An Algorithm To Explore Molecular Potential Energy Surfaces

Bhupinder Singh Metneja

A Thesis

in

The Department

of

Chemistry and Biochemistry

Presented in Partial Fulfillment of the Requirements

For the Degree of Master of Science (Chemistry) at

Concordia University

Montreal, Quebec, Canada

November 2021

CONCORDIA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: **Bhupinder Singh Metneja**

Entitled: **An Algorithm to Explore Molecular Potential Energy Surfaces**

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Chemistry)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

_____ Chair

Dr. Yves Gélinas

_____ Examiner

Dr. Gilles Peslherbe

_____ Examiner

Dr. Peter Pawelek

_____ Supervisor

Dr. Guillaume Lamoureux

_____ Supervisor

Dr. Heidi Muchall

Approved by _____

Dr. Paul Joyce, Chair of Department

Nov 24, 2021 _____

Dr. Pascale Sicotte, Dean, Faculty of Arts and Science

Abstract

An Algorithm To Explore Molecular Potential Energy Surfaces

Bhupinder Singh Metneja, M.Sc.

The work is aimed at modifying an algorithm called the Activation-Relaxation Technique (ART). The algorithm was initially envisioned by the original authors to be applied to the study of materials. This work, however, aims to take the original vision forward by applying it to the exploration of potential energy surfaces (PESs) of chemical and enzymatic reactions. To make ART more efficient in terms of the force evaluations needed and, consequently, to reduce the computational time and cost, it has been modified to include additional algorithms that significantly speed up the exploration of the PES. Each such algorithm is designed such that it allows for a strategic displacement of the atoms and subsequent sharing of atomic displacement vector information. Such a strategic displacement of atoms, for example, involves exploiting an *a priori* knowledge of the reactants in order to avoid atomic displacement moves that would not contribute to the tracing of their PES. A molecule like N-methylacetamide, for instance, which is a good model system for proteins, only undergoes a conformational change (a gradual dihedral rotation) which acts as the sole reaction coordinate for its PES and thus algorithms are included in ART that disallow for any other atomic displacement moves except for a gradual rotation of the dihedral. The sharing of atomic displacement vector information can prove helpful, for example, in enzymatic reactions where the enzyme mutants (which are similar in structure to the wild-type) can be made to follow the displacement vector information of the wild-type in order to avoid going astray along the PES and to have a good starting point from the beginning. The goal, ultimately, is to apply ART to the study of complex chemical reactions and enzymatic reactions involving large-scale rearrangements of the protein environment. As validation steps in the development of this algorithm, we illustrate the modified ART method on selected small molecules – ethane, propane, methyl acetate, N-methylacetamide, etc. – highlighting pitfalls and successes.

Acknowledgements

I would like to thank my parents, Maninder and Kanchan, and my brother Dhanvir, without whose love and support I would not have made it this far. I also want to thank both my supervisors Dr. Guillaume Lamoureux and Dr. Heidi Muchall because I got to learn so much from them. Many thanks also to my committee members Dr. Peter Pawelek and Dr. Gilles Peslherbe for their useful advice and comments. My thought also goes to Ms. Maria Dochia for always being so encouraging and instilling confidence in me. Last, but not the least, I would like to thank my lab members (both past and present): Tugba, Navjote, Philippe, Janet, Georgy, and Justin for their support and also for being such good friends.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ART   Activation Relaxation Technique

CYP450   Cytochrome P450

CYP450$_{\mathbf{BM3}}$ Cytochrome P450 from *Bacillus Megaterium*

GSM   Growing String Method

IRC   Intrinsic Reaction Coordinate

MD   Molecular Dynamics

MM   Molecular Mechanics

NEB   Nudged Elastic Band

NMA N-Methylacetamide

NPG N-Palmitoglycine

PES   Potential Energy Surface

QM   Quantum Mechanics

QST   Quadratic Synchronous Transit

RHF   Restricted Hartree-Fock

TS   Transition State

Chapter 1

<u>Introduction</u>

1.1 Motivation

The exploration of potential energy surfaces (PES) is vital as it can help reveal a wealth of information about the energetics and kinetics of a reaction. A potential energy surface (PES) is defined as the plot of the potential energy of a chemical or enzymatic reaction system with respect to the internal coordinates like the bond length, bond angle or dihedral angle. Figure 1.1 shows a contour plot of a potential energy surface. Locating the transition state on a potential energy surface describing a reaction path is at the core of studying any reaction mechanism [1]. Computational chemical and biological predictions such as reactivity predictions, regioselectivity predictions, catalyst or enzyme design all use the knowledge of transition state [2]. Predicting the outcome of a reaction, its kinetics, and the pathways leading to its products requires detailed knowledge of its energy landscape, that is, the energies of all molecular configurations encountered along the reaction. The potential energy surfaces of complex chemical and enzymatic reactions involve a multitude of peaks and valleys on their surface and thus exploring such topologies is a task in itself and there is, therefore, a need to develop algorithms that quickly and efficiently explore such potential energy surfaces. An automated method for exploring PESs is vital in the case of enzyme engineering, for instance, where one is dealing with a large number of enzyme mutants for a particular wild-type. Those mutants need to be screened for desirable properties such as the production of a certain metabolite that could act as a potential drug candidate or as a food and feed additive. Thus, it becomes imperative to quickly ascertain the PES corresponding to the biological reactions exhibited by these wild-type enzymes and their mutants. Sawayama *et al.* [3] have

demonstrated that some variants for Cytochrome P450 (CYP450), for example, from the organism *Bacillus megaterium* (CYP450$_{BM3}$) produce metabolites for marketed drugs such as verapamil and astemizole. CYP450s monooxygenate organic substrates by using molecular oxygen and electrons from reduced nicotinamide cofactors [4]. In this work, we aim to automate the exploration of energy landscapes of small organic molecules by modifying the activation-relaxation technique (ART) [5–7]. We have coupled ART with Gaussian 09 [8] and CP2K [9] – quantum chemistry software packages for *ab initio* calculations – and tested the modified code by exploring the energy surface of small organic molecules. As a proof of concept that ART could be used to study the PESs of proteins and enzymatic-substrate reactions, we have tested the code on *N*-methylacetamide which is a good model system for proteins as it contains the CONH bond that resembles the peptide bond in proteins [10]. We have modified the ART code to introduce the "information-sharing" approach whereby multiple replicas of the code run in parallel and share structural or directional information with each other. This is especially advantageous in enzyme engineering where large number of enzyme or substrate variants are typically considered and so using this approach helps cut down computational cost.
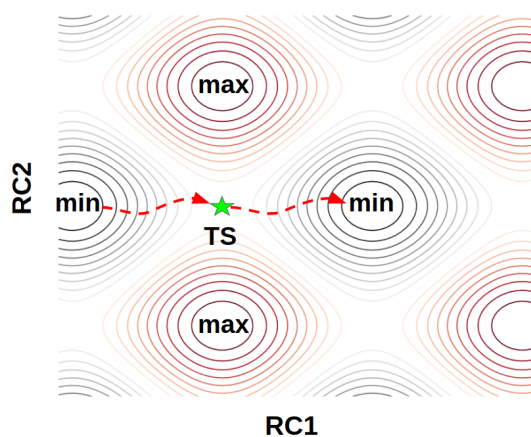

Figure 1.1: Illustration of a potential energy surface. RC1 and RC2 represent two reaction coordinates (e.g. a change in bond length, bond angle, or dihedral angle). The contours in red are representative of energy maxima while those in black are representative of energy minima. The transition state (TS) is represented by a star.

1.2   Numerical methods to explore PESs

Common numerical methods to explore potential energy surfaces include the growing string method [11], the eigenvector-following method [12], the nudged elastic band method [13], the minimum mode following method [14], the dimer method [15], the conjugate peak refinement [16], the ridge method [17], the DHS (Dewar, Healy, and Stewart) method [18], etc. Most of these methods either require a guess transition state to construct a minimum energy pathway that joins the initial energy minimum via a transition state to the final energy minimum or require the knowledge of both the initial and the final energy minimum. In contrast, the activation-relaxation technique (ART) neither requires a guess transition state nor the knowledge of a final energy minimum. It just uses the initial energy minimum as a starting point to construct an energy pathway. This is particularly useful in scenarios when the transition states are either too complicated to be guessed or the products of the reaction (final energy minimum) are not known.

1.2.1   Transition state search methods

The growing string method (GSM) works by addition of new discretized structures (also known as nodes) obtained by changing parameters such as bond length, dihedral angle or torsional angle to construct a reaction pathway. The string grows as new nodes are added and that is how the technique gets its name. The addition of nodes continues until a complete reaction pathway has been found containing a transition state and energy minima on each side of the string [19]. Since the addition of nodes is incremental, the reaction pathway constructed is very well converged as no nodes are placed at high energy regions of the PES [11]. GSM has two phases – a) the growth phase: During this phase, new nodes are added along the reaction coordinate – which can be a change of bond length, bond angle, or torsional angle – and relaxed perpendicular to the reaction coordinate. b) the optimization phase: All the nodes are optimized once the string is fully grown [19]. The growing string

method helps generate a good estimate of the transition state [11]. GSM can be either double-ended or single-ended. In a double-ended GSM, both the initial and the final states (i.e., the reactants and the products) need to be known whereas in a single-ended GSM, an energy pathway is constructed from just the initial state [19].

The nudged elastic band method [13] requires the knowledge of both the initial and final states of the system, i.e., the reactants and the products in order to be able to construct a pathway connecting these two states. A series of conformations of the system (also known as images) make up the pathway. These conformations are held by spring forces along the reaction pathway which keep the states along the path equally spaced out. The perpendicular component of the force comprises of potential forces [20]. The energy path thus behaves like an elastic band because of the spring forces that are acting parallel to the reaction pathway and that is how the technique gets its name.

The total nudged elastic band force on a given image i is given by:

$$\mathbf{F}_i^{\mathbf{NEB}} = \mathbf{F}_i^{\perp} + \mathbf{F}_i^{\mathbf{S}\parallel} \tag{1.1}$$

The component of the force perpendicular to the band and arising due to the potential, $\mathbf{F}_i^{\perp}$ can be represented as:

$$\mathbf{F}_i^{\perp} = -\nabla(\mathbf{R}_i) + \nabla(\mathbf{R}_i) \cdot \hat{\boldsymbol{\tau}}_i \hat{\boldsymbol{\tau}}_i \tag{1.2}$$

where $\hat{\boldsymbol{\tau}}$ is the normalized vector tangent to the path and pointing towards the adjacent image along the NEB path; $\mathbf{R}_i$ represents the coordinate vector of the ith state.

The parallel force acting on the band, $\mathbf{F}_i^{\mathbf{S}\parallel}$ is given by the following expression:

$$\mathbf{F}_i^{\mathbf{S}\parallel} = k(|\mathbf{R}_{i+1} - \mathbf{R}_i| - |\mathbf{R}_i - \mathbf{R}_{i-1}|)\hat{\boldsymbol{\tau}}_i \tag{1.3}$$

where k is the spring constant.

An obvious disadvantage of the nudged elastic band method is it requiring both the initial and the final energy states in order to construct a pathway connecting those states. One might not know the end products (final energy states) in some scenarios and the nudged elastic band method would not be useful in such a case.

The eigenvector-following method [12] constructs a reaction pathway starting from an initial minimum by moving along the direction corresponding to any chosen eigenvector along the reaction pathway. The curvature at each point of the reaction pathway is determined by estimating the Hessian which is a square matrix of the second-order partial derivatives with respect to the atomic coordinates of the function describing the energy surface:

$$\mathbf{H}_{i,j} = \begin{pmatrix} \frac{\partial^2 E}{\partial q_1{}^2} & \frac{\partial^2 E}{\partial q_1 \partial q_2} & \cdots & \frac{\partial^2 E}{\partial q_1 \partial q_n} \\ \frac{\partial^2 E}{\partial q_2 \partial q_1} & \frac{\partial^2 E}{\partial q_2{}^2} & \cdots & \frac{\partial^2 E}{\partial q_2 \partial q_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 E}{\partial q_n \partial q_1} & \frac{\partial^2 E}{\partial q_n \partial q_2} & \cdots & \frac{\partial^2 E}{\partial q_n{}^2} \end{pmatrix} \tag{1.4}$$

The eigenvalues of the Hessian describe the curvature at each point of the energy surface. The direction corresponding to a particular eigenvalue is the eigenvector to that eigenvalue. The choice of the eigenvector is totally random but is weighted towards the eigenvectors corresponding to the lower eigenvalues since the lower the eigenvalue, the faster the convergence to the saddle point.

A first-order saddle point representative of a transition state has all positive eigenvalues except one which is negative. The minimum mode following method [14, 21] works by following the eigenvector corresponding to this one negative eigenvalue because it is believed to converge to the saddle point (transition state) faster. This is what is referred to as the minimum mode in this method. In this method, the forces close to the minimum, $\mathbf{F}$, are transformed by subtracting the component of the force along the minimum mode:

$$\mathbf{F}^t = -(\mathbf{F} \cdot \mathbf{P}_\lambda)\mathbf{P}_\lambda \tag{1.5}$$

where $\mathbf{P}_\lambda$ represents the eigenvector that corresponds to the minimum eigenvalue mode and $\mathbf{F}^t$ represents the transformed force.

The transformed force starts pointing towards the saddle point. This transformation is needed so that the force vectors do not go astray on the energy surface and are able to successfully exit the region of the harmonic basin, i.e., the region where all eigenvalues are positive.

The forces close to the saddle point, however, are transformed by twice subtracting the component of the force along the minimum mode:

$$\mathbf{F}^t = \mathbf{F} - 2(\mathbf{F} \cdot \mathbf{P}_\lambda)\mathbf{P}_\lambda \tag{1.6}$$

The transformed force starts pointing towards a neighbouring minimum. The minimum mode following method does not need to know either a guess structure for the transition state on the final energy minimum (the end-products) in order to explore the energy surface. It can, however, be computationally very intensive in terms of the force evaluations needed in order to estimate the minimum mode [21]. These can be overcome by making modifications to the method. So for example, instead of constructing the Hessian matrix to estimate the minimum mode, one could use methods like the Lanczos algorithm [22].

1.2.2 Methods for automated exploration of PESs

Jacobson *et al.* [2] have presented an automated method for the exploration of PESs whereby minimal input from the user is needed in order to compute the pathway associated with a particular reaction. The method requires the user to feed in the information about the reaction and the initial reactant complex. The atoms involved in the reactant complex are then ordered. The coordinates of the reaction complex are then altered which leads to a change in bond distance, bond angle, or torsional angle and the end product or product complex is generated using a path of minimum energy. An estimate of the transition state

is made along this path which is then optimized using the quadratic synchronous transit (QST) approach [23]. The transition state is then validated whether it is indeed a true first-order saddle point or not and whether or not it connects the two end points – reactants and products.

Another automated algorithm for exploring reaction mechanisms and consequently their potential energy surface is PyFrag [24]. It works by taking the approximate geometries of the reactants (or reaction complex), products and transition state from the user. The reactants and products are then optimized. The transition state is optimized using the approximate transition state fed by the user as a template. All these stationary points are then validated using a frequency analysis. The reactants and products which are energy minima on the potential energy surface have all real frequencies whereas the transition state which, mathematically, is a first-order saddle point has exactly one imaginary frequency. An intrinsic reaction coordinate (IRC) approach [25] is then used to ascertain whether the transition state and the reactants and the products flanking it on either side are indeed connected with each other. An activation strain analysis [26] is then performed which estimates the energy required to deform the reactants in order to reach the transition state. The results are then collected and tabulated.

Simm *et al.* [27] have put together different approaches to exploring reaction pathways. One such approach involves relying on the curvature of the potential energy surface to explore the reaction pathways. The curvature is negative around the transition state – since mathematically the transition state is a first-order saddle point which is a maximum in one direction and hence the negative curvature. The minima, on the other hand, have positive curvature. A second approach involves using an *a priori* knowledge of chemistry to connect the reactants to intermediates or transition states. The knowledge can be, for example, a reaction involving rearrangement of bonds, or involving a conformational change. A third approach involves some sort of a tactile feedback which is detected by the system and informs it that a reaction has been accomplished. A receptor and a ligand, for

example, which have affinity for each other can initiate such a tactile or haptic feedback and thus can signal the onset of a reaction. The exploration of pathways in the second and third approaches could be crude and thus needs to be refined.

The activation relaxation technique (ART) developed by Mousseau and coworkers at the Université de Montréal is a method applicable to a broad range of organic, inorganic and biological molecules to explore their energy surface [5–7,28]. Starting from an initial energy minimum, the configuration is deformed until it is pushed out of the region close to the energy minimum, also known as the harmonic well. The eigenvalues of the Hessian are estimated at each step and as soon as one eigenvalue becomes negative, the system is out of the harmonic well. The configuration is then deformed further in the direction of the eigenvector corresponding to this negative eigenvalue until the total force becomes close to zero (magnitude of the force lower than a certain threshold). This indicates a saddle point. The configuration is then pushed slightly over the saddle point in order to minimize the odds that it will fall back to the initial minimum. The configuration then relaxes to a final minimum. This comprises one ART event. ART is so named because each ART event consists of two stages: an activation stage and a relaxation stage. Activation can itself be divided into two stages [29]:

- Leaving the harmonic well: The initial energy minimum with coordinates $\mathbf{q}$ is deformed randomly. With each deformation $\delta\mathbf{q}$, the energy of the system increases and it moves uphill in the energy profile until an inflection point in the energy profile is reached. The coordinates $\mathbf{q}$ at each step k will be:

$$\mathbf{q}_k = \mathbf{q}_{k-1} + \frac{\delta x}{||\delta\mathbf{q}||}\delta\mathbf{q} \qquad (1.7)$$

Here, δx is the amplitude of the deformation. At each step, the conformation is relaxed using the steepest descent method. However, this relaxation is in the hyperplane orthogonal to the deformation, $\delta\mathbf{q}$ in order to ensure that the conformation does not

fall back to the initial energy minimum. After the relaxation at each step, the lowest eigenvalue of the Hessian, λ, is estimated. Since evaluating the complete Hessian to find λ would be computationally expensive, the Lanczós algorithm [22] is used which helps find λ without evaluating the complete Hessian. If λ is positive, this deformation is increased one step further. Once it becomes negative (and below a certain threshold), convergence to the saddle point (transition state) starts.

- Convergence to the saddle point: Once λ is negative, convergence to the saddle point is achieved by moving along the eigenvector $\mathbf{v}$ corresponding to this negative eigenvalue. Even at this stage, relaxation in hyperplane perpendicular to $\mathbf{v}$ occurs.

$$\mathbf{q}_l = \mathbf{q}_{l-1} + \frac{\delta x_l}{||\mathbf{v}||}\mathbf{v} \qquad (1.8)$$

The amplitude of displacement δx_l at every step decreases with the square root of the number of iterations to facilitate convergence to the saddle point,

$$\delta x_l = \frac{\delta x}{\sqrt{l}} \qquad (1.9)$$

To calculate the energy of and the forces acting on the conformation at each step, ART was originally interfaced with BigDFT [5], but we have created an interface with Gaussian and CP2K. When the forces acting on the molecule become lower than a certain threshold, the conformation converges to a saddle point.

Once at the saddle point, the configuration is nudged one step further in the direction of $\mathbf{v}$ to relax it to a new minimum on the other side. The relaxation is done, for example, by performing an energy optimization on the structure with Gaussian.

1.3 ART in comparison with other methods

ART being an open-ended method gives it an obvious edge over other methods like the nudged elastic band that require the knowledge of both the initial energy minimum (the reactant or reaction complex) and the final energy minimum (the products). This can be beneficial in case where one is dealing with mutated enzymes the products of the reactions of which with the substrates might be unknown. Another area where ART stands out in comparison with other methods – both non-automated and automated – is that the atomic displacements that the algorithm makes are completely random in order to construct a pathway that would connect it to the starting material with the product via a transition state. And thus very little input is needed – in terms of an *a priori* knowledge of the starting material, for example – by the user operating this algorithm. The completely random atomic displacements can also provide insights into whether a particular reaction is (or is not) energetically feasible taking alternate pathways to reach the same products. Thus, one can better map out the routes leading from one state to another and their feasibility using ART. The Lanczos algorithm [22] incorporated in ART enables it to determine the curvature along the potential energy surface without it having to evaluate the complete Hessian matrix. This greatly helps reduce computational cost and time and makes it more advantageous to use over other methods that rely on a full determination of the Hessian.

We have incorporated algorithms in the code whereby the information about the atomic displacements can be shared with ART simulations for similar enough systems – a wild-type and a mutant enzyme or a series of similar reactants, for example, can be made to share this information with each other. This way the surface terrain explored by one system can be retraced by another. This not only helps cut down computational cost in terms of force evaluations needed but also makes sure that the correct pathway is explored right from the get-go. Another use-case of this information sharing approach could be a very crude but quick exploration of the PES for a particular system using a low level of theory and then

sharing the topological information of the terrain explored during a second run but this time using a much higher level of theory. Revisiting the same states over and over again along the energy surface can sometimes be a problem with automated methods – those that require very little input from the user – but our modified version of ART circumvents that problem with the incorporation of algorithms that check the previously visited routes and avoid going along those for subsequent runs. Finally, the validation of all the states found during an ART run is something that ensures that any stationary points that do not conform to the criteria (frequency or geometry, for example) are weeded out. For an automated method that requires minimal input, the exploration of false stationary points along the surface can be a problem but this is taken care of by the post-processing analysis scripts that we have incorporated in ART. Running these scripts can help ascertain whether an ART event is valid or not, *i.e.,* whether the minima and the transition states are indeed true minima and true transition states.

Overall, our modified ART method has great potential to be applied to exploring the PESs of large macromolecules such as enzymes and also the PESs corresponding to the interactions of these large macromolecules with ligands – such as the interaction of an enzyme with a substrate.

1.4 The Vision – Applying ART to the study of proteins

Eventually, the goal is to apply ART to the study of enzymatic reactions and the model enzyme we intend to study is Cytochrome P450 from the organism *Bacillus megaterium.* The Cytochrome P450 enzymes are so named because of their presence in the cell (cyto) membrane, the presence of a heme pigment inside these (chrome and P) that absorbs light at a wavelength of 450 nm [30]. CYP450$_{BM3}$ is a widely used and well-investigated monooxygenase due to its stability, high activity, similarity to mammalian P450s, and presence of a CYP450 reductase domain that allows the enzyme to directly

receive electrons from NADPH without a requirement for additional redox proteins [4, 31]. The CYP450 system catalyzes the following reaction:

$$RH + O_2 + NAD(P)H + H^+ \longrightarrow ROH + NADP^+ + H_2O$$
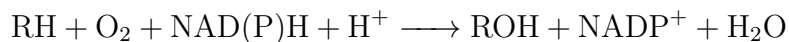
1.4.1 Understanding the reactivity of CYP450s

When using ART to simulate the reactions catalyzed by CYP450s, it is helpful to know where the reaction occurs as at that site simultaneous bond formation and breakage would be occurring. Thus, a hybrid quantum mechanics/molecular mechanics (QM/MM) approach could be used. The site of the catalysis of the substrate by the enzyme can be treated using quantum mechanics while the rest of the macromolecule can be treated using molecular mechanics. Our modified ART method can focus on the site where the reaction is occurring so that only the atoms involved in bond breakage and formation are activated. This can greatly help cut down on the number of force evaluations. Insights into the catalytically relevant active-site residues are yielded by the crystal structures of the substrate-free and substrate-bound CYP450$_{BM3}$ [32–34]. The binding of substrate to CYP450$_{BM3}$ leads to large conformational changes. These include narrowing of the binding pocket around the substrate and reorientation of the active-site side chains [33, 34]. Arg47 and Tyr51 found at the active site periphery of CYP450$_{BM3}$ help tether carboxylate groups of long chain substrates and stabilize transition state [33, 34]. Cys400 helps coordinate the heme iron. The carboxylate of the fatty acid substrate is attracted by electrostatic forces to the side chain of Arg47 at the mouth of the active site. Figure 1.2 shows the residues in and around the active site of CYP450$_{BM3}$. There is a strong driving force for the movement of the hydrophobic tail of the fatty acid from the polar solvent phase into the substrate binding channel [35]. The effects of simultaneously mutating Arg47 to alanine and Tyr51 to phenylalanine have been investigated by Ost *et al.* [35] and they have shown that these mutations significantly decrease the enzyme activity proving that both Arg47 and Tyr51 are catalytically relevant residues. Mutating these catalytically relevant residues can alter the energy surface associated with

the enzyme-substrate reaction. In enzyme engineering, when one is dealing with a large number of such mutations, it becomes imperative to quickly ascertain the PES associated with the reactions of the enzyme mutants with the substrates. Our modified ART algorithm can help quickly explore the energy terrain of such reactions. The mutants being similar in structure to the wild-type would have an equivalent energy pathway. And thus, the pathway explored for the wild-type can be followed for the mutant. This really makes the exploration of the energy surface much faster – especially if one is dealing with hundred of mutants which need to be screened for desirable properties.
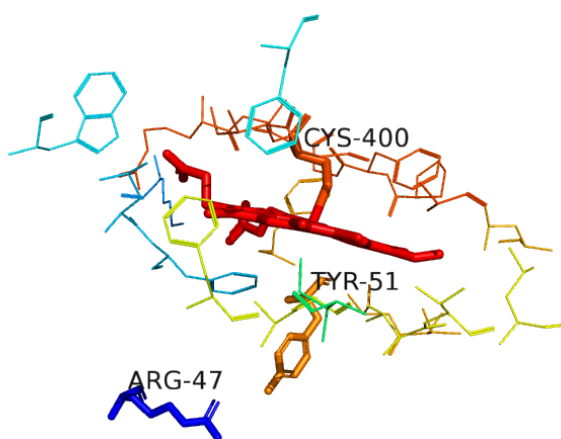

Figure 1.2: Residues in and around the active site of $CYP450_{BM3}$ (PDB ID: 1BVY). Image generated in PyMOL. [36]

1.4.2 Quantum Mechanics/Molecular Mechanics (QM/MM) modelling of reactions catalyzed by CYP450s

We aim to use ART to the study of QM/MM simulations on enzymatic reactions and for that very purpose we have coupled it with CP2K [9] – an *ab initio* package capable of performing QM/MM simulations on large macromolecules such as proteins. The objective of using a hybrid QM/MM approach is to treat a small part of a large system – one which is directly involved in a reaction (e.g. the active site of an enzyme) via QM while the rest of system via MM, thus obtaining high accuracy where it is required while avoiding the

computationally intensive QM calculations for a large number of atoms [37]. To use a hybrid QM/MM approach, it is imperative to partition the system into regions, one that will be treated using QM and the other that will be treated using MM. When such regions are not covalently bound (e.g. a solute in a solvent), the calculation of a hybrid QM/MM potential is relatively easy. However, when the regions are covalently bound as is the case with the active sites of enzymes, calculating the hybrid QM/MM potential becomes more challenging [37].

Dubey *et al.* [38] have used molecular dynamics (MD) and QM/MM simulations to predict the substrate induced gating of CYP450$_{BM3}$. The substrate-bound CYP450$_{BM3}$ with *N*-palmitoglycine (NPG) as the substrate (PDB ID: 1JPZ) was studied by them using MD simulations and the studies revealed that the substrate entrance induces a conformational change that closes the substrate channel. Their simulations showed that the major conformational transition from the open state to the closed state is exclusively induced by the NPG substrate.

Our modified ART algorithm can come in handy when the NPG substrate is bound to the enzyme, *i.e.,* when the channel is closed as it can explore the reaction pathway owing to the reaction of the substrate with the enzyme. The reaction would entail bond breakage, formation, and rearrangement at the active site and this region can be treated using Quantum Mechanical methods with our modified ART algorithm while the rest of the molecule can be treated using Molecular Mechanics.
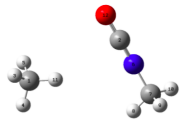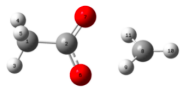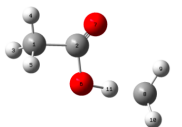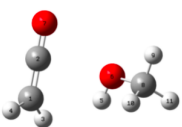
1.5 Separating conformational and reactive energy landscape exploration

The ART algorithm can deal with both conformational changes as well as reactive events which involve bond breakage, formation, rearrangement of atoms, transfer of groups and, at times, formation of highly unstable intermediates. Since there are molecular systems for which the PES is governed solely by conformational changes, it is imperative to only allow moves in the algorithm that would lead to a gradual rotation of the dihedral while

disallowing any other moves, e.g. bond stretching. We have incorporated algorithms in ART which ensure that no reactive moves e.g. stretching are allowed for a system that is supposed to only undergo conformational changes – e.g. the *cis-* to *trans-* conversion of an acetate or acetamide molecule. In such cases only a gradual rotation of the torsional angle is allowed, for example.

Since the initial atomic displacements for distorting the structure are completely random in ART, molecules like methyl acetate and N-methylacetamide when simulated using ART exhibited fragmentation, rearrangement of atoms, and new bond formation leading to the formation of highly unstable species like ketenes, isocyanates, epoxides, radicals, and carbenes. Table 1.1 shows some of these structures. Due to the formation of these rather high-energy and consequently unstable species, the energy barrier associated with exploration of PESs for these molecules rises significantly.

Table 1.1: High-energy molecular species generated by ART as a result of fragmentation. The simulations were run using the RHF/6-31G model chemistry

Species	Structure
Isocyanate	
Methyl radical	
Epoxide	
Carbene	
Ketene	

Chapter 2

Methods, Approaches, and Results

2.1 ART coupled with an *ab initio* package

The Activation-Relaxation Technique (ART) code was coupled with *ab initio* packages Gaussian and CP2K in order to evaluate the energy of and the forces acting on the structure. The ART code distorts the structure by displacing the atoms of the structure in a random fashion and makes a Gaussian or CP2K call every time it requires the energy of and the forces acting on the distorted structure. The ART user can choose which package to use and the call is made accordingly. The decision on how to displace the atoms further is taken by evaluating these energies and forces. At each point along the way on the potential energy surface that is being gradually traced, ART estimates the eigenvalues of the Hessian matrix to know the curvature along the potential energy surface. A negative curvature is indicative that the activation towards the saddle point (transition state) has begun. At the saddle point, the atoms of the structure are displaced further in the direction of the minimum on the other side of the potential energy surface. This is done in order to minimize the odds that the structure will fall back to the initial minimum. The minimum on the other side of the potential energy surface is then reached at by making a Gaussian optimization call – which relaxes the structure to a minimum. This comprises one ART event.

2.2 Validation of ART events

All stationary points found during an ART event – the initial energy minimum, the transition state (or mathematically, a first-order saddle point) and the final energy minimum are validated using a two-step validation. The first step involves optimizing the

structures found by ART. An energy optimization and a quadratic synchronous transit (QST3) optimization are done respectively on the minima and the transition state found by ART. If the optimized minima and transition state superimpose well on those found by ART, then the second validation step entails doing a frequency calculation on the ART structures. If the minima have all real frequencies, then they are indeed true minima while if the transition state has exactly one imaginary frequency, then it is indeed a true transition state. Once all the ART events are validated, the various states of the system – all the minima and transition states – are connected together in the form of a network. Such a network of states can help one determine the feasibility of a reaction or the shortest route to get from a reactant to a product.

2.3 Reducing Computational Cost

2.3.1 Information Sharing

To make the ART code more robust in terms of force evaluations needed to fully explore the potential energy surface of a structure, various information sharing strategies were incorporated. These strategies are aimed at sharing displacement vector information either across multiple ART runs going in parallel or across multiple events of the same ART run. By sharing this displacement vector information: a) ART already knows the right direction to aim at in its pursuit of the saddle point along the potential energy surface thereby cutting down greatly on force evaluations by eliminating any wasteful moves, b) New avenues along the potential energy surface can be explored and force evaluations are not wasted exploring the same avenues over and over again.

Information sharing strategies:

1. **Follow Strategy**: The *follow* strategy is so named because the direction that successfully led to a saddle point for one structure is followed for another similar structure. This helps ensure that no force evaluations are wasted making unsuccessful

17

moves that do not lead to the saddle. An application of the *follow* strategy could be in enzyme engineering, whereby, if the energy landscape exploration of a large number of enzyme mutants could be automated, it could make the process faster. The energy landscape for the wild-type enzyme could be explored by ART using default settings and for the mutants (which are similar in structure to the wild-type), the pathway explored for the wild-type could be followed – leading to a faster exploration of the landscape and with much less force evaluations. Another application of the *follow* strategy could be exploring the reaction pathway for a system at a low level of theory and then following that pathway for the same system but this time at a higher level of theory in order to more accurately explore the terrain. For designing the follow strategy, it is imperative that the structure which is going to follow the path of another aligns well with it. Figure 2.1 shows the wild-type and mutant CYP450$_{\text{BM3}}$ structures. The mutant has alanine at position 82 replaced with a hydrophobic residue phenylalanine [39]. As can be seen from the figure, the wild-type and the mutant structures align well together. The direction of atomic displacements chosen for the wild-type can thus be followed for the mutant. If the structures do not align well, the direction of atomic displacements used for one cannot be followed by the other since it will lead to a completely different pathway. We use the Kabsch algorithm [40] to check the alignment of the structures in question.

The coordinates of the two structures, $\mathbf{P}$ and $\mathbf{Q}$ (having n atoms each) that we want to check the alignment of are given by:

$$\mathbf{P} = \begin{bmatrix} p_{1x} & p_{1y} & p_{1z} \\ p_{2x} & p_{2y} & p_{2z} \\ \vdots & \vdots & \vdots \\ p_{nx} & p_{ny} & p_{nz} \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} q_{1x} & q_{1y} & q_{1z} \\ q_{2x} & q_{2y} & q_{2z} \\ \vdots & \vdots & \vdots \\ q_{nx} & q_{ny} & q_{nz} \end{bmatrix} \tag{2.1}$$

The centroids of the both structures, $\vec{c}_P$ and $\vec{c}_Q$ are calculated as:

$$c_{Px} = \frac{\sum_{i=1}^{n} p_{ix}}{n}; \quad c_{Py} = \frac{\sum_{i=1}^{n} p_{iy}}{n}; \quad c_{Pz} = \frac{\sum_{i=1}^{n} p_{iz}}{n}$$

$$(2.2)$$

$$c_{Qx} = \frac{\sum_{i=1}^{n} q_{ix}}{n}; \quad c_{Qy} = \frac{\sum_{i=1}^{n} q_{iy}}{n}; \quad c_{Qz} = \frac{\sum_{i=1}^{n} q_{iz}}{n}$$

The two structures, $\mathbf{P}$ and $\mathbf{Q}$ are then translated so that they are centred at the origin:

$$\mathbf{P}' = \begin{bmatrix} p_{1x} - c_{Px} & p_{1y} - c_{Py} & p_{1z} - c_{Pz} \\ p_{2x} - c_{Px} & p_{2y} - c_{Py} & p_{2z} - c_{Pz} \\ \vdots & \vdots & \vdots \\ p_{nx} - c_{Px} & p_{ny} - c_{Py} & p_{nz} - c_{Pz} \end{bmatrix}, \quad \mathbf{Q}' = \begin{bmatrix} q_{1x} - c_{Qx} & q_{1y} - c_{Qy} & q_{1z} - c_{Qz} \\ q_{2x} - c_{Qx} & q_{2y} - c_{Qy} & q_{2z} - c_{Qz} \\ \vdots & \vdots & \vdots \\ q_{nx} - c_{Qx} & q_{ny} - c_{Qy} & q_{nz} - c_{Qz} \end{bmatrix}$$

$$(2.3)$$

A covariance matrix, $\mathbf{M}$ is then calculated using the translated structures, $\mathbf{P}'$ and $\mathbf{Q}'$:

$$\mathbf{M} = \mathbf{P}'^{\mathrm{T}}\mathbf{Q}' \tag{2.4}$$

Since $\mathbf{P}'$ and $\mathbf{Q}'$ are both matrices of dimensions $n \times 3$ each, $\mathbf{M}$ will be a 3×3 matrix. The singular value decomposition (SVD) of $\mathbf{M}$ factorizes it as follows:

$$\mathbf{M} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{T}} \tag{2.5}$$

where $\mathbf{U}$ and $\mathbf{V}$ are 3×3 orthogonal matrices and $\boldsymbol{\Sigma}$ is a 3×3 diagonal matrix. The optimal rotation matrix, $\mathbf{R}$ that rotates one structure to align it with the other is given by:

$$\mathbf{R} = \mathbf{U}^{\mathrm{T}}\mathbf{V} \tag{2.6}$$

An alignment matrix, $\mathbf{A}$ is then calculated which is a combined matrix of all the transformations – the optimal rotation and the translation needed to align the two

structures:

$$
\mathbf{A} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & -(c_{Qx} - c_{Px}) \\ r_{21} & r_{22} & r_{23} & -(c_{Qy} - c_{Py}) \\ r_{31} & r_{32} & r_{33} & -(c_{Qz} - c_{Pz}) \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \text{where } \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} = \mathbf{R} \qquad (2.7)
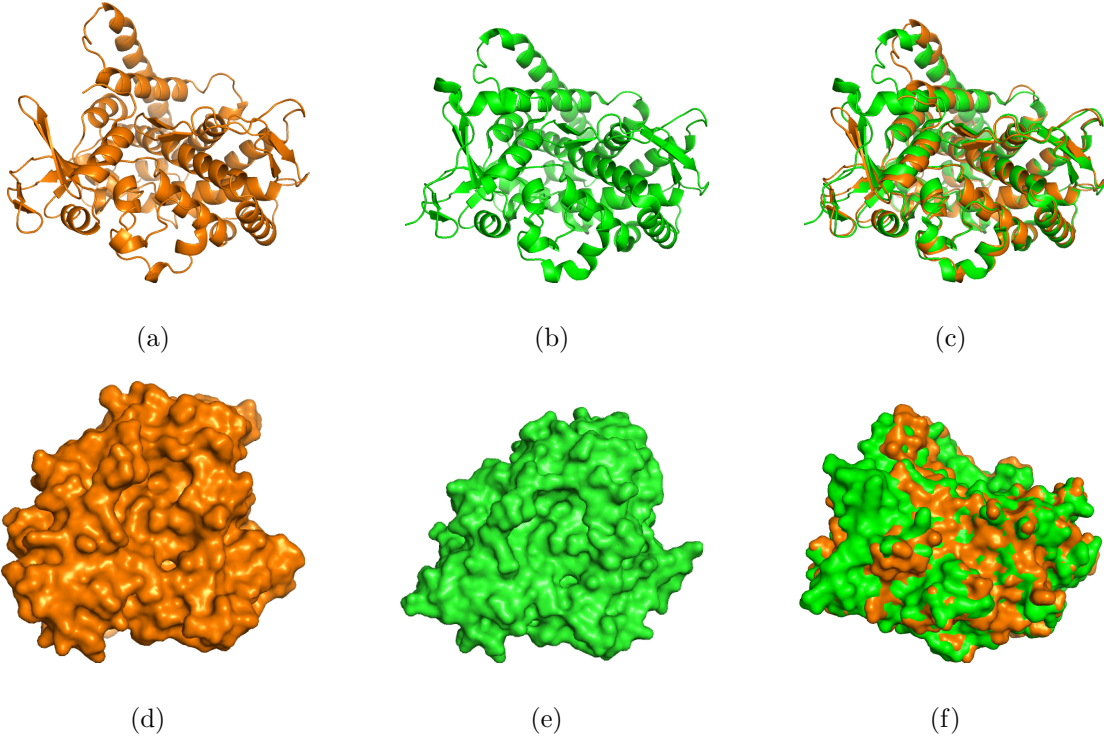$$

(a)   (b)   (c)

(d)   (e)   (f)

Figure 2.1: The figure shows CYP450$_{\mathrm{BM3}}$ wild-type (PDB ID: 1BVY) and mutant (PDB ID: 2UWH) structures . The wild-type is shown in brown and the mutant in green. (a) and (b) show the ribbon structures of the wild-type and mutant respectively while (d) and (e) show the surface structures of the wild-type and mutant respectively. The alignment of the wild-type and mutant ribbon structures is shown in (c) and surface structures is shown in (f). Images generated in PyMOL. [36]

2. **Avoid Strategy**: Another strategy is to have ART avoid the directions which have previously been explored for a system in order to explore new avenues along the energy landscape for that system. This prevents the same states from being revisited several times over and thus provides a means to explore the full energy landscape faster. As ART picks a random atomic displacement vector for distorting the system, it compares that randomly chosen vector with the displacement vectors that have already been used to explore the landscape. This is done, for instance, by computing the cosine of the angle between the two vectors. If a yet to be visited direction is deemed collinear to the one already visited, *i.e.,* the cosine of the angle between the two directions is close to 1, it is avoided and a new random direction is chosen (see Figure 2.2(b)).
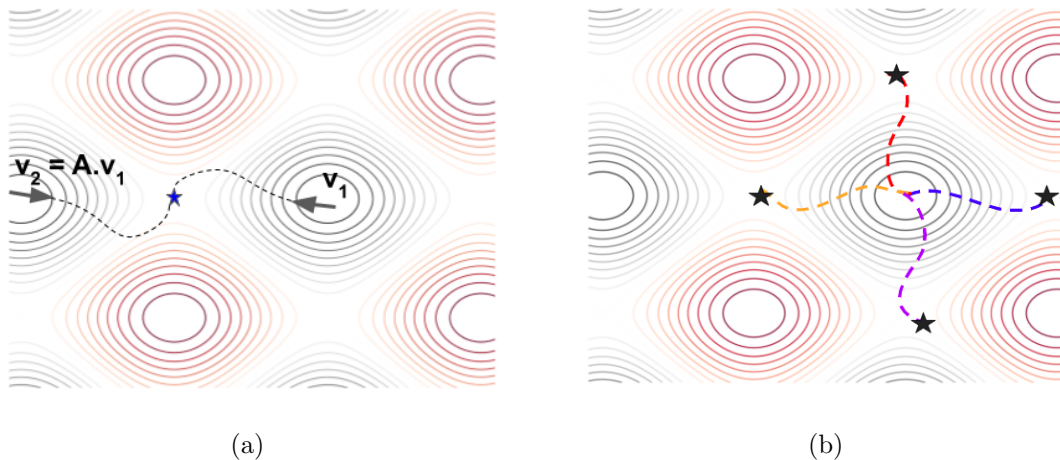

(a)   (b)

Figure 2.2: Illustration of the (a) *follow* and (b) *avoid* strategies. On the contour plots, the regions in black represent energy minima and the regions in red represent energy maxima while the transition states are represented by stars. In the (a) *follow* strategy, the atomic displacement vector $\vec{v}_1$ used to displace the atoms for one system is followed for another similar system. If the two structures align well, the transformation vector, $\mathbf{A}$, required to superimpose one structure onto the other is applied to the displacement vector that is to be followed. In the (b) *avoid* strategy, any direction that is already explored is avoided for the next run. The directions red, yellow, blue, and purple shown in the figure all explore different avenues along the energy surface.
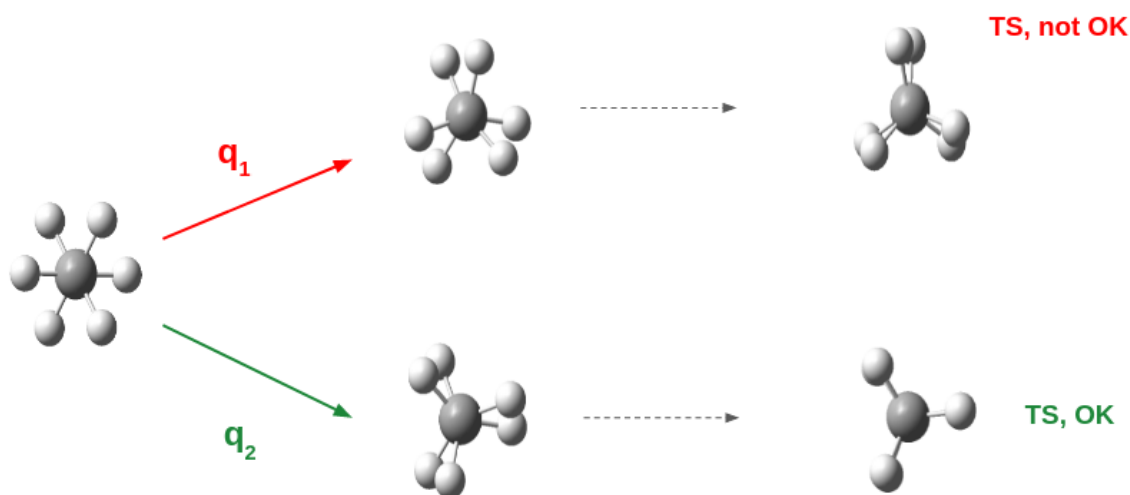
Figure 2.3: Illustration of a potential application of the *avoid* strategy. Starting from a staggered ethane conformation, an atomic displacement vector $\vec{\mathbf{q}}_1$ leads to a transition state that is not the real transition state as it is not an eclipsed conformation. Atomic displacement vector $\vec{\mathbf{q}}_2$, however, leads to an eclipsed conformation. When using the *avoid* strategy, both $\vec{\mathbf{q}}_1$ and $\vec{\mathbf{q}}_2$ will be avoided for the next run.

It is imperative to judiciously utilize the follow and avoid strategies in order to not only exploit successfully explored avenues but also to explore new avenues. The objective is to minimize the computational cost while simultaneously ensuring that new pathways are discovered. One caveat with the avoid strategy is that it can quickly run out of directions in case all possible directions accessible for a given conformation have already been explored. To circumvent this problem, we use what is called a mixed approach where ART would sometimes deploy the *follow* strategy and sometimes the *avoid* strategy. The decision to follow or avoid is totally random.

2.3.2   Exploitation of *a priori* knowledge of structures

An *a priori* knowledge of the structure can be exploited in order to faster trace its potential energy surface by disallowing atomic displacements that do not contribute

towards tracing its energy surface. For e.g. an ethane molecule only undergoes a gradual rotation of the dihedral in order to reach the transition state. Thus, in such a case any bond stretching moves would be unproductive and what is needed is simply a gradual rotation of the dihdedral. For this purpose, we have devised some strategies which are as follows:

1. **Dihedral rotation**: This strategy disallows any stretching moves – allowing only for a gradual rotation rotation of the dihedral. A rotatable bond, the rotation of which leads to a transition state, is identified in the molecule and rotated. In order to find the displacement vector that would cause a rotational distortion in the molecule, the Rodrigues' rotation formula is used:

$$\vec{\mathbf{q}}_{\text{rotated}} = \vec{\mathbf{q}}\,\cos\theta + (\mathbf{k} \times \vec{\mathbf{q}})\sin\theta + \mathbf{k}(\mathbf{k}\cdot\vec{\mathbf{q}})(1 - \cos\theta) \tag{2.8}$$

where $\vec{\mathbf{q}}$ is the vector that is rotated by an angle θ; $\mathbf{k}$ is the unit vector describing the axis of rotation; $\vec{\mathbf{q}}_{\text{rotated}}$ is the resultant vector after the rotation.

This formula helps ascertain the resultant position of the atomic displacement vector after it is rotated about a rotatable bond (see Figure 2.4). This strategy greatly helps reduce the number of force evaluations by eliminating all unproductive atomic displacements – stretching, for example – that do not lead to the transition state.
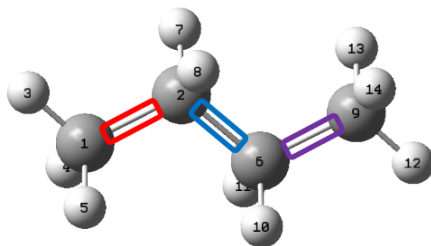

Figure 2.4: Illustration of the *dihedral rotation* strategy. The figure shows a butane molecule in which three rotatable bonds are identified. A rotatable bond is picked at random and atoms on either side of it are made to move.

2. **Bond angle bending**: Not too dissimilar to the *dihedral rotation* strategy is the *bond angle bending* strategy that disallows all atomic displacements except a gradual change in bond angle, i.e., an angle defined by three atoms connected by a pair of bonds.

3. **Focussed strategy**: As the molecular system grows larger and more complex, the number of force evaluations needed to fully explore the energy landscape increase proportionately. Thus, we have incorporated a strategy whereby atomic displacements are allowed only in a subsection of the molecular system – one that is directly involved in the reaction. In enzyme-substrate reactions, for example, the atoms in the vicinity of the active site of the enzyme are the ones directly involved in the reaction because that is where the interaction of the enzyme with the substrate occurs giving rise to products, so only in that region will the atomic displacement moves be allowed.

2.4   Testing the ART algorithm

The modified ART algorithm coupled with Gaussian was tested on ethane. This molecule was chosen to be a test system because we know what to expect in terms of its conformational states along the energy surface. Starting from a staggered conformation of ethane, the ART algorithm attempted to find a transition state – which is the eclipsed conformation (see Figure 2.5). The level of theory used for evaluating the energy of and the forces acting on the molecule was restricted Hartree-Fock (RHF). The basis set used was 6-31G(d). The energy barrier associated with the rotation of ethane from a staggered conformation to an eclipsed one has been reported by Mo *et al.* [41] to be 2.73 kcal/mol which is what ART also found. The force evaluations needed by ART to explore the reaction pathway corresponding to the rotation of ethane were 390. There was also one ART event which did not succeed in reaching the eclipsed conformation and thus ART began a new search (see Figure 2.6(a) and 2.6(b)). To cut down on the number of force evaluations, ART was run by implementing the dihedral rotation strategy. This helped ensure that any other atomic displacements except a gradual rotation of the ethane dihedral are disallowed and

this greatly cut down on the number of force evaluations which were brought down from 390 to 180. There were also no failed events in the process (see Figure 2.6(c)).
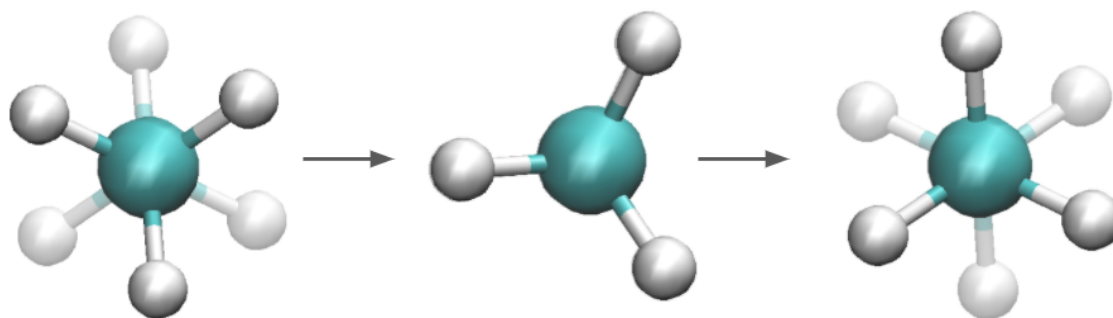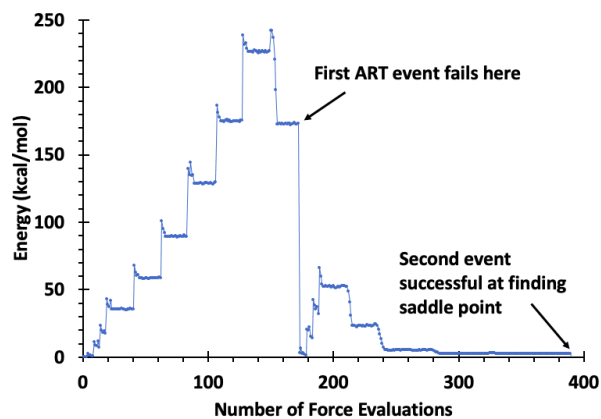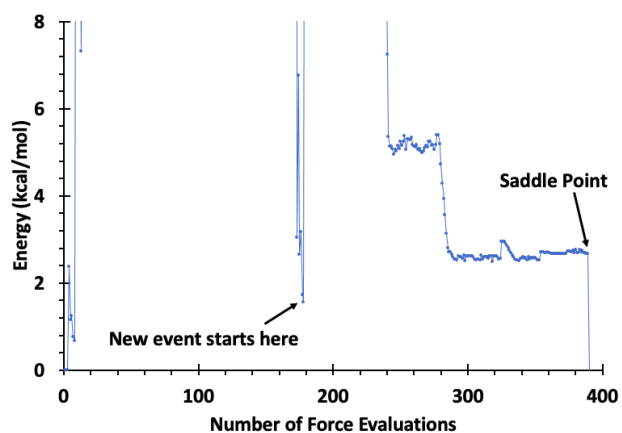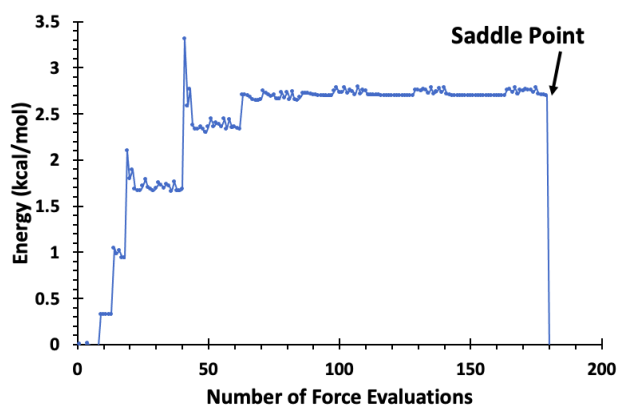

Figure 2.5: ART simulation of ethane showing the rotation of ethane from a staggered conformation to an eclipsed conformation and then relaxing back to a new staggered conformation.

(a)

(b)

(c)

Figure 2.6: Plot of energy vs number of force evaluations for an ART simulation for the ethane molecule. Panel (a) shows the plot of energy vs force evaluations for ethane using ART in the default mode. Panel (b) shows a zoomed-in version of panel (a). Panel (c) shows the plot of energy vs force evaluations when using the dihedral rotation strategy.

2.5 Structures Tested

The ART code was used to explore the potential energy landscapes of small organic molecules – e.g. esters, amides, and cyclic structures – as a proof of concept that it can successfully be used to study and explore more complex chemical and enzymatic reaction systems. The study of these organic molecules using ART was partitioned into two domains – a conformational landscape study where the tracing of the potential energy surface is governed solely by the conformational changes of the reactant or the reaction complex and a reactive landscape study where reactive events are also observed. This separation into two domains is necessitated in order to avoid unproductive atomic displacements to explore the energy landscape.

2.5.1 Conformational Landscape Exploration

1. **Methyl acetate**: We tested our modified ART algorithm on an ester such as methyl acetate since ester linkages are present in fatty acids like palmitic acid which are common substrates for CYP450s. Blom and Günthard [42] have shown via infrared analysis that the enthalpy difference between the *trans-* and *cis-* conformations of methyl acetate is 8.5 $\pm$ 1 kcal/mol.

$$\Delta H = 8.5 \pm 1 \text{ kcal/mol}$$

trans-methyl acetate   *cis*-methyl acetate
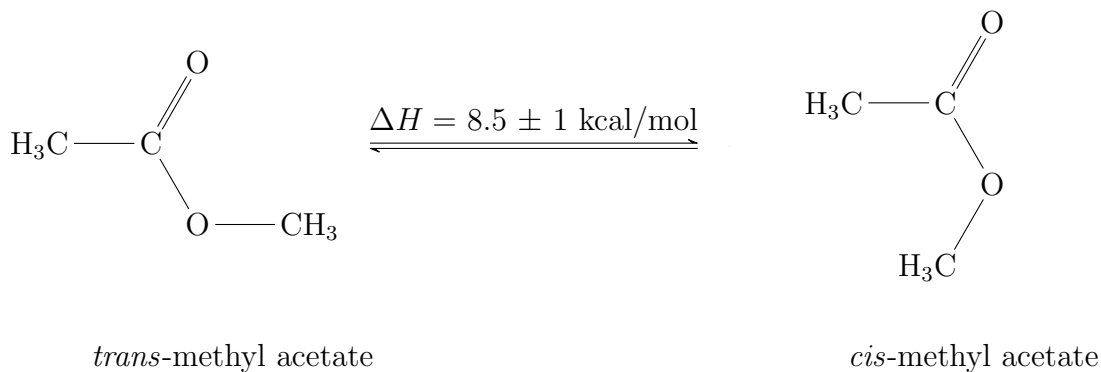
The methyl acetate molecule starting from its *trans-* conformation was simulated using ART. ART was able to find the transition state connecting the *trans-* conformation to

the *cis-* conformation. The energy barrier was found to be 24.5 kcal/mol (see Figure 2.7). The *cis/trans* energy difference was found to be 10.3 kcal/mol. The literature values report this difference to be 8.5 ± 1 kcal/mol [42, 43]. The level of theory used for calculating the energy and forces was restricted Hartree-Fock with a 6-31G basis set. The ART simulation was run in its default setting and it took 871 force evaluations for the isomerization of *trans-* methyl acetate to *cis-* methyl acetate. There were also 4 failed events in the process, *i.e.* events that did not succeed in converging to the transition state connecting the *trans-* and *cis-* conformer of methyl acetate. A Gaussian energy optimization and frequency calculation was done on all the conformations found in order to confirm whether the energy minima and the transition states are indeed true minima and true transition states and it was seen that the minima had all real frequencies while the transition state exactly one imaginary frequency as shown in the figure below:
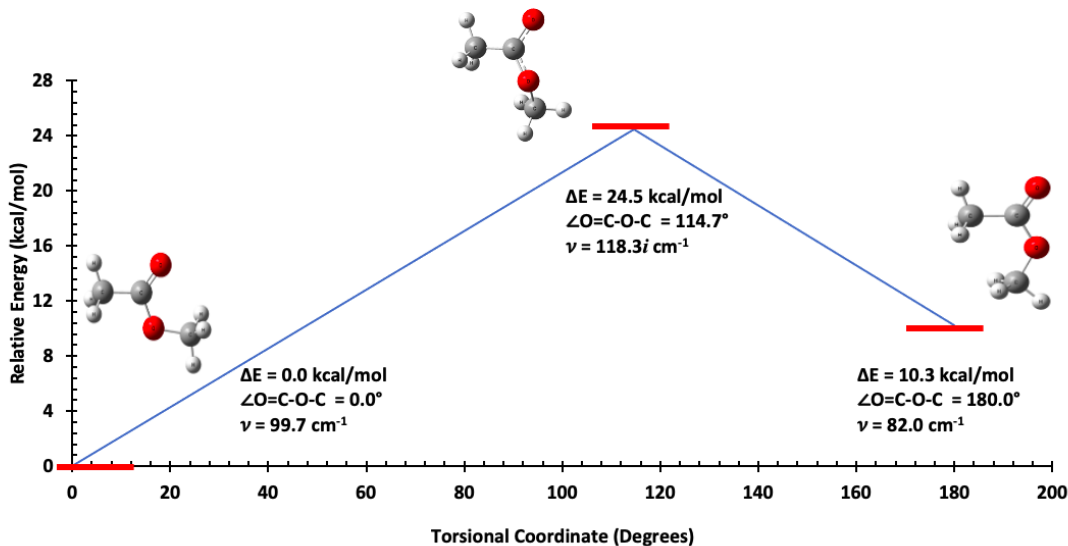

Figure 2.7: Energy diagram representing the methyl acetate conformations found by ART at RHF/6-31G model chemistry. Table A.1 in the appendix shows the total electronic energies of these conformations.

2. **N-methylacetamide (NMA)**: NMA is a good model system for proteins because

it contains the CONH bond which is also present as the peptide bond in protein backbone chains [44]. The free energy for internal rotation, $\Delta G^{\ddagger}$ for the *cis-* to *trans-*conversion of NMA as determined by NMR analysis is 18.8 ± 0.3 kcal/mol [45].

$$\Delta G^{\ddagger} = 18.8 \pm 0.3 \text{ kcal/mol}$$

trans-NMA   *cis*-NMA
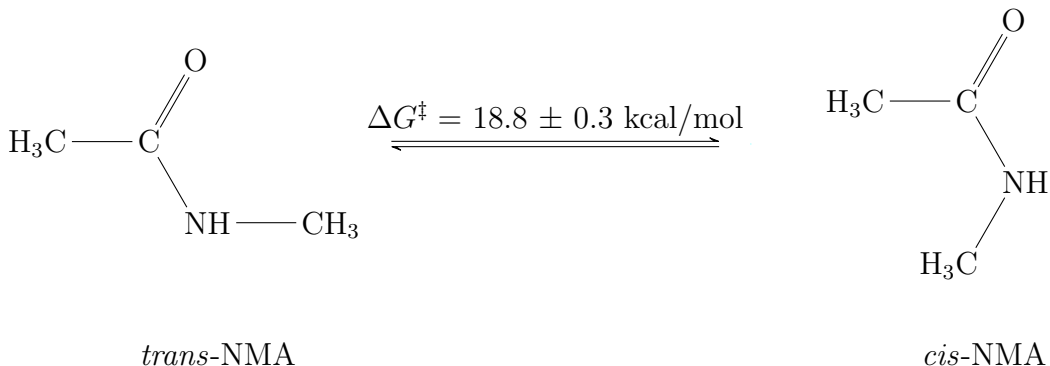
The *N*-methylacetamide molecule starting from its *cis-* conformation was simulated using ART in its default setting and using the RHF/6-31G model chemistry. ART was able to find the transition state connecting the *cis-* conformation to the *trans-*conformation. It took 10,371 force evaluations for the same and there were 70 failed events in the process. The energy barrier was found to be 26.4 kcal/mol (see Figure 2.8) which was higher than the literature value of 18.8 kcal/mol [45]. The deviation from the literature value could be attributed to the fact that the literature value is an experimental value in aqueous solution. And solvents usually tend to lower the energy barrier. The conformations found by ART were validated via a frequency calculation and it was found that the minima had all real frequencies while the transition state exactly one imaginary frequency as shown in the figure below:
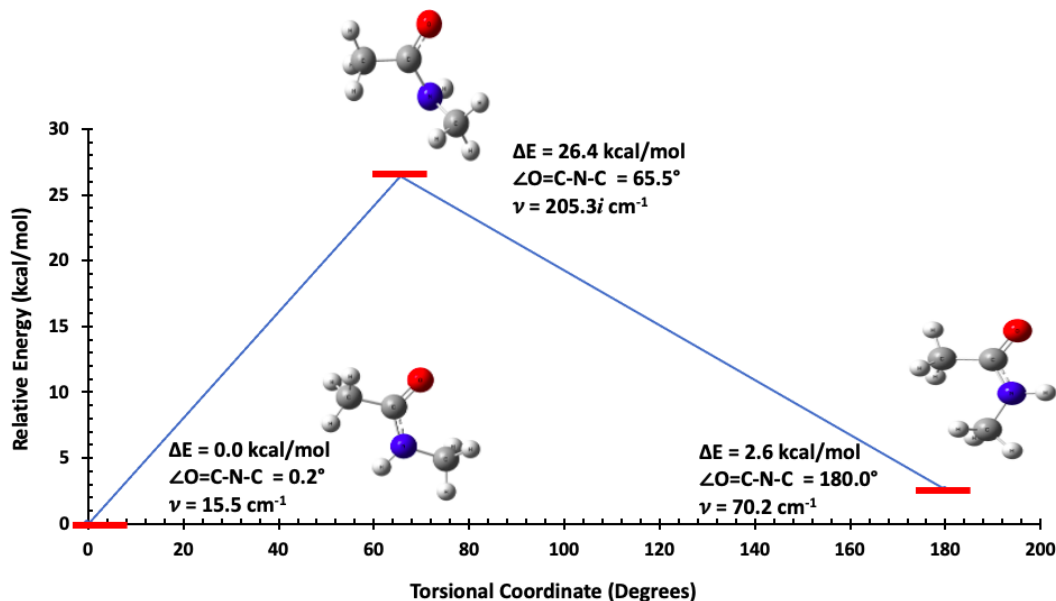
Figure 2.8: Energy diagram representing the *N*-methylacetamide conformations found by ART at RHF/6-31G model chemistry. Table A.2 in the appendix shows the total electronic energies of these conformations.

3. **Cyclohexane**: The conformational landscape for the ring inversion in cyclohexane was explored using ART in its default setting at RHG/6-31G. It took ART 160,655 force evaluations to discover all the conformations corresponding to the cyclohexane ring inversion and there were 590 failed events in the process. Starting from a cyclohexane chair conformation, ART was able to find a twist-boat conformation connected via a half-chair conformation. The rotational energy barrier to reach twist-boat conformation starting from a chair was found to be 12.2 kcal/mol. This comprised one ART event. Squillacote *et al.* [46] have shown that the $\Delta G^{\ddagger}$ for the conversion of cyclohexane chair to twist-boat via a half-chair is 10.8 kcal/mol. Now, starting from the twist-boat conformation, ART was able to find another twist-boat which was mirror image to the first twist boat but non-superimposable. The transition state connecting the two twist-boats was a boat conformation. The energy barrier associated with this event was a relatively low-lying one at 1.1 kcal/mol. Starting from the twist-

boat conformation product of the second event, ART was able to locate another chair conformation which was inverted with respect to the original chair conformation. The relative positions of the hydrogen atoms in the two chairs are inverted. The hydrogens axial to the ring in the first chair become equatorial in the second chair and vice-versa. The transition state connecting the twist-boat and the chair was a half-chair which was enantiomeric to the first-half chair. The energy barrier to go from the twist-boat to the chair via the half-chair was found to be 5.7 kcal/mol. According to Squillacote *et al.* [46], the $\Delta G^{\ddagger}$ for the twist-boat to chair isomerization via a half chair was found to be 5.3 kcal/mol.

Thus, ART was able to successfully explore the entire conformational landscape representing the ring inversion process in cyclohexane. Figure 2.9 shows the energy level diagram for this process indicating the relative energies of all the states found during the PES exploration. Table A.3 in the appendix shows all the conformations with their absolute energies and geometries and also the lowest frequency of each conformation. The frequency calculation was done in order to validate that the energy minima found along the PES are indeed true minima and the transition states indeed true transition states (first-order saddle points). It was seen that the two chair conformations each had all real frequencies and so did the two twist-boats as these states all represented energy minima on the PES. On the other hand, the two half-chair conformations and the one boat conformation all had exactly one imaginary frequency each proving that these states are indeed first-order saddle points (transition states) along the PES.
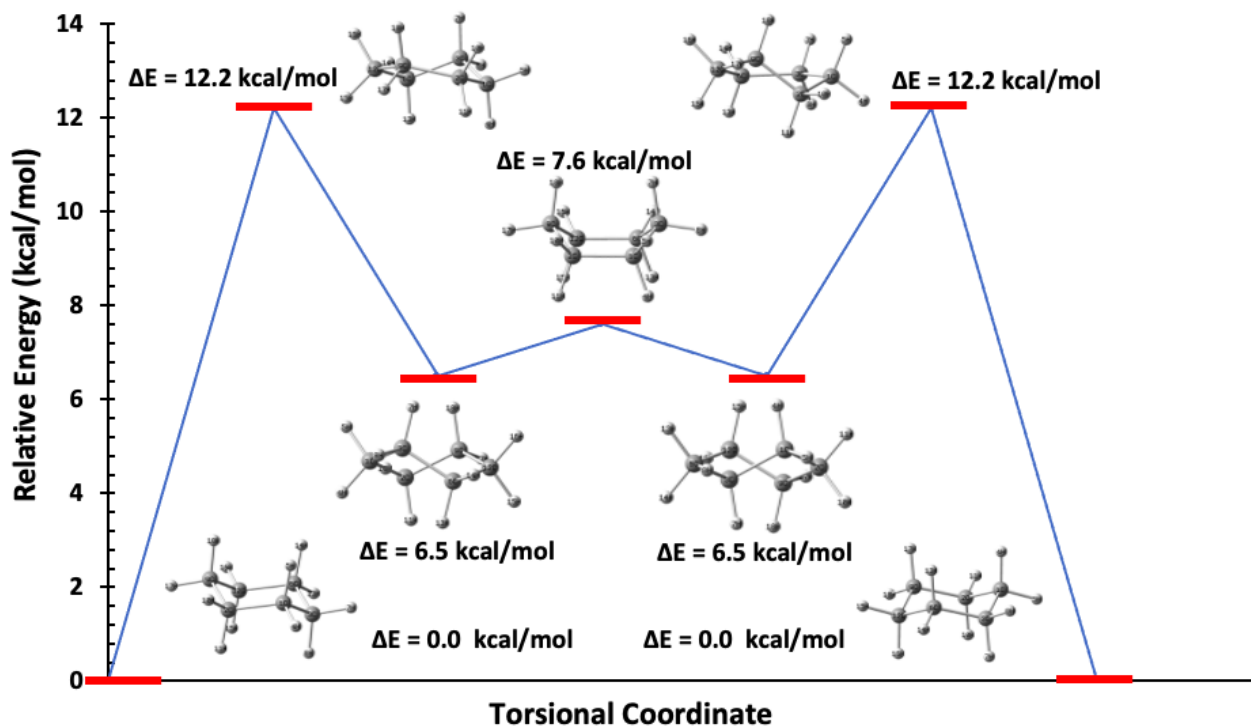
Figure 2.9: Energy diagram representing the cyclohexane conformations found by ART at RHF/6-31G model chemistry. Table A.3 in the appendix shows the total electronic energies of these conformations.
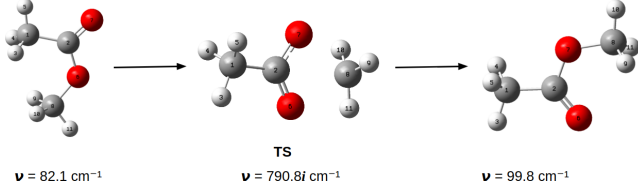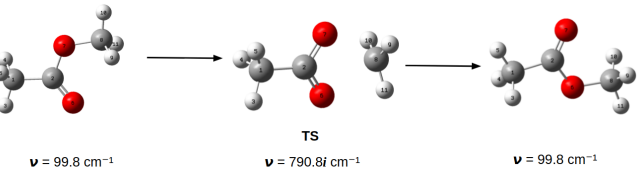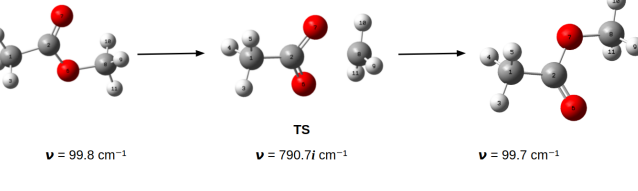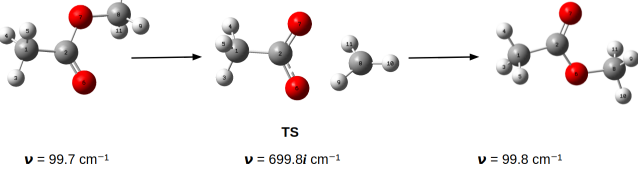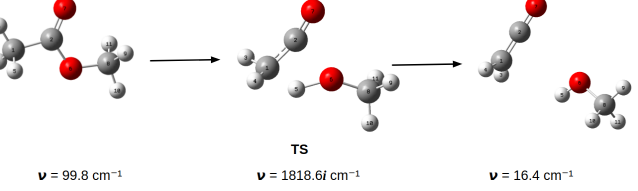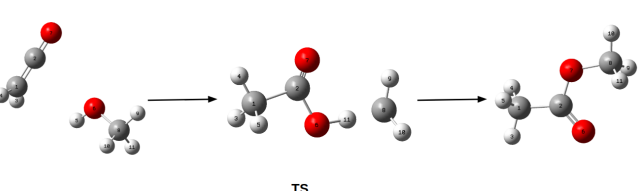
2.5.2 Reactive Landscape Exploration

1. **Methyl acetate**: The ART simulation for methyl acetate was run using the RHF/6-31G model chemistry. The maximum energy to which ART could climb was set at 184.4 kcal/mol (8 eV). The ART code was used in its default setting and six events were found where reactivities were observed (see Table 2.1). It took 200,841 force evaluations to find the conformations corresponding to these events and there were 1,598 failed events in the process. In the first such event, starting from a *cis-* conformation of methyl acetate, the methoxy group fragments leading to the formation of a methyl radical. The methyl group then attaches to the carbonyl oxygen. Thus, fragmentation and subsequent bond rearrangement are involved in this step. The energy barrier

associated with this event was 57.9 kcal/mol. The second event, starting from the product of the first event, leads to the methyl group fragmenting from the carbonyl oxygen and reattaching to the original methoxy oxygen leading to the formation of a *trans-* methyl acetate. The energy barrier associated with the event was 66.7 kcal/mol. The third event, starting from a *trans-* methyl acetate, undergoes a fragmentation of the methoxy group leading to the formation of a methyl radical which then attaches to the carbonyl oxygen. The energy barrier associated with the event was 66.4 kcal/mol. In the fourth event, the methyl group attached to the oxygen cleaves leading to the formation of a methyl radical and then binds to the other oxygen. In the fifth event, a ketene and methanol were formed as the end products. The ketene being a very unstable species places the energy barrier of the event very high at 98.0 kcal/mol. In the sixth event, starting from the ketene and methanol as a reactant complex, a *cis-* methyl acetate is formed via a transition state that comprised of acetic acid and methylene fragments. The energy barrier associated with this event is rather high at 164.20 kcal/mol and that is due to the carbene (methylene) in the transition state. A carbene is a highly unstable molecule and that is contributing to the high energy barrier.

Even though all the structures generated by ART during these six events with the exception of the transition state for Event 4 (which was found to be a second-order saddle point) when validated using a frequency calculation were found to be true minima and true transition states, on visual inspection, one can tell that the events are not representing elementary reactions. In Event 6, for example, there is a lot of simultaneous bond formation and breakage which is not possible in one elementary step. There is a very high likelihood of there being intermediate states in between the reactants and the products for Event 6 that were not found by ART.

Table 2.1: ART simulations for methyl acetate exhibiting reactive events

Event	Reaction	Energy Barrier (kcal/mol)
Event 1	ν = 82.1 cm⁻¹ TS ν = 790.8i cm⁻¹ ν = 99.8 cm⁻¹	57.9
Event 2	ν = 99.8 cm⁻¹ TS ν = 790.8i cm⁻¹ ν = 99.8 cm⁻¹	66.7
Event 3	ν = 99.8 cm⁻¹ TS ν = 790.7i cm⁻¹ ν = 99.7 cm⁻¹	66.4
Event 4	ν = 99.7 cm⁻¹ TS ν = 699.8i cm⁻¹ ν = 99.8 cm⁻¹	103.6
Event 5	ν = 99.8 cm⁻¹ TS ν = 1818.6i cm⁻¹ ν = 16.4 cm⁻¹	98.0
Event 6	ν = 16.4 cm⁻¹ TS ν = 29.1i cm⁻¹ ν = 99.8 cm⁻¹	164.2

34

2. **_N_-methylacetamide**: NMA was simulated using ART in its default setting with 230.6 kcal/mol (10 eV) being the maximum that ART could climb in terms of energy. The model chemistry used was RHF/6-31G. Table 2.2 shows the events found by this ART simulation and their respective energy barriers. It took 82,344 force evaluations to find the conformations corresponding to these 5 events and there were 529 failed events in the process. The first event was an isomerization of _cis_-_N_-methylacetamide to _trans_-_N_-methylacetamide with an energy barrier of 24.1 kcal/mol. The second event involved a methyl group transferring from the nitrogen atom to the oxygen atom resulting in a fairly high energy barrier of 91.6 kcal/mol. The third event, starting from the product of the previous one, not only led to the transfer of the methyl group from the carbonyl oxygen back to nitrogen but also led to the transfer of a hydrogen atom from the acyl methyl to nitrogen. The energy barrier associated with the event was 82.2 kcal/mol. In the fourth event, a hydrogen atom was transferred from the nitrogen to the acyl methylene while the methyl group attached to the nitrogen cleaved and bonded with the oxygen atom. The fifth event saw a transfer of a hydrogen atom from the acyl methyl group to the nitrogen atom. The methyl group bonded to the carbonyl oxygen also got transferred to the nitrogen atom. The third and fifth events, although visually similar, were energetically different with the fifth event having a higher energy barrier of 114.4 kcal/mol.

Table 2.2: ART simulations for *N*-methylacetamide exhibiting reactive behaviour. The energies were estimated using the RHF/6-31G model chemistry.

Event	Reaction	Energy Barrier (kcal/mol)
Event 1	ν = 70.9 cm^{-1} TS ν = 204.5i cm^{-1} ν = 18.2 cm^{-1}	24.1
Event 2	ν = 18.2 cm^{-1} TS ν = 824.9i cm^{-1} ν = 142.9 cm^{-1}	91.6
Event 3	ν = 142.9 cm^{-1} TS ν = 2160.0i cm^{-1} ν = 80.3 cm^{-1}	82.2
Event 4	ν = 80.3 cm^{-1} TS ν = 825.0i cm^{-1} ν = 142.9 cm^{-1}	34.7
Event 5	ν = 142.9 cm^{-1} TS ν = 2160.1i cm^{-1} ν = 80.3 cm^{-1}	114.4

Again, just like in the case of methyl acetate, even though the structures found during these events when validated using a frequency calculation were found to be valid, however, on visual inspection, some of these events do not seem to be representing elementary reactions. In Events 3, 4, and 5, for instance, the transition state does not seem to be directly leading to the reactants and the products on either side of it. It must go through some intermediate – which is not found by ART.

ART finding events that are not resolving the reaction into its elementary steps is definitely a limitation of this method and something that needs improving. A more efficient validation procedure is needed to be used that can help detect when the transition state is not connecting the minima on either side of it.

2.6   Analysis of ART simulations

The maximum energy parameter to which an ART simulation could climb was set at 5 different values: 46.1 kcal/mol (2 eV), 92.2 kcal/mol (4 eV), 138.4 kcal/mol (6 eV), 184.4 kcal/mol (8 eV), and 230.6 kcal/mol (10 eV). The ART simulations for methyl acetate and N-methylacetamide were run at these five parameter values. An analysis was done in terms of the maximum energy that ART could climb to and observation of reactive events.

It was seen that for methyl acetate, reactive events were observed starting from 138.4 kcal/mol onwards. At 46.1 kcal/mol and 92.2 kcal/mol, only conformational changes were seen for methyl acetate. These conformational changes entailed conversion from *cis*-conformation to *trans*- conformation and vice-versa. The reactive events from 138.4 kcal/mol onwards included reactions like fragmentation, rearrangement and new bond formation. Highly unstable species such as ketenes, carbenes, and radicals were witnessed. Products such as methanol and acetic acid were formed as well.

For N-methylacetamide, however, the reactive events were only witnessed starting from 184.4 kcal/mol onwards. During the ART runs at 46.1 kcal/mol, 92.2 kcal/mol, and

138.4 kcal/mol, only conformational changes involving the *cis/trans-* isomerization of *N*-methylacetamide were seen.

Since in *N*-methylacetamide, the CONH bond has a partial double bond character, it is difficult to break and consequently requires more energy. This is one plausible reason why in the energy analysis, *N*-methylacetamide is seen to start fragmenting at higher energy barriers as opposed to methyl acetate.

Table 2.3 below shows the energy cut-off for the fragmentation behaviour for both methyl acetate and *N*-methylacetamide. As can be seen from the table, fragmentation of molecular species was observed from 6 eV onwards while for *N*-methylacetamide, it was observed from 8 eV onwards.

Table 2.3: Dependence of fragmentation behaviour on the ART energy barrier threshold

Energy (kcal/mol)	Fragmentation	
	Methyl acetate	*N*-methylacetamide
46.1	✗	✗
92.2	✗	✗
138.4	✔	✗
184.4	✔	✔
230.6	✔	✔

Chapter 3

Summary and Conclusion

This work proposes a method that could potentially be used to successfully explore the potential energy landscapes of complex organic and enzymatic reactions. The ART method, an already existing algorithm, was successfully coupled with *ab initio* packages Gaussian and CP2K. The algorithm coupled with these methods was able to uncover the PESs of small organic molecules, the likes of methyl acetate, *N*-methylacetamide, and cyclohexane. The strategies incorporated in the code to speed up the process of the exploration of PESs not only managed to bring down the computational cost in terms of the force evaluations needed making the process less time intensive but also ensured a full exploration of the landscape by cutting down on moves that resulted in the exploration of the same terrain over and over again. The *follow* strategy, as envisioned, helped a structure follow the pathway traced by another structure that is structurally similar to it. The *avoid* strategy was successful in ensuring that the regions of the PES already explored are not revisited and newer avenues along the PES are explored. The strategies that exploited an *a priori* knowledge of the structures were also successful in cutting down on force evaluations and consequently the computational cost by eliminating any atomic displacement moves that did not at all contribute to the tracing of the energy surface of those structures. The *dihedral rotation* strategy was shown as constructing a potential energy surface comprising solely of torsional rotations and disallowing any bond stretching moves. It was seen that for methyl acetate, when the threshold to which ART could climb in terms of energy was set to 138.36 kcal/mol (6 eV), reactive events were observed. Such reactive events included bond breakage, new bond formation, rearrangement of atoms, etc. Formation of highly unstable species such as a ketene and methyl radicals were found as a result of the methyl acetate molecule undergoing bond stretching and rearrangement. The transfer of a methyl group

was observed which can be seen as a rearrangement reaction. End products like methanol were also formed which exhibited new bond formation. When the maximum energy that ART could climb to during the simulations was kept lower at 92.24 kcal/mol (4 eV), only conformational changes were observed for methyl acetate, *i.e.,* ones resulting in *cis-/trans* conversion of methyl acetate. The validation of these methyl acetate structures found during the ART runs at various energy barriers was done using a frequency check and it was seen that the minima were indeed true minima – having all real frequencies – and the transition states were indeed true transition states – representing first-order saddle points and thereby having exactly one imaginary frequency. For *N*-methylacetamide, the reactive events were observed as soon as ART was allowed to climb to an energy of 184.48 kcal/mol (8 eV). The acetamide molecule fragmenting using much higher energy compared to methyl acetate is in agreement with the fact that the CONH bond in acetamide is harder to break owing to its partial double bond character. Rearrangement reactions like a methyl group transferring from an oxygen atom to a nitrogen atom or vice versa were observed. Formation of products like methane and methyl isocyanate was seen. Besides, *N*-methylacetamide and methyl acetate, a ring structure like cyclohexane was also studied using ART. All the conformations featuring in the cyclohexane ring inversion process were found using ART.

In conclusion, the modified ART algorithm was successfully able to tackle small organic molecules and this shows its potential to explore the energy landscape of large biological molecules as well – which is the ultimate vision of this project. The algorithm being open-ended is really adept at handling chemical or biological systems where the end-products of a reaction are not known. This could be true in case of enzymes where the products resulting from the enzyme-substrate reaction might not be known beforehand. Thus, an open-ended method like ART has an obvious advantage in such a case over double-ended methods like the nudged elastic band as it can attempt constructing an energy pathway from just the initial energy minimum – reactant or reaction complex. Even when the end-products of an enzyme-catalyzed reaction are known, the energy barrier associated with

the catalysis can be different for a mutant than for a wild-type enzyme. If the energy barrier for a mutant-catalyzed reaction is lower, then that simply implies that the reaction is energetically more feasible and will happen faster, thus making the mutation desirable in case the design objective is to lower the energy barrier. Thus, it necessitates the need for methods that can quickly explore the energy surface associated with the reactions catalyzed by enzyme mutants. Our modified ART method stands out in relation to other methods for quickly exploring the energy surface because of its ability to share topological information about the energy surface of one system (e.g. a wild-type enzyme) with other similar systems (e.g. mutants) or even within the same system. This sharing of information not only helps quickly explore the energy terrain of the system but also ensures a full exploration without any features of the surface being explored repetitively.

The ART method, at this point of time, also stands coupled with CP2K – an *ab initio* package that is capable of performing QM/MM simulations of enzymatic reactions whereby regions around the active site of the enzyme can be treated using very accurate quantum mechanics methods while the rest of the macromolecule can be treated using molecular mechanics. The future work of this project would thus focus on taking the algorithm a step further and embark on the exploration of these massive biological molecules which are undergoing complex reactions. This study has laid all the foundation for that ambition and has highlighted the successes and pitfalls of this algorithm and how those pitfalls can be overcome.

Chapter 4

Future Work

Future work would entail applying the ART method to the study of biological molecules, e.g. enzyme-substrate reactions where the objective would be either to ascertain the products of the reaction catalyzed by mutated enzymes or to investigate the effect of a mutation on enzyme catalysis in terms of energetics of the reaction or its effect on substrate specificity, *i.e.,* whether the mutated enzyme is specific to catalyzing the conversion of just a single substrate into products or whether it shows promiscuous behaviour – catalyzing a wide variety of substrates. As a first step in that direction, it would be judicious to start with the study of small peptides like trialanine (see Figure 4.1). If ART can study the conformational energy landscape of this peptide, then it will lay the ground for it to be used to the study of larger proteins. The *dihedral rotation* strategy can be used for this purpose so that any stretching moves are disallowed and just allowing a gradual rotation of the ϕ, ψ angles in the tripeptide.
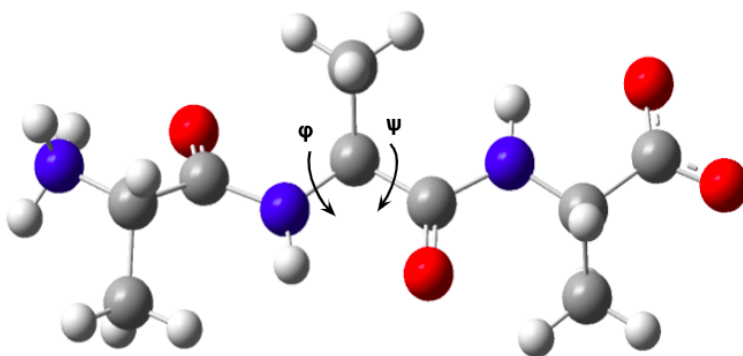

Figure 4.1: The figure shows a trialanine molecule that can act as a good model system for the application of ART to proteins.

The *focussed* strategy can be brought into play when exploring the energy

landscape of enzyme-substrate reactions via QM/MM analysis. The residues in the active site of the enzyme can be focussed upon by the strategy, thus reducing the dimensionality of a large system. This could not only cut down on force evaluations, thereby reducing computational cost, but also help make the process faster. Figure 4.2 shows the active site of CYP450$_{\text{BM3}}$.
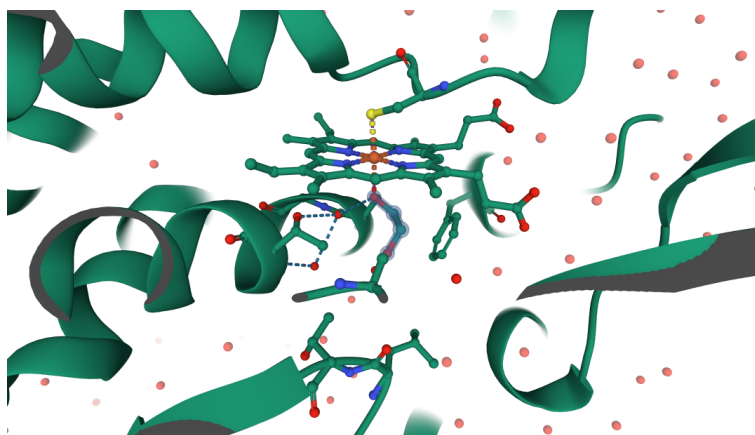

Figure 4.2: The figure shows the active site of the wild-type CYP450$_{\text{BM3}}$ (PDB ID: 1BVY). The atoms of the residues directly involved in the enzyme-substrate reactions can be found in the site and those atoms can be focussed upon.

Once the energy surface for an enzyme-substrate reaction has been explored by ART, the algorithm can then be used to explore in parallel the energy landscape of the reaction of the substrate with the mutants. The information sharing strategies incorporated into the algorithm can help achieve this. The interesting topological features in the energy terrain of one of the mutants or wild-type can be shared with the other mutant systems. The sharing of information can make the process of exploration much faster. The pathways corresponding to the reaction of CYP450$_{\text{BM3}}$ and its mutants with fatty acids such as palmitic acid can be quickly explored in this fashion.

We saw in the case of methyl acetate and *N*-methylacetamide that ART can find events that are not representative of elementary reactions, *i.e.,* a reaction where the

transition state is leading to the two minima found by ART on either side of it. The post-processing analysis that we are currently performing to validate the events found by ART are just checking whether the transition state is a true first-order saddle point or the minima are true minima by computing their frequencies and comparing the structures with the Gaussian-optimized structures. However, the validation procedure does not tell whether the transition state found is indeed connecting the two minima on either side of it. For example, Figure 4.3 shows an illustration of a hypothetical ART event where the transition state is not connecting the minima on either side and thus not resolving the reaction in terms of its elementary steps. We would like to improve the validation procedure in the future by incorporating a Gaussian intrinsic reaction coordinate (IRC) calculation in the procedure. Such an IRC calculation, starting from the transition state, helps construct both a forward and reverse reaction pathway leading to the products and reactants respectively. If the products and reactants found by the IRC calculation superimpose fairly reasonably with those found by ART, it will imply that the transition state found by ART is indeed connecting the two minima on either side of it. However, if the IRC calculation leads to a reactant or product completely different from the one found by ART, then that would mean that the event leading from the reactants to the products via the transition state corresponded to a non-elementary reaction. Now, in order to find all the intermediates in such a non-elementary reaction would entail modifying the ART algorithm in such a fashion that the optimization step performed after the configuration is displaced slightly from the saddle point does not make the configuration straightaway plummet to the products but undergoes a gradual optimization, *i.e.,* first optimizing to a any intermediate along the way.
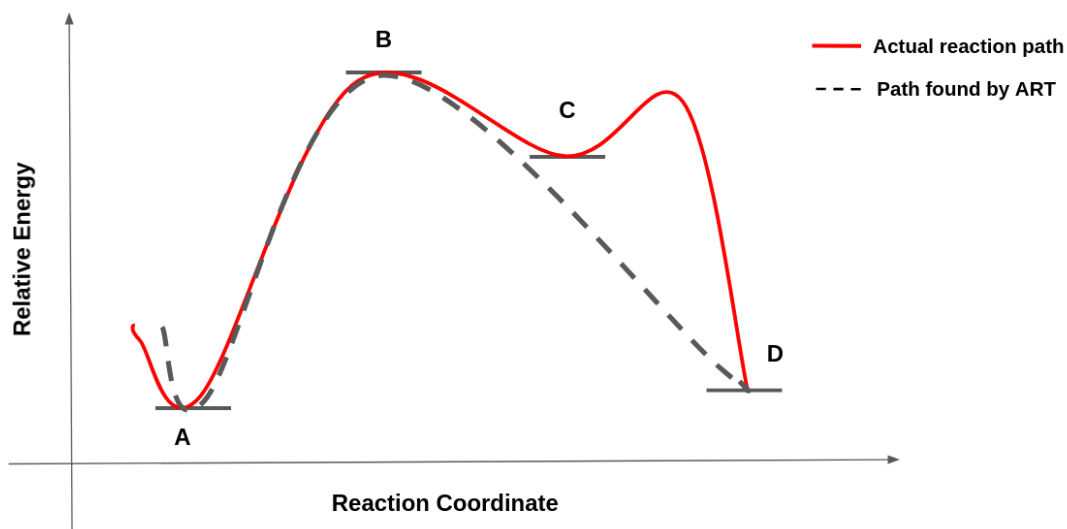
Figure 4.3: A speculation that some of the events found by ART might not be representative of an elementary reaction. In this case, for example, the ART simulation failed to find intermediate C.

References

[1] P. Bhasi, Z. P. Nhlabatsi, and S. Sitha, "Expanding the Applicability of Electrostatic Potentials to the Realm of Transition States," *Physical Chemistry Chemical Physics*, vol. 18, no. 18, pp. 13002–13009, **2016**.

[2] L. D. Jacobson, A. D. Bochevarov, M. A. Watson, T. F. Hughes, D. Rinaldo, S. Ehrlich, T. B. Steinbrecher, S. Vaitheeswaran, D. M. Philipp, M. D. Halls, and R. A. Friesner, "Automated Transition State Search and its Application to Diverse Types of Organic Reactions," *Journal of Chemical Theory and Computation*, vol. 13, no. 11, pp. 5780–5797, **2017**.

[3] A. M. Sawayama, M. M. Chen, P. Kulanthaivel, M.-S. Kuo, H. Hemmerle, and F. H. Arnold, "A Panel of Cytochrome P450 BM3 Variants to Produce Drug Metabolites and Diversify Lead Compounds," *Chemistry–A European Journal*, vol. 15, no. 43, pp. 11723–11729, **2009**.

[4] A. Hegde, D. C. Haines, M. Bondlela, B. Chen, N. Schaffer, D. R. Tomchick, M. Machius, H. Nguyen, P. K. Chowdhary, L. Stewart, C. Lopez, and J. A. Peterson, "Interactions of Substrates at the Surface of P450s can Greatly Enhance Substrate Potency," *Biochemistry*, vol. 46, no. 49, pp. 14010–14017, **2007**.

[5] E. Machado-Charry, L. K. Béland, D. Caliste, L. Genovese, T. Deutsch, N. Mousseau, and P. Pochet, "Optimized Energy Landscape Exploration Using the *ab initio* Based Activation-Relaxation Technique," *The Journal of Chemical Physics*, vol. 135, no. 3, p. 034102, **2011**.

[6] G. Barkema and N. Mousseau, "Event-Based Relaxation of Continuous Disordered Systems," *Physical Review Letters*, vol. 77, no. 21, p. 4358, **1996**.

[7] R. Malek and N. Mousseau, "Dynamics of Lennard-Jones Clusters: A Characterization of the Activation-Relaxation Technique," *Physical Review E*, vol. 62, no. 6, p. 7723, **2000**.

[8] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, "Gaussian 09 Revision E.01." Gaussian Inc. Wallingford CT **2009**.

[9] J. Hutter, M. Iannuzzi, F. Schiffmann, and J. VandeVondele, "CP2K: Atomistic Simulations of Condensed Matter Systems," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 4, no. 1, pp. 15–25, **2014**.

[10] N. Tukachev, V. Bataev, and I. Godunov, "Conformational Analysis of *N*-methylacetamide Molecule in the Ground and Excited Electronic States," *Computational and Theoretical Chemistry*, vol. 1113, pp. 82–93, **2017**.

[11] B. Peters, A. Heyden, A. T. Bell, and A. Chakraborty, "A Growing String Method for Determining Transition States: Comparison to the Nudged Elastic Band and String Methods," *The Journal of Chemical Physics*, vol. 120, no. 17, pp. 7877–7886, **2004**.

[12] J. Doye and D. Wales, "Surveying a Potential Energy Surface by Eigenvector-Following," *Zeitschrift für Physik D Atoms, Molecules and Clusters*, vol. 40, no. 1, pp. 194–197, **1997**.

[13] G. Henkelman, B. P. Uberuaga, and H. Jónsson, "A Climbing Image Nudged Elastic Band Method for Finding Saddle Points and Minimum Energy Paths," *The Journal of Chemical Physics*, vol. 113, no. 22, pp. 9901–9904, **2000**.

[14] A. Pedersen, S. F. Hafstein, and H. Jónsson, "Efficient Sampling of Saddle Points with the Minimum-Mode Following Method," *SIAM Journal on Scientific Computing*, vol. 33, no. 2, pp. 633–652, **2011**.

[15] G. Henkelman and H. Jónsson, "A Dimer Method for Finding Saddle Points on High Dimensional Potential Surfaces Using Only First Derivatives," *The Journal of Chemical Physics*, vol. 111, no. 15, pp. 7010–7022, **1999**.

[16] S. Fischer and M. Karplus, "Conjugate Peak Refinement: An Algorithm for Finding Reaction Paths and Accurate Transition States in Systems with Many Degrees of Freedom," *Chemical Physics Letters*, vol. 194, no. 3, pp. 252–261, **1992**.

[17] I. V. Ionova and E. A. Carter, "Ridge Method for Finding Saddle Points on Potential Energy Surfaces," *The Journal of Chemical Physics*, vol. 98, no. 8, pp. 6377–6386, **1993**.

[18] M. J. Dewar, E. F. Healy, and J. J. Stewart, "Location of Transition States in Reaction Mechanisms," *Journal of the Chemical Society, Faraday Transactions 2: Molecular and Chemical Physics*, vol. 80, no. 3, pp. 227–233, **1984**.

[19] M. Jafari and P. M. Zimmerman, "Reliable and efficient reaction path and transition state finding for surface reactions with the growing string method," *Journal of Computational Chemistry*, vol. 38, no. 10, pp. 645–658, **2017**.

[20] D. Sheppard, R. Terrell, and G. Henkelman, "Optimization Methods for Finding Minimum Energy Paths," *The Journal of Chemical Physics*, vol. 128, no. 13, p. 134106, **2008**.

[21] M. Plasencia Gutiérrez, C. Argáez, and H. Jónsson, "Improved Minimum Mode Following Method for Finding First Order Saddle Points," *Journal of Chemical Theory and Computation*, vol. 13, no. 1, pp. 125–134, **2017**.

[22] C. Lanczós, *Applied Analysis.* Courier Corporation, **1988**.

[23] N. Govind, M. Petersen, G. Fitzgerald, D. King-Smith, and J. Andzelm, "A Generalized Synchronous Transit Method for Transition State Location," *Computational Materials Science*, vol. 28, no. 2, pp. 250–258, **2003**.

[24] X. Sun, T. M. Soini, J. Poater, T. A. Hamlin, and F. M. Bickelhaupt, "PyFrag 2019—Automating the Exploration and Analysis of Reaction Mechanisms," *Journal of Computational Chemistry*, vol. 40, no. 25, pp. 2227–2233, **2019**.

[25] S. Maeda, Y. Harabuchi, Y. Ono, T. Taketsugu, and K. Morokuma, "Intrinsic Reaction Coordinate: Calculation, Bifurcation, and Automated search," *International Journal of Quantum Chemistry*, vol. 115, no. 5, pp. 258–269, **2015**.

[26] W.-J. van Zeist and F. M. Bickelhaupt, "The Activation Strain Model of Chemical Reactivity," *Organic & Biomolecular Chemistry*, vol. 8, no. 14, pp. 3118–3127, **2010**.

[27] G. N. Simm, A. C. Vaucher, and M. Reiher, "Exploration of Reaction Pathways and Chemical Transformation Networks," *The Journal of Physical Chemistry A*, vol. 123, no. 2, pp. 385–399, **2019**.

[28] G. Wei, N. Mousseau, and P. Derreumaux, "Exploring the Energy Landscape of Proteins: A Characterization of the Activation-Relaxation Technique," *The Journal of Chemical Physics*, vol. 117, no. 24, pp. 11379–11387, **2002**.

[29] M.-C. Marinica, F. Willaime, and N. Mousseau, "Energy Landscape of Small Clusters of Self-Interstitial Dumbbells in Iron," *Physical Review B*, vol. 83, no. 9, p. 094119, **2011**.

[30] T. Lynch and A. Price, "The Effect of Cytochrome P450 Metabolism on Drug Response, Interactions, and Adverse Effects," *American Family Physician*, vol. 76, pp. 391–6, **2007**.

[31] A. Seifert, S. Vomund, K. Grohmann, S. Kriening, V. B. Urlacher, S. Laschat, and J. Pleiss, "Rational Design of a Minimal and Highly Enriched CYP102A1 Mutant Library with Improved Regio-, Stereo-and Chemoselectivity," *Chembiochem*, vol. 10, no. 5, pp. 853–861, **2009**.

[32] K. G. Ravichandran, S. S. Boddupalli, C. Hasermann, J. A. Peterson, and J. Deisenhofer, "Crystal Structure of Hemoprotein Domain of P450BM-3, a Prototype for Microsomal P450's," *Science*, vol. 261, no. 5122, pp. 731–736, **1993**.

[33] H. Li and T. L. Poulos, "The Structure of the Cytochrome P450BM-3 Haem Domain Complexed with the Fatty Acid Substrate, Palmitoleic acid," *Nature Structural and Molecular Biology*, vol. 4, no. 2, p. 140, **1997**.

[34] M. A. Noble, C. S. Miles, S. K. Chapman, D. A. Lysek, A. C. MacKay, G. A. Reid, R. P. Hanzlik, and A. W. Munro, "Roles of Key Active-Site Residues in Flavocytochrome P450 BM3," *Biochemical Journal*, vol. 339, no. 2, pp. 371–379, **1999**.

[35] T. W. Ost, C. S. Miles, J. Murdoch, Y.-F. Cheung, G. A. Reid, S. K. Chapman, and A. W. Munro, "Rational Re-Design of the Substrate Binding Site of Flavocytochrome P450 BM3," *FEBS Letters*, vol. 486, no. 2, pp. 173–177, **2000**.

[36] The PyMOL Molecular Graphics System, Version 2.0, Schrödinger, LLC.

[37] D. M. Philipp and R. A. Friesner, "Mixed *ab initio* QM/MM Modeling Using Frozen Orbitals and Tests with Alanine Dipeptide and Tetrapeptide," *Journal of Computational Chemistry*, vol. 20, no. 14, pp. 1468–1494, **1999**.

[38] K. D. Dubey, B. Wang, and S. Shaik, "Molecular Dynamics and QM/MM calculations Predict the Substrate-Induced Gating of Cytochrome P450 BM3 and the Regio-and Stereoselectivity of Fatty Acid Hydroxylation," *Journal of the American Chemical Society*, vol. 138, no. 3, pp. 837–845, **2016**.

[39] W.-C. Huang, A. C. Westlake, J.-D. Maréchal, M. G. Joyce, P. C. Moody, and G. C. Roberts, "Filling a Hole in Cytochrome P450 BM3 Improves Substrate Binding and Catalytic Efficiency," *Journal of Molecular Biology*, vol. 373, no. 3, pp. 633–651, **2007**.

[40] W. Kabsch, "A Discussion of the Solution for the Best Rotation to Relate Two Sets of Vectors," *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, vol. 34, no. 5, pp. 827–828, **1978**.

[41] Y. Mo and J. Gao, "Theoretical Analysis of the Rotational Barrier of Ethane," *Accounts of Chemical Research*, vol. 40, no. 2, pp. 113–119, **2007**.

[42] C. Blom and H. H. Günthard, "Rotational Isomerism in Methyl Formate and Methyl acetate; A Low-Temperature Matrix Infrared Study Using Thermal Molecular Beams," *Chemical Physics Letters*, vol. 84, no. 2, pp. 267–271, **1981**.

[43] K. B. Wiberg and K. E. Laidig, "Barriers to Rotation Adjacent to Double Bonds. 3. The Carbon-Oxygen Barrier in Formic Acid, Methyl Formate, Acetic Acid, and Methyl Acetate. The Origin of Ester and Amide Resonance," *Journal of the American Chemical Society*, vol. 109, no. 20, pp. 5935–5943, **1987**.

[44] W. L. Jorgensen and J. Gao, "Cis-Trans Energy Difference for the Peptide Bond in the Gas Phase and in Aqueous Solution," *Journal of the American Chemical Society*, vol. 110, no. 13, pp. 4212–4216, **1988**.

[45] T. Drakenberg and S. Forsén, "The Barrier to Internal Rotation in Monosubstituted Amides," *Journal of the Chemical Society D: Chemical Communications*, no. 21, pp. 1404–1405, **1971**.

[46] M. Squillacote, R. Sheridan, O. Chapman, and F. Anet, "Spectroscopic Detection of the Twist-Boat Conformation of Cyclohexane. Direct Measurement of the Free Energy Difference Between the Chair and the Twist-Boat," *Journal of the American Chemical Society*, vol. 97, no. 11, pp. 3244–3246, **1975**.

Appendix A

Supplementary Tables for Chapter 2

Table A.1: Energies and geometries of the methyl acetate conformations found by ART calculated using the RHF/6-31G model chemistry
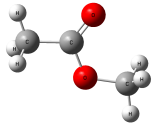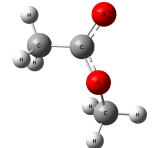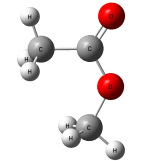
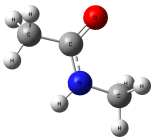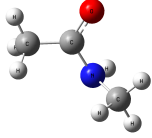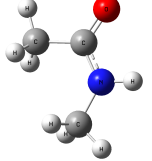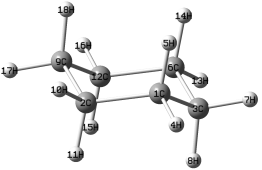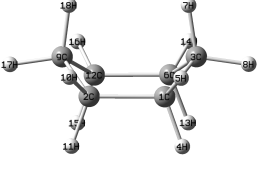| Compound	Structure	Energy (Hartrees)	Lowest Frequency (cm^{-1})	Geometry
trans- methyl acetate		-266.707346490	99.7	$\angle$ O=C–O–C = 0°
TS		-266.668302565	118.3*i*	$\angle$ O=C–O–C = 114.7°
cis- methyl acetate		-266.691013221	82.0	$\angle$ O=C–O–C = 180°

Table A.2: Energies and geometries of the *N*-methylacetamide conformations found by ART calculated using the RHF/6-31G model chemistry
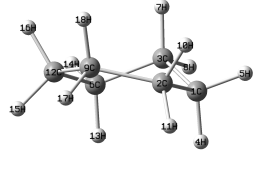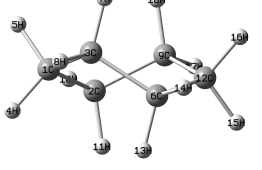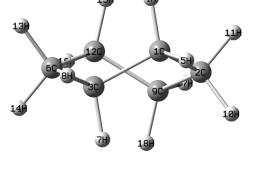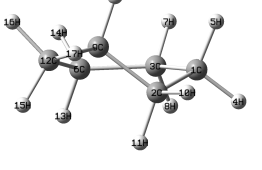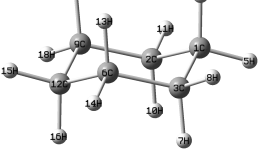
Compound	Structure	Energy (Hartrees)	Lowest Frequency (cm^{-1})	Geometry
trans- *N*- methylacetamide		-246.903921975	15.5	$\angle$ O=C–N–C = 0°
TS		-246.861779084	205.3*i*	$\angle$ O=C–N–C = 65.5°
cis- *N*- methylacetamide		-246.899817232	70.2	$\angle$ O=C–N–C = 180°

Table A.3: Energies and geometries of the cyclohexane conformations found by ART calculated using the RHF/6-31G model chemistry

Structure	Energy (Hartrees)	Lowest Frequency (cm^{-1})	Geometry
	-234.111019190	245.5	∠ 1C–2C–9C = 111.38° ∠ 2C–9C–12C = 111.43° ∠ 9C–12C–6C = 111.40° ∠ 12C–6C–3C = 111.40° ∠ 6C–3C–1C = 111.40° ∠ 3C–1C–2C = 111.45°
	-234.091615271	240.7*i*	∠ 1C–2C–9C = 117.13° ∠ 2C–9C–12C = 118.95° ∠ 9C–12C–6C = 116.63° ∠ 12C–6C–3C = 112.81° ∠ 6C–3C–1C = 109.85° ∠ 3C–1C–2C = 113.55°
	-234.100715646	115.9	∠ 1C–2C–9C = 111.07° ∠ 2C–9C–12C = 111.07° ∠ 9C–12C–6C = 112.26° ∠ 12C–6C–3C = 111.08° ∠ 6C–3C–1C = 111.12° ∠ 3C–1C–2C = 112.24°
	-234.098871363	100.4*i*	∠ 1C–2C–9C = 113.14° ∠ 2C–9C–12C = 112.47° ∠ 9C–12C–6C = 113.12° ∠ 12C–6C–3C = 113.09° ∠ 6C–3C–1C = 112.33° ∠ 3C–1C–2C = 113.08°
	-234.100715693	115.7	∠ 1C–2C–9C = 112.24° ∠ 2C–9C–12C = 111.07° ∠ 9C–12C–6C = 111.10° ∠ 12C–6C–3C = 112.19° ∠ 6C–3C–1C = 111.10° ∠ 3C–1C–2C = 111.11°
	-234.091753041	240.6*i*	∠ 1C–2C–9C = 108.39° ∠ 2C–9C–12C = 109.49° ∠ 9C–12C–6C = 114.73° ∠ 12C–6C–3C = 118.59° ∠ 6C–3C–1C = 117.54° ∠ 3C–1C–2C = 113.13°
	-234.111019163	245.4	∠ 1C–2C–9C = 111.40° ∠ 2C–9C–12C = 111.43° ∠ 9C–12C–6C = 111.38° ∠ 12C–6C–3C = 111.44° ∠ 6C–3C–1C = 111.40° ∠ 3C–1C–2C = 111.42°