On the Role of Nuclear Quantum Effects in Reactive Dynamics Simulations

Federica Angiolari

Sorbonne Université

École doctorale 388: Chimie Physique et Chimie Analytique de Paris-Centre
Laboratoire de chimie théorique (LCT)

On the Role of Nuclear Quantum Effects in Reactive Dynamics Simulations

Defended by Federica Angiolari

PhD thesis in Physical Chemistry

25 September 2023

Jury:

President	Mathieu Salanne	Professeur, Sorbonne Unviersité
Reviewers	Federica Agostini	Maitre de Conférence (HDR), Université Paris-Saclay
	Francesca Ingrosso	Maitre de Conférence (HDR), Université de Lorraine
Examiners	Enrico Bodo	Professeur, La Sapienza Universitá di Roma
Supervisor	Riccardo Spezia	Directeur de Recherche CNRS, Sorbonne Université
Co-Encadrant	Simon Huppert	Maitre de Conférence, Sorbonne Université

The thesis is based on the following articles:

1. Angiolari, F.; Huppert, S.; Spezia, R. "Quantum versus classical unimolecular fragmentation rate constants and activatoin energies at finite temperature from direct dynamics simulations", *Phys. Chem. Chem. Phys.* **2022**, *24*. Chapter 3

2. Angiolari, F.; Huppert, S.; Pietrucci, F.; Spezia, R. "Environmental and Nuclear Quantum Effects on Double Proton Transfer in Guanine-Cytosine Base Pair", *J. Phys. Chem. Letter* **2023**, *14,22*, 5102-5108. Chapter 4

3. "Thermodynamical and dynamical nuclear quantum effects in heavy atom tunneling: the prototypical case of Cope rearrangement of semi-bulvalene", in preparation

The code developed in the present work was also applied in the following article (not part of the thesis): Senyah, K.N.; Asare, P.; Wilcox, J.D.; Angiolari, F.; Spezia, R.; Angel, L.A."Extending the competitive threshold collision-induced dissociation of Zn(II) ternary complexes using traveling-wave ion mobility-mass spectrometry", *Int. J. Mass Spectrom.* **2023**, *488*,1387-3806

Abstract

This thesis mainly focuses on the study of reactive systems, exploring how nuclear quantum effects (NQEs), in particular tunneling and zero point energy (ZPE), can affect the dynamical and thermodynamical properties of a chemical reaction. In principle, in order to include NQEs, one should solve the time-dependent Schrödinger equation for the atomic nuclei, but this is possible only for systems with few degrees of freedom. For this reason, different approximate methods were developed in the past in order to take into account those effects without excessive computational effort.

In this thesis, we will discuss in detail two approaches: ring polymer molecular dynamics (RPMD), and the quantum thermal bath (QTB). The first one is based on the path integral formalism outlined by Feynman, while the second one uses a generalized Langevin equation in order to introduce zero point energy (ZPE) in an otherwise classical simulation.

First, we investigate how the ZPE of the system affects an unimolecular fragmentation, comparing rate constants and activation energies for both classical and quantum results obtained with RPMD and QTB methods. The system is described by an analytical potential, and in order to explore a different ranges of temperature, the potential was modified, changing the barrier of the reaction. Furthermore, a benchmark with the exact result is obtained for a 1D model system (a simple morse potential). Results show that RPMD is able to capture correctly the NQEs on rate constants (and therefore activation energy). Instead, QTB largely overestimates the rate constants in all the systems. The reason of this wrong behaviour, is to search in the distance distribution, where the QTB describes poorly the tail of the probability distribution which is fundamental to describe the fragmentation correctly.
Then, we study the double proton transfer in guanine-cytosine DNA base pairs. In particular we focus on how NQEs and environment can affect this reaction, thus the thermodynamics and dynamical properties. In order to study the mechanism, the free energy landscape was obtained using umbrella sampling for both classical and path integral based simulations. Furthermore, an ensemble of trajectories was performed starting from the tautomeric form and was let free to go back to the more stable form (the canonical one). The results show that for the dimer in gas phase the mechanism is concerted, and the effect of NQEs is to speed up the reaction up to a factor of 30 in comparison with classical results. When the environment is included the mechanism becomes step-wise, the NQEs still speed up the reaction, but what is more affected is the free energy landscape, for which the intermediate ceases to be a local minimum, making the proton transfer almost barrierless.
The last part concerns the study of heavy atoms tunneling for a prototypical reaction, the Cope rearrangement of semibullvalene. We focus on the reaction within transition state theory (TST) approximation, by computing the free energy barrier with umbrella sampling technique for both classical and RPMD methods. We also correct the TST by the recrossing factor, in order to estimate the accuracy of this approximation. We found that in the case of RPMD when the temperature is lower than the cross over temperature, the free energy profile in the transition state region is completely flat, showing a typical marker for tunneling. The inclusion of tunneling speeds up the reaction up to 60 order of magnitude at 25 K, and this difference disappears in the classical limit, as we expected. The computation of the recrossing factor, shows that it is significant only at the lowest temperature (25 K), where this value is about 0.6. It should be noted that even if the rate constant is lowered by the inclusion of the recrossing factor in this low temperature regime, it is not comparable with the impact of tunneling which, as we said before, speeds up the reaction up to 60 order of magnitude.

Resumé

Cette thèse se concentre sur l'étude des systèmes réactifs, en explorant comment les effets quantiques nucléaires (EQN), en particulier le *tunneling* et l'énergie du point zéro (EPZ), peuvent influencer les propriétés dynamiques et thermodynamiques d'un système chimique, en particulier lorsque des atomes légers sont impliqués. En principe, pour inclure les EQN, il faudrait résoudre l'équation de Schrödinger dépendante du temps pour les noyaux atomiques, mais cela n'est possible que pour les systèmes à quelques degrés de liberté. Pour cette raison, différentes méthodes approchées ont été développées.

Dans cette thèse, nous discuterons en détail deux approches : *ring polymer molecular dynamics* (RPMD) et le *quantum thermal bath* (QTB). La première repose sur le formalisme des intégrales de chemin développé par R. Feynman, la seconde utilise une équation de Langevin généralisée pour introduire l'EPZ dans une simulation classique.

En premier lieu, nous avons étudié comment l'EPZ du système affecte la fragmentation unimoléculaire du méthane, en comparant la constante de réaction et l'énergie d'activation dans les cas classiques et quantiques obtenus avec les méthodes RPMD et QTB. Les résultats montrent que RPMD est capable de capturer correctement les EQN sur les constantes de réaction (et donc l'énergie d'activation). En revanche, QTB surestime largement la constante de réaction dans tous les systèmes étudiés. La raison de ce comportement erroné se trouve dans la distribution des distances: le QTB décrit mal la queue de distribution de probabilité qui est fondamentale pour bien décrire la fragmentation.
Ensuite, nous avons étudié le double transfert de proton dans les paires de bases d'ADN guanine-cytosine. En particulier, nous nous concentrons sur la façon dont les EQN et l'environnement biochimique peuvent affecter cette réaction, ainsi que les propriétés thermodynamiques et dynamiques. Afin d'étudier le mécanisme, le profil d'énergie libre a été obtenu en utilisant la technique d'*umbrella sampling* (US) pour les simulations classiques et basées sur le RPMD. De plus, un ensemble de trajectoires a été effectué en partant de la forme tautomérique et en la laissant libre de revenir à la forme plus stable (la forme canonique). Les résultats montrent que pour le dimère en phase gaz, le mécanisme est concerté, et l'effet des EQN est d'accélérer la réaction d'un facteur de 30 par rapport aux cas classiques. Lorsque l'environnement est inclus, le mécanisme se déroule par étapes, les EQN accélérant toujours la réaction, mais c'est le profil d'énergie libre qui est le plus affecté, le processus devenant presque sans barrière.
La dernière partie concerne l'étude du tunneling des atomes lourds pour une réaction prototypique, le réarrangement de Cope du semi-bullvalène. Nous nous concentrons sur la réaction dans l'approximation *transition state theory* (TST), en calculant la barrière d'énergie libre avec la technique US pour les méthodes classiques et RPMD. Nous corrigeons également la TST par le facteur de *recrossing*, afin d'estimer la précision de cette approximation. Nous avons constaté que dans le cas de RPMD, lorsque la température est inférieure à la température dite de "*cross-over*", le profil d'énergie libre dans la région de l'état de transition est complètement plat, ce qui est un indicateur typique de l'effet tunnel. L'inclusion du tunneling accélère la réaction jusqu'à 60 ordres de grandeur à 25 K, et cette différence disparaît dans la limite classique (haute température), comme nous nous y attendions. Le calcul du facteur de recrossing montre qu'il est significatif uniquement à la plus basse température (25 K), où cette valeur est d'environ 0,6. Il convient de noter que même si la constante de vitesse est réduite par l'inclusion du facteur de recrossing dans ce régime de basse température, cet effet n'est pas comparable à l'impact de l'effet tunnel qui, comme nous l'avons dit précédemment, accélère la réaction jusqu'à 60 ordres de grandeur.

Contents

List of Figures

List of Tables

Introduction

Computational molecular dynamics (MD) simulations are one of the most common tools to study the equilibrium and dynamical properties of a chemical system. This approach can be useful as a complement to experimental works and as a support to their interpretation. The first system studied with MD simulations was composed of 64 particles in a 1-D model in 1955 by Fermi, Pasta and Ulam [1]. About twenty years later Karplus and coworkers [2] were able to simulate the folding of a protein (bovine pancreatic trypsin inhibitor) using an empirical potential, for a total simulation time of 8.8 ps. Nowadays, thanks to the progress in computing power and algorithms, the scientific community is able to perform MD simulations of complex systems (up to hundreds of millions of atoms) on much longer timescales (up to microseconds) [3, 4]. Depending on the process under study and the associated size / timescale, different electronic descriptions can be used. In most cases, MD simulations are based on the Born-Oppehneimer approximation [5] which separates the motion of electrons and nuclei. Within this approach, it is possible to obtain a 3N potential energy surface on which nuclei move, either according to classical or to quantum mechanics. A typical way is to perform the electronic structure calculation on-the-fly: at each step of the dynamics the energy and the gradients of the system are computed and the dynamics is thus propagated. In this way, one can obtain a trajectory from which dynamical or equilibrium properties of the system can be extracted. The most common approach in MD simulations, is to treat nuclei in a classical way, within Newton's equations of motion. This kind of approach can lead to significant errors, especially when light nuclei such as protons are involved or at low temperatures when the quantum nature of the nuclei cannot be neglected.

This thesis mainly focuses on the study of reactive systems, exploring how Nuclear Quantum Effects (NQEs), in particular tunnelling and zero point energy (ZPE), can affect the dynamical and thermodynamical properties of chemical systems [6]. Indeed, NQEs are known to potentially influence the reactivity in chemical and biological systems [7], to correctly capture the complex and anomalous behaviour of water [8, 9] and to affect hydrogen bonds [10–13].

In principle, to include NQEs in the description of chemical reactions, one should solve the time-dependent Schrödinger equation for the atomic nuclei, but this is possible only for systems with few degrees of freedom. Some refined techniques have been designed to propagate the nuclear wave function in a more efficient manner, such as the Multi Configurational Time Dependent Hartree (MCTDH) [14], but it is still limited to tens of atoms. For this reason, different approximate methods were developed in the past in order to take

into account those effects without excessive computational effort, based on different types of semi-classical approximations, that allows to treat larger molecular systems. For the calculation of reaction rates, these methods can be divided into two groups: those that involve only static calculations and those that are based on trajectories, as in the classical picture, but accounting for NQEs via different approximations. In the former, it is used the Transition State Theory (TST) approximation combined with small curvature tunnelling (SCT) [15] or semi-classical TST (SCTST) [16] to correct for quantum effects such as tunnelling and ZPE. The second category instead, exploits reactive trajectories of the system in order to estimate the rate constant. An example of this category is the quasi-classical trajectory [17–19] or Linearized semi-classical Initial Value Representation [20], which combine a sampling of initial conditions according to quantum statistics (usually Wigner sampling) with classical trajectories to compute dynamical observables. Both methods have as main issue to perform the dynamics classically, and sample only initial conditions correctly. This predisposes the system to the phenomenon of ZPE leakage.

In this thesis, we will discuss in detail two trajectory-based approaches: ring polymer molecular dynamics (RPMD)[21], and the quantum thermal bath (QTB)[22]. The first one is based on the path integrals formalism outlined by Feynman, while the second one uses a generalized Langevin equation in order to introduce zero point energy (ZPE) in an otherwise classical simulation. Both of these methods were first applied to a model system, which represents a unimolecular fragmentation, to evaluate how accurately they capture NQEs. We then focus on two more complex reactions: the double proton transfer in DNA base pairs, and the Cope rearrangement of semibullvalene at low temperature. In order to study dynamical and thermodynamical properties, we combine enhanced sampling techniques, transition state theory and direct dynamics simulations in the RPMD framework.

To carry out the unimolecular fragmentation calculations, we used VENUS [23], in which we implemented the RPMD algorithm. For the last two systems, we developed a specific code which allows the use of RPMD interfaced with DFTB+ [24] and PLUMED [25] software. For more details, see Appendix C.

The thesis is organized as follows:

Chapter 1.

This Chapter starts with a brief overview of molecular dynamics simulation. We start with a general introduction, before presenting the Langevin thermostat in detail, then the tight-binding density functional theory (DFT) method is shortly reviewed. Finally, we provide a short overview of the methods used to obtain free energy profiles, focusing on metadynamics, umbrella sampling and to compute rate constants within

the quantum and classical frameworks.

Chapter 2.

In this Chapter, we explain in detail the methods used in this thesis in order to include NQEs in the calculation of dynamical properties: ring polymer molecular dynamics, and the quantum thermal bath. A subsection on the RPMD rate theory is also included.

Chapter 3.

In this Chapter, we compared the accuracy of the two methods, QTB and RPMD, for the evaluation of the rate constant of unimolecular fragmentation. Here we studied different systems: a model potential of CH_4 is modified in order to explore different ranges of temperature, and a benchmark with exact results is performed for a 1D Morse potential. We compare different methods to compute the rate constants and the associated activation energies.

Chapter 4.

Here we study the double proton transfer reaction in DNA base pairs. We performed simulations for an isolated guanine-cytosine base pair and for the same pair within a model DNA environment. The mechanism is analyzed through the umbrella sampling method and we compare the rate constants between classical and RPMD rate constants. Furthermore, we deuterate the system in order to compare the rate constants k^H and k^D and evaluate the kinetic isotopic effect.

Chapter 5.

A simple reaction is modelled in order to explore the influence of NQEs on reactions involving heavy atoms at low temperatures. The reaction is the Cope rearrangement of semibullvalene, the aim is to evaluate the accuracy of the TST, evaluating the recrossing factor. The relative contributions of tunnelling and ZPE are also discussed.

Part I

Methods

Chapter 1

Molecular Dynamics Simulations

This Chapter begins with a general overview of MD simulation and the different thermostats that can be used to fix the temperature. Additionally, the basic principles of the electronic semiempirical method density functional tight binding (DFTB) used in this thesis will be explained. Moreover, in the last part of this Chapter, an exploration it is provided of the methods used to obtain reaction rate constants and relevant free energy profiles from MD simulations.

1.1 Overview

Molecular dynamics is a computational technique, which allows us to follow the dynamical evolution of a system composed of N classical particles subject to a potential V. Given an initial set of $3N$ cartesian coordinates and $3N$ momenta (or velocities), the evolution of the system in time is obtained through the numerical integration of Newton's equation of motion:

$$F_i(t) = m\ddot{x}_i(t) = -\frac{\partial V(\vec{x})}{\partial x_i} \tag{1.1}$$

where i runs over the number of degrees of freedom ($3N$). $F_i(t)$ is the force at time t that acts on the particle, m is the mass, $\ddot{x}_i(t)$ the acceleration and V is the potential which is only function of the nuclear positions.

The outcome of an MD simulation is a trajectory of the system, that can be defined as a time sequence of points in the phase space (i.e. a set of $3N$ coordinates and $3N$ momenta).

Knowing the dynamics of the system, it is possible to obtain all the physical observables by computing their average along the trajectory. From statistical mechanics, the average of the observable A is computed over an ensemble of all accessible phase space configurations Γ:

$$\langle A \rangle_\Gamma = \int_\Gamma A(\Gamma) P(\Gamma) d\Gamma \tag{1.2}$$

where Γ are all the possible phase space configurations of the system and $P(\Gamma)$ is the probability distribution, which depends on the chosen ensemble. In the case of the canonical ensemble (NVT), $P(\Gamma)$ is proportional to the Boltzmann factor $e^{-H(\vec{x},\vec{p})/k_b T}$ with $H(\vec{x},\vec{p})$ the Hamiltonian of the system and k_b the Boltzmann constant. Thanks to the ergodic hypothesis, the average of A can also be calculated by its integration over time:

$$\langle A \rangle_\Gamma = \langle A \rangle_t = \int_0^{t_f} A(t) dt \tag{1.3}$$

This hypothesis ensures that the two ways of computing the average value of A (Equation 1.2 and Equation 1.3) lead to the same results, for t long enough and for optimal sampling of the phase space. This hypothesis allows us to obtain statistical properties from molecular dynamics simulations.

1.2 The canonical ensemble and Langevin thermostat

The integration of the equation of motion, Equation 1.1, leads to the so-called microcanonical ensemble (NVE), where the number of particles (N), the volume (V) and the total energy (E) of the system are kept constant. On the other hand, many dynamical and statistical properties are obtained at a given temperature. This means that one should be able to simulate the canonical ensemble (NVT), fixing the temperature at a certain value during the simulation. In order to control the temperature, different approaches can be followed, like scaling (Berendsen thermostat) [26], extended systems (Nosé-Hoover) [27] or dissipative methods (Langevin) [28]. The former can be considered the simplest and quickest way to thermalize the system since it simply rescales the velocities ($\vec{v}$) of the particles according to the desired temperature:

$$\vec{v}_{\text{scaled}} \leftarrow \lambda \vec{v} \tag{1.4}$$

with λ:

$$\lambda = \left[1 + \frac{\Delta t}{\tau_T}\left(\frac{T^*}{T} - 1\right)\right]^{\frac{1}{2}} \tag{1.5}$$

where $\vec{v}_{\text{scaled}}$ and $\vec{v}$ are respectively the new and the old velocities of the system, T^* is the desired temperature, T is the instantaneous temperature and τ_T is a coupling factor, homogeneous to a time. Another approach to fix the temperature during an MD simulation involves the inclusion of additional degrees of freedom, whereby the system can exchange energy (extended systems). One of the most common thermostats of this category is the Nosé Hoover [27] thermostat. The main drawback of this approach is that it is deterministic (strictly determined by the initial condition of the extended system coordinates), furthermore, it does not work properly for small systems, as it fails for example in fixing the temperature for a single Harmonic Oscillator [29]. To overcome this problem different modifications were proposed, leading to more efficient thermostats as Nosé-Poincaré [30] and Nosé-Hoover chains methods [31].

The approach used in this thesis in order to fix the desired temperature is the Langevin thermostat. This can be described by a stochastic equation, known as the Langevin equation. It finds its origin in the classical theory of Brownian motion [32], where it represents the dynamics of a macroscopic particle subject to random impacts. Those random collisions between the particle and the environment lead to two different effects:

1. A systematic effect: the surrounding molecules create frictional forces
2. A random effect: that is represented by a random driving force on the particle

For a single particle in 1-D, the stochastic equation used to describe this motion is the Generalized Langevin Equation (GLE) [33]:

$$m\ddot{x} = -\frac{dV}{dx} - \int_{t_0}^{t} dt^{'} K(t - t^{'})\dot{x}(t^{'}) + R(t) \tag{1.6}$$

where x is the position, m the mass of the particle, V the potential, K is the kernel which is a function of time, and characterizes the dependence of the friction force on the past values of the velocity $\dot{x}$ up to the time t. The Fluctuation Dissipation Theorem (FDT) [34] expresses the fact that for a system at thermal equilibrium, the power injected by the random

force and dissipated by the friction should compensate each other. The relation between the random force $R(t)$ and the kernel $K(t)$ is expressed by the following relation:

$$\langle R(t)R(t+\tau)\rangle = k_B T K(\tau) \tag{1.7}$$

In the particular case of a Markovian system (i.e. when there is no memory effect) the Equation 1.7 becomes:

$$\langle R(t)R(t+\tau)\rangle = 2k_B T\gamma m\delta(\tau) \tag{1.8}$$

where γ is the (constant) friction coefficient of the thermostat. The Langevin equation then becomes:

$$m\ddot{x} = -\frac{dV}{dx} - \gamma\dot{x} + R(t) \tag{1.9}$$

This last equation is the one used in this thesis, in order to fix the temperature during the simulations. Since the dynamics can be affected by the use of the thermostat, a small value of friction γ is recommended if one wants to obtain dynamical properties.

1.3 Born-Oppenheimer molecular dynamics

To obtain the full quantum dynamics of a system, in principle the quantum time-dependent Schrödinger equation should be solved for both nuclei and electrons. This is not possible for more than a few degrees of freedom, and for this reason, different approximations are made to obtain the dynamics of more complex systems. One of the most common approaches is to divide the motion of electrons and nuclei: obtaining the potential energy surface for the nuclei from the ground state of the time-independent Schrödinger equation for the electrons at fixed nuclei position. Nuclei are then generally evolved classically under the forces deriving from this potential. This approach is called Born-Oppenheimer molecular dynamics. The forces can be obtained by different electronic structure methods such as density functional theory (DFT) [35], semiempirical methods (such as AM1 [36], PM3 [37, 38] and PM6 [39]) or analytical methods (such as force fields, or analytical potentials). In this thesis, the electronic structure problem is solved through the semiempirical approach called tight binding DFT.

1.3.1 Electronic Structure: density functional tight binding

The density functional tight binding (DFTB) method is a semi-empirical method based on the tight binding theory (TB)[40] and density functional theory (DFT) [35, 41]. This method takes into account the variations of the electronic density self-consistently in the DFT. From this theory, it retains only a few terms and recasts them in a tight-binding Hamiltonian.

From the DFT theory, the ground state energy of an interacting electron system is written as:

$$E[\rho(\vec{r})] = T_s + E_{ext}[\rho(\vec{r})] + E_H[\rho(\vec{r})] + E_{NN} + E_{XC}[\rho(\vec{r})] \tag{1.10}$$

where T_s is the kinetic energy for a non-interacting system, E_{ext} is the external interaction (nuclei-electron interaction), E_H is the Hartree energy, E_{NN} is the energy of nuclei-nuclei interaction and E_{XC} is the exchange-correlation energy term. The electron density $\rho(\vec{r})$ is defined as:

$$\rho(\vec{r}) = \sum_{i=1}^{occ} \psi_i^*(\vec{r})\psi_i(\vec{r}) \tag{1.11}$$

with $\vec{r}$ the vector for the position of electrons, ψ_i the wavefunction for one electron orbital and i runs over the occupied orbitals. The electronic density can be obtained through a self-consistent method, by iteration of the Kohn-Sham equations:

$$\left[-\frac{1}{2}\nabla^2 + V_{eff}(\vec{r})\right]\psi_i = \epsilon_i\psi_i \tag{1.12}$$

where:

$$\begin{aligned} V_{eff}(\vec{r}) &= V_H(\vec{r}) + V_{ext}(\vec{r}) + V_{xc}(\vec{r}) \\ V_H &= \int \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} dr' \\ V_{ext} &= -\sum \frac{Z_a}{|\vec{r} - \vec{R}_a|} \\ V_{xc} &= \frac{\partial E_{xc}}{\partial \rho(\vec{r})} \end{aligned}$$

with a over the number of nuclei. These terms correspond to the terms in Equation 1.10 as follows: V_H is the Hartree potential, V_{ext} is the external nuclei-electron potential,

R_a is the position vector a-nuclei, while V_{xc} is the exchange-correlation potential. Using Kohn-Sham (KS) Equation (Eq. 1.12) the total energy written in Equation 1.10 can be recast as [41]:

$$E[\rho(\vec{r})] = \sum_{i=1}^{occ} \langle \psi_i | -\frac{1}{2}\nabla^2 + V_{ext} + V_H + V_{xc} | \psi_i \rangle - E_H[\rho(\vec{r})] - V_{xc}[\rho(\vec{r})] + E_{xc}[\rho(\vec{r})] + E_{NN} \tag{1.13}$$

The basic idea behind DFTB theory is to express the electronic density as the superposition of a reference density $\rho_0(\vec{r})$ (considered equal to a superposition of the densities of the neutral atoms of the system) plus a fluctuation (defined as $\Delta\rho(\vec{r})$):

$$\rho(\vec{r}) = \rho_0(\vec{r}) + \Delta\rho(\vec{r}) \tag{1.14}$$

Expanding the functional of the electronic density (Eq. 1.13) up to the third order of Taylor expansion [42] in $\Delta\rho$:

$$\begin{aligned} E[\rho_0 + \Delta\rho] = & \sum_{i=1}^{occ} \langle \psi_i | -\frac{1}{2}\nabla^2 + V_{ext}[\rho_0] + V_H[\rho_0] + V_{xc}[\rho_0] | \psi_i \rangle + \\ & - \frac{1}{2}\int\int \frac{\rho_0 \rho_0^{'}}{|\vec{r} - \vec{r}^{'}|} d\vec{r} d\vec{r}^{'} - \int V_{xc}[\rho_0]\rho_0 d\vec{r} + E_{xc}[\rho_0] + E_{NN} + \\ & + \frac{1}{2}\int\int dr dr^{'} \left(\frac{1}{|\vec{r} - \vec{r}^{'}|} + \frac{\delta^2 E_{xc}[\rho_0]}{\delta\rho\delta\rho^{'}} \right) \Delta\rho\Delta\rho^{'} + \\ & + \frac{1}{6}\int\int\int dr dr^{'} dr^{''} \left(\frac{\delta^3 E_{xc}}{\delta\rho\delta\rho^{'}\delta\rho^{''}} \Delta\rho\Delta\rho^{'}\Delta\rho^{''} \right) \end{aligned} \tag{1.15}$$

where the term $\Delta\rho$ or ρ_0 stand for $\Delta\rho(\vec{r})$ and $\rho_0(\vec{r})$, $\Delta\rho^{'}$ or $\rho_0^{'}$ stand for $\Delta\rho(\vec{r^{'}})$ and $\rho_0(\vec{r^{'}})$, while $\Delta\rho^{''}$ stands for $\Delta\rho(\vec{r}^{''})$. The first order term in $\Delta\rho$ vanishes [43] because of the variational theorem, and the therms $\Delta\rho^n$ for $n > 3$ are small and can be neglected . The different terms in Equation 1.15 can be separated as a function of the order in the Taylor expansion as follows:

$$E^{DFTB1} = \sum_{i=1}^{occ} \langle \psi_i | -\frac{1}{2}\nabla^2 + V_{ext}[\rho_0] + V_H[\rho_0] + V_{xc}[\rho_0] | \psi_i \rangle - \frac{1}{2}\int\int \frac{\rho_0 \rho_0^{'}}{|\vec{r} - \vec{r}^{'}|} d\vec{r} d\vec{r}^{'} + - \int V_{xc}[\rho_0]\rho_0 dr + E_{xc}[\rho_0] + E_{NN} \tag{1.16}$$

$$E^{DFTB2} = \frac{1}{2}\int\int \left(\frac{1}{|\vec{r} - \vec{r}^{'}|} + \frac{\delta^2 E_{xc}\rho_0}{\delta\rho\delta\rho^{'}} \right) \Delta\rho\Delta\rho^{'} \tag{1.17}$$

$$E^{DFTB3} = \frac{1}{6}\int\int\int \left(\frac{\delta^3 E_{xc}}{\delta\rho\delta\rho^{'}\delta\rho^{''}} \Delta\rho\Delta\rho^{'}\Delta\rho^{''} \right) \tag{1.18}$$

We can express, in agreement with the Tight Binding theory, the total energy of the system as the sum of two contributions:

$$E_{tot}^{TB} = E_{binding} + E_{rep} \tag{1.19}$$

Where the first term on the right of the equation describes the binding energy of the system, while the second is related to all the repulsive contributions. The term $E_{binding}$ corresponds to the first term in Equation 1.16, $\sum_{i=1}^{occ} \langle \psi_i | -\frac{1}{2}\nabla^2 + V_{ext}[\rho_0] + V_H[\rho_0] + V_{xc}[\rho_0] | \psi_i \rangle$, that is the Kohn-Sham Hamiltonian at the reference density ρ_0. Therefore, its expectation value is equal to sum of the electronic eigenvalues $\sum_{i=1}^{occ} \epsilon_i$. The basis set in which ψ_i are expanded is within LCAO ansatz and only the valence orbitals are considered (frozen core approximation). The last terms in Equation 1.16, represent the electrostatic energy of the a,b atom pair and can be approximated as a sum of two centre repulsion terms:

$$-\frac{1}{2}\int\int^{'} \frac{\rho_0 \rho_0^{'}}{|\vec{r} - \vec{r}^{'}|} d\vec{r} d\vec{r}^{'} - \int V_{xc}[\rho_0]\rho_0 dr + E_{xc}[\rho_0] + E_{NN} \simeq \frac{1}{2}\sum_{a,b} V_{rep,ab}[\rho_0^a, \rho_0^b] \tag{1.20}$$

All terms related to the variation of electron density appear in Equation 1.17 and 1.18. For what concerns the second order term, which appears in Equation 1.17, it expresses the response of an atom that modifies its charge state due to the influence of the charge of another atom, and it can be approximated as:

$$E^{DFTB2} \simeq \frac{1}{2}\sum_{a,b} \gamma_{a,b}(|R_a - R_b|)\Delta q_a \Delta q_b \tag{1.21}$$

Where $\Delta q_a = Q_a - Z_a$ is the Mulliken charge of the atom a, and $\gamma_{a,b}$ is the second derivative of the Hartree and exchange-correlation contributions with respect to the atomic charge.

Summarizing all the approximations, the energy expression for DFTB2 (i.e. second-order expansion of DFT) is:

$$E^{DFTB1+DFTB2} \simeq \sum_{i}^{occ} \epsilon_i + \frac{1}{2}\sum_{a,b}\gamma_{a,b}(|R_a - R_b|)\Delta q_a \Delta q_b + \sum_{a,b} V_{rep,ab}(R_{ab}) \tag{1.22}$$

Taking into account the third order expansion in Equation 1.18 we obtain a term which considers the derivative of the $\gamma_{a,b}$, called $\Gamma_{a,b}$, with respect to the charge. The final expression for DFTB3 is:

$$E^{DFTB1+DFTB2+DFTB3} \simeq \sum_{i}^{occ} \epsilon_i + \frac{1}{2}\sum_{a,b}\gamma_{a,b}(|R_a - R_b|)\Delta q_a \Delta q_b + \sum_{a,b} V_{rep,ab}(R_{ab}) + \frac{1}{3}\sum_{a,b}\Delta q_a^2 \Delta q_b \Gamma_{a,b} \tag{1.23}$$

Up to now we have seen the energy within the DFTB method. Now we will focus on the corresponding Hamiltonian. The corresponding Kohn-Sham (KS) approximated equations derived from this formalism are:

$$\left(\hat{h}^0 + \hat{h}^1 + \hat{h}^2\right)\psi_i = \epsilon_i \psi_i \tag{1.24}$$

Where the different terms of the KS operator ($\hat{h}^0 + \hat{h}^1 + \hat{h}^2$) correspond to the Hamiltonian when it is taking into account the first, second and third order expansion. Expanding the orbital ψ_i within LCAO approach leads to:

$$\psi_i(\vec{r}) = \sum_{\nu}^{N} C_{i,\nu}\phi_\nu(\vec{r} - \vec{R}_a) \tag{1.25}$$

Where $C_{i,\nu}$ are the coefficients, and ϕ_ν is the basis set of localized orbitals centred on atom a on which the function ψ_i is expanded. This leads to the secular equations:

$$\sum_{\nu}^{N} C_{i,\nu}(H_{\mu,\nu} - \epsilon_i S_{\mu,\nu}) = 0 \tag{1.26}$$

$$H_{\mu,\nu} = H^0_{\mu,\nu} + S_{\mu,\nu}\sum_j q_j \left(\frac{1}{2}(\gamma_{a,b} + \gamma_{b,j}) + \frac{1}{3}(q_a\Gamma_{a,j} + q_b\Gamma_{b,j}) + \frac{q_j}{6}(\Gamma_{j,a} + \Gamma_{j,b})\right) \tag{1.27}$$

where $S_{\mu,\nu}$ is the overlap matrix. These equations are equivalent to the KS equations in the DFTB approach. The orbitals in Equation 1.24 are considered within the frozen core approximation (i.e. only valence atomic orbitals), and the Hamiltonian and overlap matrix contain one and two-centre contributions, as in the standard TB approach. In contrast, its elements rather than empirical, are computed starting from DFT. For this reason, they can be calculated and tabulated in advance, so there is no need to recalculate them during the simulation. Furthermore the diagonal elements of h^0 are tabulated from DFT values within PBE exchange-correlation functional, while the off-diagonal elements are approximated with two-centre contributions [44]. In practice the Equations 1.26 and 1.27 are solved within a self-consistent approach: an initial guess of Δq_a is given, this allows us to obtain $H_{\mu,\nu}$, at this point the Equation 1.26 is used in order to get the coefficient $C_{i,\nu}$ and then a new value of Δq_a is achieved. Since this iteration is done on charges, rather than density as in DFT, it is called self-consistently-charge treatment (SCC).

The atomic forces can be obtained through the gradient of Equation 1.23. For example, on the atom a the force will be calculated as follows:

$$\vec{F}_a = -\sum_{i}^{occ}\sum_{\mu,\nu} C^*_{i,\mu}C_{i,\nu}\left[\nabla_a H^0_{\mu,\nu} - (\epsilon_i - h^1_{\mu,\nu})\nabla_a S_{\mu,\nu}\right] - \Delta q_a \sum_j (\nabla_a \gamma_{a,j})\delta q_j - \nabla_a E_{rep} \tag{1.28}$$

The advantage of this approach is the low computational cost while retaining the reactivity, its main approximations can be summarized as:

1. Expansion of the DFT functional energy

2. Expansion of the orbital ψ_i within LCAO approach (considering just the valence orbitals)

These approximations enable the possibility to study systems for which the standard DFT would require a too high computational cost, while retaining the reactivity, which is an advantage of the DFT with respect to force fields. The applications of DFTB are wide: it was shown to well describe the geometries and energies of organometallic complexes, especially if the third order is included [45–47]. Furthermore, it was extensively tested on biological molecules such as base-pair and DNA [48] or anticancer drugs [49, 50] where the interaction between an external molecule and DNA was studied. Moreover, there are a huge number of applications for carbon clusters and nanomaterials [51–53].

1.4 Dynamical and equilibrium properties

This Section focuses on the methods that can be used to obtain the rate constants and the free energy surface from MD simulations. For the rate constant, we start with the quantum approach, where the rate can be directly computed from the quantum side, flux or flux-side correlation functions, then go step by step to the classical picture and to the widely used transition state theory (TST) approximation. For the free energy surface instead, we review the enhanced sampling techniques (metadynamics and umbrella sampling) used to obtain the free energy surface for rare events in the present thesis.

1.4.1 Rate constants

The thermal equilibrium rate constant of a chemical reaction can be written as [54]:

$$k(T) = \frac{1}{Q} \lim_{t\to\infty} Tr\left[\hat{F} e^{\frac{iHtc^*}{\hbar}} h(\hat{s}) e^{-\frac{i\hat{H}tc}{\hbar}}\right] \tag{1.29}$$

with $t_c = t - i\hbar\beta/2$, β equal to $1/k_bT$ (with k_b the Boltzmann constant) and Q the reactant partition function. The operator $\hat{F}$ is the flux operator which depends on the reaction coordinate $\hat{s}$, the momentum $\hat{p}$ and the mass m as follows:

$$\hat{F} = \delta(\hat{s}) \left(\frac{\hat{p}}{m}\right) \tag{1.30}$$

Finally, h is the step function defined as:

$$\begin{cases} s > 0, & h(s) = 1 \\ s < 0, & h(s) = 0 \end{cases} \tag{1.31}$$

where $s = 0$ is the dividing surface.

The quantum correlation functions ($c_{f,f}$ flux, $c_{s,s}$ side and $c_{f,s}$ flux-side) are defined as follows:

$$c_{f,f}(t) = Tr\left[\hat{F}e^{\frac{i\hat{H}t_c^*}{\hbar}}\hat{F}e^{-\frac{i\hat{H}t_c}{\hbar}}\right] \tag{1.32}$$

$$c_{s,s}(t) = Tr\left[h(-\hat{s})e^{\frac{i\hat{H}t_c^*}{\hbar}}h(\hat{s})e^{-\frac{i\hat{H}t_c}{\hbar}}\right] \tag{1.33}$$

$$c_{f,s}(t) = Tr\left[\hat{F}e^{\frac{i\hat{H}t_c^*}{\hbar}}h(\hat{s})e^{-\frac{i\hat{H}t_c}{\hbar}}\right] \tag{1.34}$$

The three correlation functions just mentioned, are related by the following math operations:

$$c_{f,f}(t) = \frac{\mathrm{d}c_{f,s}(t)}{\mathrm{d}t} \tag{1.35}$$

$$c_{f,s}(t) = \frac{\mathrm{d}c_{s,s}(t)}{\mathrm{d}t} \tag{1.36}$$

So the relation between the thermal rate constant and the different correlation functions can be written as:

$$k(T) = \frac{1}{Q}\int_0^\infty c_{f,f}(t)dt \tag{1.37}$$

$$k(T) = \frac{1}{Q}\lim_{t\to\infty}\frac{dc_{s,s}(t)}{dt} \tag{1.38}$$

$$k(T) = \frac{1}{Q}\lim_{t\to\infty} c_{f,s}(t) \tag{1.39}$$

This shows how the quantum thermal rate constant can be obtained from correlation functions. In order to have the information to reconstruct one of the three correlation functions, one should perform full quantum dynamics. An application to a simple 1D model is given in Chapter 3.

In the past, different methods were developed in order to obtain an approximation (semiclassical) to the true rate constant. These semiclassical methods take into account the nuclear quantum behaviour of nuclei. Examples of these methods are the classical Wigner model [55], forward-backwards semiclassical dynamics [56] or ring polymer molecular dynamics [21]. The latter will be explained in detail in Section 2.1.

Classical Picture

A further approximation that can be made, is to consider the system fully classically (the

nuclei are treated as classical objects). To obtain the classical counterpart of the thermal rate constant, the trace Tr must be replaced with a classical phase space average, and the operators with the corresponding classical functions. The flux operator in Eq. 1.30 can be written classically as:

$$F(p,x) = \frac{p}{m}\delta(s) \tag{1.40}$$

where, as before (see Equation 1.30) s is the reaction coordinate which is a function of the positions, and the dividing surface is located at $s = 0$. Within the classical picture the step function h remains a simple function of the reaction coordinate s.

The classical side-flux correlation function becomes:

$$c_{f,s}^{cl}(t) = \frac{1}{2\pi\hbar}\int dp_0 \int dx_0 e^{-\beta H(p_0,x_0)}\frac{p_0}{m}\delta(s_0)h(s_t) \tag{1.41}$$

Where x_0 and p_0 are the initial conditions, sampled according to Boltzmann probability density. This last equation of the classical side-flux correlation function, has some similarity with its quantum counterpart, as it is also odd and independent of the location of the dividing surface, in the limit of $t \to \infty$. An important difference occurs when it approaches $t \to 0$, since in the classical case the correlation function is discontinuous at $t = 0$, while in the quantum expression, the correlation function is continuous for $t \to 0$. Analyzing the limit for $t \to 0^+$ of the classical flux-side correlation function in Eq. 1.41 it can be shown that (for a 1D case $s \longleftrightarrow x$):

$$\begin{aligned}
\lim_{t\to 0+} c_{f,s}^{cl}(t) &= \frac{1}{2\pi\hbar}\int dp_0 \int dx_0 e^{-\beta H(x_0,p_0)}\frac{p_0}{m}\delta(x_0)h(p_0) \\
&= \frac{1}{2\pi\hbar}\int_0^\infty dp_0 \frac{p_0}{m}e^{-\beta p_0^2/2m}\int dx_0 e^{-\beta V(x_0)}\delta(x_0) \\
&= \frac{1}{2\pi\beta\hbar}e^{-\beta V(x=0)}
\end{aligned}$$

Inserting this last equation in Eq. 1.39 and choosing the dividing surface at the highest point of the potential ($x = x^\ddagger$) gives the well-known classical transition state theory (TST) [57] result:

$$k_{cl}(T)^{TST} = \frac{1}{Q}\lim_{t\to 0^+} c_{f,s}^{cl}(t) = \frac{k_B T}{h}\frac{1}{Q}e^{-\beta V(x^\ddagger)} \tag{1.42}$$

with h the Planck constant. The TST allows us to compute the approximated rate constants, it assumes that the reactant is in equilibrium with the activated complex (transition state) and that once the system passes through the dividing surface, it cannot come back (no

recrossing). The result expressed in Equation 1.42, in comparison to Equation 1.29 depends on the location of the dividing surface, due to the exponential factor.

The extension of this to a 3N degrees of freedom will become:

$$k_{cl}(T)^{TST} = \frac{k_B T}{h} \frac{Q^{\ddagger}}{Q} e^{-\beta V(s^{\ddagger})} = \frac{k_B T}{h} e^{-\beta \Delta G^{\ddagger}} \tag{1.43}$$

Where we explicit the partition function of the transition state ($Q^{\ddagger}$) and of the reactant (Q). The term $\Delta G^{\ddagger}$ refers to the free energy value at the transition state.

Applying TST theory can have two main advantages:

- The rate constant will be always the upper bond of the true rate constant, since within the TST approximation the recrossing is neglected
- The rate constant can be obtained only from the energy at the transition state ($V(s^{\ddagger})$), and the vibrational frequencies are needed to calculate the partition function Q, thus there is no need of performing trajectories

If one considers the recrossing, the rate constant can be written as:

$$k(T) = k(T)^{TST} \kappa \tag{1.44}$$

where κ is the value in the limit of $t \to \infty$ of the recrossing factor [58]. One of the main assumptions within the TST approximation is to consider all the trajectories reactive: once the system passes through the bottleneck, individuated as the dividing surface ($s = 0$) the system will go into the products and there is no chance to come back. Including this correction factor, κ, it is possible to take into account the recrossing. Considering the dividing surface as $s = 0$, the recrossing factor can be written as an average of the flux side correlation function:

$$\kappa(t) = \frac{\langle \delta(s_0) \dot{s_0} h(s_t) \rangle}{\langle \delta(s_0) \dot{s_0} h(\dot{s_0}) \rangle} \tag{1.45}$$

where at the numerator the correlation is between the flux given by $\dot{s_0}$ on the dividing surface at time 0 and the side function for the position at time t which gives 1 when the reaction coordinate is on the side of the products and 0 when it is on the side of reactants. An application of this will be explained in detail in Chapter 5 with the Cope rearrangement reaction of semibullvalene.

In this thesis, we also used a direct method to calculate the rate constants, as will be

better described in Chapter 3: an ensemble of reactive trajectories is performed, then we can determine if for each time window and for each trajectory the system is still in the reactant state. From that, we can obtain a population decay and fit it with an exponential function. This is the case of unimolecular fragmentation as discussed in Chapter 3.

1.4.2 Free energy surface

Two of the main drawbacks of MD simulations are the limited time scale and the limited sampling of conformational space. In order to follow a reaction, it is useful to reduce the dimensionality of the analysis, so that the process (as a reaction or a conformational change) is analyzed on a reduced set of variables that describe correctly the mechanism under observation. These variables are called collective variables (CVs) and are typically expressed through a function of the atomic coordinates $s(q)$. For example in case of a conformational change, a reasonable CV could be a relevant torsional angle, or in case of a unimolecular fragmentation the length of the bond as it is shown in Figure 1.1.

In the canonical ensemble when analyzing a system through a particular set of CVs, the probability $P(s)$ of a certain value s is defined as:

$$P(s) \propto \int dq \exp\left(-\frac{V(q)}{k_B T}\right) \delta(s(q) - s) \tag{1.46}$$

where $V(q)$ is the potential, k_B the Boltzmann constant and T the temperature. The relation between the free energy for a certain value called s and the potential energy is:

$$F(s) = -k_B T \log\left[\int dq \delta(s(q) - s) \exp\left(-\frac{V(q)}{k_B T}\right)\right] \tag{1.47}$$

where $F(s)$ is the free energy, the $\delta(s(q) - s)$ inside the integral selects the configurations which correspond to the value s. The free energy obtained using Equation 1.47, requires computing a multidimensional integral in 3N dimension. For this reason, $F(s)$ is usually obtained through conformational sampling using MD or Monte Carlo approaches. In this way, the free energy landscape of the mechanism can be obtained through:

$$F(s) = -k_B T \log P(s) \tag{1.48}$$

Reducing the set of variables, in order to study and obtain the free energy of a system, is not enough for mechanisms that require a long time in order to explore a certain region of the phase space. These processes can have a high barrier, or be associated with multiple

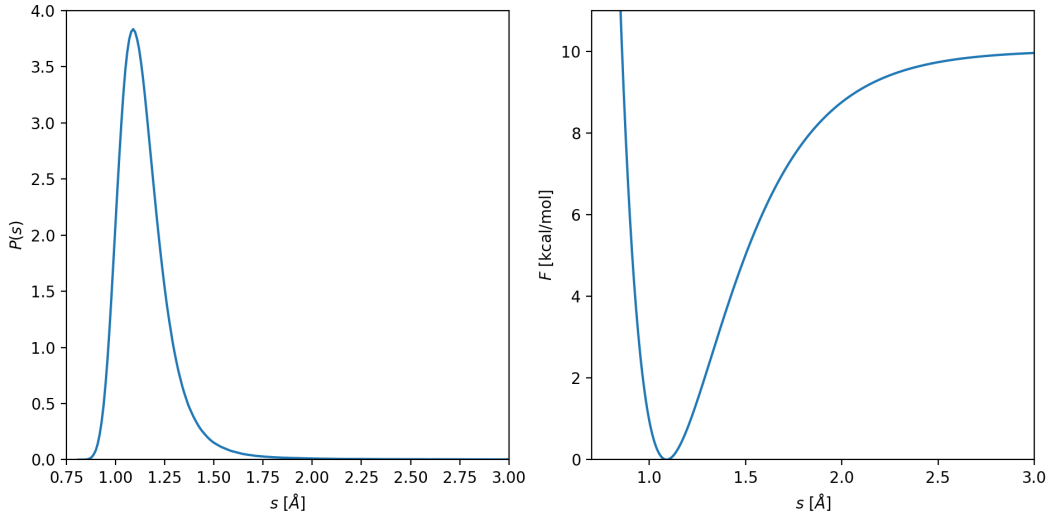

Figure 1.1: Example of free energy (right) obtained from the Probability Distribution of the CV (left). The model is a simple Morse potential. In this particular case, the CV corresponds to the atomic coordinate.

paths. To overcome these issues, a bias potential is added $\tilde{V}(s(q))$ to the unbiased potential $V(q)$ that would be used in a conventional MD simulation. This approach is used in a category of methods called "biased sampling methods". In this way, the system is forced to explore the region of the potential energy surface of interest, away from the minima. Whatever CVs are chosen to observe the process, it is fundamental to recognize the effect of the bias potential and to remove it after the simulation is performed. In particular, the biased distribution of the CV is expressed as:

$$P^{'}(s) \propto \exp\left(-\frac{V(q)+\tilde{V}(s(q))}{k_B T}\right)\delta(s(q)-s) \propto \exp\left(-\frac{\tilde{V}(s(q))}{k_B T}\right)P(s) \tag{1.49}$$

So the biased free energy landscape will be:

$$F(s) = -F^{'}(s) - \tilde{V}(s) + f \tag{1.50}$$

where $F^{'}(s)$ is the biased free energy obtained from $-k_b T \log P^{'}(s)$, while f is a constant which depends on the partition functions (biased and unbiased).

This means that the effect of the bias potential on the free energy landscape is additive, and so to recover the unbiased free energy, the bias potential must be subtracted a posteriori from the biased free energy. This chapter means to present an overview of two bias sampling methods used in this thesis: metadynamics[59][60] and umbrella sampling[61].

Metadynamics

Metadynamics simulations are a powerful technique that combines molecular dynamics simulation with statistical mechanics.

The basic idea behind this method is to add to the actual potential of the simulation an external one, called "bias" in order to promote the exploration of the potential energy surface (PES) and overcome barriers that could not be overcome in a conventional MD simulation. In order to achieve it, a time-dependent potential is added to the potential energy of the system. This bias will be tuned in order to drive the system towards the region of interest of the PES and will discourage the system from revisiting the previously sampled regions of the free energy landscape.

The time-dependent bias potential takes the form of a sum of gaussian functions, and it is added to the unbiased potential V at regular intervals of time $(1t, 2t..kt)$. For 1D case:

$$\tilde{V}(s,t) = \sum_{kt<t} W(kt) \exp\left(- \sum_{i=1}^{NumCVs} \frac{(s_i - s_i(q(kt))^2}{2\sigma_i^2} \right) \tag{1.51}$$

where $W(kt)$ is the height and σ_i is the width of the gaussian functions on the i CV. In the long time limit (when the probability becomes flat) it is possible to recover the free energy $F(s)$ from the following equation[62]:

$$F(s) = -\lim_{t\to\infty} \tilde{V}(s(q)) + C \tag{1.52}$$

The main drawback of this method is the necessity to wisely choose the parameters of the gaussian functions, and the CVs before performing the metadynamics simulation. Moreover, it is not trivial to decide when the simulation must end.

Umbrella sampling

As with metadynamics, umbrella sampling allows for obtaining the free energy landscape of a chemical process. A bias potential is applied to the system in order to sample a specific region of phase space. In this case, the bias potential is usually a harmonic one, which confines the system to a particular region of interest in phase space. The equation of the harmonic potential centred at a value s_0 of CV coordinate is:

$$\tilde{V}(s) = \frac{k}{2}(s - s_0)^2 \tag{1.53}$$

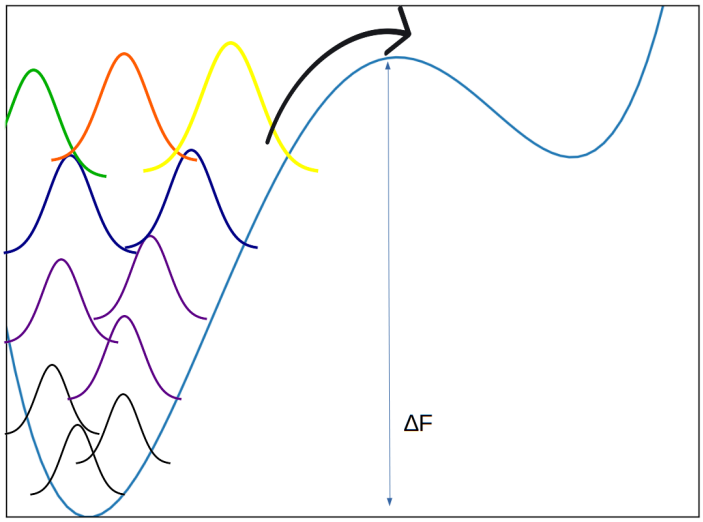

Figure 1.2: This graph presents a double well potential ($y(x) = \frac{1}{4}x^4 - x^2 + 0.5x$) as a model for a generic reaction coordinate (as can be for a single proton transfer). If $\Delta F \gg k_B T$ a bias sampling method is needed. In this case, the method depicted in this graph is metadynamics: starting from the minimum a certain number of gaussian functions is added in order to overcome the barrier.

where k is the constant of the harmonic potential, centred on the CV s_0. The corresponding probability distribution is proportional to:

$$P^{'}(s) \propto P(s) \exp\left(-\frac{k(s-s_0)^2}{2k_B T}\right) \tag{1.54}$$

The unbiased free energy landscape will be recovered using the weighted histogram analysis method (WHAM) [63]. The main advantage of umbrella sampling is that the different simulations can run independently and thus in parallel, and it is possible to add one or more windows after the first set of simulations has been performed. The only main drawback is the need to know the mechanism behind the process under observation: imagine for example a double proton transfer described by two CVs, which are represented as the difference of distances between donor-Proton and acceptor-Proton (in such a way that when this value is zero, the proton is in the middle between the donor and acceptor). If we know from the beginning that the process occurs through a synchronous and concerted mechanism, the centres of the biasing potential can be chosen only for the symmetric value of these distances. On the other hand, if the mechanism is step-wise, the potential will need to be centred for each possible combination of these two distances.

Weighted histogram analysis method (WHAM)

The WHAM algorithm is a way of reconstructing the free energy surface of a system

from umbrella sampling data. The algorithm works by first constructing a histogram of data obtained from umbrella sampling simulations, and then using this histogram to estimate the free energy as a function of the CVs. The bias potential in Equation 1.53 is centred at different values of s_0. For each value of s_0 a simulation is performed, and so each simulation (usually referred to as "window") explores a portion of a phase space in which the system is constricted through the harmonic bias potential. Each i-th window will be related to a bias potential called $\tilde{V}_i$ (same of Equation 1.53), and will be associated with a biased probability. From this, the unbiased probability $P(s)$ can be obtained by iteration [64, 65]. For details on the equation used see Ref. [63].

Chapter 2

Nuclear quantum effects in molecular dynamics simulation

This Chapter we review the two methods used in this thesis in order to include nuclear quantum effects in molecular dynamics simulations. We start with an extensive review of the path integral theory developed by Feynman, in order to explain in detail the ring polymer molecular dynamics method. Later, we present more briefly the quantum thermal bath, an approach which uses a generalized Langevin equation in order to account approximately for zero-point energy effects. For simplicity, we will use 1D notations, but the results can easily be generalized to molecular systems with multiple degrees of freedom.

2.1 Path integral theory

2.1.1 Formalism

Path integral theory was developed by Feynman in 1948 [66]. It was presented as a third approach, beside that of Schrödinger and that of Heisenberg, to describe non-relativistic quantum mechanics.

The core of this third formalism is to express the quantum time propagation operator as a sum over all possible paths, where each of them contributes with the same amplitude but a different phase. The matrix element of the propagation operator within Dirac's formalism is:

$$U(x, x^{'}; t) = \langle x^{'} | \exp\left(-\frac{i\hat{H}t}{\hbar}\right) |x\rangle \tag{2.1}$$

where $\hat{H}$ is the Hamiltonian, t is the time, and $\hbar$ is the Planck constant divided by 2π. Within the Feynman picture, the quantum time operator can be rewritten as:

$$U(x, x^{'}, t) = \sum_{\text{all paths x}(\tau)} \exp\left[\frac{i}{\hbar} S(x(\tau))\right] \tag{2.2}$$

where $S(x(\tau))$ is the classical action, the functional of the trajectory that can be obtained by integrating the Lagrangian ($L[x(\tau), \dot{x}(\tau)]$) as follows:

$$S(x(\tau)) = \int_0^t L[x(\tau), \dot{x}(\tau)] d\tau \tag{2.3}$$

Thanks to the similarity between the quantum time propagation operator expressed in Equation 2.1 and the canonical density operator $\hat{\rho}$, the path integral approach can be used to obtain the canonical partition function $Z(\beta)$ and therefore the thermodynamics properties of a chemical system.

The matrix elements of the canonical density operator are:

$$\rho(x, x^{'}; \beta) = \frac{1}{Z(\beta)} \langle x^{'} | \exp\Big(-\beta\hat{H}\Big) |x\rangle \tag{2.4}$$

where $\beta = \frac{1}{k_B T}$.

The connection between the two operators, $\hat{\rho}$ and $\hat{U}$, can be seen as a Wick rotation (Figure 2.1), which allows to do all the procedures to obtain a path integral expression for the canonical density operator. The canonical density operator can be obtained from the quantum time operator calculating it at an imaginary time, $t \longleftrightarrow -i\beta\hbar$.

In order to obtain a path integral expression for the canonical density operator in Equation 2.4, the first step is to apply the Trotter theorem and split the canonical density operator in an infinite number of slices (P) as:

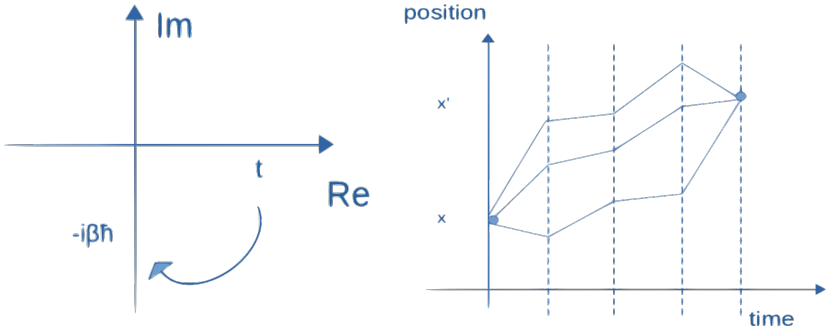

Figure 2.1: Left: Wick rotation from the real to imaginary axes. As it can be seen t and β can be considered as two coordinates in the imaginary plane. Right: Schematic representation of different paths between the initial state x and final state $x^{'}$. Figure adapted from ref [67].

$$\exp\left(-\beta\hat{H}\right) = \lim_{P\to\infty}\left[\exp\left(-\frac{\beta\hat{V}}{2P}\right)\exp\left(-\frac{\beta\hat{K}}{P}\right)\exp\left(-\frac{\beta\hat{V}}{2P}\right)\right]^{P} \tag{2.5}$$

where $\hat{V}$ is the potential which depends on the position operator $\hat{x}$, and $\hat{K}$ is the kinetic which depends on the momentum operator $\hat{p}$.

Denoting the product of the three exponential terms in Equation 2.5 as a new operator $\hat{\gamma}$ and inserting this term in Equation 2.4:

$$\rho(x, x^{'}; \beta) = \lim_{P\to\infty}\left\langle x^{'}\left|\hat{\gamma}^{P}\right|x\right\rangle \tag{2.6}$$

To expand the term in Equation 2.6, the identity operator is inserted between each $\hat{\gamma}$. A general matrix element can be written as:

$$\langle x_{k+1}|\,\hat{\gamma}\,|x_k\rangle = \langle x_{k+1}|\,[e^{-\frac{-\beta\hat{V}}{2P}}e^{-\frac{-\beta\hat{K}}{P}}e^{-\frac{-\beta\hat{V}}{2P}}]\,|x_k\rangle \tag{2.7}$$

Since $\hat{V} = V(\hat{x})$ is a function of the position operator, the $|x_k\rangle$ and $\langle x_{k+1}|$ are eigenvectors of the potential operator, with eigenvalues $V(x_k)$ and $V(x_{k+1})$. For the kinetic energy which is a function of the momentum operator, the resolution of identity must be inserted in the momentum basis. In this way, the kinetic operator will act directly on the $|p\rangle$ ket, as follows:

$$\langle x_{k+1}|e^{\frac{-\beta\hat{K}}{P}}|x_k\rangle = \int dp\langle x_{k+1}|\exp\left(-\frac{\beta\hat{K}}{P}\right)|p\rangle\langle p|x_k\rangle$$

$$= \left(\frac{mP}{2\pi\beta\hbar^2}\right)^{\frac{1}{2}} \exp\left[\left(-\frac{mP}{2\beta\hbar^2}\right)(x_{k+1}-x_k)^2\right] \tag{2.8}$$

By replacing Equation 2.1.1 in Equation 2.6 we can rewrite the canonical density matrix within the path integral approach:

$$\rho(x,x';\beta) = \frac{1}{Z(\beta)} \lim_{P\to\infty} \left(\frac{mP}{2\pi\beta\hbar^2}\right)^{P/2} \int dx_2...dx_P \exp\left[-\frac{1}{\hbar}\sum_{k=1}^{P}\left[\frac{mP}{2\beta\hbar}(x_{k+1}-x_k)^2 + \frac{\beta\hbar}{2P}(V(x_{k+1})+V(x_k))\right]\right]\Big|_{x_1=x}^{x_{P+1}=x'} \tag{2.9}$$

In Equation 2.9 the quantum kinetic energy appears as a harmonic nearest-neighbour coupling term between points along the same path, where the spring constant is equal to $mP/\beta^2\hbar^2$. Note that x_1 is equal to x (initial state) while x_{P+1} is equal to x' (final state). The integration over the intermediate terms $(x_2..x_P)$ represents all possible paths between x and x' as can be seen in the right panel of Figure 2.1.

The canonical partition function is fundamental in order to obtain the thermodynamical properties of a chemical system and it is written as the trace of the canonical density matrix:

$$Z(\beta) = Tr\left[\exp\left(-\beta\hat{H}\right)\right] = \int \langle x|\exp\left(-\beta\hat{H}\right)|x\rangle\, dx \tag{2.10}$$

Setting the cyclic condition $x_{P+1} = x_1$ in the Equation 2.9 it is possible to integrate over the diagonal element and the Equation 2.10 becomes:

$$Z(\beta) = \lim_{P\to\infty}\left(\frac{mP}{2\pi\beta\hbar^2}\right)^{\frac{P}{2}} \int dx_1..dx_P \exp\left[-\beta\phi(x_1..x_P)\right] \tag{2.11}$$

where $\phi(x_1..x_P)$ is an explicit potential, that acts on the replicas $(x_1,..,x_P)$ (as shown in Figure 2.2), defined by the following equation:

$$\phi(x_1..x_P) = \sum_{k=1}^{P}\left[\frac{1}{2}m\omega_P^2(x_k-x_{k+1})^2 + \frac{1}{P}V(x_k)\right] \tag{2.12}$$

The frequency $\omega_P = \sqrt{P}/(\beta\hbar)$ is the frequency of the harmonic springs between neighbouring replicas of the same physical atom. The Equation 2.11 is equivalent to the classical configuration partition function for P classical particles, under the potential $\phi(x_1..x_P)$. Because of the cyclic condition, the replicas of the system are usually called "beads", due to the similarity of the ring polymer with a necklace (Figure 2.2).

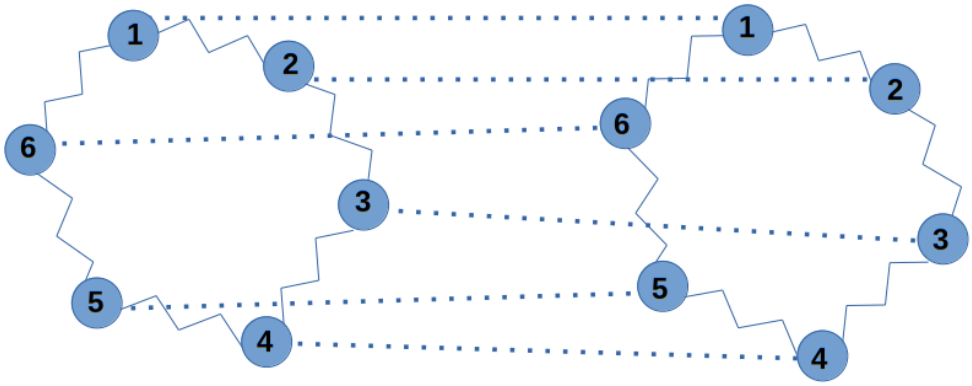

Figure 2.2: The figure describes two atoms depicted in the path integral formalism, where each atom is represented by P beads connected by harmonic springs. The dashed lines represent the physical interaction potential V that acts between beads of the same bead index. Figure adapted from ref [67]

Note that in the high-temperature limit, $T \to \infty$, $\beta \to 0$, the spring constant becomes infinite, causing the collapse of the cyclic polymer in one point (as for the classical case). In other terms, classical representation is recovered for $T \to \infty$.

From the path integral expression for the canonical partition function (Eq. 2.11), we can obtain the average value for an operator $\hat{A}$ which depends on position (local operator):

$$\langle A(\hat{x}) \rangle = \frac{1}{Z(\beta)} Tr \left[\hat{A} \exp\left(-\beta \hat{H} \right) \right] \tag{2.13}$$

which becomes:

$$\begin{aligned} \langle A(\hat{x}) \rangle = \frac{1}{Z(\beta)} \lim_{P \to \infty} \left(\frac{mP}{2\pi\beta\hbar^2} \right)^{\frac{P}{2}} \int dx_1..dx_P A_P(x_1..x_P) \\ \exp\left[-\frac{1}{\hbar} \sum_{k=1}^{P} \left[\frac{mP}{2\beta\hbar} (x_k - x_{k+1})^2 + \frac{\beta\hbar}{P} V(x_k) \right] \right] \end{aligned} \tag{2.14}$$

where $A_P(x_1..x_P)$ is defined as:

$$A_P(x) = \frac{1}{P} \sum_{k=1}^{P} A(x_k) \tag{2.15}$$

Once this classical equivalence expressed in Equation 2.11 has been constructed, the numerical methods that have been developed in the context of classical statistical mechan-

ics such as Monte Carlo (MC) or molecular dynamics (MD) can also be used in order to obtain the thermodynamical properties of the system. Those methods within the path integralformalism are called path integral Monte Carlo (PIMC)[68] or path integral molecular dynamics (PIMD) [69]. The first to be developed was PIMC, and it was used in quantum chemistry and condensed phase systems [70] but also to describe the behaviour of bosonic liquid at ultra-low temperature correctly [71].

2.1.2 Path integral molecular dynamics

In the path integral formulation to obtain the partition function (Equation 2.11), no dynamics is involved. In order to perform a molecular dynamics simulation, the momenta of the particles involved are inserted in the integral of the Equation 2.11:

$$Z(\beta) = \lim_{P\to\infty} \frac{1}{(2\pi\hbar)^P} \int dp_1..dp_P \int dx_1..dx_P \exp\left[-\beta \sum_{k=1}^{P} \frac{p_k^2}{2m'} + \phi(x_1..x_P)\right] \tag{2.16}$$

where ϕ is defined in Equation 2.12. Therefore the Hamiltonian of the system is:

$$H_P(p,x) = \sum_{j=1}^{P} \left(\frac{p_j^2}{2m'} + \frac{1}{2} m\omega_p^2 (x_j - x_{j+1})^2 + \frac{1}{P} V(x_j) \right) \tag{2.17}$$

where p and x are the vectors which contain the momenta and the positions of all the beads. From the equation 2.17 the equations of motion can be obtained:

$$\begin{aligned} \dot{x_j} &= +\frac{\partial H_P}{\partial p_j} = \frac{p_j}{m'} \\ \dot{p_j} &= -\frac{\partial H_P}{\partial x_j} = -m\omega_P^2 (2x_j - x_{j+1} - x_{j-1}) - \frac{\partial V(x_j)}{\partial x_j} \end{aligned}$$

In this way, it is possible to run a molecular dynamics simulation (PIMD [69]), and obtain (as for PIMC) the exact quantum thermodynamical properties of a chemical system, in the limit of an infinite number of beads ($P \to \infty$). In practice the number of beads required to estimate the quantum behaviour of the system depends on its frequency and on the temperature at which the simulation is performed. As a general (approximate) rule, the number of beads must be greater than the ratio between the vibrational energy (considering the highest frequency of the system) and the thermal energy $k_B T$.

2.2 Ring polymer molecular dynamics

As discussed in the previous section, both PIMC and PIMD approaches are useful to obtain the thermodynamical properties of a chemical system. If one is interested in the dynamical properties (such as diffusion coefficient or rate constants), some approximations must be done. Craig and Manolopolous in 2004 [21] developed the ring polymer molecular dynamics (RPMD) method which can be seen as one particular type of PIMD simulation, which allows computing approximate quantum dynamical properties. It exploits PIMD trajectories to obtain the approximate Kubo transform correlation function (one of the multiple possible ways to express quantum correlation functions), in this way also dynamical properties can be obtained from molecular dynamics simulation, with the inclusion of NQEs such as the zero point energy or the tunnelling.

As the first step the mass $m^{'}$, in the Equation 2.17 is fixed to the physical mass of the atoms, so we can write the Hamiltonian of the system $H_P(p,x)$, with p and x as the momenta and position of all the beads:

$$H_P(p,x) = \sum_{k=1}^{P} \left(\frac{p_k^2}{2m} + \frac{1}{2} m\omega_p^2 (x_k - x_{k+1})^2 + V(x_k) \right) \tag{2.18}$$

Within the RPMD formalism, the correlation function of two position dependent operator, $\hat{A}$ and $\hat{B}$ can be written as:

$$\langle A(0)B(t)\rangle_P = \frac{1}{(2\pi\hbar)^P Z_P} \int dp_0 \int dx_0 \exp\left[-\frac{\beta H_P(x_0,p_0)}{P} \right] A_P(x_0) B_P(x_t) \tag{2.19}$$

where A_P and B_P are defined in Equation 2.15.

For $t \to 0$ and $P \to \infty$ this correlation function is equal to the Kubo correlation function:

$$\widetilde{C}_{AB}(t) = \frac{1}{\beta Z} \int_0^\beta d\lambda Tr \left[\exp\left(-(\beta-\lambda)\hat{H}\right) \hat{A} \exp\left(-\lambda\hat{H}\right) \exp\left(+i\hat{H}t/\hbar)\right) \hat{B} \exp\left(-i\hat{H}t/\hbar\right) \right] \tag{2.20}$$

The assumption of the RPMD approach is to consider the similarity between the Equation 2.19 and 2.20 true even for $t > 0$.

It was extensively studied on analytical models such as the quartic double-well potential [72], it shows good agreement with quantum exact theory results for bimolecular collision [73] and it was applied also to gas-surface reactions [74].

The two main drawbacks of RPMD are the computational cost, equal to P times a classical simulation, and the description of vibrational spectra. This is due to a resonance problem: the excited modes of the ring polymer could be in resonance during the simulation with one of the degrees of freedom of the system. This resonance results in a splitting of certain peaks in the vibrational spectra, which are not related to the physical frequencies of the system but to the internal frequencies of the ring polymer which do not correspond to any physical resonance. In order to avoid this problem, Ceriotti and coworker [75] modified the algorithm RPMD, coupling the harmonic motion of the beads with a Langevin thermostat (T-RPMD). Applying this strong damping on the internal modes of the ring polymer, make these spurious peaks disappear from the spectra. This is the one used in this thesis, and the details of the algorithm can be found in Appendix C.

2.2.1 Rate constant from RPMD

As shown before, RPMD is an approximate method which exploits the similarity between the quantum Kubo-transformed correlation function and the classical one within an extended phase space. It therefore allows to calculate any dynamical observable that can be expressed from such correlation functions, as for example the reaction rate constants presented in Subsection 1.4.1 (Eq. 1.39).

The rate constant can be obtained from the flux-side correlation function as:

$$k(T) = \frac{1}{Q} \lim_{t\to\infty} c_{f,s}(t)$$

Taking into account that the centroid of the ring polymer is expressed as:

$$\bar{x} = \frac{1}{P}\sum_{k}^{P} x_k \tag{2.21}$$

we can write the flux-side correlation function within RPMD framework as [21, 76]:

$$\widetilde{C}_{f,s}(t) \simeq \frac{1}{(2\pi\hbar)^P} \int dp_0 \int dx_0 e^{-\beta_P H_P(x_0,p_0)} \delta(s(\bar{x_0})) \frac{\bar{p_0}}{m} h(s(\bar{x}_t)) \tag{2.22}$$

where the term $s(\bar{x_0})$, $\bar{p_0}$ and $s(\bar{x_t})$ are respectively the reaction coordinate at the centroid position at time zero, the centroid momenta at time zero and the reaction coordinate at centroid position at time t.

The rate obtained within RPMD has no variational principle (so it does not give the upper bound to the true rate constant). In the limit $t \to 0^+$ of the flux-side correlation function (Eq. 2.22), we obtain the so-called the RPMD-TST rate, which is an upper bound of the RPMD rate (in the same way as the classical TST rate is an upper bond to the full classical reaction rate). As described in Section 1.4, applying the TST means doing several approximations, in particular the assumption that once the system passes through the dividing surface it does not come back. In Section 1.4 we have seen that in order to consider this recrossing, we can multiply the classical TST rate by a factor $\kappa(t \to \infty)$ which is expressed as a correlation function (Equation 1.45). In the same way, the RPMD rate constant can be obtained by multiplying the RPMD-TST rate by a recrossing factor with a similar expression, only involving the centroid position and momenta:

$$\kappa(t) = \frac{\langle \delta(s(\bar{x_0}))\dot{s}(\bar{x_0})h(s(\bar{x_t}))\rangle}{\langle \delta(s(\bar{x_0}))\dot{s}(\bar{x_0})h(\dot{s}(\bar{x_0}))\rangle} \tag{2.23}$$

An application will be presented in Chapter 5.

2.3 Quantum thermal bath

The quantum thermal bath (QTB) method is based on the use of a generalized Langevin equation in order to introduce approximately the zero point energy of the system during the simulation. To explain it, it is useful to recall the (classical) Langevin equation in the Markovian limit (see Chapter 1, Section 1.2):

$$m\ddot{x} = -\frac{dV}{dx} - \gamma\dot{x}(t) + R(t) \tag{2.24}$$

and the equation which relates the random force with the friction γ thanks to the fluctuation dissipation theorem (FDT):

$$\langle R(t)R(t+\tau)\rangle = 2k_BT\gamma m\delta(\tau) \tag{2.25}$$

Applying the fourier transform on both sides of the previous equation we have:

$$C_{RR}(\omega) = 2k_B T\gamma m \tag{2.26}$$

The QTB exploits the FDT within a quantum framework, replacing the classical term $k_B T$ in Equation 2.26 with the average quantum energy at a given temperature for an harmonic oscillator:

$$C_{RR}(\omega) = 2m\gamma\theta(\omega, T) \tag{2.27}$$

where the term $\theta(\omega, T)$ is the average energy of a harmonic oscillator with a frequency ω at the temperature T, expressed as:

$$\theta(\omega, T) = \frac{\hbar\omega}{2} \coth\left(\frac{\hbar\omega}{2k_B T}\right) \tag{2.28}$$

This last expression is the one used by the quantum thermal bath [22] method in order to introduce the ZPE in the system through a thermostat. As it can be seen comparing Equations 1.8 and 2.27, the difference is in the random noise generated during the simulation. In the case of the classical Langevin equation, the noise is a white noise (Eq. 1.8) which goes to 0 for $T \to 0$. On the other hand, QTB is able to introduce approximate ZPE in the dynamics: the random noise generated is quantum (Eq. 2.27) and frequency dependent, so each mode is thermalized at an effective temperature that depends on its frequency to include the ZPE (of the mode). As a consequence of this approach, the system as a whole should have approximately the right ZPE.

Dammak and co-workers [22] show how QTB is able to catch the experimental trend of the lattice parameter and the heat capacity for MgO crystals. Another study was performed on diamond by Ceriotti and coworkers [77], where the Quantum Thermostat method (similar to QTB) was found in agreement with PI, but at low temperatures it displays a problem called zero point energy Leakage. This consists in the flow of the energy from high-frequency modes to low-frequency ones. This is particularly problematic for very anharmonic systems, as the classical forces that appear in the Langevin equation 1.6 efficiently drive the system towards the classical equipartition of the energy [78].

One way to mitigate this leakage would be using a large friction γ during the simulation, which represents the coupling term between the classical system and the quantum bath [78], but this affects thermodynamical and even more strongly dynamical properties. [79, 80]. Another way which has less impact on the dynamical and thermodynami-

cal properties and allows to complitely compensate for the ZPEL, is the adaptive-QTB [81]. This method uses the FDT in order to estimate the ZPEL, and adjust the γ parameter, which now becomes frequency dependent. In this way, there is a balance between the quantum bath and the classical degrees of freedom. In this way it allows to correct for the effects of the leakage on static properties, but its ability to capture dynamical rate constants had not been evaluated.

Part II

Applications

Chapter 3

Unimolecular fragmentation

The present chapter presents an investigation of how nuclear quantum effects modify the temperature-dependent rate constants and, consequently, the activation energies in unimolecular reactions. In the reactions under study, nuclear quantum effects mainly derive from the presence of large zero point energy. Thus, we analyze the behavior of methods compatible with reactive dynamics simulations, the quantum thermal bath (QTB) and ring polymer molecular dynamics (RPMD). To this end, as the first step, we present a comparison of these two methods with quantum reaction theory for a model Morse potential before extending this comparison to molecular models. Results show that, in particular in the temperature range comparable with or lower than the zero point energy of the system, the RPMD method is able to correctly capture nuclear quantum effects on rate constants and activation energies. On the other hand, although the QTB provides a good description of equilibrium properties including zero point energy effects, it largely overestimates the rate constants. The origin of these contrasted behaviors is in the distance distribution

provided by the two methods and, in particular, in the different description of the tails of such distributions. The comparison with the harmonic approximation to transition state theory (TST) shows that RPMD can be used to study the fragmentation of complex systems for which it may be difficult to determine the multiple reaction pathways and associated transition states.

3.1 Background

Unimolecular fragmentation is an elementary chemical process which is involved in many reaction mechanisms [82]. In fact, it is not only relevant per se in the direct fragmentation of chemical species but it can be present as part of a more complex mechanism. Direct fragmentation is, for example, the key process involved in tandem mass spectrometry, where an ion is activated (typically by collisions) and then fragments in two (or more) parts [83, 84]. Fragmentation in mass spectrometry has both a qualitative and quantitative application: the first in elucidating the nature of the precursor chemical species from its fragmentation signature [85–88], the second in determining binding energies [89–91]. Unimolecular fragmentation is one of the elementary steps in several reaction mechanisms. For example, the S_N2 reaction has two unimolecular fragmentation steps as part of the whole mechanism: (i) once the complex is formed and the intermediate species eventually rearranged, the products are obtained by the unimolecular fragmentation of the leaving group, (ii) the reverse process of the bi-molecular capture is a unimolecular fragmentation [92, 93].

For this reason, unimolecular fragmentation processes were largely studied both experimentally and theoretically [82]. The central theory used to describe such process in the canonical ensemble is the transition state theory, described in Section1.4.1:

$$k(T) = \frac{k_B T}{h} \frac{Q^\dagger(T)}{Q(T)} e^{-E_0/k_B T} \tag{3.1}$$

where E_0 is the potential energy at the transition state, while $Q^\dagger(T)$ and $Q(T)$ are the canonical partition functions of the TS and reactant, respectively.

The TST is a powerful theory which well describes the kinetics of unimolecular dissociation but one needs to clearly identify the reaction pathway and locate the TS. While this is straightforward for simple processes, it may be very difficult for complex and flexible systems in which different pathways are present. In this context, molecular simulations, and notably reactive dynamics, provide a useful tool to study the unimolecular fragmentation of such complex molecular systems. From such approach, it is possible not only to

discover new reaction mechanisms and/or products, but also to obtain the threshold and/or activation energies [94]. In particular, Hase and co-workers have studied relatively complex systems at different levels of theory, showing how reactive dynamics combined with unimolecular reaction theory is a powerful tool to get threshold or activation energies [95]. Note that this is particularly useful for large systems, like e.g. polypeptides, for which the localization of the corresponding TS is highly problematic [96, 97].

One critical aspect when dealing with the quantitative characterization of the reaction barrier from reactive dynamics simulations is the zero point energy (ZPE). In fact, when the ZPE is not the same for the reactant and the TS, the classical and the quantum threshold energies do not correspond. Chemical dynamics simulations are often based on quasi-classical initial conditions followed by Newtonian equations of motion numerical integration. In this way, one can observe typical non-physical effects, like ZPE leakage [17, 98] or, as it is the case for simple unimolecular fragmentation, products with incorrect energy distribution. For the unimolecular fragmentation case, this is due to the fact that simulations can generate products with vibrational energy which is less than the ZPE. This problem was recently discussed by Paul and Hase for the micro-canonical reactivity of a simple CH_4 model [99]. In this case, the unimolecular fragmentation consists of the H abstraction through a loose TS, such that one can consider only the difference between the ZPE of the reactant (CH_4) and of one product (CH_3). These authors have proposed a relatively simple method in which, if a trajectory forms a product with an energy lower than the ZPE, the trajectory is sent back to the reactant basin. This approach can be seen as the unimolecular fragmentation version of one of the first methods proposed in the past to avoid ZPE leakage in reactive dynamics simulations [17, 98]. While the rate constants obtained in this way qualitatively reproduce ZPE effects on the unimolecular rate constant, the resulting energy dependence did not show the expected Rice–Ramsperger–Kassel–Marcus (RRKM) behavior (i.e. the analogous of TST in microcanonical ensemble) [100].

Reactive dynamics simulations in the canonical ensemble represent another possibility to obtain rate constants and then, from an Arrhenius fit, the activation energy. The resulting $k(T)$ will accounts for anharmonicity and can account for NQEs if an appropriate dynamics method is used. A first approach was reported recently by Spezia and Dammak, where nuclear quantum effects were considered using the quantum thermal bath (QTB) [100]. This approach is appealing because it requires a computational effort that is very similar to classical simulations. The QTB has also recently proved capable of capturing accurately the effects of ZPE in liquid water, both on equilibrium properties and on the vibrational spectrum [101]. However, the quantum-classical activation energy differences reported in Ref. [100] were overestimated, as further analyzed below.

Another very popular class of methods to account for nuclear quantum effects, that can be used in conjunction with reactive dynamics simulations is that deriving from Path Integral theory [69, 102, 103]. In particular, ring polymer molecular dynamics (RPMD) [21] and Centroid Molecular Dynamics (CMD) [104, 105] are able to predict approximate reaction rate constants [76, 106]. While these methods have been used for isomerizations and bi-molecular simulations [76, 107–113], detailed study of how they behave in the case of unimolecular fragmentation kinetics at different temperatures has not been reported in the literature so far.

Another class of methods is based on a thermal Wigner sampling of initial conditions, but they will likely suffer from two problems in this context: (i) they usually rely on the harmonic approximation for the sampling of the Wigner density, whereas fully anharmonic approaches rapidly become computationally heavy for large systems [114–116]; (ii) since relatively long trajectories are needed for the fragmentation to occur, ZPE leakage from high to low frequencies might become problematic and bias the results. Other methods, like e.g. the Gaussian weighting in the quasi-classical trajectory method [18] which is very powerful to study molecular scattering [19, 117] or dissociation of bimolecular complexes [118, 119], are largely too expensive for medium to large molecular systems, require the knowledge of products and are generally used for well-defined state-to-state reactions. On the other hand, QTB and RPMD can be applied nowadays to relatively large molecules and sample the canonical ensemble.

In the present chapter, we studied the ability of reactive dynamics simulations with QTB and RPMD approaches to quantitatively describe unimolecular fragmentation at different temperatures and in particular to capture the difference between classical and quantum rate constants and activation energies. To this end, a simple one-dimensional (1-D) Morse model is considered, for which simulation results are compared with thermal rate constants from quantum theory [120] using a sum-of-states approach. Then we analyze the CH_4 model, which is modified in order to represent different barrier heights and to explore different temperature ranges. This relatively simple molecular model is chosen in view of the solid experimental data for CH_4 dissociation on a wide temperature range [121–125] and because an analytical model has been developed [126–128] which is known to be in agreement with experiments and kinetic theory [127, 128]. Simulation results can thus be compared with the available experimental and TST data. The simple analytical form of this model enables fast simulations and can be easily tuned to mimic molecular systems with lower dissociation energies, which can therefore be studied in reactive dynamics simulations at lower temperatures. This study also illustrates the conditions for nuclear quantum effects to have a non-negligible impact on the rate constant.

3.2 Rate Constants from Reactive Dynamics Simulations

From trajectory-based simulations the rate constants can be obtained considering the flux through the transition state hypersurface and the correlation function(s) discussed in Section 1.4.1 (using classical correlation functions in the case of classical dynamics). When the temperature (or the energy) is high enough, the reactivity can be directly sampled and rate constants obtained. This was the case of a number of unimolecular fragmentation simulations studied by Hase and co-workers, ranging from small to large systems [94, 95, 129–131].

In unimolecular fragmentation, the time evolution of the reactants ($N(t)$) has a simple behavior:

$$N(t)/N(0) = e^{-kt} \tag{3.2}$$

if transition state theory (TST) holds. In this case, the rate constant, k, is related to the lifetime (τ) of the reactants, $k = 1/\tau$, which can be obtained from reactive dynamic simulations. From the ensemble of trajectories and associated reaction times, it is thus possible to reconstruct the number of trajectories which are in the reactant state at each time frame. If the resulting curve follows a single exponential decay (as it was generally observed in previous studies [94, 96, 97, 132–134]) then τ can be directly obtained and thus k. Each trajectory is considered to be converted to products when a given threshold distance is passed. Trajectories are then immediately stopped, so that they do not allow for recrossing, as in the main assumption of transition state theory. For CH_4 previous studies have set this threshold distance at 6 Å based on variational TST and here we use the same value [99, 100]. Thus, once the rate constants for each unimolecular process are obtained at different temperatures if an Arrhenius behavior holds (and it can be easily verified from the outcomes of the different simulations), then the activation energy can be obtained by simply fitting:

$$k(T) = Ae^{-\frac{E_a}{k_B T}} \tag{3.3}$$

where A is the pre-exponential factor and E_a is the activation energy.

3.3 Simulations Set-Up

This section describes the details of the simulations performed, in terms of model potentials, determination of rate constants via Equation 1.39 and the associated computational details. The two methods to include nuclear quantum effects, RPMD and QTB, are described in detail in Chapter 2.

3.3.1 Model potentials

1-D Morse. The minimal model for unimolecular fragmentation is a one-dimensional Morse function:

$$V(x) = D_e \left[1 - e^{-B(x-x_0)}\right]^2 \tag{3.4}$$

which parameters are set with a relatively low barrier D_e=10 kcal/mol, in order to be able to observe fragmentation from reactive dynamics simulations in the temperature range of interest. The value of B was set to 3 Å^{-1} and x_0=1.09 Å . The equilibrium distance corresponds roughly to the C–H bond length and in the dynamics we use the corresponding masses. With these parameters the zero point energy is 2.2 kcal/mol and the vibrational frequency is 1511 cm^{-1}.

CH_4 Analytical Model. To study how nuclear quantum effects impact unimolecular fragmentation for a molecular system, we employ the analytical model proposed and extensively studied by Hase and co-workers some years ago for the reaction [126–128, 135]:

$$\mathrm{CH_4} \rightarrow \mathrm{CH_3}\cdot + \mathrm{H}\cdot \tag{3.5}$$

The potential energy function is represented by different terms:

$$V = V_{Morse} + V_{ang} + V_{oop} + V_{nd} \tag{3.6}$$

Where the first term is a Morse potential, the second is an angular potential, the third is the out-of-plane potential and the last one is a non-diagonal cubic term. The combination of these terms ensures to have the tetrahedral structure of the molecule, for more detail see Ref. [100]. To model different fragmentation regimes, we considered the Morse term:

$$V_{Morse} = \sum_{i=1}^{4} D_i \left[1 - e^{-B_i(r_i - r_i^0)}\right]^2 \tag{3.7}$$

where the sum runs over four C–H bonds. This term is modified in order to build three different potentials:

- Potential A, which corresponds to the original model for CH_4 fragmentation;
- Potential B, where the barrier was lowered down by about 50 %;
- Potential C, where the barrier was lowered further down to 30 kcal/mol, roughly corresponding to typical values of protonated systems;

The atomic masses are those of C and H (i.e. 12 and 1.008 amu, respectively) for the three

potentials. Notably, the two parameters, D_i and B_i, are modified for all four bonds, keeping the equilibrium geometry fixed (r_i^0 = 1.09 Å). As a consequence, the ZPEs of reactants and products are changed and thus also the "quantum" zero Kelvin barrier, D^0. The different sets of parameters are reported in Table 3.1, with the corresponding ZPEs and D^Q, while Figure 3.1 shows the corresponding Morse functions.

Table 3.1: Sets of Parameters modified with respect to the original CH_4 model [126] (corresponding to potential A) as used in the present work.

Potential	$D_i(\equiv D_0^{(Cl)})$ [kcal/mol]	B_i [Å $^{-1}$]	ZPE_R [kcal/mol]	ZPE_P [kcal/mol]	D^Q [kcal/mol]
A	109.460	1.944	29.18	19.38	99.66
B	50.000	2.500	26.76	16.74	39.98
C	30.000	3.000	25.63	16.54	20.91

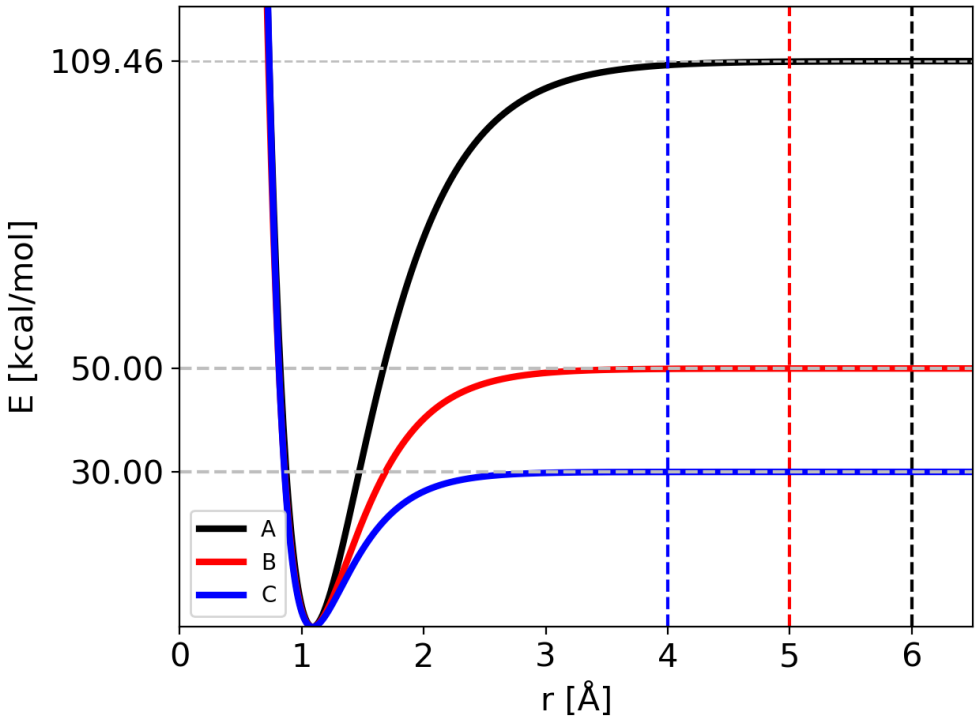

Figure 3.1: Morse functions used in the modified CH_4 molecular system: potential A is in black, potential B in red and potential C in blue. Horizontal dotted lines show the corresponding D_0 values, while vertical dashed lines show the position of the threshold distances used in trajectory simulations.

3.3.2 Sum-of-States Rate Constants

The expressions 1.37 to 1.39 for the rate constant can be developed via a discrete basis set approach. We denote E_i the eigenvalues, the eigenfunctions are denoted as $\psi_i(s)$ which

corresponds to the eigenstates $|i\rangle$ of the Hamiltonian: for a one-dimensional system, they are easily obtained by numerical diagonalization over an arbitrary set of basis functions. In this eigenstates representation, the flux correlation function in Eq. 1.37 can be rewritten as:

$$\begin{aligned} c_{f,f}(t) &= \sum_i e^{\frac{-\beta E_i}{2}} e^{\frac{-E_i t}{\hbar}} \langle i|\hat{F} e^{i\frac{Ht_c^*}{\hbar}} \hat{F}|i\rangle \\ &= \sum_{i,j} e^{\frac{-\beta(E_i+E_j)}{2}} \cos\left(\frac{(E_i-E_j)t}{\hbar}\right) |\langle i|\hat{F}|j\rangle|^2 \end{aligned}$$

where we introduced the resolution of identity: $1 = \sum_j |j\rangle\langle j|$. The rate constant obtained from the integral of $c_{f,f}$ thus becomes:

$$k = \frac{1}{Q} \sum_{i,j} e^{-\frac{\beta(E_i+E_j)}{2}} \frac{\sin\left(\frac{(E_i-E_j)t}{\hbar}\right)}{E_i - E_j} |\langle i|\hat{F}|j\rangle|^2 \tag{3.8}$$

Finally, the flux matrix element is related to the wave functions by two different, equivalent expressions:

$$\begin{aligned} |\langle i|\hat{F}|j\rangle|^2 &= \left(\frac{\hbar}{2m}\right)^2 |\psi_i'(q^\ddagger)\psi_j(q^\ddagger) - \psi_i(q^\ddagger)\psi_j'(q^\ddagger)|^2 \\ &= \frac{1}{\hbar^2} |\langle i|[H, h(q-q^\ddagger)]|j\rangle|^2 \\ &= \frac{(E_i-E_j)^2}{\hbar^2} |\langle i|h(q-q^\ddagger)|j\rangle|^2 \end{aligned}$$

where ψ_i' denotes the position derivative of the function. Note that a classical counterpart to this rate expression can easily be obtained by noting that, for the one-dimensional Morse function, the long-time limit of the flux-side correlation function is given by:

$$\begin{aligned} \lim_{t\to\infty} c_{f,s}^{cl}(t) &= \frac{1}{2\pi\hbar} \int_{\sqrt{2m(D_e - V(q^\ddagger))}}^{\infty} dp \int dq\, e^{-\beta H(q,p)} \delta(q-q^\ddagger) \frac{p}{m} \\ &= \frac{e^{-\beta D_e}}{2\pi\hbar\beta} \end{aligned} \tag{3.9}$$

3.3.3 Computational details

1-D Morse. In the case of the 1-D Morse model RPMD, QTB and classical Langevin molecular dynamics (LMD) simulations are performed as follows. We run about 2000 trajectories for each temperature, a time-step of 0.1 fs, with a friction constant (γ) on the centroid set up at different values: 0.01, 0.045 and 0.3 fs^{-1}. Note that the highest γ value corresponds

to an overdamped regime with $\gamma \simeq \omega$, with ω the typical angular frequency of the Morse potential, while the smallest friction coefficient corresponds to an underdamped regime. The maximum simulation time for each trajectory is of 5.0 ns if it does not react before: this time-length ensures a reactivity of almost 100 % in all simulations. The number of beads was chosen in order to obtain the correct quantum energy value at equilibrium, corresponding to 8 beads for higher temperatures and 32 for the lowest ones. The range of temperature is from 800 K up to 1200 K. Details on temperatures and numbers of beads employed are listed in the Appendix A, Table A.1.

CH_4 and modified potentials. Molecular dynamics simulations were performed with the three potentials described before using RPMD, QTB and LMD approaches. All simulations were done using our in-house modified version of the VENUS software [23, 136]. The trajectories were carried out for different ranges of temperature, listed in Appendix A Table A.1, to allow the fragmentation of the reactant. If the distance of one bond reaches the cut-off value, the trajectory is stopped and the lifetime is collected. The cut-off value used in ref. [99] for the potential A (6 Å) corresponds to a plateau in the Morse potential energy function: the cut-off values for potentials B and C (5 and 4 Å, respectively) were reduced since the dissociation energy is reached at shorter distances (see Figure 3.1). The maximum simulation time was set to 5 ns, with a time-step of 0.1 fs. As before, the number of beads used in RPMD simulations was chosen in order to converge the average energy: in Table A.1 are listed the number of beads used for each potential as a function of temperature.

The friction parameter γ was chosen in order to yield a fast enough temperature equilibration, as illustrated by the temperature autocorrelation function reported in Figure 3.2. Notably $\gamma = 0.01$ fs^{-1} provides an efficient thermostat for this system, whereas smaller values would result in a too-slow equilibration (here the equilibration should be faster than the typical reaction time) and too-large values can affect the dynamical results (in particular in the overdamped regime).

To obtain rate constants and the associated uncertainties from trajectory simulations, we implemented the bootstrap algorithm [137]. This statistical method randomly re-samples a single data-set, to create multiple data-sets and obtain the mean and the standard deviation from a gaussian distribution. In the present case the data set consists in the set of reaction times for each individual trajectory and for each re-sampling, the associated rate constant is computed from single exponential fitting.

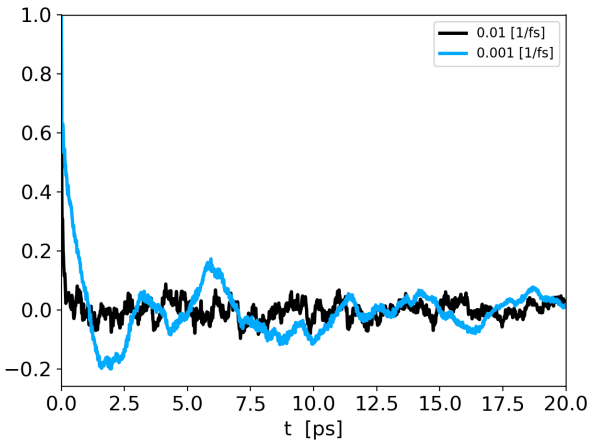

Figure 3.2: Temperature autocorrelation function obtained from LMD simulations at 3000 K with potential C for two different values of γ: 0.01 fs $^{-1}$ in black and 0.001 fs $^{-1}$ in blue.

3.4 Results

3.4.1 1-D model potential

For the 1-D Morse potential the quantum rate constants are calculated using the Sum-Of-States (SoS) approach, Equation 3.8, detailed previously. Classical rate constants can also be directly obtained. These results are used as a reference calculation to evaluate the performances of LMD, RPMD and QTB simulations.

In the SoS approach the rate constant is obtained from the $t \to \infty$ limit of the function $c_{f,s}(t)/Q$. The time dependence of this function is shown in Figure 3.3 for the 1-D Morse model at two temperatures, 800 and 1500 K. As discussed in details by Miller in the past [54, 120], for such low-dimensional system, the function actually goes to zero in the long-time limit, but the value of the rate k can still be obtained by considering the maximum (plateau) of the function in an intermediary time range. In the same plots we also report the corresponding classical rate constant as an horizontal black line.

Rate constants obtained in an extended temperature range (700-1200 K) are shown in Figure 3.4 while all the values are listed in Tables A.2 of Appendix A. It is important to notice that the rate constants show an Arrhenius-like behavior, and thus it is possible to fit them and obtain activation energies in both classical and quantum regimes. Values are reported in Table 3.2 together with the difference between classical and quantum activation energies. These results can now be used as a reference to evaluate values obtained from trajectory simulations.

Trajectory simulations (LMD, RPMD and QTB) were performed on the same temper-

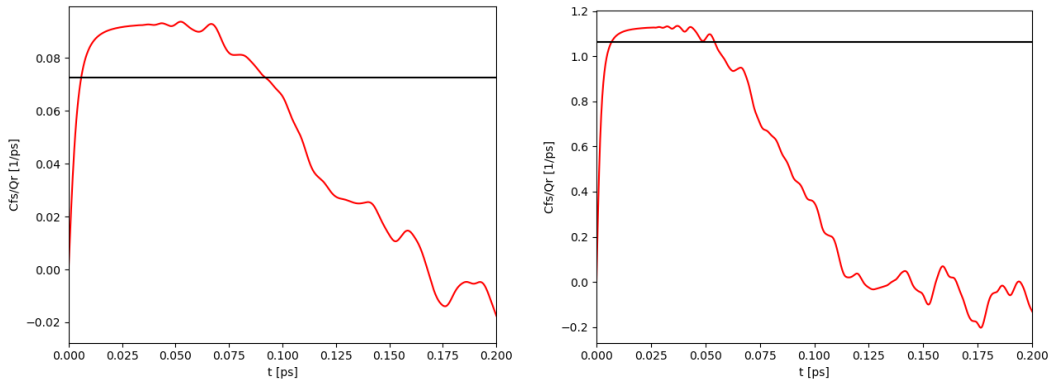

Figure 3.3: Quantum (red) and classical (black) $c_{f,s}$ obtained with the SoS method for the 1-D Morse model at two different temperatures: 800 K (left) and 1500 K (right)

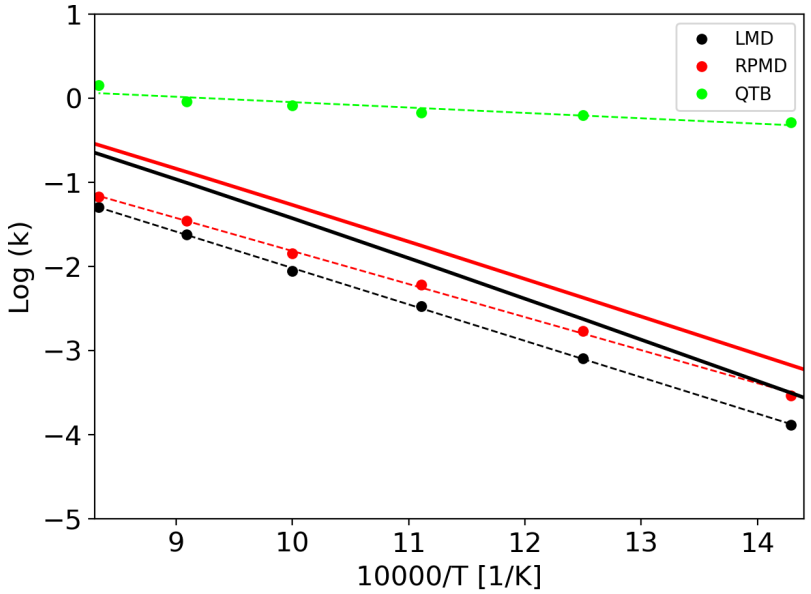

Figure 3.4: Quantum and classical rate constants as a function of temperature for the 1-D Morse. In full lines we report values obtained from sum-of-state approach (both Classical and quantum), while results from trajectory simulations (γ = 0.01 fs $^{-1}$) are reported as dots: LMD (black), RPMD (red), QTB (green).

Table 3.2: Activation energy (E_a) and pre-exponential factor (A) obtained from temperature dependent rate constants (in the 700-1200 K range) for the 1-D Morse model. The difference between classical and quantum E_a obtained with the same simulation conditions corresponds to ΔE^{Cl-Q}. E_a and ΔE^{Cl-Q} are in kcal/mol and A and γ in fs^{-1}. We also report the products-reactant energy difference obtained in the harmonic approximation, both classical and quantum.

Method	E_a	ln(A)	ΔE^{Cl-Q}
Classical reference	9.49	3.34	–
Quantum (SoS)	8.74	3.13	0.75
LMD (γ = 0.01)	8.60 $\pm$ 0.09	2.31 $\pm$ 0.05	–
RPMD (γ = 0.01)	7.8$\pm$0.1	2.12$\pm$0.08	0.8 $\pm$ 0.1
QTB (γ = 0.01)	1.3 $\pm$0.2	0.6 $\pm$ 0.1	7.3$\pm$0.2
LMD (γ = 0.045)	9.70 $\pm$ 0.03	3.09 $\pm$ 0.01	–
RPMD (γ = 0.045)	8.85$\pm$0.02	2.83$\pm$0.01	0.85 $\pm$ 0.04
QTB (γ = 0.045)	1.74$\pm$0.08	1.08$\pm$0.05	7.96 $\pm$0.08
LMD (γ = 0.3)	9.9 $\pm$ 0.2	2.1 $\pm$ 0.1	–
RPMD (γ = 0.3)	9.21$\pm$0.1	1.91$\pm$0.07	0.7 $\pm$ 0.2
QTB (γ = 0.3)	5.8$\pm$0.4	1.1$\pm$0.2	4.1$\pm$0.4
Harmonic classic	8.11	-	-
Harmonic quantum	7.33	-	0.78

ature range and for all of them we observed a single exponential decay of the reactant populations, such that rate constants could be extracted directly by fitting the equation 3.2 with the bootstrap method to assign the associated uncertainties. An example is shown in Figure 3.5, corresponding to simulations at 800 K and γ = 0.01 fs^{-1} done with LMD, RPMD and QTB methods. Similar single exponential decays are obtained at other temperatures and γ values. As we can already see, while LMD and RPMD provide quite similar decays, the QTB reaction rate is much faster. This aspect will be discussed in the following.

Rate constants obtained from trajectories as a function of temperature are shown in Figure 3.4 for γ =0.01 fs $^{-1}$ (and all the values are reported in Appendix A, Tables A.2). We can notice that LMD and RPMD simulations show a clear Arrhenius-like behavior, with a slope which is similar to that obtained via the SoS method, and rate constant values of the same order of magnitude as the reference. On the other hand, in QTB simulations, the rate constants are much higher and they show only weak variations with temperature. The corresponding activation energies and pre-exponential factors are reported in Table 3.2 where they can be compared with SoS values. Remarkably, activation energies obtained from LMD and RPMD simulations are very similar to those obtained from the SoS and classical reference approaches. In the simulations discussed previously, we use a friction parameter γ of 0.01 fs $^{-1}$. Simulations with two additional γ values (0.045 and 0.3 fs $^{-1}$) are also performed in order to investigate if there is an impact on kinetics, activation energy

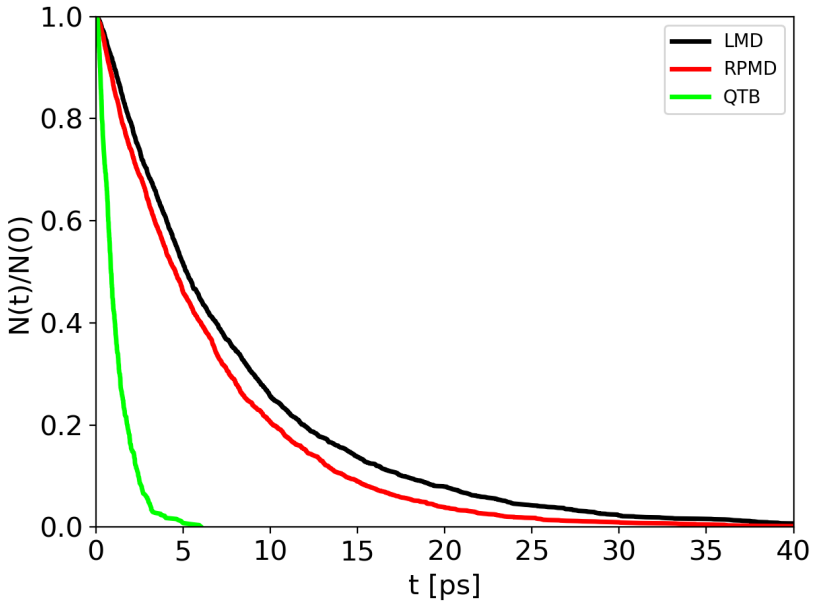

Figure 3.5: Population decay for 1-D Morse simulations at 800 K as obtained from LMD, RPMD and QTB trajectories.

and, more importantly, quantum-classical difference. Note that γ = 0.3 fs $^{-1}$ corresponds to an overdamped regime ($\gamma \sim \omega$ with ω the typical angular frequency of the Morse potential). Lifetimes and rate constants vary with γ, but the variation on the value of activation energy is very small as shown by the Arrhenius plots reported in Figure A.1 of Appendix A. From the rate constants as a function of temperature with different γ values, we have done the Arrhenius fits to obtain activation energies which display only limited dependence (apart from the QTB) on γ, as shown in Table 3.2. If now we move to the difference between the classical and quantum activation energies, it can be noticed that RPMD simulations are in very good agreement with SoS values, while QTB largely overestimating it. Furthermore, the effect of γ on this quantity almost vanishes, in particular for RPMD simulations.

Before moving to the study of more complex molecular systems, we considered the simple harmonic approximation (for both classical and quantum statistics) to estimate the barrier. Notably, the barrier can be approximated, for a simple dissociation, as the average energy difference between reactants and products. In the case of a simple 1-D Morse model:

$$E^i(T) = E^i_{POT} + E^i_{TR}(T) + E^i_{VIB}(T) \tag{3.10}$$

where i stands for reactant and product and POT, TR and VIB for potential translational and vibrational energy, respectively, where E^i_{POT} is the potential energy and thus not temperature dependent. The energy difference can be simply estimated as:

$$\Delta E(T) = \Delta E_{POT} + \Delta E_{TR}(T) - E_{VIB}^{react}(T) \tag{3.11}$$

since in the simple 1-D Morse model the products have no vibrational energy. The energy difference between the classical and quantum approaches is thus simply reduced to:

$$\Delta E^{Cl-Q}(T) = E_{VIB}^{Cl}(T) - E_{VIB}^{Q}(T) \tag{3.12}$$

where $E_{VIB}^{Cl}(T)$ and $E_{VIB}^{Q}(T)$ are the temperature-dependent classical and quantum vibrational energies of the reactant (here the products have no vibrational structure). Results are reported in the same Table 3.2. Although the average energy difference tends to slightly underestimate the activation energy in absolute value, the quantum-classical difference obtained with this method is very similar to that derived from the SoS approach (and to what is obtained comparing LMD with RPMD simulations). This will be very important for the study of a more complex molecular system for which the SoS approach is not feasible, and the simple harmonic approximation can provide a good reference to compare simulations.

3.4.2 CH_4-based potentials

In this Subsection we report the results of the rate constants obtained for the unimolecular fragmentation of the CH_4 model (potential A) as a function of temperature and how this behavior is modified by decreasing the barrier (potentials B and C).

From trajectory simulations we obtain the population decay as a function of time which shows, also in this case, a single exponential behavior. The single-exponential behavior was found for each temperature, potential and method (see a prototypical example reported in Figure 3.6). Thus, it was possible to extract lifetimes and unimolecular rate constants.

LMD, RPMD and QTB rate constants as a function of temperature are shown in Figure 3.7 for the three potentials (A, B and C), while the full set of values of rate constants are reported in Tables A.3, A.4 and A.5 of Appendix A. Experimental data are available only for potential A, corresponding to methane fragmentation. The best sets of values to be compared with unimolecular rate constants from simulations (where the colliding gas is not explicitly simulated such that they have inverse time dimensions) are those reported by Cobos and Troe [121, 122] which also provide an analytical function to express $k(T)$, originally in the 300–3000 K temperature range later extended up to 5000 K. Simulation results agree well with such experimental data as shown in Figure 3.8.

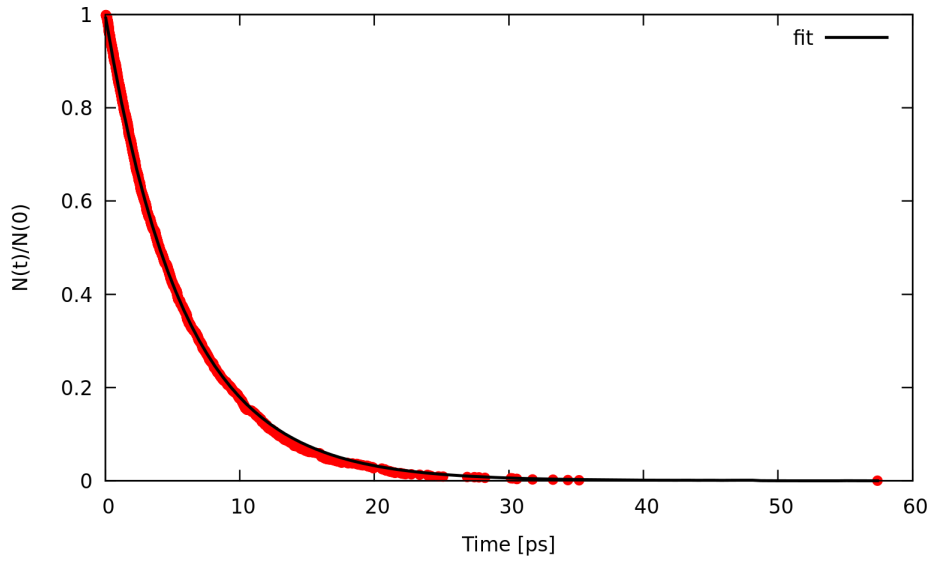

Figure 3.6: Population decay obtained from RPMD simulations with γ =0.01 fs^{-1}, P = 8 and T = 4500 K using potential C. Red dots are the simulation data, black solid line corresponds to a single-exponential fit.

The ability of the RPMD and QTB to correctly capture nuclear quantum effects can be first assessed by comparing the ratio between the quantum and classical rate constants (k^q and k^{cl}, respectively) with that obtained from transition state theory (TST), which typically holds for CH_4 fragmentation, [99, 138, 139]. Results are summarized in Table 3.3 where in order to evaluate the partition function ratio for TST we employed the rigid-rotor harmonic approximation both classical and quantum. [82] The nuclear quantum effects are negligible for potential A, while they increase for potential B and C, as expected since it was possible to simulate lower temperatures. Notably, the k^q/k^{cl} values obtained from TST are very similar to the RPMD/LMD ratios, while the QTB largely overestimates the rate constants, in agreement with what was found for the simple 1-D Morse model. The slight differences between the TST and RPMD/LMD results can be ascribed to anharmonicity effects that are present in the simulations while not accounted for by the TST harmonic approximation for the partition function. This effect is more marked for potential B and C where the barrier is lower and therefore anharmonicity is likely more important. Note that the CH_4 experimental data (Figure 3.8) are in better agreement with RPMD (and quantum TST) values than with the QTB ones.

As clearly shown in Figure 3.7, the Arrhenius-like behaviors are obtained in all simulations such that it was possible to fit them and derive the activation energies. They are reported in Table 3.4 together with the corresponding quantum-classical differences. It should be noticed that for all the potentials, the RPMD activation energies are much closer to the LMD ones than what is obtained from QTB simulations: for the potential A, which

Table 3.3: Ratio between quantum and classical unimolecular dissociation rate constants as obtained from the three molecular model potentials using transition state theory (TST) and simulations (RPMD and QTB for quantum rate constants and LMD for classical ones).

	T[K]	$k^{TST,q}/k^{TST,cl}$	k^{RPMD}/k^{LMD}	k^{QTB}/k^{LMD}
Pot. A				
	3000	1.17	1.09	3.04
	3500	1.12	1.12	2.01
	4000	1.09	1.06	1.44
	4500	1.07	1.02	1.33
	5000	1.06	1.04	1.17
Pot. B				
	1350	1.86	1.71	54.3
	1700	1.49	1.48	5.68
	2000	1.34	1.16	2.85
	2500	1.21	1.19	1.88
Pot. C				
	800	4.49	3.53	896
	1000	2.72	2.47	60.90
	1200	2.04	1.69	11.00
	1500	1.59	1.29	3.64
	1800	1.39	1.36	2.03

has a high barrier and for which, therefore, the simulations were performed at relatively high temperatures, this difference is negligible, comparable to the uncertainty, while for the QTB it is about 15 kcal/mol which is similar to the quantum-classical difference of the model at 0 K, namely the difference in ZPE between reactant and products, that is around 10 kcal/mol. Moving to potentials B and C, we observe a statistically significant quantum-classical difference from RPMD (3.1 and 3.3 kcal/mol, respectively) and, as before, larger values from the QTB. Note that the 0 K quantum-classical difference does not change much moving to potentials B and C, while QTB provides even larger values of activation energy quantum-classical differences.

As previously done for the 1-D Morse model, it is possible to use the simple harmonic approximation approach to evaluate how the reaction barrier depends on the temperature and compare with simulation results. This simple approach was shown to provide reasonable results compared with the SoS method for the 1-D Morse model and thus it can also be used to evaluate the accuracy of the RPMD and QTB results for the molecular model case.

In particular, we can estimate the temperature-dependent average energy difference

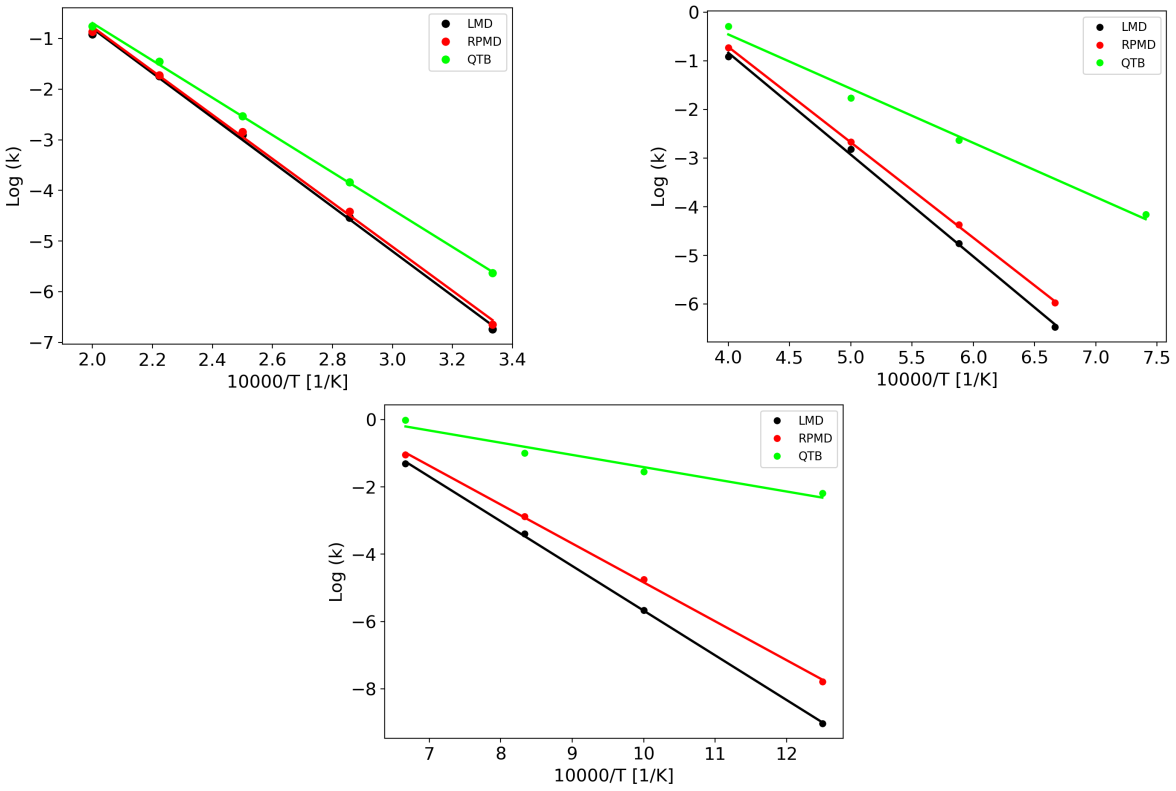

Figure 3.7: Arrhenius plots obtained from LMD, RPMD and QTB simulations for A (top-left), B(top-right) and C (bottom) potentials. Circles are data obtain from simulations while lines represent the fit results.

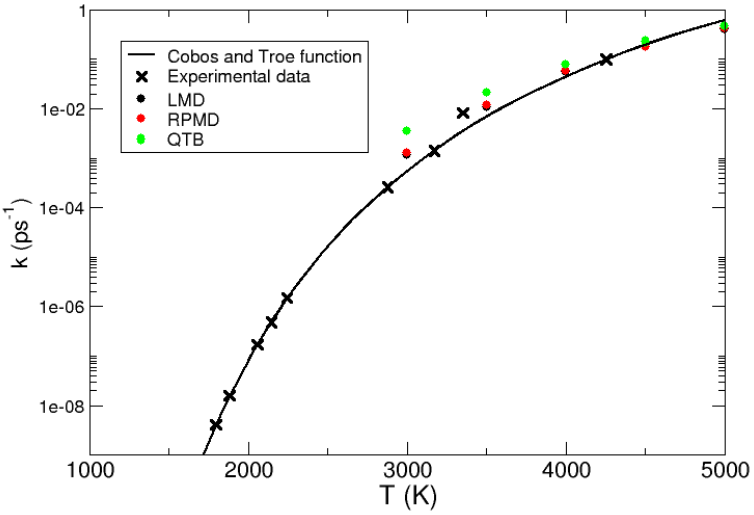

Figure 3.8: Rate constants for CH_4 fragmentation as obtained from RPMD, QTB and LMD simulations compared with experimental data and fitted functions as reported by Cobos and Troe. [121, 122]

Table 3.4: Activation energies (E_a) and quantum-classical difference (ΔE), both in kcal/mol, obtained for the three CH_4-model potentials from LMD, RPMD and QTB simulations. We report also average energy differences obtained from the simple harmonic approximation both classical ($\Delta E(\bar{T})^{Cl}$) and quantum ($\Delta E(\bar{T})^{Q}$).

	Pot. A		Pot. B		Pot. C	
Method	E_a	ΔE	E_a	ΔE	E_a	ΔE
LMD	88 ± 2	–	43.9 ± 0.2	–	26.2 ± 0.4	–
RPMD	87 ± 2	1 ± 3	40.8 ± 0.9	3.1 ± 0.9	22.9 ± 0.4	3.3 ± 0.6
QTB	73.3 ± 0.9	15 ± 2	22 ± 1	22 ± 1	7.2 ± 1.0	19 ± 1
$\Delta E(\bar{T})^{Cl}$	97.52	–	45.11	–	26.24	–
$\Delta E(\bar{T})^{Q}$	95.99	1.52	41.82	3.29	22.87	3.37

(both classical,quantum and their difference) from:

$$\Delta E^{Cl}(T) = D_0 + \frac{3}{2}k_bT + \Delta E^{Cl}_{VIB}(T) + \Delta E^{Cl}_{ROT}(T) \quad (3.13)$$

$$\Delta E^{Q}(T) = D_0 + \frac{3}{2}k_bT + \Delta E^{Q}_{VIB}(T) + \Delta E^{Q}_{ROT}(T) \quad (3.14)$$

$$\Delta\Delta E^{Cl-Q}(T) = \Delta\Delta E^{Cl-Q}_{VIB}(T) + \Delta\Delta E^{Cl-Q}_{ROT}(T) \quad (3.15)$$

where D_0 is the potential energy surface barrier (coinciding with D_i values reported in Table 3.1), ΔE^{cl}_{VIB} and ΔE^{Q}_{VIB} are the classical and quantum vibrational energy differences between products and reactant, ΔE^{Cl}_{ROT} and ΔE^{Q}_{ROT} are rotational energy differences and $\Delta\Delta E^{Cl-Q}_{VIB}(T)$ with $\Delta\Delta E^{Cl-Q}_{ROT}(T)$ are the vibrational and rotational quantum-classical differences. The $3/2k_BT$ terms come from the difference in the translational energy. Here again, both classical and quantum vibrational energies are obtained in the harmonic approximation. The rotational energy was considered classical and thus the same for reactant and product states. In Table 3.4 we report the temperature-averaged barrier values, that are slightly higher than the activation energy obtained from the Arrhenius plots, in particular for the potential A for which the simulations were performed at relatively high temperatures. However, the quantum-classical energy differences are in excellent agreement with the LMD-RPMD differences for the three potentials. In particular, while for potential A (in the temperature range considered) there is almost no difference between classical and quantum vibrations, quantum effects are larger for potentials B and C: even if the difference remains small, it is significant compared to the statistical uncertainty in the simulation reults.

In Figure 3.9 we show the products-reactant energy difference as a function of temperature as obtained for potentials A and C using both classical and quantum energies. The plots also illustrate the limit of the "low" temperature region for each potential. It corresponds to temperatures below the ZPE (shown as vertical dotted lines), which is the point from

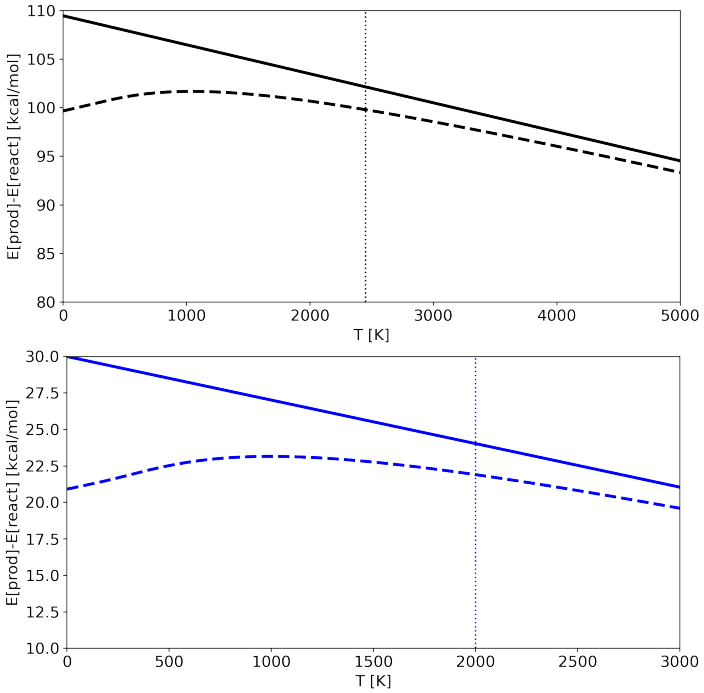

Figure 3.9: Energy differences between products and reactants as a function of temperature in the harmonic approximation as obtained from potentials A (top) and B (bottom). Full lines are classical values, while dashed lines are quantum ones. Vertical lines show the ZPE in K.

which the classical and quantum energy curves begin to diverge significantly. It should be noted that for potential A simulations were done at temperatures higher than the ZPE, in an almost classical regime (lower temperatures were not accessible due to computational limitations in trajectories time-length), while for potential C (and B) it was possible to run trajectories at temperatures that are lower than the ZPE and so in a region where nuclear quantum effects are detectable. As a consequence, for potential A the quantum-classical energy differences are almost irrelevant, while for potential C the difference becomes noticeable. RPMD simulations are able to catch this effect accurately.

It is well known that the absolute value of the activation energy is often underestimated from Arrhenius fits [140, 141], but the quantum-classical difference, which mainly reflects the ZPE difference between reactants and products absent in classical dynamics, are well reproduced by the comparison between RPMD and LMD simulations.

Clearly, QTB rate constants and activation energies are much higher than that obtained from the other methods. Even if the harmonic approximation is rather crude, the present results together with that obtained from the 1-D Morse model, suggest that QTB overestimates unimolecular fragmentation rates in this case, while RPMD gives results in very good agreement with the quantum theory.

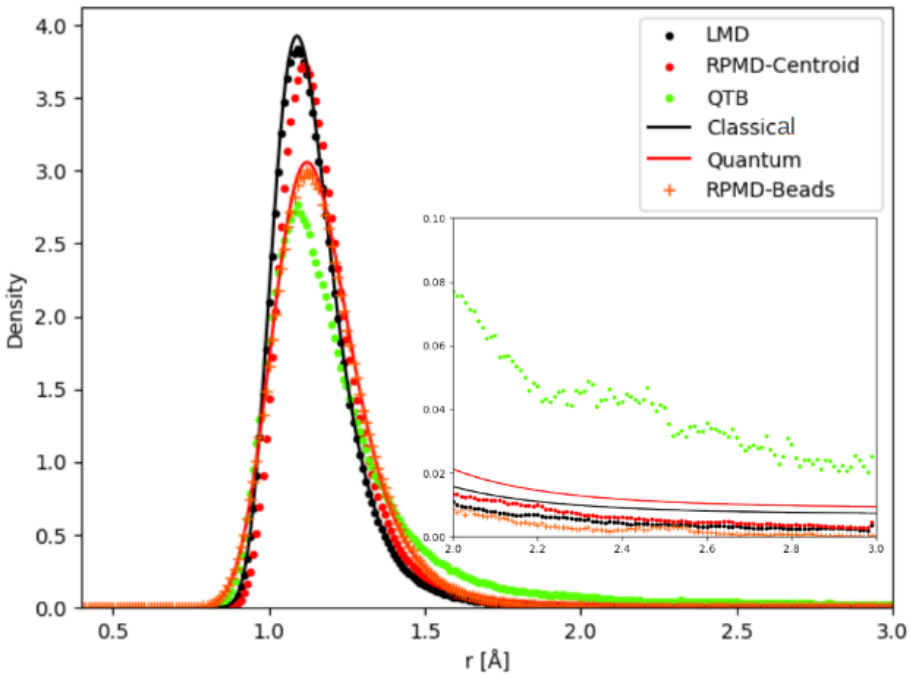

Figure 3.10: Distance distribution (as probability density) for 1-D Morse model at 1000 K. The inset shows a zoom of the distribution tail.

3.4.3 Distance distributions

Unimolecular fragmentation is obviously related to how the bond distance evolves from equilibrium to the threshold distance. Chemical reactivity in the canonical ensemble is related to energy fluctuations [142], and thus to distance fluctuations. As a consequence, even small differences in the distance distribution can cause large differences in reaction dynamics. In particular, the tail of the distribution plays a crucial role. Thus, we analyze the distance distributions obtained in different simulations (and theory for the simple 1-D Morse) at some relevant temperatures. Firstly we consider the 1-D Morse model at 1000 K: in Figure 3.10 we show the classical and quantum analytical distributions and how simulations compare with them. In the case of RPMD simulations two quantities are plotted: the distribution of the atomic distance for each bead (marked in the figure as "RPMD-Beads") as well as the distribution of the centroid of the distance, i.e. its average over all the beads (marked in the figure as "RPMD-Centroid"). As one should expect by construction, the LMD distance distribution essentially coincides with the classical theoretical distribution, while the RPMD-Beads curve is almost superimposed with the quantum reference. The quantum distribution is significantly broader than the classical one, due to the non-negligible ZPE effects, even at this relatively high temperature. The maximum of the quantum peak is also slightly displaced with respect to its classical counterpart. The QTB distribution is broadened in a similar way as the quantum reference, showing that the coloured noise thermostat

provides an approximation to ZPE effects. However, the peak maximum is slightly displaced with respect to the quantum reference (it is more in line with the maximum of the classical distribution), and more importantly, the long-distance tail of the distribution is markedly longer in the QTB simulations. These slight inaccuracies of the QTB approximation have been discussed in the literature in the non-reactive case (see for example Ref. [143]), where they usually have only minor consequences. However, the shape of the distribution tail can have a much larger impact on the unimolecular fragmentation rate and partly explain its overestimation by the QTB. The RPMD-Centroid curve is also instructive: the distribution of the centroid distance is very similar to the classical distribution, with only a slight shift of its maximum towards longer distances as for the quantum distribution. The distributions obtained for the C-H distance in the CH_4 model potential show very similar behavior. Figure 3.11 shows an example of the results as obtained for potential C at 1000 K (in this case the simulations were run with a larger distance cut-off of 5 Å , to enhance the statistics on the tail of the distribution). As in the 1-D Morse model, the centroid distribution of the distance is much sharper than that of the beads and closer to the classical results. The narrow centroid distribution explains why the RPMD reaction rate is significantly lower at this temperature than the QTB one. Indeed, in RPMD simulations, even if individual beads can undergo a momentaneously increase of the C-H bond length, the harmonic spring forces of the ring-polymer then tend to attract it back towards the equilibrium distance (where the centroid tends to remain localized), whereas no such mechanism exists in the case of the QTB.

3.5 Conclusions

In this chapter, we investigated how a unimolecular reaction is affected when passing from a classical to a quantum description of the nuclear motion. In particular, we studied how the reaction rates and activation energies are affected and how reactive dynamics simulations based on an ensemble of trajectories are able to reproduce this behavior. For the 1-D Morse model it was possible to compare the simulation results with quantum rate constant theory, and then we studied a more complex molecular model tuning the barrier height. Results show that ring polymer molecular dynamics (RPMD) method provides a satisfactory approximation to the unimolecular kinetics, and how it is impacted by NQEs. This confirms previous studies by Manolopoulos and co-workers on different reactions such as isomerizations and bi-molecular reactions [76, 107, 108]. In particular in the low-temperature regime, although the beads probability distribution (which represents the physical quantum distribution) is strongly broadened by the zero point motion, the centroid distribution remains localized in a similar way as for a classical system. Since the fragmentation pro-

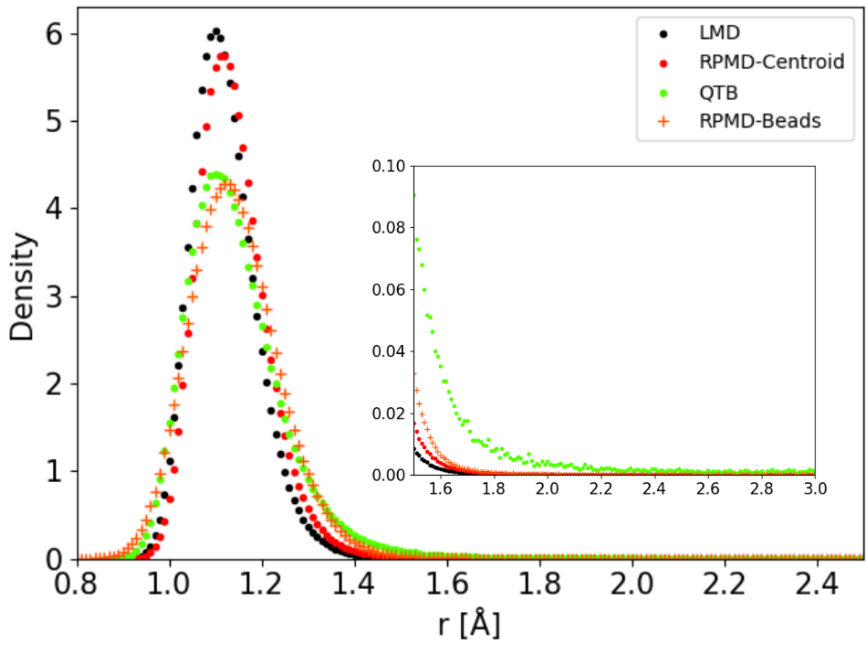

Figure 3.11: C–H distance distribution as obtained from trajectory simulations using Potential C at 1000 K

cess requires the whole polymer to go through the barrier, the localization of the centroid effectively reduces the reaction rate at low temperatures, in agreement with the quantum theory. The quantum thermal bath (QTB) on the other hand, largely overestimates the reaction rates, in particular for low barriers, or when the temperature becomes lower than the zero point energy (ZPE). The QTB approximates ZPE as an elevated effective temperature via a coloured-noise thermostat. Although this approach proves accurate to describe ZPE effects on the equilibrium distribution, it cannot be used to compute fragmentation rates in a reactive dynamics set-up.

The temperature dependence of the quantum/classical rate constant ratio shows that nuclear quantum effects can be relevant even at relatively high temperatures. In fact, we observed that, even at 800 and 1000 K, the quantum rate constants are about 4 and 2.5 times faster than the corresponding classical values, respectively. The temperature range for which the NQEs are relevant is related to the ZPE of the corresponding breaking bond. This study suggests that RPMD can be used as a trajectory-based method to investigate unimolecular reactions of complex systems. This approach can be a powerful alternative to the transition state theory (TST) to study the fragmentation of large molecules, having many fragmentation pathways and for which the determination of all the corresponding transition states may be difficult [95], while clearly, it will not supersede TST for systems in which it can be applied. For example, the fragmentation of complex peptides was recently

studied with Newtonian reaction dynamics not only to study qualitatively the fragmentation pathways, but also to evaluate quantitatively the activation energies [94, 96, 97]. Future studies could be now done using RPMD to investigate the role of NQEs on both aspects. Such reactive dynamics approach is able to provide anharmonic rate constants in which nuclear quantum effects are taken into account. In terms of TST this corresponds to (i) taking into account the ZPE effect on the barrier height and (ii) considering the quantum partition functions without any specific approximation (like the typical harmonic approximation): the effect is directly obtained on the resulting rate constant without the need of any specific correction. While all these aspects can be considered using TST (equation 3.1), one has to determine the reaction pathways which can be not obvious for large and flexible molecules: RPMD provides it directly from an ensemble of trajectories in a relatively simple way. Clearly, one critical aspect to apply RPMD to large systems is the computational cost, which is roughly multiplied by the number of beads compared to classical dynamics. However, the recent and future advances of fast and reliable reactive potentials based, for example, on machine learning techniques, will enable the study of fragmentation dynamics on much larger systems and the inclusion of nuclear quantum effects will also be computationally possible. The very good agreement between RPMD and TST on fragmentation reactions gives confidence for the use of RPMD when using TST may have practical problems, as for large and flexible molecules.

Chapter 4

Environmental and nuclear quantum effects on double proton transfer in Guanine-Cytosine base pair

In this chapter, we investigate the double proton transfer (DPT) tautomerization process in Guanine-Cytosine (GC) DNA base pairs. In particular, we study the influence of the biological environment on the mechanism, kinetics and thermodynamics of this DPT. To this end, we present a molecular dynamics (MD) study in the tight-binding density functional theory framework, and compare the reactivity of the isolated GC dimer with that of the same dimer embedded in a small DNA structure. The impact of Nuclear Quantum Effects (NQEs) is also evaluated using Path Integral based MD. Results show that in the isolated dimer, the DPT occurs via a concerted mechanism, while in the model biological environment, it turns into a step-wise process going through an intermediate structure. One of the water molecules in the vicinity of the proton transfer sites plays an important role as it changes H-bond pattern during the DPT reaction. The inclusion of NQEs has the effect of

speeding up the tautomeric-to-canonical reaction, reflecting the destabilization of both the tautomeric and intermediate forms.

4.1 Background

Base pairs mismatch is a phenomenon that occurs when the two DNA strands are not complementary. This means that the nucleotides which make up the base pairs do not match, [144] leading to an incorrect pairing in the double helix of the DNA. One of the possible mechanism for this mismatch to occur, proposed by Lowdin [145], is a double proton transfer (DPT) between the nucleotides causing an error during the replication of the DNA. Previous studies have shown that the Guanine-Cytosine (GC) pair is more prone to DPT than adenine-thymine (AT) [146, 147]. In the case of GC the most probable tautomer, as discussed in the litterature [148], is that in which H_4 is transferred from N_4 of cytosine to O_6 of guanine and H_1 from N_1 of guanine to N_3 of cytosine (see numbering in Figure 4.1A with the corresponding canonical, GC, and tautomeric, G*C*, forms). Thus, if the G*C* tautomer is present during the replication, there is a possibility to form the non-standard G*T and AC* pairs, which in turn provides a possible explanation for the conversion of GC into AT [149], as schematically shown in Figure 4.1B. This DPT reaction, which is potentially responsible for tautomerism in base pairs, has raised significant interest in the theoretical community, while only a few experiments have been reported [150, 151]. One detailed theoretical study of the reaction mechanism is reported by Ceron-Carrasco et al. [148] including the role of water micro-solvation using static DFT calculations. Notably, they found that the influence of the surrounding water molecules may change the mechanism from concerted to asynchronous, for the isolated dimer. Recently, Gheorghiu and coworkers studied the reaction pathways for tautomerism in GC and AT base pairs via quantum mechanics/molecular mechanics (QM/MM) simulations, showing that GC can form the short-lived G*C* tautomer, while A*T* tautomerism was not observed [147]. Similarly, Li et al. employed a QM/MM approach to study tautomerism, focusing on wobble GT pairs. [152] Very recently a QM/MM study by Soler-Polo and coworkers [153] using Umbrella Sampling [61] suggested that the water molecules and DNA environment destabilize the tautomeric form thus showing how nature has designed a robust base pair system. However, all these calculations do not consider the quantum nature of the proton, which is clearly an important aspect when studying proton transfer. [11–13]

The first study considering nuclear quantum effects (NQEs) on DPT in DNA base pairs is reported by Perez et al. [154] who combined Umbrella Sampling (US) with Path Integral Molecular dynamics (PIMD) [69]. They found that the inclusion of NQEs clearly desta-

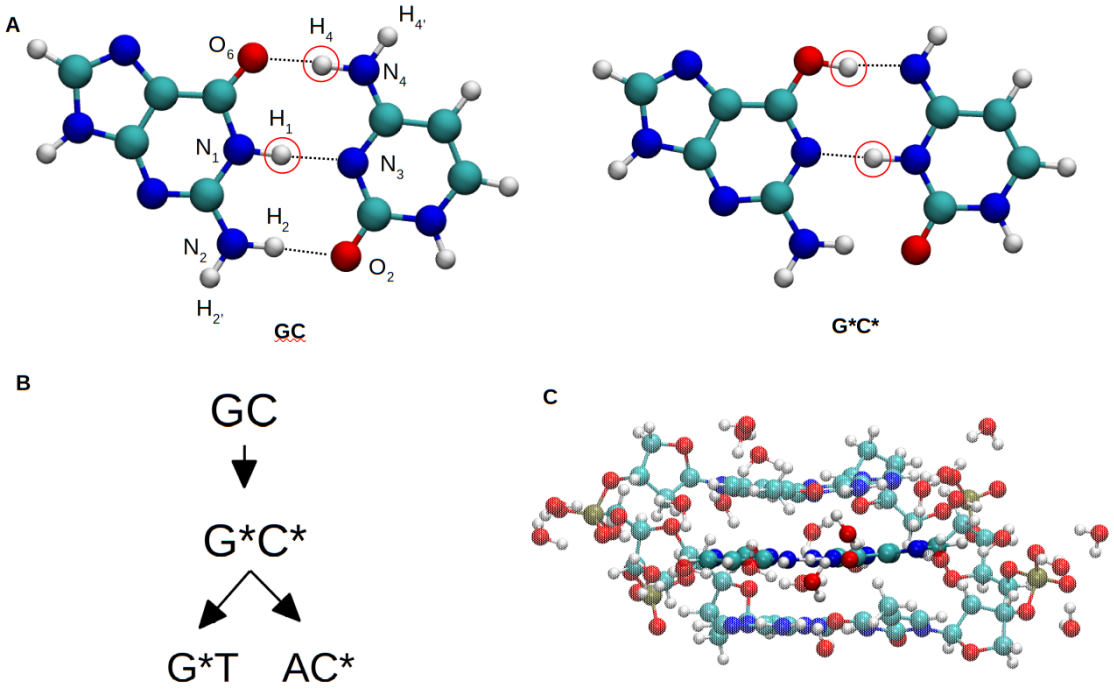

Figure 4.1: Panel A) Canonical (GC) and tautomeric (G*C*) forms of the Guanine-Cytosine base pair, with atom numbering. Transferring hydrogen atoms are shown in red circles. Panel B) Schematic representation of the tautomerism and its impact on the replication of DNA, with a GC base pair forming the G*C* tautomer and eventually leading to non-standard pairing G*T or AC* (A = Adenine and T = Thymine). Panel C) Structure of the 3BP-DNA model extracted from the 1D28 PDB structure [157] with the corresponding crystallographic water molecules.

bilizes the tautomeric form by flattening the free energy profile of the tautomeric state. However, they reduced the guanine-cytosine (GC) pair to a simpler model, taking into account only the atoms that directly take part in the mechanism. Slocombe et al. [155] studied this tautomerism using DFT and Machine Learning Nudged Elastic Band methods with a tunnelling correction to account for NQEs. They show that the G*C* tautomer has a lifetime long enough to survive during the cleavage process of DNA, while it is not the case for A*T*, which displays a very low reverse DPT barrier making the tautomeric form highly improbable. More recently, the DPT process in GC base pairs was also modelled using an open quantum system approach [156], suggesting that the tunnelling plays a central role even at biological temperature. However, all these studies that consider the quantum nature of the proton transfer, do not take into account the biological environment in the description of the proton transfer mechanism or only indirectly using a bath of harmonic oscillators to represent it [156].

4.2 Molecular Systems

In this chapter, we study the DPT reaction dynamics and thermodynamics including both environmental and nuclear quantum effects. The base pair taken into account is the GC one, since, as discussed previously, it is more prone to DPT than the AT base pair [146, 147, 155]. The reactivity is modelled using tight-binding density functional theory (DFTB) [42, 158, 159] which provides a good compromise between accuracy and computational accessibility, as was recently shown for different molecular systems [95, 133, 160]. Therefore in the following, electrons are always considered explicitly and quantum-mechanically, the label "classical" simulations refers only to the nuclear dynamics (as opposed to Path Integral simulations that include NQEs).

In particular, two models of GC base pairs are considered: (i) isolated GC, (ii) GC embedded in a DNA model composed of three base pairs with GC in the middle, denoted hereafter 3BP-DNA. The embedded model was extracted from the 1D28 PDB structure [157] and corresponds to a TGA (Thymine-Guanine-Adenine) sequence. For this last model, the crystallographic water molecules from the X-Ray structure were also included. This corresponds to a micro-solvation model that is computationally allowed in conjunction with DFTB and path integrals and can be compared to some previous literature results [147, 148]. Note that all the atoms are free to move in the simulations, including the crystallographic water molecules. The different systems studied are shown in Figure 4.1.

TGA is only one of the possible DNA sequences and in principle, the nature of the bases above and below the GC pair could affect the reactivity. However, Cerón-Carrasco and Jacquemin [161] studied all the possible DNA-trimers and found that the proton transfer energies differ only by a very small amount. For this reason, we chose a structure extracted from the crystallographic data of an actual DNA sequence. From this sequence, we extracted the aforementioned TGA trimer as well as a longer pentamer, TTGAG, used for validation of the mechanism (5BP-DNA).

Since the DPT occurs between the canonical GC and the tautomeric G*C* forms shown in Figure 4.1A, four characteristic distances can be used to describe the reaction: $r_1 = |O_6 - H_4|$, $r_2 = |N_4 - H_4|$, $r_3 = |N_1 - H_1|$ and $r_4 = |N_3 - H_1|$. Two collective variables are then typically used for DPT reactions: [10, 153, 162] $d_1 = r_2 - r_1$ and $d_2 = r_3 - r_4$. Note that the G*C* form is typically considered in the literature as the most relevant tautomeric form from all the possible configurations in alternative to the canonical (GC) form [148].

First, it is considered how the G*C* tautomer evolves dynamically once it is formed. To account for NQEs, ring polymer molecular dynamics (RPMD) simulations starting from

G*C* are performed. The RPMD approximation is based on the Path Integral formalism [21, 76] as shown in Chapter 2, and it is known to correctly describe proton transfer at room temperature [10]. It was also recently used to study the stability of base pairs and the influence of NQEs on hydrogen bonds in DNA base pairs [163]. A combined theoretical and experimental work proposed that the tautomeric form is accessible photochemically [150]: once formed in the excited state it can eventually relax back to the canonical form. It is thus useful to understand the dynamics of the reaction pathway connecting the tautomeric form to the canonical one, and the impact of NQEs and of the biological environment on this process.

4.3 Simulations set up

Atomic interactions are modeled via DFTB using the MIO set of Slater-Kostner parameters [164] modified to better describe N–H bonds (MIO:NH parametrization [165]). The third-order expansion was used and dispersion was added at the D4 level [24]. This specific set of parameters was chosen after a thorough comparison between DFTB and high-level electronic structure calculations. Indeed, the level of theory employed can clearly modify the energy profile of the reaction and consequently its mechanism and rate constant [147, 152, 155, 166]. The DFTB energies and gradients are calculated with the DFTB+ software (version 22.1) [24], and the RPMD and LMD simulations are implemented via our own in-house code.

4.3.1 DFTB Benchmarking

The isolated GC dimer in the gas phase was optimized with different DFT functionals with the 6-311++G(d,p) basis set, and notably: B3LYP [167–170], CAM-B3LYP [171] (with dispersion corrections using the D3 version with Becke-Johnson damping [172]), LC-ωHPBE [173] and ωB97XD [174]. Furthermore, a scan of the double proton transfer was performed to move from the canonical (GC) to the tautomeric (G*C*) form. To this end, the two protons are moved along the d_1 and d_2 coordinates on a 6×6 grid with a step of 0.15 Å

Since the CAM-B3LYP functional shows the best agreement with the CCSD(T)/MP2 results for the optimized GC dimer, this functional is the one used for the DPT scan and it is compared with DFTB. Results are reported in Figure 4.2. In this case the MIO:NH parameters yield the best agreement with CAM-B3LYP reference: the barrier is overestimated by 5 kcal/mol but the other DFFTB parametrizations yield even higher barriers. Furthermore, the potential energy surface obtained with the MIO:NH parameters is symmetric as found

	O_6N_4	N_1N_3	N_2O_2	E_{int}
CCSD(T)/aug-cc-pVQZ//RI-MP2/cc-pVTZ[175]	2.75	2.90	2.89	-28.2
B3LYP/6-311++G(d,p)	2.81	2.95	2.94	-25.71
CAM-B3LYP/GD3BJ/6-311++G(d,p)	2.78	2.93	2.91	-28.23
LC-ωHPBE/6-311++G(d,p)	2.80	2.94	2.93	-26.30
ωB97XD/6-311++G(d,p)	2.78	2.92	2.91	-30.21
MIO:NH	2.76	2.83	2.82	-26.57
3OB:NH	2.86	2.98	2.95	-21.79
MIO	2.82	2.96	2.89	-23.89
OB2	2.81	2.89	2.88	-26.43

Table 4.1: Distances (in Å) and interaction energy (in kcal/mol) of the GC dimer in gas phase. Atoms are numbered as in figure 4.1 panel A.

	$\Delta E^{\ddagger}$
CAM-B3LYP/GD3BJ/6-311++G(d,p)	34.39
MIO:NH	39.58
OB2	36.75
3OB:NH	46.01

Table 4.2: Energies (in kcal/mol) for the barrier connecting GC with G*C* as isolated system as obtained at different levels of theory.

from CAM-B3LYP calculations, while the other parameter sets show asymmetric profiles. Combining together the results shown here, it can be stated that MIO:NH parameters best reproduce reference calculations, slightly overestimating the reaction barrier but with a correct global behaviour. Therefore, even if the absolute value of the rate constants can be affected, the mechanism should likely not be impacted. We therefore chose MIO:NH for the reaction dynamics calculations.

All DFT calculations were performed with the Gaussian16 software [176], while the DFTB+ software [24] was employed for DFTB calculations.

4.3.2 Reaction Dynamics

To study the DPT process, reaction dynamics simulations are initiated in the tautomeric form (G*C*) and propagated on the DFTB Born-Oppenheimer surface. Nuclear quantum effects are included within the RPMD framework using a Langevin thermostat with optimal damping on the fluctuation modes of the ring polymer, corresponding to the Thermostatted RPMD (T-RPMD) algorithm [75], initially developed to improve vibrational spectrum calculations and later applied to reaction rates [177]. First, we performed short simulations (100 fs), the timestep is 1.0 fs, we used a Langevin thermostat at 300 K, with a relatively

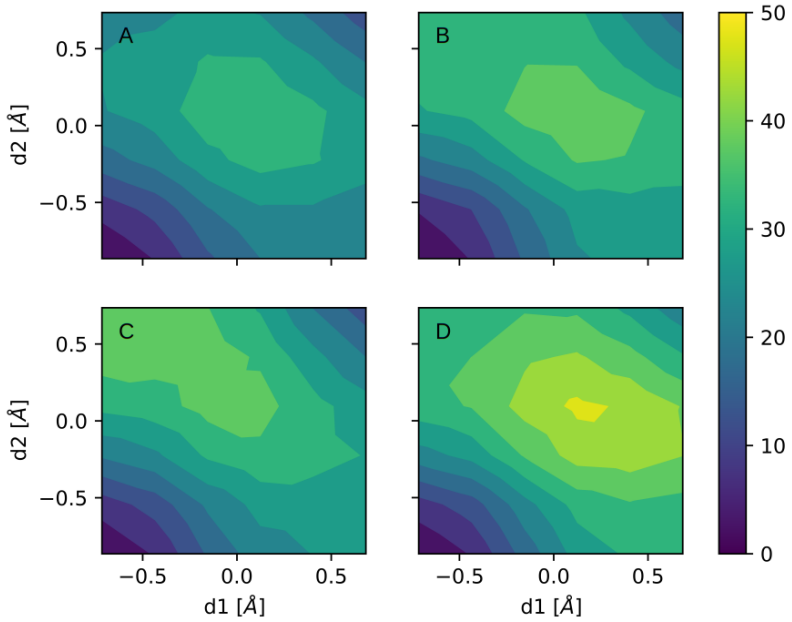

Figure 4.2: Potential energy surface of the DPT reaction as obtained from DFT calculations using CAM-B3LYP/GD3BJ/6-311++G(d,p) (panel A) and DFTB with different Slater-Kostner parameters: MIO:NH (panel B), OB2 (panel C) and 3OB:NH (panel D). Energies are in kcal/-mol, while d_1 and d_2 are the one described in the text.

high friction coefficient (γ = 100 ps $^{-1}$) to obtain initial positions and velocities for both LMD and T-RPMD [75] simulations using the BAOAB algorithm [178] (see Appendix C). We then perform an ensemble of trajectories (100 for each system) with the same Langevin thermostat with a timestep of 1.0 fs but with a lower friction coefficient (γ = 10 ps $^{-1}$) to reduce the impact of the thermostat on the reactivity. The simulation length was chosen to have above 80% of reactive events, and it is in the range of 1 to 50 ps, depending on the simulated system. The number of beads is set to 8, lower than the value used in previous studies on analogous system [10, 154], but it allows capturing the main impacts of NQEs. Further increase in the number of beads only causes a limited modification of the DPT rate constants, while keeping the mechanism unchanged, as we will show in Section 4.4. Note that in this set of simulations, the reaction pathway is not forced: in this way, we can observe where the system naturally evolves, and it could therefore form one or more final structures through potentially different mechanisms.

4.3.3 Free Energy Landscape

The free energy surfaces for the DPT process are obtained via Umbrella Sampling (US) [61] using the d_1 and d_2 collective variables (CVs) previously defined. We considered for each CV a total number of 6 windows between -0.75 and +0.75 Å (spacing of 0.3 Å), for a total of 36

points on a bi-dimensional grid. The force constant used is 12471 kJ/mol nm^2 and the time step is 1.0 fs. The runs were of 5 ps length. To obtain appropriate initial conditions for each run, we first performed a short trajectory with a high constraint on the CV. The standard Weighted Histogram Analysis Method (WHAM) [63] is used in order to reconstruct the free energy surface. In the case of PIMD simulations, 8 beads were used and the reaction coordinates d_1 and d_2 in the extended ring-polymer space are computed from the centroid atomic positions, as in previous studies of reactive dynamics using RPMD [58, 76, 179]. Results are listed in tables B.1, B.2, B.3, B.4 of Appendix B. The values are obtained using the block average method: each CV trajectory is divided into 5 blocks of 1 ps each, and 5 different FES are calculated (one from each block). The average value and the standard deviation are then obtained from these different evaluations of the FES, discarding the first picosecond, so using the last four blocks.

4.4 Results

4.4.1 Dynamics Simulations

Almost all the trajectories in the G*C* tautomeric form spontaneously end up in the canonical form during the simulation time length. However, the reaction rates and mechanisms are dramatically affected by both NQEs and the DNA environment. As discussed previously, two collective variables are typically used to describe the DPT, d_1 and d_2, where d_1 corresponds to the external proton that interacts more strongly with the surrounding water molecules (when present). Figure 4.3 shows the projection of the direct dynamics on the d_1-d_2 plane (the trajectories evolve in the full-dimensional phase space without any constraint), as obtained for the isolated base pair and for the 3BP-DNA system. In both cases, classical (LMD) and quantum (T-RPMD) results are shown. The GC canonical form corresponds to negative values (the minimum is around d_1 = -0.7 Å and d_2 = -0.8 Å), and the tautomeric G*C* form corresponds to positive values (with a minimum around d_1 = 0.7 Å and d_2 = 0.7 Å).

For the isolated system, the DPT occurs in a concerted way (i.e. along the diagonal in the d_1–d_2 plot of Figure 4.3) similarly to what was previously suggested [162]. As it can be noticed, there is a slight deviation from the exact diagonal that corresponds to a slightly asynchronous mechanism, but it disappears when taking NQEs into account. Indeed, the inclusion of NQEs has the effect of significantly accerating the DPT reaction and of making the mechanism fully synchronous. From direct dynamics simulations it was possible, as previously, to evaluate reaction rate constants (k) and the corresponding life-times ($\tau = 1/k$)

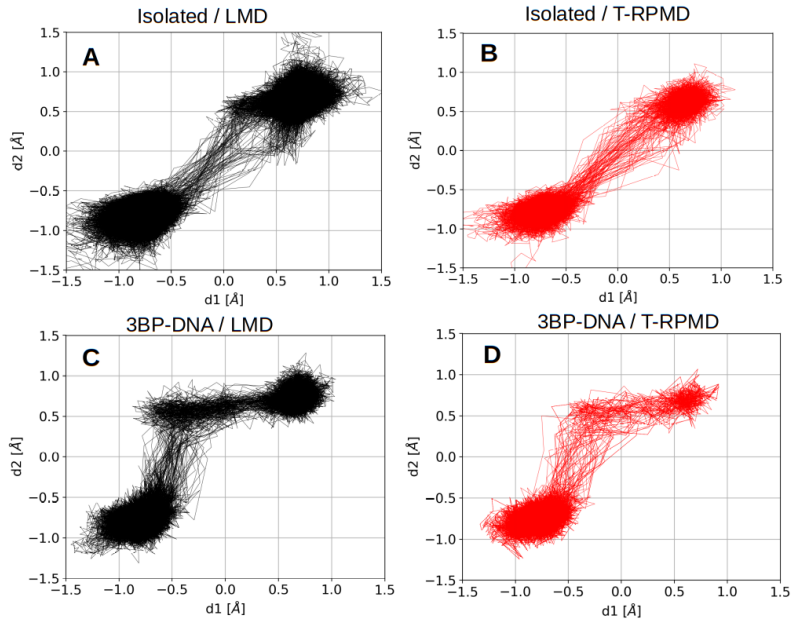

Figure 4.3: Trajectories projected on the d_1-d_2 collective variables plane, as obtained from DFTB-based direct dynamics simulations starting from the G*C* form: LMD and T-RPMD for the isolated base pair (panels A and B, respectively), LMD and T-RPMD for the DNA model environment (3BP-DNA, panels C and D)

from an exponential fit of the population decay of the initial state (here the G*C* tautomer), as shown in Figure 4.4. The rate constants and corresponding uncertainties, obtained via the bootstrap method as done for simulations presented in the previous chapter, are listed in the Table 4.3. The G*C* rate constant of the isolated system estimated from the T-RPMD trajectories (2.4 ps $^{-1}$) is about 30 times greater than the LMD one (0.0821 ps $^{-1}$). This result is in agreement with Perez et al. [154] who show that the barrier in the free energy pathway connecting G*C* with GC almost disappears when including NQEs, using Path Integral Umbrella Sampling. Note that, when increasing the number of beads (Table 4.4) the rate constant furhter increases from 2.4 to 4.0 ps $^{-1}$, but the mechanism does not change, showing that 8-beads results are not totally converged for the isolated system but already capture the correct trend. By substituting with the deuterium the hydrogen involved in the DPT, the impact of NQEs is reduced, as expected. The ratio k^H/k^D is equal to 3.8$\pm$0.8 for T-RPMD results, and to 1.2 $\pm$ 0.1 in LMD simulations.

When the model DNA environment is taken into account, the DPT mechanism clearly changes: a step-wise mechanism is observed in which H_4 moves first from O_6 to N_4, forming an intermediate structure and then, in a second step, the other proton (H_1) moves and finally forms the neutral canonical form (GC). This intermediate structure is described in more detail below.

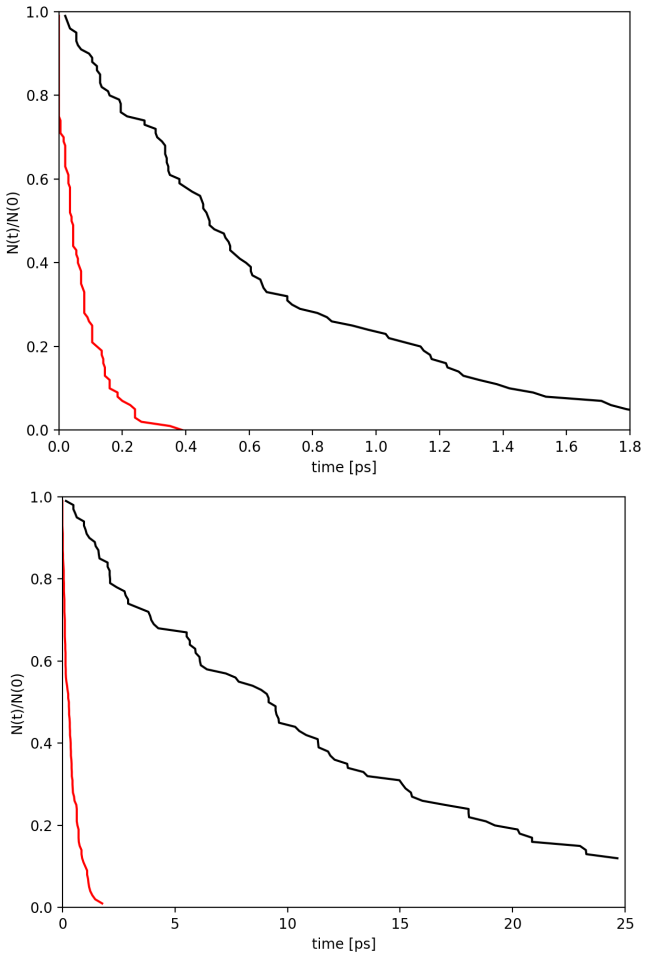

Figure 4.4: Population decay for the isolated dimer (top panel) and for the 3BP-DNA system (bottom panel). In black, we show results obtained from classical trajectories, while those from T-RPMD simulations using 8 beads are reported in red.

System	k $[ps^{-1}]$	k^{T-RPMD}/k^{LMD}	k^H/k^D
Isolated/LMD	0.0821 ± 0.009	–	–
Isolated/LMD/Deuterated	0.070 ± 0.008	–	1.2 ± 0.1
Isolated/T-RPMD	2.4 ± 0.2	29 ± 5	–
Isolated/T-RPMD/Deuterated	0.64 ± 0.07	9 ± 1	3.8 ± 0.8
3BP-DNA/LMD	1.7 ± 0.2	–	–
3BP-DNA/LMD/Deuterated	1.3 ± 0.1	–	1.3 ± 0.2
3BP-DNA/T-RPMD	18 ± 2	11 ± 2	–
3BP-DNA/T-RPMD/Deuterated	6.9 ± 0.9	5 ± 1	2.6 ± 0.5

Table 4.3: Comparison between the rate constant (k) of the reverse reaction (from tautomeric to canonic) for the two systems: the isolated GC dimer and the 3BP-DNA model as obtained from LMD and T-RPMD (using 8 beads) DFTB-based direct dynamics simulations. We report also the values obtained for deuterated systems. Ratios between LMD and T-RPMD as well as hydrogen (H) and deuterium (D) rate constants are also shown.

System	$k\ [ps^{-1}]$	$\tau\ [ps]$
Isolated/T-RPMD (16 beads)	4.0 ± 0.4	0.24 ± 0.02
3BP-DNA/T-RPMD (16 beads)	14 ± 2	0.068 ± 0.009

Table 4.4: Rate constants and corresponding life-times as obtained from DFTB-based T-RPMD direct dynamics simulations using 16 beads.

The inclusion of the DNA model environment also has the effect of further destabilizing the G*C* tautomer, as shown by the corresponding rate constants reported in Table 4.3. This result is in agreement with recent QM/MM simulations showing that the tautomeric form is thermodynamically destabilized when a DNA-like environment is included [153]. The acceleration due to NQEs is slightly reduced compared to the isolated dimer, as k^{T-RPMD}/k^{LMD} is lowered down to 11. Isotopic substitution results also reflect this finding as k^{H}/k^{D} is now reduced to 2.6. Interestingly, when performing T-RPMD simulations with 16 beads the rate constant does not increase further and remains essentially unchanged within statistical uncertainties (Table 4.4). More importantly, the mechanism is unchanged (only few trajectories do not follow a step-wise process, as it can be seen from Figure 4.5) which allows to use only 8 beads for the free energy calculations and largely reduces their computational load.

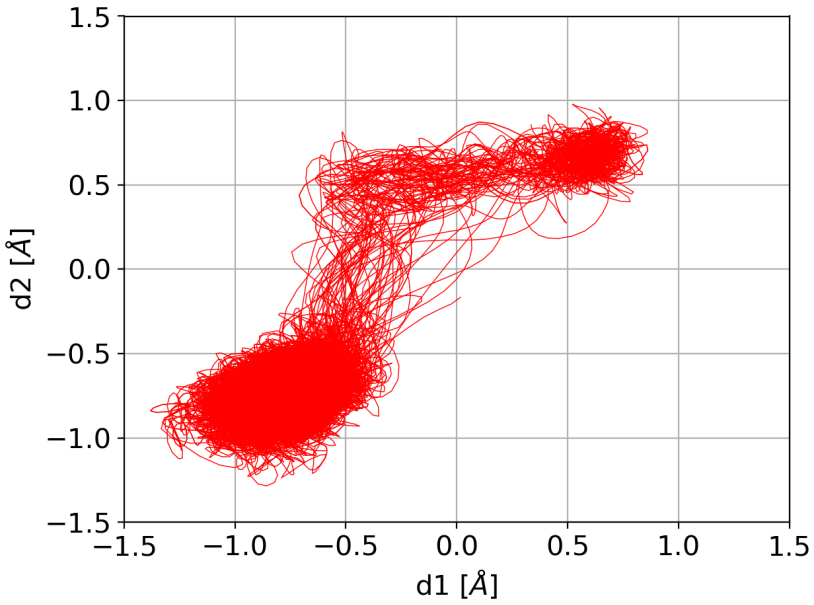

Figure 4.5: Reaction dynamics simulations: trajectories in the d_1 vs d_2 plane as obtained for 3BP-DNA system in RPMD simulations with 16 beads.

To make sure that this effect does not depend on the size of the DNA model, additional simulations are performed with 5 base pairs (TTGAG sequence with the crystallographic water molecules, corresponding to the addition of a thymine and a guanine at the beginning

and at the end of the TGA sequence, rispectively). No significant change was observed in the mechanism for these larger structures, as it can be seen from Figure 4.6.

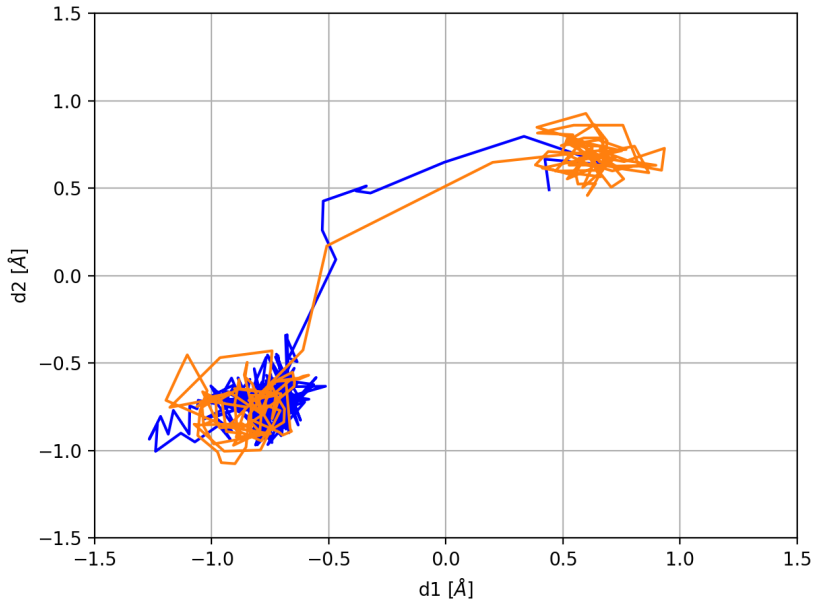

Figure 4.6: Classical reaction dynamics simulations: trajectories in the d_1 vs d_2 as obtained for a 5 base pairs system (blue line) and for a GC dimer with few water molecules next to the DPT sites (orange line).

For this reason in this project we used the 3 base pairs model (with 8 beads for T-RPMD) for further discussions and US simulations.

4.4.2 Free Energy Landscape

The Figure 4.8 shows the free energy surfaces (FES) for the isolated GC base pair and for the 3BP-DNA system, as obtained for both classical and PIMD US simulations. The classical FES of the isolated dimer is almost symmetric with respect to the diagonal with a small deviation from the diagonal, of about 0.2 Å, which is in agreement with the corresponding direct dynamics results. When including NQEs, the barrier decreases, reflecting also in this case the results of the direct dynamics and in agreement with the FES reported on a simplified model by Perez et al [154]. Furthermore, the reaction pathway becomes fully symmetric as also found in direct dynamics simulations. Table 4.6 summarizes the different free energy barriers associated with the process as obtained from DFTB-based US simulations, using both classical and Path Integral approaches. The inclusion of NQEs has an effect on both the free energy difference between GC and G*C* and on the associated barriers. Notably, the barrier associated with the reaction G*C* $\rightarrow$ GC decreases from 4.5 to 2.3 kcal/mol. However, the barrier associated with the GC $\rightarrow$ G*C* reaction remains relatively high, thus

making the process unlikely in the absence of any particular source of activation.

When considering the DNA model environment, the FES profile changes dramatically: the tautomeric form becomes much less stable and the connection with the canonical form does not follow the diagonal (and consequently it is not possible to define a single saddle point in the FES as was possible, on the contrary, in the isolated system reaction). Notably, this is in agreement with the direct dynamics simulations, performed from the tautomeric to the canonical form. An intermediate form is observed (shown in Figure 4.7) in which only one proton has been transferred, corresponding to $d_1 \simeq 0.6$ Å and $d_2 \simeq +0.6$ Å . The d_1 coordinate, corresponding to the transferring proton exposed to the solvent, is the same as in the canonical form, while the d_2 coordinate is mostly in the tautomeric configuration. This new state is lower in free energy than the tautomeric form (14.4 vs 19.1 kcal/mol) and has a particular charge distribution character. In order to obtain the charges on the two base pairs, specific simulations of the 3BP systems are performed: canonical (GC), tautomeric (G*C*) and intermediate (I). In each case, the total time of the simulation is 5 ps (with 1 fs timestep) at 5 K with a friction constant of 10.0 ps^{-1} for the Langevin thermostat. To fix each simulation in the desired state, a restraint is applied on the d_1 and d_2 distances, depending on the protonation state under observation. In this way the system is cooled down and reaches a structure of (almost) minimum energy within the applied restraint. The Mulliken partial charge on each atoms are calculated and summed up to obtain the total charge on the G and C bases. Table 4.5 reports the results obtained for the three states (tautomeric, intermediate and canonical), for the atoms numbered in Figure 4.1 panel A. Notably, while in the tautomeric and canonical form the total charge on each base pair is equal to -0.1 for the guanine and 0.0 for the cytosine, in the intermediate form the total charge is equal to -0.8 on the guanine and +0.6 on the cytosine, showing a charge separation which is stabilized by the surrounding water molecules. It should be noted that the trajectories sample a region of the (d_1,d_2) space close to the intermediate minimum in the classical FES, but never reach it. This shows that the exact location of the intermediate as a free energy minimum is difficult to reach dynamically from the G*C*, but its presence has an important impact on the shape of the FES and consequently on the dynamics.

When including NQEs, there is no major change in the free energy difference between GC and G*C* and the FES globally keeps a similar shape, but the intermediate is no more a clear minimum since the barriers connecting it to the canonical and tautomeric states are lost.

The decrease in the stability of the tautomeric form in the DNA model environment with respect to the isolated system is also visible from the donor-acceptor distributions reported in Figure 4.9(A) for LMD trajectories. The donor-acceptor distance O_6N_4, corresponding

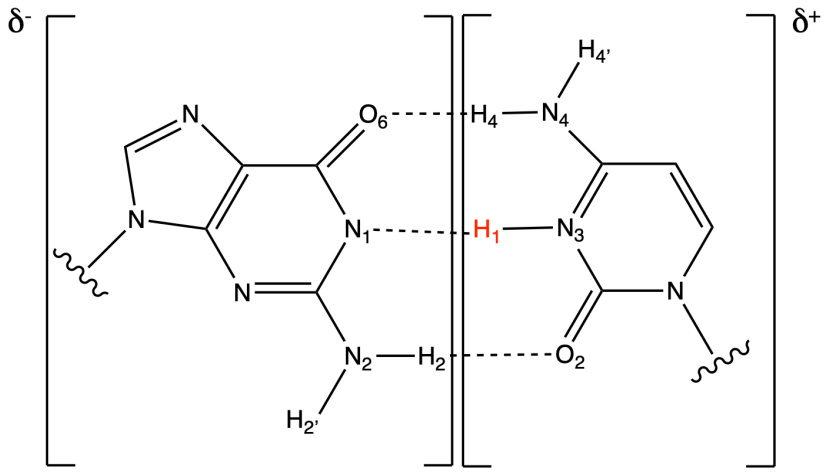

Figure 4.7: Intermediate structure obtained from Umbrella Sampling simulations in the 3BP-DNA system.

Atom	GC	I	G*C*
O_6	-0.6	-0.7	-0.4
C_6	+0.5	+0.5	+0.5
N_1	-0.2	-0.6	-0.5
H_1	+0.2	+0.3	+0.3
H_4	+0.2	+0.3	+0.2
$H_{4'}$	+0.2	+0.2	+0.2
N_4	-0.3	-0.3	-0.6
C_4	+0.3	+0.3	+0.3
N_3	-0.3	-0.2	-0.2
Base			
G	-0.1	-0.8	-0.2
C	+0.1	+0.6	0.0

Table 4.5: Mulliken charges obtained on canonical (GC), intermediate (I) and Tautomeric (G*C*) structures in 3BP-DNA simulations. We report the values obtained on atoms numbered as in Figure 4.1 and also the result of the summation on the G and C atoms.

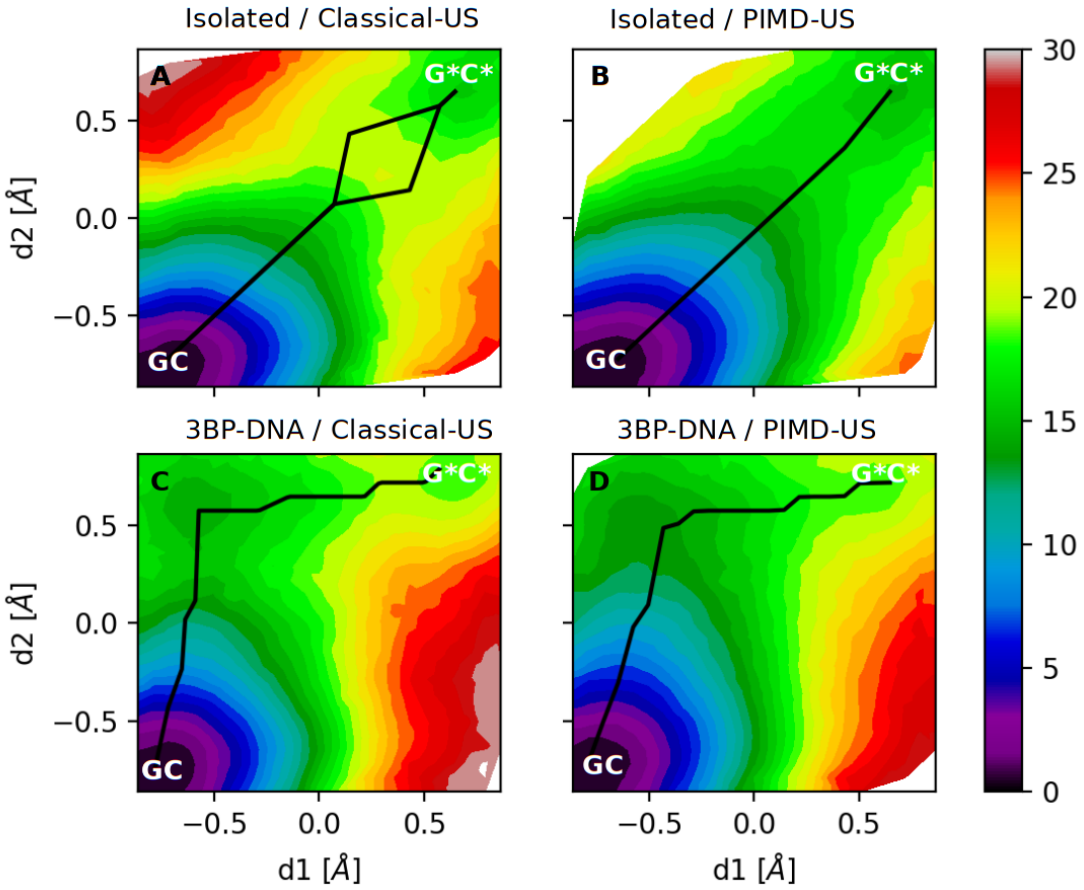

Figure 4.8: Free Energy Surfaces as a function of d_1 and d_2 collective variables as obtained from DFTB-based Umbrella Sampling (US) simulations: A) classical US of the isolated dimer; B) PIMD US of the isolated dimer; C) classical US of 3BP-DNA system; D) PIMD US of the 3BP-DNA system. Canonical (GC) and tautomeric (G*C*) states are indicated as well as the minimum free energy path connecting these two states as black solid line.

reaction	value	Isolated system	3BP-DNA
GC → G*C*	ΔF_{Cl}	15.0 ± 0.1	19.1 ± 0.3
GC → G*C*	ΔF_{PI}	14.1 ± 0.4	19.1 ± 0.2
GC → G*C*	$\Delta F^{\ddagger}_{Cl}$	19.5 ± 0.1	–
GC → G*C*	$\Delta F^{\ddagger}_{PI}$	16.4 ± 0.1	–
G*C* → GC	$\Delta F^{\ddagger}_{Cl}$	4.5 ± 0.1	–
G*C* → GC	$\Delta F^{\ddagger}_{PI}$	2.3 ± 0.4	–
GC → Intermediate	ΔF_{Cl}	–	14.4 ± 0.1

Table 4.6: Free energy differences (in kcal/mol) as obtained from US DFTB-based simulations for the isolated system and in the DNA model environment (3BP-DNA). The indices Cl and PI refer to classical and Path Integral Umbrella Sampling simulations, respectively, while ‡ denotes the free barrier barrier to pass the saddle point of a given reaction. The "Intermediate" label refers to the locally stable state observed in Classical Umbrella Sampling in the 3BP-DNA model and represented in Scheme 4.7.

to the external proton transfer, is shorter in the tautomeric structure than in the canonical one. In the isolated system, this decrease in the distance is of about 0.04 Å, while in the DNA model it is close to 0.1 Å, showing that for this latter system, it is easier for the proton H_4 to move from O_6 to N_4, causing the formation of the intermediate previously discussed.

An important finding of this study is that the DNA model environment has a crucial impact and modifies the DPT mechanism. By investigating the different trajectories we found that a key role is played by the surrounding water molecules (in the present simulations the crystallographic ones are included). More precisely, Figure 4.9 (B) shows the distance distribution between O_6 and the two hydrogen atoms of the closest water molecule (labeled HW_1 and HW_2) for the three different forms: the canonical (green), the tautomeric (blue) and the intermediate forms (yellow). In the canonical form, the O_6 atom is strongly H-bonded to this water molecule via one of its hydrogen atom (and of course to the H_4 atom of the cytosine base). In the tautomeric form, conversely, O_6 is covalently bound to the transferred H_4 atom and it is, therefore, less prone to form a hydrogen bond with the surrounding water. The nearby water molecule now interacts with O_6 via its two hydrogen atoms in a weaker way, as shown by the O_6-HW distributions. In the intermediate form, the water molecule moves back to the configuration where it makes a strong directional H-bond with O_6 (now H_4 is back to the cytosine as in the canonical form). This process is schematically shown in panels C and D of Figure 4.9. In other words the driving force that pushes the tautomeric form through the formation of the intermediate is the formation of an N–H bond and, from the point of view of the O_6 atom, the formation of two strong H-bonds: one with the cytosine base and one with the nearest water molecule.

Additional simulations of an isolated G*C* structure in which only few water molecules are included in the vicinity of the O–HN H-bond confirm this picture: the trajectories show a pathway which is similar to what is observed in the full DNA-model structure, as reported in Figure 4.6. These results, together with that of the 3BP-DNA model show the importance of micro-solvation on the reactivity. Notably, in agreement with the works by Tolosa et al. [180] and Gheorghiu et al. [147] with different methods and approaches, micro-solvation is crucial to open a new reaction pathway connecting GC with G*C* via an intermediate. An important finding of the present study is that this intermediate is partially destabilized by NQEs.

4.5 Conclusions

Summarizing, this project can be seen as an exhaustive study of how NQEs and the environment can affect the mechanism of DPT in the GC base pair, using a combination of

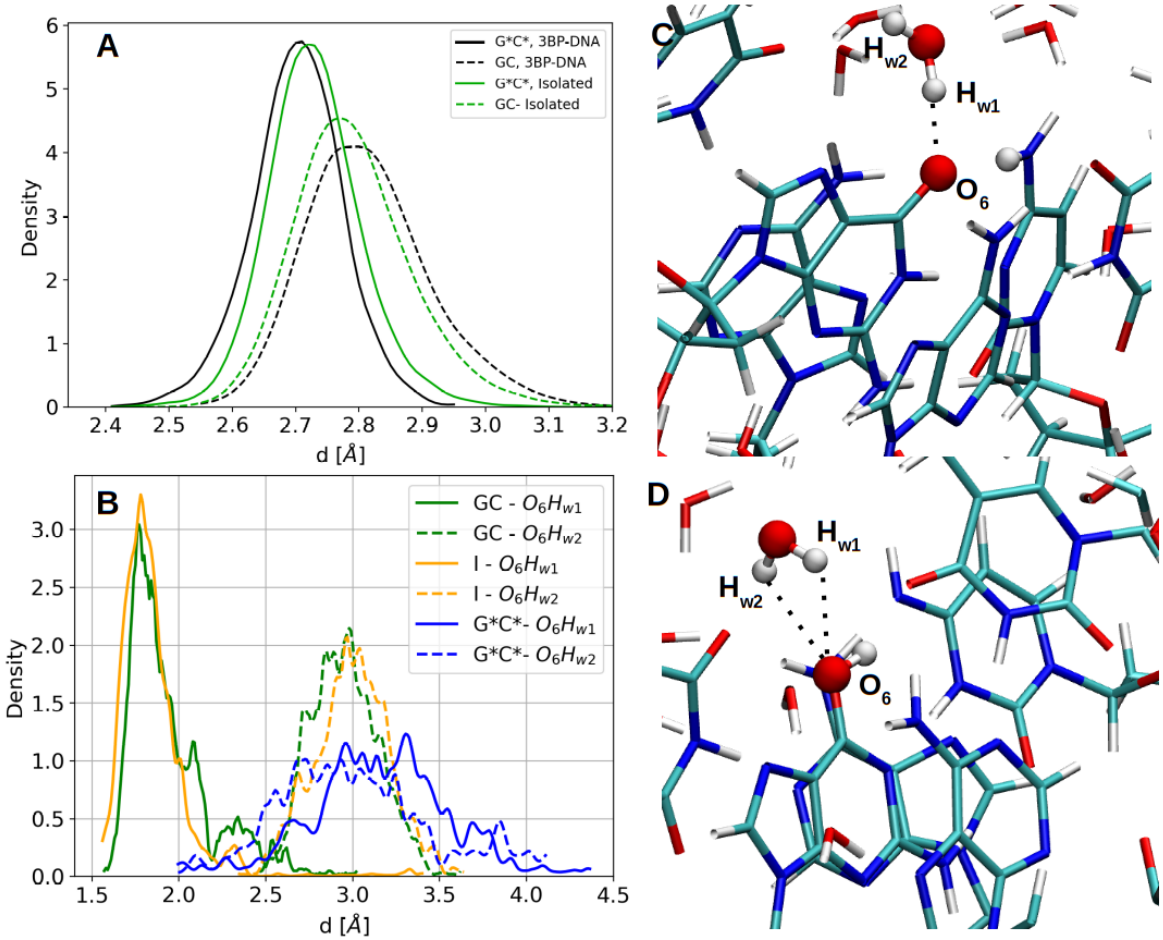

Figure 4.9: Panel A: O–N distance distributions in the isolated system (green) and in the 3BP-DNA (black) as obtained for the canonical (dashed line) and the tautomeric (continuous line) forms from DFTB-based simulations. Panel B: Distance distributions between the O_6 and hydrogen atoms of the nearest water molecule (H_{w1} and H_{w2}) obtained from DFTB-based classical US simulations: canonical (green), tautomeric (blue) and intermediate (yellow) forms. Panels C and D: two prototypical snapshots of the canonical (C) and tautomeric (D) forms where the atoms involved in DPT reaction are highlighted as balls.

direct dynamics simulations, to characterize the spontaneous decay of the tautomeric form, and Umbrella Sampling simulations to obtain the free energy surface of this reaction.

For the isolated dimer, the mechanism is concerted though slightly asynchronous and the effect of NQEs is to accelerate the reaction by approximately a factor of 30, making the mechanism fully synchronous. Indeed, while in LMD simulations the minimum free energy pathway passes about 0.2 Å away from the diagonal of the d_1-d_2 plot, it moves to follow the diagonal almost exactly when NQEs are included. When the environment is taken into account in the 3BP-DNA structure, the mechanism changes completely as the DPT becomes a step-wise reaction. In this case, we demonstrate the importance of the role of the surrounding water molecules, that stabilize an intermediate structure with opposite

partial charges on each base pair. As for the isolated base pair, NQEs speed up the DPT process (about 10 times compared to classical simulations for the 3BP-DNA structure). The free energy landscape is also deeply modified as the intermediate form ceases to be a local minimum, making the proton transfer essentially barrierless.

These results clearly show that further computational studies must include the environment at least by considering more than one base pair and (crystallographic) water molecules. This will be important to investigate if modifications in the base pairs, for example through methylation which may result from carcinogenic agents, [181, 182] have an effect on the stability of the tautomeric and/or intermediate forms. Finally, the step-wise mechanism reported for GC tautomerism could be important to unravel the biochemistry of GC-rich DNA regions which are associated with the so-called CpG islands [183] and related to the promoter region of the genome. [184, 185]

Chapter 5

Heavy Atoms tunnelling in Cope rearrangements

In the current Chapter, we focus our attention on the study of Heavy Atom tunnelling using ring polymer molecular dynamics (RPMD) in order to include nuclear quantum effects (NQEs). The reaction under study is the Cope rearrangement of semibullvalene, which was shown in the literature to display tunnelling at low temperatures. Here we compute the rate constant over a wide range of temperatures, from 25 K to 300 K, comparing classical and RPMD results. The rate constants are firstly obtained with the Eyring formulation, within transition state theory (TST) approximation. By estimating the associated recrossing factor, we evaluate the accuracy of the transition state theory. Our results show that on one side tunnelling increases the reaction rate as a consequence of the decrease of the free energy barrier, in the low-temperature regime. On the other hand, the reaction rate TST is decreased by the recrossing factor, which becomes significantly lower than 1 at low temperatures when NQEs are included, while it remains almost equal to 1 in the classical case.

5.1 Background

Quantum tunnelling is a physical process that was first predicted by Hund [186] for electrons in 1927, stating that it occurs when a light particle passes through a barrier without having enough energy to classically overcome it. Five years later, Bell [187] and Wigner [188], discussed this concept in the study of chemical reactions, suggesting that light atoms, like hydrogen, can exhibit tunnelling. In particular, after the discovery of deuterium in 1950 [189], it was possible to study the contribution of tunnelling and Zero Point Energy (ZPE) in organic reactions, thanks to the substitution of hydrogen atoms with their heavier isotope. This allows us to make a comparison between the rate constant obtained with the hydrogenated and deuterated systems. By evaluating the ratio between the two rate constants (k_H and k_D) it is possible to estimate how much ZPE or tunnelling impact the mechanism of the reaction (kinetic isotopic effect). Many publications followed [190–194] showing how tunnelling and ZPE can affect rate constants when light atoms such as hydrogen are involved.

The general idea suggested that for reactions which involve heavier atoms (for instance carbon), the tunnelling effect should not have a central role. However recent observations and theoretical studies have shown that atoms heavier than hydrogen can also undergo tunnelling [195–198]. In particular, one of the first papers that suggest the relevance of heavy atom tunnelling in an organic reaction was published in 1983 by Carpenter [197]. The approach was fully theoretical, and the self-tautomerization of cyclobutadiene was studied through an analytical potential: the energy levels were calculated within the harmonic oscillator approximation, while the transmission coefficient was obtained from Bell's formula [199]. Once the rate constant was calculated, he found that carbon tunnelling can contribute up to 97% of the total rate constant below 0 °C. A list of further organic reactions is presented in a recent Review by Castro and Karney [195], who describe in detail the theory and the reactions studied up to now.

Recently experiments and theoretical calculations showed, that a candidate for a Heavy Atom tunnelling (HAT) is the Cope rearrangement of semibullvalene [196, 198, 200, 201]. The hypothesis that tunnelling could affect the mechanism of this reaction, was first predicted theoretically by Borden et al in 2010 [198], with the small curvature tunnelling (SCT)+TST approach. The authors also suggested an experiment to provide evidence for tunnelling. This experiment was performed 10 years later by Sander and coworkers [196]. Since the reaction is symmetric (see Scheme 5.1), it is impossible to distinguish the two isomers through Infrared Spectroscopy (IR). For this reason, the first step is a deuteration (in position 2 or 4, Scheme 5.1). In this way, it is possible to recognize the product or

the reactant through IR measurements. What is done in practice, is to prepare one of the two deuterated isomers, and observe the reaction via IR at low temperatures. Sanders and coworkers found that thermodynamical equilibrium is obtained even below 10 K, which is classically impossible if we consider that the barrier (found experimentally [202]) is about 6 kcal/mol. The most probable explanation is a strong enhancement of the reaction via tunnelling.

To theoretically study this family of reactions, transition state theory (TST) including tunnelling corrections via small curvature tunnelling (SCT) [15] is typically employed [201]. This approach relies on the assumption that the tunnelling path proceeds along the Minimum Energy Path (MEP), and that if the concave curvature of the barrier is small enough, it can be described within the harmonic approximation. In this way the tunnelling probability can be obtained, and a transmission coefficient corrects the classical TST rate constant for the tunnelling effect. Another efficient method to take into account tunnelling in the rate constant, is the semi-classical TST (SC-TST) [16], which requires anharmonic constants of the system in the reactant and transition state geometry. Both of these methods are the only ones used up to now to describe tunnelling when the reaction involves heavy atoms. The main drawback of the SCT method is that one has to know in detail energy, gradients and vibrational frequencies of the region of the potential energy surface (PES) where the transition state is located, and it is highly sensitive to the height and width of the barrier, while the SC-TST theory scales poorly with the size of the system since it requires the calculation of the anharmonic frequencies.

An alternative approach is to use ring polymer molecular dynamics which naturally includes the NQEs in the dynamics of the system. Furthermore, this method was extensively used in literature to predict tunnelling when light atoms are involved [58, 72, 107, 203]. In particular, RPMD was found to give the exact rate even in the deep tunnelling regime for a parabolic barrier [72], while in the case of three specific bimolecular reactions (H+H_2, Cl+HCl and F+H_2) it gives the rate constant within a factor of 3 in comparison with exact results [107]. In general, in the deep tunnelling regime, RPMD slightly underestimates the rate constant in the case of a symmetric barrier and overestimates the rate constant in the case of an asymmetric one [204].

In the present Chapter, we want to evaluate the accuracy of TST, in classical and RPMD cases, by estimating the associated recrossing factor. We compute the reaction rates in the TST approximation with RPMD for the reaction of Cope rearrangement of semibullvalene (see Figure 5.1), and we compare it with classical results. Then, we correct the TST rate constant by including the recrossing factor. In order to do it, we have performed MD trajectories from the transition state, using a procedure previously employed by Hinsen and

Roux [58], Suleimanov[203] and Craig and Manolopolous [76].

The electronic structure was treated within a semiempirical approach: the Tight Binding DFT method [43] (see Chapter 1).

5.2 Simulation Set-Up

5.2.1 Molecular System and Reaction

In Scheme 5.1 we show the Cope rearrangement of semibullvalene which is a [3,3]-sigmatropic rearrangement. In Table 5.1 we report the potential energy barrier and imaginary frequency at TS obtained at different levels of theory: Coupled Cluster (CCSD(T)), Density Functional Theory (DFT) and Density Functional Tight Binding (DFTB) methods using various functionals and parametrizations. Similarly to the approach followed for choosing the best Slater Kostner (SK) set of parameters in the case of the Guanine-Cytosine base pair (see Chapter 4), we first compare the single-point values obtained from DFT with the ones obtained from CCSD(T) and then we make a comparison with the DFTB. As it can be noticed from Table 5.1 all the functionals except B3LYP and LC-ωPBE-D3BJ provide a potential energy barrier in agreement with the CCSD(T) results. Furthermore, the energies obtained with DFTB are in good agreement with DFT and CCSD(T). For what concerns the imaginary frequency at the TS, the best set of parameters is 3OB, which shows good agreement with the functional ωB97XD, while OB2 is in agreement with CAM-B3LYP-D3BJ.

In a second step, as was done previously in Chapter 4, we perform a qualitative analysis to see what SK set of parameters best captures the reaction mechanism. To this end, we performed a scan of the potential energy surface as a function of d_1 and d_2, which are the distances between atoms C_1-C_2 and C_3-C_4 respectively (see Figure 5.1 for atom numbering). The potential energy scan was performed with the DFT functional ωB97XD with the 6-31G basis set, which was shown to be a good compromise between accuracy and computational cost [205] and compared with the three sets of DFTB parameters (MIO [164], OB2 [206], 3OB [207–209]). The results are shown in Figure 5.2. It can be seen clearly that the most suitable SK set of parameters for this reaction is the 3OB since it shows a concerted and synchronous mechanism, in agreement with DFT results and in contrast with MIO and OB2. This is not surprising, since 3OB includes the third-order dispersion, it was especially parametrized to study organic molecules and it has a better parametrization for the vibrational stretching frequencies of the carbon double bonds [207].

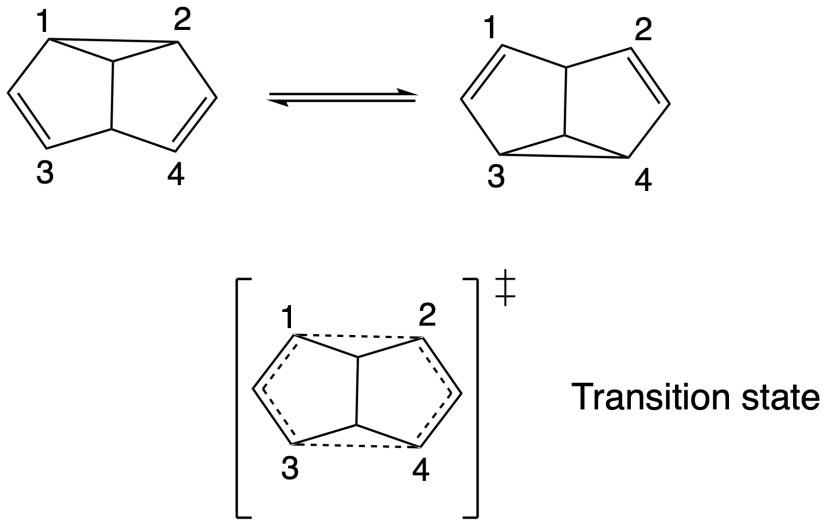

Figure 5.1: Schematic representation of the Cope rearrangement reaction of the semibullvalene

Table 5.1: Activation energies for tautomerization: potential energy barrier, potential energy including ZPE. Values are in kcal/mol. We report also the frequency of the imaginary frequency at the TS, in cm^{-1}.

Method	$\Delta E^{\ddagger}$	$\Delta^{\ddagger}E$+ZPE	$\omega^{\ddagger}$
CCSD(T)/6-311G(d,p)	9.08		
CCSD(T)/cc-pVDZ	8.87		
CCSD(T)/6-311G(d,p)//HF/aug-cc-pv5Z	8.49		
CCSD(T)/cc-pVDZ//HF/aug-cc-pv5Z	8.24		
ωB97XD/6-311++G(d,p)	8.35	7.34	401
ωB97XD/6-31G	8.50	7.46	388
M062X/6-311++G(d,p)	8.79	7.73	380
B3LYP/6-311G(d,p)	5.09	4.02	196
CAM-B3LYP-D3BJ/6-311++G(d,p)	9.12	7.90	355
LC-ωPBE-D3BJ/6-311++G(d,p)	13.9	12.68	606
MIO	10.22	9.12	278
OB2	8.4	6.99	322
3OB	10.5	9.07	413

5.2.2 Free Energy Calculations

To obtain the free energy as a function of the reaction coordinate, we proceed as follows. Firstly, we define the collective variable (CV) as:

$$d = d_1 - d_2 \tag{5.1}$$

where d_1 is the distance between C_1 and C_2 and d_2 is the distance between C_3 and C_4

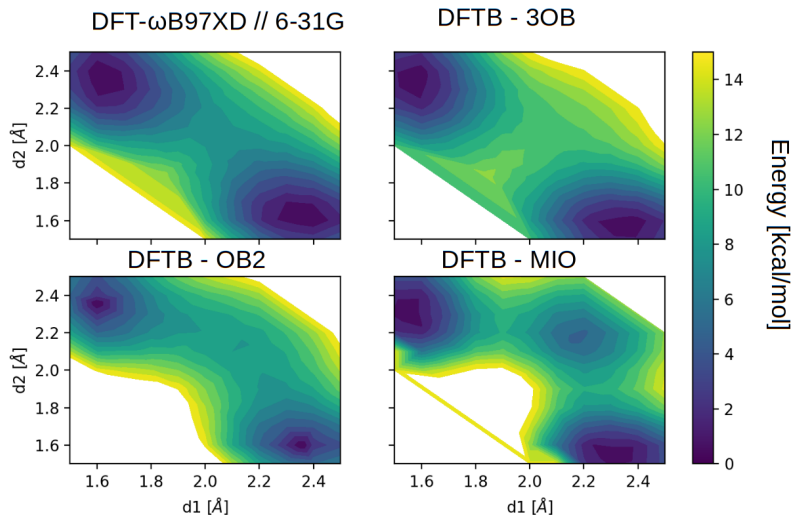

Figure 5.2: Scan of the potential energy surface as a function of two reaction coordinates, d_1 and d_2 as obtained with DFT (ωB97XD/6-31G) and different DFTB Slater-Kostner parameters.

(see atom numbering in Scheme 5.1). We then use umbrella sampling to force the system to explore the CV coordinate between -0.88 and +0.88 Å for a total of 37 equally spaced windows. For each window, we run about 50 ps of molecular dynamics simulations, with a time step of 1.0 fs and a friction coefficient of 10 ps $^{-1}$ for the Langevin thermostat. The force constant k is chosen in order to obtain a normal distribution of the umbrella sampling histogram:

$$\exp\left(-\frac{k(d-d_0)^2}{2k_bT})\right) = \exp\left(-\frac{(d-d_0)^2}{2\sigma^2}\right) \tag{5.2}$$

where d_0 represents the value of the CV centre for each constrained simulation, while σ is equal to the spacing between each d_0 in order to provide a significant overlap in the final histogram.

We have thus performed umbrella sampling with both classical (ClMD-US) and path integral based (PI-US) molecular dynamics at different temperatures: 300, 200, 100, 50 and 25 K. In the case of PI-US, we used a different number of beads (P) depending on the temperature (respectively 8, 8, 16, 32, 32) in order to describe at our best the nuclear quantum effects. At 50 K we checked the convergence of the free energy with the number of beads P, as can be seen in Figure 5.3.

In the quantum case, we have defined the collective variable from the centroids of the atomic positions [58, 76]. Once the US simulations are done, the free energy is reconstructed using the WHAM algorithm. [63]

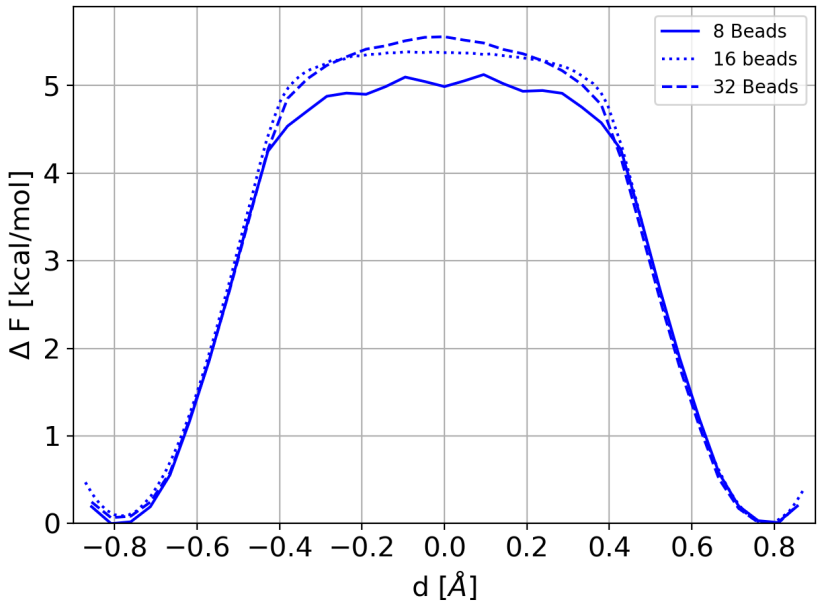

Figure 5.3: Free energy profile from PI-US simulations at 50 K as a function of the number of beads used.

Free energy calculations were performed using our own molecular dynamics code which performs path integral based molecular dynamics coupled with DFTB+ [42] software to obtain energies and gradients and with PLUMED [25] to perform umbrella sampling. The current version of our software is parallelized with OpenMP when using P beads such that the user time is almost identical for classical and PI-based simulations. Implementation details are available in Appendix C.

5.2.3 Rate Constant

The calculation of rate constants was discussed in Chapter 1, here we simply recall the main equations.

Free energy calculations can give access to the free energy barrier, $\Delta F^{\ddagger}$ of a reaction, which can be used to evaluate the reaction rate constant via the transition state theory (TST) simply by:

$$k^{TST}(T) = \frac{k_b T}{h} \exp\left(-\frac{\Delta F^{\ddagger}}{k_b T}\right) \tag{5.3}$$

where k_b is the Boltzmann constant, T the temperature and h is Planck's constant. However, there are a number of assumptions in TST and one of the most critical is that each trajectory reaching the transition state (TS) leads to the products side. However, it is

possible that a number of trajectories "recross" back to the reactants and the full rate can be written as:

$$k(T) = \kappa k^{TST}(T) \tag{5.4}$$

where κ is the so-called recrossing factor, which is temperature dependent and can be obtained from MD trajectories starting at the TS configuration [58, 203], as shown in Chapter 1:

$$\kappa(t) = \frac{\langle \delta(d_0)\dot{d}_0 h(d_t)\rangle}{\langle \delta(d_0)\dot{d}_0 h(\dot{d}_0)\rangle} \tag{5.5}$$

Practically, what is done is to generate a number of initial conditions at the transition state and then run post-TS trajectories in the microcanonical ensemble. From the analysis of those trajectories, we can obtain κ, where we define that a product is formed when d $>$ 0.0 Å otherwise if the system is in the reactant side, d $<$ 0.0 Å.

To sample the initial conditions we run, for each temperature and both for classical and RPMD based simulations, a long NVT trajectory with the CV constrained to be at the TS, i.e. d = 0.0 Å. From that, we select geometries and momenta each 200 fs, for a total number of snapshots between 200 and 500 (since for low temperatures we need a larger number of beads, the computational cost is higher). The microcanonical RPMD simulations were performed with the same number of beads as for the free energy calculations.

5.3 Results

5.3.1 Free energy profile

In Figure 5.4 we report the free energy profiles as a function of the reaction coordinate (defined in Eq. 5.1) for the different umbrella sampling simulations. From Figure 5.4, the trends with temperature can be observed. In the case of classical Langevin MD simulations, (left panel), the free energy barrier slightly decreases with increasing temperature. On the contrary, the PIMD results, provide the opposite trend: the free energy barrier increases with the temperature, while in the low-temperature regime the free energy curve becomes almost flat in the TS region (this is particularly visible at 25 K). This behaviour, is due to the fact that in PIMD simulations, the springs between the nearest beads of the same atom, become so loose at low temperatures that the ring polymer is spread in both regions of

space, reactant and product. This behaviour was already found by Roux and Hinsen [58], in the case of a proton transfer reaction, and it was qualitatively attributed to tunnelling in PIMD simulations.

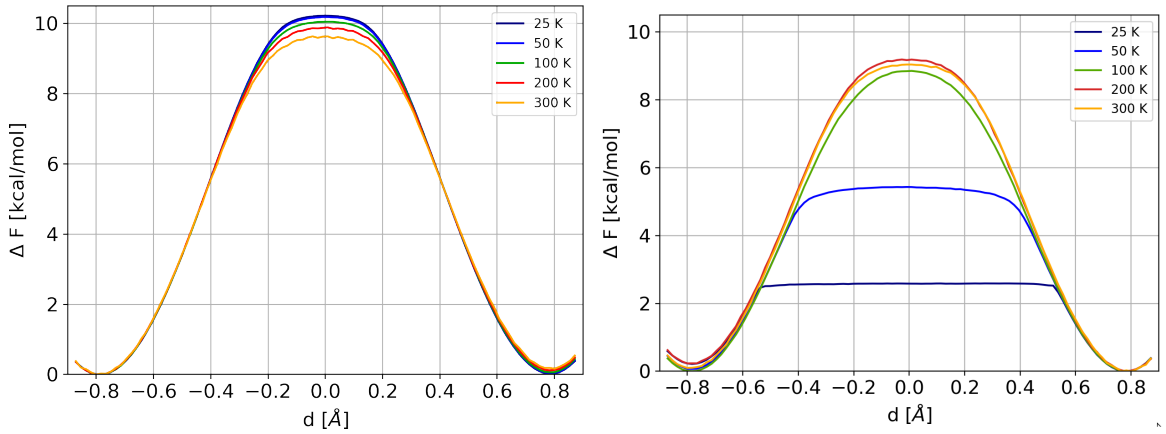

Figure 5.4: Free energy curves as function of d at different temperatures, as obtained from ClLMD-US (left panel) and PI-US (right panel) simulations.

In Table 5.2 we report the free energy barriers for both classical and PIMD simulations. Notably, the inclusion of NQEs decreases the free energy barriers, as was shown already by the plot in Figure 5.4: at 25 K the difference between the ClLMD free energy barrier and the PIMD one, is about 8 kcal/mol, while at the classical limit this difference is only 0.5 kcal/mol.

Table 5.2: Free energy barriers for ClLMD and PIMD umbrella sampling simulations in kcal/mol.

Temperature [K]	$\Delta F^{\ddagger}_{ClLMD}$	$\Delta F^{\ddagger}_{PIMD}$
25 K	10.22	2.59
50 K	10.18	5.43
100 K	10.05	8.85
200 K	9.89	9.18
300 K	9.63	9.04

One way to estimate the temperature below which tunnelling is expected to have a significant impact on the reaction rate, it is by calculating the crossover temperature [211, 212]:

$$T_c = \frac{\hbar\omega^{\ddagger}}{2\pi k_b} \tag{5.6}$$

where $\omega^{\ddagger}$ is the imaginary frequency (in absolute value) at the transition state. For semibullvalene, this value is around 400 cm^{-1} (see Table 5.1), giving a crossover tempera-

ture of about 90 K.

As NQEs, both Zero Point Energy and tunnelling can affect the free energy barriers, and together have as main effect the lowering of the free energy barrier, with respect to classical results. To disentangle the contribution of ZPE from that of tunnelling, it can be useful to calculate the difference, within the harmonic approximation (ΔH.A.), between the classical vibrational energy and the quantum vibrational energy:

$$\Delta\text{H.A.}(T) = \Delta E_{cl}^{vib}(T) - \Delta E_{q}^{vib}(T) \tag{5.7}$$

where ΔE_{cl}^{vib} and ΔE_{q}^{vib} are the vibrational energy differences between TS and reactant, calculated using classical and quantum statistics, respectively. Note that this Quantum-Classical difference probes how the ZPE modifies the barrier height. Analyzing the value of ΔH.A. at different temperatures we can provide an idea of the contribution of the Zero Point Energy of the system to the reaction rate. In Figure 5.5 we report the difference of ZPE (green line), the crossover temperature (red vertical line) and the difference $\Delta F^{\ddagger}_{ClLMD} - \Delta F^{\ddagger}_{PIMD}$ as obtained from umbrella sampling simulations. We can notice that for temperatures higher than T_c (T$>$ 90 K) the simulation results follow the harmonic approximation. However, when the temperature decreases, the difference between the free energy barrier obtained within classical and PIMD simulations becomes much higher than the harmonic approximation (which for T$\rightarrow$ 0 coincides with the difference in ZPE). This is an evidence that below T_c, tunnelling is taking place and it has an important role in the reaction.

In the next subsection we will see how the dynamical properties of such reaction are impacted by nuclear quantum effects.

5.3.2 Rate Constants

In Table 5.3 we report the rate constants obtained within the TST approximation, using Equation 5.3. As it can be noticed, thanks to the lower free energy barrier discussed previously, the rate with the inclusion of NQEs at low temperature (25 K) is more than 60 orders of magnitude lower than the classical result. As expected, at higher temperatures, the PIMD and classical results are closer and tend to be almost equal at room temperature (differing only by a factor of 3).

The basic assumption behind the TST is to neglect recrossing. As we said before (Chapter 1), the rate obtained within TST is an upper bound to the true rate constant. This is due to the fact that TST only takes into account the trajectories which, starting from the tran-

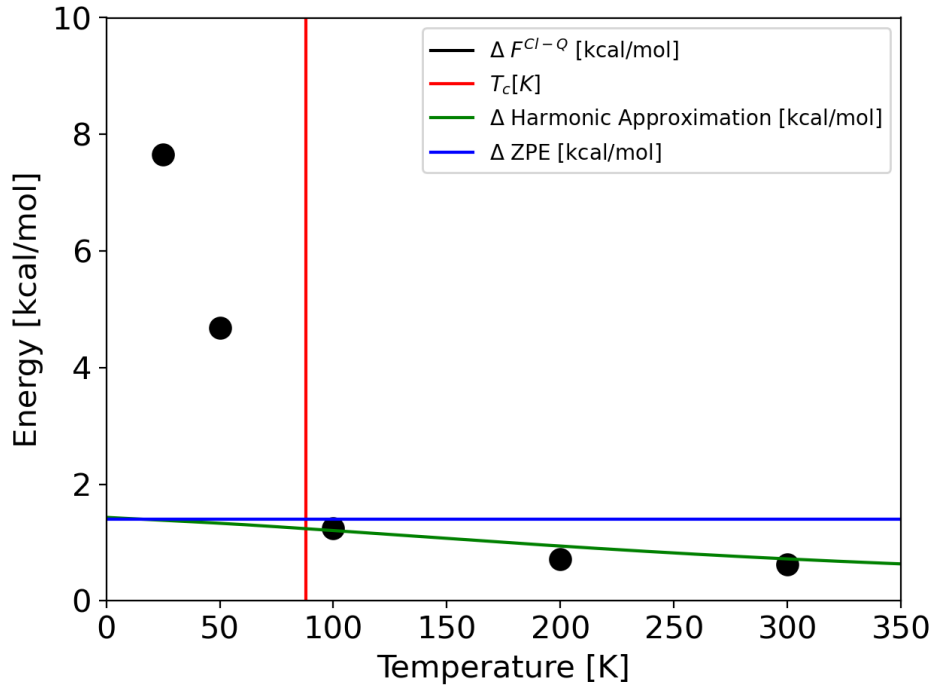

Figure 5.5: Difference between ClLMD and PIMD free energy barrier as a function of temperature. Black dots are simulation results, while the green curve is the value obtained from harmonic approximation (Equation 5.7). With the vertical line in red, we report the crossover temperature obtained from Equation 5.6.

Table 5.3: Rate constants obtained with Eyring equation 5.3 for ClLMD and PIMD umbrella sampling simulations in s^{-1}.

Temperature [K]	k^{TST}_{ClLMD}	k^{TST}_{PIMD}
25 K	2.09×10^{-78}	1.15×10^{-11}
50 K	3.13×10^{-33}	1.86×10^{-12}
100 K	2.19×10^{-10}	9.24×10^{-8}
200 K	6.40×10^{1}	3.82×10^{2}
300 K	5.98×10^{5}	1.61×10^{6}

sition state (at the dividing surface), have the momenta pointing towards the product side, and these trajectories are considered reactive. Instead, what can happen, is that a trajectory starts in the TS with the momentum pointing to the product side, but after a certain time goes back to the reactant side, due to the exchange of energy between the different modes. Also the opposite is true: a trajectory at the dividing surface could have the momentum projected in the reactant side (and so it will not be considered in the TST approximation) but, due to this flow of energy, it can end up in the product side. This approximation can be refined by running actual trajectories starting from the dividing surface to compute the recrossing factor.

The final rate constant is obtained by multiplying the TST rate constant by the recross-

ing factor in the limit of $t \to \infty$.

From simulations, the recrossing factors ($t \to \infty$) are obtained as a function of time and they are plotted in Figure 5.6, for different temperatures, for classical (left) and RPMD (right) results. For T = 100 K and above, the shapes are similar between RPMD and classical results, except for low temperatures. In particular, it can be seen from Table 5.4 where we report the values of the recrossing factors in the limit of $t \to \infty$, that in the classical case, this value approaches 1 with the decrease of the temperature. RPMD results instead, become different from classical ones at temperatures lower than T_c, where for 25 K we have the lowest value of the recrossing equal to 0.6. This has the same origin of the flat free energy profile shown before: the ring polymer of the atom involved is so spread, that it is delocalized on both sides, reactant and product. This means that the system is in both sides at the same time and that is the way in which RPMD theory represents tunnelling. Furthermore, as it can be seen from Figure 5.7 which shows the plot at 25 K for RPMD, the simulation time needed for all the trajectories to reach the product or reactant side, was 2.0 ps, while for temperatures higher than 25 K, a simulation time of 0.3 ps was sufficient.

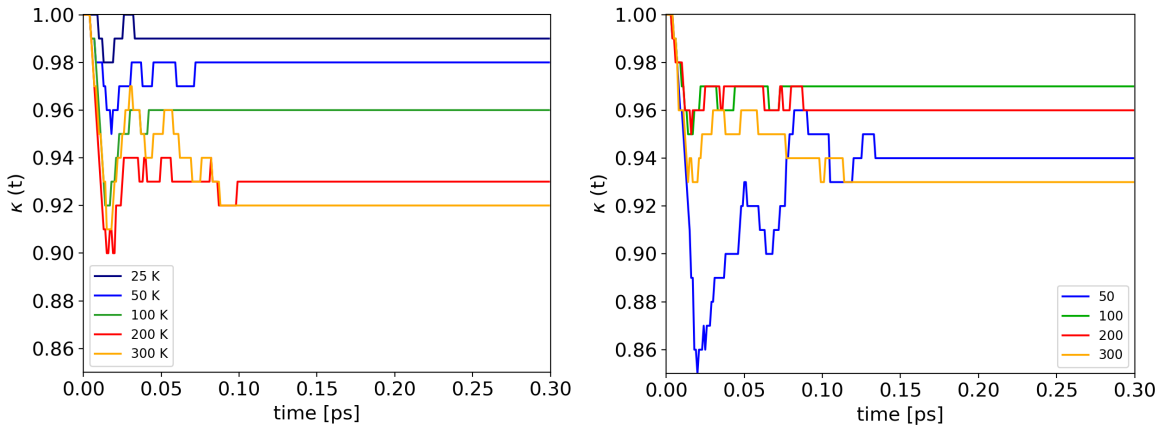

Figure 5.6: Recrossing factors as function of time at different temperature, classical results (left) and RPMD results (right)

Temperature [K]	$\kappa^{cl}_{t\to\infty}$	$\kappa^{RPMD}_{t\to\infty}$
25	0.99	0.63
50	0.98	0.94
100	0.96	0.97
200	0.93	0.96
300	0.92	0.93

Table 5.4: Factor recrossing for t$\to \infty$

The difference in the trend between the classical and RPMD results can be appreciated from the plot in Figure 5.8, where we report the value of $\kappa(t \to \infty)$ for RPMD and classical

simulations at different temperatures. It can be seen that in the classical case, the recrossing factor decreases as the temperature increases. The trend of RPMD is completely different: between 100-300 K the results are almost equal to the classical one. The difference with the classical results becomes evident only for temperature lower than the cross-over temperature when tunnelling is expected. This is not surprising, since only when we are in a regime of tunnelling, the polymer is delocalized between both sides, thus the centroid initial velocity has less importance and there is more chance that a polymer with initial velocity towards the products actually ends up in the reactants.

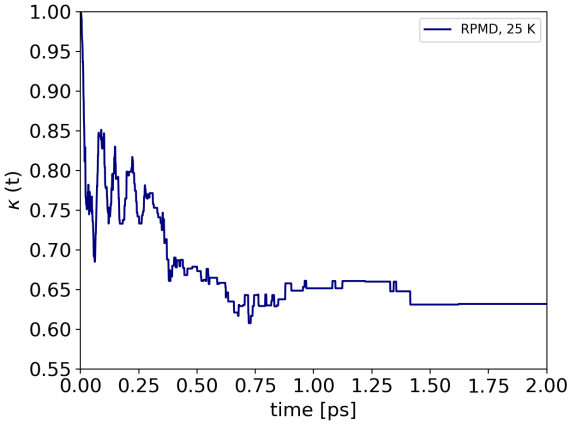

Figure 5.7: Recrossing factors as function of time for RPMD at 25 K.

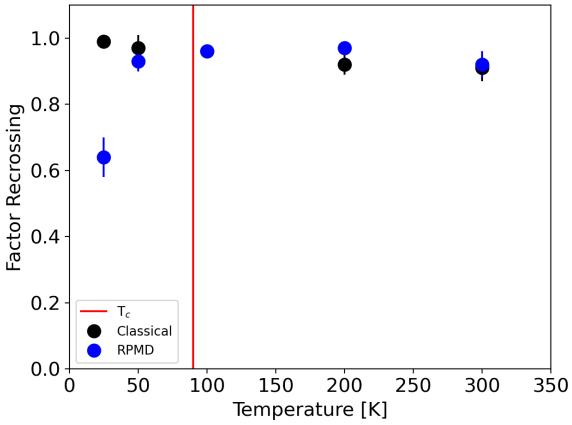

Figure 5.8: Recrossing factor for $t \to \infty$ for both RPMD (blue) and Classical (black) results. The red vertical line represents the cross-over temperature

5.4 Conclusions

Summarizing, in this Chapter we focus on how NQEs affect free energy profiles and rate constants for a model reaction which involves heavy atoms. In particular, we explore the accuracy of the transition state theory approximation, by correcting the rate obtained within this approach with the recrossing factor. Thanks to the harmonic approximation and the estimation of the cross-over temperature, we were able to identify the regime in which tunnelling plays a central role.

From the analysis of the free energy barrier and the comparison with the harmonic approximation used to probe the effect of ZPE, it is clear that, below the crossover temperature, the main effect on the free energy profile is due to the tunnelling. This effect is evident in particular at 25 K, where the path integral free energy profile is completely flat in the TS region. This leads to a huge increase of the rate constant, by 60 orders of magnitude compared to the classical one. This difference, as expected, disappears with increasing temperature.

Computing the recrossing factor shows that for temperatures greater than the crossover temperature the TST holds well for both classical and nuclear quantum results, since the value for $\kappa(t \to \infty)$ is between 0.94 and 0.99. This means that the effect of recrossing on the rate can be neglected. On the other side, at 25 K, when the tunnelling has a huge effect on the reaction, the TST becomes less accurate, since the recrossing factor was found to be around 0.6. This means that the inclusion of NQEs (and in particular tunnelling at low temperature) has two opposite effects: on one hand, it reduces the free energy barrier (which speeds up the reaction) and on the other hand, it decreases the recrossing factor by about a half (which slow the reaction down). This is still not comparable with the huge effect of tunnelling on the free energy barrier, which speeds up the reaction by 60 orders of magnitude in comparison with the classical results.

Part III

Conclusions and perspectives

Conclusions and perspectives

In this thesis, we focus on how tunnelling and zero point energy (ZPE) affect the dynamical and thermodynamical properties of reactive systems, within a direct dynamics approach. The methods used to include nuclear quantum effects (NQEs) in the simulations are the quantum thermal bath (QTB) and ring polymer molecular dynamics (RPMD). The first was developed for equilibrium properties and is able to include the zero point energy of the system through the use of generalized Langevin thermostat. The second is exact for equilibrium thermodynamical properties and can provide approximate reaction rate values that account for both ZPE and tunnelling.

The systems studied are:

1. An analytical model for the unimolecular fragmentation of CH_4 dissociation that was tuned to explore different temperature ranges and on which we compared the two methods (QTB and RPMD).
2. The double proton transfer in DNA base pairs, where we focused on how the double proton transfer mechanism is affected by ZPE and by the biochemical environment.
3. The Cope rearrangement of semibullvalene, to explore how tunnelling can speed up the reaction at low temperatures even if the atoms involved are not considered "light".

In Chapter 3, we investigate how the ZPE of the system affects an unimolecular fragmentation. To obtain the rate we used a direct approach: an ensemble of reactive trajectories is performed, and we collect the time at which each of these trajectories reacts. Results show that RPMD provides a good approximation of the impact of ZPE on rate constants, and in particular it captures very accurately the quantum-classical difference in the activation energy. Instead, the QTB largely overestimates the rate constants in all systems studied. The reason for this incorrect behaviour, is to search in the distance probability distribution: the QTB describes poorly the tail of this distribution which is fundamental to describe the fragmentation correctly. We also perform a comparison with TST theory, with which RPMD is in almost perfect agreement. This study shows how RPMD can be used to obtain rate constants for unimolecular fragmentations, especially when TST is difficult to apply, like when the reaction pathways are not well known, when the system is highly anharmonic or it has a large number of degrees of freedom.

In Chapter 4 we study the Double Proton Transfer (DPT) in the Guanine-Cytosine (GC) DNA base pairs. In particular, we focus on how NQEs and the biochemical environment af-

fect this reaction. The results show that for the isolated dimer the mechanism is concerted, and the effect of NQEs (mainly due to ZPE) is to speed up the reaction by a factor of 30 in comparison with classical results. In contrast, when the environment is included the mechanism becomes step-wise, NQEs still speed up the reaction (even if slightly less than for the isolated dimer) and the free energy landscape is strongly affected as the intermediate state ceases to be a local minimum, making the proton transfer almost barrierless. These results underline the importance of both including NQEs and fully accounting for the complexity of the environment, especially in biological systems.

In Chapter 5 we focus on the study of Heavy Atoms tunnelling for a prototypical reaction, the Cope rearrangement of semibullvalene. In literature, this kind of reaction was studied both experimentally and theoretically, using different semiclassical approaches [15, 201]. We performed a RPMD study and found that when the temperature is lower than the cross-over temperature, the free energy profile in the TS region becomes completely flat, showing a typical marker for tunnelling. The inclusion of tunnelling speeds up the reaction up to 60 orders of magnitude at 25 K, and the quantum-classical difference in the free energy barriers disappears in the classical (high temperature) limit, as we expected. The computation of the recrossing factor shows that it is relevant only at the lowest temperature (25 K), where this value is about 0.6. It should be noted that even if the rate constant is lowered by the inclusion of the recrossing factor in this low temperature regime, is not comparable with the impact of tunnelling which, as we said before, speeds up the reaction by 60 orders of magnitude.

In conclusion, we extensively investigated several crucial aspects. In the first instance, we employed an analytical model for studying an unimolecular fragmentation, which yield three significant findings. First, the utilization of RPMD (ring polymer molecular dynamics) proved effective in obtaining accurate dynamical properties. Second, the observation of the equivalent temperature of the stretching zero point energy (ZPE), particularly in the context of fragmentation, emerged as a key point for predicting when NQEs have a significant impact. Lastly, we showed the inadequacy of the QTB method to describe unimolecular fragmentation. In particular, the representation of the tail of the distance distribution was found to be a key issue.

Furthermore, we focused on two distinct reactive systems. We first showed the importance of including the environment when investigating reaction mechanisms, especially for biological systems. Subsequently, we analyzed a reaction in which heavy atoms are involved, and we observed that in a deep tunnelling regime, the tunnelling can lead to dual effects on the rate constant. On one side, this NQE speeds up the reaction due to its influence on the free energy barrier, which is lowered down. On the other side, it slows down the

reaction due to the delocalization of the system that increases the probability of recrossing of the dividing surface.

Finally, during these three years we developed a software to run such calculations which is now a user friendly resource, which can be used by future students in the group as well as researchers from other groups (the code will be soon be made available on github). This code allows the study of molecular systems within the RPMD approach, and DFTB electronic structure method, alongside enhanced sampling techniques. The simplicity of the software, enables easy modifications, like the incorporation of different subroutines which can describe other analytical models, or adding the interaction with an optical cavity in order to study strong vibrational coupling phenomena, as presently done in the group. This code, written in fortran90 and parallelized with OpenMP, will be shortly published with these new features.

Bibliography

(1) Fermi, E.; Pasta, P.; Ulam, S.; Tsingou, M. Studies of Nonlinear Problems. **1955**.

(2) McCommon, J.; Gelin, B.; Karplus, M. Dynamics of folded proteins. *Nature* **1977**, *267*, 585–590.

(3) Hospital, A.; Goni, J.; Gelpi, J Molecular dynamics simulations: advances and applications. *Adv. Appl. Bioinform. Chem.* **2015**, *8*, 37–47.

(4) Hollingsworth, S.; Dror, R. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99*, 1129–1143.

(5) Born, M.; Oppenheimer, R. Zur Quantentheorie der Molekeln. *Ann. Phys.* **1927**, *389*, 457–484.

(6) Markland, T.; Ceriotti, M. Nuclear quantum effects enter the mainstream. *Nature* **2018**, *2*.

(7) Pereyaslavets, L.; Kurnikov, I.; Kamath, G.; Butin, O.; Illarionov, A.; Leontyev, I.; Olevanov, M.; Levitt, M.; Kornberg, D.; Fain, B. On the importance of accounting for nuclear quantum effects in ab initio calibrated force fields in biological simulations. *Proc. Natl. Acad. Sci.* **2018**, 8878–8882.

(8) Nilsson, A.; Pettersson, M. The structural origin of anomalous properties of liquid water. *Nat. Commun.* **2015**, *6*, 8998.

(9) Ceriotti, M.; Fang, W.; Kusalik, P.; McKenzie, R.; Michaelides, A.; Morales, M. A.; Markland, T. Nuclear quantum effects in water and aqueous systems: experiment, theory, and current challenges. *Chem. Rev.* **2016**, *13*, 7529–7550.

(10) Ivanov, S. D.; Grant, I. M.; Marx, D. Quantum free energy landscapes from ab initio path integral metadynamics: Double proton transfer in the formic acid dimer is concerted but not correlated. *J. Comput. Phys.* **2015**, *143*, 124304.

(11) Fang, W.; Chen, J.; Feng, Y.; Li, X.; Michaelides, A. The quantum nature of hydrogen. *Int. Rev. Phys. Chem.* **2019**, *38*, 35–61.

(12) Litman, Y.; Richardson, J. O.; Kumagai, T.; Rossi, M. Elucidating the Nuclear Quantum Dynamics of Intramolecular Double Hydrogen Transfer in Porphycene. *J. Am. Chem. Soc.* **2019**, *141*, 2526–2534.

(13) Tuckerman, M. E.; Marx, D.; Klein, M.; Parrinello, M. On the Quantum Nature of the Shared Proton in Hydrogen Bonds. *Science* **1997**, *275*, 817–820.

(14) Beck, M.; Jackle, A.; Worth, G.; Meyer, H. The multiconfiguration time-dependent hartree (MCTDH) method: a highly efficient algorithm for propagating wavepackets. *Phys. Rep.* **2000**, *324*, 1–105.

(15) Skodje, R. T.; Truhlar, D. G.; Garrett, B. C. A general small-curvature approximation for transition-state-theory transmission coefficients. *J. Phys. Chem.* **1981**, *85*, 3019–3023.

(16) Miller, W. H. Semiclassical limit of quantum mechanical transition state theory for nonseparable systems. *J. Phys. Chem.* **2008**, *62*, 1899–1906.

(17) Bowman, J. M.; Gazdy, B.; Sun, Q. A method to constrain vibrational energy in quasiclassical trajectory calculations. *J. Comput. Phys.* **1989**, *91*, 2859–2862.

(18) Bonnet, L.; Rayez, J.-C. Quasiclassical trajectory method for molecular scattering processes: necessity of a weighted binning approach. *Chem. Phys. Lett.* **1997**, *227*, 183–190.

(19) Bonnet, L.; Rayez, J.-C. Gaussian weighting in the quasiclassical trajectory method. *Chem. Phys. Lett.* **2004**, *397*, 106–109.

(20) Miller, W. The semiclassical initial value representation: a potentially practical way for adding quantum effects to classical molecular dynamics simulations. *J. Phys. Chem. A* **2001**, *105*, 2942–2955.

(21) Craig, I. R.; Manolopoulos, D. E. Quantum statistics and classical mechanics: Real time correlation functions from ring polymer molecular dynamics. *J. Comput. Phys.* **2004**, *121*, 3368–3373.

(22) Dammak, H.; Chalopin, Y.; Laroche, M.; Hayoun, M.; Greffet, J.-J. Quantum Thermal Bath for Molecular Dynamics Simulation. *Phys. Rev. Lett.* **2009**, *103*, 190601.

(23) Hase, W. L.; Duchovic, R. J.; Hu, X.; Komornicki, A.; Lim, K. F.; Lu, D.-H.; Peslherbe, G. H.; Swamy, K. N.; Linde, S. R. V.; Varandas, A.; Wang, H.; Wolf, R. J. VENUS. A General Chemical Dynamics Computer Program. *QCPE Bull.* **1996**, *16*, 671.

(24) Hourahine, B. et al. DFTB+, a software package for efficient approximate density functional theory based atomistic simulations. *J. Comput. Phys.* **2020**, *152*, 124101.

(25) Tribello, G.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comp. Phys. Commun.* **2014**, *185*, 604–613.

(26) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Comput. Phys.* **1984**, *81*, 3684–3690.

(27) Evans, D. J.; Holian, B. L. The Nosé–Hoover thermostat. *J. Comput. Phys.* **1985**, *83*, 4069–4074.

(28) Ford, G. W.; Kac, M.; Mazur, P. Statistical Mechanics of Assemblies of Coupled Oscillators. *J. Math. Phys.* **1965**, *6*, 504–515.

(29) Posch, H. A.; Hoover, W. G.; Vesely, F. J. Canonical dynamics of the Nosé oscillator: Stability, order, and chaos. *Phys. Rev. A* **1986**, *33*, 4253–4265.

(30) Bond, S. D.; Benedict, J. L.; Brian, B. L. The Nosé-Poincaré Method for constant Temperature Molecular Dynamics. *J. Comput. Phys.* **1999**, 114–134.

(31) Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nosé-Hoover chains: The canonical ensemble via continuous dynamics. *J. Comput. Phys.* **1992**, *97*, 2635–2643.

(32) Langevin, P. Sur la théorie du mouvement brownien. *C. R. Acad. Sci.* **1908**, *146*, 530–533.

(33) Zwanzig, R. Nonlinear generalized Langevin equations. *J. Stat. Phys.* **1973**, *9*, 215–220.

(34) Kubo, R. The fluctuation-dissipation theorem. *Rep. Prog. Phys.* **1966**, *29*, 255.

(35) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864–B871.

(36) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.

(37) Stewart, J. J. P. Optimization of parameters for semiempirical methods I. Method. *J. Comput. Chem.* **1989**, *10*, 209–220.

(38) Stewart, J. J. P. Optimization of parameters for semiempirical methods II. Applications. *J. Comput. Chem.* **1989**, *10*, 221–264.

(39) Stewart, J. J. P. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **2007**, *13*, 1173–1213.

(40) Andersen, O. K.; Jepsen, O. Explicit, First-Principles Tight-Binding Theory. *Phys. Rev. Lett.* **1984**, *53*, 2571–2574.

(41) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.

(42) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-consistent-charge density-functional-tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **1998**, *58*, 7260–7268.

(43) Koskinen, P.; Mäkinen, V. Density-functional tight-binding for beginners. *Comput. Mater. Sci.* **2009**, *47*, 237–253.

(44) Kaminski, S.; Giese, T. J.; Gaus, M.; York, D. M.; Elstner, M. Extended Polarization in Third-Order SCC-DFTB from Chemical-Potential Equalization. *J. Phys. Chem. A* **2012**, *116*, 9131–9141.

(45) Gaus, M.; Jin, H.; Demapan, D.; Christensen, A. S.; Goyal, P.; Elstner, M.; Cui, Q. DFTB3 Parametrization for Copper: The Importance of Orbital Angular Momentum Dependence of Hubbard Parameters. *J. Chem. Theory Comput.* **2015**, *11*, 4205–4219.

(46) Simon, A.; Rapacioli, M.; Lanza, M.; Joalland, B.; Spiegelman, F. Molecular dynamics simulations on [FePAH]+ π-complexes of astrophysical interest: anharmonic infrared spectroscopy. *Phys. Chem. Chem. Phys.* **2011**, *13*, 3359–3374.

(47) Zheng, G.; Witek, H. A.; Bobadova-Parvanova, P.; Irle, S.; Musaev, D. G.; Prabhakar, R.; Morokuma, K.; Lundberg, M.; Elstner, M.; Köhler, C.; Frauenheim, T. Parameter Calibration of Transition-Metal Elements for the Spin-Polarized Self-Consistent-Charge Density-Functional Tight-Binding (DFTB) Method: Sc, Ti, Fe, Co, and Ni. *J. Chem. Theory Comput.* **2007**, *3*, 1349–1367.

(48) Elstner, M.; Frauenheim, T.; Kaxiras, E.; Seifert, G.; Suhai, S. A Self-Consistent Charge Density-Functional Based Tight-Binding Scheme for Large Biomolecules. *Phys. Status Solidi B* **2000**, *217*, 357–376.

(49) Riahi, S.; Eyonallahi, S.; Ganjali, M. R. Interaction of emodin with DNA bases: a density functional theory. *J. Theor. Comput. Chem.* **2010**, *09*, 875–888.

(50) Riahi, S.; Eynollahi, S.; Ganjali, M. R. Computational Studies on Effects of MDMA as an Anticancer Drug on DNA. *Chem. Biol. Drug Des.* **2010**, *76*, 425–432.

(51) Zheng, G.; Irle, S.; Morokuma, K. Performance of the DFTB method in comparison to DFT and semiempirical methods for geometries and energies of C20–C86 fullerene isomers. *Chem. Phys. Lett.* **2005**, *412*, 210–216.

(52) Witek, H. A.; Irle, S.; Zheng, G.; de Jong, W. A.; Morokuma, K. Modeling carbon nanostructures with the self-consistent charge density-functional tight-binding method: Vibrational spectra and electronic structure of C28, C60, and C70. *J. Comput. Phys.* **2006**, *125*.

(53) Małolepsza, E.; Lee, Y.-P.; Witek, H. A.; Irle, S.; Lin, C.-F.; Hsieh, H.-M. Comparison of geometric, electronic, and vibrational properties for all pentagon/hexagon-bearing isomers of fullerenes C38, C40, and C42. *Int. J. Quantum Chem.* **2009**, *109*, 1999–2011.

(54) Miller, W. H.; Schwartz, S. D.; Tromp, J. W. Quantum mechanical rate constants for bimolecular reactions. *J. Comput. Phys.* **1983**, *79*, 4889–4898.

(55) Yamamoto, T.; Wang, H.; Miller, W. H. Combining semiclassical time evolution and quantum Boltzmann operator to evaluate reactive flux correlation function for thermal rate constants of complex systems. *J. Comput. Phys.* **2002**, *116*, 7335–7349.

(56) Shao, J.; Makri, N. Forward–backward semiclassical dynamics in the interaction representation. *J. Comput. Phys.* **2000**, *113*, 3681–3685.

(57) Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J. Current Status of Transition-State Theory. *J. Phys. Chem.* **1996**, *100*, 12771–12800.

(58) Hinsen, K.; Roux, B. Potential of mean force and reaction rates for proton transfer in acetylacetone. *J. Comput. Phys.* **1997**, *106*, 3567–3577.

(59) Laio, A.; Parrinello, M. Escaping free-energy minima. *PNAS* **2002**, *99*, 12562–12566.

(60) Laio, A.; Gervasio, F. L. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.* **2008**, *71*, 126601.

(61) Torrie, G.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.

(62) Bussi, G.; Laio, A.; Parrinello, M. Equilibrium Free Energies from Nonequilibrium Metadynamics. *Phys. Rev. Lett.* **2006**, *96*, 090601.

(63) Hub, J. S.; de Groot, B. L.; Van der Spoel, D. g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory Comput.* **2010**, *6*, 3713–3720.

(64) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

(65) Roux, B. The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.* **1995**, *91*, 275–282.

(66) Feynman, R. Space-Time Approach to Non-Relativistic Quantum Mechanics. *Rev. Mod. Phys.* **1948**, *20*, 367–387.

(67) Tuckerman, M., *Statistical Mechanics: Theory And Molecular Simulation*, 2001.

(68) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Comput. Phys.* **1953**, *21*, 1087–1092.

(69) Marx, D.; Parrinello, M. Ab initio path integral molecular dynamics: Basic ideas. *J. Comput. Phys.* **1996**, *104*, 4077–4082.

(70) Thirumalai, D.; Hall, R. W.; Berne, B. J. A path integral Monte Carlo study of liquid neon and the quantum effective pair potential. *J. Comput. Phys.* **1984**, *81*, 2523–2527.

(71) Ceperley, D. M. Path integrals in the theory of condensed helium. *Rev. Mod. Phys.* **1995**, *67*, 279–355.

(72) Craig, I. R.; Manolopoulos, D. E. Chemical reaction rates from ring polymer molecular dynamics. *J. Comput. Phys.* **2005**, *122*, 084–106.

(73) Pérez de Tudela, R.; Aoiz, F. J.; Suleimanov, Y. V.; Manolopoulos, D. E. Chemical Reaction Rates from Ring Polymer Molecular Dynamics: Zero Point Energy Conservation in Mu + H2 $\rightarrow$ MuH + H. *J. Phys. Chem. Lett.* **2012**, *3*, 493–497.

(74) Liu, Q.; Zhang, L.; Li, Y.; Jiang, B. Ring Polymer Molecular Dynamics in Gas–Surface Reactions: Inclusion of Quantum Effects Made Simple. *J. Phys. Chem. Lett.* **2019**, *10*, 7475–7481.

(75) Ceriotti, M.; Parrinello, M.; Markland, T. E.; Manolopoulos, D. E. Efficient stochastic thermostatting of path integral molecular dynamics. *J. Comput. Phys.* **2010**, *133*, 124104.

(76) Craig, I. R.; Manolopoulos, D. E. A refined ring polymer molecular dynamics theory of chemical reaction rates. *J. Comput. Phys.* **2005**, *123*, 034102.

(77) Ceriotti, M.; Bussi, G.; Parrinello, M. Nuclear Quantum Effects in Solids Using a Colored-Noise Thermostat. *Phys. Rev. Lett.* **2009**, *103*, 030603.

(78) Brieuc, F.; Bronstein, Y.; Dammak, H.; Depondt, P.; Finocchi, F.; Hayoun, M. Zero-Point Energy Leakage in Quantum Thermal Bath Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2016**, *12*, 5688–5697.

(79) Barrat, J.; Rodney, D. Portable Implementation of a Quantum Thermal Bath for Molecular Dynamics Simulations. *J. Stat. Phys.* **2011**, *144*, 679–689.

(80) Basire, M.; Borgis, D.; Vuilleumier, R. Computing Wigner distributions and time correlation functions using the quantum thermal bath method: application to proton transfer spectroscopy. *Phys. Chem. Chem. Phys.* **2013**, *15*, 12591–12601.

(81) Mangaud, E.; Huppert, S.; Plé, T.; Depondt, P.; Bonella, S.; Finocchi, F. The Fluctuation–Dissipation Theorem as a Diagnosis and Cure for Zero-Point Energy Leakage in Quantum Thermal Bath Simulations. *J. Chem. Theory Comput.* **2019**, *15*, 2863–2880.

(82) Baer, T.; Hase, W., *Unimolecular reaction dynamics: theory and experiments*; Oxford University Press, New York: 1996.

(83) de Hoffmann, E.; Stroobant, V., *Mass Spectrometry. Principles and Applications*; Wiley et al.: 2007.

(84) Gross, J. H., *Mass Spectrometry. A Textbook*; Springer: 2004.

(85) Papayannopoulos, I. The Interpretation of Collision-Induced Dissociation Tandem Mass Spectra of Peptides. *Mass Spectrom. Rev.* **1995**, *14*, 49–73.

(86) Zaia, J. Mass Spectrometry of Oligosaccharides. *Mass Spectrom. Rev.* **2004**, *23*, 161–227.

(87) Paizs, B.; Suhai, S. Fragmentation Pathways of Protonated Peptides. *Mass Spectrom. Rev.* **2005**, *24*, 508–548.

(88) Laskin, J.; Futrell, J. H. Collisional Activation of Peptide Ions in FT-ICR Mass Spectrometry. *Mass Spectrom. Rev.* **2003**, *22*, 158–181.

(89) Rodgers, M. T.; Armentrout, P. B. Noncovalent metal-ligand bond energies as studied by threshold collision-induced dissociation. *Mass Spectrom. Rev.* **2000**, *19*, 215–247.

(90) Armentrout, P. Reactions and Thermochemistry of Small Transition Metal Cluster Ions. *Ann. Rev. Phys. Chem.* **2001**, *52*, 423–461.

(91) Armentrout, P. B. Guided Ion Beam Studies of Transition Metal-Ligand Thermochemistry. *Int. J. Mass Spectrom.* **2003**, *227*, 289–302.

(92) Merkel, A.; Havlas, Z.; Zahradnik, R. Evaluation of the Rate Constant for the SN2 Reaction $CH_3F + H^- \rightarrow CH_4 + F^-$. *J. Am. Chem. Soc.* **2003**, *110*, 8355–8359.

(93) Manikandan, P.; Zhang, J.; Hase, W. L. Chemical Dynamics Simulations of X^- + $CH_3Y \rightarrow XCH_3$ + Y^- Gas-Phase SN2 Nucleophilic Substitution Reactions. Nonstatistical Dynamics and Nontraditional Reaction Mechanisms. *J. Phys. Chem. A* **2012**, *116*, 3061–3080.

(94) Martin-Somer, A.; Macaluso, V.; Barnes, G. L.; Yang, L.; Pratihar, S.; Song, K.; Hase, W. L.; Spezia, R. Role of Chemical Dynamics Simulations in Mass Spectrometry Studies of Collision-Induced Dissociation and Collisions of Biological Ions with Organic Surfaces. *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 2–24.

(95) Yang, L.; Sun, R.; Hase, W. L. Use of Direct Dynamics Simulations to Determine Unimolecular Reaction Paths and Arrhenius Parameters for Large Molecules. *J. Chem. Theory Comput.* **2011**, *7*, 3478–3483.

(96) Homayoon, Z.; Pratihar, S.; Dratz, E.; Snider, R.; Spezia, R.; Barnes, G.; Macaluso, V.; Martin-Somer, A.; Hase, W. L. Model Simulations of the Thermal Dissociation of the TIK(H+)2 tripeptide. Mechanisms and Kinetic Parameters. *J. Phys. Chem. A* **2016**, *120*, 8211–8227.

(97) Spezia, R.; Martin-Somer, A.; Macaluso, V.; Homayoon, Z.; Pratihar, S.; Hase, W. L. Unimolecular Dissociation of Peptides: Statistical vs Non-Statistical Fragmentation Mechanisms and Time Scales. *Faraday Discuss.* **2016**, *195*, 599–618.

(98) Miller, W. H.; Hase, W. L.; Darling, C. L. A simple model for correcting the zero point energy problem in classical trajectory simulations of polyatomic molecules. *J. Comput. Phys.* **1989**, *91*, 2863.

(99) Paul, A.; Hase, W. L. Zero-Point Energy Constraint for Unimolecular Dissociation Reactions. Giving Trajectories Multiple Chances To Dissociate Correctly. *J. Phys. Chem. A* **2016**, *120*, 372–378.

(100) Spezia, R.; Dammak, H. On the Use of Quantum Thermal Bath in Unimolecular Fragmentation Simulation. *J. Phys. Chem. A* **2019**, *123*, 8542–8551.

(101) Mauger, N.; Plé, T.; Lagardéfre, L.; Bonella, S.; Mangaud, E.; Piquemal, J.; Huppert, S Nuclear Quantum Effects in Liquid Water at Near Classical Computational Cost Using the Adaptive Quantum Thermal Bath. *J. Phys. Chem. Lett.* **2021**, *12*, 8285–8291.

(102) Feynman, B. J.; Hibbs, A. R., *Quantum mechanics and path integrals*; McGraw-Hill, New York: 1965.

(103) Berne, B. J.; Thirumalai, D. On the Simulation of Quantum Systems: Path Integral Methods. *Ann. Rev. Phys. Chem.* **1986**, *37*, 401–424.

(104) Cao, J.; Voth, G. A. The formulation of quantum statistical mechanics based on the Feynman path centroid density. I. Equilibrium properties. *J. Comput. Phys.* **1994**, *100*, 5093.

(105) Cao, J.; Voth, G. The formulation of quantum statistical mechanics based on the Feynman path centroid density. II. Dynamical properties. *J. Comput. Phys.* **1994**, *100*, 5106–5117.

(106) Geva, E.; Shi, Q.; Voth, G. A. Quantum-mechanical reaction rate constants from centroid molecular dynamics simulations. *J. Comput. Phys.* **2001**, *115*, 9209–9222.

(107) Collepardo-Guevara, R.; Suleimanov, Y. V.; Manolopoulos, D. E. Bimolecular reaction rates from ring polymer molecular dynamics. *J. Comput. Phys.* **2009**, *130*, 174713.

(108) Perez de Tudela, R.; Aoiz, F. J.; Suleimanov, Y. V.; Manolopoulos, D. E. Chemical Reaction Rates from Ring Polymer Molecular Dynamics: Zero point energy conservation in Mu + H_2 to MuH + H. *J. Phys. Chem. Lett.* **2012**, *3*, 493–497.

(109) Li, Y.; Suleimanov, Y. V.; Yang, M.; Green, W. H.; Guo, H. Ring Polymer Molecular Dynamics Calculations of Thermal Rate Constants for the $O(^3P) + CH_4 \rightarrow OH + CH_3$ Reaction: Contributions of Quantum Effects. *J. Phys. Chem. Lett.* **2013**, *4*, 48–52.

(110) Suleimanov, Y. V.; Aoiz, F. J.; Guo, H. Chemical Reaction Rate Coefficients from Ring Polymer Molecular Dynamics: Theory and Practical Applications. *J. Phys. Chem. A* **2016**, *120*, 8488–8502.

(111) Suleimanov, Y.; Kong, W. J.; Guo, H.; Green, W. H. Ring-polymer molecular dynamics: Rate coefficient calculations for energetically symmetric (near thermoneutral) insertion reactions (X + H2) $\rightarrow$ HX + H(X = C(1D), S(1D)). *J. Comput. Phys.* **2014**, *141*, 244103.

(112) Navrotskaya, I.; Shi, Q.; Geva, E. Quantum-mechanical reaction rate constants from centroid molecular dynamics simulations: Barrier crossing in an asymmetrical double-well. *Isr. J. Chem.* **2002**, *42*, 225–236.

(113) Suleimanov, Y. V.; Allen, J. W.; Green, W. H. RPMDrate: Bimolecular chemical reaction rates from ring polymer molecular dynamics. *Comp. Phys. Comm.* **2013**, *184*, 833–840.

(114) Wang, H.; Sun, X.; Miller, W. H Semiclassical approximations for the calculation of thermal rate constants for chemical reactions in complex molecular systems. *J. Comput. Phys.* **1998**, *108*, 9726–9736.

(115) Liu, J.; Miller, W. H. A simple model for the treatment of imaginary frequencies in chemical reaction rates and molecular liquids. *J. Comput. Phys.* **2009**, *131*, 074113.

(116) Bose, A.; Makri, N. Wigner Distribution by Adiabatic Switching in Normal Mode or Cartesian Coordinates and Molecular Applications. *J. Chem. Theory Comput.* **2018**, *14*, 5446–5458.

(117) Banares, L.; Aoiz, F.; Honvault, P.; Bussery-Honvault, B.; Launay, J.-M. Quantum mechanical and quasi-classical trajectory study of the C(1D) + H2 reaction dynamics. *J. Comput. Phys.* **2003**, *118*, 565–568.

(118) Gonzalez-Martinez, M. L.; Rubayo-Soneira, J.; Janda, K. Quasi-classical trajectories study of Ne79Br2(B) vibrational predissociation. *Phys. Chem. Chem. Phys.* **2006**, *8*, 4550–4558.

(119) Czako, G.; Wang, Y.; Bowman, J. M. Quasiclassical trajectory calculations of correlated product-state distributions for the dissociation of (H2O)2 and (D2O)2. *J. Comput. Phys.* **2011**, *135*, 151102.

(120) Miller, W. H. Quantum mechanical transition state theory and a new semiclassical model for reaction rate constants. *J. Comput. Phys.* **1974**, *61*, 1823–1834.

(121) Cobos, C.; Troe, J. The Dissociation-Recombination System CH4 + M −> CH3 + H + M: Reevaluated Experiments from 300 to 3000 K. *Z. Phys. Chem. (N. F.)* **1990**, *167*, 129–149.

(122) Cobos, C. J.; Troe, J. The Dissociation-Recombination System CH4+ M −> CH3 + H + M: II. Evaluation of Experiments up to 5000 K and Temperature Dependence of <E>. *Z. Phys. Chem. (N. F.)* **1992**, *176*, 161–171.

(123) Kiefer, J.; Kumaran, S. Rate of CH4 Dissociation over 2800-4300 K: The Low-Pressure-Limit Rate Constant. *J. Phys. Chem.* **1993**, *97*, 414–420.

(124) Tsang, W.; Hampson, R. Chemical Kinetic Data Base for Combustion Chemistry. Part I. Methane and Related Compounds. *J. Phys. Chem. Ref. Data* **1986**, *15*, 1087–1279.

(125) Sutherland, J.; Su, M.; Michael, J. Rate Constants for H + CH4, CH3 + H2, and CH4 Dissociation at High Temperature. *Int. J. Chem. Kinet.* **1986**, *15*, 1087–1279.

(126) Duchovic, R. J.; Hase, W. L.; Schlegel, H. B. Analytic Function for the H + CH_3 = CH_4 Potential Energy Surface. *J. Phys. Chem.* **1984**, *88*, 1339.

(127) Hase, W.; Mondro, S.; Duchovic, R.; Hirst, D. Thermal Rate Constant for H + CH_3→CH_4 Recombination. 3. Comparison of Experiment and Canonical Variational Transition State Theory. *J. Am. Chem. Soc.* **1987**, *109*, 2916–2922.

(128) Hu, X.; Hase, W. Modification of the DHS Analytic Potential Energy Function for H + CH_3→CH_4. Comparison of CVTST, Trajectory and Experimental Association Rate Constants. *J. Comput. Phys.* **1991**, *95*, 8073–8082.

(129) Lourderaj, U.; McAfee, J. L.; Hase, W. L. Potential Energy Surface and Unimolecular Dynamics of Stretched n-Butane. *J. Comput. Phys.* **2008**, *129*, 094701.

(130) Kolakkandy, S.; Paul, A.; Pratihar, S.; Kohale, S.; Barnes, G.; Wang, H.; Hase, W. Energy and Temperature Dependent Dissociation of the $Na^+(benzene)_{1,2}$ Clusters. Importance of Anharmonicity. *J. Comput. Phys.* **2015**, *142*, 044306.

(131) Ma, X.; Paul, A.; Hase, W. Chemical Dynamics Simulations of Benzene Dimer Dissociation. *J. Phys. Chem. A* **2015**, *119*, 6631–6640.

(132) Malik, A.; Lin, Y.-F.; Pratihar, S.; Angel, L.; Hase, W. L. Direct Dynamics Simulations of Fragmentation of a Zn(II)-2Cys-2His Oligopeptide. Comparison with Mass Spectrometry Collision-Induced Dissociation. *J. Phys. Chem. A* **2019**, *123*, 6868–6885.

(133) Malik, A.; Spezia, R.; Hase, W. L. Unimolecular Fragmentation Properties of Thermometer Ions from Chemical Dynamics Simulations. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 169–179.

(134) Perez-Mellor, A. F.; Spezia, R. Determination of Kinetic Properties in Unimolecular Dissociation of Complex Systems from Graph-Theory Based Analysis of an Ensemble of Reactive Trajectories. *J. Comput. Phys.* **2021**, *155*, 124103.

(135) Song, K.; Hase, W. Fitting classical microcanonical unimolecular rate constants to a modified RRK expression: Anharmonic and variational effects. *J. Comput. Phys.* **1999**, *110*, 6198–6207.

(136) Hu, X.; Hase, W.; Pirraglia, T. Vectorization of the General Monte Carlo Classical Trajectory Program VENUS. *J. Comput. Chem.* **1991**, *12*, 1014–1024.

(137) Stine, R. An Introduction to Bootstrap Methods: Examples and Ideas. *Sociological Methods & Research* **1989**, *18*, 243–291.

(138) Aubane, E. E. l; Wardlaw, D. M.; Zhu, L.; Hase, W. Role of angular momentum in statistical unimolecular rate theory. *Int. Rev. Phys. Chem.* **1991**, *10*, 249–286.

(139) Miller, J. A.; Klippenstein, S. J.; Raffy, C. Solution of Some One- and Two-Dimensional Master Equation Models for Thermal Dissociation: The Dissociation of Methane in the Low-Pressure Limit. *J. Phys. Chem. A* **2002**, *106*, 4904–4913.

(140) Johnston, D. G.; Birks, J. Activation Energies for the Dissociation of Diatomic Molecules Are Less than the Bond Dissociation Energies. *Acc. Chem. Res.* **1972**, *5*, 327–335.

(141) Truhlar, D. G. Interpretation of the Activation Energy. *J. Chem. Educ.* **1978**, *55*, 309–311.

(142) Keizer, J. The coupling of density and internal energy fluctuations by isothermal chemical reactions. *J. Comput. Phys.* **1984**, *80*, 4185–4192.

(143) Huppert, S.; Plé, T.; Bonella, S.; Depondt, P.; Finocchi, F. Simulation of Nuclear Quantum Effects in Condensed Matter Systems via Quantum Baths. *Applied Sciences* **2022**, *12*.

(144) Modrich, P. DNA mismatch correction. *Ann. Rev. Biochem.* **1987**, *56*, 435–466.

(145) Lowdin, P. Proton tunneling in DNA and its biological implications. *Rev. Mod. Phys.* **1963**, *35*, 742–732.

(146) Cerón-Carrasco, J. P.; Requena, A.; Michaux, C.; Perpéte, E. A.; Jacquemin, D. Effects of hydration on the proton transfer mechanism in the adenine-thymine base pair. *J. Phys. Chem. A.* **2009**, *113*, 7892–7898.

(147) Gheorghiu, A.; Coveney, P. V.; Arabi, A. A. The influence of base pair tautomerism on single point mutations in aqueous DNA. *Interface Focus* **2020**, *10*, 20190120.

(148) Cerón-Carrasco, J. P.; Requena, A.; Zuniga, J.; Michaux, C.; Perpéte, E. A.; Jacquemin, D. Intermolecular Proton Transfer in Microhydrated Guanine-Cytosine Base Pairs: a New Mechanism for Spontaneous mutation in DNA. *J. Phys. Chem. A.* **2009**, *113*, 10549–10556.

(149) Galtier, N; Piganeau, G; Mouchiroud, D; Duret, L GC-Content Evolution in Mammalian Genomes: The Biased Gene Conversion Hypothesis. *Genetics* **2001**, *159*, 907–911.

(150) Catalán, J.; del Valle, J.; Kasha, M. Resolution of concerted versus sequential mechanisms in photo-induced double-proton transfer reaction in 7-azaindole H-bonded dimer. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 8338–8343.

(151) Pohl, R.; Socha, O.; Slavicek, P.; Sala, M.; Hodgkinson, P.; Dracinsky, M. Proton transfer in guanine-cytosine base pair analogue studied by NMR spectroscopy and PIMD simulations. *Faraday Discuss.* **2018**, *212*, 331–344.

(152) Li, P.; Rangadurai, A.; Al-Hashimi, H.; Hammes-Schiffer, S. Environmental effects on guanine-thymine mispair tautomerization explored with quantum mechanical/-molecular mechanical free energy simulations. *J. Am. Chem. Soc.* **2020**, *142*, 11183–11191.

(153) Soler-Polo, D.; Mendieta-Moreno, J. I.; Trabada, D. G.; Mendieta, J.; Ortega, J. Proton Transfer in Guanine-Cytosine Base Pairs in B-DNA. *J. Chem. Theory Comput.* **2019**, *15*, 6984–6991.

(154) Pérez, A.; Tuckerman, M. E.; Hjalmarson, H. P.; von Lilienfeld, O. A. Enol Tautomers of Watson Crick Base Pair Models Are Metastable Because of Nuclear Quantum Effects. *J. Am. Chem. Soc.* **2010**, *132*, 11510–11515.

(155) Slocombe, L.; Al-Khalili, J.; Scacchi, M. Quantum and classical effects in DNA point mutations: Watson-Crick tautomerism in AT and GC base pairs. *Phys. Chem. Chem. Phys.* **2021**, *23*, 4141.

(156) Slocombe, L.; Al-Khalili, J.; Scacchi, M. An Open Quantum Systems approach to proton tunnelling in DNA. *Commun. Phys.* **2022**, *23*, 109.

(157) Narayana, N.; Ginell, S. L.; Russu, I. M.; Berman, H. M. Crystal and molecular structure of a DNA fragment: d(CGTGAATTCACG). *Biochem.* **1991**, *30*, 4449–4455.

(158) Slater, J. C.; Koster, G. Simplified LCAO method for the periodic potential problem. *Phys. Rev.* **1954**, *94*, 1498–1524.

(159) Foulkes, W. F.; Haydock, R. Thight-binding models and density-functional-theory. *Phys. Rev. B* **1989**, *39*, 12520–12536.

(160) Nieman, R.; Spezia, R.; Jayee, B.; Minton, T.; Hase, W. L.; Guo, H. Exploring Reactivity and Product Formation in N(4S) Collisions with Pristine and Defected Graphene with Direct Dynamics Simulations. *J. Comput. Phys.* **2020**, *153*, 184702.

(161) Cerón-Carrasco, J. P.; Jacquemin, D. DNA spontaneous mutation and its role in the evolution of GC-content: assessing the impact of the genetic sequence. *Phys. Chem. Chem. Phys.* **2015**, *17*, 7754–7760.

(162) Xiao, S; Wang, L.; Liu, Y.; Lin, X.; Liang, H. Theoretical investigation of the proton transfer mechanism in guanine-cytosine and adenine-thymine base pairs. *J. Comput. Phys.* **2012**, *137*, 195101.

(163) Fang, W.; Chen, J.; Rossi, M.; Feng, Y.; Li, X.; Michaelides, A. Inverse Temperature Dependence of Nuclear Quantum Effects in DNA Base Pairs. *J. Phys. Chem. Lett.* **2016**, *7*, 2125–2131.

(164) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. Design of Density Functionals by Combining the Method of Constraint Satisfaction with Parametrization for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions.

(165) Bondar, A.; Fischer, S.; Smith, J. C.; Elstner, M.; Suhai, S. Key Role of Electrostatic Interactions in Bacteriorhodopsin Proton Transfer. *J. Am. Chem. Soc.* **2004**, *126*, 14668–14677.

(166) Brovarets, O.; Hovorun, D. Can tautomerization of the A·T Watson–Crick base pair via double proton transfer provoke point mutations during DNA replication? A comprehensive QM and QTAIM analysis. *J. Biomol. Struct. Dyn.* **2014**, *32*, 127–154.

(167) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Comput. Phys.* **1993**, *98*, 5648–5652.

(168) Lee, C.; Yang, W.; Parr, R. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.

(169) Vosko, S.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* **1980**, *58*, 1200–1211.

(170) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

(171) Yanai, T.; Tew, D.; Handy, N. A new hybrid exchange-correlation functional using the Coulomb-Attenuating Method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51–57.

(172) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comp. Chem.* **2011**, *32*, 1456–1465.

(173) Henderson, T. M.; Izmaylov, A. F.; Scalmani, G.; Scuseria, G. E. Can short-range hybrids describe long-range-dependent properties? *J. Comput. Phys.* **2009**, *131*, 044108.

(174) Chai, J.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.

(175) Sponer, J.; Jurecka, P.; Hobza, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs,and amino acid pairs. *J. Am. Chem. Soc.* **2004**, *126*, 10142–10151.

(176) Frisch, M. J. et al. Gaussian 16 Revision C.01. **2016**, Gaussian Inc. Wallingford CT.

(177) Hele, T. J.; Suleimanov, Y. V. Should thermostatted ring polymer molecular dynamics be used to calculate thermal reaction rates? *J. Comput. Phys.* **2015**, *143*, 074107.

(178) Benedict, L.; Charles, M. Robust and efficient configurational molecular sampling via Langevin dynamics. *J. Comput. Phys.* **2013**, *138*, 174102.

(179) Suleimanov, Y. V.; Collepardo-Guevara, R.; Manolopoulos, D. E. Bimolecular reaction rates from ring polymer molecular dynamics: Application to H + CH4→ H2 + CH3. *J. Comput. Phys.* **2011**, *134*, 044131.

(180) Tolosa, S.; Sanson, J.; Hidalgo, A. Mechanisms for guanine-cytosine tautomeric equilibrium in solution via steered molecular dynamics simulations. *J. Mol. Liq.* **2018**, *251*, 308–316.

(181) Peterson, L. A.; Hecht, S. S. O6-Methylguanine is a critical determinant of 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone tumorigenesis in A/J mouse lung. *Cancer Res.* **1991**, *51*, 5557–5564.

(182) Hecht, S. S. DNA adduct formation from tobacco-specific N-nitrosamines. *Mut. Res.-Fund. Mol. M.* **1999**, *424*, 127–142.

(183) B.; Bernardi, G. CpG islands, genes and isochores in the genomes of vertebrates. *Gene* **1991**, *106*, 185–195.

(184) Saxonov, S.; Berg, P.; Brutlag, D. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 1412–1417.

(185) Deaton, A. M.; Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **2011**, *25*, 1010–1022.

(186) Hund, F. Zur Deutung der Molekelspektren. I. *Zeitschfrift für Physik* **1927**, *40*, 742–764.

(187) Bell, R. P.; Hinshelwood, C. N. The application of quantum mechanics to chemical kinetics. *Proc. R. Soc. A: Math. Phys. Eng. Sci.* **1933**, *139*, 466–474.

(188) Wigner, E. Über das Überschreiten von Potentialschwellen bei chemischen Reaktionen. *Zeitschrift für Physikalische Chemie* **1932**, *19B*, 203 –216.

(189) Urey, H. C.; Brickwedde, F. G.; Murphy, G. M. A Hydrogen Isotope of Mass 2. *Phys. Rev.* **1932**, *39*, 164–165.

(190) Bell, R. P.; Le Roy, R. J. The Tunnel Effect in Chemistry. *Physics Today* **1982**, *35*, 85–86.

(191) Amnon, K. Kinetic Isotope Effects as Probes for Hydrogen Tunneling, Coupled Motion and Dynamics Contributions to Enzyme Catalysis. *Prog. React. Kin.* **2003**, *28*, 119–156.

(192) Nagel, Z. D.; Klinman, J. P. Tunneling and Dynamics in Enzymatic Hydride Transfer. *Chem. Rev.* **2006**, *106*, 3095–3118.

(193) Sheridan, R. S. In *Reviews of Reactive Intermediate Chemistry*; John Wiley & Sons, Ltd: 2007; Chapter 10, pp 415–463.

(194) Truhlar, D. G. Tunneling in enzymatic and nonenzymatic hydrogen transfer reactions. *J. Phys. Org. Chem.* **2010**, *23*, 660–676.

(195) Castro, C.; Karney, W. L. Heavy-Atom Tunneling in Organic Reactions. *Angew. Chem. Int. Ed.* **2020**, *59*, 8355–8366.

(196) Schleif, T.; Tatchen, J.; Rowen, J. F.; Beyer, F.; Sanchez-Garcia, E.; Sander, W. Heavy-Atom Tunneling in Semibullvalenes: How Driving Force, Substituents, and Environment Influence the Tunneling Rates. *Chem. Eur. J.* **2020**, *26*, 10452–10458.

(197) Carpenter, B. K. Heavy-atom tunneling as the dominant pathway in a solution-phase reaction? Bond shift in antiaromatic annulenes. *J. Am. Chem. Soc.* **1983**, *105*, 1700–1701.

(198) Gonzalez-James, O. M.; Zhang, X.; Datta, A.; Hrovat, D. A.; Borden, W. T.; Singleton, D. A. Experimental Evidence for Heavy-Atom Tunneling in the Ring-Opening of Cyclopropylcarbinyl Radical from Intramolecular 12C/13C Kinetic Isotope Effects. *J. Am. Chem. Soc.* **2010**, *132*, 12548–12549.

(199) Bell, R., *The Proton in Chemistry*; Springer New York, NY: 1973.

(200) Schleif, T.; Mieres-Perez, J.; Henkel, S.; Ertelt, M.; Borden, W. T.; Sander, W. The Cope Rearrangement of 1,5-Dimethylsemibullvalene-2(4)-d1: Experimental Evidence for Heavy-Atom Tunneling. *Angew. Chem. Int. Ed.* **2017**, *56*, 10746–10749.

(201) Zhang, X.; Hrovat, D. A.; Borden, W. T. Calculations Predict That Carbon Tunneling Allows the Degenerate Cope Rearrangement of Semibullvalene to Occur Rapidly at Cryogenic Temperatures. *Org. Lett.* **2010**, *12*, 2798–2801.

(202) Cheng, A. K.; Anet, F. A. L.; Mioduski, J.; Meinwald, J. Determination of the fluxional barrier in semibullvalene by proton and carbon-13 nuclear magnetic resonance spectroscopy. *J. Am. Chem. Soc.* **1974**, *96*, 2887–2891.

(203) Suleimanov, Y. V. Surface Diffusion of Hydrogen on Ni(100) from Ring Polymer Molecular Dynamics. *J. Phys. Chem. C* **2012**, *116*, 11141–11153.

(204) Richardson, J. O.; Althorpe, S. C. Ring-polymer molecular dynamics rate-theory in the deep-tunneling regime: Connection with semiclassical instanton theory. *J. Chem. Phys.* **2009**, *131*.

(205) Loco, D.; Chataigner, I.; Piquemal, J.; Spezia, R. Efficient and Accurate Description of Diels-Alder Reactions Using Density Functional Theory**. *Chem. Phys. Chem.* **2022**, *23*.

(206) Vuong, V. Q.; Akkarapattiakal Kuriappan, J.; Kubillus, M.; Kranz, J. J.; Mast, T.; Niehaus, T. A.; Irle, S.; Elstner, M. Parametrization and Benchmark of Long-Range Corrected DFTB2 for Organic Molecules. *J. Chem. Theory Comput.* **2018**, *14*, 115–125.

(207) Gaus, M.; Goez, A.; Elstner, M. Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.* **2013**, *9*, 338–354.

(208) Gaus, M.; Lu, X.; Elstner, M.; Cui, Q. Parameterization of DFTB3/3OB for Sulfur and Phosphorus for Chemical and Biological Applications. *J. Chem. Theory Comput.* **2014**, *10*, 1518–1537.

(209) Kubillus, M.; Kubar, T.; Gaus, M.; Rezac, J.; Elstner, M. Parameterization of the DFTB3 Method for Br, Ca, Cl, F, I, K, and Na in Organic and Biological Systems. *J. Chem. Theory Comput.* **2015**, *11*, 332–342.

(210) M, B.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R.; Parrinello, M. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comp. Phys. Commun.* **2009**, *180*, 1961–1972.

(211) Gillan, M. J. Quantum-classical crossover of the transition rate in the damped double well. *J. Phys. C: Solid State Phys.* **1987**, *20*, 3621.

(212) Álvarez Barcia, S.; Flores, J. R.; Kästner, J. Tunneling Above the Crossover Temperature. *J. Phys. Chem. A* **2014**, *118*, 78–82.

Part IV

Appendix

Appendix A

Unimolecular Fragmentation

In this section some additional computational details and results of the study of chapter 3:

- The number of beads used for each temperature and each potentials, Table A.1

  The rate constants in ps^{-1} computed for each temperature and each potentials, Tables A.2, A.3, A.4, A.5

- The Arrhenius plots obtained for different friction values of the Langevin thermostat for the 1D Morse potential, Figure A.1

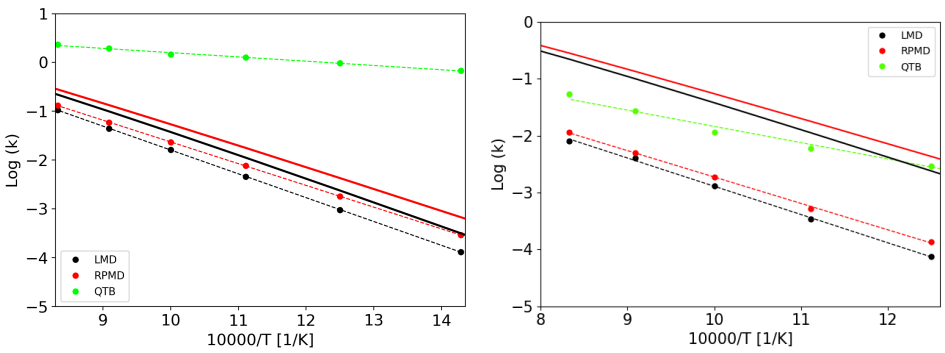

Figure A.1: Quantum and classical rate constants as a function of temperature for 1-D Morse. In full lines we report values obtained from sum-of-state approach (both classical and quantum), while results from simulations are reported as dots: LMD (black), RPMD (red) and QTB (green). On the left the results for γ = 0.045 fs^{-1}, on the right γ = 0.3 fs^{-1}

Table A.1: Simulation set-up: temperatures (in K) and number of beads (P, for RPMD simulations) used for the 1-D Morse and the three CH_4-model potentials.

	T (K)	P
1D Morse	800	32
	900	16
	1000	8
	1100	8
	1200	8
Potential A	3000	8
	3500	8
	4000	8
	4500	8
	5000	8
Potential B	1350	16
	1500	8
	1700	8
	2000	4
	2500	4
Potential C	800	16
	1000	16
	1200	8
	1500	8

Table A.2: Rate constants in ps^{-1} for the 1D-Morse potential as a function of temperature (in K).

Method	700	800	900	1000	1100	1200
Classical SoS	0.03	0.07	0.14	0.24	0.37	0.52
Quantum SoS	0.04	0.17	0.28	0.42	0.57	0.09
LMD (γ = 0.01)	0.0207	0.0456	0.0847	0.128	0.199	0.275
RPMD (γ = 0.01)	0.0293	0.0633	0.109	0.158	0.234	0.311
QTB (γ = 0.01)	0.756	0.819	0.849	0.924	0.965	1.17
LMD (γ = 0.045)	0.0207	0.0488	0.0968	0.168	0.258	0.379
RPMD (γ = 0.045)	0.0292	0.0644	0.121	0.1958	0.293	0.415
QTB (γ = 0.045)	0.845	0.989	1.109	1.18	1.336	1.441
LMD (γ = 0.3)	0.00669	0.0161	0.0312	0.0559	0.0914	0.123
RPMD (γ = 0.3)	-	0.0209	0.0376	0.06538	0.100	0.143
QTB (γ = 0.3)	0.049	0.0796	0.108	0.143	0.2099	0.280

Table A.3: Rate constants in ps^{-1} for the potential A as a function of temperature (in K)

Temperature	LMD	RPMD	QTB
3000 K	0.00118	0.00129	0.00359
3500 K	0.0107	0.0120	0.0215
4000 K	0.0549	0.0581	0.0792
4500 K	0.175	0.178	0.233
5000 K	0.402	0.420	0.471

Table A.4: Rate constants in ps^{-1} for the potential B as a function of temperature.

Temperature	LMD	RPMD	QTB
1350 K	0.000289	0.000495	0.0157
1500 K	0.00155	0.00255	-
1700 K	0.00859	0.0127	0.0721
2000 K	0.0600	0.0696	0.171
2500 K	0.397	0.472	0.748

Table A.5: Rate constants in ps^{-1} for the potential C at different temperatures (in K)

Temperature	LMD	RPMD	QTB
800 K	0.000125	0.000441	0.112
1000 K	0.00348	0.0086	0.212
1200 K	0.0336	0.0571	0.370
1500 K	0.271	0.351	0.986
1800 K	0.875	1.197	1.780
2000 K	1.595	1.901	-

Appendix B

Environmental and Nuclear Quantum Effects on Double Proton Transfer in Guanine-Cytosine Base Pair

In this section we report the detailed values of the free energy with the associated uncertainties, obtained for the double proton transfer in base pairs. These data correspond to the graphs in Figure 4.8 in Chapter 4, where the free energies have been computed on a finer grid (36x36) using the Wham algorithm.

d_2 / d_1	-0.75	-0.45	-0.15	0.15	0.45	0.75
-0.75	0.0	2.2±0.2	7.3±0.3	15.1±0.3	22.3.7±0.3	25.3±0.4
-0.45	4.2±0.1	5.5±0.1	9.3±0.2	15.1±0.4	20.2±0.3	23.2±0.3
-0.15	12.26±0.05	10.93±0.03	12.91±0.06	15.9±0.2	19.4±0.3	21.8±0.1
0.15	20.4±0.1	18.2±0.1	17.1±0.2	17.96±0.08	19.5±0.1	19.49±0.06
0.45	26.3±0.1	23.28±0.08	20.5±0.1	18.6±0.1	18.5±0.2	16.9±0.1
0.75	29.3±0.97	26.4±0.1	23.2±0.2	20.2±0.2	17.7±0.2	15.0±0.1

Table B.1: Umbrella Sampling results for the **LMD** of **isolated dimer**: free energies and associated uncertainties (in kcal/mol) as a function of d_1 and d_2 coordinates (in Å). The colours correspond to the canonical (green) and tautomeric (blue) structures.

d_2 / d_1	-0.75	-0.45	-0.15	0.15	0.45	0.75
-0.75	0.0	1.9±0.1	6.2±0.2	13.7±0.1	20.5±0.2	24.2±0.2
-0.45	3.5±0.3	4.42±0.04	7.48±0.07	18.0±0.1	20.2±0.3	21.63±0.09
-0.15	11.0±0.2	9.2±0.1	10.6±0.1	17.0±0.2	19.4±0.3	19.9±0.2
0.15	18.62	16.1±0.2	14.4±0.1	15.1±0.2	16.6±0.1	18.1±0.3
0.45	23.5	20.4±0.2	17.9±0.1	16.3±0.1	15.33±0.08	15.8±0.3
0.75	-	21.1	20.92±0.09	18.4±0.2	16.2±0.3	14.1±0.4

Table B.2: Umbrella Sampling results for the **PIMD** of **isolated dimer**: free energies and associated uncertainties (in kcal/mol) as a function of d_1 and d_2 coordinates (in Å). The colours correspond to the canonical (green) and tautomeric (blue) structures.

d_2 / d_1	-0.75	-0.45	-0.15	0.15	0.45	0.75
-0.75	0.0	2.87±0.07	10.3±0.2	18.7±0.3	26.0±0.3	29.2±0.2
-0.45	3.21±0.05	5.7±0.1	12.1 ±0.2	19.7±0.2	24.8±0.2	27.9±0.2
-0.15	9.9±0.2	11.2±0.2	14.9±0.3	21.0±0.4	25.8±0.2	28.4±0.2
0.15	13.6±0.4	14.4±0.2	17.4±0.3	20.6±0.2	23.8±0.3	26.1±0.2
0.45	14.2±0.6	14.4±0.1	16.7±0.2	19.3±0.2	21.3±0.2	22.2±0.3
0.75	14±1	13.8±0.3	15.9±0.3	18.1±0.4	19.1±0.2	19.1±0.3

Table B.3: Umbrella Sampling results for the **LMD** of **3BP-DNA** model: free energies and associated uncertainties (in kcal/mol) as a function of d_1 and d_2 coordinates (in Å). The colours correspond to the canonical (green), tautomeric (blue), and intermediate (yellow) structures.

d_2/d_1	-0.75	-0.45	-0.15	0.15	0.45	0.75
-0.75	0.0	2.9±0.1	10.12±0.08	17.9±0.2	24.8±0.5	28.4±0.4
-0.45	2.88±0.04	5.3±0.2	10.1 ±0.3	17.1±0.2	22.7±0.5	26.1±0.6
-0.15	8.2±0.1	8.9±0.3	12.5±0.3	17.9±0.1	22.5±0.2	25.6±0.3
0.15	12.3±0.2	11.8±0.4	14.6±0.4	18.1±0.2	21.2±0.1	23.8±0.2
0.45	14.7±0.5	13.7±0.7	15.7±0.6	17.9±0.5	19.7±0.2	21.2±0.1
0.75	17.1±0.7	15.9±0.8	17.8±0.9	19.6±0.8	19.0±0.3	19.1±0.2

Table B.4: Umbrella Sampling results for the **PIMD** of **3BP-DNA** model: free energies and associated uncertainties (in kcal/mol) as a function of d_1 and d_2 coordinates (in Å). The colours correspond to the canonical (green), tautomeric (blue), and intermediate (yellow) structures.

Appendix C

Code and Algorithms

Here we report the structure of a software developed during this thesis, in order to perform molecular dynamics simulation with the inclusion of NQEs. This software was used for the results shown in Chapter 4 and 5.

The language used is Fortran 90. A previous installation of the PLUMED [25] software is mandatory.

The aim of this code is to perform Thermostatted Ring Polymer Molecular Dynamics simulations, and it gives access to the basic observables from a set of trajectories. The forces can be computed by an analytical potentials (morse, harmonic oscillator) or through an external software, presently DFTB+ [24]. On the other hand, PLUMED software, called as a library, allows to couple Path Integral based simulations with enhanced sampling methods.

In the Figure C.1 we show the flowchart of the code. In order to perform a set of simulations, it is necessary to have the following files in the directory:

1. Input file: input.nml
2. Geometry: geometry.xyz
3. Input for the external softwares (DFTB+ and/or PLUMED)

Once the simulations are performed the output files in the directory will be:

1. md_out
2. traj.xyz

3. traj_beads.xyz

Each of those files will be describe in detail in the next sections. After the software is initialized, and the initial conditions are chosen, the information concerning the system and the parameters used in the simulation will be printed out. Once the dynamics started, the most computationally expensive loop in the algorithm, namely the computation of the DFTB forces for the different beads, can be run in parallel using OpenMP. In this way the computational cost of the simulation is lower, for example in the case of 4 beads the cost is lower by more than two times then the version not parallelized, as it can be seen in Table C.1.

Method	WallTime [Hours:Minutes:Seconds]	Ratio [T-RPMD/Classic]
Classical	00:00:41	-
T-RPMD 4 Beads	00:02:44	4
T-RPMD 4 Beads, OPENMP	00:01:05	1.5

Table C.1: Here we report the wall clock time of simulations performed with the Classical, T-RPMD and T-RPMD parallelized methods

Algorithm

The algorithm implemented for the integration of the T-RPMD equations of motion is the BAOAB [178], as reported by Ceriotti and coworkers [75]. For a system with N atomic degrees of freedom (labeled by the index i) and P replicas (labeled by j):

$$p_i^j \leftarrow p_i^j - \Delta t \frac{\partial V(q_1^j..q_N^j)}{\partial q_i^j} \tag{C.1}$$

$$\widetilde{p}_i^k = \sum_{j=1}^{P} p_i^j C_{j,k} \quad ; \quad \widetilde{q}_i^k = \sum_{j=1}^{P} q_i^j C_{j,k} \tag{C.2}$$

$$\begin{bmatrix} \widetilde{p}_i^k \\ \widetilde{q}_i^k \end{bmatrix} \leftarrow \begin{bmatrix} \cos\omega_k \frac{\Delta t}{2} & -m_i\omega_k \sin\omega_k \frac{\Delta t}{2} \\ [\frac{1}{m_i\omega_k}] \sin\omega_k \frac{\Delta t}{2} & \cos\omega_k \frac{\Delta t}{2} \end{bmatrix} \begin{bmatrix} \widetilde{p}_i^k \\ \widetilde{q}_i^k \end{bmatrix} \tag{C.3}$$

$$\widetilde{p}_i^k \leftarrow c_1^k \widetilde{p}_i^k + \sqrt{\frac{m_i}{\beta_P}} c_2^k \eta_i^k \tag{C.4}$$

$$, \begin{bmatrix} \widetilde{p}_i^k \\ \widetilde{q}_i^k \end{bmatrix} \leftarrow \begin{bmatrix} \cos\omega_k \frac{\Delta t}{2} & -m_i\omega_k \sin\omega_k \frac{\Delta t}{2} \\ [\frac{1}{m_i\omega_k}] \sin\omega_k \frac{\Delta t}{2} & \cos\omega_k \frac{\Delta t}{2} \end{bmatrix} \begin{bmatrix} \widetilde{p}_i^k \\ \widetilde{q}_i^k \end{bmatrix} \tag{C.5}$$

$$p_i^j = \sum_{k=0}^{NM} C_{j,k} \widetilde{p}_i^k \quad ; \quad q_i^j = \sum_{k=0}^{NM} C_{j,k} \widetilde{q}_i^k \tag{C.6}$$

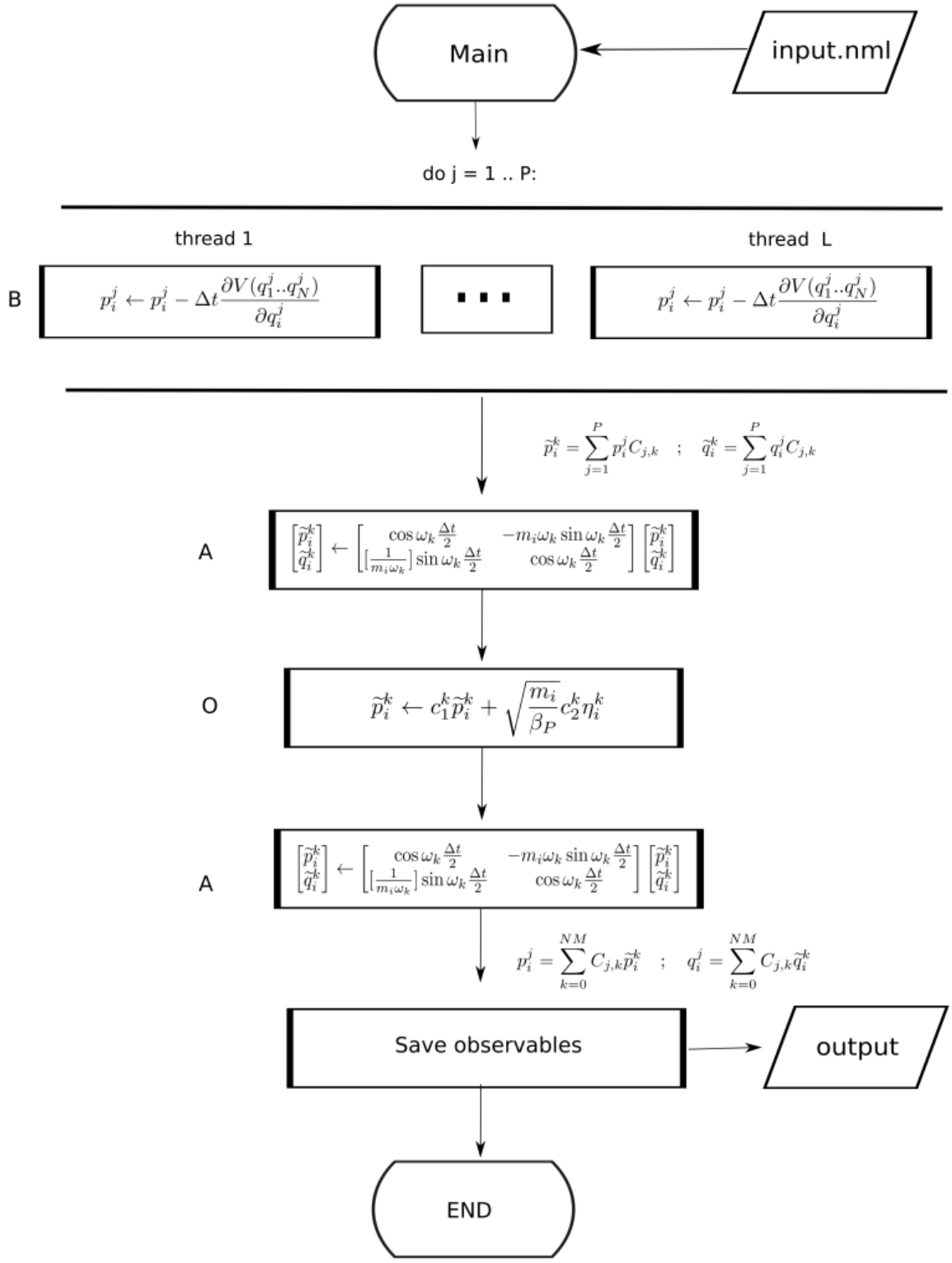

Figure C.1: Flowchart. Step B: propagate momenta; Step A: propagate springs in normal mode; Step O: Langevin thermostat.

Where k runs over the normal mode index, η_i^k are normalized gaussian random numbers and:

$$c_1^k = e^{-\Delta t \gamma^k} \tag{C.7}$$

$$c_2^k = \sqrt{1 - [c_1^k]^2} \tag{C.8}$$

The Equation C.1 is the B step of the BAOAB algorithm, and involve the propagation of the momenta in the cartesian coordinates. This is the most expensive passage in terms of computational cost, since the forces must be calculated for each replicas of the system. The schematic loop can be represented in this way:

```
do ibeads=1, Pbeads

        do m = 1, 3N
                q(m) = qbeads(m,ibeads)
        enddo

        if (potential.eq."DFTB") then
                        call write_gen(omp_flag, ibeads, array_index_label, q, 3N)
                        call dftb+(omp_flag,ibeads,path,output,pot,fxyz,3N)
        elif (potential.eq."MORSE") then
                        call analytical_morse(pot,q,fxyz,3N)
        elif (potential.eq."HO") then
                        call analytical_ho(pot,q,fxyz,3N)
        endif

        pot_array(ibeads) = pot

        do m=1, 3N
                store_forces(m,ibeads) = fxyz(m)
        enddo
enddo
```

A convenient way to propagate the springs of the Ring Polymer, is by doing it in normal modes coordinates. This can be achieved thanks to the matrix $C_{j,k}$ in Equation C.2, whose elements are defined as:

$$C_{j,k} = \begin{cases} \sqrt{1/P}, & \text{if } k = 0 \\ \sqrt{2/P}\cos(2\pi jk/P), & \text{if } 1 \leq k \leq P/2 - 1 \\ \sqrt{1/P}(-1)^j, & \text{if } k = P/2 \\ \sqrt{2/P}\sin(2\pi jk/P), & \text{if } P/2 + 1 \leq k \leq n - 1 \end{cases} \tag{C.9}$$

Once positions and momenta are in the normal mode representation, the spring can be propagated (Equation C.3). The last fundamental step of the algorithm is the step O, represented by Equation C.4, where the Langevin thermostat is applied. While the friction over the centroid is chosen arbitrarily, the friction coefficient on the normal modes of the P replicas are equal to their frequency according to the optimal friction criterion defined in Ref. [75].

Input File

The input file is as follows:

```
&calculation_parameters
        calculation_type='DFTB'/'MORSE'/'HO'
        path_calculation='path'
        plumed_flag=.FALSE./.TRUE.
        restart_flag=.FALSE./.TRUE.
        omp_flag=.FALSE./.TRUE.
/

&general_parameters
        save_steps= integer     # step to wait for saving
        temperature= integer    # [K]
        iseed= integer              # seed for the randomic number generator
        num_traj= integer           # number of trajetories to perform
        n_replicas= integer         # number of replicas or beads
/

&equilibration_parameters
        equilibration_flag=.FALSE./.TRUE.
        dt_eq= float                # [fs]
        step_eq= integer            # number of steps for the equilibration
        gamma_eq= float             # [1/fs]
/

&dynamic_parameters
        dt_dyn= float                   # [fs]
        step_dyn= integer               # number of steps for the dynamics
        flag_NVE=.FALSE./.TRUE.
        gamma_dyn= float                # [1/fs]
/
```

```
&writing_options:
        print_replicas=.FALSE./.TRUE. # print or not the position of the beads
/
```

Here we describe each keywords:

calculation_parameters:

1. *calculation_type* : the forces can be either calculated with the DFTB+ software or analytical potentials (Morse or harmonic oscillator)

2. *path_calculation* : in the case of DFTB+, the path where is located the executable of the software must be given in the input file

3. *plumed_flag* : in case of TRUE, the software PLUMED will be initialized, and the working directory must contain the input file for plumed, called "plumed.dat"

4. *restart_flag* : in case of TRUE, the working directory must contain a file called "pos_vel" which contains the positions and momenta to restart the trajectory. The cartesian coordinates must be specified in Angstorm, while for momenta the Units of measure must be (kg/mol)Å/fs^2, with the format "X Y Z P_x P_y P_z". In case of FALSE, the velocity will be generated by Maxwell Boltzmann distribution, and the cartesian coordinate will be the ones given by the user by the xyz file.

5. *omp_flag* : TRUE or FALSE, it enables the option for running in parallel

general_parameters:
here are described the general keywords to run the simulation, such as the number of steps to wait until saving, the temperature, a seed for the random number generator, the number of trajetories to perform and the number of replicas or beads.

equilibrium_parameters and **dynamic_parameters**:
the same keywords are used for both equilibrium and dynamical simulations. If the "equilibration_flag" is set to TRUE, two separate outputs will be generated: "traj_eq.xyz" and "md_out_eq". These outputs represent the trajectory of length "step_eq" with a time-step of "dt_eq" and a friction value of "gamma_eq" for the thermostat. Once this trajectory is finished, the final cartesian coordinates and momenta will be used to start a second trajectory, the one on which we do the analysis. The "dynamic_parameters" are the ones associated to the trajectory on which we will do the analysis. The option "flag_NVE", if TRUE, disables the step O in the BAOAB algorithm, and the trajectory is performed in the NVE ensemble.

writing_options:
this will generate an additional output "traj_beads.xyz" which will contain positions and momenta of all the beads.

Output files

The output file will consist in the "md_out" file, "traj.xyz", "traj_beads.xyz", "pos_vel". The format of the trajectory is xyz, and it contains also the momenta. In the case of "traj_beads.xyz" the coordinate and momenta are written for all the beads. The "pos_vel" file contains the last coordinates and momenta, that are useful to start a new trajectory with them as initial conditions. The "md_out" file contains the most basic observables from the dynamics:

1. Kinetic Energy
2. Potential Energy
3. Total Hamiltonian (see Equation 2.17)
4. Virial Energy

All the values are in kcal/mol. The Total Hamiltonian within the RPMD formalism, is useful to check the conservation of the total energy in the case of performing a trajectory without a thermostat. On the other hand, the virial energy, is a quantity which describes the total quantum energy of the system within the ring polymer formalism. For one atom, and P beads:

$$E_{virial} = \frac{Pk_BT}{2} - \frac{1}{2P}\sum_{i=1}^{P} m\omega_P^2(q_i - q_{i+1})^2 + \frac{1}{P}\sum_{i=1}^{P} V(q_i) \tag{C.10}$$