DISSERTATION

DEVELOPMENT AND APPLICATION OF CONFORMATIONAL METHODOLOGIES: ELICITING ENTHALPIC GLOBAL MINIMA AND REACTION PATHWAYS

Submitted by

Joseph T. Allison

Department of Chemistry

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2014

Doctoral Committee:

Advisor: Anthony Rappé

Steven Strauss
Nancy Levinger
Matthew Shores
Richard Slayden

ABSTRACT

DEVELOPMENT AND APPLICATION OF CONFORMATIONAL METHODOLOGIES: ELICITING ENTHALPIC GLOBAL MINIMA AND REACTION PATHWAYS

The information granted by assembling the global minimum and low-enthalpy population of a chemical species or ensemble can be utilized to great effect across all fields of chemistry. With this population, otherwise impossible tasks including (but not limited to) reaction pathway characterization, protein folding, protein-ligand docking, and constructing the entropy to characterize free energy surfaces becomes a reasonable undertaking. For very small systems (single molecule with 1-3 torsions) generating the low-enthalpy population is a trivial task. However as the system grows, the task exponentially increases in difficulty. This dissertation will detail the two sides of this problem, generating the low-energy population of larger and more complex species and then utilizing those populations to garner a greater understanding of their systems.

The first discussion describes a new model, Surface Editing Molecular Dynamics (SEMD), which aids in accelerating conformational searching by removing minima from the potential energy surface by adding Gaussian functions. Accompanying this new method are a multitude of new tools that can be utilized to aid in molecular dynamics simulations. The first of these tools, named CHILL, performs a projection of unproductive degrees of freedom from the molecular dynamics velocity to smooth atomic motions without artificially constraining those degrees of freedom. Another tool, Conjugate Velocity Molecular Dynamics (CVMD), rigorously generates a list of productive velocities via the biorthogonalization of local modes with a vector

representation of previously explored conformational minima. In addition to these tools, a new description of distance in torsional space was developed to provide a robust means of conformational uniqueness. With each of these tools working in concert, the global minimum and associated low-enthalpy population of conformations have been obtained for various benchmark species.

The second section discusses the application of conformational searching and the subsequent electronic structure calculations to characterize the reaction pathway for the ruthenium tris(2,2'-bipyridine) photocatalyzed [2+2] cycloaddition of aromatically substituted bis(enones). The APFD hybrid density functional is used along with a 6-311+g* basis and a PCM solvent model. The reaction is computed to proceed through a rate-limited formation of a cyclopentyl intermediate. Lithium tetrafluoroborate is found to facilitate initial bis(enone) reduction as well as final product distribution. In addition, aromatic substituents are found to impact both initial reduction and final product distribution.

DEDICATION

For my wife, Rebecca Allison,

my parents, Neil and Amelia Allison,

my advisor, Anthony Rappé,

and my friends, Brandon Thorne, Cameron Byers, and Nathan O’Brien.

TABLE OF CONTENTS

CHAPTER 1: AN INTRODUCTION TO CONFORMATIONAL SEARCHING & SAMPLING

There exists a class of computational problems wherein their proof must be verifiable in polynomial time with a nondeterministic Turing machine[1]. These problems are referred to as NP (nondeterministic polynomial time) problems. The most difficult of these are classified as NP-hard. One such NP-hard problem, known as the traveling salesman[2], is defined as: given a list of cities, determine the shortest path that visits every city once and then returns to the starting city (Figure 1.1). This problem applies within the realm of chemistry as related to the concept of the potential energy surface (PES) (Figure 1.2) and is known as the multiple minima problem[3]. On any given PES there exist valleys and hills, the bottom of each valley represents a local minimum. When comparing the similarities between the traveling salesman problem and chemistry you obtain the premise behind conformational searching wherein the cities become the minima and the goal becomes efficiently and reliably constructing the population of the "best" minima from the PES without regard to the starting position. The "best" minima are typically those corresponding to the relatively low-enthalpy regimes (within approximately 3 kcal/mol of the lowest-enthalpy minimum on the surface, the global minimum).

In addition to conformational searching, where collecting the global minimum and low-enthalpy population of minima is the goal, another realm of computational chemistry, known as sampling, places special importance on vibrational effects and their probabilistic sampling. The goal of sampling is to reconstruct the relevant structural populations (not just the minima) of low-enthalpy. The relevant regions can range from distributions of structures around minima or the ensemble of structures that reside on a reaction coordinate. With these populations, a wide array of metrics can be ascertained, such as the entropic contribution to the free energy, which is

used (with the enthalpy) to compute the free energy surface.

There are a laundry list of methods that perform conformational searching and sampling that can be characterized by a combination of three subclasses: heuristic, stochastic, and systematic. A purely heuristic method utilizes knowledge gained during the course of the simulation to inform decision-making, stochastic methods rely on random chance, and systematic methods utilize a pre-determined algorithm to exhaustively reach the end goal.

The research presented in this dissertation utilizes and builds upon many facets of conformational searching and sampling. The second chapter details the development and application of a new conformational searching method, called surface editing molecular dynamics (SEMD). The third chapter focuses upon the application of SEMD to construct the configurational entropy of a species. The fourth chapter discusses the application of conformational searching to enable the characterization of a photocatalyzed cycloaddition reaction via electronic structure theory. Finally, the fifth chapter proposes an extension of SEMD to enable the efficient characterization of the low-enthalpy population of multi-body systems.

This first chapter focuses on describing relevant methods that make up the current state of conformational searching and sampling while also giving context with specific examples of their application. Section 1.1 discusses conformational searching techniques and section 1.2 discusses sampling techniques.

1.1 – Conformational Searching

Reliably and efficiently obtaining the global minimum and associated, thermally accessible, conformation space needed for a free energy assessment of a molecule, reaction sequence, or ensemble of molecules remains a challenge in fields ranging from drug discovery[4–6]

to industrial catalyst development[7]. The task is particularly daunting for systems containing rings and significant conformational flexibility. Adding the intermolecular degrees of freedom found in crystal packing, active site docking, transition state modeling, and protein folding only compounds the challenge. Conformational searching protocols utilizing systematic[3,8–12], Monte Carlo[8,13–17], molecular dynamics[18–24], genetic algorithms[11,25–30], and history-dependent algorithms[31–38] have been developed and are in current use.

1.1.A – Systematic Search. The most thorough set of approaches systematically scan through conformational space, attempting to collect low-enthalpy minima. A systematic grid search[3,8–12] (Figure 1.3.A) establishes a set of variables (internal or external), illustrated here with the most common, the torsional degrees of freedom, and systematically evaluates the energy for all possible combinations of torsions, Φ, with pre-defined torsional angle spacing $\Delta\Phi$. A second variation upon this method, known as the torsion tree-searching algorithm, iteratively scans through a list of possible torsions that are associated with the well-known minima of the torsions in question. Challenges for torsion tree-searching algorithms include selection of the torsions of interest and fineness of the torsional grid. In 1990, Saunders et al.[12] attempted to search for the lowest 262 conformations of cycloheptadecane (Figure 1.4.A) via a systematic torsion tree-search, as the lowest 262 enthalpy conformations reside within 3 kcal/mol of the global minimum (Figure 1.4.B). Since cycloheptadecane contains 17 C-C-C-C torsions, the search began with a coarse-grained approach with increments of 120° with a ring closure constraint of 0.5-5 Å. This yielded 20,000 starting point geometries. Instead of performing a costly minimization on all 20,000 structures, the number of starting points was reduced by removing geometries containing any high energy g+-g- arrays (adjacent C-C-C-C torsions $\phi_1 - \phi_2 = \pm 60^\circ - \mp 60^\circ$). Upon minimization of the remaining 405, structures 58 unique conformers were

obtained within the lowest 3 kcal/mol threshold. Allowing structures with up to one g+-g- array resulted in 7265 geometries which minimized to ~200 unique minima within the 3 kcal/mol threshold. By relaxing the ring closure constraint from an upper limit of 5 Å to 7.5 Å, ~2500 more geometries were generated which resulted in 11 new unique conformers within the energy threshold. In total, 211 of the 262 lowest energy conformers of cycloheptadecane were discovered. This study, among many, exemplifies that systematic searches suffer the distinct disadvantage that cost increases exponentially as the size of the system increases, precisely by $(360^{\circ}/\Delta\Phi)^{n}$-$(360^{\circ}/\Delta\Phi)^{n-1}$ where n is the number of torsional degrees of freedom. This exponential dependence renders systematic searches unfeasible for very large systems and quite challenging for ring systems where the ring closure condition eliminates the vast majority of trials.

1.1.B – Monte Carlo. Rather than systematically generating all possible combinations of torsions one can randomly choose the combinations of torsions to be sampled or apply random displacements to individual torsion or Cartesian coordinates[8,13–17] (Figure 1.3.B). This is the basis of Monte Carlo searching procedures. Dating from the early 1950's Monte Carlo procedures have been a workhorse conformational searching procedure[14].

In the previously mentioned paper by Saunders et al.[12] Monte Carlo was also utilized in their search for the 262 low-enthalpy conformers of cycloheptadecane. Both the random walk and usage-directed methods were employed for choosing the starting structure for each Monte Carlo step. The random walk method was configured to choose the resulting structure of the Monte Carlo step if its geometry was within 3 kcal/mol of the global minimum, otherwise the previous starting geometry was reused. The usage-directed method performed random torsional displacements on each conformer (as they were discovered) within 3 kcal/mol of the global

minimum an equal number of times, preferring conformers that had been used the fewest times. The random walk search and the usage-directed methods resulted in 237 and 260 of the 262 lowest 3 kcal/mol conformers, respectively.

1.1.C – Energy Driven. Energy driven searching methods such as quenched[19] or annealed[20] molecular dynamics (MD) have had wide use in conformational searching[18–24]. By integrating Newton's equations of motion[23], MD has the advantage of preferentially sampling low energy space relative to the high-energy space that Monte Carlo or systematic searches explore. Because energy is the driving force, MD is able to treat ring and non-rings systems with equal facility.

Molecular dynamics simulates a classical system by utilizing molecular position , $\boldsymbol{r}$, velocity, $\dot{\boldsymbol{r}}$, and acceleration, $\ddot{\boldsymbol{r}}$. Molecular motion is achieved by a Taylor series truncated at the second derivative (Eqn. (**1**.1)).

$$\boldsymbol{r}(t + \Delta t) = \boldsymbol{r}(t) + \dot{\boldsymbol{r}}(t)\Delta t + \frac{\ddot{\boldsymbol{r}}(t)\Delta t^2}{2} \qquad (1.1)$$

Where Δt is the length of the time interval. This equation can be modified into a more efficient form by numerically evaluating the derivatives to yield the Verlet equation (Eqn. (1.**2**)).

$$\boldsymbol{r}(t + \Delta t) = 2\boldsymbol{r}(t) - \boldsymbol{r}(t - \Delta t) + \Delta t^2 F(t)/m \qquad (1.2)$$

The force, $F(t)$, is obtained from evaluating the gradients of the functional forms present in a force field.

There are many force fields currently available[39] and they all define their potential energy as a sum of bonded and nonbonded interactions (Eqn. (1.**3**)).

$$V = V_{stretch} + V_{bend} + V_{torsion} + V_{vdW} + V_{electrostatics} \qquad (1.3)$$

The bonded interactions consist of the bond stretch ($V_{stretch}$), angle bend (V_{bend}), and dihedral angle torsion ($V_{torsion}$) terms. The nonbonded interactions include van der Waals (V_{vdW}) and electrostatic ($V_{electrostatics}$) terms. These are the primary building blocks that make up the potential energy where additional terms such as inversion, hydrogen bonding, and cross terms may be added depending on the force field. Additionally, the functional forms of each term and parameterization can vary quite dramatically to achieve accurate results for a specific range of chemical space.

With quenched dynamics, a MD simulation is run at constant temperature and structural aliquots from the simulation are minimized with the hope that new conformational space is being explored as a function of time. Another approach is to virtually anneal the system via annealed dynamics. Raising the temperature during the annealing process increases the probability of barrier-traversing events. Lowering the temperature allows the system to settle into lower energy conformations.

Saunders et al.[12] performed two quenched dynamics simulations in addition to Monte Carlo and systematic searches in an attempt to obtain the lowest-enthalpy 262 conformers of cycloheptadecane. Simulations were started from a structure ~1.5 kcal/mol higher than the global minimum. The first simulation ran for 14 days with an overall simulation time of 1 ns with a time step of 1.25 fs and temperature of 500 K. The second simulation also ran for 14 days but with a temperature of 1,000 K and an elapsed simulation time of 250 ps at 0.33 fs per dynamic step. They sampled 4,000 structures spaced uniformly in time from each trajectory. After minimization and duplicate elimination, the low temperature and high temperature simulations obtained 169 and 156 of the 262 structures, respectively.

Disadvantages of simulated annealing and quenched dynamics methods for conformational searching include 1) overcoming PES barriers that separate minima, 2) the inherent issue of lingering in and returning to previously discovered low-energy space, and 3) knowing when a new conformational well is being sampled (the challenge of choosing an optimal minimization criterion). These challenges plague all MD-based conformational searching methods. Additionally, disadvantages 1 and 3 are coupled in that an optimal minimization criterion would be a metric that provides knowledge that a new conformation is being sampled.

In the aforementioned Saunders et al. paper, quenched MD was performed on cycloheptadecane at 500 K with a time step of 1.25 fs for 1000 ps. 4000 structures uniformly sampled from the simulation were minimized. CH bonds were constrained via SHAKE during the dynamics. Only 169 of the 4000 minimized structures were within the set of the 262 lowest 3 kcal/mol conformers. The same calculation was performed at the higher temperature of 1000 K to facilitate sampling of a wider configurational space, however the resulting minimized structures were typically higher in energy.

1.1.D – Potential Energy Surface Smoothing. Rather than taking the potential surface as it is and working with the large number of conformational degrees of freedom and local minima, a number of methods have been developed that modify the potential surface by smoothing it in order to traverse the potential energy surface more readily with molecular mechanics minimizations.

The diffusion equation approach suggested by Piela, Kostrowicki and Scheraga[40] and developed by Hart, Pappu, and Ponder[41] utilizes modified force field equations. Force field

potential deformation is provided by insertion of Gaussian scaling terms. For example, a deformable torsion potential is given in Eqn. (1.**4**).

$$V_{torsion}(\phi,t)=\frac{1}{2}\sum_{j}V_j(1+\cos(\phi j+\Delta))e^{-j^2D_{torsion}t} \qquad (1.4)$$

In Eqn. (1.**4**), j is the periodicity of the torsional potential, V_j is the potential barrier, Δ is the phase offset, and ϕ the dihedral angle. The Gaussian scaling parameter t controls the degree of scaling across all types of force field terms, and the scaling coefficient $D_{torsion}$ along with companion scaling coefficients for electrostatics (Eqn. (1.**5**)) and van der Waals (Eqn. (1.**6**)) assure comparable scaling for the various terms.

$$\mathrm{E}_{charge}=\frac{q_iq_j}{4\pi\varepsilon_0 r_{ij}}$$

$$\mathrm{E}_{charge}(r_{ij,}t)=\frac{q_iq_j}{4\pi\varepsilon_0 r_{ij}}\,\mathrm{erf}\left(\frac{r_{ij}}{2\sqrt{D_{charge}t}}\right) \qquad (1.5)$$

Where q_i and q_j are the partial charges of atoms i and j and r_{ij} is the distance between those atoms.

$$\mathrm{E}_{vdw}=\sum_{i=1}^{n_{gauss}}a_ie^{-b_ir_{ij}^2}$$

$$\mathrm{E}_{vdw}(r_{ij,}t)=\sum_{i=1}^{n_{gauss}}\frac{a_i}{(1+4b_iD_{vdw}t)^{3/2}}\times\exp\left(\frac{-b_ir_{ij}^2}{1+4b_iD_{vdw}t}\right) \qquad (1.6)$$

with

$$a_i=a_i^{\circ}\varepsilon_0, \qquad b_i=b_i^{\circ}\left(\frac{2^{1/6}}{r_0}\right)^2$$

$$\varepsilon_0=\sqrt{\varepsilon_a\varepsilon_b}, \text{ and } r_0=\sqrt{r_ar_b}$$

Within Eqn. (1.**6**) ε_x and r_x are Lennard-Jones parameters for atom x. The values $\langle a_i^\circ, b_i^\circ \rangle$ are reference parameters chosen to fit a canonical Lennard-Jones function with well depth ε=4.184 kJ/mol and radius $\sigma = 1$ Å; n_{gauss} is the number of Gaussians used in the approximation.

In 1998 Pappu, Hart, and Ponder[41] utilized their diffusion equation method to perform potential surface smoothing with the MM2 force field to attempt to find the global minimum of cycloheptadecane. A normal mode local search procedure was used in addition to smoothing. Over the course of multiple minimizations the smoothing algorithm was adjusted to reduce the surface to a single minimum. This approach discovered the second lowest enthalpy conformer in 440 full-energy minimizations.

1.1.E – Genetic Algorithms. Since the early 1990s a number of genetic algorithm approaches have been applied to conformational searching[11,25–30]. Genetic algorithms perform "natural selection" on an initial set of randomly generated/minimized conformations as well as all subsequent conformations generated through mutation. The process of natural selection is derived from a pre-defined fitness function wherein low-energy structures are more likely to be selected over high-energy structures. Performing a crossover procedure between parents, the highest-fitness (usually lowest enthalpy) conformers, then creates new generations of structures. Crossover is the process of mixing complimentary attributes of each parent to form the child structure. The remaining attributes are used to create a second child structure. During crossover, it is also possible for mutation to occur wherein a new attribute arises. The new ensemble of children is evaluated on their fitness and the process repeated until some termination condition is met.

In 2010 Long, Tran, Adams, Darwen, and Smythe[30] reported an implementation of the genetic algorithm, population-based incremental learning algorithm (PBIL), designed to perform

conformational searches. PBIL was first developed by Shumeet Baluja in 1994. The PBIL algorithm forms each generation of conformers from a probability distribution biased towards the previous generation's fittest dihedral angles. With each new generation, each structure is minimized. After 500 minimization steps, if the energy is above a pre-defined cutoff the minimization is stopped, the conformer discarded, and the minimization attempt not counted towards the total full-energy minimization count. In order to maintain conformational diversity within the algorithm, a penalty energy was applied to newly generated conformations. The penalty energy term is shown in Eqn. (1.**7**).

$$e_{i,j}^{p} = e_{i,j} + \frac{\alpha}{R_1} + \frac{\beta}{R_2} \qquad (1.7)$$

Where $e_{i,j}$ is the energy of a conformer, α and β are constants used to control the strength of the penalty, and R_1 and R_2 are the root mean square deviation (RMSD) of the conformer from the accepted list of low-enthalpy conformations and the conformations in the current generation, respectively. The penalty then propagates into the biased probability distribution that the next generation of structures is generated from. Since the algorithm cannot natively handle ring structures, a ring-closing trial is added wherein the same criteria as Saunders et al.[12] was employed requiring the bond formed from the ring closure to be greater than 0.5 Å or less than 3.5 Å. With this algorithm the full set of 264 conformations was obtained in 7662 full-energy minimizations. The number of partial minimization attempts was not reported.

1.1.F – History Dependent. Some algorithms utilize the history of the simulation to aid the exploration of previously undiscovered conformational space.[31–38] There are a handful of "meta" algorithms, such as the tabu search algorithm[34–38]. With tabu searching all conformers within a single "step" from the starting conformer are evaluated and the lowest energy becomes to the new starting conformer. A "tabu" list is generated from these starting conformers and in

the future, those structures are forbidden to be re-visited. One difficulty of a tabu search is determining the optimal size of the list. If the list is too large, excessive time is taken iterating through the list looking for matches. If the list is too small, conformationally nearby history is forgotten and time is wasted revisiting those structures.

The gradient tabu search method (GTS)[37] and the following gradient only tabu search method (GOTS)[38] were developed by Stepanenko and Engels in 2007 and 2008 respectively. GTS utilizes both the gradient and Hessian to enhance searching for global minima of differential functions. Starting from a local minimum, the diagonal elements of the Hessian are used to estimate the displacement from the current location that follows the path of 'mildest ascent'. This is similar to methods commonly referred to as an eigenvector following technique[42]. This process is continued until a barrier has been crossed. The algorithm then utilizes a combination of the Steepest Descent and Quasi-Newton methods to minimize the structure.

To prevent minimization to already discovered minima, a tabu list is generated from the minimized structures as they are discovered. From the tabu list, regions around each entry are declared as tabu (if the search enters the region a minimization does not occur). Additionally, the overlap of the normalized vectors representing the directions starting from the current minimum to the original start of the search and to the next coordinate as a result of the displacement is evaluated. This overlap is then compared to a predetermined criterion and the direction is either accepted (if it is leading towards undiscovered space) or rejected (if it is leading backwards towards the start). Finally, the tabu list of structures maintains a maximum size via the first in/first out principle.

GOTS (as the name implies) makes use of only the gradient and not the Hessian. Instead of the Hessian, all functional values (energies) that correspond to a given displacement in each degree of freedom are computed. Stepanenko and Engels found GOTS to be more efficient than GTS. As an example, they performed a search with both methods on the 50 dimensional Ackley Test Function. The GOTS method performed 3808.75 function evaluations, 225.42 gradient evaluations, 15080 sums where the GTS method performed 1335.84 function evaluations, 228.74 gradient evaluations, 114.19 Hessian evaluations, and 18482 sums (these are averages over 100 simulations).

1.2 – Sampling Phase Space

Instead of attempting to discover the low lying minima and their energies, a separate field of study, sampling, focuses on efficiently gathering an accurate ensemble of low-energy phase space. Phase space is the collection of every possible value for position and momentum. One point in phase space is defined as the position of momentum of all particles at a given time. For an *n* particle system in Cartesian space, phase space is described in Eqn. (1.**8**) as:

$$\boldsymbol{X} = (x_1, y_1, z_1, p_{x,1}, p_{y,1}, p_{z,1}, x_2, y_2, z_2, p_{x,2}, p_{y,2}, p_{z,2}, \dots, x_n, y_n, z_n, p_{x,n}, p_{y,n}, p_{z,n}) \qquad (1.8)$$

where the coordinates *x*, *y*, and *z* represent the position for each atom, and p_x, p_y, and p_z represent the momentum in each direction. To simplify the notation, the position and momentum can be separated into their respective variables:

$$\boldsymbol{q} = (x_1, y_1, z_1, x_2, y_2, z_2, \dots, x_n, y_n, z_n,)$$

$$\boldsymbol{p} = (p_{x,1}, p_{y,1}, p_{z,1}, p_{x,2}, p_{y,2}, p_{z,2}, \dots, p_{x,n}, p_{y,n}, p_{z,n})$$

$$\boldsymbol{X} = (\boldsymbol{q}, \boldsymbol{p}) \qquad (1.9)$$

For any measurable property, *A*, it is possible to obtain the average value over all space, $\langle A \rangle$, by integrating over phase space (Eqn. (1.**10**)):

$$\langle A \rangle = \iint A(\boldsymbol{q},\boldsymbol{p})P(\boldsymbol{q},\boldsymbol{p})d\boldsymbol{q}d\boldsymbol{p} \qquad (1.10)$$

where *P* is the probability of the physical system visiting specific points in phase space. *P* is defined by statistical mechanics as:

$$P(\boldsymbol{q},\boldsymbol{p}) = \frac{e^{-E(\boldsymbol{q},\boldsymbol{p})/k_B T}}{Q} \qquad (1.11)$$

$$Q = \iint e^{-E(\boldsymbol{q},\boldsymbol{p})/k_B T}\, d\boldsymbol{q}d\boldsymbol{p} \qquad (1.12)$$

Q is the canonical partition function, k_B is the Boltzmann constant, and *T* is temperature. While analytic evaluation of the integrals in Eqn. (1.**10**) and (1.**12**) is possible for simple systems, the evaluating the integral for highly complex systems becomes intractable. For systems of sufficient complexity, the goal instead turns to constructing a representative sample of low-energy phase space. This is still a daunting process for large systems however. A common illustration of this problem is the evaluation of a single point within each 'hyper-octant' of a system (evaluate every possible combination of signs for $\boldsymbol{q}$ and $\boldsymbol{p}$). Because each atom contains 6 degrees of freedom and each degree of freedom can adopt both negative and positive values the number of points in phase space to evaluate is equal to 2^{6N} where N is the number of atoms within a system. For a reasonably-sized system of 50 atoms the number of configuration comes to approximately $2\text{x}10^{90}$ points to evaluate each *A* and *E*, quite a few for a modestly sized system!

However, it should be noted that the vast majority of phase space is a gigantic high-energy mesa of incomprehensible configurations. These high-energy points in phase space have little to no contribution to the measureable property, *A*, relative to lower energy points by virtue

of Eqn. (1.**11**). Therefore sampling low-energy space is key to transform the problem into something tractable.

There are a number of methods that perform sampling, most of which can be characterized within the realms of Monte Carlo or MD methods. Only a small portion of relevant MD methods will be discussed here, as the number that span this field are too great to cover in any great detail.[43,44]

1.2.A. – Metadynamics. Laio and Parrinello in 2003[31] created a method known as metadynamics[33,45] that adds "sand" or "small pebbles" to a reduced dimension space in order to gradually fill in minima and facilitate exploration of phase space and rare events over the course of a MD simulation. The sand or small pebbles consist of Gaussian functions centered at discrete positions in the reduced dimension space (referred to as the collective variable (CV) space). As with coarse-grain models[46,47] a limited number of CVs are chosen. The CVs range from a simple atom-atom distance or angle to the functionals governing MD itself. By modifying visited CV space with Gaussians, the MD simulation is biased against returning to areas of space previously explored. Over the duration of the MD simulation, potential wells are gradually filled, the local surface explored, and a free energy description of the local surface is constructed.

In their 2003 publication, Laio and Parrinello[31] obtained the free energy surface for the dissociation of NaCl in water. In this study, the CVs were defined as the interatomic distance between the sodium and chlorine atoms, as well as the electric fields of each atom. The Gaussians were set to a height of 0.3 kcal/mol and the simulation was performed with the AMBER95 force field and TIP3P water molecules with periodic boundary conditions. The system was simulated with 0.7 fs time steps with six replicas where the replicas were equilibrated for 200 steps and then allowed to run for 500 steps, after which the forces were

averaged. From the free energy surface, two minima and a transition state were observed. The two minima corresponded to 1) the ion pair (interatomic distance of 2.6 Å) and 2) the ions dissociated into water (>5 Å separation). The transition state was characterized as a distance of 3.3 Å between the two ions.

By gradually filling in the PES, a complete description of local phase space is gathered rather trivially. Metadynamics has gone on to become a staple method in the sampling community ranging in use from something as simple as salt dissociation to something as complex as blind protein-ligand docking. However, a common issue with metadynamics is the difficult choice of defining the appropriate CVs to effectively sample space. For systems with increasing complexity, the high dimensionality of the surface makes that decision much harder. In fact, multiple papers expand upon the strategy of properly choosing CVs. In 2011, Barducci, Bonomi, and Parrinello[33] described multiple protocols for defining proper CVs but the general consensus was that the choice is highly system dependent.

B. Umbrella Sampling. Another means of elucidating information from the PES is the umbrella sampling method[48–51] developed by Torrie and Valleau[48]. The primary goal of umbrella sampling is to obtain the free energy associated with a given reaction coordinate. Utilizing the knowledge of a reaction coordinate a harmonic potential a.k.a. umbrella potential is applied to the surface to bias conformational sampling along the coordinate where MD is then performed at each position. From the resulting data, an ensemble average of this local space is generated. Over the course of multiple simulations with differing umbrella potentials along a reaction coordinate, a free energy surface of the reaction coordinate can be constructed.

In 2002, Mu and Stock[49] utilized umbrella sampling to garner a greater understanding of the conformational probability of trialanine in water. They utilized the GROMOS96 force field

with a time step of 2 fs, running dynamics for 5 ps. To perform umbrella sampling, a harmonic bias term was added to the Hamiltonian: $V_R = k(\phi - \phi_0)^2$. The force constant, k, was chosen to be 1 kJ/(mol deg^2). This restrains the sampling of the dihedral angles, ϕ, within a window around a chosen dihedral angle, ϕ_0. A step size of 8 degrees was chosen to sample along ϕ_0. From the resulting trajectories a biased distribution of phase space was obtained from which a biased probability distribution was calculated. From this distribution, umbrella integration was applied to obtain an unbiased potential mean force with a probability distribution of trialanine existing in two extended conformations 80% of the time and the remaining 20% is spent in a helical conformation.

Overall, umbrella sampling is a very useful tool with only one parameter that must be chosen, the force constant. With too large of a force constant, the harmonic potential will dominate and a narrow distribution of structures will be sampled, however too soft of a potential will result in redundancy in sampling due to too much overlap between windows. Additionally, it is generally more accurate to sample a finer grid of windows on a shorter time scale for each than it is fewer windows for longer.

1.3 – Conclusions

The research presented in this dissertation will be drawing from and building upon various searching and sampling methods discussed. Specifically, chapter 2 will discuss the development and benchmarking of SEMD that utilizes facets from each method to optimize molecular dynamics conformational searching. Additionally, chapter 4 will directly utilize select conformational searching methods to aid in eliciting the reaction coordinate of the photocatalyzed [2+2] cycloaddition of a phenyl substituted bis(enone).

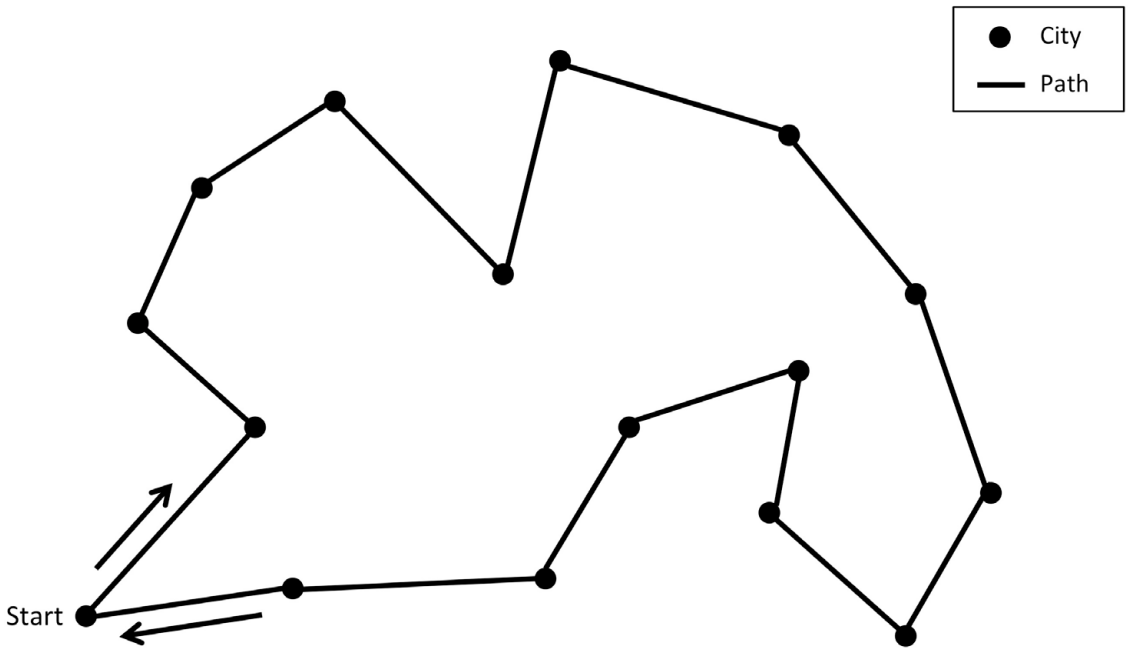


Figure 1.1. An example case of the traveling salesman problem. Starting from the "start" city (bottom left), the shortest path that visits every city and returns is obtained. The above path is not necessarily the shortest, however it is an illustration of one possible path that may be taken.

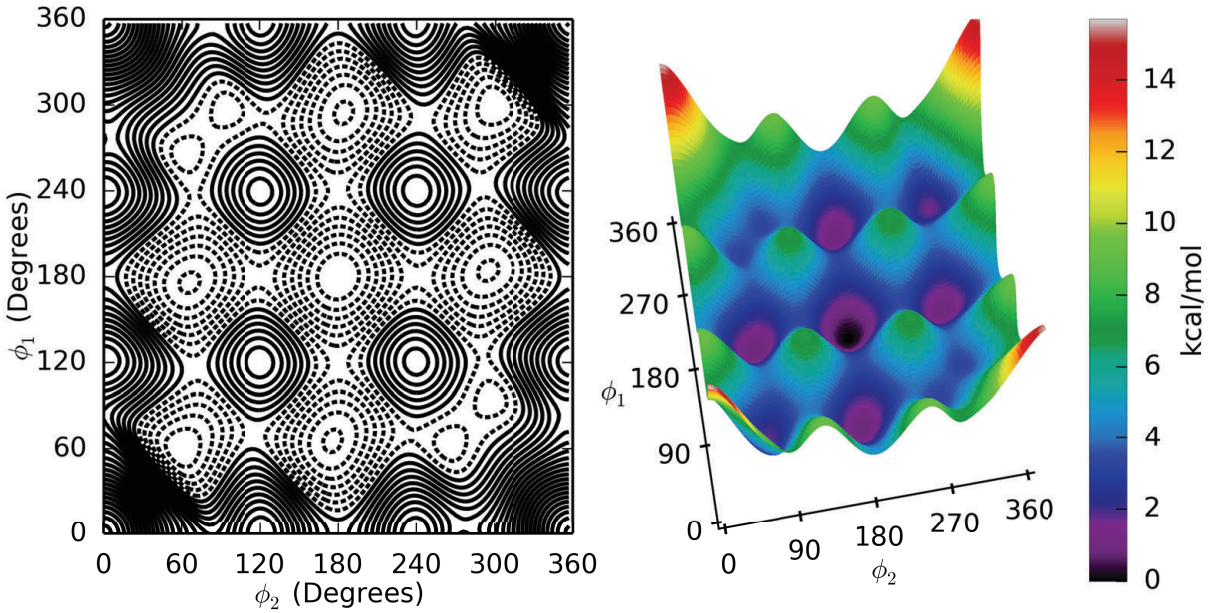


Figure 1.2. A contour plot (left) and three dimensional plot (right) of the potential energy surface of pentane as a function of dihedral angle of the two C-C-C-C torsions present within the molecule. Within the context of the traveling salesman problem, the minima are equivalent to the cities.

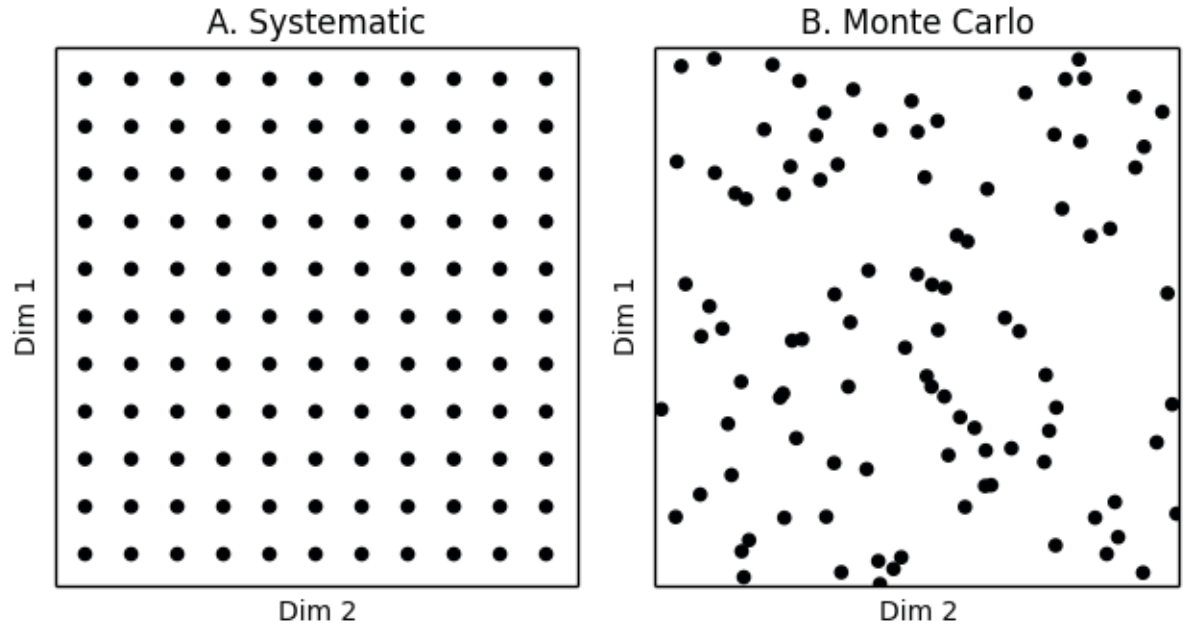


Figure 1.3. A depiction of a systematic grid search (A) and a Monte Carlo (random) search (B). The depicted systematic search scans over fixed intervals in two dimensions until all possible combinations of intervals have been exhausted. The Monte Carlo search depicted randomly generates points in two dimensional space.

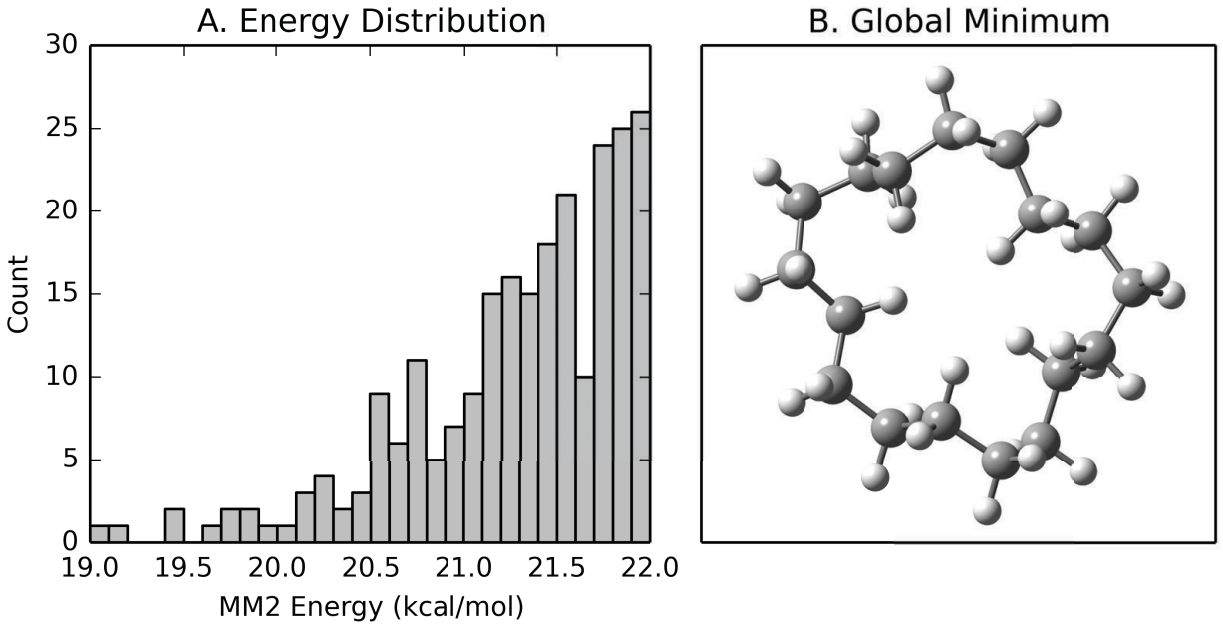


Figure 1.4. The energy distribution of lowest 3 kcal/mol conformers (A) and the global minimum (B) of cycloheptadecane. The energy distribution has a bin width of 0.1 kcal/mol.

CHAPTER 1 REFERENCES

(1) Leeuwen, J. van. *Algorithms and complexity*; Elsevier [u.a.]: Amsterdam [u.a., 1998.

(2) Applegate, D. L.; Bixby, R. E.; Chvátal, V.; Cook, W. J. *The Traveling Salesman Problem: A Computational Study*; Princeton University Press, 2011.

(3) Lipton, M.; Still, W. *J. Comput. Chem.* **1988**, *9*, 343–355.

(4) Makino, S.; Kuntz, I. D. *J. Comput. Chem.* **1997**, *18*, 1812–1825.

(5) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nat. Rev. Drug Discov.* **2004**, *3*, 935–949.

(6) Lexa, K. W.; Carlson, H. A. *Q. Rev. Biophys.* **2012**, *45*, 301–343.

(7) Giovannetti, J.; Kelly, C.; Landis, C. *J. Am. Chem. Soc.* **1993**, *115*, 4040–4057.

(8) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.* **2004**, *47*, 1739–1749.

(9) Howard, A. E.; Kollman, P. A. *J. Med. Chem.* **1988**, *31*, 1669–1675.

(10) Sherer, E. C.; Lee, C. H.; Shpungin, J.; Cuff, J. F.; Da, C.; Ball, R.; Bach, R.; Crespo, A.; Gong, X.; Welch, C. J. *J. Med. Chem.* **2014**, *57*, 477–494.

(11) Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J. *J. Comput. Aided Mol. Des.* **1996**, *10*, 293–304.

(12) Saunders, M.; Houk, K. N.; Wu, Y. D.; Still, W. C.; Lipton, M.; Chang, G.; Guida, W. C. *J. Am. Chem. Soc.* **1990**, *112*, 1419–1427.

(13) Chang, G.; Guida, W. C.; Still, W. C. *J. Am. Chem. Soc.* **1989**, *111*, 4379–4386.

(14) Metropolis, N.; Rosenbluth, A.; Rosenbluth, M.; Teller, A.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087–1092.

(15) Saunders, D.; Winnik, M. *Macromolecules* **1978**, *11*, 18–24.

(16) Sali, A.; Shakhnovich, E.; Karplus, M. *Nature* **1994**, *369*, 248–251.

(17) Verdier, P.; Stockmayer, W. *J. Chem. Phys.* **1962**, *36*, 227 – &.

(18) Stillinger, F.; Weber, T. *Phys. Rev. A* **1982**, *25*, 978–989.

(19) Legrand, B.; Treglia, G.; Desjonqueres, M.; Spanjaard, D. *J. Phys. C-Solid State Phys.* **1986**, *19*, 4463–4472.

(20) Biswas, R.; Hamann, D. *Phys. Rev. B* **1986**, *34*, 895–901.

(21) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.

(22) Ryckaert, J.; Bellemans, A. *Faraday Discuss.* **1978**, *66*, 95–106.

(23) Verlet, L. *Phys. Rev.* **1967**, *159*, 98–103.

(24) Beeman, D. *J. Comput. Phys.* **1976**, *20*, 130–139.

(25) Judson, R.; Jaeger, E.; Treasurywala, A.; Peterson, M. *J. Comput. Chem.* **1993**, *14*, 1407–1414.

(26) Mcgarrah, D.; Judson, R. *J. Comput. Chem.* **1993**, *14*, 1385–1395.

(27) Nair, N.; Goodman, J. M. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 317–320.

(28) Hartke, B. *J. Phys. Chem.* **1993**, *97*, 9973–9976.

(29) Deaven, D. M.; Tit, N.; Morris, J. R.; Ho, K. M. *Chem. Phys. Lett.* **1996**, *256*, 195–200.

(30) Long, S. M.; Tran, T. T.; Adams, P.; Darwen, P.; Smythe, M. L. *J. Comput. Chem.* **2011**, *32*, 1541–1549.

(31) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562–12566.

(32) Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. *J. Phys. Chem. B* **2005**, *109*, 6714–6721.

(33) Barducci, A.; Bonomi, M.; Parrinello, M. *Wiley Interdiscip. Rev.-Comput. Mol. Sci.* **2011**, *1*, 826–843.

(34) Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D. *Proteins-Struct. Funct. Genet.* **1998**, *33*, 367–382.

(35) Grebner, C.; Becker, J.; Stepanenko, S.; Engels, B. *J. Comput. Chem.* **2011**, *32*, 2245–2253.

(36) Morales, L. B.; Garduño–Juárez, R.; Aguilar–Alvarado, J. M.; Riveros–Castro, F. J. *J. Comput. Chem.* **2000**, *21*, 147–156.

(37) Stepanenko, S.; Engels, B. *J. Comput. Chem.* **2007**, *28*, 601–611.

(38) Stepanenko, S.; Engels, B. *J. Comput. Chem.* **2008**, *29*, 768–780.

(39) Cramer, C. J. *Essentials of computational chemistry: theories and models*; Wiley: Chichester, West Sussex, England; Hoboken, NJ, 2004.

(40) Piela, L.; Kostrowicki, J.; Scheraga, H. *J. Phys. Chem.* **1989**, *93*, 3339–3346.

(41) Hart, R. K.; Pappu, R. V.; Ponder, J. W. *Biophys. J.* **1998**, *74*, A176–A176.

(42) Kolossváry, I.; Guida, W. C. *J. Am. Chem. Soc.* **1996**, *118*, 5011–5019.

(43) Wood, W. W.; Erpenbeck, J. J. *Annu. Rev. Phys. Chem.* **1976**, *27*, 319–348.

(44) Christen, M.; van Gunsteren, W. F. *J. Comput. Chem.* **2008**, *29*, 157–166.

(45) Laio, A.; Gervasio, F. L. *Rep. Prog. Phys.* **2008**, *71*, 126601.

(46) Gear, C. W.; Kevrekidis, I. G.; Theodoropoulos, C. *Comput. Chem. Eng.* **2002**, *26*, 941–963.

(47) E, W.; Engquist, B. *arXiv:physics/0205048* **2002**.

(48) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187–199.

(49) Mu, Y.; Stock, G. *J. Phys. Chem. B* **2002**, *106*, 5294–5301.

(50) Hansen, H. S.; Huenenberger, P. H. *J. Comput. Chem.* **2010**, *31*, 1–23.

(51) Kästner, J. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 932–942.

CHAPTER 2: SURFACE EDITING MOLECULAR DYNAMICS, AN ALTERNATIVE CONFORMATIONAL SEARCHING TOOL

2.1 – Introduction

A model is described that permits selective editing or removal of potential surface minima or maxima. This surface editing accelerates various conformational searching procedures based on molecular dynamics, Monte Carlo, or genetic algorithms facilitating the determination of families of low energy structures and assemblies. This development should aid small molecule conformational search and free energy estimation, and likely impact protein-ligand binding studies, catalyst transition state modeling, and crystal polymorph prediction. The surface editing function model supports a wide variety of internal coordinate systems including but not limited to Cartesian, torsional, quaternion, or lattice representations. Application to molecular dynamics-based torsional search of cyclic and long chain hydrocarbons is illustrated here where it is found that global minima and the ensemble of low energy conformations can be assembled with fewer minimizations than previous approaches.

The relevant literature was reviewed in chapter 1, the method is described in section 2.2, results and discussion collected in section 2.3, and conclusions provided in section 2.4.

2.2 – Method

Surface editing molecular dynamics (SEMD) rests on several premises. First, it is assumed that individual minima can be efficiently and selectively removed from a potential energy expression. Secondly, out of the 3n-6 vibrational velocity directions, it is hypothesized that a select few directions are most productive for producing new conformations and that those

select few directions can be efficiently determined. Finally, it is assumed that it is possible to estimate when a new conformational potential well has been entered. If these premises are well-founded, then the global minimum as well as a representative family of low energy conformations can be efficiently assembled.

For MD, the question of when a system is in a different well/conformation can be particularly challenging due to a tendency of MD to become trapped within a well and harmonically oscillate within previously explored configurational space. With SEMD, applying Gaussian-shaped and potential well-sized modifications to the potential energy surface at minima discourages the return to and being trapped in discovered conformational space. In addition to these advantages, starting dynamics at the peak of a Gaussian instead of in the valley of a well allows for an initial period of highly directed dynamics away from a known conformation. The following describes the overall algorithm, wherein the specifics of each segment is referred to and discussed in their respective section.

To begin, a starting geometry is minimized and a Gaussian potential (section 2.2.A) is applied to the PES centered at the minimized geometry. The Gaussian exponent is taken as a linear combination of singularity and discontinuity-free torsional coordinates (section 2.2.B). A reduced family of starting velocity directions is generated (sections 2.2.C and 2.2.D) and starting from the seed (the minimized geometry), molecular dynamics is carried out in each direction for a period of time. After this period of time, a minimization on the original unmodified surface occurs (section 2.2.E). If after the minimization a non-duplicate geometry (section 2.2.F) is discovered, a Gaussian is added for this combination of torsional coordinates and the dynamics terminated in that direction. Similar to the usage-directed protocol used by Chang et al.[1] the

lowest energy conformer that has not yet served as a starting structure becomes the new seed and the process repeated once all velocity directions have been exhausted for a given seed.

2.2.A – Surface Editing Tool: Due to the specific degrees of freedom that are of interest, metadynamics limits itself to a small number of carefully chosen collective variables (CVs) to generate the free-energy surface. For the purposes of the present illustration of SEMD, the CVs are defined as all of the non-degenerate carbon-based torsions in the compound. This choice of CVs gives a complete description of the torsional potential energy surface for the isolated systems discussed in this paper, thus enabling the bias potential to exert control over the entire system.

As with metadynamics, a Gaussian function is chosen to edit the potential energy surface due to its separable sum of squares representation of "distance" space. This separability imparts the function with the property that when a single "distance" is far from the minimum the entire function vanishes. A Gaussian is described with only two parameters and dies off smoothly and quickly. The specific functional form of the Gaussian is given in Eqn. (2.1).

$$V_{SEMD} = A \prod_{n=1}^{tors} e^{-\alpha\left(f(\phi_i)-f(\phi_{i,0})\right)^2} \qquad (2.1)$$

In Eqn. (2.1), A is the scaling factor (in kcal/mol), α, the Gaussian exponent, which controls the width or distance range of the function, $f(\phi_i)$ is a function of the current torsion angle, ϕ_i, and $f(\phi_{i,0})$ is a function of a torsion angle, $\phi_{i,0}$, from a previously obtained conformation. The energy correction, Eqn. (2.1), is a product of Gaussians (or sum of Gaussian exponents) to ensure that as any single torsion twists away from a previous value the entire function approaches zero. The parameters A and α are chosen to create a single Gaussian that will sufficiently edit out minima from the potential surface. Some modicum of care must be taken

when assigning values for A and α, as a Gaussian that is too large or broad of can unintentionally deform or edit nearby minima from the surface, whereas a Gaussian that is too small or narrow will not sufficiently bias against previously discovered minima.

The ethane rotation potential surface was used to provide initial estimates for Gaussian parameters fills the minimum but not span too wide a range (A=6 and α=6) (Figure 2.1). Upon application, it was discovered that a taller Gaussian (A=12) facilitates more guided dynamics without editing out nearby minima as evidenced by the pentane potential energy surface in Figure 2.2. Values of at least 12 for A and at most 8 for α are used for the present study. It should be noted that over the course of multiple searching simulations, varying the height and width of the Gaussian did not significantly impact the efficacy of the searching protocol as long as the choices were reasonable.

2.2.B – Singularity and discontinuity-free coordinate representation: In contrast to bond angles that are uniquely defined over the range of $0^\circ \leq \theta < 180^\circ$, torsional angles are unique over the entire $0^\circ \leq \phi < 360^\circ$ range. To avoid discontinuities at 360˚ and 0˚ or 180˚ and -180˚ torsion angles and inconsistent differences between torsion angles, torsion angles are described as distances between points on a unit circle. If the first point is at x_1, y_1, ($\cos\phi_1$, $\sin\phi_1$) and the second point is at x_2, y_2 ($\cos\phi_2$, $\sin\phi_2$) the square of the distance between them is $(x_1-x_2)^2 + (y_1-y_2)^2$ or $(\cos\phi_1 - \cos\phi_2)^2 + (\sin\phi_1 - \sin\phi_2)^2$. Two equivalent representations for the Gaussian surface editing function are collected in Eqn. (2.2).

$$V_{SEMD} = A \sum_{j=1}^{confs} \prod_{i=1}^{tors} e^{-\alpha\left(\left(sin\phi_i - sin\phi_{i,0}\right)^2 + \left(cos\phi_i - cos\phi_{i,0}\right)^2\right)}$$

$$V_{SEMD} = A \sum_{j=1}^{confs} \prod_{i=1}^{tors} e^{-2\alpha\left(1 - \left(sin\phi_i sin\phi_{i,j} + cos\phi_i cos\phi_{i,j}\right)\right)} \qquad (2.2)$$

The first derivative of the sine terms in Eqn. (2.2) with respect to Cartesian coordinate contains a singularity as ϕ approaches 0 or 180˚. In 1991, Swope and Ferguson[2] reported a singularity-free representation, provided in Eqn. (2.3).

$$\nabla sin\phi_{ijkl} = -\hat{r}_{kj} \cdot (\hat{u} \times \hat{t}) \qquad (2.3)$$

where: $\hat{u} = \frac{r_{lk} \times r_{kj}}{|r_{lk}||r_{kj}|}$ and $\hat{t} = \frac{r_{ij} \times r_{kj}}{|r_{ij}||r_{kj}|}$

Eqn. (2.3) provides both an efficient means of computing $sin\phi_{ijkl}$ and a singularity free $\nabla sin\phi_{ijkl}$ representation for a torsion with atoms *i, j, k, l* where atoms *j* and *k* are the two central atoms, r_{xy} represents the vector from atom x to atom y, and the hat denotes that it is a unit vector.

2.2.C – Generation of Initial MD Velocities: Typically in MD an initial velocity vector is generated as a random combination of Cartesian displacement coordinates, the linear momentum is projected out, the resulting velocity vector scaled to a prescribed temperature, and then used as displacements to initiate Newtonian dynamics. For an isolated molecule the selected initial velocity direction is but one of 3n-6 possible vibrational directions. Further, of these 3n-6 directions a number of them may well lead toward previous minima. The final challenge is that local modes in general are not linearly independent. For example, the three local bends about an sp^2 center only represent two actual bends.

To mitigate these issues a linearly independent set of collective variable displacement coordinates (torsions) is generated, biorthogonalization is used to remove significant overlap with previous directions, and the highest eigenvalue normal mode directions used as initial velocities. The procedure proceeds as follows.

i. A singular valued decomposition[3] (SVD) is carried out in order to remove the linear dependence and determine the number of independent degrees of freedom. Given local modes,

q, and the matrix **B** that transforms the Cartesian representation, **x**, to local modes, **q**, Eqn. (2.**4**),

$$\mathbf{q}=\mathbf{B}\mathbf{x} \qquad (2.4)$$

the overlap between the local modes in Cartesian space in matrix **S**, is given in Eqn. (2.**5**).

$$\mathbf{S}=\mathbf{B}^{\mathrm{T}}\mathbf{B} \qquad (2.5)$$

A SVD analysis of **S** yields the diagonal overlap eigenvalue matrix **s** and the associated unitary transformation provided in Eqn. (2.**6**).

$$\mathbf{s} = \mathbf{U}^{\mathbf{T}}\mathbf{S}\mathbf{U} \qquad (2.6)$$

This unitary matrix, **U**, is used to prepare the canonical orthogonalization matrix **X**, in Cartesian space.

$$\mathbf{X} = \frac{1}{\sqrt{\mathbf{s}}}\mathbf{U} \qquad (2.7)$$

Vectors of **U** with zero or sufficiently small eigenvalue (<0.2) are simply left out of subsequent analysis.

ii. Previous velocity space directions, described as direction vectors from previous structures in torsional space, **D**, are transformed to the linearly independent sub-space, Eqn. (2.**8**),

$$\mathbf{D}' = \mathbf{X}^{\mathbf{T}}\mathbf{D} \qquad (2.8)$$

a density matrix is formed, Eqn. (2.**9**),

$$\mathbf{R} = \mathbf{D}'\mathbf{D}'^{\mathbf{T}} \qquad (2.9)$$

and diagonalized, Eqn. (2.**10**).

$$\mathbf{r} = \mathbf{V^T R V} \qquad (2.10)$$

The resulting non-zero eigenvalues **r** and back-transformed eigenvectors $\mathbf{V'}$, Eqn. (2.**11**),

$$\mathbf{V'} = \mathbf{XV} \qquad (2.11)$$

are used to form a symmetric matrix in the Cartesian space, Eqn. (2.**12**).

$$\mathbf{Y} = \mathbf{V'} \frac{1}{\sqrt{\mathbf{r}}} \mathbf{V'^T} \qquad (2.12)$$

Matrix **Y**, which contains the non-zero overlaps with previous directions is transformed to the linearly independent sub-space, Eqn. (2.**13**),

$$\mathbf{Y'} = \mathbf{X^T Y X} \qquad (2.13)$$

and diagonalized to find the directions that maximally and minimally overlap previous direction space, Eqn. (2.**14**). Direction vectors with an eigenvalue y' less than a threshold (0.1) are retained.

$$\mathbf{y'} = \mathbf{W^T Y' W} \qquad (2.14)$$

The resulting eigenvector matrix is backtransformed to the linearly independent Cartesian sub-space, Eqn. (2.**15**).

$$\mathbf{W'} = \mathbf{XW} \qquad (2.15)$$

iii. Finally, the Cartesian Hessian is transformed to this final sub-space, Eqn. (2.**16**),

$$\mathbf{H''} = \mathbf{W'^T H W'} \qquad (2.16)$$

and diagonalized to obtain the set of normal modes that are orthogonal to previous direction space, Eqn. (2.**17**).

$$\mathbf{H''C''} = \mathbf{C''}\varepsilon \qquad (2.17)$$

The resulting mode matrix is backtransformed to the Cartesian space, Eqn. (2.**18**),

$$\mathbf{C}^{\mathrm{v}} = \mathbf{W'C''} \qquad (2.18)$$

and the highest 10 eigenvalue normal modes used as the set of velocity vector directions in SEMD.

2.2.D – Reduction in MD Velocity degrees of freedom: Once the initial family of velocity directions is created it is useful to restrict dynamical motion to degrees of freedom that contribute to conformational exploration. In 1977 Ryckaert developed SHAKE[4], a constrained molecular dynamics algorithm wherein specific internal degrees of freedom, typically stretches and bends are constrained during dynamics, focusing kinetic energy into desired degrees of freedom.

The conventional Verlet equation is modified through the addition of a restraining force $\mathbf{g_R}(t)$, selected to prevent the motion of select degrees of freedom, Eqn. (2.**19**)

$$\boldsymbol{r}(t+\Delta t) = 2\boldsymbol{r}(t) - \boldsymbol{r}(t-\Delta t) + \frac{1}{2}\big(\boldsymbol{f}(t) + \boldsymbol{g}_R(t)\big)\Delta t^2/m \qquad (2.19)$$

In 1983 Anderson generalized SHAKE to permit internal velocity constraints as well as geometric constraints, resulting in RATTLE[5], Eqns. (2.**20**) and (2.**21**). In addition to the position constraint, $\mathbf{g}_R(t)$, a velocity constraint term, $\mathbf{g}_V(t)$ is added.

$$\boldsymbol{r}(t+\Delta t) = 2\boldsymbol{r}(t) - \dot{\boldsymbol{r}}(t)\Delta t + \frac{1}{2}\big(\boldsymbol{f}(t) + \boldsymbol{g}_R(t)\big)\Delta t^2/m \qquad (2.20)$$

$$\dot{\boldsymbol{r}}(t+\Delta t) = \dot{\boldsymbol{r}}(t) + \frac{1}{2}[\big(\boldsymbol{f}(t) + \boldsymbol{g}_R(t)\big) + \big(\boldsymbol{f}(t+\Delta t) + \boldsymbol{g}_V(t+\Delta t)\big)]\Delta t/m \qquad (2.21)$$

While constraint methods such as SHAKE or RATTLE do facilitate molecular dynamics there are numerical stability issues that limit time step size. An additional concern is that

constraining distances and angles enforce localized structural distortions found in the initial, minimized starting guess and prohibit small structural distortions that occur as torsional barriers are traversed. Here we suggest an alternative, CHILL, that, via velocity projection, Eqn. (2.**22**), dampens the motion of selected internal degrees of freedom permitting focused dynamical motion and extended time step size while not constraining internal degrees of freedom.

$$\dot{\boldsymbol{r}}^{\boldsymbol{P}} = \boldsymbol{P}\dot{\boldsymbol{r}} \qquad (2.22)$$

As previously described by Pulay[6] and Truhlar[7,8], projection for an internal degree of freedom such as stretch or bend can be constructed from its Cartesian displacement representation, **e**, Eqn. (2.**23**).

$$\boldsymbol{P} = 1 - \boldsymbol{e}\boldsymbol{e}^{\boldsymbol{T}} \qquad (2.23)$$

Projections can be sequentially applied, yielding in this case a velocity vector lacking particular degrees of freedom. A simplified description of Eqn. (2.**23**) is provided in Eqn. (2.**24**)

$$\dot{\boldsymbol{r}}^{\boldsymbol{P}}(t) = \dot{\boldsymbol{r}}(t) - \sum_{\boldsymbol{i}} \boldsymbol{e}_i \gamma_i \qquad (2.24)$$

where γ_i is given by Eqn. (2.**25**).

$$\gamma_i = \boldsymbol{e}_i \cdot \dot{\boldsymbol{r}}(t) \qquad (2.25)$$

The resulting projected velocity Verlet equations are given in Eqns. (2.**26**) and (2.**27**).

$$\boldsymbol{r}(t + \Delta t) = \boldsymbol{r}(t) - \dot{\boldsymbol{r}}^{P}(t)\Delta t + \frac{1}{2}\boldsymbol{f}(t)\Delta t^2/m \qquad (2.26)$$

$$\dot{\boldsymbol{r}}(t + \Delta t) = \dot{\boldsymbol{r}}^{P}(t) + \frac{1}{2}[\boldsymbol{f}(t) + \boldsymbol{f}(t + \Delta t)]\Delta t/m \qquad (2.27)$$

In Eqn. (2.**26**) the projected velocity term $\dot{\mathbf{r}}^{P}(t)$ does not contribute structural distortion for the selected internal degrees of freedom but the force-based acceleration term does provide motion towards a minimum.

Cartesian displacements $\mathbf{e}_i$ are not intrinsically orthogonal, and bends often contain intrinsic linear dependence, application of subsequent projections will impact previous projections. This dependency can be dealt with by iterative application of the projections or by using orthogonal displacements. A set of orthogonal displacements can be obtained by applying a SVD to the Cartesian displacement overlap matrix, ignoring vectors with small eigenvalue. Given the local nature of the overlaps sparse methodologies would accelerate the process. Alternatively, the subset of local modes centered on each non-hydrogen atom can be generated, orthogonalized, and projected sequentially. For sp^3 C there are 4 stretches and 6 bends (5 combinations of which are linearly independent), SVD over the 10 dimensional space results in 9 orthogonal projections. Both the iterative and atom-based approaches yield acceptable results.

For octane, the 72 internal degrees of freedom are reduced to 7. A 29 K molecular dynamics simulation is equivalent to a 300 K conventional simulation. A 293 K, 2 ps molecular dynamics simulation with 2 fs step size on octane yield average C-H and C-C displacements of 0.0004 and 0.0005 Å. C-C-C, H-C-H, and C-C-H angular deformations average 0.04˚, 0.19˚ and 0.05˚ respectively. If hydrogen masses are scaled to 12, then 8 fs time steps are feasible with this approach.

2.2.E – Minimization Criteria: A multifaceted approach was developed to ensure that minimization proceeds to a structure outside of the current seed's well. The first criterion utilizes the distance from the seed structure and the second is based on the local dynamics elapsed time.

To prevent artificial and potentially system-specific minimization criteria from being chosen by the user, a learning algorithm was developed to formulate a distance cutoff criterion in terms of previous successful minimizations during a given simulation. For the first twenty minimizations, a dynamics time-based criterion is used to train the distance criterion. For each minimization that yields a unique conformer, the distance from the seed structure is calculated, see Eqn. (2.**28**),

$$d_0 = \sum_{i=1}^{tors} \sqrt{1 - \left(sin\phi_i sin\phi_{i,0} + cos\phi_i cos\phi_{i,0}\right)} \qquad (2.28)$$

and included in a running average and standard deviation. Following twenty successful minimizations, the distance criterion becomes, Eqn. (2.**29**).

$$d \geq \bar{d}_{min} - 2s_d \qquad (2.29)$$

Reduction of the average of the distances leading to a successful minimization, $\bar{d}_{min}$, by two standard deviations of those distances, s_d, was used to relax the distance criterion to prevent close-by minima from being excluded.

In tandem with the distance cutoff and while the distance cutoff is being developed, a dynamics time-based criterion is utilized to allow minimizations to occur as long as the distance criterion is met. The time cutoff is defined in Eqn. (2.**30**).

$$t_{current} = \begin{cases} 2 \cdot t_{cutoff} \\ 3 \cdot t_{cutoff} \\ 4 \cdot t_{cutoff} \end{cases} \qquad (2.30)$$

Thus, a maximum of three minimizations can potentially occur for each velocity direction.

Utilization of the distance and time-based criteria in an “AND” relationship increases the probability of finding a new potential energy well.

2.2.F – Duplicate Checking: The cumulative distance, Eqn. (2.**28**) is also used to determine if a "new" minimized conformer has been previously discovered. Comparing against all history, a distance, *d*, above 0.001 defines a unique conformation. This corresponds to a greater than three degree torsional difference.

2.2.G: Techniques to Reduce Computations: A number of methods are used to reduce the number of computations to increase efficiency without sacrificing completeness. Like tabu search methods, a list of previous history is generated. This list must be scanned. Here, the Gaussian function as well as its derivative must be evaluated relative to each previous conformer in the list, at each dynamic step. As the list grows, the time spent evaluating this potential increases. However, not every conformer in the history contributes significantly at the current position on the PES. In fact, the rapid decay of Gaussians leads to many previous conformers not contributing to the editing function. Instead of performing a full evaluation of the surface editing potential and forces at every dynamic step, the Gaussian exponent of a conformer is utilized to determine if its contribution must be calculated. If the distance to a previous conformer, Eqn. (2.**28**), meets the criteria in Eqn. (2.**31**),

$$d < -\frac{ln\left(\frac{0.01}{A}\right)}{2\alpha} \qquad (2.31)$$

the conformer will contribute more than 0.01 kcal/mol to the potential energy and is added to short list of contributing conformers. This reduced list is then used to determine the surface editing potential for the next 5 dynamics steps after which the list will be re-evaluated.

A reduced list of nearby conformers is also generated immediately prior to minimization. During minimization, it is efficient to check this shorter list to determine if the structure is minimizing to a previously discovered conformer. If the structure is a duplicate, d_0 less than 0.001, the minimization terminates and dynamics resumes. If the minimization is successful, the

structure is compared to the full list of previous structures to maintain the integrity of the surface editing potential.

Symmetry can also be used to reduce computation time. For cycloheptadecane, up to 68 geometrically equivalent conformers can be generated from a single structure and a Gaussian is placed at each equivalent position on the surface, reducing accessible space by factors of up to 68. As described by Saunders[9], a geometrically equivalent geometry is created "by permuting the atom-numbering systems around the ring in both directions" in addition to changing the sign of the torsion angles. Within each geometrically equivalent set, a duplicate check is performed to ensure the set is non-redundant due to the presence of symmetry. When applicable, symmetry is also used when checking new conformers against previously discovered geometries.

To maintain reproducibility the non-random number-based Nose-Hoover[10,11] thermostat was utilized in favor of random number based thermostats.

2.3 – Results and Discussion

2.3.A – Cycloheptadecane. Dating to the 1990's cycloheptadecane (Figure 2.3) has served as the prototypically challenging benchmark case for conformational searching. Cycloheptadecane contains a large number of degrees of freedom, symmetry, and a constraining ring. In 1990, Saunders, Houk, Still[9] and coworkers established that there are 262 unique conformations within 3.0 kcal/mol of the ground state structure in vacuo on the MM2 potential energy surface. A multitude of searching protocols were employed, each finding a varying number of the 262 low-energy conformations. The number of conformations found for each method were reported as follows: Cartesian stochastic (222), torsion tree (211), Monte Carlo (260), distance geometry (176), and molecular dynamics (169). By combining the results from

every search they were able to construct the population of 262 conformations that exist within lowest 3 kcal/mol of the global minimum.

A pseudosystematic search via Ngo and Karplus was utilized to characterize the lowest 3.0 kcal/mol conformers of cycloheptadecane in 1997[12]. In the pseudosystematic search an initial random structure is generated and cyclized via a kinematic twist operator and minimized. From this initial conformation the kinematic twist operator is applied to achieve each of the 34 possible twists. Each resulting non-duplicate geometry is then stored as a possible future starting structure. Using this pseudosystematic search 258 of the 262 known geometries were obtained in approximately 10,000 minimizations.

In 1998, Pappu, Hart, and Ponder[13] reported a diffusion equation method to perform potential surface smoothing. To efficiently find the global minimum of cycloheptadecane potential smoothing was applied in tandem with a normal mode local search. The smoothing algorithm was adjusted to reduce the surface to a single minimum over the course of multiple minimizations. This approach discovered the second lowest minimum in 440 minimizations. In addition to this, a normal mode local search was performed in Cartesian space to generate the full low-energy spectrum of cycloheptadecane. A total of 20,469 unique structures were generated within 26 kcal/mol of the global minimum.

In 1999, Kolossváry and Guida reported the conformational search method *c*-LMOD[14], a ravine following method that evaluates the Hessian at every step along the PES. The Hessian and gradient are used to construct a path to approach and then pass through a saddle point along the negative eigenvector and then proceed to the bottom of a new well. In addition to using *c*-LMOD to study cycloheptadecane, the conformational search method *l*-LMOD[15] was presented and applied. This alternative *l*-LMOD method utilizes single large "leap" steps along low mode

directions from the starting structure, the leap distances selected to land past barriers. Once the leap has been taken, the resulting structure is minimized and the process repeated. Their *l*-LMOD method found all 262 conformers within 11,631 leaps and subsequent minimizations while *c*-LMOD generated 89,508 saddle points and 42,130 unique minima before all 262 structures were discovered. In a previous paper they had discovered 262 conformations within the lowest 12.6 kJ/mol (~3.01 kcal/mol) with *l*-LMOD. Using this 12.6 kJ/mol energy cutoff, they discovered 264 conformations with *l*-LMOD where the two new structures had energies 3.001557 and 3.004018 kcal/mol above the global minimum.

In 2010 Long, Tran, Adams, Darwen and Smythe generated a population-based incremental learning algorithm[16] (PBIL) to perform conformational searches. The PBIL algorithm forms each generation of conformers from a probability distribution of the fittest dihedral angles from previous generations. With each new generation, each structure is minimized. After 500 minimization steps, if the energy is above a pre-defined cutoff the minimization is stopped, the conformer discarded, and the minimization attempt not counted towards their total full-energy minimization count. With their algorithm they obtained the set of 264 conformations in 7662 full-energy minimizations. The number of partial minimization attempts was not reported.

Starting from the completely flat structure of cycloheptadecane, SEMD obtained the 262 conformers in the lowest 3.0 kcal/mol in 3622 full-energy minimizations and 620,000 dynamic steps (Δt=8.0 fs , T=100 K, t_{cutoff}=240 fs, A=22, and α=8). A total of 3622 conformers were obtained over the course of the dynamics and the highly symmetric structure enabled the SEMD algorithm to edit out a total of 243,423 minima from the surface. During this process there were 9,543 visits to previously discovered conformers as determined by partial minimizations.

Additionally, the global minimum was obtained in 84 full-energy minimizations and 7,200 dynamics steps. Because full energy minimizations, partial minimizations, and molecular dynamics all perform gradient evaluations, we tracked that metric to provide a singular measure of efficiency. As such, a total of 3,157,250 gradient evaluations were performed.

Differential round off issues in different software packages as well a limited precision in the conversion between kcal/mol and kJ/mol preclude a rigorous definition of 3 kcal/mol. The MM2 force field in version 5.2 of the Tinker software suite produces only 261 of the 262 structures within 3 kcal/mol of the global minimum,[13] the 262nd conformer being 3.0016 kcal/mol above the global minimum. A structure data file containing the geometries of all 262 conformations, obtained from Kollosváry (generated via MacroModel), was used to perform a structural similarity analysis when a new conformer was discovered. The calculation was terminated once all 262 geometries were accounted for.

While the goal of finding all 262 lowest 3 kcal/mol energy conformers in the fewest computations possible is an easily definable goal, it is not generalizable. The energy distribution of the various conformations of cycloheptadecane does not have a gap at 3 kcal/mol above the global minimum[13]. Another means of determining a method's efficacy is to measure its ability to generate a near-complete collection of low-energy conformations. SEMD can preferentially generate a wide array of unique low energy conformations. In the same calculation where all 262 lowest 3 kcal/mol conformations were discovered, a Maxwell-Boltzmann-like distribution of low-energy structures was obtained (Figure 2.4). The presence of a high-energy tail is most likely due to having filled in most of the finite low-energy space and beginning to fill in the more of the vast high energy space. Of the 3,622 conformations found, 3,565 were within 10 kcal/mol of the global minimum with the highest of the 3,622 being 13.197 kcal/mol uphill, approximately

half of the 19.093+26 kcal/mol energetic upper limit reported by Ponder[13]. The median energy of the population is 24.535 kcal/mol with a standard deviation of 1.836 kcal/mol.

Over the course of the calculation, after quickly discovering the global minimum, the running average energy (using the previous 50 energies) increases steadily from approximately 22.0 to 25.5 kcal/mol (Figure 2.5). During this rise both low and mid-energy structures are being sampled. This ability to preferentially sample low-energy space makes SEMD a versatile tool in developing the low-energy population distribution of a species.

2.3.B – $C_{39}H_{80}$. An alternative, less common, but perhaps more challenging benchmark sets as its goal finding the global minimum for $C_{39}H_{80}$ (Figure 2.6), a model small strand of polyethylene, in vacuo using the MM3 force field. Rather than adopting an all-trans extended conformation, van der Waals forces allow the molecule to compensate for the energy of less favorable dihedral angles and fold into a β-sheet-like conformation.

Utilizing *l*-LMOD, Kolossváry and Guida[14] discovered the currently known global minima of $C_{39}H_{80}$ after several 10,000 step searches. Long, Tran, Adams, Darwen and Smythe attempted to obtain the global minimum with their PBIL algorithm[16]. They were unable to obtain the global minimum after 4000 full energy minimizations, but did obtain a conformer within 0.2 kcal/mol of the global minimum.

Since the goal for $C_{39}H_{80}$ is to simply find the global minimum rather than fill in conformational space, a different SEMD protocol was used. Instead of performing a single isothermal run, two independent isothermal calculations were performed. The first calculation's goal was to obtain a wide cross section configurational space starting from the fully extended conformation. Here, dynamics was carried out at a higher temperature and dynamics allowed to proceed for an extended amount of time between minimizations. The torsional temperature was

125 K with a t_{cutoff} of 800 fs, Δt = 8.0 fs for 500,000 steps, A=22, and α=8.0. A total of 2048 structures were obtained with the lowest energy conformer found, 2.1 kcal/mol above the global minimum. The second calculation, starting from the lowest energy conformer from the first calculation, was performed to sample low-energy space. The second calculation was performed with a T of 100 K, a t_{cutoff} of 600 fs, 500,000 Δt = 8.0 fs, A=22, and α=8.0. During the course of the calculation the global minimum, with an energy 12.742 kcal/mol was obtained, along with 2410 other low-energy conformers.

The distribution of conformers from the high T calculation is reversed relative to the cycloheptadecane distribution (Figure 2.7)—that is there is a low energy tail, rather than a high energy tail. From the running average trajectory plot (Figure 2.8), the first two thirds of the dynamics is spent attempting to discover a "pathway conformer" to a population of low energy structures. Once the single "pathway conformer" is found (approx. #1,500), SEMD proceeds to follow a path to the bottom of its respective funnel[17] to discover the global minimum.

For the first two thirds of the simulation shown in Figure 2.8, SEMD discovered conformations that are fully extended with a limited number of torsional defects. This gives rise to the seemingly iso-energetic population of conformations. The amount of time spent generating these extended structures can be reduced by allowing the dynamics to run for longer lengths of time before a minimization is attempted coupled with an even higher temperature. However, this has the adverse side effect of generating conformers on pathways to higher-energy basins, resulting in time spent exploring these unproductive basins.

2.4 – Conclusions

A novel protocol has been presented that efficiently searches conformational space for global minima A product of Gaussians is used to effectively edit out minima from a potential energy surface. This surface editing, coupled with a new approach to partitioning space into orthogonal directions has shown to be a fast and efficient means of generating low-energy populations as well as isolating global minima for highly flexible molecules as well as ring systems. For classic benchmarks, SEMD was able to obtain all 262 lowest 3 kcal/mol structures of cycloheptadecane with roughly half the computational effort of current state of the art, as well as discovering the global minimum of $C_{39}H_{80}$ with record computational effort. In both cases, a low-energy Maxwell-Boltzmann-like distribution of structures was obtained. In addition, through temperature variation SEMD can be tailored to generate either a wide cross-section of conformational space or to refine to a low energy structure.

In addition to these accomplishments a novel, robust, and singularity-free approach to measuring conformational torsional angle differences was developed. This unit circle distance was used to compute Gaussian surface modifications as well as define conformational similarity. Use of the unit circle eliminates discontinuities at 0º and 360º or -180º and 180º.

SEMD facilitates new and unique conformational searching aides. For example, conformational symmetry can be used to bias against portions of undiscovered space without having to explore those regions. In the case of highly symmetric cycloheptadecane, 67 additional Gaussian can be added to the surface for each symmetry equivalent position of each discovered conformation. Another unique capability is the ability to project out degrees of freedom from the velocity permitting dynamics to focus on productive modes while still following low-energy paths across the potential energy surface. In this study, stretch and bend

were both removed focusing dynamic motion onto the torsional degrees of freedom, thus allowing larger step sizes and lower effective temperatures.

SEMD can be extended to alternative coordinates to facilitate conformational searching for protein folding or protein-ligand docking. For protein folding, the backbone torsions could be biased against while allowing the side chains to relax thus performing an initial backbone optimization. Once an optimal backbone conformation has been reached, the side chains can then be searched. With protein-ligand docking, the added complexity of rigid-body rotational degrees of freedom can be utilized by SEMD to allow ligands to explore an active-site's low-energy landscape.

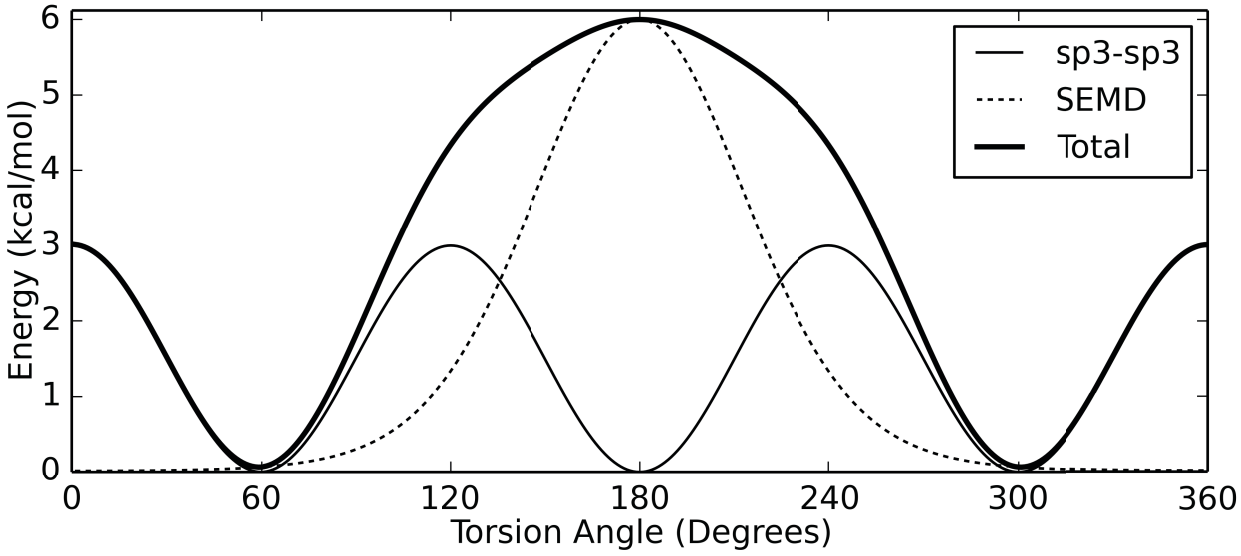


Figure 2.1. Potential energy surface of the sp^3-sp^3 torsional rotation for ethane. The dashed line represents a Gaussian parameterized with A=6 and α =6, centered at the θ=180º minimum. Summation of the carbon potential and Gaussian results in the bold potential energy surface with a maximum at 180º.

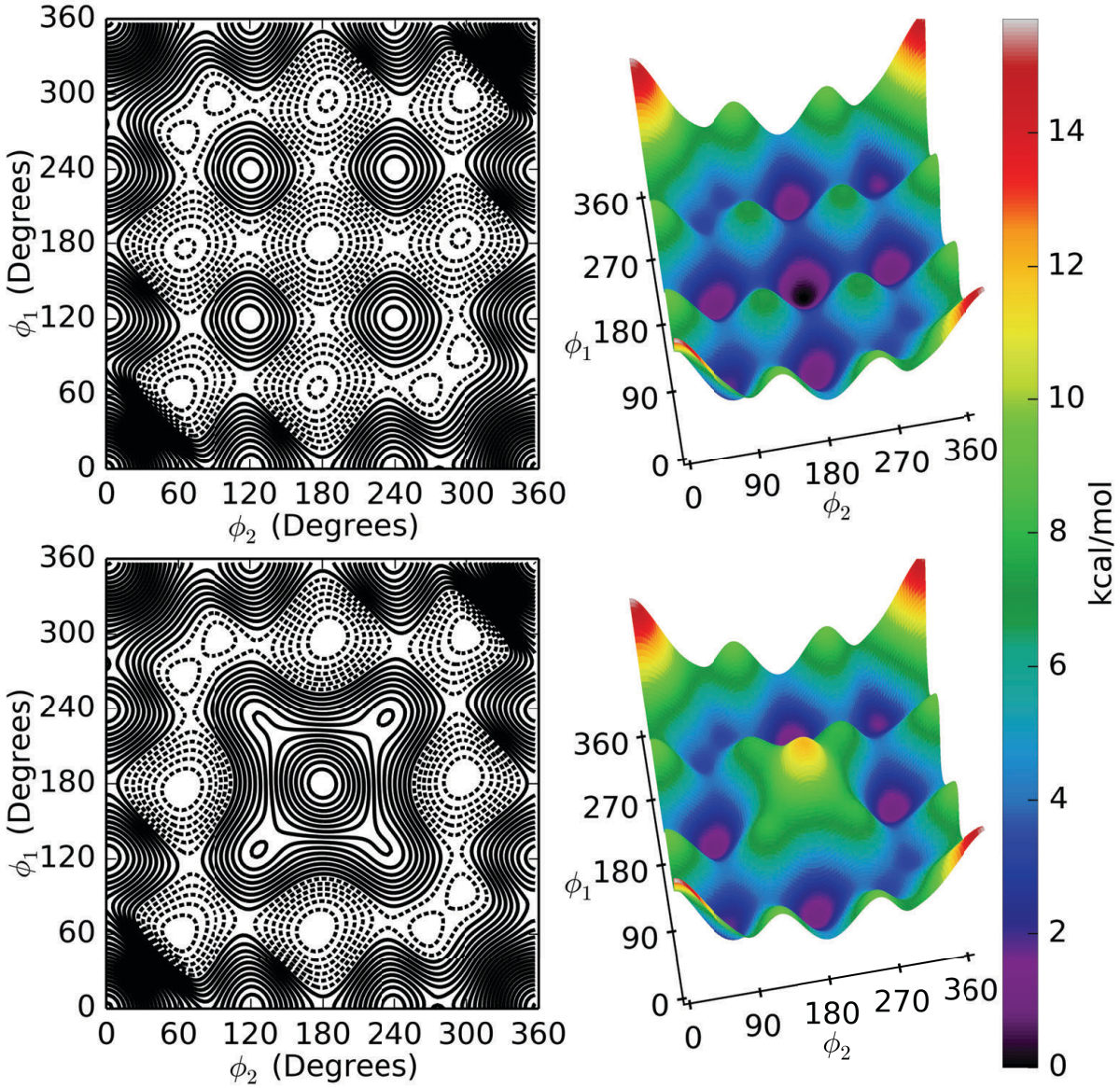


Figure 2.2. The top contour plot (left) and three-dimensional surface (right) represent the unmodified pentane torsion potential energy surface where ϕ_1 and ϕ_2 represent the two C-C-C-C torsions. The bottom contour plot (left) and three-dimensional surface (right) illustrate the impact of a Gaussian centered at the ϕ_1=180º and ϕ_2=180º minimum; parameterized with A=12 and α=6. The result is a maximum at the previous minimum without significant modification to nearby potential minima.

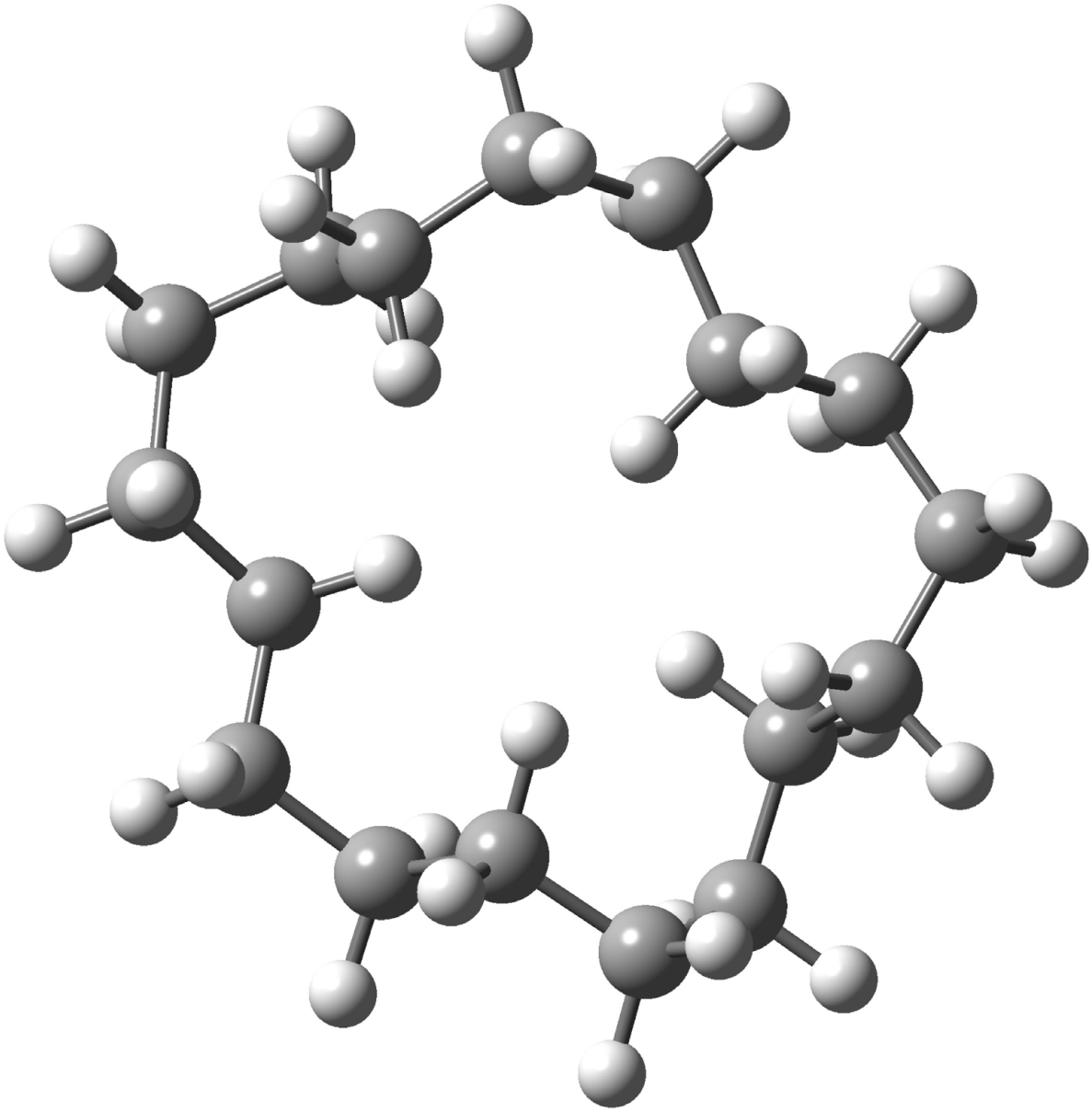


Figure 2.3. MM2 Enthalpic global minimum of cycloheptadecane.

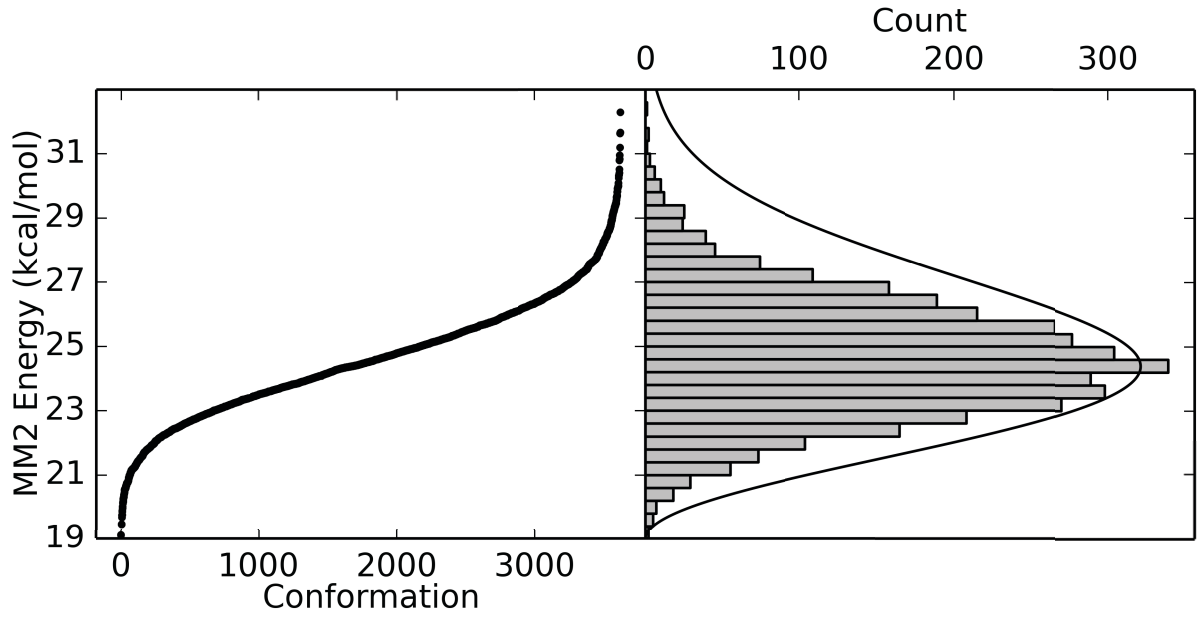


Figure 2.4. Left: Conformations of cycloheptadecane discovered by SEMD, ordered in order of increasing energy. Right: Histogram of Maxwell-Boltzmann-like distribution of discovered conformations (bin width = 0.4 kcal/mol).

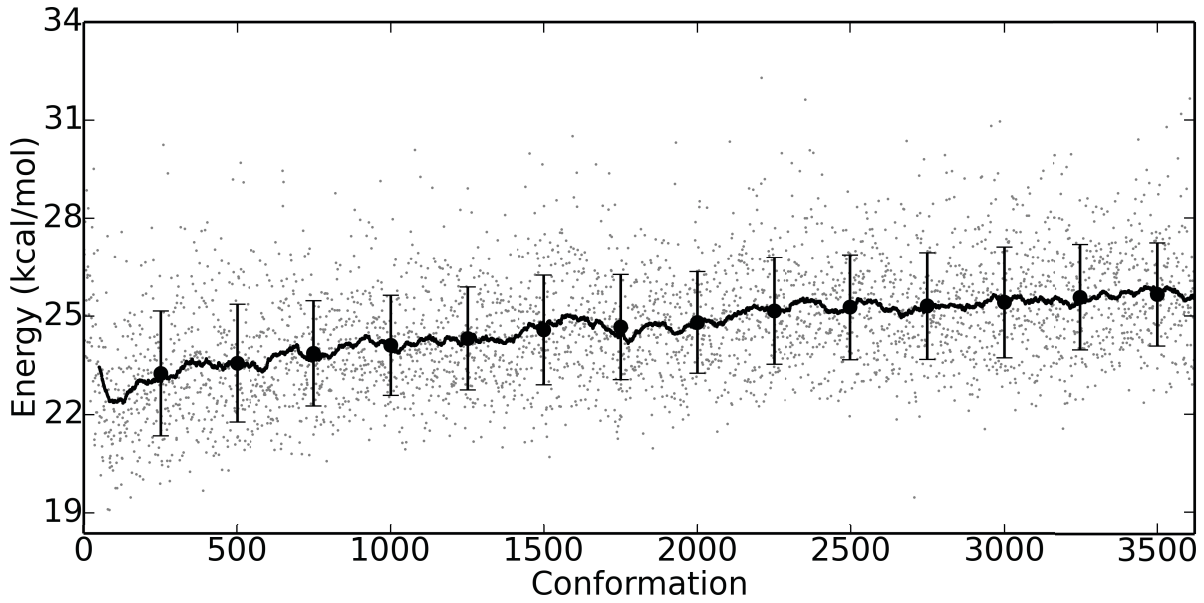


Figure 2.5. Minima of cycloheptadecane obtained over the course of an SEMD simulation. The grey points represent individual minima and the black line represents the running average (previous 50 energies) with running standard deviation (previous 250 energies). The running average gradually increases over the course of the simulation illustrating SEMD's ability to preferentially fill in the low-energy regions of the potential energy surface.

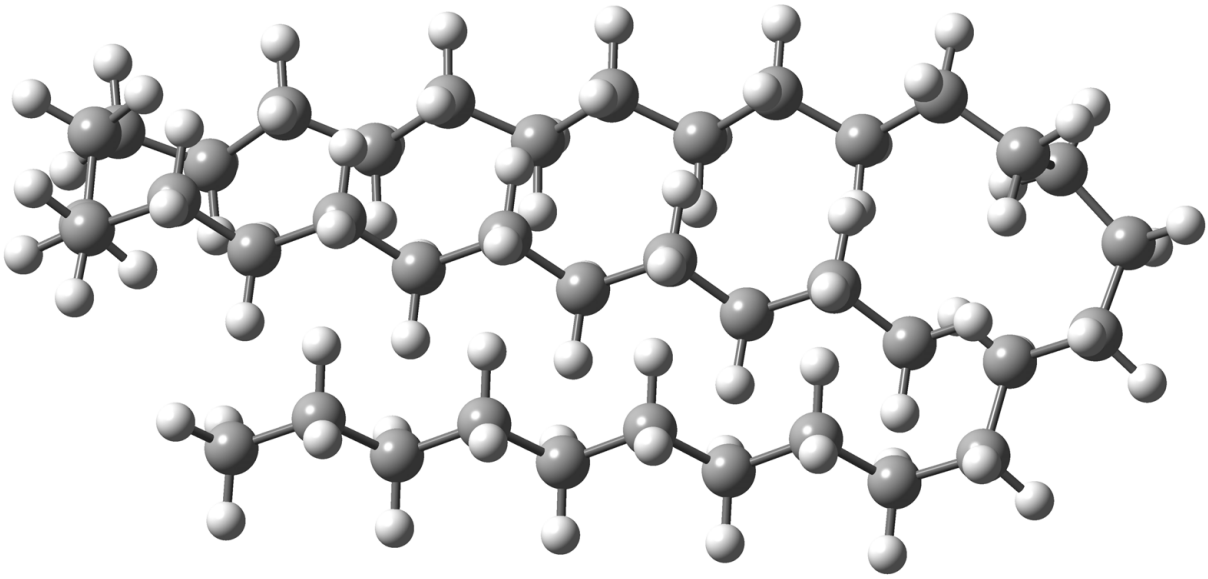


Figure 2.6. In vacuo MM3 forcefield energetic global minimum of $C_{39}H_{80}$.

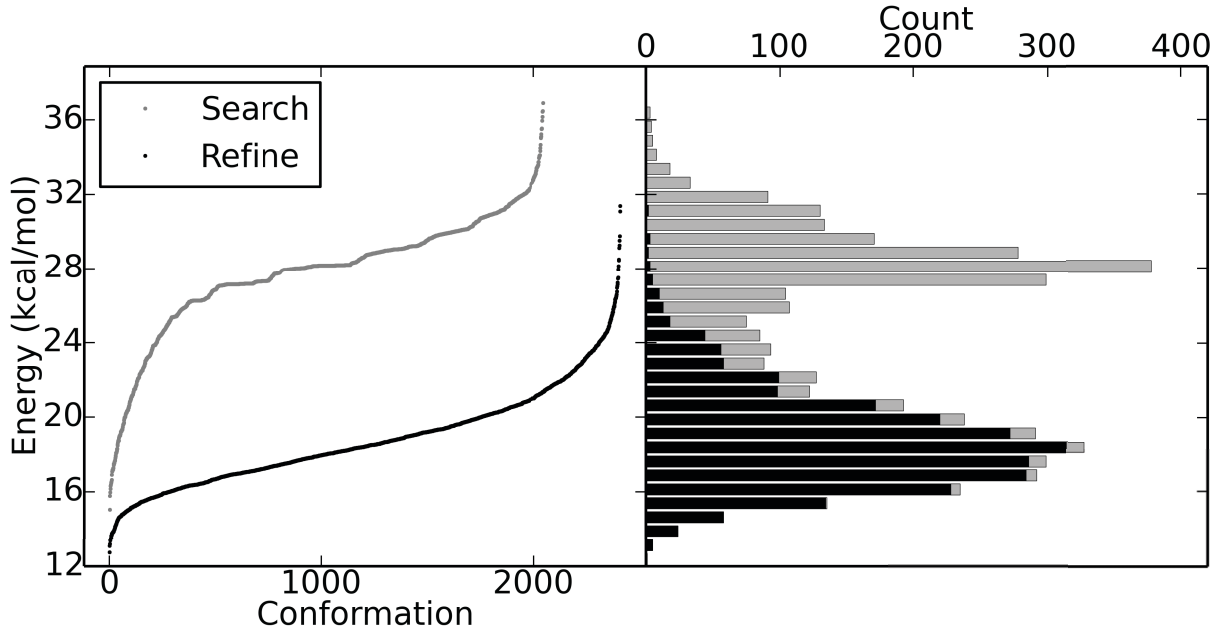


Figure 2.7. Left: Ordered (by energy) conformers of $C_{39}H_{80}$ wherein the searching simulation (gray) utilizes a high temperature (125 K) and long dynamics time before minimization (800 fs) to obtain a wide cross-section of structures. The refinement simulation (black) starting from the previous simulation's lowest energy conformer utilizes a low temperature (100 K) and short dynamics time before minimization (600 fs) to refine the structure to the MM3 global minimum. Right: Stacked histogram representing the distribution of structures for each simulation wherein a Maxwell-Boltzmann distribution can be easily observed for the refinement simulation (black).

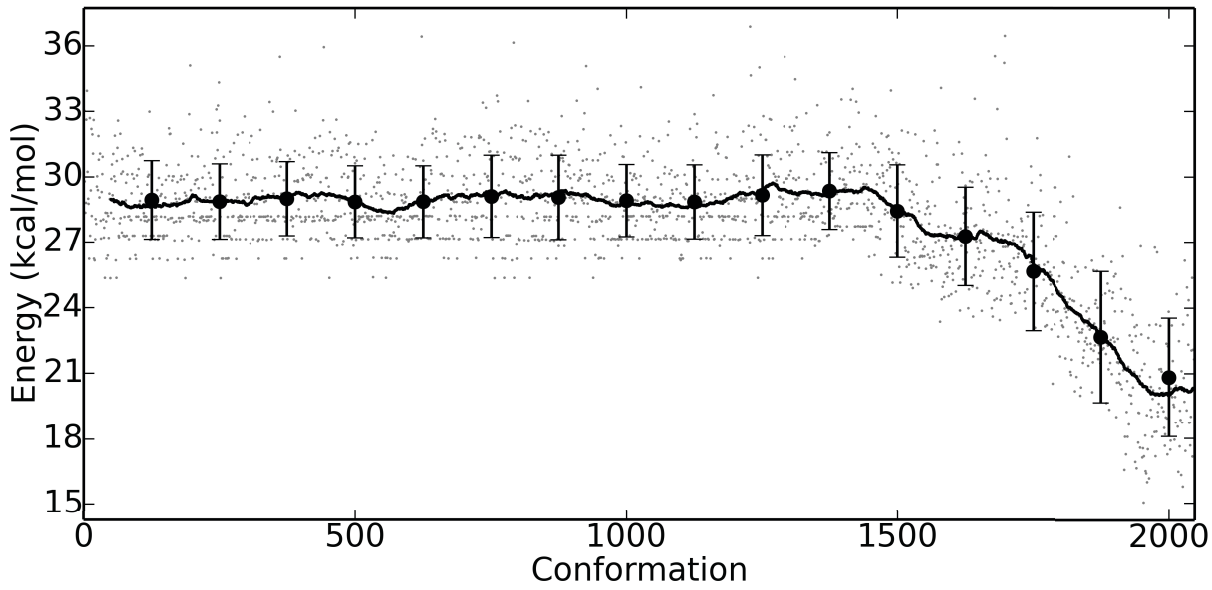


Figure 2.8. Conformational trajectory of the search simulation, the 50 conformer running average (black) with 125 conformer running standard deviation and each conformation (gray) as they are discovered are provided. The simulation discovers a conformation at approximately #1500 that leads to the global minimum subsequently discovered in the refinement run.

CHAPTER 2 REFERENCES

(1) Chang, G.; Guida, W. C.; Still, W. C. *J. Am. Chem. Soc.* **1989**, *111*, 4379–4386.

(2) Swope, W. C.; Ferguson, D. M. *J. Comput. Chem.* **1992**, *13*, 585–594.

(3) Golub, P. G. H.; Reinsch, D. C. *Numer. Math.* **1970**, *14*, 403–420.

(4) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.

(5) Andersen, H. *J. Comput. Phys.* **1983**, *52*, 24–34.

(6) Pulay, P. In *Applications of Electronic Structure Theory*; III, H. F. S., Ed.; Modern Theoretical Chemistry; Springer US, 1977; pp. 153–185.

(7) Lu, D.; Zhao, M.; Truhlar, D. *J. Comput. Chem.* **1991**, *12*, 376–384.

(8) Lu, D.; Truhlar, D. G. *J. Chem. Phys.* **1993**, *99*, 2723.

(9) Saunders, M.; Houk, K. N.; Wu, Y. D.; Still, W. C.; Lipton, M.; Chang, G.; Guida, W. C. *J. Am. Chem. Soc.* **1990**, *112*, 1419–1427.

(10) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511–519.

(11) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.

(12) Ngo, J. T.; Karplus, M. *J. Am. Chem. Soc.* **1997**, *119*, 5657–5667.

(13) Pappu, R. V.; Hart, R. K.; Ponder, J. W. *J. Phys. Chem. B* **1998**, *102*, 9725–9742.

(14) Kolossváry, I.; Guida, W. C. *J. Comput. Chem.* **1999**, *20*, 1671–1684.

(15) Kolossváry, I.; Guida, W. C. *J. Am. Chem. Soc.* **1996**, *118*, 5011–5019.

(16) Long, S. M.; Tran, T. T.; Adams, P.; Darwen, P.; Smythe, M. L. *J. Comput. Chem.* **2011**, *32*, 1541–1549.

(17) Dill, K. A.; Chan, H. S. *Nat. Struct. Mol. Biol.* **1997**, *4*, 10–19.

CHAPTER 3: PROGRESS TOWARDS ESTIMATING CONFIGURATIONAL ENTROPY OF HYDROCARBONS VIA SURFACE EDITING MOLECULAR DYNAMICS

3.1 – Introduction

As discussed in the introduction, a wide array of metrics may be obtained from the resulting low-energy population of either a sampling or searching method. Estimation of one of these metrics, relative entropy, via molecular dynamics is as van Gunsteren describes[1], "a notoriously difficult problem that currently constitutes one of the key challenges in computational chemistry". In 1981, Karplus[2] presented a means of computing the configurational entropy of a system. Specifically, he proposed computing the configurational entropy of a species via the probability distribution of the geometries of that species.

Configurational entropy, being a subset of the more general entropy term, is utilized to evaluate the relative "flexibility" of a single molecule or, more accurately described, its accessible configurational space[3]. Furthermore, it is only related to position (and not the momentum) of the species over some measurable quantity, such as potential energy.

This chapter serves as a brief illustration of the surface editing molecular dynamic method's efficacy at obtaining the conformational contribution to the configurational entropy. The relevant background is presented in section 3.2, the computational methods given in section 3.3, preliminary results with an accompanying discussion is collected in section 3.4, and conclusions in section 3.5.

3.2 – Background

3.2.A – Configurational Entropy: As proposed by Karplus and Kushick[2], configurational entropy is related to the probability distribution function of a given geometry, $P(\boldsymbol{q})$, and is of the form

$$S_{conf} = -k_B \int P(\boldsymbol{q}) \ln P(\boldsymbol{q})\, d\boldsymbol{q} \qquad (3.1)$$

where k_B is Boltzmann's contant and $P(\boldsymbol{q})$ is defined as

$$P(\boldsymbol{q}) = \frac{1}{Q} e^{-V(\boldsymbol{q})/k_B T} \qquad (3.2)$$

and $V(\boldsymbol{q})$ is the potential energy of the conformation, and T is the temperature. The partition function, Q, is defined as

$$Q = \int e^{-V(\boldsymbol{q})/k_B T} \qquad (3.3)$$

Utilizing this system of equations, a configurational entropy can be constructed which gives insight into the conformational flexibility of a system. Ideally, integration over the entire surface is preferable, however due to an infinite amount of time required to achieve this by sampling methods, in practice summation over the obtained population must suffice.

Karplus et al. later expanded upon configurational entropy[4] by partitioning the configurational entropy into two terms, vibrational (S_{vib}) and conformational ($S_{conform}$) entropy:

$$S_{conf} = S_{vib} + S_{conform} \qquad (3.4)$$

Computation of $S_{conform}$ is then computed from the probability distribution of the discrete minima, P_i, using the Shannon informational entropy[5] and is found in Eqn. (3.**5**).

$$S_{conform} = -k_B \sum_{i=1}^{n} P_i \ln P_i \quad (3.5)$$

The probability distribution, P_i, is computed similarly to Eqn. (3.**2**).

$$P_i = \frac{1}{Q_C} e^{-V_i/k_B T} \quad (3.6)$$

Where the partition function, Q_C, is

$$Q_C = \sum_{i=1}^{n} e^{-V(\boldsymbol{q})/k_B T} \quad (3.7)$$

3.2.B. Configurational Entropy Estimation of Hydrocarbons: In 2006, van Gunsteren et al.[1] compared atomic-level and coarse-grained models for estimating the configurational entropy of hydrocarbons with molecular dynamics. They utilized the GROMOS 45A3[6] for the united atom atomic-level simulations and a coarse-grained force field[7] for their course-grain simulations. They found that there was a loss of configurational entropy from the atomic-level to coarse-grain models. The decrease was on the order of 40-100 J mol^{-1} K^{-1} per bead.

3.3 – Methodology

All conformational searching presented here is performed with the SEMD method with the MM2/MM3 force fields[8,9]. The SEMD parameters A and alpha are equal to 22 and 8 respectively. Simulation length varied from system to system and is reported within each case.

3.4 – Results and Discussion

3.4.A – Cycloheptadecane: Cycloheptadecane provides perhaps the most time efficient system to determine the SEMD's ability to estimate $S_{conform}$. Cycloheptadecane's well-characterized low-energy population of minima can be utilized as a test bed for determining

convergence of the conformational contribution to the configurational entropy. Additionally, the data has already been gathered to benchmark SEMD.

Utilizing the population of structures from successful search of 262 conformers from the previous chapter, a value for $S_{conform}$ can be estimated. Specifically, the evolution of $S_{conform}$ over the course of the search for the 262 conformers is observed by applying Eqns. (3.**5**)-(3.**7**) to obtain $S_{conform}$ as a function of conformers discovered (Figure 3.1).

After 3621 conformers, $S_{conform}$ is equal to 44.2 J/(mol K) with a standard deviation of the last 50 conformers equal to 0.01 J/(mol K). This suggests that $S_{conform}$ has converged, thus systems with comparable complexity to cycloheptadecane may be treated by SEMD.

3.4.B – $C_{39}H_{80}$: Conforming to the same reasoning as with cycloheptadecane, the well characterized global minimum and already present $C_{39}H_{80}$ conformational searching data from the previous chapter can be readily analyzed to provide another test. Due to the nature of the search (finding the global minimum), the low energy population should not be entirely present.

After performing the analysis with the population of 2410 minima, it is found that $S_{conform}$ is estimated to be 41.6 J/(mol K) with a standard deviation of the last 50 conformers equal to 0.06 J/(mol K). Upon observing the evolution of $S_{conform}$ (Figure 3.2), it is apparent that the majority of low-energy population has not been discovered and the value has not quite converged. Additionally, comparing the value of $S_{conform}$ for $C_{39}H_{90}$ to the value for cycloheptadecane does not track with a qualitative assessment of flexibility. However, with this case of each species being simulated in a vacuum, their values may be similar due to the tendency of long hydrocarbon chains favoring “folded” geometries, thus excluding wide swaths of extended conformations due to the relative stability of the folded conformers.

A more accurate estimation may be achieved by running the simulation for a longer period of time to elicit a larger population of low-energy minima.

3.4.C – Solvation Effects on $S_{conform}$: To illustrate the effects of solvation on the conformational contribution to configurational entropy, a simple case was devised with the relatively "small" molecule, octane, and the implicit solvent model, solvent accessible surface area (SASA)[10]. Due to octane's short length, the folded conformations are not energetically feasible due to the lack of accessible stabilizing van der Waals (vdW) interactions relative to the destabilizing torsion angles that would be required to achieve them. However with the inclusion of implicit solvent parameterized for methane, the energetic gap between conformers should widen, favoring the fully extended conformations and destabilizing the conformations with torsional defects.

By scaling the implicit solvent quality from 0 (vacuum) to 1 (equivalent to methane solvent) in 0.2 increments and performing a SEMD search at each increment, the impact of solvent was elicited (Figure 3.3). As expected, the vacuum equivalent simulation allows for the most accessible conformations as evidenced by the relatively higher entropy, 34.0 J/(mol K). As the solvent quality improves, the number of accessible geometries decreases, reducing the entropy to 30.8 J/(mol K) for the perfectly solvated simulation.

3.5 – Conclusions

Surface editing molecular dynamics proves to be an intriguing choice for computing the conformational contribution to the configurational entropy by preferentially eliciting low-energy minima. An estimation of the vibrational contribution may be obtained by evaluating the Hessian at each low-energy minima, which is already done within SEMD to obtain the starting

velocities. Therefore by properly utilizing the information from the Hessian, SEMD will be able to estimate the full configurational entropy of a given species.

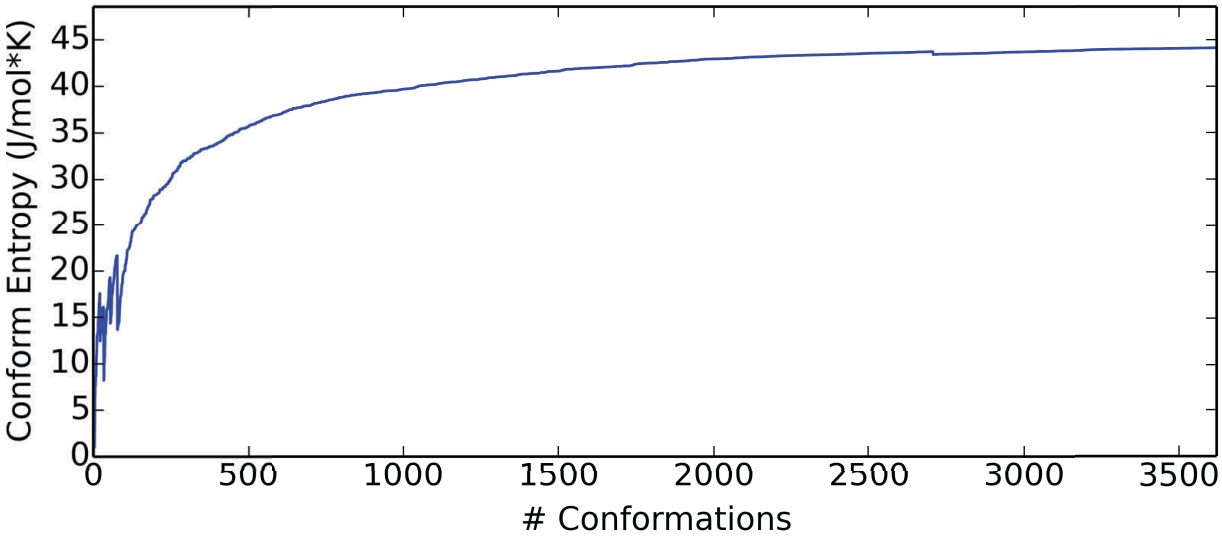


Figure 3.1. Conformational contribution to the configurational entropy for cycloheptadecane.

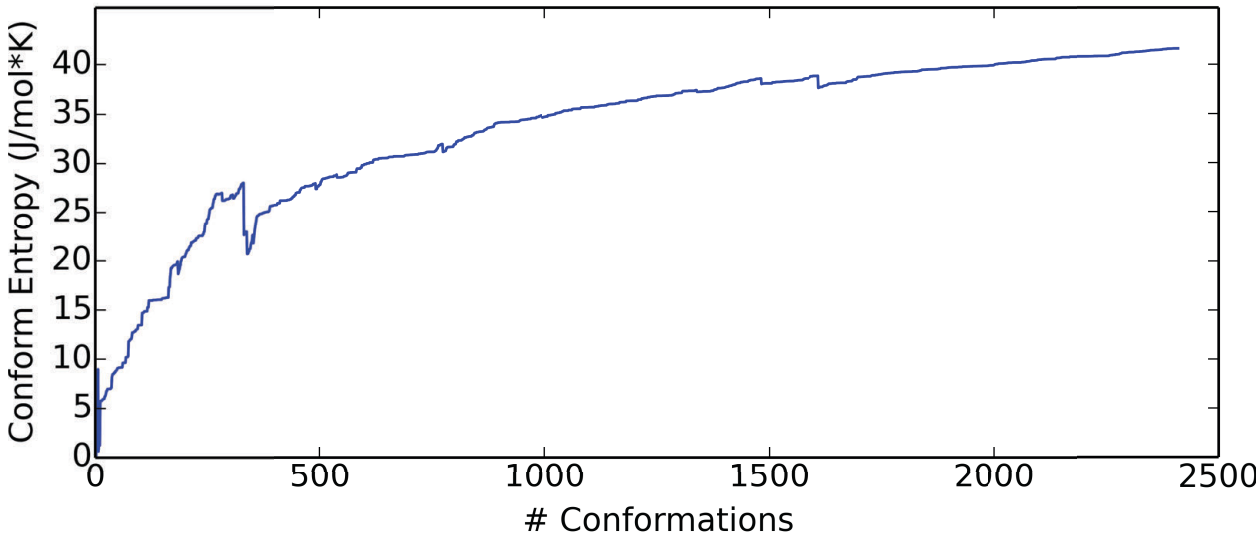


Figure 3.2. Conformational contribution to the configurational entropy for $C_{39}H_{80}$.

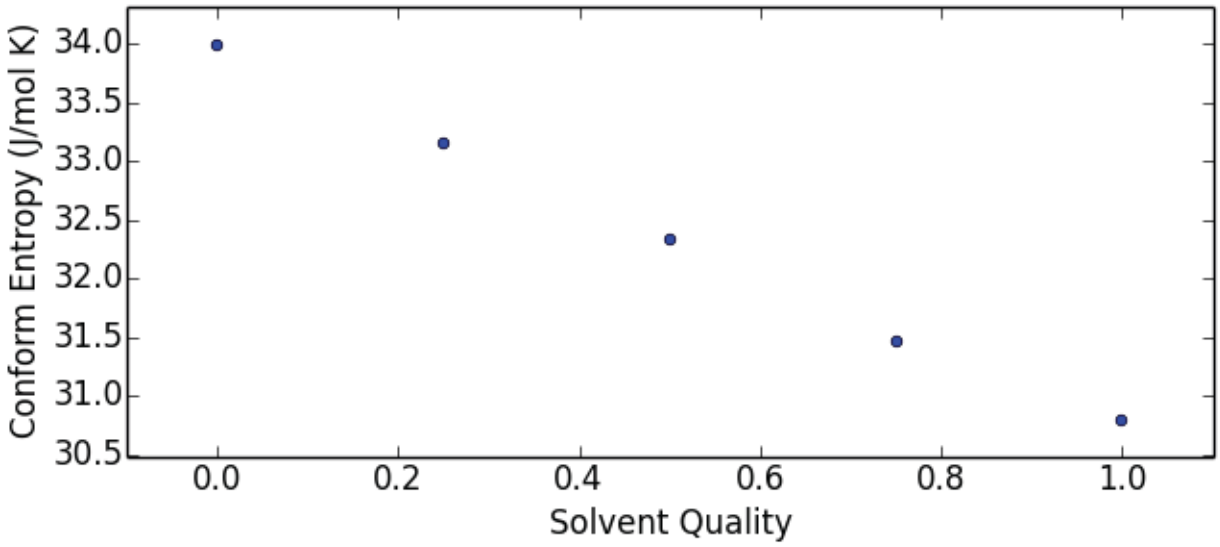


Figure 3.3. Conformational contribution to the configurational entropy for octane as a function of solvent quality.

CHAPTER 3 REFERENCES

(1) Baron, R.; de Vries, A. H.; Hünenberger, P. H.; van Gunsteren, W. F. *J. Phys. Chem. B* **2006**, *110*, 8464–8473.

(2) Karplus, M.; Kushick, J. N. *Macromolecules* **1981**, *14*, 325–332.

(3) Suárez, E.; Díaz, N.; Suárez, D. *J. Chem. Theory Comput.* **2011**, *7*, 2638–2653.

(4) Karplus, M.; Ichiye, T.; Pettitt, B. M. *Biophys. J.* **1987**, *52*, 1083–1085.

(5) Shannon, C. E.; Weaver, W. *The mathematical theory of communication*; University of Illinois Press: Urbana, 1964.

(6) Schuler, L. D.; Daura, X.; Van Gunsteren, W. F. *J. Comput. Chem.* **2001**, *22*, 1205–1218.

(7) Marrink, S. J.; de Vries, A. H.; Mark, A. E. *J. Phys. Chem. B* **2004**, *108*, 750–760.

(8) Allinger, N. *J. Am. Chem. Soc.* **1977**, *99*, 8127–8134.

(9) Allinger, N.; Yuh, Y.; Lii, J. *J. Am. Chem. Soc.* **1989**, *111*, 8551–8566.

(10) Richmond, T. J. *J. Mol. Biol.* **1984**, *178*, 63–89.

CHAPTER 4: THEORETICAL STUDY OF BIS(ENONE) PHOTOCATALYTIC [2+2] CYCLOADDITION: ALTERNATIVE PATHWAYS

4.1 – Introduction

The re-emerging area of photoredox catalysis offers the opportunity to carry out endothermic or activated transformations under mild conditions without pre-activation of reagents, thus reducing energy requirements, mitigating the use of toxic reagents, and minimizing waste. In the past five years there have been more than 150 publications where photoredox catalysis was used to carry out useful synthetic transformations, following both reductive as well as oxidative pathways. The highlights of this work are captured in several recent reviews.[1–3]

The diastereocontrolled [2+2] cycloaddition of enones (Figure 4.1) is a synthetically useful example of the reductive pathway. In this work, photo-excited $[Ru(bpy)_3]^{2+}$* is reduced by sacrificial diisopropylethyl amine in the presence of $LiBF_4$. Other examples of useful transformations have been developed by several groups, including Stephenson's[2,3] functionalization of indoles and pyrroles via intramolecular radical cyclizations and MacMillan's reductions of alkyl and α-bromo carbonyls.[4]

The Yoon photocatalyzed [2+2] cycloaddition[5] example is of particular interest as it is a relatively rare example of photocatalytic diastereocontrol. It forms two C-C bonds, does not require a leaving group, and it highlights a potentially significant specific ion effect---$LiBF_4$ is required; replacement by $Bu_4N•BF_4$ or $NaBF_4$ leads to no observable consumption of enone. In contrast, electrochemical reduction leads to a diverse product distribution.

This chapter outlines the theoretical characterization of possible mechanisms for photocatalyzed [2+2] cycloaddition involving phenyl substituted bis-enones while also investigating potential sources of specific ion effect as well as diastereocontrol. The relevant literature is reviewed in section 4.2, the computational method is in section 4.3, results are collected in section 4.4, discussion of the most likely pathway(s) is in section 4.5, and conclusions are provided in section 4.6.

4.2 – Background

Diphenyl and phenyl alkyl substituted bis(enone)s have been reduced under a wide range of conditions that have resulted in varying products and diastereoselectivities[5–9]. Aromatic bis(enone)s have been subjected to cathodic reduction, chemical induction, metal catalysis, as well as photocatalysis each yielding a different product distribution. Each of the following species and reactions correspond to entries in Figure 4.3 and Table 4.1, respectively.

4.2.A – Cathodic Reduction. Roh, Jang, Lynch, Bauld, and Krische,[6] reported the cathodic reduction of the bis(enone)s at a voltage of -0.90 V versus SCE in acetonitrile and lithium perchlorate. The *cis* and *trans* [2+2] cycloadducts (**2a** & **2b**), a [4+2] Diels-Alder adduct (**2f**), both cis and trans simple reductive cyclization products (**2g** & **2h**), and a reductive cyclization-aldolization product are observed (**2c**). No diastereoselectivity is observed as well as little product selectivity.

4.2.B – Chemical Induction. Yang, Felton, Bauld, and Krische[7] reported the chemically induced bis(enone) cycloaddition with a chrysene radical anion in THF. This yielded the *cis* [2+2] cycloadduct (**2a)**, the [4+2] Diels-Alder cycloadduct (**2f**), reductive cyclization-aldolization product (**2c**), and the simple *cis* reductive cyclization product (**2g**). When a napthyl

group was substituted for the phenyl, selectivity for the *cis* [2+2] cycloadduct was increased by a factor of two.

4.2.C – Metal Catalysis. Baik, Luis, Wang, and Krische[8] reported that catalytic reduction of the bis(enone) with cobalt 2,2,6,6-tetramethylheptane-3,5-dionate [$Co(dpm)_2$] with stoichiometric phenylmethylsilane in dichloroethane resulted in the formation the *cis* [2+2] cycloadduct (**2a**) in a 7:1 ratio over the Michael cycloadduct product (**2e**), at 50ºC.

4.2.D – [Ru(bpy)$_3$] Photocatalysis. In 2008, Ischay, Anzovina, Du, and Yoon[5] described the photocatalytic [2+2] cycloaddition of bis(enone). A nearly 10:1 diastereometric ratio of *cis* to *trans* [2+2] cycloadduct (**2a** & **2b**) was obtained utilizing ruthenium tris(bipyridine), diisopropylethylamine (DIPEA), lithium tetrafluoroborate, acetonitrile, and sunlight. It was reported that under the same conditions alkyl bis(enones) did not react.

4.2.E – Eosin Y Photocatalysis. More recently, in 2013, Neumann and Zeitler[9] reported the selective formation of the *trans* [2+2] cycloadduct (**2b**) (96% yield). Eosin Y was used as the photocatalyst, along with lithium bromide, and DIPEA in acetonitrile and with 530 nm light. Alternatively the bis(enone) produced the simple *trans* reductive cyclization product (**2h**) (93% yield) when Eosin Y was combined with the reductive quencher, Hantzcsh ester, and a thiourea organocatalyst, in dichloromethane solvent.

4.2.F – Theoretical Studies. In 2009, Zhang, Li, and Chen[10] performed a theoretical study of the intramolecular anion radical cycloaddition of the phenyl-substituted bis(enone) using the density functional B3LYP with the 6-311+G(d,p) basis set. The PCM implicit solvent model was used simulate the effects of THF. The [2+2] cycloaddition reaction was reported to be thermodynamically unfavorable *in vacuo* and in THF solvent.

In 2012, Yuan, Zhang, and Chen[11] performed another study on the bis(enone) [2+2] cycloaddition, this time reporting a substituent study. Again, *in vacuo* and PCM THF as well as acetonitrile solvated results were reported. Depending upon substituent either the product was not a stationary point or the reaction was enthalpically unfavorable. Mulliken atomic spin densities were analyzed and it was concluded that electron-withdrawing substituents were required for the [2+2] cycloadduct to form.

4.3 – Methodology

The seemingly straightforward bis(enone) [2+2] cycloaddition reaction posses several computational challenges.

1) *Conformational flexibility*. The bis(enone) reactant and intermediate possess 8 and 6 conformationally significant torsion angles, respectively. An exhaustive coupled two and three-fold torsion tree-search would generate 144 intermediate geometries and 1,296 geometries for the reactant. Excluded volume effects diminish the number of accessible conformations, but evaluating even a fractional subset of structures is intractable with electronic structure methodologies.

2) *Solvent interactions*. The bis(enone) [2+2] cycloaddition involves anion formation in a 37 debye acetonitrile solvent medium. Gas phase calculations will yield incorrect characterization of the species' charge polarization and thus incorrect relative energies between conformers and intermediates along the reaction coordinate. It is necessary to include solvent interactions via an implicit solvent model.

3) *Potential π-stacking*. The phenyl substituents of the bis(enone), at least one of which is required for reactivity, potentially participate in favorable t-shaped or parallel displaced π-

stacking. π-stacking interactions could stabilize otherwise unfavorable conformations, altering the calculated potential energy landscape. Because π-stacking is predominantly dispersive in character, the selected computational model must include dispersion to accurately describe conformational structure and conformation energy differences.

4) *Counter-ion*. Yoon cycloaddition does **not** occur without added $LiBF_4$. Inclusion of a properly solvated lithium ion is essential to accurately modeling the system in question.

All of these challenges must be considered in tandem with a characterization of potentially valid reaction pathways that lead to products (Figure 4.2). For each species in Figure 4.2, there should be a conformational search to generate a population of low energy structures, inclusion of a proper solvent model, inclusion of dispersion, and an evaluation of the energetic and conformational impact of the lithium ion. It should be noted that the reaction diagram in Figure 4.2 is stereo-agnostic and therefore does not include rotational events that could lead to the isomeric ratio of products experimentally observed. Evaluation of the rotational barriers of each species and their relative energies along with the transition states connecting each can give insight into the source of diastereoselectivity.

All DFT calculations were performed in Gaussian 09 utilizing the hybrid functional APF[12]. In order to effectively add an empirical dispersion term to a density functional model, APF is designed as a linear combination of the two hybrid functionals to minimize long-range interactions. Specifically, the combination of 41.1% B3PW91[13,14], which is repulsive at long range for methane dimer, with 58.9% PBE1PBE[15–17], which is attractive at long range for methane dimer achieved an "average" of near zero at long range. With long-range interactions removed, a specifically tailored empirical dispersion correction (APF-D) was added to accurately describes the binding curve of methane dimer as well as benzene dimer. A 6-311+G(d) was used

for all atoms in all calculations. Diffuse functions were added to provide a more balanced description of neutral and anionic species. In calculations where solvent is considered, a polarizable continuum model (PCM)[18,19] was used to approximate the effects of acetonitrile (ε=37). Unless otherwise noted the energies reported are differences in quantum mechanical Total Energy. Ideal gas statistical thermodynamic corrections were applied to the neutral products.

4.4 – Results

4.4.A – Bis(enone) conformations. Bis(enone), **1**, contains 8 low-barrier torsions (Figure 4.4). There are 2 aromatic sp^2-carbonyl sp^2 single bonds, 2 olefinic sp^2-carbonyl sp^2 single bonds, 2 olefinic sp^2-sp^3 single bonds and 2 sp^3-sp^3 bonds. Single bonds between aromatic sp^2-carbonyl sp^2 centers typically have 2-fold rotational barriers of 5 kcal/mol (for acetophenone there is a 2-fold 5.4 kcal/mol barrier). Acrolein provides an example of an olefinic sp^2-carbonyl sp^2 single bond (2-fold 6.6 kcal/mol barrier). 1-butene suggests that bonds between olefinic sp^2-sp^3 have 3-fold barriers of 3 kcal/mol, and ethane suggests that sp^3-sp^3 single bonds have 3-fold barriers of around 3 kcal/mol. This suggests $2^4 x 3^4$=1,296 conformations. Steric interactions as well as π-stacking interactions contribute to distortions away from torsional minima.

A torsion-tree search was performed on **1** utilizing the molecular mechanics force field APT (Figure 4.5). Of the set of structures obtained, 72 structures were found with a molecular mechanics energy lower than the fully extended conformation. Of those 72 structures, 5 were chosen for treatment via electronic structure methodologies. They were chosen based on their conformational diversity and their likelihood of forming conformationally unique *cis*-5-membered ring intermediates.

The neutral structures were then minimized via APF-D/6-311+g* with and without PCM acetonitrile implicit solvent (Figure 4.6) and their energies obtained (Table 4.2). Of these five structures, the two lowest energies conformers (**1a** & **1e**) were used for further minimizations in addition to the fully extended conformation (to maintain as a reference point). The two conformers primarily differ in their orientation of their ketones, where for structure **1a** the ketones are parallel whereas for **1e** the ketones are antiparallel. The unfavorable interaction of the two parallel dipoles in **1a** in vacuum is mitigated by solvent as evidenced by the widening energy difference between **1a** and **1e** from vacuum to solvent.

1a, **1e** and the fully extended conformer were minimized with B3LYP[20] to assess the impact of dispersion on the geometries and their relative energy (Figure 4.7). The π-stacking observed within **1a** when evaluated with APF-D was lost with B3LYP as evidenced by an interatomic distance between the two closest carbons from each phenyl group increasing from 3.4 to 4.9 Å. Additionally, their relative energies invert, favoring the fully extended conformation by 2.6 kcal/mol over both **1a** and **1e**.

In addition to the thermally accessible barriers discussed above, **1**, possesses two sp^2-sp^2 double bonds. Formation of diastereoisomer **2b** requires rotation about one of these double bonds at some point along the reaction coordinate (Figure 4.8).

4.4.B – Ruthenium Photocatalysis. The low energy photophysics of [$Ru(bpy)_3^{2+}$] consists of an initial metal to ligand charge transfer transition (MLCT) from a metal-centered t_{2g} orbital to a ligand centered π* orbital. This initially formed singlet excited state undergoes rapid intersystem crossing (30 fs)[21] to a low-lying, long-lived triplet excited state (890 ns lifetime)[22]. The low-lying triplet excited state can either gain an electron from an auxiliary agent or donate an electron to an auxiliary agent. In the present case, the Ru(II) reactant is reduced by *i*-Pr_2NEt.

The energetics of this reaction, including charge balancing counterions, is summarized in Eqn. (4.**1**) with the geometries shown in Figure 4.9.

$$Ru(bpy)_3 \bullet (BF_4)_2 + i\text{-}Pr_2NEt \longrightarrow Ru(bpy)_3 \bullet BF_4 + i\text{-}Pr_2NEt \bullet BF_4 \qquad (4.1)$$

Initial photo-excitation/intersystem crossing leads to a triplet excited state computed to be 50.4 kcal/mol above the ground state. Reduction of the ^{3}Ru(II) complex by *i*-Pr_2NEt is computed to be 1.1 kcal/mol downhill, in implicit acetonitrile solvent.

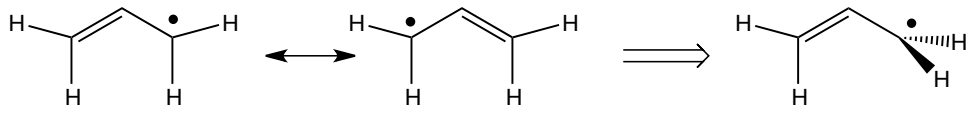
4.4.C – Bis(enone) reduction by Ru(I). Since reductive cycloaddition of **1** has been found to depend upon the nature of the counterion, reduction of **1** by the Ru(I) complex [$Ru(bpy)_3$•BF_4] was examined as a function of counterion (Eqn. (4.**2**)). Tetra methyl ammonium ion $(N(CH_3)_4)^+$ (TMA) (Figure 4.10), a weakly-coordinating counterion, and two coordinating counterions Li(acetonitrile)$_2$, and Na(acetonitrile)$_2$ were studied. For TMA, reduction is computed to be 8.6 kcal/mol downhill, for Li(acetonitrile)$_2$ 12.3 kcal/mol downhill, and for Na(acetonitrile)$_2$ 10.0 kcal/mol downhill. In addition, alkyl bis(enones) are not observed to react. Reduction of the dimethyl bis(enone) is computed to be exothermic, but only by 5.6 kcal/mol.

$$Ru(bpy)_3 \bullet BF_4 + \text{bis(enone)} + \text{cation} \bullet BF_4 \longrightarrow Ru(bpy)_3 \bullet (BF_4)_2 + \text{bis(enone)} \bullet \text{cation} \qquad (4.2)$$

Alternatively, observed reactivity could potentially proceed through the lowest triplet excited state of **1**. Energy transfer from the triplet Ru(II) complex could form this triplet excited state of the bis(enone). This energy transfer is computed to be 21.7 kcal/mol uphill. The lowest triplet state of **1** could also be formed through back electron transfer from the bis(enone) radical anion to the amine radical cation. This back electron transfer is computed to be 23.7 kcal/mol uphill for the bis(enone) ground state geometry.

4.4.D – Bis(enone) radical anion conformations. Addition of an electron to **1** leads to the formation of a discotic radical---the negative charge is localized on the carbonyl oxygens while radical character is centered on the β carbons. The structures of **1a**, **1e**, and the extended conformers do not change significantly with the addition of an electron. However the spin density for each is different due to specific conformational differences (Figure 4.11). For the parallel-displaced conformation, **1a**, the electron spin density is delocalized equally over each half of the bis(enone). For the antiparallel conformation, **1e**, the spin density is more localized on the half that coordinates to the lithium cation. Finally, the extended conformer's spin density can be characterized as being entirely localized onto the lithium coordinating half.

4.4.E – Rotational barrier for bis(enone) radical anion. One possible mechanism for diastereoinversion (formation of **2b**) is that the initially formed bis(enone) radical anion could undergo rotation about the C_α-C_β, now formally single bond (Figure 4.8). The computed barriers for rotation with TMA and Li(acetonitrile)$_4$ counterions are 25.6, and 21.7 kcal/mol, respectively. The high barriers can be attributed to allylic stabilization of the radical center, see Eqn. (4.**3**).



(4.3)

4.4.F – Transition state to formation of intermediate. The structure and spin density plot for the lowest energy, π-stacked radical anion configuration, **1a,** (Figure 4.11) suggests a straightforward ring closure forming a 5-membered ring intermediate (Figure 4.12), **IMa**. The computed barriers for ring closure with TMA and Li(acetonitrile)$_2$ counterions are 6.0, and 10.5 kcal/mol, respectively and the structures found in Figure 4.13. The bond-forming barrier is significantly smaller than the rotational barrier for the bis(enone) radical anion or back electron transfer-formation of the triplet excited state of **1** (23 kcal/mol) (Eqn. (4.**4**)).

$$\text{bis(enone)} \bullet \text{cation} + i\text{-Pr}_2\text{NEt} \bullet \text{BF}_4 \longrightarrow \text{bis(enone)}^{\text{T}} + \text{cation} \bullet \text{BF}_4 + i\text{-Pr}_2\text{NEt} \qquad (4.4)$$

4.4.G – Anionic intermediate conformations. The radical anion intermediate, **IMa**, is computed to be slightly more stable than the bis(enone) radical anion, by 0.8 kcal/mol for the TMA counterion, and 0.4 kcal/mol for Li(acetonitrile)$_2$. Formation of the five-membered ring reduces the number of available conformations, relative to the bis(enone) reactant. The intermediate contains 5 low-barrier torsions (Figure 4.14). There are 2 aromatic sp^2-carbonyl sp^2 single bonds (2-fold), 1 olefinic sp^2- Csp^3 single bond (3-fold), 1 radical sp^2-carbonyl sp^2 single bond (2-fold), 1 radical sp^2-Csp^3 single bond (3-fold) and the restricted motion of the cyclopentyl ring with envelop and twist conformations. This suggests $2^3x3^2x4=288$ conformations.

Low-lying conformations were generated via a QM conformational search. To avoid generating and minimizing all 288 possible conformations, two smaller torsion tree searches were performed on the radical and anion halves (sans the 5-membered ring) respectively (Figure 4.15). Utilizing the excluded volumes of these structures, a portion of the 288 torsional combinations are removed. Two torsions were scanned in 15 degree increments for the radical half and a single torsion was scanned, also in 15 degree increments, for the anion half. It was found that 5 minima existed on the two-dimensional potential energy surface of the radical half and 3 for the anion half resulting in 15 possible combinations of structures. The 15 structures were constructed and the nonsensical and duplicate geometries removed from further consideration. The remaining low lying structures, which are consistent with the products of Table 4.1, are shown with the TMA counter ion in Figure 4.16 and their energies reported in Table 4.3.

The counter-ion, TMA, drastically changes the relative energies of the conformers allowing for the Diels-Alder pro-**2f** conformers to be energetically competitive with the pro-**2a** conformers.

4.4.H – Rotational barriers for anionic intermediate. In contrast to the radical anion of **1**, here rotation to a precursor of the diastereodefect product involves rotation about a radical sp^2-Csp^3 single bond (Figure 4.14). Barriers less than 5 kcal/mol are found and facile collapse to product is observed.

4.4.I – Triplet and broken symmetry intermediates. As with the bis(enone) radical anion, the intermediate can undergo back electron transfer to *i*-Pr_2NEt^+ (Eqn. (4.**4**)). Back electron transfer forming a triplet intermediate is 9.1 kcal/mol and 5.3 kcal/mol downhill for TMA and Li(acetonitrile)$_2$ counterions, respectively. Energetic data for the triplet states are collected in Table 4.4 and a spin-density plot of **IMc** is provided in Figure 4.17. Alternatively, back electron transfer could form a singlet diradical intermediate (Figure 4.18), broken symmetry energies for the set of intermediates (at the triplet geometries) are provided in Table 4.5. Formation of the broken symmetry intermediate is 10 kcal/mol downhill for both the TMA and Li(acetonitrile)$_2$ counterions. The lowest triplet state of the product is 64 kcal/mol uphill. Triplet intermediates could undergo intersystem crossing to the singlet diradical states and then collapse to product. In general, the broken symmetry diradical species collapse to product without barrier.

4.4.J – Transition state from intermediate to product. Given the short non-bond distance between carbons α to the 5-membered ring in the intermediate structures, barriers to formation of radical anion product (Figure 4.19) are likely small. The barrier to formation of the radical anion of **2a** is computed to be 1.4 kcal/mol and 0.1 kcal/mol for the TMA and Li(acetonitrile)$_2$ counterions, respectively (Figure 4.20).

4.4.K – Radical anion product. Closure of the second C-C bond to form **2a** (or **2b**) further reduces the conformational space. There are 2 aromatic sp^2-carbonyl sp^2 single bonds (2-fold), 2 Csp^3-carbonyl sp^2 single bonds (3-fold), and a constrained cyclopentyl ring for a total of 36 possible conformations for **2a**. Alternative products also considered include the radical anion of **2b** and the Diels Alder product, **2f**. The structures for these three products are collected into Figure 4.21 and their energetics provided in Table 4.6. The array of products were collected as a result of MM ring closure computations on the reactants and intermediates as well as QM ring closure and broken symmetry computations on the intermediates. It was found that formation of **2a** is 7.3 and 5.3 kcal/mol downhill from the anionic intermediate with the TMA counterion and Li(acetonitrile)$_2$ counterions, respectively

4.4.L – Neutral product. Back electron transfer from the anionic products to *i*-Pr_2NEt^+ leads to formation of neutral products. Back electron transfer is 49.1 kcal/mol and 47.7 kcal/mol downhill for TMA and Li(acetonitrile)$_4$ coordinated **2a**, respectively. Relative energies of **2a**, **2b**, and **2f** are collected in Table 4.7. The diastereoisomers **2a** and **2b** are nearly isoenergetic, the Diels-Alder regioisomer **2f**, is 3 kcal/mol higher in energy. The free energies of each **2a**, **2b**, and **2f** were also calculated and are reported in Table 4.8.

4.4.M – Reduction of neutral products by Ru(I). As the neutral products build up in solution they become susceptible to reduction by Ru(I). For TMA and Li(acetonitrile)$_2$ counterions, reduction of **2a** is 0.2 kcal/mol and 1.6 kcal/mol downhill, respectively.

4.5 – Discussion

This discussion of the most probable pathway(s) will closely follow the reaction coordinate, Figure 4.22.

Reduction of **1** by Ru(I) is viable, though dependent upon the nature of the counter-ion. The lithium counterion and aromatic rings were shown experimentally to be necessary. Without knowledge of the kinetics, we can suggest that the lithium counterion provides the largest overpotential, hence the greatest likelihood for reduction occurring. Further, delocalization provided by π-stacking of the aromatic rings favors reduction, relative to methyl substituents, as well. Once formed, the radical anion bis(enone) (**1**), can either rotate introducing a diastereoerror or close to the intermediate. Rotation is not competitive with ring closure to the nearly thermoneutral 5-membered ring intermediate (**IM**) (>10 kcal/mol difference). Once the 5-membered ring intermediate is formed, the barrier for diastereoerror inducing rotation is lower than the barrier for reversion to reactant allowing for interconversion to various product differentiating intermediates. Additionally, due to a very low intermediate to product barrier and the very modest exothermicity, intermediate complexes and products are likely in a rapidly maintained equilibrium. This suggests that relative product stability controls the product distribution, which is in turn controlled by solvation and the nature of the counterion. Further, since the neutral products are energetically competent to be reduced by the photocatalyst they can be reintroduced into the interconversion dynamic equilibrium.

An alternative pathway to product formation is formation of singlet diradical intermediates by way of back electron transfer to *i*-Pr_2NEt^+. Once formed the intermediates are computed to collapse to product without barrier. Diastereoselectivity in this case would originate from the relatively fast rotational interconversion of isoenergetic intermediates (~ps time scale) occurring before the charge transfer from intermediate species to the amine proceeds (~ns time scale). It still holds, however, that as the neutral product is being formed it can be reintroduced into the interconversion dynamic equilibrium.

4.6 – Conclusions

Alternative reaction pathways for the $[Ru(bpy)_3]^{2+}$* photocatalyzed, diastereocontrolled [2+2] cycloaddition of bis(enone)s have been computed. Two energetically competent pathways involving a family of reactive intermediates have been found. Lithium ion is found to facilitate initial bis(enone) reduction, relative to alternative TMA and sodium cations. Diastereocontrol is suggested to accrue through thermodynamic product stability rather than by way of a diastereoselective transition state.

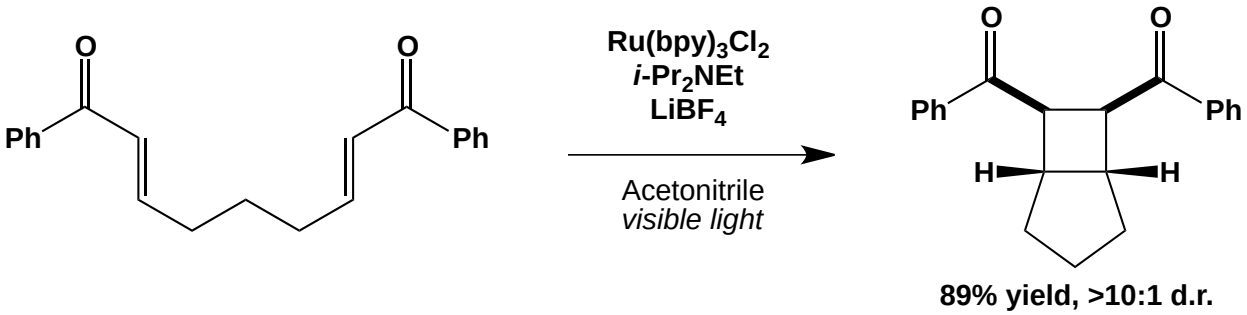


Figure 4.1. Yoon photocatalysis

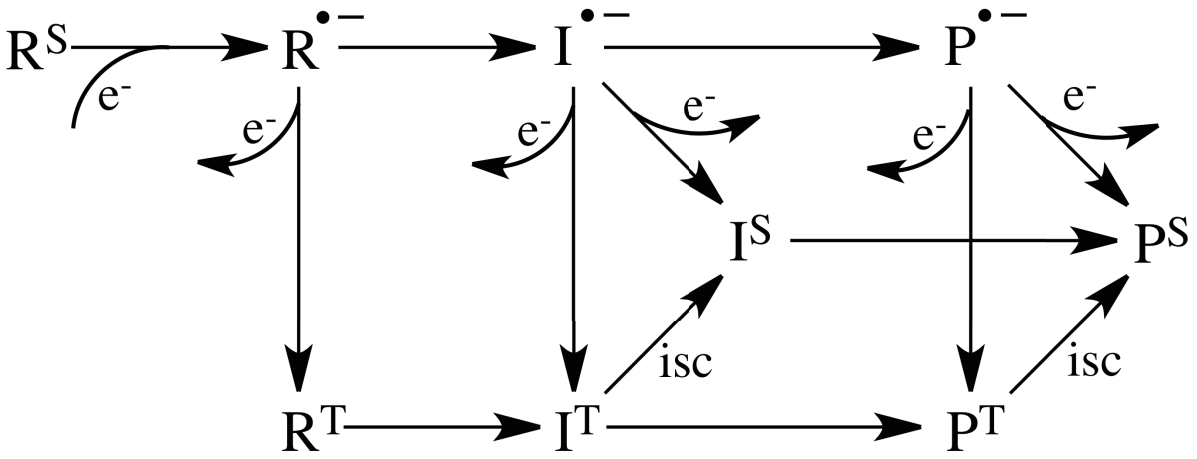


Figure 4.2. Possible reaction pathways. R, I and P represent the reactant, intermediate, and product, respectively.

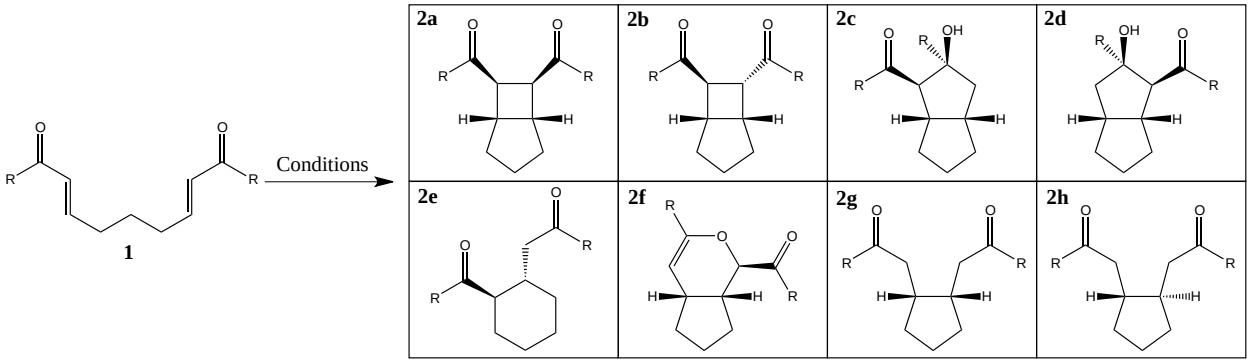
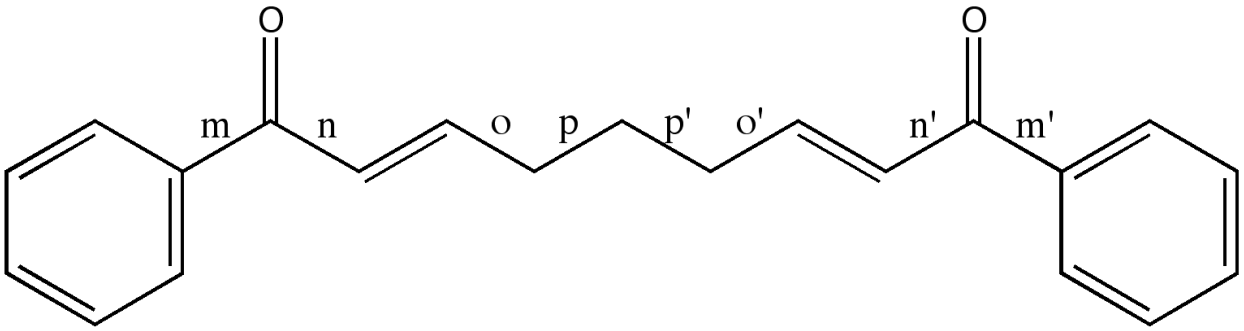


Figure 4.3. Bis(enone) and observed products

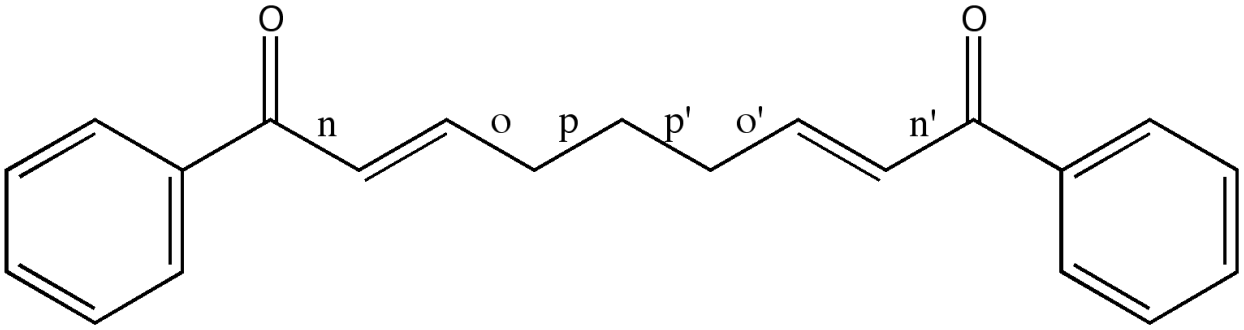
Table 4.1. Reactions of bis(enones) with varying conditions

Reaction Type	R	Conditions	Products (%)
A. Cathodic Reduction[6]	Ph	-0.90 V vs. SCE, 0.1 M $LiClO_4$, MeCN	**2a**(17), **2b**(20), **2c**(10), **2f**(8), **2g**(14), **2h**(7)
B. Chemically Induced[7]	Ph	70 mol% Chrysene, 0.086 M THF, -78 ºC	**2a**(31.5), **2d**(20.1), **2f**(9.2), **2g**(2.9)
	Np	70 mol% Chrysene, 0.086 M THF, -78 ºC	**2a**(61.7), **2d**(18.0), **2g**(10.0)
C. Metal Catalyzed[8]	Ph	10 mol% $Co(dpm)_2$, 4 eq. $PhMeSH_2$, 0.45 M DCE, 50 ºC	**2a**(72), **2e**(11)
D. $[Ru(bpy)_3]$ Photocatalysis[5]	Ph	5 mol% $Ru(bpy)_3Cl_2$, 2 eq. $LiBF_4$, 2 eq. DIPEA, 0.1 M MeCN, Sunlight	**2a**(81), **2b**(8)
	Me	5 mol% $Ru(bpy)_3Cl_2$, 2 eq. $LiBF_4$, 2 eq. DIPEA, 0.1 M MeCN, Sunlight	None
E. Eosin Y Photocatalysis[9]	Ph	0.5 mol% Eosin Y, 2 eq. LiBr, DIPEA, MeCN, 530 nm light	**2b**(96)
	Ph	2.5 mol% Eosin Y, 20 mol %thiourea, 1.1 eq. Hantzsch ester, CH_2Cl_2, light	**2h**(93)



aromatic sp^2-carbonyl sp^2 single bond: m, m'
olefinic sp^2-carbonyl sp^2 single bond: n, n'
olefinic sp^2-sp^3 single bond: o, o'
sp^3-sp^3 bond: p, p'

Figure 4.4. Rotatable bonds of the neutral bis(enone) reactant.



Torsion-tree angles searched (degrees)
n, n': 180, 0
o, o': 180, 120, 60, 0, -60, -120
p, p': 180, 60, -60

Figure 4.5. Bonds rotated (and their angles) during torsions tree scan

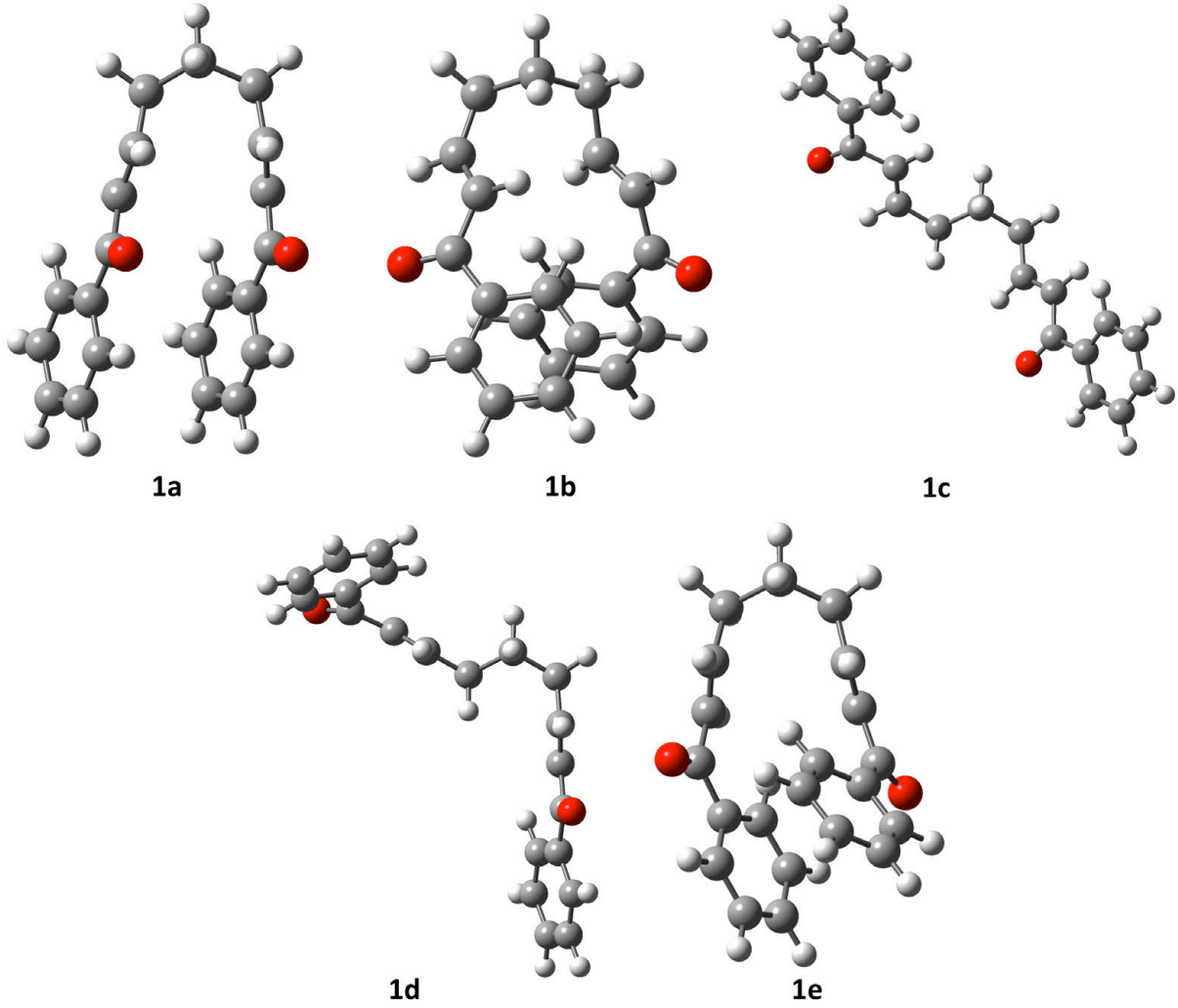


Figure 4.6. Neutral reactant Bis(enone) APFD geometries from APT results

Table 4.2. Relative energies of APFD neutral reactants

	Relative Energy (kcal/mol)	
Structure	Vacuum	PCM Solvent (Acetonitrile)
1a	0	0
1b	2.80	3.54
1c	6.82	7.09
1d	5.95	5.79
1e	0.15	1.16

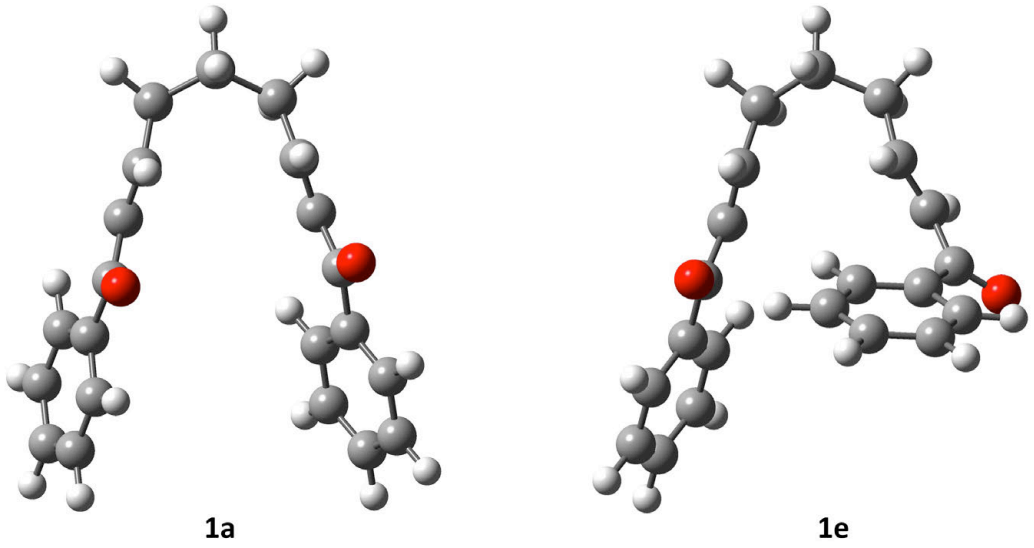


Figure 4.7. Structures 1a and 1e evaluated by B3LYP. Lack of dispersion results in a loss of π-stacking interactions.

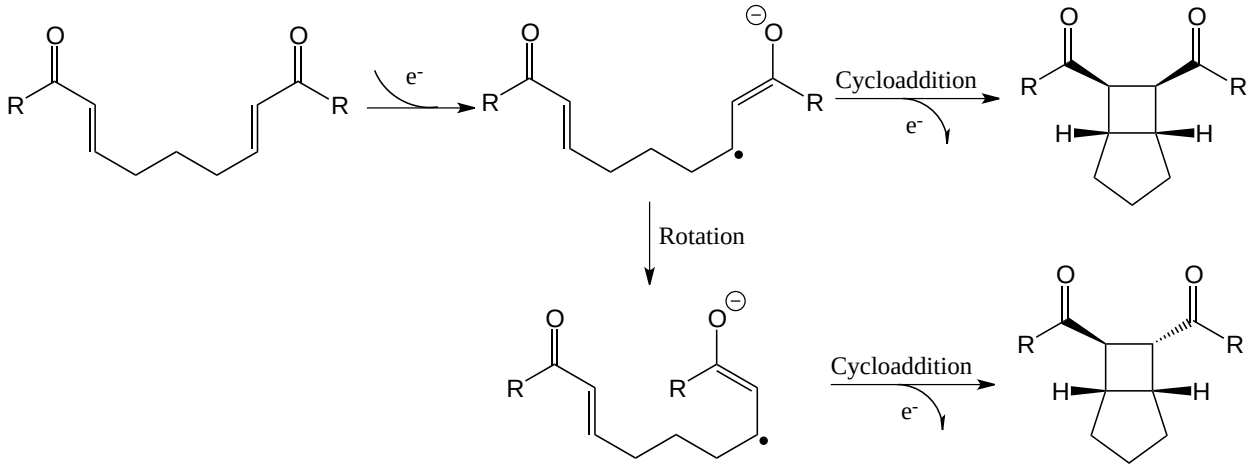


Figure 4.8. Rotation of bond in radical anion reactant to form two form the trans [2+2] cycloadduct.

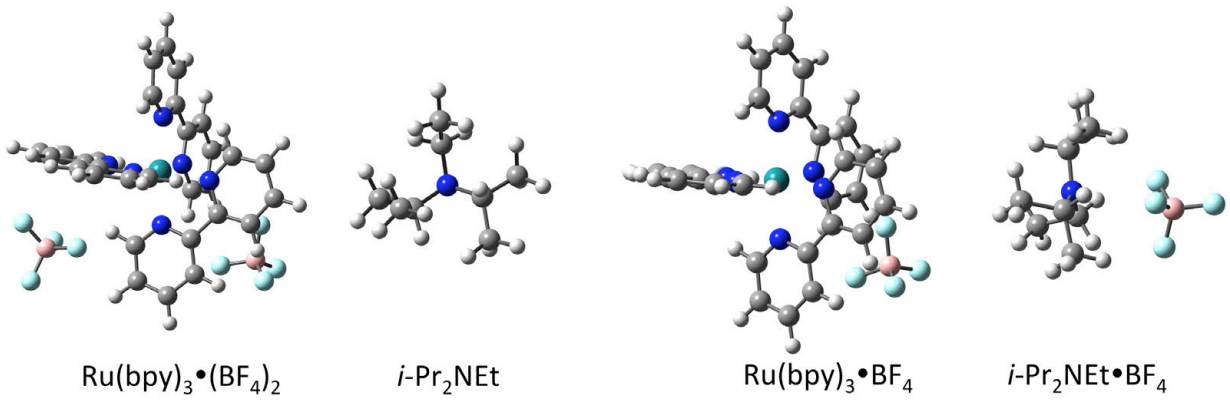


Figure 4.9. Ruthenium reduction products and reactants.

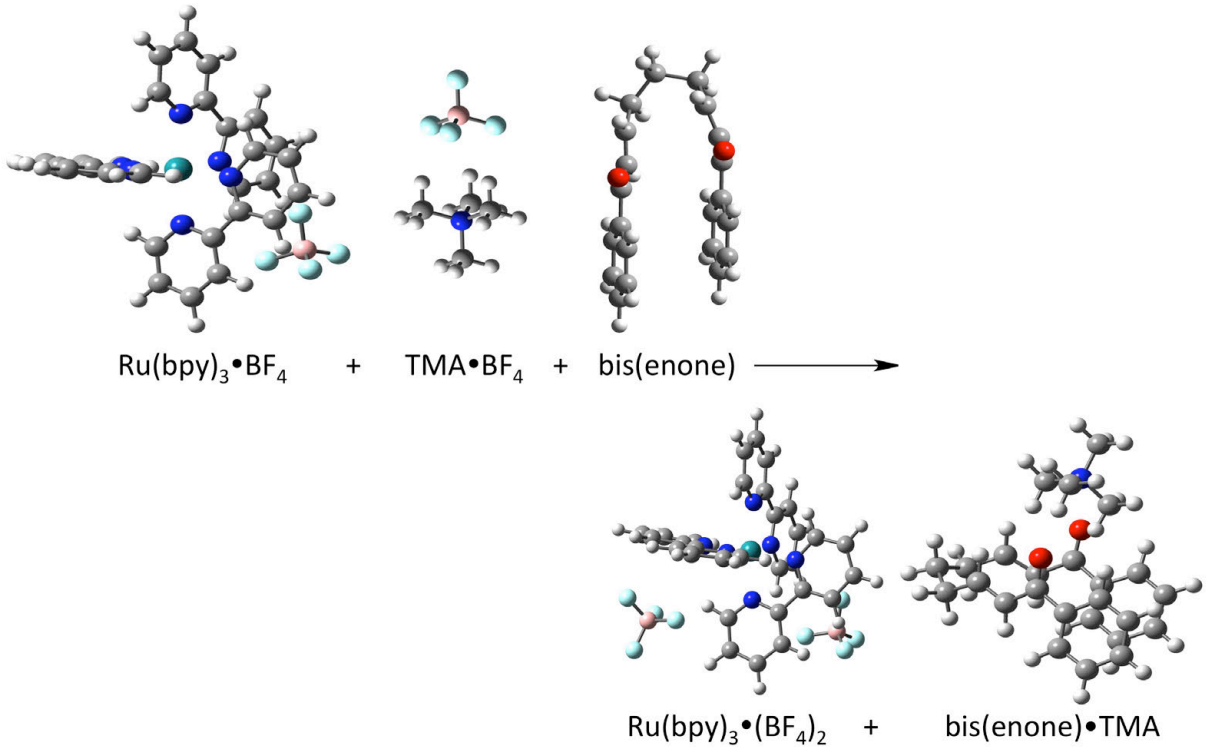


Figure 4.10. Reduction of bis(enone) via Ruthenium

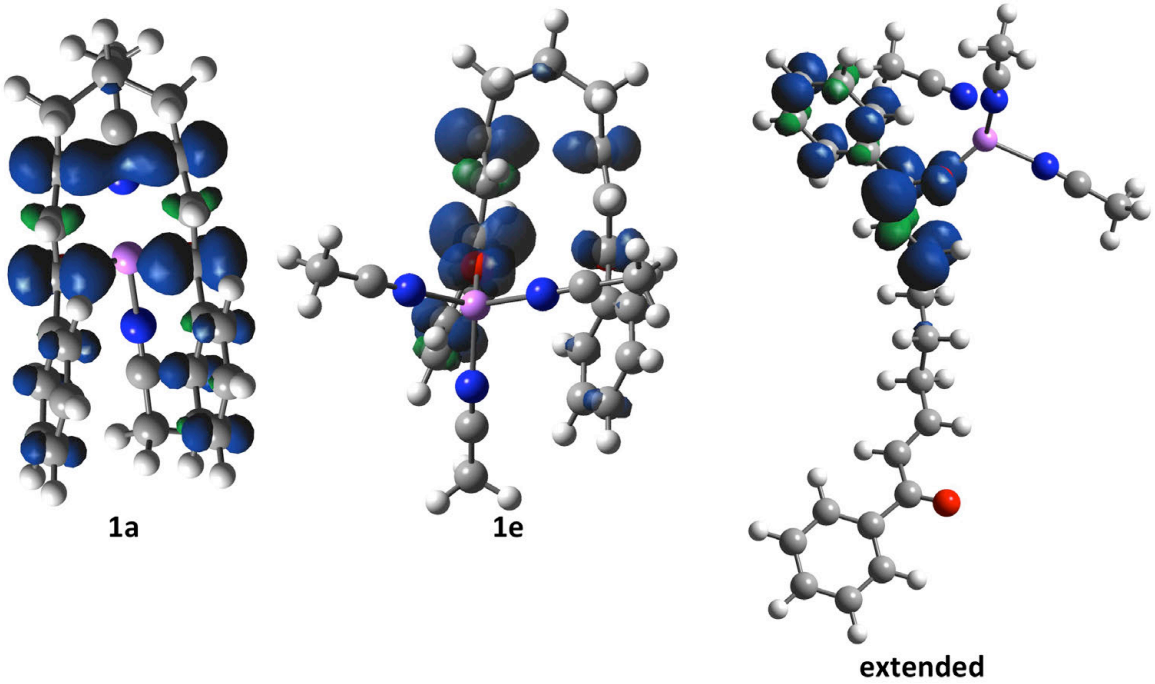


Figure 4.11. Electron spin density plots of radical anion reactant coordinated to lithium solvated by acetonitrile

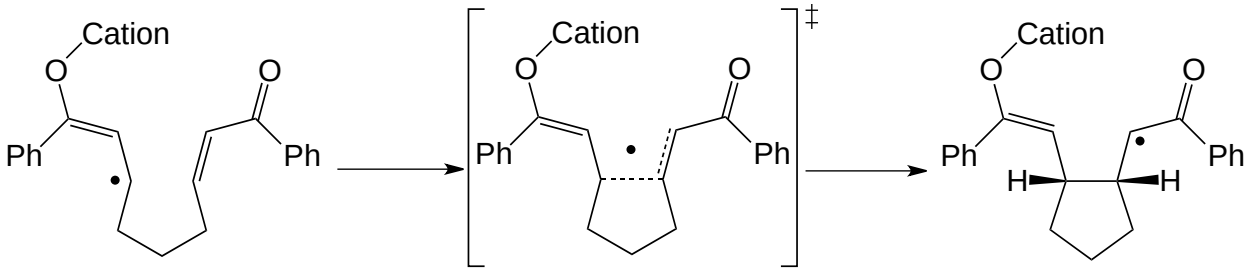


Figure 4.12. Formation of the 5-membered ring radical species.

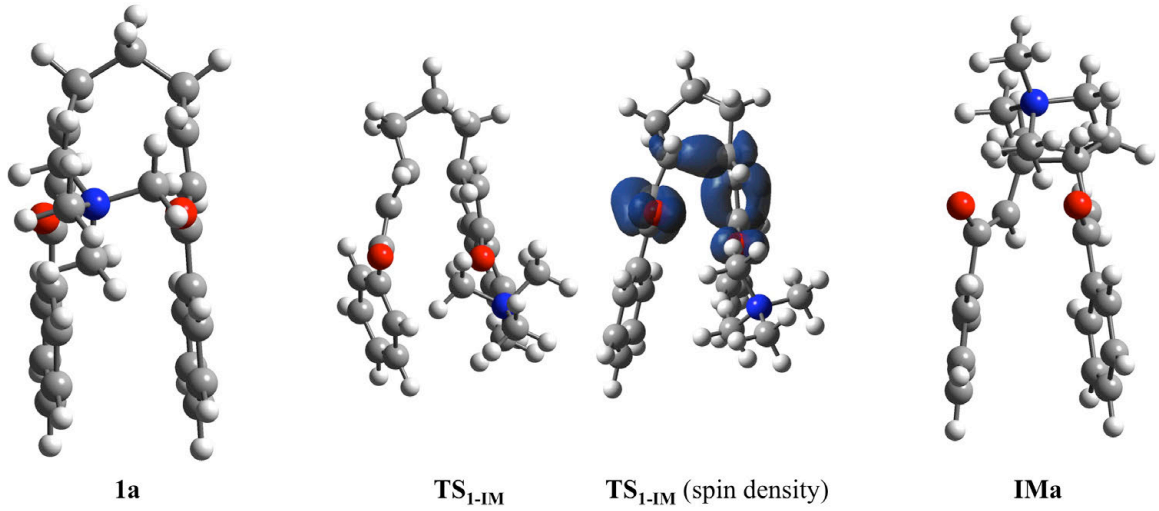
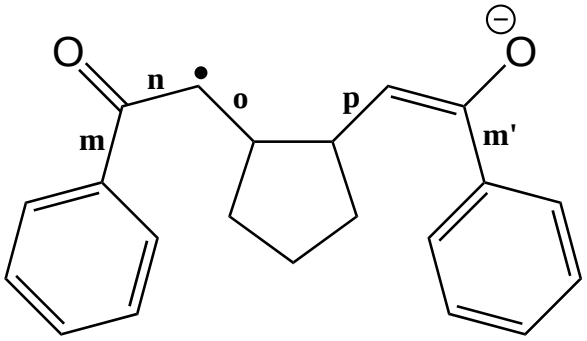


Figure 4.13. The parallel conformation of the reactant, 1a, moving through the 5-membered ring transition state, $IM_{1\text{-}IM}$, to reach the 5-membered ring intermediate, IMa.



Aromatic sp^2-carbonyl sp^2 single bonds: **m, m'**

Olefinic sp^2- Csp^3 single bond: **p**

Radical sp^2-carbonyl sp^2 single bond: **n**

Radical sp^2-Csp^3 single bond: **o**

Figure 4.14. Distribution of bond types in the radical anion intermediate

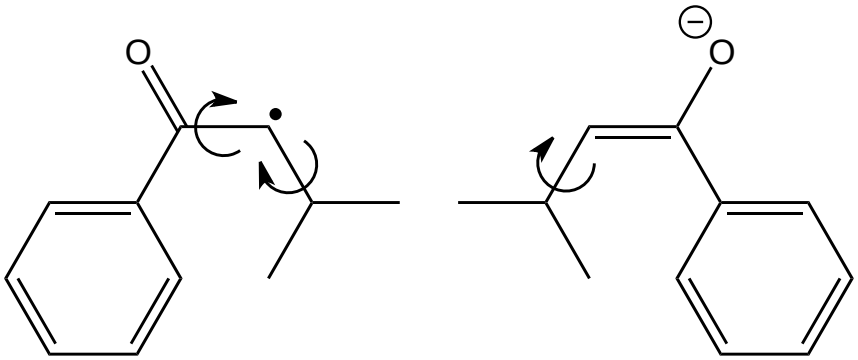


Figure 4.15. Rotated bonds for each torsion tree search.

Table 4.3. Low lying anionic intermediates

	Relative Energy (kcal/mol)		
Name	Solvated	Solvated+TMA	Predicted product
IMa	0	0	**2a**
IMb	4.9	-0.18	**2f**
IMc	4.9	1.8	**2f**
IMd	12	4.2	**2a**
IMe	9.2	5.0	**2a**
IMf	15	3.4	**2f**
IMg	4.3	-0.52	**2f**
IMh	2.8	-1.1	**2a**
IMi	2.3	0.55	**2a**

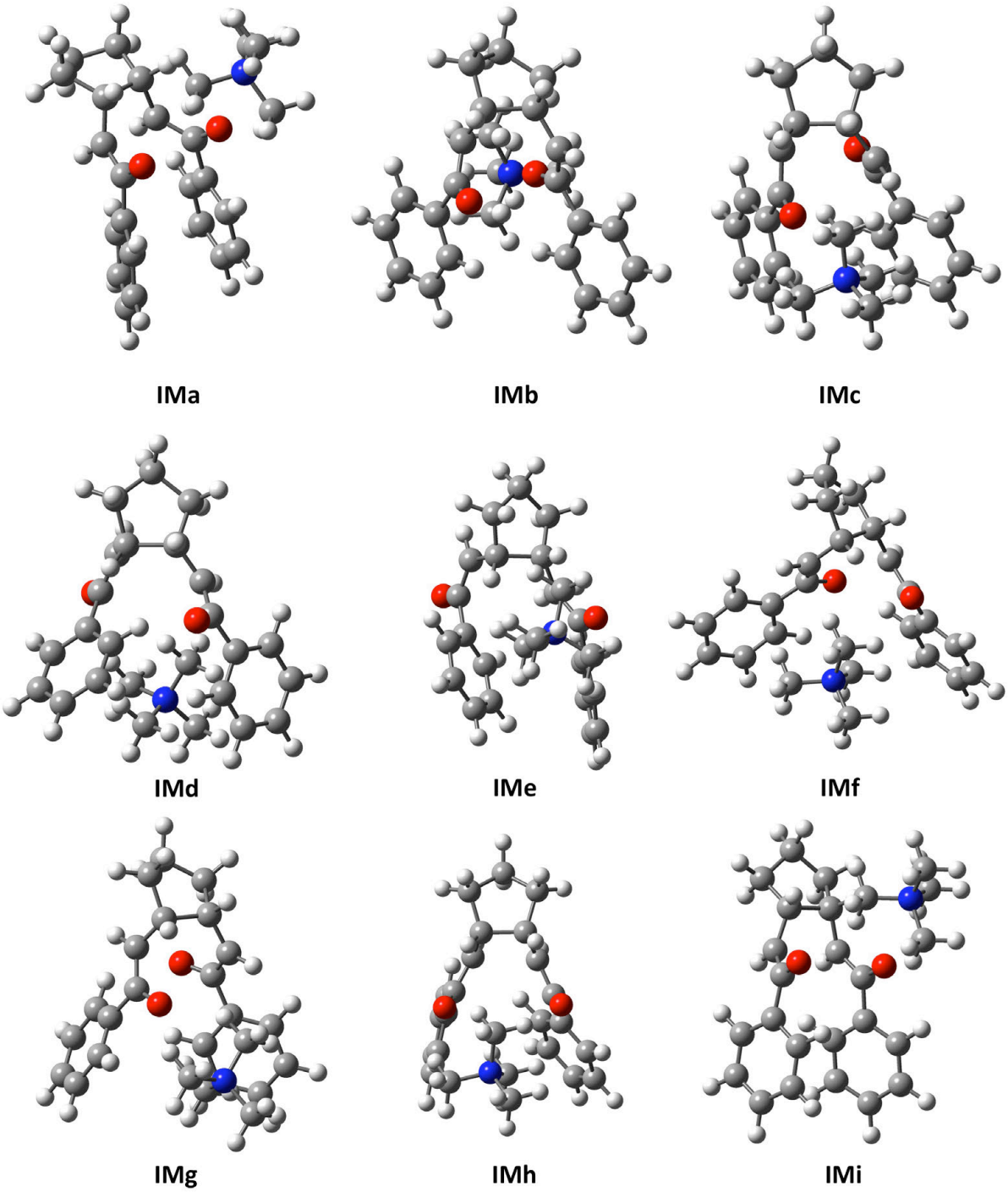


Figure 4.16. Intermediates obtained from conformational search.

Table 4.4. Relative energies of triplet state intermediates.

	Relative Energy (kcal/mol)		
Name	Vacuum	Solvated	Predicted product
IMa	0	0	**2a**
IMb	-3.0	0.21	**2f**
IMc	-2.5	-0.12	**2f**
IMd	2.4	3.2	**2a**
IMe	-0.91	0.51	**2a**
IMf	0.37	3.2	**2f**
IMg	-3.7	0.23	**2f**
IMh	0.26	0.54	**2a**
IMi	-0.97	-0.69	**2a**

Table 4.5. Energies of broken symmetry diradical intermediates relative to their respective triplet state intermediates

	Relative Energy (kcal/mol)		
Name	Vacuum	Solvated	Predicted product
IMa	-0.7	-0.88	**2a**
IMb	0.83	0.71	**2f**
IMc	0.08	0.61	**2f**
IMd	0.34	0.04	**2a**
IMe	0.15	-0.24	**2a**
IMf	-0.14	-0.16	**2f**
IMg	0.80	0.72	**2f**
IMh	0.91	0.50	**2a**
IMi	-0.07	-0.05	**2a**

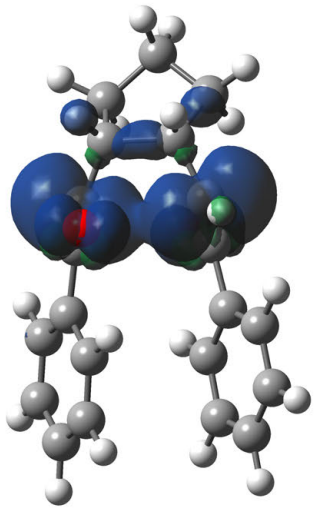


Figure 4.17. Bis(enone) neutral triplet intermediate IMc electron spin density.

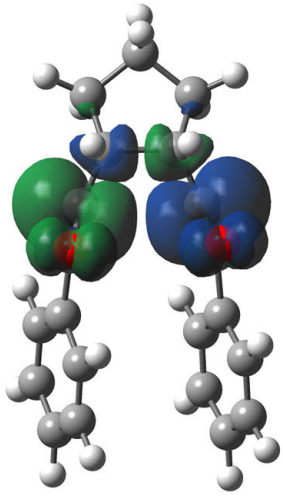


Figure 4.18. Bis(enone) diradical intermediate, IMi

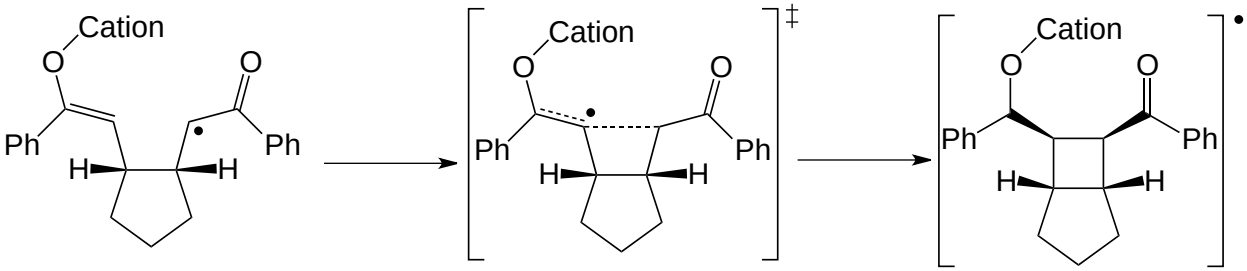


Figure 4.19. Formation of the radical product

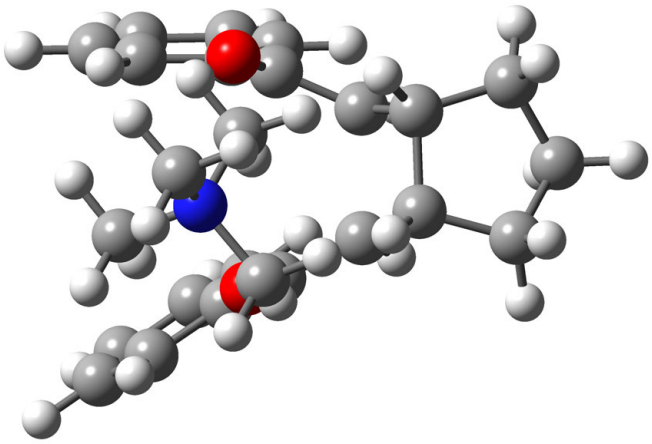


Figure 4.20. Second transition state of bis(enone)•TMA to form the 4-membered ring.

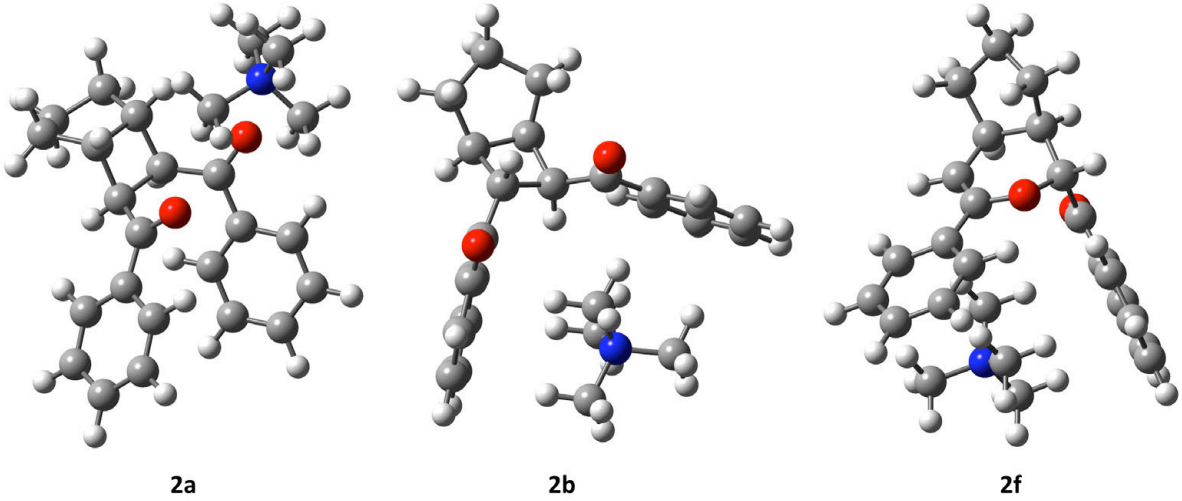


Figure 4.21. Cross-section of bisenone radical anion products coordinated to TMA cation.

Table 4.6. Relative energies of products 2a, 2b, and 2f with their respective counter ions.

	Relative Energy (kcal/mol)	
	$Li(MeCN)_4$	TMA
2a	0	0
2b	3.1	1.7
2f	4.4	-0.9

Table 4.7. Relative energies of neutral products..

	Relative Energy (kcal/mol)	
	Vacuum	Solvated
2a	0	0
2b	-1.3	0.3
2f	0.9	3.5

Table 4.8. Relative free energies of neutral products

	Relative Energy (kcal/mol)	
	Vacuum	Solvated
2a	0	0
2b	-2.1	-1.3
2f	1.5	3.7

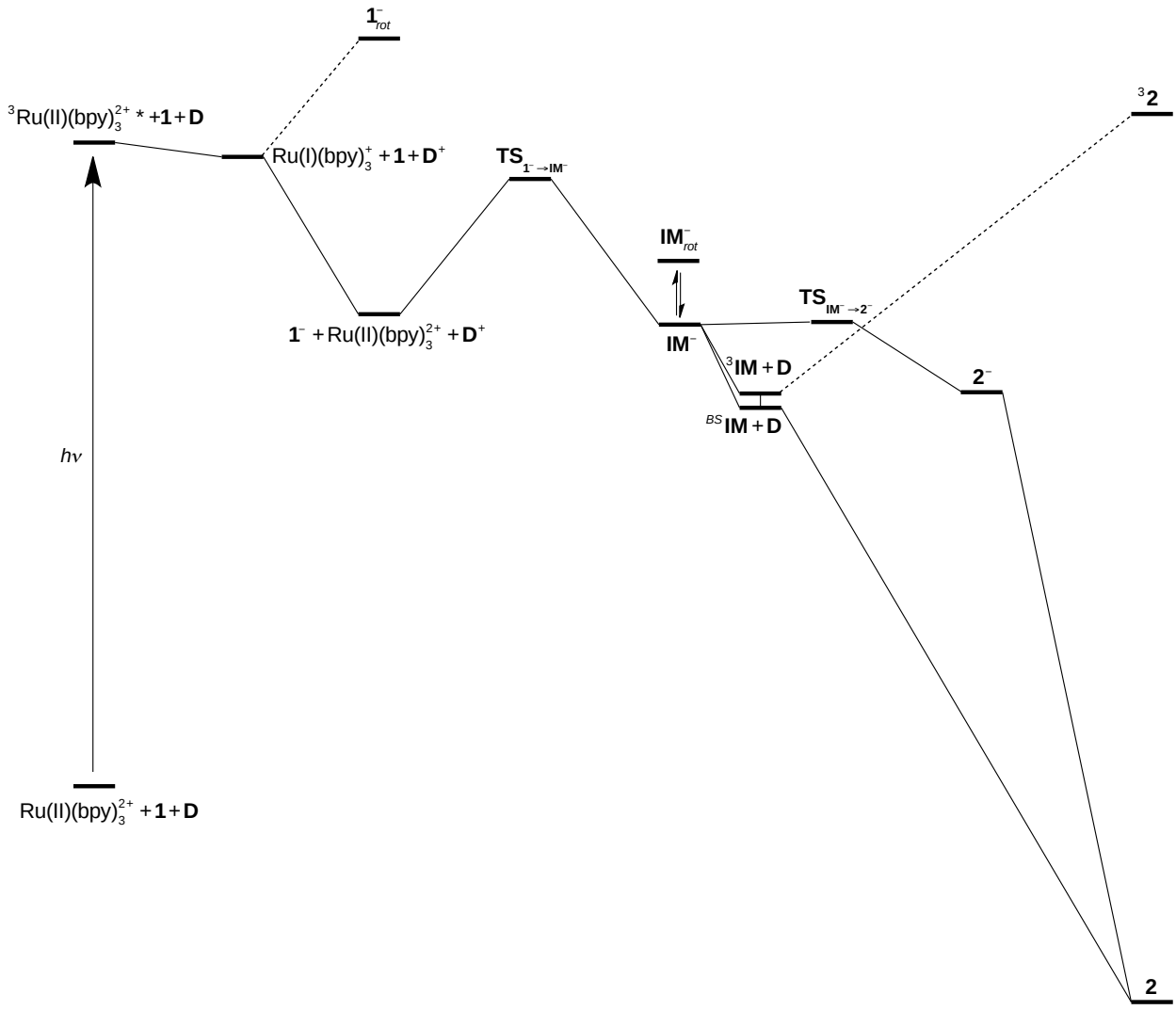


Figure 4.22. Reaction coordinate for bis(enone) coordinated to lithium acetonitrile. The counter ion is not shown.

CHAPTER 4 REFERENCES

(1) Yoon, T. P.; Ischay, M. A.; Du, J. *Nat. Chem.* **2010**, *2*, 527–532.

(2) Narayanam, J. M. R.; Stephenson, C. R. J. *Chem. Soc. Rev.* **2011**, *40*, 102–113.

(3) Tucker, J. W.; Stephenson, C. R. J. *J. Org. Chem.* **2012**, *77*, 1617–1622.

(4) Evans, R. W.; Zbieg, J. R.; Zhu, S.; Li, W.; MacMillan, D. W. C. *J. Am. Chem. Soc.* **2013**, *135*, 16074–16077.

(5) Ischay, M. A.; Anzovino, M. E.; Du, J.; Yoon, T. P. *J. Am. Chem. Soc.* **2008**, *130*, 12886–12887.

(6) Roh, Y.; Jang, H.-Y.; Lynch, V.; Bauld, N. L.; Krische, M. J. *Org. Lett.* **2002**, *4*, 611–613.

(7) Yang, J.; Felton, G. A. N.; Bauld, N. L.; Krische, M. J. *J. Am. Chem. Soc.* **2004**, *126*, 1634–1635.

(8) Baik, T.-G.; Luis, A. L.; Wang, L.-C.; Krische, M. J. *J. Am. Chem. Soc.* **2001**, *123*, 6716–6717.

(9) Neumann, M.; Zeitler, K. *Chem. - Eur. J.* **2013**, *19*, 6950–6955.

(10) Zhang, Q.; Li, Z.; Chen, B. *J. Mol. Struct. THEOCHEM* **2009**, *901*, 202–209.

(11) Yuan, J.; Zhang, Q.; Chen, B. *Comput. Theor. Chem.* **2012**, *996*, 110–116.

(12) Austin, A.; Petersson, G. A.; Frisch, M. J.; Dobek, F. J.; Scalmani, G.; Throssell, K. *J. Chem. Theory Comput.* **2012**, *8*, 4989–5007.

(13) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822–8824.

(14) Wang, Y.; Perdew, J. *Phys. Rev. B* **1991**, *44*, 13298–13307.

(15) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158–6170.

(16) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(17) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1997**, *78*, 1396–1396.

(18) Miertuš, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117–129.

(19) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999–3093.

(20) Becke, A. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(21) Yoon, S.; Kukura, P.; Stuart, C. M.; Mathies, R. A. *Mol. Phys.* **2006**, *104*, 1275–1282.

(22) *Handbook of Photochemistry, Third Edition*; 3 edition.; CRC, 2006.

CHAPTER 5: UTILIZING QUATERNIONS WITH POTENTIAL ENERGY SURFACE EDITING MOLECULAR DYNAMICS TO PERFORM MULTI-BODY CONFORMATIONAL STUDIES

5.1 – Foreword

This chapter was independently submitted as a requirement for the Colorado State University chemistry PhD program and proposes a non-trivial extension of surface editing molecular dynamics (SEMD). Minor modifications have been made to improve readability within this dissertation.

5.2 – Introduction

The ability to selectively sample low energy conformers reliably and efficiently for highly complex cases has been a challenge across a multitude of realms of computational chemistry[1–7]. This proposal discusses a new method that utilizes surface editing molecular dynamics (SEMD) and its application to highly complex multi-body systems. Specifically, I am proposing the utilization of quaternions and translational terms to aid in biasing intermolecular orientations. Their inclusion in SEMD will necessitate new and unique algorithm development. This new technology will extend the versatile and efficient SEMD method to highly complex multi-body systems such as protein/ligand docking or the polymer melt.

5.3 – Statement of Problem

Within the conformational searching and sampling communities a large focus has been placed on the ability to quickly and efficiently predict the binding of ligands to active sites of

proteins and what effect this has on the overall structure of the protein in silico via molecular dynamics[7–15]. However, conformational searching via molecular dynamics simulations is unlikely to discover protein/ligand binding modes from a neutral starting geometry on any realistic time scales especially when compared to the efficiency afforded by grid-based methods[7,8,11–14]. This is largely due to the fact that as the degrees of freedom within the system increase, the number of possible configurations for the system to occupy increase exponentially. To compound the issue, the vast majority of these searches are spent in the unproductive high-energy areas of the PES due to the high temperatures necessary to overcome potential energy barriers. Lastly, molecular dynamics does not bias against previously found geometries. Therefore a large amount of time is spent rediscovering previously visited conformations.

A new method called potential energy surface editing molecular dynamics (SEMD) has been developed to overcome these three issues and has already been applied to single, highly flexible molecules to great effect[16]. SEMD has the innate ability to bias against previously discovered conformers in tandem with aiding the dynamics in overcoming energy barriers. These advantages allow the dynamics to be run at lower temperatures and thus preferentially samples lower energy conformations. It is the purpose of this proposal to extend SEMD to multi-body systems such as protein/ligand docking by implementing a robust algorithm to utilize rotational and translational degrees of freedom inherent in such systems.

5.4 – Goals & Objectives

A new method involving SEMD will be developed in order to effectively sample conformational space of various multi-body systems such as protein/ligand docking. To accomplish this, the rotational and translational degrees of freedom must be considered to

efficiently bias against previously discovered molecular orientations in a multi-body system. It is proposed that a quaternion coordinate system to describe rotational bias will sufficiently aid SEMD in sampling rotational degrees of freedom. In addition to the quaternion coordinate system, a Cartesian center-of-mass coordinate for each body will be used to bias the translational degrees of freedom within SEMD. A number of modifications to the original SEMD method will need to be made to handle biasing translation and rigid orientation.

Once these additions are made, it is the goal of this proposal to successfully and quickly sample low energy space of a number of well-known and challenging multi-body systems. Such test cases would include:

- β-Trypsin/benzamidine system[4]: Within this system, modified SEMD would dock the benzamidine ligand to the β-trypsin protein with the Amber ff99 force field[17] with TIP3P[18] for the water molecules. Due to the relative rigidity of benzamidine and β-trypsin, this system is well-studied and would serve as an ideal test case to optimize the additions discussed in this proposal. In order to determine the efficiency of the modified SEMD method, one would measure gradient evaluations, dynamics time, computation time, and number of minimizations taken to reach the various states in which the benzamidine ligand will bind into the binding sites of β-trypsin. Those metrics would then be compared to existing methods such as metadynamics[4] and unbiased molecular dynamics[5].

- High-throughput docking to identify flaviviral protease inhibitors[19]: High throughput docking of the same diversity set of molecules by Ekonomiuk et al.[19] from the ZINC

database[20] to the active site of flaviviral protease using the CHARMM force field[21,22] with the TIP3P model for water molecules would prove to be an excellent test of SEMD's ability to quickly find global minima. To assess the accuracy of the method, the number of favorable poses of the ligands discovered would be compared to the previous method's poses.

These two goals would display SEMD's efficiency perform multi-body conformational searching as well as its inherent ability to discover global minima of highly complex species and systems quickly and reliably.

5.5 – Background

5.5.A – Past Methods: A number of computational methods have been developed in recent history to attempt to tackle complex multi-body cases[5,7,23]. One approach is to virtually anneal the system in a molecular dynamics simulation[24]. In an annealed dynamics simulation, the system is equilibrated at high temperatures for a given amount of dynamics time then raising the temperature allows the system to overcome potential energy barriers more readily. The system is then cooled and the conformer is subjected to a full energy minimization to obtain the lowest energy geometry in the current potential energy well. One disadvantage of simulated annealing is that there is still no guarantee that the system will move into undiscovered conformational space and with large systems (i.e. protein/ligand docking) the cost of "randomly" searching becomes prohibitive.

To combat the difficulties inherent in simulated annealing, a sampling method known as metadynamics[25–28] biases against previously explored areas of the PES by adding small (0.1-0.3

kcal/mol) Gaussians to pre-determined degrees of freedom, known as collective variables (CVs), on the PES as the dynamics is performed (Figure 5.1). CVs can be drawn from any degree of freedom that is relevant to the quantity being measured (Cartesian space, torsion space, spherical space, hydrogen bond distances, etc.), and it is usually a difficult to obtain the optimal set. The ambiguity of CV choices is a common criticism of metadynamics and remains to this day highly system dependent[26]. However, even with this disadvantage, if the proper CVs can be chosen this unique method enables molecular dynamics to efficiently sample pertinent swaths of space to garner a greater understanding of the surface as a whole.

Metadynamics has been used in the past to perform flexible docking[29–32]. One of the first trials by Gervasio, Liao, and Parrinello[29] was performed on the previously mentioned *β*-trypsin/benzamidine system. They defined their only collective variables as the interatomic distance between the C_7 atom on the benzamidine and the C_γ atom of the *β*-trypsin aspartate 189. Starting from the Brookhaven Protein Databank crystal structure and simulating for 0.5 ns, the method has deposited 166 Gaussian with a height of 0.48 kcal/mol. By allowing the simulation to "push" the ligand out of the active site, they were able to obtain a binding free energy of $\Delta G_{binding}$ = -6.0 kcal/mol where the experimental binding energy is $\Delta G_{binding}$ = -6.5 kcal/mol.

For the relatively simple case of the *β*-trypsin/benzamidine system, the choice of CV is relatively easy. However for more complex cases, the choice becomes less clear. With metadynamics, there is not one silver bullet CV that will solve all problems. It is the goal of this proposal to establish a single set of degrees of freedom that will be applicable within SEMD for all multi-body systems.

5.5.B – Quaternions: Quaternions were invented by William Hamilton in 1843 as a means to overcome many of the disadvantages inherent in existing approaches to rigid body

rotations[33]. Before defining quaternions however, it is useful to first explore an alternative approach, first as a means to justify the use of quaternions and second as an informative tool to better understand the rotational degree of freedom as a whole. One means of performing multi-body rotations (other than quaternions) is Euler angle rotation[34].

Euler angle rotation, developed by Leonhard Euler in 1776, is a method to perform a rotation via a sequence of rotations around each x, y, and z-axes. Its popularity is mostly due to its ease of understanding, however its simplicity is greatly offset by its many disadvantages. Euler rotation's largest disadvantage is that the alignments of the rotational axes are modified by previous rotations due to their dependence upon one another. In the formalism XYZ Euler angles, Z is dependent upon X and Y, and Y is dependent upon X. For example, in Figure 5.2 the blue, green, and red circles represent rotation about the z, y, and x-axes, respectively. If a rotation of 30 degrees occurs about the x-axis, then the y-axis is rotated with it. Once the y-axis is rotated, any rotations about that axis are no longer consistent to the absolute y-axis.

This leads to one primary issue when considering Euler angle rotations for multi-body rotations within SEMD. The primary disadvantage with utilizing Euler angle rotations for any single-body rotations is that a degree of freedom is lost when the y-axis is rotated to align with the z-axis rotation. This is easily achieved by rotating 90 degrees about the y-axis. This is a well-known issue termed "gimbal lock" (Figure 5.3). When gimbal lock occurs, there is no means of rotating about the lost degree of freedom thus making Euler rotations a poor choice for even single-body rotations.

The quaternion representation has one advantage that better lends itself to metadynamics. The advantage is that it becomes trivial and "catch-free" to calculate the minimum angle between two orientations. In this representation there are four degrees of freedom. They are represented

by the constituents q_0, q_1, q_2, and q_3 and combined into a coordinate representation (Eqn. (5.**1**)).

$$\mathbf{q} = [q_0, q_1, q_2, q_3] \quad (5.1)$$

The constituents, or coordinate, define(s) a position on the "unit" hypersphere around the body being rotated. A quaternion is expressed in unit vector form as:

$$\mathbf{q} = q_0 + \boldsymbol{i}q_1 + \boldsymbol{j}q_2 + \boldsymbol{k}q_3 \quad (5.2)$$

where q_0, q_1, q_2, and q_3 are real numbers, and $\boldsymbol{i}$, $\boldsymbol{j}$, and $\boldsymbol{k}$ are imaginary numbers belonging to the system of equations:

$$\begin{gathered} \boldsymbol{i}^2 = \boldsymbol{j}^2 = \boldsymbol{k}^2 = -1 \\ \boldsymbol{ij} = \boldsymbol{k}, \ \ \boldsymbol{jk} = \boldsymbol{i}, \ \ \boldsymbol{ki} = \boldsymbol{j} \\ \boldsymbol{ji} = \boldsymbol{ik}, \ \ \boldsymbol{kj} = -\boldsymbol{i}, \ \ \boldsymbol{ik} = -\boldsymbol{j} \end{gathered} \quad (5.3)$$

It is this system of equations that give quaternions their power.

A recent paper by Charles Karney[35] goes into great detail about the exact means of applying quaternions to molecular modeling. There is one primary section in his paper that is relevant to this proposal in which he documents how to perform a least-squares fit given two poses of a molecule, however he states that for flexible molecules this would be a poor choice due to the non-continuous nature of quaternions with flexible bodies. If one can overcome this issue, the procedure to obtain a least-squares fit as described by Karney is attractive because the least-squares rotation obtained is a proper rotation and overlapping orientations are treated properly.

5.5.C – Surface Editing Molecular Dynamics: Similar to the metadynamics method, SEMD utilizes the unique shape of a Gaussian to modify the potential energy surface. The equation for the Gaussian penalty term and how it is used in the total energy term is:

$$V_{total} = V_{SEMD} + V_{conf} \quad (5.4)$$

$$V_{SEMD} = A \prod_{i=1}^{tors} e^{-\alpha(\phi_i - \phi_{i,0})^2} \quad (5.5)$$

where A is the energy at the maximum of the Gaussian, α is the width of the Gaussian, ϕ_i is the current torsion and $\phi_{i,0}$ is the previously discovered torsion. Unlike metadynamics however, the end goal is to quickly explore conformational space rather than map out the entirety of space. This is achieved in a multitude of algorithmic ways. First, in order to quickly move throughout the potential energy surface, the Gaussian is made much taller than the energy barriers (on the order of 12-20 kcal/mol depending on the system of torsions) and wider than the potential energy wells. The larger (properly parameterized) Gaussian will fill the energy well (Figure 5.4) and allow the dynamics to overcome potential energy barriers more readily. Second, a Gaussian is only added to minima found on the unaltered potential energy surface so as to reduce the computational load of storing and computing orders of magnitudes more Gaussians while moving through the vast PES of large systems. Third, the dynamics are only run for enough time to escape the immediate energy well. This allows for a more flexible and less computationally expensive means of determining when a new minimum has been reached. Lastly, the application of initial normal mode velocities from a minimum allows more guided dynamics toward undiscovered conformational space rather than a pseudo-random march through the PES.

SEMD has had great success in both finding the low energy population of conformers and also in discovering global minima with surprisingly little computational effort. The 262 lowest 3 kcal/mol conformers of cycloheptadecane (Figure 5.5) were discovered by SEMD in 3,622 full energy minimizations, a factor of two times faster than the current leading method[36], and 84 full energy minimizations to find the global minimum from the flat conformer. Additionally, the global minimum of the highly flexible polyethylene strand, $C_{39}H_{80}$, was

obtained in 4,410 minimizations where the leading method, *l*-LMOD[37], discovered it in multiple 10,000 minimization simulations. These benchmarks exemplify that SEMD is a method that is well suited to multi-body conformational searching problems.

5.6 – Proposed Research: Multi-body SEMD

It is the goal of SEMD to be able to perform conformational searching on complex multi-body systems. However, it is advisable to initially start with simple cases and increase in complexity so as to understand the nuances of each addition as they are introduced. This section proposes a limited number of test cases from the protein-ligand docking field that are initially very simple (conformationally strained ligand & only the active site of the protein modeled) and eventually build to be incredibly complex (flexible ligand & full protein modeled) with intermediate steps to slowly introduce complexity.

5.6.A – Utilizing SEMD for Docking Conformationally Strained Ligands to a Frozen Active Site of a Protein: The simplest case to consider is first docking a conformationally constrained ligand to an active site of a protein wherein only the ligand and the active site of the protein are being modeled. This reduces the complexity of the test case to the rotational and translational degrees of freedom of the single ligand.

Biasing against rotational and translational degrees of freedom with Gaussian functions is at first glance relatively easy to consider and is achieved by adding those terms to the total product of the penalty function:

$$V_{SEMD} = V_{torsions} * V_{rotation} * V_{translation} \quad (5.6)$$

With this new penalty term, discovered minima will be concertedly biased against in torsional, rotational, and translational degrees of freedom. If the ligand is sufficiently conformationally

constrained, the torsional term need not be included. Simplifying to a test case wherein rotational and translational motion are the dominant degrees of freedom allows for a greater understanding of the impact of this modified penalty term on the potential energy surface.

As previously established, rotational orientation is most easily described in terms of quaternions. Therefore designing a Gaussian to penalize rotational orientation in quaternion space is a specific goal, wherein the general functional form would be:

$$V_{rotation} = A_{rot} e^{-\alpha_{rot}\left(RMSD_Q\right)^2} \quad (5.7)$$

where $RMSD_Q$ is the rotational root mean squared difference (RMSD) of the current rotational orientation with a previous rotational orientation defined in quaternion space. A_{rot} and α_{rot} are the height and width of the Gaussian applied to rotation respectively. This term properly biases against previously discovered orientations and would behave as expected (two different orientations would give rise to a large RMSD and therefore a small $V_{rotation}$ term and no penalty on the PES). In addition to this function, a first derivative must be defined to allow the molecular dynamics to generate a velocity with a bias against the previously discovered rotations accounted for.

Translational motion must also be concertedly considered in the penalty term to "push" the ligand into new regions of the active sites. In the case of the rigid ligand docking to a frozen active site, a simple center of mass RMSD in Cartesian coordinates will properly bias against a previously discovered conformer/pose in a region of an active site:

$$V_{translation} = A_{trans} e^{-\alpha_{trans}(\Delta r)^2} \quad (5.8)$$

where Δr is the Euclidean distance between the current center of mass and a previous centers of mass in Cartesian space. A_{trans} and α_{trans} are, respectively, the height and width of the Gaussian applied to translation. Like the rotational term, the associated first derivative will be

generated. It should be noted that there is no product operator within these singular terms as was in the torsional term (Eqn. (5.**5**)). This is due to the fact that there will only ever be one coordinate to consider for rotation and translation in this simple case.

A problem will arise when attempting to evaluate the product (Eqn. (5.**6**)) as written and becomes evident after some rearrangement:

$$V_{SEMD} = A_{tors} \prod_{i=1}^{tors} e^{-\alpha_{tors}(\phi_i - \phi_{i,0})^2} * A_{rot} e^{-\alpha_{rot}(RMSD_Q)^2} * A_{trans} e^{-\alpha_{trans}(\Delta r)^2}$$

$$V_{SEMD} = [A_{tors} * A_{rot} * A_{trans}] \left[\prod_{i=1}^{tors} e^{-\alpha_{tors}(\phi_i - \phi_{i,0})^2} * e^{-\alpha_{rot}(RMSD_Q)^2} * e^{-\alpha_{trans}(\Delta r)^2} \right] \quad (5.9)$$

The issue occurs in Eqn. (5.**9**) when $A_{tors} * A_{rot} * A_{trans}$ is evaluated. Each Gaussian has a height parameterized to the typical barrier height associated with the types of forces each term will be biasing against. For example, A_{tors} will have a height equal to or greater than 3 kcal/mol, this is estimated from the 3 kcal/mol well depth of the sp^3-sp^3 carbon torsion potential curve. When the product is evaluated, the total height of the final Gaussian that is applied to the PES will be unrealistically large. It is thus proposed that a normalized Gaussian-weighted average of the scalers be evaluated instead:

$$A_{tot} = \frac{\sum_{i=1}^{n} A_i e^{\alpha_i (x_i - x_0)^2}}{\sum_{i=1}^{n} e^{\alpha_i (x_i - x_0)^2}} \quad (5.10)$$

and applied to the overall SEMD bias term:

$$V_{SEMD} = A_{tot}[V_{torsions} * V_{rotation} * V_{translation}] \quad (5.11)$$

where the previous scalers ($A_{tors} * A_{rot} * A_{trans}$) will be removed in favor of A_{tot}. This normalized Gaussian-weighted average of the scalers results in a scaling value dependent upon the mode of dynamic travel (Figure 5.6) as well as having a first derivative that is relatively simple to evaluate.

Another, simpler, approach would be to estimate the size of the well taking into consideration all pertinent degrees of freedom and, instead of a normalized Gaussian-weighted average of scalers, define a suitable value for *A* that would generally apply. This would be the most computationally efficient means as the exponential terms in Eqn. (5.**9**) would control the width of the overall Gaussian in each degree of freedom.

Finally, with the proper means to evaluate the function, the proper parameterization must be considered. This can be accomplished by taking into consideration the intermolecular forces present and tailoring the Gaussian to overcome the typical barrier as well as the typical "distance" of the barriers from the minima to determine A and alpha respectively. One means to evaluate these is to create a representative test case and track the non-bond terms over the course of a dynamics simulation. Once these preliminary parameters are generated, evaluating their efficiency versus parameters within the same order of magnitude via rigorous trials would be a reliable means to determine their efficacy.

With a properly defined means of evaluating the rotation and translation, the question of how to define when a conformer is in a "new" orientation presents itself. In order to properly define this metric one can evaluate the values of the quaternions and center of mass as a function of the dynamic time. Utilizing the results from this analysis, one could define a distance cutoff where a minimization will occur with a high probability of minimizing to a "new" conformer/pose. This will need to be coupled with existing criteria for minimization by implementing new logic to define the cutoffs into the dynamics code.

With these considerations made, utilizing SEMD to perform docking of a rigid ligand into a frozen active site of a protein is achieveable. A specific case to test against is the previously mentioned docking by Ekonomiuk et al.[19] of various ZINC structural database species

into the active site of flaviviral protease. In this case, one would begin with conformationally constrained species and initially keep the protease active site frozen to reduce system complexity. To assess the accuracy of the method, one would compare the energy of the favorable poses of each conformationally constrained species tested with the energy of the poses of the same species obtained by Ekonomiuk et al.

5.6.*B – Utilizing SEMD for Docking Flexible Ligands to a Frozen Active Site of a Protein*: In this case, all of the considerations noted previously should be applied, however it is important to note that the quaternion function of a flexible body (flexible ligands) is not continuous as the body distorts as discussed by Charles Karney[35]. This presents an issue for molecules as they distort to form other conformers by rendering the quaternion of the previous conformer non-applicable to the new conformer (Figure 5.7). It has been suggested that a rigid section of the molecule can be used to compare the orientation of a conformer[5]. However, this is a weak remedy at best and will fail if there are no rigid sections or multiple rigid sections that are separated by flexible portions of molecules.

For our purposes it is unnecessary to evaluate the $RMSD_Q$ of a specific conformer to any other conformer, as comparing a new conformer to a previously known conformer will result in a torsional penalty term of zero. Therefore no matter what the value of $RMSD_Q$, the SEMD potential energy term will be zero. The $RMSD_Q$ is then only necessary to calculate when a known conformer is in a new orientation. This can be achieved by calculating the RMSD of the current geometry's torsions to those of previously found conformers before attempting to calculate the RMSD of the rotation. If the RMSD of the torsions is below a predetermined cutoff value then the quaternion function should be continuous and can be calculated. The challenge here will be to ascertain exactly when the quaternion is non-continuous and applying that to the

cutoff. A trial test to observe the continuity of quaternion space as a function of torsional displacement should allow for enough insight to intelligently assign a cutoff value.

This consideration enables the docking of flexible species to an active site. Specifically, one should attempt to dock the more flexible species from the ZINC structural database to the active site of flaviviral protease as done by Ekonomiuk et al.[19]. To assess the efficacy of the method, one would compare the energy of the favorable poses of each flexible species tested to the same species' energies obtained by Ekonomiuk et al.

5.6.C – Utilizing SEMD for Docking Flexible Ligands to an Active Site of a Fully Simulated Protein: With all previous considerations applied, docking a flexible ligand to a fully simulated protein becomes possible but adds the complexity of a multi-body system. With this true multi-body system, the quaternion penalty term must include a product (similar to the torsion term within Eqn. (5.**5**)) to account for both bodies:

$$V_{rotation} = \prod_{i=1}^{bodies} e^{-\alpha_{rot}\left(RMSD_{Q,i}\right)^2} \qquad (5.12)$$

where the term, $RMSD_{Q,i}$, is evaluated for each body i. The product allows the entire penalty term to move to zero as a single body rotates to a new orientation. However, there is an inherent problem with this function: as the system as a whole rotates, the quaternion penalty function will not properly penalize the set of infinitely identical relative rotations since they will both have a set of different quaternion coordinates when compared to the previous orientation (Figure 5.8.A and 5.8.B). It then becomes necessary to create a rotational penalty term wherein the quaternion RMSD of each body are coupled, relative to one another. It is then proposed that a large focus will be placed on the development of this particular function to properly bias against identical relative rotational orientations.

A second issue closely related to the coupled rotation is the handling of degenerate center of mass cases. In this two-body case, the center of mass coordinate will not sufficiently describe the surface to penalize against every degenerate case (Figure 5.8.A and 5.8.C). The solution lies within a coupling of the center of mass coordinates of each body. A center of mass difference term would fix degeneracy with all translation cases:

$$V_{translation} = e^{-\alpha_{trans}\left(\Delta r_{cm} - \Delta r_{cm,0}\right)^2} \quad (5.13)$$

Where Δr_{cm} is the center of mass distance of the current orientation and $\Delta r_{cm,0}$ is the center of mass distance of a previously discovered orientation. When this translational term is applied concertedly with a newly developed means of biasing against degenerate rotational orientations, the total penalty term will sufficiently penalize all degenerate cases with both two-body translations and rotations.

To test these additions, the previously mentioned typsin-benzamidine system would serve as a means to verify the efficiency of multi-body SEMD in determining binding modes of multiple active sites. To determine the efficiency, one would measure gradient evaluations, dynamics time, computation time, and number of minimizations taken to reach the various states in which the benzamidine ligand is bound into the active sites of trypsin. These metrics would then be compared to metrics of leading methods previously discussed[5,19].

5.6.D – Computational Efficiency Considerations: There are a number of possible means to increase the efficiency of the calculations. To prevent evaluating a full penalty function, previous minima can be initially excluded based on their "nearness" to previous conformers in translational space. For the single-body case, this would greatly increase the efficiency since the comparison is done over a single coordinate. For multi-body cases, a gain in efficiency would not be as great but should still be anticipated. After this initial comparison, the torsional portion

of the penalty function can be evaluated before the quaternion to exclude orientations wherein the torsions do not match. Unfortunately, the quaternion (also being a single coordinate) cannot be evaluated before the torsions (potentially n coordinates) as the torsion space must first be checked to determine if quaternions space is continuous or not.

5.7 – Conclusions

The SEMD method coupled with the means for rotational and translational biasing will allow for the conformational searching of large and complex systems. The inclusion of quaternions and center of mass as the equivalent of collective variables presents unique challenges such as how to scale the product of Gaussians, how to handle the non-continuous quaternion space, and the seemingly-straightforward issue of parameterization. Many solutions as well as the means to obtain the solutions to these problems have been proposed and will aid in the extension of SEMD to multi-body systems.

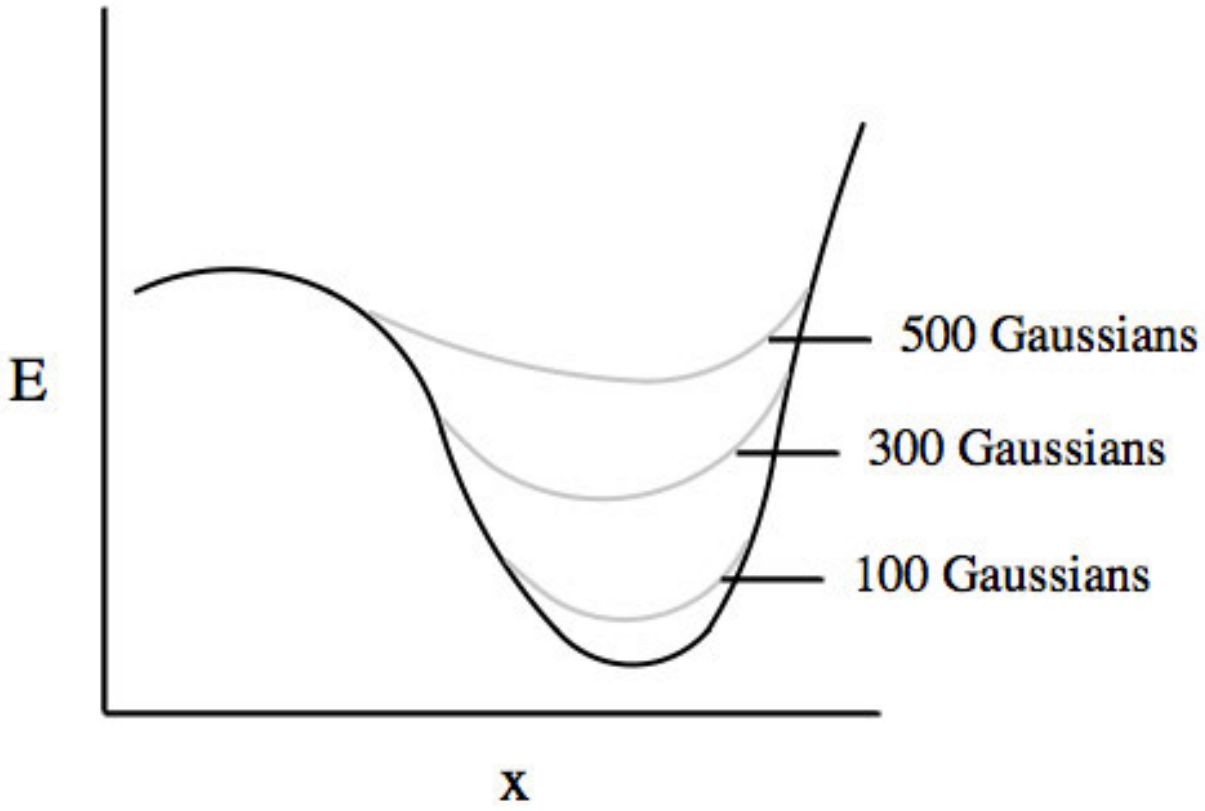


Figure 5.1. Addition of multiple Gaussians to a potential energy surface (dark line) via metadynamics. The resulting potential energy surface is shown as the gray lines as the number of Gaussian added to the surface increase.

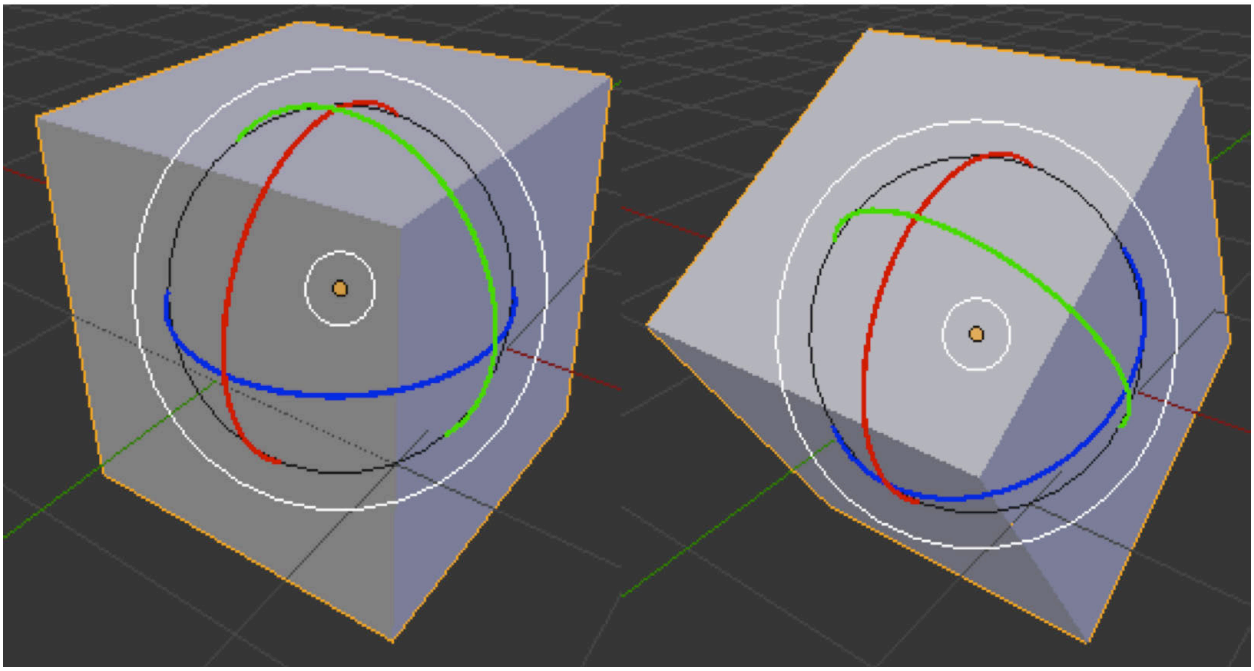


Figure 5.2. Unrotated cube (left) where each rotational axis (x=red, y=green, & z=blue circles) is aligned with their corresponding absolute axis. Cube (right) that has been rotated about the x-axis where the y and z rotational axes are now misaligned from their absolute axes.

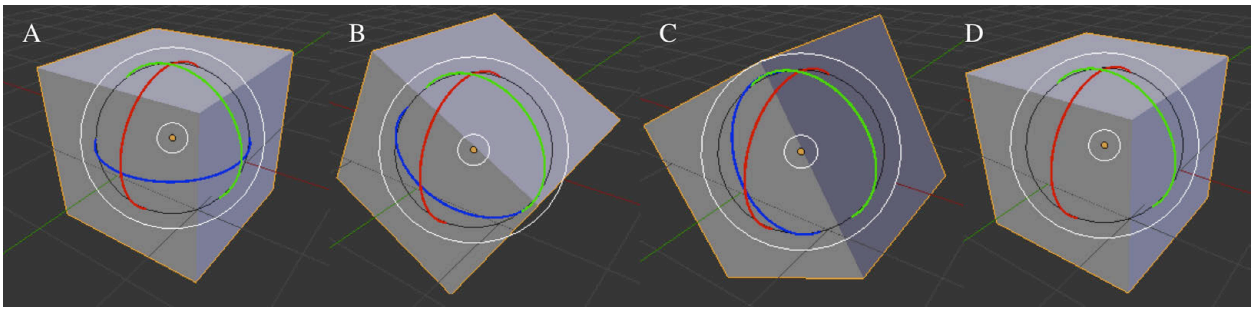


Figure 5.3. Rotation of a cube utilizing Euler angles to demonstrate gimbal lock. The unrotated cube (A) is rotated about the y-axis. Due to the rotational z-axis being dependent upon the rotational y-axis, the z-axis is rotated as well. The rotation occurs in 30 degree increments where (B) has been rotated 30 degrees, (C) rotated a total of 60 degrees, and finally (D) having been rotated 90 degrees from the starting position. Gimbal lock has occurred in D, resulting in the loss of a rotational degree of freedom (the rotational z-axis).

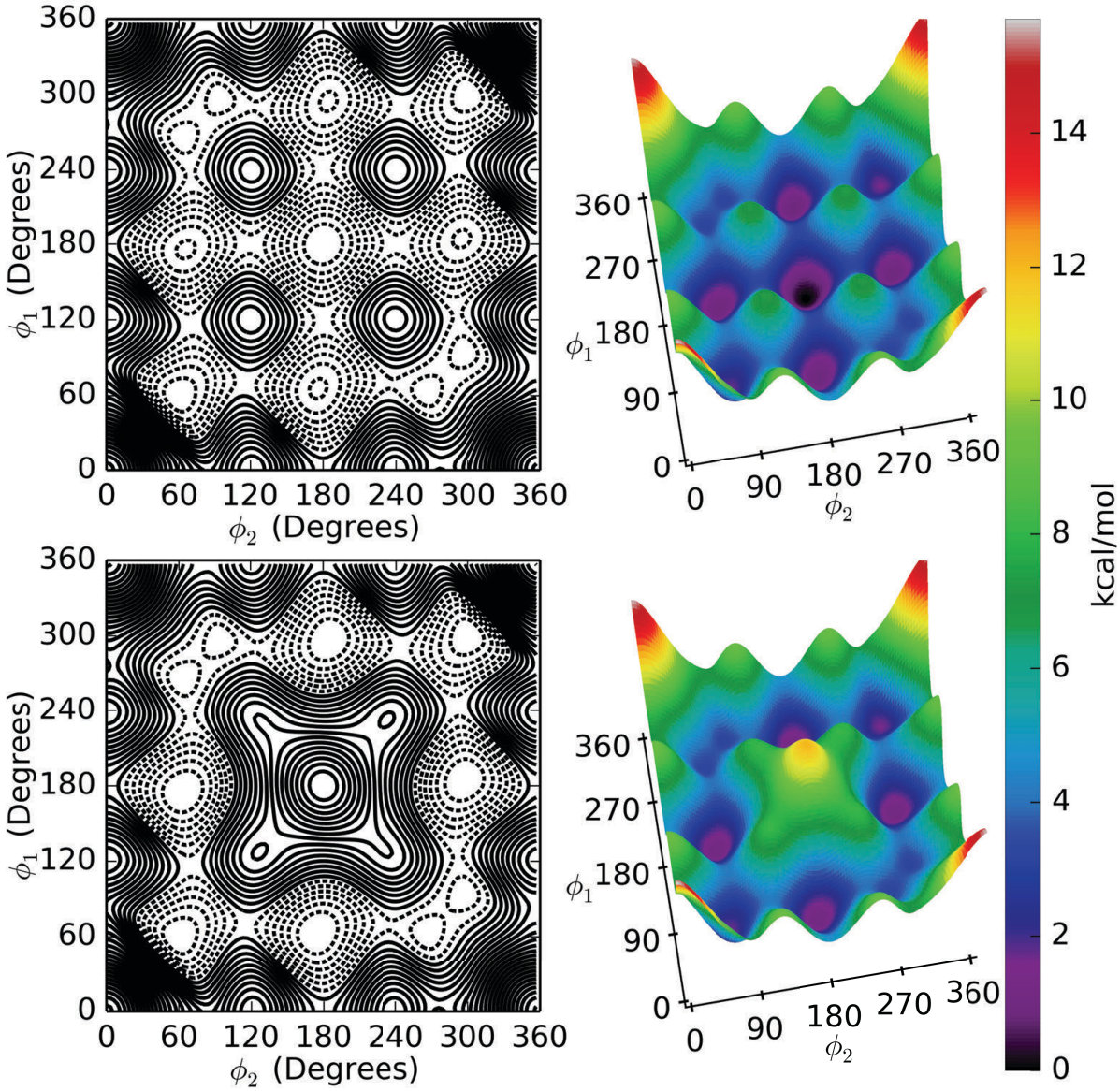


Figure 5.4. The top contour plot (left) and three-dimensional surface (right) represent the unmodified pentane torsion potential energy surface where ϕ_1 and ϕ_2 represent the two C-C-C-C torsions. The bottom contour plot (left) and three-dimensional surface (right) illustrate the impact of a Gaussian centered at the ϕ_1=180° and ϕ_2=180° minimum; parameterized with A=12 and α=6. The result is a maximum at the previous minimum without significant modification to nearby potential minima.

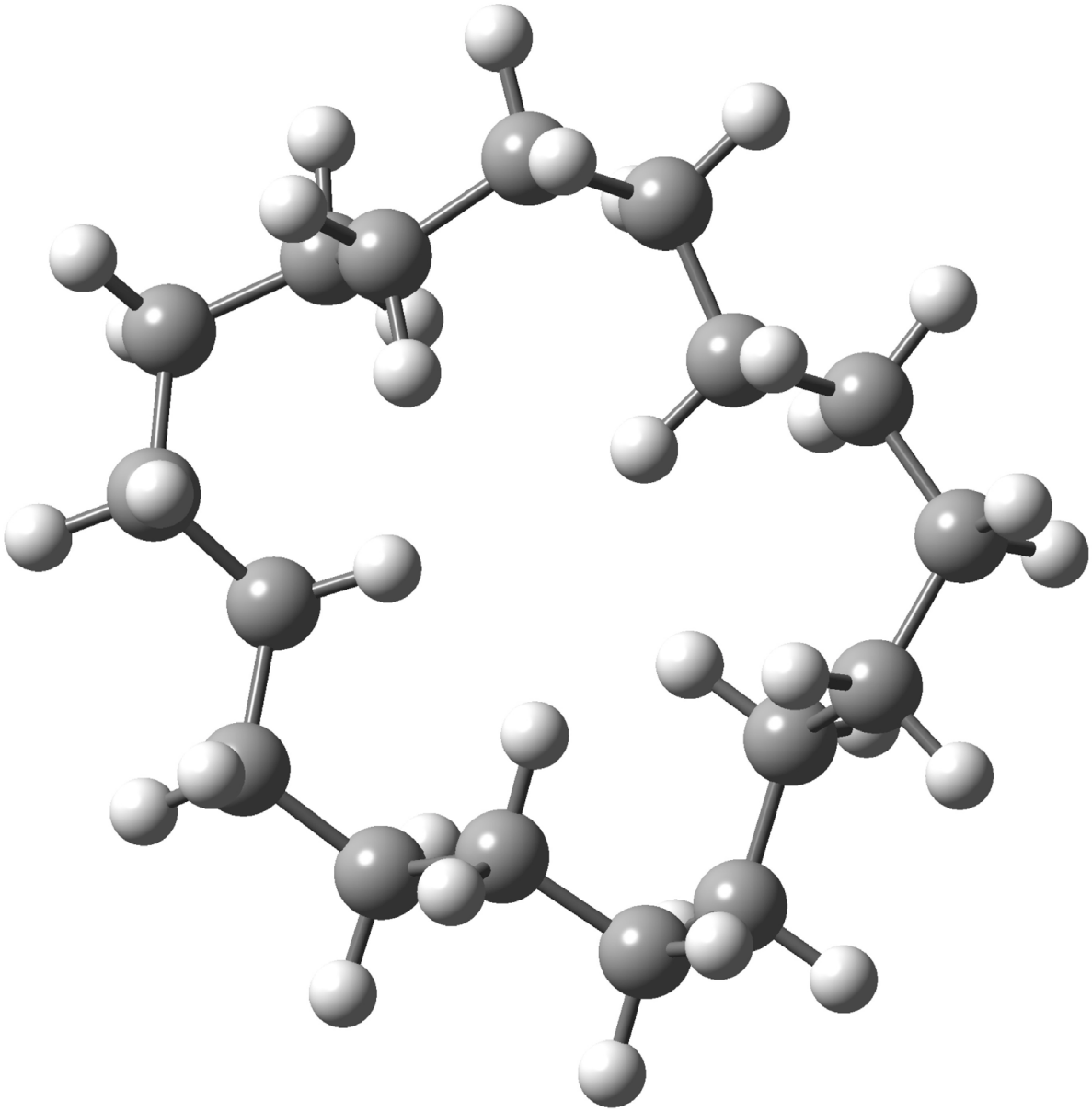


Figure 5.5. MM2 Enthalpic global minimum of cycloheptadecane.

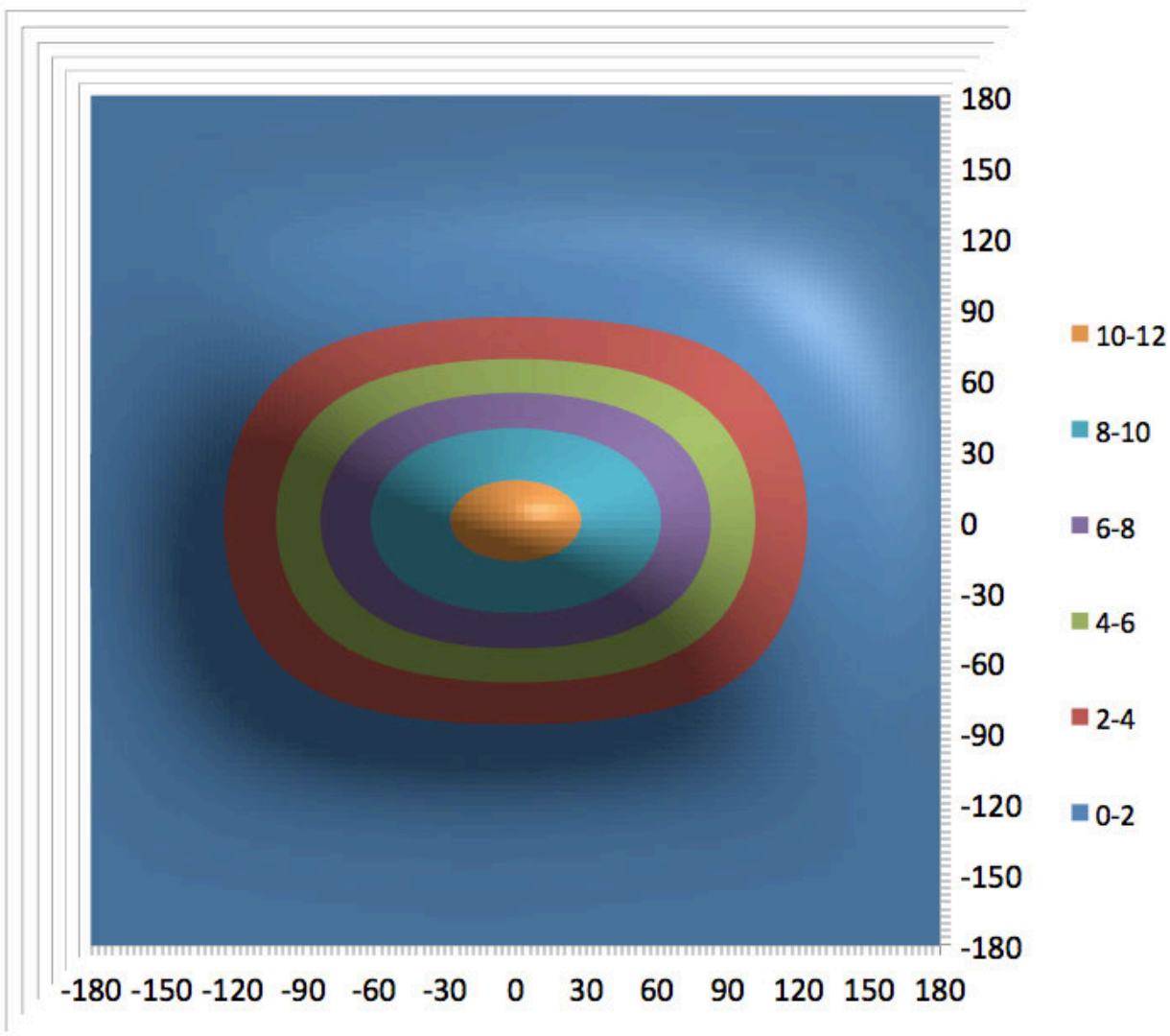


Figure 5.6. The product of two differently parameterized Gaussians (as a function of dihedral angle) utilizing the normalized Gaussian-weighted average of the scalar A. Vertical axis Gaussian: A = 1, Horizontal axis Gaussian: A = 20 arbitrary units.

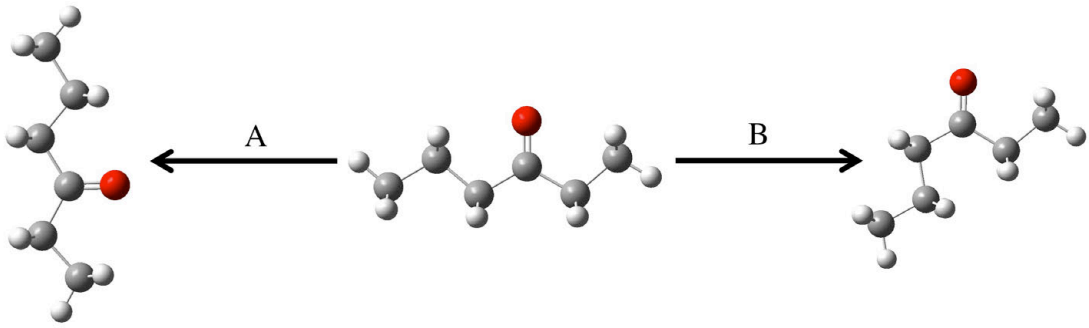


Figure 5.7. The comparison of 3-hexanone via path A can be made, as the quaternion function is continuous (the two conformers are identical). However, the quaternion cannot be evaluated via path B since the function is no longer continuous due to the conformation changing.

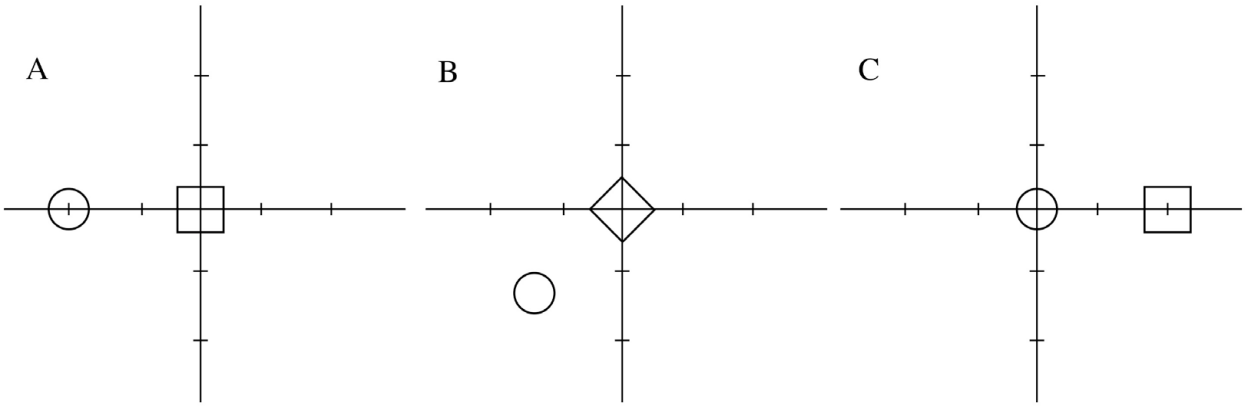


Figure 5.8. Rotational and translational orientations of a square and circle, where in each case they are relatively identical. If A is the unaltered state, B is rotated 45 degrees counter clockwise. A and B for the purposes of protein-ligand docking would be functionally identical. Additionally, comparing A to C, C is translated two units to the right. This case is relatively identical to both A and B.

CHAPTER 5 REFERENCES

(1) Lexa, K. W.; Carlson, H. A. *Q. Rev. Biophys.* **2012**, *45*, 301–343.

(2) Saunders, M.; Houk, K. N.; Wu, Y. D.; Still, W. C.; Lipton, M.; Chang, G.; Guida, W. C. *J. Am. Chem. Soc.* **1990**, *112*, 1419–1427.

(3) Giovannetti, J.; Kelly, C.; Landis, C. *J. Am. Chem. Soc.* **1993**, *115*, 4040–4057.

(4) Söderhjelm, P.; Tribello, G. A.; Parrinello, M. *Proc. Natl. Acad. Sci.* **2012**, *109*, 5170–5175.

(5) Buch, I.; Giorgino, T.; Fabritiis, G. D. *Proc. Natl. Acad. Sci.* **2011**, *108*, 10184–10189.

(6) Mu, Y.; Stock, G. *J. Phys. Chem. B* **2002**, *106*, 5294–5301.

(7) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.* **2004**, *47*, 1739–1749.

(8) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. *J. Comput. Chem.* **2009**, *30*, 2785–2791.

(9) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. *Proteins-Struct. Funct. Genet.* **2003**, *52*, 609–623.

(10) Schuttelkopf, A. W.; van Aalten, D. M. F. *Acta Crystallogr. Sect. -Biol. Crystallogr.* **2004**, *60*, 1355–1363.

(11) Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. *J. Mol. Graph. Model.* **2003**, *21*, 289–307.

(12) Wu, G. S.; Robertson, D. H.; Brooks, C. L.; Vieth, M. *J. Comput. Chem.* **2003**, *24*, 1549–1562.

(13) Osterberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. *Proteins-Struct. Funct. Genet.* **2002**, *46*, 34–40.

(14) Meng, E.; Shoichet, B.; Kuntz, I. *J. Comput. Chem.* **1992**, *13*, 505–524.

(15) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996**, *261*, 470–489.

(16) Allison, J. T.; Rappé, A. K. *Manuscr. Prep.*

(17) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.

(18) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(19) Ekonomiuk, D.; Su, X.-C.; Ozawa, K.; Bodenreider, C.; Lim, S. P.; Otting, G.; Huang, D.; Caflisch, A. *J. Med. Chem.* **2009**, *52*, 4860–4868.

(20) Irwin, J. J.; Shoichet, B. K. *J. Chem. Inf. Model.* **2005**, *45*, 177–182.

(21) Brooks, B.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.

(22) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

(23) Goedecker, S. *J. Chem. Phys.* **2004**, *120*, 9911–9917.

(24) Metropolis, N.; Rosenbluth, A.; Rosenbluth, M.; Teller, A.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087–1092.

(25) Ensing, B.; De Vivo, M.; Liu, Z. W.; Moore, P.; Klein, M. L. *Acc. Chem. Res.* **2006**, *39*, 73–81.

(26) Barducci, A.; Bonomi, M.; Parrinello, M. *Wiley Interdiscip. Rev.-Comput. Mol. Sci.* **2011**, *1*, 826–843.

(27) Laio, A.; Gervasio, F. L. *Rep. Prog. Phys.* **2008**, *71*, 126601.

(28) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562–12566.

(29) Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2005**, *127*, 2600–2607.

(30) Masetti, M.; Cavalli, A.; Recanatini, M.; Gervasio, F. L. *J. Phys. Chem. B* **2009**, *113*, 4807–4816.

(31) Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. *J. Phys. Chem. B* **2006**, *110*, 3533–3539.

(32) Leone, V.; Marinelli, F.; Carloni, P.; Parrinello, M. *Curr. Opin. Struct. Biol.* **2010**, *20*, 148–154.

(33) Hamilton, W. R. In *Proceedings of the Royal Irish Academy*; 1844; Vol. 2, pp. 424–434.

(34) Euler, L. *Novi Comment. Acad. Sci. Petropolitanae* **1776**, *20*, 189–207.

(35) Karney, C. F. F. *J. Mol. Graph. Model.* **2007**, *25*, 595–604.

(36) Long, S. M.; Tran, T. T.; Adams, P.; Darwen, P.; Smythe, M. L. *J. Comput. Chem.* **2011**, *32*, 1541–1549.

(37) Kolossváry, I.; Guida, W. C. *J. Comput. Chem.* **1999**, *20*, 1671–1684.

APPENDIX

A.1 – Abbreviations and Their Corresponding Names

CV: Collective variable

GOTS: Gradient only tabu search

GTS: Gradient tabu search

MC: Monte Carlo

MD: Molecular Dynamics

NP: Non-deterministic polynomial-time

PBIL: Population based incremental learning

PCM: Polarizable continuum model

PES: Potential energy surface

RMSD: Root mean squared difference

SEMD: Surface editing molecular dynamics

SVD: Singular value decomposition

THF: Tetrahydrofuran