Theoretical Methods and Techniques

In the work presented in the thesis, electronic structure calculations, direct dynamics simulations, and RRKM theory rate constant calculations were used to study unimolecular and bimolecular reaction dynamics. Various theoretical techniques used throughout the thesis work are briefly described in this Chapter.

2.1 ELECTRONIC STRUCTURE

By solving the electronic Schrödinger equation under the Born-Oppenheimer approximation, potential energy surface (PES) of a molecule can be determined[Levine *et al.*, 2009]. The hyper-dimensional PES describes the complicated relationship between potential energy of the molecule and the internal coordinates of the system. Motion of a molecule containing N atoms are described using $3N$ coordinates. PES depends only on the internal coordinates and invariant under translational and overall rotational motion of the molecule. Hence, the PES depends on $3N - 6$ coordinates (without the 3 translational and 3 rotational coordinates) for non-linear molecules and $3N - 5$ (without the 3 translational and 2 rotational coordinates) coordinates for linear molecules. The internal coordinates may be represented by simple stretch, bend, torsion coordinates, or symmetry-adapted linear combinations of these internal coordinates, or normal mode coordinates[Hratchian and Schlegel, 2005]. For interpretation of the experimental data and to understand chemical processes, construction of a PES is necessary[Domcke *et al.*, 2004; Zhang, 1999; Gerratt, 1985; Laganà and Riganelli, 2012; Truhlar, 2013]. However, determining the complete PES for a reaction involving many atoms is an extremely challenging, if not impossible, task. Instead, important features of a PES such as stationary points and minimum energy pathways can be determined using numerical methods[Jensen, 2017].

2.1.1 Features of PES

Potential energy $V(\vec{q})$ depends on the internal coordinates $\vec{q}(q_1, q_2, ...)$ of the system. Solving the electronic Schrödinger equation for a fixed $\vec{q}$ to obtain $V(\vec{q})$ is known as *single point energy* calculation. Stationary points on the PES are points at which the net force acting on the molecule is zero i.e.,

$$\frac{\partial V}{\partial \vec{q}} = 0 \tag{2.1}$$

Process of determining the stationary points on a PES is called *geometry optimization*[Schlegel, 1982] and is an essential tool in computational chemistry. Stationary points can be classified as local and global minimum and maximum. A global minimum is the lowest energy value in the entire PES and local minimum is the lowest energy value in a given region. Various stationary points on a PES include equilibrium geometry of a molecule, reaction intermediates, and n^{th} order saddle points. A transition state (TS) is a 1st order saddle point. An example PES is shown in Figure 2.1. At a stationary point, PES is flat with respect to coordinates. Search for a stationary point begins with a guess geometry using an algorithm such as Newton or Quasi-Newton algorithm[Schlegel,
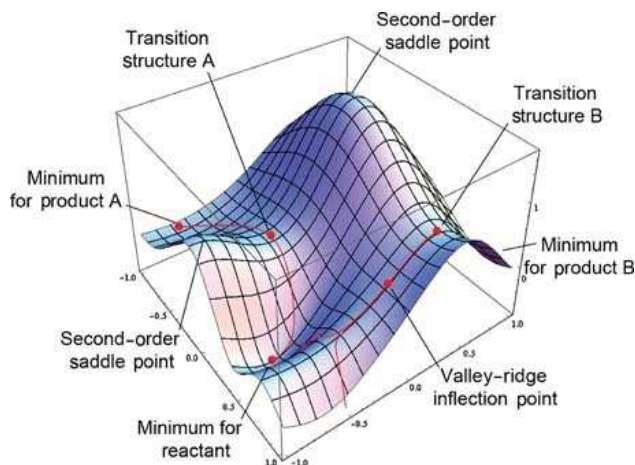
Figure 2.1: Potential energy surface[Schlegel, 2011], and its features.

2011]. Stationary point search algorithms can be classified as methods using only energy $V(\vec{q})$ and gradients $\partial V/\partial \vec{q}$ and methods based on second derivatives [Hessian $\partial^2 V/\partial q_i q_j$, $i,j = 1,2,3,...$][Scales, 1985]. Methods based on Hessian converge quickly to the target stationary point in comparison to gradient based methods, however determining the Hessian is a complicated process. Hence, gradient based methods are usually preferred in the geometry optimization process. The success of a geometry optimization process depends on how close the guess geometry is to the target geometry, optimization algorithm, and the chosen coordinate system.

At a stationary point, gradient is zero with respect to all the coordinates.

$$\frac{\partial V}{\partial q_1} = \frac{\partial V}{\partial q_2} = ... = 0 \tag{2.2}$$

After identification, the nature of a stationary point as a minimum or a saddle point can be determined using the second derivatives $\partial^2 V/\partial q_i^2$[Sheppard $et\ al.$, 2008]. At a minimum, second derivative is positive with respect to all the $3N - 6$ coordinates (or $3N - 5$ for a linear molecule) i.e., V is a minimum with respect to all the coordinates,

$$\frac{\partial^2 V}{\partial q_1^2} > 0, \quad \frac{\partial^2 V}{\partial q_2^2} > 0, ... \frac{\partial^2 V}{\partial q_{3N-6}^2} > 0 \tag{2.3}$$

This leads to $3N - 6$ real normal mode frequencies associated with the minimum (a discussion of normal modes is presented below). These points correspond to equilibrium geometries of reactants, products, and intermediates and they are the local or global minima. At a transition state, which is a first order saddle point, V is a maximum with respect to reaction coordinate[Fukui, 1970] q_{rc} and minimum with respect to all the other coordinates. Hence,

$$\frac{\partial^2 V}{\partial q_1^2} > 0, \quad \frac{\partial^2 V}{\partial q_2^2} > 0, ..., \frac{\partial^2 V}{\partial q_{rc}^2} < 0, ..., \frac{\partial^2 V}{\partial q_{3N-6}^2} > 0 \tag{2.4}$$

This leads to one imaginary and the remaining real normal mode frequencies for a first order saddle point or a transition state. For a second order saddle point, V is maximum along two coordinates and results in two imaginary frequencies. Role of second and higher order saddle points in chemical reactions is not clearly understood[Pradhan and Lourderaj, 2019].

After a TS geometry has been optimized, it is important to verify that the found TS connects the correct reactant(s) with product(s). This is done by finding the steepest descent path on both sides of the TS. Such a 'reaction path following' may also help to identify reaction

intermediates. A reaction path described in terms of a coordinate system in which all the coordinates are scaled to have the same reduced mass is known as *minimum energy path* (MEP)[Garrett and Truhlar, 1979; Kuppermann and Truhlar, 1971]. Both normal mode frequency calculations and MEP characterizations are used to identify correct TSs in electronic structure calculations. Further, frequencies are needed to compute zero point energy (ZPE) corrected stationary point energies. Relative energies computed from electronic structure calculations have to be ZPE corrected for making direct comparison to the experimental data.

2.2 AUTOMATED REACTION PATH DISCOVERY METHODS

Computational chemistry have opened the door for understanding of challenging chemical problems, using ab initio methods in conjugation with advanced potential energy surface (PES) walking algorithms (e.g., freezing string method, growing string method), which automatically locate transition states of reactions of interest. Finding saddle points, which correspond to transition states, still a challenge for systematic characterization because a transition structure optimization must step uphill in one direction and downhill in all other orthogonal directions[Suleimanov and Green, 2015]. Various methods for automated reaction pathway search have been proposed. The growing string method (GSM) is highly useful for locating reaction paths connecting two molecular intermediates. GSM approaches based on bond-forming rules, which first detect product channels of potential interest and then decide on which to apply saddle point searches. GSM has often been used in a two-step procedure to locate exact transition states (TS), where GSM creates a quality initial structure for a local TS search[Zimmerman, 2013, 2015]. The freezing-string method offers a choice between Cartesian interpolation and linear synchronous transit interpolation. It also allows to choose between conjugate gradient and quasi-Newton optimization techniques[Suleimanov and Green, 2015]. By automated reaction path discovery methods, can detect all known reaction pathways, which manually detected in the previous studies, and other new previously unknown, reaction pathways, which involve significant atom rearrangements.

2.2.1 Normal Mode Analysis

A normal mode is defined as the coherent oscillation of all atoms in a molecule maintaining same phase and frequency[Wilson *et al.*, 1980]. The $3N-6$ normal modes of a non-linear molecule (or $3N-5$ for a linear molecule) are obtained by separating out the translational and rotational motions. Under the harmonic oscillator approximation[Levine *et al.*, 2009], the vibrational frequency of a diatomic molecule is given by

$$\nu = \frac{1}{2\pi} \left(\frac{k}{\mu} \right)^{1/2} \tag{2.5}$$

where k is the force constant and μ is the reduced mass of the diatomic molecule. By treating the normal modes of a polyatomic molecule as separable harmonic oscillators, the force constant matrix or Hessian matrix H_{ij} is computed

$$H_{ij} = \frac{\partial^2 V}{\partial q_i \partial q_j} \tag{2.6}$$

where $q_1, q_2, ...$ are the normal coordinates of the system. Diagonalization of the Hessian matrix provides the $3N - 6$ force constants from which the normal mode frequencies $\nu_1, \nu_2, \nu_3, ...$ are computed. If a stationary point on a PES is a minimum, all the $3N - 6$ normal mode frequencies are real. At a TS, the reaction coordinate is a maximum and this leads to an imaginary frequency.

2.3 BORN-OPPENHEIMER DIRECT DYNAMICS

In Born-Oppenheimer direct dynamics simulations, classical equations of motion are time propagated with potentials $V(\vec{q})$ and gradients $\partial V/\partial \vec{q}$ computed on-the-fly using an appropriately selected electronic structure theory such as an *ab initio*, DFT, or a semi-empirical method[Pratihar *et al.*, 2017; Paranjothy *et al.*, 2013b]. Because of the numerous electronic structure calculations required to generate classical trajectories, direct dynamics method is computationally expensive. Due to the advancements in computer hardware and intelligent computing algorithms, direct dynamics method is becoming a popular tool for studying atomic level reaction dynamics. This method has been applied to a broad range of problems including, but not limited to, unimolecular and bimolecular reactions[Krishnan *et al.*, 2017; Xie *et al.*, 2015; Zhang *et al.*, 2010], gas-surface collision reactions[Wang *et al.*, 2003; Park *et al.*, 2009], post-TS reaction dynamics[Lourderaj *et al.*, 2008b], etc. In a direct dynamics simulation, an ensemble of classical trajectories are generated with each trajectory having different initial positions and momenta. Selection of appropriate initial conditions is a crucial part of a classical simulation and a variety of initial condition sampling algorithms are available[Peslherbe *et al.*, 1999]. Classical trajectories generated by direct dynamics are highly accurate and have been used to interpret experimental spectra and to establish atomic level reaction dynamics.
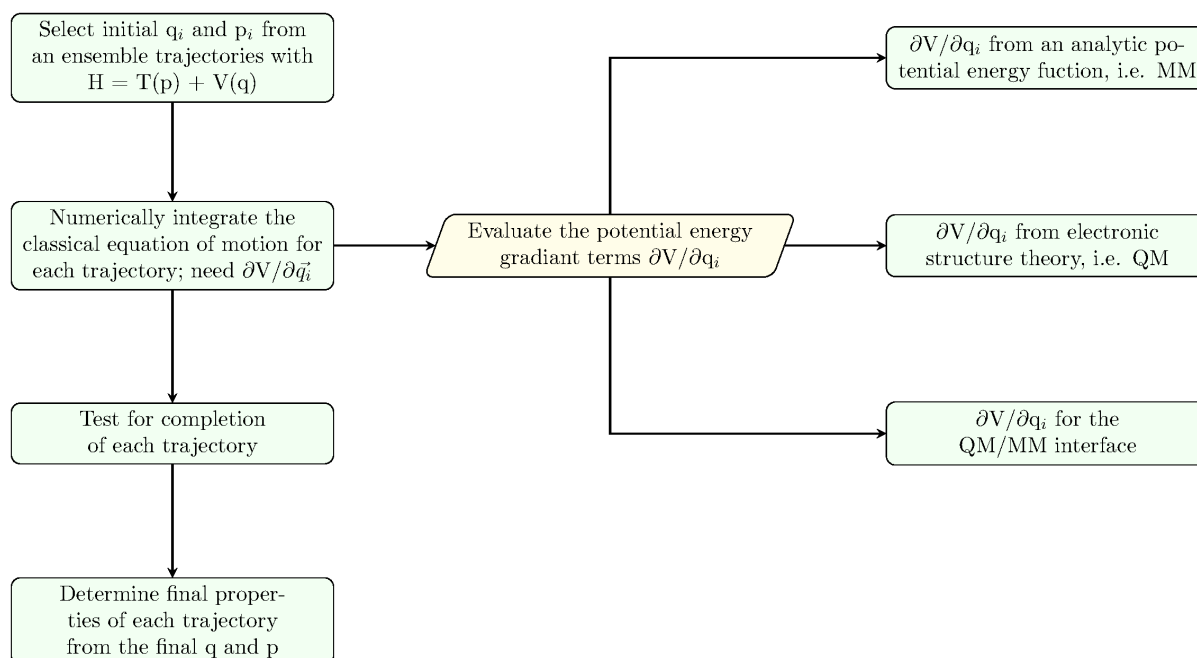

Figure 2.2: Flowchart describing the steps involved in a classical trajectory simulation.

Components of a classical trajectory simulation[Sun and Hase, 2003] are shown as a flowchart in Figure 2.2 and described below.

∘ Selection of proper initial conditions $[q_0, p_0]$ modeling an experiment or to compare with predictions by a statistical rate theory such as RRKM theory.

∘ Selection of a PES which can be analytical (molecular mechanics) or can be from quantum mechanics (direct dynamics).

∘ Integration of Newton's or Hamilton's equation of motion using an integration algorithm to obtain time evolved coordinates q_t and momenta p_t.

∘ Using q_t and p_t to analyze reaction mechanisms by direct trajectory animation and compute properties such as product energies, scattering angles, etc., to compare with experiments or a theoretical model.

2.3.1 Trajectory Initial Conditions

Selecting appropriate initial conditions is a crucial part of classical trajectory simulations. Initial conditions are selected to mimic an experiment or to compare with predictions from a rate theory such as TST or RRKM theory. There are several initial condition sampling schemes available such as classical and quantum micro-canonical sampling, thermal sampling, and sampling at the barrier[Peslherbe *et al.*, 1999; Cho *et al.*, 1992; Paranjothy *et al.*, 2013b] to model various experimental conditions. For example, classical micro-canonical sampling can be used to model gas phase experiments. Sampling schemes used to perform the direct dynamics simulations reported in the present work are briefly described below.

Classical Micro-canonical Sampling

In the classical micro-canonical sampling method[Hase and Buckowski, 1980], a molecule is assumed to be made of separable harmonic oscillators. For a system made of n normal modes, the Hamiltonian can be written as sum of energies for all separable harmonic oscillators,

$$H = E = \sum_{i=1}^{n} E_i = \sum_{i=1}^{n} \frac{P_i^2 + \omega_i^2 Q_i^2}{2} \tag{2.7}$$

Random values for the coordinates Q_i and momenta P_i are selected such that there is a uniform distribution in the constant energy shell $H(P_i, Q_i)$ in the phase space. Hence, the total energy E is a constant in all the trajectories of the ensemble. The selected energy E is added to the n normal modes of the molecule. Random values are selected for the mode energies E_i which are converted into Q_i and P_i by assigning random phase to each normal mode. The probability that mode i has energy E_i is proportional to the number of ways (which depends on the classical density of states) in which the remaining energy $E - E_i$ can be distributed among the rest of the $n-1$ oscillators. This probability is given by[Hase and Buckowski, 1980]

$$P(E_i) = \left(E - \sum_{j=1}^{i-1} Ej - Ei \right)^{n-1-i} \left\{ \int_0^{(E - \sum_{j=1}^{i-1} E_j)} \left(E - \sum_{j=1}^{i-1} Ej - Ei \right)^{n-1-i} dE_i \right\}^{-1} \tag{2.8}$$

This method has been widely used to study gas phase reaction dynamics[Godara *et al.*, 2020; Kakhiani *et al.*, 2009; Carrà *et al.*, 2019; Subramanian *et al.*, 2018]. In this sampling scheme, energy quantization and ZPE constraints are not considered. A quantum micro-canonical algorithm[Park *et al.*, 2011] can be used to account for energy quantization and ZPE effects. In the present work, only classical micro-canonical sampling method was employed. Since the total energy is a constant, this scheme can be used to study intrinsic RRKM and non-RRKM behavior associated with unimolecular dynamics[Baer and Hase, 1996].

Thermal Sampling

While micro-canonical sampling is performed at fixed total energies, a thermal sampling algorithm can be used to select initial conditions Q_0 and P_0 at fixed temperatures[Yang *et al.*, 2011]. At a vibrational temperature T_{vib}, normal mode vibrational energies are sampled by Boltzmann

distribution. Normal mode vibrational quantum numbers n_i ($i = 1, 2, 3, ... 3N - 6$) are selected randomly from a Boltzmann distribution at a given temperature. The probability distribution $P(n_i)$ for the ith mode to have vibrational quantum number n_i is given by[Levine, 2009].

$$P(n_i) = e^{-n_i h \nu_i / k_B T_{vib}}[1 - e^{h \nu_i / k_B T_{vib}}] \tag{2.9}$$

Here, ν_i is the normal mode vibrational frequency of the ith mode and k_B is the Boltzmann constant. While micro-canonical ensemble can be used to find classical $k(E)$, thermal sampling leads to $k(T)$. For large molecule with several vibrational degrees of freedom, $k(E)$ and $k(T)$ become equal at sufficiently large excitations[Lourderaj *et al.*, 2008a]. After the selection of initial conditions, they are time propagated using an integration method which is briefly discussed below.

2.3.2 Integrator

In a classical trajectory simulation, numerical integration of the Newtonian or Hamiltonian equations of motion are performed. The integrations can be performed in internal, Cartesian, or instantaneous normal mode coordinates[Wilson *et al.*, 1980; Adams and Stratt, 1990; Miller *et al.*, 1989]. Potential energy is well defined in internal coordinates but to compute kinetic energy, both coordinates and momenta are required and hence Cartesian or normal mode coordinates are more suitable. Different integration algorithms such as Gear, Adams-Moulton, Gauss-Radau, Bulirsch-Stoer, fourth-order-Range-Kutta, and sixth-order Adams-Bashforth-Moulton are used for performing integration in Cartesian coordinates[Press *et al.*, 1992; Sun and Hase, 2003; Stoer and Bulirsch, 1993; Thompson, 1998; Bolton and Nordholm, 1994]. These methods are based on predictor-corrector algorithm which is a modified Euler method. They are best suited for integrating ordinary differential equations to find unknown functions. Total energy conservation in the trajectories is not well conserved using these methods particularly for longer total integration times. For good energy conservation, methods beyond 4th order are needed. Symplectic integration methods[Gray *et al.*, 1994; Schlier and Seiter, 1998], which are based on Runge-Kutta algorithm, are a good alternate and are very good in conserving total energy in the trajectories. These methods are the best choices for direct dynamics simulations because they maintain the symplectic structure of the classical phase space $dq \wedge dp$. A comparison of energy conservation in a classical trajectory using a symplectic and a non-symplectic method is shown in Figure 2.3. Non-symplectic integrators are
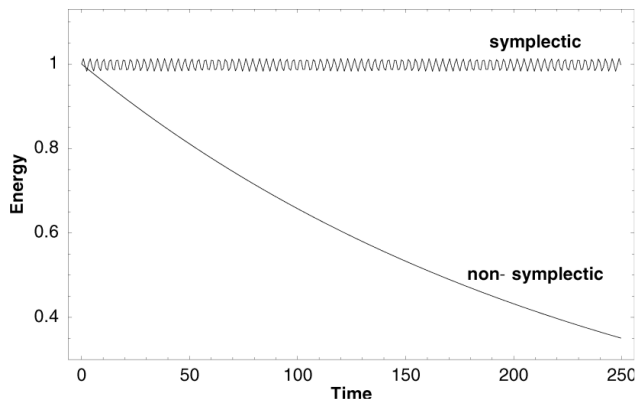

Figure 2.3: Numerical energy at each time step using a symplectic method and a non-symplectic Runge-Kutta method[Rieben *et al.*, 2004]. y axis shows energy in kcal/mol and x-axis shows time in units of 10^{-14}s.

computationally less time consuming in comparison to symplectic integrators but the total energy conservation is very good in the latter. Further, symplectic integrators are preferred in longer time integrations. The choice of the integrator depends on the size of the system, time duration, and also time-step. The idea of symplectic integrators is briefly mentioned here. Hamiltonian of a system

can be written as

$$H(q_i, p_i) = T(p_i) + V(q_i) \tag{2.10}$$

and the equations of motion are

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i} \qquad \text{and} \qquad \frac{dq_i}{dt} = \frac{\partial H}{\partial p_i} \tag{2.11}$$

For a given dynamical variable $z = (q_i, p_i)$, the Hamilton's equations can be expressed as

$$\dot{z} = \{z, H(z)\} \tag{2.12}$$

Here $\{.,.\}$ stands for the Poisson bracket. On introducing a differential operator D_H, above equation can be simplified as:

$$\dot{z} = D_H z \tag{2.13}$$

The Solution $z(t)$ from $t = 0$ to $t = \tau$ is given by a matrix exponential

$$z(\tau) = e^{(\tau D_H)} z(0) \tag{2.14}$$

Here τ is step-size.

$$z(\tau) = e^{[\tau(D_T + D_V)]} z(0) \tag{2.15}$$

Suppose, c_i and d_i are real numbers, where

$$\sum_{i=1}^{k} c_i = \sum_{i=1}^{k} d_i = 1 \tag{2.16}$$

and satisfy the given equality,

$$e^{[\tau(D_T + D_V)]} = e^{(c_i \tau D_T)} e^{(d_i \tau D_V)} + O(\tau^{n+1}) \tag{2.17}$$

Here n is the order of the integrator. For all z

$$D_T^2 z = \{\{z, T\}, T\} = \{\{\dot{q}, 0\}, T\} = (0, 0) \tag{2.18}$$

Since $\{.,.\}$ represents Poisson bracket. Hence

$$D_T^2 = 0 \tag{2.19}$$

Using a Taylor series expansion, $e^{(\alpha D_T)}$ can be expressed as

$$e^{(\alpha D_T)} = \sum_{n=0}^{\infty} \frac{(\alpha D_T)^n}{n!} \tag{2.20}$$

where α is an arbitrary real number.

$$e^{(\alpha D_T)} = 1 + \alpha D_T \tag{2.21}$$

and

$$e^{(\alpha D_V)} = 1 + \alpha D_T \tag{2.22}$$

The simplified form of equations are:

$$q_{j+1} = q_j + c_j \frac{p_j + 1}{m} t \tag{2.23}$$

$$p_{j+1} = p_j + d_i F(q_j) t \tag{2.24}$$

Here, F represents the gradient and m is the scalar quantity of mass.

In the present work, 6th order symplectic integration method and velocity-verlet (2nd order) integration method are used for trajectory integration[Schlier and Seiter, 2000; Yoshida, 1990; Verlet, 1967]. Time-step used in the integrations is an important quantity to determine the accuracy of the trajectories. Larger time-step lead to poor energy conservation and very small time-steps will enhance the computational time significantly. In practice, the largest normal mode vibrational frequency and the associated time-period are used in determining the time-step $\triangle t$. The best value for $\triangle t$ will be $(1/\nu) * 0.1$ where ν is the largest vibrational frequency in the system.

2.3.3 Bimolecular Collisions

A collision between two reactants

$$A + B \rightarrow C + D \tag{2.25}$$

is defined by the force acting between the molecules[Levine, 2009]. The collision can be reactive or unreactive. Bimolecular collisions in the gas phase depend on several factors including impact parameter and collision cross section. Reaction cross section, which is an important quantity in the study of bimolecular reaction dynamics, is a measure of the number of reactive collisions between molecules of type A and those of type B (both in well defined quantum states) which are moving towards each other with well defined relative velocity[Kuppermann and Greene, 1968].

$$\left(\frac{dn_B}{dt}\right)_v = -\left(\frac{dn_A}{dt}\right)_v = \sigma_r v n_A n_B \tag{2.26}$$

Here n_A and n_B are the number densities of A and B and σ_r and v are the reaction cross section and relative velocity, respectively. Reaction cross section has the dimensionality of area and it measures the effective size of molecules for the reaction to occur i.e., it provides the likelihood of a reaction occurring during a collision. Classical expression for the reaction cross section is given by[Sun and Hase, 2003]

$$\sigma_r = \int_0^{b_{max}} P_r(b) 2\pi b db \tag{2.27}$$

where b is the collision impact parameter which is measure of distance of closest approach between the two colliding molecules in the absence of a force. $P_r(b)$ is the probability for the reaction to occur at a given b, i.e., $P_r(b) = N_r/N$ where N_r and N are the number of reactive and total number of trajectories, respectively, in a classical trajectory simulation at a given value of b. Integration in Eq.(2.26) is between 0 and b_{max} which is the largest value of b at which a reactive collision takes place. Further,

$$\langle P_r(b) \rangle = \frac{\int_0^{b_{max}} P_r(b) 2\pi b db}{\int_0^{b_{max}} 2\pi b db} \tag{2.28}$$

$$\int_0^{b_{max}} 2\pi b db = \pi b_{max}^2 \tag{2.29}$$

$$\sigma_r = \langle P_r(b) \rangle \pi b_{max}^2 \tag{2.30}$$

To determine σ_r from a classical trajectory simulation, trajectories are sampled with b values ranging from 0 to b_{max}. For practical purposes, b_{max} is determined by running a small number of trajectories over a range of b values and b_{max} is selected based on reactivity.

2.3.4 Unimolecular Dissociation Reactions

In unimolecular dissociation reactions a molecule excited with its threshold energy. Threshold energy distributed within intramolecular bonds and gives vibrational-rotational excitation of molecule.

$$A \rightarrow A^* \rightarrow B + C \tag{2.31}$$

where the (*) denotes vibrational-rotational excitation. If unimolecular decomposition is random during the timescale i.e. A^* is initially excited with a microcanonical ensemble and intramolecular dynamics of A^* is ergodic[Hase *et al.*, 2003]. The unimolecular rate constant, for a microcanonical ensemble of reactant calculated with the Rice-Ramsperger-Kassel-Marcus (RRKM) rate constant. The trajectory lifetime is exponential[Lourderaj and Hase, 2009] for RRKM unimolecular dynamics. If the system is intrinsically non-RRKM, the lifetime distribution is multi-exponential[Lourderaj and Hase, 2009]

2.4 RRKM CALCULATIONS

RRKM (Rice-Ramsperger-Kassel-Marcus) theory is used to determine the energy dependent rate constant $k(E)$ for the unimolecular decomposition of a micro-canonical ensemble[Marcus and Rice, 1951; Marcus, 1952]. RRKM theory is based on the assumption that the internal energy in a molecule is distributed among the modes of the molecule according to equipartition theorem i.e., instantaneous IVR happening in the molecule in the timescale of reaction. Vibrational modes are treated in an adiabatic manner because of the angular momentum conservation. According to RRKM theory, micro-canonical rate constant is given by

$$k(E) = \frac{G(E^{\dagger})}{hN(E_v)} \tag{2.32}$$

where $G(E^{\dagger})$ is the sum of states at the transition state and $N(E_v)$ is the density of states of the reactant and h is the Planck's constant. Sum and the density of states can be determined in a classical or quantum mechanical or a semi-classical manner[Beyer and Swinehart, 1973].

In a classical trajectory simulation, for a micro-canonical ensemble of initial conditions, the trajectory lifetime $N(t)/N(0)$ is exponential[Lourderaj and Hase, 2009] for RRKM unimolecular dynamics. Here, $N(t)$ is the number of trajectories in which the unimolecular reactant has not dissociated at time t and $N(0)$ is the total number of trajectories.

$$\frac{N(t)}{N(0)} = e^{-k(E)t} \tag{2.33}$$

If the system is intrinsically non-RRKM, the lifetime distribution is multi-exponential[Lourderaj and Hase, 2009],

$$\frac{N(t)}{N(0)} = \sum_i f_i e^{-k_i t} \tag{2.34}$$

Here, $\sum_i f_i = 1$. Such non-RRKM behavior occurs due to trajectory recrossing the transition state region and the non-ergodic nature of the classical phase space[Manikandan and Keshavamurthy, 2014].

2.5 SOFTWARE

For performing the electronic structure calculations reported in the present work, the open-source NWChem electronic structure theory package[Valiev *et al.*, 2010] was used. The direct dynamics calculations were performed using the general chemical dynamics program VENUS[Hase *et al.*, 1996] which is interfaced with NWChem for carrying out the on-the-fly electronic structure calculations. The two codes are compiled together using a tight coupling algorithm[Lourderaj *et al.*, 2014] and run as a single executable. The RRKM rate constants $k(E)$ were computed using the computer program developed and published by Hase's research group[Zhu and Hase, 1994].