2

Theoretical Methods and Techniques

To understand the fundamental processes involved in a chemical reaction at the atomic level, the study of reaction dynamics is required. Any dynamical study of chemical reactions starts with an appropriate potential energy surface. A brief overview of the potential energy surface is provided in this chapter along with descriptions of Born-Oppenheimer direct dynamics method.

2.1 POTENTIAL ENERGY SURFACE

A chemical reaction is a dynamical process depending primarily on the interaction forces between atoms and molecules:

$$F = -\frac{dV(R)}{dR}$$

(2.1)

where V(R) is the potential energy of the system and dependent on the nuclear coordinates. V(R) is the potential energy function of a system and in a multidimensional system, it becomes *potential energy surface* (PES). In order to establish the dynamics of a chemical system, understanding of all the forces operating within a system is required. PES provides information about the interactions depending on the nuclear configuration of the system during the transformation of the reactants into products. A knowledge of PES is necessary to investigate molecular properties and the processes that occur during a chemical reaction [Domcke *et al.*, 2004; Laganà and Riganelli, 2012; Zhang, 1999; Gerratt, 1985; Varandas, 2000]. Hence an accurate description of a PES is the key step in any successful dynamical calculation to characterize the atomic level motion of the reaction system [Truhlar, 2013]. Some features of the PES are presented below.

2.1.1 Dimensionality

A polyatomic molecule containing N atoms requires $3N$ coordinates for a successful description of its vibrational, rotational, and translation motions. Rotational and translational degrees of freedom do not alter the PES which depends only on the relative (internal) coordinates of all the atoms. Hence, six (five) coordinates corresponding to overall translations and rotations can be subtracted from the 3N coordinates of non-linear (linear) molecules from the overall dimensionality [Jaquet, 1999]. Hence,

$$\text{Dimension of a PES} = \begin{cases} 3N - 5 & \text{(for linear molecules)} \\ 3N - 6 & \text{(for non-linear molecules)} \end{cases}$$

In the case of a linear molecule, only two rotational degrees of freedom exist. For a diatomic molecule AB, where $N = 2$, V(R) is called as *potential energy curve* [as presented in Figure 2.1(a)]
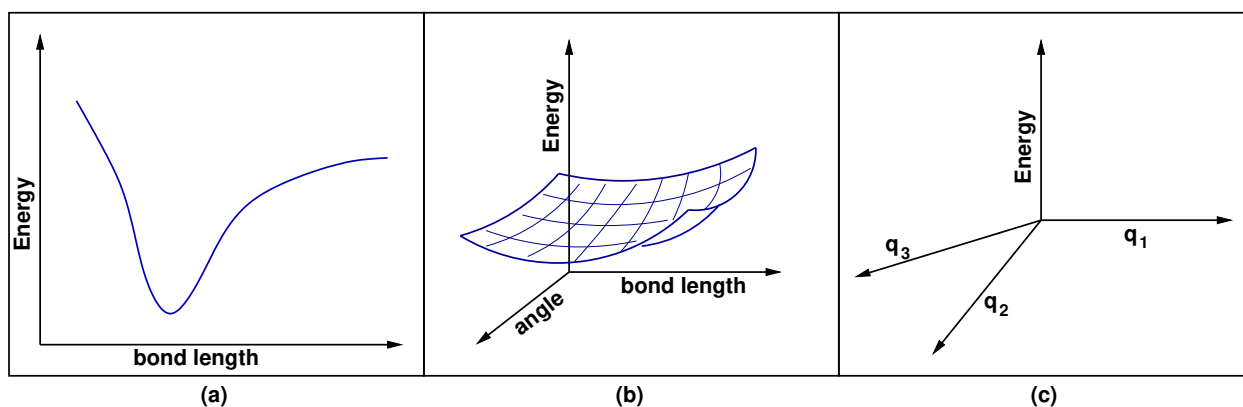
Figure 2.1. : Schematic Potential energy (a) curve for a molecule AB where PE depends upon one geometric parameter (b) surface for a ABA type molecule in which PE is a function of two geometric parameters, (c) hyper-surface for such molecules where N$\geq$3 and PE depends on more than two geometric parameters.

and, in this case, potential energy is only a function of the inter-nuclear bond distance of A-B. The potential energy graph, *V vs.* two geometric parameters i.e., bond angle and bond length, for example, is known as *potential energy surface* [Figure 2.1(b)]. Note that for a non-linear triatomic molecule, the PE depends on 3 coordinates. Figure 2.1(b) is only for representation purposes. If the potential energy depends on more than two geometric parameters, it is termed as *potential energy hyper-surface* [Figure 2.1(c)]. Such hyper-surfaces cannot be completely projected in three dimensional space. However, mathematically such high dimensional spaces can be treated with appropriate tools.

2.1.2 Geometry Optimization and Stationary Points

A standard theoretical technique for determining molecular structures is the optimization of geometries. Generally, two kinds of molecular geometries viz., equilibrium and transition state
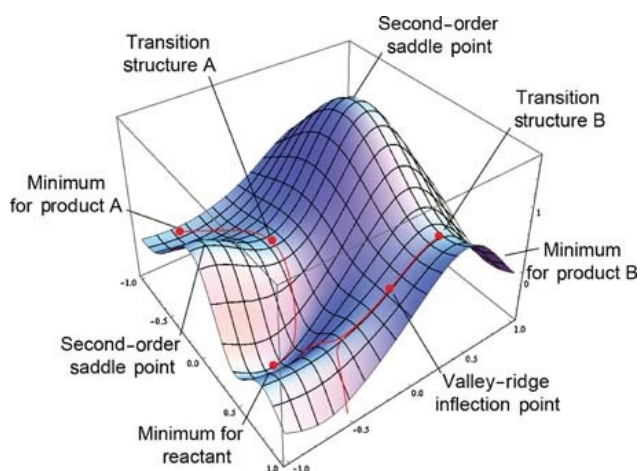

Figure 2.2. : Characterization of Points on the PES (Reprinted from Reference [Schlegel, 2011].)

(saddle point) geometry are important which are stationary points in a PES i.e., points of nuclear configurations at which dV(R)/dR=0. The process of trying to find such points in a multidimensional PES using various mathematical tools is called *geometry optimization* [Schlegel, 1982]. A *single point energy calculation* is the process of finding V(R) for a fixed nuclear configuration for a given R by solving the electronic Schrödinger equation. The process of geometry optimization starts with a 'guess' input geometry resembling the required geometry to a computer algorithm which efficiently changes the geometry until it finds a desired stationary point [Deaven and Ho, 1995]. Finding a minimum is called energy *minimization*, and finding a transition state is called *transition state optimization*. Such processes require a suitable computer algorithm such as a Newton based method [Schlegel, 2011]. The nature of a stationary point as a minimum or a saddle point is determined using the second derivatives of the potential with respect to the coordinates or the Hessian [Fernandez *et al.*, 1988]. Features of a PES are shown in Figure 2.2. A reaction coordinate or an intrinsic reaction coordinate [IRC] in a PES is the minimum energy path connecting the reactants to products [Fukui, 1970]. At the stationary points on a PES, the surface is flat i.e., horizontal or parallel with respect to the reaction coordinate. Mathematically, a stationary point is a point where the first derivative of potential energy is zero with respect to all the coordinates.

$$\frac{\delta V}{\delta q_1} = \frac{\delta V}{\delta q_2} = \cdots = 0 \tag{2.2}$$

Here, $q_1, q_2, \cdots$ are the coordinates. These stationary points can be further classified based on the second derivatives of the potential energy with respect to the coordinates.

Minimum: A minimum corresponds to stable or meta-stable species i.e., reactants, products, or reactive intermediates and the second derivative is positive with respect to all the coordinates.

$$\frac{\delta^2 V}{\delta q_i^2} > 0 \tag{2.3}$$

Transition state: Transition states are those points which are minimum in all the dimensions but maximum along the reaction coordinate q_{rc} i.e., second derivatives are positive along all the coordinates except the reaction coordinate.

$$\frac{\delta^2 V}{\delta q_i^2} > 0 \tag{2.4}$$

for all *i* except the reaction coordinate.

$$\frac{\delta^2 V}{\delta q_{rc}^2} < 0 \tag{2.5}$$

The above description is for a first order saddle point. For a second order or higher order saddle points, correspondingly the second derivatives are maxima in two or more dimensions. The stationary point which is the lowest energy point on the whole PES is called the *global minimum*.

The transition state, which connects two minima, is a maximum along IRC, but it is minimum in all other directions. TS is a first-order saddle point and has one imaginary frequency and higher order saddle points will have more than one imaginary frequencies depending upon the nature of the saddle point.

Valley-ridge inflection point: Valley-ridge inflection point on the PES is the place where an eigenvalue of the hessian orthogonal to the gradient direction changes from "+" to "-" through zero, or vice versa. A valley inflects into a ridge. Valley-Ridge Inflection point is located where at least one main curvature of the PES becomes zero i.e. the second derivative of the potential with respect to the reaction coordinate is equals to zero. These points represent non-stationary points of the PES [Quapp, 2003; Lourderaj *et al.*, 2008a].

Zero-point energy (ZPE) is the energy which a molecule has at absolute zero temperature. The energies V(R) obtained by solving above equations are classical energies without ZPE corrections. The vibrational frequency data can be used to calculate ZPE that can added to the classical energies to obtain the zero point corrected energies. The ZPE of a molecule is not often small but it tends to cancel out for the whole reaction energy profile when ZPE is added to all the reactant, transition states, and products. ZPE corrected energies are used for comparison with the experimental data. The process of determining normal mode frequencies is described below.

2.1.3 Normal Mode Analysis

After optimizing the geometry of a molecule, it is necessary to check whether the stationary point is a minimum or a saddle point (transition state). Normal mode vibrational frequency calculations are used to determine whether a stationary point is a minimum or a saddle point. In a normal mode vibration of a molecule, all the atoms in the molecule move in phase with same frequency, i.e., they all reach their equilibrium positions, maximum and minimum displacements simultaneously. Local mode vibrations of the molecule can be written as linear combinations of these vibrations. A nonlinear molecule with N atoms has 3N-6 independent normal modes of vibrational motions, and a linear molecule has 3N-5 normal modes. The normal-mode frequency for a molecule is given by:

$$\nu = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}} \tag{2.6}$$

where ν is the normal mode vibrational frequency, k is the force constant associated with the vibration, and μ is the reduced mass. The force constant k measures the stiffness of a bond. Larger is the value of the force constant, the more difficult it is to bend or stretch the molecule in that mode. Thus the frequency of a vibrational mode is related to the force constant. A matrix of second-order energy derivatives with respect to geometric coordinates (e.g., cartesian or internal coordinates) is called the force constant matrix (Hessian, H_{ij}).

$$H_{ij} = \frac{\delta^2 V}{\delta q_i \delta q_j} \tag{2.7}$$

Diagonalization of the Hessian matrix provides the eigenvectors which describe the normal mode vibrations of the molecule, and the corresponding eigenvalues are directly proportional to the square of the vibrational frequencies. Maxima and minima on the PES can be recognized by the
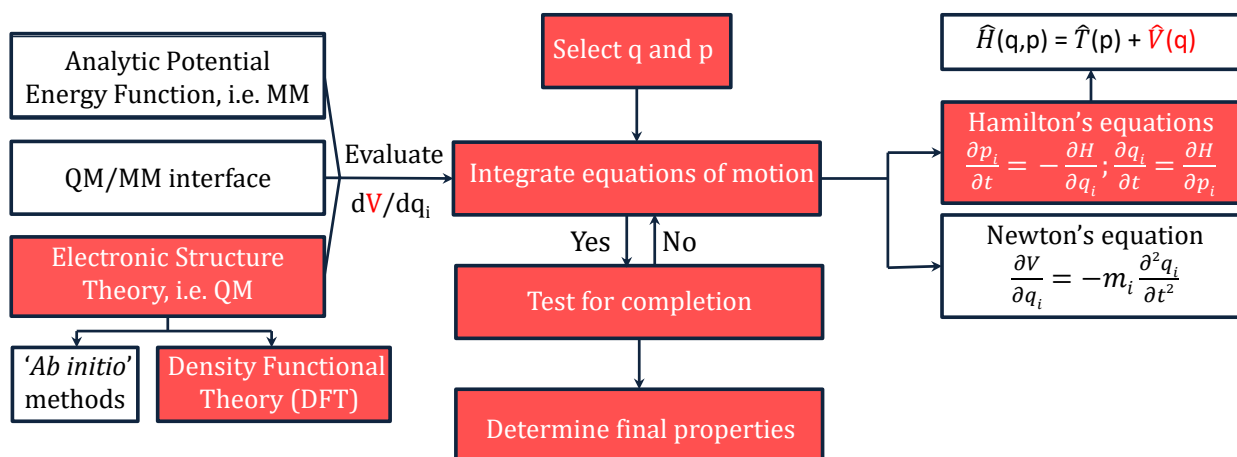
Figure 2.3. : Flowchart for the components of a classical trajectory simulation.

sign of the second derivative of energy. The maxima or saddle points (maximum in one direction but minimum in other directions) have at least one negative (imaginary) frequency whereas minima on the PES have only positive vibrational frequencies i.e. positive eigenvalues. Thus normal mode calculations must be carried out to determine the sign of the vibrational frequencies to characterize the stationary points on PES.

2.2 BORN-OPPENHEIMER MOLECULAR DYNAMICS (BOMD)

Direct dynamics simulations [Pratihar *et al.*, 2017] are widely used to interpret experimental results and study the atomistic dynamics of chemical reactions and energy transfer processes. Essential components of a classical trajectory simulation [Bunker, 2012] is shown in Figure 2.3 as a flowchart and are described below.

- Development of a potential energy surface for the chemical problem under investigation using appropriate electronic structure theory or experimental data;

- Selection of proper initial conditions (coordinates q's and the conjugate momenta p's) for the ensemble of trajectories to be calculated that mimics given experimental conditions or a chemical phenomenon;

- Numerical integration of the classical equations of motion (either Hamilton's or Newton's) to determine the motion of atoms;

- Converting the trajectory's final values of atomic coordinates and momenta to obtain properties which can be compared with theoretical models such as Rice-Ramsperger-Kassel-Marcus (RRKM) theory and/or experiments.

Various initial condition sampling schemes [Paranjothy *et al.*, 2013] are available such as classical microcanonical sampling, quantum microcanonical sampling, thermal sampling, transition state sampling, etc. After selecting the initial conditions, an ensemble of trajectories are generated. Trajectories are integrated till an appropriate t_{max} is reached or until a reaction occurred. Once the dynamics simulation is completed, the time evolved dynamical variables $[q_i(t), p_i(t)]$ are analyzed for reaction products, pathways, mechanisms, etc. Using the time evolved coordinates, a

movie on molecular motion can be established to directly visualize the reaction mechanism using softwares such as molden [Schaftenaar and Noordik, 2000] or visual molecular dynamics (VMD) program [Humphrey *et al.*, 1996]. The vibrational, rotational, and relative translational energies of the products, amount of energy in individual molecular degrees of freedom, the lifetime of a vibrationally excited molecule, and scattering angles can be computed using the time evolved coordinates and momenta [Sun and Hase, 2003].

2.2.1 Integration Methods

As discussed above, Newton's or Hamilton's equations of motion are numerically integrated in classical trajectory simulations. Coordinates required for the integration can be represented in the form of internal, cartesian, or instantaneous normal modes [Wilson *et al.*, 1980; Adams and Stratt, 1990; Miller *et al.*, 1989]. Potential energy functions are easily represented in internal coordinates. But these coordinates are efficient only for smaller systems. Since kinetic energy depends upon both the coordinates and momenta, it is efficient to integrate the classical equations of motion using instantaneous normal-mode coordinates or cartesian coordinates. There are several algorithms available for numerical integrations in cartesian coordinates such as Gear, Adams-Moulton, fourth-order Runge-Kutta, sixth-order Adams-Bashforth-Moulton methods [Sun and Hase, 2003; Thompson, 1998; Press *et al.*, 1992; Stoer and Bulirsch, 1993; Bolton and Nordholm, 1994]. These are predictor-corrector algorithms consisting of a prediction and a correction at each integration step. The predictor requires largest possible integration time-step for evaluations of the potential energy and its gradient during each time step to reduce the number of quantum chemistry single point calculations in a direct dynamics simulation. The accuracy order of these algorithms is usually at or higher than fourth order. Another class of integration method is the symplectic integrator schemes [Gray *et al.*, 1994; Schlier and Seiter, 1998], often used to integrate the Hamilton's equations of motion. These are based on Runge-Kutta methods. Symplectic integrators give good energy conservation because they preserve the symplectic structure of the phase space and are particularly useful to study long time integration of large systems. They are superior to non-symplectic methods which usually give a continuous drift in energy and hence suitable for short time integration. Although symplectic methods are slow in comparison to non-symplectic algorithms, they are favorable when integrating large systems for longer time duration. Description of symplectic integrations is given below.

Consider a system with a separable Hamiltonian i.e., H is written as sum of kinetic and potential energy

$$H(p,q) = T(p) + V(q) \tag{2.8}$$

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

$$\dot{z} = \{z, H(z)\} \tag{2.9}$$

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

$$\dot{z} = D_H z \tag{2.10}$$

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

$$z(\tau) = [\exp(\tau D_H)]z(0) \tag{2.11}$$

Here τ is step-size. Hamiltonian of the form of Eq. (2.8) can be written as:

$$z(\tau) = \exp[\tau(D_T + D_V)]z(0) \tag{2.12}$$

Suppose, c_i and d_i are real numbers, where

$$\sum_{i=1}^{k} c_i = \sum_{i=1}^{k} d_i = 1 \tag{2.13}$$

and satisfy the given equality,

$$\exp[\tau(D_T + D_V)] = \exp(c_i \tau D_T) \exp(d_i \tau D_V) + O(\tau^{n+1}) \tag{2.14}$$

Here n is an integer called the order of the integrator.
Since $D_T^2 z = \{\{z, T\}, T\} = \{\{\dot{q}, 0\}, T\} = (0,0)$ for all z

$$D_T^2 = 0 \tag{2.15}$$

Using a Taylor series expansion, $\exp(aD_T)$ can be expressed as

$$\exp[aD_T] = \sum_{n=0}^{\infty} \frac{(a(D_T)^n}{n!} \tag{2.16}$$

Using Eq. (2.15) and (2.16),

$$\exp[aD_T] = 1 + aD_T \tag{2.17}$$

Similarly,

$$\exp[aD_V] = 1 + aD_V \tag{2.18}$$

where a is an arbitrary real number. The simplified form of equations are:

$$q_{i+1} = q_i + c_i \frac{p_{i+1}}{m} t \tag{2.19}$$

$$p_{i+1} = p_i + d_i F(q_i) t \tag{2.20}$$

In Lagrangian coordinates:

$$x_{i+1} = x_i + c_i v_{i+1} t \tag{2.21}$$

$$v_{i+1} = v_i + d_i a(x_i) t \tag{2.22}$$

where $F(x)$ is the force vector at x, $a(x)$ is the acceleration vector at a, and m is the scalar quantity of mass.

Symplectic integrator uses position x_i and velocity v_i of a particle at time t and calculate the position $x_{(i+1)}$ at $(t+1)$ by adding to its velocity $v_{(i+1)}$ (previously updated) multiplied by c_i. Velocity of the particle $v_{(i+1)}$ at $(t+1)$ is obtained by adding to its acceleration (at updated position x_i) multiplied by d_i. Several higher order symplectic integrators [Schlier and Seiter, 2000; Yoshida, 1990] are available in the literature. The 6th order symplectic integrator is used in the present work.

2.2.2 Integration Time-step

Value of the integration time-step is a very important parameter while performing any type of dynamics simulation. Larger time steps accelerate the calculations, but can affect the accuracy of the simulation. The maximum recommended value depends upon the fastest movements that happen during the simulation. The selection of optimal integration time-step Δt is one of the most important pointers for a successful numerical integration of the equations of motion. It is not practical to use constant step size in a numerical integration. If the chosen time-step is too large,

the calculated solution may differ from the exact solution. When it is too small, the number of arithmetic operations, calculation time, and the calculation errors begin to increase. Thus, if the solution is changing quickly, the chosen time-step must be checked. Conversely, one can choose a longer time step if the solution is changing slowly. For selecting the best possible time-step, the largest frequency among all the vibrations in a system can be selected and the corresponding 'period' (inverse of the frequency) can be divided by 10 and can be selected as the time-step. In this way, one will have fastest motion sampled at least ten times during each vibration.

2.2.3 Trajectory Initial Conditions

Direct dynamics simulations have been used to identify decomposition mechanisms of gas phase reactions and to make comparisons with experiments and theoretical models such as Rice-Ramsperger-Kassel-Marcus (RRKM) theory. A crucial step for a successful classical trajectory simulation is to choose appropriate initial conditions to mimic experimental conditions or a theoretical model i.e., selection of coordinates and momenta for simulating an ensemble of trajectories. A detailed description of different initial condition sampling schemes is given by Peslherbe et al. [Peslherbe *et al.*, 1999] and Cho et al. [Cho *et al.*, 1992]. Two different sampling schemes were primarily used in the present work, classical microcanonical sampling and to model a collision induced dissociation (CID) experiment. A brief description of initial condition selection schemes for trajectory simulations is given below.

2.2.4 Unimolecular Dynamics

Unimolecular reaction dynamics reported in the thesis were performed using classical microcanonical sampling method which is typically used to model gas phase experiments [Sun and Hase, 2003].

Classical Microcanonical Sampling

In order to produce a microcanonical ensemble, random values for p_i and q_i are selected such that total energy is a constant in all the trajectories [Hammersley, 2013]. This sampling is appropriate for modeling experiments in which the dynamics is well represented by classical mechanics [Dobbyn *et al.*, 1995]. Trajectories are numerically integrated under the condition that total energy is constant until the dissociation products are well separated. Total energy for a molecule consisting of *s* harmonic oscillators is given by:

$$E = \sum_{i=1}^{s} E_i \tag{2.23}$$

where E_i are the individual harmonic oscillator energies. If an oscillator *i* has energy between E_i and $(E_i + dE_i)$ then the corresponding probability $P(E_i)dE_i$ is directly proportional to the (classical) density of states of the remaining $s-1$ oscillators [Hase and Buckowski, 1980]. $P(E_i)$ is given by:

$$P(E_i) = [E - \sum_{j=1}^{i-1} E_j - E_i]^{s-1-i} \times \left\{ \int [E - \sum_{j=1}^{i-1} E_j - E_i']^{s-1-i} dE_i' \right\}^{-1} \tag{2.24}$$

In the above equation, the integral is in the range of 0 to $(E - \sum_{j=1}^{i-1} E_j)$. $P(E_i)$ is evaluated by cumulative distribution function [Hammersley, 2013] or by the rejection method such as von

Neumann technique. The random values of energy is distributed to coordinates and conjugate momenta by either associating a random classical phase [Peslherbe *et al.*, 1999] to the normal mode or a quantum mechanical distribution function [Mayer and Band, 1947; Husimi, 1940; Takahashi, 1986] such as Wigner sampling. In the present work, purely classical sampling were performed.

Mass Spectrometry Simulations: Collision-Induced Dissociation (CID)

Direct dynamics simulations have been used to study the collisional activation of small organic and biological molecules. In these simulations, the reactant molecule is energized by a collision with an unreactive molecule (e.g., N_2) or a rare gas atom (e.g., Ar) at a fixed amount of collision energy and impact parameter *b* [Levine, 2009]. The reactant vibrational and rotational energies are selected from Boltzmann distributions at fixed temperatures [Peslherbe *et al.*, 1999]. The reactant molecule is oriented differently about their Euler angles in different trajectories to mimic the experimental conditions. If electronic excitations are not important which is usually the case, the collisions transfer a fraction of the translational energy to the internal vibrational and/or rotational energy of the ion in the limit of low-energy collisions so that the ion can eventually dissociate. These CID simulations play an important role [Pratihar *et al.*, 2017] in studying the mechanisms of unimolecular dissociations, predicting fragmentation mechanisms of new structures such as different isomers having same chemical formula, identifying the neutral molecules which cannot be observed in mass spectrometry experiments, and calculating the vibrational and rotational energy distributions of the excited molecules. Simulations of CID unimolecular dissociations have shown that [Martin Somer *et al.*, 2019] two types of mechanisms may occur: (1) the traditional, statistical RRKM mechanism with efficient IVR (intramolecular vibrational energy redistribution) and (2) shattering type mechanism in which dissociation occurs during the collision or within a very short time following the collision. The shattering type of mechanisms may be identified only via direct dynamics simulations.

2.2.5 Non-RRKM behavior

One of the fundamental assumptions of RRKM theory is that IVR is instantaneous and the dissociation process is random [Bunker, 1964; Bunker and Hase, 1973; Baer and Hase, 1996]. This will be the situation if a molecule *A* is microcanonically excited i.e., molecules are excited randomly at constant total energy and this total energy is conserved throughout the trajectory integration. In this process, each vibrational state of an excited molecule has an equal probability of being populated in the energy interval $E \rightarrow E + \Delta E$. For a nonlinear molecule, each state has a random energy E_i in each of the $3N - 6$ vibrational normal modes and the sum of the E_i equal to total E in the interval $E \rightarrow E + \Delta E$. For an initial microcanonical ensemble, the RRKM rate constant $k(E)$ can be expressed as

$$N(t) = N(0)e^{-k(E)t} \tag{2.25}$$

where $N(0)$ is the total number of trajectories and $N(t)$ is the number of trajectories that did not dissociate at time t. For RRKM dynamics, the lifetime distribution $[N(t)/N(0)]$ is single exponential as shown in Eq. (2.25). When the unimolecular dissociation dynamics is intrinsically Non-RRKM [Lourderaj and Hase, 2009], the lifetime distribution is multi-exponential i.e., sum of two or more exponential functions are required to fit the lifetime distribution.

$$N(t) = N(0)\sum_i f_i e^{-k_i t} \tag{2.26}$$

where $\sum_i f_i = 1$.

RRKM theory assumes that the classical phase space of the dissociating system is adequately chaotic to keep the microcanonical ensemble of states as the reactant decomposes. Classical microcanonical unimolecular RRKM rate constant is expressed as [Baer and Hase, 1996]:

$$k(E) = \frac{N(E)}{h\rho(E)} \tag{2.27}$$

where $N(E)$ is the sum of states at the transition state and $\rho(E)$ is the density of states of the excited reactant. Further, RRKM theory assumes that the trajectories do not recross once the transition barrier is crossed. However, trajectory recrossings may happen and the rate constants may deviate from the RRKM predictions [Steinfeld *et al.*, 1989; Bunker and Pattengill, 1968].

2.3 SOFTWARE

NWChem [Valiev *et al.*, 2010] quantum chemistry software package was used to perform electronic structure calculations reported in the present work. For dynamics simulations, the general chemical dynamics program VENUS developed by Hase and coworkers [Hase *et al.*, 1996] was used. This dynamics program was interfaced with NWChem to perform the single point energy calculations required for the on-the-fly integrations. The coupling is considered to be tight coupling [Lourderaj *et al.*, 2014] as the two codes are compiled together and work through routine calls very frequently. At each time step, the dynamics program makes a call to the electronic structure theory program to obtain the energies and gradients. Conservation of total energy is monitored continuously and trajectories were animated with molden [Schaftenaar and Noordik, 2000] and VMD (visual molecular dynamics) [Humphrey *et al.*, 1996] to understand the reaction mechanisms.

…