpretation of all the observations in the transient absorption spectroscopy but failed to accurately account for the change in $pK_a$ of the photoacids upon isomerization. This demonstrates in another way how computation can provide useful information for some features but fail in others.

In conclusion, this thesis presents a compelling story of how long-standing problems in photochemistry can be understood by new methods in computational photochemistry. We have identified further reason why the [4+2] photocycloaddition is forbidden, identified an elusive hula-twist conical intersection for stilbene, identified a photoproduct which has been missing for over 40 years, revealed the fine details of the mechanics of the bicycle pedal photoisomerization, and have identified the early time details of merocyanine based photoacids. However, like all good research the results reveal additional questions and possibilities. It is my hope, therefore, that the work of this thesis can be further expanded and contribute to future research and chemistry that can benefit the world.

# Appendix A Supporting information for Chapter 2

## *A.1 Seam Minimum Energy Paths*

### A.1.1 *Ethylene*

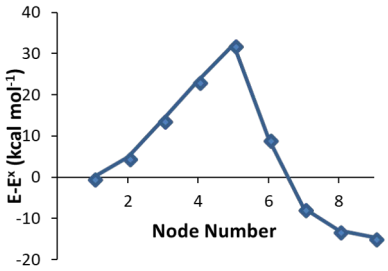



**Figure A-1 DE-GSM seam from 1a to 1b using 11 nodes.**

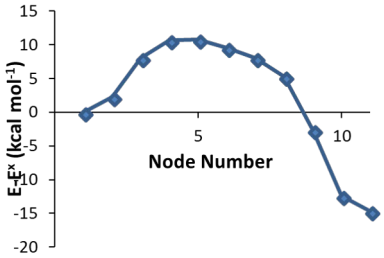



**Figure A-2 DE-GSM seam from 1a to 1b using 9 nodes.**

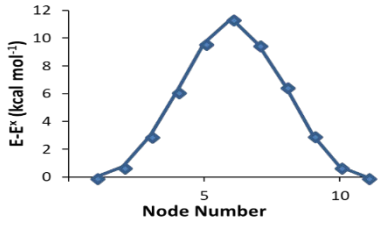



**Figure A-3 DE-GSM seam from 1a to 1a' (tilt isomer).**

### A.1.2 *Stilbene*

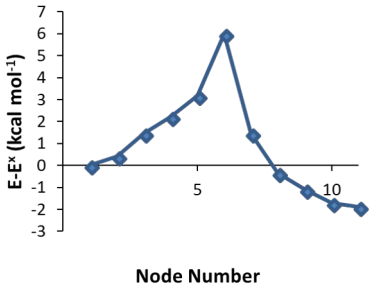



**Figure A-4 DE-GSM seam from 3a to 3b.**

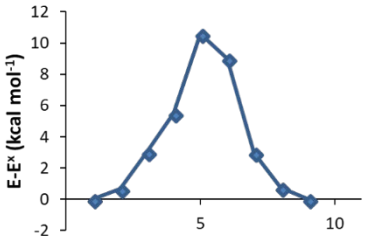



**Figure A-5 SE-GSM seam.**

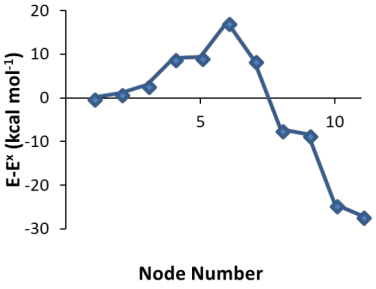



**Figure A-6 DE-GSM seam from 3a to 3c.**

## *A.2 Butadiene Dimerization*

### A.2.1 *MECI*

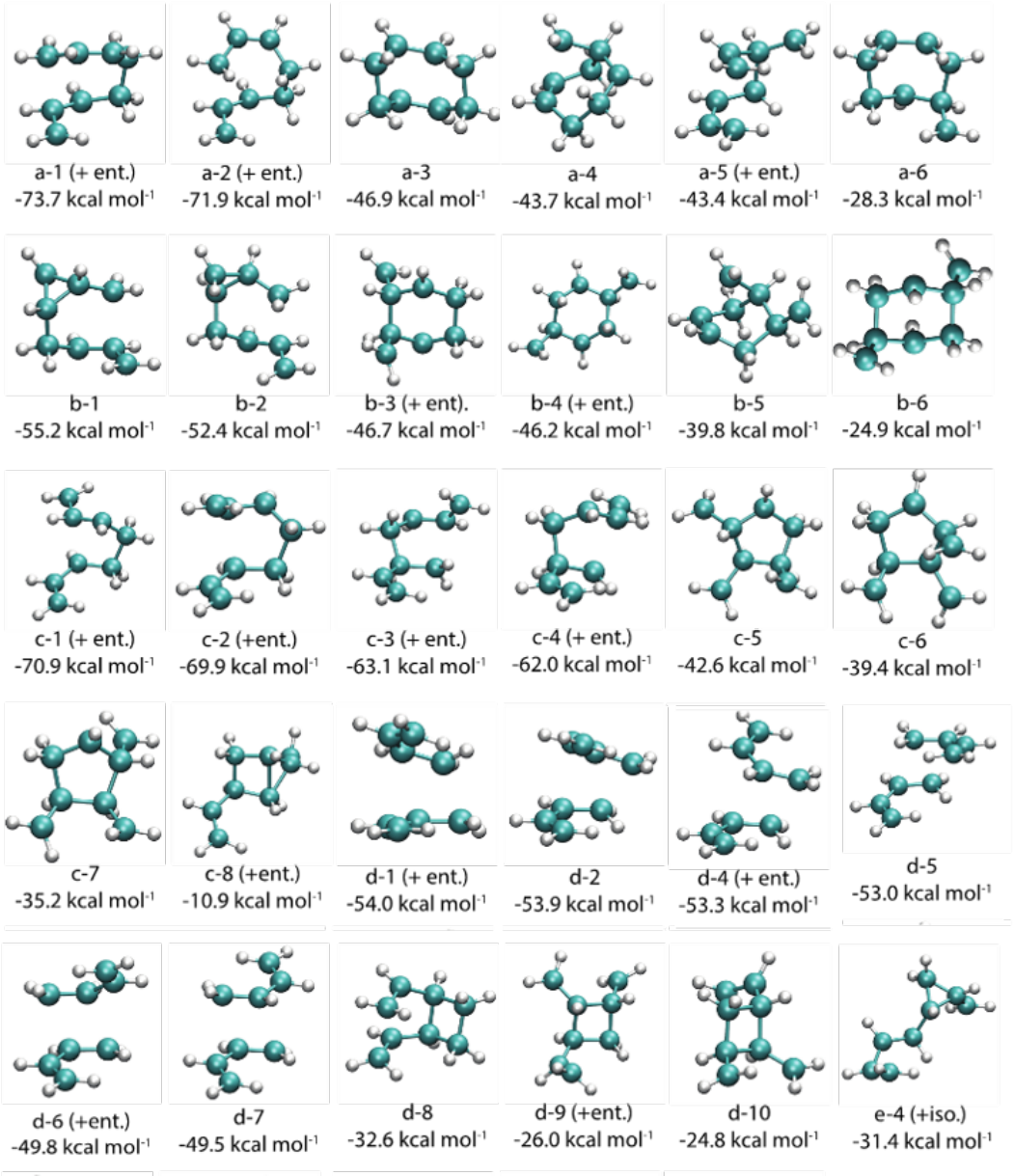



**Figure A-7 MECI discovered by ZStruct SE-GSM search**

### A.2.2 *Reaction paths*

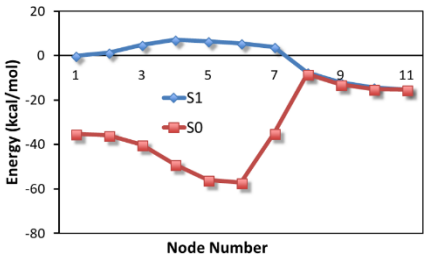



Figure A-8 4b to a-1

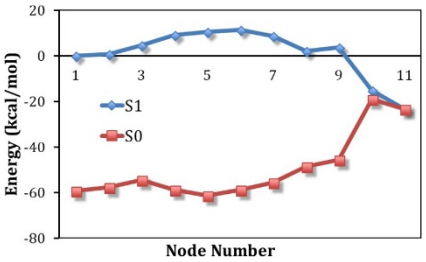



Figure A-9 4e to a-1

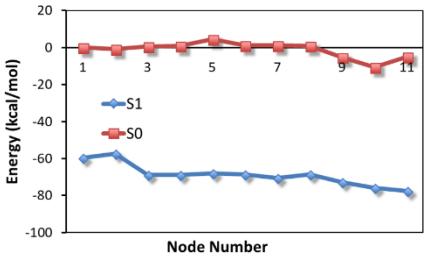



Figure A-10 4d to a-1 partial string

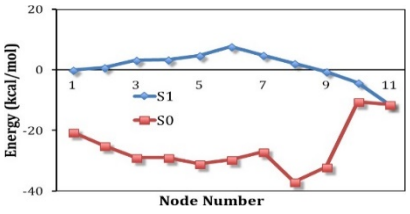



Figure A-11 4a to a-2

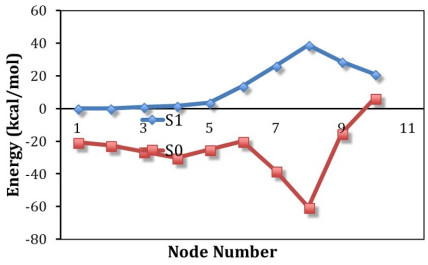



Figure A-12 4a to a-3

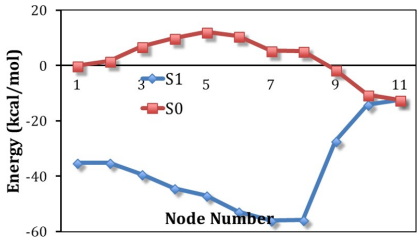



Figure A-13 4b to c-1

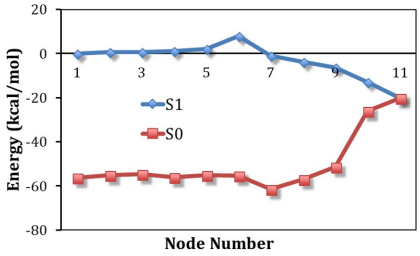



Figure A-14 4c to c-1

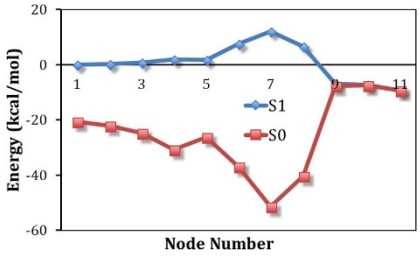



Figure A-15 4a to c-2

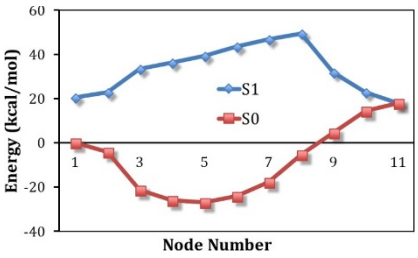



Figure A-16 4a to c-3

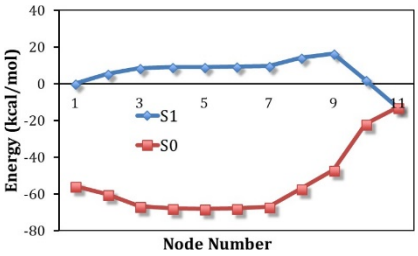



Figure A-17 4c to c-3

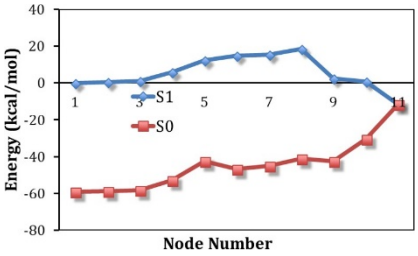



Figure A-18 4c to c-4

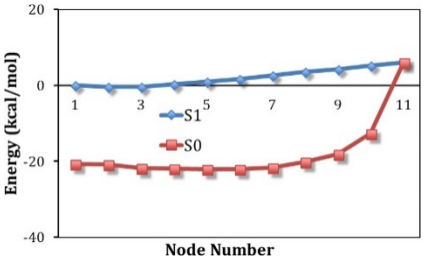



Figure A-19 5a to d-1

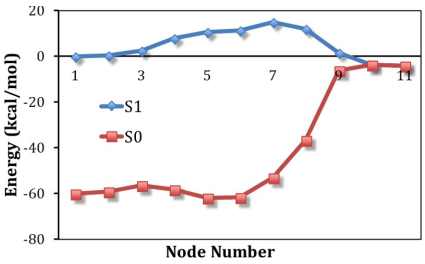



Figure A-20 4d to d-2

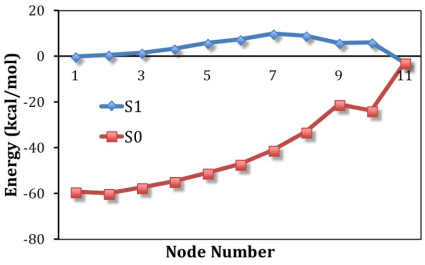



Figure A-21 4e to d-2

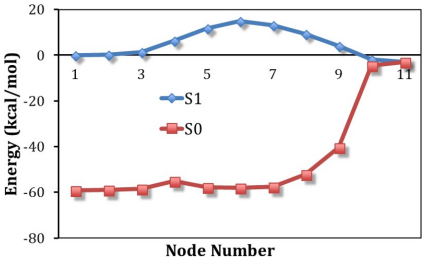



Figure A-22 4e to d-3

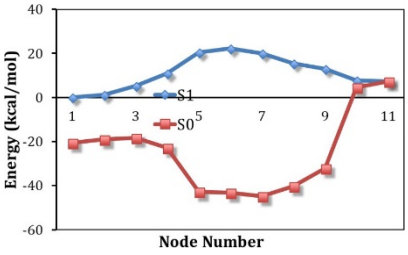



Figure A-23 4e to d-4

All reaction paths to b and e conical intersections involved multiple adiabatic barriers or had too large of reaction barriers and are not reported here.

## **Appendix B** Supporting information for Chapter 3

Table B-1 Atoms types and charges for gyroscope

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | C1 | 3.557 | 4.855 | 5.943 | ca | 1 | GYR | -0.2296 |
| 2 | H1 | 3.913 | 4.199 | 5.171 | ha | 1 | GYR | 0.176897 |
| 3 | C2 | 3.81 | 6.225 | 5.859 | ca | 1 | GYR | 0.294909 |
| 4 | C3 | 3.343 | 7.064 | 6.868 | ca | 1 | GYR | -0.2296 |
| 5 | H3 | 3.54 | 8.118 | 6.807 | ha | 1 | GYR | 0.176897 |
| 6 | C4 | 4.526 | 6.779 | 4.738 | c1 | 1 | GYR | -0.29293 |
| 7 | C5 | 5.09 | 7.274 | 3.81 | c1 | 1 | GYR | 0.114065 |
| 8 | C6 | 5.761 | 7.905 | 2.645 | c3 | 1 | GYR | -0.48985 |
| 9 | C7 | 5.692 | 9.439 | 2.781 | ca | 1 | GYR | 0.327529 |
| 10 | C8 | 4.652 | 10.062 | 3.461 | ca | 1 | GYR | -0.24276 |
| 11 | H8 | 3.901 | 9.472 | 3.948 | ha | 1 | GYR | 0.15614 |
| 12 | C9 | 4.567 | 11.451 | 3.509 | ca | 1 | GYR | -0.11439 |
| 13 | H9 | 3.756 | 11.915 | 4.04 | ha | 1 | GYR | 0.131347 |
| 14 | C10 | 5.512 | 12.224 | 2.865 | ca | 1 | GYR | -0.16537 |
| 15 | H10 | 5.442 | 13.297 | 2.896 | ha | 1 | GYR | 0.136894 |
| 16 | C11 | 6.554 | 11.618 | 2.197 | ca | 1 | GYR | -0.11439 |
| 17 | H11 | 7.301 | 12.213 | 1.703 | ha | 1 | GYR | 0.131347 |
| 18 | C12 | 6.647 | 10.231 | 2.16 | ca | 1 | GYR | -0.24276 |
| 19 | H12 | 7.466 | 9.78 | 1.636 | ha | 1 | GYR | 0.15614 |
| 20 | C13 | 5.02 | 7.429 | 1.369 | ca | 1 | GYR | 0.327529 |
| 21 | C14 | 4.579 | 8.306 | 0.393 | ca | 1 | GYR | -0.24276 |
| 22 | H14 | 4.728 | 9.361 | 0.498 | ha | 1 | GYR | 0.15614 |
| 23 | C15 | 3.931 | 7.833 | -0.749 | ca | 1 | GYR | -0.11439 |
| 24 | H15 | 3.599 | 8.533 | -1.494 | ha | 1 | GYR | 0.131347 |
| 25 | C16 | 3.711 | 6.485 | -0.913 | ca | 1 | GYR | -0.16537 |
| 26 | H16 | 3.206 | 6.121 | -1.79 | ha | 1 | GYR | 0.136894 |
| 27 | C17 | 4.139 | 5.601 | 0.058 | ca | 1 | GYR | -0.11439 |
| 28 | H17 | 3.97 | 4.545 | -0.059 | ha | 1 | GYR | 0.131347 |
| 29 | C18 | 4.792 | 6.068 | 1.188 | ca | 1 | GYR | -0.24276 |
| 30 | H18 | 5.122 | 5.369 | 1.934 | ha | 1 | GYR | 0.15614 |
| 31 | C19 | 7.241 | 7.455 | 2.636 | ca | 1 | GYR | 0.327529 |
| 32 | C20 | 7.887 | 7.051 | 1.478 | ca | 1 | GYR | -0.24276 |
| 33 | H20 | 7.35 | 6.991 | 0.552 | ha | 1 | GYR | 0.15614 |
| 34 | C21 | 9.237 | 6.722 | 1.492 | ca | 1 | GYR | -0.11439 |
| 35 | H21 | 9.716 | 6.411 | 0.581 | ha | 1 | GYR | 0.131347 |
| 36 | C22 | 9.956 | 6.798 | 2.667 | ca | 1 | GYR | -0.16537 |
| 37 | H22 | 11.001 | 6.544 | 2.681 | ha | 1 | GYR | 0.136894 |
| 38 | C23 | 9.323 | 7.198 | 3.828 | ca | 1 | GYR | -0.11439 |
| 39 | H23 | 9.874 | 7.258 | 4.75 | ha | 1 | GYR | 0.131347 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 40 | C24 | 7.979 | 7.526 | 3.816 | ca | 1 | GYR | -0.24276 |
| 41 | H24 | 7.502 | 7.84 | 4.726 | ha | 1 | GYR | 0.15614 |
| 42 | C25 | 2.624 | 6.559 | 7.935 | ca | 1 | GYR | -0.2296 |
| 43 | H25 | 2.268 | 7.215 | 8.707 | ha | 1 | GYR | 0.176897 |
| 44 | C26 | 2.371 | 5.189 | 8.02 | ca | 1 | GYR | 0.294909 |
| 45 | C27 | 2.838 | 4.35 | 7.01 | ca | 1 | GYR | -0.2296 |
| 46 | H27 | 2.64 | 3.296 | 7.07 | ha | 1 | GYR | 0.176897 |
| 47 | C28 | 1.655 | 4.636 | 9.141 | c1 | 1 | GYR | -0.29293 |
| 48 | C29 | 1.091 | 4.14 | 10.068 | c1 | 1 | GYR | 0.114065 |
| 49 | C30 | 0.42 | 3.51 | 11.234 | c3 | 1 | GYR | -0.48985 |
| 50 | C31 | 0.489 | 1.975 | 11.097 | ca | 1 | GYR | 0.327529 |
| 51 | C32 | 1.529 | 1.352 | 10.418 | ca | 1 | GYR | -0.24276 |
| 52 | H32 | 2.281 | 1.942 | 9.931 | ha | 1 | GYR | 0.15614 |
| 53 | C33 | 1.614 | -0.037 | 10.369 | ca | 1 | GYR | -0.11439 |
| 54 | H33 | 2.426 | -0.501 | 9.838 | ha | 1 | GYR | 0.131347 |
| 55 | C34 | 0.669 | -0.81 | 11.013 | ca | 1 | GYR | -0.16537 |
| 56 | H34 | 0.738 | -1.883 | 10.982 | ha | 1 | GYR | 0.136894 |
| 57 | C35 | -0.373 | -0.204 | 11.681 | ca | 1 | GYR | -0.11439 |
| 58 | H35 | -1.12 | -0.799 | 12.175 | ha | 1 | GYR | 0.131347 |
| 59 | C36 | -0.467 | 1.183 | 11.719 | ca | 1 | GYR | -0.24276 |
| 60 | H36 | -1.286 | 1.634 | 12.243 | ha | 1 | GYR | 0.15614 |
| 61 | C37 | 1.161 | 3.985 | 12.51 | ca | 1 | GYR | 0.327529 |
| 62 | C38 | 1.602 | 3.108 | 13.486 | ca | 1 | GYR | -0.24276 |
| 63 | H38 | 1.453 | 2.053 | 13.381 | ha | 1 | GYR | 0.15614 |
| 64 | C39 | 2.25 | 3.581 | 14.628 | ca | 1 | GYR | -0.11439 |
| 65 | H39 | 2.582 | 2.881 | 15.374 | ha | 1 | GYR | 0.131347 |
| 66 | C40 | 2.47 | 4.929 | 14.791 | ca | 1 | GYR | -0.16537 |
| 67 | H40 | 2.975 | 5.293 | 15.668 | ha | 1 | GYR | 0.136894 |
| 68 | C41 | 2.041 | 5.814 | 13.821 | ca | 1 | GYR | -0.11439 |
| 69 | H41 | 2.209 | 6.87 | 13.938 | ha | 1 | GYR | 0.131347 |
| 70 | C42 | 1.389 | 5.346 | 12.69 | ca | 1 | GYR | -0.24276 |
| 71 | H42 | 1.059 | 6.045 | 11.944 | ha | 1 | GYR | 0.15614 |
| 72 | C43 | -1.06 | 3.959 | 11.242 | ca | 1 | GYR | 0.327529 |
| 73 | C44 | -1.706 | 4.363 | 12.4 | ca | 1 | GYR | -0.24276 |
| 74 | H44 | -1.169 | 4.423 | 13.326 | ha | 1 | GYR | 0.15614 |
| 75 | C45 | -3.056 | 4.693 | 12.386 | ca | 1 | GYR | -0.11439 |
| 76 | H45 | -3.534 | 5.005 | 13.297 | ha | 1 | GYR | 0.131347 |
| 77 | C46 | -3.775 | 4.616 | 11.212 | ca | 1 | GYR | -0.16537 |
| 78 | H46 | -4.82 | 4.87 | 11.198 | ha | 1 | GYR | 0.136894 |
| 79 | C47 | -3.142 | 4.216 | 10.051 | ca | 1 | GYR | -0.11439 |
| 80 | H47 | -3.694 | 4.155 | 9.13 | ha | 1 | GYR | 0.131347 |
| 81 | C48 | -1.799 | 3.888 | 10.063 | ca | 1 | GYR | -0.24276 |
| 82 | H48 | -1.322 | 3.574 | 9.153 | ha | 1 | GYR | 0.15614 |

# Appendix C Supporting information for Chapter 4

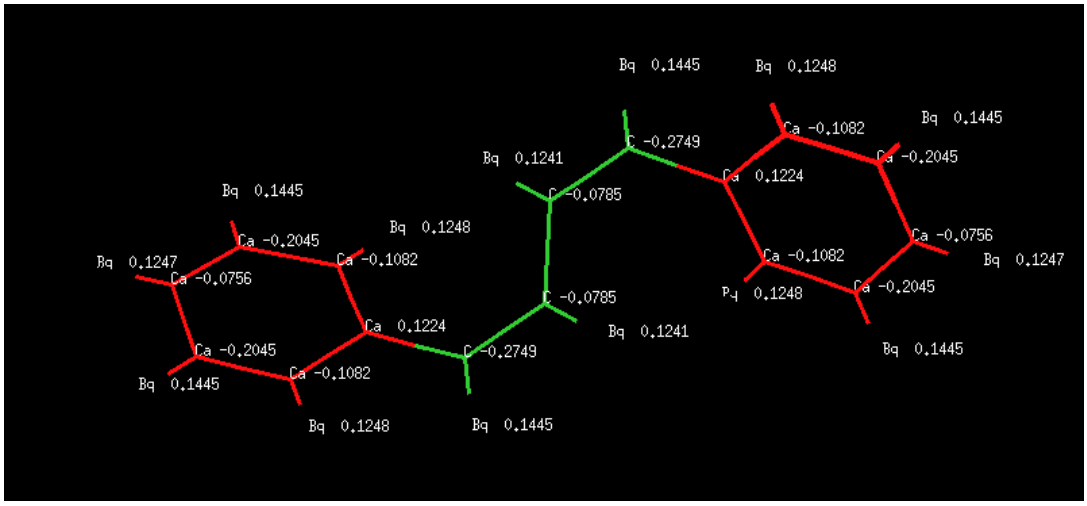



**Figure C-1 AMBER Molecular atom types (reference 1) and charges calculated with RESP which are used in all quantum-mechanics molecular mechanics (QM/MM) and molecular mechanics (MM) calculations in the main text.**

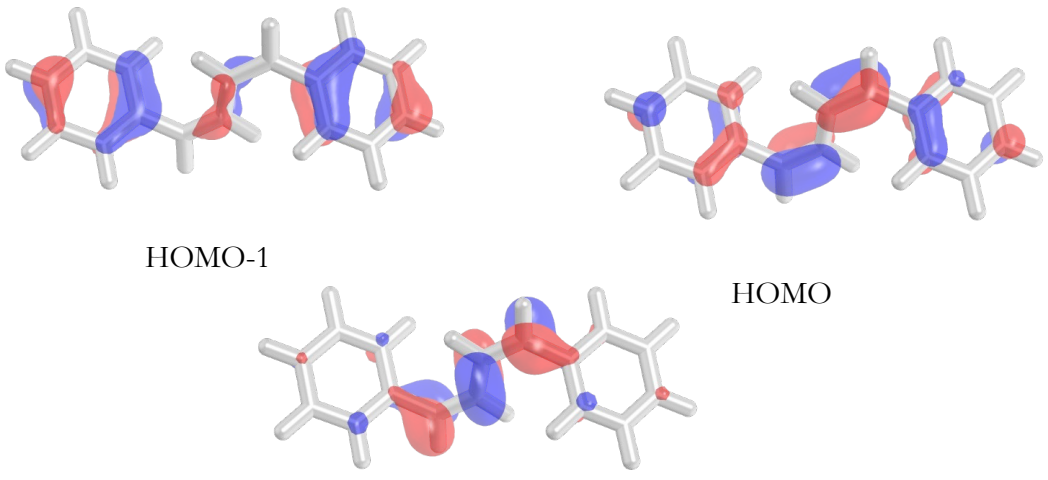



**Figure C-3 Orbitals in the active space for CAS(4,3)SCF calculations.**

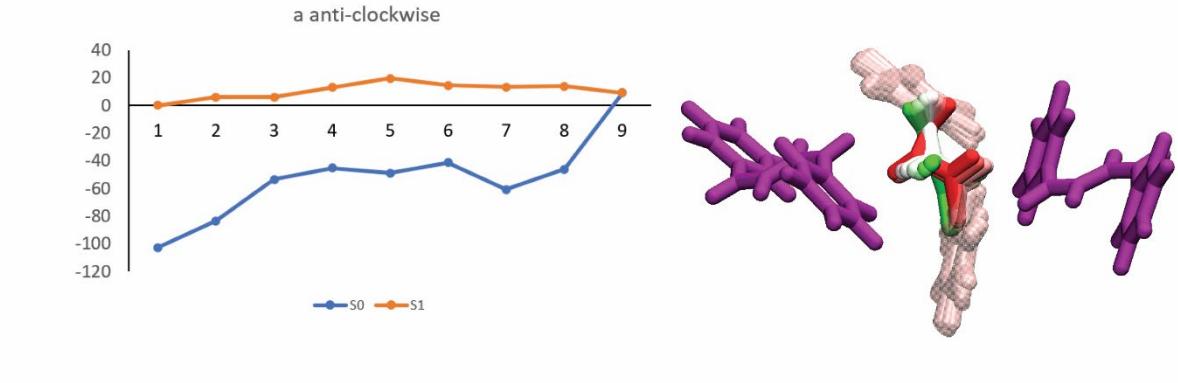



**Figure C-4 Reaction path for A counter clockwise BP rotation.**

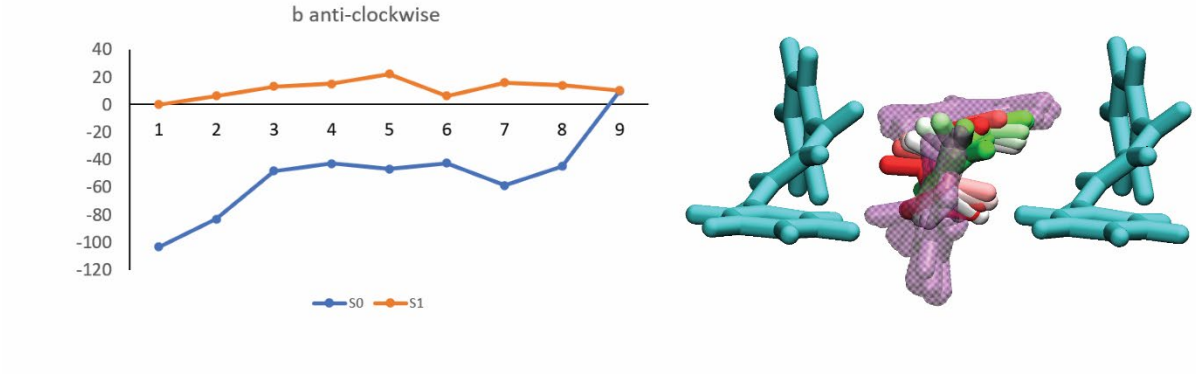



**Figure C-5 Reaction path for B counter-clockwise BP rotation.**

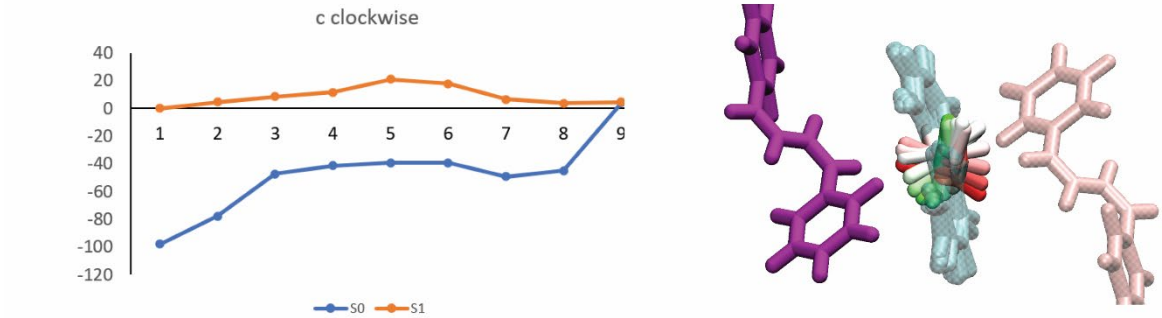



**Figure C-6 Reaction path for B clockwise BP rotation.**

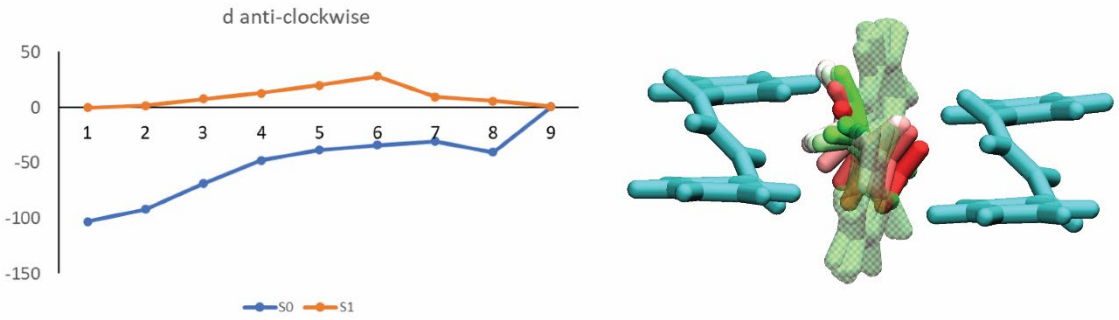



**Figure C-7 Reaction path for D anti-clockwise rotation.**

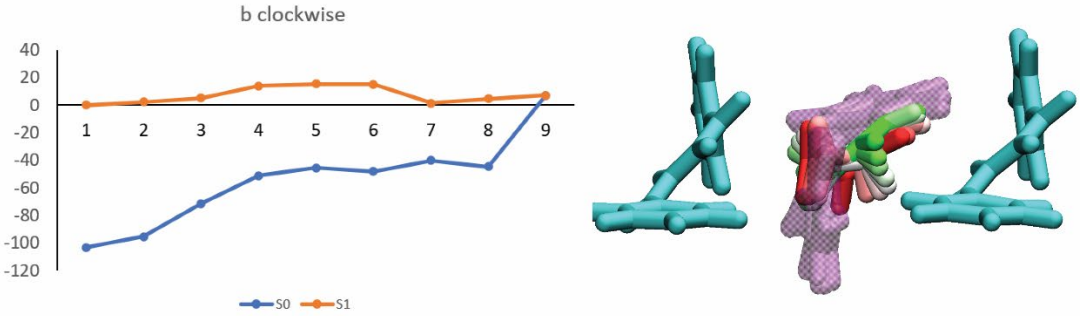



**Figure C-8 Reaction path for C clockwise BP rotation.**

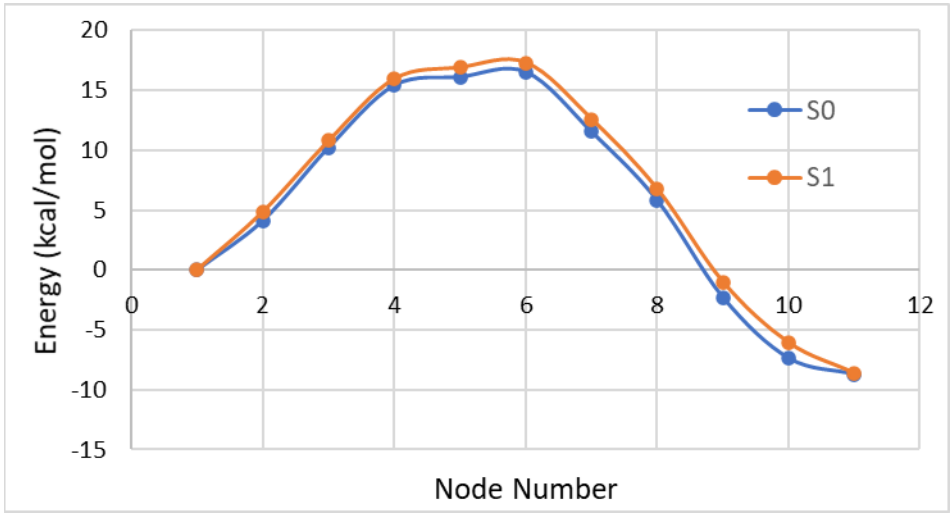



**Figure C-9 Seam path connecting cc-transoid-R to tt-transoid MECI.**

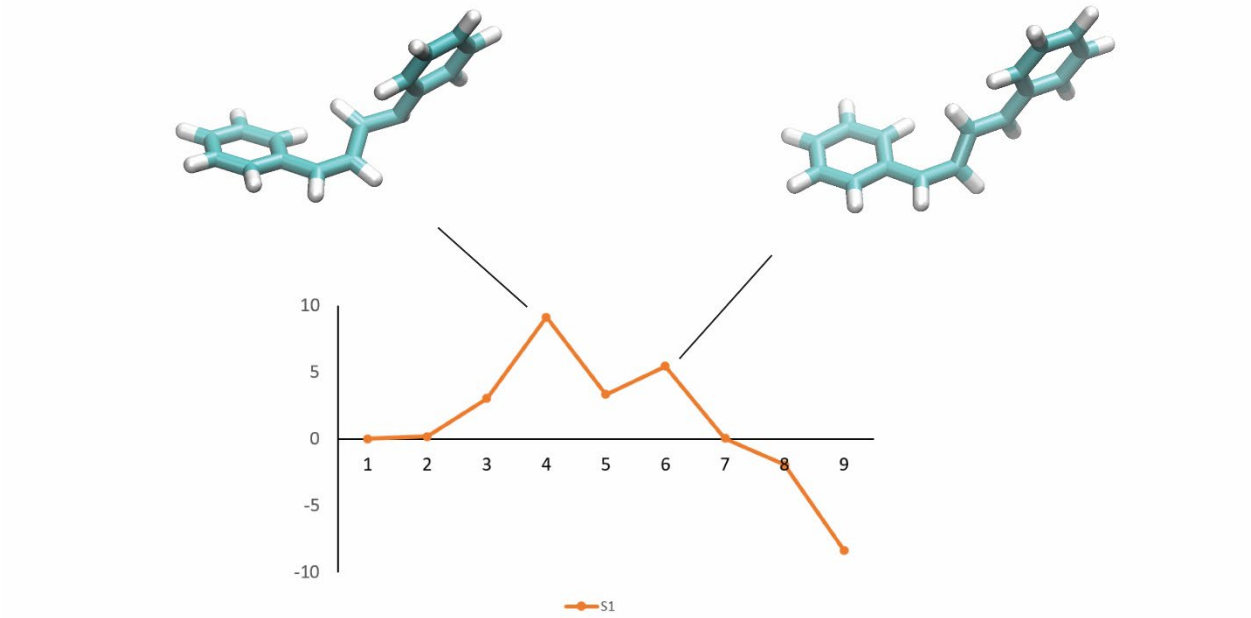



**Figure C-10 Gas-phase reaction path to the tt-transoid MECI. This reaction path is not fully optimized, due to oscillations between nodes 4 and 6.**

# Appendix D Supporting information for Chapter 5

The experimental work done in support of Chapter 5 was only achievable through the expertise of Drs. Ted Wiley, Nicholas Miller, and Professor Roseanne Sension, and are provided here to support the computational results.

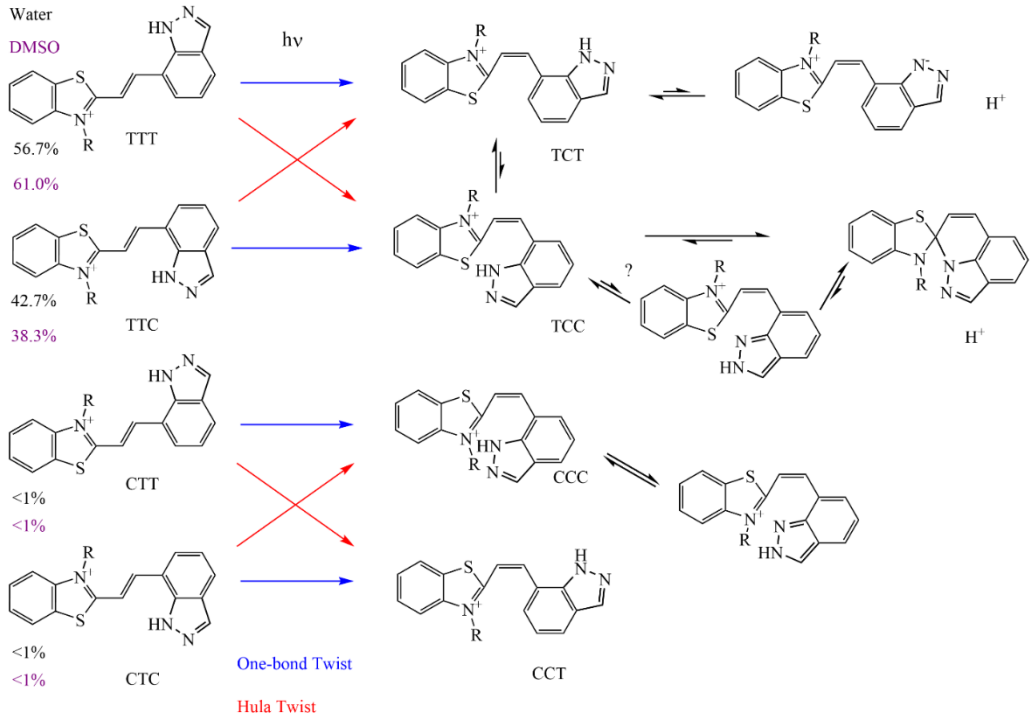



**Figure D-1 Conformations and reaction pathways for indazole-photoacids. R is the sulfonate tail**

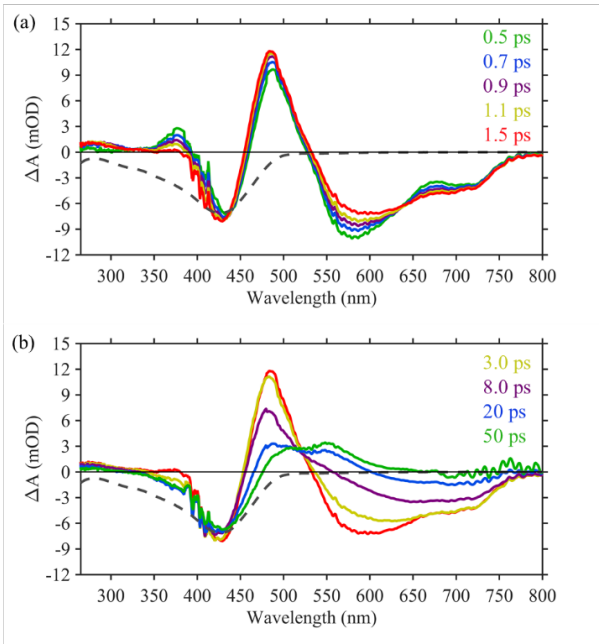



**Figure D-2** (a) TA spectra of phenylhydroxy-MCH in pH 5.5 at 0.5 ps (green), 0.7 ps (blue), 0.9 ps (purple), 1.1 ps (yellow), 1.5 ps (red), and scaled UV-vis of phenylhydroxy-MCH (grey dash). (b) TA spectra of phenylhydroxy-MCH at 1.5 ps (red), 3.0 ps (yellow), 8.0 ps (purple), 20 ps (blue) and 50 ps (green), scaled UV-vis of phenylhydroxy-MCH (grey dash).

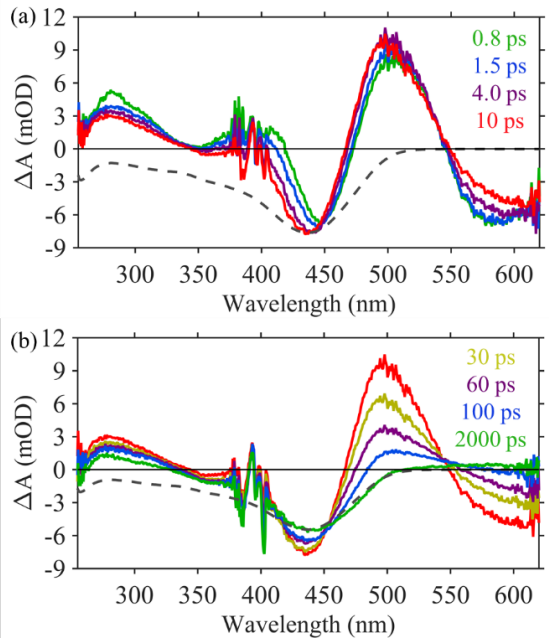
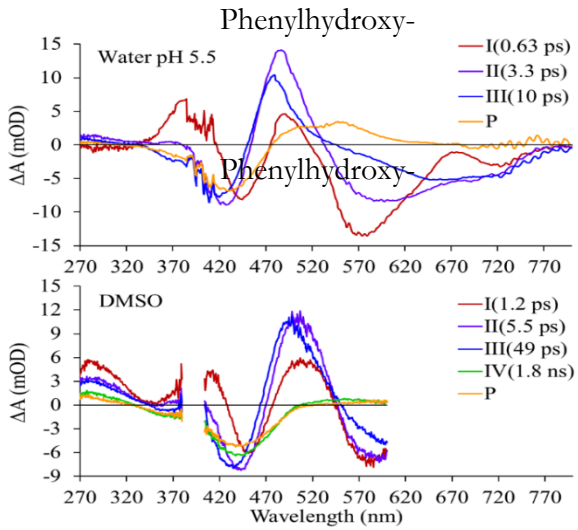



**Figure D-3** TA spectra of phenylhydroxy MCH in DMSO at early times (a) 0.8 ps (green), 1.5 ps (blue), 4.0 ps (purple), 10 ps (red). Longer times (b) 10 ps (red), 30 ps (yellow), 60 ps (purple), 100 ps (blue) and 2000 ps (green). Scaled UV-vis of phenylhydroxy MCH (grey dash).



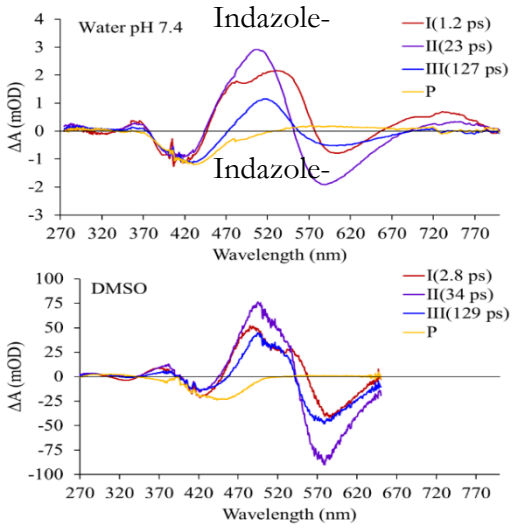




**Figure D-4** Evolution associated difference spectra (EADS) representing the time dependent evolution of the spectrum of the excited state population following excitation of left) trans-phenylhydroxy-MCH at 404 nm in aqueous solution buffered at pH 5.5 (top panel) and in DMSO (bottom panel), right) trans-indazole-MCH at 404 nm in aqueous solution buffered at pH 7.4 (top panel) and in DMSO (bottom panel)

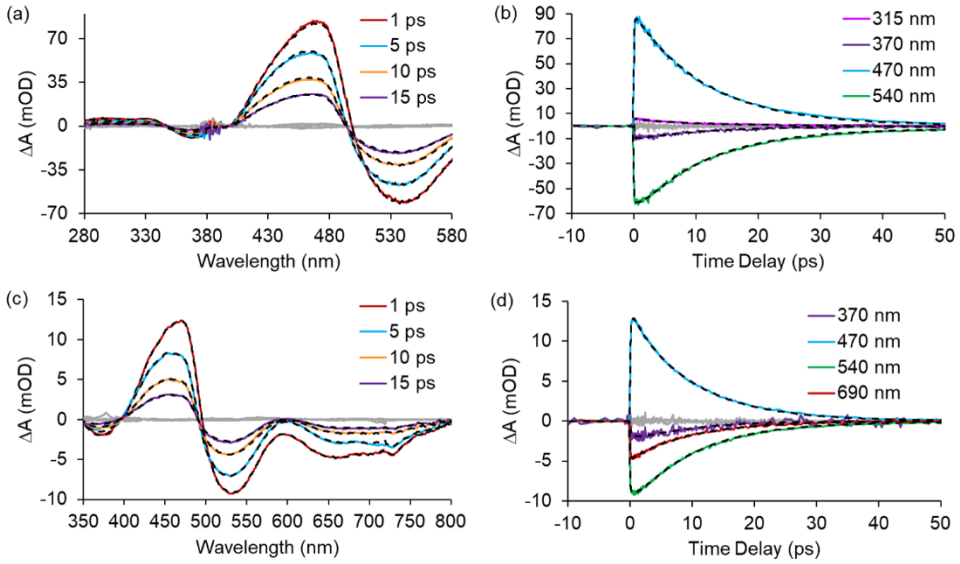



**Figure D-5** Transient absorption data obtained following excitation of the trans-phenylhydroxy-MC$^{-}$ conjugate base at 530 nm. The data are as indicated in the legends and the black dashed lines represent the fit to the data obtained in a global analysis as described in the text. (a) Transient spectra at select time delays obtained with a UV continuum. (b) Kinetic traces at select wavelengths obtained using the UV continuum. (c) Transient spectra at select time delays obtained with a visible continuum. (b) Kinetic traces at select wavelengths obtained using the visible continuum.

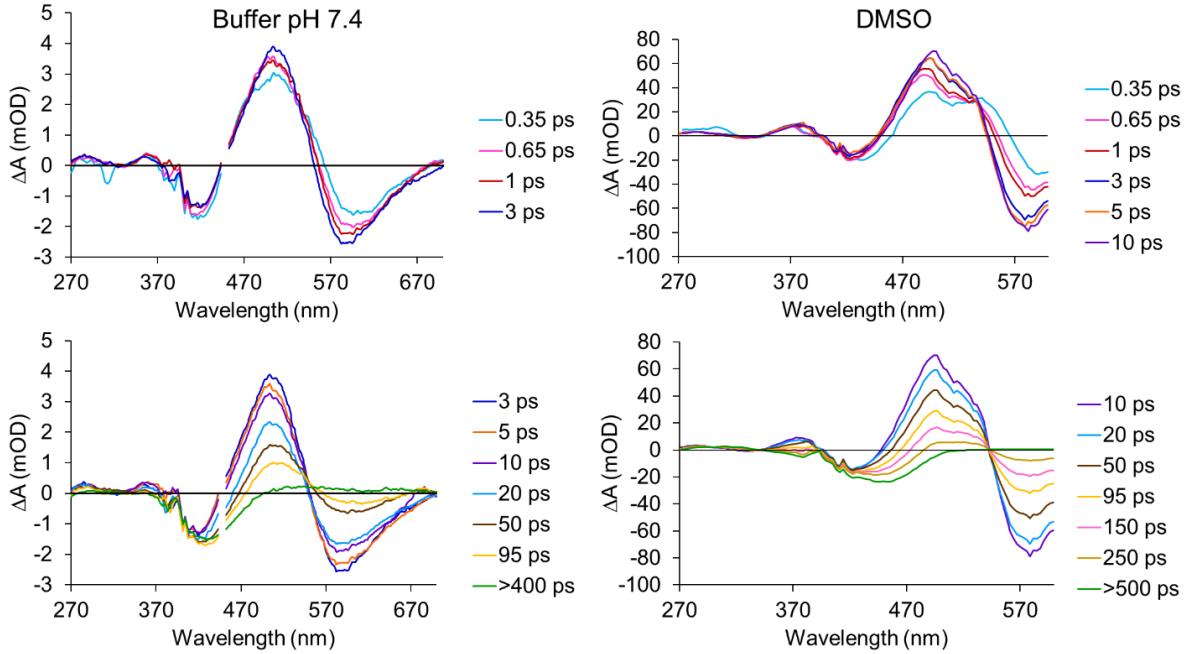



**Figure D-6** TA spectra of Indazole MCH in pH 7.4 buffer (left) and in anhydrous DMSO (right) averaged around the indicated time delays

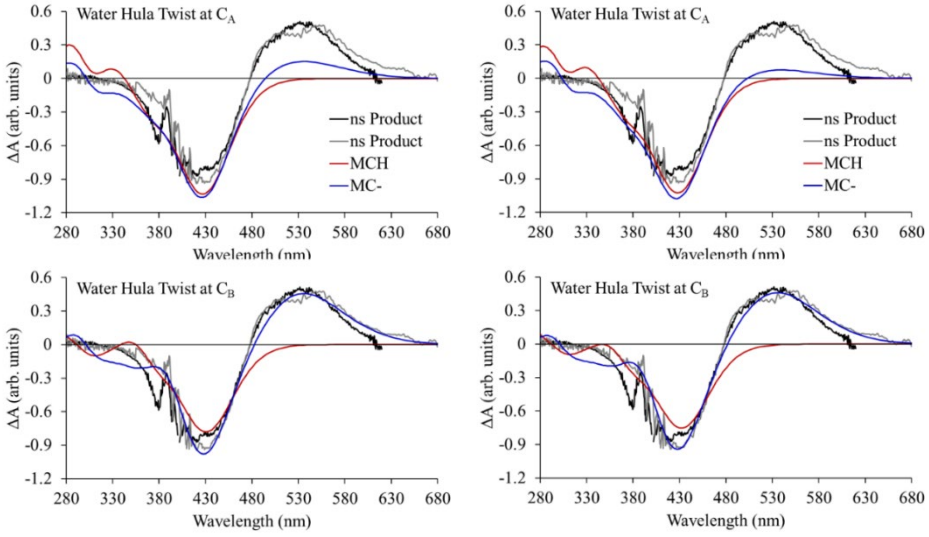



**Figure D-8** Comparison of the measured (black, gray lines) and simulated product difference spectra in water. The product spectra come from two independent measurements. The blue lines represent the simulated difference spectrum for the formation of *cis*-phenylhydroxy-MC$^-$ via the indicated hula twist mechanism. The red lines represent the simulated difference spectra for the formation of *cis*-phenylhydroxy-MCH via the same hula twist mechanism.

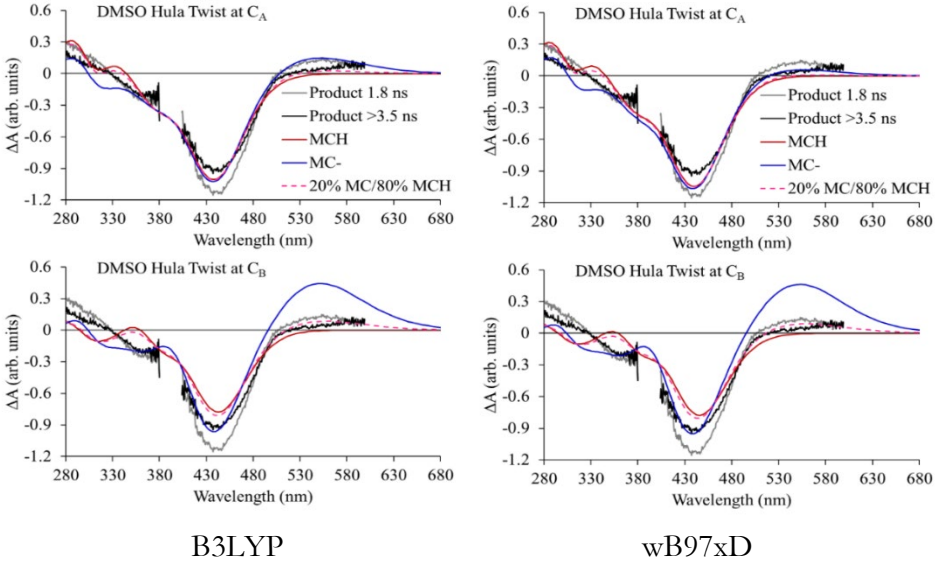



**Figure D-10** Comparison of the measured (black, gray lines) and simulated product difference spectra in DMSO. The product spectra represent the two long-lived spectral components. The blue lines represent the simulated difference spectrum for the formation of *cis*-phenylhydroxy-MC$^-$ via the indicated hula twist mechanism. The red lines represent the simulated difference spectra for the formation of *cis*-phenylhydroxy-MCH via the same hula twist mechanism. The dashed pink line in assumes 20% deprotonation.

**Table D-1 B3LYP+PCM(H2O)/6-31+G* geometry data**

| Conformer | Helicity 1 | Helicity 2 | G3LARGE | DLPNO-CCSD(T) | H | Original Entropy | Corrected Entropy | G (B3LYP) | G (DLPNO-CCSD(T)) | TOR 1 | TOR 2 | TOR 3 | Imaginary Freq |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TTT-MCH | F | F | -543541.9 | -542376.5 | 229.6 | 0.138 | 0.136 | -543352.1 | -542147.0 | -178.6 | -179.7 | -178.4 | 0 |
| CTT-MCH | P | F | -543540.5 | -542375.3 | 229.5 | 0.138 | 0.135 | -543350.4 | -542145.9 | 21.9 | 178.9 | -172.5 | 0 |
| CTT-MCH | M | F | -543540.4 | -542375.3 | 229.3 | 0.138 | 0.135 | -543350.7 | -542146.0 | -21.3 | -177.7 | 178 | 0 |
| TTC-MCH | F | F | -543541.8 | -542376.8 | 229.6 | 0.137 | 0.135 | -543351.5 | -542147.2 | -179.2 | -179.7 | 0.9 | 0 |
| CTC-MCH | P | F | -543540.3 | -542375.8 | 229.4 | 0.136 | 0.134 | -543350.1 | -542146.4 | 26.4 | 177.4 | 5.2 | 0 |
| CTC-MCH | M | F | -543540.3 | -542375.6 | 229.6 | 0.136 | 0.134 | -543349.9 | -542146.1 | -23.7 | -178 | -2.4 | 0 |
| TCT-MCH | P | P | -543531.4 | -542372.0 | 229.0 | 0.136 | 0.134 | -543341.6 | -542143.1 | 121.5 | -11.5 | 152.7 | 0 |
| TCT-MCH | M | M | -543531.4 | -542372.2 | 229.1 | 0.135 | 0.134 | -543341.4 | -542143.1 | -120 | 11.3 | -152.2 | 0 |
| CCT-MCH | P | M | -543534.9 | -542374.4 | 229.0 | 0.135 | 0.134 | -543344.9 | -542145.4 | 44.3 | 10.1 | -155.9 | 0 |
| CCT-MCH | M | P | -543534.8 | -542374.2 | 229.1 | 0.135 | 0.134 | -543344.8 | -542145.1 | -44.1 | -10.7 | 156.8 | 0 |
| TCC-MCH | P | M | -543530.6 | -542372.8 | 229.2 | 0.136 | 0.133 | -543340.3 | -542143.6 | 122.3 | -9.6 | -28.2 | 0 |
| TCC-MCH | M | P | -543530.5 | -542372.8 | 228.8 | 0.136 | 0.134 | -543340.9 | -542144.0 | -118.6 | 8.8 | 26.2 | 0 |
| CCC-MCH | P | P | -543534.3 | -542374.5 | 229.2 | 0.134 | 0.134 | -543344.2 | -542145.4 | 41.4 | 10 | 25.3 | 0 |
| CCC-MCH | M | M | -543534.2 | -542374.7 | 229.0 | 0.135 | 0.134 | -543344.2 | -542145.7 | -41.4 | -10.4 | -25 | 0 |
| TTT-MC | F | F | -543252.2 | -542081.5 | 220.6 | 0.141 | 0.137 | -543071.8 | -541861.0 | -179.1 | -179.7 | -179.1 | 0 |
| CTT-MC | F | F | -543250.6 | -542079.5 | 220.7 | 0.141 | 0.136 | -543069.8 | -541858.9 | -4 | 179.5 | 179.6 | 0 |
| TTC-MC | F | F | -543252.4 | -542081.8 | 220.7 | 0.139 | 0.136 | -543071.4 | -541861.2 | -179.6 | -179.7 | 0.5 | 0 |
| CTC-MC | P | F | -543250.5 | -542079.6 | 220.7 | 0.139 | 0.135 | -543069.4 | -541859.0 | 13 | 178.9 | 3 | 0 |
| CTC-MC | F | F | -543250.4 | -542079.5 | 220.6 | 0.139 | 0.135 | -543069.4 | -541858.9 | 0.3 | -170.9 | 8.5 | 0 |
| TCT-MC | P | P | -543237.7 | -542069.0 | 219.6 | 0.133 | 0.135 | -543057.6 | -541849.5 | 166.4 | -34.4 | 166.7 | 1 |
| TCT-MC | M | M | -543237.8 | -542068.5 | 219.5 | 0.133 | 0.135 | -543057.7 | -541849.0 | -166.9 | 35.2 | -166.2 | 1 |
| CCT-MC | P | F | -543242.8 | -542074.4 | 219.7 | 0.132 | 0.130 | -543061.1 | -541854.7 | 30.3 | 22.8 | -172.1 | 0 |
| CCT-MC | M | F | -543242.7 | -542074.3 | 219.6 | 0.133 | 0.136 | -543062.8 | -541854.8 | -30.4 | -22.7 | 172.6 | 1 |
| SP | P | F | -543249.5 | - | 220.3 | 0.131 | 0.129 | -543066.8 | - | 121 | -0.7 | -0.5 | 0 |
| SP | M | F | -543249.5 | - | 220.4 | 0.132 | 0.129 | -543067.0 | - | -122 | 1.1 | 0.7 | 0 |
| CCC-MC | P | P | -543242.9 | -542076.8 | 220.3 | 0.136 | 0.134 | -543061.6 | -541856.6 | 32.4 | 20.7 | 16.4 | 0 |
| CCC-MC | M | M | -543242.8 | -542076.8 | 220.2 | 0.137 | 0.134 | -543061.9 | -541856.6 | -31.8 | -20.8 | -15.6 | 0 |

**Table D-2 wB97xD +PCM(H2O)/6-31+G* geometry data**

| Conformer | Helicity 1 | Helicity 2 | G3LARGE | H | Original Entropy | Corrected Entropy | G (wB97xD) | TOR 1 | TOR 2 | TOR 3 | Imaginary Freq. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| TTT-MCH | F | F | -543356 | 231.8 | 0.136 | 0.135 | -543164.0 | 176.7 | 178.4 | 172.9 | 0 |
| CTT-MCH | P | M | -543355 | 232.4 | 0.134 | 0.133 | -543161.6 | 32.8 | 173.4 | -161.8 | 0 |
| CTT-MCH | M | F | -543355 | 231.7 | 0.135 | 0.134 | -543162.6 | -30.6 | -176.2 | 174.7 | 0 |
| TTC-MCH | F | F | -543356 | 232.3 | 0.134 | 0.132 | -543162.7 | 178.2 | -179.1 | -1.1 | 0 |
| CTC-MCH | P | F | -543355 | 232.1 | 0.134 | 0.132 | -543161.5 | 34.7 | 177.1 | 5.8 | 0 |
| CTC-MCH | F | M | -543354 | 231.7 | 0.136 | 0.134 | -543161.0 | -0.1 | -176.4 | -34.6 | 0 |
| | | | | | | | | | | | |
| TCT-MCH | P | P | -543352 | 232.4 | 0.131 | 0.130 | -543157.1 | 106.9 | -6 | 135.8 | 0 |
| TCT-MCH | M | M | -543351 | 232.1 | 0.132 | 0.131 | -543157.4 | -113.9 | 7.5 | -139.6 | 0 |
| CCT-MCH | P | M | -543352 | 232.4 | 0.132 | 0.131 | -543157.5 | 47.4 | 5.8 | -141.3 | 0 |
| CCT-MCH | M | P | -543354 | 231.8 | 0.133 | 0.131 | -543160.2 | -47.8 | -6.6 | 152.5 | 0 |
| TCC-MCH | P | M | -543351 | 232.4 | 0.132 | 0.131 | -543156.4 | 122 | -7.5 | -32.4 | 0 |
| TCC-MCH | M | P | -543350 | 231.5 | 0.134 | 0.132 | -543157.3 | -116.6 | 5.4 | 26.2 | 0 |
| CCC-MCH | P | P | -543353 | 232.0 | 0.133 | 0.132 | -543160.0 | 44.7 | 5.3 | 30.6 | 0 |
| CCC-MCH | M | M | -543353 | 232.8 | 0.131 | 0.129 | -543158.5 | -45.4 | -5.9 | -30.1 | 0 |
| | | | | | | | | | | | |
| TTT-MC | F | F | -543064 | 222.7 | 0.137 | 0.135 | -542880.7 | 179.4 | 176.6 | 178.9 | 0 |
| CTT-MC | F | F | -543062 | 223.0 | 0.136 | 0.134 | -542878.2 | -10 | -178.1 | 178.7 | 0 |
| TTC-MC | F | F | -543064 | 223.0 | 0.136 | 0.134 | -542880.1 | 176.8 | -178.3 | -0.5 | 0 |
| CTC-MC | P | F | -543062 | 223.6 | 0.134 | 0.132 | -542877.0 | 14.1 | 176.4 | 0.3 | 0 |
| CTC-MC | F | P | -543062 | 223.3 | 0.135 | 0.132 | -542877.5 | 0.3 | 175.2 | 29.2 | 0 |
| | | | | | | | 0.0 | | | | |
| TCT-MC | P | P | -543051 | 222.2 | 0.131 | 0.133 | -542868.2 | 146.4 | -21.6 | 159.1 | 1 |
| TCT-MC | M | M | -543052 | 222.4 | 0.129 | 0.132 | -542867.7 | -148.1 | 23.1 | -160.6 | 1 |
| CCT-MC | P | M | -543055 | 222.4 | 0.129 | 0.132 | -542871.7 | 39.3 | 16.9 | -164.4 | 1 |
| CCT-MC | M | P | -543055 | 221.9 | 0.134 | 0.136 | -542873.3 | -34.7 | -17.9 | 165.7 | 1 |
| SP | P | F | -543072 | 222.9 | 0.128 | 0.127 | -542886.2 | 138.8 | -3.6 | -7.9 | 0 |
| SP | M | F | -543073 | 223.0 | 0.128 | 0.126 | -542886.7 | -140.1 | 4.6 | 8.4 | 0 |
| CCC-MC | P | P | -543058 | 223.3 | 0.132 | 0.130 | -542873.2 | 36.7 | 16.6 | 20 | 0 |
| CCC-MC | M | M | -543058 | 222.8 | 0.135 | 0.133 | -542874.2 | -38.9 | -17 | -19.9 | 0 |

**Table D-3 wB97xD + PCM(DMSO) 6-31+G* geometry data**

| Conformer | Helicity 1 | Helicity 2 | G3LARGE | H | Original Entropy | Corrected Entropy | G (wB97xD) | TOR 1 | TOR 2 | TOR 3 | Imaginary Freq |
|---|---|---|---|---|---|---|---|---|---|---|---|
| TTT-MCH | F | F | -543356 | 231.755 | 0.136538 | 0.134862 | -543164.1 | 176.7 | 178.4 | 172.9 | 0 |
| CTT-MCH | F | M | -543355 | 232.406 | 0.134001 | 0.132582 | -543161.6 | 2.1 | 161 | -155.9 | 0 |
| CTT-MCH | M | F | -543355 | 231.665 | 0.135475 | 0.133564 | -543162.6 | -30.6 | -176.2 | 174.7 | 0 |
| TTC-MCH | F | F | -543356 | 232.309 | 0.134093 | 0.132402 | -543162.7 | 178.2 | -179.1 | -1.1 | 0 |
| CTC-MCH | P | F | -543355 | 232.109 | 0.133724 | 0.132011 | -543161.5 | 34.7 | 177.1 | 5.8 | 0 |
| CTC-MCH | F | M | -543354 | 231.726 | 0.135739 | 0.133813 | -543161.0 | -0.1 | -176.4 | -34.6 | 0 |
| TCT-MCH | P | P | -543352 | 232.355 | 0.131238 | 0.129892 | -543157.1 | 106.9 | -6 | 135.9 | 0 |
| TCT-MCH | M | M | -543351 | 232.068 | 0.132433 | 0.131081 | -543157.3 | -114 | 7.5 | -139.7 | 0 |
| CCT-MCH | P | M | -543352 | 232.43 | 0.131966 | 0.130621 | -543157.5 | 47.3 | 5.9 | -141.3 | 0 |
| CCT-MCH | M | P | -543354 | 231.851 | 0.132549 | 0.131194 | -543160.2 | -47.8 | -6.6 | 152.6 | 0 |
| TCC-MCH | P | M | -543351 | 232.37 | 0.13199 | 0.13064 | -543156.4 | 122 | -7.5 | -32.4 | 0 |
| TCC-MCH | M | P | -543350 | 231.489 | 0.133755 | 0.132392 | -543157.3 | -116.6 | 5.4 | 26.2 | 0 |
| CCC-MCH | P | P | -543353 | 231.988 | 0.133289 | 0.131934 | -543160.0 | 44.7 | 5.3 | 30.6 | 0 |
| CCC-MCH | M | M | -543353 | 232.776 | 0.130573 | 0.129235 | -543158.5 | -45.4 | -5.9 | -30.1 | 0 |
| TTT-MC | F | F | -543064 | 222.658 | 0.137135 | 0.135193 | -542880.7 | 179.4 | 176.6 | 178.9 | 0 |
| CTT-MC | F | F | -543062 | 222.985 | 0.136231 | 0.133653 | -542878.2 | -9.9 | -178.1 | 178.7 | 0 |
| TTC-MC | F | F | -543064 | 222.976 | 0.135977 | 0.133688 | -542880.1 | 176.9 | -178.3 | -0.5 | 0 |
| CTC-MC | P | F | -543062 | 223.63 | 0.133946 | 0.132321 | -542877.0 | 14 | 176.5 | 0.3 | 0 |
| CTC-MC | F | P | -543062 | 223.296 | 0.135195 | 0.132486 | -542877.5 | 0.3 | 175.2 | 29.2 | 0 |
| | | | 0 | | | | 0.0 | | | | |
| TCT-MC | P | P | -543051 | 222.218 | 0.130722 | 0.133068 | -542868.1 | 147.1 | -22.2 | 159.6 | 1 |
| TCT-MC | M | M | -543052 | 222.386 | 0.128866 | 0.13171 | -542867.8 | -149.9 | 24.3 | -161.8 | 1 |
| CCT-MC | P | M | -543055 | 222.277 | 0.129737 | 0.133003 | -542871.9 | 38.8 | 17.4 | -164.7 | 1 |
| CCT-MC | M | P | -543055 | 222.021 | 0.13122 | 0.134308 | -542872.6 | -33.8 | -18.5 | 166.2 | 1 |
| SP | P | F | -543072 | 222.885 | 0.127891 | 0.12659 | -542886.2 | 138.8 | -3.6 | -7.9 | 0 |
| SP | M | F | -543070 | 223.036 | 0.127741 | 0.126208 | -542884.2 | -140.1 | 4.6 | 8.4 | 0 |
| CCC-MC | P | P | -543058 | 223.267 | 0.132346 | 0.130313 | -542873.2 | 36.6 | 16.7 | 19.9 | 0 |
| CCC-MC | M | M | -543058 | 222.829 | 0.135639 | 0.133664 | -542874.4 | -38.8 | -17.2 | -19.8 | 0 |

**Table D-4 indazole-MCH B3LYP +PCM (H2O)/6-31+G* geometry data**

| Conformer | Helicity 1 | Helicity 2 | G3LARGE | H | Original Entropy | Corrected Entropy | G (B3LYP) | TOR 1 | TOR 2 | TOR 3 | Imaginary Freq. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ttt-MCH | F | F | -764828 | 184.553 | 0.133301 | 0.129698 | -764683 | 178.734 | 179.553 | -178.242 | 0 |
| ttt-MCH | F | M | -764828 | 183.807 | 0.128748 | 0.1314 | -764682 | 175.75 | -179.2 | -174.7 | 1 |
| ctt-MCH | P | F | -764826 | 184.632 | 0.131094 | 0.128805 | -764679 | 27.7 | 177.4 | -176.3 | 0 |
| ctt-MCH | M | P | -764826 | 184.543 | 0.131793 | 0.128944 | -764680 | -26.7 | -179.1 | 169.4 | 0 |
| ttc-MCH | M | P | -764828 | 184.495 | 0.133191 | 0.130191 | -764682 | -169.8 | -179 | 10.9 | 0 |
| ttc-MCH | F | M | -764828 | 184.542 | 0.132929 | 0.12992 | -764682 | 175.851 | 179.344 | -11.205 | 0 |
| ctc-MCH | P | P | -764825 | 184.638 | 0.130511 | 0.128264 | -764679 | 30.3 | 177.8 | 14.8 | 0 |
| ctc-MCH | M | M | -764825 | 184.57 | 0.131289 | 0.12882 | -764679 | -31 | -177.9 | -15.2 | 0 |
| | | | | | | | | | | | |
| tct-MCH | P | P | -764822 | 184.537 | 0.129674 | 0.127962 | -764675 | 151.7 | -12.1 | 146.5 | 0 |
| tct-MCH | M | M | -764822 | 184.553 | 0.129616 | 0.127994 | -764675 | -151.3 | 12.3 | -147 | 0 |
| cct-MCH | P | M | -764821 | 184.479 | 0.128572 | 0.127126 | -764674 | 49.6 | 7.3 | -153 | 0 |
| cct-MCH | M | P | -764821 | 184.477 | 0.128355 | 0.126874 | -764674 | -49.9 | -7.6 | 153.2 | 0 |
| tcc-MCH | P | M | -764821 | 184.41 | 0.130705 | 0.128759 | -764675 | 163.9 | -11.7 | -41.7 | 0 |
| tcc-MCH | M | P | -764821 | 184.439 | 0.129227 | 0.127583 | -764674 | -158.5 | 12 | 40.2 | 0 |
| ccc-MCH | | P | -764819 | 184.498 | 0.127487 | 0.126188 | -764672 | 42.1 | 7.5 | 30.8 | 0 |
| ccc-MCH | M | M | -764819 | 184.363 | 0.128551 | 0.127124 | -764672 | -42.9 | -7.3 | -30.9 | 0 |
| | | | | | | | | | | | |
| ttt-MC | F | F | -764373 | 174.832 | 0.129579 | 0.130581 | -764236 | -178.6 | -179.9 | -179.2 | 1 |
| ctt-MC | M | F | -764370 | 174.888 | 0.128234 | 0.129547 | -764233 | -21.4 | -179.1 | 175.7 | 1 |
| ctt-MC | P | F | -764370 | 174.912 | 0.127299 | 0.129338 | -764233 | 22.7 | 177.4 | -176.3 | 1 |
| ttt-MC | F | F | -764376 | 175.466 | 0.134699 | 0.13004 | -764240 | 178.7 | 179.4 | -0.1 | 0 |
| ctc-MC | P | F | -764372 | 174.947 | 0.127123 | 0.128849 | -764235 | 22.9 | 177.6 | 2.8 | 1 |
| ctc-MC | M | F | -764372 | 175.116 | 0.125961 | 0.128304 | -764234 | -19.7 | -177.1 | -1.5 | 1 |

**Table D-5 Indazole wB97xD +PCM (H2O)/6-31+G* geometry data**

| Conformer | Helicity 1 | Helicity 2 | G3LARGE | H | Original Entropy | Corrected Entropy | G (B3LYP) | TOR 1 | TOR 2 | TOR 3 | Imaginary Freq. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ttt-MCH | P | F | -764633 | 186.774 | 0.133 | 0.129 | -764484.2 | 173.4 | 179.9 | -178.7 | 0 |
| ttt-MCH | F | M | -764633 | 186.679 | 0.134 | 0.129 | -764484.5 | -178.8 | -178.7 | -169.5 | 0 |
| ctt-MCH | P | F | -764632 | 186.596 | 0.132 | 0.129 | -764482.7 | 35.3 | 177.7 | -176.6 | 0 |
| ctt-MCH | M | P | -764632 | 186.671 | 0.131 | 0.128 | -764482.3 | -36.0 | -179.0 | 164.9 | 0 |
| ttc-MCH | M | P | -764633 | 186.651 | 0.132 | 0.129 | -764484.1 | -161.8 | -178.9 | 19.5 | 0 |
| ttc-MCH | M | M | -764633 | 186.775 | 0.132 | 0.129 | -764483.9 | -167.5 | -178.7 | -18.0 | 0 |
| ctc-MCH | P | P | -764632 | 186.551 | 0.131 | 0.128 | -764482.4 | 36.9 | 176.9 | 18.0 | 0 |
| ctc-MCH | M | M | -764632 | 186.616 | 0.130 | 0.128 | -764482.4 | -38.5 | -177.3 | -21.7 | 0 |
| | | | | | | | | | | | |
| tct-MCH | P | P | -764630 | 186.787 | 0.128 | 0.127 | -764480.1 | 128.7 | -8.2 | 142.6 | 0 |
| tct-MCH | M | M | -764630 | 186.447 | 0.129 | 0.128 | -764481.1 | -129.7 | 8.0 | -142.2 | 0 |
| cct-MCH | P | M | -764631 | 186.9 | 0.126 | 0.125 | -764480.9 | 53.0 | 3.6 | -139.9 | 0 |
| cct-MCH | M | P | -764631 | 186.731 | 0.126 | 0.125 | -764481.1 | -52.5 | -3.9 | 139.8 | 0 |
| tcc-MCH | P | M | -764630 | 186.601 | 0.127 | 0.126 | -764480.3 | 137.8 | -6.8 | -47.3 | 0 |
| tcc-MCH | M | P | -764630 | 186.673 | 0.127 | 0.126 | -764480.4 | -139.4 | 7.0 | 48.5 | 0 |
| ccc-MCH | | P | -764630 | 186.615 | 0.126 | 0.125 | -764480.2 | 47.9 | 2.0 | 40.1 | 0 |
| ccc-MCH | M | M | -764630 | 186.547 | 0.126 | 0.124 | -764480.3 | -48.3 | -2.2 | -40.5 | 0 |
| | | | | | | | | | | | |
| ttt-MC | F | F | -764330 | 176.925 | 0.129 | 0.129 | -764190.8 | -174.2 | -179.8 | -176.3 | 1 |
| ctt-MC | M | F | -764328 | 176.245 | 0.121 | 0.124 | -764187.6 | -30.5 | -177.7 | 177.7 | 1 |
| ctt-MC | P | F | -764328 | 177.01 | 0.127 | 0.128 | -764188.3 | 30.2 | 177.2 | -179.0 | 1 |
| ttt-MC | F | F | -764332 | 177.453 | 0.134 | 0.129 | -764192.8 | 179.0 | 179.3 | 0.0 | 0 |
| ctc-MC | P | F | -764330 | 177.124 | 0.125 | 0.128 | -764189.8 | 32.3 | 176.8 | 2.6 | 1 |
| ctc-MC | M | F | -764330 | 176.894 | 0.126 | 0.128 | -764190.2 | -29.2 | -176.4 | -1.6 | 1 |
| | | | | | | | | | | | |
| tct-MC | P | P | -764325 | 176.998 | 0.124 | 0.127 | -764185.4 | 131.1 | -9.7 | 144.1 | 1 |
| tct-MC | M | M | -764326 | 176.741 | 0.125 | 0.128 | -764186.1 | -135.3 | 9.9 | -144.4 | 1 |
| cct-MC | P | M | -764327 | 177.084 | 0.123 | 0.127 | -764186.7 | 51.0 | 5.2 | -142.5 | 1 |
| cct-MC | M | P | -764327 | 176.842 | 0.123 | 0.126 | -764187.1 | -51.0 | -5.4 | 142.0 | 1 |
| ccc-MC | P | P | -764328 | 177.344 | 0.127 | 0.125 | -764187.2 | 48.5 | 6.0 | 25.8 | 0 |
| ccc-MC | M | M | -764328 | 177.485 | 0.127 | 0.126 | -764187.6 | -49.0 | -5.8 | -26.0 | 0 |
| tcc-MC | M | P | -764327 | 177.541 | 0.129 | 0.127 | -764186.2 | -132.9 | 9.8 | 30.5 | 0 |
| tcc-MC | P | M | -764327 | 177.371 | 0.129 | 0.127 | -764186.5 | 132.8 | -9.8 | -31.8 | 0 |